5.01 Overview and Introduction

John C. Vederas, University of Alberta, Edmonton, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

In the first edition of Comprehensive Natural Products Chemistry, Volume 4 edited by Jeffrey Kelly was entitled 'Amino Acids, Peptides, Porphyrins, and Alkaloids'. The present volume on 'Amino Acids, Peptides, and Proteins' has a different scope and focus. The intent in the current volume is to provide up-to-date reviews of salient areas dealing with biosynthesis of amino acids and peptides, as well as some proteins/enzymes that metabolize them. Clearly a volume of this size with only 20 chapters cannot be fully comprehensive in its coverage of what is a major segment of the field of biological chemistry. Inevitably there are major gaps, and topics of great importance and research activity will be missing. Partly this stems from limitations of space and partly it is caused by unavoidable circumstances that prevented the delivery of several desirable chapters at late stages of assembly of this volume. The main goal has been to focus on selected areas and provide an entry into a current understanding of the topic at the postgraduate level for the nonspecialist. The first half of this volume describes amino acids and peptides, whereas the second half focuses on proteins and their modifications.

The volume begins with an overview by Sabesan Yoganathan and John Vederas (Chapter 5.02) of uncommon L-amino acids that occur in nature, sometimes as parts of nonribosomally synthesized peptides or alkaloids, but which are generally not found in normal proteins. The intent is to update excellent prior reviews with newer examples. The following chapter (5.03) by Paul Engel and Francesca Paradisi describes biotransformations and the resolution of amino acids, both naturally occurring and unnaturally occurring. Applications to chiral synthesis, kinetic resolution, interconversion of enantiomers, and production of oxoacids are discussed along with biological screening, protein engineering, and whole-cell biocatalysis. A review then follows (Chapter 5.04) by Małgorzata Jakubowicz and Witold Nowak of 1-aminocyclopropane-1-carboxylate (ACC), its synthase (ACS), and related enzymes involved in the biosynthesis of the essential plant hormone, ethylene. Recent results from molecular genetic studies on components of the ethylene signal transduction pathway are discussed. The following article (Chapter 5.05) by William Self discusses the incorporation of selenium, which is toxic in large quantities but essential as a micronutrient, into amino acids and macromolecules, especially selenoproteins. The functions of these, as well as the modes of selenium uptake and utilization, are detailed. A chapter by Holger Barth and Bradley Stiles (Chapter 5.06) then discusses bacterially produced toxins and their mode of action on mammalian cells. This obviously has very important implications for food safety, bacterial pathogenicity, and sepsis. A review on bridging antimicrobial and immunomodulatory activities of host-defense peptides by Jason Kindrachuck, Anastasia Nijnik, and Robert Hancock (Chapter 5.07) describes a first line of defense against infection. Over 1000 such peptides have been identified, and in addition to providing fascinating insights into biological activity, they are potential targets for the development of new antimicrobials active against resistant bacteria. The lantibiotics, which are posttranslationally modified bacterial peptide antibiotics, are described in a chapter (5.08) by Lisa Cooper, Bo Li, and Wilfred van der Donk. In the last decade there has been an explosion in the understanding of the biosynthesis and mode of action of these potent antimicrobial substances that have important uses in food preservation and great potential for human therapeutic applications. Jan-Christoph Westermann and David Craik then detail peptides that act as defense substances from nonmarine plants (Chapter 5.09). Many of these have important applications and potential in biotechnology and pharmaceutical development. The latter theme is also explored for peptidic animal venoms in the review (Chapter 5.10) by Nicolas Andreotti, Besma Jouirou, Stephanie Mouhat, Ludovic Mouhat, and Jean-Marc Sabatier. Their ability to create ion channels, catalytically cleave biomolecules, or interact with receptors gives these compounds access to a large variety of therapeutic targets. A subsequent review (Chapter 5.11) by Ingolf Nes, Ola Johnsborg, and Dzung Bao Diep on signal transduction in Gram-positive bacteria shifts the focus to communication mediated by peptidic compounds. Bacterial communication through chemical signaling

controls a host of microbial properties, including virulence, gene transfer, and production of antimicrobial substances, including other peptides.

In the second part of this volume, the focus is more directly on proteins. Rémy Ricoux and Jean-Pierre Mahy review new advances in the generation of catalytic antibodies or 'abzymes' as artificial enzymes (Chapter 5.12). These biocatalysts based on monoclonal antibodies are able to catalyze a wide range of chemical reactions, although it is still problematic to achieve the high rates displayed by enzymes. In a subsequent chapter, Ute Kothe provides a summary and new insights into our current understanding of normal ribosomal protein synthesis (Chapter 5.13). This is one of the most fundamental and conserved processes in all cells, and recent crystal structures have provided valuable insights into the functioning of this molecular machine consisting of both RNA and proteins. Jonathan Huot, Jacques Lapointe, Robert Chenevert, Marc Bailly, and Daniel Kern next discuss glutaminyl-tRNA and asparaginyl-tRNA biosynthetic pathways and function (Chapter 5.14). These unusual biosynthetic routes are both fascinating and fundamentally important. Once proteins have been synthesized, many are posttranslationally modified, and the review by Keith Green and Sylvie Garneau-Tsodikova (Chapter 5.15) provides a current overview of these processes. These modifications expand the possible amino acid units in proteins and are essential for cell growth, transcriptional regulation, and metabolism. The next article (Chapter 5.16) by Hans Peter Bachinger, Kazunori Mizuno, Janice Vranka, and Sergei Boudko provides extensive coverage of a particular group of modified proteins, the 28 types of collagens. Collagens are the major components of the extracellular matrix of multicellular animals and facilitate the formation, regulation, and maintenance of tissue and organ structures. Another posttranslational modification that plays an essential role in cellular processes and can be a therapeutic target is protein lipidation, which is reviewed by Kristina Görmer, Luc Brunsveld, and Herbert Waldmann (Chapter 5.17). The attachment of lipids to proteins leads to changes in structures and physical properties of the proteins, thereby affecting their biological activity, stability, and cellular localization. Expansion of the genetic code through the incorporation of unusual amino acids into proteins is described by Angela Parrish and Lei Wang (Chapter 5.18). Over 40 unnatural amino acids with novel chemical or physical properties have been genetically encoded in *Escherichia coli*, yeast, and mammalian cells using this approach. Timothy Montavon and Steven Bruner next provide a current review (Chapter 5.19) of the mechanisms and enzymes of nonribosomal peptide biosynthesis. The large multimodular enzymes that catalyze the formation of such peptides use a thiol-templated mechanism and make a host of biologically and medicinally important compounds. In the following review (Chapter 5.20), Petrus Milne and Gareth Kilian delineate the formation and biological activity of 2,5-diketopiperazines. Members of this family of cyclic dipeptides exhibit a host of effects including immunomodulatory, modulation of glucose, hormone regulatory, antifungal, antibacterial, antitumor, and cardiac activity. The volume closes with a review (Chapter 5.21) by Ashok Hegde on ubiquitin-dependent protein degradation. This pathway plays a vital role in many physiological and pathological processes, including cancer and Alzheimer's disease.

This editor would like to express his deepest gratitude to all of the authors who have provided the incisive and comprehensive reviews of these fundamental topics. Each chapter has required extensive time and effort, and has been written to enhance the knowledge and understanding of the scientific community. This work by the authors is greatly appreciated by the editor, who hopes that the readers will also enjoy it.

Biographical Sketch

John C. Vederas is Professor of Chemistry and holds a Canada Research Chair in Bioorganic and Medicinal Chemistry. He obtained a BSc degree from Stanford University and a PhD degree with the late George Bü chi from the Massachusetts Institute of Technology. His postdoctoral work at the University of Basel (with Christoph Tamm) and at Purdue University (with Heinz Floss) inspired a continuing interest in application of organic chemistry to understanding of biological mechanisms, especially in polyketide and peptide biosynthesis. He joined the University of Alberta as an assistant professor in 1977. He has received recognition for research and teaching from the University of Alberta, including the Rutherford Award for Excellence in Undergraduate Teaching (1995), the University Cup for Research and Teaching (1998), the J. Gordin Kaplan Award for Excellence in Research (2003), the Killam Award for Excellence in Mentoring (2003), and the Klawe Prize in Teaching Large Classes (2006). He is an Alberta Centennial Medal recipient (2006), a Fellow of the Royal Society of Canada (1997) and a Fellow of The Royal Society (London, 2009). He has received the Merck Sharp Dohme Award (1986), the John Labatt Award (1991), the R. U. Lemieux Award (2002), and the Alfred Bader Award (2005) from the Canadian Society for Chemistry for his research. In 2008, he was awarded the Chemical Institute of Canada (CIC) Medal. He also served in numerous scientific organizations, was President of the Canadian Society for Chemistry (2002–03), a member of Council of the Natural Sciences and Engineering Research Council of Canada (NSERC) (2001–04), and Chair of the 2008 Annual Conference of the Canadian Society for Chemistry. He is the author of over 235 research publications and 13 patents.

5.02 Nonprotein L-Amino Acids

Sabesan Yoganathan and John C. Vederas, University of Alberta, Edmonton, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

5.02.1 Introduction

Proteins and peptides are vital biomolecules for regulating biochemical processes in living organisms, including enzymes that catalyze many biochemical reactions. There are also a number of peptides known to exhibit significant importance as hormones, pheromones, and defense substances. The basic building blocks of peptides and proteins are amino acids, which are linked through amide bonds. Peptide synthesis in living organisms can be classified as either ribosomal peptide synthesis or nonribosomal peptide synthesis (NRPS).

Ribosomal synthesis of peptides proceeds through translation of messenger ribonucleic acid (mRNA) and utilizes the 20 primary L-α-amino acids. These amino acids are incorporated with the use of specific transfer ribonucleic acid (tRNA) codons.[1] The 20 primary α-amino acids, with the exception of glycine that is achiral, are characterized by an L-configuration at the α-position (**Figure 1**). In general, most proteins are found to be composed of these 20 L-α-amino acids, as such they are referred to as protein amino acids.

NRPS involves a family of polyfunctional enzymes where both L-α-amino acids and unusual amino acids are incorporated into peptides.[2,3] The unusual amino acids identified are generally not found to be part of proteins and therefore these amino acids are referred to as nonprotein amino acids. The biosynthetic machinery of NRPS is unique, as it produces complex structures in which the amino acids are linked through side chains to give branched and cyclic peptides (**Figure 2**). Plants are rich sources of nonprotein amino acids and many of these amino acids have been isolated and characterized.[4] Many nonprotein amino acids are also found in other terrestrial organisms. Two well-known examples of these types of amino acids are 3,4-dihydroxy-L-phenylalanine (L-DOPA) and L-thyroxine, which are both produced in humans. A number of nonribosomal peptides containing unusual amino acids have also been isolated from microorganisms, such as bacteria and fungi. Several marine invertebrates have been identified as sources of nonribosomal peptides; however, these peptides may originate from symbiotic microorganisms.[5–7]

For the purpose of this chapter, all amino acids containing side chains that differ from the L-α-amino acids will be designated as either unusual or nonprotein amino acids. This is due to the observation that these unusual amino acids are generally not found to be the constituents of proteins. The scope of this chapter is limited to nonprotein L-α-amino acids. The definition of nonprotein amino acid is not absolute and, therefore, the

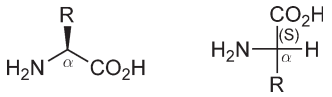


Figure 1 Stereochemical drawing and Fischer projection of an L-α-amino acid, where R is the side chain of the amino acid.

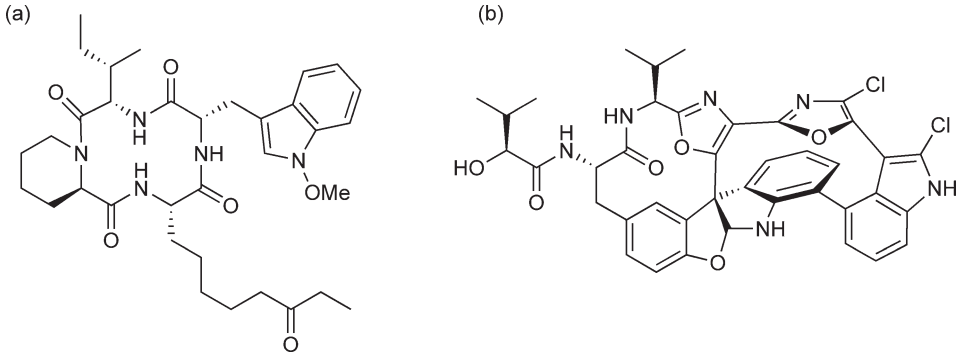


Figure 2 Nonribosomal peptides (a) apicidin[8] and (b) diazonamide A.[9–11]

unusual amino acids derived from posttranslational modifications of L-α-amino acids have been excluded. More detail about posttranslational modifications of peptides and proteins and the formation of unusual amino acids can be found in Chapter 5.15. Barrett[12] has listed a number of nonprotein amino acids, including both L- and D-amino acids. The focus of this chapter is to present an updated collection of nonprotein L-α-amino acids and this book includes some structurally unique nonprotein L-α-amino acids that may also appear in Barrett's book. The amino acids reported here are identified as either single amino acids or components of nonribosomal peptides.[2–5,13,14] The reader is encouraged to consult the book by Barrett[12] for other nonprotein amino acids that are not listed in this chapter.

There is also a significant occurrence of nonprotein D-amino acids in nature.[14] These amino acids have attractive structural variations and remarkable biological properties. However, this chapter will not include details on nonprotein D-amino acids. Selenium-containing amino acids, such as selenocysteine and selenomethionine, are also considered as unusual amino acids. Recently, much interest in selenium-containing amino acids has emerged. More information about selenium-containing amino acids is available in Chapter 5.05.

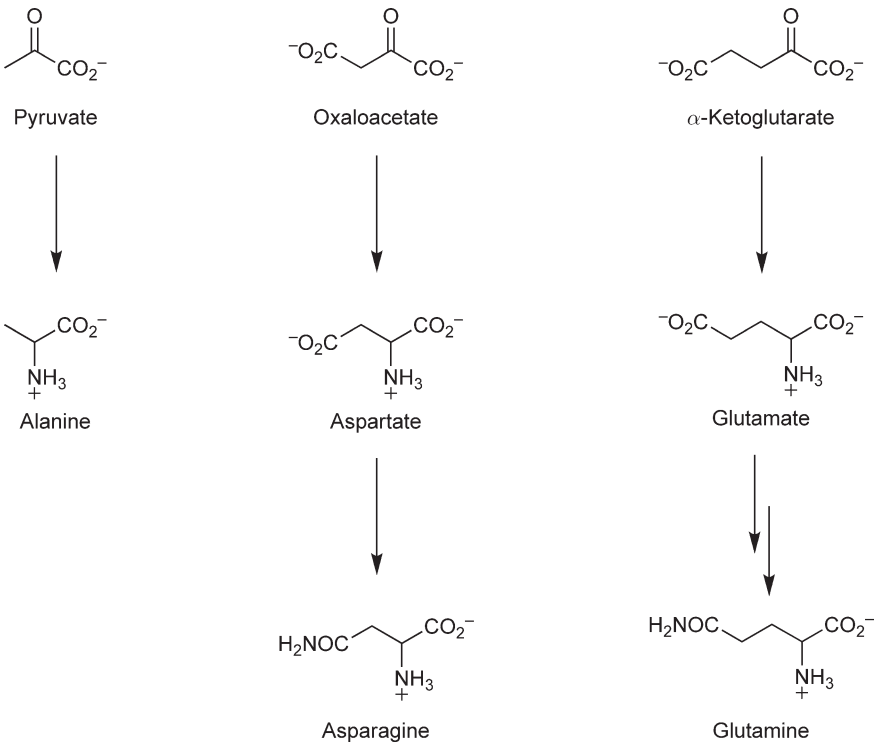
According to the report by Bell *et al.*,[4] there are about 800 nonprotein amino acids known to date in the literature. It is not practical to cite all the primary literature of the reported nonprotein amino acids. To the best of our knowledge, we have tried to cite the most relevant references and the recent review papers available. In most cases, further references for the primary literature can be obtained from these reviews. We apologize to the reader for any missed compounds and citations that might be relevant to the material.

5.02.2 Biosynthesis of Amino Acids

The biosynthetic origin of nonprotein amino acids is remarkable as it gives rise to a diversity of structures. By looking at the structures of isolated nonprotein amino acids, one can generally see the resemblance of an L-α-amino acid. This suggests that in most cases, nature makes use of the available biosynthetic machinery that produces L-α-amino acids for the synthesis of nonprotein amino acids. Therefore, before describing the biosynthetic origin of nonprotein amino acids, it is necessary to highlight the biosynthetic pathways of L-α-amino acids that lead to the production of nonprotein amino acids.

5.02.2.1 Biosynthesis of Amino Acids with Aliphatic Side Chains

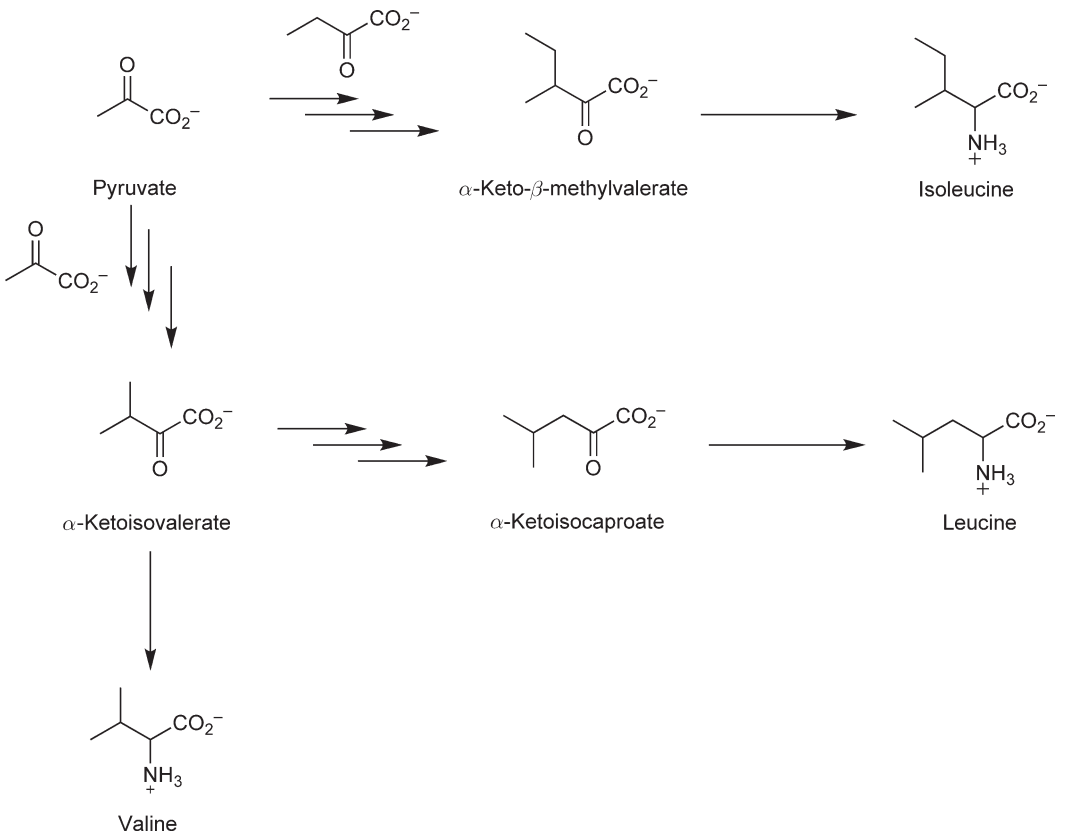
The biosynthetic pathways for many L-α-amino acids have been elucidated using the biosynthetic machineries present in plants and microorganisms.[15,16] Generally, keto acids, including pyruvate, oxaloacetate, α-ketoglutarate, and α-ketobutyrate, are identified as the precursors for the biosynthesis of amino acids with aliphatic side chains. In many cases, an enzyme-catalyzed transamination reaction takes place to convert the α-carbonyl group into an α-amine functionality. It has been reported that glutamine serves as the amine donor during many of these transamination reactions.[15] Note that a number of other biosynthetic routes are available in various organisms that produce L-α-amino acids, including α-aminoadipate,[17] diaminopimelate (DAP),[18] and tRNA-dependent transamidation[19] pathways; however, these pathways are not pertinent to this chapter. The general biosynthetic pathways for most L-α-amino acids containing aliphatic side chains are given below (**Schemes 1(a)–1(d)**).[15,20] Note that the given scheme does not include all intermediate structures of the biosynthetic pathways and more detail is available from cited references.



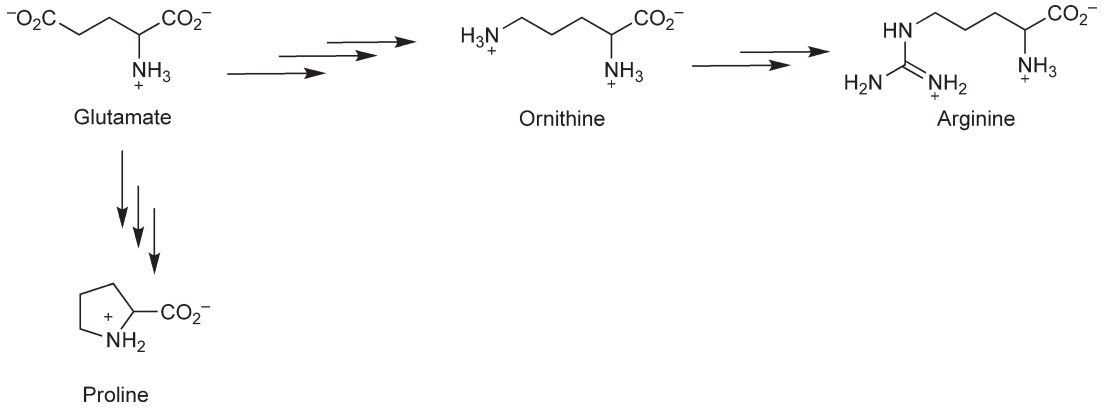
Scheme 1a Biosynthesis of L-α-amino acids – Ala, Asp, Glu, Asn, and Gln.

5.02.2.2 Biosynthesis of Amino Acids with Aryl Side Chains

Aryl side chain containing L-α-amino acids, such as phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp), are derived through the shikimate pathway.[21,22] The enzymatic transformation of phosphoenolpyruvate (PEP) and erythro-4-phosphate, through a series of reactions, yields shikimate (**Scheme 2**). Although shikimate is an important biosynthetic intermediate for a number of secondary metabolites,[22] this chapter only describes the conversion of shikimate to amino acids containing aryl side chains. In the second part of the biosynthesis, shikimate is converted into chorismate by the addition of PEP to the hydroxyl group at the C5 position. Chorismate is then transformed into prephenate by the enzyme chorismate mutase (**Scheme 3**).

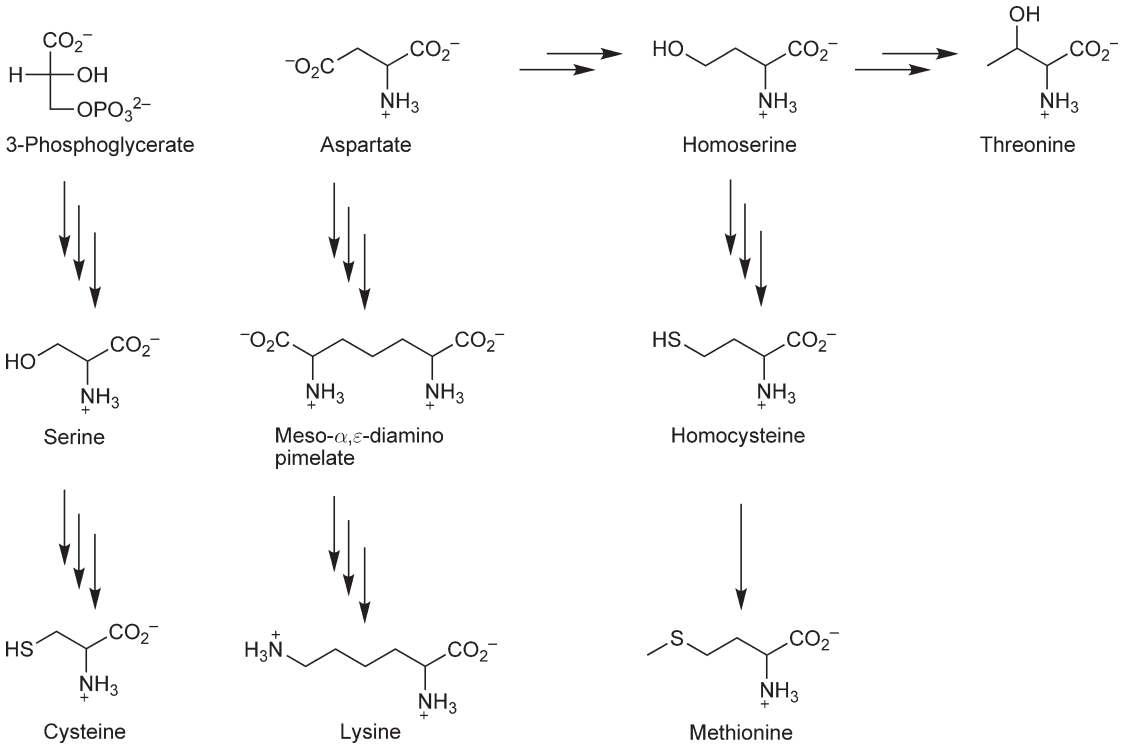


Scheme 1b Biosynthesis of L-α-amino acids – Val, Ile, and Leu.

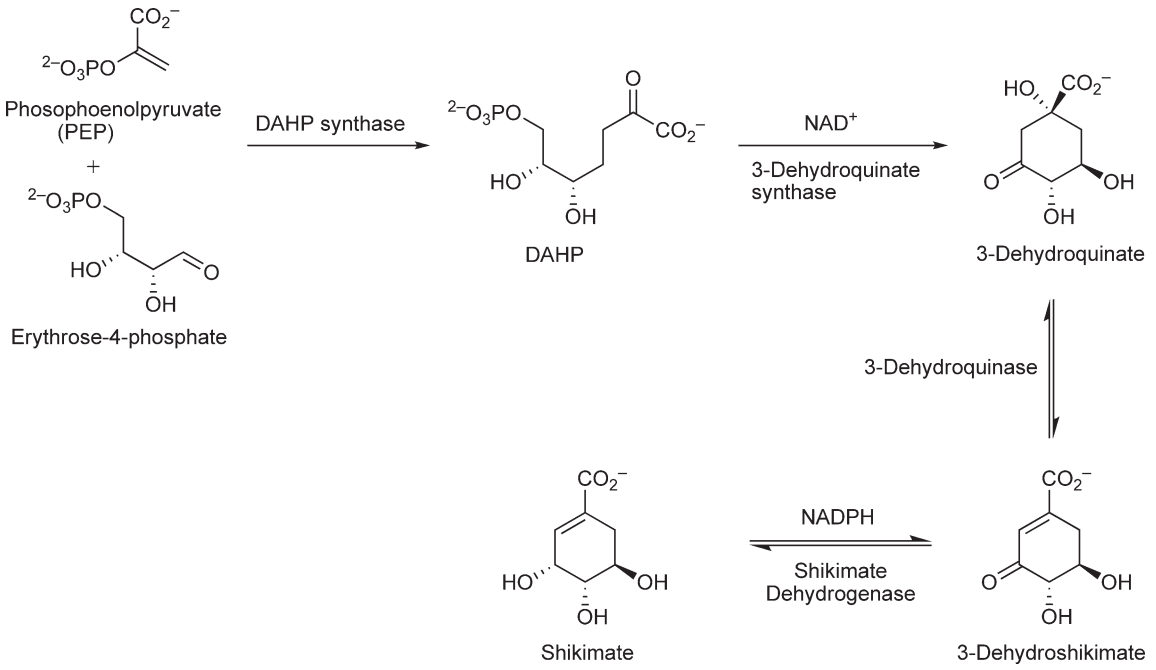


Scheme 1c Biosynthesis of L-α-amino acids – Pro and Arg.

During the biosynthetic transformation, chorismate is the point of divergence for the biosynthesis of Phe, Tyr, Trp, and other amino acids containing aromatic groups. For example, the biosynthesis of Trp begins with the conversion of chorismate to anthranilate (**Scheme 4(a)**). A sequence of amination and aromatization reactions produces anthranilate, which is then condensed with phosphoribosylpyrophosphate. The intermediate is carried through a series of reactions to yield Trp (**Scheme 4(b)**).



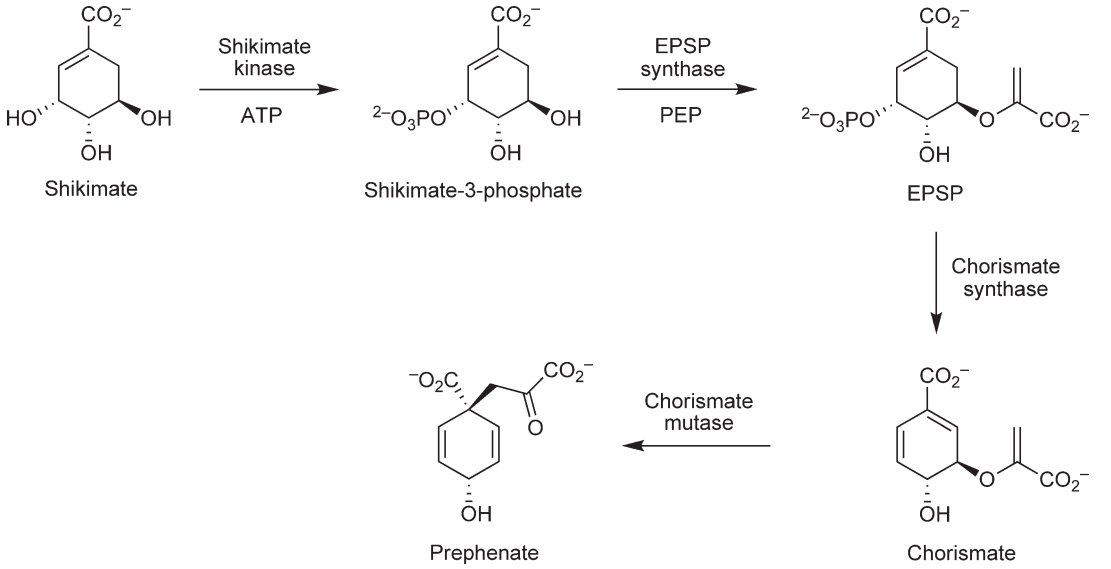
Scheme 1d Biosynthesis of L-α-amino acids – Cys, Lys, Met, and Thr.



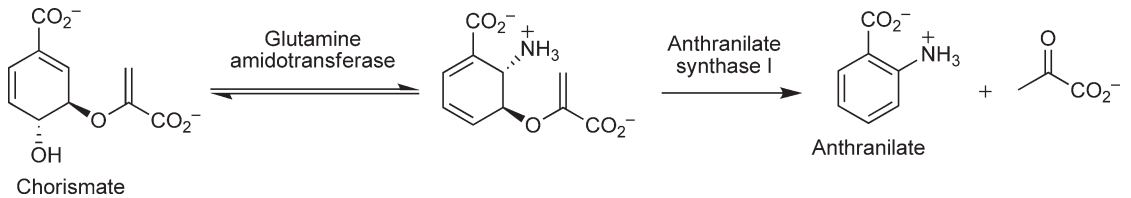
Scheme 2 Biosynthesis of shikimate.

5.02.2.3 Biosynthesis of Nonprotein Amino Acids

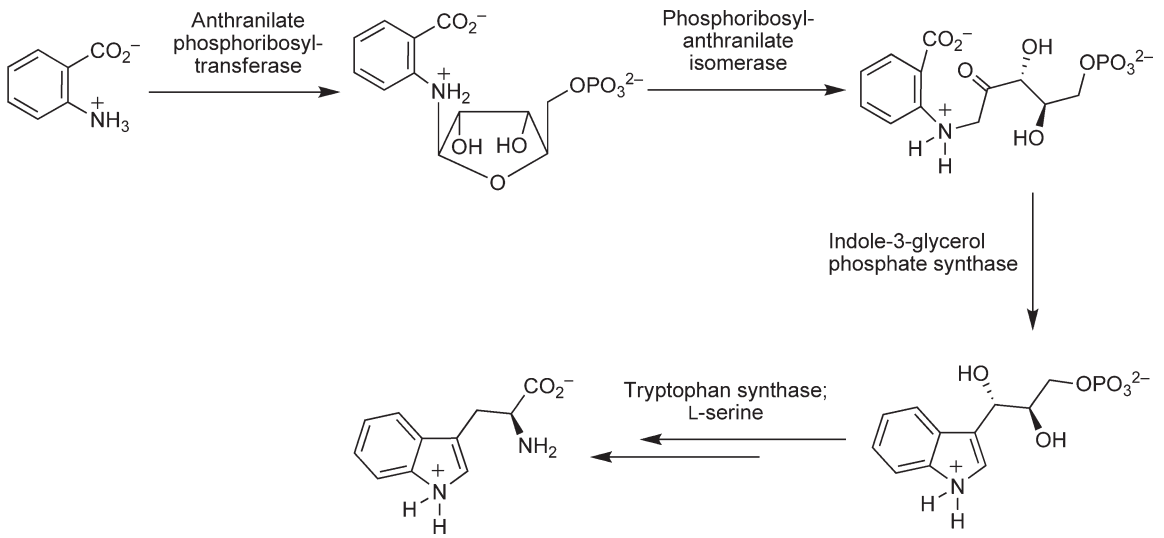
When the biosynthetic pathways given above are examined, it is apparent that several intermediates are indeed nonprotein α-amino acids. Ornithine, homoserine, homocysteine, and α-ε-diaminopimelic acid are a few examples. This shows that some nonprotein amino acids originate as intermediates during the biosynthesis of



Scheme 3 Biosynthesis of chorismate and prephenate.



Scheme 4a Biosynthesis of anthranilate.



Scheme 4b Biosynthesis of tryptophan.

known L-α-amino acids. An organism may purposefully produce these types of intermediates, or a failed biosynthetic pathway toward an L-α-amino acid may cause the accumulation of nonprotein amino acids. In many cases, it has also been observed that metabolism or degradation of L-α-amino acids produces structurally interesting nonprotein amino acids. These nonprotein amino acids are often the precursors in the biosyntheses of biologically important molecules. For example, the nonprotein amino acids *N*′-formylkynurenine and kynurenine are synthesized through the metabolic pathway of Trp and are the precursors for the biosynthesis of niacin (vitamin B_3).[23]

In addition to this, it has been reported that nonprotein amino acids could be formed by structural modifications to protein amino acids (methylation, hydroxylation, and halogenation) through modified L-α-amino acid biosynthetic pathways and through novel biosynthetic routes. Some examples of the nonprotein amino acids derived through these biosynthetic pathways are given below (**Figure 3**).[24] A detailed discussion of known biosyntheses for certain nonprotein amino acids will be discussed later in this chapter.

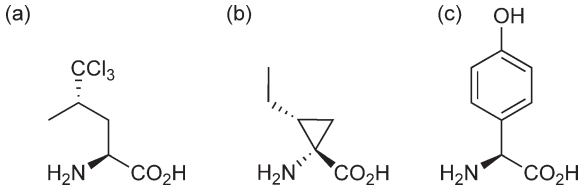
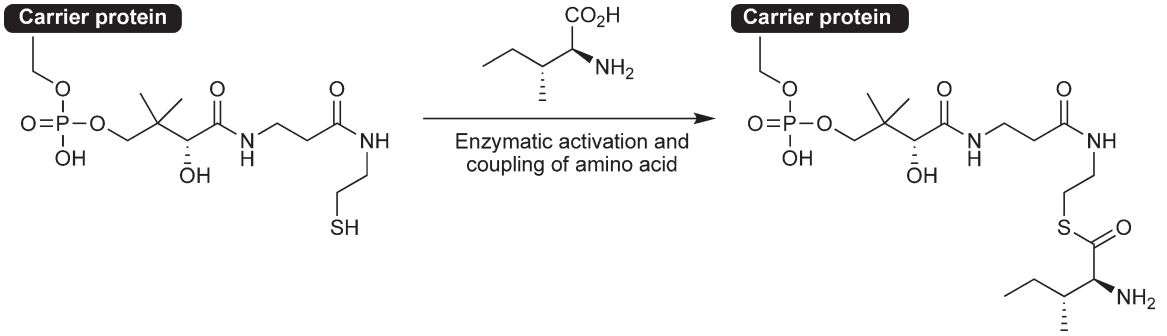


Figure 3 Nonprotein amino acids (a) 5,5,5-trichloroleucine, (b) coronamic acid, and (c) 4-OH-phenylglycine.

Earlier in this chapter, it was mentioned that many of the nonprotein amino acids are components of nonribosomal peptides. During such a biosynthesis, the peptide is attached to a carrier protein through a thioester bond, until chain termination occurs and the final product is released. The carrier protein is posttranslationally modified by the attachment of a phosphopantetheinyl group from coenzyme A. This step gives rise to the active carrier protein with a phosphopantetheine arm upon which amino acids are added to during NRPS. As an example, loading of isoleucine onto the carrier protein is depicted below (**Scheme 5**). Further details about nonribosomal peptide syntheses and enzymatic reactions can be found in Chapter 5.19.



Scheme 5 Loading of isoleucine onto the phosphopantetheine arm of a carrier protein domain.

During the biosynthesis of nonribosomal peptides, there are two ways to incorporate the nonprotein amino acids. They can be incorporated either as a single unit or as an L-α-amino acid, which then undergoes structural modifications, while attached to the carrier protein. In the case of coronamic acid, L-*allo*-isoleucine is loaded onto the carrier protein and a unique biosynthetic pathway produces a cyclopropyl group containing a nonprotein amino acid.[25] Specific examples of the biosynthesis of nonprotein amino acids will be discussed in the following sections.

5.02.3 Classification of Nonprotein Amino Acids

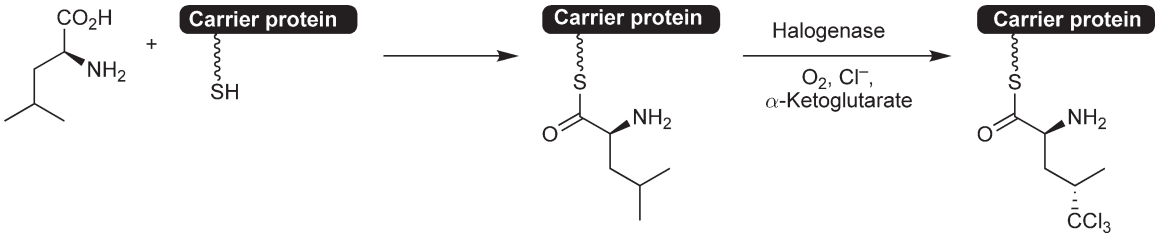
There are a vast number of nonprotein L-α-amino acids reported in the literature and each one presents unique and interesting structural features. In this chapter, we have adopted a simple classification system to categorize these amino acids so that compounds with similar side chain functionalities can be grouped together. This classification system is not absolute and the purpose is to organize these compounds into tables to visualize the structural similarities. All of the entries are classified into one of the following functional group categories: (1) alkyl, (2) alkenyl and alkynyl, (3) aryl, (4) amino, (5) hydroxyl, (6) carbonyl, (7) heterocyclic, and (8) other functional groups. When an amino acid with more than one functional group is identified, it is placed in the best-suited category. There are a number of nonprotein amino acids lacking a stereogenic center at the α-carbon and as such these structures may not be labeled as L-amino acids. However, it is possible to trace back an α-amino acid biosynthetic precursor for these amino acids. Since these amino acids exhibit interesting structural features and biological properties, they are included in this review.

As mentioned earlier, this chapter is a compilation of known nonprotein L-α-amino acids. In many cases, nonprotein amino acids and nonribosomal peptides exhibit remarkable biological activities, including cytotoxicity, antimicrobial, antitumor, and other inhibitory activities. The structure of the amino acid, its biological origin, and its known biological properties are listed in a tabular form throughout this chapter. If the nonprotein L-α-amino acids are constituents of nonribosomal peptides and are linked through amide bond, then these amino acids are listed as single units with free α-amino and α-carboxyl groups. It is important to note that in such cases, the observed bioactivity is from the peptide itself and not from the listed amino acid.

5.02.3.1 Amino Acids with Alkyl or Haloalkyl Side Chains

This section focuses on the nonprotein amino acids containing alkyl or haloalkyl groups as side chains. There are number of examples containing linear and branched alkyl chains, such as ethyl, propyl, and *t*-butyl. Some amino acids contain the cyclopropyl group as side chains (coronamic acid derivatives and BZR-cotoxin II). Furthermore, dysamides, a class of diketopiperazines, contain nonprotein amino acids with chlorinated alkyl side chains.[26]

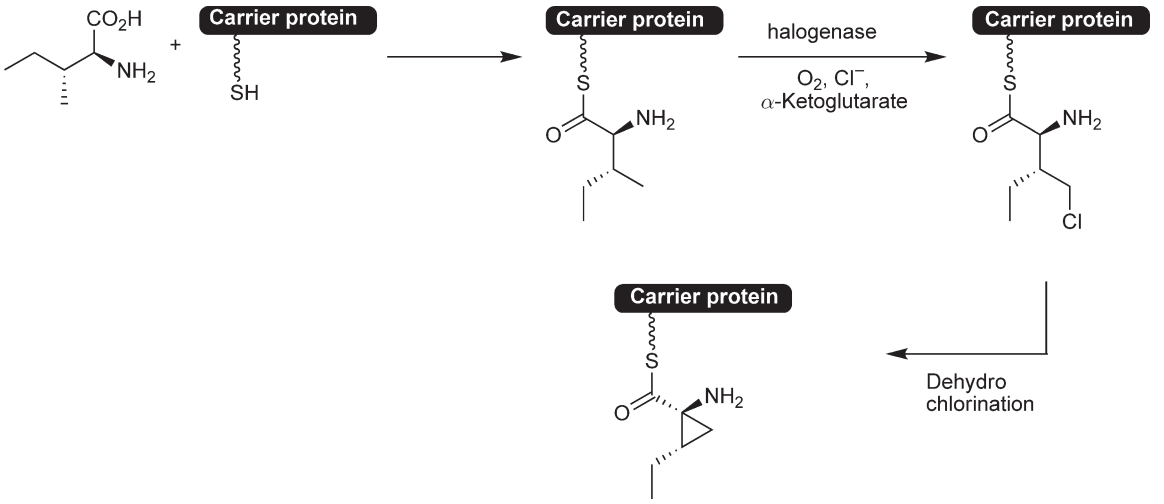
The biosynthesis of barbamide, a potent molluscicidal natural product, was reported by Walsh and coworkers.[27] Their findings support the enzymatic trichlorination of L-leucine by two genes, *barB1* and *barB2*, to yield L-5,5,5-trichloroleucine (**Scheme 6**).



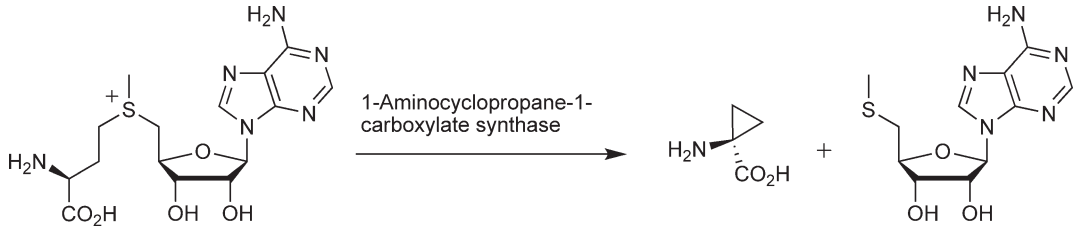
Scheme 6 Biosynthesis of trichloroleucine attached to a carrier protein.

With the use of gene clusters of the natural products coronatine and kutznerides, the biosynthetic pathway of coronamic acid has also been elucidated by Walsh and coworkers.[25,28] From the biosynthetic analyses, a nonheme Fe^{II}-dependent halogenase[29] was identified as the chlorinating enzyme that converts L-*allo*-isoleucine to γ-chloroisoleucine. A second enzyme carries out a dehydrochlorination reaction to yield coronamic acid. The general biosynthetic pathway is shown below (**Scheme 7**).

The nonprotein amino acid, 1-aminocyclopropane-1-carboxylic acid, is an intermediate of ethylene biosynthesis in plants. This amino acid is synthesized from the L-α-amino acid methionine through the intermediate *S*-adenosyl-L-methionine (SAM) (**Scheme 8**).[13]



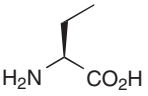
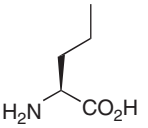
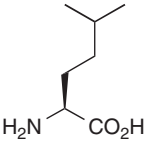
Scheme 7 The biosynthetic pathway for coronamic acid attached to the carrier protein.[25]



Scheme 8 Biosynthesis of 1-aminocyclopropane-1-carboxylate.

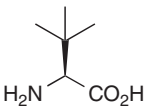
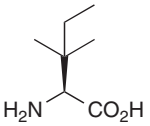
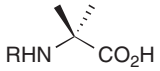
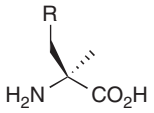
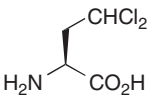
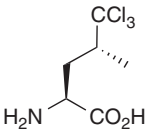
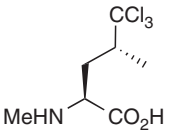
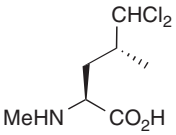
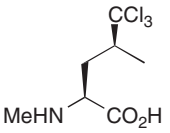
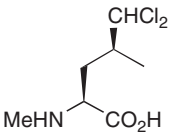
Many of the listed nonprotein amino acids have been identified as components of nonribosomal peptides from fungi and sponges. Hence, they tend to exhibit antifungal and antimicrobial properties, and in some cases potent cytotoxicity is observed. Specific details on known nonprotein amino acids are given in **Table 1**.

Table 1 Nonprotein amino acids containing alkyl, cycloalkyl, or haloalkyl side chains

Structure	*Biological origin*	*Biological properties*	*References*
	Nonribosomal peptides from marine sponge – *Theonella* sp.	Miraziridine – Cathepsin B inhibitor Nazumamide A – thrombin inhibitor	14, 30
	Fungi – *Trichoderma* sp.	Cyclosporines D and E – cytotoxic peptides	14, 31
	Marine sponge – *Theonella swinhoei*	Cyclolithistide A – antifungal peptide	14, 32
	Sponge – *Theonella* sp.	Keramamides (B, E, M, N), orbiculamide A, discobahamins A, B, and oriamide (all are cytotoxic peptides)	14, 33, 34 35 36 37
	Peptaibols from fungi – *Trichoderma* sp.	Atroviridin A – antimicrobial and antifungal peptide	38

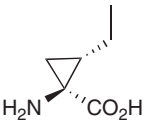
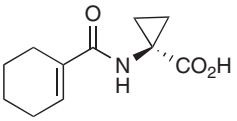
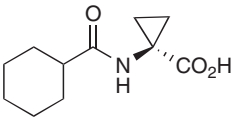
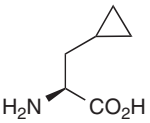
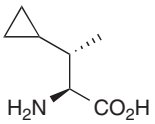
(*Continued*)

Table 1 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Marine sponge – *Discodermia kiiensis* and related species	Discodermins – antimicrobial and antifungal peptides	14, 39, 40
	Sponge – *Halichondria cylindrata*	Halicylindramides – antifungal peptides	14, 41
	Sponge – *Hemiasterella minor, Cymbastela* sp.	Hemiasterlin A, criamides A, B – cytotoxic peptides	14, 42
	Sponge – *Halichondria cylindrata*	Polydiscamide A – cytotoxic peptide	14, 40, 43
	Sponge – *Sidonops microspinosa*	Microspinosamide – peptide with HIV-1 inhibitory activity	14, 44
 R = H, Ac	Peptaibols from fungi – *Trichoderma* sp.	Trichorzin HA1, atroviridins – antimicrobial and antifungal peptides	38, 45
 R = H, Me	Fungi – *Myrothecium* sp.	MS-681a, b, c, and d – peptide-based inhibitors of mysin light chain kinase	46, 47
	Fungi – *Geotrichum candidum*	Neoefrapeptins – insecticidal peptides	48
	Actinobacteria – *Streptomyces armentosus*	Armentomycin – amino acid with antibiotic activity	49, 50
	Sponge – *Dysidea herbacea*	Dysidenins – reversible inhibitors of Na^+–I^- cotransporter	14, 51–54
	Sponge – *Dysidea herbacea*	Dysidenins – reversible inhibitors of Na^+–I^- cotransporter	14, 52, 54
	Sponge – *Dysidea herbacea*	Dysidenins – reversible inhibitors of Na^+–I^- cotransporter	14, 52, 54
	Marine sponge – *Dysidea* sp. Cyanobacteria – *Oscillatoria spongeliae*	Dysamides – chlorinated natural products	14, 26, 55
	Marine sponge – *Dysidea* sp. Cyanobacteria – *Oscillatoria spongeliae*	Dysamides – chlorinated natural products	14, 26, 55

(*Continued*)

Table 1 (Continued)

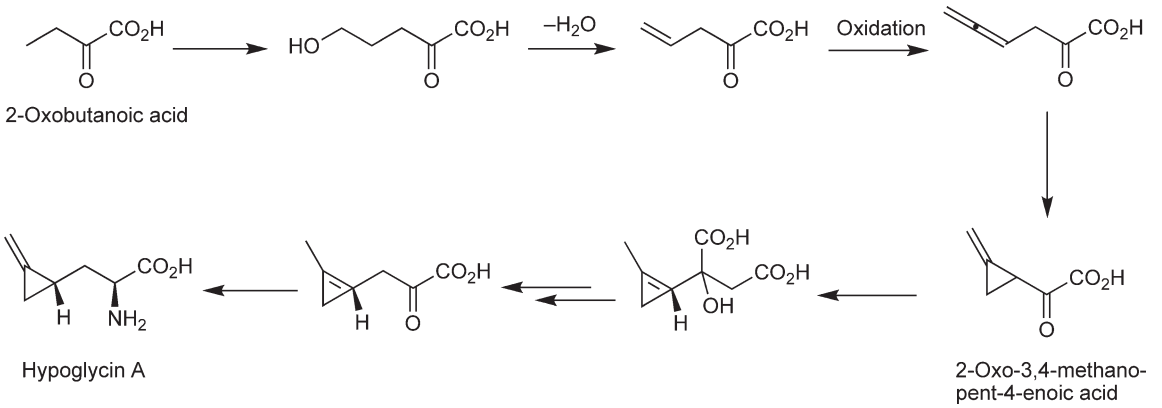
Structure	*Biological origin*	*Biological properties*	*References*
	Peptide from fungi – *Bipolaris zeicola* race 3	BZR-cotoxin II – phytotoxic peptide	56, 57
	Fungi – *Geotrichum candidum*	Neoefrapeptins – insecticidal peptides	48
	Occurs in many plants	Intermediate in ethylene biosynthesis	13
	Phytopathogenic bacteria – *Pseudomonas syringae* pv. *glycinea*	Norcoronamic acid – phytotoxin	56, 58
	Phytopathogenic bacteria – *Pseudomonas syringae* pv. *glycinea*	Coronamic acid – phytotoxin	56, 59, 60
	Nonribosomal peptide isolated from *Streptomyces* sp.	Cytotrienin A – induces apoptosis in human leukemia HL-60 cells	56, 61, 62
	Nonribosomal peptide isolated from *Streptomyces* sp.	Cytotrienin B – induces apoptosis in human leukemia HL-60 cells	56, 62
		Compound UCF116 – farnesyltransferase inhibitor	63
	Mushroom – *Amanita virgineoides* Bas.	Nonprotein amino acid with antifungal property	64, 65
	Isolated from mushroom *Amanita castanopsidis* and *Amanita cokeri*	Nonprotein amino acid with phytotoxic, antifungal, and antimicrobial activities	66–68

5.02.3.2 Amino Acids with Alkenyl or Alkynyl Side Chains

A number of nonprotein amino acids with unsaturated side chains have been isolated. Many of these contain alkene side chains, but some alkyne side chains containing amino acids have also been identified. Nonprotein dehydroamino acids do not have an α-stereocenter; these amino acids are still classified under this category. Dehydroamino acids are generally biosynthesized by the enzymatic elimination of a leaving group at the β-carbon. For example, serine and threonine are enzymatically dehydrated to give dehydroalanine and dehydrobutyrine, respectively.[69] A similar biosynthetic pathway is hypothesized for dehydroamino acids found in nonribosomal peptides, such as nodularins and microcystins.[70]

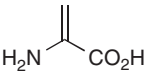
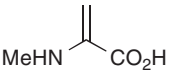
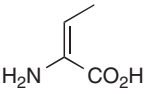
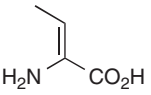
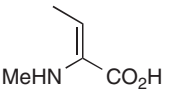
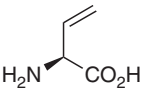
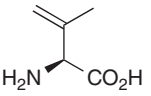
Nonprotein amino acids with unsaturated side chains are generally isolated as components of nonribosomal peptides from marine bacteria and sponges. These peptides often show cytotoxic and antifungal properties. Dysamides, a class of halogenated diketopiperazines from marine sponges, also contain unsaturated nonprotein amino acids. No known bioactivities have been reported for this class of compounds. Two interesting nonprotein amino acids that fall into this category are hypoglycin A and its γ-glutamyl dipeptide analog,

hypoglycin B. Significant amounts of these amino acids are found in the unripe fruit of the Jamaican ackee tree and they cause hypoglycemia, also known as Jamaican vomiting sickness. A proposed biosynthetic pathway is shown below (**Scheme 9**), where 2-oxobutanoic acid undergoes a series of transformations to yield hypoglycin A. Other nonprotein amino acids that fall into this category are listed with their corresponding structure in **Table 2**.



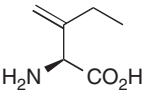
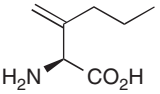
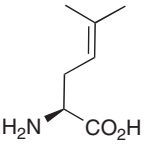
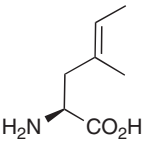
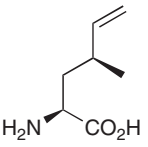
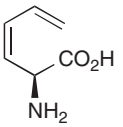
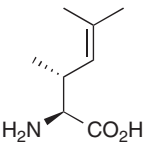
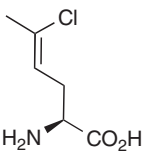
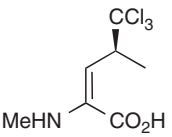
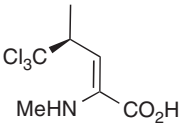
Scheme 9 Biosynthesis of hypoglycin A[13].

Table 2 Nonprotein amino acids containing alkenyl, cycloalkenyl, or alkynyl side chains

Structure	*Biological origin*	*Biological properties*	*References*
	Nonribosomal glycopeptide from *Micromonospora carbonacea*	Sch 40832 – glycopeptide antibiotic	71, 72
	Cyanobacteria – *Microcystis* and *Anabaena*	Microcystins – hepatotoxins and neurotoxins (inhibit protein phosphatases 1 and 2A)	6, 70, 73, 74
	Sponge – *Aciculites orientalis*	Aciculitins A–C and aciculitamides A–B – antifungal peptides	14, 75
	Sponge – *Theonella* sp.	Perthamide B – inhibits binding of IL-1b to thymoma cells	14, 76
	Cyanobacteria – *Nodularia spumigena*	Nodularins – hepatotoxins, inhibit protein phosphatases 1 and 2A	6, 70, 77
	Occurs in various organisms	Substrate for many enzymes and exhibits protein inhibitory activity	78
	Fungi – *Phomopsis leptostromifrmis*	Phomopsin A – antimitotic cyclic peptide	78

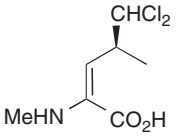
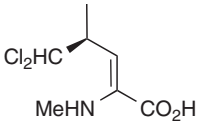
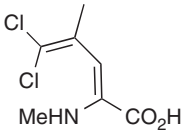
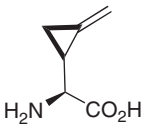
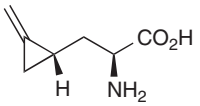
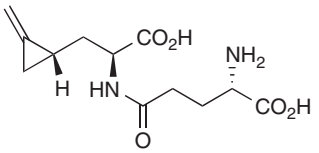
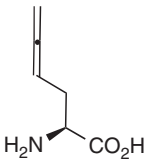
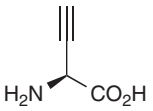
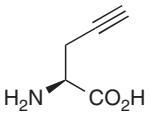
(*Continued*)

Table 2 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Lactarius helvus* and *Philadelphus coronaries*	No known bioactivity	78, 79
	Isolated from *Amanita vaginata*	No known bioactivity	78
	Mushroom – *Leucocortinarius bulbiger*	No known bioactivity	80
	Isolated from seeds of *Aesculus californica*	No known bioactivity	81
	Isolated from New Guinea *Boletus*	No known bioactivity	82
	Beetles – *Doryphorina* sp. and *Platyphora kollari* Fungi – *Clavulinopsis helvola*	Defensive secretions of beetles	78
	Marine bacteria – *Streptomyces* sp.	Cyclomarins A–C – peptides with anti-inflammatory property	83, 84
	Mushroom – *Amanita* sp.	No known bioactivity	85
	Marine sponge – *Dysidea* sp. Cyanobacteria – *Oscillatoria spongeliae*	Dysamides – chlorinated diketopiperazines with no known bioactivity	14, 26, 55
	Marine sponge – *Dysidea* sp.	Dysamides – chlorinated diketopiperazines with no known bioactivity	14, 26

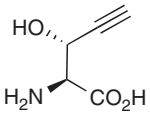
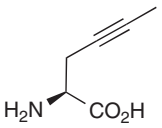
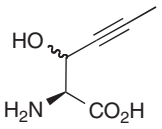
(*Continued*)

Table 2 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Cyanobacteria – *Oscillatoria spongeliae*	Didechlorodihydrodysamide C – chlorinated diketopiperazines with no known bioactivity	14, 55
	Marine sponge – *Dysidea* sp.	Dysamides – chlorinated diketopiperazines with no known bioactivity	14, 26
	Marine sponge – *Dysidea* sp.	Dysamides – chlorinated diketopiperazines with no known bioactivity	14, 26
	Isolated from seeds of *Billia hippocastanum* and fruits of *Acer pseudoplatanus*	2-(Methylenecyclopropyl)glycine – hypoglycemic activity and antimutagenic activity	65, 86–89
	Fruits from Jamaican ackee tree – *Blighia sapida, Sapindaceae* sp., and other tropical plants	Hypoglycin A – highly toxic and hypoglycemic property	13
	Fruits from Jamaican ackee tree – *Blighia sapida, Sapindaceae* sp., and other tropical plants	Hypoglycin B – highly toxic and hypoglycemic property	13
	Mushroom – *Amanita* sp.	No known bioactivity	85
	Isolated from *Streptomyces catenulae*	Amino acid with antimicrobial activity (inhibitor of alanine racemase)	90
	Isolated from *Streptomycetes* sp.	Antibacterial and antifungal amino acids (irreversible deactivation of pyridoxal-5′-phosphate (PLP)-dependent enzymes)	90

(*Continued*)

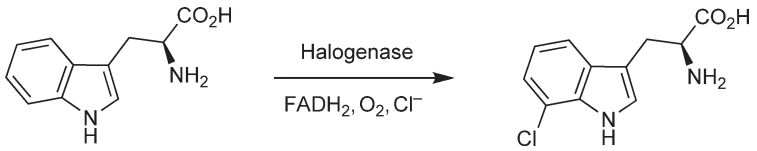
Table 2 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Fungi – *Sclerotium rolfsii*	Cytotoxic amino acid	91
	Isolated from *Tricholomopsis rutilans* and New Guinea fungus	No known bioactivity	92
 (both diastereomers)	Fungi – *Tricholomopsis rutilans*	No known bioactivity	93

5.02.3.3 Amino Acids with Aryl Side Chains

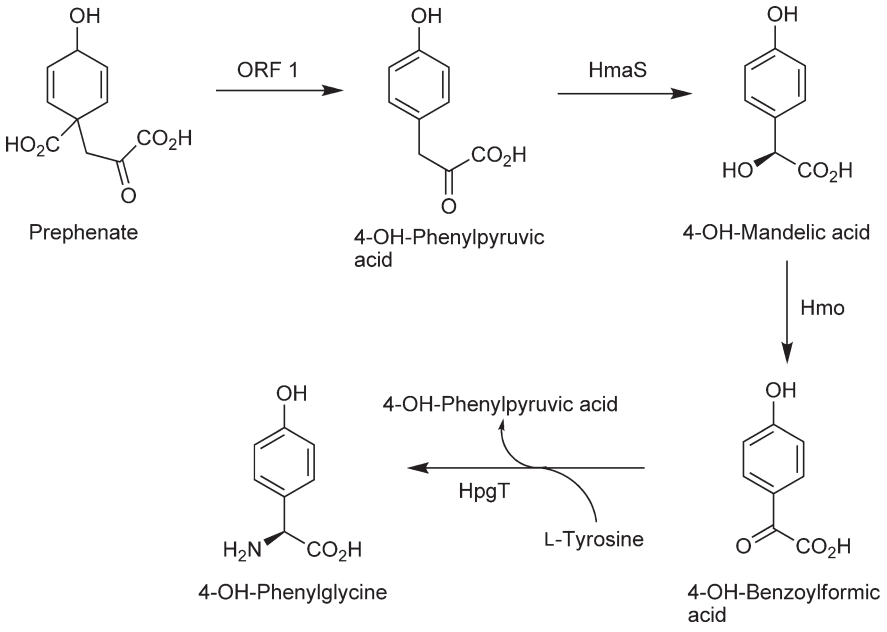
Nature utilizes the shikimate pathway for the biosynthesis of amino acids with aryl side chains. These nonprotein amino acids are often synthesized through intermediates found in the shikimate pathway. In many cases, L-α-amino acids are functionalized at different sites to yield nonprotein amino acids. These modifications include oxidation, hydroxylation, halogenation, methylation, and thiolation. In addition to these modifications, nature also utilizes modified biosynthetic pathways to produce compounds that are structurally more complex. When analyzing the structures of these nonprotein amino acids, one can generally identify the structural similarities to one of the L-α-amino acids with aromatic side chains.

Rebeccamycin is an antitumor or antibiotic agent isolated from bacteria[94] and contains a maleimide indolocarbazole framework. The Trp units in this molecule are chlorinated at the C-7 position and a flavin-dependent halogenase was identified as the enzyme that carries out this chlorination (**Scheme 10**).[24,95] There are many other halogenating enzymes known in the literature and these enzymes are responsible for the syntheses of metabolites containing bromine, chlorine, and fluorine.[29]



Scheme 10 Biosynthesis of 7-chlorotryptophan.

The glycopeptide antibiotics such as vancomycin and chloroeremomycin are complex nonribosomal peptides. One of the nonprotein amino acids found in these antibiotics is 4-hydroxyphenylglycine. The biosynthetic pathway of this nonprotein amino acid has been studied and prephenate was identified as the precursor. The biosynthetic pathway is described below (**Scheme 11**).[96] It is interesting to note that L-tyrosine is utilized during the transamination step to yield 4-OH-phenylglycine, but also 4-OH-phenylpyruvic acid that feeds back into the pathway and can be utilized in subsequent cycles.



Scheme 11 Biosynthesis of 4-hydroxyphenylglycine.[96]

Generally, most of the nonprotein amino acids containing aryl or functionalized aryl side chains are part of nonribosomal peptides isolated from bacteria, fungi, and sponges. These peptides exhibit interesting biological activities that include antimicrobial, antitumor, antifungal, and other inhibitory activities. **Table 3** describes many of these types of amino acids.

5.02.3.4 Amino Acids with Amino Groups as Part of the Side Chain

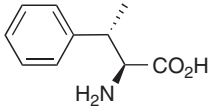
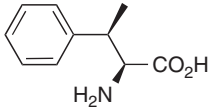
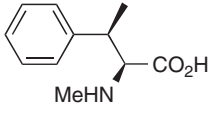
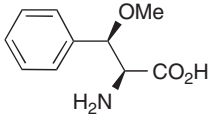
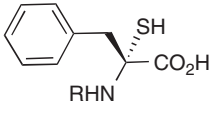
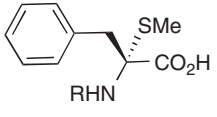
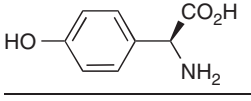
This section focuses on those nonprotein amino acids that contain amino groups in the side chain. Ornithine, which is classified as a nonprotein amino acid, was identified as an intermediate in arginine biosynthesis. Another interesting amino acid that has been isolated from certain leguminous plants is L-canavanine.[13] This nonprotein amino acid is an analog of arginine and it exhibits antimetabolic activity in bacteria and fungi. It has been proposed that the biosynthesis of canavanine begins with the conversion of L-aspartic acid to L-canaline. Then a series of enzymatic transformations gives rise to L-canavanine (**Scheme 12**).[13] In this example, nature uses a dedicated pathway for the synthesis of canavanine.

The nonprotein amino acid *β*-*N*-methylamino-L-alanine (BMAA) is a neurotoxin found in various species of marine cyanobacteria. This nonprotein amino acid occurs both as a free amino acid and bound to proteins.[167] Siderophores are secondary metabolites generally produced by bacteria under iron-deficient conditions. These molecules sequester and transport ferric ion via active transport in bacteria.[168] Some known siderophores are nonribosomal peptides that contain nonprotein amino acids with terminal amines or hydroxylamine side chains, such as exochelin MN.[169] Many of the known nonprotein amino acids of this type are tabulated in **Table 4**.

5.02.3.5 Amino Acids with Hydroxyl Groups as Part of the Side Chain

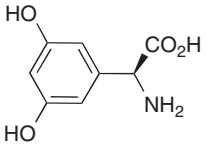
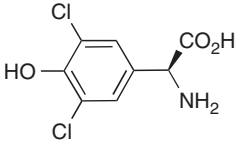
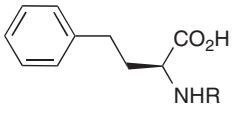
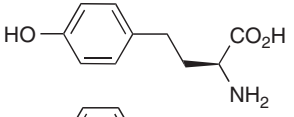
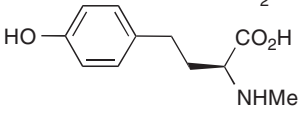
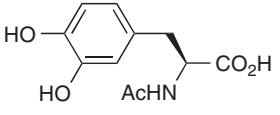
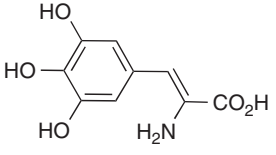
Enzymatic hydroxylation of biological molecules is often catalyzed by hydroxylases. These types of enzymes are either oxygenases or peroxidases, in which the source of oxygen is O_2 or H_2O_2, respectively. Cytochrome P-450-dependent enzymes[191] represent a common class of enzymes that carry out hydroxylation reactions. L-Carnitine is a metabolite isolated from many organisms and its biosynthesis begins with the enzymatic hydroxylation of trimethyllysine. The intermediate, 3-hydroxyl-*ε*-(*N*,*N*,*N*-trimethyl)-L-lysine, is further

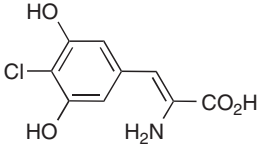
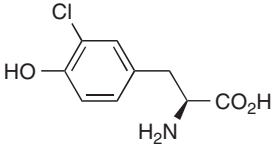
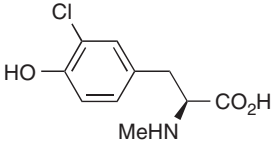
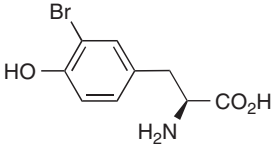
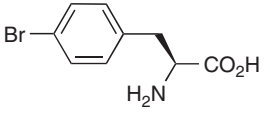
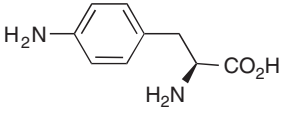
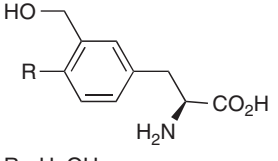
Table 3 Nonprotein amino acids containing aryl or functionalized aryl side chains

Structure	*Biological origin*	*Biological properties*	*References*
	Bacteria – *Streptomyces bottropenis*	Bottromycin A2 – peptide antibiotic	97, 98
	Bacteria – *Streptomyces* sp.	SB-203208 – peptide-based inhibitor of isoleucyl tRNA synthetase	99, 100
	Isolated from *Streptomyces griseoflavus* strain W 384	Hormaomycin – peptide with antimicrobial and antimalarial activity	65, 101, 102
	Marine bacteria – *Streptomyces* sp. and *Salinispora arenicola*	Cyclomarins A–D – peptides with anti-inflammatory property	83, 84
 R = H, CH_3	Insect pathogenic fungi – *Verticillium hemipterigenum*	Vertihemiptellides A and B – diketopiperazines with antimicrobial and cytotoxic activities	103
 R = H, CH_3	Insect pathogenic fungi – *Verticillium hemipterigenum* Fungi – *Gliocladium virens*	Diketopiperazines from fungi with antimicrobial and cytotoxic activities	103, 104
	Nonribosomal peptides from*Streptomyces* sp. WK-3429	Chloropeptins I and II – HIV inhibitors	96, 105, 106
	Streptomyces toyocaensis	A47934 – peptide antibiotic	96, 107, 108

(*Continued*)

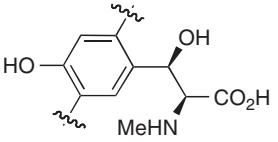
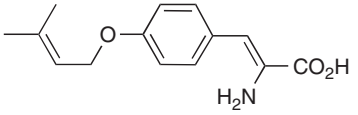
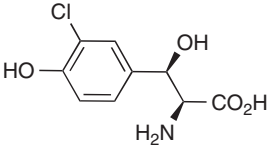
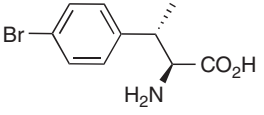
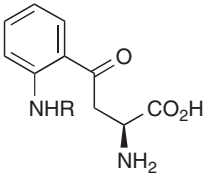
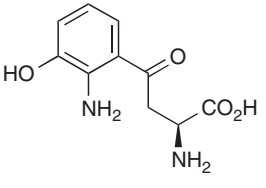
Table 3 (Continued)

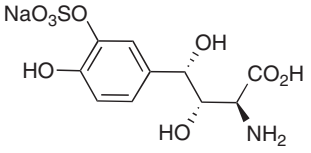
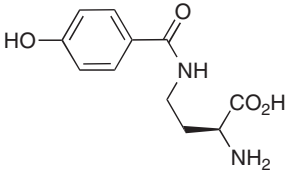
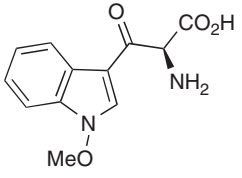
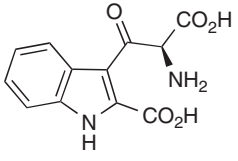
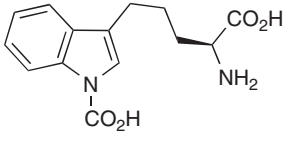
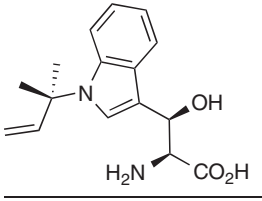
Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Streptomyces toyocaensis*	A47934 – peptide antibiotic (glycopeptide antibiotics, vancomycin and chloroeremomycin, also contain this amino acid)	107, 108 24
	Bacteria – *Streptomyces lavendulae*	Complestatin – peptide with inhibitory activity against protease	109, 110
 R = H, Me	Cyanobacteria – *Anabaena* sp.	Anabaenopeptins NZ825, NZ841, and NZ857 – nonribosomal peptides with no known biological activity	111
	Cyanobacteria – *Anabaena* spp. and *Oscillatoria* sp. (*Planktothrix* spp.)	Anabaenopeptins A–D – cytotoxic peptides	6, 112
	Cyanobacteria – *Planktothrix* spp.	Oscillamides – protein phosphatase (1 and 2A) inhibitors	14, 113
	Cyanobacteria – *Oscillatoria agardhii*	Anabaenopeptins G and H – carboxypeptidase A inhibitors	114
	Isolated from *Streptomyces akiyoshiensis*	Amino acid with cytotoxicity toward melanoma cells	115
	Patagonian sponge – *Cliona chilensis*	Celenamides – peptide-based secondary metabolites	14, 116

	Isolated from *Streptomyces platenis*	Resormycin – herbicidal, antifungal, and antibacterial peptides	117–119
	Isolated from *Streptomyces toyocaensis*	A47934 – peptide antibiotic (glycopeptide antibiotics, vancomycin and chloroeremomycin, also contain this amino acid)	107, 108 24
	Cyanobacteria – *Microcystis aeruginosa*	Micropeptins 478-A and -B – inhibitors of serine protease, plasmin	120
	Marine ascidian – *Diazona chinensis*	Diazonamide B – nonribosomal peptide with cytotoxicity	9
	Sponge – *Halichondria cylindrata*	Halicylindramides – antifungal peptides Polydiscamide A – cytotoxic peptides	14, 41 14, 43
	Plants – seeds of *Vigna* sp.	*p*-Aminophenylalanine – inhibits bacterial growth	121
 R = H, OH	Plant – seeds of *Caesalpinia tinctoria*	Not available	121–123

(*Continued*)

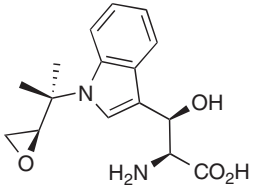
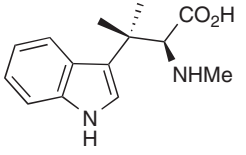
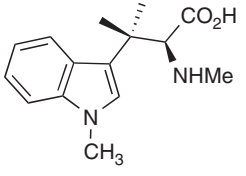
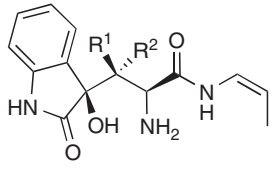
Table 3 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Rice plant pathogenic fungi – *Ustilaginoidea virens*	Ustiloxins A and F – antimitotic peptides	124–126
	Fungi – *Tolypocladium* sp.	Sch 56396 – diketopiperazine that inhibits c-fos protooncogene	127
	Isolated from *Streptomyces toyocaensis*	A47934 – peptide antibiotic (glycopeptide antibiotics, vancomycin and chloroeremomycin, also contain this amino acid)	107, 108 24
	Isolated from theonellid sponges	Theonellamides (theonegramide and theopalauamide) – cytotoxic and antifungal peptides	14, 128–130
	Plants – intermediates in tryptophan metabolism	Intermediate in the biosynthesis of niacin (vitamin B_3)	12
	Plants – intermediate in tryptophan metabolism	Not available	12

	Isolated from *Coleophoma empedri*	FR901379 – echinocandin-like lipopeptide with antifungal property	131, 132
	Sponge – *Theonella cupola*	Cupolamide A – cytotoxic peptide	14, 133
	Plant pathogenic fungi – *Fusarium pallidoroseum*	Apicidins – coccidiostats and antimalarial peptides	8
	Sponge – *Microscleroderma* sp.	Microsclerodermins A and B – antifungal peptides	134, 135
	Marine sponge – *Rhaphisia pallida*	Pallidin – secondary metabolite (diketopiperazine)	136
	Marine bacteria – *Streptomyces* sp. and *Salinispora arenicola*	Cyclomarins C and D – peptides with anti-inflammatory property	83, 84

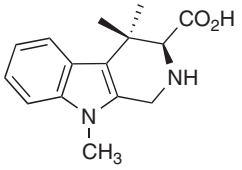
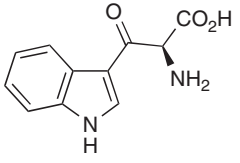
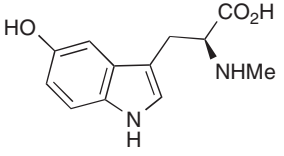
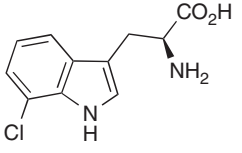
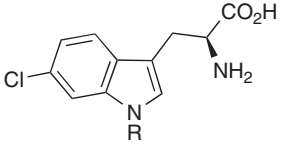
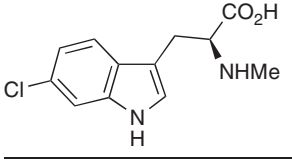
(*Continued*)

Table 3 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Marine bacteria – *Streptomyces* sp.	Cyclomarins A and B – peptides with anti-inflammatory property	83, 84
	Sponge – *Hemiasterella minor*, *Cymbastela* sp.	Hemiasterlins A and B and criamide A – cytotoxic peptides	14, 42, 137
	Sponge – *Hemiasterella minor*, *Cymbastela* sp.	Hemiasterlin C and criamide B – cytotoxic peptides	14, 42, 137
	Plant pathogenic fungi – *Apiospora montagnei* Sacc. TC 1093	TMC-95A-D – peptide-based proteosome inhibitors	138, 139

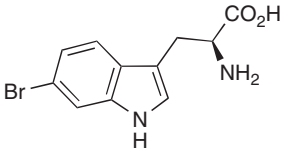
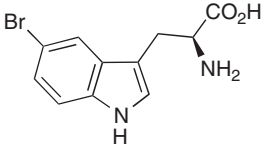
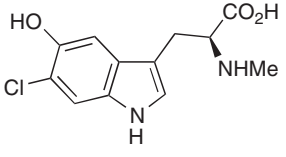
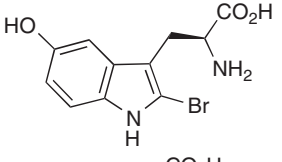
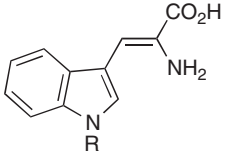
A, B: $R^1 = H$, $R^2 = OH$

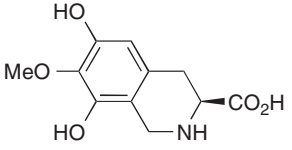
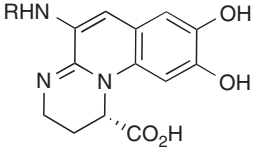
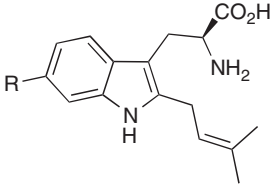
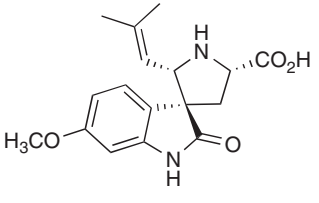
C, D: $R^1 = OH$, $R^2 = H$

	Sponge – *Auletta* cf. *constricta*	Milnamide A – cytotoxic peptide	14, 140
	Marine fungi – *Hypoxylon oceanicum*	15G256g – lipopeptide with antifungal property	141, 142
	Isolated from Theonellid sponge from Mozambique	Mozamides A and B – peptides with unknown biological activity	14, 143
	Sponge – *Theonella* sp.	Discobahamins A and B – cytotoxic peptides	36
	Bacteria – *Saccharothrix aerocolonigenes*	Rebeccamycin – antitumor and antibiotic agents (topoisomerase I inhibitor)	94, 144
 R = H, $CONH_2$	Sponge – *Microscleroderma* sp.	Microsclerodermins C and D – antifungal peptides	135
	Sponge – *Theonella* sp.	Keramamide L – cytotoxic and antifungal peptide	14, 145

(*Continued*)

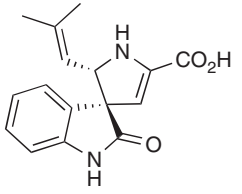
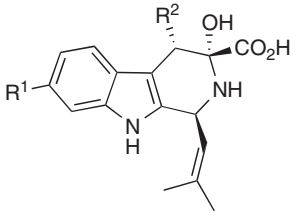
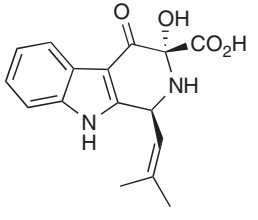
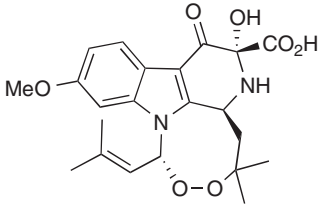
Table 3 (Continued)

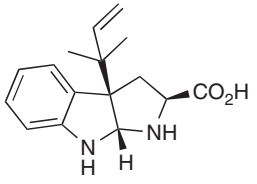
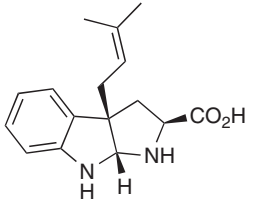
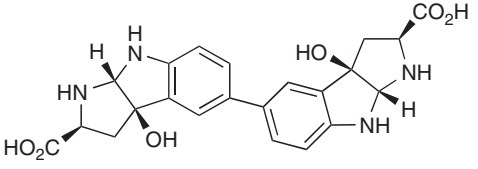
Structure	*Biological origin*	*Biological properties*	*References*
	Patagonian sponge – *Cliona chilensis*	Celenamides – peptides with unknown biological activity	14, 116
	Marine sponge – *Psammocinia* sp.	Cyclocinamide A – cytotoxic peptide	146
	Marine sponge – *Theonella* sp.	Keramamide A – cytotoxic peptide	147
	Marine sponge – *Theonella* sp.	Konbamide – peptide metabolite with calmodulin antagonistic activity Keramamides B, E, M, and N – cytotoxic peptides	14, 148 14, 33, 34
 R = H, Me	Marine sponge – *Theonella* sp. Lithistid sponge – *Microscleroderma* sp.	Keramamides F, G, and K – cytotoxic peptides Microsclerodermins F–I – antitumor and antifungal peptides	149, 150 151

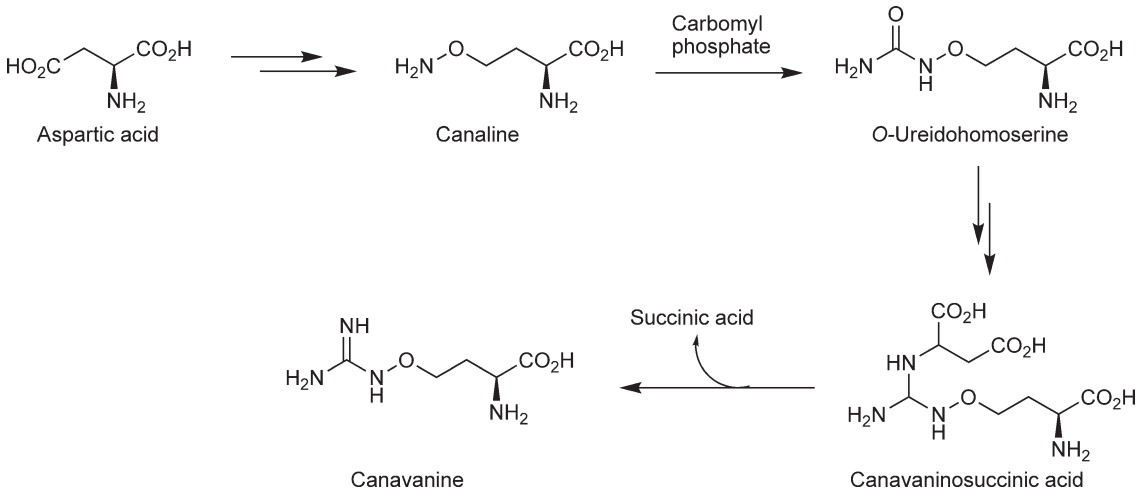
	Fungi – *Aspergillus oryzae*	TMC-2A, -2B, and -2C – dipeptidyl peptidase IV inhibitors	152, 153
 R = CH_2-CH_2-CO-CO_2H R = CH_2-CH_2-$CHNH_2$-CO_2H R = CH_2-CH_2-CO_2H	Pathogenic bacteria – *Pseudomonas aeruginosa*	Pyoverdin and other siderophores from bacteria – molecules that mediate iron uptake	154, 155
 R = H, OCH_3	Fungi – *Aspergillus fumigatus*	Tryprostatins A and B – diketopiperazines that inhibit mammalian cell cycle	156
	Fungi – *Aspergillus fumigatus*	Spirotryprostatin A – inhibitor of mammalian cell cycle	157

(*Continued*)

Table 3 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Fungi – *Aspergillus fumigatus*	Spirotryprostatin B – diketopiperazine that inhibits mammalian cell cycle	157
 A: $R^1 = OCH_3$, $R^2 = OH$ B: $R^1 = OCH_3$, $R^2 = OCH_3$ C: $R^1 = H$, $R^2 = OH$	Fungi – *Aspergillus fumigatus*	Cyclotryprostatins A–C – diketopiperazines that inhibit mammalian cell cycle at G2/M phase	158
	Fungi – *Aspergillus fumigatus*	Cyclotryprostatin D – diketopiperazine that inhibits mammalian cell cycle at G2/M phase	158
	Fungi – *Aspergillus fumigatus*	13-Oxoverruculogen – diketopiperazine with cytotoxic property Analogs of spirotryprostatins and fumitremorgin B were also isolated	159

	Fungi – *Penicillium brevicompactum* and *Aspergillus janus*	Brevicompanines A–C – peptides with antiplasmodial property and plant growth-regulating activity	160–162
	Cyanobacteria – *Microcystis aeruginosa*	Kawaguchipeptins A and B – cyclic peptides with antibacterial activity	163, 164
	Isolated from *Streptomyces hygroscopicus*	Himastatin – dimeric peptide with antitumor and antibiotic activity	165, 166



Scheme 12 Biosynthesis of canavanine.[13]

transformed to give L-carnitine (**Scheme 13**).[13] The hydroxylation is carried out by trimethyllysine dioxygenase, a nonheme Fe^{II}-dependent enzyme that uses O_2 and 2-oxoglutarate as substrates.

Sphingofungins A–F are a family of nonprotein amino acids isolated from *Aspergillus fumigatus*.[192] These amino acids have polyhydroxylated long lipid chains and exhibit antifungal properties. Salinosporamide A, an inhibitor of the 20S proteasome was isolated from marine bacteria.[193] This chlorinated nonribosomal peptide contains a nonprotein amino acid that is proposed to come from the shikimate pathway.[194] Fluorinated amino acids present another interesting class of halogenated nonprotein amino acids. The mechanism by which fluorine is incorporated into secondary metabolites is very interesting and has been intensively studied.[195] One of the known fluorinated amino acids is 4-fluorothreonine, and it was isolated from bacteria.[196] A general overview of the biosynthesis of 4-fluorothreonine is shown below (**Scheme 14**).[195] **Table 5** also contain nonprotein amino acids that have halide or thiol groups in addition to hydroxyl groups on their side chains.

5.02.3.6 Amino Acids with Carbonyl Groups as Part of the Side Chain

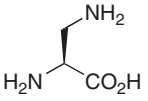
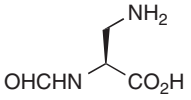
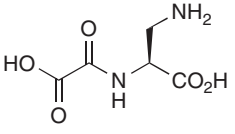
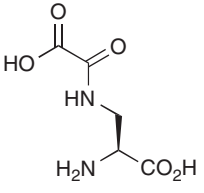
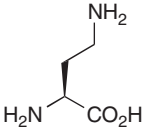
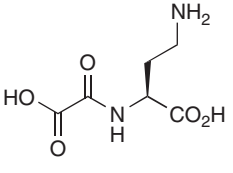
There is only a small selection of nonprotein amino acids that contain carbonyl groups in the form of ketone, aldehyde, and carboxylic acid moieties, as part of the side chain. The examples given in **Table 6** are components of nonribosomal peptides isolated from bacteria or fungi and siderophores from bacteria. The biosynthesis of these amino acids is not clear; however, some of the amino acids with carboxylic acid side chains may be traced back to the L-α-amino acids aspartic acid and glutamic acid.

5.02.3.7 Amino Acids with Heterocyclic Side Chains

The only L-α-amino acid that contains a cyclic structure (pyrrolidine ring) is proline and in this chapter, proline is classified as an amino acid with a heterocyclic side chain. Many nonprotein amino acids with a proline core have been reported. In several cases, this core is further decorated by different functional groups. Kainoid amino acids represent a family of nonprotein pyrrolidine dicarboxylic amino acids, including kainic acids, domic acids, and acromelic acids.[230] This family of amino acids exhibits cytotoxicity and insecticidal activities. Many proline analogs that have a bicyclic ring system have also been identified from various marine sources. Aeruginosins from cyanobacteria contain this type of bicyclic ring system and exhibit protease inhibitory activity. Many diketopiperazines from pathogenic fungi have been isolated with bicyclic proline analogs. One interesting structural feature of these compounds is the presence of thioether and disulfide bridges.[231,232]

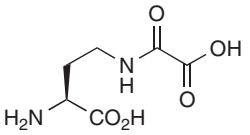
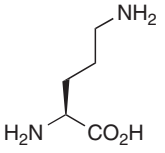
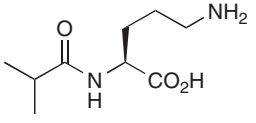
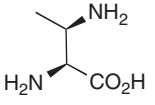
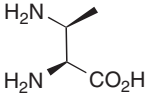
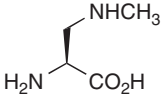
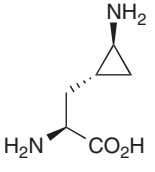
Nature has extended its ability to construct structurally unique nonprotein amino acids by synthesizing four-membered rings, six-membered rings, and pyrrole rings. Most of these ring systems have been identified as

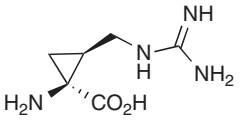
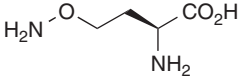
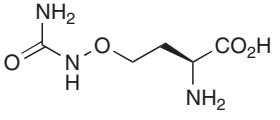
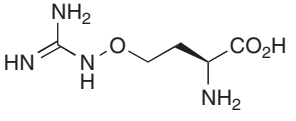
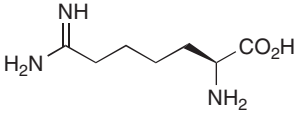
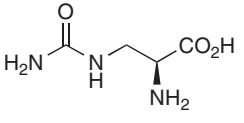
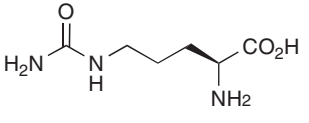
Table 4 Nonprotein amino acids containing amino groups as part of the side chain

Structure	*Biological origin*	*Biological properties*	*References*
	Nonribosomal peptides from marine sponge – *Theonella* sp.	Cyclotheonamides A, C, and D – thrombin inhibitors Keramamides F–H, J, and K – cytotoxic peptides	14, 170, 171 149, 150
	Marine sponge – *Discodermia* sp. Marine sponge – *Theonella* sp.	Pseudotheonamides A–D – weak inhibitors of thrombin and trypsin Cyclotheonamides E, E2, and E3 – thrombin inhibitors	14, 172 14, 173
	Seeds of *Lathyrus* sp.	α-ODAP (oxalyl-diaminopropionic acid) – nonprotein amino acid with no reported biological activity	121
	Seeds of *Lathyrus* sp.	β-ODAP – nonprotein amino acid with neurotoxic property	121
	Fungi – *Trichoderma atroviride*	Cupolamide A – cytotoxic peptide	14, 133
	Seeds of *Lathyrus* sp.	α-ODAB – nonprotein amino acid with neurotoxic activity	121, 174

(*Continued*)

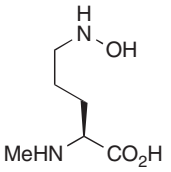
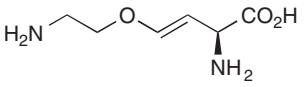
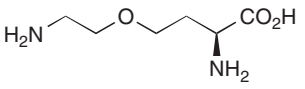
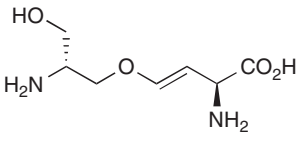
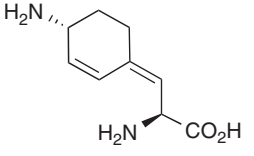
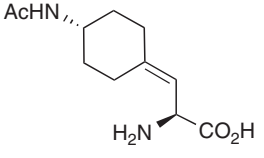
Table 4 (Continued)

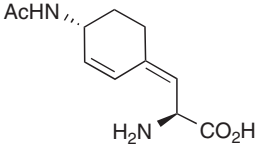
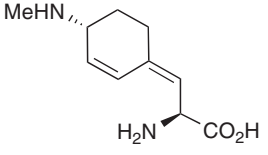
Structure	*Biological origin*	*Biological properties*	*References*
	Seeds of *Lathyrus* sp.	γ-ODAB – nonprotein amino acid that targets glutamate receptor sites (neurotoxic activity)	121, 174
	Sponge – *Theonella cupola*	Discobahamins A and B, Keramamides B–E, M, and N – cytotoxic peptides	14, 33, 34, 150
	Bacteria – *Streptomyces resistomicificus*	FR901277 – cyclic peptide that inhibits human leukocyte elastase	175, 176
	Nonribosomal peptide from sponge – *Theonella mirabilis* and *Theonella swinhoei*	Papuamides A–D – weak cytotoxicity and inhibition of HIV infection	14, 177
	Sponge – *Aciculites orientalis*	Aciculitins A–C and aciculitamides A–B – antifungal peptides	75
	Various species of marine cyanobacteria	Nonprotein amino acid – neurotoxin	167, 178, 179
	Isolated from *Streptomyces* sp.	Belactosin A – proteasome inhibitor, antiproliferative and antitumor activities Belactosin B – inactive member	180, 181

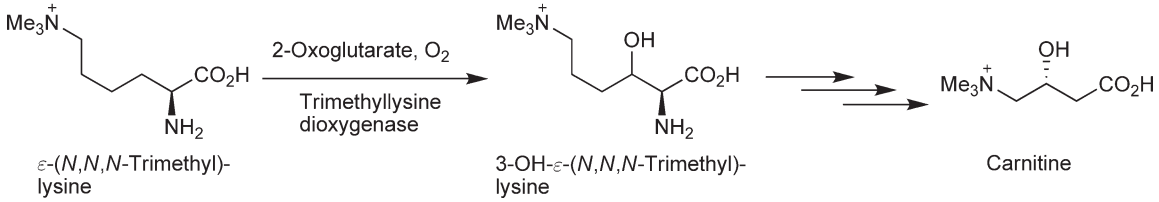
	Red algae – *Grateloupia carnosa*	Carnosadine – unnatural amino acid with anti-inflammatory activity	56, 182, 183
	Isolated from legume species	Canaline – insecticidal and inhibitor of ornithine aminotransferase	121, 184
	Seeds of *Papilionoideae* sp. and *Dioclea megacarpa*	*O*-Uredohomoserine – intermediate in canavanine biosynthesis	13
	Seeds of *Papilionoideae* sp. and *Dioclea megacarpa*	Canavanine – antimetabolite of arginine	121, 185
	Occurs in legume species *Indigofera spicata* and *Indigofera linnaei*	Indospicine – antimetabolite of arginine and shows teratogenic and hepatotoxic activity	121, 186
	Occurs in seeds and seedlings of *Albizzia julibrissin* and *Albizzia lophanta*	No known bioactivity	187
	Plants – *Citrullus* sp.	Intermediate in ornithine cycle	12

(*Continued*)

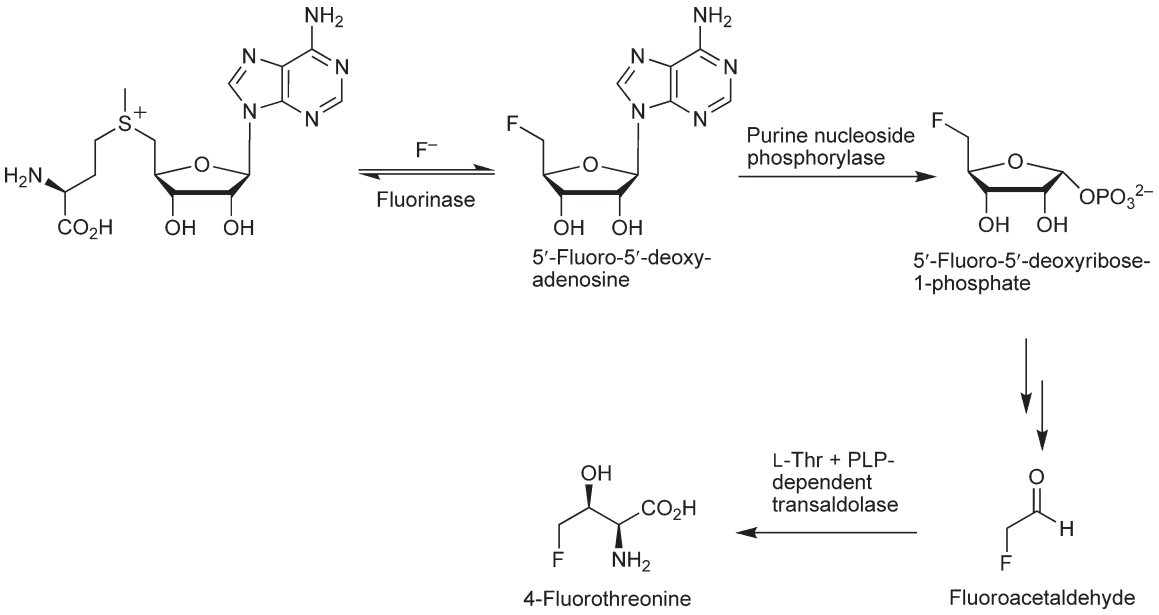
Table 4 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Mycobacterium neoaurum*	Exochelin MN – siderophore from bacteria	188
	Isolated from *Streptomyces* sp.	Inhibitor of cystathionine-processing enzymes (cystathionine β-lyase) and cystalysin	78, 189
	Isolated from *Streptomyces* sp.	Increases ethylene production in plants	190
	Bacteria – *Rhizobium japonicum* and *Pseudomonas andropogonis*	Inhibitor of cystathionine β-lyase from both *Salmonella typhimurium* and spinach	78
	Cyanobacteria – *Plectonema radiosum*	Radiosumin A – peptide-based inhibitor of trypsin	78
	Cyanobacteria – *Plectonema radiosum*	Radiosumin A – peptide-based inhibitor of trypsin	78

	Cyanobacteria – *Microcystis aeruginosa*	Radiosumin B – peptide-based inhibitor of trypsin	78
	Cyanobacteria – *Mycrocystis aeruginosa*	Radiosumin B – peptide-based inhibitor of trypsin	78



Scheme 13 Biosynthesis of L-carnitine.



Scheme 14 Biosynthesis of 4-fluorothreonine.

components of nonribosomal peptides. Papuamides are a class of nonribosomal peptides with anti-HIV property.[14] The nonprotein amino acid, pipecolic acid, was identified as a constituent of these peptides. Pipecolic acid appears to derive from lysine in the producing organisms. The first step involves an oxidative deamination of the α-amino group of lysine and further enzymatic transformations give rise to pipecolic acid (**Scheme 15**).[13]

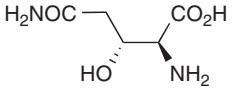
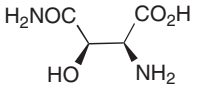
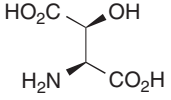
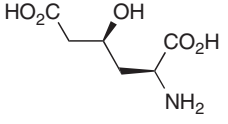
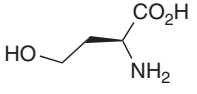
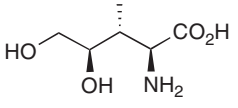
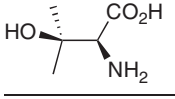
Garden beets contain significant amounts of the four-membered heterocyclic amino acid, azetidine-2-carboxylic acid. This nonprotein amino acid is cytotoxic and has teratogenic properties.[121] Other nonprotein amino acids containing an azetidine unit as part of their side chain such as nicotianamine, mugineic acid, and its analogs have also been identified from higher plants. These compounds serve as phytosiderophores that promote iron uptake. The azetidine unit is derived from SAM through an enzymatic cyclization reaction that releases 5′-methylthioade-nosine.[13]

Table 7 includes many structurally unique nonprotein amino acids with heterocyclic units as part of the structure. These amino acids are isolated from various sources and exhibit interesting bioactivities. In most cases, enzymatic modification of proline occurs through installation of hydroxyl groups, halogenated side chains, and alkyl side chains to give rise to the observed structural diversity.

5.02.3.8 Amino Acids with Other Functional Groups as Part of the Side Chain

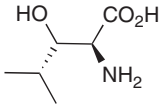
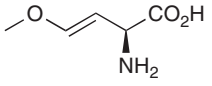
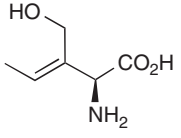
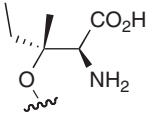
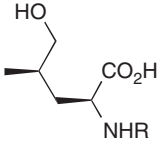
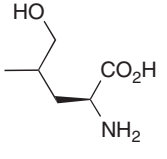
The final section of this chapter includes the nonprotein amino acids that contain functional groups that do not fall into the above-mentioned categories, such as nitro, sulfoxide, sulfoximine, nitrile, and thiol functional

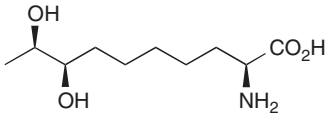
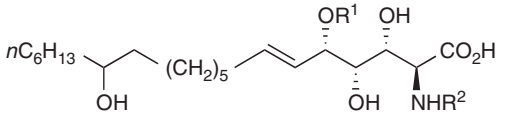
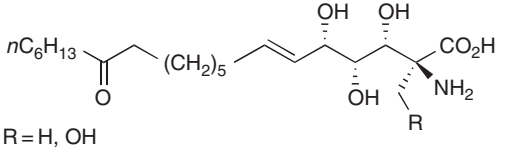
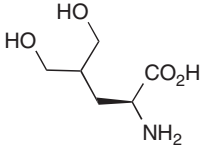
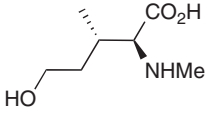
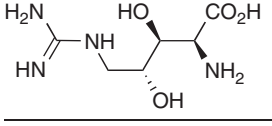
Table 5 Nonprotein amino acids containing hydroxyl groups as part of the side chain

Structure	*Biological origin*	*Biological properties*	*References*
	Sponge – *Aciculites orientalis*	Aciculitins and aciculitamides – antifungal peptides	14, 75
	Isolated from theonellid sponges	Theonellamides – cytotoxic, antifungal peptide metabolites	14, 128
	Bacteria – *Pseudomonas syringae*	Pseudomycins A–C – antifungal peptides	197, 198
	Isolated from theonellid sponges	Theonellamides – cytotoxic, antifungal peptide metabolites	14, 128
	Blue-green algae – *Scytonema* sp.	Scytonemin A – peptide-based calcium antagonist	199
	Mushroom – *Amanita phalloides*	Amanitins – peptide-based toxins	200
	Isolated from *Streptomyces platenis*	Resormycin – herbicidal, antifungal, and antibacterial peptide	117–119

(*Continued*)

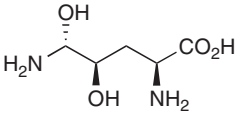
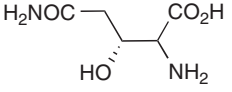
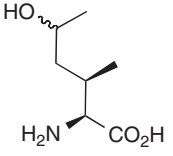
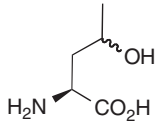
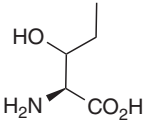
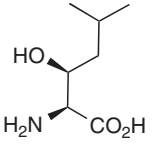
Table 5 (Continued)

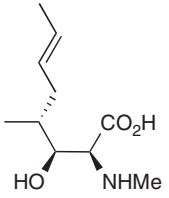
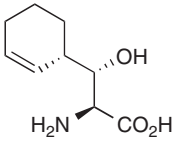
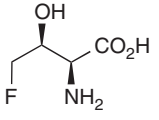
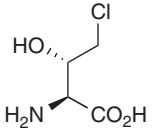
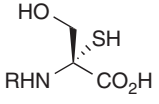
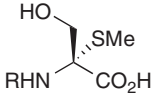
Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Streptomyces* sp.	GE3 and GE3B – antitumor and antibiotic peptides Muraymycins – inhibitors of peptidoglycan biosynthesis	201 202
	Bacteria – *Pseudomonas aeruginosa*	Substrate and inhibitor of tryptophan synthase and other enzymes	78
	Mushroom – *Bankera fulgineoalba*	No known bioactivity	78
	Rice plant pathogenic fungi – *Ustilaginoidea virens*	Ustiloxins A and F – antimitotic peptides	124–126
 R = H, Me	Marine bacteria – *Streptomyces* sp. and *Salinispora arenicola*	Cyclomarins A–D – peptides with anti-inflammatory property	83, 84
 (two diastereomers)	Fungi – *Aspergillus oryzae*	TMC-2A, -2B, and -2C – dipeptidyl peptidase IV inhibitors	152, 153

	Fungi – *Diheterospora chlamydospora* Q 58044	Diheteropeptin – peptide with TGF-β-like activity	203–205
 A: $R^1 = H$; $R^2 = C(NH)NH_2$ B: $R^1 = H$; $R^2 = H$ C: $R^1 = Ac$; $R^2 = H$ D: $R^1 = H$; $R^2 = Ac$	Fungi – *Aspergillus fumigatus*	Sphingofungins A–D – nonprotein amino acids with antifungal property	192, 206
 R = H, OH	Fungi – *Paecilomyces variotii*	Sphingofungins E and F – antifungal agents and inhibitors of serine palmitoyltransferase (SPT)	207
	Fungi – *Aspergillus oryzae*	TMC-2A, -2B, and -2C – dipeptidyl peptidase IV inhibitors	152, 153
	Sponge – *Haliclona* sp.	Helipeptins A and B – anti-inflammatory peptides	208
	Palauan sponge – *Microciona eurypa*	Eurypamide A – peptide metabolite with no known biological activity	209

(*Continued*)

Table 5 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Coleophoma empedri*	FR901379 – echinocandin-like lipopeptide with antifungal property	131, 132
	Isolated from *Coleophoma empedri*	FR901379 – echinocandin-like lipopeptide with antifungal property	131, 132
 (both diastereomers)	Isolated from *Trigonella foenumgraecum* Fungi – *Amanita phalloides*	Amino acid with antidyslipidemic and antihyperglycemic activities Amanitins – toxic peptides	12, 200, 210 12, 200
 (both diastereomers)	Isolated from seeds of *Lathyrus odoratus*	No known bioactivity	211
	Bacteria – *Streptomyces*	Cycloheptamycin – peptide antibiotic Amino acid is toxic and teratogenic	212 213
	Bacteria – *Streptomyces* sp.	Telomycin – peptide antibiotic	214

	Peptaibols from fungi – *Trichoderma* sp.	Cyclosporins D and E – antifungal peptides	38, 215
	Marine bacteria – *Salinispora tropica*	Salinosporamide A – γ-lactam that inhibits 20S proteasome	193
	Isolated from *Streptomyces cattleya*	Fluorinated amino acid with antibiotic property	195, 196
	Bacteria – *Pseudomonas syringae*	Pseudomycins A–C – antifungal peptides	197, 198
 R = H, CH_3	Insect pathogenic fungi – *Verticillium hemipterigenum*	Vertihemiptellides A and B – diketopiperazines with antimicrobial and cytotoxic activities	103
 R = H, CH_3	Insect pathogenic fungi – *Verticillium hemipterigenum* Fungi – *Gliocladium virens*	Diketopiperazines from fungi with antimicrobial and cytotoxic activities	103, 104

(*Continued*)

Table 5 (Continued)

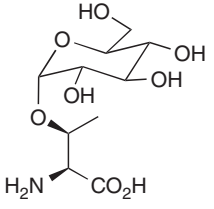
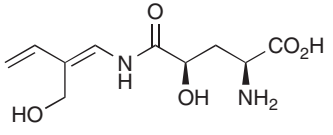
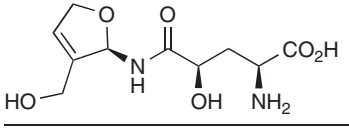
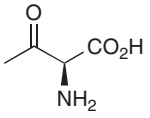
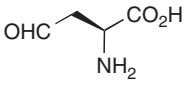
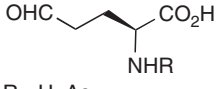
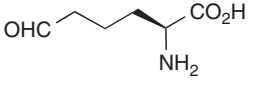
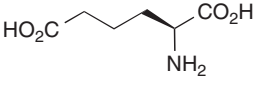
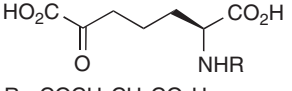
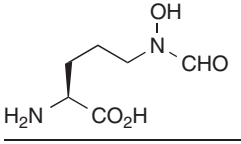
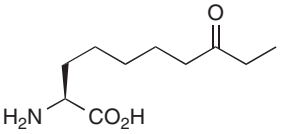
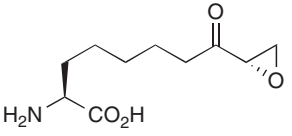
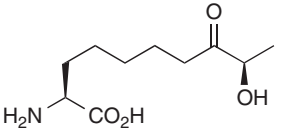
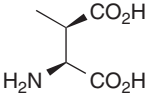
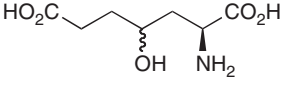
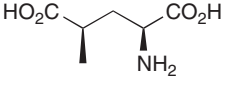
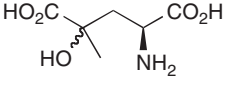
Structure	*Biological origin*	*Biological properties*	*References*
	Bacteria – *Aeromonas* sp. W-10	Sch 20562 – antifungal and antibiotic peptide	216
	Isolated from seeds of *Staphylea pinnata*	No known bioactivity	217
	Isolated from seeds of *Staphylea pinnata*	No known bioactivity	217

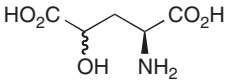
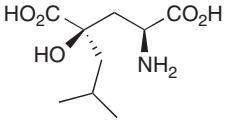
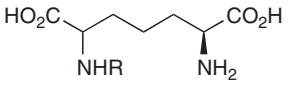
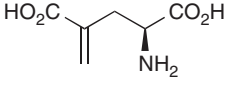
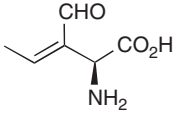
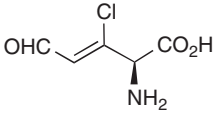
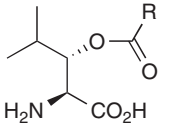
Table 6 Nonprotein amino acids containing carbonyl functional groups as part of the side chain

Structure	*Biological origin*	*Biological properties*	*References*
	Plants – intermediate in threonine metabolism	Not available	12
	Plants – intermediate in lysine biosynthesis	Not available	12
 R = H, Ac	Plants – intermediate in ornithine biosynthesis	Not available	12
	Plants – intermediate in lysine metabolism	Not available	12
	Plants – intermediate in lysine metabolism	Not available	12
 R = $COCH_2CH_2CO_2H$	Plants – intermediate in lysine biosynthesis	Not available	12
	Pathogenic bacteria – *Pseudomonas aeruginosa*	Pyoverdin and other siderophores from bacteria – molecules that mediate iron uptake	154, 155, 218

(*Continued*)

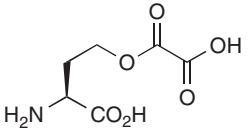
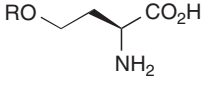
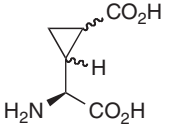
Table 6 (Continued)

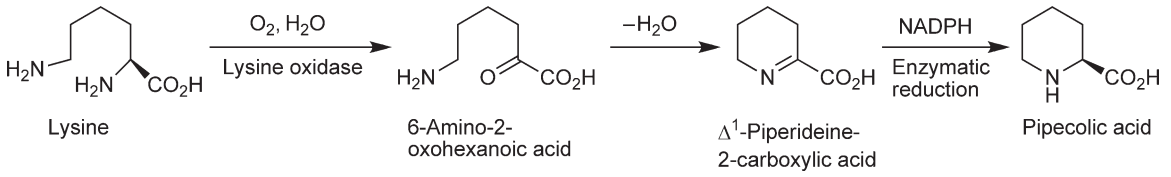
Structure	*Biological origin*	*Biological properties*	*References*
	Plant pathogenic fungi – *Fusarium pallidoroseum* Marine fungi – *Microsporum* cf. *gypseum*	Apicidins – coccidiostats and antimalarial peptides Microsporin A – histone deacetylase inhibitor	8 219
	Fungi – *Diheterospora chlamydospora*	Chlamydocin – peptide with antitumor activity	220, 221
	Fungi – *Verticillium coccosporum*	Chlamydocin analog – phytotoxic cyclic peptide	222
	Bacteria – *Actinoplanes friuliensis* sp.	Friulimicins – antimicrobial peptides that inhibit peptidoglycan synthesis	223
 (both diastereomers)	Plants – *Filicinae* sp. and subspecies of *Phyllitis scolopendrium*	No known bioactivity	224
	Plants – *Filicinae* sp. and subspecies of *Phyllitis scolopendrium*	No known bioactivity	224
 (both diastereomers)	Plants – *Filicinae* sp. and subspecies of *Phyllitis scolopendrium*	No known bioactivity	224

	Plants – *Phlox decussata*	No known bioactivity	225
	Occurs in the flowers of *Reseda odorata*	No known bioactivity	226
 R = H, $COCH_2CH_2CO_2H$	Plants – intermediate in lysine biosynthesis	Not available	12
	Plants – *Filicinae* sp. and subspecies of *Phyllitis scolopendrium*	No known bioactivity	224
	Mushroom – *Bankera fulgineoalba*	No known bioactivity	78
	Bacteria – *Streptomyces viridogenes*	No known bioactivity	78
 R = lipid chains	Bacteria – *Streptomyces* sp.	Muraymycins – inhibitors of peptidoglycan biosynthesis	202

(*Continued*)

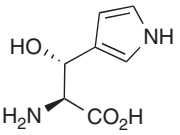
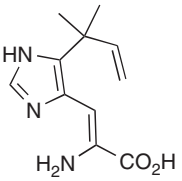
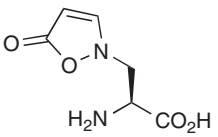
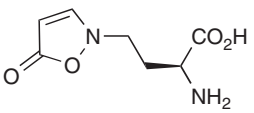
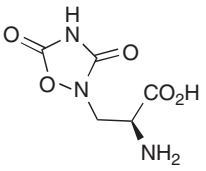
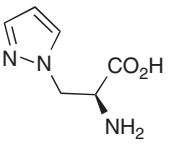
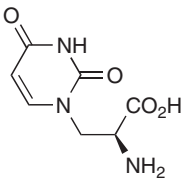
Table 6 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Leaves of the legume *Lathyrus latifolius*	*O*-Oxalylhomoserine – antifeedant for larvae of *Spodoptera littoralis*	121, 227
 R = $COCH_2CH_2CO_2H$	Plants – intermediate in the biosynthesis of methionine	Not available	12
 (three diastereomers)	Isolated from *Ephedra* species	Nonprotein amino acids from plant – with unknown bioactivity	65, 228, 229
	Immature seeds of Jamaican ackee tree – *Blighia sapida, Sapindaceae* sp., and other tropical plants	Hypoglycemic activity	13



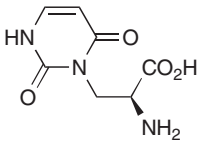
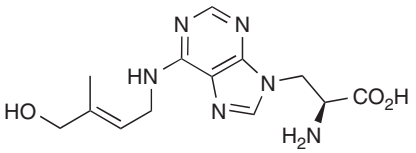
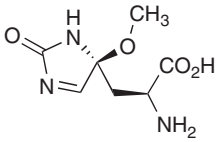
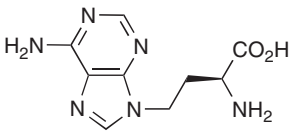
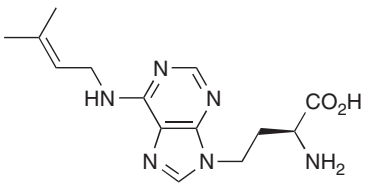
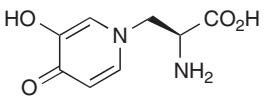
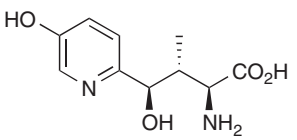
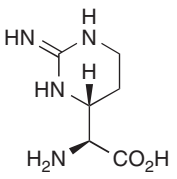
Scheme 15 Biosynthesis of pipecolic acid.

Table 7 Nonprotein amino acids containing heterocyclic side chains

Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Mycobacterium neoaurum*	Exochelin MN – siderophore from bacteria	188
	Fungi – *Aspergillus ustus*	Phenylahistin – cell cycle inhibitor	233, 234
	Seeds of *Lathyrus sativus*	Intermediate in the biosynthesis of β-ODAP, a neurotoxic amino acid	121
	Seedlings of *Lathyrus odoratus*	Secondary metabolite with no reported bioactivity	121
	Isolated from *Quisqualis* sp.	Quisqualic acid – has neuroexcitatory activity (stimulant of glutamic acid receptor)	13
	Isolated from seeds of *Citrullus* sp.	Metabolite with no known bioactivity	13
	Isolated from *Pisum*, *Acacia*, and *Fugus* species	Willardine – neuroactive amino acid	13

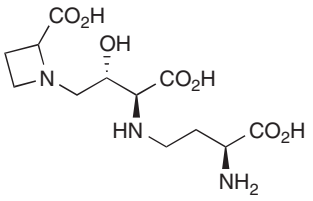
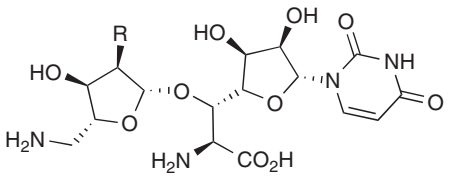
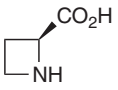
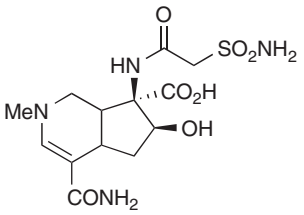
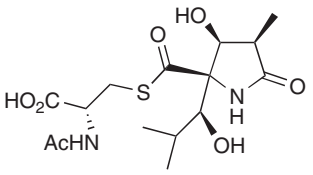
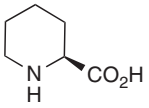
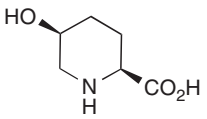
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from *Pisum* and *Crotalaria* species	Isowillardine – no known bioactivity	13
	Isolated from plants	Lupinic acid – metabolite of the cytokinin zeatin	13
	Sponge – *Aciculites orientalis*	Aciculitamides A–B – antifungal peptides	14, 75
	Fungi – *Taralomyces* sp.	NK374200 – nonprotein amino acid with insecticidal activity	235
	Fungi – *Taralomyces* sp.	NK374200 – nonprotein amino acid with insecticidal activity	235
	Isolated from *Mimosa pudica* and leaves and seeds of *Leucaena leucocephala*	Mimosine – cytotoxicity and teratogenic activity in rats and thyrotoxic	121
	Bacteria – *Streptomyces* sp.	Nikkomycins – peptide-based antibiotics	236
	Bacteria – *Streptomyces* sp.	Muraymycins – inhibitors of peptidoglycan biosynthesis	202

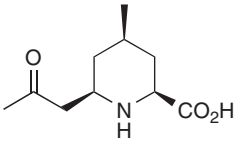
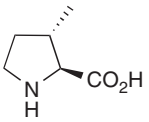
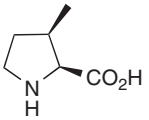
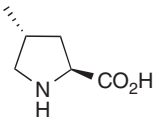
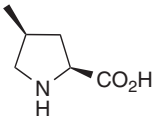
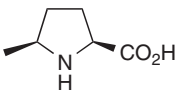
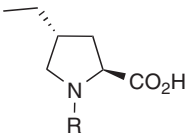
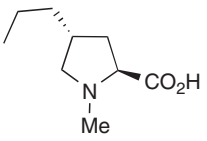
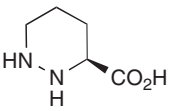
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from plants	Mugineic acid – phytosiderophore (iron chelator)	13
 R = OH, OCH_3	Bacteria – *Streptomyces* sp.	Muraymycins – inhibitors of peptidoglycan biosynthesis	202
	Garden beet – *Beta vulgaris*	Azetidine-2-carboxylic acid – cytotoxic and teratogenic properties	121, 237
	Bacteria – *Streptomyces sioyaensis*	Altemicidin – nonprotein amino acid with acaricidal and antitumor activity	238
	Bacteria – *Streptomyces* sp. OM-6519	(+)-lactacystin – γ-lactam thioester that inhibits 20S proteosome	239
	Nonribosomal peptide from sponge – *Theonella mirabilis* and *Theonella swinhoei*	Pipecolic acid – nonprotein amino acid Papuamides A–D – peptides with weak cytotoxicity and inhibition of HIV infection	14, 177
	Isolated from the leaves of *Morus alba* and the seeds of *Lathyrus japonicus*	No known bioactivity	240

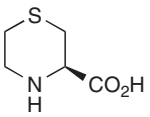
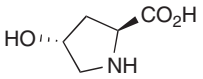
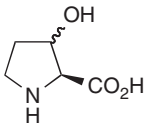
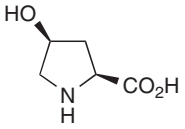
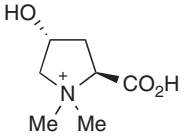
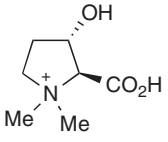
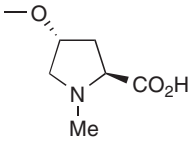
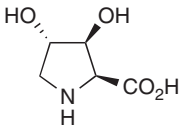
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Fungi – *Trichoderma polysporum*	No known bioactivity	217
	Blue-green algae – *Scytonema* sp. Marine fungi – *Scytalidium* sp.	Scytonemin A – peptide-based calcium antagonist Scytalidamides A and B – cytotoxic peptides	199 241
	Bacteria – *Streptomyces bottropenis*	Bottromycin A2 – antibacterial peptide	97, 98
	Bacteria – *Streptomyces sp.*	Monamycins and griselimycin – peptide antibiotics	242
	Isolated from *Paecilomyces* sp.	Leucinostatins A–D – peptide antibiotics	242
	Bacteria – *Streptomyces fradiae*	Actinomycin Z_5 – peptide antibiotic	242
 R = H, Me	Bacteria – *Actinoplanes awajinensis* Bacteria – *Streptomyces lincomycin*	Mycoplanecin A – peptide antibiotic (4-ethylproline) Lincomycin B – peptide antibiotic (*N*-methyl-4-ethylproline)	242
	Bacteria – *Streptomyces* sp.	Lincomycin and clindamycin – broad spectrum antibiotics (lincosamides)	242, 243
	Isolated from *Streptomyces* sp.	Chloptosin – dimeric nonribosomal peptide with apoptosis-inducing activity A83586 – peptide-based antibiotic	244 245

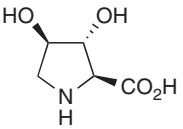
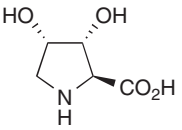
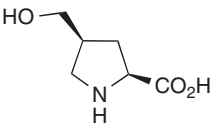
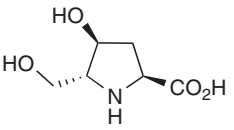
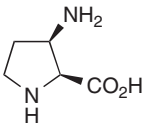
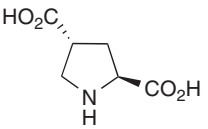
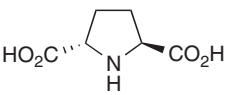
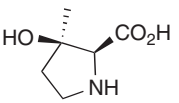
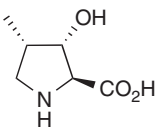
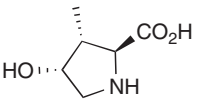
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Algea – *Heterochordaria abietina*	No known bioactivity	217
	Blue-green algae – *Anabaena laxa* Marine sponge – *Callyspongia abnormis*	Laxaphycins A and E – antifungal peptides Callynormine A – no bioactivity has been reported	246–248 249
 (both diastereomers)	Bacteria – *Streptomyces* sp. (trans isomer is also found in *Delonix regia* and many collagens)	Telomycin – cyclopeptide antibiotic	214, 242
	Fungi – *Amanita phalloides* and *Agaricus phalloides* Cyanobacteria – *Lyngbya majuscula*	Phalloidin, phallicidin, and amanitin – toxic peptides Majusculamide D – cytotoxic peptide Microcolin A – peptide with immunosuppressive, antileukemic and protein kinase C inhibitory activity	242
	Isolated from *Betonica officinalis* and *Stachys sylvatica*	Betonicine – no known bioactivity	242
	Isolated from *Courbonia virgata*	Betaine with no known bioactivity	242
	Isolated from *Petiveria alliacea*	No known bioactivity	242
	Diatom – *Navicula pelicullosa*	Component of a cell wall protein	242

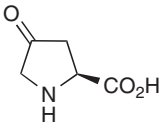
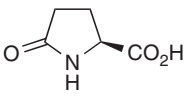
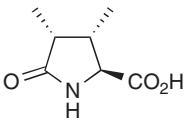
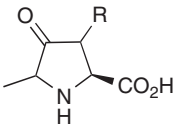
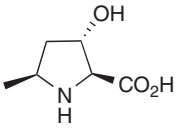
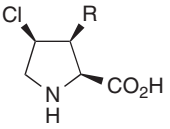
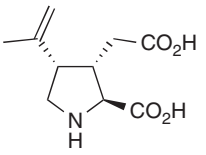
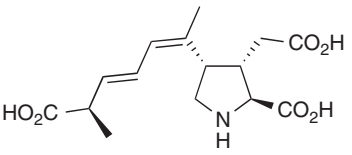
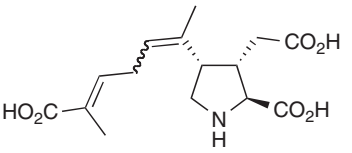
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Mushroom – *Amanita virosa*	Virotoxins – actin-binding cyclic peptides	242
	Isolated from *Coleophoma empedri*	FR901379 – echinocandin-like lipopeptide with antifungal property	131, 132
	Occurs in apples – *Malus pumila* and *Pyrus communis*	No known bioactivity	242
	Bacteria – *Pseudomonas* sp.	Bulgecins A–C – aminoglycoside antibiotics	242
	Fungi – *Morchella* sp.	No known bioactivity	242
	Isolated from the seeds of *Afzelia bella*	No known bioactivity	242
	Isolated from *Schizymenia dubyi*	No known bioactivity	242
	Isolated from *Streptomyces* sp.	Polyoxypeptin – cyclopeptide with apoptosis-inducing activity	250–252
	Isolated from *Aspergillus* sp.	Echinocandins B–D – lipopeptide antibiotics and other peptides including aculeacin A, mulundocandin, and pneumocandins	242
	Blue-green algae – *Scytonema* sp.	Scytonemin A – peptide-based calcium antagonist	199

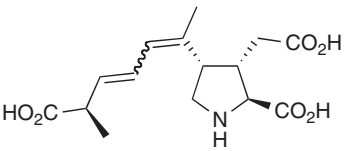
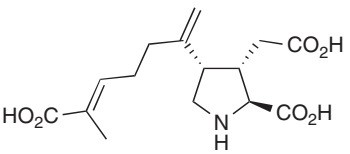
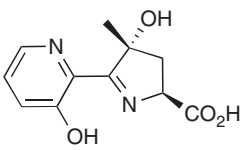
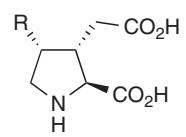
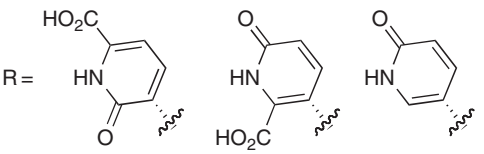
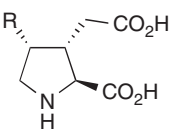
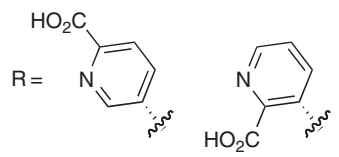
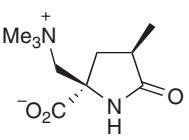
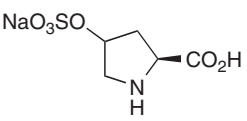
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Bacteria – *Streptomyces* sp.	Actinomycins X_2 and X_{1a} – peptide antibiotics	242
	Occurs in animal tissues	Pyroglutamic acid – component of neuropeptide hormones and glutamatergic agonist	242
	Isolated from *Callipelta* sp.	Callipeltin B – cytotoxic peptide	242, 253
 R = H, OH	Bacteria – *Streptomyces* sp.	Actinomycin Z_1 – peptide antibiotic	217
	Bacteria – *Streptomyces* sp.	Actinomycin Z_1 – peptide antibiotic	217
	Isolated form *Aster tataricus*	Astins A–C and I – cyclic peptides with antitumor activity	242, 254
	Red algae – *Digenea simplex*	α-Kainic acid – nonprotein amino acid with anthelmintic activity	230, 255
	Red algae – *Chondria armata*	Domic acid – insecticidal and neurotoxic nonprotein amino acids	230, 256
	Red algae – *Chondria armata*	Isodomic acids A and B – insecticidal nonprotein amino acids	257

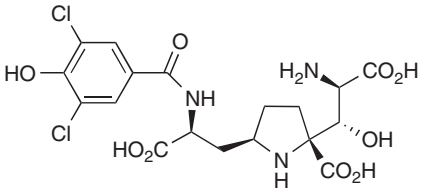
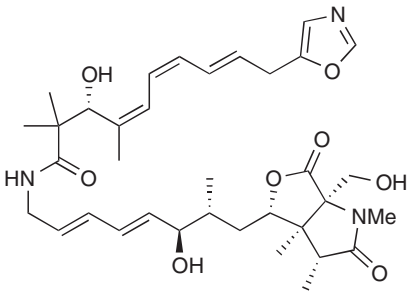
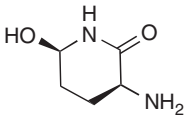
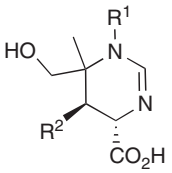
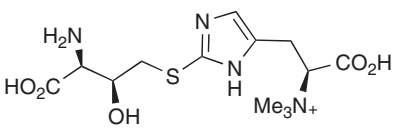
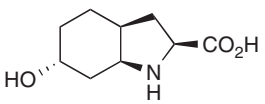
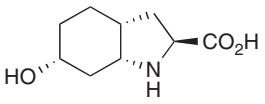
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Red algae – *Chondria armata*	Isodomic acids D–F – insecticidal nonprotein amino acids	230
	Red algae – *Chondria armata* and diatom *Pseudonitzschia australis*	Isodomic acid C – nonprotein amino acid with insecticidal activity	257, 258
	Isolated from *Norcardia species*	Siderochelin A – iron chelator with antibiotic property	242
 	Japanese mushroom – *Clitocybe acromelalga*	Acromelic acids A–C – cytotoxic nonprotein amino acids	230
 	Japanese mushroom – *Clitocybe acromelalga*	Acromelic acids D–E – cytotoxic nonprotein amino acids	230
	Marine sponge – *Dysidea herbacea*	Dysibetaine – neuroexcitotoxin (may interact with glutamate receptors)	259
	Sponge – *Theonella cupola*	Cupolamide A – cytotoxic peptide	14, 133

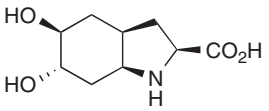
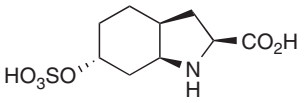
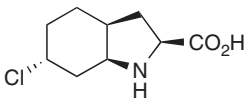
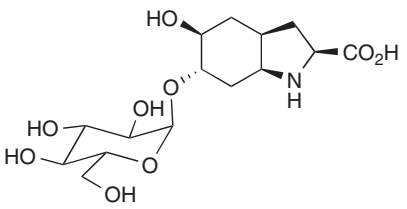
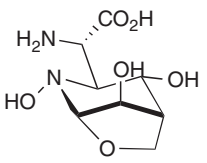
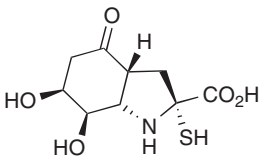
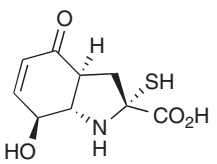
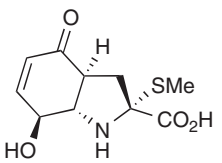
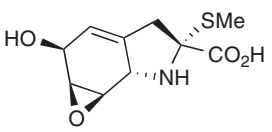
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Fungi – *Eupenicillium shearii*	Kaitocephalin – antagonist of glutamate receptor	260
	Bacteria – *Streptomyces* sp.	Neooxazolomycin – polyene metabolite with antibiotic and antitumor activity	261
	Cyanobacteria – *Microcystis aeruginosa*	Micropeptins 478-A and -B – peptide-based plasmin inhibitors Cyanopeptolins A–D – protease inhibitors Cyanopeptolin 963A – chymotrypsin inhibitor	120 6, 262 263
 R^1 = H, CH_3, R^2 = H, OH	Sponge – *Astrosclera willeyana*	Manzacidins A–D – α-adrenoreceptor blockers and serotonergic receptor antagonists	264, 265
	Toadstool – *Clitocybe acromelalga*	Clithioneine – betaine from mushroom	217
	Cyanobacteria – *Microcystis aeruginosa*	Aeruginosins 298-A and -B – thrombin and trypsin inhibitors	266–269
	Cyanobacteria – *Microcystis aeruginosa*	Aeruginosin EI461 – protease inhibitor	266, 270, 271

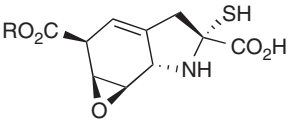
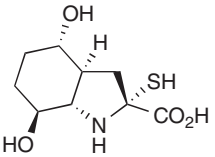
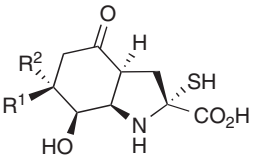
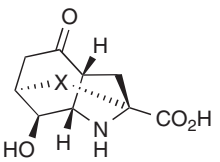
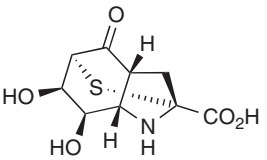
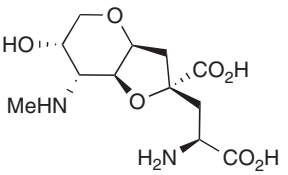
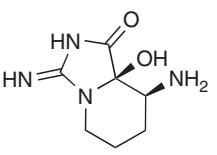
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
	Sponge – *Lamellodysidea chlorea*	Dysinosins A, C, and D – inhibitors of factor VIIa and thrombin	266, 268, 272
	Cyanobacteria – *Microcystis aeruginosa*	Aeruginosins 98-A, 98-B, 98-C, and 101 – trypsin inhibitors	266, 269, 273
	Cyanobacteria – *Oscillatoria agardhii*	Aeruginosins 205A and 205B – trypsin and thrombin inhibitors	266, 274
	Cyanobacteria – *Microcystis aeruginosa*	Dysinosin B – inhibitor of factor VIIa and thrombin	266, 272
	Bacteria – *Micromonospora* sp.	SB-219383 – a dipeptide that inhibits bacterial tyrosyl tRNA synthetase	275, 276
	Plant pathogenic fungi – *Stereum hirsutum* HKI	Epicorazine C – disulfide containing diketopiperazine with antimicrobial activity	277
	Plant pathogenic fungi – *Stereum hirsutum* HKI	Epicorazine C – disulfide containing diketopiperazine with antimicrobial activity	277
	Plant pathogenic fungi – *Exserohilum holmii*	Exserohilone – dimeric diketopiperazine with phytotoxic property	278
	Fungi – *Nigrospora sphaerica*	Epoxyexserohilone – diketopiperazine with no known bioactivity	279

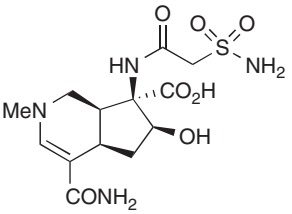
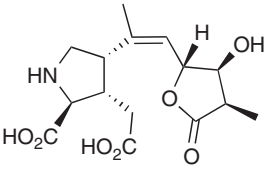
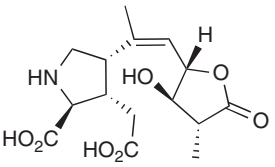
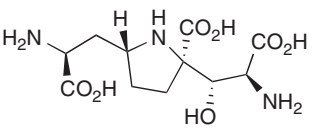
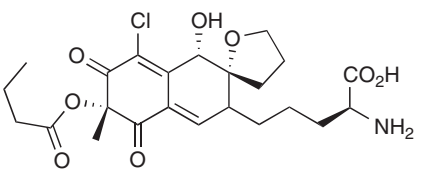
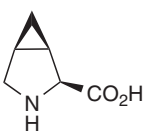
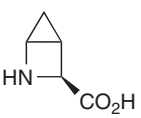
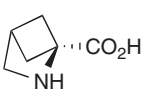
(*Continued*)

Table 7 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
 R = CH_3, $(CH_2)_2CH_3$,	Lichen – *Usnea* sp.	Ambewelamides A and B – epidithiapiperazinediones with antineoplastic property	280
	Marine fungi – *Exserohilum rostratum*	Rostratin A – disulfide containing diketopiperazine with cytotoxicity against human colon carcinoma	232
 B: $R^1 = R^2 = H$ C: $R^1 = OCH_3$, $R^2 = H$ D: $R^1 = H$, $R^2 = SH$	Marine fungi – *Exserohilum rostratum*	Rostratins C–D – disulfide containing diketopiperazines with cytotoxicity against human colon carcinoma	232
 X = S, S-S	Fungi – *Cordyceps*-colonizing isolate of *Epicoccum nigrum*	Epicoccins A–D – diketopiperazines with antimicrobial activity	231
	Fungi – *Cordyceps*-colonizing isolate of *Epicoccum nigrum*	Epicoccin B – diketopiperazine with antimicrobial activity	231
	Sponge – *Dysidea herbacea*	Dysiherbaine – nonprotein amino acid with neurotoxic properties	281, 282
	Marine sponge – *Theonella swinhoei*	Pseudotheonamides A_1, A_2, B_2, and C – peptides that inhibit serine protease	172

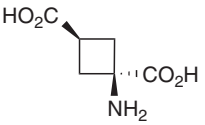
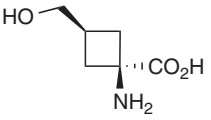
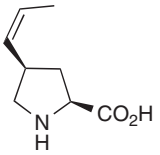
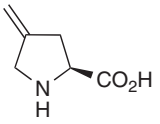
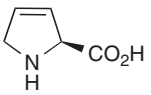
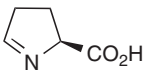
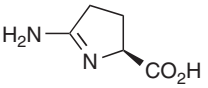
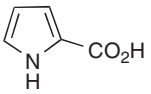
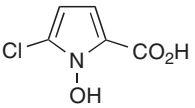
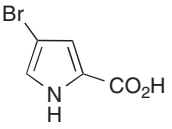
(*Continued*)

Table 7 (Continued)

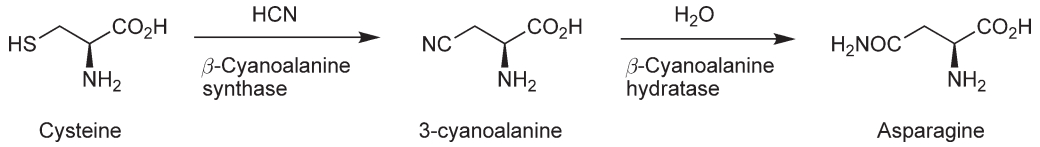
Structure	*Biological origin*	*Biological properties*	*References*
	Bacteria – *Streptomyces* sp.	SB-203207 and SB-203208 – peptide-based inhibitors of isoleucyl tRNA synthetase	99
	Red algae – *Chondria armata*	Domoilactone A – nonribosomal amino acid (domoic acid is neurotoxic)	283
	Red algae – *Chondria armata*	Domoilactone B – nonribosomal amino acid (domoic acid is neurotoxic)	283
	Fungi – *Eupenicilium shearii*	Kaitocephalin – glutamate receptor antagonist	260
	Fungi – *Fusarium* sp.	Chlorofusin – peptide-based p53-MDM2 antagonist	284–286
	Isolated from seeds of *Aesculus parviflora, Ephedra foeminea*, and *Ephedra foliata*	Unnatural amino acid from plant – with unknown bioactivity	65, 228
	Isolated from *Streptomyces zaomyceticus*	Antimicrobial agent against *Xanthomonas* sp. (antimetabolite of L-proline)	65, 287
	Isolated from legume genus *Bocoa*	2,4-Methanoproline – no known bioactivity	288

(*Continued*)

Table 7 (Continued)

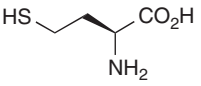
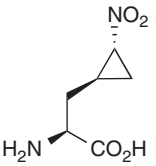
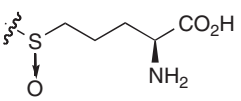
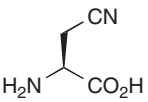
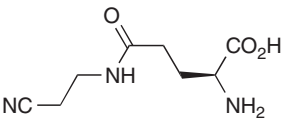
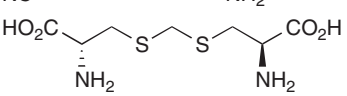
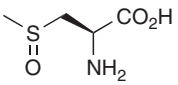
Structure	*Biological origin*	*Biological properties*	*References*
	Isolated from legume genus *Bocoa*	2,4-Methanoglutamic acid – no known bioactivity	288
	Isolated from legume genus *Bocoa*	*Cis*-1-amino-3-hydroxymethyl-cyclobutane-1-carboxylic acid – no known bioactivity	288
	Isolated from *Streptomyces griseoflavus* strain W 384	Hormaomycin – peptide with antimicrobial and antimalarial activity	65, 101, 102
	Occurs in the seeds of the loguat tree	No known bioactivity	242
	Fungi – *Phomopsis leptostromiformis*	Phomopsins A and B – mycotoxins	242
	Bacteria – *Escherichia coli*	Analog of proline with no reported bioactivity	242
	Isolated from *Streptomyces* sp. and *Nocardia formica*	Kikumycins A–B, noformicin, and anthelvencin A – peptide antibiotics	242
	Sponge – *Astrosclera willeyana*	Manzacidin D – α-adrenoreceptor blocker and serotonergic receptor antagonist	265
	Isolated from *Streptomyces griseoflavus* strain W 384	Hormaomycin – peptide with antimicrobial and antimalarial activity	65, 101, 102
	Sponge – *Astrosclera willeyana*	Manzacidins A–C – α-adrenoreceptor blockers and serotonergic receptor antagonists	264

groups. Sulfur-containing nonprotein amino acids are often derived from either cysteine or methionine and most likely undergo an enzymatic oxidation to give rise to the compounds listed below. Interestingly, nitrile moieties containing amino acids are also biosynthesized from cysteine. As an example, 3-cyanoalanine is synthesized from cysteine (**Scheme 16**), which is then converted to L-asparagine by β-cyanoalanine hydratase.[13] The reader is referred to **Table 8** for the biological origins of these amino acids and their respective bioactivities.



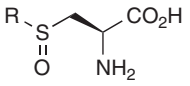
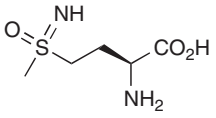
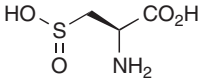
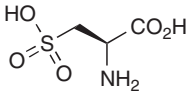
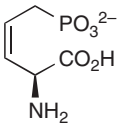
Scheme 16 Biosynthesis of L-asparagine through the intermediate L-3-cyanoalanine.[13]

Table 8 Nonprotein amino acids containing other functional groups as part of the side chain

Structure	*Biological origin*	*Biological properties*	*References*
	Plants – metabolism of methionine	Not available	12
	Isolated from *Streptomyces griseoflavus* strain W 384	Hormaomycin – peptide with antimicrobial and antimalarial activity	65, 101, 102
	Rice plant pathogenic fungi – *Ustilaginoidea virens*	Ustiloxin A – antimitotic peptide	124, 125
	Mushroom – *Clitocybe acromelalga*and seeds of *Vicia* sp.	β-cyanoalanine – neurotoxic amino acid	121, 289
	Seeds of *Lathyrus odoratus* and *Lathyrus pusillus*	Nonprotein amino acids that inhibit the cross-linking of peptides in collagen and elastin	121, 290
	Isolated from *Archidendron pauciflorum* (the djenkol bean)	Djenkolic acid – cysteine-derived amino acid with nephrotoxicity	121
	Plants – *Brassica oleracea* and other species of *Brassica*	*S*-Methylcysteine sulfoxide – precursor for hemolytic factor dimethyl disulfide in cattle, but nontoxic to nonruminant animals	121

(*Continued*)

Table 8 (Continued)

Structure	*Biological origin*	*Biological properties*	*References*
 R = ethyl, propyl, butyl, prop-1-en-1-yl and allyl	Garlic and onion – *Allium* sp.	Components that undergo decomposition to yield lachrymators	13
	Plants – *Connaraceae* sp., *Cnestis* sp., and *Rourea orientalis*	Methionine sulfoximine – nonprotein amino acid with neurotoxicity	121, 291
	Occurs in mammalian tissues	Cysteine sulfinic acid – intermediate in the biosynthesis of taurine (taurine is essential for many biological processes)	13
	Occurs in mammalian tissues	Cysteic acid – intermediate in the biosynthesis of taurine (taurine is essential for many biological processes)	13
	Bacteria – *Streptomyces plumbeus*	Constituent of antibiotics plumbemycins and rhizocticins. The amino acid is reversible inhibitor of PLP enzymes	78, 292

Abbreviations

Ac acetyl
Ala alanine
Arg arginine
Asn asparagine
Asp aspartic acid
BMAA β-*N*-methylamino-L-alanine
Cys cysteine
DAHP 7-deoxy-D-arabino-heptulosonate-7-phosphate
DAP diaminopimelate
EPSP 5-enolpyruvylshikimate-3-phosphate
$FADH_2$ flavin adenine dinucleotide, reduced
Gln glutamine
Glu glutamic acid
HmaS 4-hydroxymandelate synthase
Hmo 4-hydroxymandelate oxidase
HpgT 4-hydroxyphenylglycine transaminase
Ile isoleucine
L-DOPA 3,4-dihydroxy-L-phenylalanine
Leu leucine
Lys lysine
Me methyl

Met	methionine
NAD	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate, reduced
NRPS	nonribosomal peptide synthesis
ODAP	oxalyl-diaminopropionic acid
ORF	open reading frame
PEP	phosphoenolpyruvate
Phe	phenylalanine
PLP	pyridoxal-5′-phosphate
Pro	proline
RNA	ribonucleic acid
SAM	*S*-adenosyl-L-methionine
Thr	threonine
tRNA	transfer ribonucleic acid
Trp	tryptophan
Tyr	tyrosine
Val	valine

References

1. D. Voet; J. G. Voet; C. W. Pratt, Translation. In *Fundamentals of Biochemistry*; David Harris, Ed., John Wiley & Sons, Inc.: New York, 2002; pp 867–881.
2. D. Schwarzer; R. Finking; M. A. Marahiel, *Nat. Prod. Rep.* **2003**, *20*, 275–287.
3. R. Finking; M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.
4. E. A. Bell; A. A. Watson; R. J. Nash, *Nat. Prod. Commun.* **2008**, *3*, 93–110.
5. D. S. Bhakuni; D. S. Rawat, Bioactive Marine Peptides. In *Bioactive Marine Natural Products*; Bhakuni, D. S. and Rawat, D. S., Eds., Springer: New York, 2005; pp 278–328.
6. M. Namikoshi; K. L. Rinehart, *J. Ind. Microbiol. Biotechnol.* **1996**, *17*, 373–384.
7. D. S. Bhakuni; D. S. Rawat, Bioactive Metabolites of Marine Algae, Fungi and Bacteria. In *Bioactive Marine Natural Products*; David Harris, Ed., Springer: New York, 2005; pp 1–25.
8. S. B. Singh; D. L. Zink; J. D. Polishook; A. W. Dombrowski; S. J. DarkinRattray; D. M. Schmatz; M. A. Goetz, *Tetrahedron Lett.* **1996**, *37*, 8077–8080.
9. N. Lindquist; W. Fenical; G. D. Vanduyne; J. Clardy, *J. Am. Chem. Soc.* **1991**, *113*, 2303–2304.
10. J. Li; S. Jeong; L. Esser; P. G. Harran, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 4765–4770.
11. J. Li; A. W. G. Burgett; L. Esser; C. Amezcua; P. G. Harran, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 4770–4773.
12. S. Hunt, The Non-Protein Amino Acids. In *Chemistry and Biology of the Amino Acids*; Barrett, G.C., Ed.; Chapman and Hall Ltd.: New York, 1985; pp 55–138.
13. J. Velisek; R. Kubec; K. Cejpek, *Czech. J. Food Sci.* **2006**, *24*, 93–109.
14. S. Matsunaga; N. Fusetani, *Curr. Org. Chem.* **2003**, *7*, 945–966.
15. H. E. Umbarger, *Annu. Rev. Biochem.* **1978**, *47*, 533–606.
16. G. Riccardi; E. Derossi; A. Milano, *Plant Sci.* **1989**, *64*, 135–151.
17. H. Y. Xu; B. Andi; J. H. Qian; A. H. West; P. F. Cook, *Cell Biochem. Biophys.* **2006**, *46*, 43–64.
18. A. M. Velasco; J. I. Leguina; A. Lazcano, *J. Mol. Evol.* **2002**, *55*, 445–459.
19. M. Ibba; D. Soll, *Annu. Rev. Biochem.* **2000**, *69*, 617–650.
20. D. Voet; J. G. Voet, Amino Acid Biosynthesis. In *Biochemistry*, 2nd ed.; Bhakuni, D. S. and Rawat, D. S., Eds., John Wiley & Sons, Inc.: New York, 1995; pp 764–772.
21. P. M. Dewick, *Nat. Prod. Rep.* **1998**, *15*, 17–58.
22. A. R. Knaggs, *Nat. Prod. Rep.* **2003**, *20*, 119–136, and references therein.
23. S. Chaykin, *Annu. Rev. Biochem.* **1967**, *36*, 149–171.
24. E. S. Sattely; M. A. Fischbach; C. T. Walsh, *Nat. Prod. Rep.* **2008**, *25*, 757–793.
25. C. S. Neumann; C. T. Walsh, *J. Am. Chem. Soc.* **2008**, *130*, 14022–14023.
26. X. Fu; M. L. G. Ferreira; F. J. Schmitz; M. Kelly-Borges, *J. Nat. Prod.* **1998**, *61*, 1226–1231.
27. D. P. Galonic; F. H. Vaillancourt; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 3900–3901.
28. F. H. Vaillancourt; E. Yeh; D. A. Vosburg; S. E. O'Connor; C. T. Walsh, *Nature* **2005**, *436*, 1191–1194.
29. C. S. Neumann; D. G. Fujimori; C. T. Walsh, *Chem. Biol.* **2008**, *15*, 99–109.
30. Y. Nakao; M. Fujita; K. Warabi; S. Matsunaga; N. Fusetani, *J. Am. Chem. Soc.* **2000**, *122*, 10462–10463.
31. N. Fusetani; Y. Nakao; S. Matsunaga, *Tetrahedron Lett.* **1991**, *32*, 7073–7074.
32. D. P. Clark; J. Carroll; S. Naylor; P. Crews, *J. Org. Chem.* **1998**, *63*, 8757–8764.

33. J. Kobayashi; F. Itagaki; H. Shigemori; M. Ishibashi; K. Takahashi; M. Ogura; S. Nagasawa; T. Nakamura; H. Hirota; T. Ohta; S. Nozoe, *J. Am. Chem. Soc.* **1991**, *113*, 7812–7813.
34. M. Tsuda; H. Ishiyama; K. Masuko; T. Takao; Y. Shimonishi; J. Kobayashi, *Tetrahedron* **1999**, *55*, 12543–12548.
35. N. Fusetani; T. Sugawara; S. Matsunaga; H. Hirota, *J. Am. Chem. Soc.* **1991**, *113*, 7811–7812.
36. S. P. Gunasekera; S. A. Pomponi; P. J. McCarthy, *J. Nat. Prod.* **1994**, *57*, 79–83.
37. L. Chill; Y. Kashman; M. Schleyer, *Tetrahedron* **1997**, *53*, 16147–16152.
38. J. F. D. Daniel; E. Rodrigues, *Nat. Prod. Rep.* **2007**, *24*, 1128–1141.
39. G. Ryu; S. Matsunaga; N. Fusetani, *Tetrahedron* **1994**, *50*, 13409–13416.
40. N. Fusetani; S. Matsunaga, *Chem. Rev.* **1993**, *93*, 1793–1806.
41. H. Y. Li; S. Matsunaga; N. Fusetani, *J. Med. Chem.* **1995**, *38*, 338–343.
42. J. E. Coleman; E. D. Desilva; F. M. Kong; R. J. Andersen; T. M. Allen, *Tetrahedron* **1995**, *51*, 10653–10662.
43. N. K. Gulavita; S. P. Gunasekera; S. A. Pomponi; E. V. Robinson, *J. Org. Chem.* **1992**, *57*, 1767–1772.
44. M. A. Rashid; K. R. Gustafson; A. K. Cartner; N. Shigematsu; L. K. Pannell; M. R. Boyd, *J. Nat. Prod.* **2001**, *64*, 117–121.
45. S. Hlimi; S. Rebuffat; C. Goulard; S. Duchamp; B. Bodo, *J. Antibiot.* **1995**, *48*, 1254–1261.
46. H. Yano; S. Nakanishi; Y. Ikuina; K. Ando; M. Yoshida; Y. Saitoh; Y. Matsuda, *J. Antibiot.* **1997**, *50*, 992–997.
47. Y. Ikuina; C. Bando; M. Yoshida; H. Yano; Y. Saitoh, *J. Antibiot.* **1997**, *50*, 998–1006.
48. A. Fredenhagen; L. P. Molleyres; B. Bohlendorf; G. Laue, *J. Antibiot.* **2006**, *59*, 267–280.
49. K. Liu; R. L. White; J. Y. He; L. C. Vining, *J. Antibiot.* **1995**, *48*, 347–348.
50. A. D. Argoudel; R. R. Herr; D. J. Mason; T. R. Pyke; J. F. Zieser, *Biochemistry* **1967**, *6*, 165–170.
51. C. Charles; J. C. Braekman; D. Daloze; B. Tursch, *Tetrahedron* **1980**, *36*, 2133–2135.
52. B. L. Stapleton; G. M. Cameron; M. J. Garson, *Tetrahedron* **2001**, *57*, 4603–4607.
53. M. D. Unson; C. B. Rose; D. J. Faulkner; L. S. Brinen; J. R. Steiner; J. Clardy, *J. Org. Chem.* **1993**, *58*, 6336–6343.
54. L. Vroye; R. Beauwens; J. Van Sande; D. Daloze; J. C. Braekman; P. E. Golstein, *Pflugers Arch.* **1998**, *435*, 259–266.
55. A. E. Flowers; M. J. Garson; R. I. Webb; E. J. Dumdei; R. D. Charan, *Cell Tissue Res.* **1998**, *292*, 597–607.
56. F. Brackmann; A. de Meijere, *Chem. Rev.* **2007**, *107*, 4493–4537.
57. K. Ueda; J. Z. Xiao; N. Doke; S. Nakatsuka, *Tetrahedron Lett.* **1992**, *33*, 5377–5380.
58. R. E. Mitchell, *Phytochemistry* **1985**, *24*, 1485–1487.
59. A. Ichihara; K. Shiraishi; S. Sakamura; A. Furusaki; N. Hashiba; T. Matsumoto, *Tetrahedron Lett.* **1979**, 365–368.
60. R. E. Mitchell; S. A. Young; C. L. Bender, *Phytochemistry* **1994**, *35*, 343–348.
61. H. Kakeya; H. P. Zhang; K. Kobinata; R. Onose; C. Onozawa; T. Kudo; H. Osada, *J. Antibiot.* **1997**, *50*, 370–372.
62. H. P. Zhang; H. Kakeya; H. Osada, *Tetrahedron Lett.* **1997**, *38*, 1789–1792.
63. M. Hara; S. Soga; M. Itoh; K. Shono; J. Eishima; T. Mizukami, *J. Antibiot.* **2000**, *53*, 720–723.
64. T. Ohta; S. Nakajima; Z. Sato; T. Aoki; S. Hatanaka; S. Nozoe, *Chem. Lett.* **1986**, 511–512.
65. F. Brackmann; A. de Meijere, *Chem. Rev.* **2007**, *107*, 4538–4583.
66. H. Yoshimura; K. Takegami; M. Doe; T. Yamashita; K. Shibata; K. Wakabayashi; K. Soga; S. Kamisaka, *Phytochemistry* **1999**, *52*, 25–27.
67. W. S. Chilton; D. C. Drehmel, *Biochem. Syst. Ecol.* **2001**, *29*, 853–855.
68. D. C. Drehmel; W. S. Chilton, *J. Chem. Ecol.* **2002**, *28*, 333–341.
69. W. A. van der Donk, *J. Org. Chem.* **2006**, *71*, 9561–9571.
70. K. L. Rinehart; M. Namikoshi; B. W. Choi, *J. Appl. Phycol.* **1994**, *6*, 159–176.
71. M. S. Puar; T. M. Chan; V. Hegde; M. Patel; P. Bartner; K. J. Ng; B. N. Pramanik; R. D. MacFarlane, *J. Antibiot.* **1998**, *51*, 221–224.
72. M. C. Bagley; J. W. Dale; E. A. Merritt; A. Xiong, *Chem. Rev.* **2005**, *105*, 685–714.
73. M. Namikoshi; F. R. Sun; B. W. Choi; K. L. Rinehart; W. W. Carmichael; W. R. Evans; V. R. Beasley, *J. Org. Chem.* **1995**, *60*, 3671–3679.
74. S. Yoshizawa; R. Matsushima; M. F. Watanabe; K. Harada; A. Ichihara; W. W. Carmichael; H. Fujiki, *J. Cancer Res. Clin. Oncol.* **1990**, *116*, 609–614.
75. C. A. Bewley; H. Y. He; D. H. Williams; D. J. Faulkner, *J. Am. Chem. Soc.* **1996**, *118*, 4314–4321.
76. N. K. Gulavita; S. A. Pomponi; A. E. Wright; D. Yarwood; M. A. Sills, *Tetrahedron Lett.* **1994**, *35*, 6815–6818.
77. K. L. Rinehart; K. Harada; M. Namikoshi; C. Chen; C. A. Harvis; M. H. G. Munro; J. W. Blunt; P. E. Mulligan; V. R. Beasley; A. M. Dahlem; W. W. Carmichael, *J. Am. Chem. Soc.* **1988**, *110*, 8557–8558.
78. D. B. Berkowitz; B. D. Charette; K. R. Karukurichi; J. M. McFadden, *Tetrahedron: Asymmetry* **2006**, *17*, 869–882.
79. B. Levenber., *J. Biol. Chem.* **1968**, *243*, 6009–6013.
80. G. Dardenne; J. Casimir; J. Jadot, *Phytochemistry* **1968**, *7*, 1401–1405.
81. L. Fowden; A. Smith, *Phytochemistry* **1968**, *7*, 809–812.
82. E. Gellert; B. Halpern; R. Rudzats, *Phytochemistry* **1978**, *17*, 802–802.
83. M. K. Renner; Y. C. Shen; X. C. Cheng; P. R. Jensen; W. Frankmoelle; C. A. Kauffman; W. Fenical; E. Lobkovsky; J. Clardy, *J. Am. Chem. Soc.* **1999**, *121*, 11273–11276.
84. A. W. Schultz; D. C. Oh; J. R. Carney; R. T. Williamson; D. W. Udwary; P. R. Jensen; S. J. Gould; W. Fenical; B. S. Moore, *J. Am. Chem. Soc.* **2008**, *130*, 4507–4516.
85. T. Ohta; S. Nakajima; S. I. Hatanaka; M. Yamamoto; Y. Shimmen; S. Nishimura; Z. Yamaizumi; S. Nozoe, *Phytochemistry* **1987**, *26*, 565–566.
86. L. Fowden; H. M. Pratt, *Phytochemistry* **1973**, *12*, 1677–1681.
87. K. Melde; H. Buettner; W. Boschert; H. P. O. Wolf; S. Ghisla, *Biochem. J.* **1989**, *259*, 921–924.
88. D. Li; G. Agnihotri; S. Dakoji; E. Oh; M. Lantz; H. W. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 9034–9042.
89. H. Minakata; H. Komura; S. Y. Tamura; Y. Ohfune; K. Nakanishi; T. Kada, *Experientia* **1985**, *41*, 1622–1623.
90. S. A. Abdulganeeva; K. B. Erzhanov, *Usp. Khim.* **1991**, *60*, 1318–1342.
91. H. C. Potgieter; N. M. J. Vermeulen; D. J. J. Potgieter; H. F. Strauss, *Phytochemistry* **1977**, *16*, 1757–1759.
92. S. I. Hatanaka; K. Taniguch; Y. Niimura, *Phytochemistry* **1972**, *11*, 3327–3329.
93. Y. Niimura; S. I. Hatanaka, *Phytochemistry* **1974**, *13*, 175–178.

94. J. A. Bush; B. H. Long; J. J. Catino; W. T. Bradner, *J. Antibiot.* **1987**, *40*, 668–678.
95. S. Flecks; E. R. Patallo; X. F. Zhu; A. J. Ernyei; G. Seifert; A. Schneider; C. J. Dong; J. H. Naismith; K. H. van Pee, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 9533–9536.
96. B. K. Hubbard; M. G. Thomas; C. T. Walsh, *Chem. Biol.* **2000**, *7*, 931–942.
97. D. Schipper, *J. Antibiot.* **1983**, *36*, 1076–1077.
98. M. Kaneda, *J. Antibiot.* **1992**, *45*, 792–796.
99. C. S. V. Houge-Frydrych; M. L. Gilpin; P. W. Skett; J. W. Tyler, *J. Antibiot.* **2000**, *53*, 364–372.
100. A. L. Stefanska; R. Cassels; S. J. Ready; S. R. Warr, *J. Antibiot.* **2000**, *53*, 357–363.
101. N. Andres; H. Wolf; H. Zahner; E. Rossner; A. Zeeck; W. A. Konig; V. Sinnwell, *Helv. Chim. Acta* **1989**, *72*, 426–437.
102. K. Otoguro; H. Ui; A. Ishiyama; N. Arai; M. Kobayashi; Y. Takahashi; R. Masuma; K. Shiomi; H. Yamada; S. Omura, *J. Antibiot.* **2003**, *56*, 322–324.
103. M. Isaka; S. Palasarn; P. Rachtawee; S. Vimuttipong; P. Kongsaeree, *Org. Lett.* **2005**, *7*, 2257–2260.
104. G. W. Kirby; G. V. Rao; D. J. Robins, *J. Chem. Soc. Perkin Trans. 1* **1988**, 301–304.
105. K. Matsuzaki; H. Ikeda; T. Ogino; A. Matsumoto; H. B. Woodruff; H. Tanaka; S. Omura, *J. Antibiot.* **1994**, *47*, 1173–1174.
106. H. Gouda; K. Matsuzaki; H. Tanaka; S. Hirono; S. Omura; J. A. McCauley; P. A. Sprengeler; G. T. Furst; A. B. Smith, *J. Am. Chem. Soc.* **1996**, *118*, 13087–13088.
107. L. D. Boeck; F. P. Mertz, *J. Antibiot.* **1986**, *39*, 1533–1540.
108. M. J. Zmijewski; B. Briggs; R. Logan; L. D. Boeck, *Antimicrob. Agents Chemother.* **1987**, *31*, 1497–1501.
109. I. Kaneko; K. Kamoshida; S. Takahashi, *J. Antibiot.* **1989**, *42*, 236–241.
110. H. Seto; T. Fujioka; K. Furihata; I. Kaneko; S. Takahashi, *Tetrahedron Lett.* **1989**, *30*, 4987–4990.
111. O. Grach-Pogrebinsky; S. Carmeli, *Tetrahedron* **2008**, *64*, 10233–10238.
112. K. Harada; K. Fujii; T. Shimada; M. Suzuki; H. Sano; K. Adachi; W. W. Carmichael, *Tetrahedron Lett.* **1995**, *36*, 1511–1514.
113. T. Sano; T. Usui; K. Ueda; L. Osada; K. Kaya, *J. Nat. Prod.* **2001**, *64*, 1052–1055.
114. Y. Itou; S. Suzuki; K. Ishida; M. Murakami, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1243–1246.
115. K. C. Smith; R. L. White; Y. Z. Le; L. C. Vining, *J. Nat. Prod.* **1995**, *58*, 1274–1277.
116. J. A. Palermo; M. F. R. Brasco; E. Cabezas; V. Balzaretti; A. M. Seldes, *J. Nat. Prod.* **1998**, *61*, 488–490.
117. M. Igarashi; N. Kinoshita; T. Ikeda; M. Kameda; M. Hamada; T. Takeuchi, *J. Antibiot.* **1997**, *50*, 1020–1025.
118. M. Igarashi; H. Nakamura; H. Naganawa; T. Takeuchi, *J. Antibiot.* **1997**, *50*, 1026–1031.
119. M. Igarashi; M. Kuwada; M. Hamada; H. Inuma; H. Hayashi; K. Tsuchiya; M. Hori, *J. Antibiot.* **2001**, *54*, 1072–1079.
120. K. Ishida; H. Matsuda; M. Murakami; K. Yamaguchi, *J. Nat. Prod.* **1997**, *60*, 184–187.
121. E. A. Bell, *J. Agric. Food Chem.* **2003**, *51*, 2854–2865, and references therein.
122. P. O. Larsen; F. T. Sorensen; E. Wieczorkowska, *Phytochemistry* **1978**, *17*, 549–550.
123. R. Watson; L. Fowden, *Phytochemistry* **1973**, *12*, 617–622.
124. Y. Koiso; M. Natori; S. Iwasaki; S. Sato; R. Sonoda; Y. Fujita; H. Yaegashi; Z. Sato, *Tetrahedron Lett.* **1992**, *33*, 4157–4160.
125. Y. Koiso; Y. Li; S. Iwasaki; K. Hanaoka; T. Kobayashi; R. Sonoda; Y. Fujita; H. Yaegashi; Z. Sato, *J. Antibiot.* **1994**, *47*, 765–773.
126. Y. Koiso; N. Morisaki; Y. Yamashita; Y. Mitsui; R. Shirai; Y. Hashimoto; S. Iwasaki, *J. Antibiot.* **1998**, *51*, 418–422.
127. M. Chu; I. Truumees; R. Mierzwa; M. Patel; M. S. Puar, *J. Antibiot.* **1997**, *50*, 1061–1063.
128. S. Matsunaga; N. Fusetani, *J. Org. Chem.* **1995**, *60*, 1177–1181.
129. C. A. Bewley; D. J. Faulkner, *J. Org. Chem.* **1994**, *59*, 4849–4852.
130. E. W. Schmidt; C. A. Bewley; D. J. Faulkner, *J. Org. Chem.* **1998**, *63*, 1254–1258.
131. M. Debono; R. S. Gordee, *Annu. Rev. Microbiol.* **1994**, *48*, 471–497.
132. M. Tomishima; H. Ohki; A. Yamada; K. Maki; F. Ikeda, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1474–1477.
133. L. S. Bonnington; J. Tanaka; T. Higa; J. Kimura; Y. Yoshimura; Y. Nakao; W. Y. Yoshida; P. J. Scheuer, *J. Org. Chem.* **1997**, *62*, 7765–7767.
134. C. A. Bewley; C. Debitus; D. J. Faulkner, *J. Am. Chem. Soc.* **1994**, *116*, 7631–7636.
135. E. W. Schmidt; D. J. Faulkner, *Tetrahedron* **1998**, *54*, 3043–3056.
136. J. Y. Su; Y. L. Zhong; L. M. Zeng; H. M. Wu; X. Y. Shen; K. Ma, *J. Nat. Prod.* **1996**, *59*, 504–506.
137. R. Talpir; Y. Benayahu; Y. Kashman; L. Pannell; M. Schleyer, *Tetrahedron Lett.* **1994**, *35*, 4453–4456.
138. Y. Koguchi; J. Kohno; M. Nishio; K. Takahashi; T. Okuda; T. Ohnuki; S. Komatsubara, *J. Antibiot.* **2000**, *53*, 105–109.
139. J. Kohno; Y. Koguchi; M. Niskio; K. Nakao; M. Kuroda; R. Shimizu; T. Ohnuki; S. Komatsubara, *J. Org. Chem.* **2000**, *65*, 990–995.
140. P. Crews; J. J. Farias; R. Emrich; P. A. Keifer, *J. Org. Chem.* **1994**, *59*, 2932–2934.
141. D. Abbanat; M. Leighton; W. Maiese; E. B. G. Jones; C. Pearce; M. Greestein, *J. Antibiot.* **1998**, *51*, 296–302.
142. G. Schlingmann; L. Milne; D. R. Williams; G. T. Carter, *J. Antibiot.* **1998**, *51*, 303–316.
143. E. W. Schmidt; M. K. Harper; D. J. Faulkner, *J. Nat. Prod.* **1997**, *60*, 779–782.
144. D. E. Nettleton; T. W. Doyle; B. Krishnan; G. K. Matsumoto; J. Clardy, *Tetrahedron Lett.* **1985**, *26*, 4011–4014.
145. H. Uemoto; Y. Yahiro; H. Shigemori; M. Tsuda; T. Takao; Y. Shimonishi; J. Kobayashi, *Tetrahedron* **1998**, *54*, 6719–6724.
146. W. D. Clark; T. Corbett; F. Valeriote; P. Crews, *J. Am. Chem. Soc.* **1997**, *119*, 9285–9286.
147. J. Kobayashi; M. Sato; M. Ishibashi; H. Shigemori; T. Nakamura; Y. Ohizumi, *J. Chem. Soc. Perkin Trans. 1* **1991**, 2609–2611.
148. J. Kobayashi; M. Sato; T. Murayama; M. Ishibashi; M. R. Walchi; M. Kanai; J. Shoji; Y. Ohizumi, *J. Chem. Soc. Chem. Commun.* **1991**, 1050–1052.
149. F. Itagaki; H. Shigemori; M. Ishibashi; T. Nakamura; T. Sasaki; J. Kobayashi, *J. Org. Chem.* **1992**, *57*, 5540–5542.
150. J. Kobayashi; F. Itagaki; H. Shigemori; T. Takao; Y. Shimonishi, *Tetrahedron* **1995**, *51*, 2525–2532.
151. A. Qureshi; P. L. Colin; D. J. Faulkner, *Tetrahedron* **2000**, *56*, 3679–3685.
152. N. Nonaka; Y. Asai; M. Nishio; K. Takahashi; T. Okuda; S. Tanaka; T. Sugita; T. Ohnuki; S. Komatsubara, *J. Antibiot.* **1997**, *50*, 646–652.
153. Y. Asai; N. Nonaka; M. Nishio; K. Okamura; T. Date; T. Sugita; T. Ohnuki; S. Komatsubara, *J. Antibiot.* **1997**, *50*, 653–658.
154. M. Weber; K. Taraz; H. Budzikiewicz; V. Geoffroy; J. M. Meyer, *Biometals* **2000**, *13*, 301–309.
155. J. D. Faraldo-Gomez; M. S. P. Sansom, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 105–116.

156. C. B. Cui; H. Kakeya; G. Okada; R. Onose; M. Ubukata; I. Takahashi; K. Isono; H. Osada, *J. Antibiot.* **1995**, *48*, 1382–1384.
157. C. B. Cui; H. Kakeya; H. Osada, *Tetrahedron* **1996**, *52*, 12651–12666.
158. C. B. Cui; H. Kakeya; H. Osada, *Tetrahedron* **1997**, *53*, 59–72.
159. F. Z. Wang; Y. C. Fang; T. J. Zhu; M. Zhang; A. Q. Lin; Q. Q. Gu; W. M. Zhu, *Tetrahedron* **2008**, *64*, 7986–7991.
160. M. Kusano; G. Sotoma; H. Koshino; J. Uzawa; M. Chijimatsu; S. Fujioka; T. Kawano; Y. Kimura, *J. Chem. Soc. Perkin Trans.* **1998**, *1*, 2823–2826.
161. Y. Kimura; A. Sawada; M. Kuramata; M. Kusano; S. Fujioka; T. Kawano; A. Shimada, *J. Nat. Prod.* **2005**, *68*, 237–239.
162. K. Sprogoe; S. Manniche; T. O. Larsen; C. Christophersen, *Tetrahedron* **2005**, *61*, 8718–8721.
163. K. Ishida; H. Matsuda; M. Murakami; K. Yamaguchi, *J. Nat. Prod.* **1997**, *60*, 724–726.
164. K. Ishida; H. Matsuda; M. Murakami; K. Yamaguchi, *Tetrahedron* **1996**, *52*, 9025–9030.
165. K. S. Lam; G. A. Hesler; J. M. Mattei; S. W. Mamber; S. Forenza; K. Tomita, *J. Antibiot.* **1990**, *43*, 956–960.
166. J. E. Leet; D. R. Schroeder; J. Golik; J. A. Matson; T. W. Doyle; K. S. Lam; S. E. Hill; M. S. Lee; J. L. Whitney; B. S. Krishnan, *J. Antibiot.* **1996**, *49*, 299–311.
167. P. A. Cox; S. A. Banack; S. J. Murch; U. Rasmussen; G. Tien; R. R. Bidigare; J. S. Metcalf; L. F. Morrison; G. A. Codd; B. Bergman, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5074–5078.
168. M. Miethke; M. A. Marahiel, *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 413–451.
169. H. Drechsel; G. Jung, *J. Pept. Sci.* **1998**, *4*, 147–181.
170. N. Fusetani; S. Matsunaga; H. Matsumoto; Y. Takebayashi, *J. Am. Chem. Soc.* **1990**, *112*, 7053–7054.
171. Y. Nakao; S. Matsunaga; N. Fusetani, *Bioorg. Med. Chem.* **1995**, *3*, 1115–1122.
172. Y. Nakao; A. Masuda; S. Matsunaga; N. Fusetani, *J. Am. Chem. Soc.* **1999**, *121*, 2425–2431.
173. Y. Nakao; N. Oku; S. Matsunaga; N. Fusetani, *J. Nat. Prod.* **1998**, *61*, 667–670.
174. E. A. Bell; J. P. Odonovan, *Phytochemistry* **1966**, *5*, 1211–1219.
175. K. Fujie; Y. Shinguh; H. Hatanaka; N. Shigematsu; H. Murai; T. Fujita; M. Yamashita; M. Okamoto; M. Okuhara, *J. Antibiot.* **1993**, *46*, 908–913.
176. I. Nakanishi; T. Kinoshita; T. Tada; T. Fujita; H. Hatanaka; A. Sato, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2397–2402.
177. P. W. Ford; K. R. Gustafson; T. C. McKee; N. Shigematsu; L. K. Maurizi; L. K. Pannell; D. E. Williams; E. D. de Silva; P. Lassota; T. M. Allen; R. Van Soest; R. J. Andersen; M. R. Boyd, *J. Am. Chem. Soc.* **1999**, *121*, 5899–5909.
178. A. Vega; E. A. Bell, *Phytochemistry* **1967**, *6*, 759–762.
179. S. A. Banack; H. E. Johnson; R. Cheng; P. A. Cox, *Mar. Drugs* **2007**, *5*, 180–196.
180. A. Asai; A. Hasegawa; K. Ochiai; Y. Yamashita; T. Mizukami, *J. Antibiot.* **2000**, *53*, 81–83.
181. A. Asai; T. Tsujita; S. V. Sharma; Y. Yamashita; S. Akinaga; M. Funakoshi; H. Kobayashi; T. Mizukami, *Biochem. Pharmacol.* **2004**, *67*, 227–234.
182. T. Wakamiya; Y. Oda; H. Fujita; T. Shiba, *Tetrahedron Lett.* **1986**, *27*, 2143–2144.
183. R. J. P. Cannell, *Pestic. Sci.* **1993**, *39*, 147–153.
184. G. A. Rosenthal; D. L. Dahlman, *J. Biol. Chem.* **1990**, *265*, 868–873.
185. G. A. Rosenthal, *Q. Rev. Biol.* **1977**, *52*, 155–178.
186. M. P. Hegarty; A. W. Pound, *Aust. J. Biol. Sci.* **1970**, *23*, 831–842.
187. E. A. Bell, *FEBS Lett.* **1976**, *64*, 29–35.
188. G. J. Sharman; D. H. Williams; D. F. Ewing; C. Ratledge, *Chem. Biol.* **1995**, *2*, 553–561.
189. D. L. Pruess; J. P. Scannell; M. Kellett; H. A. Ax; J. Janecek; T. H. Williams; A. Stempel; J. Berger, *J. Antibiot.* **1974**, *27*, 229–233.
190. J. P. Scannell; D. L. Pruess; H. A. Ax; A. Jacoby; M. Kellett; A. Stempel, *J. Antibiot.* **1976**, *29*, 38–43.
191. I. G. Denisov; T. M. Makris; S. G. Sligar; I. Schlichting, *Chem. Rev.* **2005**, *105*, 2253–2278.
192. F. Vanmiddlesworth; R. A. Giacobbe; M. Lopez; G. Garrity; J. A. Bland; K. Bartizal; R. A. Fromtling; J. Polishook; M. Zweerink; A. M. Edison; W. Rozdilsky; K. E. Wilson; R. L. Monaghan, *J. Antibiot.* **1992**, *45*, 861–867.
193. R. H. Feling; G. O. Buchanan; T. J. Mincer; C. A. Kauffman; P. R. Jensen; W. Fenical, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 355–357.
194. L. L. Beer; B. S. Moore, *Org. Lett.* **2007**, *9*, 845–848.
195. D. O'Hagan, *J. Fluor. Chem.* **2006**, *127*, 1479–1483.
196. M. Sanada; T. Miyano; S. Iwadare; J. M. Williamson; B. H. Arison; J. L. Smith; A. W. Douglas; J. M. Liesch; E. Inamine, *J. Antibiot.* **1986**, *39*, 259–265.
197. L. Harrison; D. B. Teplow; M. Rinaldi; G. Strobel, *J. Gen. Microbiol.* **1991**, *137*, 2857–2865.
198. D. DiGiorgio; L. Camoni; C. Marchiafava; A. Ballio, *Phytochemistry* **1997**, *45*, 1385–1391.
199. G. L. Helms; R. E. Moore; W. P. Niemczura; G. M. L. Patterson; K. B. Tomer; M. L. Gross, *J. Org. Chem.* **1988**, *53*, 1298–1307.
200. J. Vetter, *Toxicon* **1998**, *36*, 13–24.
201. T. Agatsuma; Y. Sakai; T. Mizukami; Y. Saitoh, *J. Antibiot.* **1997**, *50*, 704–708.
202. L. A. McDonald; L. R. Barbieri; G. T. Carter; E. Lenoy; J. Lotvin; P. J. Petersen; M. M. Siegel; G. Singh; R. T. Williamson, *J. Am. Chem. Soc.* **2002**, *124*, 10260–10261.
203. Y. Masuoka; K. Shin-Ya; K. Furihata; Y. Hayakawa; H. Seto, *J. Antibiot.* **1997**, *50*, 1058–1060.
204. Y. Masuoka; K. Shin-Ya; K. Furihata; H. Matsumoto; Y. Takebayashi; K. Ngai; K. Suzuki; Y. Hayakawa; H. Seto, *J. Antibiot.* **2000**, *53*, 793–798.
205. Y. Masuoka; K. Shin-Ya; Y. B. Kim; M. Yoshida; K. Nagai; K. Suzuki; Y. Hayakawa; H. Seto, *J. Antibiot.* **2000**, *53*, 788–792.
206. F. Vanmiddlesworth; C. Dufresne; F. E. Wincott; R. T. Mosley; K. E. Wilson, *Tetrahedron Lett.* **1992**, *33*, 297–300.
207. W. S. Horn; J. L. Smith; G. F. Bills; S. L. Raghoobar; G. L. Helms; M. B. Kurtz; J. A. Marrinan; B. R. Frommer; R. A. Thornton; S. M. Mandala, *J. Antibiot.* **1992**, *45*, 1692–1696.
208. A. Randazzo; G. Bifulco; C. Giannini; M. Bucci; C. Debitus; G. Cirino; L. Gomez-Paloma, *J. Am. Chem. Soc.* **2001**, *123*, 10870–10876.
209. M. V. R. Reddy; M. K. Harper; D. J. Faulkner, *Tetrahedron* **1998**, *54*, 10649–10656.
210. T. Narender; A. Puri; Shweta; T. Khaliq; R. Saxena; G. Bhatia; R. Chandra, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 293–296.
211. L. Fowden, *Nature* **1966**, *209*, 807–808.

212. W. O. Godtfred; S. Vangedal; D. W. Thomas, *Tetrahedron* **1970**, *26*, 4931.
213. R. Louw; H. C. Potgieter; W. Vorster, *Amino Acids* **2005**, *29*, 207–212.
214. J. C. Sheehan; D. Mania; S. Nakamura; J. A. Stock; K. Maeda, *J. Am. Chem. Soc.* **1968**, *90*, 462–470.
215. R. Traber; H. R. Loosli; W. Pache; A. V. Wartburg, *Helv. Chim. Acta* **1977**, *60*, 1568–1578.
216. A. Afonso; F. Hon; R. Brambilla; M. S. Puar, *J. Antibiot.* **1999**, *52*, 398–406.
217. I. Wagner; H. Musso, *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 816–828.
218. J. M. Roosenberg; Y. M. Lin; Y. Lu; M. J. Miller, *Curr. Med. Chem.* **2000**, *7*, 159–197.
219. W. Gu; M. Cueto; P. R. Jensen; W. Fenical; R. B. Silverman, *Tetrahedron* **2007**, *63*, 6535–6541.
220. A. Closse; R. Huguenin, *Helv. Chim. Acta* **1974**, *57*, 533–545.
221. E. Bernardi; J. L. Fauchere; G. Atassi; P. Viallefont; R. Lazaro, *Amino Acids* **1994**, *6*, 315–318.
222. S. Gupta; G. Peiser; T. Nakajima; Y. S. Hwang, *Tetrahedron Lett.* **1994**, *35*, 6009–6012.
223. L. Vertesy; E. Ehlers; H. Kogler; M. Kurz; J. Meiwes; G. Seibert; M. Vogel; P. Hammann, *J. Antibiot.* **2000**, *53*, 816–827.
224. L. K. Meier; H. Sorensen, *Phytochemistry* **1979**, *18*, 1173–1175.
225. L. Fowden, *Annu. Rev. Biochem.* **1964**, *33*, 173–204.
226. P. O. Larsen; H. Sorensen; D. W. Cochran; E. W. Hagaman; E. Wenkert, *Phytochemistry* **1973**, *12*, 1713–1719.
227. E. A. Bell; K. Perera; P. B. Nunn; M. S. J. Simmonds; W. M. Blaney, *Phytochemistry* **1996**, *43*, 1003–1007.
228. A. N. Starratt; S. Caveney, *Phytochemistry* **1995**, *40*, 479–481.
229. S. Caveney; A. Starratt, *Nature* **1994**, *372*, 509.
230. A. F. Parsons, *Tetrahedron* **1996**, *52*, 4149–4174, and references therein.
231. Y. G. Zhang; S. C. Liu; Y. S. Che; X. Z. Liu, *J. Nat. Prod.* **2007**, *70*, 1522–1525.
232. R. X. Tan; P. R. Jensen; P. G. Williams; W. Fenical, *J. Nat. Prod.* **2004**, *67*, 1374–1382.
233. K. Kanoh; S. Kohno; T. Asari; T. Harada; J. Katada; M. Muramatsu; H. Kawashima; H. Sekiya; I. Uno, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2847–2852.
234. K. Kanoh; S. Kohno; J. Katada; J. Takahashi; I. Uno, *J. Antibiot.* **1999**, *52*, 134–141.
235. T. Morino; M. Nishimoto; A. Masuda; S. Fujita; T. Nishikiori; S. Saito, *J. Antibiot.* **1995**, *48*, 1509–1510.
236. K. Isono, *Pharmacol. Ther.* **1991**, *52*, 269–286.
237. E. Rubenstein; H. L. Zhou; K. M. Krasinska; A. Chien; C. H. Becker, *Phytochemistry* **2006**, *67*, 898–903.
238. A. Takahashi; D. Ikeda; H. Nakamura; H. Naganawa; S. Kurasawa; Y. Okami; T. Takeuchi; Y. Iitaka, *J. Antibiot.* **1989**, *42*, 1562–1566.
239. S. Omura; K. Matsuzaki; T. Fujimoto; K. Kosuge; T. Furuya; S. Fujita; A. Nakagawa, *J. Antibiot.* **1991**, *44*, 117–118.
240. S. I. Hatanaka; S. Kaneko, *Phytochemistry* **1977**, *16*, 1041–1042.
241. L. T. Tan; X. C. Cheng; P. R. Jensen; W. Fenical, *J. Org. Chem.* **2003**, *68*, 8767–8773.
242. A. B. Mauger, *J. Nat. Prod.* **1996**, *59*, 1205–1211.
243. J. Spizek; T. Rezanka, *Appl. Microbiol. Biotechnol.* **2004**, *64*, 455–464.
244. K. Umezawa; Y. Ikeda; Y. Uchihata; H. Naganawa; S. Kondo, *J. Org. Chem.* **2000**, *65*, 459–463.
245. T. A. Smitka; J. B. Deeter; A. H. Hunt; F. P. Mertz; R. M. Ellis; L. D. Boeck; R. C. Yao, *J. Antibiot.* **1988**, *41*, 726–733.
246. W. P. Frankmolle; L. K. Larsen; F. R. Caplan; G. M. L. Patterson; G. Knubel; I. A. Levine; R. E. Moore, *J. Antibiot.* **1992**, *45*, 1451–1457.
247. W. P. Frankmolle; G. Knubel; R. E. Moore; G. M. L. Patterson, *J. Antibiot.* **1992**, *45*, 1458–1466.
248. I. Bonnard; M. Rolland; J. M. Salmon; E. Debiton; C. Barthomeuf; B. Banaigs, *J. Med. Chem.* **2007**, *50*, 1266–1279.
249. N. Berer; A. Rudi; I. Goldberg; Y. Benayahu; Y. Kashman, *Org. Lett.* **2004**, *6*, 2543–2545.
250. K. Umezawa; K. Nakazawa; T. Uemura; Y. Ikeda; S. Kondo; H. Naganawa; N. Kinoshita; H. Hashizume; M. Hamada; T. Takeuchi; S. Ohba, *Tetrahedron Lett.* **1998**, *39*, 1389–1392.
251. K. Umezawa; K. Nakazawa; Y. Ikeda; H. Naganawa; S. Kondo, *J. Org. Chem.* **1999**, *64*, 3034–3038.
252. K. Umezawa; Y. Ikeda; O. Kawase; H. Naganawa; S. Kondo, *J. Chem. Soc. Perkin Trans. 1* **2001**, 1550–1553.
253. M. V. Dauria; A. Zampella; L. G. Paloma; L. Minale; C. Debitus; C. Roussakis; V. LeBert, *Tetrahedron* **1996**, *52*, 9589–9596.
254. H. Morita; S. Nagashima; Y. Uchiumi; O. Kuroki; K. Takeya; H. Itokawa, *Chem. Pharm. Bull.* **1996**, *44*, 1026–1032.
255. H. Watase; Y. Tomiie; I. Nitta, *Bull. Chem. Soc. Jpn.* **1958**, *31*, 714–725.
256. Y. Ohfune; M. Tomita, *J. Am. Chem. Soc.* **1982**, *104*, 3511–3513.
257. M. Maeda; T. Kodama; T. Tanaka; H. Yoshizumi; T. Takemoto; K. Nomoto; T. Fujita, *Chem. Pharm. Bull.* **1986**, *34*, 4892–4895.
258. P. T. Holland; A. I. Selwood; D. O. Mountfort; A. L. Wilkins; P. McNabb; L. L. Rhodes; G. J. Doucette; C. M. Mikulski; K. L. King, *Chem. Res. Toxicol.* **2005**, *18*, 814–816.
259. R. Sakai; C. Oiwa; K. Takaishi; H. Kamiya; M. Tagawa, *Tetrahedron Lett.* **1999**, *40*, 6941–6944.
260. K. ShinYa; J. S. Kim; K. Furihata; Y. Hayakawa; H. Seto, *Tetrahedron Lett.* **1997**, *38*, 7079–7082.
261. K. Takahashi; M. Kawabata; D. Uemura; S. Iwadare; R. Mitomo; F. Nakano; A. Matsuzaki, *Tetrahedron Lett.* **1985**, *26*, 1077–1078.
262. C. Martin; L. Oberer; T. Ino; W. A. Konig; M. Busch; J. Weckesser, *J. Antibiot.* **1993**, *46*, 1550–1556.
263. B. Bister; S. Keller; H. I. Baumann; G. Nicholson; S. Weist; G. Jung; R. D. Sussmuth; F. Juttner, *J. Nat. Prod.* **2004**, *67*, 1755–1757.
264. J. Kobayashi; F. Kanda; M. Ishibashi; H. Shigemori, *J. Org. Chem.* **1991**, *56*, 4574–4576.
265. T. Jahn; G. M. Konig; A. D. Wright; G. Worheide; J. Reitner, *Tetrahedron Lett.* **1997**, *38*, 3883–3884.
266. K. Ersmark; J. R. Del Valle; S. Hanessian, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 1202–1223.
267. M. Murakami; Y. Okita; H. Matsuda; T. Okino; K. Yamaguchi, *Tetrahedron Lett.* **1994**, *35*, 3129–3132.
268. A. R. Carroll; G. K. Pierens; G. Fechner; P. de Almeida Leone; A. Ngo; M. Simpson; E. Hyde; J. N. A. Hooper; S. L. Bostrom; D. Musil; R. J. Quinn, *J. Am. Chem. Soc.* **2002**, *124*, 13340–13341.
269. K. Ishida; Y. Okita; H. Matsuda; T. Okino; M. Murakami, *Tetrahedron* **1999**, *55*, 10971–10988.
270. N. Valls; M. Vallribera; S. Carmeli; J. Bonjoch, *Org. Lett.* **2003**, *5*, 447–450.
271. A. Ploutno; M. Shoshan; S. Carmeli, *J. Nat. Prod.* **2002**, *65*, 973–978.
272. A. R. Carroll; M. S. Buchanan; A. Edser; E. Hyde; M. Simpson; R. J. Quinn, *J. Nat. Prod.* **2004**, *67*, 1291–1294.

273. M. Murakami; K. Ishida; T. Okino; Y. Okita; H. Matsuda; K. Yamaguchi, *Tetrahedron Lett.* **1995**, *36*, 2785–2788.
274. H. J. Shin; H. Matsuda; M. Murakami; K. Yamaguchi, *J. Org. Chem.* **1997**, *62*, 1810–1813.
275. A. L. Stefanska; N. J. Coates; L. M. Mensah; A. J. Pope; S. J. Ready; S. R. Warr, *J. Antibiot.* **2000**, *53*, 345–350.
276. C. S. V. Houge-Frydrych; S. A. Readshaw; D. J. Bell, *J. Antibiot.* **2000**, *53*, 351–356.
277. P. Kleinwachter; H. M. Dahse; U. Luhmann; B. Schlegel; K. Dornberger, *J. Antibiot.* **2001**, *54*, 521–525.
278. K. Sugawara; F. Sugawara; G. A. Strobel; Y. Fu; C. H. He; J. Clardy, *J. Org. Chem.* **1985**, *50*, 5631–5633.
279. H. G. Cutler; K. Hoogsteen; R. H. Littrell; B. H. Arison, *Agr. Biol. Chem.* **1991**, *55*, 2037–2042.
280. D. E. Williams; K. Bombuwala; E. Lobkovsky; E. D. de Silva; V. Karunaratne; T. M. Allen; J. Clardy; R. J. Andersen, *Tetrahedron Lett.* **1998**, *39*, 9579–9582.
281. R. Sakai; H. Kamiya; M. Murata; K. Shimamoto, *J. Am. Chem. Soc.* **1997**, *119*, 4112–4116.
282. R. Sakai; K. Yoshida; A. Kimura; K. Koike; M. Jimbo; K. Koike; A. Kobiyama; H. Kamiya, *ChemBioChem* **2008**, *9*, 543–551.
283. M. Maeda; T. Kodama; T. Tanaka; H. Yoshizumi; T. Takemoto; K. Nomoto; T. Fujita, *Tetrahedron Lett.* **1987**, *28*, 633–636.
284. S. J. Duncan; S. Gruschow; D. H. Williams; C. McNicholas; R. Purewal; M. Hajek; M. Gerlitz; S. Martin; S. K. Wrigley; M. Moore, *J. Am. Chem. Soc.* **2001**, *123*, 554–560.
285. S. J. Duncan; S. Gruschow; D. H. Williams; C. McNicholas; R. Purewal; M. Hajek; M. Gerlitz; S. Martin; S. K. Wrigley; M. Moore, *J. Am. Chem. Soc.* **2001**, *123*, 554–560.
286. S. J. Duncan; D. H. Williams; M. Ainsworth; S. Martin; R. Ford; S. K. Wrigley, *Tetrahedron Lett.* **2002**, *43*, 1075–1078.
287. M. Shimura; M. Iwata; S. Omoto; Y. Sekizawa, *Agr. Biol. Chem.* **1979**, *43*, 2279–2281.
288. G. C. Kite; H. Ireland, *Phytochemistry* **2002**, *59*, 163–168.
289. S. Fushiya; S. Sato; G. Kusano; S. Nozoe, *Phytochemistry* **1993**, *33*, 53–55.
290. E. D. Schilling; F. M. Strong, *J. Am. Chem. Soc.* **1954**, *76*, 2848.
291. I. Murakoshi; T. Sekine; K. Maeshima; F. Ikegami; K. Yoshinaga; Y. Fujii; S. Okonogi, *Chem. Pharm. Bull.* **1993**, *41*, 388–390.
292. N. G. Ternan; J. W. Mc Grath; G. Mc Mullan; J. P. Quinn, *World J. Microbiol. Biotechnol.* **1998**, *14*, 635–647.

Biographical Sketches

Sabesan Yoganathan was born in Katchilaimadu, Sri Lanka, in 1981. He graduated with a B.Sc. degree in chemistry from McMaster University, Canada, in 2005. During the undergraduate years, he worked in the laboratory of Dr. John F. Valliant and his research involved the preparation of fluorine-18 and technetium-99m radiolabeled biomolecules. Currently, he is pursuing his Ph.D. degree in organic chemistry at the University of Alberta, Canada, under the supervision of Professor John C. Vederas. His research focuses on the chemical modifications of lantibiotics, which is a subclass of antimicrobial peptides with potent bioactivity.

John C. Vederas is University Professor of Chemistry and holds a Canada Research Chair in Bioorganic and Medicinal Chemistry. He obtained a B.Sc. from Stanford University and a Ph.D. from the Massachusetts Institute of Technology. He did postdoctoral work at the University of Basel (Switzerland) and at Purdue University prior to joining the University of Alberta in 1977. His research centers on antimicrobial peptides and fungal polyketide biosynthesis. He has received recognition for research and teaching from the University of Alberta, including the Faculty of Science Award for Teaching Excellence (1993), the Rutherford Award for Excellence in Undergraduate Teaching (1995), the University Cup for Research and Teaching (1998), the J. Gordin Kaplan Award for Excellence in Research (2003), and the Killam Award for Excellence in Mentoring (2003). He is a Fellow of the Royal Society (London) (2009), the Royal Society of Canada (1997) and an Alberta Centennial Medal recipient (2006). He received the Merck Sharp Dohme Award (1986), the John Labatt Award (1991), the R. U. Lemieux Award (2002), and the Alfred Bader Award (2005) from the Canadian Society for Chemistry for his research. He received the 2008 Chemical Institute of Canada (CIC) Medal. He has served in numerous scientific organizations; he was president of the Canadian Society for Chemistry (2002–03), and was a Member of Council of the Natural Sciences and Engineering Research Council of Canada (NSERC) (2001–04). He was the Chair of the Canadian Society for Chemistry Annual Conference in 2008.

5.03 Novel Enzymes for Biotransformation and Resolution of Alpha-Amino Acids

Paul C. Engel and Francesca Paradisi, University College Dublin, Dublin, Ireland

© 2010 Elsevier Ltd. All rights reserved.

5.03.1 Introduction

The growing demand for both L- and D-amino acids in an enantiomerically pure state as building blocks for pharmaceutical synthesis has stimulated sustained interest and effort in both chiral synthesis and racemic resolution. Although chemical approaches to the former and chromatographic approaches to the latter are constantly improving, the potential for biocatalytic solutions to both challenges is attracting ever more attention. Enzymes' green credentials provide a compelling argument in their favor in the current climate of public and political opinion, while many of the traditional objections to their use have been swept away through advances in molecular genetics, microbiology, and biotechnology (Chapter 9.19). Pure and reasonably robust biocatalysts can now be produced cheaply in large amounts by inducing 'overproduction' from a cloned gene. Alternatively, preparations of whole bacterial cells containing the relevant biocatalyst(s) may often also be used. The most obvious remaining limitation to the use of biocatalysis is that the high degree of specificity that produces the desirable and often apparently complete enantioselectivity also restricts the range of accepted substrates. This limitation is being overcome on the one hand by exhaustive searching of what the biosphere has to offer (substrate restriction in an enzyme from one biological species need not imply the same or even similar restriction in another species), and on the other by using the tools of both knowledge-based protein engineering (site-directed mutagenesis) and random mutagenesis (directed evolution) to develop new specificities, thus extending the versatility of biocatalysis.

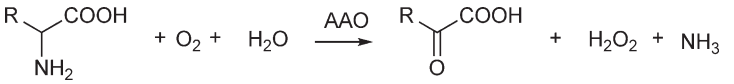
Enzymes may be used either directly for chiral synthesis of the desired enantiomer of the amino acid itself or of a derivative from which it can readily be prepared, or for kinetic resolution. Resolution of a racemate may remove the unwanted enantiomer, leaving the intended product untouched, or else the reaction may release the desired enantiomer from a racemic precursor. In either case the apparent disadvantage is that the process on its own can only yield up to 50% of the target compound. However, in a number of processes the enzyme-catalyzed kinetic resolution is combined with a second process that re-racemizes the unwanted enantiomer. This may be chemical or enzymatic, and in the latter case, the combination of two simultaneous enzymatic reactions can produce a smooth dynamic kinetic resolution leading to 100% yield.

A remarkably wide range of different enzymes has been deployed in various processes for the production of pure L- or D-enantiomers of amino acids, and in a number of these processes several enzymes are used simultaneously. In the review that follows, processes are grouped as far as possible according to the enzyme principally responsible for chiral selection in the overall process.

5.03.2 Amino Acid Oxidases

5.03.2.1 The Reaction and the Enzymes

The flavoprotein amino acid oxidases (AAOs) catalyze an essentially irreversible oxidative deamination of an amino acid. Molecular oxygen is the oxidant and the products are ammonia, the oxoacid, and H_2O_2 (Equation (1)).



(1)

Unless the last-mentioned product is removed by the inclusion of catalase, the oxoacid is liable to react further, undergoing oxidative decarboxylation to the carboxylic acid. An attractive feature of this group of enzymes in the present context is that there exist readily available representatives of both enantiospecificities. The well-studied and commercially available AAOs from vertebrate sources, such as L-AAO from snake venom and D-AAO from pig kidney, are expensive, however, and are increasingly being replaced by enzymes from microbial sources.

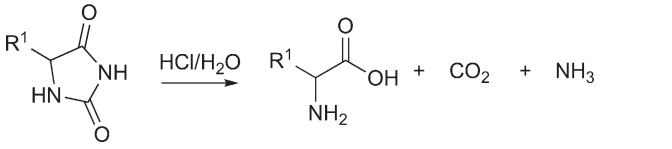
5.03.2.2 D-Amino Acid Oxidases

5.03.2.2.1 Conversion of racemic amino acids into the D-enantiomer assisted by a second enzyme

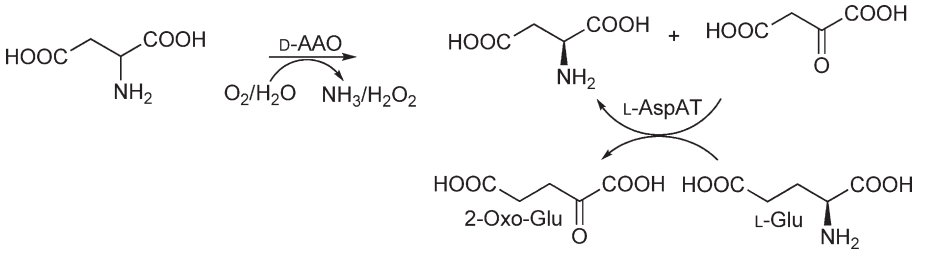
Numerous groups have described the use of D-AAO in a variety of combinations with other agents and in a variety of processes. Cheng and Wu,[1] in the context of a biotransformation of DL-aspartic acid to L-alanine, in which the key reaction works only with L-Asp, removed residual D-Asp with D-AAO and converted the resulting oxaloacetate into L-Asp by including L-aspartate aminotransferase (L-AspAT) in the reaction mixture (**Scheme 1**).

In a similar exercise with D-methionine, Findrik and Vasić-Rački[2] used the D-AAO of *Arthrobacter*, and for the second-step conversion of oxoacid into L-amino acid, used L-phenylalanine dehydrogenase (L-PheDH), which has a sufficiently broad specificity to accept L-methionine and its corresponding oxoacid as substrates. Efficient quantitative conversion in this latter reaction requires recycling of the cofactor NAD^+ into NADH, and for this the commercially available formate dehydrogenase (FDH) was used (**Scheme 2**).

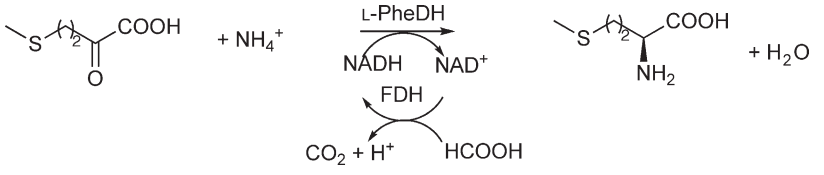
A conceptually similar approach applied in an industrial process is described by the Bristol–Myers–Squibb group, who required L-6-OH norleucine as an intermediate in the synthesis of their drug Omapatrilat. To avoid a lengthy chemical synthesis of the oxoacid, it was more convenient to start with the racemic amino acid, readily prepared by hydrolysis of the corresponding hydantoin (Equation (2)), and remove the D-isomer by oxidation using D-AAO.



(2)



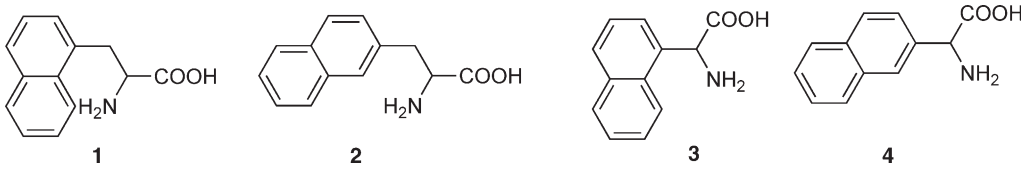
Scheme 1



Scheme 2

Hanson *et al.*[3] (also reviewed by Patel[4]) describe the use of D-AAO (either porcine kidney or *Trigonopsis variabilis*) to convert the D-enantiomer into the oxoacid. Here again, conversion into the L-amino acid was accomplished with an amino acid dehydrogenase (aaDH), in this case commercial bovine glutamate dehydrogenase (GDH). Although this enzyme is relatively specific for its dicarboxylic natural substrate, at the alkaline pH used by these workers, it becomes more inclined to the uncharged aliphatic substrates.[5] To recycle the coenzyme in this study both glucose dehydrogenase (GluDH) and FDH were found satisfactory.

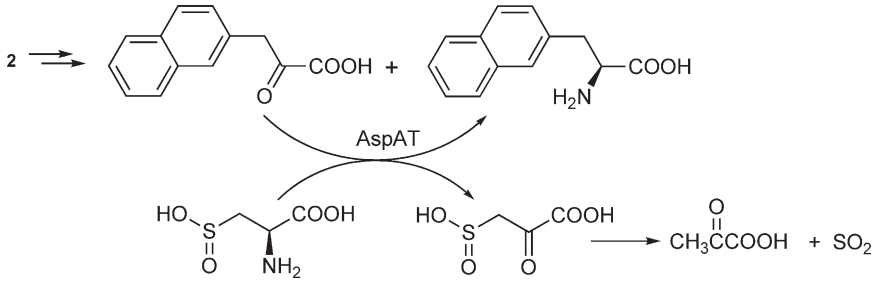
The use of D-AAO from the yeast *Rhodotorula gracilis* to deracemize naphthyl amino acids has been studied in some detail by the groups of Servi and Pollegioni,[6] who compared the kinetic properties of the enzyme with racemic 1- and 2-naphthylalanine (**1** and **2**) and 1- and 2-naphthylglycine (**3** and **4**).



1 **2** **3** **4**

D-2-Naphthylalanine was by far the best substrate of those tested, with a V_{max} comparable to that for D-alanine and a K_m over 20 times lower. D-1-Naphthylalanine was a reasonably good substrate, with V_{max} about 15-fold lower than that for the 2-substituted substrate but K_m equally low. The glycine-based analogues were much poorer substrates, mainly because of a much lower V_{max} in the case of the 2-naphthyl derivative and because of an elevated K_m in the case of 1-naphthylglycine. This is a particularly interesting study as it is one of the relatively rare cases where the investigators have gone beyond what is available in nature and used intelligent design to produce an improved catalyst by site-directed mutagenesis. Molecular modeling of substrate docking revealed a steric clash in the case of the naphthylglycine substrates, and replacement of a protein methionine residue by glycine (M213G) in a single mutation produced over 10-fold improvement in the handling of D-1-naphthylglycine, mainly through an improved K_m. With the 2-naphthyl derivative, however, a fivefold improvement in V_{max} was counterbalanced by a threefold deterioration in K_m. This interesting and detailed study underlines the value of thorough kinetic investigation. Clearly in the context of an industrial process, where substrates are typically available and supplied at high concentration, an improvement in V_{max} may be more important than changes in K_m.

The same group have used the enzyme combination employed in the aspartate deracemization cited above to deracemize 2-naphthylalanine,[7] but have made use of an interesting innovation introduced by Hélaine *et al.*[8] to pull over the poised equilibrium of the transamination reaction. Cysteine sulphinic acid was used as the amino donor in the transamination. The oxoacid product spontaneously decomposes in to pyruvic acid and SO_2 (**Scheme 3**).



Scheme 3

5.03.2.2.2 Conversion of racemic amino acids into the D-enantiomer assisted by nonenzymatic reduction

Two groups have recently combined the enantioselectivity of D-AAO with a nonselective, nonenzymatic reduction process in order to restore a racemic mixture, allowing a further round of selective oxidation and ultimately complete conversion of the racemate into the L-form. The Ingenza group at Edinburgh[9] describe cycles of chemical reduction employed in the production of a range of L-amino acids. Märkle and Lütz[10] have instead used electrochemical reduction, combining the D-AAO reaction with cyclic voltammetry to produce L-leucine starting from the oxoacid, the racemic amino acid, or from D-leucine.

5.03.2.3 L-Amino Acid Oxidases

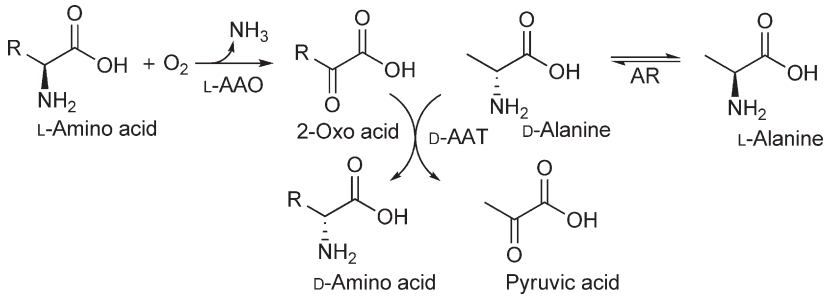
Turning to L-AAO, Pantaleone's industrial research group have reported[11] on the properties and use of an L-AAO from *Proteus myxofaciens*, overexpressed in *Escherichia coli*. This L-AAO, unusually, appears not to produce H_2O_2 in the catalytic reaction, thus making the addition of catalase unnecessary. The enzyme has a broad specificity, with a preference for nonpolar amino acids. This L-AAO was used in conjunction with a D-amino acid transaminase (D-AAT) and an alanine racemase (AR) to allow an efficient conversion of L-amino acid in to D-amino acid (**Scheme 4**).

L-AAO activity has also been used to convert a readily available L-amino acid, L-DOPA, to the corresponding oxoacid as the desired product.[12] The *Rhodococcus* L-AAO reported by Geueke and Hummel[13,14] was compared with the snake-venom enzyme (*Crotalus adamanteus*) in this application by the Zagreb group of Vasić-Rački. The cheaper bacterial enzyme, heterologously expressed in *Streptomyces*, showed better activity than the more expensive commercial snake enzyme. However, unlike the situation reported above with the *Proteus* L-AAO, in this case interference from H_2O_2 resulted in further conversion of a substantial portion of the 3,4-dihydroxyphenylpyruvic acid product into 3,4-dihydroxyphenylacetic acid unless a high level of catalase was included in the enzymatic reaction mixture.

5.03.2.4 Cyclic Amino Acids

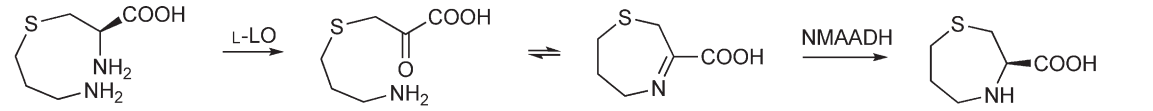
In an original application, Yasuda *et al.*[15] have used both L-AAO and D-AAO, and L-lysine oxidase to oxidize α,ω-diamino acids. The reactions produce the expected α-keto ω-amino acid products, but these then spontaneously cyclize to form cyclic α-imino acids. These compounds are then substrates for the authors' recently discovered[16] *N*-methyl amino acid dehydrogenase (NMAADH) from *Pseudomonas putida*, producing the pure L-cyclic amino acid (**Scheme 5**).

Using this procedure the authors have converted aminopropylcysteine to a new seven-membered ring cyclic amino acid [1,4]-thiazepane-3-carboxylic acid (**Scheme 6**).



Scheme 4

Scheme 5



(*S*)-*S*-Aminopropylcysteine

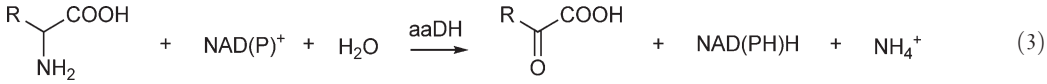
[1,4]-Thiazepane-3-carboxylic acid

Scheme 6

5.03.3 Amino Acid Dehydrogenases

5.03.3.1 The Reaction and the Enzymes

Similar to the AAOs, the aaDHs catalyze oxidative deamination, forming an oxoacid and ammonia. However, rather than using enzyme-bound FAD as the oxidant, followed by O_2, these enzymes employ nicotinamide cofactors, NAD^+ or $NADP^+$, in free solution (Equation (3)).



(3)

This use of a weaker oxidant has several consequences. First, the reaction is readily reversible. Indeed, at neutral pH and with average substrate concentrations, the equilibrium tends to lie toward amino acid formation.[17,18] Second, since the oxidant is not an ubiquitous oxygen, with a discardable product, but costly $NAD(P)^+$, forming NADPH, it becomes essential in any production process to find a way to reclaim or recycle the cofactor. Third, the absence of H_2O_2 among the products largely removes the concern about further reaction of the oxoacid through oxidative decarboxylation.

Unlike the AAOs, whose very name implies a breadth of substrate specificity, the aaDHs tend to be rather more specific and are each named after their optimal amino acid substrate. In particular, the archetype in this enzyme family, GDH,[19] shows poor activity with most substrates other than glutamate. However, the availability of X-ray crystallographic structures for several aaDHs[20–23] has opened the door to site-directed mutagenesis and altered substrate specificities.[24–27]

For a considerable period, the only aaDHs reported were the widespread GDHs and also alanine dehydrogenases, frequently found in *Bacilli* and, interestingly, entirely unrelated structurally to the GDHs. However, studies during the 1980s uncovered a number of other aaDHs in various bacterial species. These include leucine dehydrogenase, phenylalanine dehydrogenase (PheDH), and valine dehydrogenase, and the first two have found considerable use in amino acid production processes.

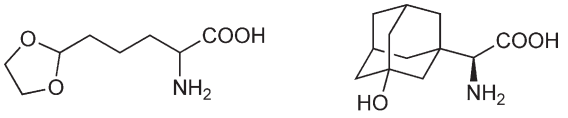
5.03.3.2 L-Phenylalanine Dehydrogenase

5.03.3.2.1 L-Phenylalanine dehydrogenase in chiral L-amino acid synthesis

Two groups in particular have pioneered the practical application of PheDH from various bacterial species, namely those of Asano *et al.* in Japan and Hummel *et al.* in Germany. Asano and his colleagues initially explored the application of PheDH to the chiral synthesis of the physiological substrate L-phenylalanine.[28] The PheDH of *Bacillus sphaericus* was overexpressed in *E. coli*. The issue of cofactor recycling was tackled by using the FDH of *Candida boidinii*. Importantly, these authors showed that both catalytic activities could be successfully

supplied by using whole cells, dispensing with the need to purify the enzymes. The same group used the PheDH of *Sporosarcina ureae*, in this case purified and crystallized, to explore substrate specificity and show that this enzyme could be used for the synthesis of a range of nonpolar natural biological amino acids.[18] Subsequently, reverting to the *B. sphaericus* enzyme,[29] they also showed that the PheDH/FDH combination could also be applied to the synthesis of a wide range of nonnatural L-amino acids and explored the long-term stability of the biocatalyst combination used repetitively in daily application (in a dialysis sac) over a period of a month with relatively little obvious decline in conversion efficiency.

The broad specificity of PheDH has been applied in an industrial context for the production of one of the amino acid building blocks in the synthesis of the antihypertensive drug Omapatrilat. The intermediate allysine ethylene acetal, with a dioxolane ring in its side chain (**5**), is produced in the pure L-form by using PheDH of *Thermoactinomyces intermedius*.[4] In an impressive application,[30] the same group, Hanson *et al.* at Bristol–Myers–Squibb, also applied this PheDH to make a more bulky amino acid, L-3-hydroxyadamantyl glycine (**6**), needed in the synthesis of an antidiabetic drug, Saxagliptin. In this case several aaDHs were screened to find the best activity. The *T. intermedius* enzyme was cloned into the methylotrophic yeast expression host *Pichia pastoris* and a fortuitous alteration of the DNA sequence in the cloned gene, resulting in a substantial alteration at the C-terminus of the enzyme, considerably enhanced the desired activity.



5 **6**

These authors also report a comparison of the efficacy of wet cells compared with heat-dried cells and with *Pichia* extracts. This appears to show the best result with the extracts, but it is difficult to assess the meaning of this study. The outcome is reported in terms of percentage yield of the final product, and it is impossible to judge whether what is being observed reflects different amounts of activity and incomplete reactions, varying stability of the biocatalyst in its various states, interfering side reactions due to other cellular activities, and so on. It is not clear to what extent reaction under each set of conditions was optimized or maximized.

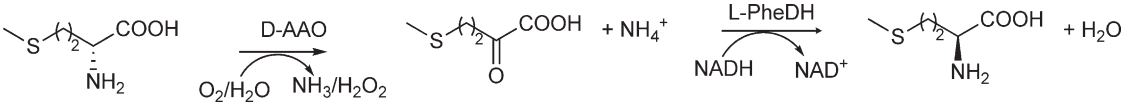
5.03.3.2.2 L-Phenylalanine dehydrogenase in chiral conversion

The broad specificity of PheDH has been applied also to the resolution of racemic amino acids or inversion of stereospecificity. Findrik and Vasić-Rački[2] report the use of PheDH of *Rhodococcus* in conjunction with D-AAO to convert D-methionine into L-methionine (**Scheme 7**).

The use of PheDH mutants in the opposite direction of reaction to accomplish the quantitative removal of the L-enantiomer of a range of substituted phenylalanine derivatives is discussed in more detail in Section 5.03.3.4.

5.03.3.2.3 L-Phenylalanine dehydrogenase in oxoacid synthesis

Most of the applications so far focus on the production of the chiral amino acid as the end product. Conversion of the chiral amino acid into the prochiral oxoacid as the end product is less common, although, for instance, Ödman *et al.*[31] describe the use of GDH to convert L-glutamate into the higher-value 2-oxoglutarate. Similarly, Findrik *et al.*[32] describe in some detail the kinetics of quantitative conversion of L-methionine into 2-oxo-4-methylthiobutyric acid. In view of the relatively unfavorable equilibrium for amino acid oxidation, thermodynamic and kinetic considerations have to be carefully balanced. A high pH favors oxidative deamination, and fortunately also the PheDH has an unusually high pH optimum, above 10. However, this in itself will not secure



Scheme 7

quantitative conversion, and it is essential to have an efficient recycling system for the coenzyme. This is achieved in the study described by the use of an NADH oxidase from *Lactobacillus*. Introduction of the second enzyme in a one-pot procedure, however, demands a compromise between the different conditions for optimum activity and for stability of the two biocatalysts. The NADH oxidase is maximally active at pH 5.5, and in order to retain efficient recycling the methionine oxidation was carried out at pH 9 rather than pH 10. Another possible benefit of the lower operating pH, though not discussed, would be the longer lifetime of the coenzyme at pH values closer to neutrality.

5.03.3.3 L-Leucine Dehydrogenase

5.03.3.3.1 L-Leucine dehydrogenase in chiral L-amino acid synthesis

Despite the preference of PheDH for aromatic rings and of LeuDH for aliphatic chains, there is in fact considerable overlap between the specificities of the two enzymes, which are closely related,[33] and, not surprisingly, there has been parallel development in their use and application. In the case of LeuDH this has been strongly influenced by the collaboration between the groups of Hummel and Kula[34] and the industrial amino acid producer Degussa AG (now Evonik). The latter adopted LeuDH as the basis for large-scale production of a number of chiral synthons such as *t*-leucine, neopentyl glycine, and homophenylalanine for the pharmaceutical industry.[35] The Degussa–Jülich collaborative consortium has provided a useful and thorough review[36] of the specificities of several different LeuDHs and also *Rhodococcus* PheDH with a range of nonnatural oxoacid substrates. The Bristol–Myers–Squibb team similarly surveyed several LeuDHs before finally selecting PheDH for their production process for 3-OH adamantyl glycine.[30] LeuDH can, of course, also be used in the production of biological amino acids, important for food and animal feed applications, and the productivity of reactor processes for this purpose is reviewed by Devaux-Basseguy *et al.*[37]

5.03.3.3.2 L-Leucine dehydrogenase in chiral conversion

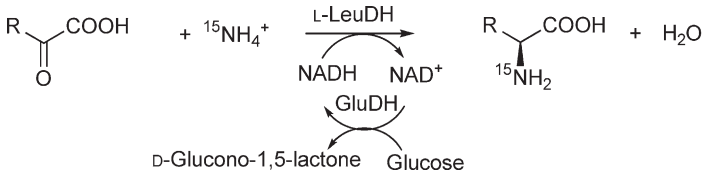
Hummel *et al.*[38] have applied the oxidative LeuDH reaction to produce D-*t*-leucine from the racemic mixture. The enzyme quantitatively removes the L-enantiomer, but this requires recycling of the NADH produced, and for this purpose Hummel *et al.* employed their *Lactobacillus* NADH oxidase. In the same paper, the authors present comparative kinetics for the oxidation of L-*t*-leucine and the natural leucine substrate, using the LeuDH of *Bacillus cereus*. It is notable that the enzyme shows a K_m approximately 10-fold higher than with the nonnatural substrate, and, perhaps more significant in a biotechnological context, a V_{max} 100-fold lower. Thus, it is not a very suitable enzyme for the purpose, and this raises a point developed in more detail in Section 5.03.3.4: in principle, it is no longer necessary to accept the limitations imposed by the range of biocatalysts existing in nature.

5.03.3.3.3 L-Leucine dehydrogenase in labeling of amino acids

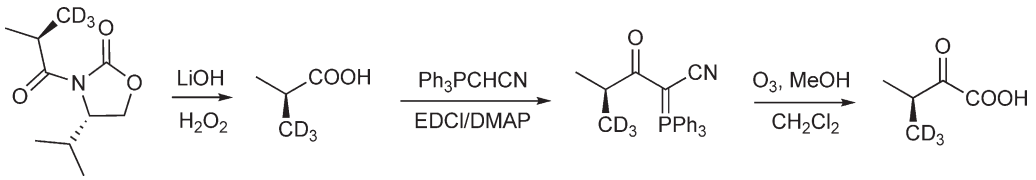
There is a continuing demand for amino acids labeled with deuterium, ^{13}C or ^{15}N, in particular for multidimensional NMR applications, and the broad specificity of LeuDHs has led to its use in several procedures for making labeled amino acids. Chiriac *et al.*[39] have recently reported an efficient procedure for ^{15}N labeling of several nonpolar amino acids using the thermostable LeuDH of *Bacillus stearothermophilus* and $^{15}NH_3$. Although this conversion runs in the thermodynamically favored direction of the reaction, it requires coenzyme recycling to ensure a complete conversion and to this end the authors used GluDH and glucose together with a mutarotase to provide a rapid supply of the glucose anomer used by GluDH (**Scheme 8**).

The Bristol group of Christine Willis, in collaboration with Amersham International, developed a procedure for deuterium (or ^{13}C) labeling of nonpolar amino acids.[40] In the chemical steps, a selectively methyl-labeled oxazolidinone is converted first into a 2-methyl carboxylic acid and then lengthened by two carbon atoms without racemization to yield an α-keto methyl ester (**Scheme 9**).

In the enzymatic part of the process, a one-pot conversion was achieved by using *Candida* lipase (Lip) to hydrolyze the ester and then LeuDH to catalyze the reductive amination (with or without ^{15}N labeling). In this case, the coenzyme recycling was accomplished by adding FDH and formate. The same group used a similar enzymatic strategy to prepare labeled L-threonine and L-allothreonine starting from the α-keto methyl ester.[41]



Scheme 8



Scheme 9

5.03.3.4 Engineered Novel Amino Acid Dehydrogenases

We have commented earlier on the seeming reluctance, particularly of industrial groups, to move beyond enzyme specificities already existing in nature. Methods have been available now for over 25 years for moving beyond the biological portfolio, adapting what nature has provided to achieve new activities. Nevertheless, for example, the Degussa process for the production of nonnatural amino acids has relied on the side specificity of natural enzymes, admittedly screened for the most favorable properties. Similarly, the Bristol–Myers–Squibb team has employed an enzyme-screening approach in their selection of biocatalysts. As we have seen, in the production of 3-OH adamantylglycine they indeed employed a mutant PheDH, but one produced only by a fortuitous accident. Such an accident must lead one to the conclusion that there is better biocatalysis to be had through protein engineering.

In the case of the aaDHs, a basis for rational modification came first from the solution of a high-resolution crystallographic structure for GDH[20,21] and then from the recognition that many of the aaDHs belong to one family, sharing a common mechanism and basic structure, so that it is possible to deduce the features of the substrate-binding pocket that have conferred individual specificities.[33] This position, permitting homology-based mutagenesis, has been further strengthened by individual structure solutions for both LeuDH[22] and PheDH.[23]

This structural information has been used as the basis for several programs of mutagenesis. Initially, modification of the glutamate-binding pocket of the original exemplar, GluDH of *Clostridium symbiosum*, produced a modified triple mutant with good activity toward L-methionine and L-norleucine.[24] Then in a homology-based program, mutations were introduced in the PheDH of *B. sphaericus*, first of all to test the validity of the structural modeling by altering the residues predicted to be responsible for the specificity differences between PheDH and LeuDH.[25] This indeed resulted in the predicted shift in specificity and was followed by a targeted alteration of Asn145, a residue predicted to be critical for interactions at the 4-position of the aromatic ring and in the first instance for discrimination between L-phenylalanine and L-tyrosine.[26] This again produced the predicted result, with substitution of Asn by Ala, Val, Leu, or Ile producing up to a 50-fold increase in discrimination between the two aromatic substrates. Subsequently, this set of mutant enzymes was tested with a range of phenylalanine derivatives and shown for several 4-substituted phenylalanine derivatives to offer a dramatic increase in activity over that of the wild-type enzyme.[42] These biocatalysts have also been shown to be quite tolerant of organic solvents[43,44] and can be used for chiral synthesis[45] or for removal of the L-isomer from a racemic mixture.[46] For chiral synthesis, the recycling of cofactor was achieved with yeast alcohol dehydrogenase and 5% ethanol, which offers the added advantage of improving substrate solubility. For resolution of a racemic mixture, the opposite requirement of recycling NADH to NAD^+ was achieved with a commercial NADH oxidase, which offers high activity at high pH.

In the context of kinetic resolution, an important issue is the effect on activity of the residual unreacted D-enantiomer. The fact that a compound is incapable of reacting does not mean that it cannot bind to the enzyme. An exploration of activities with 4 mmol l^{-1} racemic amino acid compared with 2 mmol l^{-1} pure L-amino acid[46] revealed striking differences between different mutants in terms of susceptibility to inhibition, suggesting a need for careful matching of the biocatalyst to the specific task. For example, with 4-methoxyphenylalanine the percentage inhibition by the equimolar D-enantiomer was respectively 9, 35, and 83% with three different nonpolar substitutions of the wild-type asparagine residue in the mutants N145V, N145A, and N145L.

Despite the clear value of site-directed mutagenesis in the cases cited, it is increasingly recognized that the fastest route to efficient, novel biocatalysts in many situations may be through random mutagenesis and robotic screening ('directed evolution'). Recent results in our own laboratory indicate that the aaDHs readily lend themselves to this approach in view of the efficiency of formazan-based activity screening,[47] and one may look forward to an increasing range of such mutationally optimized aaDH biocatalysts.

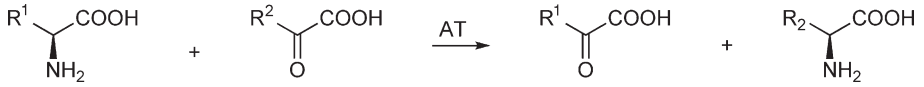
5.03.3.5 Opine Dehydrogenases

The group of Asano have reported an unusual biotransformation using an opine dehydrogenase (ODH)-like enzyme obtained from an *Arthrobacter* species.[48] This catalyzes the NAD^+-dependent oxidation of opine-type secondary amine dicarboxylic acids and thus in the reverse direction it uses NADH to reductively link pyruvate through its carbonyl group to the α-amino group of an amino group such as phenylalanine, creating a secondary amine (**Scheme 10**). The authors point out that such molecules are useful intermediates in the synthesis of ACE inhibitors. They demonstrate high levels of activity with a wide range of both natural and nonnatural amino acid substrates.

5.03.4 Aminotransferases

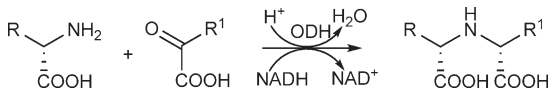
5.03.4.1 The Reaction and the Enzymes

The aminotransferases (ATs) (or transaminases) catalyze the exchange of an amino group between an amino acid and an oxoacid, so that the amino acid is converted into an oxoacid and vice versa (Equation (4)).



(4)

Usually, L-glutamate or 2-oxoglutarate provides one of the two pairs of reactants. The other pair may also be an α-amino acid and the corresponding 2-oxoacid but not necessarily so. For synthetic reactions, these enzymes have the merit in general of high turnover rates. In addition, the tightly bound pyridoxal phosphate cofactor does not need to be recycled, although it may be necessary to include added cofactor in a reaction mixture in order to ensure that the enzyme remains saturated and therefore fully active. A practical disadvantage of AT reactions is that their symmetry implies an evenly poised equilibrium, and thus complete conversion requires measures to pull the equilibrium over. We have already mentioned the use of ATs in conjunction with D-AAO for deracemization or inversion,[1] and in the second of those applications the equilibrium issue was solved by the use of cysteine sulphinic acid as the amino donor, since the corresponding 2-oxoacid spontaneously decomposes. A number of other interesting AT applications are reviewed below.

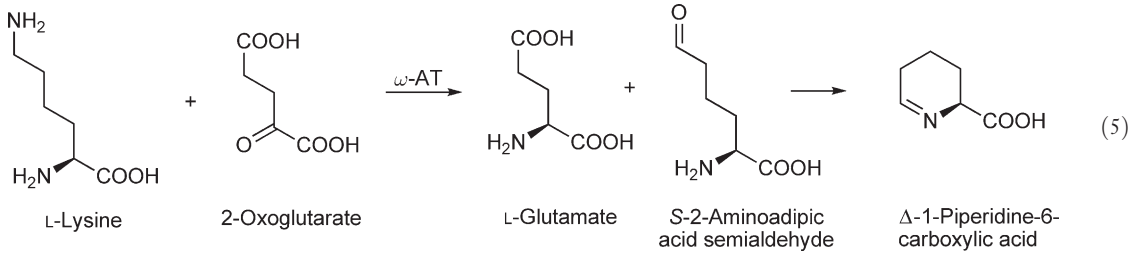


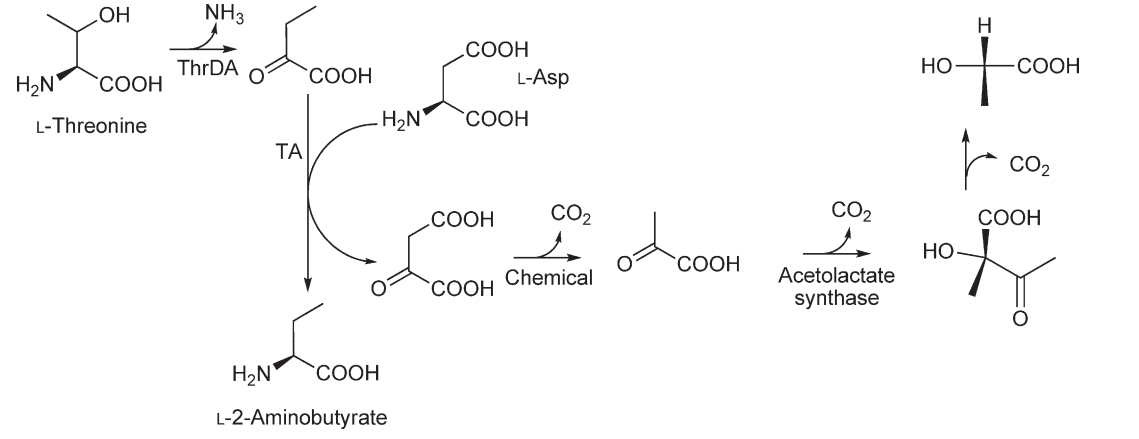
Scheme 10

5.03.4.2 Whole-Cell Procedure for Producing α-Aminobutyrate

The NSC group of Fotheringham and coworkers have used the broad-specificity AT encoded by the *ilv*A gene of *E. coli* to convert 2-oxobutyrate into α-aminobutyrate.[49] They address the fact that substrates such as 2-oxobutyrate 'are not commodity chemicals' by incorporating a second enzyme, threonine deaminase (ThrDA), which forms 2-oxobutyrate from L-threonine, which is both more stable and more readily and cheaply available. The transaminase utilizes L-aspartate as the amino donor, producing oxaloacetate, which spontaneously decarboxylates under the reaction conditions, yielding pyruvate. With this setup alone, the conversion to 2-oxobutyrate does not approach completion, and a considerable amount of L-alanine is formed as a by-product from pyruvate. In order to complete a workable process, these authors introduced a third enzyme step catalyzed by acetolactate synthase, which carries out condensation and decarboxylation on two pyruvate molecules, thus pulling the overall equilibrium over in favor of α-aminobutyrate formation (**Scheme 11**). The entire procedure is made efficient by introducing the genes for the three enzymes under the control of efficient promoters, so that all three are expressed simultaneously in the same *E. coli* strain.

This group also discuss more briefly an ingenious alternative solution to the equilibrium problem, involving a second AT. If L-glutamate rather than aspartate serves as amino donor for the main transamination reaction, then the equilibrium problem becomes one of disposing of the 2-oxoglutarate that is formed. This serves as the amino acceptor for the reaction catalyzed by L-lysine ε-AT. In this case, the overall equilibrium is tilted by virtue of the fact that the product of the second transamination, *S*-2-aminoadipic acid semialdehyde carries both an oxo group and an amino group and thus readily cyclizes to form Δ-1-piperidine-6-carboxylic acid (Equation (5)).





Scheme 11

5.03.4.3 D-β-Heterocyclic Alanine Derivatives

Kim's group in Seoul report[50] the application of another of the *E. coli* ATs, the aromatic L-AAT encoded by the *tyrB* gene, to enrich the D-component of racemic preparations of alanine substituted at the β-position with pyrazole, triazole, and imidazole. They also carried out an *in silico* investigation based on the crystal structure (PDB 3TAT) providing a reasonable rationalization (and therefore also potentially prediction) of substrate specificities.

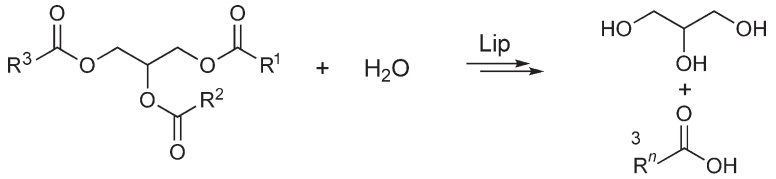
5.03.4.4 Substituted L-Glutamic Acid Analogues

Analogues of L-glutamic acid are of pharmacological interest in the characterization of CNS neuroreceptors. Bolte and colleagues[51,52] describe the chemical synthesis of 2-oxoglutarate derivatives with dimethyl, monomethyl, or methyl and hydroxyl substitution at C4, with both enantiomers where relevant. All of these have been used as substrates for glutamate oxaloacetate transaminase ('commercial' and therefore presumably of mammalian origin). All were found to react, though with very considerable differences in both K_m and V_{max} depending on both the stereochemistry and the nature of the substitution at C4. The dimethyl substitution produced much the worst substrate, with k_{cat}/K_m 2000-fold lower than with unsubstituted 2-oxoglutarate. The catalysis was nevertheless sufficient to permit synthesis of the corresponding L-glutamate derivatives in all cases, with the use of cysteine sulphinic acid as amino donor to drive the equilibrium in the required direction.

5.03.5 Lipases, Acylases, and Proteinases

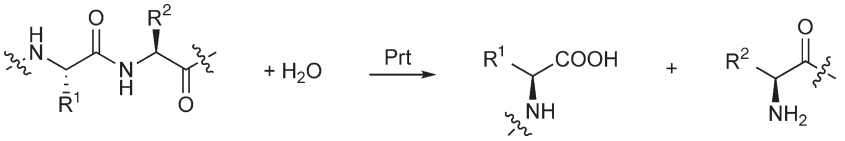
5.03.5.1 The Reactions and the Enzymes

The enzymes grouped in this section are biologically very distinct. Lip, by definition, catalyzes the hydrolysis of one or more of the three ester bonds linking fatty acids to the hydroxyl positions of glycerol (Equation (6)).



(6)

Proteinases (Prt), again by definition, are responsible for the cleavage of peptide bonds, usually with a preference for particular types of side chain on the amino acid residues on one or the other or both sides of the scissile bond (Equation (7)).



(7)

However, in both cases, although these classes of enzyme enjoy widespread industrial use in the specific context of their type reactions, their biotechnological prominence arises largely from their promiscuity. In the case of Lip, this refers not only to their ability to handle hydrolysis or transesterification of a much wider range of nonlipid substrates, but also to their catalysis of the hydrolysis or formation of amide bonds. The Lip of *Candida* yeasts have proved particularly popular.[53] In the case of Prt, conversely, there is the possibility of application not only to the cleavage of amide linkages in proteins but also to the hydrolysis of ester linkages in a wide variety of nonprotein substrates. Industrial chemists, ignoring the biologists' functional classification, have used these enzymes as a continuum toolkit of versatile catalysts. In many of the applications reviewed, the researchers have screened Prt and Lip together in the search for the most useful hydrolase available. Accordingly, we have chosen also to review them together. In the context of amino acid chemistry, Lip have

application since frequently chemical synthesis may deliver derivatives protected at both the amino and carboxyl groups. Esters may be selectively cleaved by Lip, and they are considered here together with those acylases that similarly cleave aminoacyl esters, as opposed to amino acid *N*-acylases that cleave amide linkages to the amino group.

5.03.5.2 Resolution of Epimers

Cabrele *et al.*[54] have used a chemoenzymatic approach in the preparation of the four different optically pure forms of *cis*-3-carboxycyclopentylglycine. Chemical synthesis delivered a mixture of epimeric cyclopentylglycines, protected by benzoylation of the amino group and an ethyl ester substitution on the α-carboxyl. This mixture was first separated chromatographically (allowing for the separation of two diastereomers, each being racemic). To isolate the pure enantiomers from each mixture, a wide range of organisms and also acylases, Prt, and Lip were tested, and the two giving best results were an acylase from the fungus *Aspergillus melleus* and Lip from porcine pancreas. The result was a selective cleavage of the ester, releasing one enantiomer as the free acid and allowing ready separation from the other form (**Scheme 12**) which could separately be chemically deprotected.

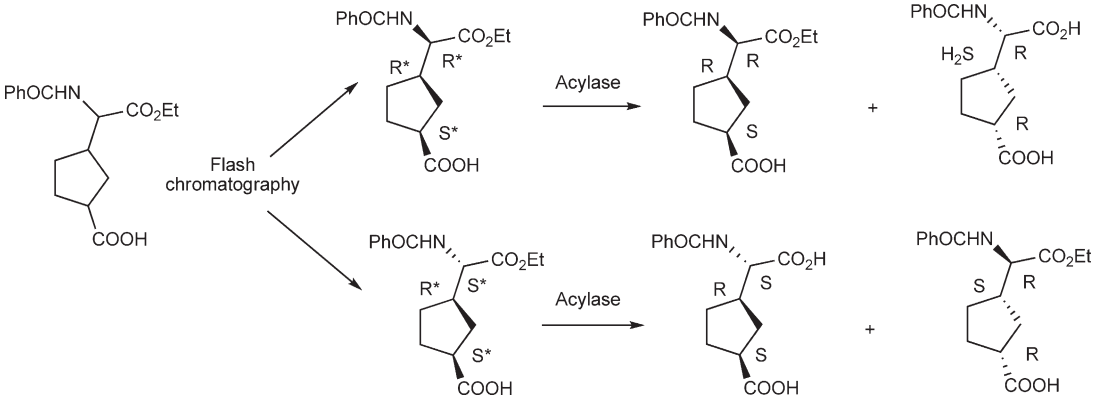
5.03.5.3 Synthesis of an Amide – *N*-Arachidonyl Glycine

Goujard *et al.*[55] have approached the task of synthesizing fatty acyl amino acids by first screening enzymes for the ability to cleave the amide bond of *N*-oleoyl glycine, testing several Prt and Lip. *Candida antarctica B* Lip gave the best result. This enzyme was then used in acetonitrile as solvent using arachidonic acid methyl ester as the acyl donor and glycine *t*-butyl ester as the receiving substrate. A 75% yield was achieved for this acyl amino acid, seen as a pharmacological candidate in the suppression of inflammatory pain.

5.03.5.4 Resolution of D- and L-Amino Acid Derivatives

5.03.5.4.1 Resolution of racemic amino acid esters

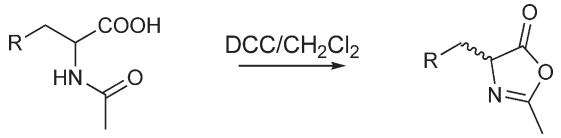
Parmar *et al.*[56] have developed a method for resolving racemic mixtures of a variety of natural and nonnatural amino acids using the ethyl ester of the amino acid protected at the amino position by the formation of a Schiff base with an aromatic aldehyde such as *p*-chlorobenzaldehyde. Both chymotrypsin and Lip such as porcine Lip gave good yields of the L-amino acid which precipitates out of solution as the amino acid ester released from the imine is cleaved by the hydrolase.



Scheme 12

5.03.5.4.2 Dynamic kinetic resolution

5.03.5.4.2(i) Resolution of heteroaryl alanine derivatives with lipase Podea *et al.*[57] have adapted a dynamic kinetic resolution method initially applied by Turner and colleagues to the resolution of racemic Phe and *t*-Leu.[58,59] The *N*-acetyl derivative of the amino acid is cyclized with the aid of dicyclohexylcarbodiimide (Equation (8)), yielding an oxazolone.



(8)

The resolution was then based on the enzymatic propanolysis of this derivative in dioxane as solvent. Lip Novozyme 435 selectively cleaves the L-form of the oxazolone producing an L-enriched (81–87% ee) 2-acetamido-3-(heteroaryl)propionic acid propyl ester, the dynamic aspect of the process being based on the continual racemization of the residual oxazolone. The propyl group was then removed with alkali and a second selective enzymatic step to remove the acetyl protecting group with Fluka Acylase 1 produced the L-amino acid at better than 99% ee (**Scheme 13**).

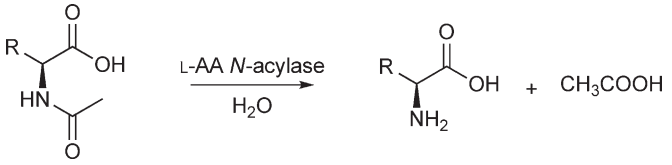
5.03.5.4.2(ii) Resolution of amino acid esters with subtilisin The commercial Prt 'Alcalase' from Novozymes, a subtilisin produced by *Bacillus licheniformis*, was used by Chen *et al.*[60] to carry out a dynamic kinetic resolution of benzyl, butyl, or propyl esters of DL-phenylalanine, tyrosine, and leucine. The hydrolysis was performed at pH 8.5 in 2-methyl-2-propanol/water (19:1) and the freed L-amino acids precipitated. The key feature bringing about continual racemization of the remaining D-amino acid esters was the inclusion of 20 mmol l^{-1} pyridoxal phosphate.

5.03.5.4.2(iii) Resolution of amino acid thioesters with subtilisin Arosio *et al.*[61] have explored another chemical solution to the racemization required for dynamic kinetic resolution. They established that the thioesters of *N*-Boc protected L-amino acids were cleavable by subtilisin. Furthermore, the thioesters, unlike the corresponding esters, showed ready deuterium exchange of the α-proton in the presence of trioctylamine. In view of this, dynamic kinetic resolution was successfully carried out in 2:1 water/methyl *t*-butyl ether on a series of *N*-Boc nonnatural amino acid ethanethiol thioesters. The released *N*-Boc L-amino acids were extracted into ethyl acetate and in all cases tested gave over 95% conversion and better than 99% ee. The amino acids in this study included 2-thienylglycine, phenylglycine, and the 2- and 4-fluoro and chloro derivatives of the latter. The authors indicate that their method should be generally applicable and not limited to β,γ-unsaturated α-amino acid derivatives.

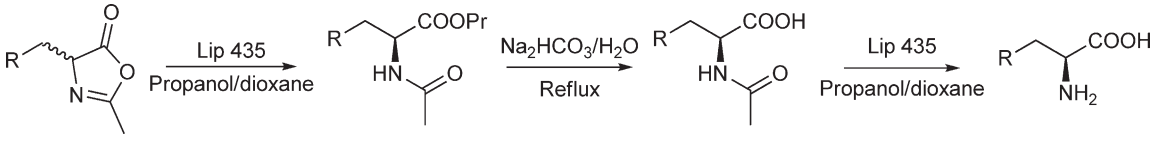
5.03.6 L-Amino Acid *N*-Acylases

5.03.6.1 The Reaction and the Enzymes

L-Amino acid acylases hydrolyze the amide bonds of *N*-acetyl-L-amino acids (Equation (9)).



(9)



Scheme 13

5.03.6.2 Dynamic Kinetic Resolution Using an *N*-Acylase and a Racemase

Hsu *et al.*[62] have cloned two enzymes from *Deinococcus radiodurans* for overexpression in *E. coli* in order to carry out a dynamic kinetic resolution to obtain L-homophenylalanine, frequently required for pharmaceutical synthesis. The starting material is the racemic mixture of *N*-acetylated homophenylalanine, and the two enzymes are an amino acid *N*-acylase, which specifically removes the acetyl group from the L-enantiomer, and a racemase, which interconverts the D- and L-forms of the *N*-acyl amino acids. The resolution was carried out successfully using whole-cell biocatalysts, with the two enzymes either expressed in separate *E. coli* strains or coexpressed in the same cells.

5.03.6.3 Synthesis of *N*-Lauroyl Amino Acids and Peptides

A *Streptomyces* enzyme that catalyzes hydrolysis of capsaicin is described by Koreishi *et al.*[63] The substrate is an *N*-vanillyl aliphatic amide, and the authors found that their enzyme also accepted *N*-lauroyl amino acids as substrates. The enzyme was used successfully to catalyze the reaction in the opposite direction, driving the equilibrium toward synthesis by running it in buffer containing 78% glycerol. Yields of 5–40% were obtained for a wide range of natural L-amino acids. In the case of L-lysine the enzyme catalyzed acylation at both amino groups, with a clear preference for the ε-NH_2.

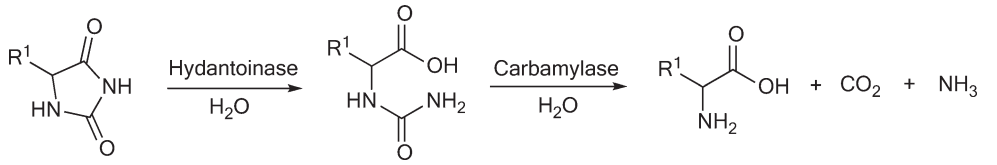
5.03.6.4 Preparation of D-Amino Acids Using the *N*-Acylase Activity of Penicillin-G Acylase

Carboni *et al.*[64] report the use of penicillin-G acylase of *E. coli* to produce the D-enantiomers of glutamic acid and glutamine from the racemic amino acids. This is based on the N-phenylacetylation of the L-enantiomer using phenylacetic acid methyl ester as donor. This process was not very satisfactory when the wild-type enzyme was employed because, although the acylated L-amino acid was produced at high ee, a substantial hydrolytic side reaction resulted in the residual D-amino acid being contaminated with the L-enantiomer. However, in studying several mutant acylases, the authors found a single mutant in which the hydrolytic action was suppressed, allowing the production of D-glutamine at 97% ee.

5.03.7 Hydantoinases and Carbamoylases

5.03.7.1 Hydantoinases and Carbamoylases

The two groups of enzymes discussed here have attracted attention because both offer a useful broad spectrum of substrate specificity. They are grouped together because in the context of amino acid synthesis they form a natural pair. Amino acid hydantoins are convenient from the standpoint of organic synthesis. The hydantoinases cleave the ring, producing the *N*-carbamoyl derivative of the amino acid. This must then be further hydrolyzed to obtain the free amino acid, and this step is likely to be strictly enantioselective (Equation (10)).



(10)

5.03.7.2 Enantioselective Synthesis of (*S*)-2-Amino-4-Phenylbutanoic Acid

Hsu's group in Taiwan have developed a procedure[65] for the synthesis of (*S*)-2-amino-4-phenylbutanoic acid, the phenylalanine homologue with one additional methylene group. Hydantoinase and L-*N*-carbamoylase genes have been cloned from different *Bacillus* species and overexpressed in *E. coli*. Both the *R*- and the *S*-enantiomers were cleaved by the hydantoinase, but only the *S*-form of the *N*-carbamoyl amino acid was hydrolyzed by the second enzyme. The reactions could be run in a single pot, with successive addition of the two enzymes, and were successful in the sense of giving a product of high chiral purity. However, the yield was

not high over the time course of the reported experiment (only 5%). The two forms of the hydantoin racemize spontaneously through keto–enol tautomerization, and, since the hydantoinase reaction is reversible, in theory a better yield should be achievable. However, as the authors say, either selection or engineering of an *S*-selective hydantoinase would probably lead to a speedier process and better yields.

5.03.7.3 Enantioselective Synthesis of D-*p*-Hydroxyphenylglycine

In view of the last report, it is interesting that Wu *et al.*[66] in Beijing have identified an organism, *Sinorhizobium morelens* S-5, that can convert the hydantoin of racemic *p*-hydroxyphenylglycine into the D-amino acid. This, similar to the process just described, involves a hydantoinase and a carbamoylase, but both appear to be strictly D-specific. These authors again draw attention to the fact that under mildly alkaline conditions, spontaneous racemization of the hydantoin should permit a 100% conversion to the final D-product.

5.03.7.4 D- and L-3-Trimethylsilylalanine

Pietzsch *et al.*[67] used immobilized *Agrobacterium* cells to transform the hydantoin of racemic 3-trimethylsilylalanine, obtaining the D-enantiomer in good yield provided the cells were kept anaerobic. The authors surmise that under aerobic conditions an aaDH may degrade the product. In addition, the 95% ee (rather than 100%) was a surprise, as an *N*-carbamoylase purified from the cells showed strict D-specificity. It was unclear whether this reflected the presence of a second carbamoylase of L-specificity or an amino acid racemase. These authors also studied the cleavage of the racemic hydantoin by various hydantoinases and note in the case of the enzyme from *Arthrobacter aurescens* that its enantiomeric preference varies with the substrate, D- in this case but L- with the hydantoins of Phe and Trp. The same organism yields an *N*-carbamoylase of strict L-specificity, allowing preparation of the pure *N*-enantiomer of the amino acid.

5.03.7.5 *N*-Carbamoylase Combined with *N*-Acyl Amino Acid Racemase to Produce L-Homophenylalanine

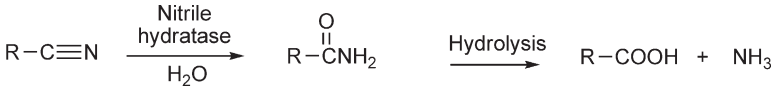
In Section 5.03.6.2, a stereoselective synthesis of L-homophenylalanine from the racemic *N*-acetylated amino acid is described. The authors, however, found that substrate solubility limited the utility of this procedure. Having found an L-*N*-carbamoylase in *Bacillus kaustophilus*, they introduced the gene for this enzyme together with that for the *N*-acyl amino acid racemase from *D. radiodurans* into *E. coli* for coexpression.[68] These cells, permeabilized with 0.5% toluene, were able to deliver L-homophenylalanine in 99% yield and were able to be used for multiple reaction cycles.

5.03.8 Nitrilases, Nitrile Hydratases, and Amidases

5.03.8.1 The Reactions and the Enzymes

The ease of the Strecker synthesis from aldehydes makes α-aminonitriles an attractive and important route to α-amino acids. Fortunately, the microbial world offers a number of enzymes for carrying out the necessary conversions, some of them highly stereoselective. Nitrilases catalyze a direct conversion of nitrile into carboxylic acid (Equation (11)), whereas nitrile hydratases catalyze formation of the amide, which can then be hydrolyzed to the carboxylic acid in a second step (Equation (12)). In a recent survey, with a view to bioremediation and synthesis, Brady *et al.*[69] have surveyed the ability of a wide range of bacteria and yeasts to grow on diverse nitriles and amides as sole nitrogen source. This provides a rich source of information on enzymes for future application.

$$R-C\equiv N \xrightarrow[2H_2O]{\text{Nitrilase}} R-COOH + NH_3 \qquad (11)$$



(12)

5.03.8.2 Conversions with Nitrile Hydratase and Amidase

Beard and Page[70] review the properties of a *Rhodococcus* strain, which gave very effective whole-cell conversions of α-aminonitriles to the corresponding *S*-amino acids. The intermediate amide accumulated and was racemic. The slower amidase activity then gave a stereoselective conversion exclusively of the *S*-amide to the amino acid, leaving the *R*-enantiomer untouched even after 2 days. The amidase activity was then tested with a wide variety of racemic α-substituted α-aminoacetamides and in all cases gave *S*-amino acid products at greater than 98% ee. Wang and Lin[71,72] have similarly used the differing enantioselectivities of these two enzymes in *Rhodococcus* to produce high yields of L-arylglycines and D-arylglycine amides. The D-amide can of course be chemically hydrolyzed[73] to provide also the D-amino acid in high optical purity. The same group have also explored the use of this whole-cell biocatalyst to produce optically active α-methylamino acids.[74] The nitrile hydratase activity did not accept the methylated substrate, but, starting from the racemic amide, a good yield of both enantiomers could again be obtained, the L- directly and the D- after chemical hydrolysis of the residual amide.

A strain of *Pseudomonas aeruginosa* has been recently described,[75] which shows the opposite enantioselectivity, converting racemic arylaminonitriles efficiently into the D-amino acids. Again, whole-cell biocatalysis worked well, the cells being entrapped in alginate beads. It is unclear whether this biotransformation involves an amide intermediate.

5.03.9 Concluding Remarks

In view of the burgeoning activity in this area of biotechnology, it would be impossible to be entirely exhaustive within the confines of a review such as this. However, we hope to have given a representative and up-to-date coverage that shows the huge range of viable methods that are now becoming available. The versatility of microbial physiology, and hence enzymes, combined with the power of modern molecular genetics, has provided a tool kit that allows industrial chemists readily to match the extraordinary enantioselectivity of biocatalysis with whatever chemistry may be convenient for an individual case. What we have reviewed is very obviously only a beginning, because the methods described here are overwhelmingly based on what is already biologically available. Once the methods of directed evolution and site-directed mutagenesis are more widely deployed, there can be little doubt that biocatalysis will occupy an increasingly central role in the synthesis and transformation of amino acids.

Nomenclature

EC1.4.3.2	L-amino acid oxidase (amino acid oxidase)
EC1.4.3.3	D-amino acid oxidase (amino acid oxidase)
EC1.4.3.20	L-lysine oxidase (amino acid oxidase)
EC1.4.1.1	alanine dehydrogenase (amino acid dehydrogenase)
EC1.4.1.3	glutamate dehydrogenase, bovine $NAD(P)^+$-dependent (amino acid dehydrogenase)
EC1.4.1.2	glutamate dehydrogenase, clostridial NAD^+-dependent (amino acid dehydrogenase)
EC1.4.1.8	valine dehydrogenase (amino acid dehydrogenase)
EC1.4.1.9	leucine dehydrogenase (amino acid dehydrogenase)
EC1.4.1.20	phenylalanine dehydrogenase (amino acid dehydrogenase)
EC1.1.1.47	glucose dehydrogenase
EC1.2.1.2	formate dehydrogenase
EC1.6.3.1	NADH oxidase

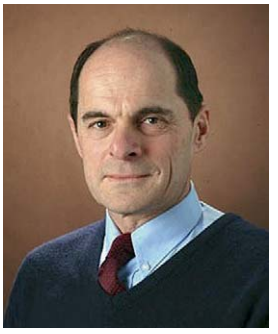
EC1.11.1.6	catalase
EC2.6.1.1	L-aspartate aminotransferase (glutamate oxaloacetate transaminase) (aminotransferase (transaminase))
EC2.6.1.57	aromatic L-amino acid transaminase (aminotransferase (transaminase))
EC2.6.1.36	L-lysine ε-aminotransferase (aminotransferase (transaminase))
EC4.3.1.19	threonine deaminase (threonine ammonia lyase)
EC5.1.3.3	mutarotase (aldose 1-epimerase)
EC5.1.1.1	alanine racemase
EC3.1.1.3	lipase (triacylglycerol lipase) (hydrolase)
EC3.4.21.62	subtilisin (bacterial serine proteinase) (hydrolase)
EC3.5.5.1	nitrilase (hydrolase)
EC3.5.1.4	amidase (hydrolase)
EC4.2.1.84	nitrile hydratase
EC4.1.3.18	acetolactate synthase

References

1. C. Cheng; S.-C. Wu, *J. Chromatogr.* **2000**, *896*, 299–310.
2. Z. Findrik; D. Vasić-Rački, *Biotechnol. Bioeng.* **2007**, *98*, 956–967.
3. R. L. Hanson; M. D. Schwinden; A. Banerjee; D. B. Brzozowski; B.-C. Chen; B. P. Patel; C. G. McNamee; G. A. Kodersha; D. R. Kronenthal; R. N. Patel; L. J. Szarka, *Bioorg. Med. Chem.* **1999**, *7*, 2247–2252.
4. R. N. Patel, *Biomol. Eng.* **2001**, *17*, 167–182.
5. J. Struck; I. W. Sizer, *Arch. Biochem. Biophys.* **1960**, *86*, 260–266.
6. A. Caligiuri; P. D'Arrigo; E. Rosini; D. Tessaro; G. Molla; S. Servi; L. Pollegioni, *Adv. Synth. Catal.* **2006**, *348*, 2183–2190.
7. A. Caligiuri; P. D'Arrigo; T. Gefflaut; G. Molla; L. Pollegioni; E. Rosini; C. Rossi; S. Servi, *Biocatal. Biotransformation* **2006**, *24*, 409–413.
8. V. Hélaine; J. Rossi; J. Bolte, *Tetrahedron Lett.* **1999**, *40*, 6577–6580.
9. I. Fotheringham; I. Archer; R. Carr; R. Speight; N. J. Turner, *Biochem. Soc. Trans.* **2006**, *34*, 287–290.
10. W. Märkle; S. Lütz, *Electrochim. Acta* **2008**, *53*, 3175–3180.
11. D. P. Pantaleone; A. M. Geller; P. P. Taylor, *J. Mol. Catal. B Enzym.* **2001**, *11*, 795–803.
12. Z. Findrik; B. Geueke; W. Hummel; D. Vasić-Rački, *Biochem. Eng. J.* **2006**, *27*, 275–286.
13. B. Geueke; W. Hummel, *Enzyme Microb. Technol.* **2002**, *31*, 77–87.
14. B. Geueke; W. Hummel, *Protein Expr. Purif.* **2003**, *28*, 303–309.
15. M. Yasuda; M. Ueda; H. Muramatsu; H. Mihara; N. Esaki, *Tetrahedron: Asymmetry* **2006**, *17*, 1775–1779.
16. H. Mihara; R. Kakutani; H. Muramatsu; M. Yasuda; M. Ueda; T. Kurihara; N. Esaki, *FEBS J.* **2005**, *272*, 1117–1123.
17. P. C. Engel; K. Dalziel, *Biochem. J.* **1967**, *105*, 691–695.
18. Y. Asano; A. Nakazawa, *Agric. Biol. Chem.* **1987**, *51*, 2035–2036.
19. E. L. Smith; B. M. Austen; K. M. Blumenthal; J. F. Nyc, In *The Enzymes*; P. D. Boyer, Ed.; Academic Press: New York, 1975; Vol. 11, pp 293–367.
20. P. J. Baker; K. L. Britton; P. C. Engel; G. W. Farrants; K. S. Lilley; D. W. Rice; T. J. Stillman, *Proteins* **1992**, *12*, 75–86.
21. T. J. Stillman; P. J. Baker; K. L. Britton; D. W. Rice, *J. Mol. Biol.* **1993**, *234*, 1131–1139.
22. P. J. Baker; A. P. Turnbull; S. E. Sedelnikova; T. J. Stillman; D. W. Rice, *Structure* **1995**, *3*, 693–705.
23. J. L. Vanhooke; J. B. Thoden; N. M. W. Brunhuber; J. S. Blanchard, *Biochemistry* **1999**, *38*, 2326–2339.
24. X.-G. Wang; K. L. Britton; T. J. Stillman; D. W. Rice; P. C. Engel, *Eur. J. Biochem.* **2001**, *268*, 5791–5799.
25. S. Y. K. Seah; K. L. Britton; D. W. Rice; Y. Asano; P. C. Engel, *Biochemistry* **2002**, *41*, 11390–11397.
26. S. Y. K. Seah; K. L. Britton; D. W. Rice; Y. Asano; P. C. Engel, *Eur. J. Biochem.* **2003**, *270*, 1–7.
27. K. Kataoka; T. Tanizawa, *J. Mol. Catal. B Enzym.* **2003**, *23*, 299–309.
28. Y. Asano; K. Endo; A. Nakazawa; Y. Hibino; N. Okazaki; M. Ohmori; N. Numao; K. Kondo, *Agric. Biol. Chem.* **1987**, *51*, 2621–2623.
29. Y. Asano; A. Yamada; Y. Kato; K. Yamaguchi; Y. Hibino; K. Hirai; K. Kondo, *J. Org. Chem.* **1990**, *55*, 5567–5571.
30. R. L. Hanson; S. L. Goldberg; D. B. Brzozowski; T. P. Tully; D. Cazzulino; W. L. Parker; O. K. Lyngberg; T. C. Vu; M. K. Wong; R. N. Patel, *Adv. Synth. Catal.* **2007**, *349*, 1369–1378.
31. P. Ödman; W. B. Wellborn; A. S. Bommarius, *Tetrahedron: Asymmetry* **2004**, *15*, 2933–2937.
32. Z. Findrik; I. Šimunović; D. Vasić-Rački, *Biochem. Eng. J.* **2008**, *39*, 319–327.
33. K. L. Britton; P. J. Baker; P. C. Engel; D. W. Rice; T. J. Stillman, *J. Mol. Biol.* **1993**, *234*, 938–945.
34. W. Hummel; M.-R. Kula, *Eur. J. Biochem.* **1989**, *184*, 1–13.
35. A. S. Bommarius; M. Schwarm; K. Stingl; M. Kottenhahn; K. Huthmacher; K. Drauz, *Tetrahedron: Asymmetry* **1995**, *6*, 2851–2888.
36. G. Krix; A. S. Bommarius; K. Drauz; M. Kottenhahn; M. Schwarm; M.-R. Kula, *J. Biotechnol.* **1997**, *53*, 29–39.

37. R. Devaux-Basseguy; A. Bergel; M. Comtat, *Enzyme Microb. Technol.* **1997**, *20*, 248–258.
38. W. Hummel; M. Kuzu; B. Geueke, *Org. Lett.* **2003**, *5*, 3649–3650.
39. M. Chiriac; I. Lupan; N. Bucurenci; O. Popescu; N. Palibroda, *J. Labelled Comp. Radiopharm.* **2008**, *51*, 171–174.
40. N. M. Kelly; R. G. Reid; C. L. Willis; P. L. Winton, *Tetrahedron Lett.* **1996**, *37*, 1517–1520.
41. A. Sutherland; C. L. Willis, *Tetrahedron Lett.* **1997**, *38*, 1837–1840.
42. P. Busca; F. Paradisi; E. Moynihan; A. R. Maguire; P. C. Engel, *Org. Biomol. Chem.* **2004**, *2*, 2684–2691.
43. G. Cainelli; P. C. Engel; P. Galletti; D. Giacomini; A. Gualandi; F. Paradisi, *Org. Biomol. Chem.* **2005**, *3*, 4316–4320.
44. S. Chen; P. C. Engel, *Enzyme Microb. Technol.* **2007**, *40*, 1407–1411.
45. F. Paradisi; S. Collins; A. R. Maguire; P. C. Engel, *J. Biotechnol.* **2007**, *128*, 408–411.
46. F. Paradisi; P. Conway; A. R. Maguire; P. C. Engel, *Org. Biomol. Chem.* **2008**, *6*, 3611–3615.
47. S. Chen; P. C. Engel, *J. Biotechnol.* **2009**, http://dx.doi.org/10.1016/j.jbiotec.2009.03.005.
48. Y. Kato; H. Yamada; Y. Asano, *J. Mol. Catal. B Enzym.* **1996**, *1*, 151–160.
49. D. J. Ager; T. Li; D. P. Pantaleone; R. F. Senkpeil; P. P. Taylor; I. G. Fotheringham, *J. Mol. Catal. B Enzym.* **2001**, *11*, 199–205.
50. B.-K. Cho; T. Y. Park; J.-H. Seo; K. Kinnera; B.-S. Lee; B. G. Kim *Biotechnol. Bioeng.* **2004**, *88*, 512–519.
51. V. Hélaine; J. Bolte, *Eur. J. Org. Chem.* **1999**, 3403–3406.
52. V. Hélaine; J. Rossi; T. Gefflaut; S. Alaux; J. Bolte, *Adv. Synth. Catal.* **2001**, *343*, 692–697.
53. S. Benjamin; A. Pandey, *Yeast* **1998**, *14*, 1069–1087.
54. C. Cabrele; F. Clerici; R. Gandolfi; M. L. Gelmi; F. Molinari; S. Pellegrino, *Tetrahedron* **2006**, *62*, 3502–3508.
55. L. Goujard; M. C. Figueroa; P. Villeneuve, *Biotechnol. Lett.* **2004**, *26*, 1211–1216.
56. V. S. Parmar; A. Singh; K. S. Bisht; N. Kumar; Y. N. Belokon; K. A. Kochetkov; N. S. Ikonnikov; S. A. Orlova; V. I. Tararov; T. F. Saveleva, *J. Org. Chem.* **1996**, *61*, 1223–1227.
57. P. V. Podea; M. I. Toşa; C. Paizs; F. D. Irimie, *Tetrahedron: Asymmetry* **2008**, *19*, 500–511.
58. S. Brown; M. C. Parker; N. Turner, *Tetrahedron: Asymmetry* **2000**, *11*, 1687–1690.
59. N. Turner; J. Winterman; R. McCague; J. Parratt; S. Taylor, *Tetrahedron Lett.* **1995**, *36*, 1113–1116.
60. S. T. Chen; W.-H. Huang; K.-T. Wang, *J. Org. Chem.* **1994**, *59*, 7580–7581.
61. D. Arosio; A. Caligiuri; P. D'Arrigo; G. Pedrocchi-Fantoni; C. Rossi; C. Saraceno; S. Servi; D. Tessaro, *Adv. Synth. Catal.* **2007**, *349*, 1345–1348.
62. S.-K. Hsu; H.-H. Lo; C.-H. Kao; D.-S. Lee; W. H. Hsu, *Biotechnol. Prog.* **2006**, *22*, 1578–1584.
63. M. Koreishi; D. Zhang; H. Imanaka; K. Imamura; S. Adachi; R. Matsuno; K. Nakanishi, *J. Agric. Food Chem.* **2006**, *54*, 72–78.
64. C. Carboni; H. G. T. Kierkels; L. Gardossi; K. Tamiola; D. B. Janssen; P. M. Quaedflieg, *Tetrahedron: Asymmetry* **2006**, *17*, 245–251.
65. H.-H. Lo; C.-H. Kao; D.-S. Lee; T. K. Yang; W. H. Hsu, *Chirality* **2003**, *15*, 699–702.
66. S. Wu; L. Yang; Y. Liu; G. Zhao; J. Wang; W. Sun, *Enzyme Microb. Technol.* **2005**, *36*, 520–526.
67. M. Pietzsch; T. Waniek; R. J. Smith; S. Bratovanov; S. Bienz; C. Syldatk, *Monatsh. Chem.* **2000**, *131*, 645–653.
68. S.-K. Hsu; H.-H. Lo; W.-D. Lin; I.-C. Chen; C.-H. Kao; W. H. Hsu, *Process Biochem.* **2007**, *42*, 856–862.
69. D. Brady; N. Dube; R. Petersen, *S. Afr. J. Sci.* **2006**, *102*, 339–344.
70. T. M. Beard; M. I. Page, *Antonie van Leeuwenhoek* **1998**, *74*, 99–106.
71. M.-X. Wang; S.-J. Lin, *Tetrahedron Lett.* **2001**, *42*, 6925–6927.
72. M.-X. Wang; S.-J. Lin; J. Liu; Q.-Y. Zheng, *Adv. Synth. Catal.* **2004**, *346*, 439–445.
73. M.-X. Wang; S.-J. Lin, *J. Org. Chem.* **2002**, *67*, 6542–6545.
74. M.-X. Wang; J. Liu; D.-X. Wang; Q.-Y. Zheng, *Tetrahedron: Asymmetry* **2005**, *16*, 2409–2416.
75. F. O. M. Alonso; O. A. C. Antunes; E. G. Oestreicher, *J. Braz. Chem. Soc.* **2007**, *18*, 566–571.

Biographical Sketches



Paul C. Engel is an enzymologist/protein engineer. Educated in Hong Kong and then as a biochemist at Oxford, he did a D.Phil. in enzyme kinetics (1968). After completing post-doctoral research on flavoproteins at the University of Michigan, he became a lecturer in Biochemistry at Sheffield (1970), a reader in 1984, and founding director of the

interdisciplinary Krebs Institute. In a 2-year interlude he became programme manager for the UK SERC Biotechnology Directorate's 'Protein Engineering Club' experiment in state/industry co-funding. Following a Royal Society Leverhulme Trust Research Fellowship back at Sheffield and a Visiting Professorship at Hong Kong University, he took up the Chair of Biochemistry at the University College Dublin (UCD) in 1994, spearheading the formation of UCD's flagship Conway Institute.

Professor Engel studies diverse aspects of oxidative enzymes, with emphasis on engineering novel biocatalysts, understanding intersubunit communication in allosteric enzymes, and explaining molecular defects in enzyme-deficiency diseases. He also has a small campus company, Enzolve Technologies, producing enzyme-based diagnostics.

He has taken keen interest in public awareness of science, promoting an annual lay presentation competition for research students. A member of the Royal Irish Academy since 2001, he chairs its Life Sciences Committee.

Dr. Francesca Paradisi graduated in chemistry in 1998 from the University of Bologna, Italy. She was awarded a Ph.D. in synthetic organic chemistry from the same university in 2002; her thesis dealt with the synthesis of amino acids and peptides with potential biological activity mimicking the diaminopimelic structure. She joined the Biochemistry Department of UCD, Ireland, in 2002 where she specialized in biocatalysis and enzymatic processes working with Professor Paul Engel. Together they have published so far eight papers in prestigious journals. In 2006, UCD School of Chemistry and Chemical Biology offered her a lectureship position in chemical biology. Her research interests are mainly in the area of biocatalysis, especially the discovery and applications of halophilic alcohol dehydrogenases and the synthesis of small molecules with biological activity.

5.04 1-Aminocyclopropane-1-Carboxylate Synthase, an Enzyme of Ethylene Biosynthesis

Małgorzata Jakubowicz and Witold Nowak, Adam Mickiewicz University, Poznan, Poland

© 2010 Elsevier Ltd. All rights reserved.

5.04.1 Introduction

Ethylene, a simple two-carbon unsaturated olefin, exists in the gaseous state under normal physiological conditions of temperature and pressure, and despite its simple structure is a critical mediator of plant growth and development. In the last three decades, a remarkable progress in genetic and molecular analyses of the ethylene synthesis and signaling pathway was made complementing earlier physiological studies. The effect of ethylene on plant growth was described in the scientific literature in the nineteenth century when the damaging influence of 'illuminating' gas on trees growing near broken gas pipes was found.[1] At the beginning of the last century, Neljubow[2] discovered ethylene as an active component of illuminating gas contaminating the laboratory air and causing a strange growth habit in etiolated pea seedlings. In 1924, ethylene present in combustion fumes from kerosene stoves was identified as the cause for the earlier lemon degreening.[3] Since the

discovery by Gane[4] in 1934 that ripe apples produce ethylene, this gas has been recognized as a plant hormone because it is a natural product of plant metabolism.

5.04.2 Ethylene Biosynthesis in Higher Plants

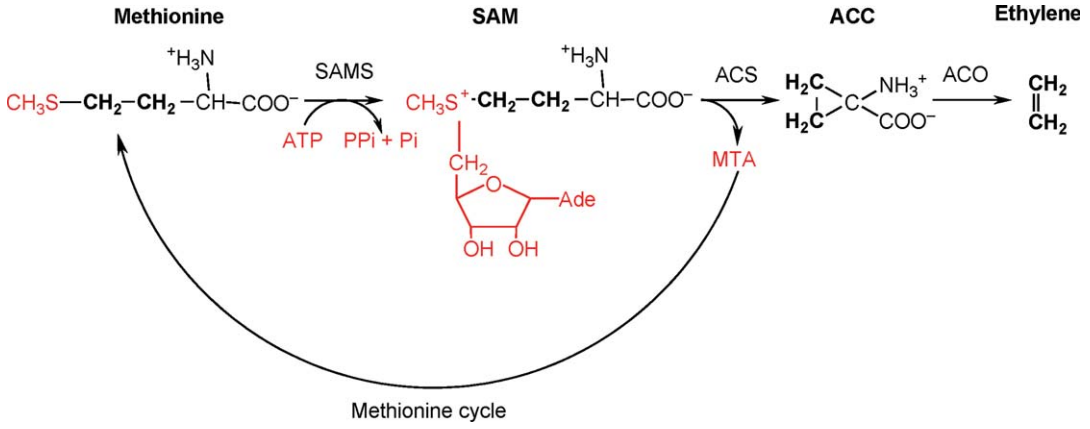
Ethylene is synthesized by higher plants in almost trace amounts up to 500 nl per gram per hour.[1] Practically, all tissues of higher plants have the capacity to produce ethylene and in some species its production is regulated in a circadian manner with a peak of synthesis at the midday. Plants respond to exogenous ethylene at concentrations ranging from 0.2 to 1000 nl per liter of air.[3]

5.04.2.1 Ethylene Biosynthesis Pathway Enzymes

The possibility that many organic compounds could potentially be precursors of ethylene was raised, but direct evidence that in apple fruit tissue ethylene derives only from carbons of methionine was provided by Lieberman[5] and was confirmed for other plant species. The pathway of ethylene biosynthesis has been well characterized during the last three decades. The major breakthrough came from the work of Yang and Hoffman,[6] who established *S*-adenosyl-L-methionine (SAM) as the precursor of ethylene in higher plants. The key enzyme in ethylene biosynthesis 1-aminocyclopropane-1-carboxylate synthase (*S*-adenosyl-L-methionine methylthioadenosine lyase, EC 4.4.1.14; ACS) catalyzes the conversion of SAM to 1-aminocyclopropane-1-carboxylic acid (ACC) and then ACC is converted to ethylene by 1-aminocyclopropane-1-carboxylate oxidase (ACO) (**Scheme 1**).

In addition to ACC, ACS produces 5′-methylthioadenosine (MTA), which is recycled through methionine cycle to methionine (see **Scheme 1**). Recycling of MTA back to methionine requires only the available ATP. A constant concentration of cellular methionine is maintained even when ethylene is rapidly synthesized or when the pool of free methionine is small. The methionine cycle involves the following subsequent intermediates: MTA, 5-methylthioribose (MTR), 5-methylthioribose-1-phosphate (MTR-1-P), 2-keto-4-methylthiobutyrate (KMB), and then the recycled methionine.[6,7]

ACO requires ascorbate as a cosubstrate and catalyzes the coupled oxidation of ACC and ascorbate to ethylene, cyanhydric acid, carbon dioxide, and dehydroascorbate, using a single nonheme ferrous ion and dioxygen.[8,9] The active site contains a single Fe(2+) ion linked to the side chains of two histidines and one aspartate. Both ACC and dioxygen are coordinated to the Fe(2+) ion, generating an Fe(3+)-ACC-superoxo



Scheme 1 The ethylene biosynthetic pathway. The enzymes catalyzing each step are shown above the arrows. SAM: *S*-adenosyl-L-methionine; SAMS: *S*-adenosyl-L-methionine synthetase; ACC: 1-aminocyclopropane-1-carboxylic acid; ACS: 1-aminocyclopropane-1-carboxylate synthase; ACO: 1-aminocyclopropane-1-carboxylate oxidase; Ade: adenine; MTA: methylthioadenosine. The atoms of SAM recycled to methionine through methionine cycle are marked in red and the atoms of methionine converted to ethylene are marked in bold. For details see text.

intermediate.[10] Oxygen is not incorporated in the products and is probably eliminated as water molecules after reduction. Moreover, carbon dioxide has been found to be a necessary activator for ACO reactivity both *in vivo* and *in vitro*, and plays a major role in protecting the enzyme from deactivation.[11] Cyanide is detoxified to β-cyanoalanine by β-cyanoalanine synthase and thereby its toxicity at high rates of ethylene biosynthesis is prevented.[12]

ACO isozymes are encoded by multigene families in various plant species.[13–15] In the systematic hierarchy, the occurrence of ACO was noted for the first time in Gnetaceae and Ephedraceae belonging to the gymnosperms.[16]

In spite of the often constitutive activity of ACO in the majority of plant tissues, an increase in its activity may regulate ethylene production especially associated with ripening and senescence of leaves, fruits, and flowers (see Sections 5.04.2.3 and 5.04.4.2.3, and **Figure 3**).

In addition to ethylene, ACC may also be converted to 1-malonyl-ACC (MACC) or glutamyl-ACC (GACC) by ACC *N*-malonyltransferase or γ-glutamyltranspeptidase.[17,18] The role of both conjugates is to decrease free ACC level in plant tissues. MACC represents the major conjugate of ACC in plant tissues, whereas GACC is a minor conjugate.[19] Under physiological conditions, this conversion is rather irreversible; however, there is some controversy with the existing data.[20]

5.04.2.2 1-Aminocyclopropane-1-Carboxylate Synthase, a Key Enzyme of Ethylene Biosynthesis

The ACS function is known only in higher plants. The activity of ACS isozymes is a key regulatory factor of ethylene biosynthesis pathway. In general, microorganisms liberate ethylene but their ethylene synthesis pathways do not involve ACC as an intermediate. *Penicillium citrinum* is the first reported microorganism that is able to synthesize ACC from SAM and to degrade it into ammonia and α-ketobutyrate, not to ethylene. ACS from *P. citrinum* shows a 100-fold higher K_m for SAM than its plant counterparts.[21,22]

It has been evidenced that auxin and ACC stimulate ethylene production in some lower plants, for example, in the moss *Funaria hygrometrica* and in the ferns *Pteridium aquilinum* and *Matteuccia struthiopteris*.[23,24] Moreover, ethylene synthesis via the ACC-dependent pathway and in the presence of endogenous ACC and its conjugate in marine unicellular *Acetabularia* algae was reported.[25] Probably during the evolution of land plants, a relatively primitive pathway of ethylene production was replaced by the ACC-dependent synthesis pathway that now predominates.

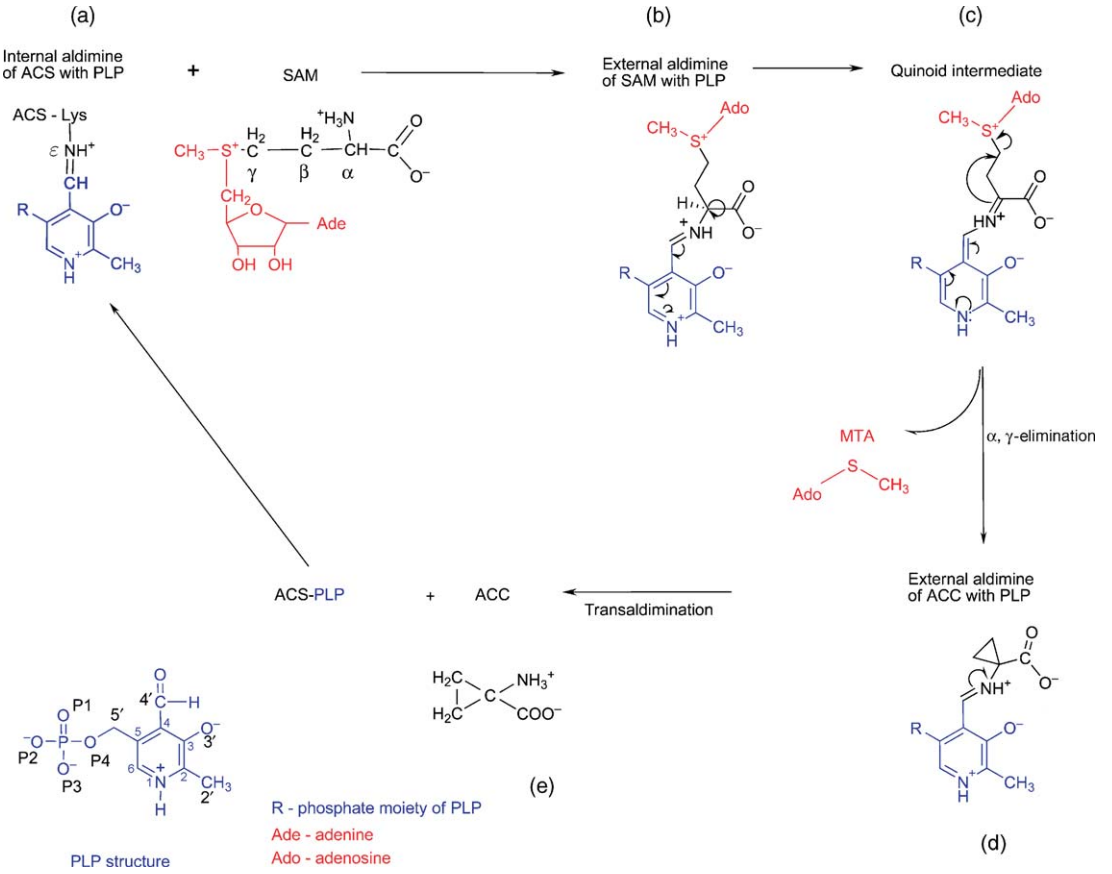
5.04.2.2.1 The reaction mechanism of the conversion of SAM to ACC

ACS isozyme utilizes pyridoxal-5′-phosphate (PLP) as a cofactor and belongs to fold type I PLP-dependent enzymes showing an absorption maximum between 422 and 431 nm, which is due to the internal aldimine.[26,27] The reaction mechanism proposed for the conversion of SAM to ACC by ACS illustrated in **Scheme 2** involves the following steps:

1. The active site lysine of ACS forms a Schiff base (internal aldimine) via its ε-amino group with the bound PLP in the unliganded enzyme (**Scheme 2(a)**).
2. The α-amino group of the substrate SAM replaces that of active site lysine as the Schiff base partner of the cofactor (external aldimine, **Scheme 2(b)**) and the C-α proton of SAM is next abstracted by the ε-NH2 function of active site lysine to form a quinoid intermediate (**Scheme 2(c)**).
3. Electrons are directed to C-γ, effecting the α, γ-elimination of MTA and formation of the ACC external aldimine (**Scheme 2(d)**).
4. ACC external aldimine undergoes transaldimination to release ACC and complete the catalytic sequence (**Scheme 2(e)**), according to McCarthy *et al.*[28]

Nevertheless, Li *et al.*[29] have proposed a model in which the Tyr152 residue (see Section 5.04.2.2.5) mediates the formation of ACC from SAM via a catalytic mechanism involving a quinoid intermediate.

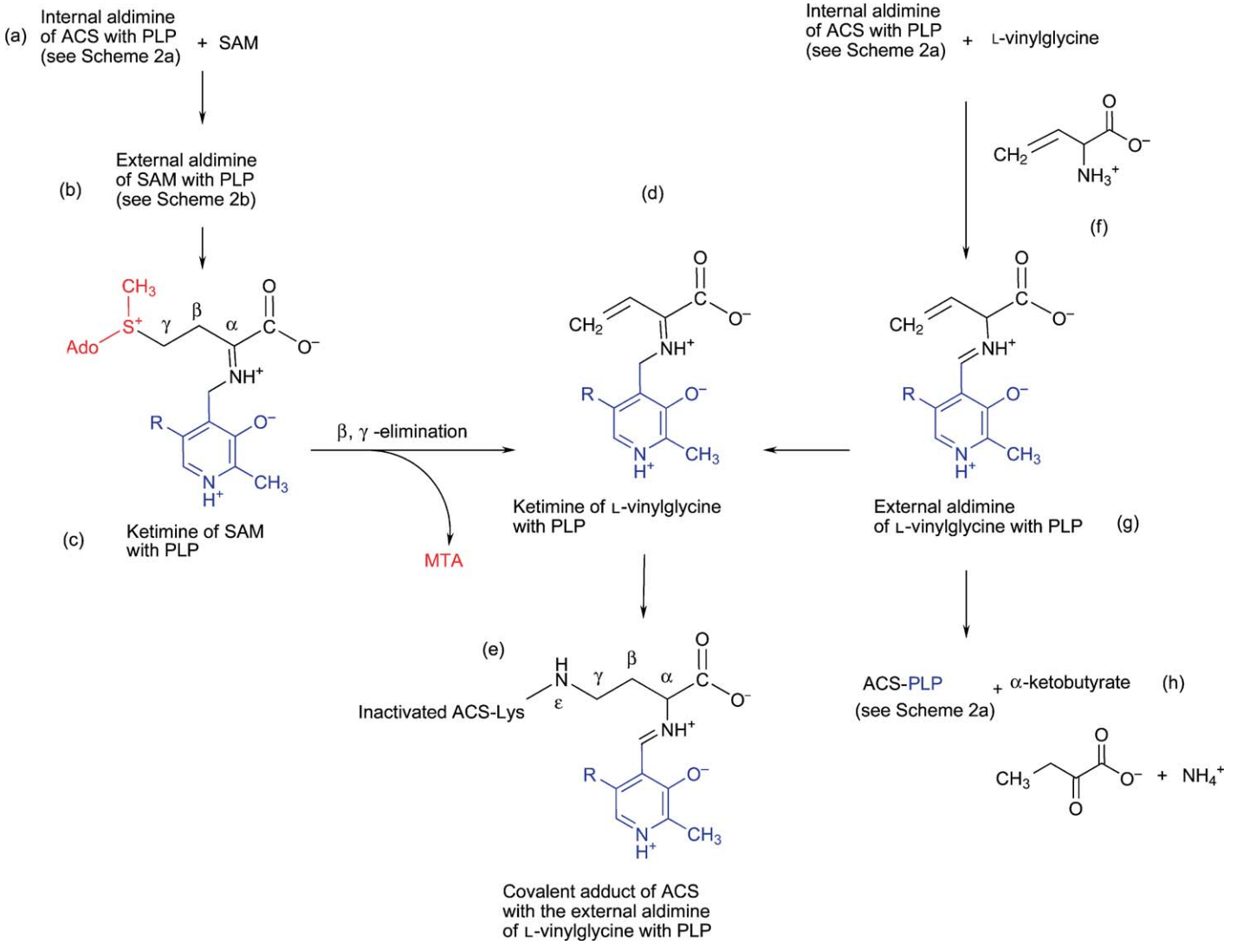
The members of the ACS isozyme family differ in their catalytical activity. The eight biochemically diverse catalytically active ACS isozymes from *Arabidopsis* show K_m values for SAM varying from 8.3 to 45 μmol; those



Scheme 2 Reaction mechanism of the conversion of SAM to ACC catalyzed by ACS. ACS-Lys denotes active site lysine of ACS, PLP denotes pyridoxal 5′-phosphate, and O_{P1}–O_{P4} denotes oxygens of the pyridoxal 5′-phosphate moiety. The atoms of SAM recycled to methionine are marked in red and the PLP atoms are marked in blue. Notations are as in **Scheme 1**. For details see text.

from tissues of tomato fruit, from 13 to 32.5 μmol; those from the hypocotyl and mesocarp of winter squash, from 12.1 to 13.3 μmol; and those from mung bean hypocotyls, 60 μmol.[30]

ACS is optimized to direct electrons from the quinonoid intermediate to the γ-carbon of its substrate, SAM (**Schemes 2(b)–2(d)**), to yield ACC and MTA. Nevertheless, ACS is inactivated by its substrate SAM in an irreversible manner (mechanism-based inactivation). Satoh and Yang[31] proposed that the mechanism-based inactivation of ACS by SAM proceeds through β, γ-elimination of MTA from SAM leading to the formation of L-vinylglycine (L-VG), which in turn would inactivate the enzyme (see **Schemes 3(a)–3(d)**). SAM exists as an *S,S*- or *R,S*-diastereomeric mixture (isomers with respect to its sulfonium center). A complete kinetic study carried out by McCarthy *et al.*[28] using enantiomerically pure (*S,S*)- and (*R,S*)-SAM as substrates has shown that (*R,S*)-SAM can undergo the same α, γ-elimination as the (*S,S*)-diastereomer, but it more often eliminates the same leaving group in a β, γ-process to produce a vinylglycine-related intermediate. The inactivation of ACS is a result of a nucleophilic attack by active site lysine on γ-carbon L-VG originating from the β, γ-elimination of MTA from the external SAM aldimine.[32,33] The mechanism-based inactivation of ACS is shown in **Schemes 3(a)–3(e)**. The crystal structure of the covalent adduct of the inactivated apple ACS was determined.[33] The active site contains an external aldimine of the adduct of L-VG with PLP. The side chain γ-carbon of L-VG is covalently bound to the ε-amino group of active site lysine (see **Scheme 3(e)**).[33] *In vivo* formation of L-VG from SAM leads to alkylation and inactivation of the enzyme at about once in 30 000 turnovers, and thus probably plays some role in decreasing the enzymatic activity in plant tissues.



Scheme 3 Model of mechanism-based inactivation of ACS by SAM: (a–e) Model of mechanism-based inactivation of ACS by L-vinylglycine (L-VG): (f-g-d-e) Conversion of L-VG to α-ketobutyrate and ammonia catalyzed by ACS: (f–h). Notations are as in **Scheme 2**. For details see text.

5.04.2.2.2 The other catalytical activities of ACS

Two compounds other than the natural substrate SAM, L-VG and *S*-methyl-L-methionine (SMM), have been described so far as both substrates and inhibitors of ACS isozymes. L-VG was isolated 30 years ago from the fungus *Rhodophyllus nidorosus*.[34] It was shown to be a mechanism-based inhibitor of aspartate aminotransferase and kynurenine aminotransferase.[33] First of all, L-VG is an alternative substrate of ACS in addition to being an inhibitor as described in the previous section.[32,33]

ACS converts L-VG mainly to α-ketobutyrate and ammonia (see **Schemes 3(f)–3(h)**). The enzyme–L-VG complex partitions to products 500 times for every inactivation event. A high molar ratio of L-VG/ACS (over 5000) is necessary for the complete inactivation of ACS[33] (see **Schemes** 3(f), 3(g), 3(d), and 3(e)).

SMM is a ubiquitous constituent of the free amino acid pool in flowering plants, occurring in leaves, roots, and other organs at levels that typically range from 0.5 to 3 mmol per 1 g of dry weight, a concentration that is often higher than that of Met or SAM.[35] SMM differs from the natural substrate SAM in a methyl group replacing the 5′-deoxyadenosyl moiety linked to the sulfur atom. The major reaction of ACS with SMM is β, γ-elimination of dimethylsulfide to yield enzyme-bound L-VG, which is subsequently converted to α-ketobutyrate and ammonia.[36] The inactivation mechanism of ACS by SMM takes place through a L-VG ketimine intermediate and leads to the same L-VG-labeled species as reported for SAM (see **Scheme 3(e)**). Likewise, SMM can be transaminated to yield 4-dimethylsulfonium-α-ketobutyrate.[36] The addition of excess of pyruvate at the end of the reaction to convert pyridoxalamine phosphate (PMP) form of ACS back to the PLP form results in about 45% recovery of the catalytical activity of ACS, which implies some non-α-ketoglutarate-coupled aminotransferase activity of ACS.[36] ACS turns over SMM about 0.5% as rapidly as it turns over SAM.[36] As no α, γ-elimination was detected for SMM, the adenosyl moiety of SAM must be utilized to force the proper alignment of the α- and γ-carbon atoms of SAM to form the strained transition state leading to the cyclopropane ring during physiologically preferred reaction.

The investigation of the aminotransferase activity of apple ACS carried out by Feng *et al.*[37] reveals that it is able to reductively aminate PLP to PMP by transamination of some L-amino acids to their corresponding α-keto acids. The enzyme has shown substrate specificity with the preference of Ala > Arg > Phe > Asp. The addition of excess pyruvate causes a conversion of the PMP form of the enzyme back to the PLP form. The quite unstable PMP form of ACS can generate apoenzyme, which captures PLP to restore its physiologically active form.

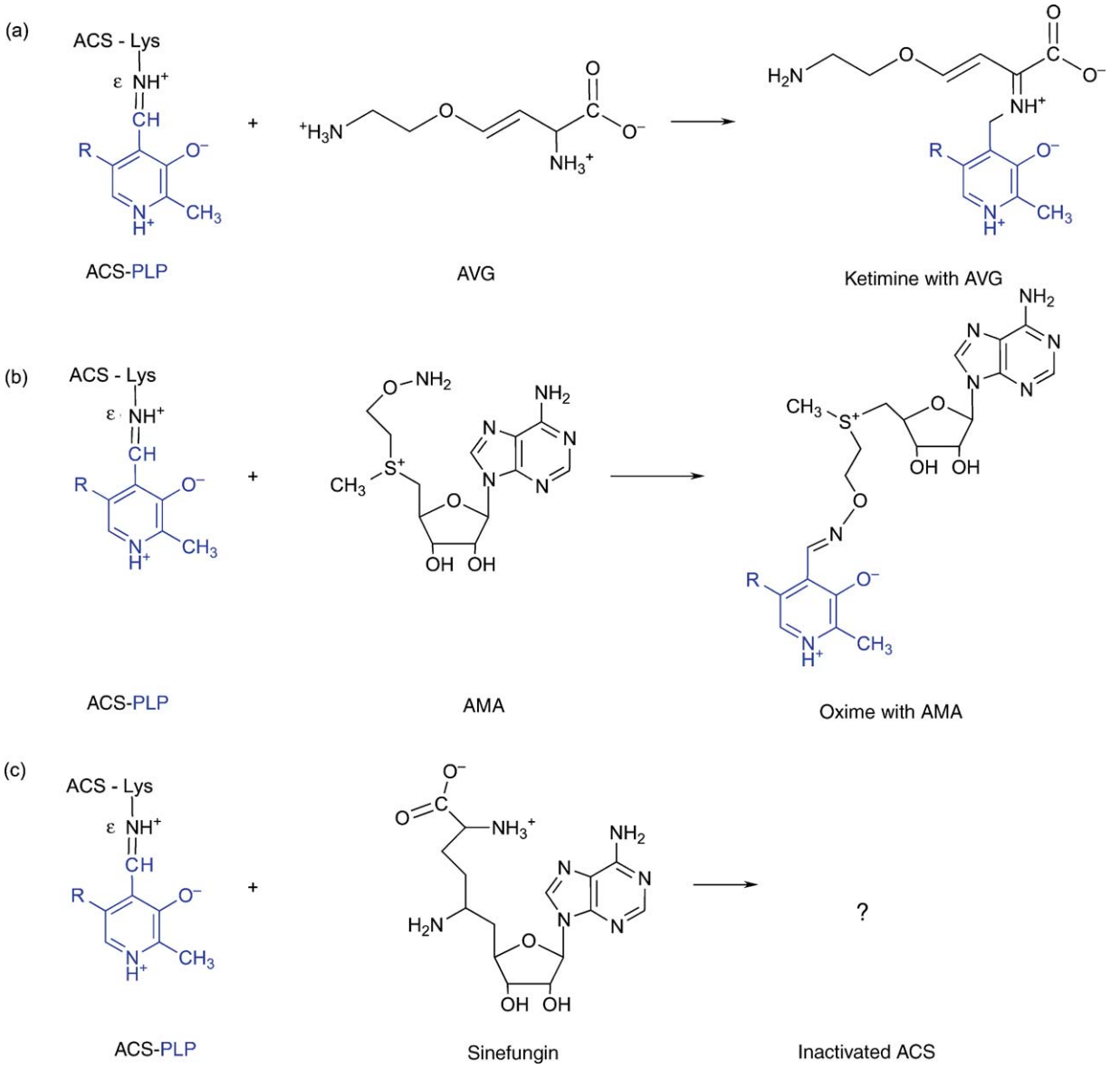
5.04.2.2.3 Inhibitors of the catalytical activity of ACS

ACS activity may be reversibly regulated by various substances associated with the methionine-recycling pathway, SAM metabolism, and polyamine synthesis, and by natural and chemical analogues of SAM or inhibitors of PLP-dependent enzymes.

At physiological pH, polyamines are fully protonated and polycationic. It is possible that the amino group of amines prevents SAM from reacting with ACS by interfering with the prosthetic groups of the enzyme. Polyamines such as putrescine, spermidine, and spermine, and methionine cycle intermediates such as MTA and KMB have been shown to have an inhibitory effect on ACS activity.[38]

The analogues of SAM that have an inhibitory effect on ACS activity can be modified in the adenine moiety and in the 5′ substituent. As follows from the studies of the inhibitory action on ACS from tomato of sinefungin (a naturally occurring antifungal antibiotic isolated from *Streptomyces griseus*), *S*-adenosylhomocysteine (SAHC), *S*-isobutyladenosine (SIBA), 3-deazaadenosylhomocysteine (3dz-SAHC), *S*-isobutyl-1-deazaadenosine (1-dzSIBA), *S*-isobutyl-3-deazaadenosine (3-dzSIBA), and *S*-isobutyl-7-deazaadenosine (7-dzSIBA), the 7-nitrogen of adenine is apparently necessary for the inhibitory activity although the 5′ substituent containing an amino group can have some inhibitory role.[39,40]

AMA ([2-(amino-oxy)ethyl](5′-deoxyadenosin-5′-yl)(methyl)sulfonium) represents an aminooxy analogue of SAM. As shown for the crystal structure of the apple ACS with AMA, AMA covalently binds to the PLP cofactor through an oxime bond but is unreactive because of the lack of an α-proton equivalent (see **Scheme 4(b)**), according to Capitani *et al.*[26] PLP–AMA adduct is linked to the inner active site near the ACS dimer interface. In the apple ACS Tyr85* residue, the neighboring subunit acts as a platform for the adenine ring of AMA and binds it through a stacking interaction.[26] Asp84* interacts with amino group of the adenine of AMA for details on the interaction between two monomers in the dimeric structure of ACS and the residues critical for catalysis, see Sections 5.04.2.2.4 and 5.04.2.2.5, respectively.



Scheme 4 Inactivation of ACS through reaction of its cofactor PLP with the inhibitors: (a) PLP forms a ketimine with aminoethoxyvinylglycine (AVG); (b) PLP forms oxime with AMA, the aminooxy analogue of SAM; (c) ACS inactivated by antifungal antibiotic sinefungin, the analogue of SAM. In this case, the structure of inhibitory adduct is still unknown (?). PLP is marked in blue. Notations are as in **Schemes 2** and **3**.

The hydroxylamine or vinylglycine analogues are potent inhibitors of PLP-dependent enzymes *in vivo* and *in vitro*. The hydroxylamine analogues inactivate the enzyme forming stable oximes with PLP.

The vinylglycine analogue, rhizobitoxine (2-amino-4-(2-amino-3-hydropropoxy)-*trans*-but-enoic acid), is synthesized by *Bradyrhizobium elkanii*, a legume symbiont, and *Burkholderia andropogonis*, a plant pathogen with a broad host range.[41–43] It inhibits β-cystathionase and ACS in the methionine and ethylene biosynthetic pathways, respectively. Both symbiotic and pathogenic bacteria produce ACS inhibitors that are structurally similar to rhizobitoxine so as to control the ethylene-induced plant response that would prevent a successful infection.

The unsaturated enol ether amino acid aminoethoxyvinylglycine (AVG), from a *Streptomyces* strain, and its methoxy analogue methoxyethoxyvinylglycine (MVG), from *Pseudomonas aeruginosa*, inhibit ethylene synthesis in higher plants.[44] AVG has the strongest inhibitory effect on ACS among the vinylglycine analogues tested.[6] Huai *et al.*[45] reported noncovalent complex of AVG with tomato ACS. According to them, AVG binds in close proximity to PLP. The α-carboxylate group of AVG forms the following three hydrogen bonds: with the backbone nitrogen of Ala54 and the guanidine nitrogen of Arg412 of tomato ACS, and with the water molecule. The α-amino group of AVG sits about 4 Å away from C4′ of PLP. In this manner, both the α-amino and α-carboxylate groups of AVG simulate binding of a substrate, SAM (see Section 5.04.2.2.5). In contrast, Capitani *et al.*[46] reported for the apple ACS that AVG inhibits its activity by forming a covalent ketimine complex with PLP (see **Scheme 4(a)**). Such an AVG–PLP adduct is tightly bound in the active site of the apple ACS.

ACS isozymes differ from each other in their sensitivity to inhibitors. For example, the eight catalytically active ACS isozymes from *Arabidopsis* have K_i values for AVG and sinefungin ranging from 0.019 to 0.8 and 0.15 to 12 µmol, whereas those from the crude extract preparation of tomato fruit have K_i values for the same inhibitors ranging from 0.2 to 10 and 1.0 to 25 µmol, respectively.[30]

5.04.2.2.4 The spatial structure of ACS isozymes: Do they function in plants as monomers, homodimers, or even heterodimers?

ACS is located in the cytoplasm and makes 0.000 015–0.001% of plant soluble proteins.[47] An opportunity to explore ACS polypeptides is provided by the expression of functional ACS cDNAs in heterologous systems such as *Escherichia coli* and yeast.

Only the N- and C-terminally truncated form of tomato ACS (11–439 aa, full-length ACS 485 aa) and the C-terminally truncated form of apple ACS (52–48 kDa) have been crystallized to date.

Huai *et al.*[45] and Capitani *et al.*[26,33,46,48] have reported two crystal structures of ripening-related tomato ACS, the first one for unliganded ACS and the second one for ACS with AVG, and four crystal structures of ACS from apple, unliganded ACS, ACS with AVG, ACS with AMA, and ACS with L-VG, respectively. Their atomic coordinates and structure factors have been deposited with PDB, with entry codes 1IAX, 1IAY, 1B8G, 1M7Y, 1M4N, and 1YNU, respectively.

There are two domains of each monomer, a large one and a small one, defined by the tertiary structure. The large domain contains the strictly conserved secondary structures typical of the families of PLP-dependent enzymes, and the more variable small domain consists of amino and carboxy regions of the ACS polypeptide (see **Figure 1**). The active site with a bound PLP cofactor lies in the cleft formed by the two domains. Some of the residues making the active site involved in catalysis (Tyr92* and Tyr85* in LE-ACS2 and Md-ACS1, respectively) are delivered from the neighboring monomer. The spatial structures of tomato and apple ACS isozymes estimated on the basis of crystallographic data are similar. According to the model proposed for tomato unliganded ACS (1IAX), the large domain of the monomer consists of a central seven-stranded β-sheet flanked by α-helices. The small domain consists of β-strands flanked by five α-helices. A schematic representation of the secondary structures building the large and the small domains in this monomer is given in **Figure 1**. The monomer of ACS from apple similar to tomato ACS represents a two-domain polypeptide with each domain composed of a central sheet of β-strands connected by α-helices packed on both sides.

Except the crystallographic asymmetric unit containing one molecule of ACS with AVG (1IAY, 1M7Y), all other structures contain two molecules of ACS, which strongly suggests a functional dimeric form of these enzymes. In the dimeric form of both tomato and apple ACS, two independent active sites are formed at the interface of a dimer and share residues from each monomer. It has been well documented that the dimeric form is the basic catalytic unit of most PLP-dependent enzymes and that both monomers within a dimer physically share an active site.[49]

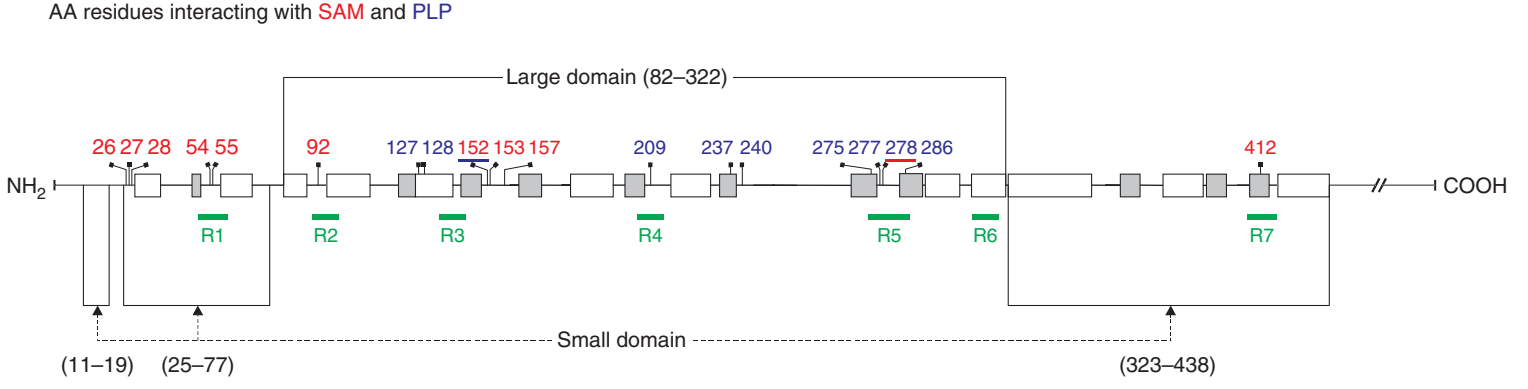


Figure 1 Schematic representation of tomato ACS polypeptide with marked α-helical and β-strand secondary structure regions (according to PDB with entry code 1IAX), residues critical for catalysis, and fragments of the polypeptide representing the large and the small domain in the spatial structure of enzyme. Open blocks denote α-helical regions and filled blocks, β-strand regions. For details see Sections 5.04.2.2.4 and 5.04.2.2.5.

Two different inactive mutants of ACS isozymes from tomato partially restore catalytic activity when they are coexpressed.[50] All ACS isozymes from *Arabidopsis* can potentially form 45 homo- or heterodimers. A series of experiments carried out in heterologous systems showed that 25 of them are functional (8 homodimers and 17 heterodimers); however, enzymatically active heterodimers are formed only among isozymes that belong to one or the other of the two phylogenetic branches (see Section 5.04.2.2.6). Probably, the shared active sites formed between the heterodimeric subunits of the same branch are structurally similar to those of the corresponding homodimers. An exception to this rule is ACS7, which is able to form functional heterodimers with some members of both branches.[51] The inactivity of certain heterodimers is probably due to structural restraint(s) that prevents the shared active sites to be functional. If indeed ACS heterodimers are formed in plants, they could provide an additional regulatory element in ethylene production. Nevertheless, some authors suggest that a monomeric form of ACS could be sufficient for catalysis.[27,45,52]

5.04.2.2.5 Specific features of ACS polypeptides and the residues critical for function

Members of the PLP-dependent family are found in four out of the six EC classes of enzymes but are characterized by significant conservation of the PLP-binding residues. It is not known whether they represent convergent evolution or have evolved divergently from a common origin. The structural alignments indicate that the plant ACS isozymes are most closely evolutionarily related to subgroup I of aminotransferases.[37,53] All but one of the only 12 amino acid residues completely conserved among aminotransferases are present at equivalent positions in all ACS. They are listed and numbered according to the LE-ACS2 from tomato (or Md-ACS1 from apple) as follows: Tyr92 (85), Pro151 (143), Asn209 (202), Pro210 (203), Gly212 (205), Asp237 (230), Tyr240 (233), Lys278 (273), Arg286 (281), Gly288 (283), and Arg412 (407).[54,55]

The plant ACS polypeptides have been divided by a majority of the authors into types 1, 2, and 3 (see **Figure 2**). All of them are characterized by seven conservative regions denoted in **Figure 1** as R1–R7, which comprise the following residues (according to LE-ACS2): R1 (49–58), R2 (89–97), R3 (136–142), R4 (205–213), R5 (274–288), R6 (304–311), and R7 (407–417). A necessary feature of the ACS polypeptide is the presence of a PLP-binding site (R5) containing a lysine residue forming the internal aldimine with C4′ of PLP (see Section 5.04.2.2.1). The other residues critical for the function have been discovered mainly using the site-directed mutagenesis method and crystallographic studies on the apple and tomato ACS (Md-ACS1, LE-ACS2).[26,28,29,33,45,46,48,51,56–59] The localization of these residues is illustrated in **Figure 1** and all of them are numbered according to LE-ACS2 (in parentheses the equivalent amino acid residue in Md-ACS1) and are characterized as follows:

Lys278 (273), lysine residue forming a Schiff base (internal aldimine) with the bound PLP in the unliganded enzyme, involved in α, γ-elimination (see **Scheme 2** and Section 5.04.2.2.1);

Asn209 (202), Tyr240 (233), and Asp237 (230), amino acid residues forming hydrogen bonds with O3′ of the pyridine ring and N1 of PLP;

Ala127 (120), Thr128 (121), Ser275 (270), Ser277 (272), and Arg286 (281), amino acid residues required for the correct orientation of PLP in the active site, and all of them interact with the oxygens of PLP phosphate moiety (see **Scheme 2**);

Tyr152 (145), mediates the formation of ACC from SAM via a catalytic mechanism involving a quinoid intermediate (according to Li *et al.*[29]) and stabilizes the pyridine ring of PLP;

Glu55 (47), putative ionic interaction between SAM and this residue;

Ala54 (46) and Arg412 (407), the α-carboxylate group of SAM forms hydrogen bonds with the nitrogen of alanine and the guanidino moiety of arginine;

Arg157 (150), O2′ and O3′ atoms of the sugar of SAM form hydrogen bonds with the guanidine group of arginine;

Pro26 (Ser18), Tyr27 (19), Phe28 (20), and Pro153 (144), residues contributing to the hydrophobic pocket for the adenine ring of SAM;

Tyr92 (85), probably acts as a platform for the adenine ring of SAM and binds it through a stacking interaction.

About a half of ACS isozymes of type 1 and those of type 3 have Pro26, whereas the remaining ones most often have Ser26 or Ala26. About one-third of ACS isozymes of type 2 have Ser127. Several ACS isozymes of type 1 have Ala153. The catalytical status of the fourth, poorly recognized group of ACS isozymes (in **Figure 2** type

AT-like) is discussed. However, it seems that a majority of them are capable of proper interaction with the adenine ring, O2′ and O3′ of sugar, and the α-carboxylate group of SAM. A majority of the residues of the nine categories listed above that interact with PLP occur in the AT-like ACS isozymes.

5.04.2.2.6 Comparative analysis of the presently found much diversified isoforms of ACS from higher plants

ACS isozymes in higher plants are represented by highly diversified isoforms divided into the well-characterized types 1, 2, and 3, and by a group denoted by the authors of this chapter as AT-like ACS isozymes (aminotransferase-like) because two members of this group, ACS10 and 12, from *Arabidopsis* have been classified as aminotransferases.[30]

Savolainen *et al.*[60] proposed a division of the angiosperms into two major groups: the eudicots (particularly asterids and rosids) and the noneudicots (monocots and Laurales, Magnoliales, Piperales, Ceratophyllales, Amborellaceae, Nymphaceae, Liliaceae). ACS isozymes used in the analysis are presented in the order proposed by molecular systematicians.

The analysis of 73 full-length ACS polypeptides representing 46 ACS isozymes from eudicots, 19 ACS isozymes from noneudicots, 7 ACS isozymes from gymnosperms, and 1 ACS-like protein from moss was performed using the neighbor-joining method and the bootstrap test with 1000 replicates.[61,62] The evolutionary distances were computed using the Poisson correction method and are in the units of number of amino acid substitutions per site.[63] Phylogenetic analyses were conducted using MEGA4 software.[64]

The tree of ACS isozymes from various plant species constructed on the basis of ACS polypeptides implies that their divergence had taken place before the eudicots and noneudicots separation. Moreover, the ACS polypeptides from conifers representing gymnosperms correspond to a distinct cluster, thus indicating that the present gymnosperms and angiosperms have diverged significantly over time since the separation of these major plant lineages. Nevertheless, the topology of the ACS tree implies a common ancestral ACS isozyme(s) for both of them.

Except the aminotransferase activity of AT-ACS10 and 12 isozymes, shown in heterologous nonplant systems, the catalytical status of a majority of members of AT-like enzymes is still unknown. The only exception is RH-ACS, whose expression has been evidenced to increase dramatically and correlate with ethylene levels in senescing petals of *Rosa hybrida*.[65] Moreover, the investigation carried out by Barnes *et al.*[66] on Pta-ACS1 from conifers suggests the putative ACS activity of this isozyme.

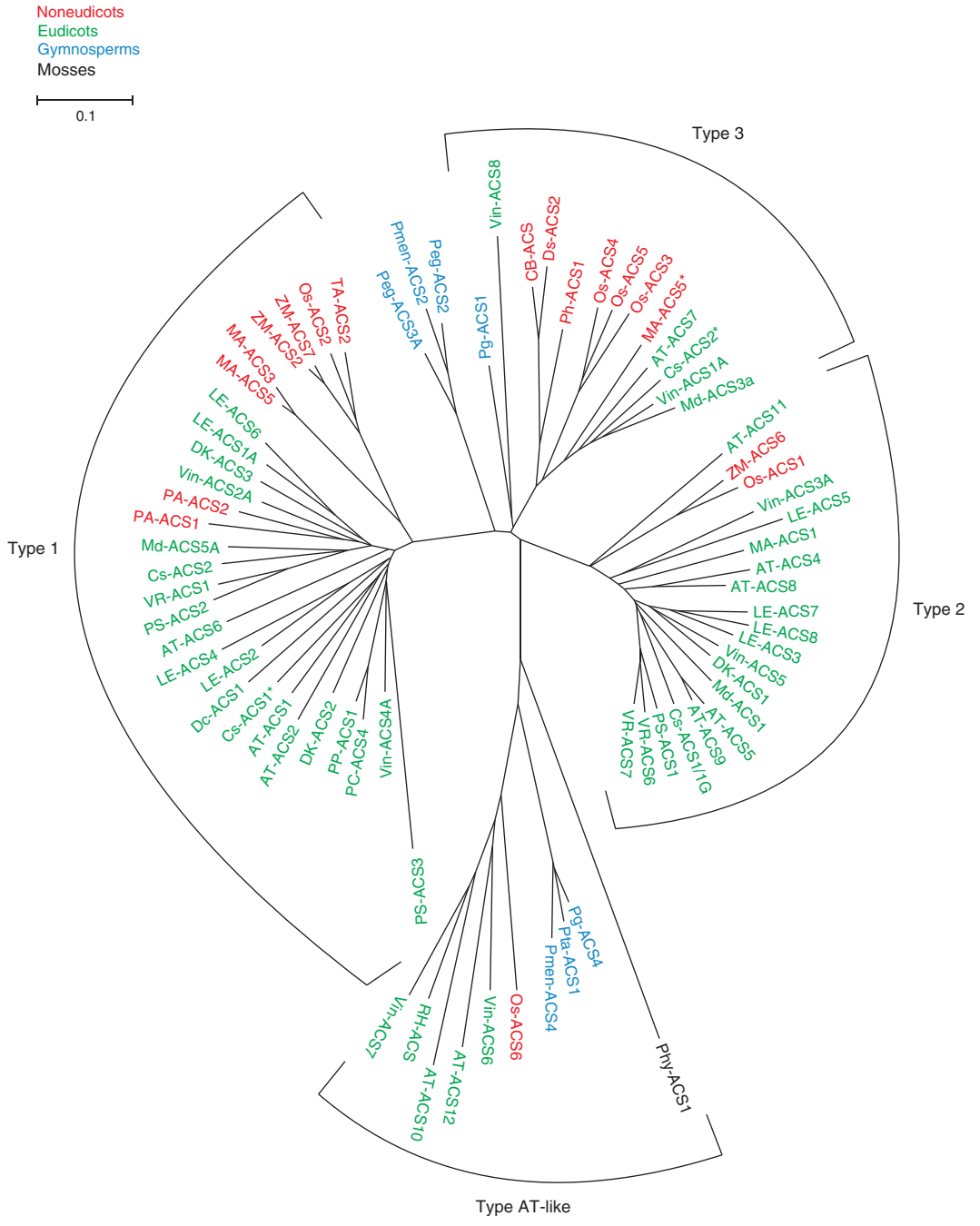
Distantly related possible homologues of plant ACS were found in the genomes of green algae, mosses, fungi, and several vertebrates including humans, according to NCBI database.[21,22,67,68] The function of the encoded protein has not yet been resolved.

5.04.2.2.7 Different types of ACS isozymes and control of their protein turnover

In addition to kinetic properties of ACS isozymes, their apparent half-life also controls ethylene production. The regulatory role of phosphorylation by calcium-dependent protein kinases (CDPKs) in ACS protein turnover has been considered for a long time. Some indirect evidence was provided by Grosskopf *et al.*,[69] Felix *et al.*,[70] Spanu *et al.*,[71] and Petruzzelli *et al.*[72] from the study of tomato cell culture, pea, and mung bean seedlings. To date, two phosphorylated regions in the C-terminus of ACS have been described: the first one represents the CDPK motif Arg-X-Ser-X and the second one the C-terminal-specific Ser residues phosphorylated by mitogen-activated protein kinase 6 (MAPK6, *Arabidopsis*) or by its orthologues in other plant species.[47,73,74]

The MAPK cascades, which are conserved signaling modules in eukaryotes, are composed of three protein kinase modules, MAPKK kinases (MAPKKKs), MAPK kinases (MAPKKs), and MAPKs, which are linked in various ways to the upstream receptors and downstream targets.[75] In the C-termini of type 1 ACS isozymes from various species, the number of putative MAPK-recognized sites is evolutionarily conserved, with most having three phosphorylation sites and a few having four sites with the last phosphorylated Ser residue situated about 4–10 amino acid residues from the last C-terminal residue.[47] In *Arabidopsis*, MAPK6 directly phosphorylates three Ser residues in the C-terminal region of ACS2 and 6 isozymes. Phosphorylation of these Ser residues slows down ACS6 protein degradation. As a consequence, ACS2 and 6 are accumulated and ethylene production increases without any alterations in the kinetic properties of the enzymes. Only the phosphorylation

of multiple C-terminal Ser residues in MAPK6-recognized motifs stabilizes ACS protein.[47] ACS2 and 6 isozymes in which these three Ser residues have been changed to three negatively charged Asp residues mimic the phosphorylated form of ACS2 and 6 and show delayed degradation. The apparent half-life of the modified ACS6 isoform having Ser residues changed to Ala residues is less than 10 min, whereas the half-life of phospho-mimicking ACS6 with Asp residues is prolonged to more than 3 h.[76] The degradation machinery targets the C-terminus of ACS6 but the negative charges of the phosphate groups introduced by MAPK6



reduce its recognition. It is also possible that phosphorylation favors phosphorylation-dependent binding of a protein that protects the ACS polypeptide from being recognized by proteasome components. The unphosphorylated ACS isozymes are rapidly degraded via the 26S proteasome pathway. Deletion of the 16 amino acid length C-terminal fragment of ACS6 polypeptide fully stabilizes the enzyme and confirms that the C-terminus contains the targeting signal for degradation.[76]

The first direct evidence that the CDPK motif Arg-X-Ser-X present in the C-terminal region of tomato ACS2 isozyme has been phosphorylated by an unidentified CDPK was provided by Tatsuki and Mori.[73] Such CDPK motifs are present in both type 1 and 2 ACS isozymes. In contrast to ACS isozymes of type 1 possessing both CDPK- and MAPK6-recognized motifs, the ACS isozymes of type 2 lack MAPK6-recognized motifs. The phosphorylation of ACS by CDPK seemed to help the accumulation of ACS proteins but has almost no effect on enzyme activity. In conclusion, ACS isozymes of type 1 are under the regulation of two kinase pathways, CDPK and MAPK6 cascades, which enable control of the ACS proteins turnover in response to environmental and developmental stimuli.

Further knowledge about the turnover of ACS proteins derives from studies on *Arabidopsis* mutants, that is, the so-called ethylene overproducers, eto1, eto2, and eto3. They produce 10- to 100-fold more ethylene than wild-type plants.[77] The recessive eto1 mutation strongly increases basic ethylene production, especially in etiolated seedlings.[78] In *Arabidopsis*, ETO1 gene and its paralogues (EOL1 and 2) encode Eto1 protein regulating the turnover of ACS5 and 9, and possibly all of type 2 ACS isozymes.[79–81] The C-terminal fragments of ACS5 and 9 polypeptides are necessary for the interaction with Eto1 protein. Eto1 acts as an adaptor that binds on one end to the substrate ACS isozyme and on the other end to E3 ligase components, which ubiquitinate ACS and direct it to proteasomal degradation.[82,83] Eto2 and eto3 mutants show a dominant missense mutation in the C-terminus of ACS5 and 9 isozymes, respectively.[84,85] The C-terminal mutations in Eto2-ACS5 and Eto3-ACS9 delay their Eto1-dependent degradation. Eto1 directly interacts with the wild-type ACS5 and 9 isozymes but not with the Eto2-ACS5 and Eto3-ACS9 isozymes.[79] It results in ethylene overproduction by the eto2 and eto3 mutant plants.[85,86]

In general, the constitutive interaction of Eto1 with the wild-type ACS isozymes of type 2 permits a rapid decline in ethylene production via directing the enzymes to the 26S proteasome-dependent protein degradation pathway.[79] Such mechanisms controlling the degradation of type 2 ACS isozymes function in other plant species, for example, the 14 amino acid length C-terminal fragment of type 2 ACS3 isozyme of tomato, are sufficient to target fusion protein for degradation.[81]

The finding of proteasomal degradation of type 2 ACS isozymes via Eto1 interaction raises the question whether there exists a modification for blocking this process. The most promising candidate for such a modification is phosphorylation of C-terminus. Probably the phosphorylation of the Ser residue (underlined)

Figure 2 The polypeptides tree of selected 73 ACS isozymes from higher plants and one ACS-like protein from moss. ACS isozymes are shown with the three letter symbol ACS preceded by the initials of each species. Angiosperms, eudicots: *Arabidopsis thaliana*: AT-ACS1, 2, 4, 5, 6, 7, 8, 9, 10, 11, and 12 (AAB60312, AAB59298, AAC49037, AAA87292, AAN18056, AAG48754, AAG50090, AAG48755, AAK15546, AAG48768, and AAG54001); *Cucumis sativus*: Cs-ACS1G, 1, 1*, 2, and 2* (ABI33818, ABI33821, BAA93714, BAA93715, and DDBJ:D89732); *Dianthus caryophyllus*: Dc-ACS1 (AAA33275); *Diospyros kaki*: DK-ACS1, 2, and 3 (BAB89348, BAB89349, and BAB89350); *Lycopersicon esculentum*: LE-ACS1A, 2, 3, 4, 5, 6, 7, and 8 (AAB17278, CAA41856, AAB48945, CAA41857, AAK72430, AAK72433, AAK72432, and AAK72431); *Malus domestica*: Md-ACS1, 3A, and 5A (AAB68617, BAE94690, and BAA92350); *Pisum sativum*: PS-ACS1, 2, and 3 (AAD04198, AAD04199, and BAB33423); *Rosa hybrida*: RH-ACS (AAQ88100); *Vitis vinifera*: Vin-ACS1A, 2A, 3A, 4A, 5, 6, 7, and 8 (CAO65393, CAO65909, CAO61519, CAN69483, CAN74730, CAN66821, CAO63606, and CAO40486); *Vigna radiata*: VR-ACS1, 6, and 7 (CAA77688, BAA33859, and AAD41083). Angiosperms, noneudicots: *Cattleya bicolor*: CB-ACS (AAR88653); *Doritaneopsis* sp.: Ds-ACS2 (AAB05849); *Musa acuminata*: MA-ACS1, 3, 5, and 5* (AAQ13435, AAD28181, AAQ13436, and CAA11158); *Oryza sativa*: Os-ACS1, 2, 3, 4, 5, and 6 (NP 001051142, BAF15551, AAV59454, AAV44081, BAB12704, and BAA84790); *Persea americana*: PA-ACS1 and 2 (AAM21682 and AAM21683); *Phalaenopsis* sp.: Ph-ACS1 (CAB01401); *Triticum aestivum*: TA-ACS2 (AAB18416); *Zea mays*: Zm-ACS2, 6, and 7 (AAR25558, AAR2560, and AAR25559). Gymnosperms: *Picea engelmanii* × *glauca*: Peg-ACS2 and 3A (ABM60748 and ABM60749); *Picea glauca*: Pg-ACS1 and 4 (ABM60747 and ABM60751); *Pseudotsuga menzenii*: Pmen-ACS2 and 4 (ABM60752 and ABM60754); *Pinus taeda*: Pta-ACS1 (ABI93270). Mosses: *Physcomitrella patens*: Phy-ACS1 (EDQ56943). ACS polypeptides from *V. vinifera* have been found in NCBI database and the authors of this chapter denoted them with subsequent isozyme numbers.

in the C-terminal CDPK motif Arg-X-Ser-X prevents ACS of type 2 from interaction with Eto1 or Eto1 paralogues and delays ACS degradation. The presence of phosphate group plays a critical role by blocking the protein–protein interaction between ACS of type 2 and Eto1. The Eto2-ACS5 isozyme represents the truncated form of ACS5 isozyme, with deletion of the C-terminal 12 residues, and thus Eto1 is unable to interact with Eto2-ACS5. The Eto3-ACS9 isozyme overproducing ethylene has Asp residue (underlined) instead of Val residue (Val residue occurs in wild-type ACS9, Arg-Val-Ser-X) in its C-terminal CDPK motif Arg-Asp-Ser-X. The change from Val residue to negatively charged Asp residue mimics the permanent presence of phosphate group, protecting Eto3-ACS9 isozyme from interaction with Eto1 and degradation.

The regulatory roles of 14-3-3 proteins are realized by their binding to the phosphothreonine or phosphoserine motifs in the targets.[87] Yao *et al.*[88] suggested that the phosphorylated Arg-X-Ser-X motif in the C-termini of ACS of types 1 and 2 is a potential binding site of 14-3-3 proteins *in vivo* and thus 14-3-3 proteins can play some role in the protection of ACS polypeptides from proteasomal degradation.

In etiolated seedlings of *Arabidopsis*, low doses of cytokinin (0.5–10 mmol l^{-1}) stimulate ethylene synthesis. It has been shown that cytokinin decreases the rapid turnover of the wild-type ACS5.[84,85,89,90] The cytokinin probably acts both by blocking the interaction of the C-terminus with Eto1 and by stabilizing ACS5 isozyme in an additional till now unknown manner.

ACS polypeptides prepared from plant tissues are mixtures of proteins of very similar molecular masses and differ from each other in the length of the C-truncated fragments. These values for *Arabidopsis*, winter squash, and tomato ACS isozymes are lower by 2–8 kDa than those calculated from the corresponding cDNA. The nature of the C-terminus of ACS makes it susceptible to cleavage.[30,52] The processing seems to occur invariably when the tissue is wounded or when cells are homogenized. Li *et al.*[91] reported a metalloprotease as the protease in the plant protein extract that removes the C-terminus of ACS. The possibility that this protease functions in specific organ, tissue, or cell types, delivering a mechanism for the cells to produce sustained high levels of ethylene, cannot be excluded.

The ACS of type 3 comprises many isozymes of tissue-specific activity restricted to generative organs. A majority of them have very short C-termini, making it impossible to control their protein turnover by the mechanisms described above. This feature probably proved physiologically favorable and was evolutionarily conserved. A consequence of this change is a significant decrease in the isoelectric point of type 3 ACS isozymes relative to the neutral and basic character of a majority of types 1 and 2 ACS isozymes, respectively.[92]

5.04.2.2.8 Structure and expression of ACS genes

ACS is encoded by a multigene family in all eudicot and noneudicot species that have been investigated till now.[54,93–103] In the completely sequenced genome of *Arabidopsis*, nine ACS genes encoding eight catalytically active (ACS2, ACS4–9, ACS11) isozymes and one nonfunctional ACS1 isozyme were found.[30,104] Recently, the ACS gene family has been described in conifers.[66,105]

A typical ACS gene consists of four exons and three introns. Usually, the last exon consists of about two-thirds of the ACS coding region. Except for a few examples of tandemly clustered ACS genes, most of them have scattered genomic organization. The highly homologous tandemly organized genes arose only by duplication followed in some cases by inversion of an ancestral gene.[54,94,106] Some of the ACS genes lack the first or the second intron, or as in the family Cucurbitaceae or Vitaceae an additional fourth intron may appear. All genes encoding isozymes of type 1 except some genes from Cucurbitaceae or Vitaceae and two genes from other plants show a typical structure. More than half of the genes encoding isozymes of types 2 and 3 lack the first or the second intron.[92] The genes of AT-like isozymes are characterized by a typical structure with one exception reported for Pta-ACS1 from conifers, which consists of two exons and one intron.[66]

Ethylene is a critical regulator of plant growth, development, and senescence, and a mediator of adaptation responses to biotic and abiotic stresses, and thus a diverse group of biotic and abiotic factors (see Section 5.04.4) have been reported as modulators of ACS gene expression, and various regulatory elements were found in the promoter regions of different ACS genes.[54,94,101,107–114] Complex hormonal and developmental networks regulate the expression of different members of the multigene ACS family in a tissue-specific manner in all species investigated till now. A comparative exploration of the model ACS gene family from *Arabidopsis* reveals spatiotemporal coexpression among the various gene family members during plant development and under

various stresses. Unique as well as overlapping expression patterns of certain members of ACS family were observed.[115]

Usually, the expression of genes of types 2 and 3 is induced by one or two kinds of stimuli (very often by auxin and/or by wounding) or is limited to generative organs, whereas many genes of type 1 are characterized as multiresponsive.

Various plant species produce both the shorter and the longer transcripts for the same ACS isozyme. For example, organ-specific production of different length transcripts was observed in carnation flowers, winter squash, and in tomato leaves, fruits, or elicitor-treated tomato cell suspension.[54,112,116–118] An auxin-induced PS-ACS1 gene from pea produces two kinds of transcripts; usually, the shorter one encoding inactive ACS isozyme appears first.[119] In general, the role of alternative transcripts is unknown but the first suggestion is that the shorter transcript could titrate away a regulatory binding protein that would otherwise inhibit translation and the second suggestion is that the truncated inactive ACS polypeptide could modulate ACS activity forming dimers with active ACS.[119] Furthermore, as reported for elicitor-induced ethylene in tomato suspension cultures, the accumulation of processed ACS transcripts does not always correlate with ACS activity or ethylene production. Despite considerable amounts of ACS transcripts, the suspension produces only basic level of ethylene. In the elicitor-treated suspension, the sum of transcripts abundance of all the expressed ACS isozymes increases threefold, whereas ethylene production increases more than 100-fold.[93] In *Arabidopsis,* cytokinin stimulates the induction of ACS5 mRNA less than twofold but ethylene level rises eightfold upon cytokinin treatment[90] see Section 5.04.2.2.7.

Auxin induces ethylene production by triggering the transcriptional activity of ACS genes and the auxin-stimulated ethylene synthesis is considered as one of the best-known hormone interactions in plant biology.[6,120] The auxin/indole-3-acetic acid (Aux/IAA) genes encode extremely short-lived nuclear proteins functioning as transcriptional regulators. According to the currently accepted model, the Aux/IAA proteins repress the activity of auxin-response factor (ARF) transcription factors. An increase in the auxin level accelerates the degradation of Aux/IAA proteins resulting in derepression of ARF activity and numerous auxin-stimulated transcriptional responses.[121] In *Arabidopsis*, the transcriptional activity of all genes encoding the eight catalytically active ACS isozymes discussed above is auxin-regulated. The transcripts of all of them are induced by a short treatment with the eukaryotic protein synthesis inhibitor cycloheximide.[30,115] Thus, it has been suggested that all auxin-regulated ACS genes may be controlled by short-lived repressor molecules such as Aux/IAA proteins.[111,122] If auxin indeed regulates ACS expression in a manner suggested above, the capability of Aux/IAA-ARF signaling apparatus to sense some other signals responsible for transcriptional activation of ACS genes cannot be completely excluded.[115]

Potamogeton pectinatus, a widespread aquatic monocot, is the first reported example of a vascular plant whose growing shoots are constitutively incapable of synthesizing ethylene. Although *P. pectinatus* does not produce ethylene, its shoot tissues contain considerable quantities of endogenous ACC. The IAA-treated *P. pectinatus* plants show a large increase in ACC levels but without accompanying the production of ethylene.[123]

5.04.2.3 The Internal Feedback Regulation of Ethylene Production

Molecular data support the concept introduced by McMurchie *et al.*[124] that there are two systems that control ethylene synthesis in plants. The ethylene autoinhibitory system 1, responsible for basic ethylene production, operates during normal vegetative growth of plant, in nonclimacteric fruits and in immature climacteric fruits. System 2 responsible for the rapid increase in ethylene production, regulated by a positive feedback mechanism, usually operates in ripening climacteric fruits and in senescing ethylene-sensitive flowers (see Section 5.04.4.2.3). Moreover, functioning of both systems in individual fruit tissues can differ temporally and spatially. Some examples of regulation of ACS and ACO genes expression by ethylene are presented below.

The four different ACS isozyme transcripts, namely, LE-ACS1A, LE-ACS2, LE-ACS4, and LE-ACS6, were detected in tomato fruit. LE-ACS1A and LE-ACS6 genes belong to system 1 but LE-ACS2 gene belongs to system 2. Each of these genes shows a distinct regulation. System 1, developmentally regulated through LE-ACS1A and LE-ACS6 genes expression, continues to act throughout fruit development until a competence to ripen is achieved. Then a transition period occurs during which the activation of RIN gene leads to an increased expression of LE-ACS1A and induction of LE-ACS4 activity. An increased expression of LE-ACSIA

and LE-ACS4 results in an increased ethylene synthesis, which decreases LE-ACS6 expression in a negative feedback manner but stimulates autocatalytically regulated expression of LE-ACS2. The catalytical activity of LE-ACS2 results in intense ethylene production, which reduces LE-ACS1A expression in an autoinhibitory feedback manner.[125] Some discrepancies exist regarding ACS isozymes regulation during tomato fruit ripening proposed by different authors, but they may result from different tomato cultivars used.[125,126] In **Figure 3(a)**, a

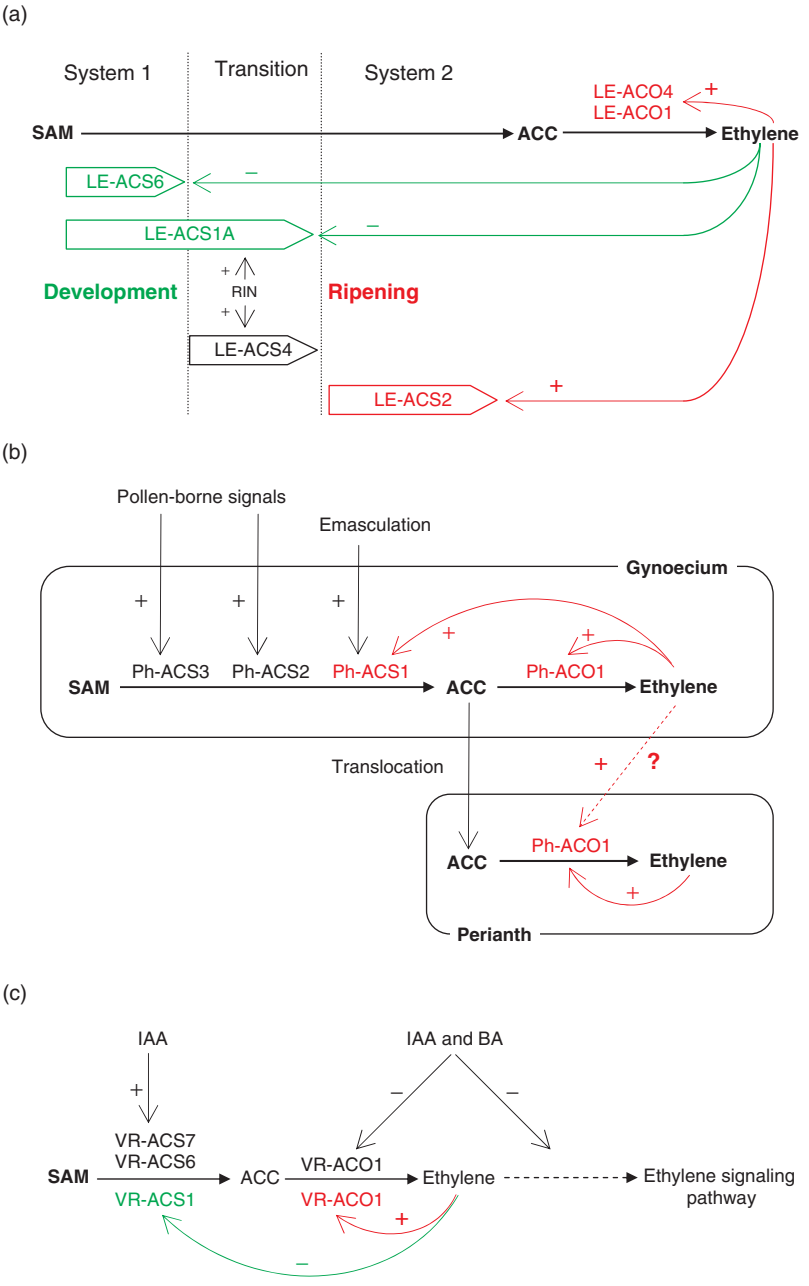


Figure 3 A model of the internal feedback regulation of ethylene production: (a) the regulation of LE-ACS and LE-ACO gene expression during the transition from system 1 to system 2 ethylene synthesis in tomato fruit; (b) the interorgan regulation of Ph-ACS and Ph-ACO gene expression in pollinated *Phalaenopsis* flower; (c) the regulation of VR-ACS and VR-ACO gene expression in etiolated *Vigna radiata* hypocotyls treated with auxin and cytokinin. For details see Section 5.04.2.3. '−' in green denotes negative regulation of ACS or ACO gene expression by ethylene; '+' in red denotes positive regulation of ACS or ACO gene expression by ethylene; '−' or '+' in black denotes negative or positive regulation of ACS or ACO gene expression by stimuli other than ethylene.

model of transition from system 1 to system 2 in ripening tomato fruit, according to Barry *et al.*[125] and Nakatsuka *et al.*,[126] is presented with respect to ACS and ACO expression, respectively.

In orchid flower (*Phalaenopsis* sp.), ethylene synthesis is regulated by the coordinated expression of three distinct ACS isozymes.[99,127,128] Pollination provides the primary signal that elicits expression of gynoecium-specific (stigma and ovary) Ph-ACS2 and 3 genes, the activity of which leads to the accumulation of ACC, which is converted to ethylene by Ph-ACO1. The production of ethylene regulates the expression of Ph-ACS1 and Ph-ACO1 genes in a positive feedback manner. A model of interorgan regulation of ethylene production in pollinated *Phalaenopsis* flower and regulation of these three ACS genes responding to primary and secondary pollination signals is illustrated in **Figure 3(b)** (see Section 5.04.4.2.3). Moreover, two of these genes, Ph-ACS2 and 3, respond to exogenous application of auxin in a manner mimicking pollination via induction of their expression in stigma and ovary, respectively.

In the mung bean (*Vigna radiata*) hypocotyls, auxin stimulates ethylene production by inducing VR-ACS6 and 7 genes expression in an ethylene-independent manner.[129] Synergistic effect of cytokinin (N6-benzyladenine (BA), a synthetic cytokinin) in the presence of auxin (IAA) on ethylene synthesis in the mung bean hypocotyls was observed. Nevertheless, the IAA and BA action inhibits ethylene signaling pathway resulting in the suppression of VR-ACO1 but induction of VR-ACS1 gene expression. The expressions of VR-ACO1 and VR-ACS1 genes are under positive and negative feedback control by ethylene, respectively. In this system, which separates the effect of IAA proper from the effect of IAA-induced ethylene, the amount of IAA-induced ethylene is enough to regulate VR-ACS1 and VR-ACO1 expression by feedback control. The complex regulation of ethylene biosynthetic genes expression by ethylene, IAA, and BA in the mung bean hypocotyls proposed by Kim *et al.*[129] is presented in **Figure 3(c)**.

5.04.3 Ethylene Signal Transduction Pathway

5.04.3.1 Ethylene Receptors as Negative Regulators Actively Suppressing the Expression of Ethylene-Inducible Genes in the Absence of Ethylene

In *Arabidopsis*, ethylene is perceived by a family of five receptors divided, on the basis of structural similarities, into ETR1-like subfamily 1 comprising ETR1 and ERS1, and ETR2-like subfamily 2 consisting of ETR2, ERS2, and EIN4.[130,131] Similar ethylene receptor subfamilies have been identified in several species.[132]

Ethylene receptors display a significant homology to a prokaryotic family of signal transducers known as two-component regulators. The signal perception by the N-terminal input domain of the sensor promotes autophosphorylation of conserved histidine residue; subsequently, phosphate is transferred from histidine to a conserved aspartate residue in the receiver domain of the cognate response regulator. The phosphorylation state of the receiver controls the attached output domain mediating downstream steps.[133] The receptors belonging to subfamily 1 have all conserved motifs necessary for histidine kinase activity, whereas the receptors from subfamily 2 except EIN4 lack a majority of these motifs. It was shown that all five receptor proteins except ETR1 autophosphorylate *in vitro* , and they predominantly autophosphorylate on serine residues.[134] Moreover, ERS1 and ERS2 lack the receiver domain. Nevertheless, the receptor mutant studies show that neither His kinase nor Ser/Thr kinase activity is essential for ethylene receptor signaling, which raises the possibility that these receptors function through a nonkinase mechanism to sense and transmit ethylene signal.[130]

Ethylene receptors are assumed to exist as disulfide-linked homodimers. In the absence of ethylene, the active plasma membrane-associated ethylene receptors suppress the expression of the ethylene-responding genes; thus, they act as negative regulators of ethylene signaling. Ethylene binding via a copper cofactor to hydrophobic pocket formed by the N-termini of dimer turns off the receptor activity and releases the expression of the ethylene-responding genes.[135] Therefore, reduced number of receptors increases tissue sensitivity to ethylene, whereas their increased number decreases tissue sensitivity to this hormone.[136,137] Ethylene responses, particularly those related to stresses, are often transitory. In order to shut down an ethylene response, synthesis of new receptors is essential. In tomato, receptor genes expression is constitutive throughout immature fruit development with high protein receptor levels in immature fruits. Ethylene binding triggers ubiquitin-dependent receptor protein degradation. If receptors are not replaced after ethylene-mediated

degradation, as occurs in immature fruit, the fruit will become more sensitive to subsequent ethylene exposure. The large increase in ethylene synthesis associated with climacteric fruit ripening depletes the receptor pool to a point at which ripening can proceed.[138,139]

A detailed analysis of ethylene receptor mutants has shown functional redundancy between them. However, in *Arabidopsis*, subfamily 1 members cannot be replaced by subfamily 2 members, in contrast to tomato in which functional redundancy between subfamily 1 and subfamily 2 members was reported.[140,141] It seems that there exists some degree of plasticity among different plant species within the ethylene signaling pathway.

In addition to ligand-mediated degradation of ethylene receptor proteins, ethylene affects the expression of various receptor genes in different ways. Some of them are ethylene inducible, whereas others are not affected by ethylene.[142,143] Such a diversity suggests the mechanism through which plants coordinate the ethylene responses.

5.04.3.2 The Other Downstream Signaling Elements Involved in the Ethylene Signal Transduction Pathway

The ethylene signal transduction pathway can be summarized as follows. Unoccupied, active receptors physically interact with a protein called constitutive triple response 1 (CTR1). CTR1 is a putative Ser/Thr MAPKKK, acting downstream of them. The active CTR1 represses further signaling. Ethylene binding to the receptor inhibits its activity and induces conformational change resulting in the inactivation of CTR1, which leads to ethylene signal transduction.[131,144] Further transduction of the ethylene signal requires the positive regulators EIN2 and EIN5 and the family of EIN3 transcription factors functioning downstream of CTR1.[130,145–148] The model of the ethylene signal transduction pathway is illustrated in **Figure 4**.

Only one CTR1 gene has been identified in the *Arabidopsis* genome. Mutations causing the loss of CTR1 activity result in the constitutive activation of all ethylene responses. CTR1 function depends on its N-terminal domain associated with endoplasmic reticulum (ER)-bound ethylene receptors and on its C-terminal Ser/Thr kinase activity. Probably, all five ethylene receptors are able to interact with CTR1 via their C-terminal kinase domains; however, subfamily 1 has a high affinity to CTR1, whereas subfamily 2 possesses a low affinity.[130] The existence of additional CTR genes, CTR 2, 3, and 4, has been evidenced in tomato.[149] The presence of multiple CTRs in plants raises many questions about the mode of transduction of the signal outputs from individual receptors. The interaction between various CTRs and the receptors, in combination with the varying ratio of receptors and different CTRs, may be a mechanism for optimization of ethylene responses in tomato and the other species with multiple CTR genes. Nevertheless, there are some data from the studies of mutants lacking CTR1 gene implying the existence of an additional branch for ethylene signaling, completely independent of the activity of CTR.[132] In this branch, ethylene may induce histidine kinase activity and autophosphorylation of ETR1. Phosphorylated ETR1 could then initiate a phosphorelay cascade involving histidine phosphotransfer protein, which ultimately results in the transcriptional regulation of ARR2 functioning as a transcription factor in the nucleus.[132]

EIN2 operating downstream of CTR1 is considered a central component in the ethylene signaling pathway and also it is assumed to be a common node mediating crosstalk of multiple hormone signal transduction pathways.[150–154]

In plants, EIN5(XRN4) exoribonuclease and its homologues are likely active in many pathways in which they degrade selected mRNAs. In general, the active EIN5(XRN4) exoribonuclease protects the next compound in ethylene signal transduction pathway, EIN3, from degradation.[145,146]

EIN3 and EIL1 (EIN3-like 1) are homodimeric transcription factors. In *Arabidopsis*, they mediate most, if not all, aspects of seedling growth responses to ethylene. EIN3 protein is synthesized constitutively and degraded rapidly by the proteasome in the absence of ethylene.[155] The more distantly related members of the EIN3 family (EIL2–5) play a minor role in the ethylene response or even function in pathways unrelated to ethylene.[130]

In *Arabidopsis*, the MAPKK4/5–MAPK6 cascade controls ethylene synthesis in the cytoplasm (see Sections 5.04.2.2.7 and 5.04.4.1), whereas the spatially distinct MAPKK9–MAPK3/6 cascade phosphorylates and stabilizes EIN3 in the nucleus.[148] The active MAPKK9–MAPK3/6 cascade positively regulates EIN3 levels and EIN3-mediated transcription (see **Figure 4**). The upstream localized CTR1 is a unique example of MAPKKK negatively stimulating the downstream MAPKK9–MAPK3/6 cascade by mechanisms unknown till today.

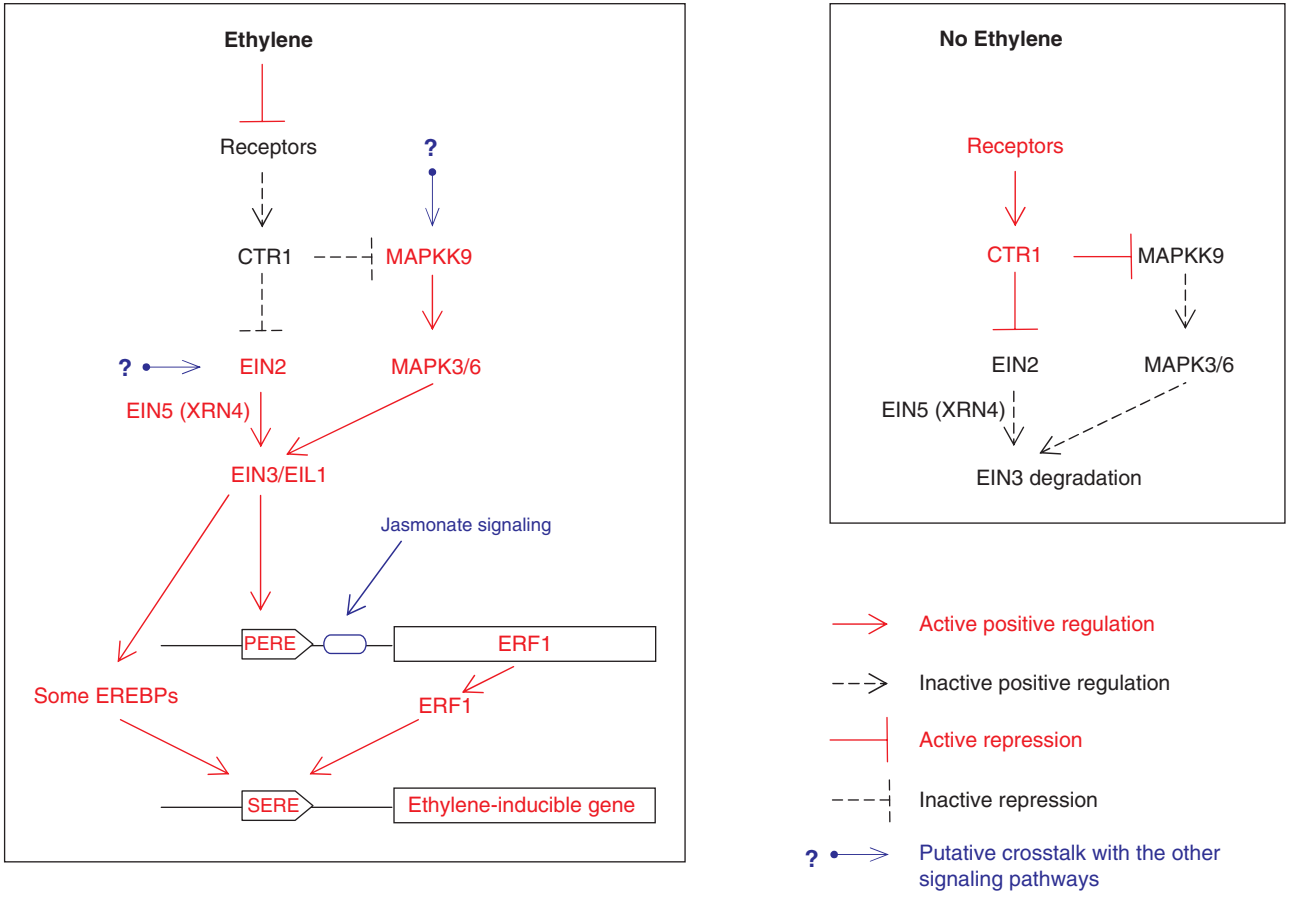


Figure 4 The ethylene signal transduction pathway. For details see Sections 5.04.3.1 and 5.04.3.2.

Furthermore, the MAPKK9–MAPK3/6 cascade can be modulated by signals other than ethylene in a CTR1-independent way.[148]

An immediate target for EIN3 is the primary ethylene-response element (PERE) in the promoter of the secondary transcription factor ethylene responsive factor 1 (ERF1) originally identified as the factor that binds to the secondary ethylene-response element (SERE), the so-called GCC box. The GCC box represents the motif present in the promoters of many ethylene- and pathogen-induced genes. ERF1 belongs to a large family of plant-specific ethylene-responsive element (GCC box)-binding transcription factors with 124 family members in *Arabidopsis*, initially referred as ethylene-response element-binding proteins (EREBPs), but later found to function in a diverse range of processes.[130] ERFs respond to extracellular signals to modulate GCC box-mediated gene expression positively or negatively. ERFs are upregulated via an ethylene-dependent pathway or an ethylene-independent pathway. In *Arabidopsis*, expression of ERF1 is directly regulated by the ethylene signal mediator EIN3, whereas expressions of other ERFs are partially regulated by ethylene signaling. ERF1, ERF2, and ERF5 activate a subset of GCC box-containing genes, whereas ERF3 and ERF4 repress the expression of these genes.[156,157] Some ERFs contain GCC boxes in their promoter, which implies that they could be targets for other members of the ERF family.[158]

5.04.4 Ethylene Action

In plants, one of the primary effects of ethylene is alteration of the expression of a broad spectrum of target genes, among which are ripening- and senescence-related and some defense-related genes such as jasmonic acid (JA)-, pathogen-, and wound-responsive genes.[159] In *Arabidopsis*, out of the 6000 genes investigated, 7% were ethylene-regulated.[160]

Phytohormones such as ethylene, salicylic acid (SA), JA, and abscisic acid (ABA) regulate responses of plants to stresses via action referred as signaling crosstalk. Moreover, reactive oxygen species (ROS), the toxic by-products of aerobic metabolism, play the important role of signaling molecules.[161,162] Usually, the defensive responses of plants depend on the interaction (positive or negative) between phytohormone signaling pathways rather than on the independent contribution of each of them.[163,164]

MAPK cascades (see Sections 5.04.2.2.7 and 5.04.3.2) modulate a myriad of internal and external signals to govern essential biological processes. They are activated by biotic and abiotic stresses, hormones, and cell division. In *Arabidopsis*, over 100 putative MAPK cascade genes have been reported and many of them are conserved in other plant species.[165–168] In *Arabidopsis* and tobacco, the activation of MAPKK4/MAPKK5–MAPK6 cascade and that of NtMEK2 (MAPKK)–SIPK/WIPK (MAPKs), respectively, increase ethylene synthesis[47,74] (see Section 5.04.2.2.7).

5.04.4.1 Ethylene as a Mediator of Adaptation Responses to a Multitude of External Stress Factors

In plants, DREB/ERF and AP2/ERF subfamily of ethylene-responsive element (GCC box)-binding transcription factors function in the regulation of abiotic stress responses and regulate disease resistance.[169] AP2/ERF members are downstream components of both ethylene and JA signaling pathways, and are key to the integration of both signals.[164,170] Overexpression of AP2/ERF family members may enhance disease resistance or confer tolerance to environmental stresses, for example, the constitutive expression of ERF1 in transgenic *Arabidopsis* confers resistance to several necrotrophic fungi and ERF1 overexpressed in transgenic tobacco leads to an increased salt tolerance.[170,171]

Expression of ERF3 and ERF4 from sugarcane and cotton, respectively, rapidly increases response to exogenous ethylene and ABA, salt, cold, and drought, whereas their overexpression in transgenic plants enhances tolerance to drought and osmotic stress.[172,173]

5.04.4.1.1 Ethylene and biotic stresses

Plants continuously challenged by pathogenic microorganisms possess efficient defense systems. The effective defense requires pathogen recognition followed by a rapid activation of plant defense gene expression. Ethylene

involvement in plant defense strongly depends on the type of pathogen and the plant species.[174] Ethylene-insensitive mutants show either increased susceptibility or increased resistance, which probably results from various infection mechanisms.[171,174–178] The discussion of ethylene involvement in the plant response to pathogens is complex because many pathogens produce ethylene through microbial-specific ethylene synthesis pathway different from the one used by plants.[179] Usually, an early burst of plant-derived ethylene is considered as resistance-associated one, whereas the pathogen-derived one seems to contribute to pathogen virulence.[180,181] Ethylene acts alone or with SA and JA in the activation of the hypersensitive response, which triggers a long-lasting one, known as systemic acquired resistance, to provide immunity against subsequent infections caused by a broad spectrum of pathogens.[182] Therefore, some *Arabidopsis* mutants defective in ethylene perception or JA signaling do not induce expression of subsets of pathogenesis-related genes and are more susceptible to some pathogens.[159]

The plant growth-promoting rhizobacteria produce microbial ACC deaminase, an enzyme hydrolyzing plant ACC, which decreases the ethylene-induced root growth inhibition during stress.[183] Nitrogen-fixing nodules formed on the roots of most legumes result from symbiotic interactions between compatible species of soil bacteria generically called rhizobia and their legumes host. Ethylene involved in plant defense responses during rhizobium infection negatively affects the nodules development and regulates their maturation; thus, plants defective in ethylene signaling show an increased number of nodules.[184,185] Moreover, some rhizobia produce rhizobitoxine, an inhibitor of plant ACS (see Section 5.04.2.2.3), and in this way they decrease ethylene production in plants.[41,186]

In contrast, certain pathogens influence ethylene production involved in plant defense for their own benefit.[187] Several of them are able to upregulate ethylene synthesis in plants as shown, for example, in virus- or bacteria-infected tobacco and citrus, respectively.[188,189]

Plants respond to the mechanical or insect herbivore damage of their tissues.[115,190,191] During herbivorous attacks, some plants emit a specific blend of volatiles, which may result in defense responses retarding development of the herbivores or attraction of herbivore enemies to feed upon them.[192] In lima bean leaves, the spider mite-induced volatiles, as well as infestation and artificial wounding, activate the ethylene and JA signaling pathways.[193]

5.04.4.1.2 Ethylene and abiotic stresses

Ethylene is involved in most of the plant defensive responses against abiotic stresses. Its action can vary depending on the kind of stress. Some examples of ethylene involvement in response to abiotic stress are presented below.

ROS are the key signaling molecules during plant response to ozone, whereas ethylene, SA, and JA function as second messengers and regulate the induction and the spread of oxidative stress symptoms. An increased ethylene and SA production promotes ozone-induced cell death, whereas JA acts as a cell-protective component and limits ozone-induced damage.[194–196]

During drought stress, ABA regulates stomatal closure, whereas increased ethylene production has an inhibitory influence on ABA action.[197,198] An inhibition of ethylene synthesis delays drought-associated chlorophyll loss, supporting the role of ethylene in drought-induced senescence.[199,200]

In contrast, it has been shown that in deepwater rice seedlings ethylene even delays senescence and chlorophyll breakdown upon complete submergence during plant adaptation to low oxygen concentrations.[109] Usually, during plant flooding, ethylene stimulates submergence tolerance and/or underwater shoot growth[201–204] (see Section 5.04.4.2.1). Osmotic and salt stress affects expression of ethylene synthesis and ethylene perception genes but ethylene involvement in response to this stress has not been well understood till now.[150,205]

5.04.4.2 Ethylene as a Critical Regulator of Plant Growth, Development, and Senescence

5.04.4.2.1 Ethylene as a regulator of seed germination and seedlings growth

A developing embryo produces ABA, which incepts and maintains mature seed dormancy.[206] Germination begins with the uptake of water by imbibition of the dry seed and is followed by embryo expansion. Usually, it is considered complete when the radicle is through all the covering layers. There is still controversy whether

embryonic produced ethylene triggers seed germination or only regulates the postgerminative processes. In a large number of species, exogenously applied ethylene can break the primary and secondary dormancy or accelerate the germination of nondormant seeds. It promotes germination probably by interfering with the action of ABA.[153,207] Ethylene-insensitive mutants of *Arabidopsis* characterized by an increased sensitivity to ABA produce more mature seeds showing primary dormancy than the wild-type plants. A reduced dormancy often occurs in seeds of ABA-insensitive mutants.[152] Gibberellin A (GA) is absolutely required for germination of the wild-type *Arabidopsis* seeds. Seeds of ethylene-insensitive and GA-insensitive mutants are supersensitive to exogenous ABA, which suggests that in imbibed seeds ethylene and GA may directly counteract the action of ABA in dormancy maintenance.[152,208]

In tomato, ERF2 involved in ethylene-mediated regulation impacts seed germination through positive regulation of germination-associated Man2 gene expression in contrast to ABA showing an opposite effect on both ERF2 and Man2 genes.[209]

In germinating chickpea and pea seeds, ethylene promotes its own synthesis by positive feedback regulation of ACO during the late phase of germination. ABA inhibits ethylene production during germination but not after radicle emergence.[210,211] In sugar beet seed, there seems to be neither a positive autoregulatory feedback loop for ethylene nor a negative impact of ABA on ethylene production.[212] ACC/ethylene treatment of sugar beet seeds does not alter the endogenous contents of ABA and ACO transcripts, whereas ABA upregulates the level of ACO mRNAs.

Seed germination of the root parasitic weeds *Striga* is induced by exogenous stimulants, collectively named strigolactones, exuded from the root of the host plants. Germination strategy of *Striga* seed involves *de novo* ethylene synthesis in response to strigolactones. This ethylene subsequently promotes seed germination in the vicinity of the host plant root. In the absence of strigolactones, exogenously applied ethylene promotes the germination of *Striga* seeds in a very effective manner.[213]

It seems that in cereals and dicots, ethylene does not influence germination in the same manner. For example, in intact caryopses of red rice or in wild oat, ethylene does not affect germination inception but rather stimulates the rate of growth of nascent seedling or speeds up the very slow growth of the radicle, respectively.[214,215]

In the dark, germinating seedlings of dicotyledonous plants, when their growth is mechanically impeded during their emergence from the soil, produce high amounts of ethylene. In response to this ethylene, they display three visible symptoms: inhibition of hypocotyl and root elongation, radial swelling of the hypocotyl, and retention and exaggeration of the curvature of the apical hook. Such a morphology of dark-grown seedlings in response to ethylene is referred to as a triple response. These growth habits are thought to protect the delicate shoot apex and facilitate penetration of the soil.[216] Seedlings of dicotyledonous plants germinating in the dark in the absence of ethylene are tall and spindly.

In *Arabidopsis*, HLS1 (HOOKLESS 1 gene) is a key regulator that integrates the ethylene and auxin signaling pathways during apical hook formation.[217] Both ethylene and light signals affect differential cell growth by acting through HLS1 to modulate ARFs. A complex interplay of ethylene, auxin, and light regulates the development of the apical hook in light- and dark-grown seedlings.[218] The crucial role of auxin in growth inhibition in a variety of dicots usually involves ethylene-mediated inhibition of root, and shoot elongation results from its influence on auxin transport and accumulation.[219,220] Promotion of ethylene synthesis following auxin treatment has been described by many authors for a long time (see Section 5.04.2.2.8), but in contrast the ethylene-mediated increase in auxin production via ethylene-stimulated expression of an enzyme involved in the auxin biosynthetic pathway, tryptophan aminotransferase TAA1, was recently reported.[221–224]

It was shown that auxin-stimulated ethylene production triggers an increase in endogenous ABA, which causally leads to growth inhibition of shoot. In such a situation, ethylene and ABA appear to function as a further hormonal second messenger of auxin in the signaling of growth inhibition.[120,225,226] Ethylene and ABA interactions are antagonistic in the control of seed dormancy but synergistic in inhibiting root growth.[152,153]

In cereals, ethylene evolution usually increases during seedling development and promotes seedling shoot growth.[120,214,215] Nevertheless, in some grasses, overproduction of cyanide, which is formed at physiologically damaging concentrations as a coproduct of ethylene biosynthesis, is implicated in phytotoxic growth inhibition.[225]

During submergence of deepwater rice, ethylene diffusion from plant tissues strongly decreases. Moreover, hypoxia enhances its endogenous synthesis in the coleoptile of rice seedlings and internodes of adult plants. Altered balance between highly accumulated ethylene, ABA, and GA regulates the rapid growth of rice shoot.[227]

In conclusion, ethylene can promote or inhibit growth depending on the cell type and plant species. Generally, it inhibits shoot elongation in most terrestrial plants, but stimulates underwater shoot extension in many amphibious, wetland, and aquatic species enabling them to keep foliage above the water and survive.[102,201,204,228]

5.04.4.2.2 Ethylene and sex expression

Most species of angiosperms produce hermaphroditic flowers. The unisexuality in some plants most often results from developmentally programmed abortion or selective reduction in sex organ primordia. In dicots, higher levels of auxins, cytokinins, and ethylene usually correlate with female sex expression and in most of them the femaleness is mainly promoted by ethylene.[229]

Ethylene acts as the major regulator responsible for the genetically controlled sex expression in cucumber flowers varying in response to environmental and hormonal cues. Cucumber plants are classified into different genotypes: monoecious, gynoecious, andromonoecious, and hermaphroditic. Two major genetic loci, F and M, control sex expression phenotypes in cucumber; the monoecious genotypes (M-ff) produce both male and female flowers, gynoecious genotypes (M-F-) produce only female flowers, the andromonoecious genotypes (mmff) produce male and bisexual flowers, whereas the hermaphroditic genotypes (mmF-) produce only bisexual flowers.[230–232] The F locus governs the ethylene biosynthesis correlated with the development of female flowers. The gynoecious plants produce more ethylene than monoecious plants, and monoecious and andromonoecious plants treated with ethylene produce an increased number of female and bisexual flowers, respectively. In addition, blocking of ethylene synthesis or action decreases the number of female or bisexual flowers. It was shown that a duplication of one of the ACS genes, Cs-ACS1, gave rise to the F locus and gynoecious cucumber plants. Cs-ACS1G is mapped to the F locus and is considered as gynoecious-specific representing an additional copy of non-sex-specific Cs-ACS1.[233,234] Moreover, the accumulation of Cs-ACS2 transcripts (in **Figure 2** designated as Cs-ACS2*) in shoot apices of gynoecious cucumbers has been shown to correlate with the appearance of female flowers.[232]

Ethylene is the major feminizing hormone in most cucurbits showing a wide range of sex phenotypes with some exception of watermelon in which a steady increase in ACS isozymes transcript level in male flowers compared with female flowers has been reported.[235]

5.04.4.2.3 Ethylene as a ripening and senescence mediator

Despite the potent influence of ethylene on the whole plant development, most often it has been recognized as a fruit-ripening and senescence-associated hormone. With respect to the gaseous nature of this hormone, the knowledge about ethylene sensitivity of fruit and flower is necessary to predict the effects of their mixed storage and transport and the usefulness of anti-ethylene treatments. Such commercial implications of ethylene have made it a topic of investigation for decades.

Ethylene-dependent mechanism terminates flower life after successful pollination as a way to benefit survival of the species. Usually, in species with short-lived flowers, such an ethylene-dependent termination apparently is not beneficial as the life of individual flowers is short anyway. Similarly, in species with numerous flowers in one flower head in which continuous visits of pollinators are required, ethylene-dependent termination is not beneficial.[236]

Flowers are divided into two groups: sensitive and insensitive to ethylene. An increase in endogenous ethylene production in the former group is either due to developmental or pollination-induced senescence. Floral abscission occurs in several families in monocotyledons and eudicotyledons in contrast to petal abscission, which is common in the latter and rare in the former group. Both are generally but not universally regulated by ethylene. With respect to petal senescence, high sensitivity to ethylene is found in Campanulaceae, Caryophyllaceae, Geraniaceae, Labiatae, Malvaceae, Orchidaceae, Primulaceae, Ranunculaceae, and Rosaceae, whereas Compositae and Iridacae, and most of Amaryllidaceae and Liliaceae,

are ethylene insensitive; however, the regulation of developmental aspects other than senescence in these flowers by ethylene is not excluded.[236–238]

Usually, flower senescence is mediated by the evolution of ethylene following contact between pollen and the stigmatal surface; however, the character of the primary signal resulting in ethylene evolution has not been established till now.[238] In the absence of pollination, corolla senescence still occurs, but is delayed, following an endogenously regulated program.

In carnation, in the absence of pollination, styles start to produce high levels of ethylene 7 days after flower opening during natural senescence, or several hours after compatible pollination, or after a treatment with exogenous ethylene. The first pollination- or senescence-induced ethylene is synthesized in gynoecium (stigma and ovary) from where it diffuses to petals and triggers its autocatalytic production leading to wilting.[239,240] In orchids, the longevity of intact unpollinated flower may reach as long as several months but pollination signals and/or emasculation (removal of the anther cap by insect or animal pollinator) initiate rapid senescence of the perianth. In some orchids, the role of ethylene as a mobile factor diffusing from the column representing fused male and female organs to the other floral organs is controversial. In pollinated *Cymbidium* flower ethylene serves as an interorgan signal, whereas in *Phalaenopsis* it is rather the ACC that is translocated from gynoecium to perianth than the ethylene itself.[241] In perianth, translocated ACC acts as a signal to commence autocatalytic senescence-related ethylene production;[99,127,128] (see **Figure 3(b)** and Section 5.04.2.3). Naturally programmed rapid senescence of tomato flower is only accelerated by pollination and ACC translocated from pistil to petals is converted to ethylene through ACO already present there. In petals, this ethylene regulates its own synthesis in a positive feedback manner by triggering the expression of ACS genes.[242]

Anatomically, fruits are swollen ovaries that may also contain associated flower parts. Their development occurs simultaneously with seed maturation. The embryo matures and the seed accumulates storage products, acquires desiccation tolerance, and loses water. The fruit then ripens. Fruits have been classified into two groups: climacteric showing a sharp increase in respiration with a concomitant surge of endogenous ethylene production at the onset of ripening; and nonclimacteric whose ripening is generally considered as a process that as a whole does not require ethylene. Climacteric fruits include tomato, avocado, pear, apple, and banana, and nonclimacteric fruits include orange, lemon, and strawberry. Moreover, the climacteric fruits accelerate their ripening in response to exogenously applied ethylene or in natural situation to ethylene diffusing from other fruits or flowers.

Ethylene coordinates the expression of genes responsible for enhanced respiratory metabolism, chlorophyll degradation, carotenoid synthesis, conversion of starch to sugars, increased activity of cell wall-degrading enzymes, aroma volatile production, and so on. All these events stimulate a series of biochemical, physiological, and structural changes making fruits mature and attractive to the consumer.

In the majority of climacteric fruits, whose life is divided into preclimacteric and climacteric stage, the transition from autoinhibitory to autocatalytic ethylene synthesis is developmentally regulated. Ripening or senescing most often commences in one region of a fruit or flower and spreads to neighboring tissues as ethylene diffuses freely from cell to cell. This autostimulatory effect of ethylene has been clearly evidenced in fruits such as apple, pear, passion fruit, avocado, and tomato.[125,243–246] The well-documented transition mechanism from system 1 (autoinhibitory) to system 2 (autocatalytic) of ethylene synthesis in ripening tomato fruit is illustrated in **Figure 3(a)** (see Section 5.04.2.3). Persimmon fruit, classified as climacteric, produces a small amount of ethylene but young persimmon fruit generates large quantities of hormone immediately after detachment from the tree. The detached fruit experiences water loss, which triggers ethylene synthesis in its calyx. This ethylene diffusing to the other part of the fruit stimulates there a burst of its autocatalytic production (system 2).[247] Some late pear varieties have to be postharvestly exposed to lower temperature for ripening. Expression of ripening-related ACS and ACO genes is controlled by two kinds of subsequent stimuli, a cold-related signal and the rewarming-stimulated ethylene synthesis.[244] The unripe, mature kiwifruits are highly sensitive to exogenously applied ethylene, which has an autostimulatory effect on SAM synthetase and ACO, whereas ACS expression is detected later with the start of endogenous ethylene synthesis only.[248] Fig shows ethylene-controlled ripening, but in fig, ethylene affects its own production in an autoinhibitory manner.[249] In contrast to the majority of the climacteric fruits, banana exhibits a sudden increase followed by a rapid decrease in ethylene production at the onset of ripening. The peel of banana fruit makes up more than

20% of a whole fruit. Ethylene regulates its own synthesis in an autostimulatory manner in the peel but it shows autoinhibitory effect in the pulp.[250]

The ripening of nonclimacteric fruits is usually considered as a process that as a whole does not require ethylene. Nevertheless, endogenous ethylene is involved at some steps of development of these fruits. In general, chlorophyll degradation in nonclimacteric fruits is believed to be ethylene stimulated, whereas the synthesis of pigments can depend on ethylene action or not.[251]

A considerable progress has been made in the understanding of ethylene biosynthesis, perception, and action in higher plants. This knowledge has enabled construction of some genetically modified plants characterized by a decreased ethylene biosynthesis or decreased ethylene sensitivity. Moreover, detailed structural and biochemical data can help to design ACS inhibitors, whose application is expected to have immense agricultural effects. The control of ethylene production and perception is of great interest for plant biotechnology because it can delay senescence and overmaturation, the processes responsible for extensive loss of vegetables and fruits on storage.

Abbreviations

1-dzSIBA	*S*-isobutyl-1-deazaadenosine
3dz-SAHC	3-deazaadenosylhomocysteine
3-dzSIBA	*S*-isobutyl-3-deazaadenosine
7-dzSIBA	*S*-isobutyl-7-deazaadenosine
ABA	abscisic acid
ACC	1-aminocyclopropane-1-carboxylic acid
ACO	1-aminocyclopropane-1-carboxylate oxidase
ACS	1-aminocyclopropane-1-carboxylate synthase
ARF	auxin-response factor
AVG	aminoethoxyvinylglycine
BA	N6-benzyladenine
CDPK	calcium-dependent protein kinase
CTR	constitutive triple response
EREBP	ethylene-response element-binding protein
GA	gibberellin A
GACC	glutamyl-ACC
IAA	indole-3-acetic acid
JA	jasmonic acid
KMB	2-keto-4-methylthiobutyrate
L-VG	L-vinylglycine
MACC	1-malonyl-ACC
MAPK6	mitogen-activated protein kinase 6
MAPKK	MAPK kinase
MAPKKK	MAPKK kinase
MAPKKK	MAP-kinase kinase kinase
MTA	5′-methylthioadenosine
MTR	5-methylthioribose
MTR-1-P	5-methylthioribose-1-phosphate
MVG	methoxyethoxyvinylglycine
PERE	primary ethylene-response element
PLP	pyridoxal-5′-phosphate
PMP	pyridoxalamine phosphate
ROS	reactive oxygen species
SA	salicylic acid
SAHC	*S*-adenosylhomocysteine
SAM	*S*-adenosyl-L-methionine

SERE	secondary ethylene-response element
SIBA	*S*-isobutyladenosine
SMM	*S*-methyl-L-methionine
TAA	tryptophan aminotransferase

References

1. S. P. Burg; E. A. Burg, *Plant Physiol.* **1962**, *37*, 179–189.
2. D. N. Neljubow, *Pflanz. Beitr. Bot. Zentralbl.* **1901**, *10*, 128–139.
3. A. Theologis, *Cell* **1992**, *70*, 181–184.
4. R. Gane, *Nature* **1934**, *134*, 1008.
5. M. Lieberman; A. Kunishi, *Plant Physiol.* **1966**, *41*, 376–382.
6. S. F. Yang; N. E. Hoffman, *Annu. Rev. Plant Physiol.* **1984**, *35*, 155–189.
7. J. H. Miyazaki; S. F. Yang, *Plant Physiol.* **1987**, *84*, 277–281.
8. Z. Zhang; J. N. Barlow; J. E. Baldwin; C. J. Schofield, *Biochemistry* **1997**, *36*, 15999–16007.
9. Y. S. Seo; A. Yoo; J. Jung; S. K. Sung; D. R. Yang; W. T. Kim; W. Lee, *Biochem. J.* **2004**, *380*, 339–346.
10. A. Yoo; Y. S. Seo; J. W. Jung; S. K. Sung; W. T. Kim; W. Lee; D. R. Yang, *J. Struct. Biol.* **2006**, *156*, 407–420.
11. J. Zhou; A. M. Rocklin; J. D. Lipscomb; L. Que, Jr.; E. I. Solomon, *J. Am. Chem. Soc.* **2002**, *124*, 4602–4609.
12. W. K. Yip; S. F. Yang, *Plant Physiol.* **1988**, *88*, 473–476.
13. X. Tang; H. Wang; A. S. Brandt; W. R. Woodson, *Plant Mol. Biol.* **1993**, *23*, 1151–1164.
14. C. S. Barry; B. Blume; M. Bouzayen; W. Cooper; A. J. Hamilton; D. Grierson, *Plant J.* **1996**, *9*, 525–535.
15. E. Lasserre; T. Bouquin; J. A. Hernandez; J. Bull; J. C. Pech; C. Balague, *Mol. Gen. Genet.* **1996**, *251*, 81–90.
16. D. J. Osborne; J. Walters; B. V. Milborrow; A. Norville; L. M. C. Stange, *Phytochemistry* **1996**, *42*, 51–60.
17. M. N. Martin; J. D. Cohen; R. A. Saftner, *Plant Physiol.* **1995**, *109*, 917–926.
18. M. N. Martin; J. P. Slovin, *Plant Physiol.* **2000**, *122*, 1417–1426.
19. G. Peiser; Y. S. Fa, *Plant Physiol.* **1998**, *116*, 1527–1532.
20. A. Klintborg; L. Eklund; C. H. Little, *Tree Physiol.* **2002**, *22*, 59–66.
21. Y. J. Jia; Y. Kakuta; M. Sugawara; T. Igarashi; N. Oki; M. Kisaki; T. Shoji; Y. Kanetuna; T. Horita; H. Matsui; M. Honma, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 542–549.
22. Y. Kakuta; T. Igarashi; T. Murakami; H. Ito; H. Matsui; M. Honma, *Biosci. Biotechnol. Biochem.* **2001**, *65*, 1511–1518.
23. B. Rohwer; M. Bopp, *J. Plant Physiol.* **1985**, *117*, 331–338.
24. F. L. Tittle, *Physiol. Plant.* **1987**, *70*, 499–502.
25. T. van den Driessche; C. Kevers; M. Collet; Th. Gaspar, *J. Plant Physiol.* **1988**, *133*, 635–639.
26. G. Capitani; A. C. Eliot; H. Gut; R. M. Khomutov; J. F. Kirsch; M. G. Grutter, *Biochim. Biophys. Acta.* **2003**, *1647*, 55–60.
27. S. Satoh; H. Mori; H. Imaseki, *Plant Cell Physiol.* **1993**, *34*, 753–760.
28. D. L. McCarthy; G. Capitani; L. Feng; M. G. Gruetter; J. F. Kirsch, *Biochemistry* **2001**, *40*, 12276–12284.
29. J. F. Li; L. H. Qu; N. Li, *J. Exp. Bot.* **2005**, *56*, 2203–2210.
30. T. Yamagami; A. Tsuchisaka; K. Yamada; W. F. Haddon; L. A. Harden; A. Theologis, *J. Biol. Chem.* **2003**, *278*, 49102–49112.
31. S. Satoh; S. F. Yang, *Plant Physiol.* **1989**, *91*, 1036–1039.
32. L. Feng; J. F. Kirsch, *Biochemistry* **2000**, *39*, 2436–2444.
33. G. Capitani; M. Tschopp; A. C. Eliot; J. F. Kirsch; M. G. Grutter, *FEBS Lett.* **2005**, *579*, 2458–2462.
34. G. Dardenne; J. Casimir; M. Marlier; P. O. Larsen, *Phytochemistry* **1974**, *13*, 1897–1900.
35. F. Bourgis; S. Roje; M. L. Nuccio; D. B. Fisher; M. C. Tarczynski; C. Li; C. Herschbach; H. Rennenberg; M. J. Pimenta; T. L. Shen; D. A. Gage; A. D. Hanson, *Plant Cell.* **1999**, *11*, 1485–1498.
36. S. Ko; A. C. Eliot; J. F. Kirsch, *Arch. Biochem. Biophys.* **2004**, *421*, 85–90.
37. L. Feng; M. K. Geck; A. C. Eliot; J. F. Kirsch, *Biochemistry* **2000**, *39*, 15242–15249.
38. H. Hyodo; K. Tanaka, *Plant Cell Physiol.* **1986**, *27*, 391–398.
39. I. Icekson; A. Apelbaum, *Biochem. Biophys. Res. Commun.* **1983**, *113*, 586–591.
40. G. A. Miura; P. K. Chiang, *Anal. Biochem.* **1985**, *147*, 217–221.
41. T. Yasuta; S. Satoh; K. Minamisawa, *Appl. Environ. Microbiol.* **1999**, *65*, 849–852.
42. M. Sugawara; S. Okazaki; N. Nukui; H. Ezura; H. Mitsui; K. Minamisawa, *Biotechnol. Adv.* **2006**, *24*, 382–388.
43. S. Okazaki; M. Sugawara; K. Yuhashi; K. Minamisawa, *Ann. Bot. (Lond.)* **2007**, *100*, 55–59.
44. A. K. Mattoo; J. D. Anderson; E. Chalutz; M. Lieberman, *Plant Physiol.* **1979**, *64*, 289–292.
45. Q. Huai; Y. Xia; Y. Chen; B. Callahan; N. Li; H. Ke, *J. Biol. Chem.* **2001**, *276*, 38210–38216.
46. G. Capitani; D. L. McCarthy; H. Gut; M. G. Grutter; J. F. Kirsch, *J. Biol. Chem.* **2002**, *277*, 49735–49742.
47. Y. Liu; S. Zhang, *Plant Cell.* **2004**, *16*, 3386–3399.
48. G. Capitani; E. Hohenester; L. Feng; P. Storici; J. F. Kirsch; J. N. Jansonius, *J. Mol. Biol.* **1999**, *294*, 745–756.
49. J. N. Jansonius, *Curr. Opin. Struct. Biol.* **1998**, *8*, 759–769.
50. A. S. Tarun; A. Theologis, *J. Biol. Chem.* **1998**, *273*, 12509–12514.
51. A. Tsuchisaka; A. Theologis, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2275–2280.
52. N. Li; A. K. Mattoo, *J. Biol. Chem.* **1994**, *269*, 6908–6917.
53. P. K. Mehta; P. Christen, *Biochem. Biophys. Res. Commun.* **1994**, *198*, 138–143.

54. W. H. Rottmann; G. F. Peter; P. W. Oeller; J. A. Keller; N. F. Shen; B. P. Nagy; L. P. Taylor; A. D. Campbell; A. Theologis, *J. Mol. Biol.* **1991**, *222*, 937–961.
55. M. F. White; J. Vasquez; S. F. Yang; J. F. Kirsch, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12428–12432.
56. Y. Li; L. Feng; J. F. Kirsch, *Biochemistry* **1997**, *36*, 15477–15488.
57. A. S. Tarun; J. S. Lee; A. Theologis, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9796–9801.
58. H. Zhou; H. W. Wang; K. Zhu; S. F. Sui; P. Xu; S. F. Yang; N. Li, *Plant Physiol.* **1999**, *121*, 913–919.
59. A. C. Eliot; J. F. Kirsch, *Biochemistry* **2002**, *41*, 3836–3842.
60. V. Savolainen; M. W. Chase; S. B. Hoot; C. M. Morton; D. E. Soltis; C. Bayer; M. F. Fay; A. Y. de Bruijn; S. Sullivan; Y. L. Qiu, *Syst. Biol.* **2000**, *49*, 306–362.
61. N. Saitou; M. Nei, *Mol. Biol. Evol.* **1987**, *4*, 406–425.
62. J. Felsenstein, *Evolution* **1985**, *39*, 783–791.
63. E. Zuckerkandl; L. Pauling, Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins*; V. Bryson, H. J. Vogel, Eds.; Academic Press: New York, 1965; pp 97–166.
64. K. Tamura; J. Dudley; M. Nei; S. Kumar, *Mol. Biol. Evol.* **2007**, *24*, 1596–1599.
65. D. Wang; J. Fan; R. S. Ranu, *Plant Cell Rep.* **2004**, *22*, 422–429.
66. J. R. Barnes; W. W. Lorenz; J. F. Dean, *Gene* **2008**, *413*, 18–31.
67. K. A. Koch; G. Capitani; M. G. Gruetter; J. F. Kirsch, *Gene* **2001**, *272*, 75–84.
68. B. R. Peixoto; Y. Mikawa; S. Brenner, *Gene* **2000**, *246*, 275–283.
69. D. G. Grosskopf; G. Felix; T. Boller, *FEBS Lett.* **1990**, *275*, 177–180.
70. G. Felix; D. G. Grosskopf; M. Regenass; C. W. Basse; T. Boller, *Plant Physiol.* **1991**, *97*, 19–25.
71. P. Spanu; D. G. Grosskopf; G. Felix; T. Boller, *Plant Physiol.* **1994**, *106*, 529–535.
72. L. Petruzzelli; M. Sturaro; D. Mainieri; G. Leubner-Metzger, *Plant Cell Environ.* **2003**, *26*, 661–671.
73. M. Tatsuki; H. Mori, *J. Biol. Chem.* **2001**, *276*, 28051–28057.
74. C. Y. Kim; Y. Liu; E. T. Thorne; H. Yang; H. Fukushige; W. Gassmann; D. Hildebrand; R. E. Sharp; S. Zhang, *Plant Cell.* **2003**, *15*, 2707–2718.
75. C. Jonak; L. Okresz; L. Bogre; H. Hirt, *Curr. Opin. Plant Biol.* **2002**, *5*, 415–424.
76. S. Joo; Y. Liu; A. Lueth; S. Zhang, *Plant J.* **2008**, *54*, 129–140.
77. J. J. Kieber; J. R. Ecker, *Trends Genet.* **1993**, *9*, 356–362.
78. P. Guzman; J. R. Ecker, *Plant Cell.* **1990**, *2*, 513–523.
79. K. L. Wang; H. Yoshida; C. Lurin; J. R. Ecker, *Nature* **2004**, *428*, 945–950.
80. H. Yoshida; M. Nagata; K. Saito; K. L. Wang; J. R. Ecker, *BMC Plant Biol.* **2005**, *5*, 14.
81. H. Yoshida; K. L. Wang; C. M. Chang; K. Mori; E. Uchida; J. R. Ecker, *Plant Mol. Biol.* **2006**, *62*, 427–437.
82. M. Dieterle; A. Thomann; J. P. Renou; Y. Parmentier; V. Cognat; G. Lemonnier; R. Muller; W. H. Shen; T. Kretsch; P. Genschik, *Plant J.* **2005**, *41*, 386–399.
83. P. Figueroa; G. Gusmaroli; G. Serino; J. Habashi; L. Ma; Y. Shen; S. Feng; M. Bostick; J. Callis; H. Hellmann; X. W. Deng, *Plant Cell.* **2005**, *17*, 1180–1195.
84. J. P. Vogel; K. E. Woeste; A. Theologis; J. J. Kieber, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4766–4771.
85. H. S. Chae; F. Faure; J. J. Kieber, *Plant Cell.* **2003**, *15*, 545–559.
86. K. E. Woeste; C. Ye; J. J. Kieber, *Plant Physiol.* **1999**, *119*, 521–530.
87. R. J. Ferl, *Physiol. Plant.* **2004**, *120*, 173–178.
88. Y. Yao; Y. Du; L. Jiang; J. Y. Liu, *Biochemistry (Mosc.)* **2007**, *72*, 1003–1007.
89. A. J. Cary; W. Liu; S. H. Howell, *Plant Physiol.* **1995**, *107*, 1075–1082.
90. J. P. Vogel; P. Schuerman; K. Woeste; I. Brandstatter; J. J. Kieber, *Genetics* **1998**, *149*, 417–427.
91. J.-F. Li; R. Qi; L.-H. Qu; A. K. Mattoo; N. Li, *J. Integr. Plant Biol.* **2005**, *47*, 1352–1363.
92. M. Jakubowicz; A. Pacak, *Acta Physiol. Plant.* **2004**, *26*, 5–27.
93. J. H. Oetiker; D. C. Olson; O. Y. Shiu; S. F. Yang, *Plant Mol. Biol.* **1997**, *34*, 275–286.
94. L. J. Destefano-Beltran; W. van Caeneghem; J. Gielen; L. Richard; M. van Montagu; D. van der Straeten, *Mol. Gen. Genet.* **1995**, *246*, 496–508.
95. X. Liang; S. Abel; J. A. Keller; N. F. Shen; A. Theologis, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 11046–11050.
96. J. R. Botella; J. M. Arteca; C. D. Schlagnhaufer; R. N. Arteca; A. T. Phillips, *Plant Mol. Biol.* **1992**, *20*, 425–436.
97. J. R. Botella; C. D. Schlagnhaufer; J. M. Arteca; R. N. Arteca; A. T. Phillips, *Gene* **1993**, *123*, 249–253.
98. J. R. Botella; C. D. Schlagnhaufer; R. N. Arteca; A. T. Phillips, *Plant Mol. Biol.* **1992**, *18*, 793–797.
99. A. Q. Bui; S. D. O'Neill, *Plant Physiol.* **1998**, *116*, 419–428.
100. F. M. Mathooko; M. W. Mwaniki; A. Nakatsuka; S. Shiomi; Y. Kubo; A. Inaba; R. Nakamura, *Plant Cell Physiol.* **1999**, *40*, 164–172.
101. E. P. Bekman; N. J. Saibo; C. A. Di; A. P. Regalado; C. P. Ricardo; C. Rodrigues-Pousada, *Planta* **2000**, *211*, 663–672.
102. T. I. Zarembinski; A. Theologis, *Mol. Biol. Cell.* **1993**, *4*, 363–373.
103. T. I. Zarembinski; A. Theologis, *Plant Mol. Biol.* **1997**, *33*, 71–77.
104. X. Liang; Y. Oono; N. F. Shen; C. Kohler; K. Li; P. A. Scolnik; A. Theologis, *Gene* **1995**, *167*, 17–24.
105. S. G. Ralph; J. W. Hudgins; S. Jancsik; V. R. Franceschi; J. Bohlmann, *Plant Physiol.* **2007**, *143*, 410–424.
106. P. L. Huang; J. E. Parks; W. H. Rottmann; A. Theologis, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7021–7025.
107. J. R. Botella; R. N. Arteca; J. A. Frangos, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1595–1598.
108. X. Liang; N. F. Shen; A. Theologis, *Plant J.* **1996**, *10*, 1027–1036.
109. D. van der Straeten; Z. Zhou; E. Prinsen; H. A. Van Onckelen; M. C. Van Montagu, *Plant Physiol.* **2001**, *125*, 955–968.
110. J. E. Lincoln; A. D. Campbell; J. Oetiker; W. H. Rottmann; P. W. Oeller; N. F. Shen; A. Theologis, *J. Biol. Chem.* **1993**, *268*, 19422–19430.
111. S. Abel; M. D. Nguyen; W. Chow; A. Theologis, *J. Biol. Chem.* **1995**, *270*, 19093–19099.
112. D. C. Olson; J. H. Oetiker; S. F. Yang, *J. Biol. Chem.* **1995**, *270*, 14056–14061.
113. Y. Ishiki; A. Oda; Y. Yaegashi; Y. Orihara; T. Arai; T. Hirabayashi; H. Nakagawa; T. Sato, *Plant Sci.* **2000**, *159*, 173–181.

114. O. Y. Shiu; J. H. Oetiker; W. K. Yip; S. F. Yang, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10334–10339.
115. A. Tsuchisaka; A. Theologis, *Plant Physiol.* **2004**, *136*, 2982–3000.
116. A. ten Have; E. J. Woltering, *Plant Mol. Biol.* **1997**, *34*, 89–97.
117. N. Nakagawa; H. Mori; K. Yamazaki; H. Imaseki, *Plant Cell Physiol.* **1991**, *32*, 1153–1163.
118. P. Spanu; T. Boller; H. Kende, *J. Plant Physiol.* **1993**, *141*, 557–562.
119. S. C. Peck; H. Kende, *Plant J.* **1998**, *14*, 573–581.
120. F. B. Abeles; P. W. Morgan; M. E. Saltveit, *Ethylene in Plant Biology*; Academic Press: San Diego, CA, 1992.
121. S. B. Tiwari; G. Hagen; T. J. Guilfoyle, *Plant Cell.* **2004**, *16*, 533–543.
122. T. Koshiba; N. Ballas; L. M. Wong; A. Theologis, *J. Mol. Biol.* **1995**, *253*, 396–413.
123. B. Jackson; J. E. Summers; L. A. C. J. Voesenek, Potamogeton Pectinatus: A Vascular Plant that Makes No Ethylene. In *Biology and Biotechnology of the Plant Hormone Ethylene*; A. K. Kanellis, C. Chang, H. Kende, D. Grierson, Eds.; Kluwer Academic Publishers: Netherlands, 1997.
124. E. J. McMurchie; W. B. McGlasson; I. L. Eaks, *Nature* **1972**, *237*, 235–236.
125. C. S. Barry; M. I. Llop-Tous; D. Grierson, *Plant Physiol.* **2000**, *123*, 979–986.
126. A. Nakatsuka; S. Murachi; H. Okunishi; S. Shiomi; R. Nakano; Y. Kubo; A. Inaba, *Plant Physiol.* **1998**, *118*, 1295–1305.
127. S. D. O'Neill; J. A. Nadeau; X. S. Zhang; A. Q. Bui; A. H. Halevy, *Plant Cell.* **1993**, *5*, 419–432.
128. N. N. Wang; S. F. Yang; Y. Charng, *Plant Physiol.* **2001**, *126*, 253–260.
129. J. H. Kim; W. T. Kim; B. G. Kang, *Plant Cell Physiol.* **2001**, *42*, 1056–1061.
130. H. Guo; J. R. Ecker, *Curr. Opin. Plant Biol.* **2004**, *7*, 40–49.
131. C. Chang; J. A. Shockey, *Curr. Opin. Plant Biol.* **1999**, *2*, 352–358.
132. Y. F. Chen; N. Etheridge; G. E. Schaller, *Ann. Bot. (Lond.)* **2005**, *95*, 901–915.
133. J. J. Kieber, *J. Exp. Bot.* **1997**, *48*, 211–218.
134. P. Moussatche; H. J. Klee, *J. Biol. Chem.* **2004**, *279*, 48734–48741.
135. W. Wang; J. J. Esch; S. H. Shiu; H. Agula; B. M. Binder; C. Chang; S. E. Patterson; A. B. Bleecker, *Plant Cell.* **2006**, *18*, 3429–3442.
136. H. J. Klee, *J. Exp. Bot.* **2002**, *53*, 2057–2063.
137. H. Klee; D. Tieman, *Physiol. Plant.* **2002**, *115*, 336–341.
138. B. M. Kevany; D. M. Tieman; M. G. Taylor; V. D. Cin; H. J. Klee, *Plant J.* **2007**, *51*, 458–467.
139. B. M. Kevany; M. G. Taylor; H. J. Klee, *Plant Biotechnol. J.* **2008**, *6*, 295–300.
140. W. Wang; A. E. Hall; R. O'Malley; A. B. Bleecker, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 352–357.
141. D. M. Tieman; M. G. Taylor; J. A. Ciardi; H. J. Klee, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5663–5668.
142. J. Q. Wilkinson; M. B. Lanahan; H. C. Yen; J. J. Giovannoni; H. J. Klee, *Science* **1995**, *270*, 1807–1809.
143. J. Hua; E. M. Meyerowitz, *Cell* **1998**, *94*, 261–271.
144. C. Chang, *Trends Plant Sci.* **2003**, *8*, 365–368.
145. G. Olmedo; H. Guo; B. D. Gregory; S. D. Nourizadeh; L. Aguilar-Henonin; H. Li; F. An; P. Guzman; J. R. Ecker, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13286–13293.
146. T. Potuschak; A. Vansiri; B. M. Binder; E. Lechner; R. D. Vierstra; P. Genschik, *Plant Cell.* **2006**, *18*, 3047–3057.
147. B. M. Binder; J. M. Walker; J. M. Gagne; T. J. Emborg; G. Hemmann; A. B. Bleecker; R. D. Vierstra, *Plant Cell.* **2007**, *19*, 509–523.
148. S. D. Yoo; Y. H. Cho; G. Tena; Y. Xiong; J. Sheen, *Nature* **2008**, *451*, 789–795.
149. L. Adams-Phillips; C. Barry; P. Kannan; J. Leclercq; M. Bouzayen; J. Giovannoni, *Plant Mol. Biol.* **2004**, *54*, 387–404.
150. W. H. Cao; J. Liu; Q. Y. Zhou; Y. R. Cao; S. F. Zheng; B. X. Du; J. S. Zhang; S. Y. Chen, *Plant Cell Environ.* **2006**, *29*, 1210–1219.
151. A. N. Stepanova; J. M. Alonso, *Sci. STKE* **2005**, *276*, cm4.
152. N. Beaudoin; C. Serizet; F. Gosti; J. Giraudat, *Plant Cell.* **2000**, *12*, 1103–1115.
153. M. Ghassemian; E. Nambara; S. Cutler; H. Kawaide; Y. Kamiya; P. McCourt, *Plant Cell.* **2000**, *12*, 1117–1126.
154. Y. Wang; C. Liu; K. Li; F. Sun; H. Hu; X. Li; Y. Zhao; C. Han; W. Zhang; Y. Duan; M. Liu; X. Li, *Plant Mol. Biol.* **2007**, *64*, 633–644.
155. H. Guo; J. R. Ecker, *Cell* **2003**, *115*, 667–677.
156. S. Y. Fujimoto; M. Ohta; A. Usui; H. Shinshi; M. Ohme-Takagi, *Plant Cell.* **2000**, *12*, 393–404.
157. M. Ohme-Takagi; K. Suzuki; H. Shinshi, *Plant Cell Physiol.* **2000**, *41*, 1187–1192.
158. R. Solano; A. Stepanova; Q. Chao; J. R. Ecker, *Genes Dev.* **1998**, *12*, 3703–3714.
159. I. A. Penninckx; B. P. Thomma; A. Buchala; J. P. Metraux; W. F. Broekaert, *Plant Cell.* **1998**, *10*, 2103–2113.
160. G. Y. Zhong; J. K. Burns, *Plant Mol. Biol.* **2003**, *53*, 117–131.
161. M. Fujita; Y. Fujita; Y. Noutoshi; F. Takahashi; Y. Narusaka; K. Yamaguchi-Shinozaki; K. Shinozaki, *Curr. Opin. Plant Biol.* **2006**, *9*, 436–442.
162. A. Pitzschke; H. Hirt, *Plant Physiol.* **2006**, *141*, 351–356.
163. J. P. Anderson; E. Badruzsaufari; P. M. Schenk; J. M. Manners; O. J. Desmond; C. Ehlert; D. J. Maclean; P. R. Ebert; K. Kazan, *Plant Cell.* **2004**, *16*, 3460–3479.
164. O. Lorenzo; R. Piqueras; J. J. Sanchez-Serrano; R. Solano, *Plant Cell.* **2003**, *15*, 165–178.
165. A. A. Ludwig; H. Saitoh; G. Felix; G. Freymark; O. Miersch; C. Wasternack; T. Boller; J. D. Jones; T. Romeis, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10736–10741.
166. S. Zhang; D. F. Klessig, *Trends Plant Sci.* **2001**, *6*, 520–527.
167. T. Zhang; Y. Liu; T. Yang; L. Zhang; S. Xu; L. Xue; L. An, *Plant Physiol. Biochem.* **2006**, *44*, 274–283.
168. H. Nakagami; A. Pitzschke; H. Hirt, *Trends Plant Sci.* **2005**, *10*, 339–346.
169. Y. Sakuma; Q. Liu; J. G. Dubouzet; H. Abe; K. Shinozaki; K. Yamaguchi-Shinozaki, *Biochem. Biophys. Res. Commun.* **2002**, *290*, 998–1009.
170. H. Zhang; Z. Huang; B. Xie; Q. Chen; X. Tian; X. Zhang; H. Zhang; X. Lu; D. Huang; R. Huang, *Planta* **2004**, *220*, 262–270.
171. M. Berrocal-Lobo; A. Molina; R. Solano, *Plant J.* **2002**, *29*, 23–32.
172. L. E. Trujillo; M. Sotolongo; C. Menendez; M. E. Ochogavia; Y. Coll; I. Hernandez; O. Borras-Hidalgo; B. P. Thomma; P. Vera; L. Hernandez, *Plant Cell Physiol.* **2008**, *49*, 512–525.

173. L. G. Jin; J. Y. Liu, *Plant Physiol. Biochem.* **2008**, *46*, 46–53.
174. T. Hoffman; J. S. Schmidt; X. Zheng; A. F. Bent, *Plant Physiol.* **1999**, *119*, 935–950.
175. C. M. Pieterse; S. C. van Wees; J. A. van Pelt; M. Knoester; R. Laan; H. Gerrits; P. J. Weisbeek; L. C. van Loon, *Plant Cell.* **1998**, *10*, 1571–1580.
176. C. Norman-Setterblad; S. Vidal; E. T. Palva, *Mol. Plant Microbe Interact.* **2000**, *13*, 430–438.
177. T. Asai; J. M. Stone; J. E. Heard; Y. Kovtun; P. Yorgey; J. Sheen; F. M. Ausubel, *Plant Cell.* **2000**, *12*, 1823–1836.
178. J. T. Greenberg; F. P. Silverman; H. Liang, *Genetics* **2000**, *156*, 341–350.
179. H. Fukuda; T. Ogawa; S. Tanase, *Adv. Microb. Physiol.* **1993**, *35*, 275–306.
180. L. C. van Loon; B. P. Geraats; H. J. Linthorst, *Trends Plant Sci.* **2006**, *11*, 184–191.
181. H. Weingart; B. Volksch, *Appl. Environ. Microbiol.* **1997**, *63*, 156–161.
182. B. P. Thomma; I. A. Penninckx; W. F. Broekaert; B. P. Cammue, *Curr. Opin. Immunol.* **2001**, *13*, 63–68.
183. B. R. Glick, *FEMS Microbiol. Lett.* **2005**, *251*, 1–7.
184. J. Prayitno; N. Imin; B. G. Rolfe; U. Mathesius, *J. Proteome Res.* **2006**, *5*, 3084–3095.
185. J. Prayitno; B. G. Rolfe; U. Mathesius, *Plant Physiol.* **2006**, *142*, 168–180.
186. B. J. Ferguson; U. Mathesius, *J. Plant Growth Regul.* **2003**, *22*, 47–72.
187. J. R. Cohn; G. B. Martin, *Plant J.* **2005**, *44*, 139–154.
188. M. Knoester; J. F. Bol; L. C. van Loon; H. J. Linthorst, *Mol. Plant Microbe Interact.* **1995**, *8*, 177–180.
189. S. Dutta; R. H. Biggs, *Physiol. Plant.* **1991**, *82*, 225–230.
190. V. H. Argandona; M. Chaman; L. Cardemil; O. Munoz; G. E. Zuniga; L. J. Corcuera, *J. Chem. Ecol.* **2001**, *27*, 53–68.
191. E. A. Schmelz; H. T. Alborn; J. H. Tumlinson, *Physiol. Plant.* **2003**, *117*, 403–412.
192. J. Kahl; D. H. Siemens; R. J. Aerts; R. Gabler; F. Kuhnemann; C. A. Preston; I. T. Baldwin, *Planta* **2000**, *210*, 336–342.
193. G. Arimura; R. Ozawa; T. Nishioka; W. Boland; T. Koch; F. Kuhnemann; J. Takabayashi, *Plant J.* **2002**, *29*, 87–98.
194. M. V. Rao; H. I. Lee; K. R. Davis, *Plant J.* **2002**, *32*, 447–456.
195. W. Moeder; C. S. Barry; A. A. Tauriainen; C. Betz; J. Tuomainen; M. Utriainen; D. Grierson; H. Sandermann; C. Langebartels; J. Kangasjarvi, *Plant Physiol.* **2002**, *130*, 1918–1926.
196. H. Tuominen; K. Overmyer; M. Keinanen; H. Kollist; J. Kangasjarvi, *Plant J.* **2004**, *39*, 59–69.
197. A. Apelbaum; S. F. Yang, *Plant Physiol.* **1981**, *68*, 594–596.
198. Y. Tanaka; T. Sano; M. Tamaoki; N. Nakajima; N. Kondo; S. Hasezawa, *J. Exp. Bot.* **2006**, *57*, 2259–2266.
199. J. Beltrano; M. G. Ronco; E. R. Montaldi, *J. Plant Growth Regul.* **1999**, *18*, 59–64.
200. T. E. Young; R. B. Meeley; D. R. Gallie, *Plant J.* **2004**, *40*, 813–825.
201. Z. Zhou; E. J. de Almeida; D. Rouan; F. Michiels; M. van Montagu; D. van der Straeten, *Plant Physiol.* **2002**, *129*, 72–84.
202. L. A. Voesenek; M. B. Jackson; A. H. Toebes; W. Huibers; W. H. Vriezen; T. D. Colmer, *Plant J.* **2003**, *33*, 341–352.
203. S. I. Mekhedov; H. Kende, *Plant Cell Physiol.* **1996**, *37*, 531–537.
204. L. A. Voesenek; J. J. Benschop; J. Bou; M. C. Cox; H. W. Groeneveld; F. F. Millenaar; R. A. Vreeburg; A. J. Peeters, *Ann. Bot. (Lond.)* **2003**, *91*, 205–211.
205. X. C. Zhao; G. E. Schaller, *FEBS Lett.* **2004**, *562*, 189–192.
206. P. Grappin; D. Bouinot; B. Sotta; E. Miginiac; M. Jullien, *Planta* **2000**, *210*, 279–285.
207. R. R. Finkelstein; S. S. Gampala; C. D. Rock, *Plant Cell.* **2002**, *14* (Suppl), 15–45.
208. M. Koornneef; L. Bentsink; H. Hilhorst, *Curr. Opin. Plant Biol.* **2002**, *5*, 33–36.
209. J. Pirrello; F. Jaimes-Miranda; M. T. Sanchez-Ballesta; B. Tournier; Q. Khalil-Ahmad; F. Regad; A. Latche; J. C. Pech; M. Bouzayen, *Plant Cell Physiol.* **2006**, *47*, 1195–1205.
210. M. Gomez-Jimenez; A. Matilla, *Plant Growth Regul.* **2006**, *50*, 101–110.
211. L. Petruzzelli; I. Coraggio; G. Leubner-Metzger, *Planta* **2000**, *211*, 144–149.
212. K. Hermann; J. Meinhard; P. Dobrev; A. Linkies; B. Pesek; B. Hess; I. Machackova; U. Fischer; G. Leubner-Metzger, *J. Exp. Bot.* **2007**, *58*, 3047–3060.
213. Y. Sugimoto; A. M. Ali; S. Yabuta; H. Kinoshita; S. Inanaga; A. Itai, *Physiol. Plant.* **2003**, *119*, 137–145.
214. A. Gianinetti; L. J. Laarhoven; S. T. Persijn; F. J. Harren; L. Petruzzelli, *Ann. Bot. (Lond.)* **2007**, *99*, 735–745.
215. J. M. Locke; J. H. Bryce; P. C. Morris, *J. Exp. Bot.* **2000**, *51*, 1843–1849.
216. J. D. Goeschl; L. Rappaport; H. K. Pratt, *Plant Physiol.* **1966**, *41*, 877–884.
217. A. Lehman; R. Black; J. R. Ecker, *Cell* **1996**, *85*, 183–194.
218. H. Li; P. Johnson; A. Stepanova; J. M. Alonso; J. R. Ecker, *Dev. Cell.* **2004**, *7*, 193–204.
219. H. Fujita; K. Syono, *Plant Cell Physiol.* **1996**, *37*, 1094–1101.
220. P. M. Chilley; S. A. Casson; P. Tarkowski; N. Hawkins; K. L. Wang; P. J. Hussey; M. Beale; J. R. Ecker; G. K. Sandberg; K. Lindsey, *Plant Cell.* **2006**, *18*, 3058–3072.
221. F. Vandenbussche; W. H. Vriezen; J. Smalle; L. J. Laarhoven; F. J. Harren; D. van der Straeten, *Plant Physiol.* **2003**, *133*, 517–527.
222. F. Vandenbussche; J. Smalle; J. Le; N. J. Saibo; P. A. De; L. Chaerle; O. Tietz; R. Smets; L. J. Laarhoven; F. J. Harren; O. H. Van; K. Palme; J. P. Verbelen; D. van der Straeten, *Plant Physiol.* **2003**, *131*, 1228–1238.
223. A. N. Stepanova; J. M. Hoyt; A. A. Hamilton; J. M. Alonso, *Plant Cell.* **2005**, *17*, 2230–2242.
224. A. N. Stepanova; J. Robertson-Hoyt; J. Yun; L. M. Benavente; D. Y. Xie; K. Dolezal; A. Schlereth; G. Jurgens; J. M. Alonso, *Cell* **2008**, *133*, 177–191.
225. H. Hansen; K. Grossmann, *Plant Physiol.* **2000**, *124*, 1437–1448.
226. M. Kraft; R. Kuglitsch; J. Kwiatkowski; M. Frank; K. Grossmann, *J. Exp. Bot.* **2007**, *58*, 1497–1503.
227. D. van der Straeten; S. Anuntalabhochai; W. van Caeneghem; Z. Zhou; J. Gielen; M. van Montagu, *Plant Sci.* **1997**, *124*, 79–87.
228. H. Saika; M. Okamoto; K. Miyoshi; T. Kushiro; S. Shinoda; Y. Jikumaru; M. Fujimoto; T. Arikawa; H. Takahashi; M. Ando; S. Arimura; A. Miyao; H. Hirochika; Y. Kamiya; N. Tsutsumi; E. Nambara; M. Nakazono, *Plant Cell Physiol.* **2007**, *48*, 287–298.
229. V. N. Khryanin, *Russ. J. Plant Physiol.* **2002**, *49*, 545–551.
230. S. Saito; N. Fujii; Y. Miyazawa; S. Yamasaki; S. Matsuura; H. Mizusawa; Y. Fujita; H. Takahashi, *J. Exp. Bot.* **2007**, *58*, 2897–2907.

231. B. Kubicki, *Genet. Pol.* **1969**, *10*, 69–143.
232. S. Kamachi; H. Sekimoto; N. Kondo; S. Sakai, *Plant Cell Physiol.* **1997**, *38*, 1197–1206.
233. T. Trebitsh; J. E. Staub; S. D. O'Neill, *Plant Physiol.* **1997**, *113*, 987–995.
234. R. R. Knopf; T. Trebitsh, *Plant Cell Physiol.* **2006**, *47*, 1217–1228.
235. A. Salman-Minkov; A. Levi; S. Wolf; T. Trebitsh, *Plant Cell Physiol.* **2008**, *49*, 740–750.
236. M. Serek; E. J. Woltering; E. C. Sisler; S. Frello; S. Sriskandarajah, *Biotechnol. Adv.* **2006**, *24*, 368–381.
237. W. G. van Doorn, *Ann. Bot. (Lond.)* **2002**, *89*, 689–693.
238. W. G. van Doorn, *Ann. Bot. (Lond.)* **2002**, *89*, 375–383.
239. H. Nukui; S. Kudo; A. Yamashita; S. Satoh, *J. Exp. Bot.* **2004**, *55*, 641–650.
240. S. Satoh; Y. Kosugi; Y. Iwazaki; K. Shibuya; K. Waki, *J. Plant Biotechnol.* **2000**, *2*, 83–87.
241. E. J. Woltering; D. Somhorst; P. van der Veer, *Plant Physiol.* **1995**, *109*, 1219–1225.
242. I. Llop-Tous; C. S. Barry; D. Grierson, *Plant Physiol.* **2000**, *123*, 971–978.
243. X. Fan; J. P. Mattheis; S. Blankenship, *J. Agric. Food Chem.* **1999**, *47*, 3063–3068.
244. J. M. Lelievre; L. Tichit; P. Dao; L. Fillion; Y. W. Nam; J. C. Pech; A. Latche, *Plant Mol. Biol.* **1997**, *33*, 847–855.
245. K. Hiwasa; Y. Kinugasa; S. Amano; A. Hashimoto; R. Nakano; A. Inaba; Y. Kubo, *J. Exp. Bot.* **2003**, *54*, 771–779.
246. S. Mita; S. Kawamura; K. Yamawaki; K. Nakamura; H. Hyodo, *Plant Cell Physiol.* **1998**, *39*, 1209–1217.
247. R. Nakano; E. Ogura; Y. Kubo; A. Inaba, *Plant Physiol.* **2003**, *131*, 276–286.
248. D. J. Whittaker; G. S. Smith; R. C. Gardner, *Plant Mol. Biol.* **1997**, *34*, 45–55.
249. W. O. Owino; Y. Manabe; F. M. Mathooko; Y. Kubo; A. Inaba, *Plant Physiol. Biochem.* **2006**, *44*, 335–342.
250. A. Inaba; X. Liu; N. Yokotani; M. Yamane; W. J. Lu; R. Nakano; Y. Kubo, *J. Exp. Bot.* **2007**, *58*, 1047–1057.
251. W. S. Wong; W. Ning; P. L. Xu; S. D. Kung; S. F. Yang; N. Li, *Plant Mol. Biol.* **1999**, *41*, 587–600.

Biographical Sketches

Małgorzata Jakubowicz holds M.Sc. and Ph.D. degrees in molecular biology from Adam Mickiewicz University. She is working as an Assistant Professor in the Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland.

Witold Nowak holds an M.Sc. degree in biotechnology and a Ph.D. degree in molecular biology from Adam Mickiewicz University. He is working as an Assistant Professor in Molecular Biology Techniques Laboratory, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland.

5.05 Specific and Nonspecific Incorporation of Selenium into Macromolecules

William T. Self, University of Central Florida, Orlando, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

5.05.1 Introduction

Selenium, first discovered by Berzelius[1] and named after the goddess of the moon Selene, is a trace metalloid present in a variety of chemical forms in the environment, both inorganic and organic. Chemists have studied selenium since the nineteenth century, including Alexander Graham Bell in his early reports of the photophone at the turn of the century.[2] For the first several decades of the twentieth century, selenium was studied for its toxicity in animals and plants; nonetheless, experiments on animals in the 1950s first indicated that selenium was a micronutrient in animals.[3] In 1973 selenium was first found in both a mammalian enzyme (glutathione peroxidase) and two bacterial enzymes (glycine reductase selenoprotein A and formate dehydrogenase), and these seminal studies led to the molecular understanding of how selenium is used in biological cells.[4,5]

Since the late 1950s, it has become quite evident that selenium plays very critical roles in biology. With chemistry analogous to sulfur, selenium plays a role in redox reactions as a potent catalyst in metabolic pathways as diverse as substrate level phosphorylation for ATP synthesis in anaerobes to DNA synthesis in humans.[6] Not all biological organisms require selenium for growth or metabolism, but for those that are capable of specific incorporation of selenium into a catalytic form within proteins and enzymes, selenium is a critical micronutrient. The unique redox chemistry of selenium apparently led to its evolution in biology in even the most primitive ancestors or life,[7] based on recent computational biology and comparative genomics. Specific incorporation of selenium into macromolecules appears widespread in strict anaerobes, suggesting that an

oxygenic environment may have altered the evolutionary pressure to maintain or lose this capability. The area of genomics has certainly opened many new windows in recent years on determining the scope of organisms that have the machinery to insert selenium specifically into nucleotides and proteins.[7–12]

The vast majority of research focused on selenium in biology (primarily in the fields of molecular biology, cell biology, and biochemistry) over the past 20 years has centered on identification and characterization of specific selenoproteins, or proteins that contain selenium in the form of selenocysteine.[6,13–15] In addition, studies to determine the unique machinery necessary for incorporation of a nonstandard amino acid (L-selenocysteine) during translation also have been central to our understanding of how cells can utilize this metalloid.[16–23] This process has been studied in bacterial models (primarily *Escherichia coli*) and more recently in mammals (*in vitro* cell culture and animal models). In this work, we will review the biosynthesis of selenoproteins in bacterial systems, and only briefly review what is currently known about parallel pathways in mammals, since a comprehensive review in this area has been recently published.[24] Moreover, we summarize the global picture of the nonspecific and specific use of selenium from a broader perspective, one that includes lesser known pathways for selenium utilization into modified nucleosides in tRNA and a labile selenium cofactor. We also review recent research on newly identified mammalian selenoproteins and discuss their role in mammalian cell biology.

When one looks at the overall use of selenium in biological systems, three major pathways now emerge (**Figure 1**). Selenium can be derived from a large number of biologically relevant species, but most often has been studied using either selenite or L-selenocysteine as a nutritional source in model systems. These two forms

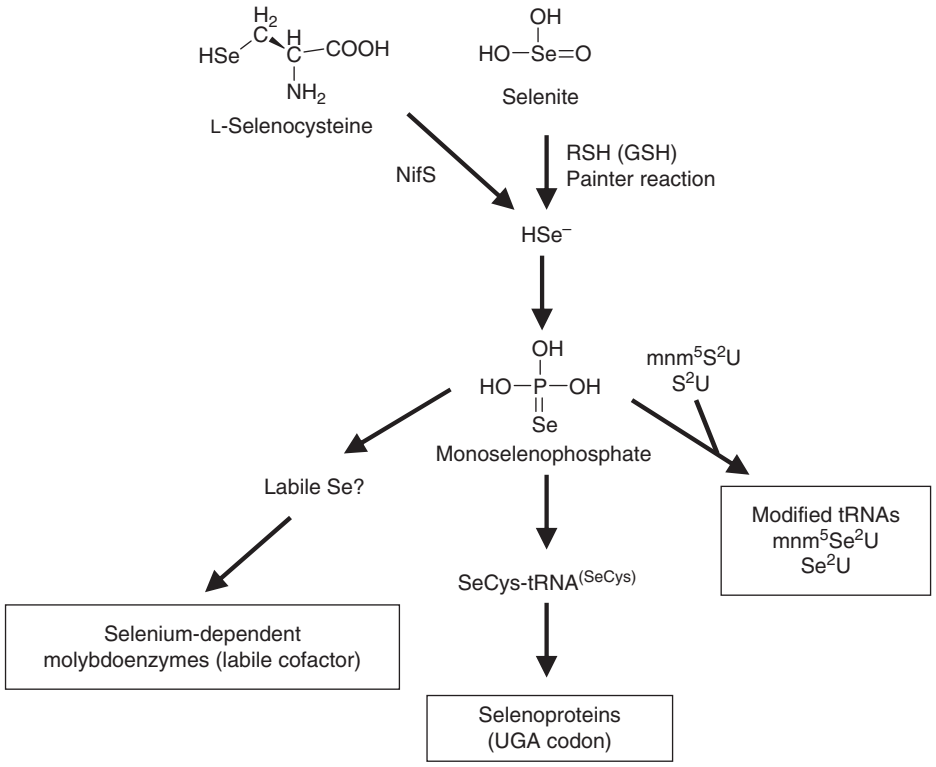


Figure 1 Overview of specific use of selenium in biological systems. Selenium can be incorporated into macromolecules in at least three separate pathways. From the reduced form of selenide, selenium is activated to selenophosphate by the action of the enzyme selenophosphate synthetase (SPS or SelD). This activated form is then used as a substrate for pathway-specific enzymes that lead to (1) insertion as selenocysteine into proteins during translation (selenoproteins), (2) incorporation into tRNA molecules as mnm^5Se^2U or Se^2U, and (3) insertion into a unique class of molybdoenzymes as a labile, but required, cofactor. The need for activation to selenophosphate has been demonstrated in all cases at the genetic and biochemical level, with the exception of the labile selenoenzymes, where activation of selenium has only been proposed based on proximity of genes within an operon encoding SPS and a molybdoenzyme.[25]

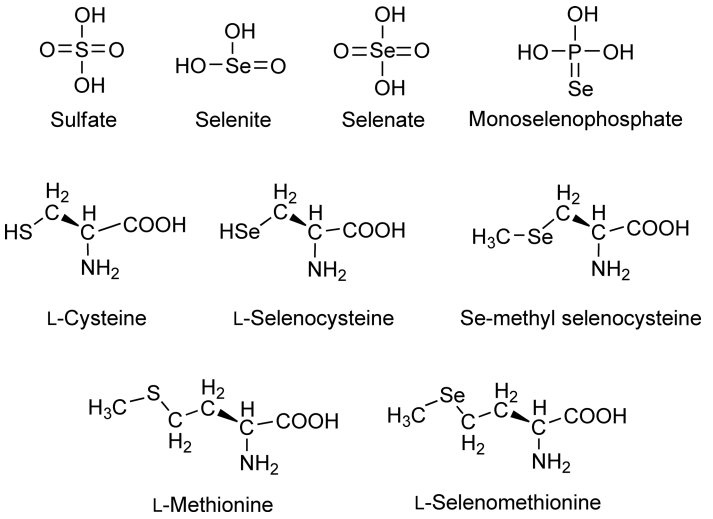


Figure 2 Selenium metabolites and critical metabolic intermediates in selenoprotein synthesis. Two-dimensional structures of each molecule are shown with common name below.

have been found to be the most efficiently used substrates for specific incorporation of selenium into macromolecules. From these forms selenium can be reduced and activated to form selenophosphate (**Figure 1**, all relevant structures shown in **Figure 2**). From this activated form, there are at least three specific pathways for the biological use of selenium. First, selenium can be incorporated into selenoproteins through translation at the ribosome, present in the polypeptide as L-selenocysteine (selenoproteins). Beyond its incorporation into selenoproteins as selenocysteine, selenium is also inserted after translation into a small set of metalloenzymes that contain a molybdenum cofactor, the so-called selenium-dependent molybdenum hydroxylases (SDMHs[9]). These enzymes are poorly understood and to date our understanding of this unique cofactor comes from biochemical studies with genetics lagging behind.[26–28] Emerging bioinformatics may propel this area in the near future.[9] Third, selenium can be specifically incorporated into a subset of transfer RNA (tRNA) molecules after sulfur is first incorporated, forming a 2-selenouridine-modified nucleotide. Although recent studies have established the proteins required for this modification,[29] the role that it plays in cellular physiology and protein synthesis has yet to be described in detail. In addition to these three specific pathways, selenium can be nonspecifically incorporated into proteins as L-selenocysteine and L-selenomethionine in place of L-cysteine and L-methionine, respectively. This nonspecific incorporation is not known to play a role in biology, but to be due to lack of specificity of enzymes in the trans-sulfuration pathway and sulfur assimilation pathway.[30] In this comprehensive review, we focus on specific versus nonspecific incorporation into proteins, as well as discussing some of the areas of selenium research that are emerging from comparative genomics.

5.05.2 Nonspecific Incorporation of Selenium into Macromolecules

5.05.2.1 Toxicology of Selenium

Selenium is a toxic metalloid that was first established as a causative agent for a condition in animals known as alkali disease.[31] During this period (1930s) the toxicity of metals and metalloids was being evaluated for animal disease and nutrition. When compared to molybdenum, tellurium, arsenic, vanadium, nickel, and tungsten, selenium was found to be the most toxic to rats given comparable doses.[32] The exact molecular mechanism to link exposure to high levels of selenium to the phenotypic symptoms of toxicity has not yet been elucidated in any model system. Nonetheless, it is clear that high concentrations of selenium are deleterious to plants, animals, and microbes. One report even suggests that selenium poisoning led to the disappearance of the dinosaurs at the end of Mesozoic era.[33]

The most well-documented case of selenium poisoning in the environment is that of the Kesterson reservoir.[34–38] This reservoir was built to drain agricultural wastewater in the San Joaquin Valley of California in the 1970s. Studies were conducted by the United States Fish and Wildlife service in the early 1980s to monitor fish species as well as aquatic birds and plants.[35] From the early data it was clear that aquatic birds such as Pied-billed grebe, American coot, mallard, gadwall, and cinnamon teal were all being affected negatively by selenium and other metals, especially with respect to reproduction capability.[35] The levels of selenium in food-chain organisms eaten by these birds (algae, rooted plants, plankton, aquatic insects) averaged from 20 to 332 $\mu g\,g^{-1}$ and were generally up to 100 times higher than a nonpolluted reference site (Volta wildlife area). Concentrations of several metalloids and metals (As, Cd, Cr, Cu, Hg, Ni, Pb, and Zn) all were elevated in the Kesterson Reservoir water, alongside higher levels of selenium. This well-documented ecological system has proven to be a model for retrospective analysis of metal poisoning to this day.

The chemical nature of the toxicity in this and other animal models remains somewhat of a mystery. One must first realize that selenium can be found in both inorganic forms (selenate and selenite), and many different organic forms (selenomethionine, selenocysteine, selenomethyl selenocysteine, **Figure 2**). Although selenium toxicity has been studied in a variety of biological models, one primary theme has emerged: selenium present in chemical forms that can efficiently give rise to high levels of hydrogen selenide (HSe^-, **Figure 1**) or abundant small molecule selenols (RSeH) leads to toxicity due to production of superoxide from redox cycling reactions.[39–41] Although it has been established that selenate, a sulfate analogue, and L-selenomethionine are less toxic among the best-studied forms of selenium, one must be cautious of the cell or cell type that is used to monitor this effect. Specifically, selenomethionine has been shown to be far less toxic in mammalian cell culture models,[41] yet is also a poor form of selenium for selenoprotein synthesis.[42] The conversion of selenomethionine into the precursor for selenoprotein synthesis, hydrogen selenide, requires the trans-sulfuration pathway.[43] Many current clinical trials studying the nutrition of selenium use selenized yeast that contains primarily selenomethionine. This form can result in an apparent increase in total plasma levels, but this is probably due to increased nonspecific incorporation of L-selenomethionine into albumin.[43,44] It is not yet known whether increased selenomethionine introduced in a nonspecific manner could underlie some of the toxic effects of selenium in biology. Taken together, there appear to be two major possible hypotheses to account for toxicity: (1) reactive oxygen species generation from redox cycling reaction from hydrogen selenide or similar reactive selenols and (2) nonspecific incorporation of selenium into amino acid residues – cysteine and methionine as selenocysteine and selenomethionine. This gap in our understanding is important since selenium is being used in clinical trials in arsenic-exposed populations and also tested for its efficacy to reduce cancer rates.[24,45–48]

5.05.2.2 Bioremediation of Selenium

The presence of dangerously high concentrations of selenium in soil, sediment, and aquifers has led to a number of research groups focusing on chemical and biological ways to metabolize selenium into a less toxic form or remove selenium by volatilization.[49–59] Selenium is present in soils and sediments in inorganic forms selenite (SeO_3^{2-}) and selenate (SeO_4^{2-}).[60] These anions dominate in aerobic environments, with elemental selenium present as the major form of selenium in anoxic aqueous environments, as well as anaerobic sediments and soils. The reduction of selenate primarily occurs by biotic processes, and the enzymes responsible for this reduction have been characterized.[61–64] A well-studied model organism, *Thauera selenatis*, which was isolated in the San Joaquin Valley in California,[65] has been described both at the bacterial level and the enzyme responsible for selenate reduction has been purified and characterized.[66] Selenate reductase is a trimeric enzyme that contains molybdenum, iron, and acid labile sulfur. Most notably the enzyme is specific for selenate, with an apparent K_m of 16 $\mu mol\,l^{-1}$. The enzyme complex does not reduce nitrate, a trait somewhat unique to this particular bacterium since many of organisms that reduce selenate are also efficient reducers of nitrate and nitrite.[64] It is clear that under evolutionary pressure, selenium-specific reduction can be obtained in the natural environment.

Selenite reduction is a far more complex situation, since selenite can itself react with reduced thiols in a reaction known as the Painter reaction.[67,68] Painter first described the reaction of selenite with cysteine, leading to production of the reduced form hydrogen selenide.[69] Ganther expanded upon this early work using glutathione as a thiol reductant, describing the reactions in a physiological context and suggesting this chemistry was fundamental to selenium metabolism.[68,70,71] Briefly, when selenite is incubated with thiols at a ratio of 1:4 (Se to S ratio),

selenotrisulfides (RSSeSR) and oxidized thiols (RSSR) are formed at approximately a 1:1 ratio. This chemistry has been later applied to other thiols for formation of unusual selenotrisulfides, such as penicillamine and alpha-lipoic acid.[42,72–74] If excess reduced thiol is present, then the selenotrisulfides are further reduced to form hydrogen selenide. Under aerobic conditions that hydrogen selenide can oxidize to form elemental selenium. This property has been utilized recently to form nanoparticles of selenium, potentially with therapeutic use.[75]

Given this reactivity, it is difficult to assess the role that biology plays in the *in situ* reduction of selenite.[60] *Bacillus selenitireducens* is a unique microorganism that has been characterized to respire on selenite for growth, using lactate or pyruvate as an electron donor.[76] In all organisms studied to date, the reduction of selenite occurs on the outside of the cytosol, perhaps at the cell's surface.[60] No definitive biochemical data exist to suggest that a selenite-specific enzyme, that is, an enzyme that prefers selenite over another anion such as nitrate, is found in any organism. However, a more recent study has identified an organism, *Stenotrophomonas maltophilia*, capable of efficiently converting both selenate and selenite to methylated (dimethyl diselenide) and volatile alkylselenides.[55] As with other similar organisms, this strain was isolated in selenium-rich drainage area of California. Overall our understanding of selenium oxyanion reduction, at the bacterial and biochemical level, has expanded rapidly over the past 10 years, yet whether this response was due to an evolutionary pressure to use selenium as an electron acceptor has yet to be firmly established.

5.05.2.3 Bacterial Metabolism of Selenium as a Sulfur Analogue

The best-studied model system for biological use of selenium remains *E. coli*. Using ^{75}Se as a tracer, studies found that selenium was incorporated both into specific selenoproteins such as formate dehydrogenase H (FDH-H) as well as nonspecifically into many other proteins based on gel electrophoresis.[77] Although it was clear that selenium was being incorporated into nearly all the proteins in the cell, the question remained early on as to whether this metalloid was incorporated nonspecifically in methionine and/or cysteine residues. Moreover, the metabolism of inorganic selenium (selenite) or organic selenium sources (selenocysteine or selenomethione) upstream of sulfur-containing amino acid biosynthesis certainly affects the final distribution of selenium in the organism. Many studies have shown that addition of large amounts of selenomethionine during overexpression of target proteins can result in efficient incorporation of selenium to assist in phase determination for X-ray crystallography.[78] However, only one clear study using *E. coli* as a model has tried to address open questions about metabolism of selenium when given as an inorganic precursor.[77]

Unexpectedly, the incorporation of selenium nonspecifically (i.e., incorporation into proteins not containing a specific UGA coded selenocysteine residue) was efficiently blocked by simple titration with L-cystine. Methionine addition contributed little to alter selenium distribution into proteins. This competition also was efficient in the absence of most of the components of the sulfate reduction pathway (conversion of sulfate into sulfide). The nonspecific incorporation of selenium into proteins was dependent on the sulfurylation step to *O*-acetylserine (cysteine biosynthesis from sulfide[77]). Later studies have shown that blocking both specific incorporation of selenium into selenoproteins (*selD* mutant), combined with a metabolic blockade of cysteine biosynthesis (*cysK* mutant), leads to a more than 20-fold reduction in the incorporation of selenium in this model.[79] Therefore, it is clear that within this model the nonspecific incorporation of selenium into proteins (when present as an inorganic form) relies on the cysteine biosynthetic route. At least in the case of *E. coli*, nonspecific incorporation into proteins is directly due to the cell's inability to differentiate hydrogen sulfide from hydrogen selenide in the presence of high concentrations of exogenous selenium.

Each bacterium may have a slightly different propensity to incorporate selenium nonspecifically, depending on the substrate specificity of enzymes from cysteine biosynthesis or methionine synthesis. An unusual finding over more than two decades ago remains a mystery. The acetoacetyl thiolase from *Clostridium kluyveri* was found to incorporate selenium nonspecifically into methionine residues when selenium was given either as selenite or selenomethionine.[80] When methionine was given in excess, the ratio of selenium to sulfur in the immunoprecipitated protein was reduced. However when selenium was given as radiolabeled selenite, addition of methionine had no effect on the level of selenium in the enzyme.[80] A clear explanation for this result, in light of the findings in *E. coli* model several years later, is lacking. Essentially, a selenomethionine-resistant phenotype exists with respect to nonspecific incorporation of selenium. Perhaps this unanswered question gives credence to the notion that some specific use of selenomethionine exists, albeit not yet understood.

5.05.2.4 Selenium Metabolism in Plants

Although there is no evidence that terrestrial plants in general utilize selenium as a micronutrient, the metabolism of selenium in plants is critical to nutrition of animals and vertebrates that produce specific selenoproteins. Moreover, a select group of plants are capable of accumulating high concentrations of selenium even in soils with low levels of selenium. These plants, with the best studied model being *Astragalus bisulcatus*, are known as selenium hyperaccumulators.[81] Even without specific selenoproteins, plants play a vital role in the cycle of selenium from the soils and sediments, aquatic selenium levels and nutrition and toxicity in vertebrates. First we discuss uptake and metabolism of selenium in plants, then turn our attention to recent work that suggests that metabolic engineering may hold hope for bioremediation as well as improved human health.

Selenium is abundant in soil in the form of selenate and selenite. Selenate is taken up by roots through a sulfate transporter and is quickly transported from the roots to the shoots.[82] This is supported by the isolation of selenate-resistant mutants of *Arabidopsis thaliana* that contain mutations in the sulfate transporter.[83] Selenite is also accumulated but apparently not by an active process.[82] Once taken up, selenate is transported to the chloroplast and treated as a sulfate analogue (based on knowledge gained on sulfate metabolism), being assimilated through the sulfate assimilation pathway. In one study of the Indian mustard (*Brassica juncea*, a selenium hyperaccumulator), selenate-treated plants only converted about 18% of the selenium into organic forms with the remaining selenium retained as selenate.[84] In contrast when selenium was given as selenite, it was rapidly converted into organic forms of selenium, including selenomethionine, *S*-(methylseleno)cysteine and methylselenocysteine with only 4.3% remaining as selenite. This study demonstrates that plant metabolism of selenium is highly dependent on the form present in the soil and subsequently taken up to the shoots. It also highlights the central nature of hydrogen selenide to most biological metabolism of selenium (**Figures 1** and **3**).

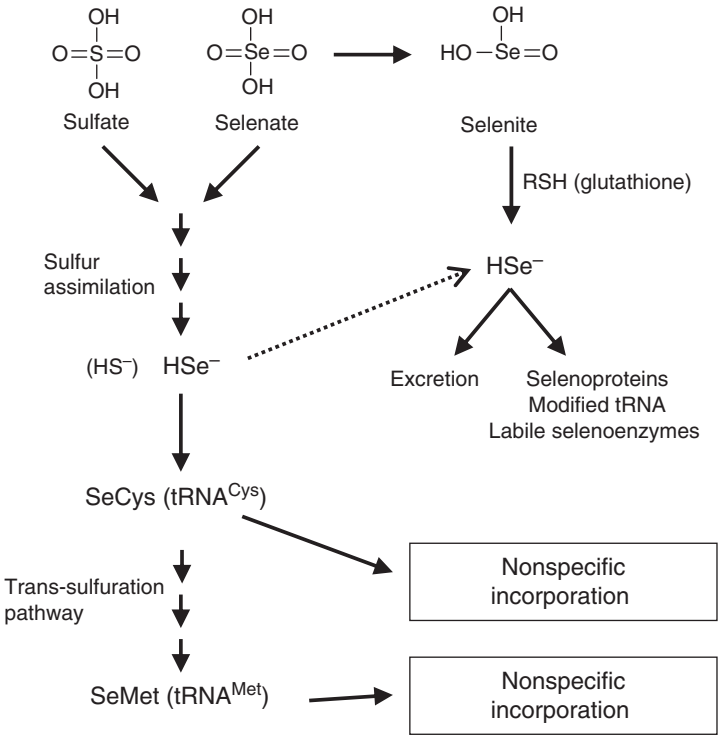


Figure 3 Specific and nonspecific pathways for incorporation of selenium into macromolecules in biological systems. Selenium derived from inorganic sources (soil, sediment) are incorporated into specific selenoproteins through the chemical conversion of selenite (through the Painter reaction) to hydrogen selenide. This is the precursor for activation of selenium to selenophosphate, leading to three specific biological macromolecules (see **Figure 1**). Selenate, in most organisms, is metabolized through sulfur assimilation pathways to hydrogen selenide and then further metabolized as an analogue of sulfur nonspecifically into selenocysteine or selenomethionine. Conversion to selenomethionine requires the presence of the trans-sulfuration pathway.

Early studies demonstrated that the toxicity of selenium in plants is dependent on incorporation of selenium into selenocysteine and selenomethionine in proteins.[85,86] It should be noted that this stands in contrast to the hypothesis that selenium toxicity is due to oxidative stress from redox cycling of hydrogen selenide.[87,88] The main metabolic difference that has been observed between selenium hyperaccumulating plants such as *Astragalus* sp. and other plants lies in the ability to methylate the free selenium-containing amino acids to prevent them from being incorporated nonspecifically into proteins.[81] When a comprehensive comparison of sulfate assimilatory enzymes and selenocysteine methyltransferase (SMT) was carried out in a variety of plant species, one clear trend arose from the data. The hyperaccumulation of selenium and sulfur-containing amino acids (methylcysteine and methylselenocysteine) correlated well with the activity of SMT[89] in plants that accumulated large quantities of selenium. Further experiments are needed to be sure that all the necessary components of the pathway for selenium-containing amino acids have been identified, yet the recent work on SMT certainly sheds light on the biochemistry of hyperaccumulators. The authors believe that an increase is sulfur and selenium flux through the sulfur assimilation pathway is the basis for the hyperaccumulation of selenium-containing amino acids. Future studies to understand this pathway could lead to metabolic engineering of such a pathway in a rapidly growing plant species that could be used to remove excess selenium from contaminated soil, such as is found in the Kesterson reservoir sediments. This has been raised by several authors in recent reviews.[90–92]

5.05.3 Specific Utilization of Selenium in Biology

In contrast to the nonspecific incorporation of selenium into amino acid metabolites that occurs in plants, and the nonspecific replacement of sulfur with selenium in methionine and cysteine residues, there are three well-established specific uses of selenium in biology. The first, and best characterized, is the incorporation of selenium into the polypeptides as selenocysteine. This specific incorporation requires cis elements and trans elements including unique elongation factors. This type of selenoprotein can be found in organisms from all three domains of life, but its presence is certainly not ubiquitous. The second specific use of selenium is the incorporation of selenium into a modified uridine residue on specific tRNA species. This pathway for insertion of selenium into this nucleotide has been studied for more than 30 years, and recent publications have identified the enzymes necessary to carry out this modification. Nonetheless, to date, we are quite unsure as to the absolute role this modification plays in cellular physiology. The third specific use of selenium is the most poorly understood of the three. Selenium is apparently incorporated after translation into a subset of molybdoenzymes and is required for enzyme activity. Computational biology has now laid the groundwork to determine how this posttranslational modification is carried out, but this work is ongoing. How widespread the use of selenium as a catalyst is in this form (labile cofactor) is an open question that needs to be answered. In the following sections, we first discuss the general pathway for specific synthesis of selenoproteins, which follow reviews of each of the best studied prokaryotic (eubacterial) selenoproteins. Later we discuss recent research on newly identified mammalian selenoproteins, many of which have implications in human health and disease. Finally, we will include the use of selenium in modified nucleotides (tRNA) and the current knowledge of the use of selenium as a labile cofactor.

5.05.3.1 Selenoprotein Synthesis – Overview

The first bacterial selenoprotein identified, selenoprotein A, was identified as a subunit of the glycine reductase complex isolated from the anaerobe *Clostridium sticklandii*.[5] **Table 1** lists the selenoproteins that have been characterized at the biochemical level from prokaryotes. Although much of the early biochemistry to determine that selenium was present in the form of selenocysteine was carried out in this model,[99,104] the specific incorporation of selenium was led using the far more genetically pliable model of *E. coli*.[108] Early bacteriological studies suggested a role for selenium in formate dehydrogenase of the *E. coli* formate hydrogenlyase,[109] and enzyme complex required for the anaerobic production of hydrogen gas during fermentation of sugars. A key discovery occurred when a series of mutants were identified that lacked formate dehydrogenase activity.[22] Complementation of these mutants led to the cloning and sequencing of the genes encoding the proteins and

Table 1 Selenium-containing enzymes in prokaryotes (Eubacteria and Archaea)

Protein name	*Role*	*Model organism*	*References*
Formate dehydrogenase (several types)	Sugar fermentation	*Escherichia coli*	93–95
Selenophosphate synthetase	Selenoprotein synthesis	*Escherichia coli*	96–98
Glycine reductase selenoprotein A	Energy metabolism	*Clostridium sticklandii*	5, 99, 100
Glycine reductase selenoprotein B	Energy metabolism	*Eubacterium acidaminophilum*	101–103
D-proline reductase	Anaerobic respiration	*Clostridium sticklandii*	104, 105
Hydrogenases (several types)	Energy metabolism	*Methanococcus jannaschii*	106
Formylmethanofuran dehydrogenase	Methanogenesis	*Methanopyrus kandleri*	107

RNA component (a unique tRNA) needed for selenoprotein biosynthesis. The genes encoding these components were later named *sel* for their need to produce an active selenoprotein (*selAB, selC*, and *selD*).

The *selC* gene, which encodes the selenocysteine-specific transfer RNA, $tRNA^{sec}$, was found to be first charged with serine, an activity catalyzed by the SerS protein.[19] The *selD* gene encodes the enzyme that activates selenium to a phosphorylated form, selenophosphate synthetase (SPS). SPS activates reduced selenium (hydrogen selenide, HSe^-) to selenophosphate ($SePO_3$) utilizing ATP as the phosphoryl group donor.[18,96,110,111] The enzyme encoded by the *selA* gene catalyzes the conversion of a seryl-$tRNA^{sec}$ into the selenocysteinyl-$tRNA^{sec}$ using selenophosphate as a selenium donor. The specific elongation factor, SelB, is a trans-acting element that recognizes the cis element (termed SECIS element) present just downstream of the alternate codon, UGA.[112] This SECIS element is a stem–loop structure that binds specifically to SelB to prevent truncation of the polypeptide.[17,113] Each of these four components have been shown to be required for the insertion of selenium into the specific selenocysteine residue in the formate dehydrogenase enzyme FDH-H. It should be noted that within this genetic screen no mutations in selenium transport or further metabolism of selenium prior to reduction to hydrogen selenide have been found. To date, no specific transport system has been identified although selenium can be efficiently taken up by cells even when present in a defined culture medium in the low nanomolar range.[114]

This pathway was defined more than two decades ago, but more recent studies have shed light on the regulation of the biosynthetic machinery. The level of each of the components (Sel gene products) is thought to be limiting, and biotechnological approaches to express mammalian selenoproteins in bacteria has proven challenging.[115,116] This is probably due in part to the fact that a SECIS element was found to reside within the promoter/operator region of the operon that encodes the SelA and SelB proteins.[117] The binding and repression of this mRNA by a low-affinity SECIS element allow the cell to control the production of these proteins, since the overexpression of an elongation factor could negatively effect the ribosome and translation of nonselenocysteine containing proteins. Whether this kind of automatic control exists in other prokaryotic or eukaryotic models is not yet known.

Mammals: The translation of the UGA codon for insertion of selenocysteine into mammalian selenoproteins is not as well described, when compared to the known bacterial system (*E. coli*), yet it is clear that nearly all or all of the necessary components have been identified over the past 10 years. Briefly, studies have demonstrated that the efficient translation of a UGA codon requires an RNA stem–loop structure located in the 3′ untranslated region of the mRNA[118–120]) as well as the SBP2 and EFSec proteins that are best described to play a role similar to that of the bacterial protein SelB.[121,122] The SBP2 protein, which was identified by cross-linking experiments using the SECIS element from the mRNA of PHGpx,[119] has been purified from testicular extracts where it is highly expressed.[118]

The stem–loop structure in the noncoding 3′ region of selenoprotein mRNAs has also been termed a SECIS element in mammals although it has a different overall structure.[10,123] *In silico* analysis of the human genome sequence, using this consensus SECIS element along with the presence of the characteristic UGA codon within an exon, has led to the discovery of several new selenoproteins, including a selenium-dependent methionine sulfoxide reductase.[124] It has been shown that a specific complex exists for selenoprotein synthesis that shuttles between the nucleus and the cytosol.[125] This possibly protects the preformed complex for nonsense-mediated decay to allow for more efficient selenoprotein synthesis. The specific tRNA needed for selenocysteine

incorporation ($tRNA^{sec}$) has been identified,[126] and mutation within this tRNA molecule results in embryonic lethality in mice.[127] Unlike the bacterial model, the seryl-charged $tRNA^{sec}$ first requires phosphorylation prior to reaction with selenophosphate. The kinase responsible for this phosphorylation was recently identified.[128]

Unlike *E. coli* and other prokaryotes, mammals express two proteins that appear to be homologues of the SelD protein (selenophosphate synthetase or SPS). Nonetheless, even in recent studies the role for each of these apparent isoenzymes is open to debate. Even more intriguing is the fact that one of these isoenzymes (SPS2) is itself a selenoenzyme with a selenocysteine residue at its active site.[10] The selenocysteine containing form (SPS2) was expressed using radioisotope ^{75}Se in *Drosophila*,[129] and found to catalyze the same reaction as does the cysteine-dependent *E. coli* enzyme. The cysteine mutant of the mouse SPS2 (SeCys to Cys) has been expressed and purified,[130] and likewise was active with the same substrates and produced selenophosphate. The human SPS2 selenoenzyme (containing selenium) has not been studied *in vitro*, due to difficulties in efficiently expressing selenoproteins.[116] The role for selenium in SPS2 (human), or any selenium-containing SPS has not yet been determined.

Human SPS1, cloned from a cDNA library, was shown to increase the levels of deiodinase (a human selenoprotein, see **Table 2** for a complete list of human selenoproteins) upon transfection of HTta-1 cells,[175] demonstrating this protein can function in selenoprotein synthesis even though it does not carry a CXC motif (**Figure 2**). However this cDNA weakly complements an *E. coli selD* mutant.[175] The nature of this weak complementation is not yet known. A more recent biochemical characterization of SPS1 and SPS2 has shown that only SPS2 can catalyze the selenide-dependent production of selenophosphate from ATP.[176] SPS1 had no catalytic activity in this study, except for significant ATP hydrolysis. It should be noted that SPS1 was also

Table 2 Selenium-containing proteins in humans

Protein name	*Location and role (based on experimental evidence)*	*References*
Glutathione peroxidase 1	Cytosolic – protection against peroxides	4, 131
Glutathione peroxidase 2	Gastrointestinal (GI-Gpx) – acts against organic hydroperoxides possibly also acting against inflammation in GI tract	132, 133
Glutathione peroxidase 3	Plasma (PGpx) – second most abundant selenoprotein in plasma, role is unknown	134, 135
Glutathione peroxidase 4	Phosphohydrolipid Gpx (phgpx) – can be present in cytosol and mitochondria – functions to reduce lipid hydroperoxides	136–138
Glutathione peroxidase 6	Function unknown, studied only at mRNA level and bioinformatics	10
Thioredoxin reductase 1	Cytosolic – TrxR1 – functions to reduce Trx and other dithiols	13, 139–141
Thioredoxin reductase 2	Mitochondria – reduction of Trx as e-donor for antioxidant enzymes	142
Thioredoxin reductase 3	Thioredoxin glutathione reductase (TGR) – capable of reducing GSSG due to glutaredoxin domain	143
Deiodonase 1	Plasma membrane – reductive removal of iodine from thyroid hormones	144–146
Deiodonase 2	ER membrane – reductive removal of iodine from thyroid hormones	147
Deiodonase 3	Deiodonates the 5′ position in tyrosyl ring of L-thyroxine	148
Selenophosphate synthetase 2	Catalyzes the activation of hydrogen selenide to selenophosphate	149, 150
Selenoprotein P	Plasma selenoprotein – selenium mobilization to tissues and organs	134, 151–153
Selenoprotein W	Present in muscle and brain – function unknown – binds glutathione	154–157
Selenoprotein H	Gpx activity – located in the nucleus – function unknown	158
Selenoprotein I	Expressed in several tissues – function unknown	10
Selenoprotein K	Expressed in heart – localized to ER – function unknown	159
Selenoprotein M	Localized in ER – function unknown	160
Selenoprotein N	Localized in ER – mutations linked to congenital muscular dystrophies	161–163
Selenoprotein O	Unknown	10
Selenoprotein R (MsrB)	Several isoforms – methionine-*R*-sulfoxide reductase	164–169
Selenoprotein S	ER – involved in protein folding – first identified as VIMP	170
Selenoprotein T	ER – neuronal development and differentiation	171
Selenoprotein V	Unknown	10
Sep15 (15 kDa selenoprotein)	ER – highly expressed in brain and other tissues	172–174

found in a large complex of proteins involved in selenoprotein synthesis, and SPS2 was not.[177] The precise role of SPS1 is clearly not yet known and deserves future study given its presence in the nucleus.

A report by Tamura *et al.*[178] demonstrated that the two human SPS enzymes may each play a different role, based on the selenium substrate available. Using *E. coli selD* mutant strain as a model system, they demonstrated better complementation by SPS1 in the presence of cells cultured with L-selenocysteine. A cysteine mutant form of human SPS2 complemented the *selD* mutation better in the presence of selenite. Although the enzymes were not purified and tested *in vitro*, this study provides intriguing evidence that SPS1 and SPS2 may function to separate an inorganic pathway (selenide) from an organic (selenocysteine) pathway for selenium utilization. Biochemical analysis of these human isoenzymes is critical in defining a role for each of these enzymes in selenium metabolism. It is unlikely, given the existing data, that SPS1 does not play any role in selenium metabolism even given the established fact that it does not produce selenophosphate from hydrogen selenide.[149,179] This is just one of many unanswered questions in the metabolic pathway for insertion of selenium into selenoproteins in mammals.

5.05.3.2 Selenoproteins

5.05.3.2.1 Eubacterial selenoproteins

5.05.3.2.1(i) Stickland reactions – overview Stickland first described the coupled fermentation of two amino acids in which one is oxidatively deaminated or decarboxylated (Stickland donor) and another amino acid is reductively deaminated or reduced (Stickland acceptor) using cell extracts of *Clostridium sporogenes*.[180] This coupled amino acid fermentation has been established as a primary metabolic pathway for energy metabolism of several nontoxigenic organisms (*C. sporogenes, C. sticklandii, Eubacterium acidaminophilum*).[104,181–185] Building upon Stickland's early research, H. A. Barker's group used redox-active dyes such as methylene blue to identify the amino acids that can act as efficient Stickland donors and acceptors. For these studies Barker's group used cell extracts of *C. sporogenes*.[182] Aliphatic amino acids (leucine, isoleucine, and valine) were found to be the best electron donors (Stickland donors). The most efficient Stickland acceptors in *C. sporogenes* were identified as glycine and proline. Subsequent work using *C. sticklandii* in Stadtman's group showed that a purified proline racemase could allow L-proline to serve efficiently as a Stickland donor, but that D-proline was the actual substrate when proline served as a Stickland acceptor.[185] This strain has been the best studied model for Stickland reactions since the 1950s both for bacteriological and biochemical analysis of these amino acid fermentation reactions.[99,100,104,105,181,185–189] It should be noted that the initial characterization of *C. sticklandii* revealed that the closest related strain was the current emerging pathogen *Clostridium difficile*.[190] More recent studies using *E. acidaminophilum* as a model system for Stickland reactions have also identified the use of glycine derivatives (betaine, sarcosine) as Stickland acceptors, using similar core components of the glycine reductase.[181]

For several decades, a requirement for selenium was not identified in these studies. Selenium can contaminate other salts in trace amounts and this contamination possibly kept researchers in the dark for the need for this metalloid. Research on the glycine reductase (GR) enzyme complex in *C. sticklandii* led to the discovery of the first selenium-containing protein, termed selenoprotein A.[5] Large-scale fermentations of *C. sticklandii* showed increased enzyme activity in extracts when sodium selenite was added to the growth medium. This was the first clue that selenium played a role in Stickland fermentations. In this seminal paper it was shown that GR protein A was labeled with radioisotope ^{75}Se, and it was later determined that the chemical form of selenium in GR was selenocysteine.[99,191] Further studies determined this selenocysteine residue was encoded by a UGA codon – normally a stop codon.[192] August Böck and his colleagues, in studies parallel to the glycine reductase selenoprotein A, determined that a stop codon, UGA, encoded selenocysteine in the *E. coli* formate dehydrogenase.[114] Using these two model systems, the foundations were laid for the study of selenoproteins in all of biology.

Once identified as a selenoprotein in this model (*C. sticklandii*), the need for selenium was also shown for *C. sporogenes*.[184] The addition of selenium to the culture medium was reported to improve the level of D-proline reductase activity as early as 1976,[104] yet the first identification of the selenoprotein component of this enzyme did not occur until more recently in 1999 by Andreesen's group.[105] It is quite clear now from data from these model systems, as well as from DNA sequence analysis of the *grd* and *prd* operons,[101,105,188,193–195] that Stickland reactions are common to many amino acid-fermenting clostridia. Those that are capable of proline reduction all

produce the end product δ-aminovaleric acid, which has been a marker for the classification of the so-called group I clostridia for nearly 40 years.[196] This group includes the toxin producers *Clostridium sordelli, C. difficile*, and *Clostridium botulinum.*

In addition to bacteria with a role in amino acid and nucleotide fermentation, one oral pathogen has been shown to use Stickland fermentations as a primary energy source.[197] Several decades of studies in oral microbiology have shown a direct association between the level of spirochetes in plaques and periodontal disease.[198] Much of the research on these organisms has focused on their ability to degrade the extracellular matrix, and on nutritional interactions and associations with *Porphyromonas gingivalis*. One report found a direct synergistic relationship between these pathogens in a study of human subjects, suggesting that *Treponema denticola* requires *P. gingivalis* for infection.[199] Although little research has been carried out on energy metabolism in *T. denticola*, one key study has shown that selenium is required for growth in a minimal medium.[197] Based on the reported genome sequence,[200] it is clear that this requirement is for the synthesis of glycine reductase selenoproteins. Since many of the uncultured spirochetes in the oral cavity are thought to be closely related to *T. denticola*, and that these spirochetes act synergistically with *P. gingivalis* to promote periodontal disease, it is probable that Stickland fermentations play a critical role in energy metabolism in this niche environment. However, the difficulty in culturing and using these organisms as a model possibly precludes in-depth analysis of this pathway in oral biology.

5.05.3.2.1(ii) Glycine reductase The reduction of glycine was first studied biochemically by Stickland, and this early study (in cell extracts) that ammonia and acetate were produced in equimolar quantities to the glycine added.[201] Further studies on Stickland reactions were carried out in H. A. Barker's laboratory, and a student, Thressa Stadtman carried on this project upon establishing a research program at the National Institutes of Health in the early 1950s. Stadtman first attempted purification of an enzyme complex that could catalyze the glycine-dependent production of ATP from ADP and inorganic phosphate using various chemical electron donors. These purifications were quite tedious, as the GR complex separated into three fractions that were discernable only by recombining fractions from chromatography steps. These fractions were then termed fractions A, B, and C.[5] The discovery that the activity of GR in extracts was significantly improved upon addition of selenium to the culture medium, and that the level of component A increased as well with added selenium, turned out to be the critical link that resulted in the purification of selenoprotein A.[5] It is ironic of course that since then this nomenclature (selenoprotein) has since been adopted for proteins that contain selenium. This first selenoprotein was not named 'A' to denote it as the first selenoprotein but to distinguish it as the component A of the GR enzyme complex.

Soon after this exciting discovery (which coincided with the discoveries of selenium in glutathione peroxidase and formate dehydrogenase in other labs in 1973), the chemical component of GR selenoprotein A was determined to be selenocysteine.[99] This newly minted 21st amino acid was also found to be exquisitely sensitive to oxygen, and to have an absorption maximum of 238 nm.[191] Biochemical characterization of component B uncovered a carbonyl group and further established in purified components the stoichiometric production of 1 mol of ATP from 1 mol of glycine, yielding 1 mol of ammonia and acetate.[202] *Clostridium sticklandii* could also use proline as a Stickland acceptor, and the uptake of glycine was actually shown to be inhibited by the presence of hydroxyproline or proline in the culture medium, indicating that the cell prefers to use one or another Stickland acceptor, but not both simultaneously.[203]

Comparative studies with purine-fermenting organisms that couple the oxidation of purines to the reduction of glycine were also carried out to better understand coupled fermentations. GR selenoprotein A from *Clostridium purinolyticum* and *C. sticklandii* had similar amino acid composition, spectral properties, and the same number of thiols.[187] Although the discovery that the stop codon (UGA) was coding for selenocysteine arose in *E. coli*, this was also confirmed by cloning and sequencing of the glycine reductase selenoprotein A in *C. sticklandii* and *C. purinolyticum.*[188,204] Certainly, the most significant finding in terms of cloning and DNA sequencing of the genes encoding these enzymes occurred when the complete region of genes encoding the GR complex uncovered the presence of two separate protein components with selenocysteine residues in *Clostridium litorale.*[195] This was further confirmed in *E. acidaminophilum* and *C. sticklandii*[101] has been subsequently confirmed in genome sequencing of *C. difficile* and *C. botulinum* among many others.[205] Sequencing the region also suggested a role of thioredoxin in the system as the biologically relevant electron donor, and the

colocalization of thioredoxin reductase in each of these reports further determines the role of pyridine nucleotide identified very early in Stickland and Barker's work.[181,206,207]

The mechanism of the GR enzyme complex was first suggested in studies by Rudnick and Abeles.[208] The critical discovery came with the understanding that acetate kinase was likely present as a contaminant and was converting acetyl-phosphate into ATP and acetate. The true end product of the GR complex was then found to be acetyl-phosphate. Although a substantial study at the biochemical level has not been undertaken since the realization of the presence of two selenoproteins and Abeles mechanism, recent work has speculated on the reaction mechanism given all the evidence in the literature.[209] A selenium-dependent peroxidase activity has also been identified to be present in the selenoprotein B component in the *E. acidaminophilum* model.[210] The role for this selenoprotein is clearly not specifically oxidative stress in general but has been proposed to be to protect the oxygen-sensitive selenocysteine residue of the abundant and critical enzyme complex. To date, biochemical characterization of the GR complex has been hindered by this oxygen sensitivity and by the purification of the intact complex, yet biotechnology may yet allow for heterologous expression of these components in the near future.[194]

5.05.3.2.1(iii) D-Proline reductase Early studies by Stickland found that the reduction of proline led to a terminal product, δ-aminovaleric acid, to be excreted into the culture medium.[201] The earliest purification of the PR complex came again from Stadtman's group using *C. sticklandii* as a model revealing a role for a racemase to interconvert L-proline and D-proline.[185] This study showed clearly that the proline reductase enzyme used D-proline as a substrate and that two equivalents of thiols (RSH) were used as electron donor. The physiological electron donor is unknown. Nearly 20 years after this initial work, Seto and Stadtman reported the purification of a homogenous preparation of D-proline reductase and found that dithiothreitol could serve as a good electron donor for the enzyme, yet NADH had no activity.[104] There was also no detectable molybdenum or iron and, possibly due to oxygen inactivation, there was no detectable selenium as well. In fact, D-proline reductase was found to contain selenocysteine based on biochemical and genetic analysis more than 20 years later in *C. sticklandii*.[105] This study also showed that one of the two protein components of this enzyme complex also is autocatalytically cleaved to yield a pyruvoyl group from a cysteine residue. Our group has recently shown that D-proline reductase is expressed in *C. difficile*, and that the enzyme does not require a divalent cation (magnesium or similar) for activity.[205] We also showed that a significant level of the protein precursor (prior to autocatalytic cleavage) is present in the purified preparation. Why such a significant level of this precursor protein exists in an active enzyme complex is a mystery.

Even with the new identification of this selenocysteine many unanswered questions remain. Specifically, what triggers the reductive autocatalytic activation of this protein subunit? Also the physiological electron donor is not yet known, though one study clearly showed that proline reduction led to growth yield increases and to the pumping of protons in the *C. sporogenes* model system.[211] Perhaps future studies with other model systems can uncover more about the biochemistry of this selenoenzyme as well as elucidating better its role in energy metabolism.

5.05.3.2.1(iv) Formate dehydrogenase Using the model organism *Clostridium thermoaceticum*, Ljungdahl's group initially found that the growth yield of the organism was enhanced with the addition of iron and molybdenum.[212,213] Even with increased growth yields the activity of the formate dehydrogenase decreased during growth, suggesting another micronutrient was missing from the culture medium. With the addition of selenium, the level of FDH activity increased, and this led them to partially purify the FDH activity in cells labeled with radioisotope selenium.[213] Further study showed that the purified protein contained tungsten, iron, and selenium in stoichiometric amounts.[214,215] These studies parallel those of Stadtman and demonstrated at nearly the same time that selenium was important in carbon and energy metabolism in eubacteria. This of course led to Andreesen's interest in selenoproteins and subsequent studies of Stickland reactions in Gottingen and later in Halle.[102,105,181,183,216,217]

The first indication that selenium might play a role in formate metabolism in the well-studied model system of *E. coli* was spotted with the isolation of mutants that did not produce hydrogen.[218,219] These mutants of course were critical to identification of the *sel* genes (see above section 5.05.3.1). Wild-type *E. coli* carries out a typical mixed acid fermentation, with acetate, lactate, ethanol, formate, and succinate as products. Formate can be

further metabolized to H_2 and CO_2 by the formate hydrogen lyase complex (FHL), or can be used as an electron donor for the nitrate reductase complex. The FHL complex includes a formate dehydrogenase (FDH-H), electron carrier proteins (Hyc) and a Ni–Fe hydrogenase (Hyc). Expression of the genes encoding the protein components of the FHL complex, activated by FhlA protein, results in an increase in the net flux of pyruvate to formate and therefore an increase in H_2 production from hydrogenase-3. The production of higher levels of CO_2 also occur when FHL activity is higher, which can reduce acidification of the culture medium during fermentations. Overall an increase in cell yield due to formate removal is also observed, given the reduced acid buildup in the culture medium. This increase in cell yield will increase the production of total H_2. It is not yet fully understood as to what is the advantage for the breakdown of formate in the gastrointestinal tract, as the benefit for gas production in batch fermentation is not probably as relevant when competing with other microbes in the host.

Three separate formate dehydrogenase isoenzymes have been identified and studied,[30] including FDH-H (hydrogenase linked), FDH-N (linked to nitrate reductase), and FDH-O (expressed under aerobic conditions). Each of these is a selenoenzyme (selenocysteine) that also contains a molybdenum (molybdopterin) cofactor. The precise role for selenium in the reaction mechanism for these enzymes has been best defined in a series of biochemical, structural, and spectroscopic studies of the FDH-H.[93,94,220,221] The purified enzyme was found to display an electron paramagnetic resonance (EPR)-active molybdenum (V) species upon reduction by formate, and enrichment of the selenocysteine with stable isotope ^{77}Se gave direct evidence for a coordination of the selenocysteine to the Mo center.[221] This was further strengthened by solving of the crystal structure of the enzyme, revealing for the first time the chemical structure of this class of molybdenum cofactor.[220] Structural analysis uncovered the presence of a 4Fe–4S cluster that was reduced from the Mo center and proposed a model for the reaction mechanism between formate, the selenocysteine residue, and molybdenum active sites. Clear evidence from several experiments showed that oxygen transfer does not take place during the oxidation of formate,[93] unlike the molybdenum hydroxylase family of selenoenzymes. Further, X-ray absorption studies probed the relationship between reduced and oxidized enzyme, especially to support the structural determination of the redox state of selenium during catalysis.[222] Some discrepancies arose between the two sets of data, specifically a lack of S–Se interaction suggested by XAS that was not present in the resolved structure. Critical differences in sample preparation hinder the accurate comparison of the data in the two studies, and further evidence is clearly required to obtain a solid reaction mechanism for this model enzyme. A recent study has suggested a new reaction mechanism for the entire class of selenium-dependent FDH enzymes, based on reanalysis of the original crystal data.[223] This study shows that a potential mistake was made during the structural analysis, and by resolving this error they combine existing EPR and XAS data to build a modified mechanism. The authors propose first the direct binding of formate to Mo, followed by attack with a reduced selenol (SeCys residue), release of CO_2, and association of the selenocysteine to the oxidized Mo atom. Further biochemical analysis of FDH-H and similar enzymes should continue to shed light on the role of selenocysteine in catalysis.

5.05.3.2.2 Newly identified mammalian selenoproteins

Glutathione peroxide was the first selenium-containing mammalian enzyme identified over 35 years ago,[4] the same year that the selenoprotein A was identified in *C. sticklandii*.[5] With the advent of the human genome sequence we now know that there are 25 selenoproteins encoded in humans,[10] with similar numbers encoded by other mammals. Some of the best-studied selenoproteins include five isoenzymes of thioredoxin reductase,[139,224–229] three deiodonase isoenzymes,[144,192,230–234] and five isoenzymes of glutathione peroxidase.[235–243] Selenoprotein W[154,244–247] and selenoprotein P[134,151,248–253] have also been the subject of many studies over more than 20 years. **Table 2** lists the known human selenoproteins based on computational biology and laboratory studies. In this section we focus our attention on the newly discovered selenoproteins that have emerged from the comparative analysis of the mouse and human genomes,[10] and will not discuss the more well-studied (and reviewed) selenoenzymes.

5.05.3.2.2(i) Selenoprotein M First identified in the seminal paper on the mammalian selenoproteome, selenoprotein M has since been studied and found to be a widely expressed protein in multiple organs and tissues.[160] The gene encoding this selenoprotein resides on chromosome 22 and based on cloned cDNA and

genomic sequence the gene contains five exons. It should be noted that this gene was not identified at all in any of the human genome sequence efforts. Northern blot analysis shows the mRNA was present in a number of tissues, but was highly expressed in the brain. Localization in transfected cells, using a green fluorescent protein (GFP) fusion, indicated this selenoprotein to be perinuclear (golgi/endoplasmic reticulum). This localization depended on a unique N-terminal signal sequence in the fusion. Although the role for this protein has not yet been established, it does have distant homology to the 15 kDa selenoprotein (Sep15, described in detail below).

Another interesting study identified genes that are upregulated in the brain upon overexpression of presenilin-2.[254] The level of mRNA encoding selenoprotein M was suppressed in cells overexpressing presenilin-2, yet the implications for such suppression is not known nor was it further followed up in this study. In another screen the mRNA encoding selenoprotein M was reduced during infection of white shrimp (*Litopenaeus vannanei*) with a virus.[255] In contrast, using another model, a fivefold increase of selenoprotein M mRNA was found in the tick (*Dermacentor variabilis*) infected with a cattle pathogen, *Anaplasma marginale*. These studies all simply identified an upregulation or downregulation of the mRNA encoding selenoprotein M, however, no evident pattern to these disparate studies give any new clues as to the role of this protein in the cell. Nonetheless, given its similarity to Sep15 and the presence of a possible redox active site (Cys-Gly-Gly-Sec) that has similarity to other selenoproteins (selenoprotein W and selenoprotein T), it is possible that this protein plays some role in redox biology and possibly protein folding in the ER.

5.05.3.2.2(ii) Selenoprotein N The first linkage of selenium nutrition to any clinical disease came in the form of studies that linked white muscle disease to selenium and vitamin E deficiency (reviewed by Oldfield[256]). This led eventually to the identification of selenoprotein W, a selenocysteine-containing protein that is abundant in muscle fibers.[155] Although identified early on, a specific role for this selenoprotein has yet to be elucidated. However with the advent of modern bioinformatics and the use of algorithms to search for SECIS-like elements in the noncoding regions of the human, mouse, and the rat genomes a novel search for selenoproteins identified a selenoprotein termed selenoprotein N.[142] This selenoprotein was identified along with several other novel selenoproteins and this report was published before the comprehensive selenoproteome was presented.[10] However in this early report no known function was proposed.

A landmark study was published in 2001 linked mutations in *SEPN1*, the gene encoding selenoprotein N, and rigid spine syndrome on human chromosome 1p35–36. This region had been previously shown to be involved in rigid spine muscular dystrophy (RSMD) disorders and this was the first linkage of a selenoprotein to a congenital mutation and disease. The authors studied 10 families and found a number of types of mutations (frameshifts, nonsense mutations, and missense mutations). One point mutation actually occurred in the codon for selenocysteine, altering the UGA codon to UAA. Reverse transcriptase PCR showed expression of the mRNA encoding selenoprotein N in muscle, brain, lung, and placenta. This study also noted that one of two selenocysteines present in selenoprotein N resides in a motif similar to that found at the C-terminus of TrxR (CUG).

A subsequent study in 2002 of 27 families with a condition known as multiminicore disease (MmD) also linked mutations in *SEPN1* to disease pathology. Multiple mutations were identified in exons 1, 5, 7, 8, 10, and 11, and the authors also mentioned that this region (RSMD) had been previously linked to MmD. Minicores are lesions by histochemistry of mitochondrial depletion within muscle tissue. The first biochemical study of selenoprotein N aimed to identify the protein localization by immunohistochemistry and found that the primary protein product of several identified mRNAs (splice variants) was a 70 kDa protein present in the endoplasmic reticulum.[161] Two potential ER targeting domains were shown to be present and the peptide expressed from the first exon was shown to be required for localization into the ER. This study also revealed that selenoprotein N was an integral membrane protein that is N-glycosylated. Expression analysis showed pronounced levels in embryonic tissue with a reduction after development and differentiation.

Several subsequent papers since this work have identified other mutations that result in congenital muscular dystrophies.[257–259] One of the most interesting is a mutation in the noncoding cis element SECIS.[258] This mutation was found in an invariant region of a non-Watson–Crick base pairing region of the SECIS element (UGAN/NGAN). This mutation blocked the binding of the elongation factor SBP2 and led to essentially no mRNA or protein produced in skin fibroblasts. It is now clear that three different but related congenital muscular dystrophies are linked to mutation in selenoprotein N.[260] One recent study has shown that a genetic

engineering approach could be beneficial.[162] The mutant UGA codon in one particular patient (changed to UAA) could be corrected by integrating an altered $tRNA^{sec}$ with an altered anticodon to decode UAA. Given the increasing knowledge of muscular dystrophies linked to this protein, further efforts in gene therapy could be used to treat patients in the near future.

5.05.3.2.2(iii) Selenoprotein S Although it was first identified as a selenoprotein using a bioinformatic approach, selenoprotein S was first characterized as a novel protein involved in ER stress response.[261] Selenoprotein S is in fact also known as TANIS or VIMP.[262] Selenoprotein S has been shown to play a role in a large membrane-associated complex that facilitates the translocation of misfolded proteins from the ER lumen to the cytosol for degradation.[261] Soon after this initial discovery and realization that VIMP (or TANIS) is in fact a selenoprotein, a study linking genetic variants in *SEPS1*, the human gene encoding selenoprotein S, with the inflammatory response. Specifically, the authors of this study found a linkage of polymorphisms to the levels of three proinflammatory cytokines (IL-6, IL1β, and TNF-α) in a pool of 522 individuals from 92 families. This study suggested that this protein must have a direct link to production of proinflammatory cytokines.

Two more recent studies have also addressed SNP polymorphisms and potential links to clinically relevant conditions.[263,264] Given the role in inflammation and cytokine expression, one study examined the potential for linkage to autoimmune diseases including type 1 diabetes, rheumatoid arthritis, and inflammatory bowel disease in a population of more than 2000 patients. No significant association was found with any of the SNPs.[263] In contrast, a strong association was found between a single polymorphism (G-105A alteration in the promoter) and women with preeclampsia.[264] Women who had preeclampsia were 1.34 times more likely to carry this mutant allele. The role that this mutant plays in the condition is unknown.

Immediately after a report confirmed that the promoter for *SEPS1* was stimulated by proinflammatory cytokines, and also showed that this promoter was the target of the critical inflammatory regulator NF-κB.[265] The data included direct binding of this transcription factor to the promoter region using gel mobility shift assays. However, the stimulation of the promoter by cytokines and activation of NF-κB did not result in synergistic production of the mRNA. The role that the cytokines or inflammation plays in regulation of expression of selenoprotein S is not yet clear.

Given its role defined in ER stress, a follow-up study showed that overexpression of selenoprotein S in macrophages could reduce cell death and apoptosis induced by tunicamycin or thapsigargin.[266] Similarly, siRNA knockdown of the mRNA encoding selenoprotein S resulted in cells with higher sensitivity to these reagents. These results clearly show that selenoprotein S has a key role in protecting macrophage from inflammatory responses that can lead to ER stress. Probably this is the role played by selenoprotein S in most cell types, although this has yet to be tested in other cell lineages.

Given the role for inflammation in coronary heart disease (CHD) and ischemic stroke, a study was conducted using two large independent Finnish cohorts to determine whether a linkage occurs in genetic variants of *SEPS1* and CHD or risk for stroke.[267] One polymorphism was linked to CHD and a second a risk for ischemic stroke, although neither reached the level of statistical significance. Nonetheless, the trend for linkage was independently confirmed in both cohorts, suggesting a weak linkage that could be found to be statistically significant given a larger study population. Future studies along these lines could be interesting as very little evidence exists for any direct link of a particular selenoprotein and CHD.

5.05.3.2.2(iv) Selenoprotein T Although identified through bioinformatics in 2003, only recently was the role for selenoprotein T uncovered.[171] Using antibodies raised against a peptide derived from the predicted protein sequence, the authors determined that selenoprotein T was expressed both during embryonic development as well as throughout adulthood in rats. By colocalization with GRP78 the authors also determined that the protein was localized primarily in the ER. This selenoprotein also carries a possible redox active motif (CVSU) similar to selenoprotein M, Sep15, and selenoprotein W. The most significant finding in this study was the fact that the expression of selenoprotein T was under control of a neuropeptide, pituitary adenylate cyclase-activating polypeptide (PACAP). This neuropeptide signals through two separate types of G-protein-coupled receptors (PAC1-R and VAPC1-R or VAPC2-R). PACAP signaling plays a number of roles in neuronal development and differentiation, and is involved in Ca^{2+} signaling.[171] To probe the role of selenoprotein T the

authors overexpressed the mRNA encoding this selenoprotein and found that Ca^{2+} levels increased. Using siRNA, knockdown of selenoprotein T reduced the ability of PACAP to induce Ca^{2+} increases and hormone secretion. The authors propose that the selenoprotein T has a critical role in Ca^{2+} homeostasis.

5.05.3.2.2(v) Sep15 (15 kDa selenoprotein) Initially identified as a radiolabeled small selenoprotein in T-cells[268] the so-named 15 kDa selenoprotein has since been termed Sep15. On the basis of its expression levels in certain tissues, correlation of lower levels in several cancer cell lines, and the overall association of low selenium status to cancer, a follow-up report suggested the Sep15 may play a role in the relationship between selenium status and cancer.[269] Further studies reveal polymorphisms in human populations[270,271] that occur in the apical loop of the SECIS element in the 3′ untranslated region of the mRNA. Two polymorphisms (C/T at nucleotide 811 and G/A at nucleotide 1125) were both found to alter the ability for the mRNA to be 'read-through' using a fusion protein system.[270] To date no conclusive evidence has arisen to link these polymorphisms to cancer or another pathological condition.

The role for this selenoprotein was identified first from a study that identified Sep15 as a ER resident protein based on confocal microscopy of a green fluorescent protein fusion. Moreover, purification of radiolabeled Sep15 showed that the protein was in complex with UDP-glucose:glycoprotein glucosyltransferase (UGTR). UGTR has a role in protein folding and quality control in the ER. Association with the ER was found to depend on an N-terminal domain, similar to other selenoproteins that reside in this subcellular compartment. The copurification of Sep15 with UGTR was found to be a stable complex through several chromatography steps and the proteins formed a stable protein during native gel electrophoresis but were separated by SDS-PAGE.[172] A later study by the same group determined that the two proteins have a binding affinity with an estimated K_d of 20 nmol l^{-1}.[272] Moreover, Sep15 immunoprecipitated with UGTR when expressed, however, no complex formation occurred with selenoprotein M, suggesting the Sep15–UGTR complex is specific. The binding of Sep15 to UGTR was also found to be mediated by a unique cluster of cysteine residues that does not occur in other selenoproteins found in the ER. Clearly this implicates Sep15 to play an integral role in protein folding in the ER in concert with the UGTR protein.

Sep15 has been shown to be one of the most highly expressed proteins in the brain, although it is also expressed in many other tissues.[12] In a tissue-specific knockout of the selenocysteine-specific tRNA ($tRNA^{sec}$), Sep15 protein levels were found to decrease along with two glutathione peroxidase isoenzymes (Gpx1 and Gpx4) when the available selenoprotein synthesis machinery had been reduced.[273] Since the knockout of the tRNAsec was found to result in embryonic lethality, this tissue-specific knockout of selenoprotein synthesis has proven to be a vital tool in determining the role of selenium in a tissue-specific manner.[273–281] These studies taken together suggest that the role for Sep15 clearly lies integral to UGTR and protein folding, however, the specific role of the redox activity cysteines and selenocysteine residues have yet to be elucidated at the molecular level. Continued studies using advanced molecular biology and biochemical approaches will probably fill this gap in our understanding in the very near future.

5.05.3.2.2(vi) MsrB (selenoprotein R or methionine-R-sulfoxide reductase) Methionine can be oxidized in proteins to methionine sulfoxide.[282] The oxidation of the amino acid results in either stereoisomer (*R* or *S*). The reduction of these oxidized residues had long been studied at the cellular and biochemical level, but only over the past 10 years it has become apparent that two distinct class of enzymes are responsible for the conversion of each stereoisomer.[282] Based on bioinformatic approaches, a vertebrate selenoprotein was first identified as selenoprotein X,[142] subsequently termed selenoprotein R and then finally methionine-*R*-sulfoxide reductase B (MsrB) based on an enzymatic activity.[10,164] MsrB from vertebrates contain an active site selenocysteine residue that was found to be required for enzymatic activity to convert methionine-*R*-sulfoxide into methionine.[124] This small (12 kDa) selenoprotein is conserved in bacteria, Archaea, and eukaryotes but only in vertebrate animals this enzyme contains an active site selenocysteine. The clue for the discovery for this activity came from phylogenetic profiles of selenoprotein R to MsrB, the cysteine containing enzyme(s) that also act on methionine-*R*-sulfoxide.[124] MsrA enzymes, of which nearly all are cysteine dependent and do not contain selenium, are responsible for the conversion of methionine-*S*-sulfoxide into methionine. More recent studies have shown that selenocysteine residues are present in both types of enzymes naturally, and that one can replace the cysteine with selenocysteine and retain stereospecificity for their respective substrate.[165,283,284]

In this section we review the biochemical studies that have been carried out on MsrB enzymes, and discuss some recent nutritional studies that suggest zinc and selenium metabolism may also affect methionine oxidation and in turn play a major role in aging.

The first biochemical analysis of a purified MsrB was that of a cysteine mutant form from the mouse, due to the difficulties of overexpression of selenoproteins in heterologous hosts.[115,116,285] Surprisingly, this enzyme preparation contained a single equivalent of zinc bound (1.08 equivalents in the as-isolated protein expressed in the *E. coli* host). The specificity for methionine-*R*-sulfoxide was determined, and to further support the catalytic role of MsrB homologues to prevent methionine oxidation, a study of yeast with knockouts in both *msrA* and *msrB* genes was carried out. Loss of either gene resulted in increased sensitivity to hydrogen peroxide, and a double mutant was substantially more sensitive. This was the first biochemical proof that the gene product of the large family of MsrB proteins catalyzes methionine-*R*-sulfoxide reduction. Little similarity at the amino acid sequence level between MsrA and MsrB proteins had precluded this assignment prior to this study.

In addition to a selenoprotein (MsrB), another MsrB isoenzyme that did not contain a selenocysteine residue had already been identified previously. This isoenzyme, termed CBS-1, was subsequently found along side a third isoenzyme during the sequencing of the mouse and the human genomes. Given the confusion a new nomenclature was suggested to include MsrB1 (selenoprotein R), MsrB2 (formerly CBS-1), and a third isoenzyme MsrB3.[286] Analysis of the predicted mRNA sequence coding for MsrB suggested that two splice variants occur, one targeted to the endoplasmic reticulum and the other targeted to the mitochondria.[286] This same study showed that the mouse MsrB3a form is targeted to the ER. It had been previously shown that MsrB2 was present in the mitochondria, with MsrB1 (SeCys) being present in the cytosol and the nucleus. The subcellular distribution and differences in the need for selenium provide a complex story that has yet to be fully unraveled.

In addition to defining the essential role of MsrB selenoenzymes, several studies have addressed the catalytic mechanism of MsrB as well as MsrA enzymes.[164,165–167,283,286–288] Subsequent structural studies have shown that although these two major classes of Msr enzymes can have both Cys and SeCys residues at their core active sites,[168] the presence of a SeCys residue alters the reaction mechanism in either case. Specifically, when a Msr enzyme uses a selenol (SeH) as a nucleophile to attack the methionine sulfoxide, a resolving cysteine residue is required to allow for the reductive release of water from the sulfenic acid–selenium intermediate.[165,166,283,284] This resolving cysteine residue also can affect the *in vitro* and *in vivo* electron donor. Dithiothreitol is commonly used *in vitro*, but all evidence points to reduced thioredoxin required in the selenium-dependent enzymes *in vivo*[165] In the case of cysteine-dependent enzymes, the resolving cysteine is not required since the electron donor reacts directly with the sulfenic acid–sulfur intermediate. However, in general, the selenium-dependent enzymes are far more catalytically active (faster turnover) when the physiologically relevant electron donor is present.[165] These studies, combined with structural data, have shed much light on this critical aspect of oxidative stress defense and the role of selenium in aging.

Although in humans only MsrB1 is a selenoprotein, the depletion of selenium from the diet of mice led to increases in both *R* and *S* stereoisomers.[289] This was not initially explained, yet a subsequent study has shown that small molecule selenols (organic selenocysteine homologues) could act as efficient electron donors *in vitro* for MsrA enzymes.[290] This effect has only been shown *in vitro*, but the possibility that small molecular selenium reductants, or more likely that some selenoproteins that contain reduced selenols (in redox-active motifs) is quite intriguing. Several small selenoproteins do not have real roles and reside in nearly all subcompartments of the cell (mitochondria, ER) where electron donors for Msr enzymes are probably critical to maintain protein stability. Low selenium nutritional status would then have a significant impact on all methionine oxidation, as was seen *in vivo*.[289] Future studies to address selenium nutrition and methionine oxidation could prove to be enlightening as to the role for selenium in catalytic reduction of methionine-*S*-sulfoxide.

5.05.3.3 Incorporation of Selenium into 2-Selenouridine (Se^2U) and 5-[(methylamino)-methyl]-2-selenouridine (mnm^5Se^2U)

The specific incorporation of selenium into nucleic acids was first identified by K. P. McConnell in the early 1970s.[291] This initial identification was based on chromatographic separation of nucleobases and sensitivity to RNases. Soon after studies to determine the specificity of selenium incorporation into tRNA was carried out in

the bacterial models of *C. sticklandii*, *E. coli*, *Salmonella typhimurium*, and the archean *Methanococcus vannielli* by T. C. Stadtman's group.[292–299] The modified tRNA species, Se^2U and mnm^5Se^2U, are found in the wobble loop of a subset of tRNA species including $tRNA^{Lys}$, $tRNA^{Glu}$, and $tRNA^{Gln}$. More recent studies have identified the selenouridine synthase enzyme, using both biochemical[299] and molecular biology approaches.[29] The positioning of the modified base strongly suggests that it plays a role in translation efficiency; however, no specific target gene has yet to be identified that is altered in regulation in response to this posttranscriptional modification to the tRNA. In the following sections, a historical overview of the incorporation of selenium into tRNAs is given, followed by a prospectus for current and future studies within this area.

5.05.3.3.1 Historic papers laying the foundation

The first report of incorporation of radiolabeled selenium into nucleotides was from Saelinger *et al.* in 1972.[291] The authors in this report present data to support the hypothesis that a tRNA molecule of some sort has been labeled with the radioisotope ^{75}Se. However this was a preliminary finding that was followed up in 1974 by the same group with evidence that selenium was incorporated into tRNA to form a 4-selenouridine.[300] Although these were clearly pioneering studies to suggest strongly that selenium was present in a modified nucleotide, subsequent more thorough biochemical studies have shown the form of selenium present is Se^2U and mnm^5Se^2U.[293,294] Although not widely cited, this work laid the initial foundation for the discoveries that lay ahead. Studies that established the need for selenophosphate for biosynthesis of selenoproteins[19,22,96,97,111] were carried out in parallel to those focused on elucidating the enzymes and intermediates in tRNA modification.[297–299]

Using first the established model system of *C. sticklandii*, Chen and Stadtman demonstrated that selenium derived either from selenocysteine or selenite could serve as a nutritional source for three unique (by chromatography) nucleotides.[301] Also quite significant were the findings that inhibition of protein synthesis or production of RNA did not significantly alter the ability of cell extracts to lead to incorporation of radioisotope selenium. This demonstrated that the molecule was probably a posttranscriptional modification of some precursor. Following up this seminal paper, the same group reported the identification of $tRNA^{Glu}$ from *C. sticklandii* as a modified tRNA.[292] A correlation with efficiency for use as an acceptor was also found to coincide with the selenium content. Even with this identification, the critical question of specificity was unknown at this time, since sulfur incorporation into tRNA molecules as mnm^5S^2U had been established.

Wittwer[302] first addressed the specificity of selenium incorporation using *E. coli* as a model. Two modified tRNA species were identified, $tRNA^{Lys}$ and $tRNA^{Glu}$, in this work. Moreover, the level of selenium incorporation into these two specific tRNA species was quantified using radioisotopes and found to be relatively stable to fluctuations in the level of sulfur and selenium in the culture medium. These results for the first time clearly showed that the cell was capable of specifically incorporating selenium into these tRNA species and this incorporation was not simply nonspecific substitution of selenium for sulfur. It was also in this work that the suggestion for the conversion of mnm^5S^2U into mnm^5Se^2U was given as an explanation for the previous findings. The definitive confirmation of the species came in the same year by the same authors when chemically synthesized mnm^5Se^2U was found to be identical to biologically produced nucleotides by both UV–visible spectroscopy and 1H-NMR methods.[294] However, this species did not account for all the selenium-modified nucleotides, and a follow-up publication had evidence to suggest the formation of Se^2U in a report quantifying the selenium-modified tRNAs from the Archaebacterium *M. vanneilii*.[293]

Clearly, the results emerging suggested that at least two nucleotides were modified, the absolute sequence position within the tRNA had yet to be established. Ching *et al.*[295] showed that a Se^2U residue was present in the wobble position of the $tRNA^{Glu}$ from *C. sticklandii*. This study confirmed a notion that the modification probably affects the translation efficiency of certain transcripts, based on the level of modification by selenium. The authors speculated that the modification to seleno-$tRNA^{Glu}$ (GAG) allowed for more efficient use of this tRNA species as compared to the $tRNA^{Glu}$ (GAA).[295] Even today, no definitive data exist to show that this modification alters the translation efficiency in these bacterial model systems. Nonetheless, these studies had established the chemical forms of Se^2U and mnm^5Se^2U, and established that they were derived from modifications to nucleotides that first required sulfur (S^2U and mnm^5S^2U), the mechanism by which selenium was inserted into the tRNA would not be definitively answered until many years later.

5.05.3.3.2 Confirmation of the role of selenophosphate and discovery of selenouridine synthase

During the 1990s, the molecular mechanism for specific incorporation of selenium into selenoproteins was defined primarily using the *E. coli* model system. [16–19,21,112,113,300,303] In studies that paralleled these discoveries, the biochemical pathway for incorporation of selenium into mnm^5Se^2U and Se^2U were also identified. The activity of the selenophosphate synthetase (SPS, or SelD) was first shown to be required for 2-selenouridine synthesis using a *S. typhimurium* mutant that had a genetic defect in selenium metabolism.[297,298] This genetic defect was found to be a mutation in the *selD* gene, and the cell extract from this strain was incapable of metabolizing selenium into tRNA *in vitro*. Addition of purified SPS allowed for efficient incorporation into unlabeled tRNA (bulk), suggesting strongly that the phosphorylation of selenium was required prior to incorporation into the tRNA nucleotides. The putative identification of selenophosphate was actually made in one of these studies,[298] and later confirmed chemically.[110] With the knowledge that selenium must first be phosphorylated, biochemical studies attempting to purify the enzyme(s) responsible for conversion of selenophosphate into Se^2U were carried out, again using the *S. typhimurium* model system.[299] A partially purified preparation of enzyme was isolated that could convert selenium from selenophosphate (generated by the addition of hydrogen selenide and SPS or chemically produced selenophosphate) to Se^2U with an apparent K_m of 17.1 $\mu mol\,l^{-1}$. Several comparable preparations were made with similar characteristics, but a low typical yield (10 μg) of enzyme limited identification of the protein at that time.

The final identification of the selenouridine synthase did not occur until a thorough analysis of sulfurtransferase mutants was carried out using an *E. coli* model.[29] Colocalization of a putative open reading frame, termed *ybbB* in *E. coli*, with *selD* in several bacterial genomes suggested that the gene product YbbB may play a role in selenium metabolism. Site-directed chromosomal mutations in this gene (*ybbB*) and other putative sulfur transferases were isolated and tested for their ability to produce Se^2U or mnm^5Se^2U. Only mutations in the *ybbB* displayed a mutant phenotype.[29] The purified YbbB protein had tRNA bound from affinity chromatography, and in the presence of active SPS this protein could facilitate the incorporation of selenium into bulk tRNA using ^{75}Se radiolabeled tracer. Further, mutation of a single cysteine residue (Cys^{97}) resulted in a lack of activity *in vivo*, although the mutant protein displayed some activity *in vitro*. It is not yet known whether another protein catalyst or selenium carrier protein is necessary for the production of the modified nucleotide; however, recent bioinformatic analysis of the distribution of YbbB and SelD strongly argue against this notion.[304]

We now have clear evidence that selenium can be introduced specifically into proteins as selenocysteine and into a subset of tRNA species as mnm^5Se^2U or Se2U. The pathways and molecular mechanisms for insertion of selenium into these molecules have been well established in several model systems, the best studied being *E. coli*. The role for selenium in selenoproteins (i.e., the need for selenium over sulfur) is thought to be its ability to act as a more reactive nucleophile and perhaps a more rapid catalyst.[305,306] However, a void in our knowledge exists for the specific need for selenium-modified tRNAs. Mutation of either *selD* or *ybbB* did not alter the growth characteristics of *E. coli*;[29] however, no thorough analysis of the bacterial stress response has been carried out in any of these mutants. Clearly, further study is needed to better define the role for selenium in wobble codon usage for a subset of tRNA species.

5.05.3.4 Labile Selenoenzymes

Beyond the use of selenium as the twenty-first amino acid and its incorporation into tRNA in a modified base, selenium has been shown to be present as a labile enzyme cofactor in a small class of molybdenum enzymes. These enzymes are within a larger cluster of molybdoenzymes known as molybdenum hydroxylases, a class that also includes the well-studied bovine xanthine dehydrogenase (XDH) (oxidase) and aldehyde oxidoreductases. Typically, these enzymes consist of a molybdenum cofactor, two 2Fe–2S clusters, and a flavin (FAD). The site of hydroxylation takes place at the molybdenum atom using a water ligand as the oxygen donor.[307] The resulting two electrons then travel from the molybdenum cofactor, through the iron–sulfur clusters to the oxidized FAD. The ultimate electron acceptor (oxygen, $NADP^+$, NAD^+) then reacts with the reduced flavin and the oxidized enzyme is regenerated. The chemical nature of the selenium cofactor is unknown, and the mechanism(s) by which selenium is incorporated into the apoenzyme also remains a

mystery. With the expansion of microbial genome sequences, computational biology techniques now may lead the way to defining the missing links in this interesting and understudied area of selenoenzymes. We discuss each of the enzymes that have been studied that contain a labile selenium cofactor, and subsequently discuss the emerging data from genome sequence analysis that should allow the complete description of this area within a few years.

5.05.3.4.1 *Nicotinic acid hydroxylase of* Eubacterium barkeri

Holcenberg and Stadtman[308] were the first to describe the purification and characterization of the nicotinic acid hydroxylase (NAH) from *Eubacterium (Clostridium) barkeri*. This organism was isolated in a defined medium with nicotinic acid as a carbon and energy source in order to better understand the metabolism of intermediary metabolites. The purified protein was found to produce hydroxylated nicotinic acid and to require pyridine nucleotide ($NADP^+$) as an electron acceptor, with the hydroxyl group being derived from water. At this time the metal cofactors were not identified. Several years later, Dilworth[26,309] published two papers describing the metals present in the enzyme, and found evidence both for molybdenum and selenium. By this time (1982), there was sufficient knowledge of the properties of formate dehydrogenase and the chemistry of selenocysteine that it became clear that the form of selenium in NAH was not covalently bound to the polypeptide, since it could be released upon denaturation. This was then a seminal report to suggest that a third possible use for selenium exists in nature.[95]

Gladyshev *et al.*[27,311] and Gladyshev and Lecchi[310] further expanded upon this work to better understand the chemical nature of the labile cofactor. Using EPR techniques, one study suggested that the labile selenium cofactor was a ligand of the molybdenum cofactor.[27] Cells grown without added selenium did produce enzyme but its activity was quite low, while the molybdenum cofactor was present. The addition of excess reducing agent (dithiothreitol) resulted in release of the labile selenium atom from the active site. Based on these results the authors suggested that the selenium cofactor is analogous to the well-studied cyanolyzable sulfur ligand of bovine xanthine oxidase.[307]

5.05.3.4.2 *Xanthine dehydrogenase of* Clostridium purinolyticum, Clostridium acidiurici, *and* Eubacterium barkeri

Rabinowitz and Barker first described the isolation of purine-fermenting anaerobes and began elucidating the purine fermentation pathways.[312] These strains, *Clostridium acidiurici* and *Clostridium cylindrosporum*, were isolated using purines (adenine, guanine, xanthine, uric acid) as sole carbon, nitrogen, and energy source. The rationale for these studies was to obtain a bacterial model system to understand purine catabolism. XDH was purified and characterized in early studies that occurred well before any knowledge of a need for selenium. Given the presence of a labile selenium cofactor in the nicotinic acid hydroxylase of the closely related strain *E. barkeri*, in the early 1980s, Dürre and Andreesen[183,313] isolated an adenine-fermenting organism, *C. purinolyticum*.[181,314] These three strains were then studied for their capabilities to ferment xanthine and the resulting products of this pathway were isolated using cell extracts.[183] One of the key end products of this pathway, glycine, is directly reduced by the glycine reductase in each of these organisms leading to a source of ATP from acetyl phosphate. Ammonia is also a product of this pathway, used arguably as an efficient nitrogen source. Purine fermentation generates energy and reduced nitrogen, yet selenium is required for the interconversion of the purines to xanthine.[28,181]

The first biochemical analysis of a selenium-containing XDH was reported in 1999 by Andreesen's group.[28] This preparation was specific for xanthine and did not hydroxylate nicotinic acid. Moreover, the enzyme contained FAD, acid-labile sulfur, iron, and a dinucleotide molybdenum cofactor. Most intriguing was the near-equimolar presence of tungsten and molybdenum. It should be noted that the culture medium contained nearly equimolar levels of these metals, making one wonder whether the specificity of this enzyme for metal may be relaxed (i.e., can use Mo or W). Selenium was also found in the preparation and could be released by treatment with cyanide indicating it was also a labile cofactor. This further confirmed the chemical nature of the cofactor from the NAH enzyme from the same strain.[27,311]

Just after this report, we published the first characterization of the XDH from *C. purinolyticum*.[315] We found that the three subunit enzymes contained molybdenum, acid-labile sulfur and iron, and an FAD prosthetic group. XDH would hydroxylate purine, hypoxanthine, or xanthine and preferred $NADP^+$ as an electron

acceptor. Based on the apparent K_m, xanthine was the preferred purine substrate. Cyanide released selenium (labeled using ^{75}Se) and resulted in loss of enzyme activity, consistent with prior reports on labile selenoenzymes. Based on this and previous reports, the selenium cofactor is analogous to the cyanolyzable sulfur found in bovine XDH.

5.05.3.4.3 *Purine hydroxylase of* Clostridium purinolyticum

Upon purification of the XDH from *C. purinolyticum*, a separate ^{75}Se-labeled peak appeared eluting from a DEAE sepharose column. This second peak also appeared to contain a flavin based on UV–visible spectrum. This peak did not use xanthine as a substrate for the reduction of artificial electron acceptors (2,6 dichloroindophenol, DCIP), and based on this altered specificity this fraction was further studied. Subsequent purification and analysis showed the enzyme complex consisted of four subunits, and contained molybdenum, iron selenium, and FAD. The most unique property of this enzyme lies in its substrate specificity. Purine, hypoxanthine (6-OH purine), and 2-OH purine were all found to serve as reductants in the presence of DCIP, yet xanthine was not a substrate at any concentration tested. The enzyme was named purine hydroxylase[315] to differentiate it from similar enzymes that use xanthine as a substrate. To date, this is the only enzyme in the molybdenum hydroxylase family (including aldehyde oxidoreductases) that does not hydroxylate the 8-position of the purine ring. This unique substrate specificity, coupled with the studies of Andreesen on purine fermentation pathways, suggests that xanthine is the key intermediate that is broken down in a selenium-dependent purine fermentation pathway.[181,183]

In an effort to better define the role of selenium in catalysis, EPR studies were carried out on PH.[316] It was determined that the FeS centers and molybdenum cofactor could be reduced using NADPH through the flavin site. Removal of selenium by cyanide completely blocked electron transfer to FeS centers and FAD with hypoxanthine as a reductant (forward reaction), but the Mo(V) signal was present. However, cyanide inactivated PH reduced by NADPH displayed reduction of FAD and both 2Fe–2S centers and yet no reduction of the Mo site. Taken together, these data strongly suggest that selenium cofactor is somehow involved in the transfer of electrons from the Mo site to the first of the two FeS centers. However, when stable isotope (^{77}Se) enriched PH was examined, no significant hyperfine coupling was observed with the reduced Mo site.[316] This is contradictory to the data obtained for the NAH using a similar experimental analysis with EPR. This may put into question the exact chemical form of the selenium cofactor.

One underlying question remains: why does this small class of microorganisms require a labile selenium cofactor in these enzymes? Few have speculated on this in the published literature. Yet one key comparison between selenium and non-selenium-dependent hydroxylases may be quite telling. The well-studied bovine XDH has a turnover rate of approximately $5\,s^{-1}$,[307] while the PH enzyme from *C. purinolyticum* has a far faster catalysis of hypoxanthine with a turnover of $450\,s^{-1}$.[316] Since these enzymes catalyze very similar reactions, it is likely that the presence of the labile selenium cofactor results in a significant improvement in catalysis of a slow step. Since this organism was isolated on adenine as a sole carbon and nitrogen source, the conversion of hypoxanthine into xanthine, subsequent to production of critical nitrogen and carbon intermediates, is probably a key and limiting step in growth. One can also speculate that this form of selenoenzyme may have arisen independent of selenoproteins that contain selenocysteine, but until we understand more about the nature and synthesis of this labile cofactor these questions will remain unanswered.

5.05.3.4.4 *Comparative genomics reveals potential widespread use of selenium in the form of a labile cofactor*

Although the metabolic pathways for incorporation of selenium into the labile selenium cofactor of the above-described enzymes has yet to be elucidated, the large number of genomic sequences that have become available in recent years may prove enlightening to this pathway. Two recent studies have posed the question *in silico* as to whether genetic biomarkers for labile selenoenzymes could be identified based on gene localization and operon structure.[9,25] If one assumes that selenium must first be activated to selenophosphate prior to incorporation into a labile cofactor (currently an unfounded assumption), then organisms can be sought that carry the *selD* gene but that do not appear to encode any of the *selABC* genes nor the genes used to incorporate selenium into 2-selenouridine (*ybbB*). Given this assumption both reports suggest

that the pathogen *Enterococcus faecalis* has the capabilities to activate selenium to selenophosphate but lacks the other genetic markers for specific use of selenium.[9,25] Moreover, an apparent operon is present in this genome that encodes a large XDH, SelD, NifS, and SirA proteins, as well as two proteins linked to selenium utilization termed HP1 and HP2. The product of the SelD gene would be predicted to produce selenophosphate, while the NifS protein would be expected to obtain selenium from L-selenocysteine. Other NifS proteins have been shown to generate elemental selenium from L-selenocysteine in a pyridoxal phosphate-dependent mechanism for use in selenium metabolism to produce selenoproteins.[317–322] Although speculative, the SirA protein may be responsible for inserting the activated selenophosphate into the apoenzyme XDH. These genes also colocalize in other organisms, such as *C. difficile*, suggesting that this pathway may be far more widespread in biology than first imagined. Ongoing research in this area should shed light both on the chemical nature and biosynthesis of the unique cofactor, and perhaps the third use of selenium in biology will reveal new insights into specific biological use of the fascinating metalloid selenium.

5.05.4 Summary

Selenium has been studied for its toxicity and now is known to be a critical micronutrient for human health. Although many of the metabolic pathways required to make specific selenoproteins have been elucidated, some critical gaps in our knowledge remain. The uptake of selenium in bacteria is largely undefined, and the specific mechanisms for tissue distribution (in mammals) of small molecule forms of selenium also are not well understood. The structure of labile selenium cofactor in a class of metalloenzymes is not yet established, nor is the pathway for its insertion. Moreover, the toxicity of selenium, specifically at the molecular level, is hotly debated depending on the model system used. Selenium is used exclusively for energy and carbon metabolism in prokaryotes, yet plays a major role in defense against oxidative stress in higher eukaryotes. Given its chemical reactivity and depending on the form, selenium is a double-edged sword between toxicology and animal nutrition that will probably remain an active area of research for decades to come.

Abbreviations

CXC	cysteine – any amino acid – cysteine motif
FDH-H	formate dehydrogenase – hydrogenase linked
FDH-N	formate dehydrogenase – nitrate reductase linked
FDH-O	formate dehydrogenase – aerobic
FHL	formate hydrogenlyase
Gpx	glutathione peroxidase
mnm^{5}Se^{2}U	5-[(methylamino)-methyl]-2-selenouridine
NAH	nicotinic acid hydroxylase
PH	purine hydroxylase
RSSeSR	selenotrisulfide (R = glutathione, cystine, etc.)
RSSR	oxidized disulfide (R = glutathione, cystine, etc.)
SDMH	selenium-dependent molybdenum hydroxylase
Se^{2}U	2-selenouridine
Sel	selenoprotein
Sep15	15 kiloDalton selenoprotein
SMT	selenocysteine methyltransferase
SNP	single-nucleotide polymorphism
SPS	selenophosphate synthetase
SPS1	mammalian selenophosphate synthetase isoenzyme 1
SPS2	mammalian selenophosphate synthetase isoenzyme 2

TGR	thioredoxin glutaredoxin reductase
tRNA	transfer RNA
Trx	thioredoxin
TrxR	thioredoxin reductase
XDH	xanthine dehydrogenase

References

1. J. J. Berzelius, *Ann. Chim. Phys.* **1818**, 239.
2. A. G. Bell, *Science* **1880**, *1*, 130.
3. C. G. Wilber, *Clin. Toxicol.* **1980**, *17*, 171.
4. L. Flohe; W. A. Gunzler; H. H. Schock, *FEBS Lett.* **1973**, *32*, 132.
5. D. C. Turner; T. C. Stadtman, *Arch. Biochem. Biophys.* **1973**, *154*, 366.
6. T. C. Stadtman, *Ann. N. Y. Acad. Sci.* **2000**, *899*, 399.
7. Y. Zhang; D. E. Fomenko; V. N. Gladyshev, *Genome Biol.* **2005**, *6*, R37.
8. S. Castellano; V. N. Gladyshev; R. Guigo; M. J. Berry, *Nucleic Acids Res.* **2008**, *36*, D332.
9. D. H. Haft; W. T. Self, *Biol. Direct* **2008**, *3*, 4.
10. G. V. Kryukov *et al.*, *Science* **2003**, *300*, 1439.
11. A. V. Lobanov *et al.*, *Nucleic Acids Res.* **2006**, *34*, 496.
12. Y. Zhang *et al.*, *J. Biol. Chem.* **2008**, *283*, 2427.
13. T. Tamura; T. C. Stadtman, *Methods Enzymol.* **2002**, *347*, 297.
14. T. C. Stadtman, *Annu. Rev. Biochem.* **2002**, *71*, 1.
15. T. C. Stadtman, *Annu. Rev. Biochem.* **1990**, *59*, 111.
16. A. Bock; K. Forchhammer; J. Heider; C. Baron, *Trends Biochem. Sci.* **1991**, *16*, 463.
17. F. Zinoni; J. Heider; A. Bock, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4660.
18. W. Leinfelder; K. Forchhammer; B. Veprek; E. Zehelein; A. Bock; *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 543.
19. W. Leinfelder; T. C. Stadtman; A. Bock, *J. Biol. Chem.* **1989**, *264*, 9720.
20. J. Heider; W. Leinfelder; A. Bock, *Nucleic Acids Res.* **1989**, *17*, 2529.
21. W. Leinfelder; E. Zehelein; M. A. Mandrand-Berthelot; A. Bock, *Nature* **1988**, *331*, 723.
22. W. Leinfelder *et al.*, *J. Bacteriol.* **1988**, *170*, 540.
23. T. C. Stadtman, *Science* **1974**, *183*, 915.
24. L. V. Papp; J. Lu; A. Holmgren; K. K. Khanna, *Antioxid. Redox Signal.* **2007**, *9*, 775.
25. Y. Zhang; A. A. Turanov; D. L. Hatfield; V. N. Gladyshev, *BMC Genomics* **2008**, *9*, 251.
26. G. L. Dilworth, *Arch. Biochem. Biophys.* **1982**, *219*, 30.
27. V. N. Gladyshev; S. V. Khangulov; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 232.
28. T. Schrader; A. Rienhofer; J. R. Andreesen, *Eur. J. Biochem.* **1999**, *264*, 862.
29. M. D. Wolfe *et al.*, *J. Biol. Chem.* **2004**, *279*, 1801.
30. J. Heider; A. Bock, *Adv. Microb. Physiol.* **1993**, *35*, 71.
31. A. L. Moxon, *S. D. Agric. Exp. Stations Bull.* **1937**, *331*, 91.
32. A. L. Moxon, *Science* **1938**, *88*, 81.
33. N. C. Koch, *J. Paleontol.* **1967**, *41*, 970.
34. D. R. Clark, Jr.; P. A. Ogasawara; G. J. Smith; H. M. Ohlendorf, *Arch. Environ. Contam. Toxicol.* **1989**, *18*, 787.
35. H. M. Ohlendorf, *Aquat. Toxicol.* **2002**, *57*, 1.
36. H. M. Ohlendorf; R. L. Hothem; C. M. Bunck; K. C. Marois, *Arch. Environ. Contam. Toxicol.* **1990**, *19*, 495.
37. H. M. Ohlendorf *et al.*, *J. Toxicol. Environ. Health* **1988**, *24*, 67.
38. C. A. Schuler; R. G. Anthony; H. M. Ohlendorf, *Arch. Environ. Contam. Toxicol.* **1990**, *19*, 845.
39. J. Chaudiere; O. Courtin; J. Leclaire, *Arch. Biochem. Biophys.* **1992**, *296*, 328.
40. C. D. Klaassen *et al.*, *Fundam. Appl. Toxicol.* **1985**, *5*, 806.
41. M. S. Stewart; J. E. Spallholz; K. H. Neldner; B. C. Pence, *Free Radic. Biol. Med.* **1999**, *26*, 42.
42. M. Alonis; S. Pinnell; W. T. Self, *Photodermatol. Photoimmunol. Photomed.* **2006**, *22*, 315.
43. R. F. Burk; B. K. Norsworthy; K. E. Hill; A. K. Motley; D. W. Byrne, *Cancer Epidemiol. Biomarkers Prev.* **2006**, *15*, 804.
44. M. P. Rayman, *Lancet* **2000**, *356*, 233.
45. R. Abdulah; K. Miyazaki; M. Nakazawa; H. Koyama, *J. Trace Elem. Med. Biol.* **2005**, *19*, 141.
46. E. A. Klein, *J. Urol.* **2004**, *171*, S50.
47. G. F. Combs, Jr., *Br. J. Cancer* **2004**, *91*, 195.
48. A. J. Duffield-Lillico *et al.*, *J. Natl. Cancer Inst.* **2003**, *95*, 1477.
49. D. R. Lovley, *J. Ind. Microbiol.* **1995**, *14*, 85.
50. A. W. Cantafio *et al.*, *Appl. Environ. Microbiol.* **1996**, *62*, 3298.
51. G. F. Combs, Jr. *et al.*, *Biol. Trace Elem. Res.* **1996**, *52*, 209.
52. M. E. Losi; W. T. Frankenberger, *Appl. Environ. Microbiol.* **1997**, *63*, 3079.
53. M. P. de Souza *et al.*, *Appl. Environ. Microbiol.* **2001**, *67*, 3785.
54. M. Roux; G. Sarret; I. Pignot-Paintrand; M. Fontecave; J. Coves, *Appl. Environ. Microbiol.* **2001**, *67*, 769.

55. R. S. Dungan; S. R. Yates; W. T. Frankenberger, Jr., *Environ. Microbiol.* **2003**, *5*, 287.
56. J. Meija; M. Montes-Bayon; J. A. Caruso; D. L. Leduc; N. Terry, *Se Pu* **2004**, *22*, 16.
57. S. Eapen; S. F. D'Souza, *Biotechnol. Adv.* **2005**, *23*, 97.
58. P. Antonioli *et al.*, *Appl. Environ. Microbiol.* **2007**, *73*, 6854.
59. Y. Zhang; Z. A. Zahir; C. Amrhein; A. Chang; W. T. Frankenberger, Jr., *J. Agric. Food Chem.* **2007**, *55*, 5714.
60. J. F. Stolz; P. Basu; R. S. Oremland, *Int. Microbiol.* **2002**, *5*, 201.
61. M. Bebien; J. Kirsch; V. Mejean; A. Vermeglio, *Microbiology* **2002**, *148*, 3865.
62. J. F. Stolz; R. S. Oremland, *FEMS Microbiol. Rev.* **1999**, *23*, 615.
63. C. A. Watts *et al.*, *FEMS Microbiol. Lett.* **2003**, *228*, 273.
64. C. A. Watts *et al.*, *Biochem. Soc. Trans.* **2005**, *33*, 173.
65. J. M. Macy *et al.*, *Int. J. Syst. Bacteriol.* **1993**, *43*, 135.
66. I. Schroder; S. Rech; T. Krafft; J. M. Macy, *J. Biol. Chem.* **1997**, *272*, 23765.
67. H. E. Ganther, *Adv. Exp. Med. Biol.* **2001**, *492*, 119.
68. H. E. Ganther, *Biochemistry* **1968**, *7*, 2898.
69. H. E. Painter, *Chem. Rev.* **1941**, *28*, 179.
70. H. S. Hsieh; H. E. Ganther, *Biochemistry* **1975**, *14*, 1632.
71. H. E. Ganther, *Biochemistry* **1971**, *10*, 4089.
72. M. Haratake; K. Fujimoto; M. Ono; M. Nakayama, *Biochim. Biophys. Acta* **2005**, *1723*, 215.
73. W. T. Self; L. Tsai; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12481.
74. T. Nakagawa *et al.*, *Biochem. Biophys. Res. Commun.* **1986**, *135*, 183.
75. J. Zhang; X. Wang; T. Xu, *Toxicol. Sci.* **2008**, *101*, 22.
76. R. S. Oremland *et al.*, *Appl. Environ. Microbiol.* **1999**, *65*, 4385.
77. S. Muller; J. Heider; A. Bock, *Arch. Microbiol.* **1997**, *168*, 421.
78. I. Uson; G. M. Sheldrick, *Curr. Opin. Struct. Biol.* **1999**, *9*, 643.
79. G. M. Lacourciere; R. L. Levine; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9150.
80. M. X. Sliwkowski; T. C. Stadtman, *J. Biol. Chem.* **1985**, *260*, 3140.
81. D. R. Ellis; D. E. Salt, *Curr. Opin. Plant Biol.* **2003**, *6*, 273.
82. T. G. Sors; D. R. Ellis; D. E. Salt, *Photosynth. Res.* **2005**, *86*, 373.
83. N. Shibagaki *et al.*, *Plant J.* **2002**, *29*, 475.
84. C. Kahakachchi; H. T. Boakye; P. C. Uden; J. F. Tyson, *J. Chromatogr. A* **2004**, *1054*, 303.
85. T. A. Brown; A. Shrift, *Plant Physiol.* **1980**, *66*, 758.
86. T. A. Brown; A. Shrift, *Plant Physiol.* **1981**, *67*, 1051.
87. J. E. Spallholz; D. J. Hoffman, *Aquat. Toxicol.* **2002**, *57*, 27.
88. J. E. Spallholz, *Biomed. Environ. Sci.* **1997**, *10*, 260.
89. T. G. Sors *et al.*, *Plant J.* **2005**, *42*, 785.
90. G. S. Banuelos; Z. Q. Lin; L. Wu; N. Terry, *Rev. Environ. Health* **2002**, *17*, 291.
91. S. J. Hamilton, *Sci. Total Environ.* **2004**, *326*, 1.
92. L. Wu, *Ecotoxicol. Environ. Saf.* **2004**, *57*, 257.
93. S. V. Khangulov; V. N. Gladyshev; G. C. Dismukes; T. C. Stadtman, *Biochemistry* **1998**, *37*, 3518.
94. V. N. Gladyshev *et al.*, *J. Biol. Chem.* **1996**, *271*, 8095.
95. T. C. Stadtman, *Annu. Rev. Biochem.* **1980**, *49*, 93.
96. A. Ehrenreich; K. Forchhammer; P. Tormay; B. Veprek; A. Bock, *Eur. J. Biochem.* **1992**, *206*, 767.
97. I. Y. Kim; Z. Veres; T. C. Stadtman, *J. Biol. Chem.* **1992**, *267*, 19650.
98. G. Sawers; J. Heider; E. Zehelein; A. Bock, *J. Bacteriol.* **1991**, *173*, 4983.
99. J. E. Cone; R. M. Del Rio; J. N. Davis; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2659.
100. T. C. Stadtman, *Fundam. Appl. Toxicol.* **1983**, *3*, 420.
101. M. Wagner *et al.*, *Eur. J. Biochem.* **1999**, *260*, 38.
102. J. R. Andreesen, *Antonie Van Leeuwenhoek* **1994**, *66*, 223.
103. M. Lubbers; J. R. Andreesen, *Eur. J. Biochem.* **1993**, *217*, 791.
104. B. Seto; T. C. Stadtman, *J. Biol. Chem.* **1976**, *251*, 2435.
105. U. C. Kabisch *et al.*, *J. Biol. Chem.* **1999**, *274*, 8445.
106. M. Rother; A. Resch; R. Wilting; A. Bock, *Biofactors* **2001**, *14*, 75.
107. J. A. Vorholt; M. Vaupel; R. K. Thauer, *Mol. Microbiol.* **1997**, *23*, 1033.
108. A. Bock *et al.*, *Mol. Microbiol.* **1991**, *5*, 515.
109. J. Pinsent, *Biochem. J.* **1954**, *57*, 10.
110. R. S. Glass *et al.*, *Biochemistry* **1993**, *32*, 12555.
111. Z. Veres; I. Y. Kim; T. D. Scholz; T. C. Stadtman, *J. Biol. Chem.* **1994**, *269*, 10597.
112. K. Forchhammer; W. Leinfelder; A. Bock, *Nature* **1989**, *342*, 453.
113. K. Forchhammer; K. P. Rucknagel; A. Bock, *J. Biol. Chem.* **1990**, *265*, 9346.
114. F. Zinoni; A. Birkmann; W. Leinfelder; A. Bock, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3156.
115. E. S. Arner; H. Sarioglu; F. Lottspeich; A. Holmgren; A. Bock, *J. Mol. Biol.* **1999**, *292*, 1003.
116. E. S. Arner, *Methods Enzymol.* **2002**, *347*, 226.
117. M. Thanbichler; A. Bock, *EMBO J.* **2002**, *21*, 6925.
118. P. R. Copeland; D. M. Driscoll, *Methods Enzymol.* **2002**, *347*, 40.
119. A. Lesoon; A. Mehta; R. Singh; G. M. Chisolm; D. M. Driscoll, *Mol. Cell Biol.* **1997**, *17*, 1977.
120. M. J. Berry *et al.*, *Methods Enzymol.* **2002**, *347*, 17.
121. P. R. Hoffmann; M. J. Berry, *Thyroid* **2005**, *15*, 769.
122. A. Lescure; D. Fagegaltier; P. Carbon; A. Krol, *Curr. Protein Pept. Sci.* **2002**, *3*, 143.
123. G. V. Kryukov; V. N. Gladyshev, *Methods Enzymol.* **2002**, *347*, 84.

124. G. V. Kryukov; R. A. Kumar; A. Koc; Z. Sun; V. N. Gladyshev, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4245.
125. L. A. de Jesus *et al.*, *Mol. Cell Biol.* **2006**, *26*, 1795.
126. B. A. Carlson; D. L. Hatfield, *Methods Enzymol.* **2002**, *347*, 24.
127. M. R. Bosl; K. Takaku; M. Oshima; S. Nishimura; M. M. Taketo, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5531.
128. B. A. Carlson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12848.
129. F. J. Martin-Romero *et al.*, *J. Biol. Chem.* **2001**, *276*, 29798.
130. I. Y. Kim; M. J. Guimaraes; A. Zlotnik; J. F. Bazan; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 418.
131. F. Ursini *et al.*, *Methods Enzymol.* **1995**, *252*, 38.
132. R. Brigelius-Flohe; A. Banning, *Free Radic. Res.* **2006**, *40*, 775.
133. A. Banning *et al.*, *Antioxid. Redox. Signal.* **2008**, *10*, 1491.
134. K. E. Hill *et al.*, *J. Biol. Chem.* **2003**, *278*, 13640.
135. R. Brigelius-Flohe, *Free Radic. Biol. Med.* **1999**, *27*, 951.
136. M. Maiorino *et al.*, *J. Biol. Chem.* **1991**, *266*, 7728.
137. X. G. Lei; J. K. Evenson; K. M. Thompson; R. A. Sunde, *J. Nutr.* **1995**, *125*, 1438.
138. P. Mauri *et al.*, *Biol. Chem.* **2003**, *384*, 575.
139. T. Tamura; V. Gladyshev; S. Y. Liu; T. C. Stadtman, *Biofactors* **1995**, *5*, 99.
140. E. S. Arner; A. Holmgren, *Eur. J. Biochem.* **2000**, *267*, 6102.
141. T. Tamura; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1006.
142. A. Lescure; D. Gautheret; P. Carbon; A. Krol, *J. Biol. Chem.* **1999**, *274*, 38147.
143. Q. A. Sun; L. Kirnarsky; S. Sherman; V. N. Gladyshev, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3673.
144. D. Behne; A. Kyriakopoulos; H. Meinhold; J. Kohrle, *Biochem. Biophys. Res. Commun.* **1990**, *173*, 1143.
145. J. R. Arthur; F. Nicol; G. J. Beckett, *Biochem. J.* **1990**, *272*, 537.
146. M. J. Berry *et al.*, *Nature* **1991**, *353*, 273.
147. O. Araki *et al.*, *Cytogenet. Cell Genet.* **1999**, *84*, 73.
148. D. Salvatore *et al.*, *J. Clin. Invest.* **1995**, *96*, 2421.
149. X. M. Xu *et al.*, *Biochem. J.* **2007**, *404*, 115.
150. M. J. Guimaraes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15086.
151. L. Schomburg *et al.*, *Biochem. J.* **2003**, *370*, 397.
152. J. G. Yang; J. Morrison-Plummer; R. F. Burk, *J. Biol. Chem.* **1987**, *262*, 13372.
153. K. E. Hill; R. S. Lloyd; J. G. Yang; R. Read; R. F. Burk, *J. Biol. Chem.* **1991**, *266*, 10050.
154. M. A. Beilstein; S. C. Vendeland; E. Barofsky; O. N. Jensen; P. D. Whanger, *J. Inorg. Biochem.* **1996**, *61*, 117.
155. P. D. Whanger, *Methods Enzymol.* **2002**, *347*, 179.
156. S. C. Vendeland; M. A. Beilstein; J. Y. Yeh; W. Ream; P. D. Whanger, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8749.
157. J. Chen; M. J. Berry, *J. Neurochem.* **2003**, *86*, 1.
158. S. V. Novoselov *et al.*, *J. Biol. Chem.* **2007**, *282*, 11960.
159. C. Lu *et al.*, *FEBS Lett.* **2006**, *580*, 5189.
160. K. V. Korotkov; S. V. Novoselov; D. L. Hatfield; V. N. Gladyshev, *Mol. Cell Biol.* **2002**, *22*, 1402.
161. N. Petit *et al.*, *Hum. Mol. Genet.* **2003**, *12*, 1045.
162. M. Rederstorff *et al.*, *Nucleic Acids Res.* **2008**, *36*, 237.
163. B. Moghadaszadeh *et al.*, *Nat. Genet.* **2001**, *29*, 17.
164. R. A. Kumar; A. Koc; R. L. Cerny; V. N. Gladyshev, *J. Biol. Chem.* **2002**, *277*, 37527.
165. D. T. Le *et al.*, *Biochemistry* **2008**, *28*, 6685–6694.
166. H. Y. Kim; V. N. Gladyshev, *Mol. Biol. Cell* **2004**, *15*, 1055.
167. H. Y. Kim; V. N. Gladyshev, *Biochem. J.* **2007**, *407*, 321.
168. W. T. Lowther; H. Weissbach; F. Etienne; N. Brot; B. W. Matthews, *Nat. Struct. Biol.* **2002**, *9*, 348.
169. A. Koc; V. N. Gladyshev, *Ann. N. Y. Acad. Sci.* **2007**, *1100*, 383.
170. Y. Gao *et al.*, *Biochem. Biophys. Res. Commun.* **2007**, *356*, 636.
171. L. Grumolato *et al.*, *FASEB J.* **2008**, *22*, 1756.
172. K. V. Korotkov; E. Kumaraswamy; Y. Zhou; D. L. Hatfield; V. N. Gladyshev, *J. Biol. Chem.* **2001**, *276*, 15330.
173. A. D. Ferguson *et al.*, *J. Biol. Chem.* **2006**, *281*, 3536.
174. V. M. Labunskyy; D. L. Hatfield; V. N. Gladyshev, *IUBMB Life* **2007**, *59*, 1.
175. S. C. Low; J. W. Harney; M. J. Berry, *J. Biol. Chem.* **1995**, *270*, 21659.
176. X. M. Xu *et al.*, *PLoS Biol.* **2006**, *5*, e4.
177. X. M. Xu *et al.*, *J. Biol. Chem.* **2005**, *280*, 41568.
178. T. Tamura *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16162.
179. A. V. Lobanov; D. L. Hatfield; V. N. Gladyshev, *Protein Sci.* **2008**, *17*, 176.
180. H. S. Stickland, *Biochem. J.* **1935**, *28*, 1746.
181. J. R. Andreesen *et al.*, *Biofactors* **1999**, *10*, 263.
182. H. A. Barker, Fermentations of Nitrogenous Organic Compounds. In *The Bacteria*; I. C. Gunsalus, R. Y. Stanier, Eds.; Academic Press: New York, 1961; Vol. 2, pp 151–188.
183. P. Durre; J. R. Andreesen, *J. Bacteriol.* **1983**, *154*, 192.
184. R. N. Costilow, *J. Bacteriol.* **1977**, *131*, 366.
185. T. C. Stadtman; P. Elliott, *J. Biol. Chem.* **1957**, *228*, 983.
186. M. X. Sliwkowski; T. C. Stadtman, *J. Biol. Chem.* **1987**, *262*, 4899.
187. M. X. Sliwkowski; T. C. Stadtman, *Biofactors* **1988**, *1*, 293.
188. G. E. Garcia; T. C. Stadtman, *J. Bacteriol.* **1992**, *174*, 7080.
189. Y. Kimura; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2189.
190. T. C. Stadtman; L. S. McClung, *J. Bacteriol.* **1957**, *73*, 218.
191. J. E. Cone; R. M. del Rio; T. C. Stadtman, *J. Biol. Chem.* **1977**, *252*, 5337.

192. T. C. Stadtman, *Annu. Rev. Biochem.* **1996**, *65*, 83.
193. A. Graentzdoerffer; A. Pich; J. R. Andreesen, *Arch. Microbiol.* **2001**, *175*, 8.
194. T. Gursinsky *et al.*, *Appl. Environ. Microbiol.* **2008**, *74*, 1385.
195. S. Kreimer; J. R. Andreesen, *Eur. J. Biochem.* **1995**, *234*, 192.
196. G. C. Mead, *J. Gen. Microbiol.* **1971**, *67*, 47.
197. M. Rother; A. Bock; C. Wyss, *Arch. Microbiol.* **2001**, *177*, 113.
198. M. N. Sela, *Crit. Rev. Oral Biol. Med.* **2001**, *12*, 399.
199. L. G. Simonson; K. T. McMahon; D. W. Childers; H. E. Morton, *Oral Microbiol. Immunol.* **1992**, *7*, 111.
200. R. Seshadri *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5646.
201. L. H. Stickland, *Biochem. J.* **1934**, *28*, 1746.
202. H. Tanaka; T. C. Stadtman, *J. Biol. Chem.* **1979**, *254*, 447.
203. V. Venugopalan, *J. Bacteriol.* **1980**, *141*, 386.
204. G. E. Garcia; T. C. Stadtman, *J. Bacteriol.* **1991**, *173*, 4908.
205. S. Jackson; M. Calos; A. Myers; W. T. Self, *J. Bacteriol.* **2006**, *188*, 8487.
206. L. Flohe; J. R. Andreesen; R. Brigelius-Flohe; M. Maiorino; F. Ursini, *IUBMB Life* **2000**, *49*, 411.
207. N. Schutze *et al.*, *Biofactors* **1999**, *10*, 329.
208. G. Rudnick; R. H. Abeles, *Biochemistry* **1975**, *14*, 4515.
209. J. R. Andreesen, *Curr. Opin. Chem. Biol.* **2004**, *8*, 454.
210. T. Grobe; M. Reuter; T. Gursinsky; B. Sohling; J. R. Andreesen, *Arch. Microbiol.* **2007**, *187*, 29.
211. R. W. Lovitt; D. B. Kell; J. G. Morris, *J. Appl. Bacteriol.* **1987**, *62*, 81.
212. J. R. Andreesen; A. Schaupp; C. Neurauter; A. Brown; L. G. Ljungdahl, *J. Bacteriol.* **1973**, *114*, 743.
213. J. R. Andreesen; L. G. Ljungdahl, *J. Bacteriol.* **1973**, *116*, 867.
214. L. G. Ljungdahl; J. R. Andreesen, *FEBS Lett.* **1975**, *54*, 279.
215. L. G. Ljungdahl; J. R. Andreesen, *Meth. Enzymol.* **1978**, *53*, 360.
216. W. Freudenberg; J. R. Andreesen, *J. Bacteriol.* **1989**, *171*, 2209.
217. D. Dietrichs; M. Meyer; M. Rieth; J. R. Andreesen, *J. Bacteriol.* **1991**, *173*, 5983.
218. L. F. Wu; M. A. Mandrand-Berthelot, *Biochimie* **1986**, *68*, 167.
219. M. A. Mandrand-Berthelot; G. Couchoux-Luthaud; C. L. Santini; G. Giordano, *J. Gen. Microbiol.* **1988**, *134*, 3129.
220. J. C. Boyington; V. N. Gladyshev; S. V. Khangulov; T. C. Stadtman; P. D. Sun, *Science* **1997**, *275*, 1305.
221. V. N. Gladyshev; S. V. Khangulov; M. J. Axley; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7708.
222. G. N. George *et al.*, *J. Am. Chem. Soc.* **1998**, *120*, 1267.
223. H. C. Raaijmakers; M. J. Romao, *J. Biol. Inorg. Chem.* **2006**, *11*, 849.
224. M. H. Yoo; X. M. Xu; B. A. Carlson; V. N. Gladyshev; D. L. Hatfield, *J. Biol. Chem.* **2006**, *281*, 13005.
225. S. R. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2521.
226. S. M. Anema *et al.*, *Biochem. J.* **1999**, *342*, 111.
227. S. Gromer; L. D. Arscott; C. H. Williams, Jr.; R. H. Schirmer; K. Becker, *J. Biol. Chem.* **1998**, *273*, 20096.
228. S. Gromer; R. H. Schirmer; K. Becker, *FEBS. Lett.* **1997**, *412*, 318.
229. V. N. Gladyshev; K. T. Jeang; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6146.
230. J. Kohrle, *Thyroid* **2005**, *15*, 841.
231. G. J. Beckett; J. R. Arthur, *J. Endocrinol.* **2005**, *184*, 455.
232. J. Kohrle, *Biochimie* **1999**, *81*, 527.
233. D. L. St. Germain; V. A. Galton, *Thyroid* **1997**, *7*, 655.
234. M. J. Berry; L. Banu; P. R. Larsen, *Nature* **1991**, *349*, 438.
235. R. A. Sunde; J. K. Evenson; K. M. Thompson; S. W. Sachdev, *J. Nutr.* **2005**, *135*, 2144.
236. M. Maiorino *et al.*, *J. Biol. Chem.* **2005**, *280*, 38395.
237. K. Wingler; C. Muller; R. Brigelius-Flohe, *Biofactors* **2001**, *14*, 43.
238. R. Brigelius-Flohe *et al.*, *Biofactors* **2001**, *14*, 101.
239. F. Ursini *et al.*, *Science* **1999**, *285*, 1393.
240. Y. Fu; W. H. Cheng; J. M. Porres; D. A. Ross; X. G. Lei, *Free Radic. Biol. Med.* **1999**, *27*, 605.
241. T. C. Stadtman, *FASEB J.* **1987**, *1*, 375.
242. I. Chambers *et al.*, *EMBO J.* **1986**, *5*, 1221.
243. Y. C. Awasthi; E. Beutler; S. K. Srivastava, *J. Biol. Chem.* **1975**, *250*, 5144.
244. J. Loflin; N. Lopez; P. D. Whanger; C. Kioussi, *J. Inorg. Biochem.* **2006**, *100*, 1679.
245. P. D. Whanger, *Cell Mol. Life Sci.* **2000**, *57*, 1846.
246. R. F. Burk; K. E. Hill, *Bioessays* **1999**, *21*, 231.
247. S. C. Vendeland *et al.*, *J. Biol. Chem.* **1993**, *268*, 17103.
248. R. F. Burk *et al.*, *J. Neurosci.* **2007**, *27*, 6207.
249. R. F. Burk; K. E. Hill; A. K. Motley, *J. Nutr.* **2003**, *133*, 1517S.
250. K. T. Suzuki; C. Sasakura; S. Yoneda, *Biochim. Biophys. Acta* **1998**, *1429*, 102.
251. R. F. Burk; K. E. Hill; R. Read; T. Bellew, *Am. J. Physiol.* **1991**, *261*, E26.
252. R. F. Burk, *J. Nutr.* **1989**, *119*, 1051.
253. M. A. Motsenbocker; A. L. Tappel, *Biochim. Biophys. Acta* **1982**, *719*, 147.
254. D. Y. Hwang *et al.*, *Neurochem. Res.* **2005**, *30*, 1009.
255. A. Clavero-Salas *et al.*, *Fish Shellfish Immunol.* **2007**, *23*, 459.
256. J. E. Oldfield, *Biol. Trace Elem. Res.* **1989**, *20*, 23.
257. H. Tajsharghi; N. Darin; M. Tulinius; A. Oldfors, *Neuromuscul. Disord.* **2005**, *15*, 299.
258. V. Allamand *et al.*, *EMBO Rep.* **2006**, *7*, 450.
259. Y. Okamoto *et al.*, *Neurogenetics* **2006**, *7*, 175.
260. M. Rederstorff; A. Krol; A. Lescure, *Cell Mol. Life Sci.* **2006**, *63*, 52.

261. Y. Ye; Y. Shibata; C. Yun; D. Ron; T. A. Rapoport, *Nature* **2004**, *429*, 841.
262. J. E. Curran *et al.*, *Nat. Genet.* **2005**, *37*, 1234.
263. A. Martinez *et al.*, *BMC Genomics* **2008**, *9*, 329.
264. E. K. Moses *et al.*, *Am. J. Obstet. Gynecol.* **2008**, *198*, 336.
265. Y. Gao *et al.*, *Cytokine* **2006**, *33*, 246.
266. K. H. Kim *et al.*, *Biochem. Biophys. Res. Commun.* **2007**, *354*, 127.
267. M. Alanne *et al.*, *Hum. Genet.* **2007**, *122*, 355.
268. V. N. Gladyshev; K. T. Jeang; J. C. Wootton; D. L. Hatfield, *J. Biol. Chem.* **1998**, *273*, 8910.
269. E. Kumaraswamy *et al.*, *J. Biol. Chem.* **2000**, *275*, 35540.
270. Y. J. Hu *et al.*, *Cancer Res.* **2001**, *61*, 2307.
271. E. Jablonska; J. Gromadzinska; W. Sobala; E. Reszka; W. Wasowicz, *Eur. J. Nutr.* **2008**, *47*, 47.
272. V. M. Labunskyy *et al.*, *J. Biol. Chem.* **2005**, *280*, 37839.
273. E. Kumaraswamy *et al.*, *Mol. Cell Biol.* **2003**, *23*, 1477.
274. R. K. Shrimali *et al.*, *J. Biol. Chem.* **2008**, *283*, 20181.
275. A. Sengupta *et al.*, *Biochem. J.* **2008**, *413*, 151.
276. A. Sengupta; B. A. Carlson; V. J. Hoffmann; V. N. Gladyshev; D. L. Hatfield, *Biochem. Biophys. Res. Commun.* **2008**, *365*, 446.
277. R. K. Shrimali *et al.*, *Neuromuscul. Disord.* **2007**, *17*, 135.
278. P. A. Sheridan *et al.*, *J. Nutr.* **2007**, *137*, 1466.
279. B. A. Carlson *et al.*, *J. Biol. Chem.* **2007**, *282*, 32591.
280. U. Schweizer *et al.*, *Biochem. J.* **2005**, *386*, 221.
281. B. A. Carlson *et al.*, *J. Biol. Chem.* **2004**, *279*, 8011.
282. H. Weissbach; L. Resnick; N. Brot, *Biochim. Biophys. Acta* **2005**, *1703*, 203.
283. H. Y. Kim; D. E. Fomenko; Y. E. Yoon; V. N. Gladyshev, *Biochemistry* **2006**, *45*, 13697.
284. H. Y. Kim; V. N. Gladyshev, *PLoS Biol.* **2005**, *3*, e375.
285. L. Johansson; G. Gafvelin; E. S. Arner, *Biochim. Biophys. Acta* **2005**, *1726*, 1.
286. H. Y. Kim; V. N. Gladyshev, *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1277.
287. A. Koc; A. P. Gasch; J. C. Rutherford; H. Y. Kim; V. N. Gladyshev, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7999.
288. H. Y. Kim; V. N. Gladyshev, *Biochemistry* **2005**, *44*, 8059.
289. J. Moskovitz; E. R. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7486.
290. D. Sagher; D. Brunell; N. Brot; B. L. Vallee; H. Weissbach, *J. Biol. Chem.* **2006**, *281*, 31184.
291. D. A. Saelinger; J. L. Hoffman; K. P. McConnell, *J. Mol. Biol.* **1972**, *69*, 9.
292. W. M. Ching; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 374.
293. W. M. Ching; A. J. Wittwer; L. Tsai; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 57.
294. A. J. Wittwer; L. Tsai; W. M. Ching; T. C. Stadtman, *Biochemistry* **1984**, *23*, 4650.
295. W. M. Ching; B. Alzner-DeWeerd; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 347.
296. A. J. Wittwer; T. C. Stadtman, *Arch. Biochem. Biophys.* **1986**, *248*, 540.
297. Z. Veres; L. Tsai; M. Politino; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6341.
298. Z. Veres *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2975.
299. Z. Veres; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8092.
300. J. L. Hoffman; K. P. McConnell, *Biochim. Biophys. Acta – Nucleic Acids Protein Synth.* **1974**, *366*, 109.
301. C. S. Chen; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1403.
302. A. J. Wittwer, *J. Biol. Chem.* **1983**, *258*, 8637.
303. C. Baron; A. Bock, *J. Biol. Chem.* **1991**, *266*, 20375.
304. H. Romero; Y. Zhang; V. N. Gladyshev; G. Salinas, *Genome Biol.* **2005**, *6*, R66.
305. M. J. Axley; A. Bock; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8450.
306. S. Gromer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12618.
307. R. Hille, *Arch. Biochem. Biophys.* **2005**, *433*, 107.
308. J. S. Holcenberg; E. R. Stadtman, *J. Biol. Chem.* **1969**, *244*, 1194.
309. G. L. Dilworth, *Arch. Biochem. Biophys.* **1983**, *221*, 565.
310. V. N. Gladyshev; P. Lecchi, *Biofactors* **1995**, *5*, 93.
311. V. N. Gladyshev; S. V. Khangulov; T. C. Stadtman, *Biochemistry* **1996**, *35*, 212.
312. J. C. Rabinowitz; H. A. Barker, *J. Biol. Chem.* **1956**, *218*, 147.
313. P. Dürre; J. R. Andreesen, *J. Gen. Microbiol.* **1982**, *128*, 1457.
314. H. Schiefer-Ullrich; R. Wagner; P. Durre; J. R. Andreesen, *Arch. Microbiol.* **1984**, *138*, 345.
315. W. T. Self; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7208.
316. W. T. Self; M. D. Wolfe; T. C. Stadtman, *Biochemistry* **2003**, *42*, 11382.
317. G. M. Lacourciere, *J. Bacteriol.* **2002**, *184*, 1940.
318. G. M. Lacourciere; H. Mihara; T. Kurihara; N. Esaki; T. C. Stadtman, *J. Biol. Chem.* **2000**, *275*, 23769.
319. G. M. Lacourciere; T. C. Stadtman, *J. Biol. Chem.* **1998**, *273*, 30921.
320. G. M. Lacourciere; T. C. Stadtman, *Biofactors* **2001**, *14*, 69.
321. H. Mihara *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6679.
322. H. Mihara; T. Kurihara; T. Watanabe; T. Yoshimura; N. Esaki, *J. Biol. Chem.* **2000**, *275*, 6195.

Biographical Sketch

William Self was born in a suburb of Birmingham, Alabama, in 1971 in a blue collar working class family whose roots lie in the coal mines and steel industry. At the age of eight his family relocated to Merritt Island, Florida, under the shadow of the US space program, where he showed a deep interest for science. He was raised in this area on the east coast and attended the University of Florida for his undergraduate studies in Microbiology. Upon graduation in 1993 he worked as a microbiologist for the State of Florida in the area of food safety for over a year, learning a great deal of bacteriology from the 'old school' of standard culture and isolation techniques. With a resolve to further his understanding of bacterial physiology, he returned to the University of Florida and received his Ph.D. in Microbiology in 1997. During his graduate work, his research contributed to our understanding of molybdenum transport, metabolism, and the regulation of transcription of genes encoding molybdoenzymes using *Escherichia coli* as a model system. K. T. Shanmugam was a mentor for his doctoral studies, a very well-respected expert in microbial physiology whose research has spanned molybdenum metabolism, studies of the molybdoenzyme nitrogenase, regulation of hydrogenases, and metabolic engineering for biofuels. Bill moved on to the National Institutes of Health in Bethesda, Maryland for a postdoctoral fellowship in the Laboratory of Biochemistry (National Heart, Lung and Blood Institute) working under the mentorship of Thressa C. Stadtman. Dr. Stadtman is well established as one of the leaders in the area of selenium metabolism and the study of selenoenzymes from bacteria to archaebacteria. During this training, Dr. Self gained an interest in studying the biological uses of selenium for labile selenoenzymes and in studying the metabolism of this metalloid in both mammals and bacterial model systems. In 2003 Dr. Self relocated to his current position at the University of Central Florida, and established an independent research program focused on selenium metabolism and selenoenzyme studies.

5.06 Protein Toxins from Bacteria

Holger Barth, University of Ulm Medical Center, Ulm, Germany

Bradley G. Stiles, Wilson College, Chambersburg, PA, USA; US Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

5.06.1 Introduction: Bacterial Toxins – An Overview

Over recent years, much progress has been made in understanding the molecular mode of action for different bacterial protein toxins, which involves new insights into their structures. This chapter provides a succinct overview on bacterial protein toxins, their structures, and their amazingly diverse modes of action. In particular, current knowledge will be presented for two different types of bacterial protein toxins, which include (1) a family of binary ADP-ribosylating toxins (AB type) that targets actin and (2) *Staphylococcus aureus* superantigens, such as the staphylococcal enterotoxins (SEs) and toxic shock syndrome toxin-1 (TSST-1), which target the immune system.

Extremely high virulence of certain bacterial strains is due to the production of toxins that elicit severe sequelae associated with a variety of diseases. Based upon chemical composition and mode of action, bacterial toxins are divided into endotoxins and exotoxins (see **Table 1**). Although the endotoxins consist of lipopolysaccharide (LPS) derived exclusively from the cell wall of Gram-negative bacteria, the exotoxins are proteins secreted by both Gram-negative and Gram-positive bacteria. One essential feature of exotoxins is that the soluble proteins are cytopathic upon targeted cells, even in the absence of offending bacteria; therefore, the cellular mode of action for exotoxins represents a real intoxication and minute quantities are usually sufficient for cell/host damage.

Generally, there are three major types of bacterial exotoxins that differ with respect to their structure and principle mode of action. First, there are toxins that attack the cell membrane and thereby damage eukaryotic

Table 1 Bacterial endotoxins and exotoxins

	Endotoxins (LPS)	*Exotoxins*
Producing bacteria	Gram negative	Gram negative and Gram positive
Release from bacteria	During cell fission (small amounts); during bacterial cell death (large amounts), which includes lysis during antibiotic treatment	Active secretion by growing bacteria
Chemical nature	Lipopolysaccharide (LPS), component of the bacterial outer membrane	Proteins
Mode of action	Activation of macrophages/monocytes: release of endogenous mediators such as lipids from arachidonic acid, reduced oxygen species, proteins	1. Pore formation in cell membranes 2. Enzymatic modification of specific substrates in the cytosol of host cells (AB-type toxins) 3. Superantigen stimulation of the immune system
Role for disease	Released mediators cause fever, hypotension, tachycardia, tachypnea, systemic shock, and multiorgan failure; not dependent upon the bacterial species	Depends upon the individual intravascular coagulation, leucopenia, toxin/bacterium (see **Table 2**)

cells. Second, there is a large group of bacterial toxins taken into eukaryotic host cells, which then modify specific substrate molecules in the cytosol. Third, a unique group of toxic proteins involves 'superantigens' produced by *S. aureus* and other bacteria/viruses that stimulate the host's immune system to dangerous, potentially life-threatening levels.

5.06.1.1 Toxins that Attack the Cell Membrane: Phospholipases and Pore-Forming Toxins

Certain bacterial toxins harbor phospholipase activity and attack eukaryotic cells by hydrolyzing membrane phospholipids. Most probably, the hydrolyzed membrane phospholipids become degraded which then leads to weakening of the cell membrane and finally cell lysis. Experimentally, the activity of bacterial phospholipases can be determined by monitoring lysis of erythrocytes. The prototype of bacterial phospholipases is the *Clostridium perfringens* α-toxin, a phospholipase C (PLC) that plays an essential role during pathogenesis of gas gangrene.[1,2] In recent years, bacterial phospholipases D (PLD) and A (PLA) have been identified and the production of phospholipases has been described for various Gram-positive bacteria, such as *Clostridium* and *Bacillus* species, as well as *S. aureus*. Gram-negative bacteria also produce phospholipases, for instance, *Pseudomonas aeruginosa*, which is a common nosocomial pathogen among patients suffering from burns and cystic fibrosis. Interestingly, intracellular pathogens such as *Listeria monocytogenes*, *Legionella pneumophila*, and *Mycobacterium tuberculosis* produce phospholipases. A possible role of the phosphatidylinositol-specific phospholipase C (PI-PLC) from *L. monocytogenes* for the pathogenesis of listeriosis was suggested by Vazquez-Boland *et al.*[3] Most probably, hydrolysis of membrane phospholipids enables this intracellular-replicating bacterium to spread from one cell to an adjacent neighbor. However, for many cases it is not completely known whether cell lysis is a direct consequence of either bacterial phospholipase activity or stress caused by modulating cellular metabolism.

Besides the enzymatic hydrolysis of membrane phospholipids, pore formation is another efficient mechanism used by certain bacterial toxins to directly damage the eukaryotic cell membrane (for review see Fivaz et al.[4]). The overall mechanism of pore-forming toxins is that the proteins first bind to the cell membrane as monomers. Pores are then formed by oligomerization of cell-bound toxin molecules with subsequent membrane insertion. Depending upon the number of monomers that form a pore, which influences the latter's diameter, these toxins can be classified into either small or large pore-forming toxins. The family of large pore-forming toxins, which have been investigated in detail by Tweten comprises of the cholesterol-dependent cytolysins (CDCs) that are produced by several species of *Clostridium*, *Listeria*, *Bacillus*, *Streptococcus*, and *Arcanobacterium* (for review see Tweten[5]). These toxins, for instance *Streptococcus pyogenes* streptolysin O (SLO) or *C. perfringens* perfringolysin O (PFO), bind as soluble monomers to cholesterol on the cell membrane. Following oligomerization, these toxins insert into the lipid membrane to form large pores.

The prototype of a small pore-forming toxin is the *S. aureus* α-toxin, also called α-hemolysin, that has been extensively investigated by Bhakdi and coworkers. Monomers of α-hemolysin (33 kDa) bind to the surface of erythrocytes, and after lateral diffusion within the lipid bilayer, seven monomers oligomerize to form pores in the cell membrane. The α-hemolysin forms mushroom-shaped pores with an outer diameter of 10 nm and an inner diameter of approximately 2.5 nm.[6] Small molecules can pass through the pore and diffuse into/out of the cytosol, along with water. As a consequence of such movement, cell homeostasis is greatly disturbed and pushed into an unhealthy state. In animals, the α-hemolysin represents a major virulence factor of *S. aureus* which causes hemolysis as well as tissue destruction.[7]

5.06.1.2 AB-Type Toxins Act as Enzymes in the Cytosol of Host Cells

'Classical' bacterial exotoxins, such as diphtheria toxin, cholera toxin, clostridial neurotoxins, and the anthrax toxins are enzymes that modify their substrates within the cytosol of mammalian cells. To reach the cytosol, these toxins must first bind to different cell-surface receptors and become subsequently internalized by the cells. To this end, many bacterial exotoxins contain two functionally different domains. The binding (B-) domain binds to a cellular receptor and mediates uptake of the enzymatically active (A-) domain into the cytosol, where the A-domain modifies its specific substrate (see **Figure 1**). Thus, three important properties characterize the mode of action for any AB-type toxin: selectivity, specificity, and potency. Because of their selectivity toward certain cell types and their specificity for cellular substrate molecules, most of the individual exotoxins are associated with a distinct disease. Because of their enzymatic nature, placement of very few A-domain molecules in the cytosol will normally cause a cytopathic effect. Therefore, bacterial AB-type exotoxins which include the potent neurotoxins from *Clostridium tetani* and *C. botulinum* are the most toxic substances known today. However, the individual AB-type toxins can greatly vary in terms of subunit composition and enzyme activity (see **Table 2**).

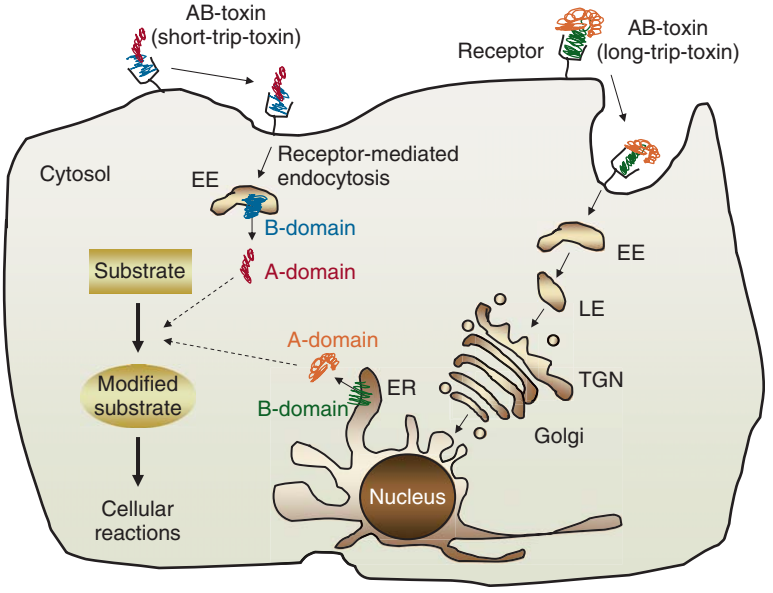


Figure 1 The mode of action for bacterial AB-type exotoxins. AB-toxins are enzymes that modify specific substrate molecules in the cytosol of eukaryotic cells. Besides the enzyme domain (A-domain), AB-toxins have a binding/translocation domain (B-domain) that specifically interacts with a cell-surface receptor and facilitates internalization of the toxin into cellular transport vesicles, such as endosomes. In many cases, the B-domain mediates translocation of the A-domain into the cytosol by pore formation in cellular membranes. By following receptor-mediated endocytosis, AB-type toxins exploit normal vesicle traffic pathways into cells. One type of toxin escapes from early acidified endosomes (EE) into the cytosol, thus they are referred to as 'short-trip-toxins'. In contrast, the 'long-trip-toxins' take a retrograde route from early endosomes (EE) through late endosomes (LE), trans-Golgi network (TGN), and Golgi apparatus into the endoplasmic reticulum (ER) from where the A-domains translocate into the cytosol to modify specific substrates.

Table 2 Examples of bacterial AB-type exotoxins

Toxin	*Source*	*Structure*	*Substrate*	*Effect on cells*	*Disease*
ADP-ribosyltransferases					
CT	*Vibrio cholerae*	AB_5	Heterotrimeric GTPases	Altered signal transduction	Cholera
LT	*Escherichia coli* (ETEC)	AB_5	Heterotrimeric GTPases	Altered signal transduction	Severe diarrhea
PT	*Bordetella pertussis*	AB_4	Heterotrimeric GTPases	Altered signal transduction	Pertussis (whooping cough)
DT	*Corynebacterium diphtheriae*	AB	Elongation Factor-2	Inhibition of protein synthesis	Diphtheria
Exotoxin A	*Pseudomonas aeruginosa*	AB	Elongation Factor-2	Inhibition of protein synthesis	Pneumonia
C2	*Clostridium botulinum*	Binary	Actin	Depolymerization of F-actin	? (see text)
Iota	*Clostridium perfringens*	Binary	Actin	Depolymerization of F-actin	Enteritis
CDT	*Clostridium difficile*	Binary	Actin	Depolymerization of F-actin	? (see text)
CST	*Clostridium spiroforme*	Binary	Actin	Depolymerization of F-actin	Enteritis
Glucosyltransferases					
TcdA	*Clostridium difficile*	AB	Rho-GTPases	Rho-signaling disturbed	Pseudomembranous colitis
TcdB	*Clostridium difficile*	AB	Rho-GTPases	Rho-signaling disturbed	Pseudomembranous colitis
Proteases					
BoNTs A-G	*Clostridium botulinum*	AB	Adaptor proteins	Block of neurotransmitter release	Botulism
TeNT	*Clostridium tetani*	AB	Adaptor proteins	Enhanced neurotransmitter release	Tetanus
Lethal toxin	*Bacillus anthracis*	Binary	MAPKK	Cell death	Anthrax
Deamidases					
CNF	*Escherichia coli*	AB	Rho-GTPases	Rho-signaling disturbed	Extraintestinal infections

For the uptake of bacterial AB-toxins into the cytosol of eukaryotic cells, a precisely concerted series of the following steps is required: (1) toxin activation; (2) receptor binding; (3) receptor-mediated endocytosis; (4) vesicle transport; and (5) translocation of the A-domain from vesicular compartments into the cytosol (for review see Montecucco[8]). Many toxins require activation of their B-domain, for example, by proteolytic cleavage to gain full biological activity. For instance, binary ADP-ribosylating toxins that modify actin require N-terminal cleavage of their B-domain to form oligomers that bind to both the receptor and A-component(s) (for review see Barth *et al.*[9]). *In vitro*, trypsin, chymotrypsin, or furin cleave the B-domains, depending on the individual binary toxin. *In vivo*, host proteases naturally found in the gastrointestinal tract probably activate these toxins. The activated toxins initially bind to unique cell-surface receptors through their B-domain (mediated by the latter's C-terminal region), followed by the receptor–toxin complex becoming internalized by receptor-mediated endocytosis. Following endocytosis, toxin is transported into the cell through vesicles (e.g., endosomes) which are part of a normal, physiologically relevant traffic route for host proteins. This implies that the toxin is protected from proteolytic degradation in the cytosol during cellular uptake. The A-domain must ultimately be delivered into the cytosol to reach its cellular substrate, and this is accomplished by two classical methods (reviewed in Sandvig *et al.*,[10] van der Goot and Gruenberg,[11] and Olsnes *et al.*[12]). A few of the AB-toxins, for instance diphtheria, anthrax, and C2 toxins, deliver their enzymatic domain from acidic endosomes into the cytosol (see **Figure 1**). These toxins are also referred to as 'short-trip-toxins'. In contrast, the 'long-trip-toxins' such as cholera toxin take a route through the endosomes, trans-Golgi network (TGN), and Golgi apparatus retrograde into the endoplasmic reticulum (ER). It is from the ER where the A-domain is finally released into the cytosol. Thus, AB-toxins exploit physiological trafficking pathways of the cell to enter the cytosol (see **Figure 1**).

5.06.1.3 Superantigens

A third group of protein toxins includes those with superantigenic properties. These molecules, produced by various bacterial pathogens that include *S. aureus* and *S. pyogenes*, are rather unique in their mode of action versus the aforementioned toxins. Superantigenic toxins are typically devoid of enzymatic properties and they are not internalized or imbedded into the membrane of a host cell. Superantigens cause profound, deleterious effects upon a host by hyperstimulating the immune system through interactions with both antigen-presenting cells and T cells. This toxin-induced 'team effort' between cells can lead to lethal shock, due to abnormally high serum levels of pro-inflammatory cytokines. Various aspects of microbial superantigens will be further described in finer detail later in this review.

5.06.2 Potent Virulence Factors Directly Attack the Actin Cytoskeleton of Mammalian Cells: Actin-ADP-Ribosylating Toxins

Several bacterial virulence factors directly attack the cytoskeleton of eukaryotic cells by mono-ADP-ribosylating actin. These toxins catalyze the covalent transfer of an ADP-ribose moiety from nicotinamide-adenine dinucleotide (NAD) onto Arg177 of G-actin, which then turns the actin molecule into a 'capping protein'.[13] This finally results in inhibition of barbed-end polymerization by nonmodified G-actin (see **Figure 2**). Most probably, ADP-ribosylation of actin at Arg177 leads to a severe steric clash and disruption of important contact sites, which hold the actin filaments (F-actin) together.[14] Moreover, ADP-ribosylation inhibits the intrinsic ATPase activity of actin.[14] When taken altogether, ADP-ribosylation of G-actin leads to a complete depolymerization of F-actin resulting in cytoskeletal disruption and rounding of cultured cell monolayers (see **Figure 2**).

On the one hand, the actin-ADP-ribosylating toxins can be further divided into the virulence factor SpvB from *Salmonella enterica*,[15] and on the other, the *Clostridium*/*Bacillus* family of binary actin-ADP-ribosylating toxins. The latter is comprised of *C. botulinum* C2 toxin,[16] *C. perfringens* iota toxin,[17] *C. spiroforme* toxin,[17,18] *C. difficile* ADP-ribosyltransferase,[19] and the VIP (vegetative insecticidal protein) toxins from *B. cereus*/*B. thuringiensis*.[20] Although SpvB is delivered from intracellular-located *Salmonella* directly into the cytosol, the binary toxins represent typical exotoxins, which then intoxicate eukaryotic target cells in the absence of toxin-producing bacteria.

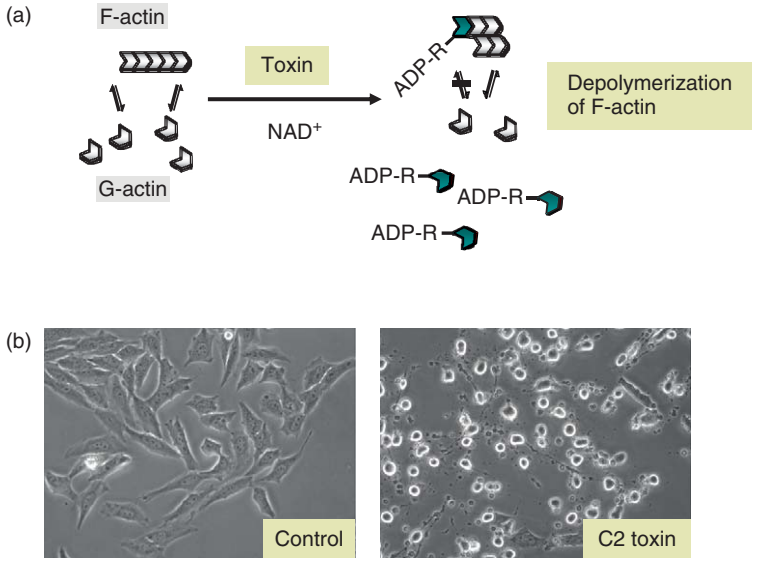


Figure 2 The actin-ADP-ribosylating toxins. (a) Molecular mode of action. The actin-ADP-ribosylating toxins covalently transfer an ADP-ribose moiety from NAD^+ onto Arg177 of G-actin in the cytosol of targeted cells. Mono-ADP-ribosylated G-actin acts as a 'capping protein' and inhibits the assembly of nonmodified actin into filaments. Thus, actin polymerization is blocked at the fast-growing ends of actin filaments (plus or barbed ends) but not at the slow growing ends (minus or pointed ends). This effect ultimately increases the 'critical concentration' necessary for actin polymerization and tends to depolymerize F-actin. Finally, all actin within an intoxicated cell becomes 'trapped' as ADP-ribosylated G-actin. (b) Cytopathic effect of actin-ADP-ribosylating toxins. Incubation of cultured cells with *Clostridium botulinum* C2 toxin (200 ng ml^{-1} of C2I + 400 ng ml^{-1} of C2IIa), an actin-ADP-ribosylating toxin, for 3 h at 37 °C results in a dramatic change (overt rounding) in cell morphology.

5.06.2.1 *Salmonella enterica* SpvB

Salmonella enterica is a Gram-negative, food-borne pathogen that causes human diseases ranging from mild gastroenteritis to severe systemic infections. For an infection to occur, the intracellular growth of *Salmonella* in macrophages is crucial. The bacteria are located in a special membrane compartment, the so-called *Salmonella*-containing vacuole (SCV).[21] Following replication, *Salmonella* escapes from the SCV and induces cell death among infected macrophages. An actin-ADP-ribosylating virulence factor (SpvB – *Salmonella* plasmid virulence B) is highly essential for intracellular growth, and thus virulence, of *S. enterica*.[15] SpvB is not an exotoxin and therefore requires the presence of *Salmonella* for its transport into the cytosol of mammalian cells. Most probably, the 65 kDa SpvB protein is directly secreted into the cytosol from intracellular growing bacteria through a type-III-secretion mechanism (see **Figure 3**). Type-III-secretion implies the formation of a bacterial protein needle (injectisome) that extends from the bacterium and through the host-cell membrane into the cytosol. The N-terminal domain of *S. enterica* SpvB shares homology with a secretory protein (TcaC) from *Photorhabdus luminescens*, an insect pathogen.[15]

The C-terminal domain of SpvB (C/SpvB) harbors an ADP-ribosyltransferase domain and shares sequence similarity with other actin-ADP-ribosylating toxins.[22] C/SpvB contains the highly conserved, essential residues Glu536 and Glu538[15] (see **Figure 3**). This motif is characteristic for the arginine-specific bacterial mono-ADP-ribosyltransferases. Recently, a crystal structure for the ADP-ribosyltransferase domain of SpvB (amino acids 390–591) was solved, revealing striking similarities to previously characterized ADP-ribosyltransferases.[14] Between the N- and C-terminal domains, there are seven proline residues which might be involved in translocating the protein into the cytosol. Recently, the molecular mode of action for SpvB was characterized and involves ADP-ribosylation of G-actin at Arg177.[14,23,24] As a consequence, the F-actin become depolymerized which adversely affects the cytoskeleton and many cell-essential functions critical for life.

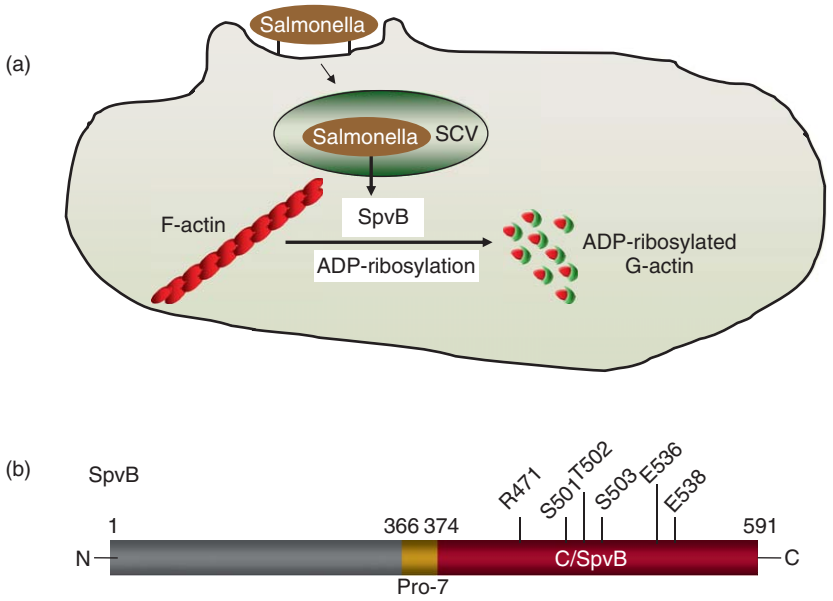


Figure 3 The SpvB virulence factor is important for intracellular growth of *Salmonella enterica*. (a) An ADP-ribosyltransferase, SpvB, is crucial for the intracellular growth of *Salmonella enterica* in macrophages and host infection. Strains without SpvB are less virulent. During infection, the bacterium replicates inside macrophages in a special membrane compartment, the *Salmonella*-containing vacuole (SCV). Most probably, SpvB is delivered directly into the cytosol from engulfed *Salmonella* through type-III-secretion. Once in the host cell cytosol, SpvB mono-ADP-ribosylates G-actin at Arg177, which then leads to depolymerization of actin filaments. (b) The ADP-ribosyltransferase SpvB. An N-terminal domain (amino acids 1–365) shares sequence homology with a secreted, insecticidal protein (TcaC) from *Photorhabdus luminescens*. The seven proline residues (Pro-7) that connect the N- with the C-terminal domains of SpvB might be involved in protein translocation. The C-terminal domain (C/SpvB) harbors the catalytic site for ADP-ribosyltransferase activity. The highly conserved residues that compose the catalytic site are depicted. In particular, Glu536 and Glu538 residues are characteristic of arginine-specific mono-ADP-ribosyltransferases from bacteria.

5.06.2.2 Binary Actin-ADP-Ribosylating Toxins

Binary toxins are unique concerning their structure because they are comprised of two individual, nonlinked proteins represented by an enzyme component and a binding/translocation component. The two components are secreted by the bacterium and assemble upon the surface of targeted eukaryotic cells to form an active toxin complex. For this to occur, both protein components of binary toxins act in a precisely concerted manner. The binding component first engages the cell-surface receptor and then mediates translocation of enzyme component(s) from the outside of a cell, through acidified endosomes, and into the host cell cytosol where it modifies the substrate (for review see Barth[25]).

Binary toxins classically include the two anthrax toxins, lethal and edema, as well as the actin-ADP-ribosylating toxins. The family of binary actin-ADP-ribosylating toxins includes the unique *C. botulinum* C2 toxin and iota-like proteins such as *C. perfringens* iota toxin, *C. difficile* toxin (CDT), and *C. spiroforme* toxin (CST). Besides the clostridial toxins, there are the related vegetative insecticidal proteins (VIP) of *B. cereus*/*B. thuringiensis* (for overview see **Table 2**). All members of this toxin family share a basic, cytotoxic mode of action as described above. In brief, the enzyme components mono-ADP-ribosylate G-actin at Arg177, leading to complete depolymerization of F-actin. Thus, mono-ADP-ribosylation of G-actin represents a very efficient mechanism by which the toxins directly attack the cytoskeleton of mammalian cells.

5.06.2.2.1 Clostridium botulinum C2 toxin: The prototype of a binary actin-ADP-ribosylating toxin

In 1980, *C. botulinum* C2 toxin was the first binary actin-ADP-ribosylating toxin to be described in the literature.[26] The C2 toxin, produced by *C. botulinum* types C and D, consists of two nonlinked components

designated as C2I and C2II. Ohishi *et al.*[27] discovered that both of these proteins are required to mediate the cytotoxic effect. Aktories *et al.*[16] demonstrated that C2I is an ADP-ribosyltransferase that exclusively modifies actin at Arg177. The discovery of this mode of action revealed a completely novel principle for how bacterial protein toxins directly damage eukaryotic cells (through the actin cytoskeleton), thus introducing a new family of bacterial toxins.

The role of C2 toxin in disease is not clear because all *C. botulinum* strains that produce C2 toxin also synthesize extremely potent neurotoxins, the effector molecules of botulism. When Simpson compared the pharmacological properties of *C. botulinum* neurotoxin type C1 with C2 toxin in detail, it became obvious that C2 toxin does not cause the flaccid paralysis symptoms attributed to classic botulism.[28] However, isolated C2 toxin is a potent enterotoxin that proves lethal in various animals: 2 pmol of C2 toxin readily kill mice, rats, guinea pigs, and chickens within 1 h after application.[28] For mice, the LD_{50} (i.v.) of C2 toxin is less than 50 fmol. Ohishi and Odagiri also reported that C2 toxin causes necrotic, hemorrhagic lesions in the intestinal wall,[29] whereas Simpson reported that C2 toxin elicits hypotension as well as fluid accumulation in the lungs.[28]

How is the C2 toxin internalized into the targeted cell's cytosol? It was observed early by Ohishi[30] that C2II requires proteolytic activation for biological activity. Barth *et al.*[31] then discovered the molecular consequences of C2II activation and its role for cellular uptake of an ADP-ribosyltransferase, C2I. The C2II protein (721 amino acids, 80.8 kDa) is cleaved at Lys181, thus resulting in an active C2IIa molecule (60 kDa).[31] Under physiological conditions, this activation occurs most probably within the gut where various host and bacterial proteases commonly exist in any healthy being. C2IIa, but not C2II, serves a dual function by forming ring-shaped heptamers that assemble with C2I and bind to the cellular receptor.[31,32] The receptor for C2 toxin has been identified as a complex, hybrid carbohydrate structure present on all yet tested eukaryotic cells.[33] Such findings easily explain the sensitivity of all tested cell types toward C2 toxin. Thus, C2IIa heptamers represent an active species of the binding component.

Recently it was found that the C2IIa–C2I complex can be formed either in solution or on the cell surface.[34] In the latter scenario, receptor-bound C2IIa serves as a docking platform on the cell for C2I. In any case, the toxin complex is internalized by receptor-mediated endocytosis and reaches endosomal compartments. Subsequently, acidification of endosomes triggers membrane insertion and thereby pore formation of C2IIa heptamers.[31] The C2IIa heptamers serve two different functions during toxin uptake, depending upon their conformation. Although adopting the pre-pore conformation, C2IIa first facilitates binding of C2I to the cell surface which forms a C2IIa–C2I complex.[34,35] Following conversion into a pore conformation, C2IIa then mediates translocation of C2I into the cytosol.[31,36] For this latter step, an unfolding of the C2I protein (50 kDa) is essential.[37] A model of the C2IIa pre-pore reveals a very narrow pore diameter of 2.7 nm.[38] However, for a native conformation of C2I to squeeze through this pore, there must be a minimum diameter of more than 4 nm.[38] The subsequent refolding of C2I in the cytosol is strictly dependent upon a chaperone molecule provided by the host, heat-shock protein (Hsp) 90.[39]

Once in the cytosol, C2I can ADP-ribosylate various G-actin types (β/γ nonmuscle and γ-smooth muscle) at position Arg177,[40] thereby inducing the cytopathic effects described above in more detail. The ADP-ribosyltransferase activity of C2I is located within the C-terminal domain which contains the 'catalytic' amino acids essential for enzymatic activity of C2I[41] (see **Figure 4**). These highly conserved residues have been identified among all ADP-ribosyltransferases. The N-terminal domain of C2I (amino acids 1–225), also referred to as C2IN, mediates both the binding to C2IIa pre-pores and later during cellular uptake, the translocation of C2I across endosomal membranes through the lumen of C2IIa pores.[42] C2IN has been successfully used as an adaptor for efficient transport of foreign cargo proteins into the cytosol of eukaryotic cells.[24,42,43] Thus, C2 toxin was the first binary actin-ADP-ribosylating toxin successfully used for protein transport into eukaryotic cells. Recombinant C2 fusion toxins have subsequently been used as cell-permeable tools to study various processes in cellular physiology (for overview see Barth *et al.*[9,44])

5.06.2.2.2 The iota-like toxins

The binary nature of iota toxin from *C. perfringens* type E was first explored in 1986 by Stiles and Wilkins.[17,45] The overall mode of action for iota toxin is widely comparable to C2 toxin. The binding/translocation component iota b (Ib) facilitates cellular uptake of the enzyme component iota a (Ia) in a like manner as previously described for C2 toxin. Ia, just as C2I, specifically mono-ADP-ribosylates G-actin at Arg177.[46]

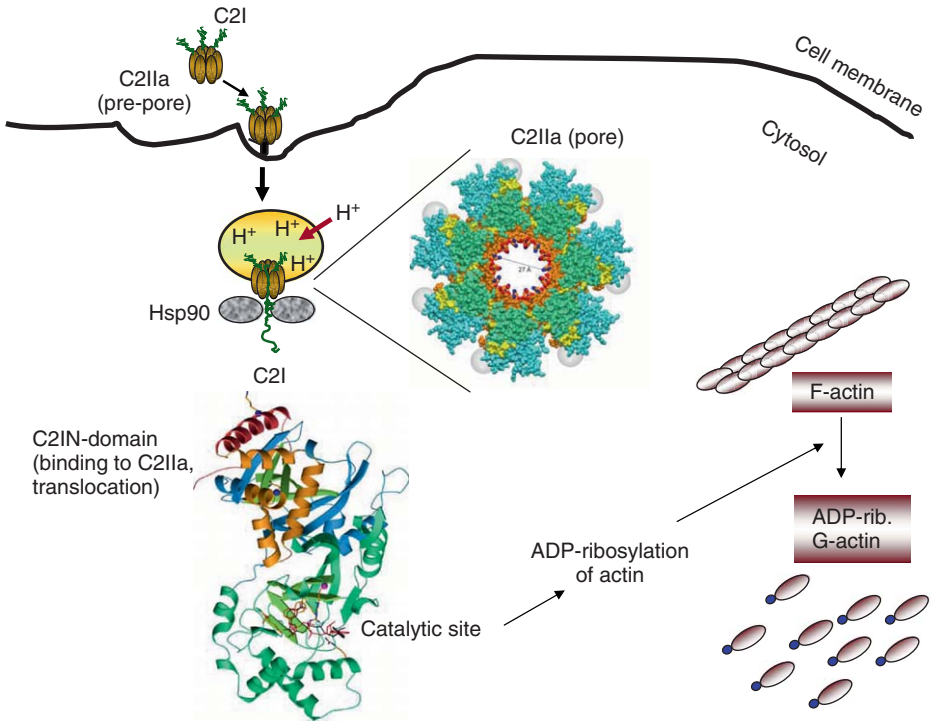


Figure 4 The internalization of binary *Clostridium botulinum* C2 toxin into eukaryotic cells. Cellular uptake of the binary C2 toxin requires a concerted interaction between two components, C2IIa and C2I. First, the activated binding/translocation component (C2IIa) mediates toxin binding to a carbohydrate receptor on the cell surface. Most probably, three molecules of C2I then bind to one heptamer of C2IIa in a pre-pore conformation. Subsequently, the toxin–receptor complex is internalized and reaches early endosomes. Following acidification of the endosomal lumen by vacuolar-ATPase located on the endosomal membranes, C2IIa changes its conformation from a pre-pore to pore and inserts into the endosomal membrane. The lumen diameter of the pore is 2.7 nm, through which C2I translocates into the cytosol. C2IIa pores, in combination with the H^+ gradient existing between the endosomes and cytosol, are absolutely essential for translocating C2I. To pass through the channel, C2I becomes partially unfolded and a host-cell chaperone, Hsp90, is required to properly refold C2I within the cytosol. Once in the cytosol, C2I ADP-ribosylates G-actin which then leads to depolymerization of actin filaments and complete destruction of the host's cytoskeleton.

However, there are some minor, yet very specific, differences that exist between the C2 and iota-like toxins. The genes for C2 toxin are located on the *C. botulinum* chromosome whereas those for *C. perfringens* iota toxin are plasmid bound.[47] The iota toxin is closely related to the *C. spiroforme* toxin CST (enzyme component Sa, binding/translocation component Sb) and the *C. difficile* toxin CDT (enzyme component CDTa, binding/translocation component CDTb). Within the group of iota-like toxins, the enzyme and binding/translocation proteins share a high degree of amino acid homology but are rather distinct from C2 toxin. The binding components of all iota-like toxins generate biologically-active chimeras when combined with enzyme from another iota-like toxin.[48–50]

All three iota-like toxins are linked to various gastrointestinal diseases in humans and/or animals.[51–53] As an example of a binary toxin linked to disease, *C. spiroforme* solely produces (from a toxin perspective) CST that causes diarrhea and lethality in rabbits.[54] In contrast, both *C. difficile* and especially *C. perfringens* produce multiple toxins that make linking any toxin to a specific disease more difficult for any investigator. Isolated iota toxin acts as an enterotoxin and causes diarrhea in calves and lambs.[17,45] The role of CDT in *C. difficile* pathogenesis is not completely understood, but there seems to be a link between the toxin and recent emergence of an epidemic strain. In humans, *C. difficile* causes a severe, recurring antibiotic-associated diarrhea and pseudomembranous colitis.[55–57] Two additional protein toxins from *C. difficile*, the so-called large cytotoxins TcdA and TcdB, which inactivate Rho-GTPases by monoglucosylation,[58,59] have been considered the primary toxin-based virulence factors involved in disease.[60] The binary CDT is produced by relatively few

C. difficile strains (approximately 16% in the United States and 6% in the United Kingdom) isolated from hospital patients.[61,62] However, many studies among hospital patients suggest that CDT is correlated with emerging, and particularly virulent, epidemic strains found around the world.[63–65] A recent epidemiological study from France suggests that strains of *C. difficile* expressing CDT are more virulent than those not expressing the binary toxin.[66] However, further work is required to elucidate the exact role played by CDT in *C. difficile* pathogenesis.

The following sections of this review will now transition into a unique group of protein toxins, the SEs and TSST-1. These proteins secreted by *S. aureus* work in a different fashion versus the aforementioned binary toxins. In fact, these staphylococcal toxins do not enter a cell and do not directly injure the targeted cell surface. Toxin damage is insidiously indirect and caused by an over zealous response executed by the host's immune system.

5.06.3 Introduction of a 'Super' Microbe

Staphylococcus aureus (Greek for 'Golden Seed Cluster') is a common microbe associated with numerous human and animal diseases.[67,68] This Gram-positive, sphere-shaped bacterium was first described as a pathogen (source being pus abscess) in the early 1880s by a Scottish surgeon named Alexander Ogston. The discovery was made possible by employing specific techniques, plus general microbial concepts, first practiced respectively in the German and French laboratories of Robert Koch and Louis Pasteur.

Staphylococcus aureus can readily colonize the skin and various mucosal surfaces, subsequently producing numerous virulence factors that promote their own survival and possible ill effects upon a host. For many individuals that are healthy and immunocompetent, colonization by most strains of *S. aureus* results in no harm. In fact, most people become asymptomatic carriers of *S. aureus* during their lifetime. However, through a natural evolutionary process, this bacterium has become rather adept at synthesizing various proteins that aid in its success as a biological entity. In addition to the many single-chain proteins called SEs, which adversely affect a host by stimulating an immune response,[69,70] *S. aureus* strains can also produce other protein-based virulence factors such as TSST-1, protein A, coagulases, hemolysins, and leukocidins.[71]

An additional reality we face today with this bacterium involves increasing resistance to antibiotics such as methicillin[72,73] (these strains are called MRSA or methicillin-resistant *S. aureus*), and our last ever-thinning line of defense, vancomycin[74,75] (these strains are called VRSA or vancomycin-resistant *S. aureus*). Recent efforts show that various MRSA strains produce protein toxins such as SEA, TSST-1, and leukocidin, which contribute to life-threatening forms of toxic shock syndrome (TSS).[76] Obviously, proper use of antibiotics is very important for fighting *S. aureus* in the clinic, as different antibiotics when improperly employed may enhance expression of various virulence factors such as TSST-1.[77] It is true that *S. aureus* represents a real health and economic concern throughout various sectors of our society.[78–80] Besides additional antibiotics (which are for various reasons relatively slow in forthcoming?) and strict adherence to infection control plans in health care facilities, one way of fighting back involves targeting the organism's protein toxins. At this time, there are sadly no effective vaccines or therapeutics (other than existing antibiotics and occasional immunoglobulins) commonly used in the clinic for targeting *S. aureus* or its toxins. Experimentally, there have been triumphs reported by various groups but many of these seemingly remain mired at the laboratory bench.

A successful microbe, such as any living entity, must adapt to its environment through a natural evolutionary process. External pressure(s) placed upon any creature produces results that involve successful adaptation, or death. The pivotal question for us clearly becomes whether *Homo sapiens*, through proper application of medical science, can effectively do the same and stay one step ahead of *S. aureus* (and other marauding microbes)?

To provide an appropriately brief background, the SEs (types A–U) are associated with a prevalent form of food poisoning readily found throughout the world.[81–83] The first documented report of human staphylococcal food poisoning was in 1914 (Philippine Islands) following consumption of contaminated milk from a cow with *S. aureus*-induced mastitis.[82] Recent gene-based studies show that mastitis strains produce one or more enterotoxins, which by definition act upon the gut and cause gastroenteric problems (i.e., diarrhea, vomiting, etc.), and a fourth of all these strains curiously carry the TSST-1 gene.[84] SE poisoning typically occurs within hours after ingestion of processed meats or dairy products previously tainted by improper handling (e.g., no gloves) and subsequent storage at elevated temperatures. Under these conditions, the newly transplanted *S. aureus* grow and concomitantly

produce one or more SEs as a toxic by-product. In humans, ingestion of only low microgram quantities of toxin in tainted food elicits emesis and diarrhea within 4 h after consumption. The SEs are rather hearty molecules as they generally resist heat, proteases, and extreme pH.[85] Ill effects due to SE intoxication are probably mediated through inflammatory compounds produced by one's own body, and these effector molecules include the prostaglandins (fatty acid derivatives of prostanoic acid) plus arachidonic acid-derived leukotrienes.[86,87] Residual effects of SE-based food poisoning can be felt upto 48 h after eating the contaminated food. This form of food poisoning is rarely fatal and most often linked to the SEA serotype. It is evident that various populations throughout the world are naturally exposed to these protein toxins, as demonstrated by SEB seroconversion. Whether one naturally develops antibodies toward preformed toxin ingested in food and/or there is transient/permanent colonization of the host by a toxin-producing strain of *S. aureus* remains a medical mystery. The fact is that all of us are constantly exposed to *S. aureus* and its protein toxins on a regular basis, and in a variety of ways.

In addition to the food-borne SEs, TSS caused by *S. aureus* TSST-1 was first described by Todd *et al.*[88] in 1978 and later linked to menstruation/improper tampon usage.[89] Originally, the TSST-1 protein was erroneously thought to be an enterotoxin and some of the early literature actually describes it as SEF.[90] However, this later proved to be a misnomer as homogeneous TSST-1 lacks enterotoxic effects upon nonhuman primates employed in a classic model for SE intoxication.[91] Unlike the SEs, TSST-1 is evidently destroyed by the stomach's acidity and/or intestinal proteases. The signs of TSST-1 intoxication are linked to an altered immune response that includes elevated serum levels of proinflammatory cytokines,[92] rash, hypotension, fever, and multisystem dysfunction.[93] Although less common, nonmenstrual TSS in men, women, and children is also attributed to SEB or SEC1 following *S. aureus* growth upon other body sites.[94] Unfortunately, all TSS patients may suffer recurring bouts unless the offending strain of *S. aureus* is eliminated or held to minimal growth by competing flora.

Antibodies play an important role in human susceptibility to TSST-1.[95] Individuals lacking, or possessing insufficient neutralizing levels of, toxin-specific antibodies frequently experience recurring bouts of TSS. A recent study shows that TSST-1-specific antibodies from patients are IgG and IgM (no IgA?), but the best neutralizers prove to be the IgG1 and IgG4 subclasses[96] that readily pass through the placenta. Related results from another laboratory surprisingly reveal that approximately 40% of pregnant Japanese women lack TSST-1-specific antibodies.[97] Perhaps these findings, and many others, further underscore the importance of experimental vaccines that may break immunological tolerance toward TSST-1 and other bacterial superantigens?[98–105]

Over the past 20 years, the *S. aureus* SEs and TSST-1 have been called 'superantigens'. This term was first proposed in the late 1980s by Marrack and Kappler[85] to describe microbial (bacterial or viral) proteins that activate large numbers of specific T cells versus conventional antigens. **Table 3** lists the various superantigens, and their diverse microbial sources, known to date. By definition, interactions of superantigens with cells of the

Table 3 Superantigens found in diverse bacterial and viral forms

Origin (bacterium/virus)	*Identified superantigen*
Helicobacter pylori	?
Mycobacterium tuberculosis	MTS
Mycoplasma arthritidis	MAM
Streptococcus dysgalactiae	SPEGG
Streptococcus equi	SEEL, SEEM, SZEL, SZEM
Streptococcus pyogenes	SPEA, C, G, H, I, J; SMEZ; and SSA
Yersinia enterocolitica	YES
Yersinia pseudotuberculosis	YPM
Cytomegalovirus	?
Epstein–Barr virus/human endogenous retrovirus	HERV-K18
Herpes virus	HVS 14
Human immunodeficiency virus	NEF
Human papillomavirus	?
Mouse mammary tumor virus	MMTV
Rabies virus	Nucleocapsid

host immune system, differ from conventional antigens through (1) direct binding of superantigens outside the peptide-binding groove of the major histocompatibility complex class II molecule; (2) superantigens exerting biological effects as an intact molecule without internalization or 'proteolytic processing' by antigen-presenting cells; and (3) superantigens not being major histocompatibility complex restricted for presentation to T cells.[85] Because of interactions with both major histocompatibility complex class II and T-cell receptor, a superantigen stimulates both antigen-presenting cells and T cells to ultimately release abnormally high levels of proinflammatory cytokines. Picomolar concentrations of these toxins activate specific Vβ-bearing T cells after binding to major histocompatibility complex class II on antigen-presenting cells. In essence, *S. aureus* SEs and TSST-1 perversely stimulate the body to turn upon itself from within through an immune system obviously meant to protect the host. This process is quite unique versus the many other nonsuperantigenic toxins produced by bacterial pathogens. We now, in greater detail, share how the superantigens elicit biological effects in various mammals as per structure/function analysis.

5.06.4 Structural Commonalities/Differences among SEs and TSST-1

The SEs and related TSST-1 are single-chain proteins (22–30 kDa) divided into common homology groups based upon amino acid sequences. There are three major groups consisting of (1) SEA, SEE, SED, SEJ, SEN, SEO, plus SEP; (2) SEI, SEK, SEL, SEM, plus SEQ; and (3) SEB, SEC, SEG, plus SEU.[106] Minor, distinct groups include SEH and TSST-1. These toxins are usually encoded by plasmids, bacteriophages, or mobile genetic elements and generally appear during the late logarithmic to stationary phases of *S. aureus* growth *in vitro*.[107] The SECs are unique in that they contain a His-Glu-X-X-His motif which represents a conserved, Zn^{2+}-binding site found in thermolysin-like metalloproteases. As diverse bacterial examples, these types of proteases are produced by *Bacillus anthracis* (lethal factor of anthrax lethal toxin) and *C. botulinum* (botulinum neurotoxins). However, unlike these latter toxins, there is no proteolytic activity evident with the SECs. Perhaps the His-Glu-X-X-His sequence in SECs represents a primordial remnant of a previous proteolytic activity associated with bacterial superantigens that now serves no obvious function for *S. aureus*. The adage touting "if you do not use it, you lose it" is perhaps appropriate. Similar to the metalloprotease sequence remnants among the SECs, superantigens from even the same microorganism are not the same in so many diverse ways. Regarding unique biological activities attributed to superantigens, one type produced by *Mycoplasma arthritidis* evidently possesses DNAse activity never described for other characterized superantigens.[108]

Besides amino acid homologies, X-ray crystallographic analysis of various *S. aureus* superantigens suggests a conserved structure containing two tightly packed domains composed of β-sheets and α-helices;[109–112] (see **Figure 5**). All of these toxins possess a common three-dimensional shape consisting of a shallow groove, considered the T-cell receptor binding site, which separates two distinct domains.[115,116] Structure–function studies using mutants plus overlapping peptides of these superantigens provide further information on critical residues, and peptide regions, important for toxin binding to host cells. For instance, the T-cell receptor binding regions for SEB involve toxin residues 22–33 (α2 helix), 55–61 (β2–β3 loop), 87–92 (β4 strand and β4 – β5 loop), plus 210–214 (β5 strand and α5 helix).[110]

Another indicator that the SEs share similar structures is evidenced by cross-reactivity and neutralization with antibodies.[117–121] Several years ago, when there were less SEs known, these molecules were considered serologically distinct entities as determined by antisera and relatively insensitive immunodiffusion assays. However, subsequent studies employing a more sensitive technique (ELISA) with polyclonal and monoclonal antibodies clearly reveal that common epitopes do indeed exist between these toxins.

5.06.4.1 The First Step Toward Superantigen-Based Intoxication: Binding to Major Histocompatibility Complex Class II

The staphylococcal superantigens initially bind to conserved elements on major histocompatibility complex class II molecules with relatively high affinity ($K_d \sim 10^{-8}$ mol l^{-1}). These receptors are found in abundance, throughout the body, on antigen-presenting cells such as macrophages and monocytes. However, each toxin

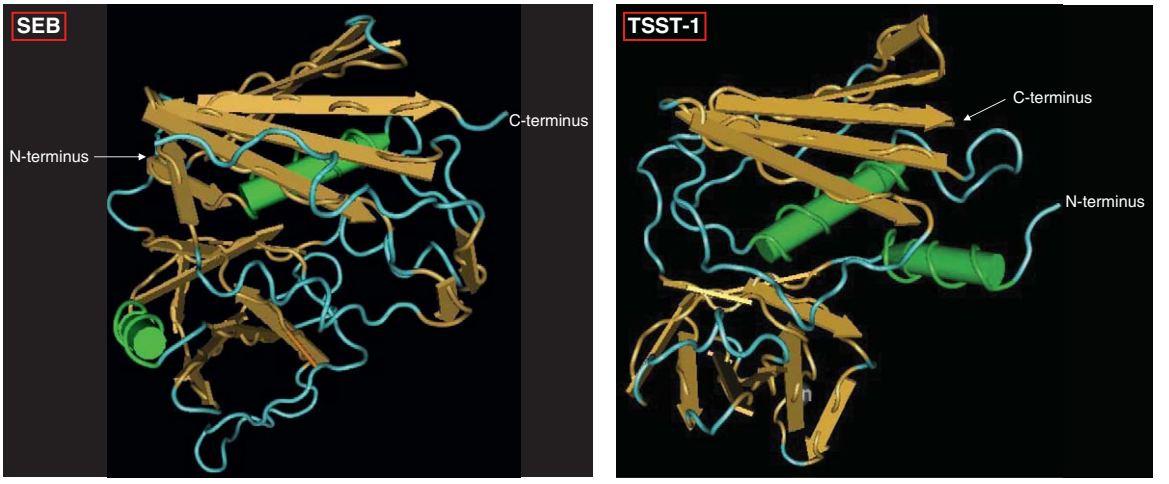


Figure 5 The crystal structures for SEB and TSST-1. These structures were constructed by using data provided by Entrez's 3-D database and software for molecular modeling.[113] Primary references for SEB and TSST-1 crystal structures are Papageorgiou *et al.*[111] and Prasad *et al.*,[114] respectively.

displays preferential binding to distinct alleles of major histocompatibility complex class II. There are to date three recognized types of superantigen interactions with class II molecules, which include (1) Zn^{2+}-facilitated binding to the β subunit of class II, (2) class II interactions through the α subunit away from the peptide binding groove, and (3) class II binding on the α subunit which encompasses the bound peptide. Overall, these observations suggest different sites and/or binding modes for the SEs and TSST-1.[122–125] As an example, binding studies show that SEA, SED, and SEE compete with SEB plus TSST-1 for human forms of major histocompatibility complex class II called *h*uman *l*eukocyte *a*ntigen (HLA)-DR; however, SEB and TSST-1 do not conversely inhibit binding of SEA, SED, or SEE.

Of the known staphylococcal superantigens, SEA has the highest affinity ($\sim$13 nmol l^{-1}) for HLA-DR and contains two major contact sites.[126–129] The higher affinity site on SEA is located within the C-terminus and binds to the HLA-DR β chain in a Zn^{2+}-dependent manner. Another unique feature of this Zn^{2+}-binding site is that His81 of the DR1β chain helps coordinate Zn^{2+} with three residues from SEA (His187, His226, and Asp227). The same Zn^{2+}-binding motif is also evident in SED as well as SEE, and coordination of Zn^{2+} in the SEA subfamily (SEA, SED, SEE) enables more efficient binding to major histocompatibility complex class II. The second binding site on SEA is of lower affinity, and similar to the binding site for SEB, located within the N-terminus (particularly involving Phe47), which interacts with Gln18 found on the invariant α chain of HLA-DR. Therefore, SEA cooperatively binds as a dimer to HLA-DR which then cross-links two HLA-DR molecules necessary for cytokine expression in monocytes.[128] Dimerization of various staphylococcal superantigens, such as SEB or TSST-1, and/or major histocompatibility complex class II molecules may play an important role in biological activity *in vivo*.[130,131]

Similar to the N-terminus of the SEA subfamily, a similar region in SEB and TSST-1 has also been identified as a binding site on major histocompatibility complex class II through studies with toxin mutants and monoclonal antibodies.[115,116] Analysis of crystal structures for SEB or TSST-1 complexed with HLA-DR1 further define differences in binding, although these toxins share the same interaction residues on the α chain.[130,131] SEB interacts exclusively with the α chain of HLA-DR1 and is not affected by associated peptide in the latter's antigen-presenting groove. This is unlike TSST-1, which interacts at 1 μmol l^{-1} affinity with the α and β chains of human HLA-DR1 or murine equivalent IA, as well as with the C-terminus of certain bound peptides. The peptide associated with major histocompatibility complex class II may not actually facilitate binding of TSST-1; however, it may effectively block TSST-1 interactions with the class II molecule.

5.06.4.2 The Second Step Toward Superantigen-Based Intoxication: T-Cell Receptor Engagement

In addition to major histocompatibility complex class II binding, and similar to other superantigens, the SEs and TSST-1 must also specifically interact with T-cell receptors to elicit a maximal effect upon a host. As mentioned before, X-ray crystallography clearly shows two distinct, conserved domains within the SEs as well as TSST-1 (see **Figure 5**). The groove formed between these domains binds the Vβ chain on the T-cell receptor[132–135] (see **Figure 6**). These toxins each interact with a distinct repertoire of Vβ-bearing T cells, thus revealing a unique biological 'fingerprint'. TSST-1 is very selective for human Vβ2; whereas, the less discriminating SEB molecule interacts with human Vβs 3, 12, 13.2, 14, 15, 17, and 20. SEB is similar to many of the other SEs that interact with multiple Vβ types. Mutational analysis of SEB identifies a conserved Asn23 (Asn25 of SEA) and amino acids 60–64 as essential for interactions with the murine Vβ chain.[116] Mutations within the binding domains of SEA for major histocompatibility complex class II differentially affect interactions with the Vβ region of a T-cell receptor.[134]

As just presented, each toxin uniquely interacts with specific Vβ regions on the T-cell receptor. Thus, for certain T-cell receptors, both the orientation and binding affinity of superantigen with the α chain of major histocompatibility complex class II affect interactions with T-cell receptor. The binding affinity between T-cell receptor and SEB is relatively weak, but this interaction is strengthened/stabilized by prior binding of toxin to major histocompatibility complex class II.[135] The same cooperative effect is also observed for SEA.[136] Therefore, the mitogenic potential of these toxins results from a cooperative process between host cells that intimately involves superantigen – class II complex binding to T-cell receptor with a higher affinity versus toxin alone. In essence, the host's own body tremendously amplifies the ill-effects of *S. aureus* superantigens. Ironically, this is precisely accomplished through collaborative efforts of the immune system which are specifically designed to protect its host. Perhaps this is an elegant evolutionary twist by *S. aureus* (and other superantigen-producing microbes!) that diverges from the 'brute force' approach employed by so many other bacterial toxins that punch holes in membranes or, once inside a targeted host cell, directly attack critical internal machinery (e.g. actin cytoskeleton, protein synthesis, etc.).

In the case of TSST-1, T-cell activation may be influenced by peptide in the antigen-binding groove of HLA-DR as per contact with the C-terminus of TSST-1.[131] Specifically, histidines 132, 135, and 140 of TSST-1 are important for T-cell receptor interactions plus stimulation of proinflammatory cytokine production. This is readily demonstrated by *in vitro* and *in vivo* studies with these toxin mutants, which also represent promising

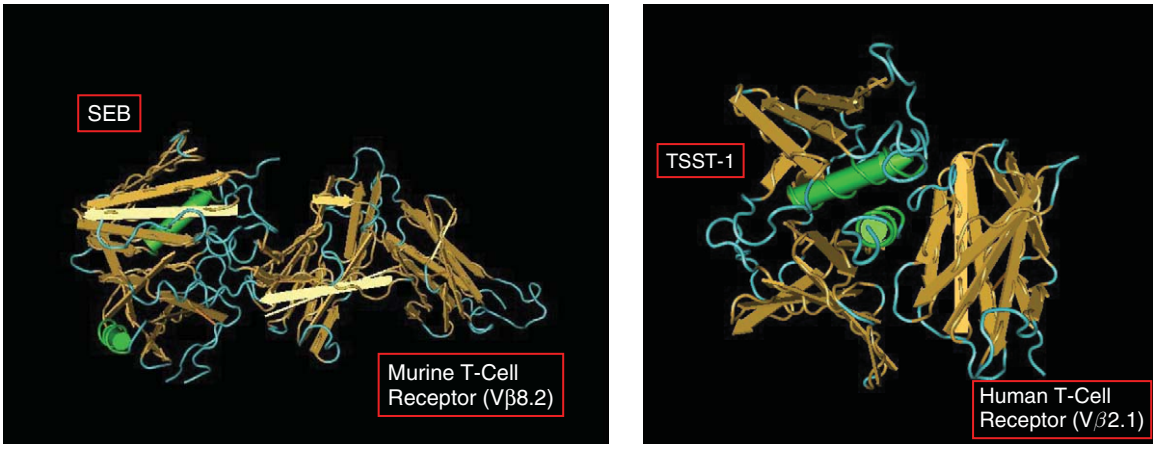


Figure 6 The binding of SEB and TSST-1 to the T-cell receptor. These structures were constructed by using data provided by Entrez's 3-D database and software for molecular modeling.[113] Primary references for these complexes are SEB plus T-cell receptor[133] and TSST-1 plus T-cell receptor.[138]

vaccine candidates.[98,104,137] The embedded surface existing between a TSST-1 – human T-cell receptor (Vβ2.1) complex encompasses approximately 1900 square angstroms, which is a much larger contact area than that for SEB or SEC3 with murine T-cell receptor (Vβ8.2).[138] The Glu61 and Lys62 residues on human Vβ2.1 play an important role in specific binding of TSST-1. Many other human T-cell receptors contain a Pro-61, which prevents binding of TSST-1. Unlike the less-discriminating SEs that generally stimulate various Vβ populations of T cells, TSST-1 is exquisitely specific for Vβ2-bearing cells.

Finally, a recent paper by Gunther *et al.*[139] shows that SEK uniquely binds to the T-cell receptor (human Vβ5.1) through a 15 amino-acid loop (α3-β8) not evident in other superantigens, except for those in the same toxin grouping (e.g., SEI). Surface plasmon resonance studies reveal a 6 μmol l^{-1} binding affinity of SEK for Vβ5.1, which fits the range for other SEs when used in this technique.

5.06.5 Signal Transduction and Cell Responses Induced by SEs and TSST-1

The coupling of superantigen–major histocompatibility complex class II to T-cell receptor swiftly results in cell-signaling cascades.[140] These staphylococcal toxins can increase levels of phosphatidyl inositol from quiescent T cells, such as other mitogens, as well as elicit intracellular Ca^{2+} movement that activates the protein kinase C (PKC) pathway important for interleukin-2 (IL-2) expression.[141] IL-2 is intimately linked to T-cell proliferation. In addition to the PKC pathway, the protein tyrosine kinase (PTK) pathway is also activated by superantigens, leading to elevated expression of various proinflammatory cytokines.[142] Staphylococcal superantigens also potently activate transcriptional factors NF-*k*B (nuclear factor kappa B) and AP-1 (activator protein-1), which subsequently elicit the synthesis of proinflammatory cytokines.[143,144]

Ironically, SE or TSST-1 concentrations that cause T-cell proliferation do not always correlate with receptor affinity.[122,126] For instance, SEE binds HLA-DR with 100-fold lower affinity relative to the very similarly structured SEA; however, SEE stimulates T-cell proliferation to equivalent levels as SEA.[126] The dose–response curves for cytokine and chemokine production *in vitro* by staphylococcal superantigen-stimulated cells are also very similar despite differences in affinity/specificity for major histocompatibility complex class II and T-cell receptor Vβ molecules.[145] Overall, these observations suggest that the biological effects of staphylococcal superantigens are induced at rather low, nonsaturating occupancy rates not readily classified by typical biokinetics.

Peripheral blood mononuclear cells are used extensively to study cellular activation by staphylococcal superantigens, and various therapeutic agents block different aspects of these pathways as described later in this text. These cells secrete various cytokines/chemokines such as IL-1, IL-2, IL-6, tumor necrosis factor alpha (TNFα), interferon gamma (IFNγ), macrophage inflammatory protein-1α (MIP-1α), MIP-1β, and monocyte chemoattractant protein-1 (MCP-1) in response to SEs or TSST-1. Although monocytes produce many chemokines plus proinflammatory cytokines such as IL-1, TNFα, and IL-6, added T cells greatly enhance levels of these mediators. Such results suggest that superantigen–class II complexed with T cells optimally contribute to cytokine and chemokine production.[145,146] There are contradictory reports of whether major histocompatibility complex class II-bearing cells, or T cells, respond to these toxins without the other cell type. Production of cytokine/chemokine mediators by human monocytic lines stimulated with superantigen (minus T cells) has been reported.[147] In contrast, others have found that IL-1 and TNFα induction by human monocytes following SEA exposure is strictly dependent upon T-cell interactions.[148] Evidently, SEA activation of human T cells is dependent upon accessory cells.[146] Purified human T cells express mRNA for several cytokines following superantigen exposure without class II-bearing cells; however, the extent of cytokine secretion and T-cell proliferation hinges upon the class II-bearing cells.[149,150] The signaling process through T-cell receptors involves clustering and lipid rafts on the T-cell surface. Therefore, it appears that the molecular triad of superantigen binding to major histocompatibility complex class II and T-cell receptor optimizes activational signaling. This altogether generates a 'cytokine storm' within the body, which can elicit a broad spectrum of effects that can be rather benign, or quite deadly. In sharp contrast to this classic paradigm for stimulation by a superantigen, presentation directly to T cells without major histocompatiblity complex class II molecules can evidently induce a tolerant, nonreactive state.[151]

There are other 'minor' cell types that respond to superantigens, which include B cells and synovial fibroblasts. Superantigen cross-linking of T-cell receptor with major histocompatibility complex class II on B cells reportedly triggers B-cell proliferation and antibody production.[152] This effect is very dose dependent, as high concentrations of superantigen actually inhibit immunoglobulin synthesis. TSST-1 also reportedly suppresses antibody secretion from B cells, thus hampering protective immunity against this toxin that is quite evident in patients suffering from recurring bouts of toxic shock.[70] Stimulation of synovial fibroblasts by superantigens induces chemokine synthesis, which triggers chemotactic responses and chronic inflammation during arthritis.[153]

5.06.5.1 Varied Effects of SEs and TSST-1 Upon the Body

In primates, SEs readily evoke vomiting and in rare cases toxic shock following ingestion of only microgram quantities. In contrast, TSST-1 (which is not an enterotoxin but erroneously described as 'SEF' in the early 1980s) does not elicit emesis after ingestion in a classic, nonhuman primate model. However, TSST-1 readily causes systemic toxic shock through *S. aureus* growth on mucosal surfaces such as the vaginal epithelium. Unlike many other bacterial enterotoxins described to date, specific cells and receptors in the intestinal tract have not been unequivocally linked to SE intoxication. After many decades of research, it is curiously odd that we do not have a better understanding of the receptor(s), trafficking, as well as overall intoxication process that leads to the enteric ill effects attributed to the SEs. Much effort indeed has elucidated the immunological (superantigenic) properties attributed to the SEs, which may (or may not!) play an important role in food poisoning by *S. aureus*.

Stimulation of mast cells by SEB and subsequent release of cysteinyl leukotrienes can cause emesis plus skin reactions in primates.[154] Specific T cells within Peyer's patches become nonresponsive following oral administration of SEB, thus suggesting an immune link.[155] Such results may also explain why ingestion of a SE in nonhuman primates yields transient resistance to an even higher toxin dose. This resistance is not evident when these animals are given another SE serotype orally, suggesting possible linkage to toxin-specific stimulation of unique Vβ-bearing T cells. In addition to toxin-specific resistance elicited by a single oral dose of SE, repeated intravenous exposure to SEA can virtually delete all Vβ-reactive T cells in mice.[156] Footpad injections of SEB in mice also cause a dose-related tolerance among SEB-reactive, Vβ8 T cells.[157] Another study in mice with previously administered SEA (intranasal), but not a recombinantly modified SEA molecule lacking superantigenicity, shows that animals become resistant to a subsequent lethal challenge with wild-type toxin.[158] Such protection is not due to toxin-specific antibody, but there is a significant increase in serum levels of IL-10. Previous *in vitro* and *in vivo* studies show that IL-10 affords protection against SE-induced effects.[159] This anti-inflammatory cytokine is produced by SEB-primed T cells, perhaps reflecting an attempt (although in severe cases of toxic shock, a feeble one!) by the host to counter the toxic consequences of other proinflammatory cytokines.

Several investigators have for many years attempted to locate a specific emetic domain within the SEs. At best, the results are limited and sometimes conflicting between different laboratories. Studies employing a human Caco-2 colon monolayer reveal transcytosis of SEA, SEB, and TSST-1. *In vivo* results from mice show that SEB enters the bloodstream more readily than SEA after oral administration.[160] These data reveal that SEs cross the gastric mucosa and subsequently circulate throughout the body. *In vitro*, these toxins are not directly cytotoxic as measured by cell leakage and inhibition of protein/nucleic acid synthesis with intestinal Henle 407 cells.[161] However, when a monolayer of another human line (T84 colonic cells) is incubated with SEB and peripheral blood mononuclear cells, an increasing ion flow suggests indirect toxin effects upon gut mucosa through the immune system.[162] As mentioned above, IL-10 affords protection against various bacterial superantigens and this cytokine dose-dependently inhibits toxin-induced permeability when added before, or concomitantly with, SEB.[159,163]

SEA (Leu-48-Gly) and SEB (Phe-44-Ser) mutants unable to bind major histocompatibility complex class II remain emetic but lack T-cell mitogenic effects.[164] A disulfide loop in SEs, which is absent in the nonenterotoxic TSST-1, may be responsible for the emetic activity of SEs but that too remains controversial.[165] Carboxymethylation of histidines on SEA[166] or SEB[167] generates superantigenic molecules devoid of enterotoxicity or skin reactivity. This chemically modified SEB also inhibits, perhaps in a competitive fashion, the

emetic/diarrheic properties of wild-type SEB in nonhuman primates when both are provided concomitantly. The lack of inherent enterotoxicity for carboxymethyl-modified SEA is not due to an altered conformation or increased susceptibility to degradation by gastric proteases.[168] Further analysis of each histidine in SEA-induced emesis and superantigenicity reveals that His61 is important for emesis, but not T-cell proliferation. These results demonstrate that emesis and superantigenicity are distinct properties.[168] However, another group has shown that antibodies which prevent SEA-induced emesis target a peptide region encompassing SEA residues 121–180.[169] This segment of SEA lacks the disulfide loop (Cys91 to Cys105) and histidines. It is possible that SEA becomes conformationally altered after antibody binding and/or the latter event sterically hinders toxin–receptor interactions.

Dual affinity of superantigens for major histocompatibility complex class II molecules and selected T-cell receptor Vβ enables these microbial toxins to perturb various effector cells of the immune system and elicit high levels of proinflammatory cytokines. Thus, the SEs and TSST-1 are pyrogenic in different mammalian species that include primates (human and nonhuman) as well as rabbits.[170] Elevated levels of proinflammatory cytokines, such as synergistically acting IL-1 and TNFα, correlate with increased temperatures.[171] Both of these cytokines are endogenous pyrogens that naturally induce fever through the hypothalamus. Additionally, the circulating levels of other T-cell-derived cytokines such as IFNγ, IL-2, and IL-6 also increase after toxin exposure. IFNγ augments immunological responses by increasing the major histocompatibility complex class II levels on antigen-presenting cells, epithelial cells, and endothelial cells. An increase in class II levels on cell surfaces clearly provides more opportunity for superantigen binding to, and yet even further stimulation of, T cells. Additionally, IFNγ also upregulates synthesis of TNFα and IL-1 receptors that respectively act in concert with their natural ligands. These cytokines enhance expression of adhesion molecules on endothelial cells, which subsequently promotes leukocyte adherence and recruitment. Shock induced by superantigens ultimately results from the cumulative biological effects of proinflammatory cytokines that adversely affect various organs.

5.06.6 Animal Models: Surprise, Mice are Not Men!

Mice are often used as models for obtaining a basic understanding of how bacterial toxins interact with mammals, which prominently features the immunological system during superantigen-mediated shock.[172–178] Although *Mus musculus* lack an emetic response, they are experimentally ideal regarding costs for *in vivo* screening of potential vaccines and therapeutics. However, mice (nontransgenic) are naturally less susceptible to SEs and TSST-1 versus humans. This is so because of an inherently low affinity of these exotoxins for murine major histocompatibility complex class II versus the human equivalent. Hundreds of micrograms for any SE or TSST-1 when given to a 'standard' inbred, or outbred, mouse causes few measureable effects that can be reliably used in different laboratories. However, this 'insensitive nature' can be overcome by administering multiple, strictly timed, high doses of toxin.[179]

It has also been discovered by various groups that potentiating agents such as D-galactosamine, actinomycin D, LPS, or even viruses greatly amplify the toxic effects of superantigens so that practical, lower amounts of toxin can be used *in vivo*. Throughout the past 15 years, many of our own SE and TSST-1 studies have employed an LPS-potentiated mouse model. Our reasoning is that various *in vitro* and *in vivo* systems from different laboratories have shown, throughout time, a natural synergy existing between these bacterial exotoxins and LPS.[180–182] As little as 2 μg of LPS causes shock-like effects in humans.[183] Bacterial superantigens such as the SEs and TSST-1 synergistically augment the effects of LPS many log-fold, as only picogram quantities of LPS with a superantigen can elicit life-threatening effects.[170] Upon considering the vast quantity of Gram-negative bacteria in normal intestinal flora, and a noticeable increase in Gram-negative vaginal flora among toxic shock patients, the odds of this synergy naturally occurring are really quite high.[170,184] There is a good correlation between increased serum levels of IL-1, IL-2, TNFα, and/or IFNγ with SEA-, SEB-, or TSST-1-induced shock.[172–182]

A relatively recent twist in studying SE toxicity has employed transgenic mice. As an example, animals expressing both human HLA-DQ6 and '*c*luster of *d*ifferentiation antigen' (CD)4-positive T cells readily succumb to normally sublethal amounts of SEB (versus nontransgenic controls) without any potentiating

agents. These transgenics, following SEB exposure, experience elevated serum levels of TNFα that correlate with lethal shock.[176] Other transgenic mice expressing human HLA-DR3 and CD4 also lethally respond to SEs, ultimately providing a less complicated model for future *in vivo* studies.[177] An earlier model involving overexpression of murine T-cell receptor (Vβ3) also reveals increased mortality linked to elevated TNF and IFNγ expression following infection by SEA-producing *S. aureus*.[185] From a scientific perspective, the fewer 'ingredients' (e.g., potentiating components) used for any model ultimately enables easier interpretation of results. This is especially so for therapeutic-based studies in which another component (e.g., drug or antibody) is added to the experimental mix.

An important, disease-pertinent variation upon a lethal endpoint for SE or TSST-1 intoxication involves temperature. Historically, rabbits have afforded an attractive model for SE- and TSST-1-induced shock with temperature or lethality endpoints.[178,186] As evidenced in humans with TSS,[180] rabbits given TSST-1 or SEB also experience elevated levels of circulating LPS. Such toxic concentrations of LPS are readily eliminated by polymyxin B, along with the clinical signs of TSS.[187] Increased levels of circulating LPS may be due to impaired liver clearance induced by these protein exotoxins.[188] More recently, SE- and TSST-1-triggered temperature fluctuations have been measured in LPS-potentiated mice implanted with a subcutaneous transponder or telemetry device.[189] Results reveal a very rapid temperature decrease (within 10 h) following intoxication. In contrast to humans and rabbits, mice do not experience elevated temperatures after SE or TSST-1 intoxication.[189] Temperature provides a quick, nonlethal parameter for investigating exotoxin – and/or endotoxin-induced effects indicative of rapid shock.

5.06.7 Strategic Countermeasures: An Eternal Battle between Man and Microbe

It is important to study how a pathogen, and in particular as per this review, how a bacterial toxin functions in the real world which on a daily basis involves morbidity and mortality. With such an understanding, then one becomes more empowered to intelligently prevent disease and death by conversely promoting health and life. This positive twist of fate begins in the laboratory with biochemistry and a true understanding of how proteins interact with receptors and/or substrates. To neutralize the ill effects of SEs and TSST-1, there are at least three important targets: (1) T-cell receptor – toxin – major histocompatibility complex class II interactions; (2) accessory, co-stimulatory, or adhesion molecules involved in activation and effector functions of T cells; as well as (3) cytokine release by activated T cells and macrophages. Successful inhibition of one or more targets has been reported both *in vitro* and *in vivo*, thus representing viable means of curbing the biological effects of these bacterial toxins. Such efforts though have largely remained within the realm of experimental laboratories. Other than IVIG (a relatively crude mix of human hyperimmune antibodies), there really are no defined/refined countermeasures used in the clinic for neutralizing the SEs and TSST-1. Perhaps human monoclonal antibodies, following rigorous characterization, may be useful reagents against the SEs and TSST-1 in the future?

Numerous efforts toward therapy of staphylococcal superantigens have been made by various groups. For example, steroids and IL-10 are possible agents for inhibiting proinflammatory cytokine production and T-cell proliferation following TSST-1 stimulation of human peripheral blood mononuclear cells. The discovery by Arad *et al.*[190] that a conserved dodecamer (Tyr-Asn-Lys-Lys-Lys-Ala-Thr-Val-Gln-Glu-Leu-Asp) derived from SEB, prevents SEA-, SEB-, or TSST-1-induced shock in mice injected intravenously 30 min after the toxin, is quite unique. Another group revealed that this same peptide prevents transcytosis of various SEs and TSST-1 across a monolayer of human colonic cells.[191] This segment of SEB is not associated with the classically defined major histocompatibility complex class II or T-cell receptor binding domains; however, this peptide may block costimulatory signals necessary for T-cell activation. Clearly, such a peptide represents a potential target for vaccines and therapeutics against the SEs and TSST-1. However, studies reported by another group question the peptide's efficacy against intoxication by staphylococcal superantigens.[192]

Several *in vitro* and *in vivo* models mentioned previously are used by various laboratories to study the best methods for preventing superantigen-induced shock. Therapeutic agents such as nitric oxide inhibitors can experimentally decrease SEA and SEB effects by subsequently inhibiting IL-1, IL-2, IL-6, TNF, and IFNγ

production.[193,194] Chimeric receptor mimics HLA-DR1 (alpha chain) plus T-cell receptor Vβ are effective *in vitro* against SEB- or TSST-1-induced T-cell proliferation.[195] More recent *in vitro* and *in vivo* studies involving T-cell receptor antagonists, the latter obtained by directed evolution plus recombinant expression of T-cell receptor domains, show efficacy and picomolar affinities toward SEB.[196] Antibodies directed against TNFα also prevent SEB-induced lethality,[172] whereas IL-10 blocks production of IL-1, TNFα, and IFNγ which ultimately reduces sequelae of superantigen-induced toxic shock.[159] As a further example of IL-10 efficacy, nasal application of SEA to mice elicits tolerance toward SEA, but not TSST-1.[158] This protection is linked to increased serum levels of IL-10 and not depletion/anergy of SEA-reactive T-cells or toxin-specific antibodies. Anti-inflammatory agents (e.g., indomethacin and dexamethasone) commonly used in clinical settings can lower SEA-induced fever in rabbits,[186] and simultaneously decrease serum concentrations of various proinflammatory cytokines. Pentoxifylline (a methylxanthine derivative that inhibits adhesion/activation of T cells) and pirfenidone (5-methyl-1-phenyl-2-(1*H*)-pyridone with antifibrotic properties) decrease SEB, as well as TSST-1, induced expression of cytokines *in vitro* and *in vivo*.[197,198] Release of proinflammatory cytokines due to SEB or TSST-1 is also suppressed *in vivo* by soluble β-glucans through an unknown mechanism.[199] Clearly, there are numerous strategies reported in the literature that experimentally inhibit superantigens *in vitro* and *in vivo*. The real (and obvious!) key is to now transition the most promising strategies from an experimental phase into clinical settings for eventual use in humans.

In addition to therapeutics, various groups have developed vaccines for the SEs and TSST-1. As mentioned before, pre-existing antibodies toward the SEs and TSST-1 play an important role in disease outcome.[95] Use of IVIG has also proven efficacious in humans after the onset of staphylococcal or streptococcal toxic shock,[200] but these reagents represent a crude mixture of immunoglobulins toward various antigens that vary from batch to batch. Given this wealth of information, it is logical that vaccination can prevent TSS due to SEs or TSST-1. Recombinantly attenuated versions of SEA, SEB, and TSST-1, which do not bind major histocompatibility complex class II and/or specific Vβ T-cell receptor molecules, represent experimentally proven vaccines that prevent toxic shock in different animal models.[99–105] These vaccines, when given either parenterally or mucosally, have proven efficacious against a toxin challenge or *S. aureus* infection.

5.06.8 Conclusions

To summarize, bacterial toxins are produced by diverse pathogens intimately involved in various human and animal diseases. It is our intended goal to share with readers the remarkably diverse nature of bacterial protein toxins. One particular interest of this article focuses upon the binary toxins, produced by *Clostridium* and *Bacillus* species, which represent potent molecules that kill host cells by damaging either their cytoskeleton or intracellular signaling network. These toxins act as a 'Trojan Horse' brought into a targeted cell through physiological pathways for protein trafficking. Much work, done by many groups, shows that these toxins possess distinct domains and share various levels of homology. This latter point emphasizes the evolution-based sharing of 'successful' virulence factors among bacteria. With additional studies, it would not be surprising to find similar binary toxins produced by other bacterial genera. As per experimental examples from different groups, the medicinal application of these toxins to traffic therapeutic proteins/peptides into cells represents a very promising twist. Domain alterations implemented among these binary toxins can switch a naturally poisonous protein into a medicinal chimera. For the binary toxins, there remains much work to be done in the future.

In a very different fashion from the clostridial/bacillus binary toxins, *S. aureus* superantigens such as the SEs and TSST-1 adversely affect host homeostasis without internalization by a targeted cell. Cleverly (from the microbe's perspective!), the immune system is 'baited' and 'tricked' by these protein toxins to respond in an overzealous, deleterious fashion. This is evidenced by the production of various immunomodulators (e.g., proinflammatory cytokines) beyond 'normally healthy' levels. Common amino acid homologies and biological activities of the multiple SEs and TSST-1 suggest the same 'endpoint' derived by divergent and/or convergent evolutionary paths. With time, more of these fascinating proteins will be discovered throughout nature and perhaps these new 'family members' will possess novel biological properties currently unrecognized in superantigens to date. Besides being produced by *S. aureus*, superantigens are now evident in

many other bacterial and viral pathogens (**Table 3**). It is commonly thought that these proteins afford a distinct advantage for the microorganism that includes delayed clearance from a host.[201] Evolutionary retention of superantigens by vastly different microbes clearly suggests a 'biological success story' leading to enhanced survival of a species.[202]

To remain a step ahead of life-threatening pathogens, it is essential to understand the chemical nature, structure, and diverse modes of action for bacterial toxins which represent some of the most toxic substances known today. Based upon this information, new vaccines can be created against specific domains of a toxin or perhaps cellular uptake of the toxins can be blocked by specific pharmacological inhibitors. Therefore, it will be of lasting impact to transfer our increasing knowledge of bacterial toxins from the laboratory bench into the clinic.

Abbreviations

ADP	adenosine diphosphate
C2	binary C2 toxin from *Clostridium botulinum*
C2I	enzyme component (ADP-ribosyltransferase) of the binary C2 toxin
C2II	binding/translocation component of the binary C2 toxin
C2IIa	activated binding/translocation component of the binary C2 toxin
C3	C3-ADP-ribosyltransferase
CD	cluster of differentiation
CDT	*Clostridium difficile* ADP-ribosyltransferase (binary toxin)
CST	*Clostridium spiroforme* ADP-ribosyltransferase (binary toxin)
EE	early endosomes
ER	endoplasmic reticulum
F-actin	actin filaments (polymeric actin)
G-actin	globular actin (monomeric actin)
HLA	human leukocyte antigen
Ia	enzyme component (ADP-ribosyltransferase) of the binary iota toxin
Ib	binding/translocation component of the binary iota toxin
IFN	interferon
IL	interleukin
LD50	lethal dose 50
LE	late endosomes
LPS	lipopolysaccharide
NAD	nicotinamide adenosine dinucleotide
PFO	perfringolysin O
PLC, -D, - A	phospholipase C, -D, -A
SE	staphylococcal enterotoxin
SLO	streptolysin O
Spv	salmonella plasmid of virulence
SpvB	ADP-ribosyltransferase SpvB
TGN	trans Golgi network
TNF	tumor necrosis factor alpha (TNFα)
TSS	toxic shock syndrome
TSST-1	toxic shock syndrome toxin-1

References

1. E. D. Williamson; R. W. Titball, *Vaccine* **1993**, *11* (12), 1253–1258.
2. M. M. Awad; A. E. Bryant; D. L. Stevens; J. I. Rood, *Mol. Microbiol.* **1995**, *15* (2), 191–202.
3. J. A. Vazquez-Boland; C. Kocks; S. Dramsi; H. Ohayon; C. Geoffroy; J. Mengaud; P. Cossart, *Infect. Immun.* **1992**, *60* (1), 219–230.

4. M. Fivaz; L. Abrami; Y. Tsitrin; F. G. Van der Goat, *Curr. Top. Microbiol. Immunol.* **2001**, *257*, 35–52.
5. R. K. Tweten, *Infect. Immun.* **2005**, *73* (10), 6199–6209.
6. A. Valeva; A. Weisser; B. Walker; M. Kehoe; H. Bayley; S. Bhakdi; M. Palmer, *EMBO J.* **1996**, *15* (8), 1857–1864.
7. S. Bhakdi; J. Tranum-Jenssen, *Microbiol. Rev.* **1991**, *55* (4), 733–751.
8. C. Montecucco, *Curr. Opin. Cell. Biol.* **1998**, *10*, 530–536.
9. H. Barth; K. Aktories; M. R. Popoff; B. G. Stiles, *Microbiol. Mol. Biol. Rev.* **2004**, *68* (3), 373–402.
10. K. Sandvig; B. Spilsberg; S. U. Lauvrak; M. L. Torgersen; T. G. Iversen; B. Van Deurs, *Int. J. Med. Microbiol.* **2004**, *293* (7–8), 483–490.
11. F. G. van der Goot; J. Gruenberg, *Trends Cell Biol.* **2006**, *16* (10), 514–521.
12. S. Olsnes; J. Wesche; P. O. Falnes, Uptake of Protein Toxins Acting Inside Cells. In *Bacterial Protein Toxins*; K. Aktories, I. Just, Eds.; Springer: Berlin, 2000; Vol. 145, Chapter 1, pp 1–19.
13. A. Wegner; K. Aktories, *J. Biol. Chem.* **1988**, *263*, 13739–13742.
14. S. M. Margarit; W. Davidson; L. Frego; C. E. Stebbins, *Structure* **2006**, *14* (8), 1219–1229.
15. M. L. Lesnick; N. E. Reiner; J. Fierer; D. G. Guiney, *Mol. Microbiol.* **2001**, *39*, 1464–1470.
16. K. Aktories; M. Bärmann; I. Ohishi; S. Tsuyama; K. H. Jakobs; E. Habermann, *Nature* **1986**, *322*, 390–392.
17. B. G. Stiles; T. D. Wilkins, *Infect. Immun.* **1986**, *54*, 683–688.
18. A. Chaudhary; W. G. King; M. D. Mattaliano; J. A. Frost; B. Diaz; D. K. Morrison; M. H. Cobb; M. S. Marshall; J. S. Brugge, *Curr. Biol.* **2000**, *10*, 551–554.
19. F. C. Fang; M. Krause; C. Roudier; J. Fierer; D. G. Guiney, *J. Bacteriol.* **1991**, *173* (21), 6783–6789.
20. S. Han; J. A. Craig; C. D. Putnam; N. B. Carozzi; J. A. Tainer, *Nat. Struct. Biol.* **1999**, *6*, 932–936.
21. D. G. Guiney; M. Lesnick, *Clin. Immunol.* **2005**, *114*, 248–255.
22. D. Tezcan-Merdol; T. Nyman; U. Lindberg; F. Haag; F. Koch-Nolte; M. Rhen, *Mol. Microbiol.* **2001**, *39*, 606–619.
23. H. Hochmann; S. Pust; G. von Figura; K. Aktories; H. Barth, *Biochemistry* **2006**, *45* (4), 1271–1277.
24. S. Pust; H. Hochmann; E. Kaiser; G. von Figura; K. Heine; K. Aktories; H. Barth, *J. Biol. Chem.* **2007**, *282* (14), 10272–10282.
25. H. Barth, *Rev. Physiol. Biochem. Pharmacol.* **2004**, *152*, 165–182.
26. M. Iwasaki; I. Ohishi; G. Sakaguchi, *Infect. Immun.* **1980**, *29* (2), 390–394.
27. I. Ohishi; M. Iwasaki; G. Sakaguchi, *Infect. Immun.* **1980**, *30*, 668–673.
28. L. L. Simpson, *J. Pharmacol. Exp. Ther.* **1982**, *223*, 695–701.
29. I. Ohishi; Y. Odagiri, *Infect. Immun.* **1984**, *43*, 54–58.
30. I. Ohishi, *Infect. Immun.* **1987**, *55*, 1461–1465.
31. H. Barth; D. Blöcker; J. Behlke; W. Bergsma-Schutter; A. Brisson; R. Benz; K. Aktories, *J. Biol. Chem.* **2000**, *275*, 18704–18711.
32. B. G. Stiles; D. Blöcker; M. L. Hale; A. M. Guetthoff; H. Barth, *Toxicon* **2002**, *40*, 1135–1140.
33. M. Eckhardt; H. Barth; D. Blöcker; K. Aktories, *J. Biol. Chem.* **2000**, *275*, 2328–2334.
34. E. Kaiser; G. Haug; M. Hliscs; K. Aktories; H. Barth, *Biochemistry* **2006**, *45* (44), 13361–13368.
35. D. Blöcker; H. Barth; E. Maier; R. Benz; J. T. Barbieri; K. Aktories, *Infect. Immun.* **2000**, *68*, 4566–4573.
36. D. Blöcker; K. Pohlmann; G. Haug; C. Bachmeyer; R. Benz; K. Aktories; H. Barth, *J. Biol. Chem.* **2003**, *278* (39), 37360–37367.
37. G. Haug; C. Wilde; J. Leemhuis; D. K. Meyer; K. Aktories; H. Barth, *Biochemistry* **2003**, *42*, 15284–15291.
38. C. Schleberger; H. Hochmann; H. Barth; K. Aktories; G. E. Schulz, *J. Mol. Biol.* **2006**, *364* (4), 705–715.
39. G. Haug; J. Leemhuis; D. Tiemann; D. K. Meyer; K. Aktories; H. Barth, *J. Biol. Chem.* **2003**, *278* (34), 32266–32274.
40. J. Vandekerckhove; B. Schering; M. Bärmann; K. Aktories, *J. Biol. Chem.* **1988**, *263*, 696–700.
41. H. Barth; J. C. Preiss; F. Hofmann; K. Aktories, *J. Biol.Chem.* **1998**, *273*, 29506–29511.
42. H. Barth; F. Hofmann; C. Olenik; I. Just; K. Aktories, *Infect. Immun.* **1998**, *66* (4), 1364–1369.
43. H. Barth; R. Roebling; M. Fritz; K. Aktories, *J. Biol. Chem.* **2002**, *277* (7), 5074–5081.
44. H. Barth; D. Blöcker; K. Aktories, *Naunyn Schmiedebergs Arch. Pharmacol*. **2002**, *366*, 501–512.
45. B. G. Stiles; T. D. Wilkins, *Toxicon* **1986**, *24*, 767–773.
46. J. Vandekerckhove; B. Schering; M. Bärmann; K. Aktories, *FEBS Lett.* **1987**, *225*, 48–52.
47. S. Perelle; M. Gibert; P. Boquet; M. R. Popoff, *Infect. Immun.* **1993**, *61*, 5147–5156.
48. I. Gulke; G. Pfeifer; J. Liese; M. Fritz; F. Hofmann; K. Aktories; H. Barth, *Infect. Immun.* **2001**, *69* (10), 6004–6011.
49. S. Perelle; S. Scalzo; S. Kochi; M. Mock; M. R. Popoff, *FEMS Microbiol. Lett.* **1997**, *146*, 117–121.
50. M. R. Popoff, *Infect. Immun.* **1987**, *55*, 35–43.
51. M. Braun; C. Herholz; R. Straub; B. Choisat; J. Frey; J. Nicolet; P. Kuhnert, *FEMS Microbiol. Lett.* **2000**, *184*, 29–33.
52. S. P. Borriello; R. J. Carman, *J. Clin. Microbiol.* **1983**, *17* (3), 414–418.
53. B. Stoddart; M. H. Wilcox, *Curr. Opin. Infect. Dis.* **2002**, *15* (5), 513–518.
54. L. L. Simpson; B. G. Stiles; H. H. Zepeda; T. D. Wilkins, *Infect. Immun.* **1987**, *55* (1), 118–122.
55. C. P. Kelly; C. Pothoulakis; J. T. LaMont, *N. Engl. J. Med.* **1994**, *330*, 257–262.
56. J. G. Bartlett, *N. Engl. J. Med.* **2002**, *346*, 334–339.
57. S. P. Borriello, *J. Antimicrob. Chemother.* **1998**, *41*, 13–19.
58. I. Just; J. Selzer; M. Wilm; C. Von Eichel-Streiber; M. Mann; K. Aktories, *Nature* **1995**, *375*, 500–503.
59. I. Just; M. Wilm; J. Selzer; G. Rex; C. Von Eichel-Streiber; M. Mann; K. Aktories, *J. Biol. Chem.* **1995**, *270*, 13932–13936.
60. C. P. Kelly; J. T. LaMont, *Annu. Rev. Med.* **1998**, *49*, 375–390.
61. S. Stubbs; M. Rupnik; M. Gibert; J. Brazier; B. Duerden; M. Popoff, *FEMS Microbiol. Lett.* **2000**, *186*, 307–312.
62. B. Geric; S. Johnson; D. N. Gerding; M. Grabnar; M. Rupnik, *J. Clin. Microbiol.* **2003**, *41*, 5227–5232.
63. D. B. Blossom; L. C. McDonald, *Clin. Infect. Dis.* **2007**, *45* (2), 222–227.
64. V. G. Loo; L. Poirier; M. A. Miller; M. Oughton; M. D. Libman; S. Michaud; A. M. Bourgault; T. Nguyen; C. Frenette; M. Kelly; A. Vibien; P. Brassard; S. Fenn; K. Dewar; T. J. Hudson; R. Horn; P. Rene; Y. Monczak; A. Dascal, *N. Engl. J. Med.* **2005**, *353* (23), 2442–2449.
65. L. C. McDonald; G. E. Killgore; A. Thompson; R. C. Owens, Jr.; S. V. Kazakova; S. P. Sambol; S. Johnson; D. N. Gerding, *N. Engl. J. Med.* **2005**, *353* (23), 2433–2441.

66. F. Barbut; B. Gariazzo; L. Bonne; V. Lalande; B. Burghoffer; R. Luiuz; J. C. Petit, *Infect. Control Hosp. Epidemiol.* **2007**, *28* (2), 131–139.
67. B. L. Kotzin; D. Y. M. Leung; J. Kappler; P. A. Marrack, *Adv. Immunol.* **1993**, *54*, 99–166.
68. S. R. Monday; G. A. Bohach, Properties of *Staphylococcus aureus* Enterotoxins and Toxic Shock Syndrome Toxin-1. In *The Comprehensive Sourcebook of Bacterial Protein Toxins*, 2nd ed.; J. E. Alouf, J. H. Freer, Eds.; Academic Press: London, 1999; pp 589–610.
69. B. Fleischer; R. Gerardy-Schahn; B. Metzroth; S. Carrel; D. Gerlach; W. Kohler, *J. Immunol.* **1991**, *146*, 11–17.
70. N. J. Poindexter; P. M. Schlievert, *J. Infect. Dis.* **1986**, *153*, 772–779.
71. T. J. Foster, *J. Clin. Invest.* **2004**, *114*, 1693–1696.
72. B. Kreiswirth; J. Kornblum; R. Arbeit; W. Eisner; J. Maslow; A. McGeer; D. Low; R. P. Novick, *Science* **1993**, *259*, 227–230.
73. G. Durand; M. Bes; H. Meugnier; M. C. Enright; F. Forey; N. Liassine; A. Wenger; K. Kikuchi; G. Lina; F. Vandenesch; J. Etienne, *J. Clin. Microbiol.* **2006**, *44*, 847–853.
74. CDC, *MMWR Morb. Mortal. Wkly. Rep.* **1997**, *46* (27), 624–626.
75. F. C. Tenover; L. C. McDonald, *Curr. Opin. Infect. Dis.* **2005**, *18*, 300–305.
76. A. Tristan; T. Ferry; G. Durand; O. Dauwalder; M. Bes; G. Lina; F. Vandenesch; J. Etienne, *J. Hosp. Infect.* **2007**, *65*, 105–109.
77. D. L. Stevens; Y. Ma; D. B. Salmi; E. McIndoo; R. J. Wallace; A. E. Bryant, *J. Infect. Dis.* **2007**, *195*, 202–211.
78. T. Kim; P. I. Oh; A. E. Simor, *Infect. Cont. Hosp. Epidemiol.* **2001**, *22*, 99–104.
79. M. Vriens; H. Blok; A. Fluit; A. Troelstra; C. van der Werken; J. Verhoef, *Eur. J. Clin. Microbiol. Infect. Dis.* **2002**, *21*, 782–786.
80. B. Capitano; O. A. Leshem; C. H. Nightingale; D. P. Nicolau, *J. Am. Geriatr. Soc.* **2003**, *51*, 10–16.
81. A. A. Wieneke; D. Roberts; R. J. Gilbert, *Epidemiol. Infect.* **1993**, *110*, 519–531.
82. M. A. Barber, *Philippine J. Sci.* **1914**, *9*, 515–519.
83. Y. Le Loir; F. Baron; M. Gautier, *Genet. Mol. Res.* **2003**, *2*, 63–76.
84. V. Srinivasan; A. A. Sawant; B. E. Gillespie; S. J. Headrick; L. Ceasaris; S. P. Oliver, *Foodborne Pathog. Dis.* **2006**, *3*, 274–283.
85. P. Marrack; J. Kappler, *Science* **1990**, *248*, 705–711.
86. M. Jett; W. Brinkley; R. Neill; P. Gemski; R. Hunt, *Infect. Immun.* **1990**, *58*, 3494–3499.
87. P. H. Scheuber; C. Denzlinger; D. Wilker; G. Beck; D. Keppler; D. K. Hammer, *Eur. J. Clin. Invest.* **1987**, *17*, 455–459.
88. J. Todd; M. Fishaut; F. Kapral; T. Welch, *Lancet* **1978**, *2*, 1116–1118.
89. K. N. Shands; G. P. Schmid; B. B. Dan; D. Blum; R. J. Guidotti; N. T. Hargrett; R. L. Anderson; C. V. Bromme; J. D. Band; D. W. Fraser, *N. Engl. J. Med.* **1980**, *303*, 1436–1442.
90. M. S. Bergdoll; B. A. Crass; R. F. Reiser; R. N. Robbins; P. J. Davis, *Lancet* **1981**, *1*, 1017–1021.
91. R. F. Reiser; R. N. Robbins; G. P. Khoe; M. S. Bergdoll, *Biochemistry* **1983**, *22*, 3907–3912.
92. J. Parsonnet, *Rev. Infect. Dis.* **1989**, *11*, S263–S269.
93. J. D. Freedman; D. J. Beer, *Adv. Intern. Med.* **1991**, *36*, 363–397.
94. M.-M. Andrews; E. M. Parent; M. Barry; J. Parsonnet, *Clin. Infect. Dis.* **2001**, *32*, 1470–1479.
95. P. F. Bonventre; C. Linnemann; L. S. Weckbach; J. L. Staneck; C. R. Buncher; E. Bigdorth; H. Ritz; D. Archer; B. Smith, *J. Infect. Dis.* **1984**, *150*, 662–666.
96. R. Kansal; C. Davis; M. Hansmann; J. Seymour; J. Parsonnet; P. Modern; S. Gilbert; M. Kotb, *J. Clin. Immunol.* **2007**, *27*, 327–338.
97. N. Takahashi; M. Hattori; K. Miwa; H. Nishida, *Am. J. Perinatol.* **2006**, *23*, 413–419.
98. B. G. Stiles; T. Krakauer; P. F. Bonventre, *Infect. Immun.* **1995**, *63*, 1229–1234.
99. B. G. Stiles; A. R. Garza; R. G. Ulrich; J. W. Boles, *Infect. Immun.* **2001**, *69*, 2031–2036.
100. I. M. Nilsson; M. Verdrengh; R. G. Ulrich; S. Bavari; A. Tarkowski, *J. Infect. Dis.* **1999**, *180*, 1370–1373.
101. S. Bavari; B. Dyas; R. G. Ulrich, *J. Infect. Dis.* **1996**, *174*, 338–345.
102. J. K. McCormick; T. J. Tripp; S. B. Olmsted; Y. V. Matsuka; P. J. Gahr; D. H. Ohlendorf; P. M. Schlievert, *J. Immunol.* **2000**, *165*, 2306–2312.
103. M. Roggiani; J. A. Stoehr; S. B. Olmsted; Y. V. Matsuka; S. Pillai; D. H. Ohlendorf; P. M. Schlievert, *Infect. Immun.* **2000**, *68*, 5011–5017.
104. J. Gampfer; V. Thon; H. Gulle; H. M. Wolf; M. M. Eibl, *Vaccine* **2002**, *20*, 1354–1364.
105. R. G. Ulrich; M. A. Olson; S. Bavari, *Vaccine* **1998**, *16*, 1857–1864.
106. T. Uchiyama; K. Imanishi; T. Miyoshi-Akiyama, Staphylococcal Superantigens and the Diseases they Cause. In *The Comprehensive Sourcebook of Bacterial Protein Toxins*, 3rd ed.; J. E. Alouf, M. R. Popoff, Eds.; Academic Press: Paris, 2006; pp 830–843.
107. M. J. Betley; D. W. Borst; L. B. Regassa, *Chem. Immunol.* **1992**, *55*, 1–35.
108. M. Diedershagen; S. Overbeck; S. Arit; B. Plumakers; M. Lintges; L. Rink, *FEMS Immunol. Med. Microbiol.* **2007**, *49*, 266–271.
109. B. R. Singh; F. Fen-Ni; D. N. Ledoux, *Struct. Biol.* **1994**, *1*, 358–360.
110. S. Swaminathan; W. Furey; J. Pletcher; M. Sax, *Nature* **1992**, *359*, 801–806.
111. A. C. Papageorgiou; H. S. Tranter; K. R. Acharya, *J. Mol. Biol.* **1998**, *277*, 61–79.
112. A. C. Papageorgiou; R. D. Brehm; D. D. Leonidas; H. S. Tranter; K. R. Acharya, *J. Mol. Biol.* **1996**, *260*, 553–569.
113. J. Chen; J. B. Anderson; C. DeWeese-Scott; N. D. Fedorova; L. Y. Geer; S. He; D. I. Hurwitz; J. D. Jackson; A. R. Jacobs; C. J. Lanczycki; C. A. Liebert; C. Liu; T. Madej; A. Marchler-Bauer; G. H. Marchler; R. Mazumder; A. N. Nikolskaya; B. S. Rao; A. R. Panchenko; B. A. Shoemaker; V. Simonyan; J. S. Song; P. A. Thiessen; S. Vasudevan; Y. Wang; R. A. Yamashita; J. J. Yin; S. H. Bryant, *Nucleic Acids Res.* **2003**, *31*, 474–477.
114. G. S. Prasad; R. Radharkrishnan; D. T. Mitchell; C. A. Earhart; M. M. Dinges; W. J. Cook; P. M. Schlievert; D. H. Ohlendorf, *Protein Sci.* **1997**, *6*, 1220–1227.
115. J. M. Hurley; R. Shimonkevitz; A. Hanagan; K. Enney; E. Boaen; S. Malmstro; B. L. Kotzin; M. Matsumura, *J. Exp. Med.* **1995**, *181*, 2229–2235.
116. J. W. Kappler; A. Herman; J. Clements; P. Marrack, *J. Exp. Med.* **1992**, *175*, 387–396.
117. W. W. Kum; A. W. Chow, *J. Infect. Dis.* **2001**, *183*, 1739–1748.
118. S. Bavari; R. G. Ulrich; R. D. LeClaire, *J. Infect. Dis.* **1999**, *180*, 1365–1369.

119. N. E. Thompson; M. J. Ketterhagen; M. S. Bergdoll, *Infect. Immun.* **1984**, *45*, 281–285.
120. L. Spero; B. A. Morlock; J. F. Metzger, *J. Immunol.* **1984**, *120*, 86–89.
121. G. A. Bohach; C. J. Hovde; J. P. Handley; P. M. Schlievert, *Infect. Immun.* **1988**, *56*, 400–404.
122. J. A. Mollick; M. Chintagumpala; R. G. Cook; R. R. Rich, *J. Immunol.* **1991**, *146*, 463–468.
123. J. Yagi; J. S. Rath; C. A. Janeway, *J. Immunol.* **1991**, *147*, 1398–1405.
124. T. Herrmann; R. S. Acolla; H. R. MacDonald, *Eur. J. Immunol.* **1989**, *19*, 2171–2174.
125. A. Herman; G. Croteau; R. P. Sekaly; J. Kappler; P. Marrack, *J. Exp. Med.* **1990**, *172*, 709–717.
126. M. M. Chintagumpala; J. A. Mollick; R. R. Rich, *J. Immunol.* **1991**, *147*, 3876–3881.
127. K. R. Hudson; R. E. Tiedemann; R. G. Urban; S. C. Lowe; J. L. Strominger; J. Fraser, *J. Exp. Med.* **1995**, *182*, 711–720.
128. R. E. Tiedemann; J. D. Fraser, *J. Immunol.* **1996**, *157*, 3958–3966.
129. R. G. Ulrich; S. Bavari; M. A. Olson, *Nat. Struct. Biol.* **1995**, *2*, 554–560.
130. T. S. Jardetzky; J. H. Brown; J. C. Gorga; L. J. Stern; R. G. Urban; Y. I. Chi; C. Stauffacher; J. L. Strominger; D. C. Wiley, *Nature* **1994**, *368*, 711–718.
131. J. Kim; R. G. Urban; J. L. Strominger; D. C. Wiley, *Science* **1994**, *266*, 1870–1874.
132. B. A. Fields; E. L. Malchiodi; H. Li; X. Ysern; C. V. Stauffcher; P. M. Schlievert, *Nature* **1996**, *384*, 188–192.
133. H. Li; A. Llera; D. Tsuchiya; L. Leder; X. Ysern; P. M. Schlievert, *Immunity* **1998**, *9*, 807–816.
134. D. W. Newton; M. Dohlsten; C. Olsson; S. Segren; K. E. Lundin; P. Land; T. Kalland; M. Kotb, *J. Immunol.* **1996**, *157*, 3988–3994.
135. A. Seth; I. J. Stern; T. H. Ottenhoff; I. Engel; M. J. Owen; J. R. Lamb; R. D. Klausner; D. C. Wiley, *Nature* **1994**, *369*, 324–327.
136. S. Redpath; S. Alam; C. Lin; A. M. O'Rourke; N. R. Gascoigne, *J. Immunol.* **1994**, *163*, 6–10.
137. P. F. Bonventre; H. Heeg; C. K. Edwards; C. M. Cullen, *Infect. Immun.* **1995**, *63*, 509–515.
138. B. Moza; A. K. Varma; R. A. Buonpane; P. Zhu; C. A. Herfst; M. J. Nicholson; A.-K. Wilbuer; N. P. Seth; K. W. Wucherpfenning; J. K. McCormick; D. M. Kranz; E. J. Sundberg, *EMBO J.* **2007**, *26*, 1187–1197.
139. S. Gunther; A. K. Varma; B. Moza; K. J. Kasper; A. W. Wyatt; P. Zhu; A. K. Rahman; Y. Li; R. A. Mariuzza; J. K. McCormick; E. J. Sundberg, *J. Mol. Biol.* **2007**, *371*, 210–221.
140. T. Chatila; R. S. Geha, *Immunol. Rev.* **1993**, *131*, 43–59.
141. T. Chatila; N. Wood; J. Parsonnet; R. S. Geha, *J. Immunol.* **1988**, *140*, 1250–1255.
142. P. R. Scholl; N. Trede; T. A. Chatila; R. S. Geha, *J. Immunol.* **1992**, *148*, 2237–2241.
143. N. S. Trede; A. V. Tsytsykovo; A. E. Goldfeld; R. S. Geha, *J. Immunol.* **1995**, *155*, 902–908.
144. A. Sundstedt; M. Sigvardsson; T. Leanderson; G. Hedlund; T. Kalland; M. Dohlsten, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 979–984.
145. T. Krakauer, *J. Leuk. Biol.* **1999**, *66*, 158–164.
146. R. Carlsson; H. Fischer; H. O. Sjogren, *J. Immunol.* **1988**, *140*, 2484–2488.
147. N. S. Trede; R. S. Geha; T. Chatila, *J. Immunol.* **1991**, *146*, 2310–2315.
148. H. Fischer; M. Dohlsten; U. Andersson; G. Hedlund; P. Ericsson; J. Hansson; H. O. Sjogren, *J. Immunol.* **1990**, *144*, 4663–4669.
149. A. Lagoo; S. Lagoo-Deenadayalan; H. Lorenz; J. Byrne; W. Barber; K. Hardy, *J. Immunol.* **1994**, *152*, 1641–1652.
150. P. A. Lando; C. Olsson; T. Kalland; D. Newton; M. Kotb; M. Dohlsten, *J. Immunol.* **1996**, *157*, 2857–2863.
151. C. Hewitt; J. Lamb; J. Hayball; M. Hill; M. Owen; R. O'Hehir, *J. Exp. Med.* **1992**, *175*, 1493–1499.
152. W. Stohl; J. E. Elliott; P. S. Linsley, *J. Immunol.* **1994**, *153*, 117–127.
153. P. H. Wooley; B. Cingel, *Ann. Rheum. Dis.* **1995**, *54*, 298–304.
154. P. H. Scheuber; C. Denzlinger; D. Wilker; G. Beck; D. Keppler; D. K. Hammer, *Int. Arch. Allergy Appl. Immunol.* **1987**, *82*, 289–291.
155. G. M. Spiekermann; C. Nagler-Anderson, *J. Immunol.* **1998**, *161*, 5825–5831.
156. J. E. McCormack; J. E. Callahan; J. Kappler; P. C. Marrack, *J. Immunol.* **1993**, *150*, 3785–3792.
157. T. Miethke; C. Wahl; H. Gaus; K. Heeg; H. Wagner, *Eur. J. Immunol.* **1994**, *24*, 1893–1902.
158. L. V. Collins; K. Eriksson; R. G. Ulrich; A. Tarkowski, *Infect. Immun.* **2002**, *70*, 2282–2287.
159. A. G. Bean; R. A. Freiberg; S. Andrade; S. Menon; A. Zlotnik, *Infect. Immun.* **1993**, *61*, 4937–4939.
160. A. R. Hamad; P. Marrack; J. W. Kappler, *J. Exp. Med.* **1997**, *185*, 1447–1454.
161. S. Buxser; P. F. Bonventre, *Infect. Immun.* **1981**, *31*, 929–934.
162. J. Lu; D. J. Philpott; P. R. Saunders; M. H. Perdue; P.-C. Yang; D. M. McKay, *J. Pharm. Exp. Ther.* **1998**, *287*, 128–136.
163. G. Hasko; L. Virag; G. Egnaczyk; A. L. Salzman; C. Szabo, *Eur. J. Immunol.* **1998**, *28*, 1417–1425.
164. T. O. Harris; D. Grossman; J. W. Kappler; P. Marrack; R. R. Rich; M. J. Betley, *Infect. Immun.* **1993**, *61*, 3175–3183.
165. C. J. Hovde; J. C. Mar; M. L. Hoffmann; S. P. Hackett; Y. I. Chi; K. K. Crum; D. L. Stevens; C. V. Stauffacher; G. A. Bohach, *Mol. Microbiol.* **1994**, *13*, 897–909.
166. G. N. Stelma; M. S. Bergdoll, *Biochem. Biophys. Res. Commun.* **1982**, *105*, 121–126.
167. G. Alber; D. K. Hammer; B. Fleischer, *J. Immunol.* **1990**, *144*, 4501–4506.
168. M. Hoffman; M. Tremaine; J. Mansfield; M. Betley, *Infect. Immun.* **1996**, *64*, 885–890.
169. D.-L. Hu; K. Omoe; M. Saleh; K. Ono; S. Sugii; A. Nakane; K. Shinagawa, *J. Vet. Med. Sci.* **2001**, *63*, 237–241.
170. P. M. Schlievert, *J. Infect. Dis.* **1993**, *167*, 997–1002.
171. S. Okusawa; J. A. Gelfand; T. Ikejima; R. J. Connolly; C. A. Dinarello, *J. Clin. Invest.* **1988**, *81*, 1162–1172.
172. T. Miethke; C. Wahl; K. Heeg; B. Echtenacher; P. H. Krammer; H. Wagner, *J. Exp. Med.* **1992**, *175*, 91–98.
173. B. G. Stiles; S. Bavari; T. Krakauer; R. G. Ulrich, *Infect. Immun.* **1993**, *61*, 5333–5338.
174. J. Y. Chen; Y. Qiao; J. L. Komisar; W. B. Baze; I. C. Hsu; J. Tseng, *Infect. Immun.* **1994**, *62*, 4626–4631.
175. S. R. Sarawar; M. A. Blackman; P. Doherty, *J. Infect. Dis.* **1994**, *170*, 1189–1194.
176. R. S. Yeung; J. M. Penninger; T. Kundig; W. Khoo; P. S. Ohashi; G. Kroemer; T. W. Mak, *Eur. J. Immunol.* **1996**, *26*, 1074–1082.
177. L. DaSilva; B. Welcher; R. Ulrich; M. Aman; C. David; S. Bavari, *J. Infect. Dis.* **2002**, *185*, 1754–1760.
178. M. M. Dinges; P. M. Schlievert, *Infect. Immun.* **2001**, *69*, 7169–7172.
179. R. Plaza; J. L. Rodriquez-Sanchez; C. Juarez, *Infect. Immun.* **2007**, *75*, 306–313.
180. R. L. Stone; P. M. Schlievert, *J. Infect. Dis.* **1987**, *155*, 682–689.
181. D. W. Beno; M. R. Uhing; M. Goto; Y. Chen; V. A. Jiyamapa-Serna; R. E. Kimura, *Crit. Care Med.* **2003**, *31*, 1154–1159.

182. A. H. Dalpke; K. Heeg, *J. Endotoxin Res.* **2003**, *9*, 51–54.
183. C. Sauter; C. Wolfensberger, *Lancet* **1980**, *2*, 852–853.
184. A. W. Chow; K. Bartlett; R. Percival-Smith; B. Morrison, *J. Infect. Dis.* **1984**, *150*, 80–84.
185. Y.-X. Zhao; A. Abdelnoour; T. Kalland; A. Tarkowski, *Infect. Immun.* **1995**, *63*, 4463–4469.
186. W. T. Huang; M. T. Lin; S. J. Won, *Infect. Immun.* **1997**, *65*, 2656–2662.
187. J. C. De Azavedo; J. P. Arbuthnott, *Infect. Immun.* **1984**, *46*, 314–317.
188. H. Fujikawa; H. Igarashi; H. Usami; S. Tanaka; H. Tamura, *Infect. Immun.* **1986**, *52*, 134–137.
189. K. D. Vlach; J. W. Boles; B. G. Stiles, *Comp. Med.* **2000**, *50*, 160–166.
190. G. Arad; S. Ley; D. Hillman; R. Kaempfer, *Nat. Med.* **2000**, *6*, 414–421.
191. J. W. Shupp; M. Jett; C. H. Pontzer, *Infect. Immun.* **2002**, *70*, 2178–2186.
192. G. Rajagopalan; M. M. Sen; C. S. David, *Infect. Immun.* **2004**, *72*, 6733–6737.
193. S.-J. Won; W.-T. Huang; Y.-S. Lai; M.-T. Lin, *Infect. Immun.* **2000**, *68*, 2003–2008.
194. R. D. LeClaire; W. Kell; S. Bavari; T. Smith; R. E. Hunt, *Toxicology* **1996**, *107*, 69–81.
195. B. Saha; D. M. Harlan; K. P. Lee; C. H. June; R. Abe, *J. Exp. Med.* **1996**, *183*, 2675–2680.
196. R. A. Buonpane; H. R. Churchhill; B. Moza; E. J. Sundberg; M. L. Peterson; P. M. Schlievert; D. M. Kranz, *Nat. Med.* **2007**, *13*, 725–729.
197. T. Krakauer; B. G. Stiles, *Clin. Diagn. Lab. Immunol.* **1999**, *6*, 594–598.
198. M. L. Hale; S. B. Margolin; T. Krakauer; C. J. Roy; B. G. Stiles, *Infect. Immun.* **2002**, *70*, 2989–2994.
199. J. Soltys; M. T. Quinn, *Infect. Immun.* **1999**, *67*, 244–252.
200. J. Darenberg; B. Soderquist; B. H. Normark; A. Norrby-Teglund, *Clin. Infect. Dis.* **2004**, *38*, 836–842.
201. O. Rott; B. Fleischer, *J. Infect. Dis.* **1994**, *169*, 1142–1146.
202. M. Llewelyn; J. Cohen, *Lancet Infect. Dis.* **2002**, *2*, 156–162.

Biographical Sketches

Holger Barth studied microbiology at the Technical University of Darmstadt and completed his Ph.D. thesis at the German Cancer Research Center in Heidelberg on the influence of toxic substances on the regulation of the cell division cycle. In 1995, he joined the laboratory of Klaus Aktories at the Institute of Experimental and Clinical Pharmacology and Toxicology in Freiburg, first as a post-doc and later as an assistant professor. Since 2004, he is a professor for pharmacology and toxicology at the Institute for Pharmacology and Toxicology at the University of Ulm. As a trained toxicologist, he is a member of the German Register of Toxicologists and Eurotox-Registered-Toxicologist. His research interests are on the interaction of bacterial protein toxins that target the actin cytoskeleton with eukaryotic cells and on the use of recombinant bacterial toxins as pharmacological tools. Together with Klaus Aktories, he characterized the mode of action of binary actin-ADP-ribosylating toxins and developed a cellular protein delivery system on the basis of a binary clostridial toxin. His most recent work focuses on the role of host cell chaperones for the translocation process of bacterial toxins across cell membranes into the host cell cytosol.

Bradley G. Stiles was formally trained in microbiology/biochemistry at Pennsylvania State University (B.S.) and Virginia Tech (Ph.D.). His research interests primarily involve protein/peptide toxins of bacterial and venom origins. Graduate studies were focused upon the purification and initial characterization of iota/iota-like toxins from *Clostridium perfringens* and *C. spiroforme*. After his post-doctoral investigations with *Treponema pallidum* and binding to RGD-containing proteins at the University of Texas Health Science Center (San Antonio), he moved to the US Army Medical Research Institute for Infectious Diseases (Frederick, Maryland) where he has been a principal investigator for the past 20 years. Early work in his laboratory involved the development of nonradioactive receptor binding assays for post-synaptic neurotoxins. These techniques led to the discovery of new toxins in coral and brown tree snake venoms, as well as the development/characterization of antivenoms which include monoclonal antibodies. Subsequent efforts included the development of a murine model for *S. aureus* superantigens and discovery of new vaccines plus therapeutics. The most recent work in his laboratory employs the use of recombinant antibodies as potential therapeutics toward staphylococcal and streptococcal toxins. Besides research, he teaches microbiology to undergraduates and spends 'free' time on old motorcycles.

5.07 Host Defense Peptides: Bridging Antimicrobial and Immunomodulatory Activities

Jason Kindrachuk, Anastasia Nijnik, and R. E. W. Hancock, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

This review is dedicated to the memory of Aaron W. Wyatt who passed away suddenly on 24 December 2008. Aaron was not only a superb colleague but also a good friend.

5.07.1 Introduction

Currently, one of the greatest medical challenges imposed is that of antibiotic- and multidrug-resistant bacteria. Although the introduction of antibiotics in the twentieth century marked the dawn of one of the most revolutionary periods in the history of modern medicine, the almost simultaneous identification of antibiotic-resistant bacterial strains foreshadowed the gradual incapacitation of these novel agents. Indeed, the half-life of maximal effectiveness of conventional antibiotics is normally 1–2 decades, with only limited numbers of novel antimicrobials being introduced within the same time periods. As a testament to this, the introduction of Linezolid in 2000 was the first novel class of antibiotics introduced since the 1960s. This disconnection between the effective lifetimes of conventional antibiotics and the discovery of replacement therapies has made the search for novel anti-infective strategies increasingly more urgent. Thus, an alternative set of agents, the cationic host defense peptides, represent a template for the development of a possible new class of antimicrobial therapies.

Short amphiphilic cationic host defense peptides are present in virtually every organism and play a central role in host defenses. The primary role of host defense peptides in organisms lacking an adaptive immune system is the resolution of microbial infection. Therefore, these peptides have traditionally been characterized based on their direct antimicrobial activities.[1] In this review, the term antimicrobial peptides is reserved for those molecules for which it has been proven that their primary biological function is the direct killing of microbes under physiological conditions. Recently, it has become increasingly appreciated that mammalian cationic peptides play a pivotal role in the regulation of mammalian immunity *in vivo* and often their direct antimicrobial activity is quite weak. Therefore, these molecules are referred to here as 'host defense peptides' in an attempt to encompass their broader contributions to the innate immune system in addition to their potential or demonstrated direct resolution of infection. These activities are not mutually exclusive, however; a single peptide may express different activities at different tissue sites. For example, mammalian α-defensins are almost certainly capable of direct killing in the granules of neutrophils or the crypts of the intestine where they may be found at $\mathrm{mg\,ml^{-1}}$ concentrations; however, at mucosal sites where they are found at nonbactericidal $\mu\mathrm{g\,ml^{-1}}$ concentrations it is likely that their immunomodulatory activity is important. Therefore for clarity, in this review we refer to these cationic peptides as host defense peptides to encompass both their known immunomodulatory activities and as a group name, and antimicrobial peptides when referring to their direct action in killing microbes.

Currently, more than 1000 natural host defense peptides have been identified. While not all of these have potent antimicrobial activity (some likely being principally modulators of innate immunity), many have been isolated or synthesized and several demonstrated to possess broad antimicrobial activities that include inhibitory actions on bacteria, viruses, and fungi. As precedents that indicate the therapeutic potential of these biomolecules, cationic peptide antibiotics such as the lipopeptide polymyxin B and the lantibiotic nisin have been used for decades as antimicrobials in medical and food processing applications, respectively, and several cationic peptides are currently moving through clinical trials. Although substantial data have been gathered regarding the antimicrobial activities of host defense peptides, many aspects of their activity are still poorly understood. This review discusses the relevant literature regarding host defense peptides from the perspectives of their structures, their associated antimicrobial and immunomodulatory activities, and their current and future clinical relevance.

5.07.2 Defining Cationic Host Defense Peptides

Host defense peptides are ubiquitous defense biomolecules found in virtually all forms of life – including bacteria, fungi, plants, invertebrates, and vertebrate species[2] with more than 1000 different peptides being either identified or predicted from nucleic acid sequences.[3] That host defense peptides have been retained within the innate immune system during the coevolution of eukaryotic hosts and prokaryotic pathogens indicates that they play a pivotal role in the innate immune response. Lower organisms lack an adaptive immune response and it is likely in these species that host defense peptides form a central component of the innate immune system against pathogen invasion. Within the immune systems of higher organisms, comprised of both innate

and adaptive immune components, these peptides are still likely to serve an essential role in infection resolution. Indeed, the demonstration that mice deficient in the murine cathelicidin, cathelin-related antimicrobial peptide (CRAMP), are significantly more susceptible to infection offered the first evidence that host defense peptides are an integral component of the innate immune arsenal.[4] This is also highlighted by the observation that human disorders such as specific granule deficiency syndrome,[5] atopic dermatitis,[6] and morbus Kostmann syndrome,[7] which are associated with the decreased production of one or more host defense peptides, all lead to an increase in the frequency and severity of bacterial and viral infections. Host defense peptides of higher organisms are expressed in many different cell types, and found on all body surfaces, tissues, and fluids. The localized expression of specific peptides has been postulated to target the activities of these molecules to the specific physiological niche occupied.[8] It is anticipated that this localized activation of cationic peptides might reduce their potential cytotoxic contact with susceptible host cells.[8]

Cationic host defense peptides are gene-encoded peptides that are synthesized ribosomally and produced constitutively or inducibly (through pathogen or inflammatory recognition events). Host defense peptides are often produced as inactive precursor peptides that are processed to an active form following the proteolytic cleavage of a prepro-peptide region.[9] As such, host defense peptides are characteristically defined based on a set of criteria: they are short (10–50 amino acids), carry an overall positive charge (+2 to +10) due to their high concentrations of lysine and/or arginine residues, and contain a high proportion of hydrophobic residues ($\geq$30%).[1] Generally, host defense peptides remain unstructured in aqueous solution with the adoption of an amphipathic or amphiphilic structure upon interaction with membranes,[1] an attribute that may be crucial both for activity and for reducing cytotoxicity. Although many of the natural host defense peptides have conserved physical properties, they are highly variable in terms of primary sequence demonstrating little sequence homology. They are normally classified into structural families based on secondary structures.

5.07.2.1 Antimicrobial Proteins

The term 'antimicrobial peptide' is generally limited to short cationic peptides. Larger polypeptides and proteins are known to also possess antimicrobial activity, and some of these, especially the degradation products of these antibacterial proteins, are often referred to as antimicrobial peptides. For example, lactoferrin, an 80 kDa glycoprotein of the transferrin family, and its by-products, the lactoferricins (Lfcins), have been demonstrated to have moderate direct antimicrobial activity.[10,11] These small, cationic Lfcins are derived from the N-terminal region of lactoferrin following treatment with pepsin under acidic conditions[10] and possibly naturally in the gut. Proteolytic digestion products of the complement protein C3 have also been demonstrated to have antimicrobial activity. C3a derivatives are microbicidal against both Gram-positive and Gram-negative organisms and interestingly a C3a derivative was able to limit *Streptococcus pyogenes* infection *in vivo* in mice.[12] However, it is worth mentioning that many of these by-products including C3a and lactoferrin are known immunomodulators.

Interestingly, biomolecules with demonstrated immunomodulatory activities such as cytokines and chemokines have also been found to possess direct antimicrobial activities. It should be noted however that the assessed activities are quite weak and measured under nonphysiological conditions; therefore, these activities may not be physiologically relevant but instead reflect their similar structural properties to other host defense peptides. Some microbicidal chemokines, or kinocidins, are direct chemoattractants that have recently been shown to have antimicrobial activity within a specific C-terminal domain of the protein.[13,14] The C-terminal domain is similar in structure and composition to many α-helical antimicrobial peptides and is typically cationic and amphipathic.[8] It should be noted that kinocidins lacking this general C-terminal cationic tail were also demonstrated to be antimicrobial.[13] Interestingly, the secondary structure adopted by these noncationic tail kinocidins demonstrates the formation of a large, cationic electrostatic patch on the surface of the molecule. Thus it would appear that the formation of a cationic domain in these molecules provides the attributes necessary for their antimicrobial activity. Kinocidins have also been found to contain iterations of the γ-core motif, a conserved domain present in a broad range of antimicrobial polypeptides lending credence to an evolutionary relationship between antimicrobial peptides, kinocidins, and related chemokines.[15] Thus, the direct microbicidal activity of the kinocidins, in combination with their ability to amplify the antimicrobial activity of leukocytes, and the corresponding chemokine activity of many host defense peptides may reflect a bridge between the innate and adaptive immune responses.

5.07.3 Host Defense Peptides Distribution and Class

5.07.3.1 Host Defense Peptide Distribution

As stated earlier, host defense peptides are thought to be central components in the immune system of virtually every organism. Although there are a limited number of structural categories within these molecules, there are a large number of groupings based on primary and/or secondary structural characteristics, host organism, or mode of action and these are overviewed in the following sections.

In mammals, two groups of host defense peptides are prominent: the cathelicidins (notably human LL-37, porcine protegrin, and bovine indolicidin) and the defensins (human α- and β-defensins). These two classes of peptides are involved in both intracellular and extracellular defensive responses and as individual peptides can be synthesized and stored in either active or inactive form within cells or secreted through degranulation into the surrounding environment.[16]

Cathelicidins, characterized by a conserved N-terminal 'cathelin-like' domain, are a large and diverse group of host defense peptides and were among the first mammalian antimicrobial peptides to be demonstrated to have broad-spectrum antimicrobial activity,[17] although their direct antimicrobial activities range from quite weak (e.g., LL-37) to very potent (protegrin). Although cathelicidin family members have been found in every mammalian species investigated, recent work has demonstrated their presence in chickens and hagfish.[18,19] Many domesticated mammals, like cattle, sheep, and pigs, express multiple cathelicidin genes that share a leader peptide and a homologous approximately 100 residue N-terminal cathelin precursor domain but contain a variable C-terminal domain, generally between 12 and 79 amino acid residues in length.[9]

Cathelicidins are synthesized in myeloid cells in an inactive prepro-peptide state and, following cleavage by signal peptidase to generate the pro-peptide, are sequestered within secretory granule subsets or specific granules of neutrophils.[20] They are also synthesized within the gastrointestinal, respiratory, and genitourinary tracts, and also by keratinocytes.[21] The pro-peptides remain inactive until further proteolytic processing by neutrophil elastase or proteinase-3 within neutrophils to generate active peptides as they are released into the extracellular fluid.[21,22] Recently, the inactivity of the N-terminal domain of cathelicidins has been challenged. Anderson *et al.*[23] have demonstrated that following proteolytic processing, the N-terminal cathelin domains of ovine cathelicidins display antimicrobial activity, whereas the full-length pro-peptide did not in ovine crude cell extracts. Similarly, the cathelin domain of LL-37 has also been demonstrated to have antimicrobial activity following proteolytic processing (**Figure 1**).[24] The authors demonstrated that recombinant cathelin was active against bacterial strains resistant to LL-37 at concentrations of 16–32 $\mu mol\,l^{-1}$ *in vitro*. This challenges the postulate that the only functions of the cathelin domain are suppression of the activity of the C-terminal region

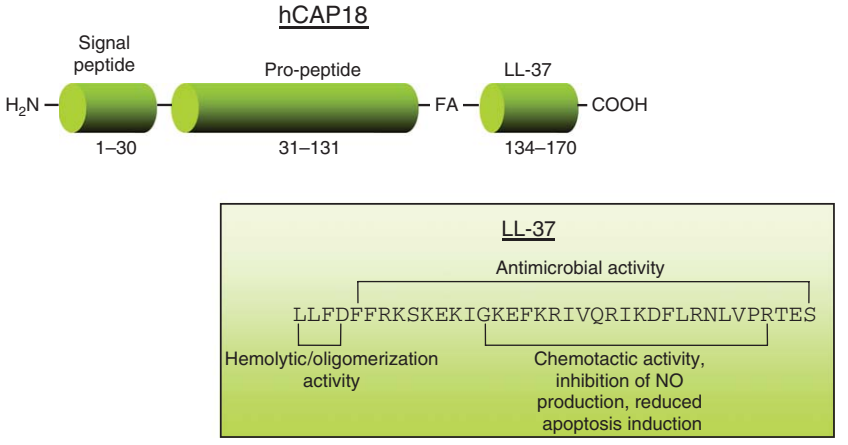


Figure 1 Cartoon representation of the human cathelicidin hCAP18. The amino acid residues that define each of the three hCAP18 domains have been added. The sequence of LL-37, the host defense peptide derived from hCAP18, is found in the inset. Structure–activity relationship studies have identified regions within the peptide that are necessary for both antimicrobial and immunomodulatory activities.

of the peptide and/or for assisting with peptide folding.[25] Many members of the cathelicidin family are α-helical peptides, but a rich diversity of other structures are found including β-hairpin structures like protegrin, the boat-like tryptophan-rich structure of indolicidin, the cyclic β-turn structure of bactenecin, and those forming extended polyproline-type structures.[26] The cathelicidins can be divided into three groups on the basis of their primary and secondary structures.[27] First, the most prominent group of cathelicidins are the type I linear amphipathic α-helical cathelicidins that are devoid of cyteine residues (e.g., CRAMP, LL-37, and SMAP-29). The second group of cathelicidins are the type II β-sheet, or β-hairpin, cathelicidins and are characterized by an even number of cysteine residues (e.g., protegrin). And third, the type III cathelicidins which contain a high proportion of one or two amino acids such as proline–arginine or tryptophan–proline (e.g., PR-39, Bac7, and indolicidin).

The second prominent mammalian host defense peptide group is the cyclic defensins, a family of peptides that are widely distributed across vertebrate species. Indeed, the naming of the defensin host defense peptide family is derived from the activities of these peptides within host defense settings.[28] The defensin family is divided into three subfamilies: (1) the α-defensin, (2) the β-defensins, and (3) the θ-defensins, a subfamily that is structurally distinct from both the α- and β-defensin subfamilies. All defensins are characterized based on the pattern of disulfide bonding between six conserved cysteines, with differing lengths of peptide segments between these conserved residues. Structurally, the α- and β-defensin subfamilies are characterized by triple-stranded β-sheets with a distinctive 'defensin fold'.[28] The θ-defensin family, found within several species of monkey and ape, are subgrouped based on their macrocyclic nature formed through the fusion of two truncated precursors.[2,29] Although the human genome contains multiple-θ-defensin genes, the presence of premature stop codons within both the genes and their transcripts aborts peptide translation.[30] Within humans there are six types of α-defensins (ranging in size from 3.5 to 4 kDa) and six types of β-defensins (normally 4–6 kDa). The α-defensins are normally comprised of 29–35 amino acids with a characteristic triple-stranded β-sheet structure in which β-hairpins contain the cationic residues.[30] The α-defensins are generally encoded as a tripartite prepro-peptide sequence of approximately 90–100 amino acids in length with the canonical sequence $x_{1-2}CXCRx_{2-3}Cx_3Ex_3GxCx_3Gx_5CCx_{1-4}$.[31] This is comprised of an N-terminal signal sequence (~19 residues), an anionic propiece (~45 residues), and the active C-terminal cationic defensin domain (~30 residues). It has been postulated that the anionic propiece may be important in folding processes of the mature peptide and/or in suppressing the activity of the cationic C-terminal domain. The β-defensins differ only in the length of the anionic propiece, as it is generally short or absent in this class and this lack of homology between α- and β-defensins has not yet been explained, and the canonical sequence $x_{2-10}Cx_{5-6}(G/A)xCX_{3-4}Cx_{9-14}Cx_{4-7}CCx_n$.[31] Intriguingly, the antimicrobial activities of the α- and β-defensins are abrogated by physiological salt conditions, whereas those of the θ-defensins are retained.[2] The human defensins, similar to the cathelicidins, are normally found within cells associated with immune functions (neutrophil granules, macrophages, NK cells) or at epithelial surfaces (respiratory mucosal surfaces, intestinal epithelial cells, and urogenital tract mucosal surface).[32] In particular in the intestinal tract a set of α-defensins, known as cryptidins, are highly expressed by Paneth cells and contribute to immune defenses of intestinal crypts. The expression of the defensins may be either constitutive or inducible; human beta defensin 1 (HBD-1) appears to be synthesized constitutively in the urinary tract, whereas HBD-2 synthesis in the skin is induced by microbial contact.[33] Although the absolute concentration of defensins within particular physiological niches has not yet been determined, it has been estimated that the concentration of these peptides may reach >10 mg ml^{-1} within the intestinal crypts or granules of leukocytes.[34,35]

In addition to the cathelicidins and defensins, humans also utilize a variety of other host defense peptides. Examples of these include the anionic dermcidins,[36] found in human sweat and possessing potent antimicrobial activity in a broad range of pH and salt concentrations, and the histatins, a histidine-rich host defense peptide family found in humans and higher primate species. The histatins are normally found in saliva and utilize an alternative mechanism to bacterial membrane lysis for their antimicrobial activity.[29]

Although this review will focus primarily on host defense peptides of mammalian origin, there is a wealth of information available on peptides from a diversity of species. An overwhelming number of host defense peptides of amphibian origin have been characterized with approximately 500 being described to date.[2] These amphibian-derived peptides are produced in the skin (e.g., magainin, dermaseptin) and as well in the mucosa of the stomach (e.g., buforin), where they have been postulated to play a central role in protection from ingested

pathogens. Of these the magainins are amongst the most well characterized. The magainins are α-helical in nature and have strong membrane-permeabilizing activities toward Gram-positive and Gram-negative organisms, fungi, yeast, and viruses.[2] As a consequence of the extensive characterization of the magainins, these peptides have served as a template for the development of the first therapeutic antimicrobial peptide Pexiganan.[37] The dermaseptins, isolated from the skin of South American frogs, are typified by lysine-rich linear peptides of 28–34 residues in length.[38]

The lack of an adaptive immune system within invertebrates requires that the resolution of infection can be carried out by the innate immune system alone. As in the vertebrate innate immune system, host defense peptides also play a central role in the immune defense of invertebrates. That invertebrates have been conserved throughout evolution lends credence to the effectiveness of the invertebrate immune system. Amongst the most well characterized invertebrate immune system is that from *Drosophila melanogaster*. Indeed, the identification and characterization of mammalian innate immune system components such as the Toll-like receptors (TLRs) is a consequence of information garnered from *Drosophila*. Within invertebrates, antimicrobial peptides have been found in the hemolymph, phagocytes, and in some epithelial cells and may be constitutively or inducibly expressed. The most abundant group of invertebrate host defense peptides are the open-ended cyclic defensins, containing three to four disulfide bridges.[2] The invertebrate host defense peptides also include the α-helical cecropins (*Drosophila*) and melittin (bee venom) and the β-hairpin-like polyphemusin and tachyplesin (horseshoe crab).

Antimicrobial peptides have also been postulated to play a central role in the plant innate immune response to microbial infection[2] and include the thionins, the defensins, and the cyclotides. It has been demonstrated that physiological concentrations of the thionins are active against bacteria and fungi *in vitro* and the heterologous expression of thionins in a transgenic plant model confer protection against bacterial challenge.[39,40]

Antimicrobial peptides are also utilized by prokaryotic organisms as a means of defense against competing microbes. The ribosomally synthesized bacteriocins are small heat-stable antimicrobial peptides with narrow or broad spectrum of activity and to which the host has a specific resistance.[41] Prominent examples of bacterially produced host defense peptides include nisin, a lantibiotic, used in food preservation, the topical antimicrobial therapeutics polymyxin B[42] and gramicidin S,[43] and the nonbacteriocin nonribosomal antimicrobial glycopeptide vancomycin.[44] The antimicrobial activity of bacteriocins is most active against low-GC Gram-positive species although limited activity against Gram-negative species has been demonstrated.[41] Many bacteriocins belong to those produced by food-grade lactic-acid bacteria (LAB) and these are defined by five main classes; it has recently been proposed by Cotter *et al.*[41] that this be revised to three primary classes. The class I bacteriocins are comprised of the lanthionine-containing lantibiotics, with nonlanthionine-containing bacteriocins in class II, and the bacteriolysins comprising class III. Amongst these, the lantibiotics have been the best characterized. Lantibiotics are typically small (19–38 amino acids), ribosomally synthesized peptides and are characterized structurally by the presence of lanthionine residues. Prominent members of this class of bacteriocins include nisin, subtilin, gallidermin, and Pep5.[45] There have been more than 40 lantibiotics reported and these can be further classified based on the modification enzymes involved in posttranslation modifications. Group A lantibiotics are comprised of elongated, amphipathic members and may be subgrouped as group AI lantibiotics which include those modified by LanB (which catalyzes serine and threonine dehydration) and LanC (which catalyzes (methyl)lanthionine ring formation), and the group AII lantibiotics in which LanM enzymes perform both modification reactions.[45] The group B lantibiotics are characterized by an overall globular structure and utilize LanM exclusively for processing.[45] Posttranslational modification common to all lantibiotics includes the dehydration of serine/threonine in the propeptide followed by the intramolecular addition of cysteine residues to form a lanthionic or methyllanthionine ring.[3] The lanthionine rings provide protease resistance but also impart defined spatial structures. The characteristic polycyclic structures of these molecules result from extensive posttranslational modifications involving the dehydration of Ser and Thr residues to 2,3-didehydroalanine (Dha) and (*Z*)-2,3-didehydrobutyrine (Dhb), respectively. Following this, the addition of Cys thiols to the unsaturated amino acids results in the formation of lanthionine (in the case of Ser) and methyllanthionine (in the case of Thr) bridges.[46] The lantibiotics are synthesized as pre-peptide with a 23–59 amino acid leader sequence and, following modification, the peptide is exported and the leader sequence is removed by proteolysis although not always in this order.[46] As with the leader sequences of host defense peptides found amongst higher organisms, there is a high degree of homology amongst lantibiotic leader sequences and in a similar fashion it is proposed that the lantibiotic leader sequence may act as a control for limiting lantibiotic activation by the host.[47] The most well-characterized lantibiotic is nisin, a 34-amino acid lantibiotic.

Nisin, discovered in 1928, is produced by *Lactococcus lactis* and is effective against a broad spectrum of Gram-positive organisms.[48] Gram-negative organisms appear to be intrinsically resistant to nisin as a result of the outer membrane barrier.[49,50] Importantly, nisin has been used as a food preservative for over 50 years in a broad range of foods ranging from dairy and bakery products to vegetables, meat, and fish.[30,51] Biosynthesis of nisin is achieved through autoregulation of the two-component regulatory system NisRK. Secreted nisin can bind to the histidine sensor kinase NisK resulting in autophosphorylation of the sensor and subsequent transfer of a phosphoryl group to the response regulator NisR resulting in transcriptional activation of the *nisABTCPRK* and *nisFEG* operons for nisin biosynthesis and self-protection of the producer strain, respectively.[52–54] Two naturally occurring variants of nisin have been characterized: nisin A and nisin Z.[55,56] These two variants differ by a single amino acid residue at position 27 (His in nisin A and Asn in nisin Z) and have been demonstrated to have almost identical antimicrobial activity.[57] Structurally, nisin consists of N- and C-terminal domains that are defined by the ring structures encompassed by each. The N-terminal domain contains three rings (rings A, B, and C) and is predominantly hydrophobic and the hydrophilic C-terminal domain contains rings D and E. These rings are connected by a flexible hinge region that has been demonstrated to play a central role in the antimicrobial activity of the peptide.[58] Recently, the potential of nisin as an anti-infective therapeutic has been proposed based on the targeting of lipid II (undecaprenylpyrophosphate-MurNAc(pentapeptide)GlcNAc), a molecule involved in the translocation of peptidoglycan subunits across the cell membrane thus pivotal for peptidoglycan synthesis.[59] It has been demonstrated that the defined structure of nisin provides a lipid II-binding motif through coordination of the lipid II pyrophosphate moiety by the N-terminal backbone amide residues in the cage-like structure of nisin.[60] Thus, nisin exerts its antimicrobial activity both through the inhibition of cell wall synthesis and via direct pore formation.

5.07.3.2 Host Defense Peptide Classes

Host defense peptides are highly diverse in terms of both primary sequence and host organism; however, recurring three-dimensional topologies have been demonstrated such that these molecules may be categorized based on their structural similarities (**Figure 2**).[8] It should be noted however that although many host defense peptides share similar three-dimensional structures there is little, if any, relationship between structural class and biological activity.[1]

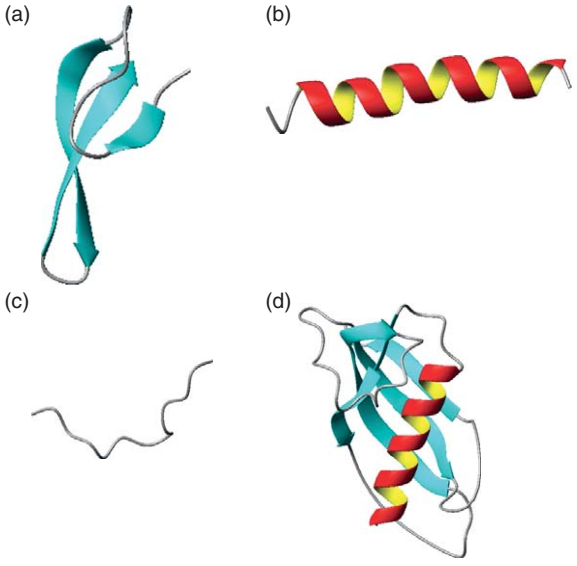


Figure 2 Selected structures of select host defense peptides representing the major structural classes: (a) β-sheet class (HNP-3; PDB 1DFN); (b) linear α-helical class (magainin; PDB 2MAG); (c) extended class (indolicidin; 1G89); (d) cysteine-stabilized α–β (protegrin-3; 1PFP). All structures were made with MOLMol and the color schema are as follows: red/yellow, α-helical propensity; aqua, β-sheet propensity; gray, extended or coil.

The amphipathic α-helical class of host defense peptides is the most abundant and most well-characterized class. Upon interaction with the hydrophobic membrane environment, the largely unstructured peptide adopts an amphipathic α-helical conformation with one helical face containing the majority of the hydrophobic residues, the opposite containing a large proportion of the polar residues.[1] These peptides are often short (<40 amino acids), devoid of cysteine residues, and found to be unstructured or linear in nonhydrophobic environments. Peptides found within this class include the antimicrobial peptide alamethicin,[61] bee venom melittin,[62] the magainins,[63] and the human cathelicidin LL-37.[21]

The β-stranded cationic peptides are the second largest class of host defense peptides. These peptides are highly diverse at the level of primary structure. Commonly, the members of the β-stranded class of host defense peptides are comprised of several antiparallel β-strands with stabilization by one or more disulfide bonds.[8] Recently, a common motif that integrates all disulfide-stabilized host defense peptides has recently been elucidated by multidimensional proteomic analysis. The γ-core, characterized by the presence of two antiparallel β-sheets as well as basic residues along its axis, has been found to recur in all major classes of cysteine-stabilized host defense peptides, across all biological kingdoms, and may represent a unifying structural motif.[64] The β-stranded class includes the highly antimicrobial tachyplesins[65] and polyphemusins,[66] protegrins,[67] and the commercialized antibiotic gramicidin S.[68]

The third class of host defense peptides, the extended peptide class, is defined by the relative absence of a defined secondary structure.[1] These peptides normally contain high proportions of amino acids such as histidine, tryptophan, or proline and tend to adopt an overall extended conformation upon interaction with hydrophobic environments. Examples of peptides belonging to the extended class include indolicidin, a bovine neutrophil peptide, and the porcine peptide fragment, tritpticin.[1] These structures are stabilized by hydrogen bonding and van der Waals forces as a result of contact with lipids in contrast to the intramolecular stabilization forces found in the former peptide classes.

Lastly, the loop or β-hairpin class of host defense peptides is characterized by a single disulfide bond and/or peptide chain cyclization as characterized by the bovine host defense peptide bactenecin.

5.07.4 Host Defense Peptide Target Selection and Self-Promoted Uptake

5.07.4.1 Host Defense Peptide Target Selection

Of central importance to the role occupied by host defense peptides within the immune system is the ability of these molecules to distinguish host cells from microbes. It is widely accepted that specific structures and/or functions within the host and microbial cells contribute to this selectivity although host defense peptide localization and regulated expression may also be contributing factors for reducing the toxicity associated with host defense peptides. The following sections discuss the recent literature highlighting this selectivity.

5.07.4.1.1 Membrane selectivity

Microbial and host cell membranes are comprised of phospholipid bilayers. Overall, the cell membrane bilayer is an amphipathic moiety with segregated hydrophobic and hydrophilic domains. Although the macromolecular structure of this bilayer is the same between prokaryotes and eukaryotes, the overall compositions of the two membranes are different. Eukaryotic membranes are enriched in neutrally charged lipids such as phosphatidylethanolamine (PE), phosphatidylcholine (PC), and sphingomyelin (SM).[8] In particular, the outer membrane leaflet of human erythrocytes is typically concentrated in PC (~20%), and SM (30%), and approximately 10% cholesterol. When present, anionic phospholipids have been found to be localized to the cytoplasmic leaflet. Consequently, eukaryotic membranes have an overall asymmetric distribution of phospholipids in the bilayer. Thus, the relative affinity of host defense peptides for eukaryotic biomembranes would be anticipated to be reduced. The interaction of host defense peptides with the host cell membrane has been postulated to amplify dissymmetry and phospholipid remodeling in target membranes. Lasch *et al.*[69] have demonstrated that the interactions of host defense peptides with microbial membranes induces 1,2-dimyristoyl-PE to segregate from lipolysaccharide (LPS) into distinct, well-defined domains. It has also been demonstrated that the asymmetric distribution of phospholipids may influence membrane phase transition and fluidity.[70,71] It must be appreciated however that the data derived from erythrocyte permeabilizations is biased

by the conditions of the assay and how this mimics the complexity of physiological environments. In contrast, microbial membranes are highly electronegative due to a predominance of phosphatidylglycerol (PG) and cardiolipin (CL) (~30% total content) and phosphatidylserine (PS) (70%).[1] It is generally accepted that the differences in the electronegativity between these two cell types is the distinguishing factor for host defense peptides although sterol content within membranes is also likely involved in target differentiation.[72] In accordance with this, Welling *et al.*[73] have demonstrated that the discrimination between microbial cells and host tissues by host defense peptides extends to *in vivo* circumstances. It was found that host defense peptides accumulated to a greater extent in bacterial-infected and *C. albicans*-infected lesions in mice and rabbits, compared to noninfected but inflamed tissues. Similar studies have also demonstrated that radiolabeled peptides efficiently discriminated between infected and noninfected tissue with up to fivefold differences in peptide binding between target and nontarget tissues within 1 h of peptide administration.[74] Thus, it has been speculated that the rapid localization and accumulation of host defense peptides at infection sites may be due to a preferential affinity for microbial cell surfaces rather than host tissues.[8]

In addition to the differences in phospholipid content between microbial and host cell membranes, it has been demonstrated that disparity exists between the transmembrane potentials of both organisms. The transmembrane potential is defined by the proton flux between the inner and outer bilayers of the cytoplasmic membrane and ranges from −90 to −110 mV in normal mammalian cells in contrast to transmembrane potentials of −130 to −150 mV for logarithmic phase microbes. The differences in these electrochemical gradients have been postulated to drive the influx of peptides into the cell and thus act as a crucial barrier for defining host defense peptide selectivity.[75]

In addition to these determinants, the success of immune system molecules such as host defense peptides relies upon the selectivity for the pathogen rather than the host. Host defense peptides have been optimized throughout evolution to selectively target a broad range of pathogens. This selectivity is defined by a number of specific biophysical themes, namely, conformation, charge, amphipathicity, hydrophobic moment (M_H), hydrophobicity, and polar angle.[8]

5.07.4.1.2 Host defense peptides and cationic charge

One of most well-defined characteristics of host defense peptides is the presence of cationic amino acid residues, namely, arginine and lysine residues, conferring an overall cationicity to the peptide. The excess of lysine and/or arginine residues confers an overall positive charge that ranges from +2 to +10, depending on pH and the number of basic residues.[76] The cationic residues of host defense peptides provide the means for the initial electrostatic interaction with the highly electronegative cell envelopes of bacteria.[76] Unlike the cell membranes of eukaryotes which are comprised of zwitterionic and neutral membrane lipids, bacterial cell membranes are comprised of highly electronegative phospholipids including PG, PS, and CL. These components confer an overall negative charge to the microbial cell surface and thus are naturally attractive target for cationic host defense peptides. Additionally, microbial membranes contain acidic compounds such as LPS (Gram-negative bacteria) and teichoic and teichuronic acids (Gram-positive bacteria) that impart additional negative charge and therefore additional attraction for the host defense peptides.[8]

It has also been demonstrated that the transmembrane potentials of prokaryotes are typically 50% greater than their eukaryotic counterparts. This chemiosmotic potential has been proposed to act electrophoretically on host defense peptides attached to the microbial surface.[75] It should also be noted that although increased cationicity of the host defense peptides is generally associated with increased antimicrobial activity there is a threshold by which selectivity between host and microbial cells is lost at the behest of increased cationicity.

5.07.4.1.3 Host defense peptide hydrophobicity and polar angle

Host defense peptide hydrophobicity (H) is defined as the proportion of hydrophobic amino acids within a peptide.[77] Typically, these peptides are comprised of ≥30% hydrophobic residues and this governs the ability of a host defense peptide to partition into the lipid bilayer, an essential requirement for antimicrobial peptide–membrane interactions. Typically, the hydrophobic and hydrophilic amino acids of natural peptides are segregated to create specific regions or domains that allow for optimal interaction with microbial membranes.[78] This likely represents evolutionary optimization to maximize the selectivity of these defense molecules. It has been established that increasing antimicrobial peptide hydrophobicity above a specific threshold correlates

with increased host cell toxicity and diminished antimicrobial activity.[79,80] Wieprecht *et al.*[79] have studied the functional correlations between H and membrane permeabilization using magainin analogues. It was demonstrated that changes to mean peptide H, with maintenance of cationicity, helicity, and M_H, resulted in increased permeabilization of neutral membranes but not those enriched in anionic phospholipids. Dathe *et al.*[81] have also demonstrated through systematic amino acid substitutions in an α-helical peptide model that peptide hydrophobicity and M_H are the most important parameters in defining cell membrane selectivity.

The polar angle of an antimicrobial peptide is defined by the relative proportion of polar vs. nonpolar regions within an amphipathic α-helix.[8] Thus, an α-helix comprised exclusively of polar residues on one face and nonpolar residues on the other face would have an overall polar angle of 180°. Polar angles within most natural α-helical peptides range from 140° to 180°[82] and it has been demonstrated that a smaller polar angle correlates with increased membrane permeabilizing activities, translocation, and pore formation rates.[79,83] However, it has been demonstrated that pore collapse is increased with decreasing polar angle and suggests the formation of less stable pores with such peptides.[8] Thus, the interactions between host defense peptides and hydrophobic membranes are influenced by the relative polar angle found within a host defense peptide.

5.07.4.1.4 Peptide amphipathicity and hydrophobic moment

In addition to cationicity and hydrophobicity, the specificity of interaction between host defense peptides and the microbial membrane is also influenced by amphipathicity. Traditionally, host defense peptides form amphipathic structures upon contact with the hydrophobic environment of the microbial membrane. Peptide amphipathicity reflects the polarization and relative proportions of hydrophobic and hydrophilic elements within a peptide. Although many protein conformations can achieve amphipathic structures, the α-helix represents one of the simplest. Most α-helices have a periodicity of approximately 3–4 residues per turn.[77] The resulting structure, with the polar side chains aligned along one side of the helix and the apolar residues lining the opposite side, is optimized for interaction with the amphiphilic microbial membrane.[84] It has been demonstrated that the amphipathicity of a host defense peptide correlates with both antimicrobial activity and cytoxicity.[8] M_H is a quantitative measure of peptide amphipathicity and is the vectorial sum on the hydrophobicities of individual amino acids normalized to an ideal helix.[8] As with increased amphipathicity, increasing M_H relates to a significant increase in host cell permeabilization and hemolytic activity. Increased M_H in magainin derivatives had little effect on calcein release from vesicles composed of electronegative PG; however, significant dye release was noted in more neutral membranes.[79] It was also noted that small increases in M_H correlated with a large reduction in the concentration of peptide needed for hemolysis of human erythrocytes. However, a high degree of peptide helicity and/or amphipathicity has been correlated with increased toxicity toward the neutral cell membrane of the host.[85,86] However, the importance of amphipathicity to activity is not limited to α-helical antimicrobial peptides. The amphipathicity of β-sheet antimicrobial peptides is characterized by the rigid segregation of charged and nonpolar regions of the peptide, a feature that is central to the mechanism of action of these peptides on membranes.[8] It has been postulated that the conformational flexibility of α-helical peptides allows for the adoption of amphipathic structures only within local settings such as membranes.[78] Thus, these peptides can maintain solubility within the aqueous extracellular matrix through limitation of hydrophobic domain formation; however, the rigidity of the β-sheet host defense peptides requires alternative mechanisms for maintenance of solubility. In this case, it is likely that the peptides form oligomeric structures that maximize hydrophobic contacts between peptides in a manner analogous to the localization of hydrophobic residues within the interior of a folded protein. Overall, host defense peptides have optimized amphipathicity that contributes to targeting of the microbial membrane; however, an increase in M_H above a critical threshold increases nonspecific interactions with host cell membranes.

5.07.4.2 Host Defense Peptides and Self-Promoted Uptake

As discussed in the earlier sections, host defense peptides have been optimized to interact with the bacterial membrane. Structurally, the Gram-negative bacterial outer membrane is an asymmetric bilayer, with the inner leaflet dominated by phospholipid and a highly divergent outer leaflet primarily composed of LPS.[1] Owing to a high content of phosphates and acidic sugars, LPS molecules carry a negative charge. The juxtaposition of LPS

molecules in the microbial membrane creates the potential for significant destabilizing repulsive forces between the LPS molecules. As a stabilization mechanism, divalent cations, such as Mg^{2+} or Ca^{2+}, are bridged between LPS molecules to partially neutralize the electrostatic repulsion between the negatively charged groups.[87] Thus, the initial interaction between a host defense peptide and the microbial membrane involves the binding of the peptide at the divalent cation binding sites on LPS at the cell surface.[1] The affinities of host defense peptides for LPS have been demonstrated to be at least three orders of magnitude higher than those of divalent cations.[75] This displacement of the divalent cation bridges results in local destabilization of the outer membrane. As a consequence, local transient 'cracks' are formed within the membrane with subsequent uptake of the peptide through the destabilized membrane culminating in the 'self-promoted uptake'.[1]

As host defense peptides are membrane-active molecules, safety mechanisms must be employed to avoid deleterious contacts with host cells. These mechanisms may involve the limitation of peptide activation to specific environments or niche-specific amplification. That most α-helical peptides remain unstructured in aqueous solution and undergo conformational transitions to an activated state within hydrophobic environments supports this postulate. It has also been postulated that the order of anionic phospholipids in microbial plasma membranes likely induces optimal periodicity of polar residues within host defense peptides at the membrane surface.[78]

5.07.4.3 Host Defense Peptide Attraction

The association of host defense peptides with lipid bilayers has been observed to be directly related to the ratio of peptide to lipid.[8] At low peptide/lipid ratios, peptides are oriented parallel to the membrane. As the ratio increases, the peptides reorient themselves perpendicular to the membrane, ultimately inserting into the bilayer. Following membrane insertion transmembrane pores are formed.[8] The insertion of peptides into the lipid membrane and subsequent translocation of peptides into the cytoplasm or formation of transmembrane pores has been described by multiple models of host defense peptide insertion.

5.07.4.3.1 Membrane insertion

The barrel–stave model (**Figure 3(b)**) describes the insertion of helical peptide bundles or 'barrels' in a ring around an aqueous pore.[88] The hydrophobic regions of the peptides are aligned with the lipid core regions and the hydrophilic peptide regions form the interior region of the pore. The 'stave' moniker refers to the individual transmembrane spokes within the 'barrel'. It is of note that alamethicin is the only host defense peptide that has been demonstrated to interact with membranes through this model.[89,90] In the second model, the accumulation of a high density of peptides at the bilayer surface results in the formation of a 'peptide carpet' (**Figure 3(a)**). During this associative phase, the cationic regions of the peptide interact with the anionic phospholipid head groups. As a threshold concentration of peptide is reached, unfavorable membrane energetics result in membrane disruption. In contrast to the barrel–stave model, membrane dissolution occurs in a dispersion-like manner rather than through the formation of transmembrane pores.[8] The third model, the 'torroidal pore' model, is functionally similar to the barrel model (**Figure 3(c)**). Although both models lead ultimately to the formation of transmembrane pores, the torroidal pore model differs in the nature of the physical interaction between peptides and the lipid bilayer. In the torroidal pore model, peptides are continuously associated with the polar head groups of the lipids due to the induction of lipid monolayer bending throughout the pore.[88] The resulting transmembrane pore is formed by the intercalation of peptides with lipid and the pore lining formed by the hydrophilic regions of the inserted peptides and the lipid head groups. This model of membrane disruption has been proposed as the primary mechanism of membrane interaction for analogues of magainin, protegrin, and LL-37.[78,91,92] The 'aggregate model' is similar to the carpet model of host defense peptide membrane insertion. Following binding of the peptides to the membrane interface (at a threshold concentration of peptide), the peptides reorientate themselves with the concomitant formation of micelle-like particles that span the lipid bilayer. These complexes can form channels within the membrane with subsequent leakage of ions and cytoplasmic contents or disintegrate enabling peptide translocation into the cytoplasm where the peptides may act on internal targets.[93] Additional models of host defense peptide-mediated membrane disruption have also been proposed. The 'molecular electroporation model' is defined by the creation of an electropotential difference across the membrane following peptide–membrane association (**Figure 3(d)**).

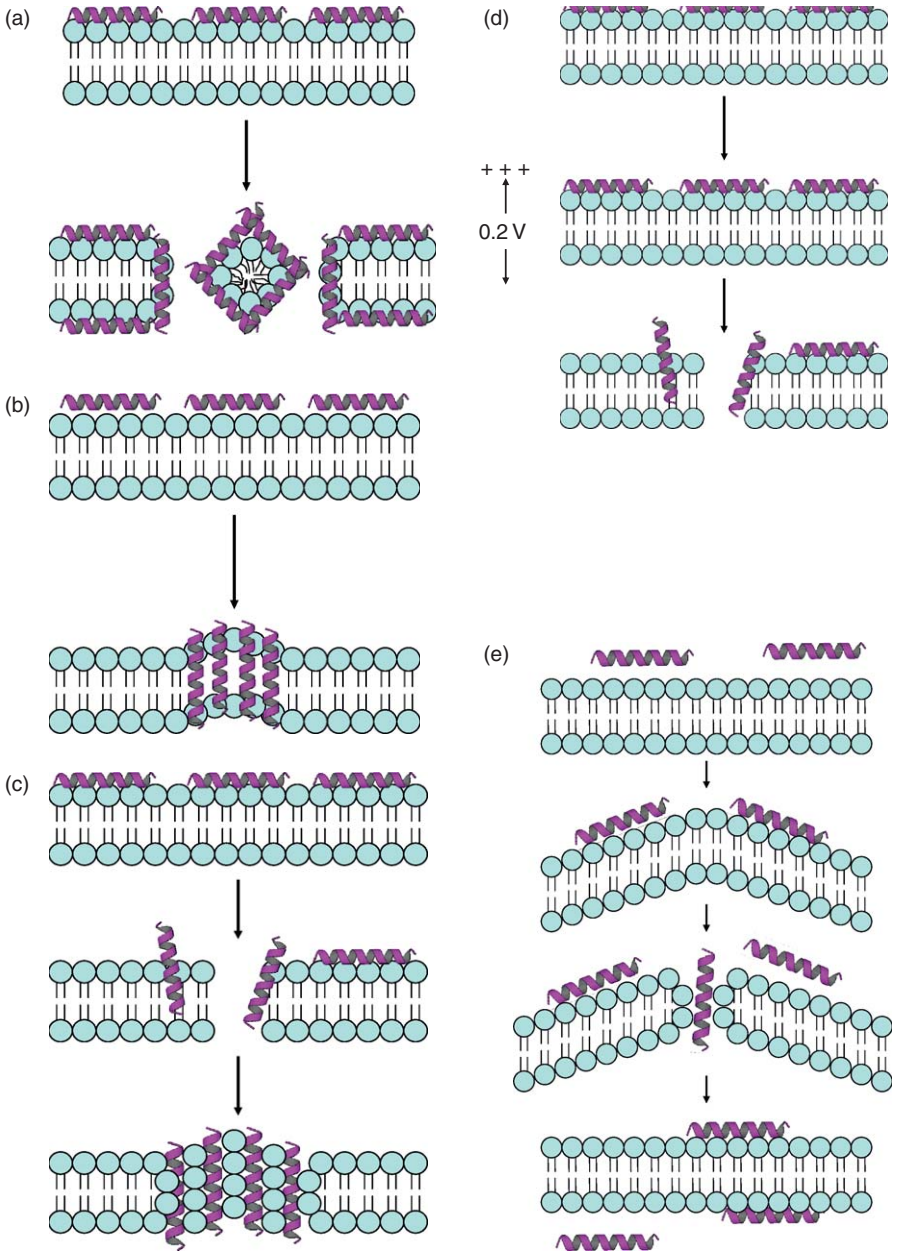


Figure 3 Host defense peptide membrane insertion models. (a) The carpet model. (b) The barrel-stave model. (c) The torroidal pore model. (d) Molecular electroporation model. (e) Sinking raft model.

As the electrochemical potential difference approaches 0.2 V, a pore is generated in the membrane leading to the leakage of cellular contents and eventual cell death.[30] Additionally, the 'sinking raft model' has also been proposed in which peptide translocation is driven by changes to membrane curvature (**Figure 3(e)**). As peptides associate with the cell membrane a mass imbalance is created resulting in increased membrane curvature and translocation of peptides across the bilayer. The self-association of peptides within the membrane results in transient pore formation allowing for peptides to localize on either side of the leaflet prior to reformation of intact membrane.[30]

Although there is evidence to support all of the above models, it has been suggested that membrane permeabilization by host defense peptides likely results from elements of all the proposed models.[88]

5.07.5 Host Defense Peptides and Cell Death

The paradigm for the antimicrobial activity of host defense peptides has long been considered to be a consequence of insurmountable defects in the microbial membrane. Thus, cell death would ensue based on the leakage of ions and metabolites, depolarization, defective biopolymer synthesis, and loss of cellular respiration.[8] Although it is likely that these effects contribute to cell death, recent observations suggest dissociation between membrane perturbation and microbial death.[1,8] Koo *et al.*[94] have demonstrated that host defense peptide-mediated membrane permeabilization does not invariably result in microbial death and Xiong *et al.*[95] demonstrated that the microbicidal activity of human neutrophil peptide (HNP-1) and human microbicidal protein 1 (tPMP-1) were a consequence of interference with cytoplasmic processes. It has also been demonstrated that depolarization of *Pseudomonas aeruginosa* following gramicidin S exposure does not effectively kill the microbe.[96] The exposure of *Escherichia coli* to fragments of Bac7, a proline-rich cathelicidin, led to a 2–5 log reduction in bacterial numbers but did not lead to permeabilization of the cells.[97]

Thus, the direct microbicidal activity of host defense peptides has been speculated to involve perturbation of both extracellular and intracellular architecture. In support of this, host defense peptides have been demonstrated to disrupt peptidoglycan synthesis upon membrane association.[98] Similarly, the disruption of intracellular processes as a consequence of exposure to host defense peptides has also been observed. Incubation of *Staphylococcus* with peptides has been demonstrated to inhibit DNA and/or RNA synthesis prior to cell death.[95] The authors speculated that this was due to the electrostatic interaction between the cationic residues of host defense peptides and the negative charge of the phosphate backbone of nucleic acids as has been observed for histone proteins. Also supporting a nonlytic mechanism of cell death, PR-39 has been demonstrated to bind directly to DNA and RNA, thus inhibiting the processes of transcription and translation.[99] The targeting of intracellular biomolecules by host defense peptides is not limited to nucleic acids as the insect host defense peptides pyrrhocidin, drosocin, and apidaecin have been demonstrated to bind to the bacterial heat-shock protein DnaK.[100] An investigation by del Castillo *et al.*[101] has also demonstrated a direct interaction between microcin B17 and DNA gyrase. The inhibition of RNA synthesis and protein synthesis and modest effects on DNA synthesis has been demonstrated in *S. aureus* following incubation with indolicidin and indolicidin derivatives.[102] This has also been demonstrated for the bactenecin derivative Bac2A as incubation with peptide concentrations of 2 and 10 times the minimal inhibitory concentration (MIC) led to inhibition of RNA, DNA, and protein synthesis.[1] This would support the postulate that the mechanism of antimicrobial peptide activity is not limited to destruction of the bacterial membrane and may also involve targeting of intracellular processes. Supportive of this, electron microscopy studies have demonstrated that the cellular damage lags substantially behind antimicrobial killing.[88] Indeed, Lehrer *et al.*[103] demonstrated the formation of membrane blebs on the surface of *E. coli* exposed to HNP-1. Kalfa *et al.*[104] have also shown that SMAP29 and CAP18 continued to exert cellular damage approximately 8 h following *P. aeruginosa* killing. Importantly, immunoelectron microscopy demonstrates that both peptides localized within the cytoplasm. Similarly, buforin II, a linear α-helical peptide with a proline hinge, accumulated in the cytoplasm but did not demonstrate membrane permeabilization.[105] The interactions between host defense peptides and bacteria have also been demonstrated to induce morphological changes to bacterial cells as PR-39 and indolicidin induce filamentation of *Salmonella typhimurium* and *E. coli*, respectively.[106,107]

5.07.6 Structure–Activity Relationship Studies of the Antimicrobial Activities of the Host Defense Peptides

Structure–activity relationship studies investigating the antimicrobial activities of host defense peptides have primarily sought the characterization of the specific sequence/structural motifs that dictate antimicrobial and cytotoxic activities. Perhaps unsurprisingly these activities appear to be dictated by a delicate balance of cationicity, hydrophobicity, amphipathicity, and ultimately the structural characteristics of the peptides.

The effect of increased cationicity on cytotoxicity and antimicrobial activity has been investigated in magainin II. Dathe *et al.*[108] have demonstrated that increased cationicity in magainin results in an increased

hemolytic propensity and concomitant loss of antimicrobial activity. Magainin II derivatives comprising overall cationic charges ranging from +3 to +7 were constructed and it was demonstrated that antimicrobial activity could be increased when overall charge was increased to a threshold of +5; however, charges above +5 resulted in a loss of antimicrobial activity. Importantly, the correlation between the maintenance of overall peptide hydrophobicity with optimized activity supports the existence of a delicate balance between hydrophobicity and cationicity within host defense peptides. The authors note that the loss in activity above the threshold charge of +5 likely represents reduced helicity within the peptides or excessive electrostatic interactions between the cationic peptides and the anionic phospholipid head groups. The relationship between peptide amphipathicity and antimicrobial activity has also been investigated using the cyclic β-sheet antimicrobial peptide gramicidin S, a peptide with high permeabilization activity toward neutral membranes. Lee *et al.*[86] have modified peptide amphipathicity while maintaining sequence, charge, and intrinsic hydrophobicity. Decreased peptide amphipathicity using diastereomeric substitutions (D- and L-amino acids) resulted in increased antimicrobial activity and decreased hemolytic activity.

Earlier investigations have indicated that the network of disulfide linkages found within β-defensins is not required for antimicrobial activity.[109] An investigation by Klüver *et al.*[110] has demonstrated that disulfide connectivity in a particular context is not required for antimicrobial activity. Indeed, HBD-3 derivatives lacking disulfide bonds were found to be as active as their cyclic analogues. This is in agreement with a previous study by Hoover *et al.*[111] in which a linear variant of full-length HBD-3 retained antimicrobial activity against a broad range of organisms. Interestingly, the authors also demonstrate that the antimicrobial activity of the peptide was dependent on overall cationicity as linear derivatives of the C-terminus, which houses the antimicrobial domain of the peptide, showed increasing antimicrobial activities that correlated with increasing basicity of the peptides. Derivatives of HBD-3 with Trp substitutions at the N-terminus also increased antimicrobial activity. This is perhaps unsurprising as previous studies of the Trp-rich peptides such as lactoferricin[112] and indolicidin[113] have identified the necessity of this amino acid for antimicrobial activity.

Tryptophan residues have been demonstrated to facilitate the most stable associations between peptides and membranes of all the amino acids. Trp residues are able to form hydrogen bonds with both water and lipid bilayer components upon imbedding into the interfacial region and as well as disrupt the strong hydrophobic interactions between the lipid acyl chains.[114] As a testament to the importance of Trp residues within antimicrobial peptides, it has been demonstrated that a parallel arrangement between Arg and Trp residues does not impede the formation of hydrogen bonds between water molecules and the Arg side chain.[114] As Arg and Trp residues are found in high proportions within many host defense peptides, it would be anticipated that interactions between these two molecules would occur. In contrast to Arg, Lys residues cannot hydrogen bond while in cation–π interactions with aromatic amino acids and this difference is likely responsible for the increased antimicrobial activity of Arg-rich peptides over their Lys variants.[114]

As many natural host defense peptides exhibit toxicity to host cells, it is prudent to investigate the characteristics that define cell selectivity between microbes and host cells as this will aid in the development of synthetic peptides with optimized activities. Melittin, a 26-residue antimicrobial peptide found within bee venom, has strong antimicrobial and cytolytic activities.[115] Indeed, structure–function studies have sought to define the molecular mechanisms behind the poor cell selectivity of the peptide. The three-dimensional structure of melittin indicates a tetrameric helix-bend-helix structure with the first helix encompassing residues 1–10, the second between residues 13–26, and a short bend region within residues 11–12.[116] Mutation of the first 20 amino acids with an alternative helix-forming sequence had little effect on either the cytolytic or antimicrobial activities of the peptide indicating a central importance of the helix to both activities.[117] Further, the individual deletion of residues Leu-6, Lys-7, Val-8, Leu-9, Leu-13, Leu-16, Ile-17, Trp-19, and Ile-20 resulted in the significantly reduced cytotoxicity of the peptide and, to a lesser extent, antimicrobial activity.[118] Recently, Asthana *et al.*[119] have demonstrated the presence of a leucine zipper motif between residues 6 and 20 of melittin. Interestingly, replacement of heptadic leucine with single- or double alanine resulted in significant reduction to the hemolytic activity of the peptide; however, antimicrobial activity was comparable to the unmodified peptide for both mutants. The authors noted that the membrane permeability of the analogues contrasted between neutral and negatively charged membranes and that this likely reflects different mechanisms of action for the peptides in the two environments. It was also noted that the substitutions of leucine to alanine in melittin decreased the ability of the peptide to self-associate and was postulated that self-association

is required for the partitioning of peptides into the cell membrane and eventual lysis.[119] The requirement of peptide self-association for hemolytic activity has also been demonstrated for other host defense peptides. Truncation of LL-37 has been demonstrated to reduce both hemolytic activity and oligomerization capability without affecting antimicrobial activity.[120] Interestingly, D-amino acid substitutions within host defense peptides have been demonstrated to reduce peptide oligomerization and hemolytic activities.[121] Although there has been little mechanistic evidence to explain this phenomenon, it is likely that the reduced potential for D-amino-substituted peptides to self-associate results in reduced hemolytic activity. In support of this postulate, the correlation between reduced overall hydrophobicity and hemolytic activity also likely results from a decreased oligomerization propensity.

Structure–activity relationship studies of host defense peptide antimicrobial activities have also extended beyond the spectrum of natural peptides. A recent investigation by Zelezetsky and Tossi[122] describes a methodology for the design of optimized α-helical peptides with optimized antimicrobial activity based on sequence patterns from natural host defense peptides. Through the analysis of 150 peptide sequences of varying length, it was found that most natural peptides contain the following: net charge range of +4 to +9, 40–60% hydrophobic residues, and a relative amphipathicity of 50–60% optimal. Interestingly, the authors conclude that antimicrobial activity is related to overall cationicity rather than a particular sequence context or motif. It has also been demonstrated that helical structuring of antimicrobial peptides is indicative of cytotoxic or antimicrobial activities; indeed, peptides that were less prone to structure maintained moderate antimicrobial activity with significantly decreased toxicity.[123] Dathe *et al.*[124] also have demonstrated using a series of variants of the idealized amphipathic α-helical peptide KLALKLALKALKAAKLA-NH_2. Reductions to overall peptide helicity as a consequence of substitutions within the middle region of the peptide led to marked reduction in hemolytic activities but only moderately affected antimicrobial activity. It would be anticipated that this disparity in activities is the result of the differences between host and microbial cell membranes such that the initial electrostatic interaction between the peptide and bacterial cell membrane would be unaffected by changes to helicity, whereas the ability to fold into an amphipathic conformation likely drives the insertion of peptides into host cell membranes.[1]

Thus, the optimization of the antimicrobial activity and minimization of cytotoxicity associated with host defense peptides through natural or synthetic means require the balance of a variety of physicochemical properties rather than the optimization of a single attribute.

5.07.7 Bacterial Host Defense Peptide Resistance

The ability of pathogenic microorganisms to survive host defense peptide exposure results from either constitutive (passive) resistance or inducible (adaptive) resistance. Constitutive resistance mechanisms, which bestow host defense peptide resistance as a result of ubiquitous expression of resistance factors, have been demonstrated for a variety of bacterial species including *Serratia* sp., *Proteus* sp., and *Providencia* sp.[77] These resistance mechanisms are predominantly due to alterations in structural features that decrease the attraction of peptides for the microbial membrane. For example, *Staphylococcus aureus* has a unique cell membrane lipid composition with enrichment of unsaturated menaquinones and lysyl-phosphatidylglycerol, a less electronegative derivative of PG.[125] Capsule production as a means of combating opsonization and phagocytosis has also been demonstrated in many virulent bacterial and fungal pathogens.[102] It is speculated that anionic carbohydrate and phosphate moieties of the capsule sequester host defense peptides thus conferring host defense peptide resistance.[102] It has also been postulated that pathogens may exploit particular physiological niches for conferring additional host defense peptide resistance. For example, *P. aeruginosa* preferentially colonizes tissue with abnormal osmotic and ionic strengths that ablate the antimicrobial activity of host defense peptides.[77] As these are evolutionarily conserved host defense molecules, it is perhaps unsurprising that bacteria have also coevolved resistance mechanisms. Conversely, inducible (adaptive) resistance refers to mechanisms that are induced in response to host defense peptides. Central to inducible resistance mechanisms are the two-component sensory systems. These systems induce diverse adaptive responses and provide pathogens with increased survival capabilities.[8] The physiological relevance of such defensive modifications is demonstrated in isolates of *P. aeruginosa* from cystic fibrosis patients. These isolates possess adaptive resistance

responses such as Lipid A modifications and N4-aminoarabinose-modified Lipid A changes, which have been postulated to increase the stability of the outer membrane.[126] Structural changes such as these would likely decrease the outer membrane permeabilization by host defense peptides. For example, sublethal concentrations of host defense peptides have been correlated with increased peptides resistance in *Pseudomonas*.[127] It has been speculated that the physiological conditions found within the lungs of cystic fibrosis patients would result in sublethal concentrations of peptides and may aid in the generation of resistant phenotypes.[127] The induction of host defense peptide resistance in *P. aeruginosa* can be attained with multiple passage of the bacterium against sublethal peptide concentrations;[128] however, it must be appreciated that the resistance patterns were modest, with only a two- to fourfold increases in MICs.[129] Conversely, resistance to the aminoglycoside gentamicin following bacterial passage under the same conditions against sublethal antibiotic concentrations induced resistance by 190-fold.[130]

Although rare, host defense peptide resistance has been observed in nature in microbial species such as *Staphylococci*, *Streptococci*, and *Salmonellae*.[131–133] It has been demonstrated that such species possess genes directly related to peptide resistance.[134,135] That host defense peptides have remained as an integral component of the innate immune system questions the concerns regarding peptide resistance within bacteria. Several peptides such as polymyxin B, gramicidin S, and nisin have been used in over-the-counter products without significant impact on the development of resistant strains. Much of the success of host defense peptides in combating pathogens is due to the metabolic cost of constitutive resistance as mutations of the microbial structures targeted by these peptides are not well tolerated by microbes.[136]

5.07.7.1 The PhoPQ Two-Component System of Salmonella

Although host defense peptide resistance has remained rare within microbes, concerns have been raised regarding the potential for clinical applications of peptides to provoke peptide-resistant phenotypes. Indeed, the introduction of antibiotics as novel microbial eradicators was met almost instantaneously with bacteria that had manifested various forms of resistance.[137] Thus, precautions should be taken with potential therapeutic application of host defense peptides to minimize similar resistance as seen with antibiotics. Many bacterial species rely on the use of two-component sensory systems to invoke phenotypes that increase virulence and recently one such system, the Salmonella PhoPQ system, has been demonstrated to respond directly to host defense peptides.[138,139]

The PhoPQ two-component system, found across a wide range of Gram-negative bacteria, has been identified as the master regulator of virulence within many pathogenic bacteria[140] and is involved in the governance of virulence characteristics, adaptation to Mg^{2+}-limiting conditions, and regulation of numerous cellular activities.[140] It has been demonstrated that PhoP-activated genes are maximally expressed within the phagosome[141] and *phoP*- and *phoQ*-null mutants are attenuated for virulence.[142] The PhoPQ system controls the expression of several genes required for microbial survival and increased virulence in Mg^{2+}-limiting environments including Mg^{2+} transporters,[140] LPS modifications,[140] and secreted proteases.[143,144] The direct and indirect expression of several transcription factors required for virulence, including SsrB,[145] SlyA,[146,147] PmrA,[148] and RpoS,[149] are also regulated by PhoPQ. In addition, PhoPQ also controls other two-component regulatory systems at transcriptional, posttranscriptional, and posttranslational levels including PmrAB.[150] It has been generally accepted that low (micromolar) Mg^{2+} concentrations activate PhoPQ following phagocytosis of bacteria, whereas high (millimolar) Mg^{2+} concentrations are PhoPQ repressive.[151–153] This hypothesis has recently been challenged with the demonstration that *in vivo* divalent cation concentrations are PhoPQ repressive following acidification of the phagosome.[154–156] Recently, Bader *et al.*[138] and Kindrachuk *et al.*[139] have suggested that host defense peptides may act as physiological ligands of PhoPQ.

5.07.7.2 PhoPQ and Host Defense Peptides

As many of the phenotypic changes induced following PhoPQ activation involve host defense peptide resistance characteristics, it is postulated that host defense peptides serve as an endogenous ligand for PhoQ.[138] Indeed, the exposure of *S. typhimurium* to sublethal concentrations of cationic antimicrobial peptides results in activation of the PhoPQ regulon and host defense peptide resistance phenotypes.[138,139,141] Supporting

a role for host defense peptides as physiological ligands of PhoQ, it has been demonstrated that the activation of PhoQ by peptides is similar to that of μmol l^{-1} Mg^{2+} concentrations, the widely accepted physiological ligand for PhoQ.[138,139] It is noteworthy that the PhoQ activation associated with host defense peptides is species-specific; a *Salmonella* mutant expressing a chimeric PhoQ constructed with the periplasmic domain of *P. aeruginosa* PhoQ failed to respond to antimicrobial peptides although both receptors responded similarly to environmental Mg^{2+} concentrations.[138] This has been suggested to reflect niche-specific evolution of the respective microbial species.[138] It was also postulated that an acidic periplasmic patch of the receptor represents a potential peptide binding domain. This region was confirmed as the endogenous Mg^{2+} binding site and was suggested to also act as the binding site for host defense peptides.[157] The authors hypothesize that Mg^{2+} serves to bridge the acidic periplasmic domain of PhoQ to the electronegative bacterial cell membrane. As the divalent cations shield the repulsive forces between the receptor and the membrane, the displacement of cations by peptide binding would be predicted to result in a significant repulsive force between the two moieties. Thus, it is postulated that this repulsion leads to a conformational change in PhoQ, resulting in activation.[157] This hypothesis was corroborated by the authors through mutational analysis of the highly acidic region of the PhoQ periplasmic sensor domain. It should be mentioned, however, that the recent demonstration of a binding domain on the sensor kinase PhoQ rather than a highly specific binding site *per se* would argue that the specificity of classical receptor–ligand relationships is deviated in PhoQ; a binding domain would allow for a more dynamic relationship between the sensor kinase and a potential ligand.[157] An investigation by Kindrachuk *et al.*[139] has also lent support for the activation of bacterial virulence characteristics through the binding of host defense peptides to PhoQ. This investigation proposes a third activity for host defense peptides: the activation of bacterial defensive responses and virulence characteristics by way of PhoPQ. Thus, the selection of host defense peptides based on their inability to initiate PhoQ activation would provide a rational basis for the evaluation of potential peptide candidates.

5.07.8 Direct Antimicrobial Activity of Host Defense Peptides *In Vivo*

Although host defense peptides have traditionally been regarded as molecules that act as direct antimicrobial agents, the relevance of this activity to peptide functions *in vivo* has been called into question.

First, under normal conditions the physiological concentrations of host defense peptides are far lower than those required to exert antimicrobial activity.[158] As a testament to this, the constitutive expression of LL-37 results in peptide concentrations of approximately 5 μg ml^{-1} in the bronchoalveolar lavage of healthy infants, a concentration difference of 3–6-fold in comparison with the MIC for LL-37 against a wide range of microbes *in vitro*.[159,160] Furthermore, the concentration of HBD-2 in the lung during inflammatory episodes is estimated to be approximately 10-fold lower than the concentration required to kill common pathogenic bacteria.[161,162] This lack of direct antimicrobial activity associated with subsets of host defense peptides under normal physiological conditions might also represent a niche-specificity for this particular activity. This has been postulated for LL-37 antimicrobial activity within the gastrointestinal system as the MICs for enteropathic strains of *E. coli* and *Salmonella* in the presence of physiological salt concentrations are <10 μg ml^{-1}, although the concentration of LL-37 found within this particular niche remains unclear.[163]

Second, the physiological niches in which host defense peptides are normally sequestered would be highly antagonistic to antimicrobial activity. Many of the studies directed at host defense peptide antimicrobial activities utilize assay conditions that do not directly mimic physiological conditions and it has been demonstrated that physiological concentrations of salt and cations are mildly to severely suppressive of peptide-associated antimicrobial activities.[102,158] The antimicrobial activities of many host defense peptides have been assessed in buffers devoid of physiological concentrations of divalent cations such as Mg^{2+} or Ca^{2+} (1–2 mmol l^{-1}), monovalent cations such as Na^{+} and K^{+} (100 mmol l^{-1}), and polyanionic biomolecules such as glycosaminoglycans and mucins.[3] Investigations of the particular ions or media components responsible for this antagonism have demonstrated that divalent cations are far more antagonistic than monovalent cations.[102] For example, the addition of 1 mmol l^{-1} $MgCl_2$ was found to be more antagonistic to the activity of model α-helical peptides on *P. aeruginosa* than 200 mmol l^{-1} NaCl.[102] A recent investigation by Raimondo *et al.*[164] postulated that the inhibitory effect could be due to either monovalent/divalent cation interactions with host defense peptides or the nonspecific binding of

serum components to the peptides. The authors demonstrate that the presence of $CaCl_2$ or NaCl ablated activity, whereas other divalent cations (Mg^{2+}, Zn^{2+}, or Fe^{3+}) had little to no effect. Interestingly, the anionic serum protein human serum albumin was also demonstrated to inhibit the antimicrobial activity of host defense peptides adding more complexity to this list of potential inhibitors.[164,165] Proteases and peptidases would also be presumed to have an effect on the stability, and therefore activity, of host defense peptides. Indeed, the prevalence of basic residues within host defense peptides represents an ideal target for trypsin-like proteases. Pini *et al.*[166] have demonstrated that the linear amphibian antimicrobial peptide magainin could be completely inactivated in the presence of serum within 2 h and attributed this loss in activity to peptide degradation. The bovine cathelicidin indolicidin had a broader range of stability and was not degraded within the same time period although it was no longer detectable at 24 h.[166] The lability of natural host defense peptides within serum has been correlated with the presence of the enzymes trypsin and proteinase K.[38] The overall stability of host defense peptides to proteases and peptidases is rooted in both the primary and secondary structures of these biomolecules. Indolicidin has been demonstrated to be partially resistant to both trypsin and chymotrypsin owing to the high concentration of tryptophan residues (nearly 40%) within the peptide possibly representative of an evolutionary tactic to avoid protease inactivation.[30] In a similar manner the increased presence of proline residues within many cathelicidins has been correlated with increased resistance to serine proteases.[129] Recently, Maisetta *et al.*[167] have evaluated the inhibitory components of human serum on the antimicrobial activities of HBD-3 in an attempt to identify the reason for inhibition of bactericidal activity in the presence of serum. The addition of 20% human serum inhibits the antimicrobial activity of HBD-3 and the removal of salts, the inactivation of proteases or peptidases, or a combination of both did not restore activity. That incubation of HBD-3 with physiological concentrations of trypsin for extended periods did not completely degrade the peptide was postulated to be due to HBD-3 oligomerization.[164] It has also been demonstrated that bacterial proteases and peptidases can also inactivate host defense peptides. A large proportion of microbial species produce proteases or peptidases which have been demonstrated to inactivate endogenous host defense peptides thus contributing to disease pathogenesis. The human pathogens Group A *Streptococci*, *P. aeruginosa*, *Proteus mirabilis*, and *Enterococcus faecalis* have been demonstrated to target and inactivate LL-37 through secretion of proteases.[168] Indeed, the addition of specific protease inhibitors to cell culture supernatants resulted in preservation of LL-37 antimicrobial activity *in vitro*.[168] Sieprawska-Lupa *et al.*[143] have also demonstrated that aureolysin production in *S. aureus* is inversely correlated to LL-37-mediated killing.

Third, that virtually every cationic peptide sequence possesses some degree of antimicrobial activity *in vitro* suggests that this activity may be a general phenomenon associated with all peptides rather than solely to host defense peptides.[158] Indeed, it has been demonstrated that a threshold amount of hydrophobicity and/or cationicity are sufficient to bestow antimicrobial activity within a peptide.[38,169]

Although numerous arguments can be made against the physiological relevance of the direct antimicrobial activity of host defense peptides, there have also been observations that corroborate such a role for these peptides. Indeed, it can be argued that in the appropriate physiological setting host defense peptides would be expected to have meaningful direct microbicidal activity. The α-defensins present within azurophilic granules of neutrophils have been estimated to reach mg ml^{-1} concentrations, far greater than the MIC values associated with these peptides.[170] Furthermore, the concentration of α-defensins in the intestinal crypts following microbial stimulation is also estimated to be in the mg ml^{-1} range, and thus more than sufficient for direct antimicrobial activity.[35] That host defense peptide genes are subject to positive selection, and belong to the most rapidly evolving group of mammalian proteins, has also been suggested to indicate a physiological role for host defense peptides in host immune defense;[136,171] however, it is not known whether this evolution is driven by selection for direct antimicrobial properties or for other activities, such as the diverse immunomodulatory properties of the peptides, discussed in the next section.

5.07.9 The Role of Peptides in Mammalian Immunity – The *In Vivo* Evidence

Despite the unresolved issues regarding the relative importance and the exact physiological roles of the direct antimicrobial and immunomodulatory activities of host defense peptides, there is overwhelming evidence that the peptides play a pivotal role in the mammalian innate immune response. Human disease

known as the specific granule deficiency syndrome, which results in a loss of neutrophil α-defensins, is associated in persistent bacterial infections.[5] The deficiency of α-defensins in human neutrophils has been demonstrated to increase susceptibility to recurrent infections.[158] Further supporting a role for host defense peptides in innate immunity, the deficiency of the human cathelicidin LL-37 and human α-defensins (HNP1-3) has been correlated with frequent oral bacterial infections and severe periodontal disease in morbus Kostmann.[7] Furthermore, low levels of intestinal α-defensins correlate with increased susceptibility to diarrhea in humans.[172] In rodent models the deficiency of matrilysin, a protease required for the cleavage and activation of murine intestinal α-defensins, led to increased susceptibility to oral *S. typhimurium* challenge in transgenic mice.[173] The importance of host defense peptides to the resolution of infection also extends beyond the spectrum of bacteria. CRAMP knockout mice demonstrated significantly greater vaccinia pox formation as compared to wild-type mice suggesting a role for CRAMP in the inhibition of orthopox virus replication.[17] In a similar fashion HNP 1-3 and HBD-5 have been demonstrated to antagonize infection by human papilloma virus (HPV) in both cutaneous and mucosal papillomavirus types, suggestive of an association between a deficiency in α-defensin expression and predisposition to HPV infection.[174] Adding further evidence for a central role of host defense peptides within the innate immune system, deficiencies in LL-37 or HBD-2 within the epithelium have been associated with atopic dermatitis and an increase in the incidence of *S. aureus* infections.[175] Recently, aberrant host defense peptide expression has also been implicated in the pathogenesis of inflammatory bowel disease (IBD) as IBD patients display reduced inducible expression of HBD-2 and -3 within the intestinal mucosa that may aid disease progression.[176] The demonstration that the pig cathelicidin PR-39 directly stimulates the synthesis of syndecan-1 and -4 from fibroblasts suggests that the peptide may play a role in wound repair as the absence of either glycosaminoglycan is associated with abnormal wound repair.[177,178] Furthermore, both PR-39 and LL-37 promote angiogenesis through direct activity on vascular endothelium, and the CRAMP-deficient mice show impaired angiogenesis during wound repair.[179,180] These examples highlight the importance of host defense peptides within the innate immune system above and beyond purely antimicrobial mechanisms. Although these examples support a central role for host defense peptides within the innate immune system, this correlation does not extend to all endogenous host defense peptides. Mice generated with a deficiency in the cathelicidin CRAMP were only partially compromised in their ability to combat skin infections, while β-defensin 1 knockout mice demonstrated delayed pulmonary clearance of *H. influenzae*.[4,181] It has been suggested that this might represent the redundancy associated with the activities amongst the murine β-defensins or that the inactivation of a constitutively expressed peptide may be muted as compared to one that is induced upon infection.[158] As many natural host defense peptides have overlapping activities with respect to both antimicrobial and immunomodulatory activities, the loss of a single peptide may not be detrimental to the host. In a similar fashion, polymorphisms amongst another central component of the innate immune system, the TLRs, result in only modest changes to infection susceptibility due to the overlapping activities of these receptors.[182]

5.07.10 Immunomodulatory Activities of Host Defense Peptides

Recently, the activities of host defense peptides related to the resolution of infection have been suggested to result in part from nondirect antimicrobial activities.[183] It has been postulated that immunomodulation may represent the primary action of these peptides *in vivo* as the immunomodulatory activities are retained under physiological conditions in contrast to the direct antimicrobial activities of most natural mammalian host defense peptides.[184] These immunomodulatory activities include, but are not limited to, direct chemotactic activity, induction of chemokines and other immune mediators, stimulation of leukocyte degranulation and other microbicidal activities, effects on leukocyte and epithelial cell survival and apoptosis, stimulation of epithelial and endothelial cell proliferation, promotion of wound healing and angiogenesis, antiendotoxic and anti-inflammatory activities, and adjuvant functions.[158] These will be described in detail in the following sections and a summary is found in **Table 1**.

Table 1 Immunomodulatory activities and cellular locations of a subset of natural host defense peptides

Peptide	*Origin*	*Cellular location*	*Immunomodulatory activity*	*Proposed receptors*
CRAMP	Murine	Immature neutrophils, spleen, respiratory and intestinal epithelia, keratinocytes, testis	Cell proliferation/differentiation; chemotaxis; antiendotoxic; adaptive immune polarization; *in vivo* protection; cytokine/chemokine induction	FPRL1, murine FPRL2
Defensins	Mammals, birds, invertebrates, plants, fungi	Membrane permeabilization, macromolecular synthesis inhibition	Cell proliferation/differentiation; chemotaxis; induction of gene expression; adaptive immune polarization; *in vivo* protection; cytokine/chemokine induction	CCR6, TLRs -1, -2 and -4
LL-37	Human	Immature neutrophils, monocytes, mast cells, NK, B, T cells, keratinocytes, respiratory, gastrointestinal, genitourinary epithelia, synovial membrane, nasal mucosa	Cell proliferation/differentiation; angiogenesis; induction of gene expression, mast cell degranulation, antiendotoxic; cytokine/chemokine induction	FPRL1, P2x7, non-FPRL1 Gi-coupled receptors
PR-39	Porcine	Immature neutrophils; lymphoid organs	Cell proliferation/differentiation; chemotaxis; induction of gene expression	$p130^{Cas}$, 26S proteasome

5.07.10.1 Angiogenesis and Cell Proliferation

The process of angiogenesis, during which new blood vessel formation occurs, was reported to be stimulated by a number of host defense peptides via direct activity on endothelial cells. For example, LL-37 and the mouse homologue CRAMP have been observed to increase endothelial cell proliferation and formation of vessel-like structures.[185] The authors demonstrate that stimulation with LL-37 resulted in the induction of blood vessel growth in a chorioallantoic membrane assay. It was also noted that collateral vessel growth and blood flow were increased in response to LL-37 in a rabbit hind-limb model.[185] Furthemore, CRAMP knockout mice demonstrate reductions in vascular structure formation at sites of tissue damage.[185] An investigation of the molecular mechanisms by which LL-37 mediates such angiogenic activities has demonstrated that angiogenesis is stimulated by the direct action of LL-37 on endothelial cells and is mediated by the formyl peptide receptor-like 1 (FPRL1) receptor.[185] Porcine peptide PR-39 was also reported to promote angiogenesis in a number of *in vitro* and *in vivo* model systems, and this activity was mediated via the inhibition of the proteosome-mediated degradation of hypoxia-inducible factor (HIF)-1α in endothelial cells.[179]

Host defense peptides were also reported to affect survival, proliferation, and migration of epithelial cells. For example, LL-37 is reported to induce keratinocytes migration *in vitro*, via a mechanism involving indirect transactivation of epidermal growth factor receptor (EGFR) and transcription factor signal transducer and activator of transcription-3 (STAT3).[186] Furthermore, adenoviral delivery of LL-37 gene to wounded tissue in mice significantly accelerated wound closure, although in this case the activity was reported to be EGFR independent.[187] HNP1-3 and HBD2 were similarly reported to promote wound healing in epithelial cells in several *in vitro* models.[188,189] Recently, an investigation by von Haussen *et al.*[190] also suggested that LL-37 could act as a growth factor for lung cancer cells, as overexpression of LL-37 is associated with increased tumor growth in animal models. These events also appear to involve the EGFR pathway. von Haussen *et al.* further demonstrated that an all-D-amino acid LL-37 derivatives retained the tumor growth-promoting activity. As chirality is of inherent importance to receptor–ligand interactions, the activation of the EGFR pathway in this case might occur via an indirect mechanism rather than via a direct receptor–peptide binding.

LL-37 and other peptides have also been demonstrated to modulate pathways regulating cell survival and apoptosis in various cell types. The effects of LL-37 on neutrophil survival remain controversial. Barlow *et al.*[191] and Nagaoka *et al.*[192] both report that LL-37 inhibits neutrophil apoptosis; however, in the first study this is shown to depend on the receptors $P2X_7$ and FPRL1, Barlow *et al.* suggest that $P2X_7$ and a non-FPRL1 Gi-protein-coupled receptor are involved. In contrast, a recent study by Zhang *et al.*[193] suggests that LL-37 over the same concentration range actually promotes neutrophil loss via secondary necrosis. LL-37 is similarly reported to promote apoptosis of primary airway epithelial cells.[191]

5.07.10.2 Chemotaxis and Chemokine Induction

Host defense peptide-mediated induction of chemokine synthesis and direct chemotactic activities have been well documented. That chemokines and host defense peptides have been demonstrated to have modest overlapping activities led to the postulation that host defense peptides may have evolved through chemokine gene duplication.[194,195] Compared to chemokines, host defense peptides have weaker chemotactic activity but appear to act over a broad range of cell types and species.[158] Supporting this, LL-37 has been demonstrated to be chemotactic for rat mast cells,[196] mouse monocytes, and neutrophils,[197] and a variety of human cells[198] including peripheral blood monocytes, neutrophils, and CD4 T cells *in vitro*.[198,199] Recently, using intravital microscopy it was also shown that neutrophil-derived LL-37 is critical for extravasation and recruitment of inflammatory monocytes in a mouse model *in vivo*.[200] LL-37 was also shown to promote mast cell degranulation, resulting in release of histamine and prostaglandin D2.[201] Owing to the many immunomodulatory activities associated with LL-37, it has been suggested that LL-37 might promote cell recruitment by activating amplification loops, as degranulation of mast cells in response to LL-37 would result in the release of inflammatory mediators, thus increasing vascular permeabilization and potentially the infiltration of neutrophils to the site of inflammation.[20] These direct chemotactic activities have not been limited to LL-37; however, PR-39 has been demonstrated to be chemotactic for neutrophils.[202] Further, HBD-1 and HBD-2 are chemotactic for immature dendritic cells (iDCs) and memory T cells,[203] an activity recently demonstrated to be mediated via chemokine receptor 6 (CCR6).[158]

Host defense peptides have also been demonstrated to induce the expression and secretion of chemokines by a variety of cell types. For example, LL-37 interacts with airway epithelial cells and at physiological concentrations induces chemokine (C-X-C motif) ligand (CXCL)-8/interleukin (IL)-8 chemokine secretion.[204] Similarly, in human monocytes LL-37 induced chemokines IL-8, chemokine (C-C motif) ligand (CCL)-2/monocyte chemoattractant protein (MCP)-1 and CCL7/MCP-3, and this activity was further synergistically enhanced in the presence of granulocyte macrophage colony-stimulating factor (GM-CSF).[205] Microarray analysis of RAW264.7 and THP1 macrophage cell lines and of primary human monocytes stimulated with LL-37 further demonstrated broad changes in gene expression, with an induction of a range of chemokines, chemokine receptors (CCRs), and other molecules involved in cell adhesion, communication, and motility.[206,207]

In an effort to elucidate the mechanisms by which host defense peptides mediate such higher-order immunomodulatory activities, investigations have sought the identification of the receptor(s) involved in these responses. Much of the information gathered on the interactions between host defense peptides and cell receptors has been derived from studies of LL-37. Yang and coworkers have demonstrated that LL-37-mediated monocyte chemotaxis and Ca^{2+} mobilization are dependent on FPRL1.[198] Indeed, the mobilization of Ca^{2+} could be inhibited by pertussis toxin and cross-desensitized by an FPRL1-specific agonist and chemotaxis by LL-37 was specific for FPRL1-positive cells including monocytes, neutrophils, and T cells. That the differentiation of monocytes into iDCs results in the ablation of LL-37-induced chemotaxis and Ca^{2+} mobilization by LL-37 was postulated to be due to the loss of functional FPRL1 during the differentiation. Alternatively, Niyonsaba *et al.*[196] have shown that LL-37 employs two non-FPRL1 Gi-coupled receptors on mast cells, the so-called high- and low-affinity receptors. As treatment with phospholipase C inhibitor suppressed LL-37-mediated mast cell chemotaxis, these results suggest the involvement of a phospholipase C signaling pathway. Investigations of β-defensin chemotactic activities have also identified potential cellular receptors involved in mediating these activities. Pazgier *et al.*[208] and De and coworkers have demonstrated that human β-defensin-induced migration of HEK293 cells was dependent on the expression of CCR6.[203] Indeed, the chemotactic activities of the β-defensins could be ablated by treatment with antibodies against CCR6.[203] It appears however that the chemotactic activities of other β-defensins are not limited to CCR6; HBD-3 is chemotactic for cells that do not express CCR6, indicating an as-of-yet-unidentified β-defensin receptor.[209] This has also been demonstrated in a recent investigation by Soruri *et al.*[210] in which HBD-2 and HBD-3 failed to induce the migration of a cell line stably transfected with CCR6.

Receptors mediating the LL-37-induced chemokine production are also extensively investigated. In lung epithelial cells, using a combination of inhibitors, it was established that LL-37 activates a metalloproteinase, which indirectly stimulates (EGFR) possibly by cleaving a membrane-anchored EGFR ligand, and this is required for the activation of downstream signaling pathways and IL8 production by the cells.[204] LL-37 was also reported to act through the $P2X_7$ receptor in promoting posttranscriptional processing of IL-1β by caspase-1 in LPS-primed monocytes;[211] however, it was unclear whether the interaction of LL-37 with $P2X_7$ was direct.

Although many of the investigations aimed at identifying host defense peptide receptors have focused on those localized to the cell surface, it has also been demonstrated that these peptides may utilize intracellular receptors. For example, a 15-residue active fragment of porcine peptide PR-39 has been demonstrated to bind cytosolic adaptor protein $p130^{Cas}$ resulting in its association with the cytoskeleton and activation of downstream signaling pathways, including PI3K.[212,213] Bao *et al.*[214] have also demonstrated that the PR-39 peptide directly interacts with the α7 subunit of the 26S proteasome, and that this interaction results in a selective suppression of nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, α (IkBα) degradation and inhibition of the nuclear factor kappa B (NFκB) pathway, in several cell lines and *in vivo* in mouse models of pancreatitis and myocardial infarction. Selective inhibition of proteasome activity by the PR-39 peptide is also responsible for the suppressed degradation of the hypoxia transcriptional regulator HIF1α and the angiogenic properties of the PR-39 peptide.[179] Alternative mechanisms for mediating the immunomodulatory activities of LL-37 that involve cell entry have also been investigated. It has been noted that the primary sequence of LL-37 shares similarities to nuclear targeting sequences and that this peptide cotransports molecules across the cell membrane, thus providing an alternative mechanism for the induction of cellular responses.[215,216] Indeed, LL-37 enters lung epithelial cells and localizes to the perinuclear region and the inhibition of peptide internalization inhibits IL-8 induction, suggesting the importance of this mechanism for LL-37 biological activity.[215] Recently, an investigation by Lande *et al.*[217] further demonstrated that LL-37 can

bind and promote the cellular uptake of short nucleic acid strands, enhancing their ability to serve as TLR-9 ligands leading to the induction of cellular responses such as type I interferon production. Importantly, these cellular responses were distinct from those elicited by the peptide or nucleic acid alone. Thus, the immunomodulatory activities of host defense peptides appear to extend past the stimulation of cellular responses at the cell membrane.

Investigations of the immunomodulatory properties of host defense peptides have also extended beyond the realm of natural peptides. A study by Scott *et al.*[218] investigated the immunomodulatory properties of the 13-residue synthetic peptide innate defense regulator-1 (IDR-1). Importantly, IDR-1 is a nonhelical extended peptide owing to the presence of residues incompatible with helix formation, and thus is considered to be devoid of direct antimicrobial activity. Despite this, IDR-1 conferred protection in *S. aureus* challenges in mice with intraperitoneal (i.p.), intravenous (i.v.), and subcutaneous (s.c.) routes of infection, with the doses of 8–24 mg kg^{-1} being protective even when administered at a site distinct from the infectious challenge. Pertinent to the therapeutic potential of host defense peptides, IDR-1-mediated protection was demonstrated for peptide administration before (−48 h) or after (+4 h) infection. That IDR-1 offers protection when administered at 24- or 48-h before bacterial challenge would be anticipated to reflect a nondirect antimicrobial means of protection; the peptide MIC against *S. aureus* was found to be >128 μg ml^{-1}, far greater than that anticipated to be present within the bloodstream upon challenge. Offering further support for the therapeutic potential of IDR-1, the tolerable peptide dose was found to range from 100 to 125 mg kg^{-1} when administered i.v. (and much higher when administered i.p.) and did not result in high levels of proinflammatory cytokines response. IDR-1-mediated protection to bacterial challenge appears to be at least partially mediated through the activation of specific cytokine and chemokine productions within the immune system. In particular, IDR1 stimulation of human monocytes *in vitro* resulted in approximately fivefold upregulation of chemokines (MCP-1 and MCP-3), cytokines (IL-10 and IL-19), and the dual-specificity kinase MEK6 and subsequent kinetic studies of chemokine induction verified that this response was concentration dependent.

5.07.10.3 Antiendotoxin Activities of Host Defense Peptides

Excessive induction of inflammatory mediators in response to infection can result in potentially lethal tissue damage and may be as detrimental as the infection itself.[1] Systemic inflammatory syndrome or sepsis is estimated to be responsible for over 200 000 deaths per year in the United States.[219] Although conventional antibiotics are able to selectively target infectious organisms, they are unable to address the pathological inflammation that may result from these infections. In contrast, a number of host defense peptides have been demonstrated to neutralize the potentially deleterious effects of endotoxin both *in vivo* and *in vitro*. These peptides exhibit the unique ability to selectively suppress pro-inflammatory responses while boosting many aspects of protective immunity, and are thus of high interest as potential anti-inflammatory therapeutics.[220]

Peptides of diverse sequence, derived from many host organisms, as well as many synthetic host defense peptides, share the ability to modulate LPS-induced responses. LL-37 is one of the peptides for which the antiendotoxic properties have been most widely investigated. Owing to its cationic nature, LL-37 binds LPS and prevents its interaction with LPS-binding protein (LBP),[221–223] an interaction that is essential for TLR4 activation. However, the direct interaction with LPS accounts for only a small part of the antiendotoxin activity of LL-37, as shown by the following studies. First, the ability of LL-37 to neutralize LPS-mediated proinflammatory responses is not dependent on the coadministration of these molecules. Mookherjee *et al.*[206] demonstrated that LL-37 pretreatment and removal prior to LPS stimulation similarly result in a powerful suppression of tumor necrosis factor (TNF)-α production. Furthermore, peptides such as MBI-28 and polymyxin B have been shown to neutralize TNF-α synthesis in mice when administered 60 min prior to LPS exposure.[224] Second, host defense peptides have been demonstrated to selectively suppress the expression of a subset of LPS-inducible genes in monocytes and macrophages.[206,225] The genes that were most strongly suppressed included many proinflammatory mediators, while expression of anti-inflammatory regulators and genes involved in cell migration and interactions were retained. This phenomenon has been demonstrated for the synthetic cecropin–melittin hybrid CEMA,[225] as well as for LL-37.[206] In contrast, in the event of a direct interaction between LPS and a host defense peptide, it would be anticipated that LPS-induced gene expression would be globally suppressed. Third, the anti-inflammatory activity of LL-37 is not limited to modulation of

responses to LPS. For example, in human peripheral blood mononuclear cells, LL-37 also suppresses proinflammatory responses to lipoteichoic acid (LTA),[206] and in dendritic cells (DCs) responses to LPS, LTA, and flagellin.[226] Importantly, host defense peptides have also been shown to offer protection against endotoxemia *in vivo* in animal models.[225,227,228] Thus, the control of inflammatory responses to invading pathogens and of the immune homeostasis with the bacterial microflora may be an important physiological function of host defense peptides. The unique ability of the peptides to suppress hyperinflammatory responses while maintaining protective effector functions of the immune response makes them attractive candidates for the development of anti-inflammatory therapeutics.

5.07.11 Host Defense Peptides and the Adaptive Immune Response

The innate and the adaptive arms of the immune systems of higher organisms are bridged by a complex set of cellular and molecular mediators. The immunomodulatory activities of host defense peptides, such as the direct chemotactic activity, cytokine and chemokine induction, and the stimulation of differentiation and activation of effector cells, may contribute to this process. Host defense peptides may also influence the polarization of adaptive immune response, for example, through their effects on cytokine production.[229]

DCs are central to the bridging of innate and adaptive immunity and are the key antigen-presenting cells (APCs) that initiate and orchestrate T-cell responses following antigen capture and presentation.[230] The induction of adaptive immunity requires antigen uptake by iDC and DC activation and maturation by microbial products and host proinflammatory stimuli. DC maturation results in increased expression of costimulatory molecules and in changes in chemokines receptor expression, which leads to DC migration to the local lymph nodes or submucosal lymphoid tissues. There antigen presentation to T cells takes place, leading to T-cell activation and clonal expansion. The maturation status and cytokine secretion by DCs also play a key role in establishing the Th1/Th2 polarization of the immune response.

LL-37 has been demonstrated to affect DC differentiation from monocytes.[229] Thus, monocytes primed with LL-37 demonstrated upregulated phagocytic capacity, and when induced to differentiate to DCs exhibited increased costimulatory molecule expression, enhanced Th1-cytokine production, and improved ability to activate Th1-polarized T-cell immunity. A subsequent investigation by Kandler *et al.*[226] working with monocyte-derived iDCs demonstrated that for these cells LL-37 inhibited responses to LPS and a number of other TLR ligands, consistent with the widely recognized antiendotoxic activities of LL-37. In contrast, for plasmacytoid DCs LL-37 was reported to strongly augment responses to self-DNA and class B CpG oligonucleotides. This activity resulted in increased production of type I interferons and was suggested to contribute to the immune pathology of psoriasis.[217] Thus, the precise effect of LL-37 on DC activation is complex and depends on the class and differentiation status of the DCs and the context and timing of LL-37 exposure.

Several studies suggested a function for β-defensins in DC activation as endogenous TLR ligands. Murine β-defensin 2 was reported to act as a TLR4-ligand, promoting DC maturation and inducing strong Th1-polarization of the immune response *in vitro* and *in vivo*.[231] Similarly, HBD-3 was reported to activate monocytes and DCs, acting via TLR-1 and -2 receptors and myeloid differentiation primary response gene 88 (MyD88)- and IL receptor-associated kinase-1 (IRAK1)-dependent signaling pathways.[232] Additionally, human α-defensins HNP1–3 enhanced *in vitro* proliferation and secretion of Th1 cytokines by anti-CD3-stimulated T cells, and this was associated with adjuvant activity *in vivo*.[233]

Various host defense peptides have been shown to enhance antigen-specific responses at both humoral and cellular levels. For example, a number of human defensins have demonstrated adjuvant activities. Co-administration of HNP1–3 with ovalbumin (OVA) resulted in enhanced production of OVA-specific immunoglobulin G (IgG) antibodies when delivered intranasally and in the generation of OVA-specific $CD4^+$ T cells in mouse models.[234] Coadministration of HBD-1 or HBD-2 with OVA also resulted in enhanced OVA-specific IgG antibodies.[234] Similarly, DNA vaccines, consisting of murine β-defensin 2 fused with tumor antigens, or with gp120 of human immunodeficiency virus (HIV), induced cell-mediated Th1-polarized responses in immunized mice.[235,236] LL-37 and murine cathelicidin CRAMP have also been shown to act as adjuvants *in vivo*, when administered together with test antigen OVA or as a part of a DNA-based mouse antitumor vaccine.[237,238] It should be noted that a model peptide $KLKL_5KLK$ was also demonstrated to be an

effective adjuvant when administered with OVA or added as a component to a commercial influenza vaccine.[239] This raises the possibility that such activities may be seen with a broad range of host defense peptides.[32] Overall, these examples suggest that host defense peptides may act as endogenous adjuvants for the enhancement of immune responses to particular antigens. Thus, host defense peptides may be regarded as mediators of both the innate and the adaptive immune responses.

5.07.12 Structure–Activity Relationship Studies of Host Defense Peptides – Immunomodulatory Activities

Recently, structure–activity relationship studies of human β-defensin have investigated the structural characteristics necessary for chemoattractive activities. Importantly, the native structure of HBD-3 has three disulfide bonds between cysteine residues C1–C5, C2–C4, C3–C6 and in this configuration is able to induce migration of CCR6 HEK293 cells at concentrations as low as 10 ng ml^{-1}.[222] An investigation by Wu *et al.*[240] demonstrated that the pattern of disulfides within HBD-3 is a critical determinant of chemotactic ability. Creation of topological analogues with different disulfide bonding patterns decreased this activity by 10–100–fold, whereas replacement of all cysteines decreased this activity by 1000-fold implying that the configuration of disulfides is critical to functional interaction with the receptor.[240] Although the overall disulfide linkages within HBD-3 are indispensable for immunomodulatory activity, the various isoforms evoke unique chemokine activities; however, these disulfides are not essential for antimicrobial activity. As chemotactic activity is a receptor-mediated event it is anticipated that the loss of activity within the linear analogue of HBD-3 is due to the destabilization of the peptide ligand as CCR binding/activation is dependent upon stable ligand structures.[240,241] In contrast, Taylor *et al.*[242] demonstrated that the chemotactic properties of HBD-3 are retained with substitution of five of the six cysteines found within the molecule. Interestingly, activity is not dependent on the simple introduction of a cysteine residue into the peptide; introduction of cysteine into position five of the six cysteine motif restores chemotactic activity, whereas a derivative with a single cysteine at position one is inactive.[242] Indeed, synthetic structural analogues of HBD-3 with noncanonical cysteine connectivities induced migration of CCR6-expressing HEK293 cells at concentrations of 10- to 100-fold higher than those of HBD-3 with canonical cysteine connectivity.[240] An HBD-3 analogue in which the cysteines were substituted with α-aminobutyric acid did not show any chemotactic activities at concentrations of 1000-fold higher than the parent peptide. It appears that the chemotaxis activity of HBD-3 is not merely a consequence of cysteine connectivity or stable three-dimensional structure; however, as chemotactic activity is evident in HBD-3 derivatives devoid of disulfides but still possessing cysteines.[243] In agreement with Wu *et al.*, the authors demonstrate that peptides in which all cysteine residues are replaced with alanine retain antimicrobial activity. Thus, it would appear for this host defense peptide that immunomodulatory and antimicrobial activities are independent of one another.

It has been demonstrated that the β-defensin family contributes to both innate and adaptive immune responses based on chemotactic and activating potentials for a wide variety of blood cells. In particular, HBDs interact directly with the CCR6 of iDCs and memory T cells.[203] X-ray crystallography and NMR spectroscopy have demonstrated that defensins of both the α- and β-families form a common overall fold characterized by the triple-stranded antiparallel β-sheet which is constrained via the three intramolecular disulfide bonds.[244] On the basis of the proposed activation of a shared receptor it is anticipated that the natural chemokine ligand for CCR6, CCL20/macrophage inflammatory protein-3 (MIP-3), and the human β-defensin would share similar structural features. While these signaling molecules share very little sequence similarity, β-defensins adopt a similar tertiary structure as CCL20/MIP-3α in the form of an antiparallel triple-stranded β-sheet with a C-terminal α-helix.[245] Specifically the Asp^4-Leu^9 motif in the β-defensins is a structural analogue of the Asp^5-Asp^8 in CCL20 which is believed to be essential in the interaction with CCR6.[243] This motif may be a crucial determinant of the ability of these host defense peptides to exert a chemotactic effect.

Many structure–activity relationship studies of host defense peptides have also focused on LL-37. Importantly, LL-37 has a broad spectrum of activities within the immune system but has also displayed cytotoxic activities that would diminish the therapeutic application of this peptide. Thus, structure–activity studies of LL-37 have focused on this cytotoxicity. A study by Nagaoka *et al.*[246] has demonstrated that the cytotoxic and hemolytic activities of LL-37 may reside within different regions of the peptide. Using an 18-mer

LL-37 derivative LLKKK, the authors showed no associated cytotoxicity in murine macrophages and have suggested that this region of the peptide cannot be attributed to the overall cytotoxicity previously demonstrated for LL-37 but rather only to its hemolytic activity. It is noteworthy that N-terminal truncation derivatives showed reduced hemolysis of red blood cells although both peptides contained the highly hemolytic 18-mer region. Thus it can be suggested that it is not the presence of a particular sequence that determines peptide activity but rather the context of the sequence within the full-length peptide.

Recently, Ciornei *et al.*[247] demonstrated that N-terminal truncations of LL-37 reduced hemolytic activity and suppressed LPS-induced nitrous oxide production *in vitro*. Truncation of the first six N-terminal residues also resulted in a significant reduction in proapoptotic activity of the peptide in cultured human vascular smooth muscle cells, as measured by DNA fragmentation, but was less pronounced in a 2-residue N-terminal truncation derivative suggesting that the removal of these hydrophobic residues reduces the cytotoxic activities of the peptide. Interestingly, an 18-mer derivative of LL-37, LLKKK, comprising residues 15–32 of LL-37 with E16L, Q22K, K25L, D26K, and N30K substitutions, had much higher hemolytic activity than the full-length parent peptide. Importantly, chemotactic activity was conserved in all three LL-37 derivatives. In contrast, a recent investigation by Sigurdardottir *et al.*[248] demonstrated that a peptide comprised of residues 14–34 of LL-37 had reductions in hemolytic activity and apoptosis induction in human cultured smooth muscle cells. In agreement with the investigation by Ciornei *et al.*, this LL-37 fragment retained chemotactic activities and inhibition of LPS-induced nitric oxide production as compared to the full-length peptide.

Interestingly it was reported that truncated and processed forms of LL-37 are naturally present in the human skin and sweat.[249] Braff *et al.*[250] investigated the properties of these naturally occurring LL-37 derivatives. They report that several of the truncated LL-37 peptides have increased microbicidal activity against a broad spectrum of microorganisms compared to the full-length LL-37, and suggest that the microbicidal activity resides within amino acids 11–30 of LL-37. Importantly, the investigators show that the antimicrobial activity of the peptides does not correlate with their immunomodulatory properties. Instead the immunomodulatory properties, specifically the induction of cytokine production in keratinocytes, were dependent on EGFR and Gi-coupled receptors, but independent of the stereochemistry of the peptide with both L and D forms showing strong activity. The investigators conclude that the immunomodulatory activity of the LL-37-derived peptides can be structurally dissociated from the microbicidal properties, is independent of a stereospecific interaction of the peptides with a receptor but relies on an indirect activation of EGFR and Gi signaling pathways through an unknown mechanism. In a recent publication the team further reports that abnormalities in the proteolytic processing of LL-37 in human skin are associated with inflammatory disorder rosacea.[251] The study suggests that increased activity of proteases in the skin of rosacea patients results in the presence of unusual LL-37 derivatives that induce cytokine production by keratinocytes and contribute to the inflammatory condition.

Structure–activity relationship studies have also aimed to unravel the complexity of the host defense peptide antiendotoxin activity. Recently, Rosenfeld *et al.*[252] have investigated the structure–function relationships within synthetic host defense peptides from the perspectives of both LPS binding and non-LPS binding-mediated antiendotoxin activities. Using a 15-mer parent peptide constructed solely of K and L residues (and which folds into an ideal amphipathic α-helix), the authors investigated, scrambled, and segregated derivative peptides constructed of either L- or D-amino acids. The authors demonstrated that all peptides retained comparable abilities to bind LPS directly thus demonstrating an independence of the interaction to peptide secondary structure. In contrast, the ability to suppress LPS-mediated inflammation was demonstrated to be highly dependent on secondary structure. Although all peptides inhibited LPS-mediated inflammatory response in macrophages it was noted that the scrambled- and segregated-D-amino acid derivatives were significantly less active than the other peptides. Indeed, at high LPS concentrations the most active detoxifying peptides were found to be those that adopted α-helices in solution or retained an amphipathic fold in the all-L or the D,L forms.[252]

5.07.13 Therapeutic Applications of Host Defense Peptides

The therapeutic development of conventional antibiotics has often been biased toward a 'single-target, single-action' clinical mechanism. Since their inception more than a half-century ago, antibiotic therapies have been a cornerstone in the treatment of infectious disease. The extensive application of conventional

antibiotics, in addition to the limited introduction of new antibiotic classes, has served as a selective pressure for the emergence of multidrug-resistant microbial strains. This has provoked the exploration for new antimicrobial agents and strategies of which host defense peptides are regarded as a leading candidate for alternative antimicrobial therapies.

Host defense peptides have significant advantages over conventional antibiotics. Perhaps the most compelling is that host defense peptides are multifunctional components of the innate immune system with a multitude of host and pathogen targets. Indeed, host defense peptides have demonstrated direct antimicrobial activity, immunomodulatory and/or endotoxin-neutralizing activities, and synergism with conventional antibiotic therapies.[253] Additionally, these molecules have retained their activities throughout evolution in the face of continual microbial pressure.

Although host defense peptides are an intriguing alternative to conventional infectious disease therapies questions have been raised regarding their efficacy. In particular, the potential for host defense peptide-mediated induction of microbial resistance through selective pressure. It would be anticipated that the development of resistance to such a central component of the innate immune system would have severe consequences for the host; however, this must be tempered with an appreciation for several issues that mitigate this concern. First, host defense peptide knockouts in animals have demonstrated that the ablation of a single peptide species has only modest effects on infection susceptibilities or animal health likely due to the host defense peptide activity redundancy. Second, as host defense peptides have a multiplicity of targets within both the pathogen and host the generation of resistance phenotypes would require the microbe to significantly change many essential biomolecules, an energetically unfavorable task.[3] Third, host defense peptide-resistant mutants display limited cross-resistance to other peptides.[254] Finally, the therapeutic application of host defense peptides that are optimized for immunomodulatory rather than antimicrobial activities would also severely limit the selective pressure for host defense peptide resistance patterns.

As a result of these favorable activities many host defense peptides are advancing through discovery, development, and clinical trials (**Table 2**). For example, Migenix is investigating the therapeutic potential of an indolicidin derivative, MX-226, in the treatment of catheter-associated infections.[3] Although MX-226 did not achieve success in the prevention of catheter-associated infections peptide administration resulted in a 49% reduction of local catheter site infections and a 21% reduction of catheter colonization. At present MX-226 is in confirmatory phase IIIb clinical trials.[3,255] MX-594AN (Migenix; topical Omiganan), a formulated cationic antimicrobial peptide, is currently in phase II clinical trials as a topical therapeutic for mild-to-moderate acne vulgaris and has demonstrated efficacy against a range of acne lesions.[256] Interestingly, Migenix has also developed MX-594AN as a topical treatment for rosacea under the name CLS001. Upon completion of phase II clinical trials there was no statistically significant difference between those in the treatment group as compared to the control group for the primary endpoint; however, a statistically significant difference was noted between the two groups for the secondary endpoint and it is anticipated that CLS001 will enter phase III clinical trials. Protegrin-1 (PG-1; IntraBiotics Pharmaceuticals, Inc.), a cysteine-rich 18 amino acid β-sheet porcine peptide, is currently in phase III clinical trials as a treatment for peritoneal infections by *P. aeruginosa*,

Table 2 Host defense peptides currently in therapeutic development

Peptide	*Company*	*Target*	*Clinical stage*
MX-226	Migenix (Vancouver, BC, Canada)	Catheter-associated infections	Phase IIIb
MX-594AN	Migenix (Vancouver, BC, Canada)	Acne vulgaris	Phase II
CLS001	Migenix (Vancouver, BC, Canada)	Rosacea	Phase II
PG-1	IntraBiotics (Mountain View, CA, USA)	Peritoneal infections	Phase III
IB 367	IntraBiotics (Mountain View, CA, USA)	Chronic respiratory infections	Phase II
Plectasin (fungal defensin)	Novozymes A/S (Bagsvaerd, Denmark)	Systemic (anti-Gram-positive) infections	Preclinical
Neuprex (rBP121)	XOMA (Berkeley, CA, USA)	Meningococcaemia and Crohn's Disease; stem cell transplants	Phase II/III; Phase I/II

S. aureus, and MRSA.[256] PG-1 has been demonstrated to be efficacious in murine infection models against *P. aeruginosa* and *S. aureus* when administered i.p. or i.v.[130] Interestingly, IntraBiotics Pharmaceuticals had previously developed Iseganan (IB 367) as a rinse for the treatment and prevention of oral mucositis in high-risk patients; however, the product failed two phase III clinical trials.[257] IB 367 is currently in phase II clinical trials as an aerosolized treatment for chronic respiratory infections in patients with cystic fibrosis.[257] Plectasin, a fungal defensin peptide, has entered preclinical phase studies. Previously, plectasin was efficacious in the treatment of systemic infections; plectasin was comparable to vancomycin in mouse peritonitis model and to penicillin in a pneumonia model.[258] Plectasin also appears to be tolerated at high doses with tolerated doses exceeding 125 mg kg^{-1} and remained active following 24 h incubation in 90% serum.[258] P-113, a 12 amino acid derivative of histatin 5, is currently in phase I/II clinical trials as a mouth rinse for treatment of plaques and gingivitis.[256] P-113 (Dermegen) has demonstrated anticandidal activity comparable to the parent peptide histatin 5 and has activity against gingivitis and plaques with minimal associated toxicity in an experimental human model.[259] Pexiganan (MSI-78; Genaera, PA, USA), a C-terminally modified magainin 2 derivative, has been investigated as a topical antimicrobial therapy for mild-to-moderate diabetic foot ulcer infections.[1] Although it was demonstrated to have equivalent wound-healing capabilities and clinical outcomes to ofloxacin, a fluoroquinolone antibiotic, and limited associated toxicities, it was rejected in 1999 by the FDA as pexiganan did not offer any additional benefit from ofloxacin.[260,261] Neuprex (rBPI21; XOMA, Berkeley, CA, USA), a 21 kDa recombinant form of bactericidal/permeability-increasing (BPI) protein, has demonstrated endotoxin neutralizing activities and has reached phase II/III clinical trials as a therapeutic for meningococcaemia and Crohn's Disease.[256] In 2007 XOMA announced the initiation of phase I/II clinical trials for the use of Neuprex in patients undergoing allogeneic hematopoietic stem cell transplants (http://www.xoma.com, XOMA). Interestingly, XOMA has also investigated a retro-inverso (RI) 9-amino acid derivative (with the substitution of two residues with napthyl-alanine) of BPI termed XMP.629 as a topical therapeutic for acne.[257] Although the peptide had potent antimicrobial activity against *P. acnes* it failed in a phase II clinical trial (http://www.xoma.com, XOMA).

Potential therapeutic applications of host defense peptides also include the lantibiotic nisin. Indeed, nisin has had an impressive history as a food preservative with FDA approval in 1988 for use in pasteurized, processed cheese spreads.[41] The attractiveness of nisin as a potential therapeutic is also enhanced due to its relative resistance to proteases and broad spectrum Gram-positive antimicrobial activity including multidrug-resistant strains.[3,41] Biosynexus Inc. has licensed the use of nisin for human clinical applications and Immucell Corp. has licensed the use of Mast Out, an antimastitic nisin-containing product, to Pfizer Animal Health.[41] Indeed, nisin formulations have been used as an active agent in the topical therapies Mast Out and Wipe-Out for bovine mastitis, an inflammatory disorder of the udder that is the most persistent disease in dairy cows.[41]

To date only two nonbacterially produced peptides, Omiganan and Pexiganan, have demonstrated efficacy in phase III clinical trials; only four cationic peptides have advanced into phase III clinical trials (Pexiganan, Iseganan, Neuprex, and Omiganan).[3] Although research and development of host defense peptides as therapeutics have been limited, there is an ongoing effort to produce new peptide candidates for a variety of clinical applications.

Thus far, host defense peptide clinical trials have focused on topical applications to address surface infection and have not addressed the concerns associated with peptide cytotoxicities upon systemic administration. As such, there is limited information regarding these concerns. A phase I study of the protegrin analogue IB 367 demonstrated no clinically significant adverse effects at peptide concentrations that reduced oral microflora.[262] Structure–activity studies of host defense peptide cytotoxicities have focused on strategies to minimize systemic toxicity concerns through the evaluation of peptide-mediated red blood cell hemolysis. The recent demonstrations that host defense peptides share features with eukaryotic nuclear localization signal peptides[215] and can translocate freely into cells[216] argue for more focus on subtle toxicities.

Therapeutic applications of host defense peptides have also been limited by the high costs of manufacturing associated with peptide synthesis. The clinical application of peptides cost in the range of \$100 and \$600 g^{-1} (the average daily dose for most systemic therapeutics) and contrasts sharply with the low cost of conventional antibiotics (aminoglycosides cost \$0.80 g^{-1}).[255] Thus, large-scale peptide production platforms, such as recombinant DNA methodologies, are being sought for the development and testing of large numbers of peptide derivatives. Novozyme Inc. has reported the large-scale production of the fungal peptide plectasin at the scale

and purity necessary for therapeutic administrations through the use of a proprietary fungal expression system.[258] In addition to this, recent investigations have focused on the reduction of host defense peptide size and decreased peptide degradation thus reducing dose quantities and frequency.[263]

The lability of host defense peptides to proteases and peptidases has been investigated as a method for increasing the therapeutic efficacy of these molecules. The high concentration of basic residues in host defense peptides increases the lability of these molecules to endogenous trypsin-like enzymes and represents a significant therapeutic impediment. As host defense peptides are significantly sensitive to proteases, with half-lives of minutes *in vivo*,[183] peptides that are protease resistant while retaining activity would offer a potential resolution to this problem.

Enatiomeric peptides comprised of all-D-amino acids and RI peptides have also been investigated as potential mechanisms to increase peptide stability. Ideally, these modifications decrease the lability of peptides to degradation while retaining the activities of the natural peptide. Investigations by Hamamoto *et al.*[264] and Chen *et al.*[67] have demonstrated that enantiomeric peptides meet these criteria and display negligible cytotoxicity. As enantiomeric peptides retain the antimicrobial acitivities of their peptide parents it is anticipated that the interactions between peptides and the bacterial membrane are not stereoselective. The exploitation of enatiomeric peptides for immunomodulatory purposes would be anticipated to be limited as these activities appear to be mediated through stereoselective receptor-mediated interactions.[158] RI modifications, which maintain amino acid side chain topology while bestowing protease resistance, could be of potential use in host defense peptide immunomodulatory applications.

RI peptides are directional isomers of natural peptides in which the residues are assembled in reverse order from that of the natural peptide and have inverted stereochemistry at all alpha carbon stereocenters.[265] As such, the overall side-chain topology of the RI analogue is analogous to that of the natural peptide and would be anticipated to accommodate the binding of an RI-peptide within a binding site (**Figure 4**). RI peptides do not universally mimic native peptides; however, the reversal of the direction of the peptide bond results in inversion of the backbone amide bonds thus shifting hydrogen bond patterns from that of the parent peptide.[266–270] The derivatization of natural peptide hormones to their RI analogues has been demonstrated as not well tolerated.[265] For example, investigations of RI-modified osteogenic growth peptide, which promotes bone anabolism and hematopoieses, demonstrated that the RI-modified peptide lacked biological activity.[271]

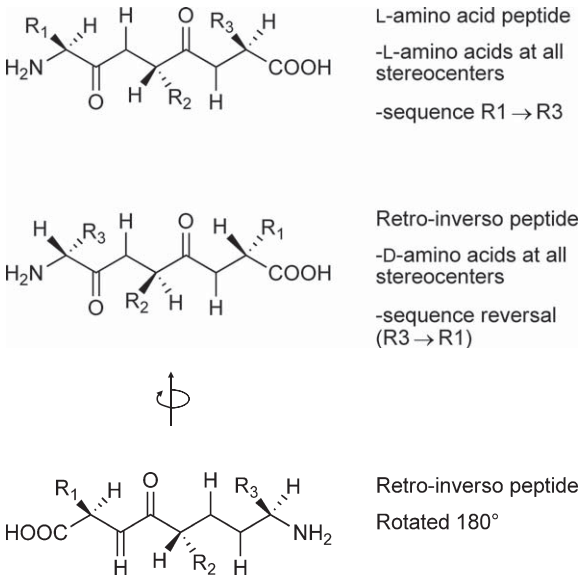


Figure 4 Retro-inversion of host defense peptides. Synthesis of RI peptides is achieved by substituting D-amino acids at all stereocenters within a peptide and reversal of peptide sequence (R1 → R3 in the L-peptide and R3 → R1 in the RI-peptide). By rotating the RI-peptide at 180° it can be seen that the three-dimensional space occupied by the amino acid functional (R) groups is retained in comparison to the L-peptide although the peptide backbone has been reversed.

In contrast, Snyder[272] has demonstrated that an RI derivative of p53 can restore endogenous p53 activity; the RI derivative induced apoptosis by activation of endogenous p53 and by restoration of function to several p53 DNA contact mutants. RI-modified peptides have also been investigated as potential alternatives for host defense peptides. Indeed, a RI derivative of indolicidin retained the antimicrobial and antiendotoxic acitivities of the natural peptide.[273] As the antiendotoxin activities of the RI-derivative were conserved these results might indicate the conservation of multiple immunomodulatory activities upon retro-inversion, highlighting the potential of this modification for therapeutic applications.

In addition to retro-inversion as a tool for increasing the therapeutic potential of host defense peptides recent investigations have highlighted the use of branched host defense peptides in such applications. Branched host defense peptide derivatives have been explored as a method for increasing the half-life of host defense peptides.[274,275] These molecules are constructed by the addition of peptide sequences onto radially branched lysine residues stemming from a peptidyl core.[276,277] Pini *et al.*[166] have constructed a tetrabranched antimicrobial peptide, M6, that had low MICs for several clinically relevant pathogenic Gram-negative microbes, remained stable in blood and serum for greater than 24 h (as compared to approximately 2 h for natural peptides) and directly bound *E. coli* LPS *in vitro*. The authors also demonstrated that M6 did not potentiate a detectable antibody response in mice and had an LD50 of approximately 125 mg kg^{-1} when administered nonsystemically and of 37.5 mg kg^{-1} when administered systemically.

The previous examples highlight some of the present technologies being applied for construction of host defense peptides with increased stability and retention of natural peptide activity. Similarly, the examination of host defense peptide candidates has also focused on the identification of methodologies that increase the cost effectiveness of present screening techniques and are also able to examine peptides in a high-throughput fashion.

5.07.14 Strategies for Host Defense Peptide Selection and Optimization

One of the largest impediments to the use of host defense peptides in clinical applications is directly related to their prohibitive costs, potential toxicities, and lack of high-throughput screening analyses. The expense of HDP therapies reflects both high costs of production as well as large dose quantities necessitated by the biological instability of HDPs. These higher dose quantities and frequencies promote dose-dependent toxicity.

5.07.14.1 Single-Position Substitution Analysis and High-Throughput Screening Analysis

The sequential modification of host defense peptides through single-position substitution analysis has been employed for identifying particular sequence patterns that contribute to superior activities. For example, Nagaoka *et al.*[278] have used substitution analysis of cationic and hydrophobic residues within LL-37 to identify particular residues that contribute to increased antimicrobial activity. This analysis generated peptides with increased activity against a broad range of Gram-positive and Gram-negative organisms. Substitution analyses have also been utilized for the identification of specific residues that contribute to therapeutic indices of host defense peptides. Lee *et al.*[86] identified specific L- and D-amino acid substitutions within GS14, a cyclic 14-residue derivative of Gramicidin S, which increased or decreased the overall therapeutic index of the parent peptide.

As synthesis of many peptide derivatives is time consuming and expensive, investigators have turned to alternative methodologies to identify improved peptide derivatives of parent peptide molecules through substitution analysis. Peptide array technology has proven an economical, high-throughput strategy for creation of large libraries of peptide derivatives. Using SPOT synthesis on cellulose supports, it is possible to create peptides of up to 50 amino acids per spot and approximately 8000 spots per cellulose sheet.[263] These peptides are not limited to gene-encoded amino acids and may be cleaved from the cellulose sheet for studies of soluble peptides or left tethered to the support. An investigation by Hilpert *et al.*[263] applied SPOT synthesis to create a library of peptide derivatives from the peptide Bac2A in the search for peptides with optimized direct antimicrobial activity. The authors created a library of 228 peptide derivatives and for antimicrobial activity screening employed a *P. aeruginosa* strain that constitutively expresses a *luxCDABE* (luciferase) gene cassette.

The antimicrobial activity of the peptide library could be quantified based on light production; peptides that are microbicidal ceases ATP production and will thus result in a quantifiable decrease in light production. Thus, this assay allows for the high-throughput screening of large numbers of peptide derivatives while minimizing the total quantity of peptide required for testing. From this screen it was found that approximately 50% of all peptide variants had improved or equivalent activity to the parent peptide with Cys, Trp, Arg, His, and Lys representing preferred substitutions. This approach provides an economical alternative for creation and testing of novel synthetic peptide derivatives for substitution analysis. It is anticipated that an analogous screening technique to the *P. aeruginosa* lux strain for measuring host cell peptide responses would provide a methodology for the evaluation of immunomodulatory peptide activities on a high-throughput scale.

5.07.14.2 *In Silico* Screening of Host Defense Peptides through Virtual Screening and Computational Methods

The utilization of *in silico* methods focusing on the rapid development of potentially superior peptide derivatives has been sought as a predictive alternative to normal prediction and synthesis procedures. Theoretically, the mutation of each residue of a 20 amino acid peptide with each of the 20 conventional amino acids would give rise to 20^{20} different combinations for screening, an economically insurmountable task. Thus, alternative methodologies have been employed that can evaluate large libraries of compounds while reducing the manufacturing costs and analysis time associated with present screening techniques. As such, virtual screening methodologies allow for the *in silico* screening of large compound libraries in a less expensive fashion. Thus far, virtual screening of host defense peptides has been limited to α-helical peptides. Generally, these studies follow a specific approach. First, physicochemical variables that are derived from previously reported relevant structural and functional information are generated. Second, a predictive model is constructed based on the relationship between the physicochemical parameters and biological data. Third, host defense peptides with a broad range of activities are run through the model and the predicted peptides with the most potent activities are evaluated experimentally.[279] Recently, quantitative structure–activity relationship (QSAR) studies have been employed for prediction of novel active host defense peptides.

QSAR methodology utilizes the structural and functional information gathered from host defense peptide investigations and examines this through computational and mathematical modeling.[280] Through the structural/functional peptide information unique physicochemical parameters, or descriptors, are defined and the specific nature of these descriptors dictates the quality and reliability of the QSAR model.[279] The most common descriptors are defined by information on hydrogen bonding field, molecular electrostatic field and partial surface area, topological, geometrical, or electrostatic information. A primary advantage of QSAR models for host defense peptides is that only primary structural information is needed rather than defined molecular structures.[279] In the case of host defense peptide QSAR modeling, the development of theoretical amino acid 'z scores' based on principal component analysis of 29 experimental descriptors for each amino acid by Hellberg *et al.*[281] has been particularly important. Thus, peptides may be described by z scores correlating to their biological activities for inclusion in the QSAR model. Although this has proven to be an advantageous tool for modeling host defense peptides, it is limited by the restriction for peptide of the same length as the z scores are directly related to peptide length and the complexity of the property model.[279] Lejon *et al.*[282] have used QSAR modeling to predict novel pentadecapeptides with improved antimicrobial activity based on data derived from Lfcins. More recently, the combination of neural network modeling for prediction of molecular properties with a QSAR model and genetic algorithms has been used for the prediction of optimized peptides. Indeed, 90 out of 100 synthetic peptides in the final population were found to be acceptable.[283] Similarly, Jenssen *et al.*[280] recently described the use of novel descriptors of contact energies between residue neighbors to correctly predict active antimicrobial peptides with 84% accuracy. An investigation of the 18 amino acid α-helical peptide Novispirin G10 was used to predict mutations that would increase antimicrobial activity. In the QSAR model all possible single residue mutants were tested and structural modeling and molecular dynamics optimization of all 360 mutants was performed. On the basis of this information, 16 analogues were created, 11 of which demonstrated increased antimicrobial activity.[279]

Indeed, the growing interest in host defense peptides as potential anti-infective therapies has potentiated the use of *in silico* techniques for the discovery or modeling of these peptides.[284] AMSDb (http://www.bbcm.units.it,

Department of Biochemistry, Biophysics and Macromolecular Chemistry; Antimicrobial Peptides Laboratory), a host defense peptide database, has been established for the compilation of a generalized set of peptide sequences. As of 2004, this set houses over 890 sequences of host defense peptides as both mature functional peptides and as prosequences, and encompasses all major host defense peptide classes. Unfortunately, the use of sequence analysis strategies for the discovery of novel host defense peptides has been limited to those peptides containing cysteine residues. Fjell *et al.*[284] have used profile-based hidden Markov models (HMM) in combination with sequence clustering and protein structure annotation in an attempt to aid the discovery of novel host defense peptides from unannotated sequences. As a result of this analysis, the authors have created a new database and automated discovery tool, AMPer (http://www.cnbi2.com, CiteULike), for investigating host defense peptide sequence diversity and as well for the discovery of novel gene-encoded peptide candidates. Utilities such as the HMMER HMM files for the prediction and classification of host defense peptides have been made available online for the scanning of sequences by the user. Fjell *et al.*[284] also note that AMPer offers the categorization of the submitted sequences and the ability to download the HMM models used in the analysis. Additional tools offered by AMPer include web pages for viewing the peptides and the corresponding properties such as peptide length, charge, hydrophobicity, as well as the consensus peptide sequence. As AMPer does not contain information on prokaryotic host defense peptides, a database has also been developed for these bacterially produced antimicrobial peptides (bacteriocins). BACTIBASE[285] (http://bactibase.pfba-lab.org, BACTIBASE) is a freely available database containing the calculated or predicted physicochemical properties of 123 bacteriocins from both Gram-positive and Gram-negative organisms. This database has been constructed to enable the prediction of structure–function relationships within the bacteriocins as well as target organisms as the current bacteriocin genome mining tool BAGEL[286] (http://bioinformatics.biol.rug.nl, BIOSUPPLYNET) is limited to the detection of putative bacteriocin gene clusters and does not provide information regarding the application of physicochemical properties.[287]

5.07.14.3 Sequence Scrambling of Host Defense Peptides

Investigations of host defense peptide structure–functions have also focused on methods to elucidate the relationship between activity and residue arrangement. Sequence scrambling of peptide sequences has been used to examine the effect of alterations to residue positions within a peptide. In such cases the physicochemical characteristics such as cationicity and hydrophobicity will be maintained. If conservation of physicochemical characteristics is the only requirement for maintenance of peptide activity all scrambled peptide derivatives should have activities equal to the parent peptide. An investigation by Pag *et al.*[287] demonstrated that sequence scrambling of a synthetic amphipathic α-helical antimicrobial peptide resulted in retained antimicrobial activity. Hilpert *et al.*[288] recently utilized sequence scrambling to create 49 variants of the linear bactenecin derivative Bac2A using a nonbiased random computational methodology. The peptide derivatives fell into six activity classes ranging from superior activity to complete loss of antimicrobial activity. Derivatives that maintained antimicrobial activity deviated substantially from Bac2A demonstrating that antimicrobial activity was specific for particular arrangements or contexts of amino acid residues within a peptide rather than just their absolute presence in the sequence. Thus, sequence scrambling may represent a methodology for the rapid development of potentially optimized active peptide derivatives.

5.07.15 Conclusions

The survival of an organism relies on the rapid deployment of an immune response upon encountering an infectious organism as well as the development of long-term specific adaptive immune responses. Indeed, this defensive response must be rapid and nonspecific so as to delay potential deleterious damage to the host. The innate immune system fulfills this requirement through a variety of mechanisms and bestows a blanket response to a broad range of microbial organisms. Host defense peptides are an important component of this blanket response and additionally in the development of long-term immunity (**Figure 5**).

Host defense peptides are ubiquitous in nature and have been found in every living organism that has been examined for their presence. These short, amphipathic cationic peptides are multifaceted host defense

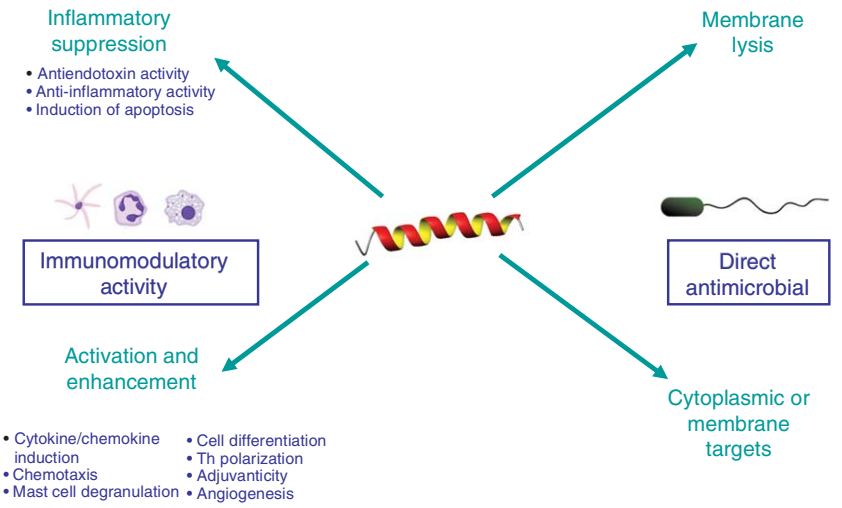


Figure 5 Direct antimicrobial and immunomodulatory activities of host defense peptides. Cationic host defense peptides exert their anti-infective activities through either direct antimicrobial activity or through modulation of the host immune response. HDP-mediated direct antimicrobial activity has been demonstrated *in vivo* for those HDPs that are either present at physiological concentrations that match their respective MIC values or are not inhibited by high salt or divalent cation concentrations. Most natural HDPs have also been demonstrated to be involved in the induction of innate and adaptive immune responses within the host as well as the selective suppression of proinflammatory responses. Ultimately, HDP-associated activities promote the resolution of infection.

molecules that offer protection from Gram-negative and Gram-positive bacteria, viruses, fungi, and other parasites.[2] Functionally, this protection is mediated through a variety of actions that include, but are not limited to, direct disruption of microbial membranes, targeting of microbial cytoplasmic biomolecules, cytokine/chemokine induction, direct chemotaxis, angiogenesis and wound repair, modulation of inflammation, and the bridging of these immediate immune responses into long-term adaptive immunity. Importantly, these activities are not mutually exclusive; thus, host defense peptides may utilize multiple activities within the same context to resolve infection. These properties, in addition to the limited resistance they promote within pathogenic bacteria, make host defense peptides a particularly enticing therapeutic alternative; however, defining the mechanism of actions of these peptides, as well as the elucidation of the structural parameters that define these characteristic activities, is central to their clinical application.

Host defense peptides were initially described as being directly antimicrobial in nature; however, the activities of these molecules have recently been extended to include immunomodulatory functions as well. From the perspective of their antimicrobial activity, host defense peptides are attractive as this activity is not confined to a single target but rather appears to work as a 'dirty drug' in which a variety of microbial biomolecules are targeted. On the basis of numerous structure–activity investigations of host defense peptide antimicrobial activity, a number of parameters have been identified that confer peptide activity, namely, peptide hydrophobicity, amphipathicity, and helicity.[8] It has been demonstrated that threshold amounts of each of these parameters allow for the optimal interaction of peptides with microbes while limiting their nonspecific targeting of host cell membranes. These parameters will help guide the rational design of synthetic host defense peptides with optimized antimicrobial activities while limiting potentially deleterious side effects to the host.

The recent elucidation of the immunomodulatory activities of host defense peptides is perhaps the most intriguing to the development of these molecules as therapeutics. Importantly, many peptides that have been demonstrated to be devoid of antimicrobial activity under physiological conditions have maintenance of immunomodulatory activities. That host defense peptides have also recently been demonstrated to bridge innate and adaptive immune responses and as well act as adjuvants extends their therapeutic potential beyond the resolution of infection.[88,158] Recent investigations have also sought to define the cell receptors involved in the immunomodulatory activities of host defense peptides in an effort to define the mechanism of action of these molecules and further the understanding of how these molecules are able to elicit such a broad range of

responses. It remains uncertain as to whether host defense peptides utilize a limited number of specific receptors for all activities or multiple and distinct receptors that initiate unique cellular responses upon binding of a distinct peptide. The recent identification that bacterial receptors such as PhoQ can bind a broad range of host defense peptides may reveal that the receptors of both microbial and host origin can bind many different host defense peptides due to the structural flexibility and physicochemical parameters associated with these molecules.

It is intriguing that host defense peptides with little to no primary structure homology are able to induce similar immunomodulatory responses; thus, a deeper understanding of the structural characteristics that define these characteristics is needed. It is anticipated that these studies will allow for the optimization and design of synthetic peptides with optimized activities. Unfortunately, the development of host defense peptides as therapeutics is limited by the high costs of manufacturing associated with these molecules and as well as the lack of convenient high-throughput methodologies for studying their activities.

Recently, host defense peptide investigations have sought to counteract these concerns through a variety of methodologies. First, *in silico* aided approaches such as QSAR have been applied to host defense peptides in an effort to define the structural parameters that bestow peptide activities.[280] Such approaches will provide researchers with optimized host defense peptide sequences that can be subsequently confirmed as compared to current optimization techniques such as single-substitution analysis. High-throughput screening methodologies such as that established by Hilpert *et al.*[263] provide conformational tools for screening peptide activities while limiting the amounts of reagent used. High-throughput technologies for immunomodulatory activities will allow for broader screening and further understanding of how host defense peptides function within the confines of the innate immune system.

That host defense peptides have been conserved across a wide range of organisms throughout evolution is a testament to the importance of these peptides as defense biomolecules. In the shadow of a looming crisis due to antibiotic resistance new therapeutics are desperately needed to aid in the battle against infectious disease. The multifaceted activities of host defense peptides make them ideal candidates for this purpose.

Abbreviations

APC	antigen-presenting cell
CCL	chemokine (C-C motif) ligand
CCR	chemokine receptor
CL	cardiolipin
CRAMP	cathelin-related antimicrobial peptide
CXCL	chemokine (C-X-C motif) ligand
DC	dendritic cell
Dha	2,3-didehyroalanine
Dhb	(*Z*)-2,3-didehydrobutyrine
DNA	deoxyribonucleic acid
EGFR	epidermal growth factor receptor
FPRL1	formyl-peptide receptor-like 1
GC	guanosine-cytosine
HBD	human beta-defensin
HIF	hypoxia-inducible factor
HIV	human-immunodeficiency virus
HMM	hidden Markov model
HNP	human neutrophil peptide
HPV	human papilloma virus
IBD	inflammatory bowel disease
iDC	immature dendritic cell
IDR-1	innate defense regulator-1
Ig	immunoglobulin
IL	interleukin
IRAK1	interleukin-1 receptor associated kinase-1

i.p.	intraperitoneal
i.v.	intravenous
LAB	lactic-acid bacteria
LPS	lipopolysaccharide
LTA	lipoteichoic acid
M_H	hydrophobic moment
MCP	monocyte chemoattractant protein
MIC	minimum inhibitory concentration
MIP-3	macrophage inflammatory protein-3
MyD88	myeloid differentiation primary response gene 88
NFκB	nuclear factor kappa B
OVA	ovalbumin
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
QSAR	quantitative structure–activity relationship
RI	retro-inverso
RNA	ribonucleic acid
SM	sphingomyelin
STAT	signal transducer and activator of transcription
TLR	Toll-like receptor
TNF-α	tumor necrosis factor-alpha
tPMP	human microbicidal protein

References

1. J. B. McPhee; R. E. Hancock, *J. Pept. Sci.* **2005**, *11* (11), 677–687.
2. H. Jenssen; P. Hamill; R. E. Hancock, *Clin. Microbiol. Rev.* **2006**, *19* (3), 491–511.
3. R. E. Hancock; H. G. Sahl, *Nat. Biotechnol.* **2006**, *24* (12), 1551–1557.
4. V. Nizet; T. Ohtake; X. Lauth; J. Trowbridge; J. Rudisill; R. A. Dorschner; V. Pestonjamasp; J. Piraino; K. Huttner; R. L. Gallo, *Nature* **2001**, *414* (6862), 454–457.
5. T. Ganz; J. A. Metcalf; J. I. Gallin; L. A. Boxer; R. I. Lehrer, *J. Clin. Invest.* **1988**, *82* (2), 552–556.
6. P. Y. Ong; T. Ohtake; C. Brandt; I. Strickland; M. Boguniewicz; T. Ganz; R. L. Gallo; D. Y. Leung, *N. Engl. J. Med.* **2002**, *347* (15), 1151–1160.
7. K. Pütsep; G. Carlsson; H. G. Boman; M. Andersson, *Lancet* **2002**, *360* (9340), 1144–1149.
8. N. Y. Yount; M. R. Yeaman, *Protein Pept. Lett.* **2005**, *12* (1), 49–67.
9. C. L. Bevins; G. Diamond, Mammalian Antimicrobial Peptides. In *Peptide Antibiotics. Discovery, Modes of Action, and Applications*; C. J. Dutton, M. A. Haxell, H. A. I. McArthur, R. G. Wax, Eds.; Marcel Dekker, Inc.: New York, 2002; pp 145–192.
10. J. L. Gifford; H. N. Hunter; H. J. Vogel, *Cell Mol. Life Sci.* **2005**, *62* (22), 2588–2598.
11. K. S. Hoek; J. M. Milne; P. A. Grieve; D. A. Dionysius; R. Smith, *Antimicrob. Agents Chemother.* **1997**, *41* (1), 54–59.
12. E. A. Nordahl; V. Rydengard; P. Nyberg; D. P. Nitsche; M. Mörgelin; M. Malmsten; L. Björck; A. Schmidtchen, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (48), 16879–16884.
13. D. Yang; Q. Chen; D. M. Hoover; P. Staley; K. D. Tucker; J. Lubkowski; J. J. Oppenheim, *J. Leukoc. Biol.* **2003**, *74* (3), 448–455.
14. A. Bjorstad; H. Fu; A. Karlsson; C. Dahlgren; J. Bylund, *Antimicrob. Agents Chemother.* **2005**, *49* (9), 3889–3895.
15. N. Y. Yount; M. T. Andrés; J. F. Fierro; M. R. Yeaman, *Biochim. Biophys. Acta* **2007**, *1768* (11), 2862–2872.
16. D. A. Devine; R. E. Hancock, *Curr. Pharm. Des.* **2002**, *8* (9), 703–714.
17. K. Radek; R. Gallo, *Semin. Immunopathol.* **2007**, *29* (1), 27–43.
18. Y. Xiao; Y. Cai; Y. R. Bommineni; S. C. Fernando; O. Prakash; S. E. Gilliland; G. Zhang, *J. Biol. Chem.* **2006**, *281* (5), 2858–2867.
19. G. Basañez; A. E. Shinnar; J. Zimmerberg, *FEBS Lett.* **2002**, *532* (1–2), 115–120.
20. A. Panyutich; J. Shi; P. L. Boutz; C. Zhao; T. Ganz, *Infect. Immun.* **1997**, *65* (3), 978–985.
21. M. Zanetti, *Curr. Issues Mol. Biol.* **2005**, *7* (2), 179–196.
22. M. Zanetti; G. Del Sal; P. Storici; C. Schneider; D. Romeo, *J. Biol. Chem.* **1993**, *268* (1), 522–526.
23. R. C. Anderson; M. Rehders; P. L. Yu, *Biotechnol. Lett.* **2008**, *30* (5), 813–818.
24. M. Zaiou; V. Nizet; R. L. Gallo, *J. Invest. Dermatol.* **2003**, *120* (5), 810–816.
25. M. Zanetti; R. Gennaro; D. Romeo, *FEBS Lett.* **1995**, *374* (1), 1–5.
26. M. Zanetti, *J. Leukoc. Biol.* **2004**, *75* (1), 39–48.

27. W. L. Zhu; K. S. Hahm; S. Y. Shin, *J. Pept. Sci.* **2007**, *13* (8), 529–535.
28. T. Ganz, *Nat. Rev. Immunol.* **2003**, *3* (9), 710–720.
29. L. Zhang; T. J. Falla, *Expert Opin. Pharmacother.* **2006**, *7* (6), 653–663.
30. R. M. Dawson; C. Q. Liu, *Crit. Rev. Microbiol.* **2008**, *34* (2), 89–107.
31. D. Yang; A. Biragyn; D. M. Hoover; J. Lubkowski; J. J. Oppenheim, *Annu. Rev. Immunol.* **2004**, *22*, 181–215.
32. D. M. Bowdish; D. J. Davidson; R. E. Hancock, *Curr. Top. Microbiol. Immunol.* **2006**, *306*, 27–66.
33. L. Liu; A. A. Roberts; T. Ganz, *J. Immunol.* **2003**, *170* (1), 575–580.
34. T. Ganz; M. E. Selsted; D. Szklarek; S. S. Harwig; K. Daher; D. F. Bainton; R. I. Lehrer, *J. Clin. Invest.* **1985**, *76* (4), 1427–1435.
35. T. Ayabe; D. P. Satchell; C. L. Wilson; W. C. Parks; M. E. Selsted; A. J. Ouellette, *Nat. Immunol.* **2000**, *1* (2), 113–118.
36. I. Cipakova; J. Gasperik; E. Hostinová, *Protein Expr. Purif.* **2006**, *45* (2), 269–274.
37. H. M. Lamb; L. R. Wiseman, *Drugs* **1998**, *56* (6), 1047–1052; discussion 1053–1054.
38. Y. Shai; A. Makovitzky; D. Avrahami, *Curr. Protein Pept. Sci.* **2006**, *7* (6), 479–486.
39. M. J. Carmona; A. Molina; J. A. Fernández; J. J. López-Fando; F. García-Olmedo, *Plant J.* **1993**, *3* (3), 457–462.
40. P. Epple; K. Apel; H. Bohlmann, *Plant Cell* **1997**, *9* (4), 509–520.
41. P. D. Cotter; C. Hill; R. P. Ross, *Nat. Rev. Microbiol.* **2005**, *3* (10), 777–788.
42. Z. Yuan; V. H. Tam, *Expert Opin. Investig. Drugs* **2008**, *17* (5), 661–668.
43. J. A. Killian, *Biochim. Biophys. Acta* **1992**, *1113* (3–4), 391–425.
44. J. E. Geraci; P. E. Hermans, *Mayo Clin. Proc.* **1983**, *58* (2), 88–91.
45. J. Lubelski; R. Rink; R. Khusainov; G. N. Moll; O. P. Kuipers, *Cell Mol. Life Sci.* **2008**, *65* (3), 455–476.
46. J. M. Willey; W. A. van der Donk, *Annu. Rev. Microbiol.* **2007**, *61*, 477–501.
47. J. R. van der Meer; J. Polman; M. M. Beerthuyzen; R. J. Siezen; O. P. Kuipers; De; W. M. Vos, *J. Bacteriol.* **1993**, *175* (9), 2578–2588.
48. L. A. Rogers, *J. Bacteriol.* **1928**, *16* (5), 321–325.
49. M. Terebiznik; R. Jagus; P. Cerrutti; M. S. de Huergo; A. M. Pilosof, *J. Food Prot.* **2002**, *65* (8), 1253–1258.
50. M. L. Cabo; L. Pastoriza; G. Sampedro; M. P. González; M. A. Murado, *J. Food Prot.* **2001**, *64* (12), 1943–1948.
51. J. Delves-Broughton; P. Blackburn; R. J. Evans; J. Hugenholtz, *Antonie Van Leeuwenhoek* **1996**, *69* (2), 193–202.
52. M. Qiao; S. Ye; O. Koponen; R. Ra; M. Usabiaga; T. Immonen; P. E. Saris, *J. Appl. Bacteriol.* **1996**, *80* (6), 626–634.
53. O. P. Kuipers; M. M. Beerthuyzen; P. G. de Ruyter; E. J. Luesink; W. M. de Vos, *J. Biol. Chem.* **1995**, *270* (45), 27299–27304.
54. P. G. de Ruyter; O. P. Kuipers; M. M. Beerthuyzen; I. van Alen-Boerrigter; W. M. de Vos, *J. Bacteriol.* **1996**, *178* (12), 3434–3439.
55. C. Kaletta; K. D. Entian, *J. Bacteriol.* **1989**, *171* (3), 1597–1601.
56. J. W. Mulders; I. J. Boerrigter; H. S. Rollema; R. J. Siezen; W. M. de Vos, *Eur. J. Biochem.* **1991**, *201* (3), 581–584.
57. W. M. de Vos; J. W. Mulders; R. J. Siezen; J. Hugenholtz; O. P. Kuipers, *Appl. Environ. Microbiol.* **1993**, *59* (1), 213–218.
58. O. McAuliffe; R. P. Ross; C. Hill, *FEMS Microbiol. Rev.* **2001**, *25* (3), 285–308.
59. E. Breukink; B. de Kruijff, *Nat. Rev. Drug Discov.* **2006**, *5* (4), 321–332.
60. S. T. Hsu; E. Breukink; E. Tischenko; M. A. Lutters; B. de Kruijff; R. Kaptein; A. M. Bonvin; N. A. van Nuland, *Nat. Struct. Mol. Biol.* **2004**, *11* (10), 963–967.
61. J. W. Payne; R. Jakes; B. S. Hartley, *Biochem. J.* **1970**, *117* (4), 757–766.
62. H. Raghuraman; A. Chattopadhyay, *Biosci. Rep.* **2007**, *27* (4–5), 189–223.
63. M. Zasloff, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84* (15), 5449–5453.
64. N. Y. Yount; M. R. Yeaman, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (19), 7363–7368.
65. D. Yang; O. Chertov; J. J. Oppenheim, *J. Leukoc. Biol.* **2001**, *69* (5), 691–697.
66. F. Miyata; T. Tokunaga; K. Yoneya; S. Yoshikawa; M. Iwanaga; T. Niwa; Y. Takao; T. Shimonishi, *J. Biochem. (Tokyo)* **1989**, *106* (4), 663–668.
67. Y. Chen; A. I. Vasil; L. Rehaume; C. T. Mant; J. L. Burns; M. L. Vasil; R. E. Hancock; R. S. Hodges, *Chem. Biol. Drug. Des.* **2006**, *67* (2), 162–173.
68. A. C. Gibbs; L. H. Kondejewski; W. Gronwald; A. M. Nip; R. S. Hodges; B. D. Sykes; D. S. Wishart, *Nat. Struct. Biol.* **1998**, *5* (4), 284–288.
69. P. Lasch; C. P. Schultz; D. Naumann, *Biophys. J.* **1998**, *75* (2), 840–852.
70. T. J. McIntosh; H. Lin; S. Li; C. Huang, *Biochim. Biophys. Acta* **2001**, *1510* (1–2), 219–230.
71. A. J. Verkleij; J. A. Post, *J. Membr. Biol.* **2000**, *178* (1), 1–10.
72. E. M. Tytler; G. M. Anantharamaiah; D. E. Walker; V. K. Mishra; M. N. Palgunachari; J. P. Segrest, *Biochemistry* **1995**, *34* (13), 4393–4401.
73. M. M. Welling; A. Lupetti; H. S. Balter; S. Lanzzeri; B. Souto; A. M. Rey; E. O. Savio; A. Paulusma-Annema; E. K. Pauwels; P. H. Nibbering, *J. Nucl. Med.* **2001**, *42* (5), 788–794.
74. A. Lupetti; A. Paulusma-Annema; M. M. Welling; S. Senesi; J. T. van Dissel; P. H. Nibbering, *Antimicrob. Agents Chemother.* **2000**, *44* (12), 3257–3263.
75. R. E. Hancock, *Lancet* **1997**, *349* (9049), 418–422.
76. A. Giangaspero; L. Sandri; A. Tossi, *Eur. J. Biochem.* **2001**, *268* (21), 5589–5600.
77. M. R. Yeaman; N. Y. Yount, *Pharmacol. Rev.* **2003**, *55* (1), 27–55.
78. N. Y. Yount; A. S. Bayer; Y. Q. Xiong; M. R. Yeaman, *Biopolymers* **2006**, *84* (5), 435–458.
79. T. Wieprecht; M. Dathe; M. Beyermann; E. Krause; W. L. Maloy; D. L. MacDonald; M. Bienert, *Biochemistry* **1997**, *36* (20), 6124–6132.
80. T. Tachi; R. F. Epand; R. M. Epand; K. Matsuzaki, *Biochemistry* **2002**, *41* (34), 10723–10731.
81. M. Dathe; J. Meyer; M. Beyermann; B. Maul; C. Hoischen; M. Bienert, *Biochim. Biophys. Acta* **2002**, *1558* (2), 171–186.
82. A. S. Bayer; R. Prasad; J. Chandra; A. Koul; M. Smriti; A. Varma; R. A. Skurray; N. Firth; M. H. Brown; S. P. Koo; M. R. Yeaman, *Infect. Immun.* **2000**, *68* (6), 3548–3553.
83. N. Uematsu; K. Matsuzaki, *Biophys. J.* **2000**, *79* (4), 2075–2083.
84. A. H. Dathe; T. Wiepricht, *Biochim. Biophys. Acta* **1999**, *1462* (1–2), 71–87.

85. M. Jelokhani-Niaraki; L. H. Kondejewski; S. W. Farmer; R. E. Hancock; C. M. Kay; R. S. Hodges, *Biochem. J.* **2000**, *349* (Pt. 3), 747–755.
86. D. L. Lee; R. S. Hodges, *Biopolymers* **2003**, *71* (1), 28–48.
87. T. I. Nicas; R. E. Hancock, *J. Bacteriol.* **1980**, *143* (2), 872–878.
88. K. A. Brogden, *Nat. Rev. Microbiol.* **2005**, *3* (3), 238–250.
89. M. S. Sansom, *Prog. Biophys. Mol. Biol.* **1991**, *55* (3), 139–235.
90. F. Y. Chen; M. T. Lee; H. W. Huang, *Biophys. J.* **2002**, *82* (2), 908–914.
91. S. Yamaguchi; T. Hong; A. Waring; R. I. Lehrer; M. Hong, *Biochemistry* **2002**, *41* (31), 9852–9862.
92. K. A. Henzler-Wildman; G. V. Martinez; M. F. Brown; A. Ramamoorthy, *Biochemistry* **2004**, *43* (26), 8459–8469.
93. J. D. Hale; R. E. Hancock, *Expert Rev. Anti Infect. Ther.* **2007**, *5* (6), 951–959.
94. S. P. Koo; A. S. Bayer; M. R. Yeaman, *Infect. Immun.* **2001**, *69* (8), 4916–4922.
95. Y. Q. Xiong; A. S. Bayer; M. R. Yeaman, *J. Infect. Dis.* **2002**, *185* (3), 348–356.
96. L. Zhang; P. Dhillon; H. Yan; S. Farmer; R. E. Hancock, *Antimicrob. Agents Chemother.* **2000**, *44* (12), 3317–3321.
97. R. Gennaro; M. Zanetti, *Biopolymers* **2000**, *55* (1), 31–49.
98. S. N. Chitnis; K. S. Prasad, *FEMS Microbiol. Lett.* **1990**, *60* (3), 281–284.
99. H. G. Boman; B. Agerberth; A. Boman, *Infect. Immun.* **1993**, *61* (7), 2978–2984.
100. G. Kragol; S. Lovas; G. Varadi; B. A. Condie; R. Hoffmann; L. Otvos, Jr., *Biochemistry* **2001**, *40* (10), 3016–3026.
101. F. J. del Castillo; I. del Castillo; F. Moreno, *J. Bacteriol.* **2001**, *183* (6), 2137–2140.
102. C. Friedrich; M. G. Scott; N. Karunaratne; H. Yan; R. E. Hancock, *Antimicrob. Agents Chemother.* **1999**, *43* (7), 1542–1548.
103. R. I. Lehrer; D. Szklarek; A. Barton; T. Ganz; K. J. Hamann; G. J. Gleich, *J. Immunol.* **1989**, *142* (12), 4428–4434.
104. V. C. Kalfa; H. P. Jia; R. A. Kunkle; P. B. McCray, Jr.; B. F. Tack; K. A. Brogden, *Antimicrob. Agents Chemother.* **2001**, *45* (11), 3256–3261.
105. C. B. Park; K. S. Yi; K. Matsuzaki; M. S. Kim; S. C. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (15), 8245–8250.
106. J. Shi; C. R. Ross; M. M. Chengappa; M. J. Sylte; D. S. McVey; F. Blecha, *Antimicrob. Agents Chemother.* **1996**, *40* (1), 115–121.
107. C. Subbalakshmi; N. Sitaram, *FEMS Microbiol. Lett.* **1998**, *160* (1), 91–96.
108. M. Dathe; H. Nikolenko; J. Meyer; M. Beyermann; M. Bienert, *FEBS Lett.* **2001**, *501* (2–3), 146–150.
109. M. Pazgier; D. M. Hoover; D. Yang; W. Lu; J. Lubkowski, *Cell Mol. Life. Sci.* **2006**, *63* (11), 1294–1313.
110. E. Kluver; S. Schulz-Maronde; S. Scheid; B. Meyer; W. G. Forssmann; K. Adermann, *Biochemistry* **2005**, *44* (28), 9804–9816.
111. D. M. Hoover; Z. Wu; K. Tucker; W. Lu; J. Lubkowski, *Antimicrob. Agents Chemother.* **2003**, *47* (9), 2804–2809.
112. M. B. Strøm; B. E. Haug; O. Rekdal; M. L. Skar; W. Stensen; J. S. Svendsen, *Biochem. Cell Biol.* **2002**, *80* (1), 65–74.
113. P. Staubitz; A. Peschel; W. F. Nieuwenhuizen; M. Otto; F. Götz; G. Jung; R. W. Jack, *J. Pept. Sci.* **2001**, *7* (10), 552–564.
114. M. P. Aliste; J. L. MacCallum; D. P. Tieleman, *Biochemistry* **2003**, *42* (30), 8976–8987.
115. C. E. Dempsey, *Biochim. Biophys. Acta* **1990**, *1031* (2), 143–161.
116. T. C. Terwilliger; D. Eisenberg, *J. Biol. Chem.* **1982**, *257* (11), 6016–6022.
117. W. F. DeGrado; G. F. Musso; M. Lieber; E. T. Kaiser; F. J. Kézdy, *Biophys. J.* **1982**, *37* (1), 329–338.
118. S. E. Blondelle; R. A. Houghten, *Biochemistry* **1991**, *30* (19), 4671–4678.
119. N. Asthana; S. P. Yadav; J. K. Ghosh, *J. Biol. Chem.* **2004**, *279* (53), 55042–55050.
120. Z. Oren; J. Hong; Y. Shai, *J. Biol. Chem.* **1997**, *272* (23), 14643–14649.
121. Y. Shai; Z. Oren, *J. Biol. Chem.* **1996**, *271* (13), 7305–7308.
122. I. Zelezetsky; A. Tossi, *Biochim. Biophys. Acta* **2006**, *1758* (9), 1436–1449.
123. S. Pacor; A. Giangaspero; M. Bacac; G. Sava; A. Tossi, *J. Antimicrob. Chemother.* **2002**, *50* (3), 339–348.
124. M. Dathe; M. Schumann; T. Wieprecht; A. Winkler; M. Beyermann; E. Krause; K. Matsuzaki; O. Murase; M. Bienert, *Biochemistry* **1996**, *35* (38), 12612–12622.
125. M. R. Nahaie; M. Goodfellow; D. E. Minnikin; V. Hájek, *J. Gen. Microbiol.* **1984**, *130* (9), 2427–2437.
126. R. K. Ernst; E. C. Yi; L. Guo; K. B. Lim; J. L. Burns; M. Hackett; S. I. Miller, *Science* **1999**, *286* (5444), 1561–1565.
127. J. B. McPhee; S. Lewenza; R. E. Hancock, *Mol. Microbiol.* **2003**, *50* (1), 205–217.
128. L. Zhang; J. Parente; S. M. Harris; D. E. Woods; R. E. Hancock; T. J. Falla, *Antimicrob. Agents Chemother.* **2005**, *49* (7), 2921–2927.
129. V. Nizet, *Curr. Issues Mol. Biol.* **2006**, *8* (1), 11–26.
130. D. A. Steinberg; M. A. Hurst; C. A. Fujii; A. H. Kung; J. F. Ho; F. C. Cheng; D. J. Loury; J. C. Fiddes, *Antimicrob. Agents Chemother.* **1997**, *41* (8), 1738–1742.
131. A. Peschel, *Trends Microbiol.* **2002**, *10* (4), 179–186.
132. R. K. Ernst; T. Guina; S. I. Miller, *Microbes Infect.* **2001**, *3* (14–15), 1327–1334.
133. V. Nizet, Antimicrobial Peptide Resistance in Human Bacterial Pathogens. In *Antimicrobial Peptides in Human Health and Disease*; R. L. Gallo, Ed.; Horizon Bioscience: Norfolk, VA, 2005; pp 277–304.
134. C. Weidenmaier; S. A. Kristian; A. Peschel, *Curr. Drug Targets* **2003**, *4* (8), 643–649.
135. S. I. Miller; R. K. Ernst; M. W. Bader, *Nat. Rev. Microbiol.* **2005**, *3* (1), 36–46.
136. A. Peschel; H. G. Sahl, *Nat. Rev. Microbiol.* **2006**, *4* (7), 529–536.
137. F. C. Tenover, *Am. J. Med.* **2006**, *119* (6 Suppl. 1), S3–S10; discussion S62–S70.
138. M. W. Bader; S. Sanowar; M. E. Daley; A. R. Schneider; U. Cho; W. Xu; R. E. Klevit; Le H. Moual; S. I. Miller, *Cell* **2005**, *122* (3), 461–472.
139. J. Kindrachuk; N. Paur; C. Reiman; E. Scruten; S. Napper, *Antimicrob. Agents Chemother.* **2007**, *51* (12), 4374–4381.
140. E. A. Groisman, *J. Bacteriol.* **2001**, *183* (6), 1835–1842.
141. M. W. Bader; W. W. Navarre; W. Shiau; H. Nikaido; J. G. Frye; M. McClelland; F. C. Fang; S. I. Miller, *Mol. Microbiol.* **2003**, *50* (1), 219–230.
142. S. I. Miller; A. M. Kukral; J. J. Mekalanos, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (13), 5054–5058.
143. M. Sieprawska-Lupa; P. Mydel; K. Krawczyk; K. Wójcik; M. Puklo; B. Lupa; P. Suder; J. Silberring; M. Reed; J. Pohl; W. Shafer; F. McAleese; T. Foster; J. Travis; J. Potempa, *Antimicrob. Agents Chemother.* **2004**, *48* (12), 4673–4679.
144. T. Guina; E. C. Yi; H. Wang; M. Hackett; S. I. Miller, *J. Bacteriol.* **2000**, *182* (14), 4077–4086.

145. J. J. Bijlsma; E. A. Groisman, *Mol. Microbiol.* **2005**, *57* (1), 85–96.
146. V. A. Norte; M. R. Stapleton; J. Green, *J. Bacteriol.* **2003**, *185* (12), 3508–3514.
147. Y. Shi; M. J. Cromie; F. F. Hsu; J. Turk; E. A. Groisman, *Mol. Microbiol.* **2004**, *53* (1), 229–241.
148. L. F. Kox; M. M. Wösten; E. A. Groisman, *Embo J.* **2000**, *19* (8), 1861–1872.
149. E. A. Groisman; C. Mouslim, *Nat. Rev. Microbiol.* **2006**, *4* (9), 705–709.
150. J. L. Bishop; B. B. Finlay, *Trends Mol. Med.* **2006**, *12* (1), 3–6.
151. E. Garcia Vescovi; F. C. Soncini; E. A. Groisman, *Cell* **1996**, *84* (1), 165–174.
152. A. Kato; H. Tanabe; R. Utsumi, *J. Bacteriol.* **1999**, *181* (17), 5516–5520.
153. F. C. Soncini; E. Garcia Vescovi; F. Solomon; E. A. Groisman, *J. Bacteriol.* **1996**, *178* (17), 5092–5099.
154. C. M. Alpuche Aranda; J. A. Swanson; W. P. Loomis; S. I. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (21), 10079–10083.
155. K. A. Christensen; J. T. Myers; J. A. Swanson, *J. Cell Sci.* **2002**, *115* (Pt. 3), 599–607.
156. N. Martin-Orozco; N. Touret; M. L. Zaharik; E. Park; R. Kopelman; S. Miller; B. B. Finlay; P. Gros; S. Grinstein, *Mol. Biol. Cell* **2006**, *17* (1), 498–510.
157. U. S. Cho; M. W. Bader; M. F. Amaya; M. E. Daley; R. E. Klevit; S. I. Miller; W. Xu, *J. Mol. Biol.* **2006**, *356* (5), 1193–1206.
158. D. M. Bowdish; D. J. Davidson; R. E. Hancock, *Curr. Protein Pept. Sci.* **2005**, *6* (1), 35–51.
159. S. Schaller-Bals; A. Schulze; R. Bals, *Am. J. Respir. Crit. Care Med.* **2002**, *165* (7), 992–995.
160. R. Bals; X. Wang; M. Zasloff; J. M. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (16), 9541–9546.
161. P. K. Singh; H. P. Jia; K. Wiles; J. Hesselberth; L. Liu; B. A. Conway; E. P. Greenberg; E. V. Valore; M. J. Welsh; T. Ganz; B. F. Tack; P. B. McCray, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (25), 14961–14966.
162. J. Harder; U. Meyer- Hoffert; L. M. Teran; L. Schwichtenberg; J. Bartels; S. Maune; J. M. Schröder, *Am. J. Respir. Cell Mol. Biol.* **2000**, *22* (6), 714–721.
163. J. Turner; Y. Cho; N. N. Dinh; A. J. Waring; R. I. Lehrer, *Antimicrob. Agents Chemother.* **1998**, *42* (9), 2206–2214.
164. D. Raimondo; G. Andreotti; N. Saint; P. Amodeo; G. Renzone; M. Sanseverino; I. Zocchi; G. Molle; A. Motta; A. Scaloni, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (18), 6309–6314.
165. P. Yadava; C. Zhang; J. Sun; J. A. Hughes, *Int. J. Antimicrob. Agents* **2006**, *28* (2), 132–137.
166. A. Pini; A. Giuliani; C. Falciani; M. Fabbrini; S. Pileri; B. Lelli; L. Bracci, *J. Pept. Sci.* **2007**, *13* (6), 393–399.
167. G. Maisetta; Di M. Luca; S. Esin; W. Florio; F. L. Brancatisano; D. Bottai; M. Campa; G. Batoni, *Peptides* **2008**, *29* (1), 1–6.
168. A. Schmidtchen; I. M. Frick; E. Andersson; H. Tapper; L. Björck, *Mol. Microbiol.* **2002**, *46* (1), 157–168.
169. B. Lopez-Garcia; W. Ubhayasekera; R. L. Gallo; J. F. Marcos, *Biochem. Biophys. Res. Commun.* **2007**, *356* (1), 107–113.
170. T. Ganz, *Infect. Immun.* **1987**, *55* (3), 568–571.
171. A. I. Maxwell; G. M. Morrison; J. R. Dorin, *Mol. Immunol.* **2003**, *40* (7), 413–421.
172. P. Kelly; M. Bajaj-Elliott; M. Katubulushi; I. Zulu; R. Poulsom; R. A. Feldman; C. L. Bevins; W. Dhaliwal, *J. Infect. Dis.* **2006**, *193* (10), 1464–1470.
173. C. L. Wilson; A. J. Ouellette; D. P. Satchell; T. Ayabe; Y. S. López-Boado; J. L. Stratman; S. J. Hultgren; L. M. Matrisian; W. C. Parks, *Science* **1999**, *286* (5437), 113–117.
174. C. B. Buck; P. M. Day; C. D. Thompson; J. Lubkowski; W. Lu; D. R. Lowy; J. T. Schiller, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (5), 1516–1521.
175. P. Y. Ong; T. Ohtake; C. Brandt; I. Strickland; M. Boguniewicz; T. Ganz; R. L. Gallo; D. Y. Leung, *N. Engl. J. Med.* **2002**, *347* (15), 1151–1160.
176. J. Wehkamp; K. Fellermann; K. R. Herrlinger; S. Baxmann; K. Schmidt; B. Schwind; M. Duchrow; C. Wohlschläger; A. C. Feller; E. F. Stange, *Eur. J. Gastroenterol. Hepatol.* **2002**, *14* (7), 745–752.
177. F. Echtermeyer; M. Streit; S. Wilcox-Adelman; S. Saoncella; F. Denhez; M. Detmar; P. Goetinck, *J. Clin. Invest.* **2001**, *107* (2), R9–R14.
178. R. L. Gallo; M. Ono; T. Povsic; C. Page; E. Eriksson; M. Klagsbrun; M. Bernfield, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (23), 11035–11039.
179. J. Li; M. Post; R. Volk; Y. Gao; M. Li; C. Metais; K. Sato; J. Tsai; W. Aird; R. D. Rosenberg; T. G. Hampton; F. Sellke; P. Carmeliet; M. Simons, *Nat. Med.* **2000**, *6* (1), 49–55.
180. R. Koczulla; G. von Degenfeld; C. Kupatt; F. Krötz; S. Zahler; T. Gloe; K. Issbrücker; P. Unterberger; M. Zaiou; C. Lebherz; A. Karl; P. Raake; A. Pfosser; P. Boekstegers; U. Welsch; P. S. Hiemstra; C. Vogelmeier; R. L. Gallo; M. Clauss; R. Bals, *J. Clin. Invest.* **2003**, *111* (11), 1665–1672.
181. C. Moser; D. J. Weiner; E. Lysenko; R. Bals; J. N. Weiser; J. M. Wilson, *Infect. Immun.* **2002**, *70* (6), 3068–3072.
182. B. Berghofer; T. Frommer; I. R. König; A. Ziegler; T. Chakraborty; G. Bein; H. Hackstein, *Clin. Exp. Allergy* **2005**, *35* (9), 1147–1154.
183. B. B. Finlay; R. E. Hancock, *Nat. Rev. Microbiol.* **2004**, *2* (6), 497–504.
184. D. M. Bowdish; D. J. Davidson; M. G. Scott; R. E. Hancock, *Antimicrob. Agents Chemother.* **2005**, *49* (5), 1727–1732.
185. R. Koczulla; G. von Degenfeld; C. Kupatt; F. Krötz; S. Zahler; T. Gloe; K. Issbrücker; P. Unterberger; M. Zaiou; C. Lebherz; A. Karl; P. Raake; A. Pfosser; P. Boekstegers; U. Welsch; P. S. Hiemstra; C. Vogelmeier; R. L. Gallo; M. Clauss; R. Bals, *J. Clin. Invest.* **2003**, *111* (11), 1665–1672.
186. S. Tokumaru; K. Sayama; Y. Shirakata; H. Komatsuzawa; K. Ouhara; Y. Hanakawa; Y. Yahata; X. Dai; M. Tohyama; H. Nagai; L. Yang; S. Higashiyama; A. Yoshimura; M. Sugai; K. Hashimoto, *J. Immunol.* **2005**, *175* (7), 4662–4668.
187. M. Carretero; M. J. Escámez; M. García; B. Duarte; A. Holguín; L. Retamosa; J. L. Jorcano; M. D. Río; F. Larcher, *J. Invest. Dermatol.* **2008**, *128* (1), 223–236.
188. J. Li; M. Raghunath; D. Tan; R. R. Lareu; Z. Chen; R. W. Beuerman, *Invest. Ophthalmol. Vis. Sci.* **2006**, *47* (9), 3811–3819.
189. J. M. Otte; I. Werner; S. Brand; A. M. Chromik; F. Schmitz; M. Kleine; W. E. Schmidt, *J. Cell Biochem.* **2008**, *104* (6), 2286–2297.
190. J. von Haussen; R. Koczulla; J. von Haussen; R. Koczulla; R. Shaykhiev; C. Herr; O. Pinkenburg; D. Reimer; R. Wiewrodt; S. Biesterfeld; A. Aigner; F. Czubayko; R. Bals, *Lung Cancer* **2008**, *59* (1), 12–23.
191. P. G. Barlow; Y. Li; T. S. Wilkinson; D. M. Bowdish; Y. E. Lau; C. Cosseau; C. Haslett; A. J. Simpson; R. E. Hancock; D. J. Davidson, *J. Leukoc. Biol.* **2006**, *80* (3), 509–520.
192. I. Nagaoka; H. Tamura; M. Hirata, *J. Immunol.* **2006**, *176* (5), 3044–3052.

193. Z. Zhang; G. Cherryholmes; J. E. Shively, *J. Leukoc. Biol.* **2008**, *84* (3), 780–788.
194. M. Durr; A. Peschel, *Infect. Immun.* **2002**, *70* (12), 6515–6517.
195. D. Yang; A. Biragyn; L. W. Kwak; J. J. Oppenheim, *Trends Immunol.* **2002**, *23* (6), 291–296.
196. F. Niyonsaba; K. Iwabuchi; A. Someya; M. Hirata; H. Matsuda; H. Ogawa; I. Nagaoka, *Immunology* **2002**, *106* (1), 20–26.
197. O. Chertov; D. F. Michiel; L. Xu; J. M. Wang; K. Tani; W. J. Murphy; D. L. Longo; D. D. Taub; J. J. Oppenheim, *J. Biol. Chem.* **1996**, *271* (6), 2935–2940.
198. Y. De; Q. Chen; A. P. Schmidt; G. M. Anderson; J. M. Wang; J. Wooters; J. J. Oppenheim; O. Chertov, *J. Exp. Med.* **2000**, *192* (7), 1069–1074.
199. O. Chertov; H. Ueda; L. L. Xu; K. Tani; W. J. Murphy; J. M. Wang; O. M. Howard; T. J. Sayers; J. J. Oppenheim, *J. Exp. Med.* **1997**, *186* (5), 739–747.
200. O. Soehnlein; A. Zernecke; E. E. Eriksson; A. G. Rothfuchs; C. T. Pham; H. Herwald; K. Bidzhekov; M. E. Rottenberg; C. Weber; L. Lindbom, *Blood* **2008**, *112* (4), 1461–1472.
201. F. Niyonsaba; A. Someya; M. Hirata; H. Ogawa; I. Nagaoka, *Eur. J. Immunol.* **2001**, *31* (4), 1066–1075.
202. H. J. Huang; C. R. Ross; F. Blecha, *J. Leukoc. Biol.* **1997**, *61* (5), 624–629.
203. D. Yang; O. Chertov; S. N. Bykovskaia; Q. Chen; M. J. Buffo; J. Shogan; M. Anderson; J. M. Schröder; J. M. Wang; O. M. Howard; J. J. Oppenheim, *Science* **1999**, *286* (5439), 525–528.
204. G. S. Tjabringa; J. Aarbiou; D. K. Ninaber; J. W. Drijfhout; O. E. Sørensen; N. Borregaard; K. F. Rabe; P. S. Hiemstra, *J. Immunol.* **2003**, *171* (12), 6690–6696.
205. D. M. Bowdish; D. J. Davidson; D. P. Speert; R. E. Hancock, *J. Immunol.* **2004**, *172* (6), 3758–3765.
206. N. Mookherjee; K. L. Brown; D. M. Bowdish; S. Doria; R. Falsafi; K. Hokamp; F. M. Roche; R. Mu; G. H. Doho; J. Pistolic; J. P. Powers; J. Bryan; F. S. Brinkman; R. E. Hancock, *J. Immunol.* **2006**, *176* (4), 2455–2464.
207. M. G. Scott; D. J. Davidson; M. R. Gold; D. Bowdish; R. E. Hancock, *J. Immunol.* **2002**, *169* (7), 3883–3891.
208. M. Pazgier; A. Prahl; D. M. Hoover; J. Lubkowski, *J. Biol. Chem.* **2007**, *282* (3), 1819–1829.
209. J. J. Oppenheim; D. Yang, *Curr. Opin. Immunol.* **2005**, *17* (4), 359–365.
210. A. Soruri; J. Grigat; U. Forssmann; J. Riggert; J. Zwirner, *Eur. J. Immunol.* **2007**, *37* (9), 2474–2486.
211. A. Elssner; M. Duncan; M. Gavrilin; M. D. Wewers, *J. Immunol.* **2004**, *172* (8), 4987–4994.
212. Y. R. Chan; R. L. Gallo, *J. Biol. Chem.* **1998**, *273* (44), 28978–28985.
213. K. Tanaka; Y. Fujimoto; M. Suzuki; Y. Suzuki; T. Ohtake; H. Saito; Y. Kohgo, *Jpn. J. Cancer Res.* **2001**, *92* (9), 959–967.
214. J. Bao; K. Sato; M. Li; Y. Gao; R. Abid; W. Aird; M. Simons; M. J. Post, *Am. J. Physiol. Heart Circ. Physiol.* **2001**, *281* (6), H2612–H2618.
215. Y. E. Lau; A. Rozek; M. G. Scott; D. L. Goosney; D. J. Davidson; R. E. Hancock, *Infect. Immun.* **2005**, *73* (1), 583–591.
216. S. Sandgren; A. Wittrup; F. Cheng; M. Jönsson; E. Eklund; S. Busch; M. Belting, *J. Biol. Chem.* **2004**, *279* (17), 17951–17956.
217. R. Lande; J. Gregorio; V. Facchinetti; B. Chatterjee; Y. H. Wang; B. Homey; W. Cao; Y. H. Wang; B. Su; F. O. Nestle; T. Zal; I. Mellman; J. M. Schröder; Y. J. Liu; M. Gilliet, *Nature* **2007**, *449* (7162), 564–569.
218. M. G. Scott; E. Dullaghan; N. Mookherjee; N. Glavas; M. Waldbrook; A. Thompson; A. Wang; K. Lee; S. Doria; P. Hamill; J. J. Yu; Y. Li; O. Donini; M. M. Guarna; B. B. Finlay; J. R. North; R. E. Hancock, *Nat. Biotechnol.* **2007**, *25* (4), 465–472.
219. N. C. Riedemann; R. F. Guo; P. A. Ward, *J. Clin. Invest.* **2003**, *112* (4), 460–467.
220. N. Mookherjee; R. E. Hancock, *Cell Mol. Life Sci.* **2007**, *64* (7–8), 922–933.
221. J. W. Larrick; M. Hirata; R. F. Balint; J. Lee; J. Zhong; S. C. Wright, *Infect. Immun.* **1995**, *63* (4), 1291–1297.
222. M. G. Scott; A. C. Vreugdenhil; W. A. Buurman; R. E. Hancock; M. R. Gold, *J. Immunol.* **2000**, *164* (2), 549–553.
223. I. Nagaoka; S. Hirota; F. Niyonsaba; M. Hirata; Y. Adachi; H. Tamura; D. Heumann, *J. Immunol.* **2001**, *167* (6), 3329–3338.
224. M. Gough; R. E. Hancock; N. M. Kelly, *Infect. Immun.* **1996**, *64* (12), 4922–4927.
225. M. G. Scott; C. M. Rosenberger; M. R. Gold; B. B. Finlay; R. E. Hancock, *J. Immunol.* **2000**, *165* (6), 3358–3365.
226. K. Kandler; R. Shaykhiev; P. Kleemann; F. Klescz; M. Lohoff; C. Vogelmeier; R. Bals, *Int. Immunol.* **2006**, *18* (12), 1729–1736.
227. A. Giacometti; O. Cirioni; R. Ghiselli; F. Mocchegiani; G. D'Amato; M. S. Del Prete; F. Orlando; W. Kamysz; J. Lukasiak; V. Saba; G. Scalise, *Gut* **2003**, *52* (6), 874–878.
228. K. Fukumoto; I. Nagaoka; A. Yamataka; H. Kobayashi; T. Yanai; Y. Kato; T. Miyano, *Pediatr. Surg. Int.* **2005**, *21* (1), 20–24.
229. D. J. Davidson; A. J. Currie; G. S. Reid; D. M. Bowdish; K. L. MacDonald; R. C. Ma; R. E. Hancock; D. P. Speert, *J. Immunol.* **2004**, *172* (2), 1146–1156.
230. J. Banchereau; F. Briere; C. Caux; J. Davoust; S. Lebecque; Y. J. Liu; B. Pulendran; K. Palucka, *Annu. Rev. Immunol.* **2000**, *18*, 767–811.
231. A. Biragyn; P. A. Ruffini; C. A. Leifer; E. Klyushnenkova; A. Shakhov; O. Chertov; A. K. Shirakawa; J. M. Farber; D. M. Segal; J. J. Oppenheim; L. W. Kwakm, *Science* **2002**, *298* (5595), 1025–1029.
232. N. Funderburg; M. M. Lederman; Z. Feng; M. G. Drage; J. Jadlowsky; C. V. Harding; A. Weinberg; S. F. Sieg, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (47), 18631–18635.
233. J. W. Lillard, Jr.; P. N. Boyaka; O. Chertov; J. J. Oppenheim; J. R. McGhee, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (2), 651–656.
234. K. A. Brogden; M. Heidari; R. E. Sacco; D. Palmquist; J. M. Guthmiller; G. K. Johnson; H. P. Jia; B. F. Tack; P. B. McCray, *Oral Microbiol. Immunol.* **2003**, *18* (2), 95–99.
235. A. Biragyn; I. M. Belyakov; Y. H. Chow; D. S. Dimitrov; J. A. Berzofsky; L. W. Kwakm, *Blood* **2002**, *100* (4), 1153–1159.
236. A. Biragyn; M. Surenhu; D. Yang; P. A. Ruffini; B. A. Haines; E. Klyushnenkova; J. J. Oppenheim; L. W. Kwak, *J. Immunol.* **2001**, *167* (11), 6644–6653.
237. L. L. An; Y. H. Yang; X. T. Ma; Y. M. Lin; G. Li; Y. H. Song; K. F. Wu, *Leuk. Res.* **2005**, *29* (5), 535–543.
238. K. Kurosaka; Q. Chen; F. Yarovinsky; J. J. Oppenheim; D. Yang, *J. Immunol.* **2005**, *174* (10), 6257–6265.
239. J. H. Fritz; S. Brunner; M. L. Birnstiel; M. Buschle; A. Gabain; F. Mattner; W. Zauner, *Vaccine* **2004**, *22* (25–26), 3274–3284.
240. Z. Wu; D. M. Hoover; D. Yang; C. Boulègue; F. Santamaria; J. J. Oppenheim; J. Lubkowski; W. Lu, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (15), 8880–8885.
241. E. J. Fernandez; E. Lolis, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 469–499.
242. K. Taylor; P. E. Barran; J. R. Dorin, *Biopolymers* **2008**, *90* (1), 1–7.

243. K. Taylor; B. McCullough; D. J. Clarke; R. J. Langley; T. Pechenick; A. Hill; D. J. Campopiano; P. E. Barran; J. R. Dorin; J. R. Govan, *Antimicrob. Agents Chemother.* **2007**, *51* (5), 1719–1724.
244. A. Szyk; Z. Wu; K. Tucker; D. Yang; W. Lu; J. Lubkowski, *Protein Sci.* **2006**, *15* (12), 2749–2760.
245. D. M. Hoover; C. Boulegue; D. Yang; J. J. Oppenheim; K. Tucker; W. Lu; J. Lubkowski, *J. Biol. Chem.* **2002**, *277* (40), 37647–37654.
246. I. Nagaoka; S. Hirota; F. Niyonsaba; M. Hirata; Y. Adachi; H. Tamura; S. Tanaka; D. Heumann, *Clin. Diagn. Lab. Immunol.* **2002**, *9* (5), 972–982.
247. C. D. Ciornei; T. Sigurdardottir; A. Schmidtchen; M. Bodelsson, *Antimicrob. Agents Chemother.* **2005**, *49* (7), 2845–2450.
248. T. Sigurdardottir; P. Andersson; M. Davoudi; M. Malmsten; A. Schmidtchen; M. Bodelsson, *Antimicrob. Agents Chemother.* **2006**, *50* (9), 2983–2989.
249. M. Murakami; B. Lopez-Garcia; M. Braff; R. A. Dorschner; R. L. Gallo, *J. Immunol.* **2004**, *172* (5), 3070–3077.
250. M. H. Braff; M. A. Hawkins; A. Di Nardo; B. Lopez-Garcia; M. D. Howell; C. Wong; K. Lin; J. E. Streib; R. Dorschner; D. Y. Leung; R. L. Gallo, *J. Immunol.* **2005**, *174* (7), 4271–4278.
251. K. Yamasaki; A. Di Nardo; A. Bardan; M. Murakami; T. Ohtake; A. Coda; R. A. Dorschner; C. Bonnart; P. Descargues; A. Hovnanian; V. B. Morhenn; R. L. Gallo, *Nat. Med.* **2007**, *13* (8), 975–980.
252. Y. Rosenfeld; H. G. Sahl; Y. Shai, *Biochemistry* **2008**, *47* (24), 6468–6478.
253. M. Zasloff, *Nature* **2002**, *415* (6870), 389–395.
254. O. Samuelsen; H. H. Haukland; H. Jenssen; M. Krämer; K. Sandvik; H. Ulvatne; L. H. Vorland, *FEBS Lett.* **2005**, *579* (16), 3421–3426.
255. A. K. Marr; W. J. Gooderham; R. E. Hancock, *Curr. Opin. Pharmacol.* **2006**, *6* (5), 468–472.
256. M. Zaiou, *J. Mol. Med.* **2007**, *85* (4), 317–329.
257. Y. J. Gordon; E. G. Romanowski; A. M. McDermott, *Curr. Eye Res.* **2005**, *30* (7), 505–515.
258. P. H. Mygind; R. L. Fischer; K. M. Schnorr; M. T. Hansen; C. P. Sönksen; S. Ludvigsen; D. Raventós; S. Buskov; B. Christensen; L. De Maria; O. Taboureau; D. Yaver; S. G. Elvig-Jørgensen; M. V. Sørensen; B. E. Christensen; S. Kjaerulff; N. Frimodt-Moller; R. I. Lehrer; M. Zasloff; H. H. Kristensen, *Nature* **2005**, *437* (7061), 975–980.
259. D. W. Paquette; D. M. Simpson; P. Friden; V. Braman; R. C. Williams, *J. Clin. Periodontol.* **2002**, *29* (12), 1051–1058.
260. H. M. Lamb; L. R. Wiseman, *Drugs* **1998**, *56* (6), 1047–1052.
261. H. M. Lamb; L. R. Wiseman, *Drugs* **1998**, *56* (6), 1053–1054.
262. D. A. Mosca; M. A. Hurst; W. So; B. S. Viajar; C. A. Fujii; T. J. Falla, *Antimicrob. Agents Chemother.* **2000**, *44* (7), 1803–1808.
263. K. Hilpert; R. Volkmer-Engert; T. Walter; R. E. Hancock, *Nat. Biotechnol.* **2005**, *23* (8), 1008–1012.
264. K. Hamamoto; Y. Kida; Y. Zhang; T. Shimizu; K. Kuwano, *Microbiol. Immunol.* **2002**, *46* (11), 741–749.
265. P. M. Fischer, *Curr. Protein Pept. Sci.* **2003**, *4* (5), 339–356.
266. J. Howl; Z. Prochazka; M. Wheatley; J. Slaninová, *Br. J. Pharmacol.* **1999**, *128* (3), 647–652.
267. E. V. Apletalina; M. A. Juliano; L. Juliano; I. Lindberg, *Biochem. Biophys. Res. Commun.* **2000**, *267* (3), 940–942.
268. N. Beglova; S. Maliartchouk; I. Ekiel; M. C. Zaccaro; H. U. Saragovi; K. Gehring, *J. Med. Chem.* **2000**, *43* (19), 3530–3540.
269. A. Sahu; A. M. Soulika; D. Morikis; L. Spruce; W. T. Moore; J. D. Lambris, *J. Immunol.* **2000**, *165* (5), 2491–2499.
270. F. J. Warner; P. Mack; A. Comis; R. C. Miller; E. Burcher, *Biochem. Pharmacol.* **2001**, *61* (1), 55–60.
271. I. Bab; M. Chorev, *Biopolymers* **2002**, *66* (1), 33–48.
272. E. L. Snyder; B. R. Meade; C. C. Saenz; S. F. Dowdy, *PLoS Biol.* **2004**, *2* (2), E36.
273. S. Nagpal; K. J. Kaur; D. Jain; D. M. Salunke, *Protein Sci.* **2002**, *11* (9), 2158–2167.
274. A. Pini; A. Giuliani; C. Falciani; M. Fabbrini; S. Pileri; B. Lelli; L. Bracci, *J. Pept. Sci.* **2007**, *13* (6), 393–399.
275. C. Falciani; L. Lozzi; A. Pini; F. Corti; M. Fabbrini; A. Bernini; B. Lelli; N. Niccolai; L. Bracci, *Chem. Biol. Drug Des.* **2007**, *69* (3), 216–221.
276. L. Bracci; C. Falciani; B. Lelli; L. Lozzi; Y. Runci; A. Pini; De M. G. Montis; A. Tagliamonte; P. Neri, *J. Biol. Chem.* **2003**, *278* (47), 46590–46595.
277. L. Bracci; L. Lozzi; A. Pini; B. Lelli; C. Falciani; N. Niccolai; A. Bernini; A. Spreafico; P. Soldani; P. Neri, *Biochemistry* **2002**, *41* (32), 10194–10199.
278. I. Nagaoka; K. Kuwahara-Arai; H. Tamura; K. Hiramatsu; M. Hirata, *Inflamm. Res.* **2005**, *54* (2), 66–73.
279. D. Raventos; O. Taboureau; P. H. Mygind; J. D. Nielsen; C. P. Sonksen; H. H. Kristensen, *Comb. Chem. High Throughput Screen.* **2005**, *8* (3), 219–233.
280. H. Jenssen; T. Lejon; K. Hilpert; C. D. Fjell; A. Cherkasov; R. E. Hancock, *Chem. Biol. Drug Des.* **2007**, *70* (2), 134–142.
281. S. Hellberg; M. Sjostrom; B. Skagerberg; S. Wold, *J. Med. Chem.* **1987**, *30* (7), 1126–1135.
282. T. Lejon; M. B. Strom; J. S. Svendsen, *J. Pept. Sci.* **2001**, *7* (2), 74–81.
283. S. Patel; I. P. Stott; M. Bhakoo; P. Elliott, *J. Comput. Aided Mol. Des.* **1998**, *12* (6), 543–556.
284. C. D. Fjell; R. E. Hancock; A. Cherkasov, *Bioinformatics* **2007**, *23* (9), 1148–1155.
285. R. Hammami; A. Zouhir; J. Ben Hamida; I. Fliss, *BMC Microbiol.* **2007**, *7*, 89.
286. A. de Jong; S. A. van Hijum; J. J. Bijlsma; J. Kok; O. P. Kuipers, *Nucleic Acids Res.* **2006**, *34* (Web Server issue), W273–W279.
287. U. Pag; M. Oedenkoven; N. Papo; Z. Oren; Y. Shai; H. G. Sahl, *J. Antimicrob. Chemother.* **2004**, *53* (2), 230–239.
288. K. Hilpert; M. R. Elliott; R. Volkmer-Engert; P. Henklein; O. Donini; Q. Zhou; D. F. Winkler; R. E. Hancock, *Chem. Biol.* **2006**, *13* (10), 1101–1107.

Biographical Sketches

Jason Kindrachuk is a postdoctoral fellow at the University of British Columbia (UBC) in the laboratory of Professor R. E. W. Hancock. Jason received his Ph.D. from the University of Saskatchewan in 2007 where his research focused on host and pathogen sensory systems. During his study he specially focused on TLR-9 receptor–ligand interactions and the interactions between host defense peptides and the PhoPQ two-component sensory system of *Salmonella typhimurium*. In 2008 Jason received the Canadian Cystic Fibrosis Foundation Kin Canada Fellowship for his research in the area of alternative therapies for treatment of antibiotic- and multidrug-resistant bacteria. Currently his research is focused on the investigation of structure–activity relationships amongst natural and synthetic host defense peptides from the perspective of associated immunomodulatory activities and as well as vaccine formulation strategies.

Anastasia Nijnik is a postdoctoral research fellow in the group of Professor R. E. W. Hancock, at the University of British Columbia. She was originally from Ukraine and was trained at Oxford University in the United Kingdom. During her Ph.D. she worked under the supervision of Professor Richard Cornall on a Wellcome Trust funded ENU mutagenesis program, studying genes involved in lymphocyte development, autoimmunity, and stem cell function. She joined UBC at the end of 2006, and is now funded by postdoctoral fellowships from CIHR and the Michael Smith Foundation. Her current work is in the Pathogenomics of Innate Immunity and mechanisms of action of immunomodulatory host defense peptides.

R. E. W. Hancock is a professor of Microbiology & Immunology at UBC and holds the Canada Research Chair. His research interests include antibiotic uptake and resistance, the functional genomics of innate immunity and of the prominent opportunistic pathogen *Pseudomonas aeruginosa*, and the development of small cationic peptides as novel antimicrobials and modulators of innate immunity. He has published more than 430 papers and reviews, is one of the Institute for Scientific Information's highly cited authors in Microbiology with more than 15 000 citations, and has 25 patents awarded. He has won many awards, including the 2003 Aventis Pharmaceuticals Award, the leading award worldwide for antimicrobial research, and Canada's three top awards for health research, the 2005 McLaughlin Medal of the Royal Society of Canada, 2006 CIHR Michael Smith Prize as Health Researcher of the Year and the 2007 Killam Prize in Health Sciences. In 2001 he was inducted as an Officer of the Order of Canada. He is a cofounder of Migenix Inc, which has taken product concepts from his lab into Phase IIIb clinical trails for the prevention of catheter-associated infections, of Inimex Pharmaceuticals Inc, which is developing a novel approach to treating infections through selective boosting of innate immunity, and of the BC Centre for Drug Research and Development (CDRD). Bob is the director of a large multinational Genome Canada-funded grant titled the Pathogenomics of Innate Immunity and a key participant on two Grand Challenges in Global Health Research grants aimed at delivering antimicrobial therapies and effective vaccines to developing countries.

5.08 Biosynthesis and Mode of Action of Lantibiotics

Lisa E. Cooper, Bo Li, and Wilfred A. van der Donk, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

5.08.1 Introduction

Lantibiotics are peptide-derived antimicrobial agents that are ribosomally synthesized and posttranslationally modified to their biologically active forms. Their name was introduced in 1988 as an abbreviation for lanthionine-containing antibiotic peptides.[1] Lanthionines consist of two alanine residues that are linked at their β-carbons by a thioether bridge (**Figure 1**). In lantibiotics, these lanthionines are imbedded within cyclic peptides. All lantibiotics that have been characterized with respect to the stereochemistry of the thioether linkage contain (2*S*,6*R*)-lanthionines (Lan), with many family members also containing (2*S*,3*S*,6*R*)-3-methyllanthionines (MeLan; **Figure 1**). In addition, they typically (but not always) contain the unsaturated amino acids 2,3-dehydroalanine (Dha) and (*Z*)-2,3-dehydrobutyrine (Dhb). In all, no less than 15 different posttranslational modifications have been identified in lantibiotics (**Figure 1**),[2,3] and it is likely that other modifications remain to be discovered. These modifications release the peptides from the structural and functional constraints typically imposed on naturally occurring ribosomal peptides.

Lantibiotics are produced by both low and high G + C Gram-positive bacteria including the lactic acid bacteria (LAB), *Bacillus*, *Enterococcus*, *Micrococcus*, *Streptococcus*, *Staphylococcus*, and actinomycetes, and recent genome database searches show that their gene clusters are also present in select Gram-negative bacteria. Nisin, the most-studied lantibiotic, is produced by *Lactococcus lactis* and has been used extensively as a preservative against foodborne pathogens without substantial development of bacterial resistance.[4] Nisin was discovered in 1928,[5,6] and is one of the oldest known antibacterial agents; however, its structure was not determined until 1971 (**Figure 2**).[7] In 1969, nisin was conferred generally recognized as safe (GRAS) status by the World Health Organization. Since then, a steady stream of lantibiotics has been reported, each varying in size, structure, and biological activity, and with the advent of complete genome sequences, it has become clear that the lantibiotic

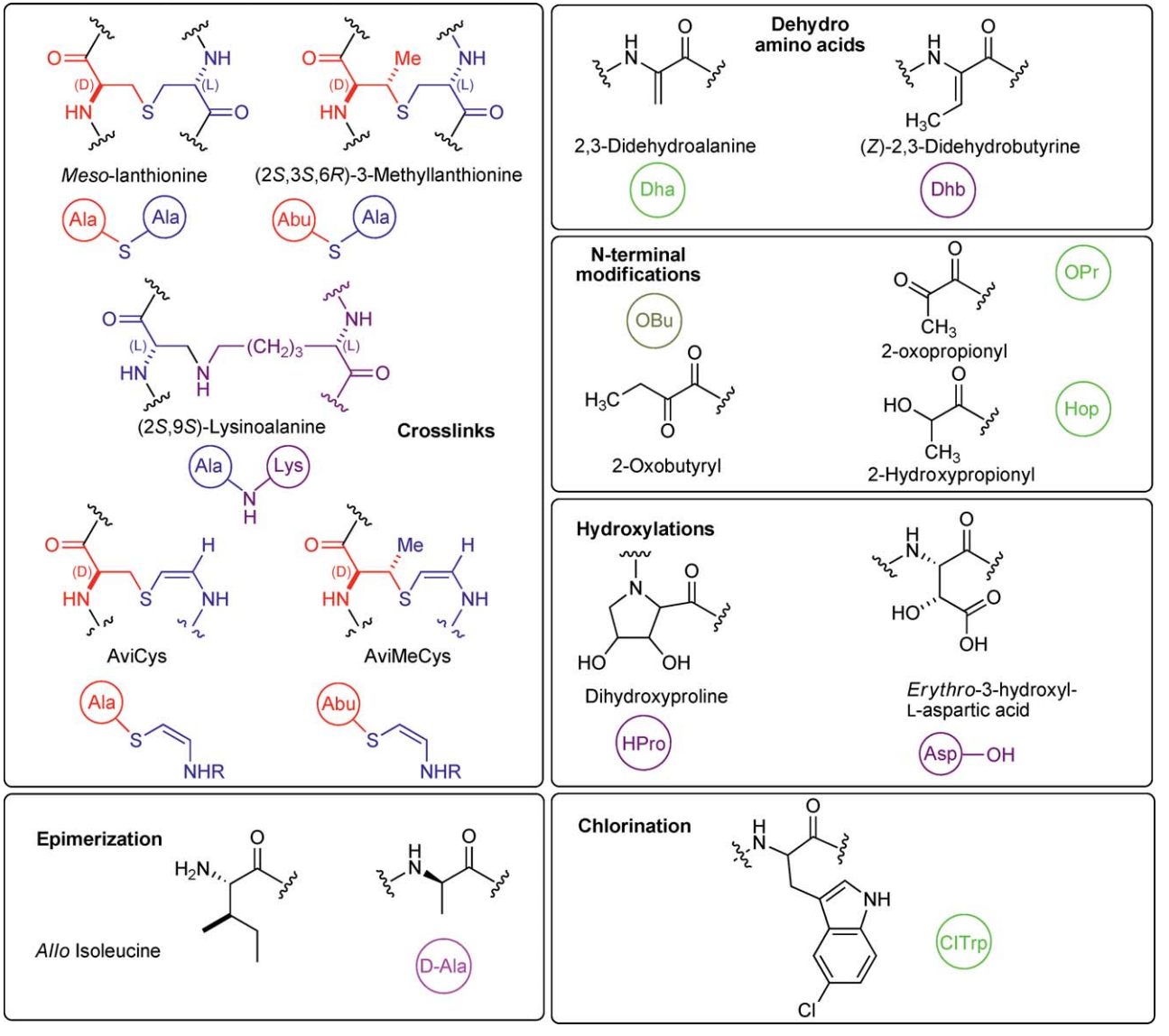


Figure 1 Posttranslational modifications that have been reported to date for the lantibiotic family. The shorthand notation that will be used in other figures in this review is listed below each structure.

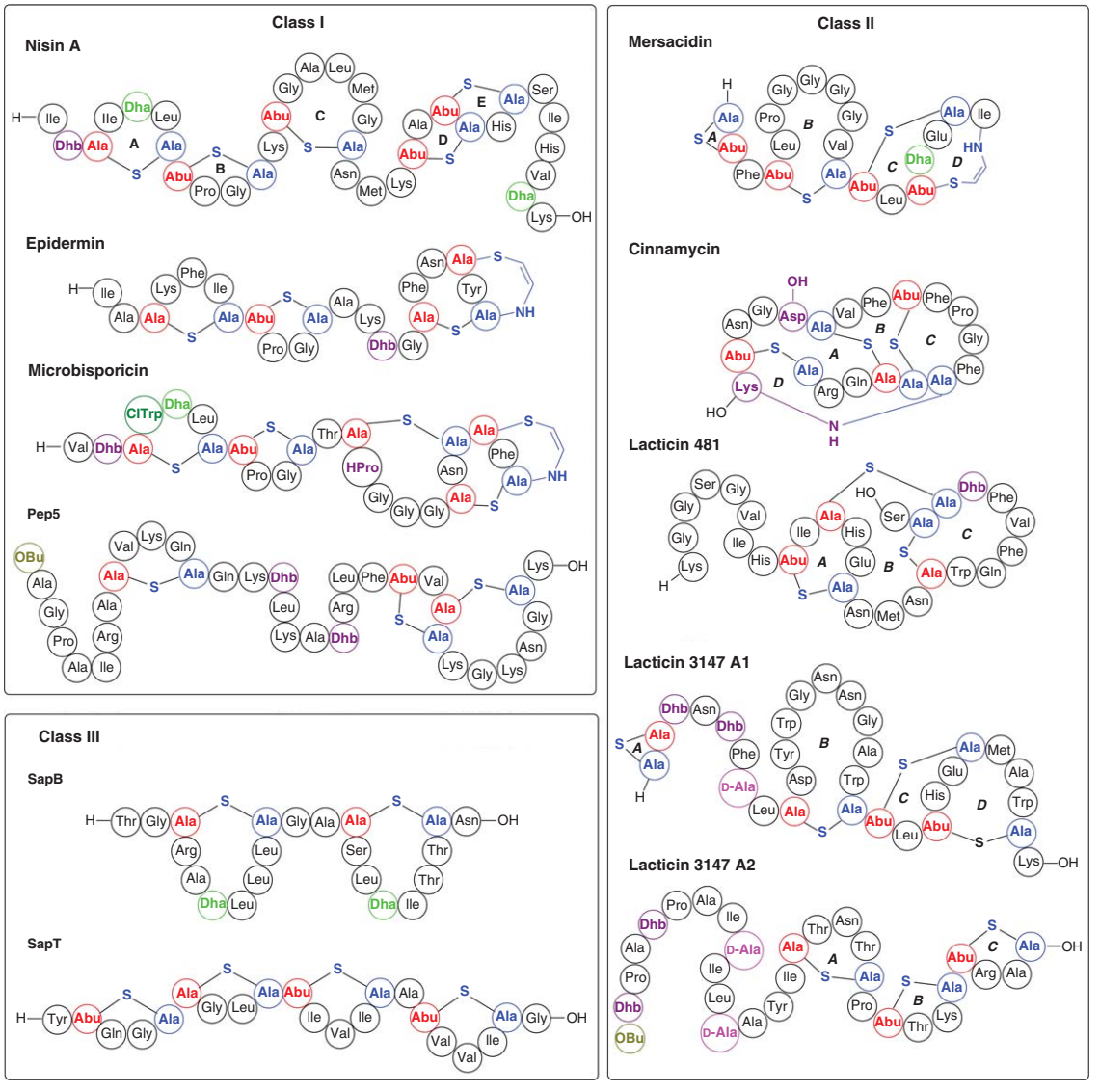


Figure 2 Representative examples of the three classes of lantibiotics. The same color-coding and shorthand notation is used as defined in **Figure 1**. For Lan and MeLan structures, the segments derived from Ser/Thr are in red whereas those derived from Cys are in blue. The ring numbering is shown for some members and is typically alphabetical from the N- to C-terminus.

gene clusters are ubiquitous and widespread. For instance, polymerase chain reaction (PCR) analysis of actinomycetes not known to produce lantibiotics revealed a large fraction of strains with the genetic capability to produce these compounds.[8] Furthermore, lantibiotics, which were originally believed to be limited to bactericidal activity against closely related species, have been shown recently to have a wide spectrum of biological activities, including immunomodulatory[9,10] and morphogenetic functions.[11]

The lantibiotics are ribosomally synthesized as precursor peptides (prepeptides) with an N-terminal leader peptide and the C-terminal structural peptide (also called propeptide) that undergoes processing to the mature active compound. As such, the lantibiotics are examples of a growing group of natural products including microcins (See Chapters 2.16 and 7.19),[12–16] patellamides (see Chapter 2.16),[17,18] and conotoxins[19] (see Chapter 5.10) for which a leader peptide appears to direct the maturation process. This chapter will focus on the structures, biosynthetic mechanisms, and biological activities of lantibiotics. For in-depth coverage of their gene clusters, regulation, applications, and self-immunity, the reader is referred to a number of reviews.[2,3,9,10]

5.08.2 Overview of Lantibiotic Structures

At present, more than 50 different lantibiotics are known displaying a striking diversity in structure, size, ring topology, and mode of action. A representative collection is depicted in **Figure 2** illustrating the high level of posttranslational modifications that typically amount to structural changes to about one-third of all amino acids in the peptide. The conformational constraints induced by lanthionines are essential for nisin's antimicrobial activity through binding to the cell wall biosynthetic intermediate lipid II[20,21] and forming pores in the bacterial cell membrane[22] (see Section 5.08.4). These multiple modes of bactericidal action are believed to account for the observed high efficacy of nisin (minimum inhibitory concentration (MIC) in the range of nanomolar) as well as the slow emergence of resistance.[23] Other lantibiotics with entirely different primary and three-dimensional structures such as mersacidin and cinnamycin also recognize with high affinity their targets, lipid II[24] and phosphatidylethanolamine[25] (PE), respectively, suggesting that the lanthionine motif is a naturally privileged architecture for constraining peptides into a bioactive conformation.[2] In addition, the lanthionine moiety provides improved chemical, proteolytic, and metabolic stability. For the majority of currently known lantibiotics, the targets are unknown, but based on the current knowledge on the biological activities of nisin, mersacidin, cinnamycin, lacticin 3147, and their structural analogues, it is highly likely that most, if not all, lantibiotics recognize specific targets with high affinity.

As depicted in **Figure 1**, a wide variety of posttranslational modifications are introduced into individual lantibiotics. In addition to the characteristic thioether cross-links Lan and MeLan, some lantibiotics, for example, epidermin and mersacidin (**Figures 1 and 2**), contain other cross-links such as *S*-aminovinyl-D-cysteine (AviCys) or *S*-aminovinyl-3-methyl-D-cysteine (AviMeCys). Yet another type of cross-link is found in cinnamycin in which a Lys residue is connected through its side chain ε-amino group to an Ala (lysinoalanine). In addition to these modifications that result in cyclic structures, a large number of structural modifications occur to amino acids that are not involved in ring formation. Among these are hydroxylations of Asp in cinnamycin and the duramycins,[26,27] hydroxylation of Pro and chlorination of Trp in microbisporicin,[28] and epimerizations resulting in *allo*-Ile in cypemycin[29] and D-Ala in lacticin 3147[30] and lactocin S.[31] Furthermore, Dha and Dhb residues that are N-terminally exposed after leader peptide processing spontaneously hydrolyze to yield a 2-oxopropionyl (OPr) moiety present in lactocin S,[31] and a 2-oxobutyryl (OBu) group present in Pep5.[32] OPr may be further modified by reduction to a 2-hydroxypropionyl (Hop) residue found in epilancin 15X,[33] epilancin K7,[34] and epicidin 280.[35]

The presence of posttranslational modifications has required the development of specialized methods for the structural determination of lantibiotics. Edman degradation is typically ineffective and therefore chemical fragmentation and derivatization techniques have been employed. Cyanogen bromide (CNBr) digestion followed by fast atom bombardment mass spectrometry (FAB-MS) was used to determine the thioether bridging pattern in lacticin 481.[36] The CNBr cleaves peptide bonds C-terminal to Met residues, resulting in the N-terminal portion of the peptide terminating with a homoserine lactone in place of Met. Additional chemical derivatization techniques focused on disruption of Lan/MeLan rings such that standard Edman degradation protocols could be utilized. Removal of an N-terminal OBu group has been achieved efficiently and successfully for Pep5 and lacticin 3147 A2 by reaction with 1,2-diaminobenzene in aqueous acetic acid.[37–39] Treatment of lantibiotics with an alkaline ethanethiol solution results in elimination reactions of the thioethers and addition of ethanethiol to the originally bridging residues.[38] However, because ethanethiol also adds to the nonbridging Dha/Dhb residues, it is not possible to distinguish Dha/Dhb from the thioether rings in the original structure. A more recently developed approach overcomes this drawback.[39] The protocol, involving the desulfurization of Lan/MeLan rings and the simultaneous reduction of Dha/Dhb using nickel boride (Ni_2B) with sodium borodeuteride ($NaBD_4$) in deuterium-labeled methanol/water (CD_3OD/D_2O, 1:1), was used for the structural determination of both peptides of lacticin 3147. A single deuterium atom was incorporated at the β-carbon of each residue that was linked in a Lan/MeLan whereas a deuterium atom was introduced at both the α- and β-carbon of former Dha/Dhb residues. Although bridging patterns are not revealed by this technique, it facilitates rapid differentiation between Dha/Dhb and Lan/MeLan

residues. In addition to derivatization techniques, a number of lantibiotic structures have been investigated by nuclear magnetic resonance (NMR) spectroscopy (e.g., nisin,[40–42] subtilin,[43] Pep5,[44] lacticin 481,[45] mersacidin,[46] actagardine,[47] cinnamycin,[48] and lacticin 3147[39]) and several NMR structures have been solved in the presence of their biological target (e.g., nisin with lipid II,[49] mersacidin with lipid II,[50] cinnamycin with PE).[51,52] When sizeable amounts of material are not available for NMR studies, tandem MS methodologies have proven useful. Although fragmentation within thioether rings does not result in new fragment ions and hence cannot be used to determine ring connectivities,[53] disruptions of the rings by site-directed mutagenesis allow the use of tandem MS to determine ring topology.[54]

5.08.3 Lantibiotic Biosynthesis

Lantibiotics are ribosomally synthesized as precursor peptides (prepeptides) with an N-terminal leader peptide and a C-terminal structural peptide (also called propeptide). The unifying lanthionine structural motif present in all lantibiotics is introduced into the structural region by dehydration of Ser and Thr residues to the corresponding Dha and Dhb structures and subsequent intramolecular attack by Cys residues onto the dehydro amino acids (**Figure 3**). These two reactions can be catalyzed either by a separate dehydratase LanB and cyclase LanC or by a bifunctional enzyme LanM. After introduction of the cyclic structures, the leader peptide that has not undergone modification is removed proteolytically by either a LanP Ser protease or a LanT Cys protease. A representative example of the posttranslational modification process is shown for nisin in **Figure 4**. The efficiency of this biosynthetic pathway is remarkable with the dehydratase catalyzing the cleavage of 16 chemical bonds and the cyclase generating 10 new chemical bonds with control over regio-, stereo-, and chemoselectivity. The regulation of the biosynthetic process is not the focus of this chapter, but an example of the best-understood system is shown in **Figure 5** for nisin.

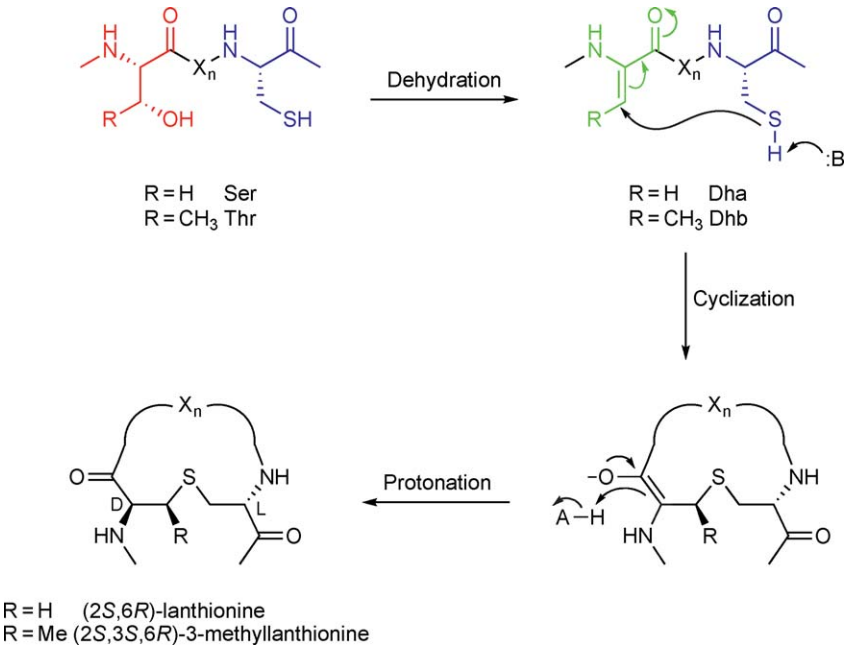


Figure 3 The two-step process resulting in the formation of lanthionines from Ser and Cys and methyllanthionines from Thr and Cys.

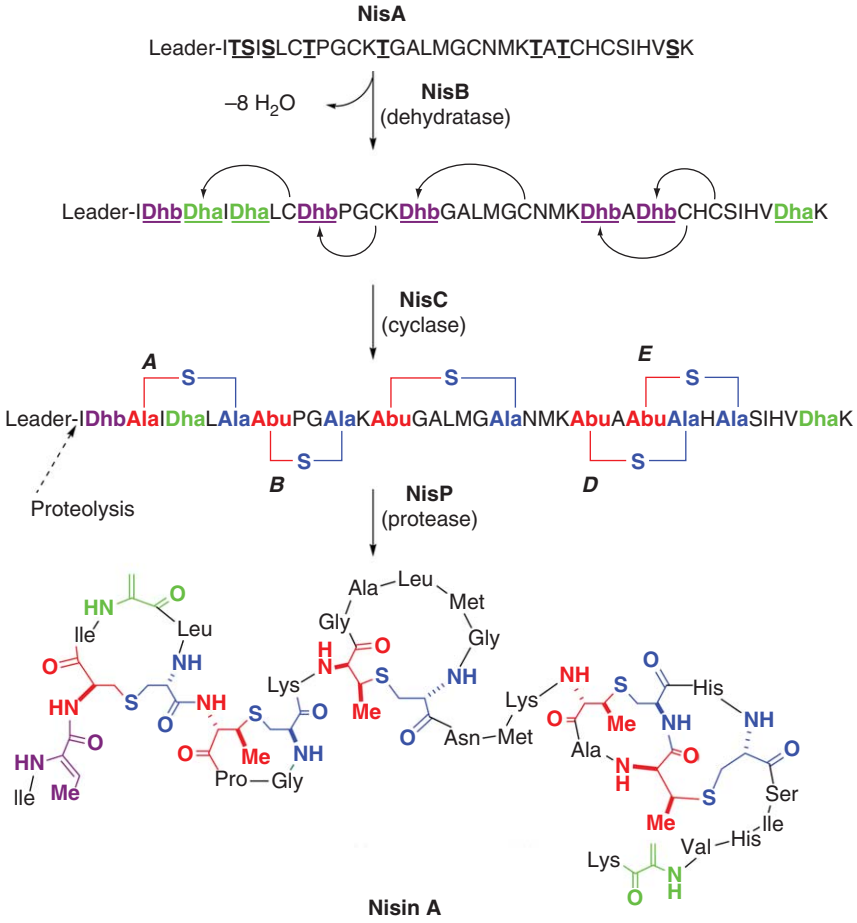


Figure 4 The biosynthesis of nisin A as a representative example of the posttranslational maturation process of lantibiotics. Following ribosomal synthesis, NisB dehydrates serine and threonine residues in the structural region of the prepeptide NisA. NisC subsequently catalyzes intramolecular addition of cysteine residues onto the dehydro amino acids in a stereo- and regioselective manner. Subsequent transport of the final product across the cell membrane by NisT and proteolytic cleavage of the leader sequence by NisP produces the mature lantibiotic. For the sequence of the leader peptide, see **Figure 6**. Adapted with permission from J. M. Willey; W. A. van der Donk, *Annu. Rev. Microbiol*. **2007**, *61*, 477–501.

5.08.3.1 Classification Schemes: Class I, II, and III Lantibiotics

Initially, lantibiotics were categorized as type A and type B based on their three-dimensional structures and their biological activities.[55] Type A lantibiotics were grouped based on an elongated, screw-shaped amphipathic structure with a net positive charge. Members of this group, including nisin and epidermin, act to form pores in cell membranes. On the other hand, type B lantibiotics such as mersacidin and cinnamycin are compact and globular with no net charge or a negative charge. These lantibiotics were believed to act by inhibition of enzymatic functions. Type A lantibiotics were later divided into type AI and AII based on their biosynthesis.[56] Type AI lantibiotics have gene clusters that contain the *lanBC* genes (e.g., nisin) and type AII clusters instead contain *lanM* genes (e.g., lacticin 481). This categorization scheme began to break down when nisin (type A) and mersacidin (type B) were both found to bind to the peptidoglycan precursor lipid II and inhibit cell wall biosynthesis.[20,21,57–60]

An alternative classification scheme was introduced in 2002 and this divides lantibiotics into two subgroups, class I and class II.[23] This scheme primarily classifies lantibiotics according to their biosynthetic enzymes and sequence homology of their leader peptides. Similar to most biosynthetic pathways in bacteria, the genes for lantibiotic biosynthesis are clustered. They have been designated the generic locus symbol *lan*, with a more

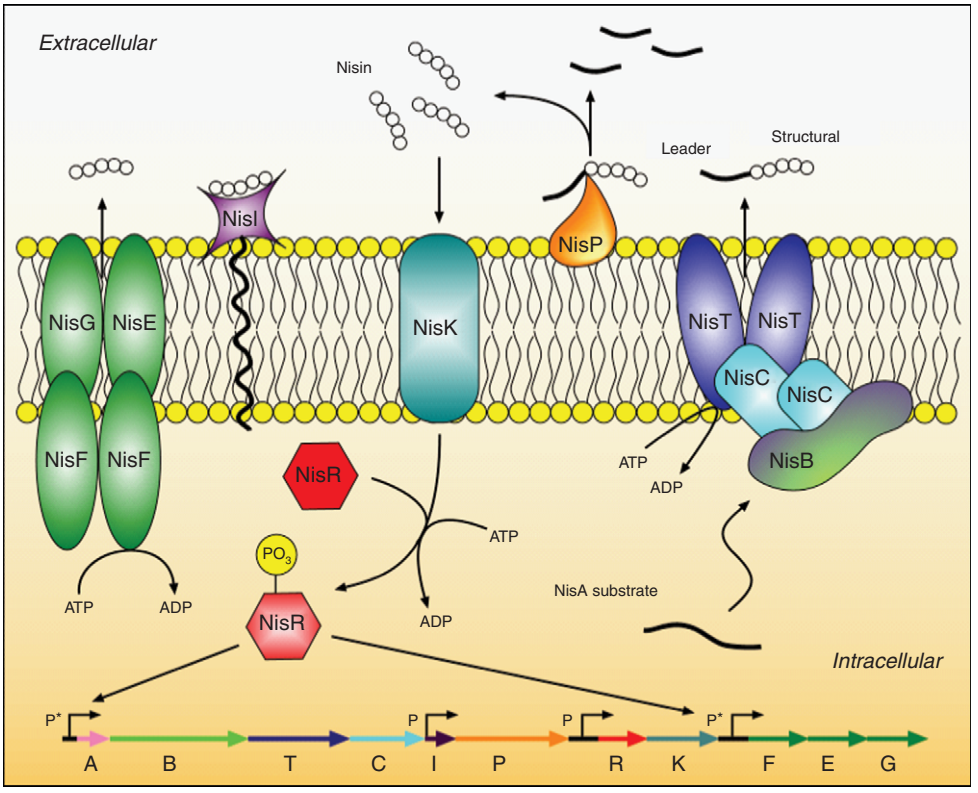


Figure 5 The structural region of the NisA prepeptide is modified by a putative multienzyme complex consisting of the dehydratase NisB, the cyclase NisC, and the transporter NisT.[121] After export, the leader peptide is removed by NisP, which is anchored to the cell wall. Mature nisin activates the two-component response regulatory system NisRK, and phosphorylated NisR serves as a positive regulator of *nisA* and the biosynthetic and immunity operons expressing NisABTC and NisFEG, respectively.

specific genotypic designation for each lantibiotic member (e.g., *nis* for nisin, *lct* for lacticin 481, *cin* for cinnamycin, *spa* for subtilin, *ltn* for lacticin 3147). For both class I and class II lantibiotics, the genes encoding the precursor peptides are designated as *lanA*. The LanA peptides in class I lantibiotics are modified by two enzymes, the dehydratase LanB and the cyclase LanC. The peptide is secreted through a dedicated ATP-binding cassette (ABC) transporter, LanT. Leader sequence removal is carried out by a lantibiotic-specific LanP serine protease. Sequence comparison of the leader peptides of class I lantibiotics revealed a common 'FNLD' motif between residues −20 and −15 (**Figure 6**),[61] and a Pro residue is usually found in the −2 position.

Class II gene clusters do not contain the *lanBCP* genes of class I lantibiotics. Instead, both the dehydration and cyclization activities are carried out by the product of a single gene, *lanM*. The C-terminal domains of the LanM proteins share 20–30% pairwise sequence identity to the LanC enzymes,[62] but no homology with the LanB enzymes has been found. Transport is carried out by a LanT protein in both class I and class II lantibiotics, but the class II LanT proteins contain an additional N-terminal cysteine protease domain that is responsible for leader peptide removal.[23] Sequence comparison also indicates that class II leader sequences contain a GlyGly or GlyAla/Ser cleavage recognition site.[61] This 'double-glycine motif'[45,61,63,64] resides at the junction between the leader peptide and the structural region of the LanA precursor peptide. To date, all two-peptide lantibiotic systems, which consist of two precursor peptides that are each posttranslationally modified and act in synergy to provide antibacterial activity, are part of class II lantibiotics. Seven such systems have been characterized: lacticin 3147 (**Figure 2**),[30] haloduracin,[53,65] staphylococcin C55,[66,67] cytolysin,[68] plantaricin W,[69] Smb,[70] and BHT-A.[71] Given the large number of structurally similar natural variants of lantibiotics, they have also been grouped

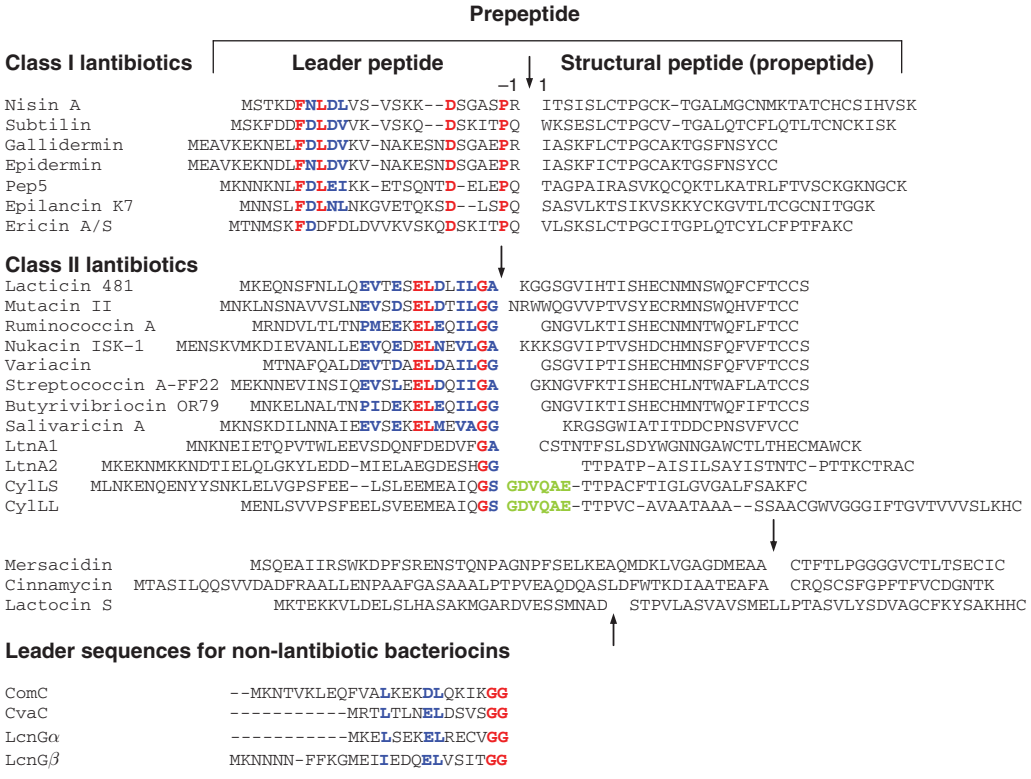


Figure 6 Sequence alignment of lantibiotic and nonlantibiotic bacteriocin prepeptides. The residues in red indicate those positions that are fully conserved within that class, and those in blue are highly conserved. For the nonlantibiotic bacteriocins, only the leader sequences are shown. The site of proteolysis is indicated by the arrow. For cytolysin, the additional six residues removed by CylA are indicated in green.

according to similarities in primary sequence and ring topology. These groups are named for the best-studied member (e.g., nisin, lacticin 481, mersacidin, cinnamycin, and lacticin 3147).[2,9]

Lantibiotics have been defined as lanthionine-containing antibiotics;[1] however in recent years, nonantibiotic peptides containing lanthionines have been discovered. The most-noted examples are the morphogenetic peptides SapB and SapT (**Figure 2**), produced by *Streptomyces coelicolor* and *Streptomyces tendae*, respectively.[11,72–74] A few others produced by *Streptomyces avermitilis*, *Streptomyces griseus*, and *Streptomyces scabies* have also been identified.[11,75] In order to accommodate these new peptides, a recently proposed classification system introduced a third group of lantibiotics (class III).[3] Members were defined as lanthionine-containing peptides that perform functions in the producing cells (e.g., signal morphological changes such as during sporulation), but lack significant antibiotic activity. The biosynthetic enzymes appear to differ from those typically found in class I and class II lantibiotics and it is not definitively known whether they also utilize dehydro amino acids as intermediates. The SapB modification enzyme, RamC, contains an N-terminal Ser/Thr kinase domain, a dimerization domain, and a LanC-like domain,[72] but the LanC-like domain lacks several metal-binding residues that are fully conserved in LanC proteins and essential for their activity (see Section 5.08.3.5).

As more information is uncovered regarding biosynthetic gene clusters, mature three-dimensional structures, modification enzymes, and the diversity of biological activities, any classification scheme is bound to continue further evolution, and certain exceptions to the general classification rules will always be encountered. For instance, the clusters of subtilin[76] (class I) and cinnamycin (class II)[77,78] lack lantibiotic-specific proteases (LanP or LanT type). For subtilin, three extracellular serine proteases that are not in the biosynthetic gene cluster – subtilisin, WprA, and Vpr – can remove the leader peptide.[79] Cinnamycin's leader peptide, the longest currently known at 59 amino acids, lacks the usual 'double-glycine motif' of class II lantibiotics, but instead ends in the AXA motif characteristic of the general secretory pathway

(*sec*) type I signal peptidases.[80] Also, mersacidin[81,82] and lactocin S[31] are exceptions to the class I and II leader peptide consensus rules[61] and neither leader sequence terminates with the anticipated class II double-glycine motif. Furthermore, despite containing multiple negatively charged residues, these two leader peptides do not appear to align well with other known class II leader sequences (**Figure 6**). The lactocin S gene cluster is also unusual in that it contains both a class II LanM protein and a class I LanP subtilisin-like protease.[83] Another apparent difficulty with respect to classification into class I or class II is posed by the salivaricins, which contain the typical leader peptide signature for class II lantibiotics but were reported to be modified by SalB and SalC proteins.[84] However, correction of a sequencing error revealed that *salBC* is actually a single *lanM*-like gene.[85,86]

5.08.3.2 Precursor Peptides

The ribosomally synthesized lantibiotic precursor peptides consist of a structural region (19–37 residues) that is modified to the mature compound and an N-terminal leader sequence (23–59 residues) that is not modified (e.g., see **Figures 4** and **6**).[87–89] The role of the leader peptide is still under investigation with several functions supported by experimental data. The modified structural region shows no antibiotic activity as long as the leader sequence is attached, and therefore the leader peptide may keep the lantibiotic inactive inside the cell, thereby protecting the producing bacterial strain.[54,90,91] In addition, the leader peptides appear to be important for molecular recognition by the lanthionine incorporation machinery. For instance, nonlantibiotic peptides attached to the leader peptide of nisin were dehydrated *in vivo* by the NisB enzyme,[92–94] and nonlantibiotic peptides attached to the leader peptides of lacticin 481 were dehydrated and cyclized by LctM *in vitro*.[95,96] Furthermore, whereas NisC correctly cyclized a dehydrated NisA peptide, no enzymatic cyclization was observed when the leader peptide was removed.[91] The leader peptide also appears to be essential for recognition by transporter systems that carry the lantibiotic across the cell membrane because nonlantibiotic peptides attached to the leader peptide of subtilin or nisin were exported.[97,98] And finally, the leader peptide is important for recognition by the proteases that remove it because mutations in the leader sequence result in loss of protease activity.[90,99,100] It should be emphasized that the segment of the precursor peptides termed the leader peptide is defined by the protease cleavage site, but it is currently not known whether it is also this same segment that is recognized by the biosynthetic enzymes that precede the proteolytic processing step. It is certainly possible that these enzymes recognize a smaller portion of the leader peptide or that part of the structural region is included for substrate binding.

The precise mechanism of leader peptide recognition by the various lantibiotic biosynthetic enzymes is at present not understood. The leader sequences within each lantibiotic class share several conserved residues that could contribute to these functions (**Figure 6**). For class I lantibiotics, the leader peptides contain several negatively charged amino acids and positively charged amino acids, mostly lysines, whereas for class II lantibiotics, a preponderance of Asp and Glu residues is typically encountered. Investigation of the importance of conserved residues has mostly focused on *in vivo* mutagenesis studies in which the leader sequence was changed in a lantibiotic-producing bacterial strain. For class I lantibiotics, such studies on nisin biosynthesis demonstrated that the highly conserved leader residues in the −18 to −15 region were essential for lanthionine incorporation into the precursor peptide, whereas Ala−4 and Arg−1 were essential for leader peptide removal by NisP (**Figure 7**).[90] This same region of the Pep5 precursor peptide, however, appeared unimportant for correct modification to the mature species, but was important for biosynthetic efficiency.[101] Similar investigations on the class II lantibiotic mutacin II showed that mutation of the conserved double-glycine motif abolished the production of the final product and resulted in accumulation of dehydrated premutacin.[99] Whether cyclization had also occurred was not determined. These observations suggested that mutacin II production was disrupted because of the inability of the protease to remove the leader sequence of modified MutA. This hypothesis is supported by recent *in vitro* data on mutants of the double-glycine motif of the precursor peptide for lacticin 481 (LctA) that were fully dehydrated and cyclized by the synthetase LctM[102] but were not processed by the protease domain of LctT (**Figure 7**).[100]

Of the various MutA leader peptide mutations generated *in vivo*, only Leu−7Lys and Ile−4Asp fully disrupted mutacin II biosynthesis, but it could not be determined what step(s) in the overall maturation process had been perturbed.[99] Recent *in vitro* studies with mutants of LctA at Leu−7 showed that incorporation of any charged residue strongly perturbed lacticin 481 synthetase activity whereas the Leu−7Ala mutant was still

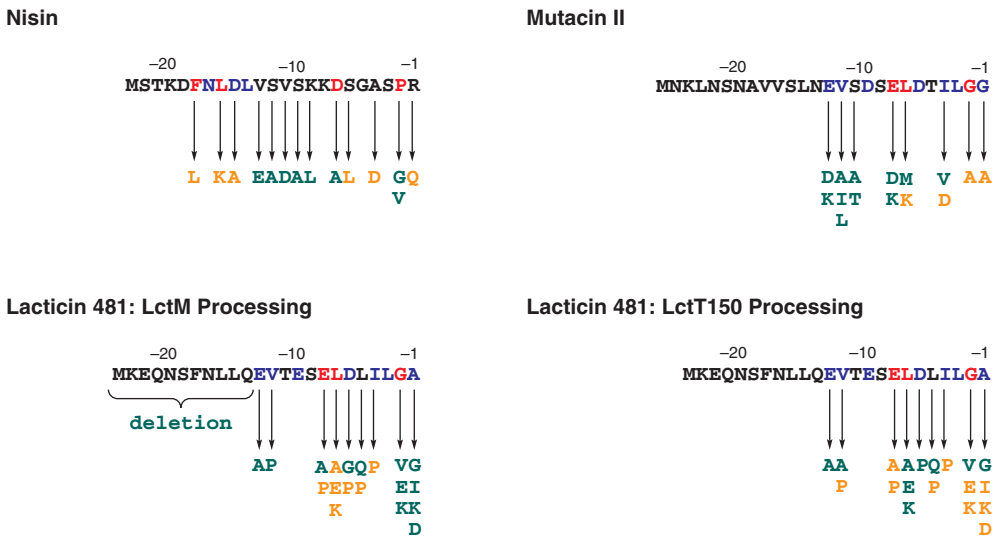


Figure 7 Sequence requirements of the leader peptides of nisin, mutacin II, lacticin 481, and lacticin 3147 as determined by site-directed mutagenesis. For nisin and mutacin II, mutants that still resulted in full processing of the prepeptides are shown in green, whereas mutants that resulted in abolished lantibiotic production are shown in orange. For lacticin 481, the mutants shown in green were good substrates *in vitro* for either the bifunctional synthetase LctM or the protease domain of LctT, whereas the mutants in orange were poor substrates. Conserved residues in the leader peptides of subgroups of lantibiotics are indicated in blue and red as described in **Figure 6**.

dehydrated and cyclized.[102] Collectively, the *in vivo* and *in vitro* mutagenesis data on the leader peptides of class II lantibiotics show that the LanM synthetases LctM and MutM tolerate almost all point mutations (**Figure 7**), including residues that are fully conserved among class II lantibiotic peptides. The fact that no single mutation abolished their activity suggests that multiple residues may be required for binding and raises the possibility that these enzymes recognize a secondary structure of the leader peptide.

Circular dichroism experiments on the leader peptides of nisin, subtilin, Pep5, epidermin, and gallidermin indicated an α-helical structure in solutions containing trifluoroethanol.[103,104] Indeed, current structure prediction tools all anticipate helical character for certain stretches of amino acids in the leader peptides, although this prediction has never been confirmed in aqueous buffer. It is possible that lantibiotic biosynthetic enzymes induce or capture a helical conformation of their substrates. The possibility of a helical conformation of the leader peptide for dehydration is supported by Glu−8Pro, Asp−6Pro, Leu−5Pro, and Ile−4Pro mutations in LctA that strongly decreased the efficiency of *in vitro* modification by LctM (**Figure 7**). These mutations are in regions that are predicted to be helical by several bioinformatics tools.

The hydroxy amino acids Ser and Thr are present in both the leader sequence and the structural region; however, Cys residues are found only in the structural region. For most lantibiotics, the number of Cys residues is smaller than the number of Ser and Thr residues in the structural region. The surplus hydroxy amino acids that are not involved in ring formation with the Cys residues are either present as dehydro amino acids in the final structures or remain unmodified. The biological importance of these residues appears to differ on a case-by-case basis.[9] Both Dhb24 in lacticin 481 and Dha33 in nisin were not essential for antibiotic activity.[95,105] On the other hand, mutation of Thr14 to Ser in gallidermin yielded a mixture of products containing either Ser14 or Dha14.[106] The Dha14 analogue retained near-wild-type activity, whereas the level of activity of the Ser analogue was reduced two- to fivefold. Dha5 in both nisin and subtilin has been reported to be essential for the prevention of spore outgrowth of bacilli and clostridia, as this activity was abolished in Dha5Ala mutants.[107–109] However, these conclusions were disputed in a more recent study that reported several mutants at this position that retained inhibition of spore outgrowth.[110] For sublancin 168 and the α-peptide of the two-component lantibiotic haloduracin,[53] the Cys residues outnumber the hydroxy residues in the structural region of the prepeptide.[53,65,111] It is unclear whether the two disulfide cross-links present in sublancin 168 are

essential for activity.[111,112] The disulfide in Halα is not required for antibiotic activity[53] but may protect the peptide from degradation by extracellular proteases.[113] In addition, the α-peptide of the two-component lantibiotic plantaricin W (Plwα) contains two Cys residues that were predicted to form a cystine linkage.[69] Its partner Plwβ has one free Cys, a highly unusual feature in lantibiotics.

5.08.3.3 LanB Dehydratases

In class I lantibiotics, the *lanB* genes encode proteins of ~120 kDa (~1000 residues) that have no homology with other proteins in the database.[114,115] Even among the LanB family, they share only ~25% pairwise sequence identity.[116] These proteins are predominantly hydrophilic with patches of hydrophobicity and were predicted to have a high helical content.[116,117] The LanB proteins were proposed to catalyze the dehydration of specific Ser/Thr residues in the class I precursor peptides. The first indirect evidence for this hypothesis was obtained from *in vivo* gene disruption experiments in a Pep5-producing strain.[118] Inactivation of *pepC* resulted in the isolation of dehydrated but not cyclized pre-Pep5 fragments, and in addition, intracellular accumulation of pre-Pep5 was observed. Similarly, whereas the dehydrated NisA prepeptide was recovered from strains lacking NisC activity, unmodified nisin precursor NisA was isolated from strains in which NisB activity was impaired, thus indicating the importance of NisB for dehydration in the biosynthesis of nisin.[119] The most conclusive evidence for the role of LanB proteins as dehydratases was obtained via studies expressing *nisABT* in a nonproducing *L. lactis* strain that yielded dehydrated prenisin without thioether rings.[98] The dehydrated NisA product so produced was later shown to be a substrate for NisC *in vitro* (*vide infra*).[91]

The enzymatic reaction mechanism of LanB dehydratases remains elusive. NisB (nisin) and SpaB (subtilin) cosedimented with membrane vesicles, suggesting that these proteins are likely membrane-associated despite their overall hydrophilic sequences and the absence of predicted transmembrane helices.[120] The difficulties encountered with reconstitution of their dehydration activity *in vitro* could be the result of compulsory formation of a multienzyme complex involving LanB, LanC, and LanT proteins, as suggested by yeast two-hybrid[121,122] and immunoprecipitation[121–123] data. However, *in vivo*, NisB is able to dehydrate substrate peptides in the absence of NisT and NisC, and *in vitro* NisC has been shown to catalyze cyclization of dehydrated NisA in the absence of NisB and NisT. Hence, multienzyme complex formation does not appear to be required for enzymatic activity. In a very recent study, *in vitro* expression of the genes *nisA*, *nisB*, and *nisC*, using a commercially available rapid translation system (RTS), generated a mixture of products that were recognized by an antibody against nisin A.[124] This product mixture displayed antimicrobial activity after trypsin treatment to remove the leader peptide. The presence of fully modified nisin was further supported through activation of the nisin-inducible *nisF* promoter in a green fluorescent protein (GFP) -based bioassay by the *in vitro*-generated product mixture. Thus, this study appears to show that NisB can act *in vitro*. However, despite many attempts, heterologous expression and purification of LanB proteins have not yet resulted in characterization of the dehydration activity.[123] As discussed in detail below, the dehydratase activity of LanM proteins has been reconstituted *in vitro* for several class II lantibiotics. However, since LanB proteins have no homology to the dehydration domain of LanM enzymes,[116] no extrapolations can be made regarding the mechanism of catalysis by LanB proteins.

Although *in vitro* reconstitution of LanB activity has been challenging, much research has been carried out on the characterization of their *in vivo* substrate specificity. As described in more detail in Section 5.08.6, many lantibiotic expression systems have been established and used for the production of analogues of class I lantibiotics. Collectively, these studies have shown the substrate promiscuity of LanB enzymes. In addition, the low substrate specificity has been explored in a nonproducing *L. lactis* strain by overexpressing nisin dehydratase and transport proteins along with chimera consisting of the nisin leader peptide fused at its C-terminus to nonlantibiotic peptides including fragments of therapeutic peptides and proteins such as an analogue of the angiotensin heptapeptide AsnArgSerTyrIleCysPro, the erythropoietin mimetic TyrAlaSerHisPheGlyProLeuGlyTrpValCysLys, and a mutant fragment of the adrenocorticotropic hormone SerTyrSerMetGluCysPheArgTrpGly.[92–94,125] Dehydrations were observed in the exported chimera, suggesting that dehydratase function is independent of the nature of the structural region as long as the leader peptide is present.

5.08.3.4 LanM Dehydratase Domains

The biosynthetic gene clusters of class II lantibiotics include *lanM* genes that code for proteins of ~115–120 kDa (900–1200 residues). They share sequence homology at the C-terminus with LanC cyclases (20–30% identity), including Cys and His metal ligands that coordinate to a Zn^{2+} in LanC proteins (*vide infra*).[126,127] Because they have no homology with LanB dehydratases, LanM proteins did not evolve from a simple fusion of the LanB and LanC enzymes found in class I lantibiotics.[116] Based on the absence of *lanB* and *lanC* genes in the class II lantibiotic clusters, LanM enzymes were proposed to catalyze both dehydration and cyclization of their LanA substrates.[116,128] Experiments to confirm this hypothesis initially focused on disruption of the *lanM* gene *in vivo* and analysis for the presence of modified LanA peptide. *Lactococcus lactis* cells expressing *lctATFEG* did not produce lacticin 481.[129] Instead, the *lctM*, *lctT*, and *lctA* genes were necessary to produce mature lacticin 481, suggesting a role in dehydration and cyclization for LctM.[128,130] The recent *in vitro* reconstitution of the dehydration and cyclization activities of LctM for the production of lacticin 481[54] and HalM1 and HalM2 for the generation of the two peptides of haloduracin[53] conclusively demonstrated that the LanM enzymes are indeed responsible for dehydration and thioether formation. In two-peptide lantibiotics, each LanA is modified by a designated LanM[53,131] except in the case of cytolysin for which the gene cluster possesses a single LanM (CylM) that is responsible for processing of both substrate peptides ($CylL_L$ and $CylL_S$).[132] This is likely a special case since $CylL_L$ and $CylL_S$ share 90% identity (**Figure 6**).

The bifunctional LanM enzymes likely possess two active sites, but the details of their interaction are unclear. It is possible that the enzymes catalyze all of the dehydration reactions before ring cyclization is initiated. A second option is that LanM enzymes pass the substrate between the two active sites and alternate between dehydration and cyclization. Most of the available data substantiate the former model, but the latter cannot be ruled out. In 2006, *in vitro* activity assays with the lacticin 481 synthetase LctM in combination with chemical modification and Fourier transform mass spectrometry (FTMS) were used to analyze the order of dehydration and ring formation.[133] The data revealed the presence of partially or noncyclized, yet completely dehydrated, species, thus providing support that dehydratase function is independent of cyclization and that the rate of Ser/Thr conversion to Dha/Dhb was considerably faster than Lan/MeLan formation.[133] However, these studies may have been biased toward efficient dehydration and slow cyclization because the assays contained concentrations of LctA that are likely higher than in physiological settings. As a result, the enzyme encountered a large excess of substrate for dehydration whereas only a very small concentration of dehydrated peptide was generated as the substrate for cyclization. As first shown for the LanC cyclases NisC and SpaC, the cyclization enzymes including the C-terminal domains of the LanM bifunctional proteins contain three strictly conserved metal ligands that coordinate a zinc ion (see Section 5.08.3.5). When one of the three zinc ligands in the LctM cyclase active site was mutated (Cys781Ala and Cys836Ala), ring formation was reduced or abolished, yet the dehydration activity remained unaffected.[134] These mutants further demonstrated that LctM-catalyzed dehydration is not coupled to thioether formation. On the other hand, it is possible that the binding site for the leader peptide is shared between the dehydration and cyclization activities in the bifunctional LanM proteins.

Much research has focused on the substrate specificity of dehydration by LanB and LanM enzymes. A comprehensive examination of both class I and class II lantibiotic structures revealed a weak consensus in the neighboring amino acids of Ser/Thr residues targeted for dehydration.[93] It was noted that Ser residues escape dehydration more often than Thr. This tenet holds true for the newly identified lantibiotic haloduracin, which retains a nondehydrated Ser residue in each of the two mature peptides.[113] Furthermore, the comparison of lantibiotic structures revealed that Ser/Thr residues were typically flanked by at least one hydrophobic amino acid, and that carboxylate-containing residues were rarely observed, with a particular aversion to Asp located N-terminally to Ser/Thr in combination with a C-terminal Arg (Asp-Ser/Thr-Arg). These observations may reflect the substrate specificity of the dehydratase enzymes but also the relative paucity of Asp and Glu residues in the predominantly cationic lantibiotics. Indeed, whereas NisB did not dehydrate Ser residues flanked by Asp and Arg in nonlantibiotic peptides,[93] LctM was capable of fully dehydrating a Ser residue situated within an AspSerArg sequence that was engineered into LctA.[95] This difference is not unexpected given the absence of homology between LanM and LanB enzymes.

Both LanM and LanB proteins in general show remarkable substrate promiscuity.[2,9,93–95,135,136] As discussed in Section 5.08.6, this promiscuity has been exploited for the preparation of analogues of many lantibiotics

including nisin,[105,137–139] subtilin,[109] epidermin and gallidermin,[106] cinnamycin,[80] Pep5,[140] mersacidin,[141] lacticin 3147,[135,136,142] and mutacin II.[143] The promiscuous nature of the modification enzymes is also illustrated by a series of studies with chimeric substrates. *In vitro*, LctM correctly dehydrated and cyclized peptides consisting of the LctA leader sequence and the structural region of other class II lantibiotics such as nukacin ISK-1, mutacin II, and ruminococcin A.[102] These peptides contain between 5 and 11 mutations in the prepeptide compared to LctA (**Figure 6**). Similar findings were reported earlier in *in vivo* studies conducted on chimeras of the class I lantibiotics nisin and subtilin.[144] Furthermore, when the leader peptides are sufficiently homologous, LanM and LanB enzymes can also dehydrate chimeric substrates containing the leader peptide of another lantibiotic, as demonstrated *in vitro* by the dehydration by LctM of a peptide consisting of the mutacin II leader peptide and lacticin 481 structural peptide,[102] and *in vivo* by the processing of a chimera of the leader sequence of subtilin and the structural peptide of nisin in a nisin-producing *L. lactis* strain.[145]

5.08.3.4.1 Catalytic mechanism

Enzymes that catalyze multiple transformations on a macromolecule can be either processive or distributive. Processive enzymes remain bound to their substrate through multiple rounds of catalysis, whereas distributive enzymes release intermediate products after each catalytic event. To investigate whether lantibiotic synthetases are processive or distributive, rapid-quench single-turnover enzymatic assays were carried out monitoring the dehydration of LctA by LctM.[133] These studies demonstrated that all four dehydrations of LctA were completed before the release of the product, suggesting that LctM dehydrates Ser/Thr in a highly processive manner.[133] More recent studies have demonstrated, however, that these observations are the result of slow deaggregation of the poorly soluble substrate peptide rather than a processive enzymatic reaction.[102] Whether the cyclization reaction catalyzed by LctM is processive is at present unknown and is difficult to determine experimentally as the intramolecular additions of cysteines to Dha/Dhb do not result in a change in mass and are therefore not readily monitored by mass spectrometry.

The current hypothesis for the role of the leader peptide in dehydratase activity and processivity is shown in **Figure 8** and is based on the results from several studies. LctM is proposed to recognize a certain secondary structure, possibly helical (see Section 5.08.3.2), adopted by the C-terminal segment of the leader peptide.[102] Leader peptide binding is then postulated to bring the structural region of the substrate in close proximity to

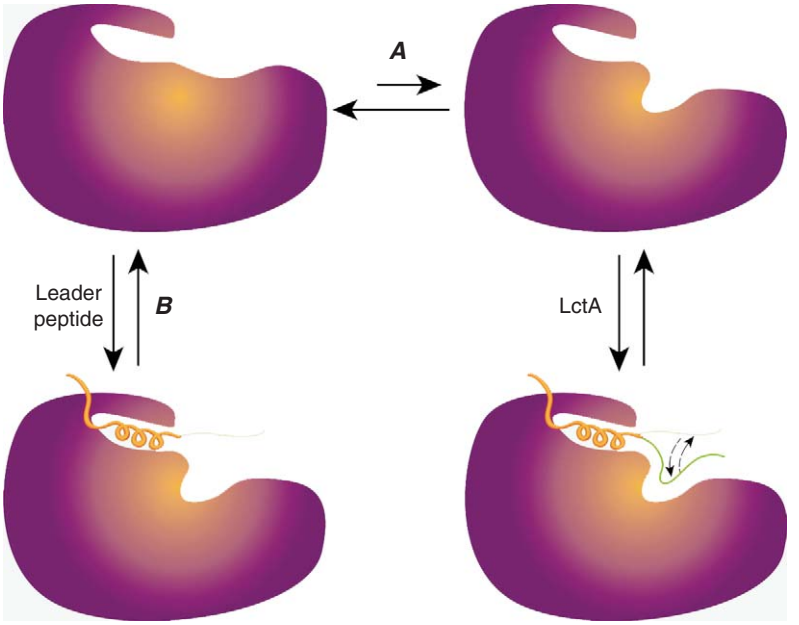


Figure 8 The proposed role of the leader peptide in LctM activity. Leader peptide binding is proposed to shift the equilibrium between inactive and active enzyme toward the latter. Reprinted with permission from G. C. Patton; M. Paul; L. E. Cooper; C. Chatterjee, W. A. van der Donk, *Biochemistry* **2008**, *47*, 7342–7351.

the active site resulting in the experimentally observed highly promiscuous dehydration. Ser/Thr residues that are too close to the leader peptide are not dehydrated as they cannot reach the active site, consistent with experimental observations.[95] Leader peptide binding also results in overall increased efficiency of dehydration because a leaderless substrate is processed much slower than the full-length prepeptide.[146] However, the observation that the prepeptide without the leader is a substrate at all demonstrates that the leader peptide is not absolutely required for dehydration.[146] These studies therefore argue against an induced fit model in which leader peptide binding triggers a compulsory conformational change to convert inactive LctM to an active dehydratase (**Figure 8**, pathway B). Instead they suggest a mechanism in which the leader stabilizes an active form of LctM that is present as a minor species (pathway A). **Figure 8** also accounts for the observed loss of directionality when the leader peptide and prepeptide are provided *in trans*.[146] Under these conditions, the leader peptide can still increase the amount of active dehydratase but because the product of each successive dehydration is not tethered after product dissociation, directionality is lost. Recent studies have shown that LctM is directional, first dehydrating Ser/Thr residues close to the leader peptide and subsequently residues located further toward the C-terminus.[146] It should be noted that it is currently not known whether binding of the leader peptide is static as shown in **Figure 8** or could involve a sliding mechanism.

The enzymatic mechanism of dehydration by LctM has been investigated in several studies. The enzyme utilizes ATP and Mg^{2+} to phosphorylate the Ser and Thr residues that are targeted for dehydration (Chapters 5.15 and 8.09). In a subsequent step in the dehydratase active site, the protein eliminates the phosphate to produce Dha and Dhb, respectively (**Figure 9**).[147] This second step requires deprotonation at the α-carbon of phospho-Ser and phospho-Thr. It is currently unknown whether α-carbon deprotonation and phosphate elimination occur through an enolate intermediate (as drawn in **Figure 9**) or in a concerted process. The requirement for ATP[53,54] was surprising as sequence alignments failed to identify a traditional ATP-binding motif.[148] Nevertheless, site-directed mutagenesis and subsequent analysis of the phenotypes of LctM mutants suggest a mechanism of phosphorylation that is strikingly similar to that of Ser/Thr kinases (**Figure 9**). Currently, no structural information is available for any lantibiotic dehydratase. If the fold of these enzymes indeed turns out to resemble that of kinases, then lantibiotic production may have evolved from a Ser/Thr kinase that picked up an active site base to catalyze the *anti* elimination step. Site-directed

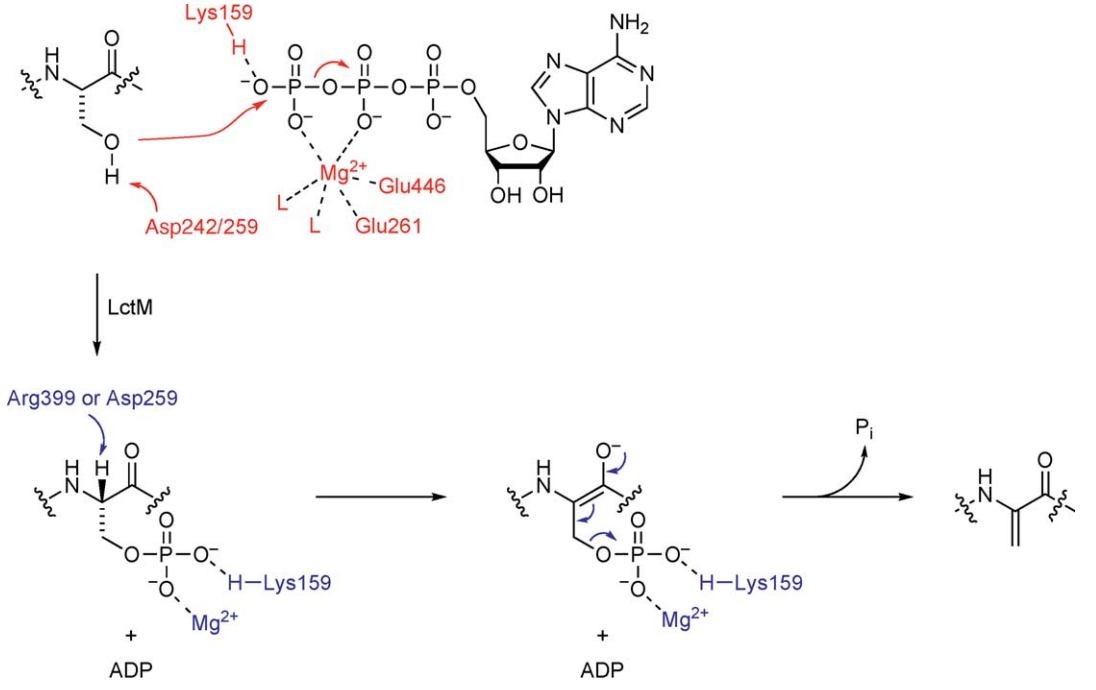


Figure 9 Proposed mechanism for dehydration of Ser and Thr residues in LctA by LctM.

mutagenesis studies suggest that Asp259 and Arg399 are important for phosphate elimination. For instance, the mutant LctM-R399M still phosphorylated the structural region of LctA but no longer eliminated the phosphate.[148] Although Arg residues have typically not been invoked for roles as catalytic acid or base, several examples have recently been reported.[149–156] Other proteins that catalyze phosphate elimination from Ser/Thr in peptides include the OspF family (OspF, SpvC, and HopAI1) of type III secretion system effector proteins[157] that possess phosphothreonine lyase activity.[158,159] OspF, SpvC, and HopAI1 selectively eliminate a phosphate group from Thr in a phospho-Thr-X-phospho-Tyr (pT-X-pY) motif that is highly conserved in mitogen-activated protein kinases (MAPKs). The crystal structure of SpvC in a complex with a phosphopeptide substrate revealed a positively charged binding pocket for phosphothreonine insertion.[160] However, despite the mechanistic connection between LanM proteins and the OspF family, they have no sequence homology.

The putative lantibiotic biosynthesis protein involved in the dehydration and/or cyclization of the morphogenetic class III lantibiotic SapB was initially identified based on sequence homology with the C-termini of LanM proteins.[72] The RamC protein has a kinase domain and a domain homologous with the LanC proteins and the C-terminus of the LanM enzymes. Interestingly, the zinc ligands are not conserved in the latter domain, suggesting that the substrate is activated for cyclization in a different manner than in all other lantibiotic cyclases (Section 5.08.3.5). The presence of a kinase domain on the other hand is consistent with the mechanism of LanM proteins and suggests that dehydration of the two serine residues in the SapB precursor also involves a phosphorylation step followed by an elimination step. Alternatively and speculatively, since the stereochemistry at C2 of the lanthionines in SapB has not been established unambiguously, it is also possible that the Cys residues directly displace the phosphate group of the phosphoserine intermediates, which would result in lanthionines with 2*R*-stereochemistry.

5.08.3.5 LanC Cyclases and LanM Cyclase Domains

The formation of thioether bridges in lantibiotics is catalyzed by either a cyclase LanC or a bifunctional enzyme LanM. As mentioned previously, the involvement of LanC in the cyclization step was first suggested by genetic knockout studies in a Pep5-producing strain of *Staphylococcus epidermis*. Disruption of the *pepC* gene resulted in the formation of full-length dehydrated prepeptide as well as smaller fragments thereof, none of which contained the correct thioether linkages.[118] In a parallel study on nisin, a NisC-deficient strain showed production of dehydrated NisA; however, no cyclization products were observed.[119] Direct evidence of the role of the LanC proteins was recently provided by *in vitro* reconstitution of the cyclization activity of the NisC enzyme. The substrate for NisC, eightfold dehydrated NisA, was obtained from an engineered *L. lactis* strain.[98] Recombinant NisC installed all five thioether rings in the dehydrated NisA peptide, and removal of the leader peptide led to the production of mature nisin.[91] Successful *in vitro* cyclization has also been performed with the LanM proteins LctM,[54] HalM1, and HalM2,[53] involved in lacticin 481, haloduracin α, and haloduracin β biosynthesis, respectively. MS/MS analysis of the LctA prepeptide modified by LctM verified the cyclization pattern characteristic of lacticin 481. These studies not only confirmed that LanC and LanM proteins were solely responsible for lantibiotic thioether ring formation, but also opened the door to further investigation of the mechanisms of catalysis.

The NisC and SpaC proteins have been shown to contain one zinc ion per polypeptide by inductively coupled plasma mass spectroscopy (ICP-MS).[126] The recently solved X-ray crystal structure of NisC confirmed the presence of a zinc ion on top of an α/α barrel fold (**Figure 10**).[91] Zinc is coordinated by two cysteines, one histidine, and a water molecule. The crystal structure is consistent with a previously proposed working model in which the zinc functions as the docking site for the cysteines of the substrate and promotes thiolate formation (**Figure 11**).[126] Deprotonation of the substrate thiols is highly advantageous because the rate constant of the addition of free thiols to α,β-unsaturated centers is 10^{10}-fold smaller than that of the corresponding thiolates.[161] In other zinc enzymes that catalyze thiol alkylation reactions,[162–165] it is thought that the zinc acts as a Lewis acid to decrease the pK_a of the substrate thiol. Cobalt substitution of zinc in protein farnesyl transferase has provided direct evidence of metal coordination of the thiolate.[164] Furthermore, the pH dependence of substrate binding to farnesyl transferase showed that the pK_a of the cysteine in the substrate was lowered from 8.3 in the free peptide to approximately 6.4 upon binding to the enzyme.[165] Additional support for

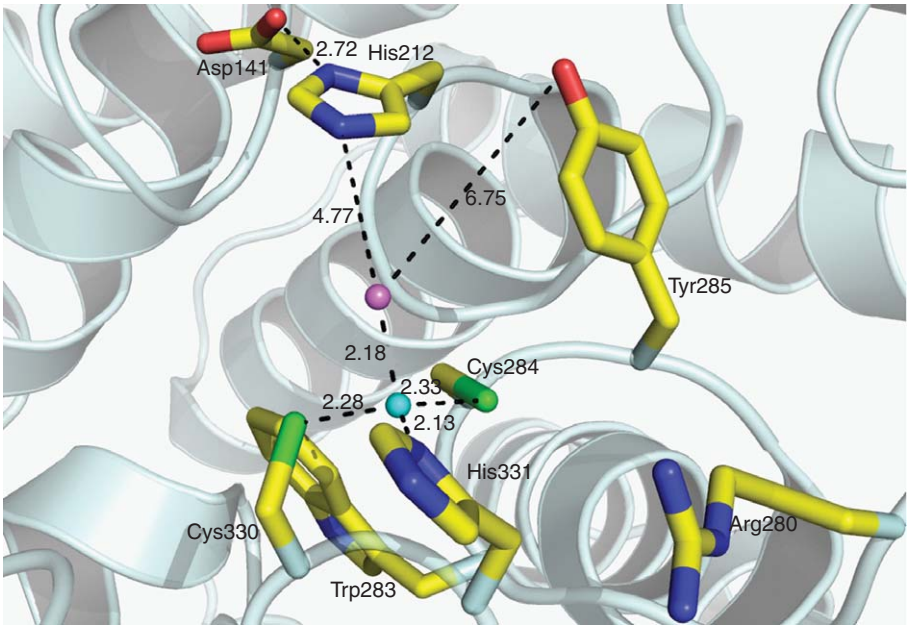


Figure 10 Active site of NisC. Conserved active site residues and zinc ligands are shown in color (carbon in yellow, oxygen in red, nitrogen in blue, and sulfur in green), the zinc ion is shown in cyan, and the putative water molecule that is coordinated to zinc is shown in pink. Adapted with permission from B. Li; W. A. van der Donk, *J. Biol. Chem*. **2007**, *282*, 21169–21175.

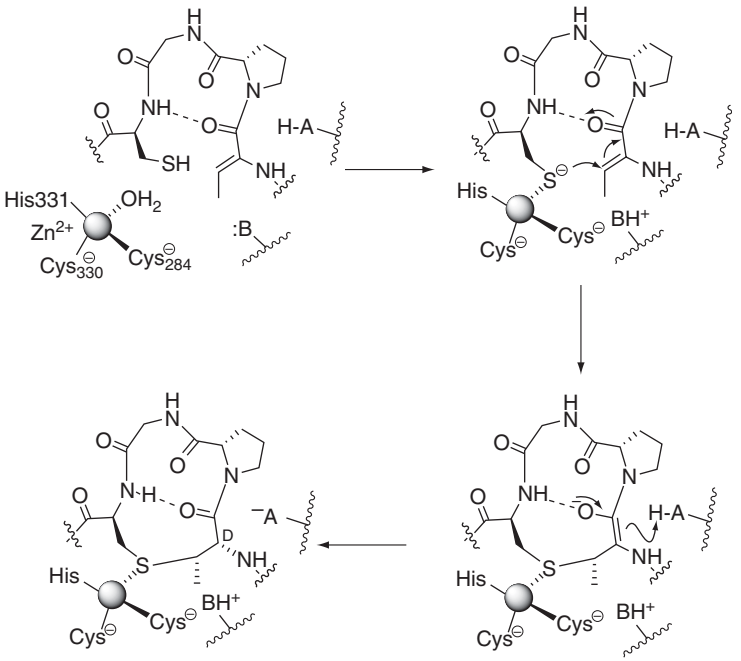


Figure 11 Proposed mechanism of cyclization of dehydrated NisA by NisC. The cyclization reaction shown results in the formation of the B-ring of nisin. The possible stabilization through a β-turn-like structure via hydrogen bonding between the amide NH of Cys and the carbonyl of Dha/Dhb is shown and may explain the high stereoselectivity observed in nonenzymatic cyclizations involving four amino acids as discussed in the text. Reprinted with permission from B. Li; W. A. van der Donk, *J. Biol. Chem*. **2007**, *282*, 21169–21175.

the role of the zinc is provided by several model studies on small molecule zinc complexes that demonstrated that increasing the number of sulfur donors to zinc accelerated the rate of thiolate alkylation.[166–175] In the case of NisC, once substrate binds, three thiolates are coordinated to zinc, including two protein ligands and one from the substrate. The resulting net negative charge at the zinc site is believed to activate the thiolate of the substrate for nucleophilic attack onto the Dha or Dhb residues.

Mutagenesis studies have been carried out to identify the essential catalytic residues in the LanC proteins.[127,176] All mutations of the zinc ligands of NisC (Cys284, Cys330, and His331) resulted in abolishment of *in vitro* cyclization activity and diminished zinc content. This result strongly supported the involvement of zinc in catalysis. Because of the proximity of His212 to the water molecule that is bound to zinc in the X-ray structure (**Figure 10**), His212 was originally postulated to be the catalytic base that deprotonates the thiol substrate. Indeed, mutation of the conserved His212 to Asn or Phe abolished the activity of NisC. Replacement of Asp141, which is hydrogen bonded to His212, by Asn also resulted in inactive NisC.[127] These findings are consistent with the His–Asp pair functioning as the active site base that deprotonates the thiol of the incoming Cys substrate. However, as previously mentioned, the zinc will lower the pK_a of the thiol of Cys such that it can be readily deprotonated at neutral pH, and a dedicated active site base may not be necessary. Indeed, no base is found in the crystal structures of farnesyl transferase,[177] methionine synthase,[178,179] and betaine-homocysteine methyltransferase,[180] three proteins that utilize a zinc ion to activate a thiol in their substrates toward alkylation. Therefore, the His212/Asp141 dyad may instead be involved in the protonation of the enolate or the electrophilic activation of the carbonyl group of Dha/Dhb. Two other residues, Arg280 and Tyr285, that are located close to the zinc site are conserved among the LanC proteins but not the LanM proteins. Both Met and Ala mutants of Arg280 retained their ability to cyclize dehydrated NisA, showing that this residue is not critical. Replacement of Tyr285 with an alanine generated an inactive enzyme; however, the Phe mutant of Tyr285 was still active. The activity of the Tyr285 Phe mutant suggests a critical role of an aromatic ring at this position, but these mutagenesis studies ruled out roles of Arg280 or Tyr285 as the active site acid that protonates the enolate intermediate.

Site-directed mutants of the subtilin cyclase gene *spaC* have been used to substitute the wild-type gene in a subtilin-producing strain. The effect of SpaC mutations was examined by monitoring subtilin production.[176] Alanine mutants of His231 (equivalent to His212 in NisC), Tyr304 (Tyr285 in NisC), and Trp302 (Trp283 in NisC) were unable to support subtilin production *in vivo*. Mutation of the zinc ligands of SpaC to alanine also halted lantibiotic production. When Arg299 (Arg280 in NisC) was mutated to alanine, subtilin was still produced. These findings are generally in good agreement with the *in vitro* results with NisC described above. In the study on SpaC, it was also shown that the ericin A/S cyclase EriC could complement a SpaC knockout strain to produce mature subtilin, whereas NisC could not. EriC and SpaC both originate from *Bacillus subtilis* and share 80% sequence identity, whereas NisC and SpaC are only 31% identical.

The C-termini of LanM proteins share ~20% identity with the LanC proteins. The conserved residues include the zinc ligands and the residue corresponding to His212. These amino acids were mutated in recombinant LctM and mutant activity was tested with full-length or truncated LctA peptide substrates *in vitro*.[134] The alanine mutants of the zinc ligands Cys781 and Cys836 retained dehydration activity, but their cyclization ability was compromised, similar to the analogous mutagenesis studies on NisC and SpaC. Mutation of His725 to Asn resulted in a protein that still dehydrated the substrate peptide, but cyclized the substrate at a decreased level compared to that of the wild-type enzyme. This investigation showed that cyclization activity is independent of dehydration activity in LctM.

The substrate specificity of NisC has been studied *in vivo* using a wide variety of designed and nisin-unrelated peptides fused to the leader peptide of NisA.[181] The capability of NisC to cyclize these peptides was evaluated by comparing the peptide products secreted by *L. lactis* containing genes for the fusion peptide, NisB, and NisT in the presence or absence of NisC. Reaction of a Dha with C-terminally located cysteines occurred spontaneously in hexapeptides fused to NisA leader even in the absence of NisC, whereas formation of MeLan from Dhb and Cys in these peptides required NisC.[181] These observations are consistent with nonenzymatic model studies that have demonstrated facile intramolecular reactions of Cys with Dha,[182–184] but much decreased reactivity of Cys toward Dhb residues.[185,186] A statistical analysis of all known lantibiotic structures suggested that certain preferred residues flanking the cysteines were involved in ring formation.[181] N-terminally flanking Glu and C-terminally flanking Lys residues appear to be favorable for

cyclization, whereas Trp occurs very rarely as the N-terminal flanking residue and Glu is never found as the C-terminal flanking residue.[93] In agreement with these observations, *in vivo*, NisC cyclized the nonlantibiotic peptides AlaDhbValGluCysLys and IleDhbProGlyCysLys attached to the NisA leader peptide, but not AlaDhbValTrpCysGlu. NisC also catalyzed the formation of two intertwined rings in the peptides LysDhbAlaDhbCysHisCysDhaLys and AlaDhbValAlaDhbCysLysGlyCysLys, and four consecutive rings in a 24-amino-acid peptide (IleDhbProGlyCysLys–AlaDhbValGluCysLys)$_2$.[181] Thus, this study demonstrates the versatility of NisC for the cyclization of nonlantibiotic peptides under the direction of the NisA leader peptide.

Investigations of the substrate specificity of the lacticin 481 synthetase LctM with respect to cyclization have provided similar results. Nonlantibiotic peptides attached to the leader peptide of LctA were both dehydrated and cyclized.[95,96] The products so obtained included a cyclic enkephalin analogue that had been previously prepared by synthetic chemistry and that has ~10 000-fold higher *in vivo* activity than morphine as an agonist for the μ- and δ-opioid receptors in mouse models. Degradation studies attributed the subnanomolar analgesic potencies to increased biostability due to the incorporation of a lanthionine ring structure.[187] Owing to the *in vitro* nature of the LctM assays, the possibility of cyclization of nonproteinogenic cysteine analogues has also been investigated. Substrate peptides containing D-cysteine, L-selenocysteine, L-homocysteine, and (*S*)-β^3-homocysteine were cyclized by LctM resulting in stereoisomers as well as homologated isomers of the naturally occurring lanthionines and methyllanthionines.[95,96,188,189] On the other hand, (*2R,3R*)-methylcysteine was not accepted as a Cys analogue, showing that substituents on the β-carbon of Cys are not tolerated in the cyclization active site of LctM.

Michael additions of a nucleophilic thiolate onto an electrophilic dehydro amino acid occur reasonably fast. In the context of lantibiotics with multiple ring structures, the challenge for the cyclase is the control of stereo- and regiochemistry. Several model studies have investigated the origin of stereoselectivity. Biomimetic formation of the lanthionine analogue of the epidermin B-ring showed that cyclization took place nonenzymatically producing a single, naturally occurring diastereomer.[182] Similar studies on the lanthionine-containing analogues of the B-ring and E-ring of subtilin also demonstrated the biomimetic formation of single stereoisomers.[183,184] Although the absolute stereochemistry of the cyclization products was not determined, computational modeling studies suggested that the naturally occurring diastereomer was most likely formed.[183] The stereochemistry of methyllanthionine formation requires an additional level of selectivity, *Si*-face selectivity by the addition of cysteine to the β-carbon of dehydrobutyrine. A linear peptide precursor for the subtilin B-ring was synthesized and tested for nonenzymatic cyclization. The resulting single MeLan diastereomer was shown to have the same configuration as the naturally occurring methyllanthionine.[185] These model studies suggested a strong intrinsic preference of the prepeptides of lantibiotics for the stereochemistry observed in the final natural products. The cyclizations discussed above all involve the addition of a cysteine to a Dha or Dhb that is N-terminal to the cysteine. Some lantibiotics such as cinnamycin, mersacidin, and the duramycins also undergo cyclization in the opposite direction. The stereochemical outcome in these compounds is again the formation of (*2S,6R*)-Lan and (*2S,3S,6R*)-3-MeLan. On the other hand, model studies of cyclizations in the N-to-C terminal direction generated a mixture of stereoisomers.[185]

Several biomimetic studies have investigated whether the prepeptides also have a natural tendency to cyclize with the same regiochemistry seen in natural lantibiotics when multiple Dha/Dhb residues are available for reaction with cysteine residues. Nonenzymatic cyclization of a precursor to the subtilin A-ring containing two dehydroalanines and one cysteine (DhaGluDhaLeuCys) produced two products in a 3:1 ratio as shown by reversed-phase high-performance liquid chromatography (RP-HPLC).[183] NMR analysis ascertained that both isomers had the same ring topology as the natural subtilin A-ring, but displayed different stereochemistry at the newly formed stereocenter. The stereochemistry of the products could not be assigned by NMR spectroscopy, but molecular modeling suggested that the major product had the same stereochemistry as the natural lanthionine. Compared to the result obtained for the formation of four-amino-acid rings described above (subtilin B-ring and E-ring), the inherent stereoselectivity of formation of a ring consisting of five amino acids was not strong. These observations may reflect the possibility to form a β-turn-like structure in the transition state of the former (e.g., see **Figure 11**). Additional studies focused on the regiochemistry of biomimetic formation of multiple rings using a linear precursor peptide for the A- and B-rings of nisin,

IleDhbDhaIleDhaLeuCysDhbProGlyCysAla.[186] NMR analysis of the product revealed that the two dehydroalanines reacted with the two Cys residues to form two lanthionine rings rather than the Lan A-ring and MeLan B-ring found in nisin (compare **Figure 4**). Thus, whereas the stereochemistry of individual rings can be obtained through nonenzymatic biomimetic cyclization, the inherently higher reactivity of Dha residues over Dhb residues prevents the correct nonenzymatic formation of multiple rings. Therefore, one important role of the cyclization enzymes is to overcome the chemoselectivity that favors lanthionine ring formation over methyllanthionine ring formation.[185]

5.08.3.6 LanP Proteases

Leader peptide removal after completion of the posttranslational modifications is generally achieved by a dedicated protease encoded in the lantibiotic gene cluster. In class I lantibiotics, the LanP enzymes that carry out the proteolysis are classified as subtilisin-type serine proteases containing a conserved catalytic triad (Asp, His, and Ser) and an Asn involved in oxyanion hole formation.[190] Peptidase localization, and thus the timing of leader processing, varies widely among individual class I lantibiotics. The enzymes can be found intracellularly, attached to the outside of the cell wall, or extracellularly. The nisin protease, NisP, contains an N-terminal *sec*-signal including a 220-residue prosequence, which is removed by a peptidase after extracellular transport, as well as a C-terminal cell wall anchor containing a consensus sequence important in anchoring surface proteins in Gram-positive bacteria (LeuProXxxThrGly).[191] The modified NisA precursor peptide with the leader still attached was cleaved by cell extracts of *Escherichia coli* overexpressing NisP and by whole cells of the nisin-producing strain *L. lactis* 9800. However, neither the medium supernatant nor the membrane free extracts of *L. lactis* 9800 cells removed the leader peptide.[190] These findings supported the model that NisP is located at the outside of the cell membrane and anchored on the cell wall surface. Several studies have investigated the substrate specificity of NisP. When a hybrid gene encoding the subtilin leader fused to the nisin Z structural region was expressed in a nisin A-producing *L. lactis* strain, a peptide consisting of subtilin leader attached to mature nisin Z was produced, indicating that NisP was unable to remove the subtilin leader.[145] The leader peptides of subtilin and nisin are 57% identical, but the last residue of the nisin leader peptide is Arg whereas it is a Gln in subtilin. The importance of Arg−1 was further established by mutation to Gln in NisA in a nisin-producing strain, which led to extracellular accumulation of fully modified nisin with the leader still attached.[90] A similar result was seen for the Ala−4Asp mutant. However, mutations of conserved residues Pro−2 and Asp−7 of NisA did not affect the processing of the leader sequence (**Figure 7**). More recently, the requirement of ring formation in the structural region of NisA for NisP cleavage was demonstrated *in vivo*.[98] Neither the unmodified precursor peptide nor the dehydrated peptide was processed by NisP. These results suggested that NisP is specific toward thioether-containing prenisin.

The subtilin biosynthetic cluster does not contain any genes encoding proteases,[76] prompting a series of studies investigating the specifics of the proteolysis step in the subtilin producer *B. subtilis* ATCC6633. Fusion of the SpaS leader to the NisA structural peptide led to the production of processed prepeptide in the extracellular space but not mature nisin (note, in an exception to the usual nomenclature for lantibiotics, the prepeptide for subtilin is called SpaS). However, a chimeric peptide consisting of residues 1–7 of the SpaS leader peptide, followed by residues 8–23 of the NisA leader and the full-length NisA structural region (1–34) was converted into mature nisin.[192] In a separate study, *B. subtilis* ATCC6633 expressing a peptide composed of the SpaS leader peptide, followed by residues 1–11 of the NisA structural region and residues 12–32 of the SpaS structural region generated bioactive peptide encompassing residues 1–11 of nisin and residues 12–32 of subtilin as demonstrated by NMR spectroscopy.[144] Collectively, these observations suggested that proteolytic processing during subtilin maturation was not completely specific for SpaS. To test whether only subtilin-producing *B. subtilis* strains have the required protease(s) for subtilin production, fully processed presubtilin with the entire leader attached was purified and incubated with the culture supernatant of a non-subtilin-producing strain *B. subtilis* 168. The leader peptide was processed and mature subtilin was produced.[193] This result, together with the lack of a protease-encoding gene in the subtilin gene cluster, suggested the involvement of one or more proteases elsewhere on the *B. subtilis* genome. The genome of *B. subtilis* encodes at least five extracellular serine proteases, three of which (subtilisin, WprA, and Vpr) were demonstrated subsequently to cleave the leader peptide and release mature subtilin.[79]

EpiP, the protease involved in epidermin biosynthesis, is initially generated as a pre-pro-enzyme and the N-terminal 99 residues that are absent from the mature protein contain a 25-residue signal peptide and a 74-residue prosequence. The absence of a C-terminal cell wall anchor in EpiP suggests that EpiP is an extracellular protein. Incubation of unmodified epidermin precursor peptide EpiA with the supernatant of *Staphylococcus carnosus* harboring the *epiP* gene indeed resulted in leader peptide removal, confirming that EpiP, unlike NisP, is not anchored on the cell wall. This study also shows that EpiP does not require the posttranslational modifications of EpiA for cleavage, another difference compared to NisP.[194] The replacement of Arg−1 in EpiA with Gln abolished the ability of EpiP to remove the leader. This result is in good agreement with the observation that NisP could not cleave Arg−1Gln NisA. An *epiP*-deficient epidermin-producing strain generated fully modified pre-epidermin attached to N-terminally truncated leader peptides.[195] Removal of the residual leader with trypsin released mature epidermin. Interestingly, this study also showed that the *agr* quorum-sensing system regulates epidermin production by controlling EpiP activity. Similar to the observations with epidermin, its structural homologue gallidermin attached to truncated leader peptides was observed when the protease GdmP was knocked out in the gallidermin-producing strain *Staphylococcus gallinarium* Tü3928, a close relative of the epidermin-producing strain *S. epidermis* Tü3298.[196] It is unclear which protease removes part of the leader in the absence of EpiP or GdmP.

PepP,[118] ElkP,[34] and LasP,[86] the proteases for Pep5, epilancin K7, and lactocin S production, respectively, lack the N-terminal prosequence and the C-terminal cell wall anchor. Therefore, they are proposed to function intracellularly, possibly as part of a large membrane-bound biosynthetic complex. When PepP was inactivated by exchanging the active site conserved His for Pro, accumulation of incorrectly proteolytically processed prepeptide was observed inside the cells but not in the culture medium. This result demonstrated the importance of PepP for correct removal of the leader and also suggested the inability of the Pep5 transporter PepT to shuttle the incorrectly processed prepeptide across the cell membrane.[118]

Class II lantibiotic gene clusters do not contain *lanP* genes. The LanT transporters from these clusters are generally around 700 amino acids in length, about 100 residues larger than the class I lantibiotic transporters. They encode not only the ABC transporter domain composed of a six-helix transmembrane domain and a C-terminal ATP-binding domain, but also an additional N-terminal peptidase domain. This N-terminal domain was proposed to remove the leader peptide from the fully modified precursor peptides with concomitant export. Sequence analysis revealed that this domain contains conserved Cys, His, and Asp residues that are also found in the protease domains of several nonlantibiotic bacteriocin ABC transporters, such as LagD,[197] ComA,[198] and CvaB.[199] The substrates of these proteins, lactococcin G precursor, ComC, and colicin V precursor, respectively, all contain leader peptides that have sequence homology with the class II lantibiotic leader peptides (**Figure 6**). *In vitro* proteolytic activity of these nonlantibiotic protease domains toward their natural substrates demonstrated cleavage after a Gly–Gly or Gly–Ala/Ser sequence, also known as the double-glycine motif (**Figure 6**).[45,63,64,200] Of the known bifunctional lantibiotic transporters, only the predicted protease domain from LctT (residues 1–150) has been overexpressed heterologously and shown to cleave both LctM-modified and unmodified lacticin 481 precursor peptide LctA at the double-glycine site.[100] Based on sequence homology with papain-like cysteine proteases, the protease domains of class II lantibiotic LanT proteins are thought to utilize conserved Cys and His for catalysis. Replacement of the conserved Cys with Ser or Ala in the LctT protease domain (LctT150) abolished proteolytic activity. However, mutation of the putative catalytic His to Ala resulted in a mutant that still cleaved LctA but after Gly−2 instead of Ala−1, suggesting a role of His in regioselectivity. Mutation of the conserved Asp in LctT150 did not affect the cleavage, indicating that it is not essential for catalysis.

The substrate specificity of proteolysis by class II lantibiotic LanT proteins has been probed both *in vivo* and *in vitro*. The importance of the double-glycine motif was revealed in two separate studies. Mutation of Gly−1 or Gly−2 to Ala in MutA in *Streptococcus mutans* blocked cleavage by MutT since accumulation of dehydrated premutacin II was observed in the cytoplasm.[99] *In vitro*, the protease domain of LctT150 also did not process the Ala−1Asp, Ala−1Lys, or Ala−1Ile mutants of LctA, illustrating the importance of this residue in the S1 position (**Figure 7**).[100] Replacement of Gly−2 in LctA with a charged residue (Glu or Lys) also prevented LctT150 from proteolytic processing. Much reduced enzyme activity of LctT150 was observed toward the Gly−2 to Val mutant of LctA, and only the Ala−1Gly mutation was tolerated by the enzyme. Collectively, these studies show that substrates with charged residues or bulky hydrophobic residues at the double-glycine

site were not accepted by LctT150. As discussed in Section 5.08.3.2, the leader peptide of LctA has been proposed to possess α-helical character between Asn−17 and Leu−3, which has been suggested to be a key recognition element for posttranslational modifications, proteolytic cleavage, and possibly export. As was observed with respect to dehydration by LctM, three Pro mutants of LctA at Leu−5, Glu−8, or Val−12 were poor substrates for LctT (**Figure 7**), in agreement with this hypothesis. On the other hand, mutants in which Leu−7 was changed to Lys or Glu, which negatively affected dehydration by LctM, were processed equally well as the wild-type LctA by LctT,[100] illustrating that residues that are important for one posttranslational modification are not necessarily important for a different modification during lantibiotic maturation.

A series of peptides were tested to investigate the substrate specificity of the LctT150 protease domain with respect to the P′ residues. Formation of the lanthionine rings was not required for proteolysis by LctT150, suggesting that the rings of the structural peptide are not important for substrate recognition. Nonconservative mutations of residues in the P1′–P5′ positions also did not affect proteolysis by LctT150, and several of the chimeric peptides (see Section 5.08.3.2) containing multiple simultaneous mutations in these positions were likewise processed. Thus, the residues C-terminal to the cleavage site appear not to be important, which also explains why two-component lantibiotic biosynthetic gene clusters contain only a single LanT protein despite dramatic differences in the sequences of the two structural peptides of their substrates (see for instance the sequences of the LtnA1 and LtnA2 peptides, **Figure 6**).[53,201]

Cytolysin is a two-component lantibiotic with hemolytic activity that is produced by pathogenic *Enterococcus faecalis*.[202] A serine protease designated CylA encoded in the cytolysin gene cluster shares 26% identity with EpiP and was postulated to be the enzyme responsible for proteolytic activation of cytolysin.[203] The CylA enzyme purified from the culture supernatant of *E. faecalis* FA2-2 lacked the N-terminal 95 amino acids of the primary translation product. Like EpiP, CylA was suggested to undergo autoproteolytic processing during maturation and did not contain a C-terminal cell wall anchor. Interestingly, when the *cylA* gene was disrupted, the *E. faecalis* mutant was defective in the production of cytolysin precursor $CylL_L$ or $CylL_S$. Instead, the mutant strain produced peptides much smaller than the expected prepeptides.[68] N-terminal sequencing revealed that both peptides lacked an N-terminal portion of the predicted precursors and started with an identical sequence: GlyAspValGlnAlaGlu (see **Figure 6**). The two peptides were named $CylL_L'$ and $CylL_S'$, and low-level hemolytic activity was observed when the two peptides were combined in high concentration. The 24- and 36-amino-acid peptides that were removed from the N-terminus of the cytolysin precursors to generate $CylL_L'$ or $CylL_S'$ contained a highly homologous C-terminal region that ended in Gly–Ser. The ABC transporter CylB encoded in the cytolysin gene cluster was proposed to be responsible for removing these N-terminal peptides, as it shares homology with the class II lantibiotic LanT proteins and contains an N-terminal protease domain. Removal of the remaining hexapeptide GlyAspValGlnAlaGlu from $CylL_L'$ and $CylL_S'$ to produce mature $CylL_L$ and $CylL_S$ was shown to be performed by CylA. Furthermore, incubation of CylA with synthetic substrates consisting of the last 17 residues of the cytolysin leader peptides attached to the first two residues in the structural region resulted in correct cleavage between Glu1 and Thr1 (for sequence, see **Figure 6**). This result confirmed the role of CylA in the final proteolytic step and suggested that neither the complete structural region of cytolysin nor the posttranslational modifications were necessary for CylA cleavage.

The precursor for the class III lantibiotic SapB contains a 21-amino-acid leader peptide.[72] Two ABC-transporters RamA and RamB, consisting of 636 and 608 amino acids, respectively, are encoded by the SapB gene cluster. RamA and RamB share 31% sequence identity; however, neither contains a serine protease or cysteine protease domain. Since no other candidate proteases are present in the gene cluster, the identity of the protease that removes the leader peptide remains elusive.

5.08.3.7 LanD Decarboxylases

The posttranslationally modified cross-link *S*-[(*Z*)-2-aminovinyl]-D-cysteine (AviCys; **Figure 1**) has been found at the C-terminus of the lantibiotics epidermin,[204] gallidermin,[205] cypemycin,[29] and mutacin III.[206] A group of decarboxylases, called LanD proteins, are encoded in the lantibiotic gene clusters and are responsible for this modification. The first LanD protein characterized was EpiD from the epidermin-producing strain *S. epidermis*.[207] EpiA and EpiD were separately overexpressed in *E. coli* and purified,

and EpiD was identified as a flavoenzyme containing a flavin mononucleotide (FMN) cofactor. Treatment of EpiA with EpiD produced a peptide that was 46 Da smaller in mass than EpiA corresponding to the loss of carbon dioxide and two hydrogen atoms.[208] The oxidative decarboxylation activity of EpiD was also demonstrated *in vivo* when the protein was coexpressed with EpiA in *E. coli*.[209] These results indicated that the dehydration or cyclization of EpiA was not necessary for EpiD modification. In addition, EpiD not only processed the structural region of EpiA without the leader peptide, but also accepted as substrate a truncated peptide corresponding to the C-terminal seven amino acids of EpiA (SerPheAsnSerTyrCysCys), and even SerTyrCysCys.[210] A library containing single amino acid alterations of the EpiA C-terminal heptapeptide was constructed to probe the substrate specificity of EpiD.[210] The results of incubation of this library with EpiD led to the formulation of a three-amino-acid consensus sequence for peptides that EpiD could process: [Val/Ile/Leu/Phe/Tyr/Trp]-[Ala/Ser/Val/Thr/Cys]-Cys.[210–212] A C-terminal cysteine residue was an absolute requirement as peptides with serine or homocysteine at this position were not decarboxylated. Furthermore, the cysteine had to contain both a free carboxylate and a free thiol, since the peptides SerPheAsnSerTyrCysCys-NH_2 and SerPheAsnSerTyrCysCys(SEt) were not processed by EpiD.[213] In order to obtain mechanistic information, a model peptide LysLysSerPheAsnSerTyrThrCys was ^{13}C-labeled at the β-carbon of the C-terminal cysteine.[214] After treatment with EpiD, the peptide was reduced by 46 Da in mass and NMR analysis of the product revealed the presence of an unusual (*Z*)-enethiol. The pK_a value of the enethiol group was determined to be 6.0, about 3 units lower than the thiol group of an unmodified C-terminal cysteine.[215] The addition of this enethiol to a Dha would yield AviCys, and this step is presumably catalyzed by LanC proteins. These results implied that decarboxylation occurred prior to cyclization.

Site-directed mutagenesis of EpiD identified Pro81, Ser83, Asn85, Gly93, and Asp96 as residues crucial for FMN binding.[216] Based on the crystal structure of EpiD, the strictly conserved His67 located in the active site is believed to be the active site base.[213] An X-ray structure of EpiD His67Asn complexed with a pentapeptide (AspSerTyrThrCys) showed that the substrate formed a β-strand embraced by a substrate-binding clamp composed of residues Pro143 to Met162 in a highly twisted antiparallel β-sheet conformation.[213] The precise geometry at the substrate recognition clamps of crystallographically independent monomers differed slightly, which implied flexibility to bind different residues and explained the broad substrate specificity of EpiD. The cocrystal structure also revealed the thiol group of the substrate C-terminal cysteine as the only group in proximity of N5 of the cofactor FMN, where oxidative attack typically occurs. Therefore, it was proposed that the cysteine thiol group is first oxidized to a thiolaldehyde followed by decarboxylation to form the enethiolate (**Figure 12**).

The lantibiotic mersacidin contains a C-terminal *S*-[(*Z*)-2-aminovinyl]-3-methyl-D-cysteine (AviMeCys), which was proposed to be synthesized in a similar mechanism as AviCys featuring oxidative decarboxylation of the C-terminal cysteine to an enethiol followed by its addition to a Dhb. The activity of MrsD, the decarboxylase encoded in the mersacidin gene cluster, was demonstrated *in vitro* toward the mersacidin precursor peptide MrsA.[211] Although MrsD shares 30% sequence identity with EpiD, MrsD differs from EpiD by its coenzyme requirement and substrate specificity. MrsD is a flavin adenine dinucleotide (FAD)-dependent enzyme and it did not process EpiA. The C-terminal peptide ThrLeuThrSerGluCysIleCys of MrsA could not be modified by MrsD either. Conversely, EpiD did not modify MrsA or the C-terminal peptide of MrsA. Like EpiD, MrsD is also a dodecamer, and the X-ray crystal structure of MrsD superimposed very well with that of EpiD.[217] Mutation of the conserved His to Asn completely eliminated the decarboxylation activity of MrsD, supporting a role as an active site base. Based on their similarity to bacterial Dfp proteins[218] and the salt tolerance protein AtHAL3a from *Arabidopsis thaliana*,[219] LanD proteins are considered members of the homooligomeric flavin-containing Cys decarboxylase (HFCD) family.

5.08.3.8 Hydroxylation

L-Asp hydroxylated at the β-carbon to generate *erythro*-3-hydroxyl-L-aspartic acid has so far only been detected in cinnamycin and the duramycins, lantibiotics produced by actinomycetes.[27,220] This modification has also been found in mammalian proteins, such as the vitamin K-dependent protein C, and the epidermal growth factor (EGF)-like domain in human plasma factor IX. Both bovine and human aspartyl-β-hydroxylases have been purified and characterized and their *in vitro* hydroxylation activity has been shown using proteins

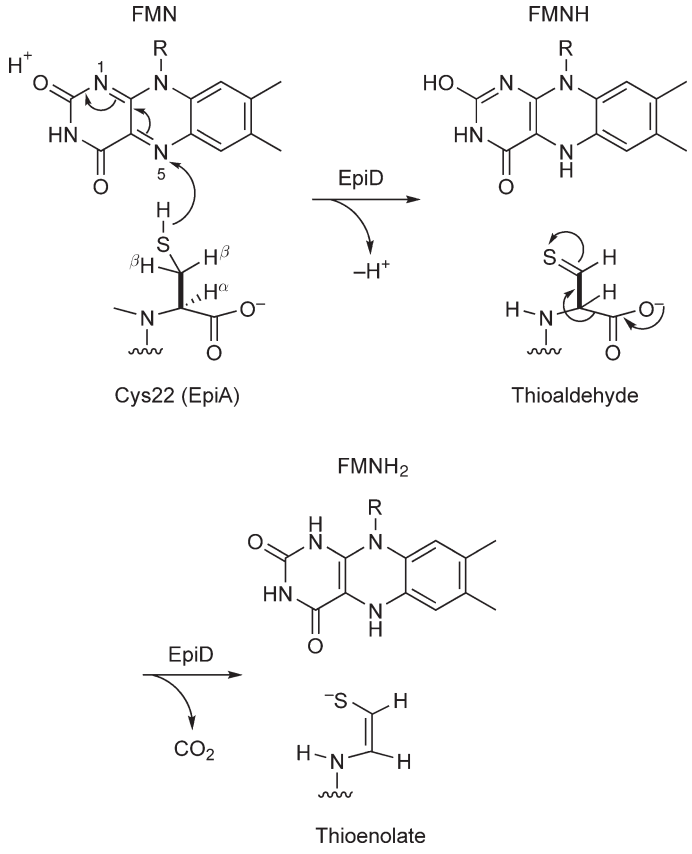


Figure 12 Proposed mechanism of decarboxylation and dehydrogenation of the C-terminal Cys of EpiA catalyzed by EpiD. Adapted with permission from C. Chatterjee; M. Paul; L. Xie; W. A. van der Donk, *Chem. Rev*. **2005**, *105*, 633–684.

containing EGF-like domains as substrates.[221–224] The mammalian aspartyl-β-hydroxylases are O_2/Fe(II)/α-ketoglutarate (α-KG)-dependent enzymes and release a stoichiometric amount of CO_2 per Asp hydroxylated.[221] They also hydroxylate Asn residues to produce *erythro*-3-hydroxyl-L-asparagine. The enzyme CinX encoded in the cinnamycin gene cluster has been identified as the hydroxylase responsible for the synthesis of the β-hydroxylated Asp both *in vivo* (M. J. Bibb, personal communication) and *in vitro* (E. Fogle and W. A. van der Donk, unpublished data). Recently, 3,4-dihydroxyproline and 4-hydroxyproline have been discovered in a novel lantibiotic microsporicin from *Microbispora* sp. ATCC PTA-5024.[28] The function of these hydroxylated prolines as well as their stereochemistry in microsporicin remains to be elucidated.

5.08.3.9 Other Modifications

In addition to the characteristic Lan and MeLan and dehydro amino acids, many additional and more rare posttranslational modifications have been detected in lantibiotics (**Figure 1**). The N-terminal modifications OPr and OBu are products of nonenzymatic hydrolysis of N-terminal Dha and Dhb, respectively, that are exposed after the leader peptide is removed. The OPr functionality has been found in lactocin S,[31] and the OBu group is present at the N-terminus of Pep5.[34,225] Epilancin K7 contains an N-terminal Hop group as demonstrated by NMR spectroscopy.[225] A Ser residue was found at this position in the precursor peptide sequence. Hence, the Hop group likely results from reduction of an OPr group, but the stereochemistry of the reduction is currently not known. Another lantibiotic, epicidin 280, has been proposed to possess an N-terminal Hop group as indicated by NMR analysis and the presence of a putative oxidoreductase EciO in the gene cluster, which was hypothesized as the enzyme catalyzing the reduction of an OPr group.[35]

Structural analysis of lactocin S produced by *Lactobacillus sake* L45 revealed the presence of three D-Ala residues.[31] Unlike the D-Ala in the natural peptide opioid dermorphin isolated from frog skin,[226,227] which is introduced via posttranslational modification of L-Ala, the D-Ala residues in lactocin S arise from L-Ser in the precursor peptide. One D-Ala was also found in the A1 peptide of the two-component lantibiotic lacticin 3147 and two D-Ala residues in the A2 peptide (**Figure 2**).[30,39] The mechanism of D-Ala synthesis has been shown to involve a two-step process featuring the dehydration of L-serine followed by stereospecific hydrogenation of the resulting Dha. The protein LtnJ encoded in the lacticin 3147 gene cluster has been postulated to catalyze the hydrogenation of Dha based on its similarity to zinc-containing alcohol dehydrogenases.[228] Deletion of the *ltnJ* gene *in vivo* led to the production of peptides 2 and 4 Da smaller in mass than LtnA1 and LtnA2, respectively. This result indicated the presence of Dha intermediates at the positions of D-Ala and supported the proposed mechanism and the role of LtnJ as the dehydrogenase. Single or double substitution of the D-Ala precursor L-Ser with L-Thr led to the production of peptides with Dhb at these positions, indicating that LtnJ cannot reduce Dhb.

In addition to the hydroxylated aspartic acid discussed in Section 5.08.3.8, a lysinoalanine (LysAla) bridge was also discovered in cinnamycin and the duramycins (see **Figures 1** and **2**). Furthermore, a posting in the UniProtKB/Swiss-Prot database (entry P38655) suggests that ancovenin, a lantibiotic structurally related to cinnamycin, which initially was not reported to contain the LysAla linkage,[229,230] also contains a LysAla bridge. The LysAla is probably formed by the addition of the ε-amine of Lys19 to Dha6, but because the stereochemistry at C2 of the LysAla is (*S*) (i.e., the L-configuration; **Figure 1**), it cannot be ruled out that the linkage is formed directly from Ser without the intermediacy of a Dha. The identity of the enzyme catalyzing this cross-linking reaction is currently unclear. LysAla is widely present in food and is believed to be generated by chemical dehydration of Ser during food processing (heat, high pH, etc.) and conjugate addition of Lys to the resulting Dha to produce both diasteromers.[231] LysAla also occurs naturally in body organs and tissues where its formation is proposed to be involved in the aging process.

The lantibiotic cypemycin isolated from a *Streptomyces* strain exhibits bis-methylation at Ala1 (Me_2N-Ala) and an L-*allo*-isoleucine at position 13.[29] The gene for its precursor peptide has not been reported, and hence the amino acid that is modified to generate the *allo*-Ile is not known. If it is introduced posttranslationally like all other modifications in lantibiotics, its most likely precursors would be either Ile or Leu and the posttranslational modification may involve a radical mechanism as neither amino acid is activated at Cβ or Cγ for heterolytic chemistry.

Several lantibiotics are believed to contain disulfides. The precursor peptide of sublancin 168 produced by *B. subtilis* 168 contains one Ser, one Thr, and five Cys residues in its structural region.[111] It is uncommon for lantibiotic precursors to have more Cys than Ser and Thr combined. Structural analysis of mature sublancin 168 revealed the presence of one Dha and one MeLan, leaving four Cys residues engaged in two disulfide linkages. The thiol-disulfide oxidoreductase BdbB is essential for sublancin 168 production and may be responsible for correct formation of its disulfide bonds.[112] The origin of the MeLan on the other hand is unclear, as a search of the fully sequenced genome of *B. subtilis* 168 reveals the gene for the sublancin prepeptide but no distinguishable dehydratase or cyclase. Sublancin 168 has also been proposed to contain an additional modification, as mass spectrometry determined a molecular mass 164.48 Da greater than expected from the amino acid composition. How the MeLan is introduced into sublancin and what the structure is of the additional modification remains to be resolved. The presence of a single disulfide bridge has also been reported for the Halα peptide of a recently discovered two-component lantibiotic, haloduracin.[53] Reduction of this disulfide did not affect the antimicrobial activity of the peptides. Bovicin HJ50 displayed a mass increase of 2.4 Da upon treatment with the reducing agent dithiothreitol (DTT), and therefore was predicted to possess a disulfide as well.[232]

A number of very rare modifications have been reported. Subtilin can undergo N-succinylation at late stages of cell growth resulting in a reduction in bioactivity.[233] A chlorinated tryptophan has been discovered in microsporicin[234] and is likely introduced by a flavin-dependent halogenase.[235] Furthermore, actagardine has been reported to contain a sulfoxide resulting from the oxidation of a C-terminal thioether.[236] The actagardine gene clusters from *Actinoplanes garbadinensis* and *Actinoplanes liguriae* have recently been sequenced, and two monooxygenases ActO and LigO were proposed to catalyze the oxidation.[237] A novel variant of actagardine with a nonoxidized MeLan was also isolated from *A. liguriae* and demonstrated similar antimicrobial activity to actagardine, suggesting that the sulfoxide may not be important for activity.

5.08.4 Biological Activities of Lantibiotics

Nisin is active at low concentrations (MICs in the low nanomolar range) against many strains of Gram-positive bacteria,[4] including drug-resistant strains[20] and the foodborne pathogens *Clostridium botulinum* and *Listeria monocytogenes*.[238–241] For many years, nisin was believed to be a cationic pore former but it was not clear why it was so much more active than typical pore-forming peptides. As described in detail below, it is now known that nisin recognizes lipid II in the bacterial membrane and that it forms pores that are made up of lipid II and nisin.[59] This docking model explains the selectivity that nisin displays toward bacterial membranes as well as its high potency. Although at present the molecular targets have only been determined for nisin, cinnamycin, and mersacidin and their structural relatives, it is believed that the conformational constraints imposed on lantibiotics by the thioether ring structures are also used for recognition of specific targets by family members for which the mode of action is currently not known. The thioether rings are proposed to confer structural stability, both in favor of their biological activities and against susceptibility to protease-mediated degradation. Loss of cyclic structures typically results in reduced or abolished antibiotic activity. For instance, studies in which one of the Cys or Ser/Thr residues was replaced with Ala in nisin,[105] Pep5,[242] epidermin,[106] mutacin II,[143,243] lacticin 481,[95] lacticin 3147,[136,142] and haloduracin[113] resulted generally in greatly reduced or completely abolished antimicrobial activity.

5.08.4.1 Lipid II Binding and Pore Formation

The prototypic lantibiotic nisin has been utilized in the food industry for over 40 years without development of widespread bacterial resistance. This observation is likely due to nisin's dual mode of action: binding to lipid II, a crucial precursor in peptidoglycan biosynthesis (**Figure 13**), and formation of pores within the cell membrane that are made up of lipid II and nisin.[20,58] A solution NMR structure of a 1:1 complex of nisin and a lipid II variant with a shortened isoprenoid membrane anchor demonstrated that the N-terminal amides of Dhb2, Ala3, Ile4, Dha5, and Abu8 in the A- and B-rings make hydrogen-bonding contacts with the pyrophosphate moiety of lipid II (**Figure 14**).[49] As the N-terminus of nisin sequesters lipid II and disrupts cell wall biosynthesis, its C-terminus inserts into the membrane in a perpendicular orientation with respect to the bilayer surface[244] to form stable pores. Pyrene fluorescence and circular dichroism studies have shown that each pore contains four lipid II and eight nisin molecules.[245,246] The distance between two labeled lipid II molecules was estimated to be about 18 Å[246] with a pore diameter of 2 nm.[247] Lipid II is also the target of other peptide antibiotics such as

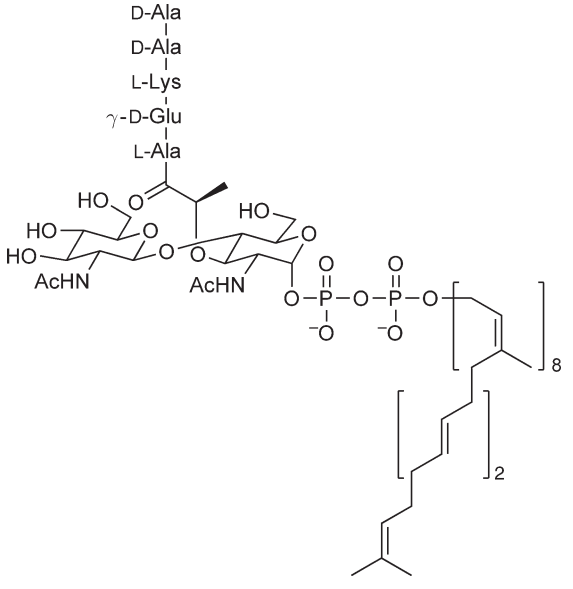


Figure 13 The structure of lipid II.

Figure 14 Solution structure of nisin complexed to a lipid II variant containing a shortened prenyl chain. Lipid II is shown in spheres with its pyrophosphate moiety colored in blue (phosphorus) and red (oxygen). Nisin is represented in sticks (A-ring carbons in cyan, B-ring carbons in magenta, C-ring carbons in green, D-ring carbons in blue, E-ring carbons in orange, and the remaining carbons in gray). In all five rings, the color scheme is nitrogen in blue, oxygen in red, and sulfur in yellow. The figures were generated using PyMOL (http://www.pymol.org/).

vancomycin and ramoplanin (see Chapter 2.05). As would be expected for a common target, treatment of *Micrococcus luteus* with ramoplanin prevented pore formation by nisin and epidermin,[21] and strongly decreased mersacidin binding to the cell surface.[57] Similarly, vancomycin protected *Micrococcus flavus* cells against membrane leakage induced by nisin but not by magainin.[20] However, as shown by the nisin–lipid II structure, the lantibiotics bind to a different part of lipid II than vancomycin, which interacts with the L-Lys-D-Ala-D-Ala segment of the pentapeptide.[248,249] Indeed, nisin and mersacidin are active against vancomycin-resistant enterococci.[20,24] Although nisin, vancomycin, teicoplanin, and ramoplanin interact with the same target, nisin is unique in that it subsequently forms pores that include lipid II as an essential constituent.[246] For instance, when lipid II is present in membranes, nisin's pore-forming efficiency is increased 1000-fold, an increase that is not seen with other pore-forming peptides like magainin.[20] Binding of an antibiotic to a complex biosynthetic intermediate like lipid II has certain advantages over binding to a single enzyme involved in peptidoglycan assembly because changing the structure of lipid II is much more demanding on a microbe than changing the structure of the active site of one enzyme, thereby decreasing the odds of bacterial resistance. For instance, eight successive enzymes are required for the biosynthesis of lipid II from UDP-GlcNAc.[250,251] That resistance can nevertheless develop has been demonstrated in vancomycin resistance when bacteria change the D-Ala-D-Ala unit of lipid II to D-Ala-lactate.[252,253]

The ability to generate nisin analogues discussed in Section 5.08.6 allowed the mechanism of action of nisin to be investigated in great detail. Mutation of Val32 to Lys or Glu, thereby introducing an additional positive or negative charge and also preventing dehydration of Ser33,[138] had very modest effects with respect to antimicrobial activity against certain test strains.[60] Therefore, the C-terminus is relatively unimportant, in keeping with the observation that epilancin K, which shares a very similar C-terminal double-ring system with nisin, does not appear to interact with lipid II.[21] On the other hand, several findings support the conclusions from the NMR structure that the N-terminus of nisin is essential for lipid II binding. An inactive nisin fragment (nisin 1–12) antagonizes the bactericidal activity of nisin, suggesting that it competes for the same binding site.[254] Moreover, opening of the A-ring results in more than 500-fold reduction of biological activity,[255] whereas complete removal of the D- and E-rings by proteolysis results in only a 100-fold decrease in potency.[254] Mutagenesis studies on amino acids between the C- and D-rings of nisin identified a hinge region, consisting of three residues (Asn20, Met21, and Lys22), that gives nisin the conformational flexibility to traverse

the lipid bilayer.[256] Some of the mutants in this region lost pore-forming ability but were still potent antibiotics due to inhibition of cell wall biosynthesis, thereby showing that this activity in itself is sufficient for cytotoxicity. A model has been proposed to explain the extensive experimental data that have been collected regarding pore formation by nisin.[60,244,245] The N-terminal rings of nisin are believed to bind to the disaccharide-pyrophosphate region of lipid II, whereas the positively charged C-terminus initially interacts with the headgroups of the lipids in the membrane bilayer. Multiple molecules of the lipid II–nisin complex[246] subsequently aggregate and form a pore of defined uniform structure. Whereas the stoichiometry of lipid II to nisin in solution is 1:1,[49,60] the stoichiometry in the pore is 1:2 as it is made up of four lipid II and eight nisin molecules.[245] How the different stoichiometry in the pore affects the structure of the lipid II–nisin complex that was determined by NMR spectroscopy remains to be established. Interestingly, pore formation and disruption of cell wall biosynthesis appear to take place in patches in the cell membrane. Based on studies with fluorescently labeled nisin, the molecule not only binds lipid II but also removes it from its functional location in the cell, thereby impairing cell growth and cell division.[257]

Several other lantibiotics such as epidermin,[21] mutacin 1140,[258] lacticin 3147,[259,260] actagardine,[21,261] plantaricin,[262] and mersacidin[50,263] also bind to lipid II and inhibit peptidoglycan biosynthesis. Some of these compounds such as epidermin, actagardine, and mutacin 1140 have very similar A- and B-rings as nisin (**Figure 2**) whereas plantaricin and the A1 peptide of lacticin 3147 have structural similarity with mersacidin (**Figure 15**). Unlike nisin, mersacidin does not form pores within cell membranes. An NMR study of

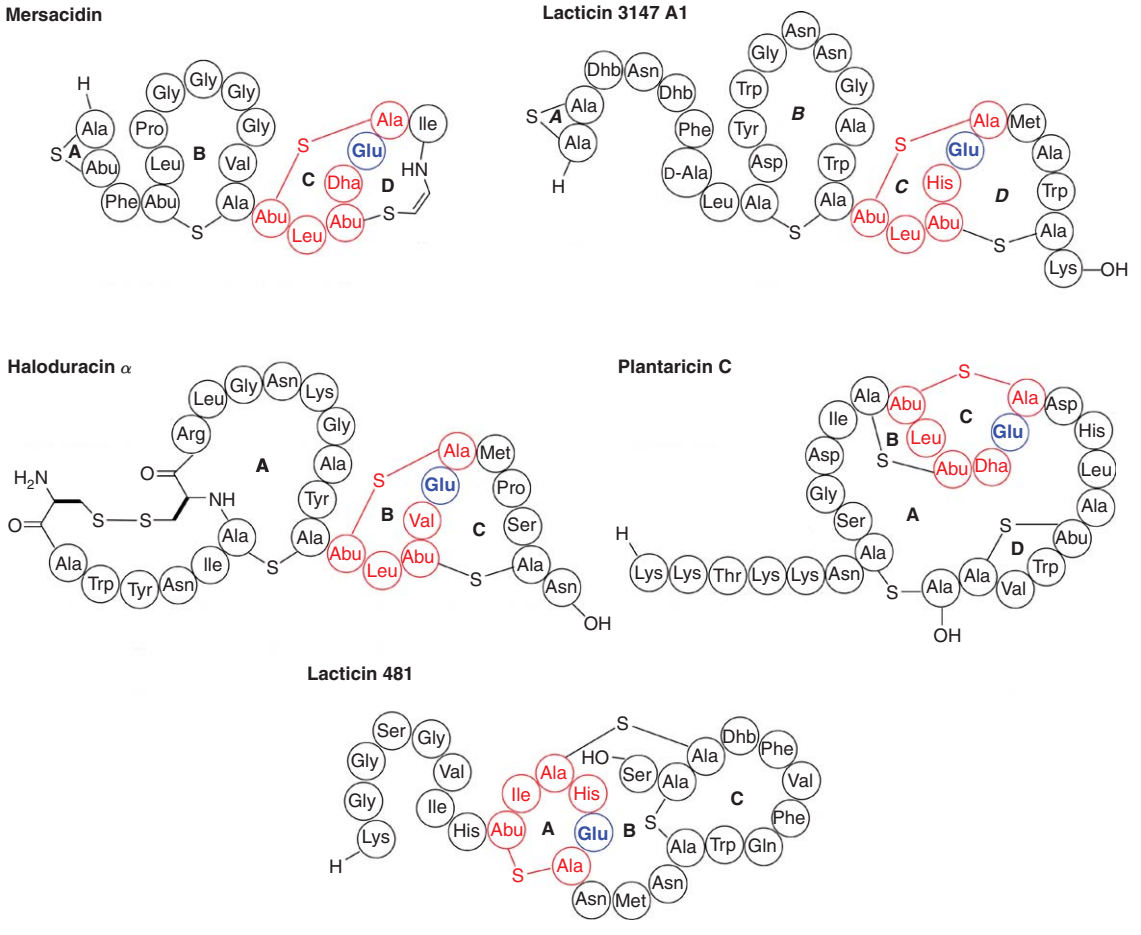


Figure 15 Structures of lantibiotics containing a conserved methyllanthionine ring displayed in red. The glutamate that is essential for both mersacidin and haloduracin antimicrobial activity is shown in blue. The structure of plantaricin C has not yet been unambiguously determined and the proposed structure is shown. The stereochemistry of the lanthionine between residues 7 and 27 is not known.[264]

dodecylphosphocholine micelles demonstrated a change in the conformation of mersacidin in the presence of lipid II, suggesting a direct interaction between the two molecules.[50] By binding to the disaccharide-pyrophosphate region of lipid II,[57] mersacidin is believed to prevent the transglycosylation step of cell wall biosynthesis. In the case of the two-component lantibiotic lacticin 3147, binding to lipid II and pore formation require two posttranslationally modified peptides to act in synergy to affect their bactericidal properties. In a recent model, lacticin 3147 A1, with a C-ring that has similarities to that of mersacidin (**Figure 15**), was proposed to first bind lipid II, causing a change in its conformation such that lacticin 3147 A2 can bind the LtnA1–lipid II complex. The A2 peptide then promotes deeper insertion into the membrane and pore formation.[259] Interestingly, the A2 peptide could be substituted with staphylococcin C55β, the β-peptide from a related two-component lantibiotic (55% sequence identity between lacticin 3147 A2 and staphylococcin C55β), resulting in nanomolar bioactivity of the hybrid pair.[265] Similarly, staphylococcin C55α and lacticin 3147 A2 constituted an active two-component system. The C-rings of lacticin 3147 A1 and mersacidin are very similar to the C-ring in plantaricin C, the B-rings in the α-peptides of other two-component lantibiotics such as haloduracin, and the A-rings of the lacticin 481 subgroup of lantibiotics (**Figure 15**). Plantaricin C was recently shown to indeed interact with lipid II and its biosynthetic precursor lipid I (lacking the GlcNAc unit).[262] The observed lipid II-dependent activities of mersacidin, lacticin 3147, and plantaricin C have led to the speculation that the A-ring of the lacticin 481 subgroup may be similarly important for antimicrobial activity,[3,95,136] with the glutamate in this ring being crucial for the antimicrobial properties of mersacidin.[141] To address the importance of this glutamate in lacticin 481, the mutant LctA-Glu13Ala was incubated with LctM resulting in the expected four dehydrations. Subsequent proteolytic removal of the leader peptide resulted in a lacticin 481 analogue that surprisingly still displayed bioactivity against a *L. lactis* reporter strain.[102] Hence, whether the A-ring of the lacticin 481 group of lantibiotics is indeed interacting with lipid II in a similar manner as mersacidin remains to be confirmed.

5.08.4.2 Cinnamycin: Phosphatidylethanolamine Binding

Some lantibiotics have other activities in addition to their antimicrobial capacity. In particular, the members of the cinnamycin group may have potential medical applications for the treatment of respiratory, immune, and cardiovascular health disorders. The cinnamycin-like lantibiotics include cinnamycin (**Figure 2**), duramycin, duramycin B, duramycin C, and ancovenin. These 19-amino-acid peptides contain one Lan, two MeLan, an *erythro*-3-hydroxyl-L-aspartate residue, and a LysAla linkage. They share extensive sequence homology, differing in, at most, six amino acids. Both cinnamycin and ancovenin inhibit angiotensin-converting enzyme,[266,267] a well-known regulator of blood pressure. The duramycins and cinnamycin inhibit phospholipase A2 (PLA2) with an IC_{50} (half maximal inhibitory concentration of a substance) of 1 μ mol l^{-1} by binding to its substrate PE.[25,27,51,266,268–270] PE is one of the most abundant phospholipids in the membrane bilayers of animals, plants, and microbes. Liberation of arachidonic acid from PE by PLA2 leads to the biosynthesis of prostaglandins and other mediators of inflammation; thus, it represents a potential target to regulate inflammation and combat cardiovascular diseases.

Specific peptide–phospholipid interactions are rare; thus the precise features of cinnamycin–PE binding are of great interest. Cinnamycin selectively binds to PE in an equimolar (1:1) complex.[25,51,52,268,270,271] How cinnamycin would gain access to PE was unclear since the latter typically resides on the inner leaflet of the membrane. It was reported that cinnamycin has the ability to induce transbilayer phospholipid movement, for PE and other lipids as well.[272] Cinnamycin-induced lipid flip-flop was accomplished in a PE concentration-dependent manner, such that no rearrangement occurred when PE was absent. How this is achieved without an initial interaction with PE on the outer leaflet is not known.

Structural details of the cinnamycin–PE complex including glycerophosphoethanolamine headgroup orientation and interactions with the amino acid side chains of cinnamycin have been determined using NMR spectroscopy. The NMR structure of cinnamycin bound to lysophosphatidylethanolamine (lysoPE) showed a cylindrical complex with a diameter of 11 Å and length of 26 Å. In this structure, cinnamycin did not appear to directly contact the lipid chain.[51] A specific ionic interaction was observed between the carboxylate group of the β-hydroxy-Asp15 residue in cinnamycin and the ammonium ion of the glycerophosphoethanolamine headgroup. This headgroup was bound in a hydrophobic pocket created by a folded structure in the central

part of cinnamycin. In particular, Gly8, Pro9, and Val13 directly contact the glycerol moiety. The limited amount of space in this pocket is thought to prevent binding of larger molecules, and thus contributes to the high selectivity of cinnamycin for PE. The secondary amine in the LysAla linkage was initially suspected to bind to the phosphate group of the lipid;[51] however, these groups were shown to be too far apart for ionic interaction (11.7 Å).[52]

High-sensitivity isothermal titration calorimetry (ITC) has been used to measure the thermodynamic binding parameters upon cinnamycin–PE complex formation.[269,273] The authors concluded that the binding constant ranged from 10^6 to 10^8 mol l^{-1} depending on the lipid matrix. Both the enthalpy and molar heat capacity for complex formation were indicative of a hydrophobic binding mechanism. Overall, an entropy–enthalpy compensation mechanism was proposed in which complex formation was entropy driven at lower temperatures (10 °C) and enthalpy driven at higher temperatures (50 °C).[273] These ITC data also suggested that cinnamycin does interact with the hydrocarbon acyl chains of PE. At least one acyl chain was required for binding and the optimum chain length was found to be eight methylene groups.[269] Diacylphosphatidylethanolamine (diacylPE) bound more strongly than lysoPE.

5.08.4.3 Morphogenetic Activities

The most recent new activity was discovered in the peptides SapB and SapT produced by *S. coelicolor* and *S. tendae*, respectively.[11] These peptides are believed to self-assemble at air–water interfaces and are important for the formation of nascent aerial hyphae during sporulation.[11,73,274] The molecular details of this activity remain to be elucidated.

5.08.5 Potential Applications of Lantibiotics

In addition to the widespread application of nisin in the food industry,[4] some interesting clinical applications have been proposed and/or are in clinical trials and a brief overview is provided here. The reader is referred to other reviews for a more extensive listing.[9,10,275] The lantibiotic mutacin 1140 is produced by *S. mutans*, the major causative agent of dental caries. The compound is effective against many strains of the same species, prompting interest in this peptide as an agent to prevent dental caries.[276,277] Along similar lines, lozenges seeded with a salivaricin A-producing strain of *Streptococcus salivarius* have been introduced as a probiotic to combat *Streptococcus pyogenes* in the oral cavity and prevent halitosis.[86,278] Lacticin 3147 has also been reported to eliminate *S. mutans* from human saliva. Moreover, this compound shows strong activity against multidrug-resistant strains[279,280] and was also shown recently to have spermicidic activity.[281] Perhaps, the most promising compound discovered to date is the recently reported microbisporicin (formerly 107891)[234] with very potent bactericidal activity against clinical staphylococcal and enterococcal isolates as well as several Gram-negative bacteria.[28] Using the same screening procedure that resulted in the discovery of microbisporicin, planosporicin produced by the uncommon actinomycetes *Planomonospora* sp. was recently identified. The compound resembles mersacidin and has a promising spectrum of *in vivo* activity against Gram-positive pathogens including multidrug-resistant clinical isolates.[282] Not all applications of lantibiotics are in antimicrobial settings, however, as duramycin has been shown to increase chloride transport in nasal epithelial cells of cystic fibrosis patients,[283] which in turn increases the fluidity of mucus in the lungs and airway, decreasing the patient's susceptibility to infections.

5.08.6 Lantibiotic Engineering and Structure–Activity Studies

The cloning of the gene clusters involved in the biosynthesis of many lantibiotics laid the foundation for genetic protein engineering aimed at *in vivo* production of novel compounds with potentially interesting properties. Many studies have indicated the feasibility of changing the molecular structures of lantibiotics by mutagenesis of the prelantibiotic genes.[105] So far, engineering of the nisin structure has been most extensively investigated[105] and the resulting variants have contributed much to our current understanding of its mode of action

(Section 5.08.4). Engineered expression systems have also been established for subtilin,[109] Pep5,[140] epidermin and gallidermin,[106] mutacin II,[143] lacticin 3147,[135,136,142] and mersacidin.[141,284] In these studies, the immunity genes were typically also required for generating successful expression systems.

Replacement of Dha by Dhb and *vice versa* has been reported for several lantibiotics. Replacement of Ser at position 5 by Thr led to the production of Dhb instead of Dha in mature nisin Z.[137] The mutant exhibited increased resistance to chemical degradation, but this was accompanied by a 2- to 10-fold reduction in bioactivity toward various indicator strains. In contrast, replacement of Dha at position 2 in nisin Z with Dhb resulted in a mutant that was twice as active as native nisin Z.[105] The gallidermin variant Dhb14Dha did not exhibit any noticeable decrease in activity,[105] and the Dhb10Dha mutant of mutacin II also showed similar activity as wild type.[143] The promiscuity of the biosynthetic enzymes is also well illustrated by the production of a nisin mutant with a Dhb residue at position 18 in place of Gly after introduction of a Thr codon in *nisA*,[137] and the analogous introduction of a novel Dha in place of a Lys at position 18 in Pep 5.[242]

Generally, removal of dehydro amino acids in lantibiotics reduces their biological activity. For instance, strains expressing nisin in which either Dha33 was replaced by Ala or both Dha5 and Dha33 were substituted with Ala resulted in greatly reduced activity (about 1% of wild-type nisin-producing strains). Unfortunately, it was not established whether this reduction of activity was due to a less active antimicrobial peptide or due to reduced production in the engineered system.[285] Replacement of Dhb at positions 16 and 20 in Pep5 by Ala also reduced its activity,[242] and changing Dha16 in mersacidin to Ile greatly reduced its activity toward several indicator strains.[141] Not all amino acid substitutions are tolerated by the biosynthetic machinery, however, and sometimes mutations lead to abolished lantibiotic production.[106,136,143]

Alterations in the Lan and MeLan structures have also been accomplished. As mentioned in Section 5.08.4, removal of the thioether rings in lantibiotics typically results in loss of or greatly reduced antimicrobial activities.[105,106,136,140,143,242,243] In a few studies, Lan rings have been replaced with MeLan by substituting a Thr in place of a Ser residue. In the cases of epidermin and gallidermin, antibiotic activity was reduced but not lost,[106] and conversion of individual Lan to MeLan and MeLan to Lan yielded lacticin 481 analogues that were still able to inhibit bacterial growth.[95] Similar results were obtained *in vivo* with mutacin II when mutation of Thr10 to Ser successfully changed a MeLan to a Lan linkage and retained wild-type level activity.[143] On the other hand, substitution of Ser at position 3 in nisin Z with Thr gave rise to MeLan instead of Lan with a dramatic reduction in activity.[105] Replacement of Thr13 with Cys produced a disulfide in place of MeLan in nisin Z resulting in reduced activity.[286] Intriguingly, a fourth thioether bridge (MeLan) was introduced between positions 16 and 19 in Pep5 by the mutation Ala19Cys.[242] This methyllanthionine increased proteolytic stability against the proteases chymotrypsin and Lys-C but also resulted in a significant decrease in antimicrobial activity. This decrease in activity may be due to rigidification of the flexible central region that is thought to aid in pore formation by Pep5. Other Pep5 analogues in which ring structures had been deleted displayed a pronounced susceptibility toward proteolysis. Collectively, these studies on changing the thioether bridges reiterate the importance of the Lan/MeLan rings for antibiotic activity.

In addition to substitutions of the residues that are posttranslationally modified, mutants have been reported in which other amino acids in the polypeptide were replaced. For instance, two nisin variants with higher solubility than the parent compound were produced by substitution of Asn27 or His31 with Lys,[139] and the introduction of two additional Cys residues in nisin Z (Ser5Cys and Met17Cys) resulted in an engineered disulfide that required a reducing agent for bioactivity.[286] Several nisin mutants have been reported in which the residues in the so-called hinge region (Asn20, Met21, Lys22) were altered,[52,60] resulting in very informative changes in the bactericidal activities (see Section 5.08.4). Exchanging Glu4 of subtilin for Ile increased the biological activity three- to fourfold compared to the wild type and significantly slowed chemical modification of Dha5.[107,109,287] Interestingly, the Leu6Val gallidermin mutant was twice as active as the wild type against *M. luteus*, while the mutants Dhb14Pro and Ala12Leu showed increased resistance to proteolytic degradation.[106] Two mutants of MrsA were expressed in an engineered host[141] and the corresponding mersacidin analogue Glu17Ala-mersacidin had strongly reduced activity whereas Phe3Leu-mersacidin displayed activity closer to the wild-type lantibiotic. Heterologous expression in *Streptomyces lividans* of the *cin* cluster containing mutated *cinA* genes resulted in the production of Arg2Lys- and Phe10Leu-cinnamycin, which correspond to duramycin and duramycin B.[80] The most comprehensive mutagenesis study of a lantibiotic has been reported for the two peptides of lacticin 3147 in which every residue in each of the peptides was replaced by Ala or Gly.[136]

Remarkably, no less than 36 of the total 59 amino acids could be replaced without complete loss of bioactivity. Interestingly, among the amino acids that were not required for activity were the residues involved in the A-ring of lacticin 3147 A1. On the other hand, disruption of the six other cross-links resulted in the elimination of bioactive compounds produced by the engineered strain. Whether the loss of bioactivity was due to abolishment of lantibiotic production or lack of activity of the resulting mutant could not be distinguished. This study was further expanded through the generation of random mutants of the two peptides of lacticin 3147, resulting in 18 more mutant peptides.[142] These two studies illustrate once more the promiscuity of the biosynthetic machinery with respect to substrate recognition and processing. The importance of the D-Ala residues in lacticin 3147 has also been probed by site-directed mutagenesis. The activity was significantly decreased when D-Ala was replaced with L-Ala in either peptide, and the relative activity was reduced to an even greater extent when a larger residue L-Val was introduced. Peptides in which D-Ala was changed to nonchiral Dhb or Gly residues were produced in low amounts and their relative bioactivity was also decreased.

The important advancements in *in vivo* protein engineering of the lantibiotics have greatly contributed to a better understanding of lantibiotic biosynthesis and antimicrobial activity. However, to date, very few mutant lantibiotics have been generated with improved antimicrobial activities. These observations suggest that nature may already have optimized the antimicrobial activity of these compounds using the same tools, that is, mutagenesis with 20 amino acids, although the optimal activity from the perspective of the producing strain may be very different from optimal activity from a therapeutic viewpoint. Another contributing factor to the absence of more potent compounds produced by genetically engineered lantibiotic producers may lie in the breakdown of self-immunity in cases where more active compounds are actually generated, resulting in shutdown of production. Finally, the nisin–lipid II structure in **Figure 14** may provide a further rationalization why to date no significantly improved analogues have been reported in the bioengineering efforts as all contacts are made by main-chain amides. However, it is important to remember that currently no structural information is available for the lipid II–nisin and nisin–nisin interactions in pores and that these may provide opportunities for engineering improved variants. Furthermore, improving lantibiotics for therapeutic use does not necessarily require increased potency as enhancing other properties such as bioavailability, metabolic stability, and therapeutic index is equally important.

Engineering of the lantibiotic biosynthetic processes *in vitro* with expressed and purified biosynthetic proteins has several conceptual advantages over genetic engineering of lantibiotics. Perhaps most obviously, the structures of the prepeptides are not limited by the physiological amino acids and an expanded functional space can be explored through the use of solid-phase peptide synthesis. In addition, peptide synthesis is particularly amenable to combinatorial techniques thereby dramatically increasing the number of rapidly accessible substrate candidates. Because of the *in vitro* nature of the approach, degradation of products is not a problem, nor will cytotoxic or regulatory properties of the products be a concern. Intriguingly, based on recent experiments using nonpeptidic linkers between the leader peptide and the structural region that were accepted by lacticin 481 synthetase,[96,146] it may prove possible to use nonpeptide structures in part of the structural region in order to produce even more stable molecules. It should be noted, however, that some advantages of genetic engineering are lost using *in vitro* methods. Most notably, molecular biology approaches produce rapidly renewable sources of manipulated genes and organisms, which is not true for chemically synthesized molecules and purified enzymes. A potential solution might be the utilization of the *in vivo* amber codon suppression methodology[288] to generate mutant substrates with unnatural amino acids at desired positions.

The *in vitro* approach to lantibiotic engineering has thus far focused on exploring the substrate specificity of the lacticin 481 synthetase LctM. Nonproteinogenic amino acids have been incorporated into its substrate LctA by using expressed protein ligation (EPL)[289,290] (see Chapter 5.12) and more recently through the use of copper-catalyzed ligation strategies ('click' chemistry).[96] These studies showed that LctM was able to dehydrate Thr analogues in which the methyl group on the β-carbon was substituted by vinyl, ethyl, ethynyl, and propynyl groups.[291] The enzyme also catalyzed the regio- and chemoselective cyclization of a series of Cys analogues including D-cysteine, L-selenocysteine, L-homocysteine, and (*S*)-β^3-homocysteine resulting in novel cross-links.[95,188,189] The ability of LctM to catalyze the cyclization of homocysteine onto Dha was utilized for the preparation of a conotoxin analogue in which the disulfide in the natural product was replaced by a homolanthionine structure that unlike the disulfide is redox stable.[96]

5.08.7 Outlook

Many questions still remain with respect to both the biosynthesis and mode of action of lantibiotics. Whereas the targets of the nisin, mersacidin, and cinnamycin groups are now known, the mechanism of action of many other lantibiotics (e.g., lacticin 481, Pep5, and sublancin) is still unclear. Additionally, the biosynthetic pathways still hold many unresolved questions, including the molecular recognition that, on the one hand, allows the synthetases their high level of substrate promiscuity, and at the same time provides exquisite control over the regioselectivity of cyclization. Similarly, the biosynthetic enzymes for many of the 15 currently known posttranslational modifications in lantibiotics have not yet been identified, with the rate of discovery of new lantibiotics and new modifications recently outpacing the characterization of new biosynthetic enzymes. Currently, the most common commercial application for lantibiotics is as a preservative in the food industry to combat foodborne pathogens and spoilage bacteria, but many other uses are under investigation. Furthermore, several lantibiotics have shown potent activities against multidrug-resistant pathogenic bacterial strains. Combined with the development of new techniques to alter the structures of lantibiotics, the future will likely see detailed structure–activity relationship (SAR) studies that may result in improved variants.

The promise of using lantibiotic synthetases is not limited to the production of lantibiotic analogues. These enzymes may also find application in installing dehydro amino acids or lanthionine rings into other synthetic targets. Cyclic lanthionine-containing peptides have found use as mimics of natural products that contain disulfide bridges or as structures that limit the conformational flexibility of bioactive compounds. The use of the lantibiotic biosynthetic machinery as an alternative to synthetic organic chemistry for the preparation of such compounds is increasingly being demonstrated.

Abbreviations

αKG	α-ketoglutarate
ABC transporter	ATP-binding cassette transporter
Abu	L-α-aminobutyric acid
ATCC	American Type Culture Collection
AviCys	*S*-aminovinyl-D-cysteine
AviMeCys	*S*-aminovinyl-3-methyl-D-cysteine
CBD	chitin binding domain
Dha	2,3-dehydroalanine
Dhb	(*Z*)-2,3-dehydrobutyrine
diacylPE	diacylphosphatidyl ethanolamine
double-glycine motif	an amino acid sequence consisting of GlyGly, GlyAla, or GlySer that is proposed to be recognized by some lantibiotic proteases
DTT	dithiothreitol
EGF	epidermal growth factor
EPL	expressed protein ligation
FAB-MS	fast atom bombardment mass spectrometry
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FTMS	Fourier transform mass spectrometry
GRAS	generally recognized as safe
HFCD	homooligomeric flavin-containing Cys decarboxylase
Hop	2-hydroxypropionyl
IC_{50}	half maximal inhibitory concentration of a substance
ICP-MS	inductively coupled plasma mass spectrometry
ITC	isothermal titration calorimetry
LAB	lactic acid bacteria
Lan	(2*S*,6*R*)-lanthionine

LanA	lantibiotic precursor peptide
LanB	lantibiotic dehydratase
LanC	lantibiotic cyclase
LanD	lantibiotic decarboxylase
LanM	lantibiotic bifunctional dehydratase and cyclase
LanP	lantibiotic-specific serine protease
LanT	lantibiotic-specific ABC transporter (sometimes with a protease domain)
LysAla	lysinoalanine
lysoPE	lysophosphatidyl ethanolamine
MAPK	mitogen-activated protein kinase
MeLan	(2*S*,3*S*,6*R*)-3-methyllanthionine
MIC	minimum inhibitory concentration
OBu	2-oxobutyryl
OPr	2-oxopropionyl
PE	phosphatidylethanolamine
PLA2	phospholipase A2
PMSF	phenylmethylsulfonylfluoride, serine protease inhibitor
RTS	rapid translation system, a transcription/translation system used for polypeptide production *in vitro*
Sec	general secretory pathway
UDP-GlcNAc	uridine diphosphate *N*-acetylglucosamine

References

1. N. Schnell; K.-D. Entian,; U. Schneider; F. Götz; H. Zahner; R. Kellner; G. Jung, *Nature* **1988**, *333*, 276–278.
2. C. Chatterjee; M. Paul; L. Xie; W. A. van der Donk, *Chem. Rev.* **2005**, *105*, 633–684.
3. J. M. Willey; W. A. van der Donk, *Annu. Rev. Microbiol.* **2007**, *61*, 477–501.
4. J. Delves-Broughton; P. Blackburn; R. J. Evans; J. Hugenholtz, *Antonie van Leeuwenhoek* **1996**, *69*, 193–202.
5. L. A. Rogers, *J. Bacteriol.* **1928**, *16*, 321–325.
6. L. A. Rogers; E. O. Whittier, *J. Bacteriol.* **1928**, *16*, 211–229.
7. E. Gross; J. L. Morell, *J. Am. Chem. Soc.* **1971**, *93*, 4634–4635.
8. H. Dodd; M. Gasson; M. Mayer; A. Narbad, Identifying Lantibiotic Gene Clusters and Novel Lantibiotic Genes. WO 2006/111743 A2, October 26, 2006.
9. P. D. Cotter; C. Hill; R. P. Ross, *Curr. Protein Pept. Sci.* **2005**, *6*, 61–75.
10. P. D. Cotter; C. Hill; R. P. Ross, *Nat. Rev. Microbiol.* **2005**, *3*, 777–788.
11. J. M. Willey; A. Willems; S. Kodani; J. R. Nodwell, *Mol. Microbiol.* **2006**, *59*, 731–742.
12. R. S. Roy; A. M. Gehring; J. C. Milne; P. J. Belshaw; C. T. Walsh, *Nat. Prod. Rep.* **1999**, *16*, 249–263.
13. K. Severinov; E. Semenova; A. Kazakov; T. Kazakov; M. S. Gelfand, *Mol. Microbiol.* **2007**, *189*, 8772–8785.
14. S. Duquesne; D. Destoumieux-Garzon; S. Zirah; C. Goulard; J. Peduzzi; S. Rebuffat, *Chem. Biol.* **2007**, *14*, 793–803.
15. G. Vassiliadis; J. Peduzzi; S. Zirah; X. Thomas; S. Rebuffat; D. Destoumieux-Garzon, *Antimicrob. Agents Chemother.* **2007**, *51*, 3546–3553.
16. T. Kazakov; A. Metlitskaya; K. Severinov, *J. Bacteriol.* **2007**, *189*, 2114–2118.
17. M. S. Donia; B. J. Hathaway; S. Sudek; M. G. Haygood; M. J. Rosovitz; J. Ravel; E. W. Schmidt, *Nat. Chem. Biol.* **2006**, *2*, 729–735.
18. B. F. Milne; P. F. Long; A. Starcevic; D. Hranueli; M. Jaspars, *Org. Biomol. Chem.* **2006**, *4*, 631–638.
19. O. Buczek; G. Bulaj; B. M. Olivera, *Cell Mol. Life Sci.* **2005**, *62*, 3067–3079.
20. E. Breukink; I. Wiedemann; C. van Kraaij; O. P. Kuipers; H. Sahl; B. de Kruijff, *Science* **1999**, *286*, 2361–2364.
21. H. Brötz; M. Josten; I. Wiedemann; U. Schneider; F. Götz; G. Bierbaum; H.-G. Sahl, *Mol. Microbiol.* **1998**, *30*, 317–327.
22. E. Breukink; B. de Kruijff, *Biochim. Biophys. Acta* **1999**, *1462*, 223–234.
23. U. Pag; H. G. Sahl, *Curr. Pharm. Des.* **2002**, *8*, 815–833.
24. H. Brötz; G. Bierbaum; A. Markus; E. Molitor; H.-G. Sahl, *Antimicrob. Agents Chemother.* **1995**, *39*, 714–719.
25. F. Märki; E. Hanni; A. Fredenhagen; J. van Oostrum, *Biochem. Pharmacol.* **1991**, *42*, 2027–2035.
26. H. Kessler; S. Steuernagel; M. Will; G. Jung; R. Kellner; D. Gillessen; T. Kamiyama, *Helv. Chim. Acta* **1988**, *71*, 1924–1929.
27. A. Fredenhagen; G. Fendrich; F. Marki; W. Marki; J. Gruner; F. Raschdorf; H. H. Peter, *J. Antibiot. (Tokyo)* **1990**, *43*, 1403–1412.
28. F. Castiglione; A. Lazzarini; L. Carrano; E. Corti; I. Ciciliato; L. Gastaldo; P. Candiani; D. Losi; F. Marinelli; E. Selva; F. Parenti, *Chem. Biol.* **2008**, *15*, 22–31.
29. Y. Minami; K. Yoshida; R. Azuma; A. Urakawa; T. Kawauchi; T. Otani; K. Komiyama; S. Omura, *Tetrahedron Lett.* **1994**, *35*, 8001–8004.

30. M. P. Ryan; R. W. Jack; M. Josten; H. G. Sahl; G. Jung; R. P. Ross; C. Hill, *J. Biol. Chem.* **1999**, *274*, 37544–37550.
31. M. Skaugen; J. Nissenmeyer; G. Jung; S. Stevanovic; K. Sletten; C. I. M. Abildgaard; I. F. Nes, *J. Biol. Chem.* **1994**, *269*, 27183–27185.
32. R. Kellner; G. Jung; M. Josten; C. Kaletta; K. D. Entian; H. G. Sahl, *Angew. Chem.* **1989**, *101*, 618–621.
33. M. B. Ekkelenkamp; M. Hanssen; S. T. Danny Hsu; A. de Jong; D. Milatovic; J. Verhoef; N. A. van Nuland, *FEBS Lett.* **2005**, *579*, 1917–1922.
34. M. van de Kamp; H. W. van den Hooven; R. N. H. Konings; G. Bierbaum; H.-G. Sahl; O. P. Kuipers; R. J. Siezen; W. de Vos; C. W. Hilbers; F. J. van de Ven, *Eur. J. Biochem.* **1995**, *230*, 587–600.
35. C. Heidrich; U. Pag; M. Josten; J. Metzger; R. W. Jack; G. Bierbaum; G. Jung; H. G. Sahl, *Appl. Environ. Microbiol.* **1998**, *64*, 3140–3146.
36. H. W. van den Hooven; F. M. Lagerwerf; W. Heerma; J. Haverkamp; J. C. Piard; C. W. Hilbers; R. J. Siezen; O. P. Kuipers; H. S. Rollema, *FEBS Lett.* **1996**, *391*, 317–322.
37. H. E. Meyer; M. Heber; B. Eisermann; H. Korte; J. W. Metzger; G. Jung, *Anal. Biochem.* **1994**, *223*, 185–190.
38. M. Sunde; M. J. Sparkes; H. B. Dixon, *Biochim. Biophys. Acta* **1998**, *1388*, 45–52.
39. N. I. Martin; T. Sprules; M. R. Carpenter; P. D. Cotter; C. Hill; R. P. Ross; J. C. Vederas, *Biochemistry* **2004**, *43*, 3049–3056.
40. F. J. Van de Ven; H. W. Van den Hooven; R. N. Konings; C. W. Hilbers, *Eur. J. Biochem.* **1991**, *202*, 1181–1188.
41. H. W. Van den Hooven; C. C. Doeland; M. Van de Kamp; R. N. Konings; C. W. Hilbers; F. J. M. Van de Ven, *Eur. J. Biochem.* **1996**, *235*, 382–393.
42. L. Y. Lian; W. C. Chan; S. D. Morley; G. C. Roberts; B. W. Bycroft; D. Jackson, *Biochem. J.* **1992**, *283*, 413–420.
43. W. C. Chan; B. W. Bycroft; M. L. Leyland; L. Y. Lian; J. C. Yang; G. C. Roberts, *FEBS Lett.* **1992**, *300*, 56–62.
44. S. Freund; G. Jung; W. A. Gibbons; H. G. Sahl, NMR and Circular Dichroism Studies on Pep5. In *Nisin and Novel Lantibiotics*; G. Jung, H. G. Sahl, Eds.; ESCOM: Leiden, The Netherlands, 1991; pp 103–112.
45. J. C. Piard; O. P. Kuipers; H. S. Rollema; M. J. Desmazeaud; W. M. de Vos, *J. Biol. Chem.* **1993**, *268*, 16361–16368.
46. T. Prasch; T. Naumann; R. L. Markert; M. Sattler; W. Schubert; S. Schaal; M. Bauch; H. Kogler; C. Griesinger, *Eur. J. Biochem.* **1997**, *244*, 501–512.
47. N. Zimmermann; G. Jung, *Eur. J. Biochem.* **1997**, *246*, 809–819.
48. H. Kessler; S. Seip; T. Wein; S. Steuernagel; M. Will, Structure of Cinnamycin (Ro 09-0198) in Solution. In *Nisin and Novel Lantibiotics*; G. Jung, H.-G. Sahl, Eds.; ESCOM: Leiden, The Netherlands, 1991; pp 76–90.
49. S. T. Hsu; E. Breukink; E. Tischenko; M. A. Lutters; B. De Kruijff; R. Kaptein; A. M. Bonvin; N. A. Van Nuland, *Nat. Struct. Mol. Biol.* **2004**, *11*, 963–967.
50. S. T. Hsu; E. Breukink; G. Bierbaum; H. G. Sahl; B. de Kruijff; R. Kaptein; N. A. van Nuland; A. M. Bonvin, *J. Biol. Chem.* **2003**, *278*, 13110–13117.
51. K. Wakamatsu; S. Y. Choung; T. Kobayashi; K. Inoue; T. Higashijima; T. Miyazawa, *Biochemistry* **1990**, *29*, 113–118.
52. K. Hosoda; M. Ohya; T. Kohno; T. Maeda; S. Endo; K. Wakamatsu, *J. Biochem.* **1996**, *119*, 226–230.
53. A. L. McClerren; L. E. Cooper; C. Quan; P. M. Thomas; N. L. Kelleher; W. A. van der Donk, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17243–17248.
54. L. Xie; L. M. Miller; C. Chatterjee; O. Averin; N. L. Kelleher; W. A. van der Donk, *Science* **2004**, *303*, 679–681.
55. G. Jung, *Angew. Chem. Intl. Ed. Engl.* **1991**, *30*, 1051–1068.
56. W. M. de Vos; O. P. Kuipers; J. R. van der Meer; R. J. Siezen, *Mol. Microbiol.* **1995**, *17*, 427–437.
57. H. Brötz; G. Bierbaum; K. Leopold; P. E. Reynolds; H. G. Sahl, *Antimicrob. Agents Chemother.* **1998**, *42*, 154–160.
58. E. Breukink; H. E. van Heusden; P. J. Vollmerhaus; E. Swiezewska; L. Brunner; S. Walker; A. J. Heck; B. de Kruijff, *J. Biol. Chem.* **2003**, *278*, 19898–19903.
59. E. Breukink; B. de Kruijff, *Nat. Rev. Drug Discov.* **2006**, *5*, 321–332.
60. I. Wiedemann; E. Breukink; C. van Kraaij; O. P. Kuipers; G. Bierbaum; B. de Kruijff; H. G. Sahl, *J. Biol. Chem.* **2001**, *276*, 1772–1779.
61. I. F. Nes; J. R. Tagg, *Antonie van Leeuwenhoek* **1996**, *69*, 89–97.
62. S. Rumpel; A. Razeto; C. M. Pillar; V. Vijayan; A. Taylor; K. Giller; M. S. Gilmore; S. Becker; M. Zweckstetter, *EMBO J.* **2004**, *23*, 3632–3642.
63. L. S. Håvarstein; H. Holo; I. F. Nes, *Microbiology* **1994**, *140* (Pt 9), 2383–2389.
64. T. R. Klaenhammer, *FEMS Microbiol. Rev.* **1993**, *12*, 39–85.
65. E. M. Lawton; P. D. Cotter; C. Hill; R. P. Ross, *FEMS Microbiol. Lett.* **2007**, *267*, 64–71.
66. M. A. Navaratna; H. G. Sahl; J. R. Tagg, *Appl. Environ. Microbiol.* **1998**, *64*, 4803–4808.
67. M. A. Navaratna; H. G. Sahl; J. R. Tagg, *Infect. Immun.* **1999**, *67*, 4268–4271.
68. M. C. Booth; C. P. Bogie; H.-G. Sahl; R. J. Siezen; K. L. Hatter; M. S. Gilmore, *Mol. Microbiol.* **1996**, *21*, 1175–1184.
69. H. Holo; Z. Jeknic; M. Daeschel; S. Stevanovic; I. F. Nes, *Microbiology* **2001**, *147*, 643–651.
70. H. Yonezawa; H. K. Kuramitsu, *Antimicrob. Agents Chemother.* **2005**, *49*, 541–548.
71. O. Hyink; M. Balakrishnan; J. R. Tagg, *FEMS Microbiol. Lett.* **2005**, *252*, 235–241.
72. S. Kodani; M. E. Hudson; M. C. Durrant; M. J. Buttner; J. R. Nodwell; J. M. Willey, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11448–11453.
73. S. Kodani; M. A. Lodato; M. C. Durrant; F. Picart; J. M. Willey, *Mol. Microbiol.* **2005**, *58*, 1368–1380.
74. J. Guijarro; R. Santamaria; A. Schauer; R. Losick, *J. Bacteriol.* **1988**, *170*, 1895–1901.
75. K. Ueda; K. Oinuma; G. Ikeda; K. Hosono; Y. Ohnishi; S. Horinouchi; T. Beppu, *J. Bacteriol.* **2002**, *184*, 1488–1492.
76. C. Klein; C. Kaletta; N. Schnell; K.-D. Entian, *Appl. Environ. Microbiol.* **1992**, *58*, 132–142.
77. H. Kessler; S. Steuernagel; D. Gillessen; T. Kamiyama, *Helv. Chim. Acta* **1987**, *70*, 726–741.
78. R. G. Benedict; W. Dvonch; O. L. Shotwell; T. Pridham; L. A. Lindenfelser, *Antibiot. Chemother.* **1952**, *2*, 591–594.
79. C. Corvey; T. Stein; S. Dusterhus; M. Karas; K. D. Entian, *Biochem. Biophys. Res. Commun.* **2003**, *304*, 48–54.
80. D. A. Widdick; H. M. Dodd; P. Barraille; J. White; T. H. Stein; K. F. Chater; M. J. Gasson; M. J. Bibb, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4316–4321.

81. S. Chatterjee; S. Chatterjee; S. J. Lad; M. S. Phansalkar; R. H. Rupp; B. N. Ganguli; H. W. Fehlhaber; H. Kogler, *J. Antibiot.* **1992**, *45*, 832–838.
82. M. D. Limbert; D. Isert; N. Klesel; A. Markus; G. Seibert; S. Chatterjee; D. K. Chatterjee; R. H. Jani; B. N. Ganguli, Chemotherapeutic Properties of Mersacidin *In vitro* and *In vivo*. In *Nisin and Novel Lantibiotics*; G. Jung, H.-G. Sahl, Eds.; ESCOM: Leiden, The Netherlands, 1991; pp 448–456.
83. M. Skaugen; C. I. Abildgaard; I. F. Nes, *Mol. Gen. Genet.* **1997**, *253*, 674–686.
84. M. Upton; J. R. Tagg; P. Wescombe; H. F. Jenkinson, *J. Bacteriol.* **2001**, *183*, 3931–3938.
85. A. Dufour; T. Hindre; D. Haras; J. P. Le Pennec, *FEMS Microbiol. Rev.* **2007**, *31*, 134–167.
86. P. A. Wescombe; M. Upton; K. P. Dierksen; N. L. Ragland; S. Sivabalan; R. E. Wirawan; M. A. Inglis; C. J. Moore; G. V. Walker; C. N. Chilcott; H. F. Jenkinson; J. R. Tagg, *Appl. Environ. Microbiol.* **2006**, *72*, 1459–1466.
87. A. Hurst, *Adv. Appl. Microbiol.* **1981**, *27*, 85–123.
88. H.-G. Sahl; R. W. Jack; G. Bierbaum, *Eur. J. Biochem.* **1995**, *230*, 827–853.
89. T. Kupke; F. Götz, *Antonie van Leeuwenhoek* **1996**, *69*, 139–150.
90. J. R. van der Meer; H. S. Rollema; R. J. Siezen; M. M. Beerthuyzen; O. P. Kuipers; W. M. de Vos, *J. Biol. Chem.* **1994**, *269*, 3555–3562.
91. B. Li; J.-P. J. Yu; J. S. Brunzelle; G. N. Moll; W. A. van der Donk; S. K. Nair, *Science* **2006**, *311*, 1464–1467.
92. R. Rink; J. Wierenga; A. Kuipers; L. D. Kluskens; A. J. M. Driessen; O. P. Kuipers; G. N. Moll, *Appl. Environ. Microbiol.* **2007**, *73*, 1792–1796.
93. R. Rink; A. Kuipers; E. de Boef; K. J. Leenhouts; A. J. Driessen; G. N. Moll; O. P. Kuipers, *Biochemistry* **2005**, *44*, 8873–8882.
94. L. D. Kluskens; A. Kuipers; R. Rink; E. de Boef; S. Fekken; A. J. Driessen; O. P. Kuipers; G. N. Moll, *Biochemistry* **2005**, *44*, 12827–12834.
95. C. Chatterjee; G. C. Patton; L. Cooper; M. Paul; W. A. van der Donk, *Chem. Biol.* **2006**, *13*, 1109–1117.
96. M. R. Levengood; W. A. van der Donk, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3025–3028.
97. G. Izaguirre; J. N. Hansen, *Appl. Environ. Microbiol.* **1997**, *63*, 3965–3971.
98. A. Kuipers; E. De Boef; R. Rink; S. Fekken; L. D. Kluskens; A. J. Driessen; K. Leenhouts; O. P. Kuipers; G. N. Moll, *J. Biol. Chem.* **2004**, *279*, 22176–22182.
99. P. Chen; F. X. Qi; J. Novak; R. E. Krull; P. W. Caufield, *FEMS Microbiol. Lett.* **2001**, *195*, 139–144.
100. L. A. Furgerson Ihnken; C. Chatterjee; W. A. van der Donk, *Biochemistry* **2008**, *47*, 7352–7363.
101. S. Neis; G. Bierbaum; M. Josten; U. Pag; C. Kempter; G. Jung; H. G. Sahl, *FEMS Microbiol. Lett.* **1997**, *149*, 249–255.
102. G. C. Patton; M. Paul; L. E. Cooper; C. Chatterjee; W. A. van der Donk, *Biochemistry* **2008**, *47*, 7342–7351.
103. A. G. Beck-Sickinger; G. Jung, Synthesis and Conformational Analysis of Lantibiotic Leader-, Pro-, and Pre-Peptides. In *Nisin and Novel Lantibiotics*; G. Jung, H.-G. Sahl, Eds.; ESCOM: Leiden, 1991; pp 218–230.
104. H. P. Weil; A. G. Beck-Sickinger; J. Metzger; S. Stevanovic; G. Jung; M. Josten; H. G. Sahl, *Eur. J. Biochem.* **1990**, *194*, 217–223.
105. O. P. Kuipers; G. Bierbaum; B. Ottenwälder; H. M. Dodd; N. Horn; J. Metzger; T. Kupke; V. Gnau; R. Bongers; P. van den Bogaard; H. Kosters; H. S. Rollema; W. M. de Vos; R. J. Siezen; G. Jung; F. Götz; H. G. Sahl; M. J. Gasson, *Antonie van Leeuwenhoek* **1996**, *69*, 161–169.
106. B. Ottenwälder; T. Kupke; S. Brecht; V. Gnau; J. Metzger; G. Jung; F. Götz, *Appl. Environ. Microbiol.* **1995**, *61*, 3894–3903.
107. W. C. Chan; H. M. Dodd; N. Horn; K. Maclean; L. Y. Lian; B. W. Bycroft; M. J. Gasson; G. C. Roberts, *Appl. Environ. Microbiol.* **1996**, *62*, 2966–2969.
108. H. M. Dodd; N. Horn; W. C. Chan; C. J. Giffard; B. W. Bycroft; G. C. Roberts; M. J. Gasson, *Microbiology* **1996**, *142* (Pt 9), 2385–2392.
109. W. Liu; J. N. Hansen, *J. Biol. Chem.* **1992**, *267*, 25078–25085.
110. R. Rink; J. Wierenga; A. Kuipers; L. D. Kluskens; A. J. Driessen; O. P. Kuipers; G. N. Moll, *Appl. Environ. Microbiol.* **2007**, *73*, 5809–5816.
111. S. H. Paik; A. Chakicherla; J. N. Hansen, *J. Biol. Chem.* **1998**, *273*, 23134–23142.
112. R. Dorenbos; T. Stein; J. Kabel; C. Bruand; A. Bolhuis; S. Bron; W. J. Quax; J. M. Van Dijl, *J. Biol. Chem.* **2002**, *277*, 16682–16688.
113. L. E. Cooper; A. L. McClerren; A. Chary; W. A. van der Donk, *Chem. Biol.* **2008**, *15*, 1025–1045.
114. J. Augustin; R. Rosenstein; B. Wieland; U. Schneider; N. Schnell; G. Engelke; K.-D. Entian; F. Götz, *Eur. J. Biochem.* **1992**, *204*, 1149–1154.
115. J. N. Hansen, *Annu. Rev. Microbiol.* **1993**, *47*, 535–564.
116. R. J. Siezen; O. P. Kuipers; W. M. de Vos, *Antonie van Leeuwenhoek* **1996**, *69*, 171–184.
117. Z. Gutowski-Eckel; C. Klein; K. Siegers; K. Bohm; M. Hammelmann; K.-D. Entian, *Appl. Environ. Microbiol.* **1994**, *60*, 1–11.
118. C. Meyer; G. Bierbaum; C. Heidrich; M. Reis; J. Süling; M. I. Iglesias-Wind; C. Kempter; E. Molitor; H.-G. Sahl, *Eur. J. Biochem.* **1995**, *232*, 478–489.
119. O. Koponen; M. Tolonen; M. Qiao; G. Wahlstrom; J. Helin; E. J. Saris Per, *Microbiology* **2002**, *148*, 3561–3568.
120. G. Engelke; Z. Gutowski-Eckel; M. Hammelmann; K.-D. Entian, *Appl. Environ. Microbiol.* **1992**, *58*, 3730–3743.
121. K. Siegers; S. Heinzmann; K.-D. Entian, *J. Biol. Chem.* **1996**, *271*, 12294–12301.
122. P. Kiesau; U. Eikmanns; Z. Gutowski-Eckel; S. Weber; M. Hammelmann; K.-D. Entian, *J. Bacteriol.* **1997**, *179*, 1475–1481.
123. L. Xie; C. Chatterjee; R. Balsara; N. M. Okeley; W. A. van der Donk, *Biochem. Biophys. Res. Commun.* **2002**, *295*, 952–957.
124. F. Cheng; T. M. Takala; P. E. Saris, *J. Mol. Microbiol. Biotechnol.* **2007**, *13*, 248–254.
125. A. Kuipers; J. Wierenga; R. Rink; L. D. Kluskens; A. J. Driessen; O. P. Kuipers; G. N. Moll, *Appl. Environ. Microbiol.* **2006**, *72*, 7626–7633.
126. N. M. Okeley; M. Paul; J. P. Stasser; N. Blackburn; W. A. van der Donk, *Biochemistry* **2003**, *42*, 13613–13624.
127. B. Li; W. A. van der Donk, *J. Biol. Chem.* **2007**, *282*, 21169–21175.
128. A. Rincé; A. Dufour; S. Le Pogam; D. Thuault; C. M. Bourgeois; J. P. Le Pennec, *Appl. Environ. Microbiol.* **1994**, *60*, 1652–1657.
129. P. Uguen; J. P. Le Pennec; A. Dufour, *J. Bacteriol.* **2000**, *182*, 5262–5266.
130. P. Uguen; T. Hindré; S. Didelot; C. Marty; D. Haras; J. P. Le Pennec; K. Vallee-Rehel; A. Dufour, *Appl. Environ. Microbiol.* **2005**, *71*, 562–565.

131. O. McAuliffe; C. Hill; R. P. Ross, *Microbiology* **2000**, *146*, 2147–2154.
132. M. S. Gilmore; R. A. Segarra; M. C. Booth; C. P. Bogie; L. R. Hall; D. B. Clewell, *J. Bacteriol.* **1994**, *176*, 7335–7344.
133. L. M. Miller; C. Chatterjee; W. A. van der Donk; N. L. Kelleher, *J. Am. Chem. Soc.* **2006**, *128*, 1420–1421.
134. M. Paul; G. C. Patton; W. A. van der Donk, *Biochemistry* **2007**, *46*, 6268–6276.
135. P. D. Cotter; L. A. Draper; E. M. Lawton; O. McAuliffe; C. Hill; R. P. Ross, *Appl. Environ. Microbiol.* **2006**, *72*, 4492–4496.
136. P. D. Cotter; L. H. Deegan; E. M. Lawton; L. A. Draper; P. M. O'Connor; C. Hill; R. P. Ross, *Mol. Microbiol.* **2006**, *62*, 735–747.
137. O. P. Kuipers; H. S. Rollema; W. M. Yap; H. J. Boot; R. J. Siezen; W. M. de Vos, *J. Biol. Chem.* **1992**, *267*, 24340–24346.
138. C. van Kraaij; E. Breukink; H. S. Rollema; R. J. Siezen; R. A. Demel; B. De Kruijff; O. P. Kuipers, *Eur. J. Biochem.* **1997**, *247*, 114–120.
139. H. S. Rollema; O. P. Kuipers; P. Both; W. M. de Vos; R. J. Siezen, *Appl. Environ. Microbiol.* **1995**, *61*, 2873–2878.
140. G. Bierbaum; M. Reis; C. Szekat; H. G. Sahl, *Appl. Environ. Microbiol.* **1994**, *60*, 4332–4338.
141. C. Szekat; R. W. Jack; D. Skutlarek; H. Farber; G. Bierbaum, *Appl. Environ. Microbiol.* **2003**, *69*, 3777–3783.
142. D. Field; B. Collins; P. D. Cotter; C. Hill; R. P. Ross, *J. Mol. Microbiol. Biotechnol.* **2007**, *13*, 226–234.
143. P. Chen; J. Novak; M. Kirk; S. Barnes; F. Qi; P. W. Caufield, *Appl. Environ. Microbiol.* **1998**, *64*, 2335–2340.
144. A. Chakicherla; J. N. Hansen, *J. Biol. Chem.* **1995**, *270*, 23533–23539.
145. O. P. Kuipers; H. S. Rollema; W. M. de Vos; R. J. Siezen, *FEBS Lett.* **1993**, *330*, 23–27.
146. M. R. Levengood; G. C. Patton; W. A. van der Donk, *J. Am. Chem. Soc.* **2007**, *129*, 10314–10315.
147. C. Chatterjee; L. M. Miller; Y. L. Leung; L. Xie; M. Yi; N. L. Kelleher; W. A. van der Donk, *J. Am. Chem. Soc.* **2005**, *127*, 15332–15333.
148. Y. O. You; W. A. van der Donk, *Biochemistry* **2007**, *46*, 5991–6000.
149. L. Hedstrom; L. Gan, *Curr. Opin. Chem. Biol.* **2006**, *10*, 520–525.
150. Y. V. Guillén Schlippe; L. Hedstrom, *Biochemistry* **2005**, *44*, 11700–11707.
151. Y. V. Guillén Schlippe; L. Hedstrom, *Arch. Biochem. Biophys.* **2005**, *433*, 266–278.
152. C. G. Mowat; R. Moysey; C. S. Miles; D. Leys; M. K. Doherty; P. Taylor; M. D. Walkinshaw; G. A. Reid; S. K. Chapman, *Biochemistry* **2001**, *40*, 12292–12298.
153. M. K. Doherty; S. L. Pealing; C. S. Miles; R. Moysey; P. Taylor; M. D. Walkinshaw; G. A. Reid; S. K. Chapman, *Biochemistry* **2000**, *39*, 10695–10701.
154. S. J. Charnock; I. E. Brown; J. P. Turkenburg; G. W. Black; G. J. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12067–12072.
155. R. T. Bossi; A. Negri; G. Tedeschi; A. Mattevi, *Biochemistry* **2002**, *41*, 3018–3024.
156. G. Tedeschi; S. Ronchi; T. Simonic; C. Treu; A. Mattevi; A. Negri, *Biochemistry* **2001**, *40*, 4738–4744.
157. J. E. Galan; H. Wolf-Watz, *Nature* **2006**, *444*, 567–573.
158. H. Li; H. Xu; Y. Zhou; J. Zhang; C. Long; S. Li; S. Chen; J. M. Zhou; F. Shao, *Science* **2007**, *315*, 1000–1003.
159. J. Zhang; F. Shao; Y. Li; H. Cui; L. Chen; H. Li; Y. Zou; C. Long; L. Lan; J. Chai; S. Chen; X. Tang; J. M. Zhou, *Cell Host Microbe* **2007**, *1*, 175–185.
160. Y. Zhu; H. Li; C. Long; L. Hu; H. Xu; L. Liu; S. Chen; D. C. Wang; F. Shao, *Mol. Cell* **2007**, *28*, 899–913.
161. R. A. Bednar, *Biochemistry* **1990**, *29*, 3684–3690.
162. J. Penner-Hahn, *Curr. Opin. Chem. Biol.* **2007**, *11*, 166–171.
163. D. B. Rozema; C. D. Poulter, *Biochemistry* **1999**, *38*, 13138–13146.
164. C. C. Huang; P. J. Casey; C. A. Fierke, *J. Biol. Chem.* **1997**, *272*, 20–23.
165. K. E. Hightower; C. C. Huang; P. J. Casey; C. A. Fierke, *Biochemistry* **1998**, *37*, 15555–15562.
166. J. J. Wilker; S. J. Lippard, *J. Am. Chem. Soc.* **1995**, *117*, 8682–8683.
167. J. J. Wilker; S. J. Lippard, *Inorg. Chem.* **1997**, *36*, 969–978.
168. S.-J. Chiou; J. Innocent; C. G. Riordan; K.-C. Lam; L. Liable-Sands; A. L. Rheingold, *Inorg. Chem.* **2000**, *39*, 4347–4353.
169. U. Brand; M. Rombach; J. Seebacher; H. Vahrenkamp, *Inorg. Chem.* **2001**, *40*, 6151–6157.
170. S. J. Chiou; C. G. Riordan; A. L. Rheingold, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3695–3700.
171. G. Parkin, *Chem. Rev.* **2004**, *104*, 699–767.
172. M. M. Morlok; K. E. Janak; G. Zhu; D. A. Quarless; G. Parkin, *J. Am. Chem. Soc.* **2005**, *127*, 14039–14050.
173. M. M. Ibrahim; J. Seebacher; G. Steinfeld; H. Vahrenkamp, *Inorg. Chem.* **2005**, *44*, 8531–8538.
174. M. Ji; B. Benkmil; H. Vahrenkamp, *Inorg. Chem.* **2005**, *44*, 3518–3523.
175. J. G. Melnick; G. Zhu; D. Buccella; G. Parkin, *J. Inorg. Biochem.* **2006**, *100*, 1147–1154.
176. M. Helfrich; K. D. Entian; T. Stein, *Biochemistry* **2007**, *46*, 3224–3233.
177. H. W. Park; S. R. Boduluri; J. F. Moomaw; P. J. Casey; L. S. Beese, *Science* **1997**, *275*, 1800–1804.
178. R. Pejchal; M. L. Ludwig, *PLoS Biol.* **2005**, *3*, e31.
179. J. C. Evans; D. P. Huddler; M. T. Hilgers; G. Romanchuk; R. G. Matthews; M. L. Ludwig, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3729–3736.
180. J. C. Evans; D. P. Huddler; J. Jiracek; C. Castro; N. S. Millian; T. A. Garrow; M. L. Ludwig, *Structure* **2002**, *10*, 1159–1171.
181. R. Rink; L. D. Kluskens; A. Kuipers; A. J. Driessen; O. P. Kuipers; G. N. Moll, *Biochemistry* **2007**, *46*, 13179–13189.
182. P. L. Toogood, *Tetrahedron Lett.* **1993**, *34*, 7833–7836.
183. S. Burrage; T. Raynham; G. Williams; J. W. Essex; C. Allen; M. Cardno; V. Swali; M. Bradley, *Chem. Eur. J.* **2000**, *6*, 1455–1466.
184. N. M. Okeley; Y. Zhu; W. A. van der Donk, *Org. Lett.* **2000**, *2*, 3603–3606.
185. H. Zhou; W. A. van der Donk, *Org. Lett.* **2002**, *4*, 1335–1338.
186. Y. Zhu; M. Gieselman; H. Zhou; O. Averin; W. A. van der Donk, *Org. Biomol. Chem.* **2003**, *1*, 3304–3315.
187. Y. Rew; S. Malkmus; C. Svensson; T. L. Yaksh; N. N. Chung; P. W. Schiller; J. A. Cassel; R. N. DeHaven; M. Goodman, *J. Med. Chem.* **2002**, *45*, 3746–3754.
188. X. Zhang; W. Ni; W. A. van der Donk, *Org. Lett.* **2007**, *9*, 3343–3346.
189. X. Zhang; W. A. van der Donk, *Methods Enzymol.* **2009**, in press.
190. J. R. van der Meer; J. Polman; M. M. Beerthuyzen; R. J. Siezen; O. P. Kuipers; W. M. de Vos, *J. Bacteriol.* **1993**, *175*, 2578–2588.
191. O. Schneewind; A. Fowler; K. F. Faull, *Science* **1995**, *268*, 103–106.
192. H. Rintala; T. Graeffe; L. Paulin; N. Kalkkinen; P. E. J. Saris, *Biotechnol. Lett.* **1993**, *15*, 991–996.

193. T. Stein; K. D. Entian, *Rapid Commun. Mass Spectrom.* **2002**, *16*, 103–110.
194. S. Geissler; F. Götz; T. Kupke, *J. Bacteriol.* **1996**, *178*, 284–288.
195. S. Kies; C. Vuong; M. Hille; A. Peschel; C. Meyer; F. Götz; M. Otto, *Peptides* **2003**, *24*, 329–338.
196. G. Valsesia; G. Medaglia; M. Held; W. Minas; S. Panke, *Appl. Environ. Microbiol.* **2007**, *73*, 1635–1645.
197. L. S. Håvarstein; D. B. Diep; I. F. Nes, *Mol. Microbiol.* **1995**, *16*, 229–240.
198. S. Ishii; T. Yano; H. Hayashi, *J. Biol. Chem.* **2006**, *281*, 4726–4731.
199. K. H. Wu; P. C. Tai, *J. Biol. Chem.* **2004**, *279*, 901–909.
200. G. Dirix; P. Monsieurs; B. Dombrecht; R. Daniels; K. Marchal; J. Vanderleyden; J. Michiels, *Peptides* **2004**, *25*, 1425–1440.
201. B. A. Dougherty; C. Hill; J. F. Weidman; D. R. Richardson; J. C. Venter; R. P. Ross, *Mol. Microbiol.* **1998**, *29*, 1029–1038.
202. M. S. Gilmore; R. A. Segarra; M. C. Booth; C. P. Bogie; L. R. Hall; D. B. Clewell, *J. Bacteriol.* **1994**, *176*, 7335–7344.
203. R. A. Segarra; M. C. Booth; D. A. Morales; M. M. Huycke; M. S. Gilmore, *Infect. Immun.* **1991**, *59*, 1239–1246.
204. H. Allgaier; G. Jung; R. G. Werner; U. Schneider, *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 1051–1053.
205. R. Kellner; G. Jung; T. Hoerner; H. Zaehner; N. Schnell; K.-D. Entian; F. Goetz, *Eur. J. Biochem.* **1988**, *177*, 53–59.
206. F. Qi; P. Chen; P. W. Caufield, *Appl. Environ. Microbiol.* **1999**, *65*, 3880–3887.
207. T. Kupke; S. Stevanovic; H. G. Sahl; F. Götz, *J. Bacteriol.* **1992**, *174*, 5354–5361.
208. T. Kupke; C. Kempter; V. Gnau; G. Jung; F. Götz, *J. Biol. Chem.* **1994**, *269*, 5653–5659.
209. T. Kupke; F. Götz, *FEMS Microbiol. Lett.* **1997**, *153*, 25–32.
210. D. G. Schmid; F. Majer; T. Kupke; G. Jung, *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1779–1784.
211. F. Majer; D. G. Schmid; K. Altena; G. Bierbaum; T. Kupke, *J. Bacteriol.* **2002**, *184*, 1234–1243.
212. T. Kupke; C. Kempter; G. Jung; F. Götz, *J. Biol. Chem.* **1995**, *270*, 11282–11289.
213. M. Blaesse; T. Kupke; R. Huber; S. Steinbacher, *EMBO J.* **2000**, *19*, 6299–6310.
214. C. Kempter; T. Kupke; D. Kaiser; J. W. Metzger; G. Jung, *Angew. Chem. Intl. Ed. Engl.* **1996**, *35*, 2104–2107.
215. T. Kupke; F. Götz, *J. Biol. Chem.* **1997**, *272*, 4759–4762.
216. T. Kupke; M. Uebele; D. Schmid; G. Jung; M. Blaesse; S. Steinbacher, *J. Biol. Chem.* **2000**, *275*, 31838–31846.
217. M. Blaesse; T. Kupke; R. Huber; S. Steinbacher, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *D59*, 1414–1421.
218. T. Kupke, *J. Biol. Chem.* **2001**, *276*, 27597–27604.
219. P. Hernandez-Acosta; D. G. Schmid; G. Jung; F. A. Culianez-Macia; T. Kupke, *J. Biol. Chem.* **2002**, *277*, 20490–20498.
220. C. Kaletta; K. D. Entian; G. Jung, *Eur. J. Biochem.* **1991**, *199*, 411–415.
221. Q. P. Wang; W. J. VanDusen; C. J. Petroski; V. M. Garsky; A. M. Stern; P. A. Friedman, *J. Biol. Chem.* **1991**, *266*, 14004–14010.
222. S. Jia; W. J. VanDusen; R. E. Diehl; N. E. Kohl; R. A. F. Dixon; K. O. Elliston; A. M. Stern; P. A. Friedman, *J. Biol. Chem.* **1992**, *267*, 14322–14327.
223. S. Jia; K. McGinnis; W. J. VanDusen; C. J. Burke; A. Kuo; P. R. Griffin; M. K. Sardana; K. O. Elliston; A. M. Stern; P. A. Friedman, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7227–7231.
224. F. Korioth; C. Gieffers; J. Frey, *Gene* **1994**, *150*, 395–399.
225. M. van de Kamp; L. M. Horstink; H. W. Van den Hooven; R. N. H. Konings; C. W. Hilbers; A. Frey; H.-G. Sahl; J. W. Metzger; F. J. M. van de Ven, *Eur. J. Biochem.* **1995**, *227*, 757–771.
226. K. Richter; R. Egger; G. Kreil, *Science* **1987**, *238*, 200–202.
227. K. Richter; R. Egger; L. Negri; R. Corsi; C. Severini; G. Kreil, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4836–4839.
228. P. D. Cotter; P. M. O'Connor; L. A. Draper; E. M. Lawton; L. H. Deegan; C. Hill; R. P. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18584–18589.
229. T. Wakamiya; Y. Ueki; T. Shiba; Y. Kido; Y. Motoki, *Tetrahedron Lett.* **1985**, *26*, 665–668.
230. T. Wakamiya; Y. Ueki; T. Shiba; Y. Kido; Y. Motoki, *Bull. Chem. Soc. Jpn.* **1990**, *63*, 1032–1038.
231. M. Friedman, *J. Agric. Food. Chem.* **1999**, *47*, 1295–1319.
232. H. Xiao; X. Chen; M. Chen; S. Tang; X. Zhao; L. Huan, *Microbiology* **2004**, *150*, 103–108.
233. W. C. Chan; B. W. Bycroft; M. L. Leyland; L. Y. Lian; G. C. Roberts, *Biochem. J.* **1993**, *291* (Pt 1), 23–27.
234. A. Lazzarini; L. Gastaldo; G. Candiani; I. Ciciliato; D. Losi; F. Marinelli; E. Selva; F. Parenti, Antibiotic 107891, Its Factors A1 and A2, Pharmaceutically Acceptable Salts and Compositions, and Use Thereof. WO2005/014628 A1, 2005.
235. K. H. van Pee; E. P. Patallo, *Appl. Microbiol. Biotechnol.* **2006**, *70*, 631–641.
236. N. Zimmermann; J. W. Metzger; G. Jung, *Eur. J. Biochem.* **1995**, *228*, 786–797.
237. S. Boakes; J. Bargallo Cortes; M. J. Dawson, Lantibiotic Biosynthetic Gene Clusters from *A. garbadinensis* and *A. liguriae*. WO 2007/083112 A2, 2007.
238. L. J. Harris; H. P. Fleming; T. R. Klaenhammer, *J. Food Prot.* **1991**, *54*, 836–840.
239. T. J. Montville; A. M. Rogers; A. Okereke, *J. Food Prot.* **1992**, *55*, 444–448.
240. M. E. Bruno; A. Kaiser; T. J. Montville, *Appl. Environ. Microbiol.* **1992**, *58*, 2255–2259.
241. N. Benkerroum; W. E. Sandine, *J. Dairy Sci.* **1988**, *71*, 3237–3245.
242. G. Bierbaum; C. Szekat; M. Josten; C. Heidrich; C. Kempter; G. Jung; H. G. Sahl, *Appl. Environ. Microbiol.* **1996**, *62*, 385–392.
243. R. E. Krull; P. Chen; J. Novak; M. Kirk; S. Barnes; J. Baker; N. R. Krishna; P. W. Caufield, *J. Biol. Chem.* **2000**, *275*, 15845–15850.
244. H. E. van Heusden; B. de Kruijff; E. Breukink, *Biochemistry* **2002**, *41*, 12171–12178.
245. H. E. Hasper; B. de Kruijff; E. Breukink, *Biochemistry* **2004**, *43*, 11567–11575.
246. E. Breukink; H. E. van Heusden; P. J. Vollmerhaus; E. Swiezewska; L. Brunner; S. Walker; A. J. Heck; B. de Kruijff, *J. Biol. Chem.* **2003**, *278*, 19898–19903.
247. I. Wiedemann; R. Benz; H. G. Sahl, *J. Bacteriol.* **2004**, *186*, 3259–3261.
248. D. H. Williams, *Nat. Prod. Rep.* **1996**, *13*, 469–477.
249. J. C. Barna; D. H. Williams, *Annu. Rev. Microbiol.* **1984**, *38*, 339–357.
250. J. van Heijenoort, *Glycobiology* **2001**, *11*, 25R–36R.
251. J. van Heijenoort, *Nat. Prod. Rep.* **2001**, *18*, 503–519.
252. C. T. Walsh; S. L. Fisher; I. S. Park; M. Prahalad; Z. Wu, *Chem. Biol.* **1996**, *3*, 21–28.
253. V. L. Healy; I. A. Lessard; D. I. Roper; J. R. Knox; C. T. Walsh, *Chem. Biol.* **2000**, *7*, R109–R119.

254. W. C. Chan; M. Leyland; J. Clark; H. M. Dodd; L. Y. Lian; M. J. Gasson; B. W. Bycroft; G. C. Roberts, *FEBS Lett.* **1996**, *390*, 129–132.
255. W. C. Chan; B. W. Bycroft; L. Y. Lian; G. C. K. Roberts, *FEBS Lett.* **1989**, *252*, 29–36.
256. J. Yuan; Z. Z. Zhang; X. Z. Chen; W. Yang; L. D. Huan, *Appl. Microbiol. Biotechnol.* **2004**, *64*, 806–815.
257. H. E. Hasper; N. E. Kramer; J. L. Smith; J. D. Hillman; C. Zachariah; O. P. Kuipers; B. de Kruijff; E. Breukink, *Science* **2006**, *313*, 1636–1637.
258. L. Smith; H. Hasper; E. Breukink; J. Novak; J. Cerkasov; J. D. Hillman; S. Wilson-Stanford; R. S. Orugunty, *Biochemistry* **2008**, *47*, 3308–3314.
259. I. Wiedemann; T. Bottiger; R. R. Bonelli; A. Wiese; S. O. Hagge; T. Gutsmann; U. Seydel; L. Deegan; C. Hill; P. Ross; H. G. Sahl, *Mol. Microbiol.* **2006**, *61*, 285–296.
260. S. M. Morgan; P. M. O'Connor; P. D. Cotter; R. P. Ross; C. Hill, *Antimicrob. Agents Chemother.* **2005**, *49*, 2606–2611.
261. S. Somma; W. Merati; F. Parenti, *Antimicrob. Agents Chemother.* **1977**, *11*, 396–401.
262. I. Wiedemann; T. Bottiger; R. R. Bonelli; T. Schneider; H. G. Sahl; B. Martinez, *Appl. Environ. Microbiol.* **2006**, *72*, 2809–2814.
263. H. Brötz; G. Bierbaum; P. E. Reynolds; H. G. Sahl, *Eur. J. Biochem.* **1997**, *246*, 193–199.
264. D. L. Turner; L. Brennan; H. E. Meyer; C. Lohaus; C. Siethoff; H. S. Costa; B. Gonzalez; H. Santos; J. E. Suarez, *Eur. J. Biochem.* **1999**, *264*, 833–839.
265. E. B. O'Connor; P. D. Cotter; P. O'Connor; O. O'Sullivan; J. R. Tagg; R. P. Ross; C. Hill, *BMC Microbiol.* **2007**, *7*, 24.
266. G. Jung, Relatedness Between the Two-Component Lantibiotics Lacticin 3147 and Staphylococcin C55 Based on Structure, Genetics and Biological Activity. In *Nisin and Novel Lantibiotics*; G. Jung, H. G. Sahl, Eds.; ESCOM: Leiden, The Netherlands, 1991; pp 1–34.
267. Y. Kido; T. Hamakado; T. Yoshida; M. Anno; Y. Motoki; T. Wakamiya; T. Shiba, *J. Antibiot. (Tokyo)* **1983**, *36*, 1295–1299.
268. S. Y. Choung; T. Kobayashi; K. Takemoto; H. Ishitsuka; K. Inoue, *Biochim. Biophys. Acta* **1988**, *940*, 180–187.
269. G. Machaidze; J. Seelig, *Biochemistry* **2003**, *42*, 12570–12576.
270. J. Navarro; J. Chabot; K. Sherrill; R. Aneja; S. A. Zahler; E. Racker, *Biochemistry* **1985**, *24*, 4645–4650.
271. S. Y. Choung; T. Kobayashi; J. Inoue; K. Takemoto; H. Ishitsuka; K. Inoue, *Biochim. Biophys. Acta* **1988**, *940*, 171–179.
272. A. Makino; T. Baba; K. Fujimoto; K. Iwamoto; Y. Yano; N. Terada; S. Ohno; S. B. Sato; A. Ohta; M. Umeda; K. Matsuzaki; T. Kobayashi, *J. Biol. Chem.* **2003**, *278*, 3204–3209.
273. G. Machaidze; A. Ziegler; J. Seelig, *Biochemistry* **2002**, *41*, 1965–1971.
274. D. S. Capstick; J. M. Willey; M. J. Buttner; M. A. Elliot, *Mol. Microbiol.* **2007**, *64*, 602–613.
275. E. M. Lawton; R. P. Ross; C. Hill; P. D. Cotter, *Mini Rev. Med. Chem.* **2007**, *7*, 1236–1247.
276. J. D. Hillman, Antimicrobial Polypeptides and Methods of Use. U.S. Patent 7,067,125, 2006.
277. B. Y. Wang; H. K. Kuramitsu, *Appl. Environ. Microbiol.* **2005**, *71*, 354–356.
278. J. P. Burton; C. N. Chilcott; C. J. Moore; G. Speiser; J. R. Tagg, *J. Appl. Microbiol.* **2006**, *100*, 754–764.
279. M. Galvin; C. Hill; R. P. Ross, *Lett. Appl. Microbiol.* **1999**, *28*, 355–358.
280. E. B. O'Connor; B. O'Riordan; S. M. Morgan; H. Whelton; D. M. O'Mullane; R. P. Ross; C. Hill, *J. Appl. Microbiol.* **2006**, *100*, 1251–1260.
281. L. Silkin; S. Hamza; S. Kaufman; S. L. Cobb; J. C. Vederas, *Bioorg. Med. Chem. Lett.* **2007**, *18*, 3103–3106.
282. F. Castiglione; L. Cavaletti; D. Losi; A. Lazzarini; L. Carrano; M. Feroggio; I. Ciciliato; E. Corti; G. Candiani; F. Marinelli; E. Selva, *Biochemistry* **2007**, *46*, 5884–5895.
283. P. L. Zeitlin; M. P. Boyle; W. B. Guggino; L. Molina, *Chest* **2004**, *125*, 143–149.
284. M. J. Dawson; J. Cortes Bargallo; B. A. M. Rudd; S. Boakes; G. Bierbaum; A. Hoffmann; S. Schmitz, Production of Mersacidin and Its Variants in SigH and/or MRSA Negative Bacillus Host Cells. WO 2005/093069 A2, 2005.
285. H. M. Dodd; N. Horn; M. J. Gasson, *Gene* **1995**, *162*, 163–164.
286. C. van Kraaij; E. Breukink; H. S. Rollema; R. S. Bongers; H. A. Kosters; B. de Kruijff; O. P. Kuipers, *Eur. J. Biochem.* **2000**, *267*, 901–909.
287. W. Liu; J. N. Hansen, *Appl. Environ. Microbiol.* **1993**, *59*, 648–651.
288. L. Wang; P. G. Schultz, *Chem. Commun.* **2002**, 1–11.
289. T. W. Muir; D. Sondhi; P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710.
290. T. C. Evans, Jr.; M. Q. Xu, *Biopolymers* **1999**, *51*, 333–342.
291. X. Zhang; W. A. van der Donk, *J. Am. Chem. Soc.* **2007**, *129*, 2212–2213.

Biographical Sketches

Lisa E. Cooper was born and raised in Renton, Washington. She received a B.S. in Biochemistry in 2004 from Western Washington University under the guidance of Professor Christopher J. A. Daley. Subsequently, she joined the laboratory of Professor Wilfred A. van der Donk at the University of Illinois at Urbana-Champaign, where she studied lantibiotic biosynthetic enzymes. After successfully engineering an *in vitro* production system for the novel two-component lantibiotic haloduracin, she characterized the structural features essential for the activity of this antibiotic.

Bo Li was born in China. She received a B.S. in Life Sciences in 2004 from Beijing University. In the fall of 2004, she moved to the University of Illinois at Urbana-Champaign and joined the laboratory of Professor Wilfred A. van der Donk, where she studied the catalytic mechanism of the cyclase involved in the biosynthesis of nisin. Currently, she is focusing on the discovery of novel lantibiotics and the study of unprecedented lantibiotic biosynthetic enzymes.

Wilfred van der Donk was born in Culemborg, the Netherlands and received his B.S. and M.S. from Leiden University. He moved to the United States in 1989 to pursue his Ph.D. under Kevin Brugess at Rice University. After postdoctoral work at MIT with JoAnne Stubbe as a Jane Coffin Child fellow, he joined the faculty at the University of Illinois in 1997, where he is currently the Richard E. Heckert Professor of Chemistry and an Investigator of the Howard Hughes Medical Institute. The research in his laboratory focuses on using organic chemistry and molecular biology to gain a better understanding of the molecular mechanisms of enzyme catalysis, with a particular focus on natural products biosynthesis.

5.09 Plant Peptide Toxins from Nonmarine Environments

Jan-Christoph Westermann and David J. Craik, University of Queensland, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

5.09.1 Introduction

Plants use a variety of approaches for their defense, ranging from physical means such as thorns and bark or visual deception of herbivores to biochemical substances that make the plant less attractive by smell or taste or act as toxins creating adverse effects on the attacker, ranging from mild illness to death. The molecules involved in biochemical plant defense range from small organic molecules, including examples such as atropine, digoxin, and terpenoids, to peptides and proteins. One could argue that because plants lack mobility and hence lack the fight-or-flight defensive options of animals their biochemical defenses must be particularly effective in terms of potency, speed of action, and diversity.

These plant defense molecules are thus of interest for a number of reasons, including their potential applications in the discovery of new pharmacological substances, their adaptation to nonproducing species, for example, for protection of crop plants from insect pests delivered either topically or via incorporation into transgenic plants, and as new structural scaffolds for protein engineering approaches.

In this review we focus on the toxic molecules from plants that are peptidic in nature. Peptides exhibit a wide range of biochemical activities and relative to classic organic chemicals, have the advantage of relatively easy synthetic access and analytical identification. Identification of peptides can be directly done from plant extracts, via nucleic acid sequences isolated from plant tissue, or from genomic data, as opposed to organic molecules, which are generally detected only at the molecular rather than genetic level. Knowledge of the gene coding for a plant defense peptide or protein also allows the identification of the expression pattern of the gene product as a response to an external challenge such as predation, injury, or infection.

We commence with some definitions. Although the word toxin is commonly understood to refer to any harmful substance, a formal definition of toxin is "a poisonous substance, especially one produced by a living organism." (*The American Heritage Science Dictionary*, 2005 by Houghton Mifflin Company.) The poisonous effect may be discomfort, disease, or death for a target organism and can be achieved via several mechanisms. Typically, peptide toxins elicit their poisonous potential by targeting receptors, cell membranes, or enzymes crucial to a cell's or organism's metabolism.

Peptides are defined less stringently. *The American Heritage Science Dictionary* (2005) defines a peptide as "a chemical compound that is composed of a chain of two or more amino acids and is usually smaller than a protein." There is no exact cutoff size when a peptide becomes a protein. For the purpose of this review, and reflecting the term peptide in its title, an arbitrary cutoff of 100 amino acid residues was chosen as the upper size limit of peptides that will be covered, that is, we will not describe examples of plant-based proteins that are larger than 100 amino acids, even though there are many examples in the plant kingdom. In some cases the literature cited in this review may refer to some of the discussed toxins as proteins, reflecting the fact that individual researchers have their own preferences on peptide/protein nomenclature, but in all cases we limit coverage to polypeptides of fewer than 100 amino acid residues.

Plants face assaults from a diverse range of pests and pathogens, including microbes such as fungi and bacteria, insects, or higher animals. **Figure 1** summarizes some of the classes of peptides that defend against these attacks. From these examples it is clear that plant peptide toxins feature a vast array of sizes, targets, and biological modes

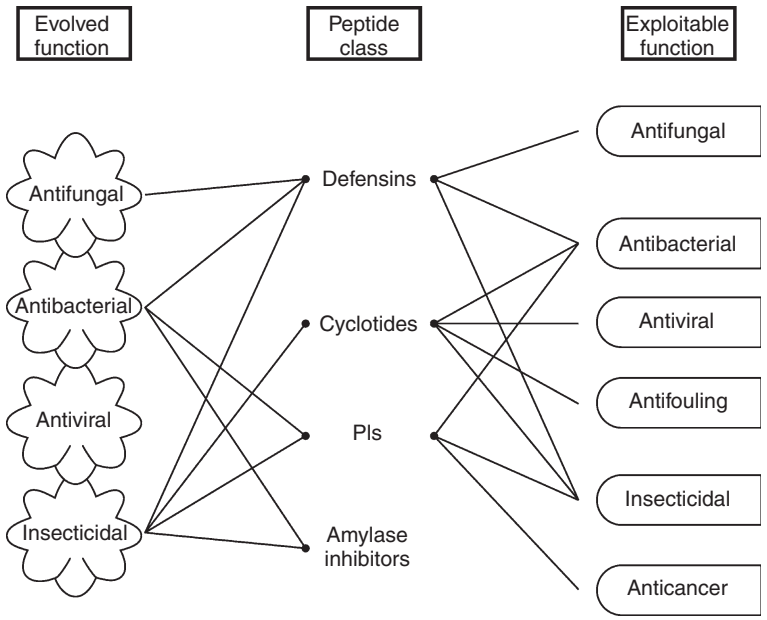
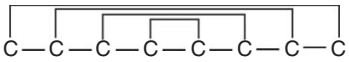
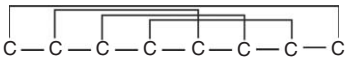
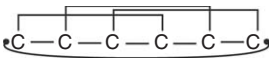
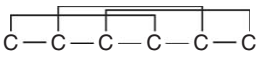
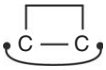
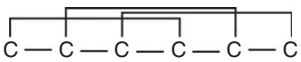


Figure 1 Examples of functions of different classes of plant defense peptides.

Table 1 Examples of plant defense peptides

Toxin	*Size (AA)*	*Source*	*Disulfide framework*	*Target*	*Mode of action*
α-Purothionin	46	*Triticum aestivum* (wheat)		Antimicrobial	Membrane binding
Rs-AFP2	50	*Raphanus sativus* (radish)		Antimicrobial	Membrane binding
Kalata B1	29	*Oldenlandia affinis* (kalata kalata)		Insecticidal	Membrane binding
CMCTI-1	29	*Cucumis melo*			Enzyme inhibitor
SFTI-1	14	*Helianthus annuus* (common sunflower)			Trypsin inhibitor
AAI	32	*Amaranthus hypochondriatus*		Insecticidal	α-Amylase inhibitor

The disulfide bridges and, if applicable, head-to-tail backbone cyclization are indicated on top and bottom, respectively, of the generic sequence.

of action, making them a broad field of research. Many of the peptide families discussed in this article are disulfide-rich peptides, as highlighted in **Table 1**. The occurrence of disulfide bridges in peptides often results in well-defined tertiary structures and many plant defense peptides have been structurally characterized by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. Reflecting our interests and expertise we thus place a substantial emphasis in this article on structures and structure–activity relationships of peptides.

In addition to the literature cited herein the reader might find it helpful to access publicly available databases on peptide and protein sequences to gain additional insight into the peptide families discussed here. These databases include PROSITE,[1] pfam,[2] and Swiss-Prot/TrEMBL.[3] Many of these databases and related analysis tools are accessible via the expert proteomics analysis system ExPASy.[4] As the number of peptide sequences grow as a result of discovery *in planta* or from genome screening, these resources will provide an ongoing update of knowledge on plant peptides and proteins.

The review is structured around major classes of toxic plant proteins. Each peptide family will be introduced, giving background information on discovery, distribution, and general biological activity. We then describe the structural features of the peptide families followed by a description of their biological activities with an emphasis on the mode of action to achieve their activity. First, the classes of plant defense peptides that are defined by a common architecture of the mature peptide, that is, thionins, plant defensins, and cyclotides will be described. These peptide families utilize different mechanisms of defense. We then describe examples of peptides that have defined toxic activity but originate from different peptide families. These examples include enzyme inhibitors targeting proteases or amylases. This review also covers the potential applications of plant-derived peptide toxins in crop protection and medical applications, either using transgenic or engineered and synthetic peptide approaches.

5.09.2 Plant Defense Peptides

5.09.2.1 Thionins

Thionins are cystine-rich, cationic small peptides (~5 kDa) found in monocots and eudicots.[5] They are divided into the families of α/β-thionins and γ-thionins. As is now generally accepted practice, we will refer to γ-thionins as plant defensins, as they are structurally more closely related to mammalian and insect defensins

(see Section 5.09.2.2) than to α/β-thionins. There are five classes of α/β-thionins (I–V) found in different tissues and different plants, and having different properties. Type I α/β-thionins comprise 45 amino acids and feature four disulfide bonds. Type II thionins are 46–47 amino acids in length. Type I and II thionins differ in the number of basic residues in their sequences, typically 10 for type I and 7 for type II. Type II thionins also have four disulfide bonds. Type III thionins feature three disulfide bonds and are 45–46 amino acids long. They are as basic as type II thionins. Thionins of type IV are approximately 46 amino acids long and of neutral charge. They have three disulfide bonds.[5,6] Type V thionins are reported to have evolved from type I thionins by a process of accelerated evolution.[7] They have shorter sequences, are of neutral charge, and their toxicity has not been fully elucidated.[5]

Thionins are expressed as precursor proteins that are processed to yield the mature peptide. The precursor peptide consists of an N-terminal signal sequence, the mature peptide sequence, and an acidic C-terminal protein.[6,8–10] Expression of thionins is inducible by external stimuli. For example, high levels of type II thionin mRNA are present in seedlings of barley grown in darkness. Upon exposure to light, levels of mRNA drop significantly but thionins expressed prior to light exposure remain stable.[11] Assault with fungal pathogens elicits transient expression of thionins under illumination[12,13] and chemical stress induces longer-lasting responses, as has been shown upon challenge with salts of magnesium, zinc, manganese, and cadmium.[14]

5.09.2.1.1 Structural aspects of thionins

Figure 2 shows a global alignment of selected thionin sequences. The background color for a given residue indicates the degree of conservation of that residue in a particular position in the sequence.[15,16] It is clear that thionins are highly conserved over different species. Based on their conserved sequences and similar three-dimensional structures it is reasonable to assume a common mode of action for all thionins.

Although thionins encompass a range of sequence diversity, and different types have different numbers of disulfide bridges, their overall structure is fairly similar, as revealed by either X-ray or NMR structures. The structural architecture consists of an N-terminal short β-strand linked to two antiparallel α-helices that are connected by a short random turn motif, followed by another β-strand forming an antiparallel β-sheet and a C-terminal coiled region. An overview of α/β-thionin structures published to date was reported in a recent review by B. Stec.[5] One of the best-studied peptides in the thionin family is crambin and very high-resolution structures of this peptide have been published with resolution as high as 0.54 Å.[17] Crambin differs from other thionins in its neutral charge, high hydrophobicity, and nontoxicity. **Figure 3** highlights the conserved structural features of thionins with the structures of α-purothionin,[18] β-purothionin,[19] crambin,[17] and viscotoxins A3[20] and B.[21]

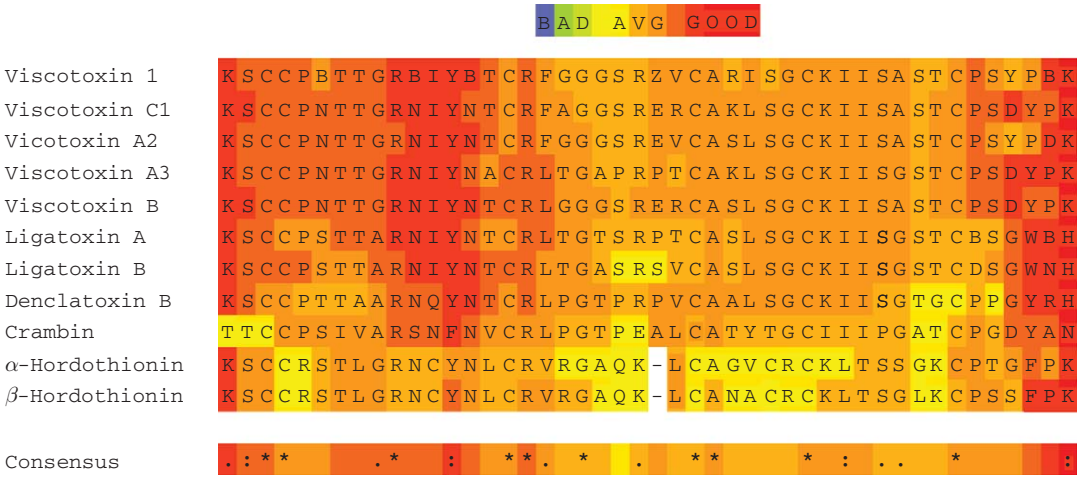


Figure 2 A global alignment of selected peptides from the thionin family reveals the similarities in their sequences. Increasing redness of the background indicates higher scores, that is, higher probability of the respective residue at that position (B: Asp or Asn, Z: Glu or Gln).

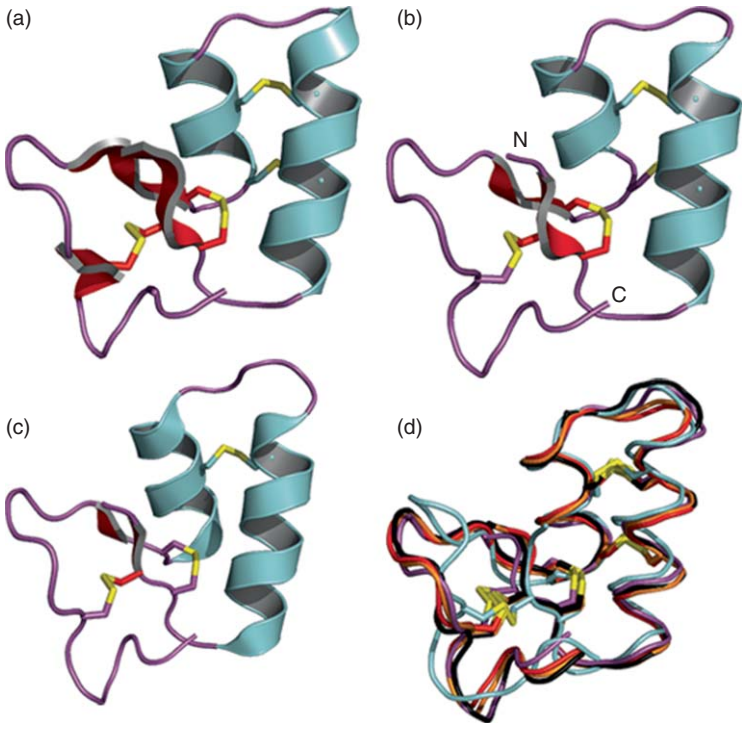


Figure 3 Comparison of several thionins reveals their structural similarity. The structures are color coded for their secondary structure, cyan: α-helix, red: β-strand, magenta: random coil/turn. (a) α-Purothionin (2plh) and (b) β-purothionin (1bhp) have four disulfide bonds. (c) Crambin (1ejg) has three disulfide bonds. Despite its thionin fold it lacks antimicrobial or other toxic activity. (d) An overlay of crambin (1ejg, black), α-purothionin (2plh, red), β-purothionin (1bhp, orange), viscotoxin A3 (1ed0, magenta), and viscotoxin B (1jmp, cyan) from *Viscum album* reveals the conserved structure of the peptide backbone.

5.09.2.1.2 Biological activity and mechanism of action

The toxic members of the thionin family have been shown to act on a variety of cells.[13,22–25] The findings support the assumption that these peptides are part of plant defense mechanisms after the first observations were made that thionins from wheat endosperm showed activity against various plant pathogens.[26] It is generally accepted that thionins exert their toxic effects via interactions with membranes[5,6,27] and many *in vitro* and *in vivo* studies support this.[28–40] Specifically, it is thought that thionins interact with the negatively charged head groups of phospholipids.[29,35] A striking feature of thionins is the occurrence of a conserved tyrosine at position 13 located in the first α-helix. In crambin this residue is phenylalanine (see **Figure 2**) and it has been suggested that this is a reason for lack of toxicity of crambin.[5] This residue is also less conserved in type IV thionins.[27] Chemical modifications of this tyrosine in purothionin results in decreased toxicity.[41]

Several crystal structures of thionins display the presence of polar molecules in the groove between β-strands and α-helices (**Figure 4**).[19,42] These molecules represent common head groups of lipids. In their bound position they are able to form a hydrogen bond with the hydroxyl group of Tyr13. A lack of capacity to form this hydrogen bond has been proposed as the reason for crambin's lack of toxicity. After initial binding, thionins either induce the formation of pores in the membrane[35] or they bind to rafts on the cell surface.[27,31] As a result of peptides binding to the membrane, cells suffer from loss of membrane integrity causing lysis, increased calcium permeability, and proteolytic degradation of membrane lipids.[43,44] Analysis of the time course and concentration dependence of thionin binding has suggested that all these effects can be explained with the carpet model of membrane action.[5] In this model, peptides accumulate on the membrane surface at high concentrations, resulting in changes of physical properties, for example, membrane fluidity causing strain on and ultimately disruption of the membrane.[45] In one of the more direct studies of the mode of action of peptides on cells the effect of purothionin on mammalian cells was determined by electron microscopy by Oka *et al.*[46]

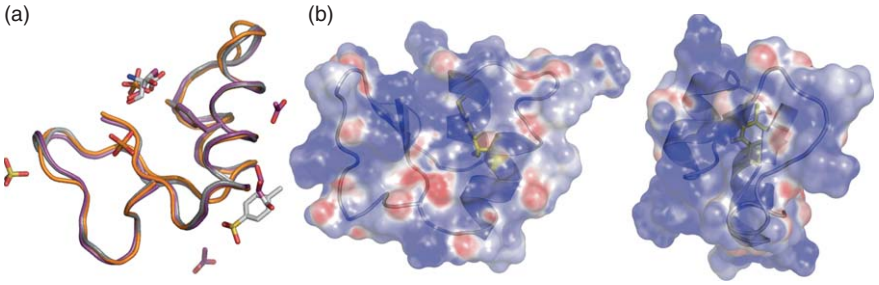


Figure 4 (a) Several X-ray structures of thionins reveal a conserved binding site in the groove between the α-helices and β-sheets. In this groove hydrophilic ligands like serine (hordothionin, 1wuw, gray), phosphate (viscotoxin A3, 1okh, orange), and glycerol (β-purothionin, 1bhp, magenta) bind preferentially. In the structure of hordothionin *p*-sulfo-toluene binds to the hydrophobic α-helices. Tyr13 is highlighted in bold and seems to be involved in ligand binding. (b) The isosurface of hordothionin reveals the distribution of positive (blue) and negative (red) charges. Positive charges dominate, especially the solvent-exposed surface of the α-helices (right).

An alternative mechanism for loss of membrane integrity is via the solubilization of phospholipids by thionins.[47] This revised version of the carpet model is based on the observation of a conserved binding site for phospholipids head groups in thionins, their tendency to form oligomers,[48,49] and small-angle X-ray scattering data.[47] Stec *et al.* propose that after an electrostatically driven approach to the membrane the hydrophobic α-helical face of the thionin inserts into the membrane, leading to the formation of patches of negatively charged lipids close to the highly positively charged thionin. This ultimately leads to the observed changes in physical properties and the formation of proteo-lipid complexes.[5,47]

5.09.2.2 Plant Defensins

Plant defensins are cystine-rich, cationic peptides ranging in size from 45 to 54 amino acids, of which eight are cysteine. They were first discovered in wheat and barley[50] and were proposed to form a novel subclass of thionins, the γ-thionins.[51] As it became clear that they closely resemble mammalian and insect defensins in primary and secondary structure, the term plant defensins was introduced to describe these peptides.[52,53] It is generally assumed that all plants express plant defensins[54–57] and that they are expressed in a wide range of plant tissue, that is, leaves,[53,58,59] floral tissue,[60–65] tubers,[61] bark,[66] root,[67] pods,[68] and seeds,[56,69] with seeds in particular being from where most plant defensins have been isolated.[52]

Plant defensins are expressed as precursor proteins of two different classes. The first class comprises the majority of plant defensins and the precursor consists of an N-terminal endoplasmic reticulum (ER) signal sequence followed by the mature peptide sequence.[63] The second class is found in solanaceous species and in these the precursor comprises an additional C-terminal prodomain of approximately 33 residues. Peptides belonging to this class have been found in floral tissue and fruit where they are expressed constitutively.[52,62,63,65] The C-terminal prodomain is rich in acidic and hydrophobic residues. At neutral pH, the charges of the prodomain counter the charges of the mature peptide domain. The C-terminal prodomain does not contain consensus sequences associated with sorting signals for extracellular or vacuolar targeting, but a high content of negatively charged and hydrophobic residues is common in vacuolar sorting determinants.[52,63] The defensin NaD1, which contains a C-terminal prodomain, has been shown to localize in vacuoles[63] whereas defensins originating from precursors expressed without the C-terminal prodomain, such as Rs-AFP2, are found in the outer cell layers and between the organs of seed.[53]

5.09.2.2.1 Structural aspects of plant defensins

Figure 5 shows a global alignment of the sequences of plant defensins described in this section. The background color highlights the conservation of residues throughout the sequence.[15,16] It is clear from this representation that in comparison to thionins (**Figure 2**) plant defensin sequences are more diverse. The sequences shown in **Figure 5** display several variations, including gaps and insertions of amino acids between the cysteine residues, resulting in variable loop sizes. It is reasonable to assume that the broad diversity of sequence and loop sizes in

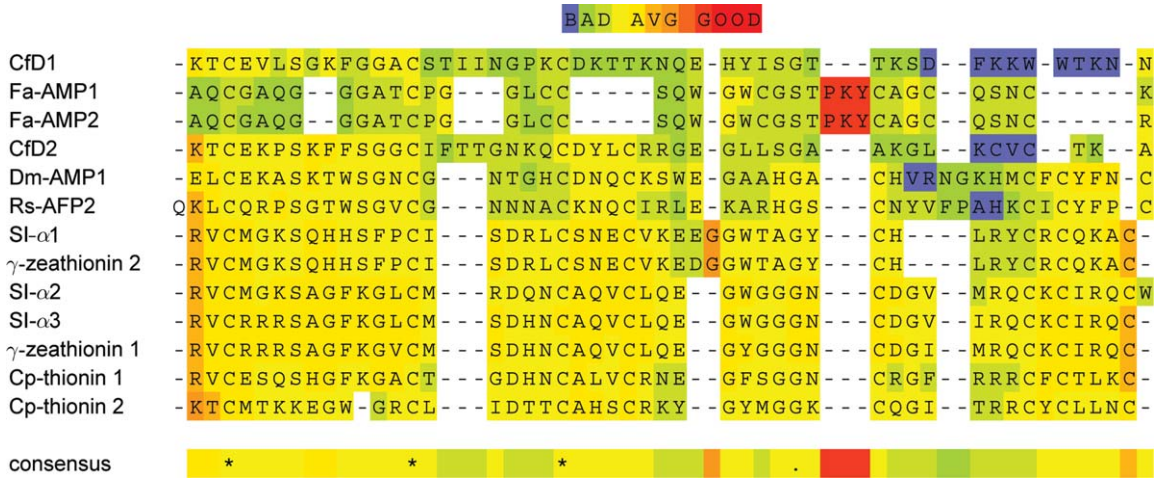


Figure 5 The increased sequence diversity of plant defensins in comparison to α/β-thionins (**Figure 2**) becomes clear in this sequence alignment. Peptides CfD1 and CfD2 show highly divergent residues toward the N-terminus. The PKY motif in Fa-AMP1 and Fa-AMP2 is unique to these peptides and the high content of glycine residues lowers the score in comparison to other plant defensins.

the plant defensins is one reason for the different modes of action for peptides of this family as the three-dimensional structure and surface chemistry of the peptides is different in the different members. For example, the C-terminal part of CfD1 and CfD2, described in the following section, are highly divergent from other plant defensin sequences, as is their biological activity.

The three-dimensional structural architecture of plant defensins is exemplified by the structure of Rs-AFP,[70] which comprises an N-terminal β-strand followed by an α-helix and two β-strands ($\beta\alpha\beta\beta$ configuration). The β-strands form a triple-stranded antiparallel β-sheet. The three-dimensional structure is stabilized by three disulfide bonds. In general, in plant defensins two disulfide bonds form between the α-helix and the central β-strand. A third disulfide bond stabilizes the structure by linking the β-strand after the helix to the coiled part after the α-helix. This motif is called the cysteine-stabilized $\alpha\beta$-motif (CS$\alpha\beta$)[45,70,71] and also occurs in toxins isolated from insects, spiders, and scorpions.[72–76] The fourth disulfide bond links the C-terminal end of the peptide with the N-terminal β-strand. Two plant defensins, PhD1 and PhD2, feature a fifth disulfide bond and have been proposed to be the prototypes of a new subclass within plant defensins.[45,71] As a result of these structural features the global structure of plant defensins is notably different from α/β-thionins, which is one of the reasons for the different nomenclature. The structures of plant defensins Rs-AFP[70] and NaD1[77] are shown in **Figure 6**, where they are compared to the thionin β-purothionin[19] and the structurally more related drosomycin[78] and charybdotoxin.[75]

5.09.2.2.2 Biological activity and mode of action

Plant defensins have a range of biological activities and target a wide range of plant pests, including fungi, bacteria, and insects. Additionally, their modes of action are more diverse than those observed for thionins.[52] Plant defensin expression is often induced upon infection with plant pathogens.[52,53,65,68,79]

Among their activities, plant defensins are potent antifungal agents[54,57,63] and this function is well characterized for a range of plant defensins.[53,57,80,81] In general, antifungal activity is characterized by the inhibition of fungal growth. The antifungal activity is exerted at the level of the fungus' plasma membrane, although the molecular target remains uncertain in many cases. Application of Rs-AFP2 from radish or Dm-AMP1 from dahlia at inhibitory concentration caused rapid influx of Ca^{2+} and efflux of K^{+}.[31,34]

Molecular targets have been elucidated for Dm-AMP1 and Rs-AFP2. Dm-AMP1 was found to bind plasma membranes from *Neurospora crassa* and *Saccharomyces cerevisiae* in a saturable manner and it competed with closely related defensins for binding.[82] Mutational studies with *S. cerevisiae* identified lipid rafts containing sphingolipids as a molecular target[69,83,84] while glycosylphosphatidylinositol (GPI)-anchored proteins could be

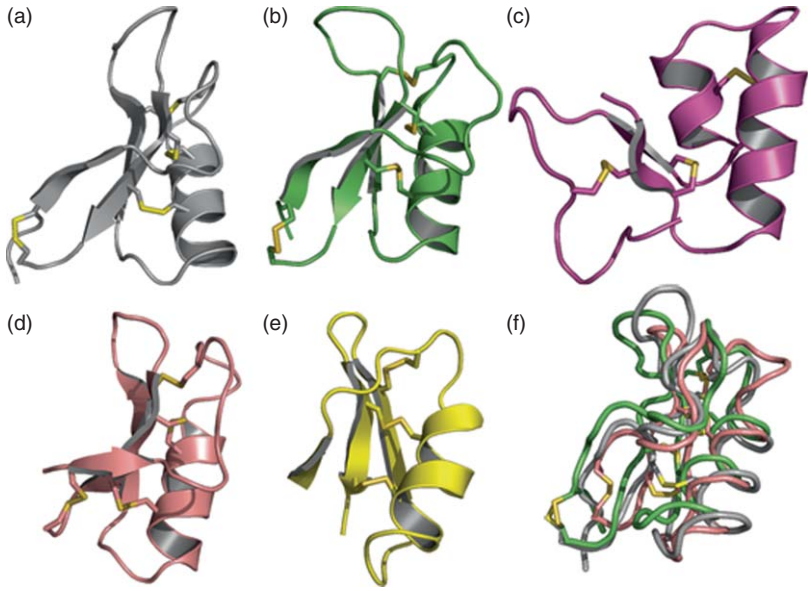


Figure 6 Comparison of the plant defensin structures Rs-AFP ((a), gray, 1ayj) and NaD1 ((b), green, 1mr4) with β-purothionin ((c), magenta, 1bhp) reveals the structural differences between plant defensins and α/β-thionins. The architecture resembles that of insect defensins, for example, drosomycin ((d), pink, 1myn) or the scorpion toxin charybdotoxin ((e), yellow, 2crd). The structural similarities become clear in the overlay of Rs-AFP, NaD1, and drosomycin ((f), colors as before).

ruled out as targets.[85] Rs-AFP2 on the other hand does not target *S. cerevisiae*. Mutational studies with *Pichia pastoris* identified glucosylceramide as the molecular target and the effect was specific for glucosylceramide from fungus while glucosylceramide from mammalian and plant sources were determined not to be binding partners of Rs-AFP2.[86] Interestingly, it was found in the same study that the mutant peptide Rs-AFP2[Y38G] does bind to glucosylceramide but does not have antifungal activity.

Several plant defensins have also been found to be inhibitors of various enzymes in plant pests. The plant defensins SIα1, SIα2, and SIα3 were the first plant defensins where inhibition of α-amylase was shown at low concentration[87] whereas purothionins inhibited α-amylase activity only at high concentrations.[88] BIα1 and BIα2 isolated from barley are two more representatives of proteinaceous α-amylase inhibitors in the plant defensin family.[89]

Inhibition of trypsin is another mechanism of activity recently discovered in plant defensins. CfD1 and CfD2 from *Cassia fistula* were the first plant defensins to be identified as trypsin inhibitors.[90] Cp-thionin from cowpea was more recently discovered to have inhibitory potency against trypsin.[91] Searches of protein sequence databases have yielded a number of other plant proteins annotated as trypsin inhibitors or potential trypsin inhibitors.[52] These annotations were most likely made on the basis of sequence similarities with other known trypsin inhibitors, namely the Bowman–Birk trypsin inhibitor. Since the actual framework of the disulfide bonds is not known, it is possible that structure and therefore activity differ from this prototype framework.[52,91]

Another interesting activity of plant defensins has been described by Kushmerick *et al.* for two defensins called γ1- and γ2-zeathionin from *Zea mays*. These two peptides showed fast and reversible inhibition of sodium channels in rat tumor cell lines.[92] Owing to an overall conservation of ion channels in eukaryotic cells it is feasible to assume that sodium channels in plant pests may be targets for these plant defensins, although their activities have not been determined *in vivo*.[52,92]

Fa-AMP1 and Fa-AMP2 isolated from buckwheat are distinguishable from other plant defensins in that they have a high content of glycine at 25%.[93] In contrast to other plant defensins and defensins from nonplant sources that target Gram-positive bacteria, Fa-AMP1 and Fa-AMP2 incorporate the characteristics from glycine-rich antimicrobial peptides and are also active on Gram-negative bacteria.[93,94]

5.09.2.3 Cyclotides

Like thionins and plant defensins, cyclotides[95–98] are disulfide-rich peptides. However, they have some unique features compared to other disulfide-rich plant defense peptides. In particular, they are distinguished from other plant peptides by their so-called cyclic cystine knot (CCK) motif.[99] This motif consists of a head-to-tail cyclized peptide backbone and three disulfide bonds that form a cystine knot. The cystine knot motif comprises a ring formed by two disulfide bonds and their connecting backbone segments that is penetrated by a third disulfide bond. In the cyclotides the ring-forming cysteine residues incorporates Cys^{I}–Cys^{IV} and Cys^{II}–Cys^{V}. The disulfide bond penetrating this ring is formed between Cys^{III} and Cys^{VI}. The structural features of the CCK motif are highlighted in **Figure 7**, which also shows the location of six loops that comprise the backbone segments between successive cysteine residues.

The backbone cyclization and cystine knot motif together render cyclotides as a class of structurally well-defined peptides that show exceptionally high stability against chemical, thermal, or enzymatic degradation.[100–105] This stability has meant that they have been proposed to have applications as protein engineering templates.[103]

So far cyclotides have been discovered in plants from the Violaceae (violet), Rubiaceae (coffee), and Cucurbitaceae (cucumber) families[104] and have been divided mainly into two structural subfamilies called the Möbius and bracelet cyclotides. These two cyclotide subfamilies are distinguished by the presence of a *cis*-proline residue in loop 5 for the Möbius subfamily.[95,96,106–108] On the basis of their trypsin inhibitory activity, the two cyclotides MCoTi-I and MCoTi-II from the seeds of the tropical vine *Momordica cochinchinensis*[109–111] form a third subfamily, referred to as the trypsin inhibitor subfamily of cyclotides. No other cyclotides have this activity.

Cyclotides have been isolated from many plant tissues, including leaves, stems, flowers, roots, and bark.[97,98,112–114] Recently, the levels of the Möbius-type cyclotide kalata B1 were quantified in various plant tissues of the Rubiaceae family plant *Oldenlandia affinis* by Seydel and Dörnenburg.[115] They found raised levels of kalata B1 in shoot tips and leaves, while flowers and stems contained significantly lower levels of the peptide.

Cyclotide genes encode precursor proteins that either contain one, two, or three copies of a single mature peptide domain, for example, Oak1 (*O. affinis* kalata B1) encodes one copy of kalata B1 and Oak4 encodes three copies of kalata B2,[116] or domains for different cyclotides, for example, Oak2 encodes both kalata B6 and kalata B3.[117] The precursors also contain an N-terminal ER signal sequence, a pro region, one or more copies of a highly conserved N-terminal repeat (NTR) region just upstream of the mature peptide domain, and a C-terminal tail.[104,117] When synthesized as an isolated peptide fragment, the NTR sequence adopts an α-helical structure, leading to the suggestion that it may play a structural role in folding or processing but no structures of any intact cyclotide precursor protein have yet been reported.[117] *In vivo* formation of the disulfide bridges may be facilitated by a protein disulfide isomerase (PDI) recently isolated from *O. affinis*,[118] while backbone cyclization is mediated by an asparaginyl endopeptidase.[119,120] Support for the latter proposal

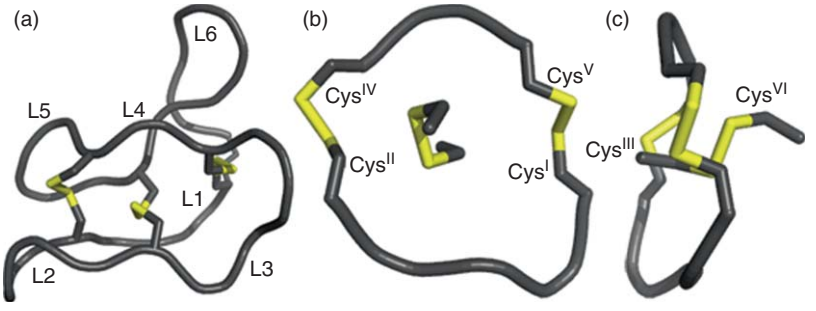


Figure 7 (a) The global topology of cyclotides is defined by the structure of kalata B1. The cysteine residues act as borders to the loops, labeled L1 to L6. (b) The cystine knot motif is defined by four cysteine residues that form a ring with the peptide backbone and disulfide bonds that is penetrated by the disulfide bond between Cys^{III} (pointing toward observer) and Cys^{VI} (in the back). The penetration is obvious when changing perspective by 90 degrees ((c), Cys^{III} left, Cys^{VI} right, Cys^{I}–Cys^{IV} bond pointing toward observer).

comes from the fact that for cyclization of the peptide backbone a C-terminal Asn residue is essential.[116,119,121] Violacin A, a recently discovered cyclotide derivative from the plant *Viola odorata*, has a point mutation at the C-terminal residue of the mature peptide (Lys instead of Asn) and as a result is a linear peptide. Its lack of a circular backbone leads to flexibility at the N-terminus, although it features a cystine knot motif and is structurally similar to cyclotides.[122]

5.09.2.3.1 Structural aspects of cyclotides

An overview of naturally occurring cyclotides is given in **Table 2**, with the sequences aligned based on the conserved cysteine residues.

As mentioned before, the cyclic backbone and cystine knot motif result in cyclotides being highly stable against thermal, chemical, and enzymatic degradation. The structure of kalata B1 was determined by NMR and reported by Saether *et al.* in 1995. This structure first revealed the, then surprising, CCK motif and showed that it was associated with a triple-stranded β-sheet and several turns (**Figure 7**).[140] The structure was refined in 2002 by Skjeldal[141] and in 2003 by Rosengren *et al.*,[108] again using NMR. In the latter study Rosengren *et al.* confirmed the cystine knot motif by using distance and dihedral angle restraints acquired for the cysteine residues in the sequence. When applying these restraints in simulated annealing methods used to determine the three-dimensional structure, it was clear that a structure comprising the cystine knot motif is by far more favorable in terms of energy and bond angles than an alternative structure comprising a laddered motif for the cysteine connectivity that had been proposed by Skjeldal.[141] The structural role of noncysteine residues was also elucidated by Rosengren *et al.*[108] Restraints data from NMR allowed the identification of a cis-peptide bond between Trp19 and Pro20, confirming the Möbius nature of the peptide's backbone in accordance with data reported by Skjeldal *et al.*[141] In contrast, cycloviolacin O1 has only two proline residues in its sequence and both adopt a trans-peptide bond as shown by NMR, and thus cycloviolacin O1 is a member of the bracelet family of cyclotides.[108]

Figure 8 gives an overview of the structural features of cyclotides exemplified by kalata B2 (Möbius subfamily),[107] cycloviolacin O1 (bracelet subfamily),[108] and MCoTi-II (trypsin inhibitor subfamily).[110] For comparison the noncyclic analogue violacin A is also shown.[122]

5.09.2.3.2 Biological activity and mode of action

The initial interest in cyclotides arose because of reports of their uterotonic,[142] anti-HIV,[143] and antineurotensin activity.[127] Cyclotides have more recently been reported to possess weak antimicrobial activity.[144] Specifically, kalata B1 and circulin A are inactive against Gram-negative bacteria; circulin B and cyclopsychotride are active against both Gram-positive and Gram-negative bacteria.[144] Impaired antimicrobial activity of cyclotides at higher salt concentration is indicative of ionic interactions. This effect of activity linked to ionic strength in the buffer has also been described for thionins and plant defensins and is an indicator for activity of the respective peptides at the membrane level. Cyclotides also possess hemolytic activity[138] as well as a range of other activitites.[96–98]

It is likely that the main functional role of cyclotides in plants is as defense against insect pests. Jennings *et al.*[116] showed that as part of an artificial diet, cyclotides had an impact on consumption of the diet by the caterpillars *Helicoverpa punctigera* and *Helicoverpa armigera*. In addition to this antifeedant effect, if *H. punctigera* and *H. armigera* are forced to consume diet containing kalata B1 and B2, respectively, *H. punctigera* failed to develop past the first instar, while *H. armigera* suffered from retarded growth and 25% loss in viability of larvae. Unlike other defense peptides with insecticidal activity kalata B1 and B2 do not inhibit enzymes in the digestive tract of insects[116] and the molecular basis of insecticidal activity seems to be of a different nature.

Recent studies suggest that the cyclotides kalata B1 and B6 bind to lipids and it is likely that membrane binding is a common mode of toxic action of all cyclotides.[145] The structure of kalata B1 in contact with lipid micelles has been determined by NMR.[146] The proposed interface of kalata B1 with a Dodecylphosphocholine (DPC) micelle is shown in **Figure 9**, and as expected coincides with a surface-exposed patch of hydrophobic residues. Consistent with the membrane binding hypothesis being general for cyclotides, Svangard *et al.*[147] demonstrated that cycloviolacin O2 is capable of disrupting lipid membranes.

Interaction with cell membranes is also supported by recent experimental data from Barbeta *et al.*[148] In feeding trials using artificial diets and larvae from *Helicoverpa armigera* it was shown that kalata B1 has a

Table 2 Sequences of currently known naturally occuring cyclotides

	Sequence								
Cyclotide subfamily	*Loop 1*	*Loop 2*	*Loop 3*	*Loop 4*	*Loop 5*	*Loop 6*	*Source*	*Reference*
Bracelet								
Circulin A	GI..P..**C**GES...**C**VWIP.**C**I.S.AAL.G.**C**S**C**	KN....KV**C**YR..N	*Chassalia parvifolia*	123
Circulin B	GV.IP..**C**GES...**C**VFIP.**C**I.ST.LL.G.**C**S**C**	KN....KV**C**YR..N	*C. parvifolia*	123
Circulin C	GI..P..**C**GES...**C**VFIP.**C**I.TS.VA.G.**C**S**C**	KS....KV**C**YR..N	*C. parvifolia*	123
Circulin D	KI..P..**C**GES...**C**VWIP.**C**V.TS.IF.N.**C**K**C**	EN....KV**C**YH..D	*C. parvifolia*	123
Circulin E	KI..P..**C**GES...**C**VWIP.**C**L.TS.VF.N.**C**K**C**	EN....KV**C**YH..D	*C. parvifolia*	123
Circulin F	AI..P..**C**GES...**C**VWIP.**C**I.S.AAI.G.**C**S**C**	KN....KV**C**YR...	*C. parvifolia*	123
Hyfl A	SIS....**C**GES...**C**VYIP.**C**TVT ALV.G.**C**T**C**	KD....KV**C**Y..LN	*Hybanthus floribundus E*	124
Hyfl B	G.S.PIQ**C**AET...**C**FIGK.**C**Y.TEEL..G.**C**T**C**	TA....FL**C**MK..N	*H. floribundus E*	124
Hyfl C	G.S.PRQ**C**AET...**C**FIGK.**C**Y.TEEL..G.**C**T**C**	TA....FL**C**MK..N	*H. floribundus E*	124
Hypa A	GI..P..**C**AES...**C**VYIP.**C**TIT.ALL.G.**C**S**C**	KN....KV**C**Y...N	*H. parviflorus*	125
Cycloviolin A	GV.IP..**C**GES...**C**VFIP.**C**I.SAAI..G.**C**S**C**	KN....KV**C**YR..N	*Leonia cymosa*	126
Cycloviolin B	GT.A...**C**GES...**C**YVLP.**C**F.T.V...G.**C**T**C**	TS...SQ.**C**FK..N	*L. cymosa*	126
Cycloviolin C	GI..P..**C**GES...**C**VFIP.**C**L.TTVA..G.**C**S**C**	KN....KV**C**YR..N	*L. cymosa*	126
Cycloviolin D	GF..P..**C**GES...**C**VFIP.**C**I.S.AAI.G.**C**S**C**	KN....KV**C**YR..N	*L. cymosa*	126
Kalata B5	G.T.P .**C**GES...**C**VYIP.**C**I.SGVI..G.**C**S**C**	TD....KV**C**Y..LN	*Oldenlandia affinis*	95
Kalata B16	GI..P..**C**AES...**C**VYIP.**C**TIT.A.LLG.**C**K**C**	QD....KV**C**Y...D	*O. affinis*	95
Kalata B17	GI..P..**C**AES...**C**VYIP.**C**TIT.A.LLG.**C**K**C**	KD....QV**C**Y...N	*O. affinis*	95
Cyclopsychotride A	SI..P..**C**GES...**C**VFIP.**C**TVT.A.LLG.**C**S**C**	KS....KV**C**YK..N	*Psychotria longipes*	127
Vico A	G.SIP..**C**AES...**C**VYIP.**C**F.TGIA..G.**C**S**C**	KN....KV**C**Y..YN	*Viola cotyledon*	128
Vico B	G.SIP..**C**AES...**C**VYIP.**C**I.TGIA..G.**C**S**C**	KN....KV**C**Y..YN	*V. cotyledon*	128
Vhl-1	SIS....**C**GES...**C**AMISF**C**F.TEVI..G.**C**S**C**	KN....KV**C**Y..LN	*V. hederaceae*	129
Vhr1	GI..P..**C**AES...**C**VWIP.**C**TVT ALL.G.**C**S**C**	SN....KV**C**Y...N	*V. hederaceae*	130
Cycloviolacin H1	GI..P..**C**GES...**C**VYIP.**C**L.TS.AI.G.**C**S**C**	KS....KV**C**YR..N	*V. hederaceae*	95
Cycloviolacin H2	SA.IA..**C**GES...**C**VYIP.**C**FI..P...G.**C**S**C**	RN....RV**C**Y..LN	*V. hederaceae*	129
Cycloviolacin H4	GI..P..**C**AES...**C**VWIP.**C**TVT.ALL.G.**C**S**C**	SN....NV**C**Y...N	*V. hederaceae*	131
Cycloviolacin O1	GI..P..**C**AES...**C**VYIP.**C**TVT.ALL.G.**C**S**C**	SN....RV**C**Y...N	*V. odorata*	95
Cycloviolacin O2	GI..P..**C**GES...**C**VWIP.**C**I.SSAI .G.**C**S**C**	KS....KV**C**YR..N	*V. odorata*	95
Cycloviolacin O3	GI..P..**C**GES...**C**VWIP.**C**L.TSAI..G.**C**S**C**	KS....KV**C**YR..N	*V. odorata*	95
Cycloviolacin O4	GI..P..**C**GES...**C**VWIP.**C**I.SSAI..G.**C**S**C**	KN....KV**C**YR..N	*V. odorata*	95
Cycloviolacin O5	G.T.P..**C**GES...**C**VWIP.**C**I.SSAV..G.**C**S**C**	KN....KV**C**YK..N	*V. odorata*	95
Cycloviolacin O6	G.TLP..**C**GES...**C**VWIP.**C**I.S.AAV.G.**C**S**C**	KS....KV**C**YK..N	*V. odorata*	95
Cycloviolacin O7	SI. P..**C**GES...**C**VWIP.**C**TIT.ALA.G.**C**K**C**	KS....KV**C**Y...N	*V. odorata*	95
Cycloviolacin O8	G.TLP..**C**GES...**C**VWIP.**C**I.SSVV..G.**C**S**C**	KS....KV**C**YK..N	*V. odorata*	95
Cycloviolacin O9	GI..P..**C**GES...**C**VWIP.**C**L.TSAV..G.**C**S**C**	KS....KV**C**YR..N	*V. odorata*	95
Cycloviolacin O10	GI..P..**C**GES...**C**VYIP.**C**L.TSAV .G.**C**S**C**	KS....KV**C**YR..N	*V. odorata*	95

(*Continued*)

Table 2 (Continued)

Cyclotide subfamily	*Sequence*						*Source*	*Reference*
	Loop 1	*Loop 2*	*Loop 3*	*Loop 4*	*Loop 5*	*Loop 6*		
Cycloviolacin O11	G . T L P . . **C** G E S . . . **C** V W I P . **C** I . S . A V V . G . **C** S **C**	K S . . . . K V **C** Y K . .	N	*V. odorata*	95
Cycloviolacin O13	G I . . P . . **C** G E S . . . **C** V W I P . **C** I . S . A A I . G . **C** S **C**	K S . . . . K V **C** Y R . .	N	*V. odorata*	132
Cycloviolacin O17	G I . . P . . **C** G E S . . . **C** V W I P . **C** I . S . A A I . G . **C** S **C**	K N . . . . K V **C** Y R . .	N	*V. odorata*	132
Cycloviolacin O18	G I . . P . . **C** G E S . . . **C** V Y I P . **C** T V T . A L A . G . **C** K **C**	K S . . . . K V **C** Y . . .	N	*V. odorata*	132
Cycloviolacin O19	G . T L P . . **C** G E S . . . **C** V W I P . **C** I . S S . V V . G . **C** S **C**	K S . . . . K V **C** Y K . .	D	*V. odorata*	132
Cycloviolacin O20	G I . . P . . **C** G E S . . . **C** V W I P . **C** L . T S A I . . G . **C** S **C**	K S . . . . K V **C** Y R . .	D	*V. odorata*	132
Cycloviolacin O25	D I . F . . . **C** G E T . . . **C** A F I P . **C** I . T H V P . . G T **C** S **C**	K S . . . . K V **C** Y . . F	N	*V. odorata*	132
Cycloviolacin Y4	G V . . P . . **C** G E S . . . **C** V F I P . **C** I . T G . V I . G . **C** S **C**	S S . . . . N V **C** Y . . L	N	*V. yedoensis*	133
Cycloviolacin Y5	G I . . P . . **C** A E S . . . **C** V W I P . **C** T V T . A L V . G . **C** S **C**	S D . . . . K V **C** Y . . .	N	*V. yedoensis*	133
Kalata B8	G . S V L N . **C** G E T . . . **C** L L G T . **C** Y . T T . . . . G . **C** T **C**	N K . . Y . R V **C** T K . .	D	*O. affinis*	134
Kalata B9	G . S V F N . **C** G E T . . . **C** V L G T . **C** Y . T . P . . . G . **C** T **C**	N T . . Y . R V **C** T K . .	D	*O. affinis*	114
Palicourein	G . D . P T F **C** G E T . . . **C** R V I P V **C** T Y S . A A L G . **C** T **C**	D D R S . D G L **C** K R . .	N	*P. condensata*	135
Vitri A	G I . . P . . **C** G E S . . . **C** V W I P . **C** I . T S A I . . G . **C** S **C**	K S . . . . K V **C** Y R . .	N	*V. tricolor*	136
Cycloviolacin Y1	G G T . I F D **C** G E T . . . **C** F L G T . **C** Y . T . P . . . G . **C** S **C**	G N . . . . G L **C** Y G T .	N	*V. yedoensis*	133
Cycloviolacin Y2	G G T . I F D **C** G E S . . . **C** F L G T . **C** Y . T . A . . . G . **C** S **C**	G N . . W . G L **C** Y G T .	N	*V. yedoensis*	133
Cycloviolacin Y3	G G T . I F D **C** G E T . . . **C** F L G T . **C** Y . T . A . . . G . **C** S **C**	G N . . W . G L **C** Y G T .	N	*V. yedoensis*	133
Tricyclon A	G G T . I F D **C** G E S . . . **C** F L G T . **C** Y . T K . . . . G . **C** S **C**	G E . . W . K L **C** Y G T .	N	*V. tricolor, V. arvensis*	121
Möbius								
Varv peptide A^{Z}	G L . . P V . **C** G E T . . . **C** V G G T . **C** N . T . P . . . G . **C** S **C**	S . . . W . P V **C** T R . .	N	*V. arvensis, V. odorata, V. tricolor*	137
Varv peptide B	G L . . P V . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** S **C**	D P W . P M **C** S R . .	N	*V. arvensis*	113
Varv peptide C	G V . . P I . **C** G E T . . . **C** V G G T . **C** N . T . P . . . G . **C** S **C**	S . . . W . P V **C** T R . .	N	*V. arvensis*	113
Varv peptide D	G L . . P I . **C** G E T . . . **C** V G G S . **C** N . T . P . . . G . **C** S **C**	S . . . W . P V **C** T R . .	N	*V. arvensis*	113
Varv peptide E^{Y}	G L . . P I . **C** G E T . . . **C** V G G T . **C** N . T . P . . . G . **C** S **C**	S . . . W . P V **C** T R . .	N	*V. arvensis, V. tricolor*	113
Varv peptide F	G V . . P I . **C** G E T . . . **C** T L G T . **C** Y . T . A . . . G . **C** S **C**	S . . . W . P V **C** T R . .	N	*V. arvensis*	113
Varv peptide G	G V . . P V . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** S **C**	D . . P W . P V **C** S R . .	N	*V. arvensis*	113
Varv peptide H	G L . . P V . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** S **C**	E T . . W . P V **C** S R . .	N	*V. arvensis*	113
Violapeptide 1	G L . . P V . **C** G E T . . . **C** V G G T . **C** N . T . P . . . G . **C** S **C**	S R . . . . P V **C** T X . .	N	*V. arvensis*	138
Kalata B1	G L . . P V . **C** G E T . . . **C** V G G T . **C** N . T . P . . . G . **C** T **C**	S . . . W . P V **C** T R . .	N	*O. affinis, V. odorata*	95
Kalata B2	G L . . P V . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** S **C**	T . . . W . P I **C** T R . .	D	*O. affinis*	95
Kalata B3	G L . . P T . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** T **C**	D . . P W . P I **C** T R . .	D	*O. affinis*	95
Kalata B4	G L . . P V . **C** G E T . . . **C** V G G T . **C** N . T . P . . . G . **C** T **C**	S . . . W . P V **C** T R . .	D	*O. affinis*	95
Kalata B6	G L . . P T . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** S **C**	S S . . W . P I **C** T R . .	N	*O. affinis*	116
Kalata B7	G L . . P V . **C** G E T . . . **C** T L G T . **C** Y . T Q . . . . G . **C** T **C**	S . . . W . P I **C** K R . .	N	*O. affinis*	116
Kalata B10	G L . . P T . **C** G E T . . . **C** F G G T . **C** N . T . P . . . G . **C** S **C**	S S . . W . P I **C** T R . .	D	*O. affinis*	114

Kalata B11	GL..PV.**C**GET...**C**FGGT.**C**N.T.P...G.**C**S**C** TD....PI**C**TR..D	*O. affinis*	114
Kalata B12	G.S.L..**C**GDT...**C**FVLG.**C**ND.S.....S**C**S**C** N...Y.PI**C**VK..D	*O. affinis*	114
Kalata B13	GL..PV.**C**GET...**C**FGGT.**C**N.T.P...G.**C**A**C** D..PW.PV**C**TR..D	*O. affinis*	114
Kalata B14	GL..PV.**C**GES...**C**FGGT.**C**N.T.P...G.**C**A**C** D..PW.PV**C**TR..D	*O. affinis*	114
Kalata B15	GL..PV.**C**GES...**C**FGGS.**C**Y.T.P...G.**C**S**C** T...W.PI**C**TR..D	*O. affinis*	114
Kalata S^Z	GL..PV.**C**GET...**C**VGGT.**C**N.T P...G.**C**S**C** S...W.PV**C**TR..N	*O. affinis*	95
Vhl-2	GL..PV.**C**GET...**C**FTGT.**C**Y.TN....G.**C**T**C** D..PW.PV**C**TR..N	*V. hederaceae*	129
Cycloviolacin H3	GL..PV.**C**GET...**C**FGGT.**C**N.T.P...G.**C**I**C** D..PW.PV**C**TR..N	*V. hederaceae*	129
Cycloviolacin $O12^Y$	GL..PI.**C**GET...**C**VGGT.**C**N.T.P...G.**C**S**C** S...W.PV**C**TR..N	*V. odorata*	95
Cycloviolacin O14	G.SIPA.**C**GES...**C**FKGK.**C**Y.T.P...G.**C**S**C** SK..Y.PL**C**AK..N	*V. odorata*	132
Cycloviolacin O15	GL.VP..**C**GET...**C**FTGK.**C**Y.T.P...G.**C**S**C** S...Y.PI**C**KK..N	*V. odorata*	132
Cycloviolacin O16	GL..P..**C**GET...**C**FTGK.**C**Y.T.P...G.**C**S**C** S...Y.PI**C**KK.IN	*V. odorata*	132
Cycloviolacin O21	GL..PV.**C**GET...**C**VTGS.**C**Y.T.P...G.**C**T**C** S...W.PV**C**TR..N	*V. odorata*	132
Cycloviolacin O22	GL..PI.**C**GET...**C**VGGT.**C**N.T.P...G.**C**T**C** S...W.PV**C**TR..N	*V. odorata*	132
cycloviolacin O23	GL..PT.**C**GET...**C**FGGT.**C**N.T.P...G.**C**T**C** DSS.W.PI**C**TH..N	*V. odorata*	132
Cycloviolacin O24	GL..PT.**C**GET...**C**FGGT.**C**N.T.P...G.**C**T**C** D..PW.PV**C**TH..N	*V. odorata*	132
Vodo M	GA..PI.**C**GES...**C**FTGK.**C**Y.T.V...Q.**C**S**C** S...W.PV**C**TR..N	*V. odorata*	139
Vodo N	GL..PV.**C**GET...**C**TLGK.**C**Y.T.A...G.**C**S**C** S...W.PV**C**YR..N	*V. odorata*	139
Trypsin Inhibitor			
MCoTI-I	GG.V...**C**PKILQR**C**RRDSD**C**....P...GA**C**I**C** RG...NGY**C**GSGSD	*Momordica cochinchinensis*	111
MCoTI-II	GG.V...**C**PKILKK**C**RRDSD**C**....P...GA**C**I**C** RG...NGY**C**GSGSD	*M. cochinchinensis*	111
MCoTI-IIb	GG.V...**C**PKILKK**C**RRDSD**C**....P...GA**C**I**C** RG...NGY**C**GSGSD	*M. cochinchinensis*	110
MCoTI-IIs	GG.V...**C**PKILKK**C**RRDSD**C**....P...GA**C**I**C** RG..NGY**C**GSGSD	*M. cochinchinensis*	110
Linear derivatives			
Violacin A	SA.I.S.**C**GET...**C**FKFK.**C**Y.T.P...R.**C**S**C** S...Y.PV**C**K....	*V. odorata*	132

Z,Y – these peptides have the same sequence but different names. The six cysteine residues conserved in all cyclotides are highlighted in bold.

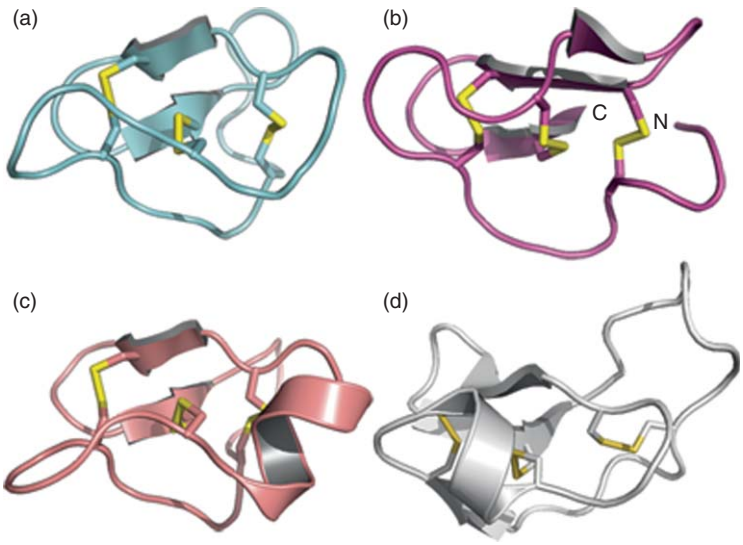


Figure 8 Comparison of structures of various cyclotides reveals their conserved structural characteristics. All peptides have been structurally aligned to the cysteine residues (a) kalata B2 (1pt4). (b) Violacin A (2fqa) is similar to cyclotides though it lacks the head-to-tail cyclized backbone; N and C denote the N- and C-termini. (c) Cycloviolacin O1 (1nbj) is a member of the bracelet family and its structure contains a small α-helical motif in loop 2. (d) MCoTi-II (1ha9) is a member of the trypsin inhibitor family; the peptide has a small α-helical motif in loop 4.

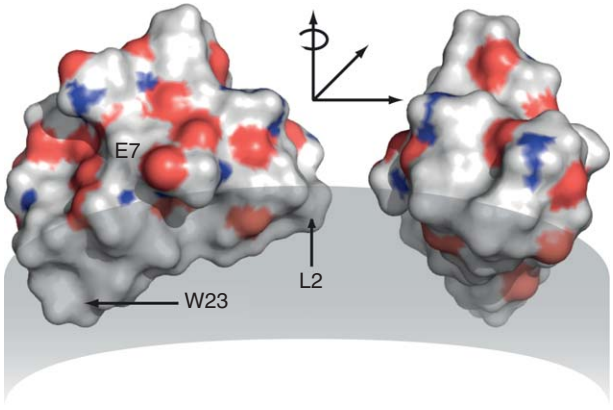


Figure 9 The proposed binding interface of kalata B1-binding DPC micelles is located on the hydrophobic surface of kalata B1. In the surface representation, oxygen is highlighted in red, nitrogen in blue, and carbon and hydrogen in white. The micelle is indicated in gray. It is clear that the surface of kalata B1 between Leu2 and Trp23 lacks polar residues and is ideally suited to interact with lipids. Glu7 forms part of a putative binding site for metal ions.

concentration-dependent effect on growth and the urge to feed as described above. Ingestion of kalata B1 had a pronounced effect on the diameter of the midgut, which was three times bigger in larvae on control diet. Reduction of gut diameter was attributed to swelling of the layer of columnar and goblet cells lining the gut observed by light and electron microscopy. Epithelial cells of kalata-fed caterpillars were elongated and protruding into the lumen of the gut. While at a low concentration of kalata B1 the epithelial cells beneath the columnar and goblet cells remained intact, higher concentration of the peptide had a more dramatic effect: epithelial cells were gone, columnar cells had lysed, and the peritropic membrane's morphology had changed.[148] Regardless of peptide concentration, damage was more pronounced anterior than posterior. It was concluded that malnutrition as a result of damage to the gut was responsible for retarded development.[148] In the same study it was also found that the cyclic backbone of cyclotides is essential for insecticidal activity. Larvae

on a diet of an acyclic mutant of kalata B1 consumed the diet without problems and their guts showed only minor damage compared to the native peptide's effect.[148]

These observations make a strong case for the interaction of cyclotides with cell membranes, particulary in the insect gut, as the basis for their biological activity *in planta*. Although there is strong evidence for membrane binding as the mode of action for cyclotides, it has been shown that kalata B1 does not penetrate the cell wall to internalize into cells, whereas MCoTi-II is internalized by macrophages and breast cancer cells *in vitro*.[149]

5.09.2.4 Plant Proteinase Inhibitors

The peptides discussed so far are defined by a common genetic pattern or architectural feature, such as their sequence or disulfide bond pattern. In this section we discuss peptides that share a common mode of action but may arise from different peptide families. Proteinase inhibitors (PIs) come in an astounding range of sizes, from the smallest gene-encoded cyclic peptide known to date, sunflower trypsin inhibitor 1 (SFTI-1),[150,151] a 14-residue cyclic peptide with a single disulfide bond, to squash inhibitors[109] that are approximately 30 residues in size and feature the cystine knot motif, to 53-residue PIs found in *Nicotiana alata*.[152,153]

Based on their sequence homology, disulfide connectivity, and cysteine location within the sequence and chemistry of the reactive site, PIs can be assigned to distinct families, as classified by Laskowski and Kato.[154] Kunitz-type, Bowman–Birk-type, Potato type I and type II, and squash inhibitors are members of these families shown in **Table 3**. For inhibitors not falling into these classifications more families have been proposed.[155,156] PIs can also be classified by their target/mode of action.[157] Plants have been found to express PIs that target serine proteinases, cysteine proteinases, aspartic proteinases, and metallo-proteinases. Serine and cysteine protease inhibitors are the best-studied PIs.[157]

5.09.2.4.1 Kunitz-type inhibitors

Kunitz-type protease inhibitors usually are large polypeptides of more than 120 amino acids and will not be discussed here, based on the focus of the review being peptides of fewer than 100 amino acids in length. Searches of publicly available databases yield a few short sequences, which are fragments from larger proteins that have only been partially sequenced.

5.09.2.4.2 Bowman–Birk inhibitors

Bowman–Birk protease inhibitors (BBIs) are among the best-studied serine protease inhibitors. They are found abundantly in dicotyledonous and monocotyledonous plants, with the former species expressing inhibitors of approximately 8 kDa in size with two reactive sites (double headed) and the latter expressing 8 kDa inhibitors with one reactive site and 16 kDa inhibitors with two reactive sites. Dicot BBIs feature 14 cysteine residues involved in disulfide bonds; monocot BBIs have 10 cysteine residues.[158]

5.09.2.4.3 Structural aspects of Bowman–Birk inhibitors

The reactive site of BBIs usually consists of a seven-residue loop held in position by a disulfide bond from cysteine residues at the termini of the loop. The loop itself forms the tip of a double-stranded β-sheet extending

Table 3 Proteinase inhibitor families found in plants

Inhibitor family	*Name*	*Pfam ID*[2]
I	Soybean trypsin inhibitor (Kunitz-type)	PF00197
II	Soybean trypsin inhibitor (Bowman–Birk-type)	PF00228
III	Potato I inhibitor	PF00280
IV	Potato II inhibitor	PF02428
V	Squash inhibitor	PF00299
VI	Other families	

Adapted from S. K. Haq; S. M. Atif; R. H. Khan, *Arch. Biochem. Biophys.* **2004**, *431*, 145–159.

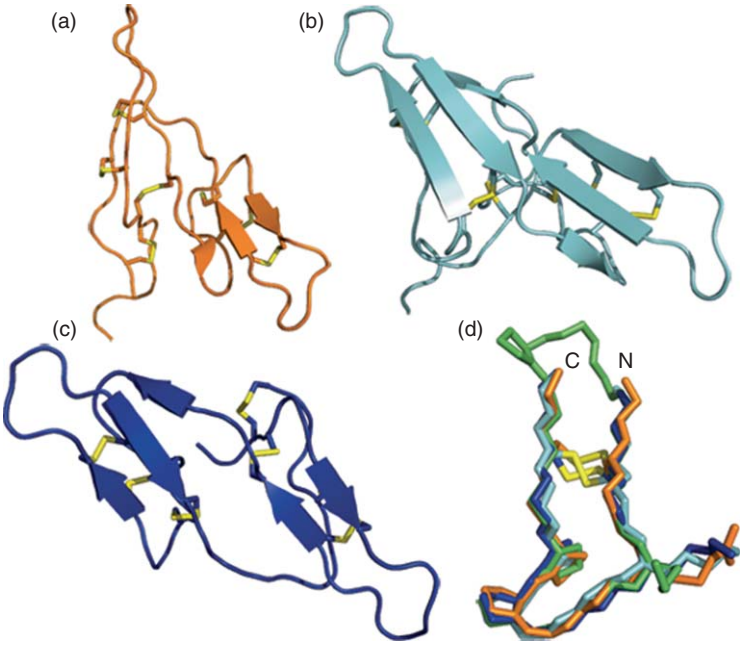


Figure 10 The global structures of Bowman–Birk inhibitors are diverse, but the reactive sites are structurally highly conserved. Reactive sites are oriented to the bottom right corner. The double-stranded β-sheets neighboring the reactive sites are clearly visible. (a) 1mvz, (b) 1bbi, and (c) 1h34. (d) The overlay of the backbone traces of the three BBIs ((a)–(c), colors as before) reveals their structural similarity. Additionally SFTI-1 (green, 1jbl) adopts the same fold in its active site. The P_1 residues (Arg in 1mvz, Lys for other peptides) are also shown; N and C denote the N- and C-termini of the truncated loops.

the termini of the reactive site. The P_1 residue of the first reactive site usually is a basic amino acid, that is, lysine or arginine and is responsible for its specificity for trypsin. The second reactive site can feature broader specificity against other proteases. Different residues at position P_1 result in other specificities. For example, Ala instead of Lys or Arg leads to elastase specificity, and Phe is the optimal residue for chymotrypsin inhibition.[159] The reactive sites of BBIs show remarkable structural similarities. The structural characteristics of this inhibitor loop have also been adopted by the inhibitor loop of the small cyclic trypsin inhibitor SFTI-1, which, owing to its single disulfide bond does not belong to the Bowman–Birk inhibitor family. **Figure 10** displays a few examples of Bowman–Birk inhibitors, including an inhibitor peptide from *Medicago scutellata* seeds,[160] the prototypical inhibitor peptide from soybean,[161] and an inhibitor isolated from *Phaseolus lunatus*[42] and depicts the structural conservation of the reactive site compared with SFTI-1.[162]

5.09.2.4.4 Potato type I and type II proteinase inhibitors

In tomato and potato plants, two families of PIs have been identified that show wound-inducible expression.[163–166] Within 48 h after wounding of leaves of these plants, up to 2% of leaf protein content is made up of protease inhibitors.[167] Potato type I (PotI) inhibitors are not only expressed in wounded leaves but also in tissue of leaves away from the site of attack.[166] This finding prompted the search for a hormone or inducing factor that carries the signal for enhanced inhibitor expression through the plant system.[168] Systemin, an 18-residue peptide hormone, has since been reported to induce more than 15 genes of defense nature in tomatoes.[169]

PotI inhibitors differ from other protease inhibitors, and from all other defense peptides mentioned thus far, in their relative lack of disulfide bonds. This means that the loop with the reactive site is not fixed, as it is in the Bowman–Birk inhibitors, yet they still form a stable fold, as shown in **Figure 11**. An interesting feature of some PotI inhibitors is their tendency to form stable, noncovalently bound oligomers. This has, for example, been shown for chymotrypsin inhibitor I from tomato.[170] This peptide has a monomer weight of 8300 Da under dissociating sodium dodecyl sulfate (SDS) gel conditions. Gel filtration and ultracentrifugal analysis revealed a

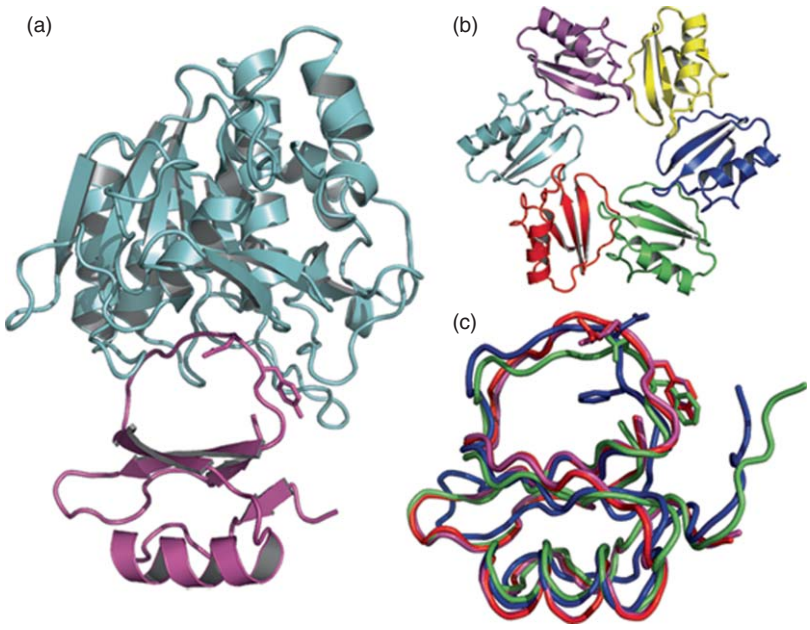


Figure 11 (a) CI2 (magenta) from *Hordeum* sp. in complex with subtilisin novo (*Bacillus amyloliquefaciens*, cyan) shows the reactive site of the inhibitor in the active site of the enzyme. Amino acids flanking the cleavage site in the inhibitor are shown with their side chains (Glu60 and Tyr61, 2sni). (b) chymotrypsin inhibitor 2 (CI2) from *Hordeum vulgare* forms a hexamer in the crystal structure (1coa). The structural motif of PotI inhibitors becomes obvious in the overlay of several structures (c). The side chains of amino acids flanking the cleavage site are shown (2sni, cyan; 1coa red, Glu60 and Tyr61; 1dwm, *Linum usitatissimum* trypsin inhibitor, green, Asp46 and Phe47; 1mit, *Cucurbita maxima* trypsin inhibitor V, blue, Asp45 and Phe46).

species at 41 000 Da, which was assigned to the pentameric form of the peptide.[170] The crystal structure of chymotrypsin inhibitor 2 from *Hordeum vulgare* revealed a hexameric oligomer (**Figure 11**).[171]

Potato type II (PotII) inhibitors are disulfide-rich peptides of approximately 50 amino acids in size. They were first discovered in leaves, seeds, and other organs of Solanaceae[172] and are a source of much interest as plant defense proteins. Recently, Barta *et al.*[173] analyzed expressed sequence tag (EST) and genomic data and discovered 11 genes that code for PotII inhibitors in various monocotyledonous and dicotyledonous plants. PotII inhibitors are expressed as large precursor proteins that contain up to eight sequence repeats of the inhibitor precursor.[173] In one particularly fascinating case from the ornamental tobacco (*N. alata*), the precursor adopts a circular permuted structure.[174,175] Barta *et al.* observed that genes outside the Solanaceae family seem to preferentially contain a single repeat unit, which has been predicted for the ancestral gene.[173–176] From gene sequences and species distribution Barta *et al.* proposed an accelerated evolution of the gene family. They proposed that multiplication of the repeat sequence happened prior to gene duplication.

PotII inhibitors inhibit several proteases. For example, serine protease inhibitors isolated from *Capsicum anuum* seeds named PSI-1.1 and PSI-1.2 are strong inhibitors of trypsin and only one order of magnitude less active on chymotrypsin. PSI-1.1 also has considerable activity against thrombin, whereas factor Xa is inhibited to a lesser extent by PSI-1.1 and PSI-1.2.[177,178]

5.09.2.4.5 *Proteinase inhibitors from* Nicotiana alata

Ornamental tobacco (*N. alata*) produces a series of 6 kDa peptides with inhibitory potency against trypsin and chymotrypsin named T1, T2, T3, T4, C1 and C2 respectively. The peptides derive from an approximately 40 kDa circular precursor protein (NaProPI), which after proteolytic cleavage yields five single-chain peptides (NaPIs).[179] A sixth two-chain peptide with chymotrypsin (C2) activity is formed from the N- and C-terminal domains of the circular precursor, linked via three disulfide bonds (**Figure 12**).[175] The sequence of NaProPI features six highly similar domains. Interestingly though, the cleavage of the precursor occurs within the

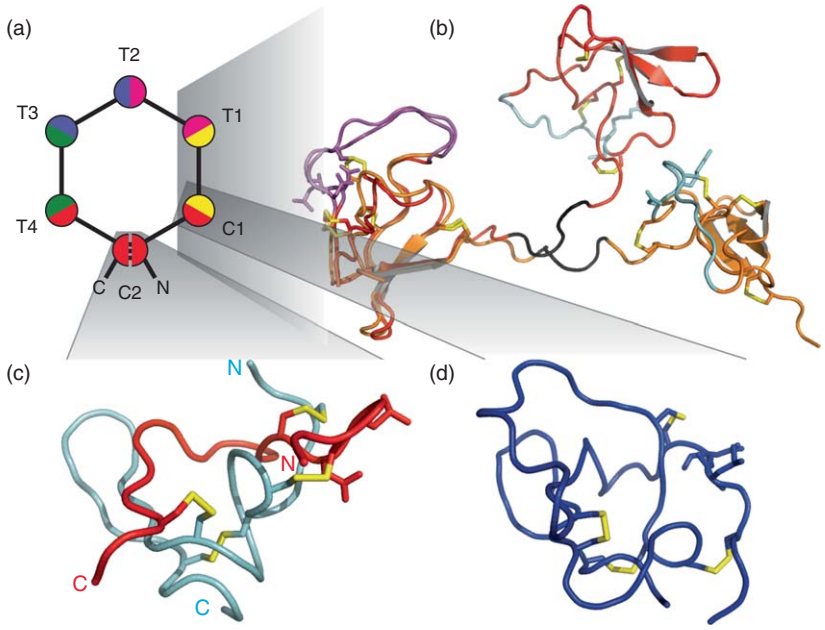


Figure 12 Schematic representation of NaProPI, its processing, and resulting structures. (a) A schematic representation of the precursor protein and the shift in the sequence repeats between the domains (domain swapping) indicated by color coding. The EEKKN motif excised during processing of the precursor is shown in black. (b) The flexibility of the precursor fragment of C1 and T1 (1fyb) with the first (orange) and the second (red) structure from the family of low-energy structures determined by NMR. The loops with the reactive sites are color coded, chymotrypsin in magenta and trypsin in cyan. (c) Structure of the two-chain chymotrypsin inhibitor C2 (1qh2). The N-terminal peptide is red, the C-terminal peptide cyan, and N- and C-termini are indicated in the respective colors. (d) Structure of C1 (1ce3).

repeats, resulting in each structural PI domain consisting of two halves of the sequence repeats, a phenomenon that has been referred to as intramolecular domain swapping.[153]

The reactive domains are cleaved from the precursor at a highly conserved repeat domain with the sequence EEKKN. Identification of small quantities of peptides with extended or shortened termini has been taken as an indicator for the involvement of unspecific proteases in the cleavage of the precursor.[180] Structural studies of a C1–T1 construct derived from the precursor protein using NMR indicate that the domains within the precursor fold independently from each other and that no interdomain interactions are detectable on long-term scale.[176]

5.09.2.4.6 Structural aspects of PotI, PotII, and related peptidic proteinase inhibitors

Despite their lack of stabilizing disulfide bridges PotI inhibitors feature a common, stable fold. The N-terminus is coiled, although in some structures a small β-strand has been identified. After a turn the structure adopts an α-helical structure, followed by a turn and an other β-strand. The sequence then features an extended turn or loop motif that contains the reactive site of the inhibitor before it proceeds with a β-strand running almost parallel to the β-strand after the α-helix. After another turn and coiled motif a short β-strand antiparallel to the other β-strands precedes the coiled C-terminus. Usually the N-terminal residue in the reactive site is an acidic residue followed by an aromatic amino acid, that is, tyrosine or phenylalanine. **Figure 11** shows the complex of chymotrypsin inhibitor (CI) 2 with subtilisin,[181] the hexamer of CI 2 from *H. vulgare*[171] and a structural comparison with a trypsin inhibitor from *Linum usitatissimum*.[182]

The global structure of PotII inhibitors is stabilized by disulfide bridges. The N-terminus features a coiled structure followed by two antiparallel β-strands connected with a turn motif. The second β-strand is linked to the N-terminal region via a disulfide bond. After the second β-strand the structure adopts an extended coiled loop that is attached to the N-terminal part through two disulfide bonds. This loop also contains the reactive site, which is positioned between the two disulfide bridges. The structure continues with a third β-strand

forming a triple-stranded β-sheet. The coiled C-terminus is connected to the N-terminal part of the first β-strand with a fourth disulfide bond. The global fold of PotII inhibitors from, for example, potato[183] and structurally related PIs from NaProPI from *N. alata*[179] is conserved. **Figure 12** depicts the structure of NaProPI, the structure of a two-domain construct, C1–T1 derived from NaProPI,[176] and the structures of the chymotrypsin inhibitor domains C2[175] and C1.[174]

5.09.2.4.7 Proteinase inhibitors from the squash family

Trypsin inhibitors in cucumber were first found in a study by Walker-Simmons *et al.*[184] after wounding of leaves and treatment with proteinase inhibitor-inducing factor (PIIF). The amino acid sequence of two inhibitors isolated from *Cucurbita maxima* (winter squash) were determined by Wilusz *et al.*[185] The peptides named ITD I and ITD III each comprised a 29-residue sequence with six cysteine residues. The only difference between the two peptides is in position 9, which is lysine in ITD I and glutamic acid in ITD III. The reactive site is located at the peptide bond between Arg5 and Ile6. Owing to their discovery and distribution in Cucurbitaceae the inhibitor family has been named squash inhibitors. Since the initial discoveries many other members of the squash family have been found.

5.09.2.4.8 Structural aspects of proteinase inhibitors from the squash family

The three-dimensional structure of trypsin inhibitor 1 from *C. maxima* was determined in 1989 by X-ray crystallography in complex with bovine trypsin[186] and by NMR[187] in aqueous solution. The three-dimensional structure revealed a cystine knot motif for the arrangement of disulfide bonds, as described above for the cyclotides. In contrast to cyclotides squash inhibitors do not, with the exception of MCoTi-II, have a cyclized peptide backbone. As shown for MCoTi-II in **Figure 8(d)**, squash inhibitors do not comprise extended secondary structure features. The occurrence of a short 3_{10}-α-helical motif is conserved over the family. The loop containing the reactive site is structurally similar to other small trypsin inhibitors.[186]

5.09.2.4.9 Biological activities and mode of action of protease inhibitory peptides

Heath *et al.*[188] studied the effect of PIs C1 and T1 to T4 from *N. alata* on enzymes occurring in guts of various insect pests as well as in feeding trials with *H. punctigera* larvae. PIs from *N. alata* showed activity on all enzymes tested with different efficacy. In the feeding trials using an artificial diet containing PIs from *N. alata* at biologically relevant concentrations there was retarded development of larvae compared to an artificial diet without PIs. All larvae on the PI-containing diet weighed less than the control group at the end of the feeding trial. In the same study cricket nymphs (*Teleogryllus commodus*) were fed artificial diet without any PIs, with PIs from *N. alata* and BBI from soybean. In these trials PIs from *N. alata* proved to have a stronger effect on gaining weight of the nymphs than the Bowman–Birk inhibitor. Survival was not significantly affected by the inhibitors, although PIs from *N. alata* seemed to have an effect on mobility of the nymphs.[188]

The activity of protease inhibitors is based on their ability to bind to a protease active site with high affinity while proteolysis of the reactive bond is slow compared to endogenous substrates.[189] Binding of inhibitors can be distinguished as being substrate canonical or product canonical.[189] The toxic effect of PIs was linked to a feedback effect in the intestine rather than malnutrition from blocked proteolysis.[190] As a result of trypsin inhibition organisms resort to hyperproduction of proteases, resulting in depleted levels of essential amino acids in the intestine, which in turn causes retarded development in larvae. The biosynthesis of proteins and peptides essential for development, for example, neuropeptides, also depends on trypsin. Impairment of these processes is likely to have an impact on development and growth.[157]

5.09.2.5 Proteinaceous Glycosidase Inhibitors from Plants

Glycosidases belong to a large family of enzymes involved in the proteolytic modification of carbohydrates in organisms. This includes buildup of carbohydrate structures and degradation as part of detoxification or metabolism. At first glance it may seem paradoxical to employ proteinaceous inhibitors to disrupt carbohydrate–enzyme interactions, but in fact plants and other organisms have evolved a vast range of these inhibitors to manage glycosidase activity and defend against pests.

A number of proteinaceous glycosidase inhibitors are proteins rather than peptides and are not covered here. The physiological and defense role of larger glycosidase inhibitor proteins has been reviewed by Bellincampi *et al.*[191]

α-Amylases and their inhibitors are best studied with respect to plant defense and glycosidases. These peptides are involved in the metabolism of carbohydrates by hydrolyzing α1-4 glycosidic bonds in carbohydrates, namely starch. Proteinaceous inhibitors of α-amylases have been found in many organisms, including plants. Peptide inhibitors target α-amylases from various organisms, including bacteria, insects, and mammals.[192] A major class of proteinaceous α-amylase inhibitors in plants includes proteins sharing similarities with Kunitz-type trypsin inhibitors, capable of inhibiting subtilisin and α-amylases. These α-amylase/subtilisin inhibitors have been found in barley (BASI),[193–195] wheat (WASI),[196] and rice (RASI).[197]

There are also some peptides exhibiting α-amylase inhibitory potency. The plant defensins SIα1, SIα2, and SIα3 from *Sorghum bicolor*[87] and BIα1 and BIα2 from barley[89] have been mentioned before. The knottin-type α-amylase inhibitor peptide AAI from *Amaranthus hypochondriacus*[198] is the smallest peptide α-amylase inhibitor known to date.[192]

5.09.2.5.1 Structural aspects of proteinaceous α-amylase inhibitor peptides

The solution structure of AAI has been solved by NMR, from which the knottin fold of the peptide was confirmed.[199] In an interesting example of protein engineering Lu *et al.* attributed α-amylase inhibitor activity to residues located in loop 1 (numbering taken from Lu *et al.*) by grafting the sequence IPKWNR from AAI into a structurally related spider toxin. As a result, the spider toxin chimera displayed α-amylase inhibitory activity, whereas the native peptide did not.[199] Martins *et al.*[200] studied the *cis–trans* isomerization occurring between Val15 and Pro16 in AAI using NMR. They found that 30% of the peptide featured *cis*-proline at that position and 70% of the peptide were in trans-conformation. Pereira *et al.*[201] were able to crystallize AAI bound to α-amylase from *Tenebrio molitor* (yellow mealworm) and resolve the X-ray structure (**Figure 13**).

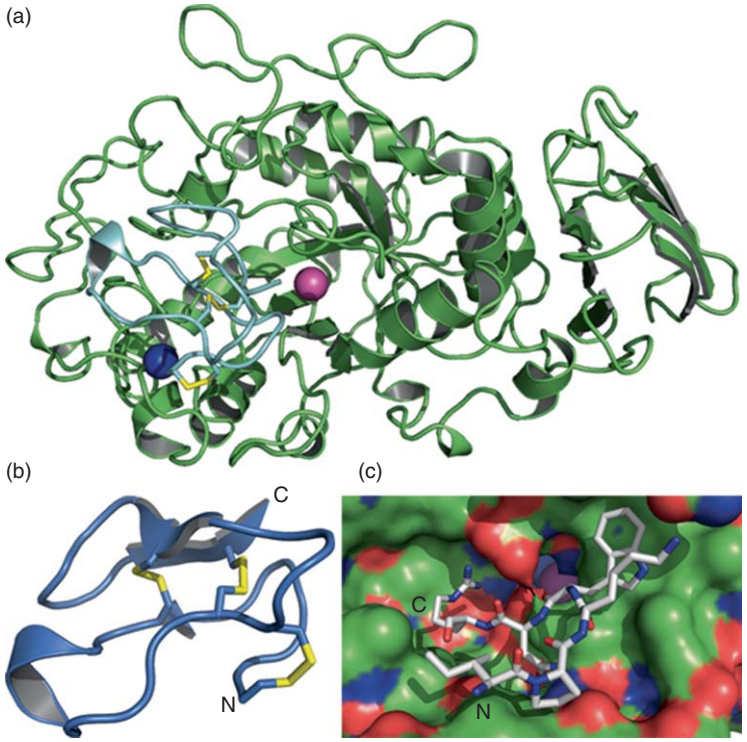


Figure 13 (a) Cartoon representation of *Tenebrio molitor* amylase (green) in complex with *Amaranthus hypochondriacus* α-amylase inhibitor (cyan). The structurally essential Ca^{2+} (blue) and catalytically required Cl^- (magenta) ions are shown as spheres (1clv). (b) Structure of AAI. N- and C-termini are indicated (1qfd). (c) Position of the binding interface (Ile2 to Arg7) in the binding site. N- and C-termini of the fragment are indicated. The receptor surface is color coded, green for carbon, red for oxygen, and blue for nitrogen atoms; the chloride ion is shown as magenta sphere. The ligand has white for carbon, and oxygen and nitrogen are color coded as in the receptor.

5.09.2.5.2 Biological activity and mode of action

As already mentioned, α-amylase inhibitors (proteins and peptides) target enzymes from many organisms that feed on plants, that is, fungi, bacteria, insects, and mammals.[192,199] Inhibition of α-amylase activity intervenes with metabolism of starch, which forms the main source of nutrition for organisms feeding on plants. Organisms exposed to α-amylase inhibitors therefore suffer from reduced availability of carbohydrates that serve as energy resource.[202] The peptides discussed here target α-amylase in insect guts and do not affect α-amylases in mammals.

To understand the inhibition of α-amylase by peptide inhibitors it is crucial to first understand the native substrate–enzyme interaction. The active site and the reaction mechanism of α-amylases have been identified from several X-ray structures of human and pig pancreatic amylases in complex with carbohydrate-based inhibitors. The structural aspects of proteinaceous α-amylase inhibition have been reviewed by Payan.[202] The sequence, architecture, and structure of α-amylases from mammals and insects are fairly homologous and mechanistic insights from mammalian enzymes can be used to elucidate inhibitor function with respect to insect enzymes. The architecture of α-amylases comprises three domains. Domain A contains the residues responsible for catalytic activity. It complexes a calcium ion, which is essential to maintain the active structure of the enzyme and the presence of a chloride ion close to the active site is required for activation.

The reaction mechanism of α-amylases is referred to as retaining, which means that the stereochemistry at the cleaved bond of the carbohydrate is retained. Hydrolysis of the glycosidic bond is mediated by an acid hydrolysis mechanism, which is in turn mediated by Asp197 and Glu233 in pig pancreatic amylase. These interactions have been identified from X-ray crystallography.[203] The aspartate residue has been shown to form a covalent bond with the C1 position of the substrate in X-ray structure of a complex formed by a structurally related glucosyltransferase.[204] The glutamate residue is located in vicinity to the chloride ion and acts as the acidic catalyst in the reaction.[205] The catalytic site of α-amylases is located in a V-shaped depression on the surface of the enzyme.

In the complex of AAI bound to α-amylase from *T. molitor*,[201] the peptide inhibitor binds in this depression and the general interactions are similar to those observed for other complexes of proteinaceous α-amylase inhibitors reviewed by Payan.[202] In more detail, 18 residues of the peptide inhibitor make contact with 24 residues of the enzyme, blocking access to the active site for subtrates.[201] *Tenebrio molitor* α-amylase (TMA) can accommodate up to six carbohydrate residues in its active site and four of those binding sites are blocked by AAI. The aspartate residue of the enzyme's active site interacts with AAI-Arg7 via a salt bridge and the glutamate residue forms a hydrogen bond with Arg7 of AAI mediated through a water molecule. Other residues in vicinity to the active site are also involved in a network of water-mediated hydrogen bonds. In contrast to other α-amylase protein inhibitor complexes AAI lacks extensive hydrophobic interactions. Based on structural comparisons Pereira *et al.*[201] were also able to elucidate a structural basis for the lack of AAI activity toward mammalian α-amylases. A model based on crystal structures from pig pancreatic α-amylase (PPA) and their own crystal data revealed six hydrogen bonds that could not be formed in the hypothetical AAI/PPA complex compared to the AAI/TMA complex.[201] The structures of AAI from NMR data and in complex with TMA are shown in **Figure 13**.

5.09.2.6 Other Peptides with Toxic Properties

Describing the whole kingdom of plant defense peptides in depth is beyond the scope of this chapter. In this section a few more peptide toxin activities implied in plant defense will be mentioned that do not fit into the categories mentioned before.

5.09.2.6.1 Chitin-binding peptides

The arsenal of plant defense peptides contains members capable of binding carbohydrate residues, namely β1-4 linked *N*-acetyl glucosamine residues that form the biopolymer chitin. The actual mode of action remains unclear. Antifungal and antimicrobial activity has been shown *in vitro*. For example Ac-AMP2 is a small disulfide-rich chitin-binding peptide isolated from the seeds of *Amaranthus caudatus* with antimicrobial activity.[206] It differs from Ac-AMP1 by one additional arginine residue at the C-terminus. The structure was determined by NMR and contains a cystine knot motif.[207] Ac-AMP2 displays a so-called hevein domain partly

characterized by the cysteine pattern, disulfide architecture, and the occurrence of a conserved aromatic residue in the domain, Phe18 in case of Ac-AMP2. The aromatic residue plays a key role in carbohydrate recognition by forming CH-π interactions.[208] Using NMR and chemically altered aromatic residues in position 18 the structure of the carbohydrate–peptide complex was elucidated.[209]

5.09.2.6.2 Lipid transfer peptides

Lipid transfer peptides and proteins occur in eukaryotic and prokaryotic cells. *In vitro* they possess the ability to transfer phospholipids between lipid membranes. Plant lipid transfer peptides are unspecific in their substrate selectivity. They bind phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, and glycolipids. Some of these peptides have shown antifungal activity *in vitro*.[210–212] The sequences of lipid transfer proteins and peptides contain 91–95 amino acids, are basic, and have eight cysteine residues forming four disulfide bonds. They do not contain tryptophan residues. About 40% of the sequence adopts a helical structure with helices linked via disulfide bonds. The tertiary structure comprises four α-helices. The three-dimensional structure of a lipid transfer peptide from *H. vulgare* in complex with palmitate has been solved by NMR.[213] In this structure the fatty acid is caged in a hydrophobic cavity formed by the helices.

5.09.3 Plant Peptide Toxins in Biotechnology and Pharmaceutical Applications

The broad spectrum of activities, in conjunction with the high efficacy of plant peptide toxins against crop pests or microbes, in general has led to a huge interest in the commercial exploitation of these molecules for crop protection or pharmaceutical applications. However, peptides have some disadvantages, making their application in agriculture or as pharmaceuticals difficult. First, peptides are susceptible to degradation by heat, chemical influences, for example, pH variations, and enzymatic digestion. However, to balance this, many of the plant peptides mentioned in this chapter are more stable to degradation than peptides from other sources of biological interest. Second, the synthetic availability of peptides can be limited. Many of the peptides discussed here are relatively large with more than 40 amino acid residues and sometimes complex disulfide bridging patterns. In general this makes these peptides difficult to obtain by means of chemical peptide synthesis. However, shorter peptides with a cystine knot motif have been shown to be amenable to chemical synthesis with yields satisfactory on the laboratory scale, as reviewed by Gunasekera *et al.*[214] for cyclotides. Indeed, many other peptides of the knottin family are also amenable to chemical synthesis. The alternative to chemical synthesis is the transfer of one or several genes coding for a plant peptide toxin into a plant of interest.

5.09.3.1 Transgenic Plants

Several successful examples are found in the literature where defense-related genes have been transformed into other plants, both model systems and crop plants. Some examples of transgenic plants and their conferred resistances are summarized in **Table 4**.

Table 4 Examples of transgenic plants, the gene sources, and conferred resistance

Defense peptide/ gene	*Source organism*	*Target organism*	*Attained resistance*	*Reference*
Thi2.1	*Arabidopsis thaliana*	Tomato	*Fusarium oxysporum, Ralstonia solanacearum*	215
α-Thionin	*Hordeum vulgare* (barley)	Tobacco	*Pseudomonas syringae*	216
AlfAFP	*Medicago sativa* (Alfalfa)	Potato	*Verticillium dahliae*	81
Rs-AFP2	*Raphanus sativus* (Radish)	Tobacco	*Alternaria longipes*	53
MTI-2	*Sinapis alba* (White mustard)	Tobacco, *Arabidopsis*, Oilseed rape	*Plutella xylostella, Mamestra brassicae, Spodoptera littoralis*	217

The thionin-encoding gene Thi2.1 from *Arabidopsis thaliana* was successfully transferred into tomato (*Lycopersicon esculentum*) plants. Thionin expression was detected in roots and leaves but not in fruit due to a fruit inactive promoter used to circumvent potential thionin toxicity in the fruit. Transgenic tobacco expressing α-thionin from barley displayed increased resistance against *Pseudomonas syringae*, a bacterial plant pathogen.[216]

Plant defensin genes have also been transformed into different plants. Rs-AFP2 has been transformed into tobacco[53] leading to enhanced protection against *Alternaria longipes*. The plant defensin alfAFP from *Medicago sativa* was potent *in vitro* against *Verticillium daliae*. This potency was also present in transgenic potato plant under greenhouse and field conditions.[81]

Genes coding for PIs are among the most commonly used for transgenic plants designed to have increased pest resistance. A comprehensive overview of protein and peptide genes transformed is given by Haq *et al.*[157] In general, serine protease inhibitors are toxic to *Lepidoptera*[218,219] and cysteine protease inhibitors are toxic to *Coleoptera*.[220,221] The effect of mustard trypsin inhibitor-2 (MTI-2) gene expressed in leaves from tobacco, *Arabidopsis*, and oilseed rape on several leptidopteran insect pest larvae has been studied by de Leo *et al.*[217] Tobacco and *Arabidopsis* especially exhibited the best levels of protection.

5.09.3.1.1 Nonplant expression systems

Some of the peptide families here have been successfully expressed in standard laboratory bacterial or yeast expression systems. This approach allows the production of larger amounts of peptides not amenable to chemical synthesis and opens the potential of large-scale production for biotechnological uses, for example, as plant protection agent applied externally.

Wisniewski *et al.*[66] transformed cDNA coding for the defensin PpDFN1 from *Prunis persica* into *P. pastoris*. The recombinant peptide displayed antimicrobial activity against *Penicillium expansum* and *Bortrytis cinerea*. Da-Hui *et al.*[222] expressed the defensin TDEF1 from *Trichosanthes krilowii* as a fusion protein fused to thioredoxin in *Escherichia coli*. The fusion protein exhibited inhibition of growth against *Fusarium oxysporum*. Olli and Kirti expressed TfgD1 from *Trigonella foenum-graecum* in *E. coli* as His- and S-tagged peptide.[223] The recombinant TfgD1 was active against *Rhizoctonia solani* and *Phaeoisariopsis personata*. Anaya-Lopez *et al.*[224] were able to express a defensin from *Capsicum chinense* in bovine endothelial cells. The recombinant peptide exhibited inhibitory activity against *Candida albicans* and cytotoxic effects against HeLa cells.

An early example of expression of a BBI was reported by Flecker.[225] The gene coding for the inhibitor was chemically synthesized with two mutations introduced to the sequence, that is, Lys16 > Arg and Met27 > Ile. The peptide was expressed as a fusion protein with β-galactosidase, cleaved and refolded. The resulting peptide showed comparable dissociation constants to the wild-type peptide. In a recent example, Vogtentanz *et al.*[226] were able to express Bowman–Birk trypsin inhibitor from soybean in *Bacillus subtilis* as a fusion protein fused to a cellulase-binding domain from cellulose.

In another recent promising development, cyclotides were expressed in *E. coli* despite the lack of native cystine knot peptides in the bacterial genome.[227–229]

5.09.3.2 Plant Peptide Toxins as Pharmaceutical Tools and Agents

The antimicrobial and cytotoxic activities described here for the peptide toxins from plants make them interesting compounds in the search for the treatment of diseases in humans.

5.09.3.2.1 Therapeutic applications of plant peptides

The use of defensins from insects and plants with antifungal activity for therapy was recently reviewed by Thevissen *et al.*[230] Several species of *Candida* were tested for their survival against defensins, including Dm-AMP1 and Rs-AFP2. The tests were also carried out with *Aspergillus flavus* and *Fusarium solani*. Depending on the target organism survival could be reduced to less than 1%. However, it must be kept in mind that some fungi are resistant to certain defensins, for example, *Candida krusei* was unaffected by Dm-AMP1.

Cyclotides have shown a vast array of activities with therapeutic potential. For example, as already mentioned, they encompass uterotonic,[112,142] antitumor,[231] antimicrobial,[144] and anti-HIV[133,143,232] activities, making them interesting peptides in a wide range of drug research.[233] The antifouling effect of cycloviolacin O2[234] is mentioned here for completeness, although not of therapeutic importance.

5.09.3.2.2 Stable plant peptides as carriers for bioactive epitopes

Owing to stable folding supported by disulfide bridges, many of the peptides discussed in this review show considerable stability against degradation compared to other peptides. This characteristic is very pronounced in peptides with head-to-tail cyclization, that is, SFTI-1 and cyclotides, which makes them interesting scaffolds for grafting, especially since both peptide classes can be chemically synthesized in an effective way. Grafting refers to the insertion of peptide epitopes with biological activity into an existing peptide framework, replacing a respective number of native residues. It can be considered as a multiple chemical mutation. The cyclic peptide acts as a donor of chemical and structural stability, and the epitope confers novel biological activity.

The concept of grafting peptide fragments into loops of cyclotides was reviewed recently[233] and the impact of modifications of the sequence on the structure of cyclotides has been studied by Clark *et al.*[106] In related studies Li *et al.*[235] used SFTI-1 to generate novel inhibitors of matriptase on the basis of the SFTI-1 framework. In their study they were able to create inhibitors with strong inhibitory effects and varying selectivity for proteases depending on the sequence.

5.09.3.3 Peptide Fragments from Plant Proteins

Many large proteins exert biological activity of pharmaceutical interest. In the drug design process the biological activity is often associated with an interaction mediated by certain residues or domains. Peptides made from these residues or domains often exhibit the same activity and can be used to derive novel compounds.

One example of a peptide derived from a plant protein is puroA from puroindoline A. The peptide comprises of the tryptophan-rich domain of the parent protein, which has antimicrobial activity *in vitro* and in transgenic rice.[236,237] The peptide FPVTWRWWKWWKG (puroA) from the Trp-rich domain in puroindoline A displayed activity against Gram-positive and Gram-negative bacteria. PuroA interacted with model membranes suggesting that membrane interactions were the biological mode of action.[238]

5.09.4 Concluding Remarks

As is clear from the preceding discussion, plants express a vast array of toxic peptides in their defense that are interesting as structural models, active compounds in crop protection, and active compounds in pharmaceutical applications. In the plant kingdom the defense mechanisms involved have produced peptides for many kinds of microorganisms and predators, and in future it is possible that for any new problem in bacterial, viral, or fungal infection, a solution can be found on the basis of leads from plant defense molecules.

References

1. N. Hulo; A. Bairoch; V. Bulliard; L. Cerutti; E. De Castro; P. S. Langendijk-Genevaux; M. Pagni; C. J. Sigrist, *Nucleic Acids Res.* **2006**, *34*, D227–D230.
2. R. D. Finn; J. Mistry; B. Schuster-Bockler; S. Griffiths-Jones; V. Hollich; T. Lassmann; S. Moxon; M. Marshall; A. Khanna; R. Durbin; S. R. Eddy; E. L. L. Sonnhammer; A. Bateman, *Nucleic Acids Res.* **2006**, *34*, D247–D251.
3. A. Bairoch; R. Apweiler; C. H. Wu; W. C. Barker; B. Boeckmann; S. Ferro; E. Gasteiger; H. Huang; R. Lopez; M. Magrane; M. J. Martin; D. A. Natale; C. O'Donovan; N. Redaschi; L. S. Yeh, *Nucleic Acids Res.* **2005**, *33*, D154–D159.
4. E. Gasteiger; A. Gattiker; C. Hoogland; I. Ivanyi; R. D. Appel; A. Bairoch, *Nucleic Acids Res.* **2003**, *31*, 3784–3788.
5. B. Stec, *Cell. Mol. Life Sci.* **2006**, *63*, 1370–1385.
6. F. Garcia-Olmedo; A. Molina; J. M. Alamillo; P. Rodriguez-Palenzuela, *Biopolymers* **1998**, *47*, 479–491.
7. A. Castagnaro; C. Marana; P. Carbonero; F. Garcia-Olmedo, *J. Mol. Biol.* **1992**, *224*, 1003–1009.
8. F. Ponz; J. Paz-Ares; C. Hernandez-Lucas; F. Garcia-Olmedo; P. Carbonero, *Eur. J. Biochem.* **1986**, *156*, 131–135.
9. C. Hernandez-Lucas; J. Royo; J. Paz-Ares; F. Ponz; F. Garcia-Olmedo; P. Carbonero, *FEBS Lett.* **1986**, *200*, 103–106.
10. F. Ponz; J. Paz-Ares; C. Hernandez-Lucas; P. Carbonero; F. Garcia-Olmedo, *EMBO J.* **1983**, *2*, 1035–1040.
11. U. Reimann-Philipp; S. Behnke; A. Batschauer; E. Schafer; K. Apel, *Eur. J. Biochem.* **1989**, *182*, 283–289.
12. F. Ebrahim-Nesbat; S. Behnke; A. Kleinhofs; K. Apel, *Planta* **1989**, *179*, 203–210.
13. H. Bohlmann; S. Clausen; S. Behnke; H. Giese; C. Hiller; U. Reimann-Philipp; G. Schrader; V. Barkholt; K. Apel, *EMBO J.* **1988**, *7*, 1559–1565.
14. R. Fischer; S. Behnke; K. Apel, *Planta* **1989**, *178*, 61–68.

15. O. Poirot; E. O'Toole; C. Notredame, *Nucleic Acids Res.* **2003**, *31*, 3503–3506.
16. C. Notredame; D. G. Higgins; J. Heringa, *J. Mol. Biol.* **2000**, *302*, 205–217.
17. C. Jelsch; M. M. Teeter; V. Lamzin; V. Pichon-Pesme; R. H. Blessing; C. Lecomte, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3171–3176.
18. U. Rao; B. Stec; M. M. Teeter, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51*, 904–913.
19. B. Stec; U. Rao; M. M. Teeter, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51*, 914–924.
20. S. Romagnoli; R. Ugolini; F. Fogolari; G. Schaller; K. Urech; M. Giannattasio; L. Ragona; H. Molinari, *Biochem. J.* **2000**, *350* (Pt 2), 569–577.
21. A. Coulon; A. Mosbah; A. Lopez; A. M. Sautereau; G. Schaller; K. Urech; P. Rouge; H. Darbon, *Biochem. J.* **2003**, *374*, 71–78.
22. B. P. Cammue; M. F. De Bolle; H. M. Schoofs; F. R. Terras; K. Thevissen; R. W. Osborn; S. B. Rees; W. F. Broekaert, *Ciba Found. Symp.* **1994**, *186*, 91–101; Discussion 101–106.
23. B. P. Cammue; M. F. De Bolle; F. R. Terras; P. Proost; J. Van Damme; S. B. Rees; J. Vanderleyden; W. F. Broekaert, *J. Biol. Chem.* **1992**, *267*, 2228–2233.
24. U. Reimann-Philipp; G. Schrader; E. Martinoia; V. Barkholt; K. Apel, *J. Biol. Chem.* **1989**, *264*, 8978–8984.
25. H. Bohlmann; A. Vignutelli; B. Hilpert; O. Miersch; C. Wasternack; K. Apel, *FEBS Lett.* **1998**, *437*, 281–286.
26. R. Fernandez de Caleya; B. Gonzalez-Pascual; F. Garcia-Olmedo; P. Carbonero, *Appl. Microbiol.* **1972**, *23*, 998–1000.
27. D. E. Florack; W. J. Stiekema, *Plant Mol. Biol.* **1994**, *26*, 25–37.
28. J. Evans; Y. D. Wang; K. P. Shaw; L. P. Vernon, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5849–5853.
29. V. R. Osorio e Castro; L. P. Vernon, *Toxicon* **1989**, *27*, 511–517.
30. W. Huang; L. P. Vernon; J. D. Bell, *Toxicon* **1994**, *32*, 789–797.
31. K. Thevissen; A. Ghazi; G. W. De Samblanx; C. Brownlee; R. W. Osborn; W. F. Broekaert, *J. Biol. Chem.* **1996**, *271*, 15018–15025.
32. W. Huang; L. P. Vernon; L. D. Hansen; J. D. Bell, *Biochemistry* **1997**, *36*, 2860–2866.
33. H. A. Wilson; W. Huang; J. B. Waldrip; A. M. Judd; L. P. Vernon; J. D. Bell, *Biochim. Biophys. Acta* **1997**, *1349*, 142–156.
34. K. Thevissen; F. R. Terras; W. F. Broekaert, *Appl. Environ. Microbiol.* **1999**, *65*, 5451–5458.
35. P. Hughes; E. Dennis; M. Whitecross; D. Llewellyn; P. Gage, *J. Biol. Chem.* **2000**, *275*, 823–827.
36. A. Coulon; E. Berkane; A. M. Sautereau; K. Urech; P. Rouge; A. Lopez, *Biochim. Biophys. Acta* **2002**, *1559*, 145–159.
37. J. A. Richard; I. Kelly; D. Marion; M. Pezolet; M. Auger, *Biophys. J.* **2002**, *83*, 2074–2083.
38. P. Llanos; M. Henriquez; J. Minic; K. Elmorjani; D. Marion; G. Riquelme; J. Molgo; E. Benoit, *Eur. Biophys. J.* **2004**, *33*, 283–284.
39. J. A. Richard; I. Kelly; D. Marion; M. Auger; M. Pezolet, *Biochemistry* **2005**, *44*, 52–61.
40. P. Llanos; M. Henriquez; J. Minic; K. Elmorjani; D. Marion; G. Riquelme; J. Molgo; E. Benoit, *FEBS J.* **2006**, *273*, 1710–1722.
41. K. Wada; Y. Ozaki; H. Matsubara; H. Yoshizumi, *J. Biochem.* **1982**, *91*, 257–263.
42. J. E. Debreczeni; G. Bunkoczi; B. Girmann; G. M. Sheldrick, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 393–395.
43. C. K. Angerhofer; W. T. Shier; L. P. Vernon, *Toxicon* **1990**, *28*, 547–557.
44. W. S. Fracki; D. Li; N. Owen; C. Perry; G. H. Naisbitt; L. P. Vernon, *Toxicon* **1992**, *30*, 1427–1440.
45. M. Giudici; R. Pascual; L. de la Canal; K. Pfuller; U. Pfuller; J. Villalain, *Biophys. J.* **2003**, *85*, 971–981.
46. T. Oka; Y. Murata; T. Nakanishi; H. Yoshizumi; H. Hayashida; Y. Ohtsuki; K. Toyoshima; A. Hakura, *Mol. Biol. Evol.* **1992**, *9*, 707–715.
47. B. Stec; O. Markman; U. Rao; G. Heffron; S. Henderson; L. P. Vernon; V. Brumfeld; M. M. Teeter, *J. Pept. Res.* **2004**, *64*, 210–224.
48. T. A. Egorov; T. I. Odintsova; V. A. Pukhalsky; E. V. Grishin, *Peptides* **2005**, *26*, 2064–2073.
49. A. G. Milbradt; F. Kerek; L. Moroder; C. Renner, *Biochemistry* **2003**, *42*, 2404–2411.
50. F. J. Colilla; A. Rocher; E. Mendez, *FEBS Lett.* **1990**, *270*, 191–194.
51. H. Bohlmann, *Crit. Rev. Plant Sci.* **1994**, *13*, 1–16.
52. F. T. Lay; M. A. Anderson, *Curr. Protein Pept. Sci.* **2005**, *6*, 85–101.
53. F. R. Terras; K. Eggermont; V. Kovaleva; N. V. Raikhel; R. W. Osborn; A. Kester; S. B. Rees; S. Torrekens; F. Van Leuven; J. Vanderleyden; B. P. A. Cammue; W. F. Broekaert, *Plant Cell* **1995**, *7*, 573–588.
54. W. F. Broekaert; B. P. A. Cammue; M. F. C. De Bolle; K. Thevissen; G. W. De Samblanx; R. W. Osborn, *Crit. Rev. Plant Sci.* **1997**, *16*, 297–323.
55. P. R. Shewry; J. A. Lucas, *Adv. Bot. Res.* **1997**, *26*, 135–192.
56. W. F. Broekaert; F. R. Terras; B. P. Cammue; R. W. Osborn, *Plant Physiol.* **1995**, *108*, 1353–1358.
57. R. W. Osborn; G. W. De Samblanx; K. Thevissen; I. Goderis; S. Torrekens; F. Van Leuven; S. Attenborough; S. B. Rees; W. F. Broekaert, *FEBS Lett.* **1995**, *368*, 257–262.
58. A. Segura; M. Moreno; A. Molina; F. Garcia-Olmedo, *FEBS Lett.* **1998**, *435*, 159–162.
59. K. M. Kragh; J. E. Nielsen; K. K. Nielsen; S. Dreboldt; J. D. Mikkelsen, *Mol. Plant Microbe Interact.* **1995**, *8*, 424–434.
60. H. C. Park; Y. H. Kang; H. J. Chun; J. C. Koo; Y. H. Cheong; C. Y. Kim; M. C. Kim; W. S. Chung; J. C. Kim; J. H. Yoo; Y. D. Koo; S. C. Koo; C. O. Lim; S. Y. Lee; M. J. Cho, *Plant Mol. Biol.* **2002**, *50*, 59–69.
61. M. Moreno; A. Segura; F. Garcia-Olmedo, *Eur. J. Biochem.* **1994**, *223*, 135–139.
62. S. B. Milligan; C. S. Gasser, *Plant Mol. Biol.* **1995**, *28*, 691–711.
63. F. T. Lay; F. Brugliera; M. A. Anderson, *Plant Physiol.* **2003**, *131*, 1283–1293.
64. B. Karunanandaa; A. Singh; T. H. Kao, *Plant Mol. Biol.* **1994**, *26*, 459–464.
65. Q. Gu; E. E. Kawata; M. J. Morse; H. M. Wu; A. Y. Cheung, *Mol. Gen. Genet.* **1992**, *234*, 89–96.
66. M. E. Wisniewski; C. L. Bassett; T. S. Artlip; R. P. Webb; W. J. Janisiewicz; J. L. Norelli; M. Goldway; S. Droby, *Physiol. Plant.* **2003**, *119*, 563–572.
67. P. Sharma; A. Lonneborg, *Plant Mol. Biol.* **1996**, *31*, 707–712.
68. C. C. Chiang; L. A. Hadwiger, *Mol. Plant Microbe Interact.* **1991**, *4*, 324–331.
69. B. P. Thomma; B. P. Cammue; K. Thevissen, *Curr. Drug Targets Infect. Disord.* **2003**, *3*, 1–8.
70. F. Fant; W. Vranken; W. Broekaert; F. Borremans, *J. Mol. Biol.* **1998**, *279*, 257–270.
71. B. J. Janssen; H. J. Schirra; F. T. Lay; M. A. Anderson; D. J. Craik, *Biochemistry* **2003**, *42*, 8214–8222.

72. J. H. Pease; D. E. Wemmer, *Biochemistry* **1988**, *27*, 8491–8498.
73. F. Bontems; C. Roumestand; B. Gilquin; A. Menez; F. Toma, *Science* **1991**, *254*, 1521–1523.
74. Y. Kobayashi; H. Takashima; H. Tamaoki; Y. Kyogoku; P. Lambert; H. Kuroda; N. Chino; T. X. Watanabe; T. Kimura; S. Sakakibara; L. Moroder, *Biopolymers* **1991**, *31*, 1213–1220.
75. F. Bontems; B. Gilquin; C. Roumestand; A. Menez; F. Toma, *Biochemistry* **1992**, *31*, 7756–7764.
76. B. Cornet; J. M. Bonmatin; C. Hetru; J. A. Hoffmann; M. Ptak; F. Vovelle, *Structure* **1995**, *3*, 435–448.
77. F. T. Lay; H. J. Schirra; M. J. Scanlon; M. A. Anderson; D. J. Craik, *J. Mol. Biol.* **2003**, *325*, 175–188.
78. C. Landon; P. Sodano; C. Hetru; J. Hoffmann; M. Ptak, *Protein Sci.* **1997**, *6*, 1878–1884.
79. I. A. Penninckx; K. Eggermont; F. R. Terras; B. P. Thomma; G. W. De Samblanx; A. Buchala; J. P. Metraux; J. M. Manners; W. F. Broekaert, *Plant Cell* **1996**, *8*, 2309–2323.
80. F. R. Terras; H. M. Schoofs; M. F. De Bolle; F. Van Leuven; S. B. Rees; J. Vanderleyden; B. P. Cammue; W. F. Broekaert, *J. Biol. Chem.* **1992**, *267*, 15301–15309.
81. A. G. Gao; S. M. Hakimi; C. A. Mittanck; Y. Wu; B. M. Woerner; D. M. Stark; D. M. Shah; J. Liang; C. M. Rommens, *Nat. Biotechnol.* **2000**, *18*, 1307–1310.
82. K. Thevissen; R. W. Osborn; D. P. Acland; W. F. Broekaert, *Mol. Plant Microbe Interact.* **2000**, *13*, 54–61.
83. K. Thevissen; B. P. Cammue; K. Lemaire; J. Winderickx; R. C. Dickson; R. L. Lester; K. K. Ferket; F. Van Even; A. H. Parret; W. F. Broekaert, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9531–9536.
84. Y. J. Im; J. Idkowiak-Baldys; K. Thevissen; B. P. Cammue; J. Y. Takemoto, *FEMS Microbiol. Lett.* **2003**, *223*, 199–203.
85. K. Thevissen; I. E. Francois; J. Y. Takemoto; K. K. Ferket; E. M. Meert; B. P. Cammue, *FEMS Microbiol. Lett.* **2003**, *226*, 169–173.
86. K. Thevissen; D. C. Warnecke; I. E. Francois; M. Leipelt; E. Heinz; C. Ott; U. Zahringer; B. P. Thomma; K. K. Ferket; B. P. Cammue, *J. Biol. Chem.* **2004**, *279*, 3900–3905.
87. C. Bloch, Jr.; M. Richardson, *FEBS Lett.* **1991**, *279*, 101–104.
88. B. L. Jones; P. Meredith, *Cereal Chem.* **1982**, *59*, 321.
89. N. Zhang; B. L. Jones; H. P. Tao, *Cereal Chem.* **1997**, *74*, 119–122.
90. R. Wijaya; G. M. Neumann; R. Condron; A. B. Hughes; G. M. Polya, *Plant Sci.* **2000**, *159*, 243–255.
91. F. R. Melo; D. J. Rigden; O. L. Franco; L. V. Mello; M. B. Ary; M. F. Grossi de Sa; C. Bloch, Jr., *Proteins* **2002**, *48*, 311–319.
92. C. Kushmerick; M. de Souza Castro; J. Santos Cruz; C. Bloch, Jr.; P. S. Beirao, *FEBS Lett.* **1998**, *440*, 302–306.
93. M. Fujimura; Y. Minami; K. Watanabe; K. Tadera, *Biosci. Biotechnol. Biochem.* **2003**, *67*, 1636–1642.
94. N. Sitaram, *Curr. Med. Chem.* **2006**, *13*, 679–696.
95. D. J. Craik; N. L. Daly; T. Bond; C. Waine, *J. Mol. Biol.* **1999**, *294*, 1327–1336.
96. D. J. Craik; N. L. Daly; J. Mulvenna; M. R. Plan; M. Trabi, *Curr. Protein Pept. Sci.* **2004**, *5*, 297–315.
97. U. Goransson; E. Svangard; P. Claeson; L. Bohlin, *Curr. Protein Pept. Sci.* **2004**, *5*, 317–329.
98. K. R. Gustafson; T. C. McKee; H. R. Bokesch, *Curr. Protein Pept. Sci.* **2004**, *5*, 331–340.
99. D. J. Craik; N. L. Daly; C. Waine, *Toxicon* **2001**, *39*, 43–60.
100. M. L. Colgrave; D. J. Craik, *Biochemistry* **2004**, *43*, 5965–5975.
101. D. J. Craik, *Science* **2006**, *311*, 1563–1564.
102. S. S. Nair; J. Romanuka; M. Billeter; L. Skjeldal; M. R. Emmett; C. L. Nilsson; A. G. Marshall, *Biochim. Biophys. Acta* **2006**, *1764*, 1568–1576.
103. D. J. Craik; M. Cemazar; N. L. Daly, *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 176–184.
104. C. W. Gruber; M. Cemazar; M. A. Anderson; D. J. Craik, *Toxicon* **2007**, *49*, 561–575.
105. P. B. Pelegrini; B. F. Quirino; O. L. Franco, *Peptides* **2007**, *28*, 1475–1481.
106. R. J. Clark; N. L. Daly; D. J. Craik, *Biochem. J.* **2006**, *394*, 85–93.
107. C. V. Jennings; K. J. Rosengren; N. L. Daly; M. Plan; J. Stevens; M. J. Scanlon; C. Waine; D. G. Norman; M. A. Anderson; D. J. Craik, *Biochemistry* **2005**, *44*, 851–860.
108. K. J. Rosengren; N. L. Daly; M. R. Plan; C. Waine; D. J. Craik, *J. Biol. Chem.* **2003**, *278*, 8606–8616.
109. L. Chiche; A. Heitz; J. C. Gelly; J. Gracy; P. T. Chau; P. T. Ha; J. F. Hernandez; D. Le-Nguyen, *Curr. Protein Pept. Sci.* **2004**, *5*, 341–349.
110. A. Heitz; J. F. Hernandez; J. Gagnon; T. T. Hong; T. T. Pham; T. M. Nguyen; D. Le-Nguyen; L. Chiche, *Biochemistry* **2001**, *40*, 7973–7983.
111. J. F. Hernandez; J. Gagnon; L. Chiche; T. M. Nguyen; J. P. Andrieu; A. Heitz; T. Trinh Hong; T. T. Pham; D. Le Nguyen, *Biochemistry* **2000**, *39*, 5722–5730.
112. L. Gran, *Lloydia* **1973**, *36*, 209–210.
113. U. Goransson; T. Luijendijk; S. Johansson; L. Bohlin; P. Claeson, *J. Nat. Prod.* **1999**, *62*, 283–286.
114. M. R. Plan; U. Goransson; R. J. Clark; N. L. Daly; M. L. Colgrave; D. J. Craik, *ChemBioChem* **2007**, *8*, 1001–1011.
115. P. Seydel; H. Dörnenburg, *Plant Cell, Tissue Organ Cult.* **2006**, *85*, 247–255.
116. C. Jennings; J. West; C. Waine; D. Craik; M. Anderson, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10614–10619.
117. J. L. Dutton; R. F. Renda; C. Waine; R. J. Clark; N. L. Daly; C. V. Jennings; M. A. Anderson; D. J. Craik, *J. Biol. Chem.* **2004**, *279*, 46858–46867.
118. C. W. Gruber; M. Cemazar; R. J. Clark; T. Horibe; R. F. Renda; M. A. Anderson; D. J. Craik, *J. Biol. Chem.* **2007**, *282*, 20435–20446.
119. I. Saska; A. D. Gillon; N. Hatsugai; R. G. Dietzgen; I. Hara-Nishimura; M. A. Anderson; D. J. Craik, *J. Biol. Chem.* **2007**, *282*, 29721–29728.
120. A. D. Gillon; I. Saska; C. V. Jennings; R. F. Guarino; D. J. Craik; M. A. Anderson, *Plant J.* **2008**, *53*, 505–515.
121. J. P. Mulvenna; L. Sando; D. J. Craik, *Structure* **2005**, *13*, 691–701.
122. D. C. Ireland; M. L. Colgrave; P. Nguyencong; N. L. Daly; D. J. Craik, *J. Mol. Biol.* **2006**, *357*, 1522–1535.
123. K. R. Gustafson; L. K. Walton; R. C. Sowder, Jr.; D. G. Johnson; L. K. Pannell; J. H. Cardellina, Jr.; M. R. Boyd, *J. Nat. Prod.* **2000**, *63*, 176–178.
124. S. M. Simonsen; L. Sando; D. C. Ireland; M. L. Colgrave; R. Bharathi; U. Goransson; D. J. Craik, *Plant Cell* **2005**, *17*, 3176–3189.

125. A. M. Broussalis; U. Goransson; J. D. Coussio; G. Ferraro; V. Martino; P. Claeson, *Phytochemistry* **2001**, *58*, 47–51.
126. Y. F. Hallock; R. C. Sowder, II; L. K. Pannell; C. B. Hughes; D. G. Johnson; R. Gulakowski; J. H. Cardellina, II; M. R. Boyd, *J. Org. Chem.* **2000**, *65*, 124–128.
127. K. M. Witherup; M. J. Bogusky; P. S. Anderson; H. Ramjit; R. W. Ransom; T. Wood; M. Sardana, *J. Nat. Prod.* **1994**, *57*, 1619–1625.
128. U. Goransson; A. M. Broussalis; P. Claeson, *Anal. Biochem.* **2003**, *318*, 107–117.
129. B. Chen; M. L. Colgrave; N. L. Daly; K. J. Rosengren; K. R. Gustafson; D. J. Craik, *J. Biol. Chem.* **2005**, *280*, 22395–22405.
130. M. Trabi; E. Svangard; A. Herrmann; U. Goransson; P. Claeson; D. J. Craik; L. Bohlin, *J. Nat. Prod.* **2004**, *67*, 806–810.
131. B. Chen; M. L. Colgrave; C. Wang; D. J. Craik, *J. Nat. Prod.* **2006**, *69*, 23–28.
132. D. C. Ireland; M. L. Colgrave; D. J. Craik, *Biochem. J.* **2006**, *400*, 1–12.
133. C. K. Wang; M. L. Colgrave; K. R. Gustafson; D. C. Ireland; U. Goransson; D. J. Craik, *J. Nat. Prod.* **2008**, *71*, 47–52.
134. N. L. Daly; R. J. Clark; M. R. Plan; D. J. Craik, *Biochem. J.* **2006**, *393*, 619–626.
135. H. R. Bokesch; L. K. Pannell; P. K. Cochran; R. C. Sowder, II; T. C. McKee; M. R. Boyd, *J. Nat. Prod.* **2001**, *64*, 249–250.
136. E. Svangard; U. Goransson; Z. Hocaoglu; J. Gullbo; R. Larsson; P. Claeson; L. Bohlin, *J. Nat. Prod.* **2004**, *67*, 144–147.
137. P. Claeson; U. Goransson; S. Johansson; T. Luijendijk; L. Bohlin, *J. Nat. Prod.* **1998**, *61*, 77–81.
138. T. Schöpke; M. I. Hasan Agha; R. Kraft; A. Otto; K. Hiller, *Sci. Pharm.* **1993**, *61*, 145–153.
139. E. Svangard; U. Goransson; D. Smith; C. Verma; A. Backlund; L. Bohlin; P. Claeson, *Phytochemistry* **2003**, *64*, 135–142.
140. O. Saether; D. J. Craik; I. D. Campbell; K. Sletten; J. Juul; D. G. Norman, *Biochemistry* **1995**, *34*, 4147–4158.
141. L. Skjeldal; L. Gran; K. Sletten; B. F. Volkman, *Arch. Biochem. Biophys.* **2002**, *399*, 142–148.
142. L. Gran, *Acta Pharmacol. Toxicol.* **1973**, *33*, 400–408.
143. K. R. Gustafson; R. C. Sowder; L. E. Henderson; I. C. Parsons; Y. Kashman; J. H. Cardellina; J. B. McMahon; R. W. Buckheit; L. K. Pannell; M. R. Boyd, *J. Am. Chem. Soc.* **1994**, *116*, 9337–9338.
144. J. P. Tam; Y. A. Lu; J. L. Yang; K. W. Chiu, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8913–8918.
145. H. Kamimori; K. Hall; D. J. Craik; M. I. Aguilar, *Anal. Biochem.* **2005**, *337*, 149–153.
146. Z. O. Shenkarev; K. D. Nadezhdin; V. A. Sobol; A. G. Sobol; L. Skjeldal; A. S. Arseniev, *FEBS J.* **2006**, *273*, 2658–2672.
147. E. Svangard; R. Burman; S. Gunasekera; H. Lovborg; J. Gullbo; U. Goransson, *J. Nat. Prod.* **2007**, *70*, 643–647.
148. B. L. Barbeta; A. T. Marshall; A. D. Gillon; D. J. Craik; M. A. Anderson, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1221–1225.
149. K. P. Greenwood; N. L. Daly; D. L. Brown; J. L. Stow; D. J. Craik, *Int. J. Biochem. Cell Biol.* **2007**, *39*, 2252–2264.
150. S. Luckett; R. S. Garcia; J. J. Barker; A. V. Konarev; P. R. Shewry; A. R. Clarke; R. L. Brady, *J. Mol. Biol.* **1999**, *290*, 525–533.
151. M. L. Korsinczky; H. J. Schirra; D. J. Craik, *Curr. Protein Pept. Sci.* **2004**, *5*, 351–364.
152. A. H. Atkinson; R. L. Heath; R. J. Simpson; A. E. Clarke; M. A. Anderson, *Plant Cell* **1993**, *5*, 203–213.
153. H. J. Schirra; D. J. Craik, *Protein Pept. Lett.* **2005**, *12*, 421–431.
154. M. Laskowski, Jr.; I. Kato, *Annu. Rev. Biochem.* **1980**, *49*, 593–626.
155. P. Ascenzi; M. Ruoppolo; A. Amoresano; P. Pucci; R. Consonni; L. Zetta; S. Pascarella; F. Bortolotti; E. Menegatti, *Eur. J. Biochem.* **1999**, *261*, 275–284.
156. C. A. Ryan, *Annu. Rev. Phytopathol.* **1990**, *28*, 425–449.
157. S. K. Haq; S. M. Atif; R. H. Khan, *Arch. Biochem. Biophys.* **2004**, *431*, 145–159.
158. R. F. Qi; Z. W. Song; C. W. Chi, *Acta Biochim. Biophys. Sin.* **2005**, *37*, 283–292.
159. S. Odani; T. Ono, *J. Biochem.* **1980**, *88*, 1555–1558.
160. M. Catalano; L. Ragona; H. Molinari; A. Tava; L. Zetta, *Biochemistry* **2003**, *42*, 2836–2846.
161. M. H. Werner; D. E. Wemmer, *Biochemistry* **1992**, *31*, 999–1010.
162. M. L. Korsinczky; H. J. Schirra; K. J. Rosengren; J. West; B. A. Condie; L. Otvos; M. A. Anderson; D. J. Craik, *J. Mol. Biol.* **2001**, *311*, 579–591.
163. J. S. Graham; G. Pearce; J. Merryweather; K. Titani; L. H. Ericsson; C. A. Ryan, *J. Biol. Chem.* **1985**, *260*, 6561–6564.
164. J. S. Graham; G. Pearce; J. Merryweather; K. Titani; L. Ericsson; C. A. Ryan, *J. Biol. Chem.* **1985**, *260*, 6555–6560.
165. T. R. Green; C. A. Ryan, *Plant Physiol.* **1973**, *51*, 19–21.
166. T. R. Green; C. A. Ryan, *Science* **1972**, *175*, 776–777.
167. G. Gustafson; C. A. Ryan, *J. Biol. Chem.* **1976**, *251*, 7004–7010.
168. C. A. Ryan, *Plant Mol. Biol.* **1992**, *19*, 123–133.
169. C. A. Ryan; G. Pearce, *Annu. Rev. Cell Dev. Biol.* **1998**, *14*, 1–17.
170. G. Plunkett; D. F. Senear; G. Zuroske; C. A. Ryan, *Arch. Biochem. Biophys.* **1982**, *213*, 463–472.
171. S. E. Jackson; M. Moracci; N. elMasry; C. M. Johnson; A. R. Fersht, *Biochemistry* **1993**, *32*, 11259–11269.
172. J. Bryant; T. R. Green; T. Gurusaddaiah; C. A. Ryan, *Biochemistry* **1976**, *15*, 3418–3424.
173. E. Barta; A. Pintar; S. Pongor, *Trends Genet.* **2002**, *18*, 600–603.
174. M. J. Scanlon; M. C. Lee; M. A. Anderson; D. J. Craik, *Structure* **1999**, *7*, 793–802.
175. M. C. Lee; M. J. Scanlon; D. J. Craik; M. A. Anderson, *Nat. Struct. Biol.* **1999**, *6*, 526–530.
176. H. J. Schirra; M. J. Scanlon; M. C. Lee; M. A. Anderson; D. J. Craik, *J. Mol. Biol.* **2001**, *306*, 69–79.
177. N. Antcheva; A. Pintar; A. Patthy; A. Simoncsits; E. Barta; B. Tchorbanov; S. Pongor, *Protein Sci.* **2001**, *10*, 2280–2290.
178. N. Antcheva; A. Patthy; A. Athanasiadis; B. Tchorbanov; S. Zakhariev; S. Pongor, *Biochim. Biophys. Acta* **1996**, *1298*, 95–101.
179. R. L. Heath; P. A. Barton; R. J. Simpson; G. E. Reid; G. Lim; M. A. Anderson, *Eur. J. Biochem.* **1995**, *230*, 250–257.
180. K. J. Nielsen; R. L. Heath; M. A. Anderson; D. J. Craik, *Biochemistry* **1995**, *34*, 14304–14311.
181. C. A. McPhalen; M. N. James, *Biochemistry* **1988**, *27*, 6582–6598.
182. T. Cierpicki; J. Otlewski, *J. Mol. Biol.* **2000**, *302*, 1179–1192.
183. H. M. Greenblatt; C. A. Ryan; M. N. James, *J. Mol. Biol.* **1989**, *205*, 201–228.
184. M. Walker-Simmons; C. A. Ryan, *Plant Physiol.* **1977**, *59*, 437–439.
185. T. Wilusz; M. Wieczorek; A. Polanowski; A. Denton; J. Cook; M. Laskowski, Jr., *Hoppe Seylers Z. Physiol. Chem.* **1983**, *364*, 93–95.
186. W. Bode; H. J. Greyling; R. Huber; J. Otlewski; T. Wilusz, *FEBS Lett.* **1989**, *242*, 285–292.
187. T. A. Holak; D. Gondol; J. Otlewski; T. Wilusz, *J. Mol. Biol.* **1989**, *210*, 635–648.

188. R. L. Heath; G. McDonald; J. T. Christeller; M. Lee; K. Bateman; J. West; R. Van Heeswijck; M. A. Anderson, *J. Insect Physiol.* **1997**, *43*, 833–842.
189. D. A. Estell; K. A. Wilson; M. Laskowski, Jr., *Biochemistry* **1980**, *19*, 131–137.
190. R. M. Broadway; S. S. Duffey, *J. Insect Physiol.* **1986**, *32*, 827–833.
191. D. Bellincampi; L. Camardella; J. A. Delcour; V. Desseaux; R. D'Ovidio; A. Durand; G. Elliot; K. Gebruers; A. Giovane; N. Juge; J. F. Sorensen; B. Svensson; D. Vairo, *Biochim. Biophys. Acta* **2004**, *1696*, 265–274.
192. B. Svensson; K. Fukuda; P. K. Nielsen; B. C. Bonsager, *Biochim. Biophys. Acta* **2004**, *1696*, 145–156.
193. R. J. Weselake; A. W. Macgregor; R. D. Hill; H. W. Duckworth, *Plant Physiol.* **1983**, *73*, 1008–1012.
194. R. J. Weselake; A. W. Macgregor; R. D. Hill, *Plant Physiol.* **1983**, *72*, 809–812.
195. J. Mundy; I. Svendsen; J. Hejgaard, *Carlsberg Res. Commun.* **1983**, *48*, 81–90.
196. J. Mundy; J. Hejgaard; I. Svendsen, *FEBS Lett.* **1984**, *167*, 210–214.
197. K. Ohtsubo; M. Richardson, *FEBS Lett.* **1992**, *309*, 68–72.
198. A. Chagolla-Lopez; A. Blanco-Labra; A. Patthy; R. Sanchez; S. Pongor, *J. Biol. Chem.* **1994**, *269*, 23675–23680.
199. S. Lu; P. Deng; X. Liu; J. Luo; R. Han; X. Gu; S. Liang; X. Wang; F. Li; V. Lozanov; A. Patthy; S. Pongor, *J. Biol. Chem.* **1999**, *274*, 20473–20478.
200. J. C. Martins; M. Enassar; R. Willem; J. M. Wieruzeski; G. Lippens; S. J. Wodak, *Eur. J. Biochem.* **2001**, *268*, 2379–2389.
201. P. J. Pereira; V. Lozanov; A. Patthy; R. Huber; W. Bode; S. Pongor; S. Strobl, *Structure* **1999**, *7*, 1079–1088.
202. F. Payan, *Biochim. Biophys. Acta* **2004**, *1696*, 171–180.
203. M. Qian; V. Nahoum; J. Bonicel; H. Bischoff; B. Henrissat; F. Payan, *Biochemistry* **2001**, *40*, 7700–7709.
204. J. C. Uitdehaag; R. Mosi; K. H. Kalk; B. A. van der Veen; L. Dijkhuizen; S. G. Withers; B. W. Dijkstra, *Nat. Struct. Biol.* **1999**, *6*, 432–436.
205. E. H. Rydberg; C. Li; R. Maurus; C. M. Overall; G. D. Brayer; S. G. Withers, *Biochemistry* **2002**, *41*, 4492–4502.
206. W. F. Broekaert; W. Marien; F. R. Terras; M. F. De Bolle; P. Proost; J. Van Damme; L. Dillen; M. Claeys; S. B. Rees; J. Vanderleyden; B. P. A. Cammue, *Biochemistry* **1992**, *31*, 4308–4314.
207. J. C. Martins; D. Maes; R. Loris; H. A. Pepermans; L. Wyns; R. Willem; P. Verheyden, *J. Mol. Biol.* **1996**, *258*, 322–333.
208. M. Muraki, *Protein Pept. Lett.* **2002**, *9*, 195–209.
209. M. I. Chavez; C. Andreu; P. Vidal; N. Aboitiz; F. Freire; P. Groves; J. L. Asensio; G. Asensio; M. Muraki; F. J. Canada; J. Jimenez-Barbero, *Chemistry* **2005**, *11*, 7060–7074.
210. J. C. Kader, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 627–654.
211. A. Molina; A. Segura; F. Garcia-Olmedo, *FEBS Lett.* **1993**, *316*, 119–122.
212. F. R. Terras; I. J. Goderis; F. Van Leuven; J. Vanderleyden; B. P. Cammue; W. F. Broekaert, *Plant Physiol.* **1992**, *100*, 1055–1058.
213. M. H. Lerche; F. M. Poulsen, *Protein Sci.* **1998**, *7*, 2490–2498.
214. S. Gunasekera; N. L. Daly; M. A. Anderson; D. J. Craik, *IUBMB Life* **2006**, *58*, 515–524.
215. Y. L. Chan; V. Prasad; Sanjaya; K. H. Chen; P. C. Liu; M. T. Chan; C. P. Cheng, *Planta* **2005**, *221*, 386–393.
216. M. J. Carmona; A. Molina; J. A. Fernandez; J. J. Lopez-Fando; F. Garcia-Olmedo, *Plant J.* **1993**, *3*, 457–462.
217. F. De Leo; M. Bonade-Bottino; L. R. Ceci; R. Gallerani; L. Jouanin, *Insect Biochem. Mol. Biol.* **2001**, *31*, 593–602.
218. K. W. Yeh; M. I. Lin; S. J. Tuan; Y. M. Chen; C. J. Lin; S. S. Kao, *Plant Cell Rep.* **1997**, *16*, 696–699.
219. M. T. McManus; D. W. R. White; P. G. McGregor, *Transgenic Res.* **1994**, *3*, 50–58.
220. A. Lecardonnel; L. Chauvin; L. Jouanin; A. Beaujean; G. Prévost; B. Sangwan-Norreel, *Plant Sci.* **1999**, *140*, 71–79.
221. J. C. Leplé; M. Bonadé-Bottino; S. Augustin; G. Pilate; V. D. Lê Tân; A. Delplanque; D. Cornu; L. Jouanin, *Mol. Breed.* **1995**, *1*, 319–328.
222. L. Da-Hui; J. Gui-Liang; Z. Ying-Tao; A. Tie-Min, *Appl. Microbiol. Biotechnol.* **2007**, *74*, 146–151.
223. S. Olli; P. B. Kirti, *J. Biochem. Mol. Biol.* **2006**, *39*, 278–283.
224. J. L. Anaya-Lopez; J. E. Lopez-Meza; V. M. Baizabal-Aguirre; H. Cano-Camacho; A. Ochoa-Zarzosa, *Biotechnol. Lett.* **2006**, *28*, 1101–1108.
225. P. Flecker, *Eur. J. Biochem.* **1987**, *166*, 151–156.
226. G. Vogtentanz; K. D. Collier; M. Bodo; J. H. Chang; A. G. Day; D. A. Estell; B. C. Falcon; G. Ganshaw; A. S. Jarnagin; J. T. Kellis, Jr.; M. A. Kolkman; C. S. Lai; R. Meneses; J. V. Miller; H. de Nobel; S. Power; W. Weyler; D. L. Wong; B. F. Schmidt, *Protein Expr. Purif.* **2007**, *55*, 40–52.
227. J. A. Camarero; R. H. Kimura; Y. H. Woo; A. Shekhtman; J. Cantor, *ChemBioChem* **2007**, *8*, 1363–1366.
228. R. H. Kimura; A. T. Tran; J. A. Camarero, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 973–976.
229. R. H. Kimura; K. Krishnan; J. A. Camarero, *Biopolymers* **2005**, *80*, 537.
230. K. Thevissen; H. H. Kristensen; B. P. Thomma; B. P. Cammue; I. E. Francois, *Drug Discov. Today* **2007**, *12*, 966–971.
231. P. Lindholm; U. Goransson; S. Johansson; P. Claeson; J. Gullbo; R. Larsson; L. Bohlin; A. Backlund, *Mol. Cancer Ther.* **2002**, *1*, 365–369.
232. D. C. Ireland; C. K. Wang; J. A. Wilson; K. R. Gustafson; D. J. Craik, *Biopolymers* **2008**, *90*, 51–60.
233. D. J. Craik; R. J. Clark; N. L. Daly, *Expert Opin. Investig. Drugs* **2007**, *16*, 595–604.
234. U. Goransson; M. Sjogren; E. Svangard; P. Claeson; L. Bohlin, *J. Nat. Prod.* **2004**, *67*, 1287–1290.
235. P. Li; S. Jiang; S. L. Lee; C. Y. Lin; M. D. Johnson; R. B. Dickson; C. J. Michejda; P. P. Roller, *J. Med. Chem.* **2007**, *50*, 5976–5983.
236. K. Krishnamurthy; C. Balconi; J. E. Sherwood; M. J. Giroux, *Mol. Plant Microbe Interact.* **2001**, *14*, 1255–1260.
237. J. E. Blochet; C. Chevalier; E. Forest; E. Pebay-Peyroula; M. F. Gautier; P. Joudrier; M. Pezolet; D. Marion, *FEBS Lett.* **1993**, *329*, 336–340.
238. W. Jing; A. R. Demcoe; H. J. Vogel, *J. Bacteriol.* **2003**, *185*, 4938–4947.

Biographical Sketches

Jan-Christoph Westermann is a research officer at the Institute for Molecular Bioscience, The University of Queensland. He received his doctorate in chemistry from the Institute of Organic Chemistry at the University of Hamburg in 2006. He is working in the field of peptide structure–activity relationships and biomolecular NMR.

David J. Craik is an Australian Research Council Professorial Fellow at the Institute for Molecular Bioscience, The University of Queensland. He is on the editorial boards of the *Journal of Biological Chemistry*, *Current Medicinal Chemistry*, *The Handbook of Magnetic Resonance*, *Biopolymers: Peptide Science*, *Current Proteomics*, *Research Letters in Biochemistry*, and *Protein and Peptide Letters*. He is the editor of the book *NMR in Drug Design*. His contributions to medicinal chemistry and NMR have been recognized by the award of the Adrien Albert medal of the Royal Australian Chemical Institute, the ANZMAG medal of the Australian and New Zealand Society of Magnetic Resonance, and the Royal Australian Chemical Institute's H. G. Smith Medal. His research interests focus on the application of NMR in drug design and development, and on toxins. His group has a particular focus on structural studies of disulfide-rich proteins, and on the discovery and applications of circular proteins and novel protein topologies. He has trained more than 50 Ph.D. students and is the author of over 380 scientific publications.

5.10 Therapeutic Value of Peptides from Animal Venoms

Nicolas Andreotti, Besma Jouirou, Stephanie Mouhat, Ludovic Mouhat, and Jean-Marc Sabatier, Université de la Méditerranée – Ambrilia Biopharma Inc., Marseille Cedex, France

© 2010 Elsevier Ltd. All rights reserved.

5.10.1 Introduction

Animal toxins targeting ion channels (i.e., K^+, Na^+, Ca^{2+}, and Cl^-) of excitable cells are mainly short to medium size polypeptides (<10 kDa) that are found in venoms of a variety of species, such as snakes, spiders, scorpions, marine cone snails, sea anemones, worms, and insects. These toxins are largely employed as molecular probes to identify and characterize the various ion channel types and subtypes that differ in their structural properties, ion selectivity, and cellular function. Nowadays, the urgent need is to study in detail, analyze, and compare both the structural and functional characteristics of animal toxins because they provide some obvious research avenues in terms of protein/polypeptide engineering, biotechnological potential, and therapeutic applications for example, in the treatment of neurologic and cardiovascular diseases, cancer, diabetes, chronic pain, and autoimmune disorders. Compiling and collating the data from previous structure–function studies on animal toxins, it is clear that there is not any striking correlation between pharmacology, toxin fold, and disulfide bridge framework, indicating that the spatial distribution of key 'functional' amino acid residues is actually crucial to toxin pharmacology (**Figure 1**).[1] Such structural 'optimization' of toxins in order to improve the recognition of a particular ion channel can be achieved through slight or marked alterations in toxin structures, which would help to unravel the molecular basis of toxin–ion channel recognition and interaction, as well as to develop peptide drugs.

5.10.2 Peptides from Scorpion Venoms

About 1500 species of scorpions (Phylum Arthropoda, Class Arachnida, Order Scorpionida) are known (**Figure 2**). Species belonging to the taxonomic families Buthidae and Chactidae are the most venomous. Among the various nontoxic and toxic products contained in scorpion venoms are the salts, mucoproteins, lipids, nucleotides, histamine, glycoaminoglycans, 5-hydroxy-tryptamine or serotonin, biogenic amines, enzymes (e.g., ribonuclease, acetylcholinesterase, acid phosphatase, phospholipase A2), toxins targeting diverse ion channels, and other noncharacterized molecules. Many scorpion peptides, as well as their molecular targets, have been structurally and functionally studied in the last two decades. At present, it is estimated that approximately 400 polypeptides were isolated from scorpion venoms, and were either partly or 'fully' characterized. In parallel, serious efforts have been made to improve the knowledge on their molecular mechanisms of action, highlighting the formal existence of toxins acting as 'pore blockers' or 'gating modifiers.' Basically, scorpion venoms are rich and complex mixtures of products generally involved in defense against predators or immobilization of preys. The scorpion peptides characterized hitherto are relatively low-mass

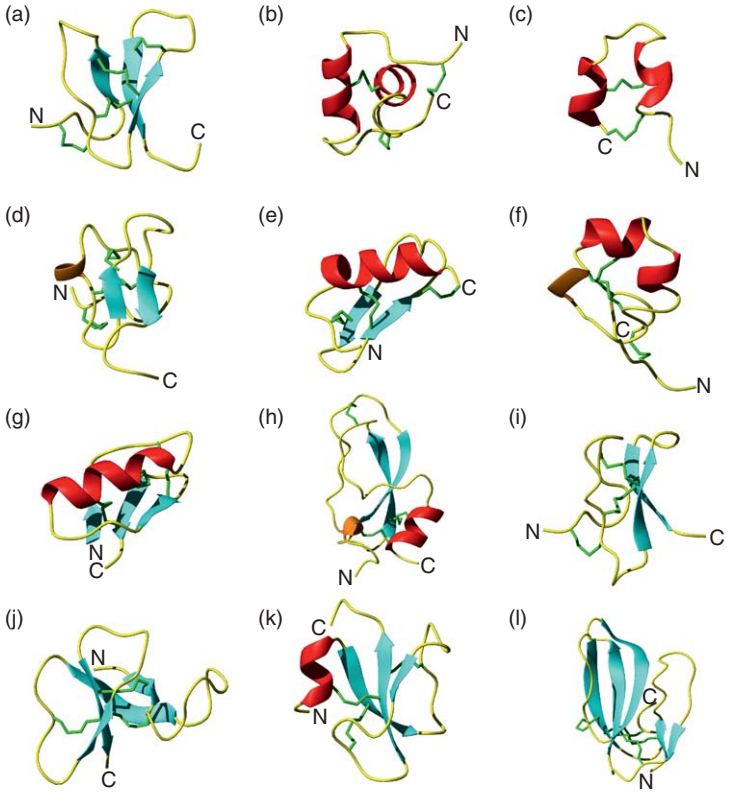


Figure 1 3-Dimensional structures of the most common animal toxin scaffolds. Different types of fold of toxins from various animal species are shown: (a) $\beta\beta\beta$; (b), and (c) two types of $\alpha\alpha$; (d) $3_{10}\beta\beta$; (e) $\alpha\beta\beta$; (f) $3_{10}\alpha\alpha$; (g) $\beta\alpha\beta\beta$; (h) $3_{10}\beta\beta\alpha$; (i) $\beta\beta$; (j) $\beta\beta\beta\beta$; (k) $\alpha\beta\beta\beta$, and (l) $\beta\beta\beta\beta\beta$. Helical structures are shown in red (α-helix) or orange (3_{10} helix), strands of β-sheet are blue, and disulfide bridges are green. The Cα peptide backbone trace is depicted in yellow. The toxin N- and C-terminal extremities are indicated. All of the 3-D structures of toxins presented originate from experimental data obtained in the original publications. The corresponding toxins are from (a) to (l) ACTX Hi:OB4219 (pdb code 1KQH), BgK (1BGK), κ-hefutoxin (1HP9), hanatoxin 1 (1D1H), maurotoxin (1TXM), ShK (1ROO), charybdotoxin (1CRD), dendrotoxin I (1DTX), huwentoxin-IV (1MB6), ATX Ia (1ATX), crotamine (1H5O), and FS2 (1TFS).

compounds (20–76 amino acid residues; <8 kDa) that are widely represented in animal venoms, with over 100 peptides evidenced by proteomic analyses. This number of peptides is actually underestimated mainly because of the rather 'limited' sensitivity of the experimental method used. Toxins from scorpions are classified into four groups (I–IV) based on their pharmacological targets: sodium (I), potassium (II), chloride (III), and calcium (IV) channels.[1] Most studied toxins belonged to groups I and II although there is recent interest in studying toxins from groups III and IV. The scorpion peptides are cross-linked between two and four disulfide bridges. Despite the great diversity of scorpion toxins, there are only three main characterized toxin folds, which are referred to as α helical hairpin-like motif,[2,3] α/β scaffold (or CS$\alpha\beta$ for 'cysteine-stabilized α-helix and β-sheet') and 'inhibitor cystine knot' (ICK) architectural motif.[4,5] The most common fold is the α/β scaffold. These types of fold contain between two and four well-defined elements of secondary structure: (1) $\alpha\alpha$ for α helical hairpin-like motif (two α-helices are arranged in an antiparallel manner with well-defined half-cystine pairings depending on the number of disulfide bridges),[2] (2) $\alpha\beta\beta$ and $\beta\alpha\beta\beta$ for α/β scaffold (an helix connected to a two- or three-stranded β-sheet structure),[6–8] and (3) $\beta\beta\beta$ for ICK motif (a ring of residues formed by disulfide bridges C1–C4 and C2–C5 through which the third disulfide bridge C3–C6 penetrates to form a cystine knot).[5,7,9] Such toxin architectures are very stable and reportedly associated with the presence of consensus amino acid sequences, such as XnCXnCX$_3$CXn(G/A/S)XCXnCXCXn (standard structural motif of α/β scaffold) and its variant XnCXnCX$_3$CXnCXn(G/A/S)XCXnCXCXn observed in short-chain scorpion toxins reticulated by four disulfide bridges (variant structural motif of α/β scaffold). An 'updated' consensus

Figure 2 Scorpions with venom peptides/toxins of potential therapeutic value. (a) Scorpion *Androctonus mauretanicus*; (b) scorpion *Androctonus crassicauda*; (c) scorpion *Buthus occitanus*; (d), scorpion *Leiurus quinquestriatus*; (e) pedipalpes with claws and facial close-up, and (f) telson with aculeus of scorpion *Centruroides vittatus*. Photos from (a) to (d), by Clocker (Creative Commons Attribution ShareAlike License), P.-A. Olsson (GNU free documentation license), F. Turmoq (Creative Commons Attribution ShareAlike License), E. Inbar (Creative Commons Attribution ShareAlike License) and (e), (f) by Sandax (Creative Commons Attribution ShareAlike License), respectively.

sequence of α/β scaffold has been proposed in 2004, that is, $CX_{2-5}CX_3CX_{5-11}CX_{4-5}CXC$.[1,7,10] In the case of ICK fold, the consensus sequence has evolved in the past years to finally give $CX_{3-7}CX_{3-8}CX_{0-7}CX_{1-6}CX_{4-13}C$.[5] In ICK fold, the side chains of functionally important amino acid residues are protruding to produce a variety of pharmacological activities by varying the inter-cystine residues of each loop. It is worth mentioning that one can anticipate the existence of other types of fold for scorpion peptides. Interestingly, it appears that peptides sharing the same type of fold can exert their action on several types and subtypes of ion channels.[11,12] Conversely, a particular ion channel can be the target of toxins that possess unrelated folds. The scorpion toxins mostly target ion channels, which are membrane proteins containing several transmembrane domains and a loop that forms the channel selectivity filter. The ion channels are tetrameric (K^+ channels) or monomeric (Na^+ and Ca^{2+} channels) structures. The monomeric structures of later channels contain four similar repeats that are equivalent to a monomer or subunit of K^+ channel. Shorter scorpion toxins (low-mass toxins of ~2–4 kDa) generally recognize K^+ channels and act as pore blockers by binding to the outer vestibule of the ion-conducting pathway (thereby blocking the ion flux through the channel).[13,14] Medium-size toxins (~4–6 kDa) usually act on Ca^{2+} or Cl^- channels, whereas large-size toxins (~7–8 kDa) alter Na^+ channel functioning by affecting its gating properties.[11,15,3] It should be noted that scorpion toxins are often not selective of a given ion channel; they can be pharmacologically active on several ion channel subtypes and/or types albeit with distinct potency. Scorpion toxins can also be used as probes to identify and characterize novel or specific types of ion channels, or as leads in the design of peptidomimetics. Since toxins clearly exhibit some therapeutic, pharmacological, and/or biotechnological potential depending on their targets, they can be derived to produce (by chemical synthesis or genetic engineering) structural analogs with the required characteristics in terms of selectivity, affinity, and stability. Finally, there is an increasing number of potential medical applications of ion channel-acting scorpion toxins

(and their bioactive derivatives) for the treatment or curing of cancer (e.g., colon, breast, prostate, lung, pancreas),[16] neural disorders (e.g., epilepsy, analgesia, apoplexy, Alzheimer's disease, paralysis),[17] autoimmune diseases (e.g., multiple sclerosis, type I diabetes),[18,19] and microbial (e.g., Gram-positive/Gram-negative bacteria and fungi) infections.[9,7,20] A few scorpion toxins, or derivatives, are under progress to be developed into drugs (e.g., chlorotoxin targeting chloride channels in glioma cancer therapy) by major pharmaceutical companies.

5.10.3 Peptides from Snake Venoms

The snakes (Phylum Chordata, Class Sauropsida, Order Squamata) currently represent approximately 2500 indexed species among which 400 are venomous. They are divided into four formal categories depending on their jaws, and subdivided into many families and subfamilies.[21] The most toxic or lethal families of snakes are Elapidae (e.g., cobras, coral snakes, and kraits) and Viperidae (vipers and pit vipers) (**Figure 3**). These snakes possess very potent venoms in their glands that are mixtures of compounds exhibiting enzymatic or nonenzymatic activities. Basically, snake venoms contain a variety of Ser proteases, organic compounds (e.g., purins),[22] phospholipase inhibitors, and toxins with different pharmacological actions (i.e., ion channel modulators (e.g., myotoxins), cardiotoxins, inhibitors of acetylcholinesterase or fasciculins). Among toxin effects hemolysis was described (e.g., disintegrins, sarafotoxins, natriuretic peptides, and CRISP toxins), anticoagulation, hypotension (e.g., inhibitors of voltage-dependent Ca^{2+} channels), inhibition of platelet aggregation (e.g., disintegrin peptides and C-type lectin-like proteins) and blockage of neurotransmission (e.g., toxins acting on voltage-gated Na^{+} (Nav) or voltage-gated K^{+} (Kv) channels, muscarinic toxins, waglerins, and

Figure 3 Elapidae and Viperidae snakes. (a) Western diamondback rattlesnake (*Crotalus atrox*); (b) western green mamba (*Dendroaspis viridis*); (c) west African gaboon viper (*Bitis gabonica* (Public domain)); (d) desert horned viper (*Cerastes cerastes* (http://www.pythonsnake.com)); (e) eastern diamondback rattlesnake fang with venom drop (*Crotalus adamantus* (http://www.Animalpicturesarchive.com)); (f) Mozambique spitting cobra projecting venom (*Naja nigricollis* (http://www.figtree.squarespace.com)). Photos by G. Stolz (Public domain) (a) and P. Coin (Creative Commons Attribution ShareAlike License) (b). See websites for photo credits (c–f).

bradykinin-potentiating peptides).[23] Therefore, the nature of products characterized from snake venom is coherent with the selected strategies of prey immobilization (i.e., hypotension, paralysis, and digestion). It is worth mentioning that, apart from noncharacterized compounds, the precise functional roles of some 'characterized' molecules present in venom are actually poorly understood and remain to be addressed. Examples are provided with the angiogenesis-stimulating vascular endothelial growth factors (VEGFs) (from snake species Vipera and Bothrops) whose exact functions are unknown although it has been suggested that they might enhance venom distribution by increasing vascular permeability after a snake bite.

The structural features of snake toxins allow to distinguish between three main groups: (1) 'three-finger' toxins (e.g., voltage-gated Ca^{2+} (Cav1) channel blockers such as calciseptine),[24,25] (2) peptides homologous to Kunitz Ser protease inhibitors (Cav1 channel modulators such as calcicludine,[26,27] and Kv1.X channel blockers such as dendrotoxins[28–30]), and (3) myotoxins of the crotamine type (Nav channel modulators such as crotamine).[31] It is noteworthy that, snake Kv1.3 channel-acting Natrin,[32] which belongs to cysteine-rich secretory protein (CRISP) toxins, possesses a cysteine-rich domain resembling the three-dimensional structures of two sea anemone toxins, ShK and BgK, both acting on the same ion channel subtype. The 'three-finger' toxins, which are widely represented in venoms of mambas, kraits, cobras, and sea snakes, are folded according to the β-sheet structures with loops. These toxins are reticulated by four or five disulfide bridges, with four of them being conserved in this structural group. Therefore, 'three-finger' toxins exhibit three β-stranded loops, which extend from a central core containing the four conserved disulfide bridges. It is worthy of note that the 'three-finger' architectural motif is not limited to elapid or hydrophid toxins. Snake toxins from other structural groups are, independent of their pharmacological target(s), folded by more complex combinations of β-sheets and α and/or 3_{10} helices.[1]

At the level of medical applications, there are several snake venom peptides or derivatives (not active on ion channels) that are candidate drugs or were actually developed into effective drugs in humans. These products are especially invaluable in the treatment of cardiovascular diseases, such as the following anticoagulants or thrombolytic compounds[33]: integrilin/barbourin (acute coronary syndrome and angioplasty), captopril (hypertension, renal syndromes such as scleroderma and diabetic nephropathy, congestive heart failure), echistatin/aggrastat, ancrod/viprinex (acute ischemic stroke), crotavirin (infectious endocarditis), fibrolase (peripheral arterial occlusions), natriuretic-like peptides (congestive heart failure), and dendoaspin/mambin. Other snake peptides with antitumor activities might be developed as potential chemotherapeutic agents in oncology,[34,35] such as contortrostatin (prevention of metastasis) and jerdonin. Finally, a therapeutic potential in analgesia has also been reported for some snake toxins/peptides with strong analgesic properties,[36] such as cobrotoxin, crotamine, and hannalgesin.

5.10.4 Peptides from Sea Anemone Venoms

The sea anemones (Phylum Cnidaria, Class Anthozoa, Order Actiniaria) are toxic species that are not lethal in humans (**Figure 4**). They possess numerous tentacles containing cnidocytes, which are specialized stinging cells involved in defense against predators, intraspecific aggression, and capture of prey (fish and crustaceans). The cnidocytes are equipped with capsule-like organelles that are referred to as nematocysts. These are capable of everting upon either chemical or mechanical stimulations. The venom is basically a complex mixture of compounds with various functions and pharmacological activities,[37,38] such as protease inhibitors (including Kunitz and cysteine proteinase inhibitors), neurotransmitters or neuromodulators, phospholipases A2, epidermal growth factor (EGF)-like peptides (gigantoxin), actinoporins or cytolysins (16–20 kDa), and finally ion channel modulators.[39] The cytolysins/actinoporins are involved in pore formation in the membrane lipid bilayer leading to osmotic imbalance promoting cell lysis,[40] whereas ion channel-acting toxins target voltage-gated and Ca^{2+}-activated K^+ channels (K^+ channel blockers of 5 kDa) as well as H^+-dependent and voltage-gated Na^+ channels (Na^+ channel modulators).[41] The sea anemone toxins acting on K^+ channels are classified into two categories depending on their amino acid sequence identities. In the first category, toxins (e.g., ShK,[42] HmK,[43] BgK,[44] kaliseptine[45]) are folded according to combined helices of α and/or 3_{10} type(s). In the second category, toxins (e.g., BDS-I and II blood depressing substances,[46] APETx1[47]) are folded with an arrangement of β-sheets (β-defensin type). As for other venomous animal species, a number of characterized

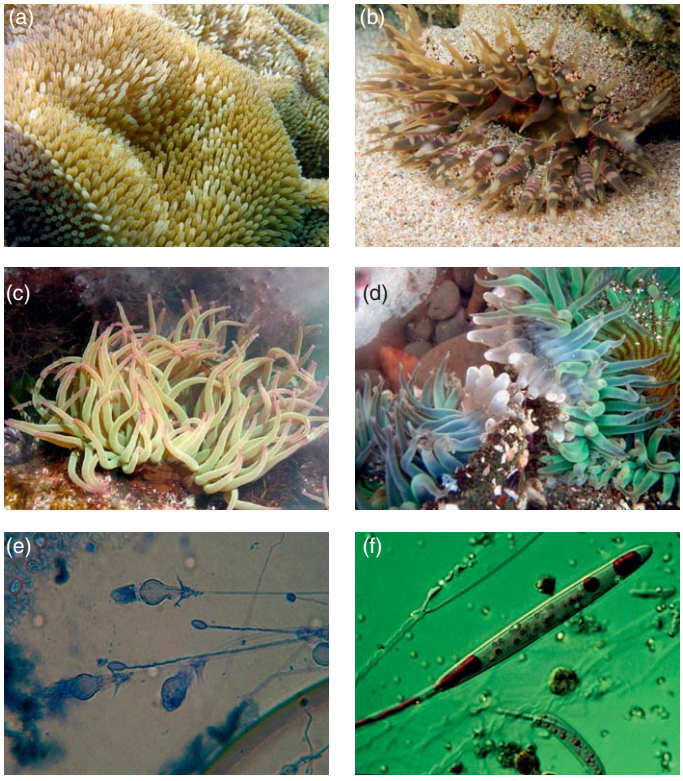


Figure 4 Venomous sea anemones. (a) Sun anemone (*Stichodactyla helianthus*); (b) *Bunodosoma granulifera*; (c) snakelocks anemone (*Anemonia viridis*); (d) starburst anemone (*Anthopleura sola*); (e) discharged nematocysts; (f) a Nomarski micrograph of a ruthenium red-stained nematocyst from *Aiptasia pallid*. The red dye stains the polyanionic venom proteins found inside the partially discharged nematocyst. Photos from (a) to (f), by F. Charpin (http://www.Florent.us), Pline (Creative Commons Attribution ShareAlike License), M. Zinkova (Creative Commons Attribution ShareAlike License), J. Engman (The cnidarian lab), and D. Brand (Public domain), respectively.

sea anemone toxins (e.g., ShK and BgK) are of therapeutic value as immunomodulators (immunosuppressants to treat autoimmune diseases).[48,49] These toxins block the voltage-gated Kv1.3 channel, which is critical for the activation of effector memory T cells, and a targeted channel in the prevention of graft rejection and treatment of (T-cell-mediated) autoimmune diseases such as multiple sclerosis.[18]

5.10.5 Peptides from Spider Venoms

The spiders (Phylum Arthropoda, Class Arachnida, Order Araneae) comprise approximately 40 000 characterized species. They are classified into two families (or infraorders) according to the position of their chelicerae: Araneomorphae and Mygalomorphae (**Figure 5**). The spider venoms are particularly complex mixtures of biologically active and biologically inactive molecules, among which were found some nucleic acids, polyamine toxins, free amino acids, monoamines, inorganic salts, proteins, peptides, and enzymes of various types (protease, phospholipase A, ATPase, collagenase, peptide isomerase, esterase, phosphodiesterase, alkaline phosphatase, hyaluronidase).[50,51] The venom peptides are rather short (with a size range of 4–10 kDa) and are cross-linked by three to seven disulfide bridges.[52,53] Post-translationally modified products were also described (e.g., O-palmitoylated PLTX-II) adding to the intrinsic complexity of spider peptides. The spider peptides/toxins modulate neurotransmission, in both vertebrates and invertebrates, by acting on various receptors, transporters, and ion channels.[54,55] The great diversity of polypeptides in spider venoms suggests they have actually developed combinatorial libraries of peptides with various activities (e.g., antimicrobial, cytolytic, enzymatic, neurotoxic).

Figure 5 Venomous spiders. (a) Mygalomorphae-type chelicerae from Goliath bird-eating spider (*Theraphosa blondi*); (b) Araneomorphae-type chelicerae from yellow sac spider (*Cheiracanthium punctorium*); (c) Venezuelan suntiger (*Psalmopoeus irminia*); (d) black widow (*Latrodectus hesperus*); (e) red-knee tarantula (*Brachypelma smithi*); (f) mouse spider (*Missulena bradleyi*). Photos from (a) to (f) by J. Smith (Creative Commons Attribution ShareAlike License), R. Altenkamp (Creative Commons Attribution ShareAlike License), B. Smith (Creative Commons Attribution ShareAlike License), Fir0002 (GNU free documentation license), B. Smith (Creative Commons Attribution ShareAlike License) and Fir0002 (GNU free documentation license), respectively.

Because the functional role of venom is to neutralize prey or predators, spider venoms are naturally rich sources of (1) ion channel-acting toxins (neurotoxins) affecting particular ion conductance, (2) presynaptic toxins affecting neurotransmitter release/exocytosis, and (3) postsynaptic toxins altering binding of neurotransmitters to their cellular targets.[56] The peptide activity could be strictly selective for insects[57] or vertebrates such as humans. Most studies were originally focused on spider toxins acting on voltage-gated Na^+ (gating modifiers and blockers) and Ca^{2+} channels. Interestingly, it was found that some Na^+ channel-acting spider toxins (e.g., δ-atracotoxin-Hv1) were able to compete with the binding of scorpion α-toxins or sea anemone toxins to site 3 of the Na^+ channel. In the case of voltage-gated K^+ channels (Kv), some spider toxins are of particular interest because they behave as the sole peptide blockers of the Kv2.x and Kv4.x channel subfamilies characterized so far. The well-studied toxins are heteroscodratoxins (HmTx) 1 and 2 acting respectively on Kv4.x and Kv2.x subtypes,[58] hanatoxins (HaTx) 1 and 2 targeting Kv2.1 channels,[59] heteropodatoxins (HpTx) 1 to 3 and phrixotoxins (PaTx) 1 and 2 acting on Kv4.x channels.[60,61] Poorly specific toxins were identified, such as stromatoxin 1 (ScTx1) acting on both Kv2.2 and Kv4.2 subtypes, as well as jingzhaotoxin III and protoxin 1 acting respectively on two (i.e., $Na_v1.5$ and Kv1.2)[62] and three channel types (i.e., $Na_v1.5$, Kv2.1, and $Ca_v3.1$).[63] At the structural level, spider toxins are generally folded according to the ICK (with or without a short 3_{10} helix) or 'disulfide-directed β-hairpin' (DDH) architectural motif that involves the formation of β-sheet structures.[3] The $\beta\beta$ (e.g., huwentoxin-IV), $\beta\beta\beta$ (e.g., ACTX-Hi:OB4219), and helical-ended $\beta\beta\beta3_{10}$ (e.g., δ-atracotoxin-Hv1) types of fold are three variants of the ICK motif. The DDH motif consists of an antiparallel β-sheet, which is stabilized by two disulfide bridges but lacks the cystine knot. The consensus sequence thereof is: $CX_{5–19}CX_2G/PX_2CX_{6–19}C$. To date, the main potential applications of spider toxins/peptides in therapy are for treating cardiovascular diseases (e.g., tarantula GsMtx-4 inhibiting

mechanosensitive channels for suppression of atrial fibrillation), microbial infection (e.g., oxyopinins or lycotoxins I and II), arrhythmy (e.g., GsMtx-4 that blocks MSC in ventricular myocytes)[64] and pain (e.g., psalmotoxin for acid-induced pain).[65] However, other medical applications might exist due to the variety of biological actions of spider toxins/peptides, that is, modulators of Kv, Na_v, Ca_v (P/Q-, R-, L-, and N-types), acid-sensing (ASIC) and mechanosensitive (MSC) ion channels, effects on neurotransmitter release (presynaptic toxins), transporters, glutamate receptors, and so on.[50] To illustrate this point, H^+-gated Na^+ channels (ASIC) would have some crucial roles in nociception, taste transduction, synaptic plasticity, learning, and memory processes to cite a few, whereas MSC are thought to be important in transducing a number of internal (e.g., local control of blood flow, dilation-induced heart rate changes and regulation of cell volume) and external (e.g., touch, sound, and vibration) stimuli. Apart from those potential medical applications of spider compounds in humans, it should be noted that spiders, especially mygalomorphs, contain many insect-selective toxins that would be 'leads' for developing novel biopesticides. Generally, these toxins are phyla-specific and target voltage-gated ion channels which are ubiquitous among insects. The molecular basis of toxin specificity (insect vs. mammal ion channels) is still poorly understood and need to be clarified for an effective development of insecticidal toxins (or derivatives). The later would help to produce nonpeptide mimetics that opened the way to foliar sprays.

5.10.6 Peptides from Cone Snail Venoms

The cone snails from all marine environments (Phylum Mollusca, Class Gastropoda, Order Sorbeoconcha) represent a large genus of approximately 700 carnivorous predator species (**Figure 6**).[66,67] They are classified into three groups, referred to as molluscivore, vermivore, and piscivore.[68] To date, evidence of over 50 000

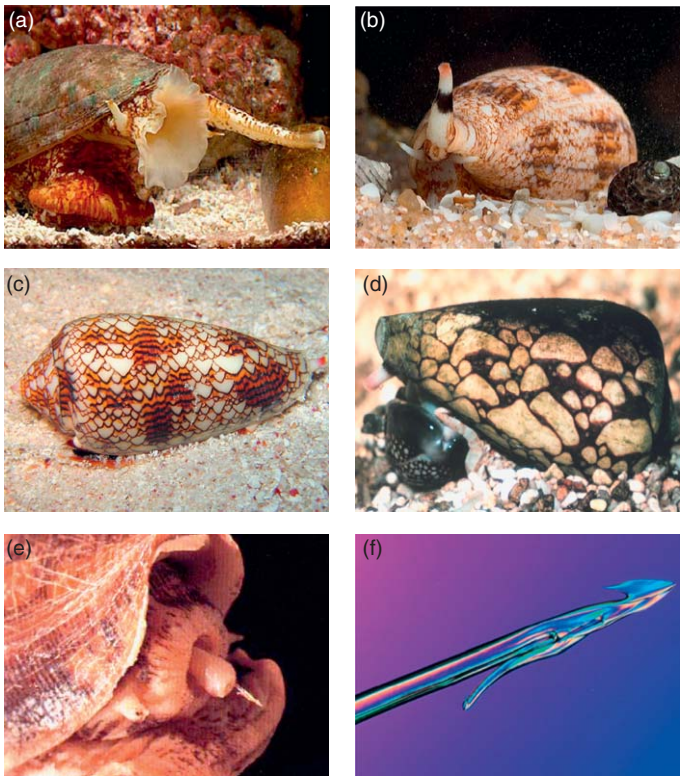


Figure 6 Venomous marine cone snails. (a) Geography cone (*Conus geographus* (Public domain)); (b) Queen Victoria cone (*Conus victoriae*); (c) cloth of gold cone (*Conus textile*); (d) marble cone (*Conus marmoreus* (Public domain)); (e) closing view of proboscis with toxoglossan cone snail; (f) microscope picture of toxoglossan from *Conus consors*. Photos (b), (c), (e), (f) by D. Paul (www:biochemistry.unimelb.edu.au), R. Ling (GNU free documentation license), Rudy (Public domain), and J.-J. Soin (http://www.Conco.eu), respectively. See websites for photo credits (a, d).

different peptides have been gathered from these groups, but only a minority have been biochemically and functionally characterized. The vast majority of peptides (several thousands) is expected to be short, highly rigid, disulfide-bridged 10- to 20-mer products. They target specific enzymes, neurotransmitter transporters (e.g., noradrenaline/norepinephrine transporter targeted by χ-conopeptides), G-protein-coupled membrane receptors, voltage-gated ion channels (Na^+, Ca^{2+}, and K^+ channels) and ligand-activated ion channels (nicotinic acetylcholine receptor of both neuronal and neuromuscular subtypes, $5HT_3$ serotonin receptor, glutamate-type receptor such as *N*-methyl-D-aspartate/NMDA receptor). Interestingly, it has been highlighted that several venom components act simultaneously to paralyze the prey and alter neuromuscular transmission.[69] Indeed, these molecules act in concert on targeted presynaptic Ca^{2+} channels,[70] postsynaptic nicotinic receptors, and voltage-gated Na_v channels to 'maximize' venom toxic effects. These effective combinations of cone snail peptides/toxins are known as 'toxin cabals.' The combination of products that interferes with neuromuscular transmission is named 'motor cabal.' The ion channel-acting peptides/toxins are now well-studied, low-molecular mass, molecules reticulated by several disulfide bridges. The cone snail peptides (conopeptides) can be 'structurally' distinguished from other venomous animal peptides because they often possess unusual D-amino acid residue(s) (e.g., D-Val, D-Phe, D-Trp, and D-Leu), and/or exhibit some post-translational modifications, such as hydroxylation (i.e., D-γ-hydroxyvaline, 5-hydroxylysine, and 4-*trans*-hydroxyproline), N-terminal cyclization of Gln residue, C-terminal amidation, O-glycosylation of Ser and/or Thr residue(s), bromination of Trp residue(s), O-sulfation of Tyr residue(s), γ-carboxylation of Glu residue(s), and so forth.[71,72] The presence of such 'unusual' or 'modified' residues in the peptide structures is considered an hallmark of cone snail peptides/toxins. The conopeptides are synthesized as 70- to 120-residue pre-propeptide precursors from genes expressed in epithelial cells of cone snail venom ducts. The precursors are composed of: (1) a pre-region which is well-conserved among members of the same superfamily, (2) a more or less conserved pro-region, and (3) the conopeptide amino acid sequence itself. The conopeptides are organized in superfamilies depending on signature signal sequences of precursors, as well as families depending on disulfide bridges frameworks and pharmacology.[73,74] The conotoxins formally refer to as cone snail peptide toxins cross-linked by two or more disulfide bridges, and acting on voltage-gated ion channels (K^+, Ca^{2+}, and Na^+ channels). Through a number of structural analyses achieved by ^{1}H-NMR in solution, conotoxins were generally found to be folded in pleated β-sheets, with an ICK architectural motif. The conotoxins are categorized on the basis of their particular patterns of half-cystine pairings, where CC are two adjacent half-cystine residues (not necessarily connected with each other by a disulfide bridge), and where C–C are two half-cystine residues separated by one or more amino acid residues (i.e., '-' equals 'Xn'): (1) the four cysteine/two-loop framework CC–C–C (gene superfamily A) observed in α-conotoxins active on neuromuscular and neuronal nicotinic acetylcholine receptors,[75,76] (2) the six cysteine/three-loop framework CC–C–C–CC (gene superfamily M) found in Na^+ channel-acting μ-conotoxins,[77] neuromuscular nicotinic acetylcholine receptor-acting ψ-conotoxins and K^+ channel-acting κM-conotoxins, and (3) the six cysteine/four-loop framework with its two variants: C–C–CC–C–C (gene superfamily O) found in δ-/μO-, ω-, and κ-conotoxins respectively targeting Na^+, Ca^{2+}, and K^+ channels, and CC–C–C–C–C (gene superfamily A) found in Na^+ channel-acting μ-PnIVA/B. Among all frameworks, the more abundant is: C–C–CC–C–C. Other cysteine frameworks were described for conotoxins, that is, CC–CC and CC–CPC, C–C–C–C, C–C–C–C–C–C, C–C–C–C–C–C–C–C–C, C–C–CC–CC–C–C, C–C–C–C–CC–C–C, C–C–C–CC–C–C–C, C–C–CC–C–C–C, and C–C–CC–C–CC–C, corresponding respectively to gene superfamilies T, J/L, P, S, I, TxX, De13a, It15.1, vi15a, and ca16.a.[78] Similar to peptides/toxins from other venomous animal species, the conotoxins are of diagnostic and therapeutic value, one being approved by the Food and Drug Administration (FDA) in 2004 as an antinociceptive drug (i.e., ω-MVIIA conotoxin (Cav2.2 inhibitor) named as Prialt/Ziconotide, and originating from *Conus magus*, cone snail venom) for treating severe chronic pain.[79,80] Additional cone snail peptides have been found interesting in treating pain: (1) a glycosylated neurotensin analog named contulakin-G (neurotensin receptor agonist), which exhibits potent analgesic properties (~100-fold more potent than neurotensin as an analgesic) in several animal models, (2) conantokin-G and -T, with specific antagonist activity of the *N*-methyl-D-aspartate (NMDA) receptor, and (3) χ-conopeptides (gene superfamily T) acting as noradrenaline transporter inhibitors. By inhibiting noradrenaline transporter, the later might have antidepressant and/or psychostimulant effects, influencing learning and memory processes. Also, χ-conopeptides might be useful to treat cardiovascular disorders and urinary incontinence. Finally, ρ-conopeptides that act as selective noncompetitive inhibitors of α_1-adrenoceptor might be developed as candidate drugs for hypertension.[81]

5.10.7 Peptides from Insect Venoms

The bees, wasps, and ants are venomous hymenoptera insects (Phylum Arthropoda, Class Insecta, Order Hymenoptera) with about 120 000 known species out of an estimate of 1–3 million hymenoptera species (**Figure 7**). They are divided into two groups: social and solitary. The venoms of social hymenoptera are mainly used in defense to protect parent insect colonies from natural predators. Although generally not lethal, these venoms caused some inflammatory and/or immunological responses upon injections (stings) in preys. Reported effects/symptoms are edema, pain, bradycardia or tachycardia, swelling, and headache. Kidney and/or respiratory failure(s) is (are) also common systemic effect(s). The social hymenoptera peptides (<7 kDa) represent from 60% to 70% of lyophilized venom; they are often highly positively charged (basic) cytolytic compounds of amphipathic nature, exhibiting a high content of helical secondary structures.[82] Additionally, toxins targeting ion channels (Na^+ and Ca^{2+} channels) and nicotinic acetyl choline receptors have been evidenced in the defensive or offensive chemical weaponry, together with polyamines and a variety of neurotransmitters. Apart from proteins and peptide toxins, other important compounds of venoms from the social bees and wasps are enzymes (damage to tissues) and low-molecular mass organic products. The venoms of solitary hymenoptera induce – via action of specific molecules – paralysis of preys (spiders and insects) allowing egg laying inside the prey body. Additional venom components act to prevent infections of food and progenies. To date, most studied hymenoptera peptides/toxins are from the social bees (e.g., apamin,[83] mast-cell degranulating (MCD) peptide,[84,85] melittin,[86,87] tertiapin,[88] secapin,[84] and so on), social wasps (e.g., mastoparans, chemotactic and kinin-like peptides, crabrolin, sylverin)[89–91], solitary wasp (e.g., bradykinin-like peptides,[92] anoplin,[93] pompilidotoxins,[94–96] eumenine mastoparan-AF,[97–99]), and ant (e.g., poneratoxins,[100] ponericins,[101,102] ectatomin,[103]

Figure 7 Venomous hymenoptera insects. (a) Common honeybee (*Apis mellifera*); (b), eastern yellowjacket (*Vespula maculifrons*); (c) European hornet (*Vespa crabro*); (d) bull ant (*Myrmecia esuriens*); (e) Asian giant hornet (*Vespa mandarinia japonica*,); (f) wasp stinger. Photos from (a) to (f) by Autan (Creative Commons Attribution ShareAlike License), E. Begin (Creative Commons Attribution ShareAlike License), N. Jones (Creative Commons Attribution ShareAlike License), Nuytsia (Creative Commons Attribution ShareAlike License), Netman (Creative Commons Attribution ShareAlike License), and M. Halldin (GNU free documentation license), respectively.

pilosulins[104]) venoms. It is worth mentioning that ants generally possess only trace amounts of venomous polypeptides. The later are toxins and/or short basic peptides with cytolytic, hemolytic, or antimicrobial properties. In some cases, they could act on mast cells to induce histamine release. As for other venomous animal peptides, a number of potential therapeutic applications exists for a few more or less well-characterized molecules (not targeting ion channels), such as ant pilosulins (fungal and bacterial infections), wasp protonectin or agelaia or anoplin, hornet cabrolin, silk moth cecropins and fly drosocin (bacterial infections), bee melittin (rheumatoid arthritis and fungal, viral, bacterial, and protozoal infections), and ant myrmexins (inflammatory). Of note, the three-dimensional structures of these compounds are generally unknown, except a few (e.g., cecropin, melittin, mastoparan-X).

5.10.8 Peptides from Worm Venoms

The carnivorous marine ribbon worms or nemertines (Phylum Nemertea or Phylum Rhyncocoela) are predators of snails, small crustaceans, mollusks, and polychaetes. About 1000 species of nemertines are described hitherto; they are classified into two groups, that is, the hoplonemertines (possess a proboscis piercing stylet to inject venom in order to capture or subdue their preys), and anoplans (heteronemertines and paleonemertines) lacking the proboscis apparatus.[105] The hoplonemertine toxins are used in chemical defense (against predators) or offense (against preys), whereas anoplan toxins are mainly employed to repel predators. The nemertines produce a variety of organic (alkaloid) and peptide toxins, among which a few are characterized. The anoplan heteronemertines (unarmed nemerteans) secrete low-molecular mass basic peptide toxins targeting ion channels involved in the generation of action potentials, as well as peptide cytolysins (not found so far in hoplonemertines).[106] Other parasitic worms such as nematodes and helminths use peptides with cytolytic properties (pore forming or detergent-like peptides) to digest host cells/tissues, whereas some predatory marine annelids show protein toxins (e.g., glycerotoxin) targeting particular synaptic Ca^{2+} channels (N-type Cav channels) that stimulate neurotransmitter release at amphibian – but not mammalian – neuromuscular synapses.[107] Because invertebrate toxins generally target invertebrate receptors (since they deal with invertebrate predators and preys), there is presumably little (if any) therapeutic value of such molecules in humans. However, by acting selectively on invertebrate nervous system, worm peptides and toxins provide clearly invaluable research tools and probes to better understand the molecular determinants that are involved in selectivity toward animal species (invertebrate vs. vertebrate). Additionally, these compounds behave as potential 'leads' or structural templates in the development of highly selective antiparasitic drugs and pesticides.

5.10.9 Other Venom Peptides and Toxins of Interest

Basically, several other animal species (**Figure 8**) possess in their venoms a number of peptides and toxins with chemotherapeutic potential as candidate anticoagulant, thrombolytic, immunomodulatory, anti-infectious, anticancer, antihypertensive, antiarrhythmic, antidiabetes, and analgesic compounds, to cite the main clinical applications.[108] In the potential treatment of cardiovascular diseases, candidate venom peptides/toxins would be bat desmoteplase that activates plasminogen (acute ischemic stroke), medicinal leech hirudin targeting thrombin (heparin-induced thrombocytopenia), frog ranatensin or margaratensin exhibiting neurotensin-like activity (hypertension). In the case of microbial infections, they would be frog magainins,[109] dermaseptins,[110–112] esculentin-1 (antifungal) and -2,[113] brevinin-1 and -2,[114] tigerinins,[115] temporin L,[116] japonicin-1,[117] palustrin-3,[118] and ranalexin (antiparasite),[119] as well as solitary tunicate halocidin[120] and dicynthaurin.[121] A candidate immunomodulator is frog ranamargarin (anti-inflammatory) that targets tachykinin.[122] Among antitumor peptides frog bombesin[123] and its analogs, and mollusk kahalalides (prostate cancer) are some of them.[124,125] The potential analgesics are frog ceruletide and caerulein (antinociceptive and anticonvulsive),[126] apart from frog alkaloid epibatidine. In the treatment of diabetes mellitus type-2, frog skin insulinotropic peptide (FSIP)[127] and lizard exendin-3 and -4[128] have been found of particular interest. Notably, the synthetic form of exendin-4, named exenatide, possesses potent antidiabetic and antiobesity activities by acting on glucagon-like peptide 1 (GLP-1) receptors.[129,130] This peptide was approved by the FDA in 2005.

Figure 8 Venomous animals with chemotherapeutic potential. (a) Vampire bat (*Desmodus rotundus* (http://www.Animalpicturesarchive.com)); (b) medicinal leech (*Hirudo medicinalis*); (c) oriental fire-bellied toad (*Bombina orientalis*); (d) phantasmal poison frog (*Epipedobates tricolor*); (e) solitary tunicate sea peach (*Halocynthia aurantium*); (f) gila monster (*Heloderma suspectum*). Photos (b)–(f) by I. Boyd (Creative Commons Attribution ShareAlike License), H. Yan (Public domain), L. Ghoul (Creative Commons Attribution ShareAlike License), A. Rode (Creative Commons Attribution ShareAlike License) and Arpingstone (Public domain), respectively. See website for photo credit (a).

Abbreviations

ω-MVIIA	omega-conotoxin MVII A from marine cone snail *Conus magnus*
^{1}H-NMR	proton-nuclear magnetic resonance
A (Ala)	alanine or alanyl
ASIC	acid-sensing ion channels
ATPase	adenosine triphosphatase
BDS I/II	blood depressive substance I/II
BgK	K^+ channel-acting toxin from sea anemone *Bunodosoma granulifera*
C (Cys)	cysteine (reduced form) or half-cystine residue (oxidized form)
Ca_v	voltage-gated Ca^{2+} channel
DDH	disulfide-directed β-hairpin
FDA	Food and Drug Administration
FSIP	frog skin insulinotropic peptide from frog *Agalychnis litodryas*
G (Gly)	glycine or glycyl
GLP-1	glucagon-like peptide 1
GsMtx	mechanosensitive toxin from spider *Grammostola spatulata*
HaTx	hanatoxin from spider *Grammostola rosea*
HmK	K^+ channel-acting toxin from spider *Heteractis magnifica*

HmTx	heteroscodratoxin from spider *Heteroscodra maculata*
HpTx	heteropodatoxin from spider *Heteropoda venatoria*
ICK	inhibitor cystine knot
kDa	kilodalton
K_v	voltage-gated K^+ channel
MCD	mast-cell degranulating peptide from bee *Apis mellifera*
MSC	mechanosensitive ion channel
MTX	maurotoxin from scorpion *Scorpio maurus*
Na_v	voltage-gated Na^+ channel
NMDA	*N*-methyl-D-aspartate
PaTx	phrixotoxin from spider *Phrixotrichus auratus*
PLTX-II	toxin II from spider *Plectreurys tristes*
S (Ser)	serine or seryl
ScTx1	stromatoxin from spider *Stromatopelma calceata*
ShK	K^+ channel-acting toxin from sea anemone *Stichodactyla helianthus*
VEGF	vascular endothelial growth factor

References

1. S. Mouhat; N. Andreotti; B. Jouirou, J. M. Sabatier, *Curr. Pharm. Des.* **2008**, *24*, 2503.
2. L. D. Possani; R. C. Rodriguez de la Vega, Scorpion Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Burlington, MA, 2006; p 339.
3. S. Mouhat; B. Jouirou; A. Mosbah; M. De Waard; J. M. Sabatier, *Biochem. J.* **2004**, *378*, 717.
4. F. Bontems; C. Roumestand; B. Gilquin; A. Ménez; F. Toma, *Science* **1991**, *254*, 1521.
5. P. K. Pallaghy; K. J. Nielsen; D. Craik; R. S. Norton, *Protein Sci.* **1994**, *3*, 1833.
6. E. Blanc; J. M. Sabatier; R. Kharrat; S. Meunier; M. El Ayeb; J. Van Rietschoten; H. Darbon, *Proteins* **1997**, *3*, 321.
7. F. Bontems; B. Gilquin; C. Roumestand; A. Ménez; F. Toma, *Biochemistry* **1992**, *31*, 7756.
8. D. Housset; C. Habersetzer-Rochat; J. P. Astier; J. C. Fontecilla-Camps, *J. Mol. Biol.* **1994**, *238*, 88.
9. D. A. Oren; O. Froy; E. Amit; N. Kleinberger-Doron; M. Gurevitz; B. Shaanan, *Structure* **1998**, *6*, 1095.
10. S. Zhu; I. Huys; K. Dyason; F. Verdonck; J. Tytgat; *Proteins* **2004**, *54*, 361.
11. R. S. Chuang; H. Jaffe; L. Cribbs; E. Perez-Reyes; K. J. Swartz, *Nat. Neurosci.* **1998**, *1*, 668.
12. T. Olamendi-Portugal; B. I. García; I. López-González; J. Van Der Walt; K. Dyason; C. Ulens; J. Tytgat; R. Felix; A. Darszon; L. D. Possani, *Biochem. Biophys. Res. Commun.* **2002**, *299*, 562.
13. M. L. Garcia; Y. Gao; O. B. McManus; G. J. Kaczorowski, *Toxicon* **2001**, *39*, 739.
14. T. C. Tenenholz; K. C. Klenk; D. R. Matteson; M. P. Blaustein; D. J. Weber, *Rev. Physiol. Biochem. Pharmacol.* **2000**, *140*, 135.
15. G. Lippens; J. Najib; S. J. Wodak; A. Tartar, *Biochemistry* **1995**, *34*, 13.
16. S. A. Lyons; J. O'Neal; H. Sontheimer, *Glia* **2002**, *39*, 162.
17. T. C. Tenenholz; K. C. Klenk; D. R. Matteson; M. P. Blaustein; D. J. Weber, *Rev. Physiol. Biochem. Pharmacol.* **2000**, *140*, 135.
18. C. Beeton; K. G. Chandy, *Neuroscientist* **2005**, *11*, 550.
19. K. G. Chandy; H. Wulff; C. Beeton; M. Pennington; G. A. Gutman; M. D. Cahalan, *Trends Pharmacol. Sci.* **2004**, *25*, 280.
20. A. Torres-Larios; G. B. Gurrola; F. Z. Zamudio; L. D. Possani, *Eur. J. Biochem.* **2000**, *16*, 5023.
21. A. L. Harvey, Snake Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 355.
22. S. D. Aird, *Toxicon* **2002**, *40*, 335.
23. D. C. Koh; A. Armugam; K. Jeyaseelan, *Cell Mol. Life Sci.* **2006**, *63*, 3030.
24. B. G. Fry; W. Wüster; R. M. Kini; V. Brusic; A. Khan; D. Venkataraman; A. P. Rooney, *J. Mol. Evol.* **2003**, *57*, 110.
25. J. R. de Weille; H. Schweitz; P. Maes; A. Tartar; M. Lazdunski, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2437.
26. S. C. Stotz; R. L. Spaetgens; G. W. Zamponi, *J. Membr. Biol.* **2000**, *174*, 157.
27. H. Schweitz; C. Heurteaux; P. Bois; D. Moinier; G. Romey; M. Lazdunski, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 878.
28. A. L. Harvey; B. Robertson, *Curr. Med. Chem.* **2004**, *11*, 3065.
29. J. M. Lancelin; M. F. Foray; M. Poncin; M. Hollecker; D. Marion, *Nat. Struct. Biol.* **1994**, *1*, 246.
30. C. G. Benishin; R. G. Sorensen; W. E. Brown; B. K. Krueger; M. P. Blaustein, *Mol. Pharmacol.* **1988**, *34*, 152.
31. A. Kerkis; I. Kerkis; G. Rádis-Baptista; E. B. Oliveira; A. M. Vianna-Morgante; L. V. Pereira; T. Yamane, *FASEB J.* **2004**, *18*, 1407.
32. F. Wang; H. Li; M. N. Liu; H. Song; H. M. Han; Q. L. Wang; C. C. Yin; Y. C. Zhou; Z. Qi; Y. Y. Shu; Z. J. Lin; T. Jiang, *Biochem. Biophys. Res. Commun.* **2006**, *351*, 443.
33. J. I. Weitz; M. A. Crowther, *Am. J. Cardiovasc. Drugs* **2003**, *3*, 201.
34. C. Marcinkiewicz, *Curr. Pharm. Des.* **2005**, *11*, 815.
35. R. S. Yang; C. H. Tang; W. J. Chuang; T. H. Huang; H. C. Peng; T. F. Huang; W. M. Fu, *Toxicon* **2005**, *45*, 661.
36. W. Rajendra; A. Armugam; K. Jeyaseelan, *Toxicon* **2004**, *44*, 1.
37. R. S. Norton, Sea Anemone Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 363.

38. L. Beress, *Pure Appl. Chem.* **1982**, *54*, 1981.
39. C. J. Grimmelikhuijzen; I. Leviev; K. Carstensen, *Int. Rev. Cytol.* **1996**, *167*, 37.
40. G. Anderluh; P. Macek, *Toxicon* **2002**, *40*, 111.
41. R. S. Norton, *Toxicon* **1991**, *29*, 1051.
42. M. W. Pennington; M. E. Byrnes; I. Zaydenberg; I. Khaytin; J. de Chastonay; D. S. Krafte; R. Hill; V. M. Mahnir; W. A. Volberg; W. Gorczyca; W. R. Kem, *Int. J. Pept. Protein Res.* **1995**, *46*, 354.
43. G. S. Gendeh; M. C. Chung; K. Jeyaseelan, *FEBS Lett.* **1997**, *418*, 183.
44. A. Aneiros; I. García; J. R. Martínez; A. L. Harvey; A. J. Anderson; D. L. Marshall; A. Engström; U. Hellman; E. Karlsson, *Biochim. Biophys. Acta* **1993**, *1157*, 86.
45. H. Schweitz; T. Bruhn; E. Guillemare; D. Moinier; J. M. Lancelin; L. Béress; M. Lazdunski, *J. Biol. Chem.* **1995**, *270*, 25121.
46. S. Diochot; H. Schweitz; L. Béress; M. Lazdunski, *J. Biol. Chem.* **1998**, *273*, 6744.
47. S. Diochot; E. Loret; T. Bruhn; L. Béress; M. Lazdunski, *Mol. Pharmacol.* **2003**, *64*, 59
48. R. S. Norton; M. W. Pennington; H. Wulff, *Curr. Med. Chem.* **2004**, *11*, 3041.
49. H. Wulff; M. Pennington, *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 438.
50. G. M. Nicholson, Spider Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 369.
51. P. Escoubas; L. Rash, *Toxicon* **2004**, *43*, 555.
52. E. Grishin, *Eur. J. Biochem.* **1999**, *264*, 276.
53. L. D. Rash; W. C. Hodgson, *Toxicon* **2002**, *40*, 225.
54. M. E. Adams, *Toxicon* **2004**, *43*, 509.
55. Y. A. Ushkaryov; K. E. Volynski; A. C. Ashton, *Toxicon* **2004**, *43*, 527.
56. H. Jackson; T. N. Parks, *Annu. Rev. Neurosci.* **1989**, *12*, 405.
57. H. W. Tedford; B. L. Sollod; F. Maggio; G. F. King, *Toxicon* **2004**, *43*, 601.
58. P. Escoubas; S. Diochot; M. L. Celerier; T. Nakajima; M. Lazdunski, *Mol. Pharmacol.* **2002**, *62*, 48–57.
59. K. J. Swartz; R. MacKinnon, *Neuron* **1995**, *15*, 941.
60. M. C. Sanguinetti; J. H. Johnson; L. G. Hammerland; P. R. Kelbaugh; R. A. Volkmann; N. A. Saccomano; A. L. Mueller, *Mol. Pharmacol.* **1997**, *51*, 491.
61. S. Diochot; M. D. Drici; D. Moinier; M. Fink; M. Lazdunski, *Br. J. Pharmacol.* **1999**, *126*, 251.
62. Z. Liao; C. Yuan; K. Peng; Y. Xiao; S. Liang, *Toxicon* **2007**, *50*, 135.
63. R. E. Middleton; V. A. Warren; R. L. Kraus; J. C. Hwang; C. J. Liu; G. Dai; R. M. Brochu; M. G. Kohler; Y. D. Gao; V. M. Garsky; M. J. Bogusky; J. T. Mehl; C. J. Cohen; M. M. Smith, *Biochemistry* **2002**, *41*, 14734.
64. F. Bode; F. Sachs; M. R. Franz, *Nature* **2001**, *409*, 35.
65. X. Chen; H. Kalbacher; S. Gründer, *J. Gen. Physiol.* **2005**, *126*, 71.
66. B. M. Olivera, Conus Snail Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 381.
67. P. Bouchet; P. Lozouet; P. Maestrat; V. Heros, *Biol. J. Linn. Soc.* **2002**, *75*, 421.
68. D. Röckel; W. Korn; A. J. Kohn. In *Book Manual of the Living Conidae*; Verlag Christa Hemmen: Wiesbaden, Germany, 1995.
69. H. Terlau; K. J. Shon; M. Grilley; M. Stocker; W. Stühmer; B. M. Olivera, *Nature* **1996**, *381*, 148.
70. D. R. Hillyard; V. D. Monje; I. M. Mintz; B. P. Bean; L. Nadasdi; J. Ramachandran; G. Miljanich; A. Azimi-Zoonooz; J. M. McIntosh; L. J. Cruz; J. S. Imperial; B. M. Olivera, *Neuron* **1992**, *9*, 69.
71. A. G. Craig; P. Bandyopadhyay; B. M. Olivera, *Eur. J. Biochem.* **1999**, *264*, 271.
72. U. C. Marx; N. L. Daly; D. J. Craik, *Magn. Reson. Chem.* **2006**, *44*, 41.
73. D. J. Espiritu; M. Watkins; V. Dia-Monje; G. E. Cartier; L. J. Cruz; B. M. Olivera, *Toxicon* **2001**, *39*, 1899.
74. J. M. McIntosh; R. M. Jones, *Toxicon* **2001**, *39*, 1447.
75. J. M. McIntosh; A. D. Santos; B. M. Olivera, *Annu. Rev. Biochem.* **1999**, *68*, 59.
76. A. D. Santos; J. M. McIntosh; D. R. Hillyard; L. J. Cruz; B. M. Olivera, *J. Biol. Chem.* **2004**, *279*, 17596.
77. H. Terlau; B. M. Olivera, *Physiol. Rev.* **2004**, *84*, 41.
78. L. Liu; X. Wu; D. Yuan; C. Chi; C. Wang, *Toxicon* **2008**, *51*, 1331.
79. G. P. Miljanich, *Curr. Med. Chem.* **2004**, *11*, 3029.
80. L. R. Webster; K. L. Fakata, *Pract. Pain Manage.* **2005**, *5*, 5.
81. C. Beeton; G. A. Gutman; K. G. Chandy, Targets and Therapeutic Properties of Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 403.
82. M. S. Palma, Insect Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 389.
83. G. Hahn; H. Leditschke, *Dtsch. Chem. Ges.* **1937**, *70*, 1637.
84. B. E. C. Banks; R. A. Shipolini, *Biochemical, Pharmacological and Behavioral Aspects*; Academic Press: London, UK, 1986; p 329.
85. P. Walde; H. Jäckle; P. L. Luisi; C. E. Dempsey; B. E. C. Banks, *Biopolymers* **1981**, *20*, 371.
86. A. F. Drake; R. C. Hider, *Biochim. Biophys. Acta* **1979**, *555*, 371.
87. C. R. Dawson; A. F. Drake; R. C. Hider, *Biochim. Biophys. Acta* **1978**, *510*, 75.
88. A. I. Miroshnikov; A. V. Boikov; L. G. S. E. Snezhkova; S. E. Sverin; V. I. Shvet; S. M. Dudkin, *Bioorg. Khim.* **1983**, *9*, 26.
89. V. Krishnakumari; R. Nagaraj, *Pept. Res.* **1997**, *50*, 88.
90. A. Argiolas; J. J. Pisano, *J. Biol. Chem.* **1984**, *259*, 10106.
91. K. Dohtsu; K. Okumura; K. Hagiwara; M. S. Palma; T. Nakajima, *Nat. Toxins* **1993**, *1*, 271.
92. K. Konno; M. S. Palma; I. Y. Hitara; M. A. Juliano; L. Juliano; T. Yasuhara, *Toxicon* **2002**, *40*, 309.
93. K. Konno; M. Hisada; H. Naoki; Y. Itagaki; T. Yasuhara; M. A. Juliano; L. Juliano; M. S. Palma; T. Yamane; T. Nakajima, *Toxicon* **2001**, *39*, 1257.
94. K. Konno; M. Hisada; Y. Itagaki; H. Naoki; N. Kawai; A. Miwa; T. Yasuhara; H. Takayama, *Biochem. Biophys. Res. Commun.* **1998**, *250*, 612.
95. Y. Sahara; M. Gotoh; K. Konno; A. Miwa; H. Tsubokawa; H. P. Robinson; N. Kawai, *Eur. J. Neurosci.* **2000**, *12*, 1961.
96. T. Kimura; K. Yamaoka; E. Kinoshita; H. Maejima; T. Yuki; M. Yakehiro; I. Seyama, *Mol. Pharmacol.* **2004**, *60*, 865.

97. F. Canduri; P. Delatorre; V. Fadel; C. C. Lorenzi; J. H. Pereira; J. R. Olivieri; J. Ruggiero Neto; K. Konno; M. S. Palma; T. Yamane; W. F. de Azevedo, Jr., *Acta Crystallogr. D Biol. Crystallogr.* **2000**, *56*, 1434.
98. K. Konno; M. Hisada; H. Naoki; Y. Itagaki; N. Kawai; A. Miwa; T. Yasuhara; Y. Morimoto; Y. Nakata, *Toxicon* **2000**, *38*, 1505.
99. M. L. Sforça; S. Oyama, Jr.; F. Canduri; C. C. Lorenzi; T. A. Pertinhez; K. Konno; B. M. Souza; M. S. Palma; J. Ruggiero Neto; W. F. Azevedo, Jr.; A. Spisni, *Biochemistry* **2004**, *43*, 5608.
100. T. Piek; A. Duval; B. Hue; H. Karst; B. Lapied; P. Mantel; T. Nakajima; M. Pelhate; J. O. Schmidt, *Comp. Biochem. Physiol. C* **1991**, *99*, 487.
101. J. Orivel; V. Redeker; J. P. Le Caer; F. Krier; A. M. Revol-Junelles; A. Longeon; A. Chaffotte; A. Dejean; J. Rossier, *J. Biol. Chem.* **2001**, *276*, 17823.
102. J. Orivel; A. Dejean, *Toxicon* **2001**, *39*, 195.
103. K. Pluzhnikov; E. Nosyreva; L. Shevchenko; Y. Kokoz; D. Schmalz; F. Hucho; E. Grishin, *Eur. J. Biochem.* **1999**, *262*, 501.
104. H. Inagaki; M. Akagi; H. T. Imai; R. W. Taylor; T. Kubo, *Arch. Biochem. Biophys.* **2004**, *428*, 170.
105. W. R. Kem. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier, 2006; p 397.
106. W. R. Kem, *J. Biol. Chem.* **1976**, *251*, 4184.
107. F. A. Meunier; Z. P. Feng; J. Molgó; G. W. Zamponi; G. Schiavo; *EMBO J.* **2002**, *24*, 6733.
108. C. Beeton; G. A. Gutman; K. G. Chandy, Targets and Therapeutic Properties of Venom Peptides. In *Handbook of Biologically Active Peptides*; J. Kastin, Ed.; Elsevier: Amsterdam, 2006; p 403.
109. M. Zasloff, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5449.
110. M. Amiche; F. Ducancel; E. Lajeunesse; J. C. Boulain; A. Ménez; P. Nicolas, *Biochem. Biophys. Res. Commun.* **1993**, *191*, 983.
111. A. Mor; K. Hani; P. Nicolas, *J. Biol. Chem.* **1994**, *50*, 31635.
112. A. Mor; P. Nicolas, *Eur. J. Biochem.* **1994**, *15* (1–2), 145.
113. J. Goraya; Y. Wang; Z. Li; M. O'Flaherty; F. C. Knoop; J. E. Platz; J. M. Conlon, *Eur. J. Biochem.* **2000**, *267*, 894.
114. C. R. Bevier; A. Sonnevend; J. Kolodziejek; N. Nowotny; P. F. Nielsen; J. M. Conlon, *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **2004**, *139*, 31.
115. K. P. Sai; M. V. Jagannadham; M. Vairamani; N. P. Raju; A. S. Devi; R. Nagaraj; N. Sitaram, *J. Biol. Chem.* **2001**, *276*, 2701.
116. A. C. Rinaldi; M. L. Mangoni; A. Rufo; C. Luzi; D. Barra; H. Zhao; P. K. Kinnunen; A. Bozzi; A. Di Giulio; M. Simmaco, *Biochem. J.* **2002**, *368*, 91.
117. T. Isaacson; A. Soto; S. Iwamuro; F. C. Knoop; J. M. Conlon, *Peptides* **2002**, *23*, 419.
118. Y. J. Basir; F. C. Knoop; J. Dulka; J. M. Conlon, *Biochim. Biophys. Acta* **2000**, *1543*, 95.
119. T. Halverson; Y. J. Basir; F. C. Knoop; J. M. Conlon, *Peptides* **2000**, *21*, 469.
120. W. S. Jang; C. H. Kim; K. N. Kim; S. Y. Park; J. H. Lee; S. M. Son; I. H. Lee, *Antimicrob. Agents Chemother.* **2003**, *47*, 2481.
121. I. H. Lee; Y. S. Lee; C. H. Kim; C. R. Kim; T. Hong; L. Menzel; L. M. Boo; J. Pohl; M. A. Sherman; A. Waring; R. I. Lehrer, *Biochim. Biophys. Acta* **2001**, *1527*, 141.
122. Y. Q. Tang; S. H. Tian; S. X. Wu; J. C. Hua; G. F. Wu; E. M. Zhao; Y. A. Lu; Y. Q. Zhu; G. Zou; K. Tsou, *Sci. China B* **1989**, *32*, 570.
123. Z. Fathi; J. W. Way; M. H. Corjay; J. Viallet; E. A. Sausville; J. F. Battey, *J. Cell. Biochem. Suppl.* **1996**, *24*, 237.
124. J. Jimeno; G. Faircloth; J. M. Fernandez Sousa-Faro; P. J. Scheuer; K. Rinehart, *Mar. Drugs* **2004**, *1*, 14.
125. J. Jimeno; J. A. López-Martín; A. Ruiz-Casado; M. A. Izquierdo; P. J. Scheuer; K. Rinehart, *Anticancer Drugs* **2004**, *15*, 321.
126. G. Zetler, *Peptides* **1985**, *6*, 33.
127. L. Marenah; C. Shaw; D. F. Orr; S. McClean; P. R. Flatt; Y. H. Abdel-Wahab, *Regul. Pept.* **2004**, *120*, 33.
128. A. Thum; K. Hupe-Sodmann; R. Göke; K. Voigt; B. Göke; G. P. McGregor, *Exp. Clin. Endocrinol. Diabetes* **2002**, *110*, 113.
129. N. C. Kesty; J. D. Roth; D. Maggs, *Expert Opin. Biol. Ther.* **2008**, *8*, 1733.
130. A. J. Krentz; M. B. Patel; C. J. Bailey, *Drugs* **2008**, *68*, 2131.

Biographical Sketches

Nicolas Andreotti is a Ph.D. student under the supervision of Dr. Sabatier at the ERT 62 laboratory. He also has a permanent position in a biopharmaceutical company. He works on animal peptide toxins and candidate drugs, and has contributed to approximately 10 scientific articles and 10 communications.

Besma Jouirou has a Ph.D. in neurosciences. She is affiliated to the ERT 62 laboratory and holds a position as a researcher in a biopharmaceutical company. She works in the field of animal toxins, antitumor compounds, and antivirals. She has contributed to 10 scientific articles, 6 communications, and 2 patents.

Stephanie Mouhat is an engineer and has a Ph.D. in biology. She is affiliated to the ERT 62 laboratory and holds a position as a researcher in a biopharmaceutical company. She works in the field of therapeutic peptides derived from venomous animal toxins, and has contributed to more than 10 scientific articles, above 20 communications, and 2 patents in the field. She won the prize for the best Ph.D. thesis at the Université de la Méditerranée in 2006.

Ludovic Mouhat is an engineer in bioinformatics. He is affiliated to the ERT 62 laboratory and holds a position as a researcher in a biopharmaceutical company. He is involved in the design and chemical production of candidate therapeutic peptide drugs.

Jean-Marc Sabatier has a Ph.D. and HDR in biochemistry. He is the director of research at the French 'Centre National de la Recherche Scientifique' (CNRS). He heads a research laboratory (ERT 62) entitled 'Engineering of Therapeutic Peptides' at the Université de la Méditerranée, in Marseilles, France. He also holds the position of a senior director (discovery research – peptides) for a public company in Canada. Dr. Sabatier works in the field of animal toxins, and leads the venom peptide group of the International Neuropeptide Society. He also designs immunomodulatory and antiviral drugs, as well as contributes to the field of peptide and protein engineering. He has contributed more than 100 scientific articles, 180 communications, and 43 patents. He is a member of several scientific advisory boards of journals (e.g., *Peptides*, *Biochemical Journal*), and has reviewed articles submitted for publication in more than 30 specialized international journals.

5.11 Signal Transduction in Gram-Positive Bacteria by Bacterial Peptides

Ingolf F. Nes, Ola Johnsborg, and Dzung Bao Diep, Norwegian University of Life Sciences, Aas, Norway

© 2010 Elsevier Ltd. All rights reserved.

5.11.1 Introduction

Bacterial communication has been recognized for years and most often takes place through chemical signaling. Intercellular communication mediated by small molecules controls numerous important microbial properties, including virulence, horizontal genetic transfer, and production of bacteriocins. In one important form of microbial cell–cell communication, quorum sensing (autoinduction), all members of a population consisting of a single cell type participate in both signal production and response. The chemical nature of signaling molecules differs between Gram-positive (G+) and Gram-negative (G−) bacteria. In G+ bacteria such communication is mediated through the release of a peptide pheromone (PP) while in G− bacteria it occurs most often through various δ-lactone molecules. This chapter will cover peptide-signaling transduction in G+ bacteria.

5.11.2 Peptide Pheromone Dependent Signaling Systems in Bacteriocin Production

The regulation of bacteriocin synthesis by PPs goes back to mid-1990s when it was shown that the Class II bacteriocins (nonmodified heat-stable bacteriocins) as well as Class I bacteriocins most often referred to as lantibiotics (post-translatory modified containing most frequently lanthionine, methyl-lanthionine, dehydrated serine, and threonine) were regulated by small ribosomally synthesized peptides either resembling a peptide

bacteriocin (plantaricin A) or being one (nisin).[1–4] The regulation of nisin and plantaricin differs in one major aspect. The antimicrobial peptide nisin molecule itself acts also as the signaling molecule (also termed induction peptide, induction factor, PP, etc.),[5] which is different from the plantaricin system where the induction peptide, PlnA, is a separate molecule that operates solely as a PP.[6]

The first PP signaling system for regulation of bacteriocin synthesis in lactic acid bacteria (LAB) was described in *Lactobacillus plantarum* C11 where it was shown that bacteriocin production was controlled by a quorum sensing system through plantaricin A.[6] Since then the sequencing of the *L. plantarum* WCFS1 genome revealed that this strain also carried a functional plantaricin system identical to the one previously found in the C11 strain except with a few changes in the nucleotide sequence.[7] Recently, it was shown that the same plantaricins were also produced in *L. plantarum* NC8.[8] But more interesting was the observation that the three-component regulatory system (composed of an inducer peptide, histidine protein kinase (HPK), and a response regulator (RR)) controlling plantaricins in NC8 shared relatively low homology to the ones found in C11 and WCFS1 except for the RRs while the bacteriocins were the same. The two PPs did not share any significant homology to each other (see **Tables 1** and **2**).

Table 1 Peptide pheromones (PPs) involved in regulation of Class II bacteriocin biosynthesis: Overview of the various PPs involved in different bacteriocin systems

Bacteriocin	*Peptide pheromone*	*Length of amino acid residues*	*Reference(s)*	*Producer*
Plantaricins	Pln A	26	6	*Lactobacillus plantarum C11*
Plantaricins	pINC8IF	28	8	*L. plantarum NC8*
Sakacin P	IP-673	19	9	*L. sakei*
Sakacin A	IP-706	23	10	*L. curvatus/sakei*
Carnobacteriocin A	IP-LV17	24	11, 12	*Carnobacterium maltaromaticum*
Enterocin A/B	Ent F	25	13	*Enterococcus faecium*
Blp bacteriocin	Pheromone BlpC	30 and 19	14	*Streptococcus termophilus*
Sakacin X/T	IP-TX	27	15	*L. sakei 5*
Piscicolin 126	PisN	24	16	*C. maltaromaticum*
Lactacin B	IP LBA-1800	25	17	*L. acidophilus*
Penocin A	Pen I	23	18	*Pediococcus pentosaceus*
ABP-18	Abp IP	21	19	*L. salivarius*
SmbAB	CSP	21 and 18	20	*S. mutants*

Table 2 Peptide pheromones (PPs) involved in regulation of Class II bacteriocin biosynthesis: Amino acid sequences of the PPs

PlnA	KSSAYSLQMGATAIKQVKKLFKKWGW
pINC8IF	KTKTISLMSGLQVPHAFTKLLKALGGHH
IP-673	MAGNSSNFIHKIKQIFTER
IP-706	TNRNYGKPNKDIGTCIWSGFRHC
IP-LV17	SKNSQIGKSTSSISKCVFSFFKKC
Ent F	AGTKPQGKPASNLVECVFSLFKKCN
BlpC	SGWMDYINGFLKGFGGQRTLPTKDYNIPQA
IP-TX	TPGGFDIISGGPHVAQDVLNAIKDFFK
PisN	NKSVIKGNPASNLAQCVFSFFKKC
IPL BA-1800	KKAPISGYVGRGLWENLSNIFKHHK
PenI	IKKKLKLEATKKLLVGIFKWLSQ
AbpIP	ATKKGGFKRWQCIFTFFGVCK
CSP	SGSLSTFFRLFNRSFTQALGK

Since the initial discovery of quorum sensing regulation of bacteriocin synthesis in LAB, numerous systems have been described both among Class I and Class II bacteriocins. An overview of the best-characterized PPs of the various three-component systems among Class II bacteriocins is presented in **Tables 1** and **2**. It should be emphasized that not all Class II bacteriocins are regulated through such a quorum sensing system and several bacteriocins seem to be constitutively produced.

In the first part of this review, the regulation of plantaricin expression in *L. plantarum* will be used as the model to describe the signaling system of bacteriocin biosynthesis and where relevant, it will be extended to other similar systems to emphasize differences between the various systems.

5.11.3 The Signaling Pathway of Plantaricin C11 System and Other Class II Bacteriocins

During the investigation of bacteriocin production in *L. plantarum* C11, an isolate originating from cucumber fermentation, it was proved in 1995 that the bacteriocin biosynthesis was regulated by a peptide, originally termed plantaricin A.[6,21] Plantaricin A was shown to be genetically located in an operon structure composed of four genes encoding a protein with strong homology to HPKs and two genes encoding homologous RRs.[21] It was pertinently presented that this operon encodes a three-component regulatory system because three players were involved in the regulation: the PP, the HPK (a membrane-located receptor), and the RRs (DNA-binding proteins).[6] The regulatory mechanism was found to involve bacterial communication through a cell density-dependent accumulation of the PP. At critical threshold concentration of PP, the three-component regulatory circuit was greatly autoactivated, which in turn triggered expression of other operons involved in bacteriocin synthesis.[6,22,23] This process, termed quorum sensing, allows the bacterial population to control the expression of genes, coordinately.

In the plantaricin system, five operons are coordinately controlled by the plantaricin A signaling system. In addition to the regulatory operon, two operons encode bacteriocin precursors and immunity proteins, one operon encodes the transport system of the bacteriocins and the PP itself, and the last operon contains four genes of unknown function (it could be a bacteriocin encoding operon but no activity has been found so far).[22] The key operon encoding the three-component signaling pathway (composed of a PP, an HPK, and two RRs) will be discussed below.

5.11.3.1 Peptide Pheromones

A few characteristics of the PPs from different bacteriocin systems are easily recognized. They are secreted by the same ATP-binding cassette (ABC) transporter as their bacteriocins.[24] All Class II PPs are processed from precursors containing an N-terminal double-glycine leader[25] that allows both the secretion and the removal of the N-terminal leader by a dedicated ABC transporter system to take place, and this process is identical for the bacteriocins. The resulting PPs are shorter (varying between 18 and 30 amino acid residues) than their bacteriocins but share some of the physicochemical properties of the bacteriocins being cationic and amphiphilic/hydrophobic. While other PPs do not encompass antimicrobial activity, plantaricin A does. However, unlike Class II bacteriocins, no dedicated immunity protein is genetically linked to the PP (plantaricin A) and its antimicrobial activity seems to be unspecific probably due to its amphiphilic/cationic properties. In addition, plantaricin A does not need any specific receptor on the target organisms to exert its antimicrobial activity like most Class II bacteriocins do.[26]

The membrane-embedded HPK (PlnB) serves as the ligand-binding receptor for plantaricin A that through its specific interaction induces a phosphorylation relay resulting in the phosphorylation of the cognate RRs PlnC and PlnD. The interaction between plantaricin A and its cognate HPK was found dependent on chirality of the peptide as only the L- but not the D-enantiomeric form had induction ability.[27] The phosphorylated RRs in turn bind to pairwise direct repeats of the dedicated promoters (in the vicinity of poorly defined −35 regions) and activate the gene expression of the operons in the *pln* bacteriocin locus including the autoregulated three-component operon.[3,28–30] The binding of PPs is very specific to their cognate receptors (HPK).

It was shown the induction activity of the PlnA peptide was not affected by removal of the four N-terminal amino acids.[6] When additional four residues were removed from the N-terminus and the three C-terminal amino acid residues were removed, the truncated plantaricin A molecule (15-mer) still exerted between 2 and 10% of the original induction activity.[6] In contrast to these observations, it was surprising that one can detect some induction activity in the very short N-terminal 5-mer of PlnA.[27] The regulation of enterocins A and B has also been shown to be controlled by a three-component regulatory system and the EntF peptide is the acting PP. When the residue Cys-16 in EntF was replaced by Phe, its induction activity was almost abolished (at least 100 000 fold reduction),[13] which strongly indicates a highly specific interaction between a ligand EntF (PP) and its cognate HPK.

In a recent study, a structure–function analysis of the highly homologous pheromones CbaX and EntF inducing bacteriocin production in *Carnobacterium maltaromaticum* LV17A and *Enterococcus faecium* CTC492, respectively, was carried out.[12] It was observed that cross-induction of bacteriocin production could take place at high concentration of CbaX in *E. faecium* CTC492 though EntF was not able to cross induce bacteriocin production in *C. maltaromaticum* LV17A.

Induction and competition experiments that included use of hybrid pheromones and synthesized pheromone fragments (10-mers) suggested a model of interaction between the PPs and their cognate HPKs. It was suggested that an unspecific and weak interaction by the C-terminal part followed by a highly specific recognition of the N-terminal part of the pheromone to its HPK takes place.[12]

The three-dimensional structure of the 26-mer plantaricin A pheromone (PlnA) has been published and simply illustrated in **Figure 1(a)**. The structure was obtained by nuclear magnetic resonance (NMR) analysis in dodecyl phosphocholine micelle-forming solvent. It was shown that PlnA was unstructured in water, but a large part of the peptide had a defined structure upon exposure to the micelle-forming solvent.[27] The structuring leads to a well-defined α-helical conformation in the amphiphilic region between amino acid residue 12 and 22 while other parts of PlnA remain mostly unstructured. The α-helix of PlnA is postulated to be positioned parallel to membrane lipid with the hydrophobic residues dipping into the membrane while the hydrophilic part is directed toward the membrane–water interphase. The interaction with the membrane is believed to help the unstructured PP become structured and will then be able to specifically interact with its cognate receptor (HPK).

5.11.3.2 The Receptor – Histidine Protein Kinase

The environmental concentration of plantaricin A is monitored by the membrane-bound HPK, PlnB, which triggers phosphorylation of its cognate RRs PlnC and PlnD. This regulation subsequently activates transcription of the bacteriocin genes. HPKs can be divided into distinct subfamilies based on their degree of amino acid homology in the kinase domain. The majority of Class II PP-activated HPKs come under one distinct subfamily, the HPK_{10}. In all of these systems, the signaling molecules perceived by their HPK sensors are PPs produced by the bacteria themselves. Interestingly, it was shown that all known PP-activated HPKs except SpaK, ComP, and NisK come under the HPK_{10} subfamily.[32] The HPK_{10} subfamily includes, among others, VirS from *Clostridium*,[33] PlnB from *Lactobacillus*,[21] ComD from *Streptococcus*, AgrC from *Staphylococcus*,[34,35] and CbnK from *Carnobacterium*.[11] All members of the HPK_{10} subfamily belong to the orthodox kinases each

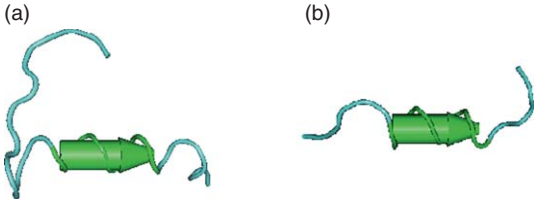


Figure 1 Peptide pheromone structures determined by NMR. (a) Structure of peptide pheromone (PlnA) from *Lactobacillus plantarum* required for plantaricin biosynthesis.[27] (b) Structure of peptide pheromone (ComC) from *Streptococcus pneumonia* required for competence development.[31]

consisting of a membrane-spanning N-terminal domain and a C-terminal cytoplasmic kinase domain.[32,36] However, they differ from other membrane-localized HPKs in two aspects. First, while the core domain in other HPKs usually contains a set of conserved regions (the N, D, F, and G boxes) that are involved in nucleotide binding, the HPK_{10} nucleotide-binding domain apparently lacks a D box and contains only one asparagine in the N box.[34] Second, it is predicted that the N-terminal membrane-associated domain of the HPK_{10} family contains 5–7 transmembrane segments (TMSs), whereas the majority of other HPKs contain only two TMSs.[36]

To investigate the mechanism underlying pheromone-mediated activation of the HPK_{10} subfamily, the membrane topology of PlnB from *L. plantarum* was determined using reporter fusion assay.[37] It was shown that PlnB is anchored to the cytoplasmic membrane through seven N-terminal located TMSs. By domain switching between HPK_{10} members, it was demonstrated that the determinants for pheromone binding and specificity are contained within the N-terminal transmembrane domain where the very N-terminal extracytoplasmic loop plays a key role.[38] Computer prediction suggests that the membrane domains of HPK_{10} kinases adopt similar topology structures and combined with gene fusion experiments, it is likely that the mechanism of signal transduction (the interaction between the induction peptides and the receptor transmembrane domains of HPKs) is conserved between members of the HPK_{10} subfamily.[38] However, details in this interaction still have to be resolved.

In addition a site-directed mutational analysis (alanine substitution) of more than 20 amino acid residues positioned in the N-terminal PlnB membrane domain was performed.[39] Most of the mutations were targeted in the loops. In this study, the receptor functionality of various mutagenized PlnB membrane domains was investigated by employing an *in vivo* reporter assay. The various mutants were evaluated for their ability to interact with PP-PlnA and subsequently activate the cytoplasmic kinase domain. The two mutations found to reduce PlnB receptor functionality were located in the most N-terminal extracytoplasmic loop (D54 and S68) of HPK. The results indicated that important determinants for receptor function are located in the most N-terminal loop of the membrane domain directed toward the exterior. Combined with the data also suggesting that the last TMS is also involved in the receptor ligand (PlnB-PlnA)-binding specificity, the present model suggests that several TMSs and their extracytoplasmic loops are involved in this interaction. Structural studies based on a three-dimensional analysis are probably needed to uncover the specificity of the PlnA and PlnB interaction.

5.11.3.3 The Response Regulator Protein

The final player in the signal transduction pathway of a three-component regulatory circuit is the RR. After being phosphorylated by its cognate HPK, the RR binds strongly to the regulated promoters to activate transcription of the operons involved in the bacteriocin synthesis. The ability of an RR to bind to regulated promoters has been shown for several bacteriocin systems but most thoroughly investigated in the plantaricin system of *L. plantarum* C11. It should be emphasized that while all other bacteriocin regulatory systems have only one RR, the plantaricin C11 regulon contains two homologous tandemly located response regulatory genes encoding PlnC and PlnD, which share 70% identity to each other. Both RRs seem to serve a regulatory function in bacteriocin production. On the basis of experimental results the present regulatory model is as follows: PlnC is the positive regulator that turns on the transcription; PlnD on the other hand accumulates to a certain critical threshold amount sometime later in the bacterial growth and eventually counteracts with PlnC to turn off transcription of all the operons involved in the plantaricin production.[23,40]

Already back in 1996 it was postulated that a DNA sequence containing two direct repeats, separated by 12- to 13-bp spacer, and in the vicinity of a poor −35 promoter sequence was the binding site of RRs in several bacteriocin regulatory systems.[3] The regulatory DNA sequence in the plantaricin system was experimentally confirmed and refined, and a consensus sequence of the two imperfect repeats was established.[29,30] Nucleotide substitutions in the consensus sequence, particularly those in invariant positions, either abolished or significantly reduced RR binding to its binding site.[29] The RRs bind as homodimers to DNA fragments containing a complete set of regulatory elements, while removal of either repeat, or alterations in the length of the spacer region, significantly weakened dimer RR binding. Detailed mutation studies of the

promoter regions combined with binding ability, transcriptional reporter assay, and DNA footprinting gave a consensus DNA-binding sequence of C11-RRs.[29,30]

The existence of specific binding sites for RR was also experimentally demonstrated in sakacin P production in *Lactobacillus sake* LTH673 that involves at least four regulated operons.[28] The promoters controlling bacteriocin operons were shown to be strictly regulated, and their activity increased more than 1000-fold upon activation by a PP.

5.11.4 Peptide-Dependent Regulation of Lantibiotics

In the same way as regulation of Class II bacteriocins, several lantibiotics are shown to be regulated by a three-component peptide-dependent regulatory system. The regulation of nisin production is the best-studied system among the lantibiotics. Already in 1995 it was published that nisin is regulated by a regulatory circuit composed of an HPK, an RR, and nisin itself being the PP, which is a major difference from regulation of Class II bacteriocin synthesis (the latter being dependent on a dedicated PP; see above). This autoregulatory system is well described and basically it follows a signaling transduction mechanism similar to that described for Class II bacteriocins (see above and Kleerebezem[41]). The antimicrobial and signaling activities of such small peptides like nisin and their alike are independent activities. It has been shown that some mutations in the nisin molecule inactivated one of the functions while still retaining the other.[5] In addition to the two linear lantibiotics nisin[5] and subtilin,[42,43] the biosynthesis of salivaricin A and streptin is also autoregulated through such a regulatory circuit.[44,45] Recently, the globular lantibiotic mersacidin was shown to autoregulate itself through a similar mechanism mediated by an HPK and an RR.[46] It is worth to note that the genetic organization of lantibiotic regulatory systems differs from that observed for the regulation of Class II bacteriocins. While the regulatory determinants of the non-lantibiotic system are always located within the same operon in this order PP, HPK, and RR, their counterparts in the lantibiotic systems are not necessarily located within the same operon and the genetic order of the signal transduction system is normally opposite, namely, with an RR gene in front of an HPK gene.

5.11.5 Induction of Competence for Natural Genetic Transformation in Streptococci

Several streptococcal species from the mitis phylogenetic group, such as *S. pneumoniae*, *S. mitis*, *S. gordonii*, and *S. sanguinis*, are known for their ability to take up naked DNA from the surroundings and incorporate it into their genome.[47,48] More than seven decades ago it was demonstrated that this phenotype, today termed competence for genetic transformation, was a transient property of *S. pneumoniae*.[49] Further investigations led to the conclusion that development of competence *in vitro* occurred in a coordinated manner throughout the bacterial population once the cells reached a certain density, indicating that competence is controlled by a cell–cell signaling mechanism.[50–52] These early studies also demonstrated that signaling was mediated by secretion of a proteinaceous activator molecule. However, the exact nature of the pneumococcal competence activator remained elusive until 1995, when it finally was identified as an unmodified ribosomally synthesized 17-residue basic PP, which was termed the competence stimulating peptide (CSP).[53] It has later been shown that the other competent members of the mitis phylogenetic group secrete variants of this peptide.[48] Thus, production and secretion of CSP is part of a conserved mechanism regulating competence development in the mitis group of bacteria.

5.11.5.1 The Competence Stimulating Peptide

At present, more than 40 CSP variants with different primary structures have been identified from various strains and species from the mitis group.[37] In general, a CSP with a certain primary structure is not able to induce competence in a strain or species that produce a CSP with a different primary structure.[31,54] Hence, on basis of a particular CSP pheromone, they respond to mitis group bacteria that can be divided into pherotype

groups. Despite differences in primary structure, all CSPs from the mitis phylogenetic group contain a conserved sequence fingerprint composed of a negatively charged N-terminal residue, an arginine residue in position 3, and a positively charged C-terminal tail.[37] In contrast, the central region of the peptides displays a high level of sequence variation. The CSP variants are encoded by various alleles of the *comC* gene and are synthesized with a characteristic N-terminal double glycine leader peptide that acts as a secretion signal. The leader is recognized and removed concomitantly with secretion by the ComAB ABC transporter, which acts as a dedicated secretion machinery.[53,55,56] Genetic studies of the chromosomal regions surrounding the *comC* locus has revealed that *comC* is cotranscribed with two additional genes, *comD* and *comE*, encoding an HPK and an RR, respectively.[57] ComD belongs to HPK subfamiliy 10, which is characterized by HPKs that contain a large noncatalytic membrane domain consisting of 5–7 predicted transmembrane helices coupled to a cytoplasmatic histidine kinase domain. While the kinase domain is highly conserved, the receptor domains of various ComDs show a high degree of sequence variation. Indeed, it has been demonstrated that the variable N-terminal part of the membrane receptor domain is responsible for the specificity of CSP recognition.[58] Interestingly, the ComD receptors of *S. pneumoniae* strain R6 (ComD-1) and *S. pneumoniae* A66 (ComD-2) only differ in 12 amino acid positions.[39] These positions are located within the 60 N-terminal amino acids and mainly involve positions with hydrophobic amino acids. Despite the high level of identity, the corresponding CSPs, CSP-1, and CSP-2 are unable to cross induce the noncognate receptor in an efficient manner.[31,54] Hence, it is likely that the receptor–ligand interaction involves hydrophobic contacts. This hypothesis is also supported by other observations. By use of CD spectroscopy, it has been demonstrated that the pheromones adopt an α-helical structure upon exposure to membrane-mimicking environments (**Figure 1**), indicating that structuring of the pheromones is initiated upon interaction with the membrane of target bacteria.[12] NMR spectroscopy analysis of the predominant CSP variant from *S. pneumoniae*, CSP-1, demonstrated that the helical region is made up of the less conserved central region of the peptide.[31] Interestingly, the α-helix is highly amphiphilic, with the nonpolar residues Phe7, Phe8, Phe11, and Ile12 facing one side of the helix and Lys6, Arg9, and Asp10 facing the opposite side (**Figure 1(b)**). This amphiphilic type of structure appears to be a common trait in all CSPs from mitis group bacteria, as inspection of the primary sequence of other CSPs shows that these peptides also might have the ability to form similar amphiphilic helices in this region. Alanine scanning mutagenesis of CSP-1 has previously indicated that the hydrophobic residues Phe7, Phe8, and Phe11 are essential for pheromone bioactivity. Recently, homologue scanning mutagenesis of CSP-2, in which certain residues of CSP-2 were swapped with the corresponding residues of CSP-1, demonstrated that the presence of CSP-1 phenylalanine residues Phe7, Phe8, together with Leu4, is essential for binding and activation of the ComD-1 receptor.[31] The exact mechanism by which the hydrophobic residues facilitate receptor recognition and activation is not known. However, it has been speculated that the hydrophobic interaction functions to correctly position the CSP N-terminus, containing the conserved Arg3, into a conserved pocket of the ComD receptor. Binding of CSP to the ComD receptor results in autophosphorylation of the cytoplasmatic ComD kinase domain. At present little is known as to how the receptor domain is able to relay the information of CSP binding across the bacterial membrane to activate the kinase domain. However, it is reasonable to assume that CSP binding results in structural changes in the receptor domain, and that this shift in conformation is sensed by the intracellular kinase domain.

5.11.5.2 The Competence Regulon

Once phosphorylated, the ComD kinase domain activates the ComE RR, presumably by transferring the phosphate group to the latter protein.[57] Once phosphorylated, ComE acts as a DNA-binding protein that binds and activates a subset of promoters that regulate the expression of a group of genes referred to as the early competence genes.[59–61] The *comAB* and *comCDE* genes are found within this class of genes, and phosphorylation of ComE thus leads to a rapid increase in synthesis and export of CSP. As a consequence of the increased extracellular level of CSP, the intracellular level of phosphorylated ComE will increase. Once a threshold concentration of phosphorylated ComE is reached another subset of competence genes, the so-called late genes are activated. Activation of the late genes is regulated by the alternative sigma-factor ComX, whose expression is activated when the level of phosphorylated ComE reaches a certain threshold level.[62,63] In *S. pneumoniae*, ComX controls the expression of about 80 late genes. Among these are the genes that encode the DNA uptake apparatus as well as genes that encode proteins participating in homologous recombination of the incoming DNA.

5.11.5.3 An Evolutionary Model for Competence Stimulating Peptide Diversity

There is a remarkable diversity with respect to the CSPs produced by bacteria from the mitis phylogenetic group. How did this diversity arise? As a specific CSP must coevolve with its cognate ComD receptor to retain its ability to act as a ligand, most random mutations affecting the CSP would be selected against. It is therefore highly unlikely that the CSP diversity has arisen as a result of genetic drift. Rather, statistical analysis has shown that the diversity is likely to have evolved due to a positive selection pressure.[64] As mentioned previously, ComX regulates the expression of about 80 late genes. However, only 14 of these genes encode proteins that are essential for DNA uptake and recombination, indicating that other bacterial traits might be coregulated with the competence phenotype. It has recently been established that competent *S. pneumoniae* are able to attack and rupture the cell wall of noncompetent siblings by a process termed fratricide.[65–68] In liquid media, this mechanism absolutely requires production of the putative murein hydrolase CbpD, which is encoded by one of the late genes.[66] Fratricide also depends on production of one of the two autolysins LytA and LytC.[69] To protect themselves from their own lysins, competent pneumococci produce an immunity factor, ComM, which is encoded by one of the early genes.[70] Very recently it has been demonstrated that this killing mechanism not only function to lyse noncompetent but otherwise isogenic cells, but also function to kill cells belonging to other species of the mitis phylogenetic group including both *S. mitis* and *S. oralis* (O. Johnsborg *et al.*, unpublished data). Homologues to the genes encoding CbpD and ComM have previously been identified in the genomes of both *S. mitis* and *S. oralis*,[71] and it has now been confirmed that CbpD-mediated cell lysis is utilized by competent *S. mitis* (O. Johnsborg, unpublished data). It thus appears that fratricide is a general mechanism employed by competent bacteria from the mitis phylogenetic group. The regulation of fratricide through CSP-mediated cell–cell communication provides cells of the same pherotype with means to coordinate their production of lysins and immunity, enabling them to attack and kill related bacteria inhabiting the same ecological niche. This strategy can only succeed if the bacteria belonging to a specific pherogroup communicate with a pheromone that cannot be detected by the ComD receptor of the targeted bacteria, since targeted bacteria that could sense the foreign CSP would be induced to competence and hence start to express the *comM* immunity gene. Thus, there is a positive selection pressure that stimulates the evolution of novel CSP variants that are not detected by competing strains. From the perspective of the targeted bacteria, a positive selection pressure exists that favors the evolution of promiscuous ComD receptors that are able to detect CSPs produced by other pherogroups. It is likely that these opposing selection pressures have been the main driving force creating the observed CSP diversity.

5.11.5.4 Competence Stimulating Peptide Communication and Horizontal Gene Transfer

In principle, competent streptococci will bind and import DNA from any source. However, certain recombination barriers must be overcome for the incoming DNA to be incorporated into the genome. One major obstacle to recombination in competent cells is a high degree of DNA sequence divergence between incoming DNA and the recipient chromosome).[72] Low homology DNA will recombine with low efficiency and can often be more harmful than beneficial to the recipient. The fact that induction of natural competence is controlled by CSP-mediated cell–cell signaling probably helps to ensure the presence of homologous donor DNA. CSP communication allows the competent cells of the same pherotype to mount an attack against closely related bacteria, resulting in lysis of the target cells. Such lysis has been shown to result in a burst of released DNA into the bacterial growth medium[66] and has very recently been demonstrated to dramatically increase the efficiency of gene flow from the lysed cells into the competent attackers (O. Johnsborg *et al.*, unpublished data). Since the lysis mechanism appears to be restricted to function only against members of the mitis phylogenetic group, it is likely that it has evolved to increase the presence of homologous donor DNA during competence.

5.11.6 Virulence Regulation by Peptide Signaling in *Staphylococcus aureus*

Staphylococcus aureus is an important human pathogen that causes a variety of clinical manifestations, ranging from benign skin infections to life-threatening infections such as septicemia, endocarditis, osteitis, and toxic shock syndrome.[73] The virulence has been ascribed to a coordinated production of a large set of different toxins,

hydrolytic enzymes, such as lipases, proteases, and others, that cause damage, lyse host cells, or interfere with the immune system. The onset of the virulence genes, which takes place during post-exponential growth phase, is controlled by several regulatory loci including *agr*, *sar*, *sigB*, *sae*, *arl*, and *sarA* homologues in an interactive manner for some of the loci.[74,75] Among these, *agr* is known to involve a peptide-regulated quorum sensing.

5.11.6.1 The *agr* Regulon

The *agr* locus is the key player in the staphylococcal virulence and also is one of the best-studied peptide-regulated quorum sensing systems in G+ bacteria. Two divergent operons regulated by promoters P2 and P3 are involved in this locus (**Figure 2**).[76,77] The P2-regulated operon contains four genes (*agrABCD*) encoding all components necessary for the quorum sensing network: *agrA* and *agrC* encode an RR and an HPK, respectively, that together ensemble a classic two-component regulatory system, whereas *agrD* encodes a pheromone precursor that is processed and modified to form a 7–9 cyclic peptide (often referred to as autoinducing peptide, AIP) prior to its export by the gene product of *agrC*, a membrane protein.[75] The P3-regulated operon encodes a small transcript, normally referred to as RNAIII, that itself serves as the actual effector regulating the expression of target genes.[78,79] In addition, RNAIII harbors a gene (*hld*) coding for a small peptide called δ-hemolysin that constitutes part of the staphylococcal virulence.[80]

The HPK protein AgrC has been predicted, by hydrophobicity analysis and partially confirmed by *phoA* fusions, to contain five transmembrane helices in the N-terminal domain.[81] Deletion study suggested that the last extracellular loop is involved in AIP binding that triggers a series of phosphorylation reactions and eventually results in phosphorylation of AgrA. The latter in turn binds to regulated promoters P2 and P3 to activate gene expression.[82] Interestingly, these regulated promoters each contain a pair of 9-nt direct repeats separated by a 12-nt spacer, an arrangement that also have been observed for the regulated promoters of other quorum sensing networks such as *pln* and *spp* in bacteriocin production[22,83] and *comDE* in streptococcal competence.[61] In *pln*- and *spp*-regulated promoters, the involving regulators, which share relatively high sequence homology to AgrA,[21] have been shown to bind on the pairwise DNA repeats as dimers in a cooperative manner.[28–30] Thus, it is likely that the regulator AgrA involves a similar mechanism on promoter binding to activate gene expression. In addition to AgrA, expression from P2 and P3 in the *agr* locus is controlled by at least two other transcriptional regulators, SarA and SarR, both being activators on the *agr* locus but SarR somehow acting as a repressor on *sarA* expression.[84,85] Whether AgrA, SarA, and SarR act independently or in an interactive manner on the *agr* promoters is not known.

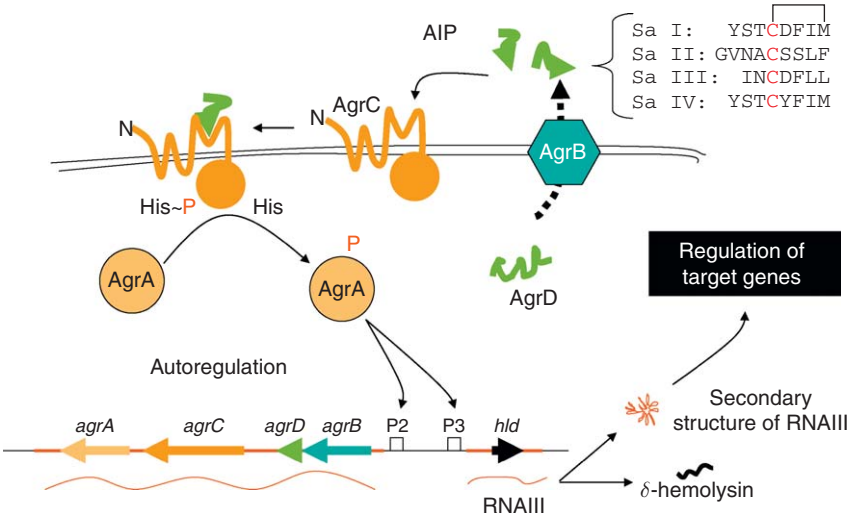


Figure 2 The *agr* regulon is depicted, showing different players and their actions in the signaling network. See text for detail.

5.11.6.2 The Effector RNAIII

Activation of the *agr* locus, which is efficiently enhanced by the *agr* autoregulatory circuit via the action of AgrA on its own promoter (P2), primarily leads to an elevated expression of the divergent P3 transcript, which in turn regulates expression of the majority of virulence genes and repression of many surface protein genes (e.g., *spa* and *coa* encoding protein A and coagulase, respectively). The regulatory function of RNAIII was discovered relatively early in the 1990s[75,79,80] but nevertheless, some detailed knowledge of the mechanism(s) by which it regulates target genes has been gained just recently. RNAIII appears to be a specific regulator. Deletion of its operon caused pleiotropic effect on several virulence factors as *agrA* mutations, and complementation with a cloned RNAIII determinant defective in δ-hemolysin production could restore the normal *agr* virulence in an RNAIII deletion mutant[80] or in an *agr*-null strain[79] evidencing that the RNAIII transcript itself, but not the gene product of *hld*, is the actual regulator in the *agr* regulon. RNAIII regulates target genes at both the transcriptional level and the translational level.[75] By computer analysis combined with chemical and enzymatic probes, the RNAIII molecule is predicted to form an extensive secondary structure containing 14 hairpin motifs connected by unpaired nucleotides.[86] Some of the hairpins (e.g., H7, H13, and H14) have a relatively high cytosine content (an unusual feature in an AT-rich organism) and these have been suggested to directly regulate translation of target genes by complementary pairing with the guanine-rich translation initiation region (TIR) occluding ribosomal binding site (RBS). For instance, both hairpins H13 and H14 have been found to be necessary for repression of the protein A gene (*spa*)[86] and the sequence of H13 shows strong complementarity to the TIR of *spa*.[87] Deletion and chemical analyses revealed that the loop of H13 could initiate base-pairing with *spa* transcript forming a complex that prevents translation initiation of *spa* and concurrently induces rapid *in vivo* mRNA degradation carried out by a double-strand-specific endoribonuclease III (RNase III).[88] Another study also suggested the pleiotropic transcriptional regulator gene *rot*, an antagonist of *agr*, to be sequestered by RNAIII and RNase III, in a similar manner as described for *spa*.[89] In some cases, RNAIII can also serve as a factor facilitating translation initiation. In the absence of RNAIII, translation of *hla* (encoding α-hemolysin) is prevented by the involvement of its RBS in an intramolecular base-pairing. Hybridization with RNAIII resolves this intramolecular complex and makes the *hla* transcript accessible for translation initiation.[78] Furthermore, the 3′end of RNAIII has been shown to have a regulatory effect on the translation of *hld*, which is part of the RNAIII transcript.[90] Translation of *hld* is delayed by 1 h following transcription and this delay is abolished by removal of the 3′-half of RNAIII.

As mentioned RNAIII has been suggested, on the basis of various deletion/mutation studies, to regulate gene expression at the transcriptional level.[79,80] Nevertheless, no study has yet reported any detailed mechanisms as to how RNAIII directly regulates transcription of target genes. It is possible that RNAIII might do so in an indirect manner, for instance, by facilitating or sequestering the translation of some selected transcriptional regulators that, in turn, directly regulate transcription of target genes, as in the case for the transcriptional regulator gene *rot* described above. On the other hand, it cannot rule out the possibility that RNAIII might regulate gene expression by allosterically modulating the activity of a gene regulator.

5.11.6.3 Autoinducing Peptides and Their Interference

agr AIPs produced by various strains of *S. aureus* are structurally well conserved.[75] They are small in size, 7–9 amino acids, and each contain a thiolactone ring structure, in which the α-carboxyl group of the C-terminal amino acid is linked to the sulfhydryl group of a cysteine, which is always located five residues before the C-terminus. Their precursors, of approximately 50 amino acids in size, are processed both N-terminally and C-terminally by at least one type I signal peptidase[91] and their cognate AgrB proteins; the latter is also needed to carry out the circularization modification (to form thiolactone ring structure) and the following export of the mature peptides.[92] AIPs from *S. aureus* have been classified into four functional groups based on their specificity,[93,94] called Sa I–IV (see **Figure 2**). Thus, bacterial strains from the different AIP groups can only be induced by their cognate AIPs but not by an AIP from other groups. Furthermore, *S. aureus* strains appear to compete with each other at the level of *agr* expression as individual AIPs inhibit *agr* expression in strains from other groups.[95] This type of bacterial interference is uncommon because it represses expression of a set of genes rather than inhibiting bacterial growth.

The nature behind autologous activation and heterologous inhibition by AIPs has been accessed in great detail by Novick and his coworkers. Using a chemical approach, an alanine screening substitution was carried out on Sa II-AIP[96] to evaluate the biological function of the different residues. Asn3 (in the N-terminal tail) and two residues within the cyclic structure (Leu8 and Phe9) were found important for intragroup activation while only the endocylic residues (Leu8 and Phe9) are important for intergroup inhibition. A truncated analogue of Sa-II AIP lacking the N-terminal tail (thus containing only the thiolactone ring structure) was found to inhibit strains from all four groups.[95] These results not only point out the importance of the N-terminal tail in autologous activation but also represent a key step in the development of a universal *agr* antagonist against staphylococcal virulence.

Linear analogues of AIPs, obtained from a synthetic approach, have no activity in *agr* activation or inhibition, suggesting that the thiolactone ring is indispensable for biological activity.[34] The involvement of the thiolactone ring structure in receptor binding is still elusive. As ester groups (as found in AIP thiolactone) are relatively good acylating agents, it has been suggested that the thiolactone group serves as an acyl donor for covalent modification of the ligand-binding sensor AgrC.[96] However, whether an acylation reaction takes place upon AIP binding by AgrC is questionable as another study showed that a synthetic AIP analogue in which the ring sulfur atom was replaced with nitrogen (forming a lactam ring that is inactive as acylating agent) still retained activation activity, albeit with an activity significantly (about 1000-fold) less than the thiolactone counterpart.[97] Interestingly, natural AIPs with a non-thiolactone ring also exist albeit outside of the species of *S. aureus*. In *S. intermedius*, which involves a similar regulation of *agr* virulence, a serine is in place of the conserved AIP cysteine, giving rise to AIPs with a lactone ring. These lactone AIPs are self-activators in *S. intermedius* and as expected, they act as inhibitors in *agr* regulation for all strains of *S. aureus*.[98]

As subtle differences in the AIP primary sequence (e.g., four functional Sa-AIP groups) and the type of ring structure (e.g., thiolactam and lactam) are important for the signaling function (both as autologous activators and as heterologous inhibitors), the dedicated protein components involved in their biosynthesis and in transmitting AIP signaling must be coevolved with the development of AIPs in order to retain the molecular specificity between the involved players. In indeed, both AgrB (involved in biosynthesis) and AgrC (in ligand binding) contain a hypervariable region in their primary sequence.[99] For AgrC, this region is located in the N-terminal membrane-located part containing the AIP-binding site. For AgrB, it is in C-terminal membrane-located half, which is responsible for group-specific processing of the substrate.[100]

5.11.7 Virulence Regulation in *Enterococcus faecalis* by Peptide Signaling

The two/three-component regulatory systems have been shown to control several virulence traits in *E. faecalis*. It has been demonstrated that both cytolycin and gelatinase biosynthesis are regulated by a PP compound.

Cytolycin: Cytolycin is a two-peptide lantibiotic that is produced by *E. faecalis* strains. In this context, it is interesting to note that though cytolycin is a cytolytic/hemolytic peptide, it also encompasses a strong antibacterial activity and is therefore often referred to as a bacteriocin.[5,6] The regulation of cytolycin biosynthesis in *E. faecalis* has been studied thoroughly.[101,102] This two-peptide lantibiotic toxin is shown to be regulated by a quorum sensing mechanism. Of the two peptides constituting the cytotoxic effect the smallest (CylL S) also serves a second purpose; it feeds back to the cytolycin producer by being the peptide inducer in a three-component regulatory system.[102]

Gelatinase: The *fsrABC* system of *E. faecalis*, a homologue to *agrABCD* in *S. aureus*, was shown to activate the virulence-related proteases, the gelatinase and serine protease[103,104] as well as biofilm formation[105] and possibly other genes related to virulence.[106] An autoinducing cyclic peptide-lactone (termed gelatinase biosynthesis-activating pheromone or GBAP, see **Table 3**) has been identified as the entity triggering a two-component regulatory system (*fsrABC*) and thereby activating expression of genes required for gelatinase production.[108]

Table 3 Peptide pheromone (PP) lactone signaling quorum sensing systems. The processed peptides shaded in yellow have to be post-translationally modified to become active

Organisms	*Peptide pheromone precursor: Autoinduction peptide (AIP or PP) sequence in yellow*	*Regulatory gene system*	*Regulation*	*References*
E. faecalis	MKFGKKIIKNVIEKRVAKVSDGVGTKPRLNQN**S**PNIFGQWMGQTEKPKKNIEK	*fsr*	Gelatinase	60
S. aureus	MNTLFNLFFDFITGILKNIGNIAAYST**C**DFIMDEVEVPKELTQLHE	*agr*	Virulence	107
L. plantarum	MKQKMYEAIAHLFKYVGAKQLVMC**C**VGIWFETKIPDELRK	*lam*	Biofilm production/sugar metabolism	36

5.11.8 Peptide Signaling Regulation of Carbohydrate Metabolism in *Lactobacillus plantarum*

In *L. plantarum* WCFS1 it has been identified an *agr*-like regulatory system, termed *lam* (*Lactobacillus agr*-like module), that is autoregulated by a cyclic peptide.[109] The inducing peptide (PP) was identified as a cyclic thiolactone pentapeptide derived from a large precursor peptide (LamD) (**Figure 3**). In addition to the autoregulation of the regulatory *lam* operon, it regulated expression of a surface polysaccharide biosynthesis gene cluster and some other cell wall and carbohydrate utilization genes.[110]

5.11.9 *agr*-Like Quorum Sensing Gene Clusters Identified in Other G+ Bacteria

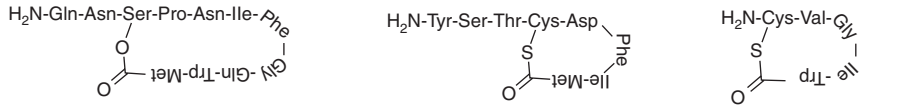
It has also been identified *agr*-like gene clusters in *Listeria monocytogenes* that seem to be involved in bacterial virulence.[111] Also in the genome of *Clostridium acetobutylicum* ATCC 824 a gene locus (*cac0078, 0079, 0080*, and *0081*) has been identified with strong homology to the *agr* system.[112] Both clusters seem to have the same organization as seen in the *lam* cluster of *L. plantarum* and in both clusters putative PP encoding *orfs* (*cac0079* and *lmo2250*) are found.

5.11.10 Pheromone-Responding Conjugative Plasmids in *Enterococcus faecalis*

Bacterial communication systems have so far been described when the signal is sensed by the organism that produces the response. However, there is a second type of bacterial communication: those in which the signal is detected solely by organisms other than the signal producer. This communication system is thoroughly investigated in the conjugative plasmid systems of *E. faecalis*.

The enterococcal, conjugative, virulence plasmids encoding a mating response to a peptide sex pheromone have been known for many years.[113] About 30 years ago the first conjugative plasmid (pAD1) was identified in a clinical isolate of *E. faecalis*[114,115] and several others have been found since (**Table 4**).

The enterococcal mating pheromone systems engage a donor cell and a recipient cell. The recipient cell produces a mating PP that leads to a response of the donor cell in an activation of its conjugative genes encoded by a plasmid that also leads to cell aggregation of the cells involved in conjugation. Consequently, an efficient horizontal transfer of the plasmid to the recipient will take place. The pheromone-responsive plasmid is efficiently transferred to pheromone producing cells via conjugation. A number of such plasmids have been



FsrA-derived PP (AIP or GBAB) AgrD-derived PP (AIP) LamD-derived PP (AIP)

Figure 3 Secondary structure of FsrA-derived PP, AgrD-derived PP, and LamD-derived PP.

Table 4 Conjugal plasmids and their peptide pheromones in *Enterococcus faecalis*

Conjugal plasmid	*Pheromone signal sequence*[a]	*Designation of the active pheromone*	*Size of precursor (aa)*	*Reference(s)*
pAD1	LFSLVLAG	cAD1	308	116
pCF10	LVTLVFV	cCF10	275	107, 117
pPD1	FLVMLSG	cPD1	234	118
pAM373	AIFILAS	cAM373	166	119, 120
pOB1	VAVLVLGA	cOB1	272	121

[a] The pheromone amino acids sequences are modified into a peptide lactone.

characterized. It has been shown that plasmid-free cells are able to secrete different pheromones and at least six specific for different donor plasmids;[113] each pheromone is specific for a particular plasmid or group of plasmids.[122] Different peptide propheromone amino acid sequences are shown in **Table 4**.

The PPs involved in conjugation are small peptides corresponding to 7–8 amino acid residues. They are derived from a larger peptide encoded by the bacterial chromosome of recipient. The precursor of these PPs is part of signal sequences of proteins that apparently are surface lipoproteins. The processing of the signal sequence (21–24 amino acid residues long) results in the pheromone signals derived from the C-terminal part. The critical event in the upregulated expression of the genes responsible for the conjugation of plasmid in the donor cells is the binding of the internalized PP to the transcription regulator. The structure of the regulator (PrgX) in the pCF10 plasmid has been determined. It has been shown earlier to be a negative regulator and it has been proposed to act as a tetramer in a similar manner as the C_I repressor of λ phage and by the binding of its dedicated pheromone (cCF10) PrgX is released and thereby activates the conjugative genes of pCF10[123]

5.11.11 Perspectives

Two-component regulatory systems are abundant among G+ bacteria and among these the peptide sensing systems constitute a minor but important subgroup. However, the peptide signaling system may be underestimated for various reasons. Such peptides can be encoded by small genes that may not be discovered during annotation of genome sequences or they could be difficult to identify if they are formed by processing and modification from larger proteins.

Regulation of virulence as well as conjugal transfer of DNA by peptide signaling seems to be important in many G+ bacteria including bacilli that have not been discussed in this review. Bacteriocin production in LAB is considered as a positive trait in environments such as food and alimentary canal and they are often regulated by PP-dependent three-component regulatory systems.

Bacterial genome sequences have been scrutinized for putative two-component regulatory systems, and numerous such regulatory genes are identified in G+ bacteria. Just to mention the analyses/annotation of a few G+ genome sequences: Nine two-component systems are found in the *L. plantarum* WCFS1 genome, 11 in *E. faecalis* V583 (a total of 18 RR),[124,125] *S. aureus* NU50 harbors at least seven regulator pairs, and *S. pneumonia* and *B. subtilis* genomes encode at least 13 and 29 putative HPK–RR regulator pairs, respectively. Though not many *orfs* encoding putative PPs have been identified so far, further analysis supported by experiments may identify new peptide-signaling dependent regulations. This was recently performed in *L. plantarum* WCFS1.[126] Due to their importance in regulation of pathogenic factors as well as positive properties of bacteria, it is important to learn more about the extend they exist and what genetic traits they control.

Abbreviations

AIP autoinducing peptide
CSP competence stimulating peptide
HPK histidine protein kinase
PP peptide pheromone
RR response regulator
TMS transmembrane segment

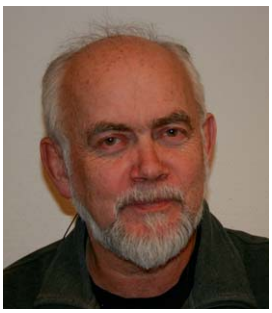
References

1. M. Kleerebezem; W. M. de Vos; O. P. Kuipers, The Lantibiotics Nisin and Subtilin Act as Extracellular Regulators of Their Own Biosynthesis. In *Cell–Cell Signaling in Bacteria*; G. M. Dunny, S. C. Winans, Eds.; ASM Press: Washington, DC, 1999; pp 159–174.
2. M. Kleerebezem; L. E. Quadri; O. P. Kuipers; W. M. de Vos, *Mol. Microbiol.* **1997**, *24*, 895–904.
3. I. F. Nes; D. B. Diep; L. S. Havarstein; M. B. Brurberg; V. Eijsink; H. Holo, *Antonie Van Leeuwenhoek* **1996**, *70*, 113–128.
4. I. F. Nes; V. G. H. Eijsink, Regulation of Group II Peptide Bacteriocin Synthesis by Quorum-Sensing Mechanisms. In *Cell–Cell Signaling in Bacteria*; G. M. Dunny, S. C. Winans, Eds.; ASM Press: Washington, DC, 1999; pp 175–192.
5. O. P. Kuipers; M. M. Beerthuyzen; P. G. de Ruyter; E. J. Luesink; W. M. de Vos, *J. Biol. Chem.* **1995**, *270*, 27299–27304.
6. D. B. Diep; L. S. Havarstein; I. F. Nes, *Mol. Microbiol.* **1995**, *18*, 631–639.
7. M. Kleerebezem; J. Boekhorst; R. van Kranenburg; D. Molenaar; O. P. Kuipers; R. Leer; R. Tarchini; S. A. Peters; H. M. Sandbrink; M. W. Fiers; W. Stiekema; R. M. Lankhorst; P. A. Bron; S. M. Hoffer; M. N. Groot; R. Kerkhoven; M. de Vries; B. Ursing; W. M. de Vos; R. J. Siezen, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1990–1995.
8. A. Maldonado; R. Jimenez-Diaz; J. L. Ruiz-Barba, *J. Bacteriol.* **2004**, *186*, 1556–1564.
9. M. B. Brurberg; I. F. Nes; V. G. Eijsink, *Mol. Microbiol.* **1997**, *26* (2), 347–360.
10. D. B. Diep; L. Axelsson; C. Grefsli; I. F. Nes, *Microbiology* **2000**, *146* (Part 9), 2155–2160.
11. L. E. Quadri; M. Kleerebezem; O. P. Kuipers; W. M. de Vos; K. L. Roy; J. C. Vederas; M. E. Stiles, *J. Bacteriol.* **1997**, *179*, 6163–6171.
12. M. J. van Belkum; D. J. Derksen; C. M. Franz; J. C. Vederas, *Microbiology* **2007**, *153*, 3660–3666.
13. T. Nilsen; I. F. Nes; H. Holo, *J. Bacteriol.* **1998**, *180* (7), 1848–1854.
14. L. Fontaine; C. Boutry; E. Guedon; A. Guillot; M. Ibrahim; B. Grossiord; P. Hols, *J. Bacteriol.* **2007**, *189*, 7195–7205.
15. A. Vaughan; V. G. Eijsink; D. Van Sinderen, *Appl. Environ. Microbiol.* **2003**, *69*, 7194–7203.
16. L. J. Gursky; N. I. Martin; D. J. Derksen; M. J. van Belkum; K. Kaur; J. C. Vederas; M. E. Stiles; L. M. McMullen, *Arch. Microbiol.* **2006**, *186*, 317–325.
17. A. E. Dobson; R. B. Sanozky-Dawes; T. R. Klaenhammer, *J. Appl. Microbiol.* **2007**, *103*, 1766–1778.
18. D. B. Diep; L. Godager; D. Brede; I. F. Nes, *Microbiology* **2006**, *152*, 1649–1659.
19. S. Flynn; D. van Sinderen; G. M. Thornton; H. Holo; I. F. Nes; J. K. Collins, *Microbiology* **2002**, *148*, 973–984.
20. F. C. Petersen; G. Fimland; A. A. Scheie, *Mol. Microbiol.* **2006**, *61*, 1322–1334.
21. D. B. Diep; L. S. Havarstein; J. Nissen-Meyer; I. F. Nes, *Appl. Environ. Microbiol.* **1994**, *60*, 160–166.
22. D. B. Diep; L. S. Havarstein; I. F. Nes, *J. Bacteriol.* **1996**, *178*, 4472–4483.
23. D. B. Diep; R. Myhre; O. Johnsborg; A. Aakra; I. F. Nes, *Mol. Microbiol.* **2003**, *47*, 483–494.
24. L. S. Havarstein; D. B. Diep; I. F. Nes, *Mol. Microbiol.* **1995**, *16*, 229–240.
25. L. S. Havarstein; H. Holo; I. F. Nes, *Microbiology* **1994**, *140* (Part 9), 2383–2389.
26. D. B. Diep; M. Skaugen; Z. Salehian; H. Holo; I. F. Nes, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2384–2389.
27. P. E. Kristiansen; G. Fimland; D. Mantzilas; J. Nissen-Meyer, *J. Biol. Chem.* **2005**, *280*, 22945–22950.
28. P. A. Risoen; M. B. Brurberg; V. G. Eijsink; I. F. Nes, *Mol. Microbiol.* **2000**, *37*, 619–628.
29. P. A. Risoen; L. S. Havarstein; D. B. Diep; I. F. Nes, *Mol. Gen. Genet.* **1998**, *259*, 224–232.
30. P. A. Risoen; O. Johnsborg; D. B. Diep; L. Hamoen; G. Venema; I. F. Nes, *Mol. Genet. Genomics* **2001**, *265*, 198–206.
31. O. Johnsborg; P. E. Kristiansen; T. Blomqvist; L. S. Havarstein, *J. Bacteriol.* **2006**, *188*, 1744–1749.
32. T. W. Grebe; J. B. Stock, *Adv. Microb. Physiol.* **1999**, *41*, 139–227.
33. M. Lyristis; A. E. Bryant; J. Sloan; M. M. Awad; I. T. Nisbet; D. L. Stevens; Rood J. I, *Mol. Microbiol.* **1994**, *12*, 761–777.
34. G. Ji; R. C. Beavis; R. P. Novick, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12055–12059.
35. G. J. Lyon; R. P. Novick, *Peptides* **2004**, *25*, 1389–1403.
36. J. B. Stock; M. G. Surett; M. Levit; P. Park, Two-Component Signal Transduction System: Structure–Function Relationship and Mechanisms of Catalysis. In *Two-Component Signal Transduction*; J. A. Hoch, T. J. Silhavy, Eds.; ASM Press: Washington, DC, 1995; pp 25–51.
37. O. Johnsborg; T. Blomqvist; M. Kilian; L. Havarstein, Biologically Active Peptides in Streptococci. In *Molecular Biology of Streptococci*; R. Hakenbeck, S. Chatwal, Eds.; Horizon Scientific Press: Norfolk, VA, 2007; pp 25–59.
38. O. Johnsborg; D. B. Diep; I. F. Nes, *J. Bacteriol.* **2003**, *185*, 6913–6920.
39. O. Johnsborg; L. H. Godager; I. F. Nes, *Arch. Microbiol.* **2004**, *182*, 450–457.
40. D. Straume; M. Kjos; I. F. Nes; D. B. Diep, *Mol. Genet. Genomics* **2007**, *278*, 283–293.
41. M. Kleerebezem, *Peptides* **2004**, *25*, 1405–1414.

42. M. Kleerebezem; R. Bongers; G. Rutten; W. M. de Vos; O. P. Kuipers, *Peptides* **2004**, *25*, 1415–1424.
43. T. Stein; S. Borchert; P. Kiesau; S. Heinzmann; S. Kloss; C. Klein; M. Helfrich; K. D. Entian, *Mol. Microbiol.* **2002**, *44*, 403–416.
44. M. Upton; J. R. Tagg; P. Wescombe; H. F. Jenkinson, *J. Bacteriol.* **2001**, *183*, 3931–3938.
45. P. A. Wescombe; J. R. Tagg, *Appl. Environ. Microbiol.* **2003**, *69*, 2737–2747.
46. S. Schmitz; A. Hoffmann; C. Szekat; B. Rudd; G. Bierbaum, *Appl. Environ. Microbiol.* **2006**, *72*, 7270–7277.
47. J. P. Claverys; L. S. Havarstein, *Front. Biosci.* **2002**, *7*, d1798–d1814.
48. O. Johnsborg; V. Eldholm; L. S. Havarstein, *Res. Microbiol.* **2007**, *158*, 767–778.
49. M. H. Dawson; R. H. Sia, *J. Exp. Med.* **1931**, *54*, 681–699.
50. A. Tomasz, *Nature* **1965**, *208*, 155–159.
51. A. Tomasz; R. D. Hotchkiss, *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *51*, 480–487.
52. A. Tomasz; J. L. Mosser, *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 58–66.
53. L. S. Havarstein; G. Coomaraswamy; D. A. Morrison, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11140–11144.
54. G. Pozzi; L. Masala; F. Iannelli; R. Manganelli; L. S. Havarstein; L. Piccoli; D. Simon; D. A. Morrison, *J. Bacteriol.* **1996**, *178*, 6087–6090.
55. F. M. Hui; D. A. Morrison, *J. Bacteriol.* **1991**, *173*, 372–381.
56. F. M. Hui; L. Zhou; D. A. Morrison, *Gene* **1995**, *153*, 25–31.
57. L. S. Havarstein; P. Gaustad; I. F. Nes; D. A. Morrison, *Mol. Microbiol.* **1996**, *21*, 863–869.
58. E. V. Pestova; L. S. Havarstein; D. A. Morrison, *Mol. Microbiol.* **1996**, *21*, 853–862.
59. A. Dagkessamanskaia; M. Moscoso; V. Henard; S. Guiral; K. Overweg; M. Reuter; B. Martin; J. Wells; J. P. Claverys, *Mol. Microbiol.* **2004**, *51*, 1071–1086.
60. S. N. Peterson; C. K. Sung; R. Cline; B. V. Desai; E. C. Snesrud; P. Luo; J. Walling; H. Li; M. Mintz; G. Tsegaye; P. C. Burr; Y. Do; S. Ahn; J. Gilbert; R. D. Fleischmann; D. A. Morrison, *Mol. Microbiol.* **2004**, *51*, 1051–1070.
61. O. Ween; P. Gaustad; L. S. Havarstein, *Mol. Microbiol.* **1999**, *33*, 817–827.
62. P. Luo; H. Li; D. A. Morrison, *Mol. Microbiol.* **2003**, *50*, 623–633.
63. P. Luo; D. A. Morrison, *J. Bacteriol.* **2003**, *185*, 349–358.
64. H. Ichihara; K. Kuma; H. Toh, *J. Bacteriol.* **2006**, *188*, 6429–6434.
65. S. Guiral; T. J. Mitchell; B. Martin; J. P. Claverys, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8710–8715.
66. L. Kausmally; O. Johnsborg; M. Lunde; E. Knutsen; L. S. Havarstein, *J. Bacteriol.* **2005**, *187*, 4338–4345.
67. H. Steinmoen; E. Knutsen; L. S. Havarstein, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7681–7686.
68. H. Steinmoen; A. Teigen; L. S. Havarstein, *J. Bacteriol.* **2003**, *185*, 7176–7183.
69. M. Moscoso; J. P. Claverys, *Mol. Microbiol.* **2004**, *54*, 783–794.
70. L. S. Havarstein; B. Martin; O. Johnsborg; C. Granadel; J. P. Claverys, *Mol. Microbiol.* **2006**, *59*, 1297–1307.
71. J. P. Claverys; B. Martin; L. S. Havarstein, *Mol Microbiol.* **2007**, *64*, 1423–1433.
72. J. Majewski; P. Zawadzki; P. Pickerill; F. M. Cohan; C. G. Dowson, *J. Bacteriol.* **2000**, *182*, 1016–1023.
73. F. D. Lowy, *N. Engl. J. Med.* **1998**, *339*, 520–532.
74. B. Said-Salim; P. M. Dunman; F. M. McAleese; D. Macapagal; E. Murphy; P. J. McNamara; S. Arvidson; T. J. Foster; S. J. Projan; B. N. Kreiswirth, *J. Bacteriol.* **2003**, *185*, 610–619.
75. R. P. Novick, *Mol. Microbiol.* **2003**, *48*, 1429–1449.
76. E. Morfeldt; L. Janzon; S. Arvidson; S. Lofdahl, *Mol. Gen. Genet.* **1988**, *211*, 435–440.
77. P. Recsei; B. Kreiswirth; M. O'Reilly; P. Schlievert; A. Gruss; R. P. Novick, *Mol. Gen. Genet.* **1986**, *202*, 58–61.
78. E. Morfeldt; K. Tegmark; S. Arvidson, *Mol. Microbiol.* **1996**, *21*, 1227–1237.
79. R. P. Novick; H. F. Ross; S. J. Projan; J. Kornblum; B. Kreiswirth; S. Moghazeh, *EMBO J.* **1993**, *12*, 3967–3975.
80. L. Janzon; S. Arvidson, *EMBO J.* **1990**, *9*, 1391–1399.
81. G. Lina; S. Jarraud; G. Ji; T. Greenland; A. Pedraza; J. Etienne; R. P. Novick; Vandenesch F, *Mol. Microbiol.* **1998**, *28*, 655–662.
82. R. L. Koenig; J. L. Ray; S. J. Maleki; M. S. Smeltzer; B. K. Hurlburt, *J. Bacteriol.* **2004**, *186*, 7549–7555.
83. V. G. Eijsink; M. B. Brurberg; P. H. Middelhoven; I. F. Nes, *J. Bacteriol.* **1996**, *178* (8), 2232–2237.
84. Y. Liu; A. C. Manna; C. H. Pan; I. A. Kriksunov; D. J. Thiel; A. L. Cheung; G. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2392–2397.
85. A. C. Manna; A. L. Cheung, *Mol. Microbiol.* **2006**, *60*, 1289–1301.
86. Y. Benito; F. A. Kolb; P. Romby; G. Lina; J. Etienne; F. Vandenesch, *RNA* **2000**, *6*, 668–679.
87. S. Boisset; T. Geissmann; E. Huntzinger; P. Fechter; N. Bendridi; M. Possedko; C. Chevalier; A. C. Helfer; Y. Benito; A. Jacquier; C. Gaspin; F. Vandenesch; P. Romby, *Genes Dev.* **2007**, *21*, 1353–1366.
88. E. Huntzinger; S. Boisset; C. Saveanu; Y. Benito; T. Geissmann; A. Namane; G. Lina; J. Etienne; B. Ehresmann; C. Ehresmann; A. Jacquier; F. Vandenesch; P. Romby, *EMBO J.* **2005**, *24*, 824–835.
89. E. Geisinger; R. P. Adhikari; R. Jin; H. F. Ross; R. P. Novick, *Mol. Microbiol.* **2006**, *61*, 1038–1048.
90. N. Balaban; R. P. Novick, *FEMS Microbiol. Lett.* **1995**, *133*, 155–161.
91. J. S. Kavanaugh; M. Thoendel; A. R. Horswill, *Mol. Microbiol.* **2007**, *65*, 780–798.
92. L. Zhang; L. Gray; R. P. Novick; G. Ji, *J. Biol. Chem.* **2002**, *277*, 34736–34742.
93. S. Jarraud; G. J. Lyon; A. M. Figueiredo; L. Gerard; F. Vandenesch; J. Etienne; T. W. Muir; R. P. Novick, *J. Bacteriol.* **2000**, *182*, 6517–6522.
94. G. Ji; R. Beavis; R. P. Novick, *Science* **1997**, *276*, 2027–2030.
95. G. J. Lyon; P. Mayville; T. W. Muir; R. P. Novick, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13330–13335.
96. P. Mayville; G. Ji; R. Beavis; H. Yang; M. Goger; R. P. Novick; T. W. Muir, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1218–1223.
97. P. M. Dowell; Z. Affas; C. Reynolds; M. T. Holden; S. J. Wood; S. Saint; A. Cockayne; P. J. Hill; C. E. Dodd; B. W. Bycroft; W. C. Chan; P. Williams, *Mol. Microbiol.* **2001**, *41*, 503–512.
98. G. Ji; W. Pei; L. Zhang; R. Qiu; J. Lin; Y. Benito; G. Lina; R. P. Novick, *J. Bacteriol.* **2005**, *187*, 3139–3150.
99. P. Dufour; S. Jarraud; F. Vandenesch; T. Greenland; R. P. Novick; M. Bes; J. Etienne; G. Lina, *J. Bacteriol.* **2002**, *184*, 1180–1186.
100. L. Zhang; G. Ji, *J. Bacteriol.* **2004**, *186*, 6706–6713.

101. M. C. Booth; C. P. Bogie; H. G. Sahl; R. J. Siezen; K. L. Hatter; M. S. Gilmore, *Mol. Microbiol.* **1996**, *21*, 1175–1184.
102. W. Haas; B. D. Shepard; M. S. Gilmore, *Nature* **2002**, *415*, 84–87.
103. X. Qin; K. V. Singh; G. M. Weinstock; B. E. Murray, *J. Bacteriol.* **2001** *183*, 3372–3382.
104. X. Qin; K. V. Singh; G. M. Weinstock; B. E. Murray, *Infect Immun.* **2000**, *68*, 2579–2586.
105. L. E. Hancock; M. Perego, *J. Bacteriol.* **2004**, *186*, 5629–5639.
106. A. Bourgogne; S. G. Hilsenbeck; G. M. Dunny; B. E. Murray, *J. Bacteriol.* **2006**, *188*, 2875–2884.
107. M. Mori; Y. Sakagami; Y. Ishii; A. Isogai; C. Kitada; M. Fujino; J. C. Adsit; G. M. Dunny; A. Suzuki, *J. Biol. Chem.* **1988**, *263*, 14574–14578.
108. J. Nakayama; S. Chen; N. Oyama; K. Nishiguchi; E. A. Azab; E. Tanaka; R. Kariyama; K. Sonomoto, *J. Bacteriol.* **2006**, *188*, 8321–8326.
109. M. H. Sturme; J. Nakayama; D. Molenaar; Y. Murakami; R. Kunugi; T. Fujii; E. E. Vaughan; M. Kleerebezem; W. M. de Vos, *J. Bacteriol.* **2005**, *187*, 5224–5235.
110. M. H. J. Sturme; J. Nakayama; D. Molenaar; Y. Murakami; R. Kunugi; T. Fujii; E. E. Vaughan; M. Kleerebezem; W. M. de Vos, *J. Bacteriol.* **2005**, *187*, 5224–5235.
111. N. Autret; C. Raynaud; I. Dubail; P. Berche; A. Charbit, *Infect. Immun.* **2003**, *71*, 4463–4471.
112. J. Nölling; G. Breton; M. V. Omelchenko; K. S. Makarova; Q. Zeng; R. Gibson; H. M. Lee; J. Dubois; D. Qiu; J. Hitti; Y. I. Wolf; R. L. Tatusov; F. Sabathe; L. Doucette-Stamm; P. Soucaille; M. J. Daly; G. N. Bennett; E. V. Koonin; D. R. Smith, *J. Bacteriol.* **2001**, *183*, 4823–4838.
113. D. B. Clewell; G. M. Dunny, Conjugation and Genetic Exchange in Enterococci. In *The Enterococci. Pathogenesis, Molecular Biology, and Antibiotic Resistance*; M. S. Gilmore, Ed.; ASM Press: Washington, DC, 2002; pp 265–300.
114. G. M. Dunny; B. L. Brown; D. B. Clewell, *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3479–3483.
115. P. K. Tomich; F. Y. An; S. P. Damle; D. B. Clewell, *Antimicrob. Agents Chemother.* **1979**, *15*, 828–830.
116. M. Mori; Y. Sakagami; M. Narita; A. Isogai; M. Fujino; C. Kitada; R. A. Craig; D. B. Clewell; A. Suzuki, *FEBS Lett.* **1984**, *178*, 97–100.
117. M. H. Antiporta; G. M. Dunny, *J. Bacteriol.* **2002**, *184*, 1155–1162.
118. A. Suzuki; M. Mori; Y. Sakagami; A. Isogai; M. Fujino; C. Kitada; R. A. Craig; D. B. Clewell, *Science* **1984**, *226*, 849–850.
119. S. E. Flannagan; D. B. Clewell, *Mol. Microbiol.* **2002**, *44*, 803–817.
120. M. Mori; H. Tanaka; Y. Sakagami; A. Isogai; M. Fujino; C. Kitada; B. A. White; F. Y. An; D. B. Clewell; A. Suzuki, *FEBS Lett.* **1986**, *206*, 69–72.
121. J. Nakayama; Y. Abe; Y. Ono; A. Isogai; A. Suzuki, *Biosci. Biotechnol. Biochem.* **1995**, *59*, 703–705.
122. G. M. Dunny; M. H. Antiporta; H. Hirt, *Peptides* **2001**, *22*, 1529–1539.
123. K. Shi; C. K. Brown; Z. Y. Gu; B. K. Kozlowicz; G. M. Dunny; D. H. Ohlendorf; C. A. Earhart, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18596–18601.
124. L. E. Hancock; M. Perego, *J. Bacteriol.* **2004**, *186*, 7951–7958.
125. F. Teng; E. C. Nannini; B. E. Murray, *J. Infect. Dis.* **2005**, *191*, 472–480.
126. M. H. J. Sturme; C. Francke; R. J. Siezen; W. M. de Vos; M. Kleerebezem, *Microbiology* **2007**, *153*, 3939–3947.

Biographical Sketches



Ingolf F. Nes graduated from University of Bergen in 1973. He has during the last 20 years been working mainly with antimicrobial peptides from Gram-positive bacteria as well as bacterial gene regulation. He has been professor in biotechnology at the Norwegian University of Life Sciences since 1993.

Ola Johnsborg is a research scientist. He received his Ph.D. in 2003 from the Norwegian University of Life Sciences. His Ph.D. work was on molecular mechanisms underlying regulation of bacteriocin production in *Lactobacillus plantarum*. Since then he has been studying horizontal gene transfer by means of competence for natural transformation in the human pathogenic bacterium *Streptococcus pneumoniae*.

Dzung B. Diep received his Ph.D. in molecular microbiology in 1996, from the Norwegian University of Life Sciences (UMB). He spent 2 years as post doc in University of Victoria (Canada) and then 2 years in University in Oslo. In 2008 he was appointed professor in Microbiology at UMB.

His major scientific interests are bacterial virulence and various aspects of bacteriocins and toxins including quorum sensing regulation, mode of action, and mechanisms involved in immunity and receptor targeting.

5.12 A New Generation of Artificial Enzymes: Catalytic Antibodies or 'Abzymes'

Rémy Ricoux and Jean-Pierre Mahy, Université de Paris-sud 11, Orsay Cedex, France

© 2010 Elsevier Ltd. All rights reserved.

5.12.1 Introduction

In the quest for reactions that would allow the synthesis of more or less complex target molecules that are of huge interest in industry, fine chemistry, or pharmacology, the chemist is always looking for new catalysts that are able to perform some key reactions, most efficiently and selectively as possible, under mild pressure and temperature conditions. This requires from the chosen entity, not only catalytic properties, but also acute molecular recognition properties. These properties are typical of enzymes that catalyze under biological conditions a wide range of reactions and are able to convert specifically their substrate, which is extremely precisely positioned in their active site thanks to a network of multiple interactions with the side chains of its amino acids: hydrogen bonds, van der Waals, and electrostatic and hydrophobic interactions. Conscious about this, since about 20 years, the chemist does not hesitate anymore to use enzymes as catalysts in organic synthesis (bioconversions), and numerous works are now directed either toward the search for new activities in extreme media (extremophiles) or toward the modulation or the creation of new activities by directed evolution following genetic or genomic approaches.

On the other hand, when not using enzymes themselves, the researchers try to elaborate systems that are able to reproduce as well as possible their properties. So original systems, elaborated from antibodies, were born during these last 15 years: the catalytic antibodies or abzymes (AB: antibody, ZYME: enzyme). Indeed, scientists have realized that the immune system is a rich source of intriguing and highly efficient catalysts for common organic synthesis reactions. These catalysts are antibodies that have been identified in the immune system using small molecules known as haptens. The hapten is an organic molecule that is rationally designed for a given targeted chemical reaction in the hope that the antibody it elicits will catalyze that reaction. Antibodies that perform the desired catalysis are then identified as catalytic antibodies or 'abzymes.' Until now, catalytic antibodies have been shown to catalyze a wide range of various chemical

processes, inducing specificity, stereoselectivity, and even the ability to route a reaction through a disfavored chemical pathway. Some antibodies have even been discovered for reactions for which there exists no known natural enzyme. Many steps are required for the elaboration of a successful catalytic antibody for a given chemical transformation. The rational design of the hapten that will be used for immunization is a crucial step, and several strategies, which will be presented in the followings paragraphs of this chapter, have been developed. One of the most widely applied ones has been the use of antigens that are designed as transition state analogues (TSAs) of the target reactions. Alternatively, haptens carrying a point charge have been employed in order to recruit a complementary charged amino acid in the antibody active site to perform catalysis, which has been named the 'bait and switch' strategy. Recently, reactive immunization has been developed as another powerful strategy to provide a means for the selection of antibody catalysts *in vivo* on the basis of their reactivity.

Once the hapten has been designed and prepared, it is conjugated with a carrier protein to induce the best immunogenicity as possible to elicit an immune response in the animal (most commonly a mouse) in which it is inoculated. The antibodies produced by the defense mechanism of the adaptive immune system that specifically recognizes the hapten are then isolated, overproduced, and purified for testing their catalytic activity toward the targeted chemical reaction.

A good number of reviews were already dedicated to these new biocatalysts[1–8] and this chapter will focus on recently discovered catalytic antibody generated against haptens designed by a variety of strategies (*vide supra*). It will also constitute a chapter at the address of the chemists, to describe the concept of the catalytic antibodies, the diverse technological evolutions envisaged to go toward the best possible performances, and, in the light of the obtained results and of the structural knowledge and already realized mechanistic studies, what can the chemists hope from these abzymes.

5.12.2 The Concept of Catalytic Antibodies or Abzymes

Antibodies or immunoglobulins (Igs) are glycoproteins of about 150 kDa produced by the immune system cells, such as mature B lymphocytes, in response to an aggression of the organism by an exogenous agent or antigen: parasite, bacteria, virus, protein, and so on. They bind tightly to these antigens and precipitate them as an antibody–antigen immune complex that is formed thanks to the specific recognition of a part of the antigen, or epitope, by the N-terminal variable part of the antibody or paratope. The process of recognition of an antigen by an antibody is strongly analogous to that of a substrate by an enzyme and occurs thanks to the same kind of network of multiple interactions: hydrogen bonds, van der Waals, and electrostatic and hydrophobic interactions between the antigen and the side chains of the amino acids that build up the binding site of the antibody. This analogy thus made antibodies a tool of choice for the elaboration of new biocatalysts, with the only condition that was to know how to resolve this key problem: How to transform an antibody capable of fixing a molecule in its most stable state, the antigen (or hapten in the case of a small molecule), into an abzyme capable not only of fixing a hapten but also of catalyzing its selective transformation? The answer to this question came in three steps. Indeed, as early as in 1947, Linus Pauling expressed the principle of enzymatic catalysis,[9] in which he suggested that enzymes could accelerate the rate of the reactions by binding at their active site their transition state with a better affinity than the substrate in its fundamental state (**Figure 1**) and, by doing so, decreasing the energy of activation.

Being inspired by this principle, Jenks suggested then in 1969 that by generating antibodies raised against a stable analogue of the TS of the reaction that one wished to catalyze, one could obtain antibodies endowed with catalytic activity.[10] It was necessary to wait for another 17 years for this hypothesis to be independently demonstrated by two Californian groups, the groups of Lerner[11] and Schultz,[12] who for the first time obtained antibodies capable of accelerating the hydrolysis of esters. Thus, they showed that antibodies raised against phosphonates, stable analogues of the tetrahedral TS formed during the hydrolysis of esters (**Figure 2**), could catalyze this reaction. The observed kinetics were of Michaëlien type, charaterized by values of $k_{cat} = 4.5 \times 10^{-3}\,s^{-1}$ and of $K_M = 260\,\mu mol\,l^{-1}$; however, the acceleration factors, k_{cat}/k_{uncat}, remained lower than 10^3.

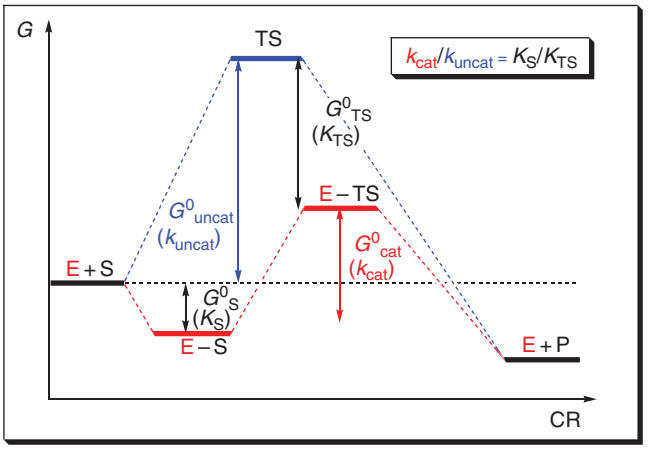


Figure 1 Energy profile of a chemical reaction with (- - -, red) or without enzymatic catalysis (- - -, blue).

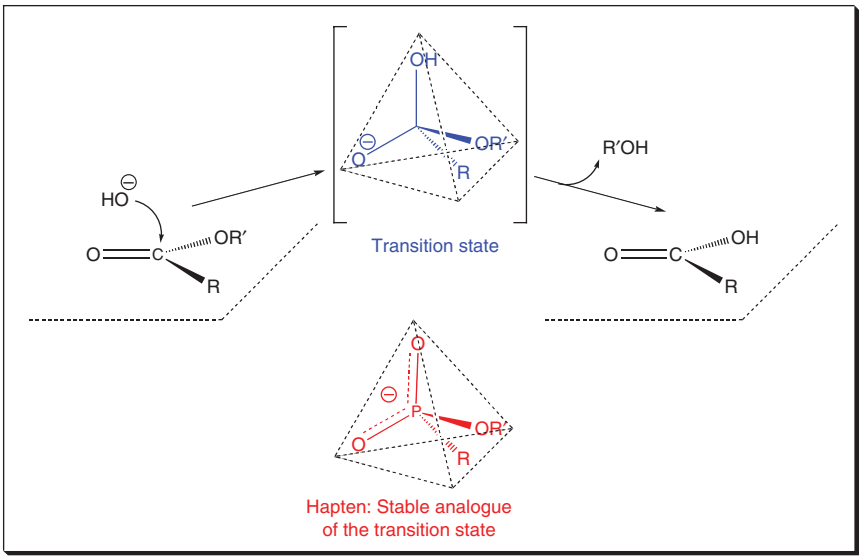


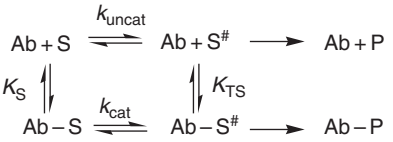
Figure 2 Ester hydrolysis. Design of a stable analogue of the tetrahedral transition state to be used as a hapten to generate catalytic antibodies with an esterase activity.

5.12.3 Catalytic Antibodies Generated against Transition State Analogues

Following the first successful examples of catalytic antibodies raised against haptens as transition state analogues (TSAs) reported by Lerner and Schultz, the TSA approach has been applied in a large number of studies in order to generate new biocatalysts for many chemical transformations. According to the transition state theory,[13–16] the catalytic efficiency (k_{cat}/k_{uncat}) of a given enzymatic reaction can be deduced from the thermodynamic cycle (**Scheme 1**) under ideal conditions.[16,17]

From this cycle, the following equations can be established

$$\frac{K_{TS}}{k_{uncat}} = \frac{K_S}{k_{cat}} \Rightarrow \frac{k_{cat}}{k_{uncat}} = \frac{K_S}{K_{TS}}$$



Scheme 1

in which K_{TS} represents the dissociation constant for the antibody–transition state complex, k_{uncat} represents the rate constant for the uncatalyzed reaction, K_S represents the dissociation constant for the antibody–substrate complex, and k_{cat} the rate constant for the reaction in the presence of the catalytic antibody. If the hapten used for immunization is a good analogue of the TS of the target reaction, the rate enhancement can be predicted from the ratio K_S/K_{TSA} where K_{TSA} represents the affinity of the antibody for the transition state analogue.

$$\frac{k_{cat}}{k_{uncat}} = \frac{K_S}{K_{TS}} = \frac{K_S}{K_{TSA}}$$

Haptens are often designed to mimic the high-energy intermediate on the chemical pathway that leads to the desired product – such haptens are referred to as transition state analogues. According to the transition state theory, a TS is a short-lived theoretical species that is believed to occur at the energetic peak in a reaction pathway, being transient with no finite lifetime. Ideally, it is this structure that a hapten should mimic. In practice, however, it is easier, and sometimes only possible, to mimic a structure that is isolable and the closest as possible to the TS, that is, a reaction intermediate. Finally, manipulating bond lengths and charge distributions may also lead to mimics of the TS, but obviously not exact replicas.

The strategy consisting in producing antibodies against TSAs was the most used by far in the 15 years that followed the results of the groups of Lerner and Schultz, and it allowed to obtain a first generation of abzymes capable of accelerating numerous reactions.[1–8] These reactions can be as well monomolecular – such as the rearrangement of Claisen of chorismate into prephenate, the oxy-Cope rearrangement of one 1,5-disubstituted hexadiene, or the isomerization of steroids – than bimolecular: hydrolysis of the carboxylic acid derivatives, esters, amides, carbonates, and so on. Among all these reactions, many are also catalyzed by enzymes and from the point of view of chemists, in particular those working in bioconversions, it is interesting to note that, as in the case of the bioconversions, in which more than 50% of the enzymes used are esterases, a wide majority of abzymes obtained until now possess a hydrolytic activity. There are however a certain number of reactions catalyzed by the abzymes for which is obtained a selectivity that is different from that observed with the enzyme, or even, there exists no equivalent enzyme. In this respect, one of the most striking examples for the chemist is unquestionably the reaction of Diels–Alder. Indeed, while a single enzyme capable of catalyzing this reaction was recently discovered,[18] not less than four abzymes were described as being able to realize this reaction, the first one of them having been obtained in 1989 by the team of Hilvert.[1–7] Indeed, from a tricyclic hapten they have been able to produce a monoclonal antibody 1E9 capable of accelerating approximately 100 times the condensation of tetrachlorothiophene dioxide with *N*-ethylmaleimide to lead to a tricyclic intermediate that decomposes into sulfur dioxide and *N*-ethyl-3,4,5,6-tetrachlorophtalimide.

5.12.3.1 Hydrolysis of Carboxylic Acid Derivatives

As already mentioned in the introduction of this chapter, ester and amide hydrolysis are known to involve high-energy tetrahedral intermediates that decompose into the corresponding carboxylic acid and alcohol or amine, respectively. This high-energy intermediate, that is close in energy to the corresponding TS, can be mimicked by several stable chemical structures such as phosphonates, phosphonamidates, arsonates, and sulfonates, all of which contain the key tetrahedral structural motif and are often employed in the design of TSA haptens for immunization. Catalytic antibodies displaying protease and esterase activities have been reviewed by Tanaka in 2002.[15] Most of these antibodies fulfill the relationship $k_{cat}/k_{uncat} = K_S/K_{TSA}$. The mechanism of antibody-catalyzed reactions is in general validated by showing that the transition state analogue used to elicit the antibody inhibits catalysis in a competitive fashion and, furthermore, binds with a higher

affinity than the corresponding substrate. Amide hydrolysis remains a difficult task due to the fact that an amine is a poor leaving group. In the design of amidase antibodies, phosphinates and phosphonamidates – which are ionized at physiological pH – came to the fore as the preferred haptens. However, few amide-cleaving antibodies have been obtained. One notable example is antibody 43C9, which not only catalyzes the hydrolysis of aromatic amides and esters, but also shows an exceptional rate acceleration (2.5×10^5 over background reaction).[19,20] Antibody 43C9 was induced with a tetrahedral TS mimic, phosphonamidate (**1**) (**Figure 3**). The unusually high rate enhancement of this antibody suggested a more intricate catalytic mechanism than mere proximity of reactive groups accompanied by modest TS stabilization. In 1994, Roberts *et al.*[21,22] constructed a computational model to investigate the structural basis for the catalytic activity of antibody 43C9. This model implicated Arg L96 in stabilizing the TS and His L91 as a neutral nitrogen nucleophile. In order to investigate the precise catalytic mechanism of 43C9, the crystallographic structure of the 43C9 scFv (single chain variable fragment) has been determined.[23] **Figure 3** presents the structure of 43C9 scFv in which the hapten **1** has been docked.These data fully support the computational model, indicating that transition state stabilization through an extensive hydrogen bonding network is important in catalysis. However, 43C9 is unique in having a nucleophile well positioned at the binding pocket, which may be the key for amide hydrolysis.

Yet, the active site of 43C9 is similar to those of other esterase antibodies, such as CNJ206, 48G7, and particularly 17E8.[23] Recently, Chong *et al.*[24] carried out quantum mechanical calculations, molecular dynamics simulations, and free energy calculations to assess the mechanism involving direct hydroxide attack for antibody 43C9, with the results supporting this mechanism. Furthermore, they suggested that this direct hydroxide attack mechanism is plausible for other antiphosphonate antibodies tailored for the hydrolysis of *para*-nitrophenyl esters. Amidase antibody 312D6 was obtained against the sulfonamide hapten **2** (**Figure 4**), which mimics the tetrahedral intermediate as well as the related TS of a distorted amide hydrolysis.[25] It appeared that even though sulfonamides adequately reproduce the geometry and conformation of tetrahedral intermediates for the base-catalyzed hydrolysis of amides, they do not provide the same charge distribution. It was speculated, however, that for highly reactive amides, a neutral hydrolysis pathway corresponding to the uncatalyzed addition of water might operate at near-neutral pH. Therefore, it is reasonable to assume that a sulfonamide is a better mimic for the neutral species involved along the hydrolytic pathway. By using the sulfonamide as a hapten, it is also anticipated that antibodies generated upon binding the substrates will force them to adopt a twisted conformation in which the highly distorted amide bond would be more susceptible toward hydrolysis. Hapten **2** was used for immunization as a KLH (keyhole limpet hemocyanin) conjugate, and hapten **3** was used for ELISA (enzyme-linked immunosorbent assay) screening as a BSA (bovine serum albumin) conjugate. Two antibodies showed hydrolytic activity above background levels, and 312D6 proved to be the best catalyst.

5.12.3.2 Cationic Cyclization

Antibody HA519A4 catalyzes the tandem cationic cyclization of a polyene substrate (**Figure 5**).[26]

To date, this antibody is the only one that has been analyzed at atomic resolution. X-ray crystallographic data of the Fab fragment of HA519A4 cocrystallized with eliciting hapten **4**, designed as a TSA, suggested that the hapten is deeply buried within a hydrophobic pocket.[27] The antibody-combining site provides a highly complementary fit as well as multiple aromatic residues. Therefore, it appears that upon binding the polyene, the active site of the antibody forces it into the productive chair–chair conformation.

Recently published crystal structures of antibody 4C6, an antibody that catalyzes another cationic cyclization reaction (**Figure 6**),[28] revealed that this antibody has exquisite shape complementarity to its eliciting hapten **5**.[29] The active site contains multiple aromatic residues which shield the high-energy intermediate from solvent and stabilize the carbocation intermediates through cation–π interactions.

5.12.3.3 Disfavored Ring Closure

One dramatic feature of antibody catalysts is their ability to reroute reaction pathways, thereby achieving disfavored chemical transformations instead of favored low-energy chemical processes. An archetypal example is the antibody-catalyzed disfavored 6-endo-intramolecular cyclization reaction of *trans*-epoxyalcohol

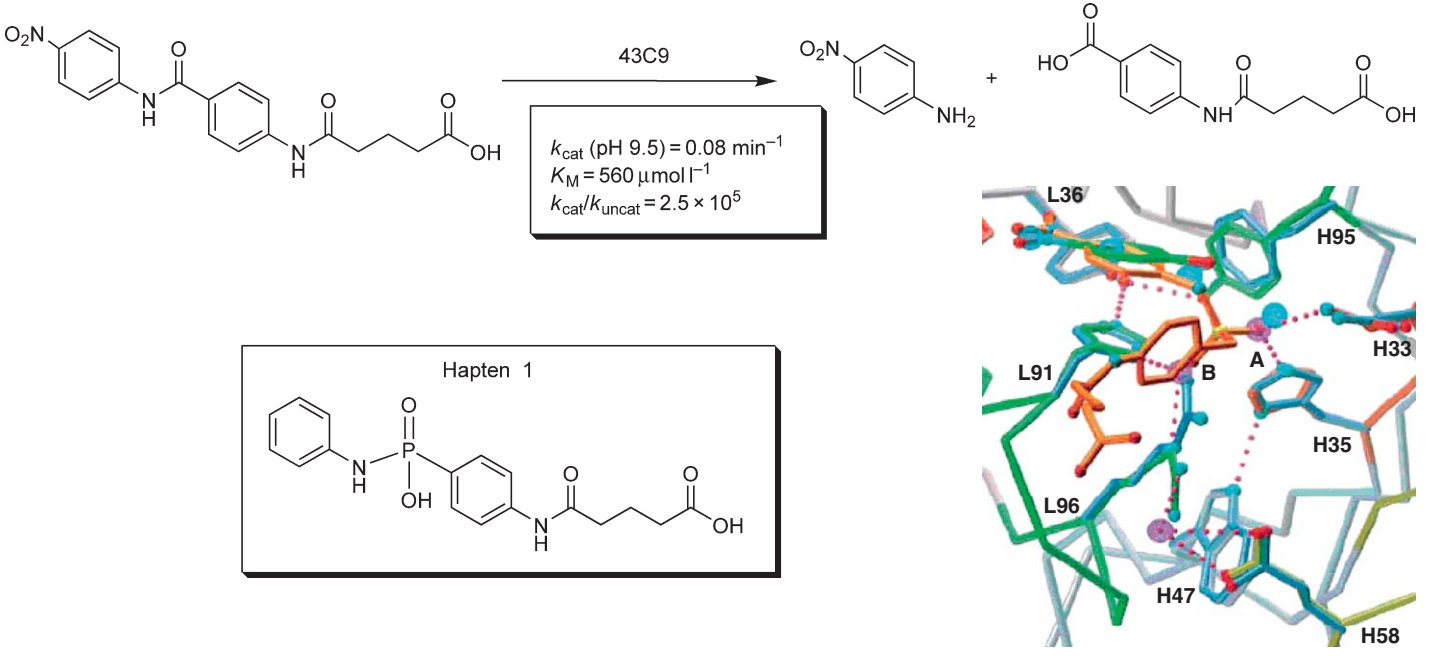


Figure 3 Amide hydrolysis catalyzed by the antibody 43C9 raised against phosphonamidate **1**.

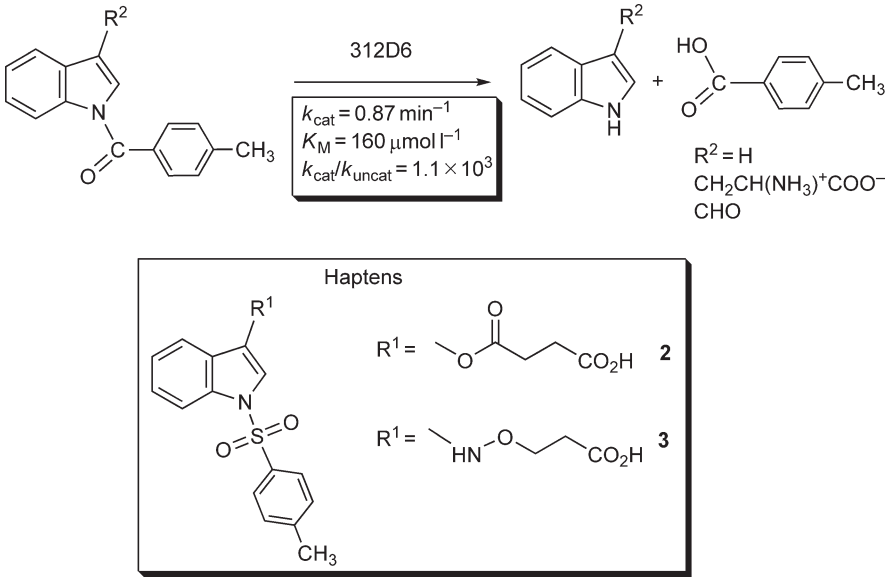


Figure 4 Amide hydrolysis catalyzed by the antibody 312D6 raised against sulfonamide **2**.

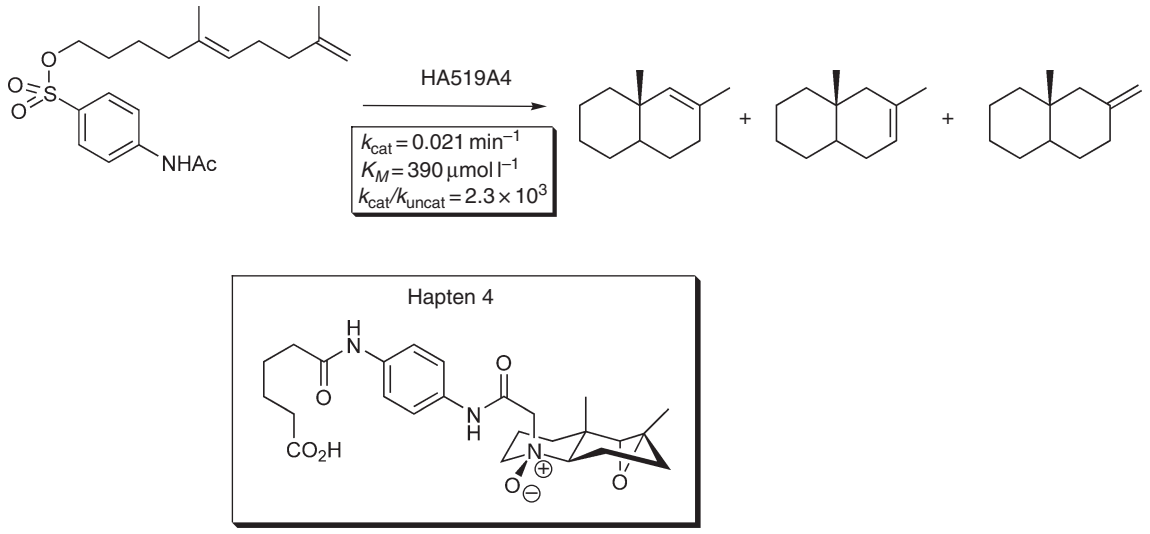


Figure 5 Tandem cationic cyclization catalyzed by antibody HA519A4 raised against hapten **4**.

(**Figure 7**). Due to the significant stereoelectronic constraints predicted by Baldwin's rules,[30,31] the uncatalyzed cyclization of *trans*-epoxyalcohol proceeds via the 6-exo pathway, affording tetrahydrofuran. Hapten **6a** was designed to mimic the stereoelectronic features of the disfavored 6-endo TS, where the N-oxide functional moiety mimics the electronic polarization of the epoxide in the TS. The piperidinium ring provides the required pyran chair conformation of the disfavored product.[32] Antibody 26D9, which was able to reroute the reaction process, was obtained, yielding tetrahydropyran as the only product. Alternatively, *N*-methyl ammonium (**6b**) was used in the immunization. Antibody 5C8 was discovered as a catalyst for the regio- and enantio-selective disfavored endo-ring opening of the substrate. The X-ray structures of two complexes of Fab 5C8 with the eliciting hapten and with an inhibitor were recently published.[33] The active site of the antibody contains a putative catalytic diad, consisting of AspH95 and HisL89 that perform general acid–base catalysis.

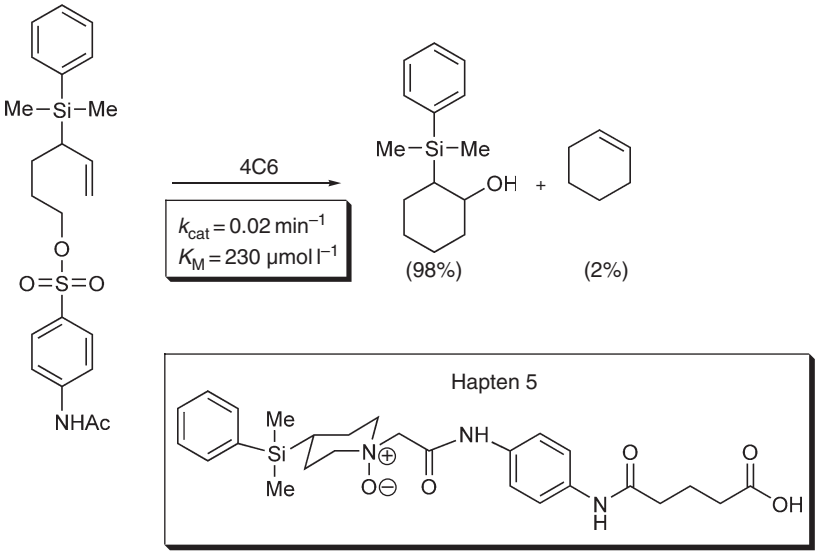


Figure 6 Tandem cationic cyclization catalyzed by antibody 4C6 raised against hapten **5**.

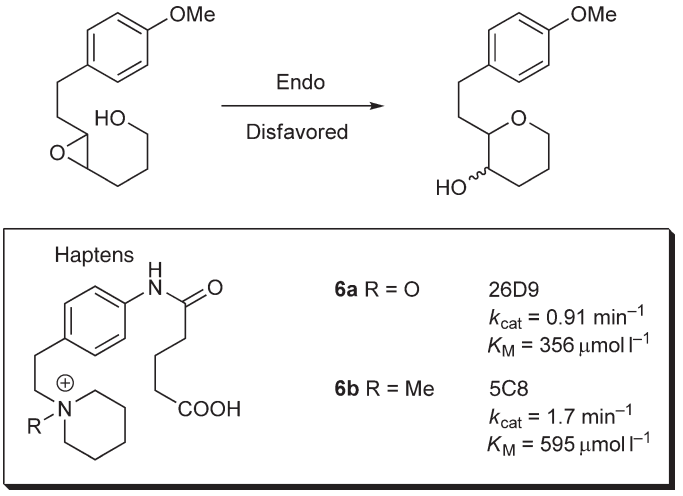


Figure 7 Intramolecular cyclization of *trans*-epoxyalcohol catalyzed by antibodies 26D9 and 5C8 raised against cyclic ammonium haptens **6a** and **6b**, respectively.

5.12.3.4 Diels–Alder Reaction

The Diels–Alder reaction is of particular interest for chemists not only because it is a rare reaction in nature, but also because of the fact that it proceeds via an entropically disfavored, highly organized TS.[34] To date, a number of antibody Diels–Alderases have been reported. Antibody 1E9 was elicited against the endo-hexachloronorbornene derivative **7**, which is a stable analogue of the high-energy TS for the cycloaddition between tetrachlorothiophene dioxide and *N*-ethylmaleimide (**Figure 8**).[35] Since this reaction liberates SO_2 spontaneously and oxidizes to form the aromatic product, which is structurally dissimilar to the hapten, no product inhibition was detected. From the X-ray crystallographic data of Fab fragment of 1E9, it was revealed that the antibody-binding pocket is preorganized to provide significant shape complementarity with the hapten through Van der Waals contacts, π-stacking with the maleimide functional moiety, and a hydrogen bond with AsnH35.[36]

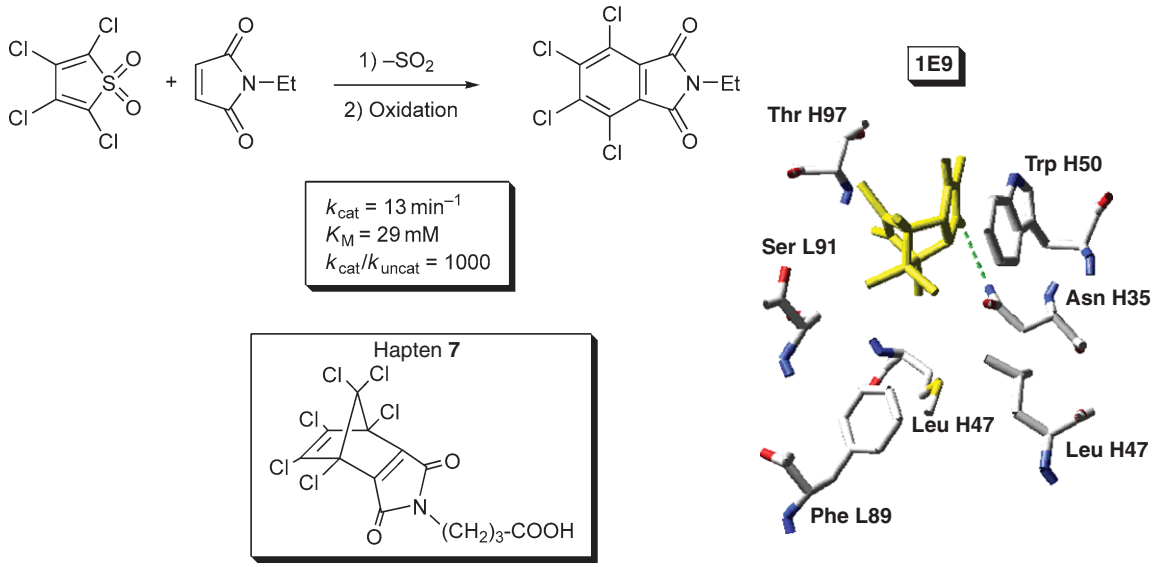


Figure 8 Diels–Alder reaction catalyzed by the antibody 1E9 raised against endo-hexachloronorbornene **7** and X-ray structure of antibody 1E9.

In a recent study, Kim *et al.*[37] surveyed Diels–Alder reactions catalyzed by noncovalent binding to synthetic, protein, and nucleic acid hosts. Antibody 1E9 was revealed as the most effective catalyst of the noncovalent catalyst systems studied. This extraordinary catalytic capability has been explained by theoretical calculations, and results indicated that 1E9 has a high degree of shape complementarity, consistent with the X-ray crystallographic data.[36] Antibody 39A11 was raised against bicyclo[2.2.2]octane hapten **8**, which was designed as a mimic of the proposed boat-like TS of the 4p + 2p cycloaddition between the corresponding diene and dienophile (**Figure 9**).[38] Product inhibition is circumvented by the structural disparity between the product cycloadduct and the pseudoboat form of the hapten employed for immunization. The X-ray crystallographic data of the Fab fragment of 39-A11[39] as a complex with the hapten has been reported.[40] It was revealed that the antibody binds the diene and the dienophile in a reactive conformation and presumably reduces translational and rotational degrees of freedom. The binding of enantiomeric haptens by antibody 39A11 was studied theoretically by Zhang *et al.*[41] in an investigation of mechanism of stereoselective hapten binding by this Diels–Alderase antibody using docking simulations and quantum mechanical models. Based on these data, they

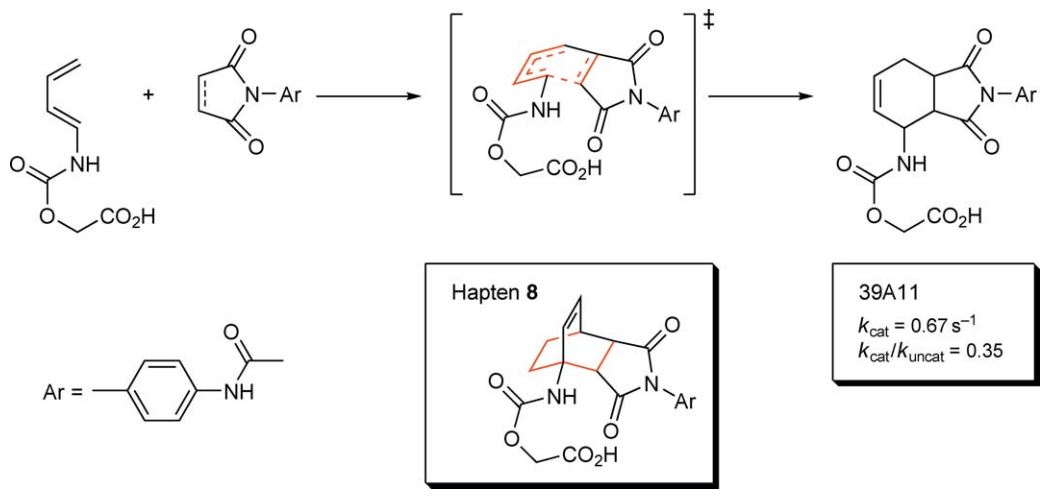


Figure 9 Diels–Alder reaction catalyzed by antibody 39A11 raised against the bicyclic hapten **8**.

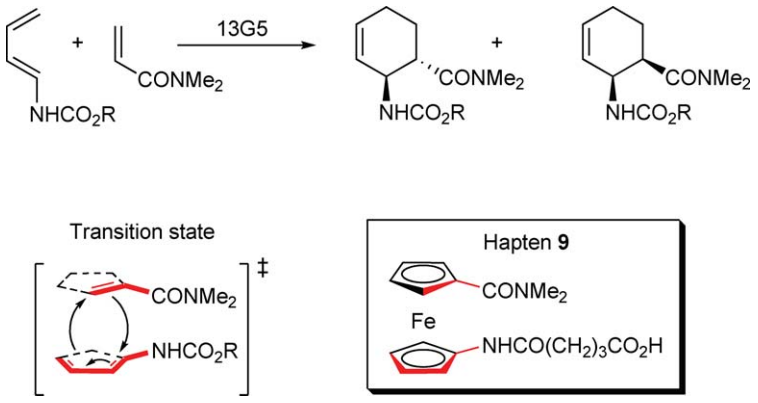


Figure 10 Diels–Alder reaction catalyzed by antibody 13G5 raised against a ferrocene hapten **9**.

predicted that the stereoselectivity of 39A11 was accomplished by two strategically positioned hydrogen bonds and π-stacking of the maleimide with a tryptophan at the antibody-binding site.

As a new approach to produce Diels–Alderase antibodies that could catalyze the formation of either the exo- or endo-cycloadducts, Janda and co-workers[42] have employed a ferrocenyl hapten **9**. This hapten is highly flexible with the cyclopentadienyl rings able to rotate freely in solution. Antibody 13G5, raised against hapten **9**, is able to catalyze the disfavored exo-cycloaddition reaction between the corresponding diene and dienophile in high regio-, diastereo-, and enantio-selectivity (**Figure 10**). The crystal structure of the 13G5 Fab complexed to an attenuated form of hapten was determined.[43] It was shown that the ferrocene moiety is completely buried in the antibody-combining site, and the ferrocene ring rotation is restricted by the steric restraints imposed by specific hydrogen bonding interactions with the antibody-binding pocket. Cannizzaro *et al.*[44] synthesized the enantiomerically pure Diels–Alder adduct obtained by antibody catalysis with 13G5 and other monoclonal antibodies elicited during the same immunization process. Based on this information, they established enantioselectivity of these antibodies and the effects of different catalytic residue arrangements on the TSs were modeled quantum mechanically. The results provided an explanation of the origin of the observed enantioselectivity of 13G5. It was depicted in their study that the hapten molecule **9** used for screening antibodies for binding to TS analogues resembles the Van der Waals complex between the reactants more closely than the TS; this selection process yields binders that preferentially recognize the rotamer of the hapten that mimics the (*R, R*) TS analogues.

5.12.3.5 Rearrangements

5.12.3.5.1 Chorismate–prephenate rearrangement

The antibodies can also act like entropy traps by stabilizing a particular conformation of a substrate that is favorable to the formation of the TS. It is the case of the antibody 1F7 catalyzing the transformation of chorismate into prephenate,[45] which stabilizes, thanks to several hydrogen bonds and an ionic bond between an arginine (Arg H95) and a carboxylate substituent of the substrate, the conformation of the chorismate which will give rise to the TS in a chair conformation for this reaction (**Figure 11**).

5.12.3.5.2 Oxy-Cope rearrangement

The oxy-Cope rearrangement is a [3,3] sigmatropic rearrangement, which occurs via a highly organized chair-like TS.[46,47] An antibody catalyst for such a reaction must be able to bind the substrate and orient the ground state into this productive chair-like conformation. Antibody AZ28 was raised against the chair-like TSA **11**, which catalyzes the oxy-Cope rearrangement of **12** to produce **13** (**Figure 12**).[48]

Product inhibition and chemical modification of the antibody are prevented by *in situ* generation of the oxime. Surprisingly, the germline precursor of AZ-28, which has a much lower affinity toward the eliciting hapten, accelerated the reaction 164 000-fold over the uncatalyzed background reaction. In order to study the

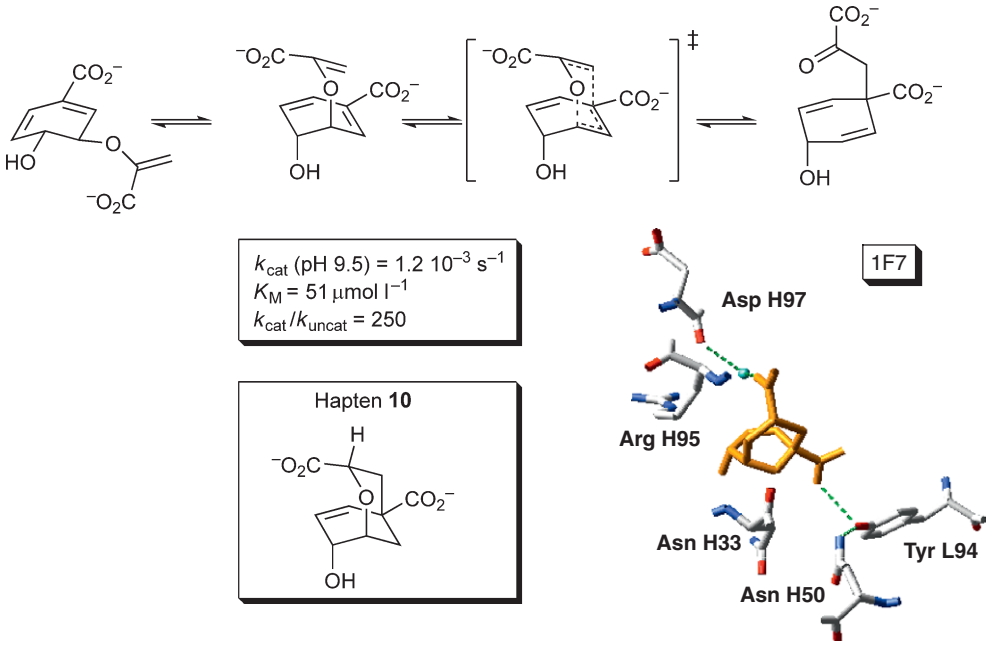


Figure 11 Chorismate–prephenate rearrangement catalyzed by antibody 1F7 raised against a bicyclic hapten that mimics the chair conformation of the transition state of the reaction and X-ray structure of the corresponding antibody 1F7–hapten complex.

structural basis for binding and catalysis, the X-ray crystal structures of AZ-28, apo-form and complexed with the eliciting hapten, have been determined.[49] In the antibody–hapten complex, the TSA is fixed in a catalytically unfavorable conformation by a combination of Van der Waals and hydrogen bonding interactions (**Figure 12**). In contrast, the active site of the germline precusor of AZ-28 appears to have a much higher degree of flexibility; specifically, CDRH3 moves 4.9 Å outwards from the active site upon binding the eliciting hapten. It was initially proposed that this conformational flexibility in the germline antibody allows dynamic changes that lead to enhanced orbital overlap and increased rate acceleration. This mechanistic explanation was further substantiated by a molecular dynamics simulation study by Asada *et al.*[50]

5.12.3.6 Biological Like Reductions

Alcohol dehydrogenases catalyze the oxidation of alcohols to aldehydes and simultaneously reduce the nicotinamide derivatives NAD^+ and $NADP^+$ to the corresponding 1,4-dihydronicotinamides.[51,52] Compounds **14** and **15** were designed and prepared as stable TSAs for the hydride transfer process between the nicotinamide derivative **16** and an aliphatic aldehyde to distinguish specific antibody catalysis from contaminant dehydrogenases (**Figure 13**).[52] These haptens incorporated a rigid [3.2.2] bicyclic structure that contain a 3-piperidone oxime, wherein the oxime motif mimics the carboxamide group in nicotinamide. The piperidone is held in the boat conformation corresponding to the TS by a three-atom lactam bridge. The aldehyde carbon in the TS was mimicked by a methylene group in **14**, while in hapten **15**, it is mimicked by a sulfonyl group. The production of monoclonal antibodies against these novel haptens and their kinetic characterization has not been reported to date.

5.12.3.7 Limits of the TSA-Based Approach

In general, this first generation of abzymes obtained from TSAs behave like enzymes, present saturation kinetics, substrate specificity, a stereoselectivity and competitive inhibition phenomena. However the acceleration factors obtained, k_{cat}/k_{uncat}, remain weak and are limited in theory by the ratio of the constant of

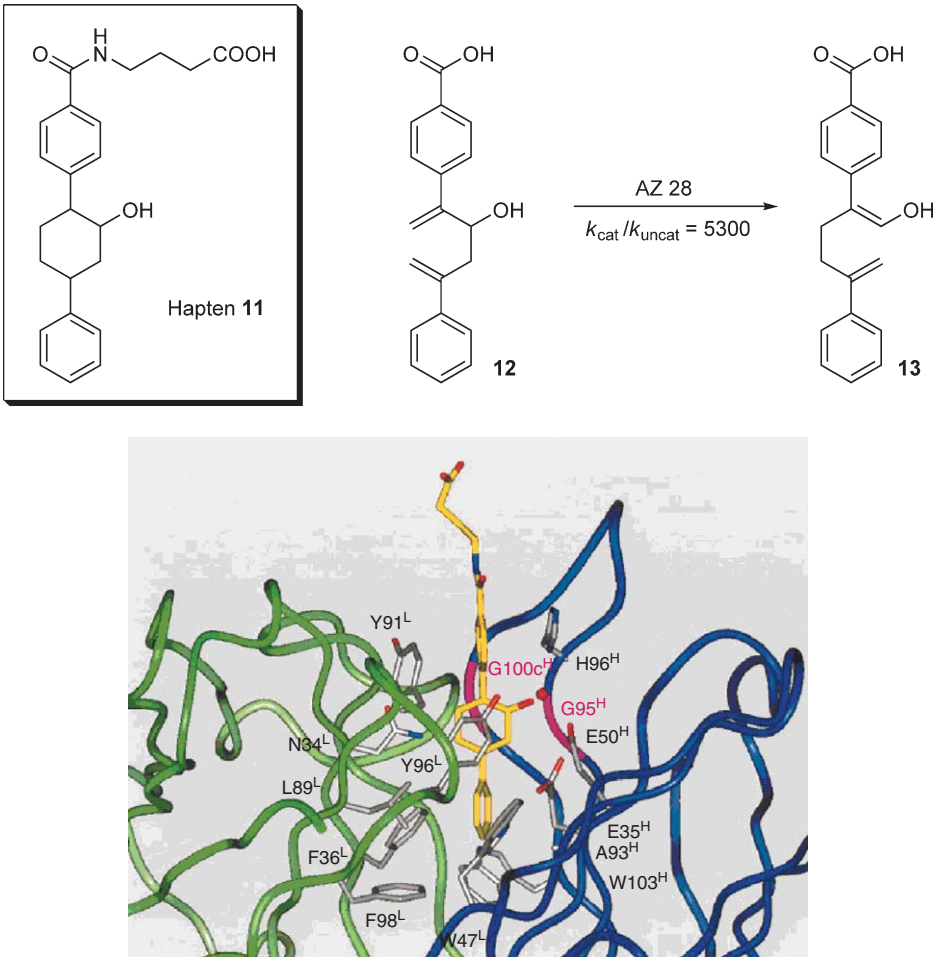


Figure 12 Oxy-Cope reaction catalyzed by antibody AZ28 raised against the hapten **y** and X-ray structure of the antibody AZ28–hapten **y** complex.

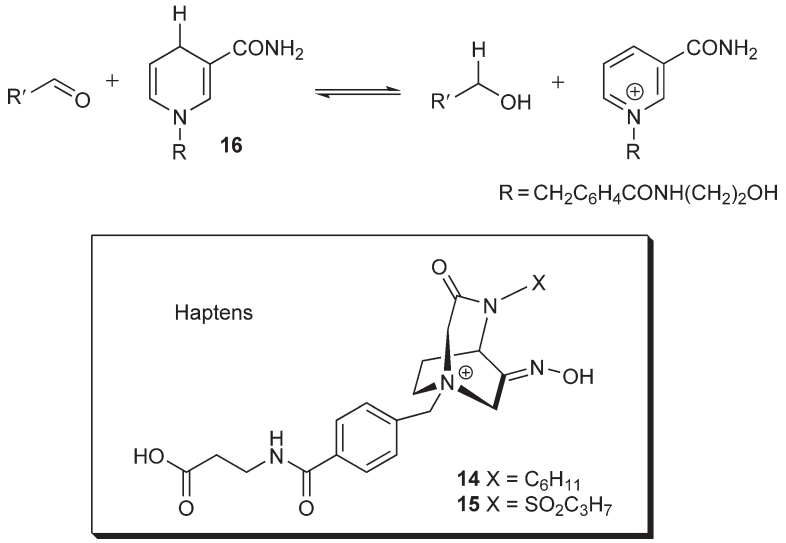


Figure 13 Structure of haptens **14** and **15** designed to elicit monoclonal antibodies that would be able to catalyze the reduction of carbonyl compounds by nicotinamide derivatives.

Michaelis of the reaction catalyzed on the constant of inhibition by the hapten, K_M/K_i. Actually, acceleration factors ranging from 10^3 to 10^8 were observed, which is far from the values of the acceleration factors observed for the best enzymes which can reach 10^{17}.[53] It has also appeared that the efficiency of antibody catalysts to accelerate chemical reactions ($(k_{cat}/K_M)/k_{uncat}$ or $1/K_{TS}$) is much lower than that of natural enzymes.[17,54,55] Indeed, k_{cat}/K_M values reported for catalytic antibodies range from 10^2 to 10^4 $mol^{-1}\,l\,s^{-1}$, while those of natural enzymes range from 10^6 to 10^8 $mol^{-1}\,l\,s^{-1}$. In addition, a significant fraction of hapten binders failed to catalyze the target reactions.

Attempts can be made to explain the apparent limitations of catalytic antibodies raised against TSA. First of all, during the immune response, the maturation of the immune response is driven so as to obtain the best hapten affinity rather than catalytic activity. Consequently, somatic mutations can raise antibodies that favor tighter hapten binding but are deleterious for catalysis. Indeed, it is crucial that antibody catalysts raised against TSAs have to release the product to be efficient, but in a number of cases, the products of the reaction bear a high degree of structural similarity to the TSAs used in immunization. As a consequence, slow release of product can be observed, inducing product inhibition at the antibody-combining site, which has appearred as a major contributing factor for the low efficacy of hydrolytic antibodies.[56–58] One second possible explanation for the poor performance of some TSA-based antibody catalysts may be an inability to design a stable organic compound that could reproduce some of the structural characteristics of the TS such as the fractional bond orders, extended bond lengths, expanded valences, distorted bond angles, and charge distributions. Considering that TSs can share recognition elements with ground state molecules, it seems obvious that the TSAs would never achieve geometries and charge distributions of TSs precisely.[59] For example, a phosphonate is far from being a perfect analogue of the tetrahedral TS involved in the hydrolysis of carboxylic acid derivatives, although it has proven to be very useful to design haptens to elicit hydrolytic antibodies.[60,61] Indeed, there are obvious discrepancies in reproducing the exact bond length and charge distribution compared to the real TS structure evolved during the reaction process.[62] Recently, Tantillo and Houk[63] examined the properties of TSs and the high-energy intermediates involved in the hydrolysis of phenyl and *p*-nitrophenyl acetate and compared them with that of haptens designed according to the TSA strategy. The results suggested that even though aryl phosphonates mimic some of the geometric features of the TSs and intermediates, they less faithfully mimic the asymmetry and electrostatic properties of these stationary states involved in the reaction process. In particular, what the haptens mimic the best are molecular size, bond lengths, and bond angles that are relatively similar to those of the elimination TS and tetrahedral intermediate, but they bear little resemblance to the addition TS or the reactant or product complexes.

As already mentioned in the introduction of this chapter, the majority of catalytic antibodies recognize antigens through interactions including Van der Waals interactions, hydrogen bonding, and other electrostatic and hydrophobic forces, which is very similar to substrate binding in the enzyme active site. In a recent review, Houk *et al.*[64] compared the binding affinities of host–guest, protein–ligand, and protein–TS complexes. In the case of catalytic antibodies, comparison of kinetic data[65] revealed that the TSs for typical antibody-catalyzed reactions are bound 10^3 times more strongly than the substrates. Indeed, dissociation constants for catalytic antibody–TS complexes of $10^{-6.6}$ to $10^{-8.6}$ $mol\,l^{-1}$ were reported, whereas the dissociation constants for the corresponding catalytic antibody–substrate were about 2–6 orders of magnitude higher: $10^{-2.4}$ to $10^{-4.6}$ $mol\,l^{-1}$. This selectivity in guest recognition provided the basis for tailoring antibodies as catalysts. However, the dissociation constants for catalytic antibody–TS complexes are still far from those observed for typical enzyme–TS complexes that have dissociation constants in the 10^{-12} to 10^{-14} $mol\,l^{-1}$ range, demonstrating the strongest binding that involves enzymes and TSs of the reactions they catalyze. This massive binding constant of enzymes are attributed to billions of years of evolution, which progressively incorporated binding interactions between the enzyme and the substrate such as strong electrostatic interactions as well as metal cofactors, acid–base chemistry, and most importantly covalent interactions between the enzyme and TS.[17]

Another possible explanation for the limitations of catalytic antibodies raised against TSA can be found in the different accessibility of the active site. In the case of natural enzymes, it is that their catalytic machinery and bound substrates are often buried. This feature isolates from the solvent the reactive functionalities that mediate chemical transformations.[17,66] On the contrary, in antibody catalysis, the moieties of the bound haptens that mimic the TS are often positioned near the entrance of the antibody-combining site. This disparity in the overall architecture of natural enzymes and catalytic antibodies is undoubtedly a factor in the lower catalytic

proficiency of the latter. A relationship between average binding constants and the average surface areas buried upon binding has been proposed in Houk's review. Based on the thermodynamic data available for protein–substrate complexes, he pointed out a clear trend between the buried areas of the guests and binding affinities – 67% of accessible surface area of guest is buried for 1 kcal mol^{-1} binding energy.[64]

From the results reported to date, it seems that the manner in which haptens are attached to carrier proteins leads to significant differences in certain cases.[67] Clearly, haptens designed with aromatic moieties between the linkage to the immunogenic carrier protein and the TSA motif often have better antibody recognition. Recently, Hilvert[68] pointed out that on both micro and macro levels, mechanistic improvements arise as a function of time. The differences in time scales for the evolution of natural enzymes and antibodies – millions of years versus weeks or months – also appear to be an explanation of the low efficiency of antibody catalysts. He also highlighted that the unique immunoglobulin fold has not been adopted by nature as one of the common scaffolds on which to build enzyme catalytic machinery. Therefore, antibody structure itself places limitations on the kind of reactions amenable to catalysis.

5.12.4 'Bait and Switch' Strategy

The 'bait and switch' strategy was developed for the first time by Shokat *et al.*[69] in 1989 in order to generate within an antibody the residues of desired functionality that would not inevitably have been induced by the 'transition state analogue' approach. For that it is necessary to design a hapten which has not only certain structural characteristics close to the substrate, in order to ensure a sufficient recognition of the substrate by the antibody, but also functions carrying a permanent charge suitable for inducing in the antibody a function of opposite sign. This strategy involves the placement of a point charge on the hapten in close proximity to, or in direct substitution for, a chemical functional group that is expected to transform the corresponding substrate. The haptenic charge is expected to induce a complementary charge at the active site. The charged amino acid residues thereby recruited contribute to catalysis as general acid–base or nucleophilic catalysts. Since the haptens designed according to this strategy serve as a 'bait' for eliciting catalytic functions during the immunization process, which is then 'switched' for the substrate, the strategy has been named 'bait and switch'.[70,71] Until now, only haptens involving an ammonium or an amidinium function bearing a positive charge have been used for this strategy. Thus, the use of a hapten carrying a quaternary ammonium **17** by Shokat *et al.* has allowed to induce in the site of an antibody a glutamate residue whose carboxylate function plays the basic part able to abstract a proton in α of a ketone allowing the elimination of HF[72] (**Figure 14(a)**).

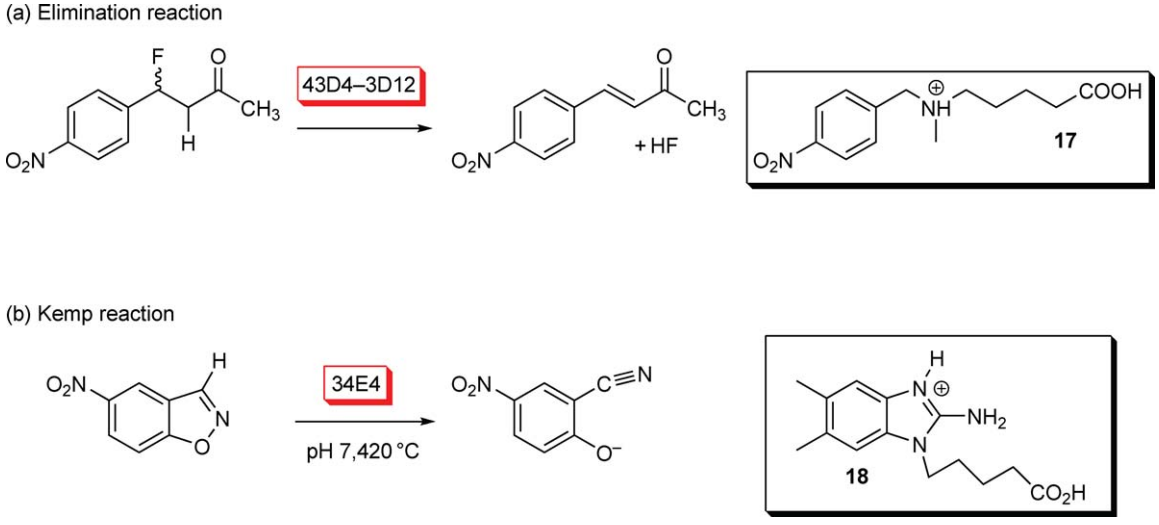


Figure 14 (a) Elimination reaction catalyzed by antibodies 43D4 and 3D12 raised against the quaternary ammonium **17** and (b) Kemp reaction catalyzed by antibody 34E4 raised against the amidinium **18**.

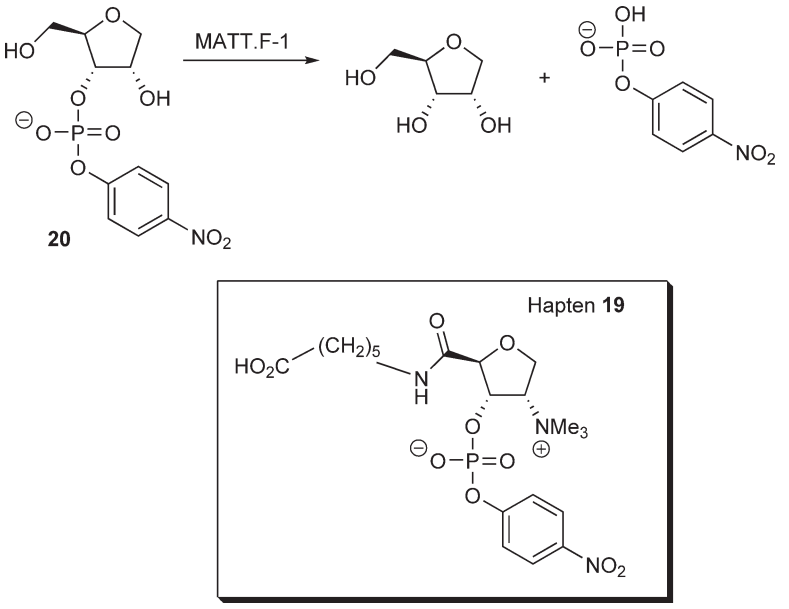


Figure 15 Hydrolysis of phosphodiester bonds catalyzed by antibody MATT.F-1 raised against hapten **19** according to the 'bait and switch' strategy.

This strategy allowed a little later the group of Hilvert,[72] thanks to the synthesis of a hapten **18** carrying an amidinium function, to generate antibodies that are able to very effectively accelerate the ring opening reaction of a benzisoxazole cycle ($k_{cat}/k_{uncat} = 3 \times 10^8$) (**Figure 14(b)**).

The hydrolysis of phosphodiester bonds, which are often found in DNA and RNA, is a reaction of significant importance in living systems. Antibody MATT.F-1, which was raised against a quaternary ammonium hapten, is the most proficient antibody catalyst generated for the phosphodiester hydrolysis reaction (**Figure 15**).[71] Hapten **19** was designed according to the 'bait and switch' strategy with the objective to incorporate a general base in an antibody-binding site proximal to the 20 hydroxy of substrate **20** to facilitate nucleophilic attack of this hydroxyl group on the adjacent phosphoryl center. In this case then, the transition state mimicry is sacrificed and replaced by a point charge. MATT.F-1, which has a catalytic efficiency ($(k_{cat}/K_M)/k_{uncat}$) of $1.6 \times 10^7\ mol^{-1}\ l$, which is higher than that reported for 2G12, elicited a TSA for the same reaction. Perhaps most impressive, however, is that the proficiency of MATT.F-1 is only three orders of magnitude lower than that of the naturally occurring enzyme RNaseA for the same substrate.

4B2 is an antibody that catalyzes the allylic isomerization of β,γ-unsaturated ketones (**Figure 16**) as well as the Kemp reaction.[73]

The hapten **25**, a substituted cyclic amidinium, is expected to elicit, by charge complementarity, an acidic residue (Asp or Glu) in the combining site that would catalyze the α-proton exchanges on the same substrate, through a dienol intermediate involved in the reaction process. The presence of Asp and Glu in the complementarity-determining region (CDR) was later confirmed by cloning and sequencing the light and heavy chains of the 4B2. Finally, a 3D structure of the 4B2–hapten complex was obtained (**Figure 16**) that showed the presence in the binding site of the antibody 4B2 of a glutamate residue face to the amidinium function,[2] which validates the use of the 'bait and switch' strategy.

5.12.5 Reactive Immunization

The antibody catalysts described in the previous paragraphs, generated either by the TSA or by the 'bait and switch' strategy, have been generated by immunization with chemically inert antigens that mimic the geometric and/or electronic features of a reaction's TS. A new hapten design strategy – reactive immunization – provides

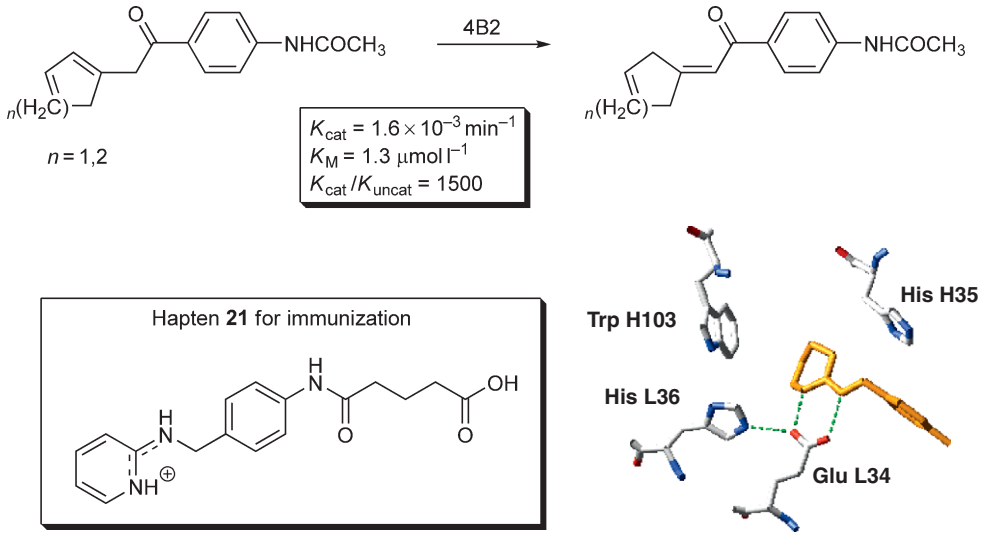


Figure 16 Isomerization reaction catalyzed by antibody 4B2 raised against the cyclic amidinium **21** and X-ray structure of the antibody 4B2–hapten **21** complex.

a chance for catalytic antibodies to approach the catalytic efficiency of natural enzymes through the use of reactive immunogens.[74] This term of reactive immunization indicates the strategy in which the hapten used comes to react with a residue of the active site of the antibody and to engage a covalent bond with it. The major advantage of this strategy lies in the fact that, in this manner, are naturally only selected the clones that secrete the antibodies having in their recognition site this residue correctly placed and able to react with the hapten.

In 1995 Janda, Lerner, and co-workers pioneered the use of a highly reactive antigen that undergoes a chemical reaction in the antibody-combining site during immunization.[75] In the first attempt to examine this strategy, they designed an organophosphorus diester hapten **22** as the primary reactive immunogen for immunization. This hapten can be either hydrolyzed at physiological pH or trapped by a nucleophile at the B-cell level of the immune response affording the monoester **23**, an analogue of the TS (**Figure 17**). Nineteen mAbs (monoclonal antibodies) were isolated, 11 of which were able to catalyze the target acyl-transfer reaction, hydrolysis of phosphonate diester **24** to generate the corresponding monoester, in a one-turnover inactivation of the antibody (**Figure 17**). Among these 11 catalytic antibodies, SPO49H demonstrated the best catalytic activity; it effectively catalyzes the hydrolysis of the activated ester **25** to yield carboxylic acid **31** with a k_{cat} of 31 min^{-1} and a rate acceleration (k_{cat}/k_{uncat}) of 6700 at pH 8.0. The first direct comparison between reactive immunization and TSA hapten manifolds for catalytic antibody production was demonstrated by comparing esterase antibodies elicited against a TSA, phosphonate monoester **26**, with the ones raised by a reactive immunogen, phosphonate diester **27**.[76,77] Hapten **27** was initially designed for the purpose of resolving a racemic mixture of naproxen esters.

Antibody 15G2 raised against hapten **27** catalyzed the homochiral production of the anti-inflammatory agent Naproxen, rac-29, from **28**, incorporating stereoselective activity and disposition, the *S*-(+)-enantiomer **29a** of naproxen being formed 28 times more than the *R*-(−)-enantiomer **29b** (**Figure 18**). This antibody catalyzed the hydrolysis of *S*-(+)-**28a** and gave *S*-(+)-**29a** with a $k_{cat} = 28\ min^{-1}$, $K_M = 300\ \mu mol\ l^{-1}$ at pH 8.0, and $k_{cat}/K_M = 9.3 \times 10^4\ mol^{-1}\ l\ s^{-1}$. Meanwhile, antibody 6G6, raised against the TSA **26** catalyzed the same reaction with a $k_{cat} = 81\ min^{-1}$, $K_M = 890\ \mu mol\ l^{-1}$ at pH 8.0, and $k_{cat}/K_M = 4.5 \times 10^4\ mol^{-1}\ l\ s^{-1}$. The TSA approach provided good biocatalysts in terms of turnover numbers and enantiomeric discrimination, albeit with varying degrees of product inhibition by the phenol by-product. Reactive immunization has generated biocatalysts that are ultimately more proficient because they combine an efficient catalytic mechanism, improve substrate recognition, and do not suffer from product inhibition. In practice, these hydrolytic antibodies generated by reactive immunization have also been applied to the hydrolysis of polyesters.[78,79]

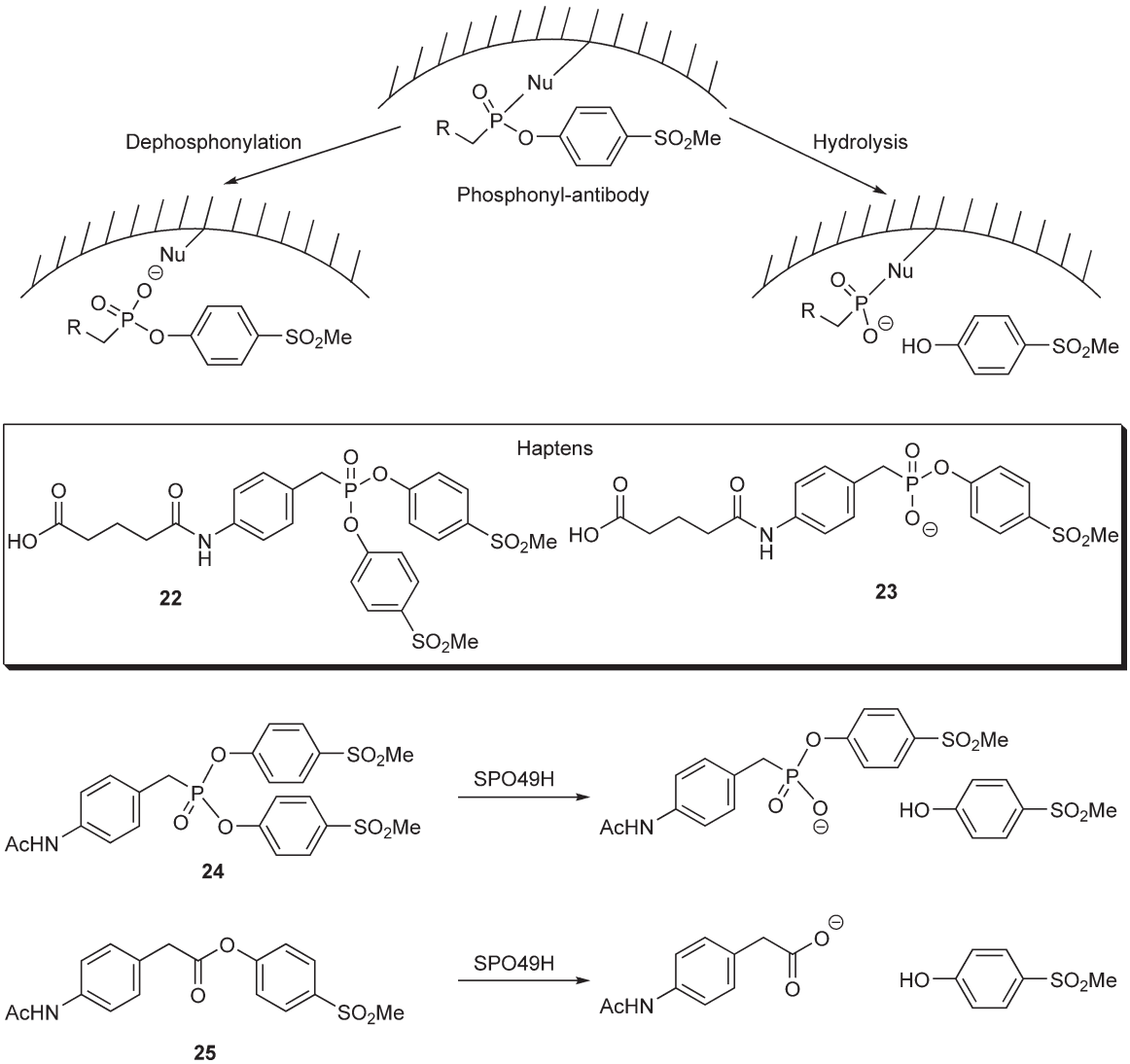
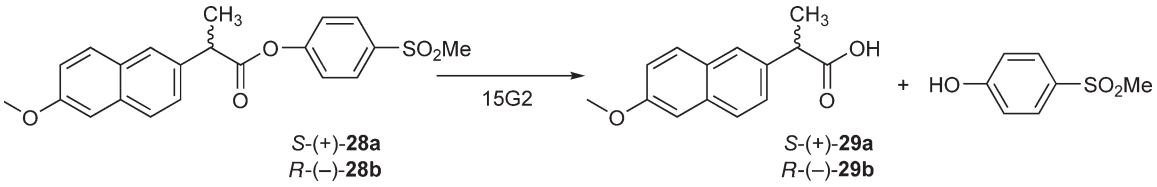


Figure 17 Generation, by the reactive immunization strategy, of antibody SPO49H that catalyzes the hydrolysis of phosphonate diester **24** and that of ester **25**.

The reactive immunization strategy has also been successfully used by Richard Lerner and co-workers[80,81] to obtain abzymes with aldolase activity. Thus, using hapten **30**, equipped with a moderately reactive β-1,3-diketone functionality, they could easily select the antibodies having a lysine sufficiently accessible in their binding site to be able to form, after reaction with the hapten, a cyclic β-imino-ketone directly detectable in UV–visible spectroscopy ($\lambda_{max} = 316$ nm) (**Figure 19**). The β-1,3-diketone moiety thus demonstrated its ability to trap a lysine side chain amine that could directly participate in the mechanism of the aldol reaction in the active site of the antibody (**Figure 19**). The antibodies thus selected could, like the aldolases, form a donor iminium by reaction of this lysine with a donor ketone which was then able to condense with an acceptor aldehyde. Two of the selected antibodies, 38C2 and 33F12, were able to catalyze the reactions of aldolization and retro-aldolization as efficiently ($k_{cat}/k_{uncat} = 5 \times 10^6$) as the natural aldolases, but are able to accept a much broader range of substrates. These two aldolases were shown to be extremely robust and they have been involved in major steps of the total synthesis of epothilones A–F.[82,83] These properties, added to the remarkable enantioselectivity displayed by these abzymes, led to their use on a preparative scale in laboratories and one of the two abzymes, 38C2, was the first antibody to have been commercialized (Fluka-Aldrich).



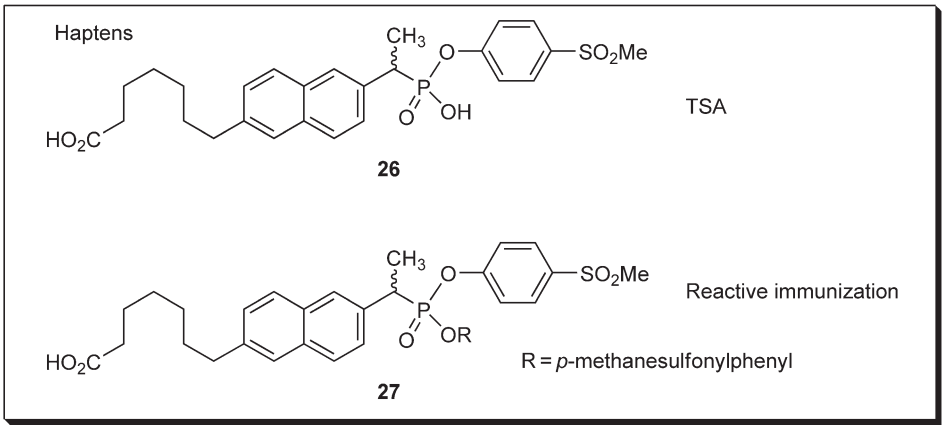


Figure 18 Kinetic resolution of racemic naproxen ester **28** by antibody 15G2 raised against hapten **27**.

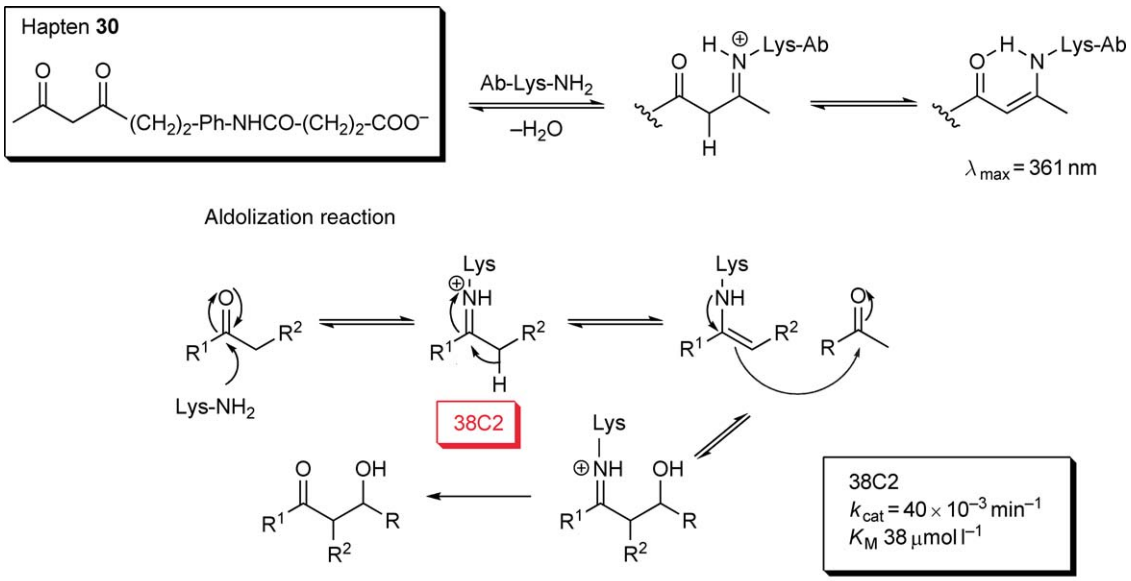


Figure 19 A success of the reactive immunization strategy. Aldolization reaction catalyzed by antibody 38C2 raised against a β-1,3-diketone hapten.

The X-ray structure of 33F12 revealed that the catalytic mechanisms of this antibody is significantly dependent on LysH93, which initiates catalysis by forming a stable covalent conjugated enamine with the ketone substrate that becomes the aldol donor.

In order to improve aldolase antibodies, Zong *et al.*[84] employed reactive immunization in combination with transition state theory. Based on hapten **30**, a hybrid, hapten **31**, was designed, recruiting not only a sulfone

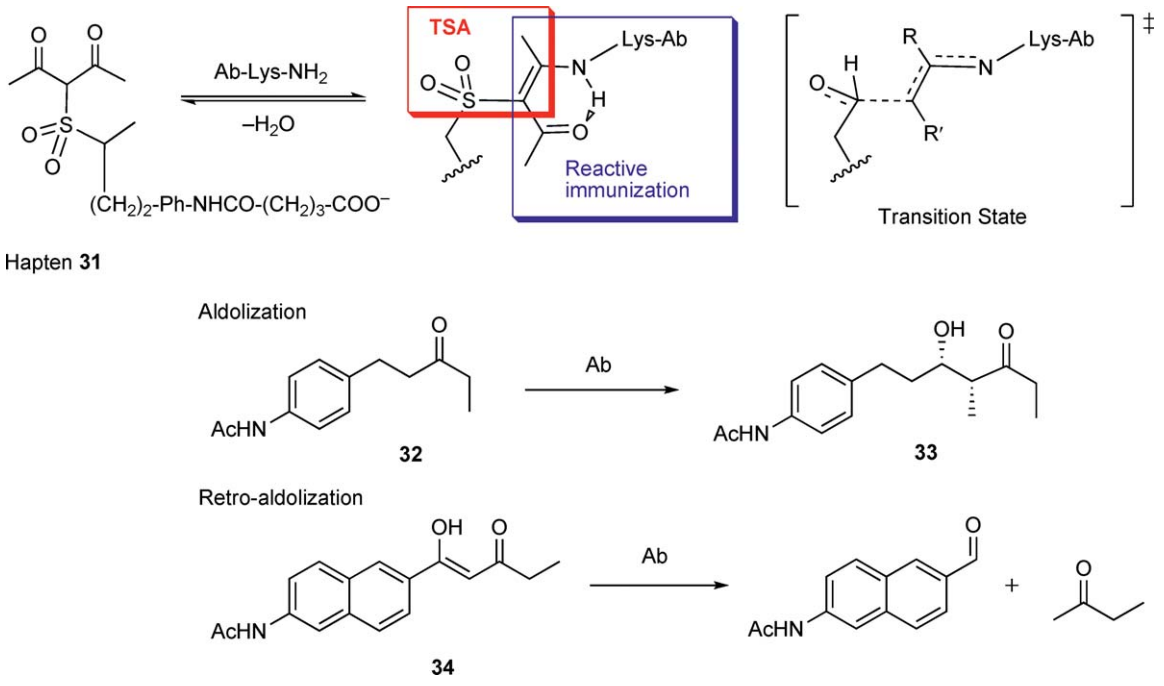


Figure 20 Aldolization and retro-aldolization reactions catalyzed by antibodies raised against hapten **31**.

functionality to establish the tetrahedral motif that is present in the TS, but also a β-diketone for trapping a lysine side chain at the active site (**Figure 20**).[84] Two aldolase antibodies, 93F3 and 84G3, were isolated. In the aldol reaction of **32** with 3-pentanone, antibody 93F3 provided syn-aldol (**33**) with 90% de and 99% enantiomeric excess (ee), while antibody 38C2, elicited to hapten **30**, afforded only 62% de and 59% ee. When using these aldolase antibodies for the kinetic resolution of (±)-**34**, antibodies 93F3 and 84G3 showed a 10^3-fold increase in proficiency over the antibody 38C2 (**Figure 20**).

Taking advantage of insights gained from the existing aldolases 38C2 and 33F12, obtained by reactive immunization, a V gene shuffling strategy was established by Tanaka *et al.*[85] for the reconstruction of aldolase antibodies with improved substrate specificity and turnover. The crystal structure of antibody 33F12 revealed that sequences of the LysH93, HCDR3, and LCDR3 are critical with respect to the mechanism of catalysis, as well as the hydrophobic nature of the combining sites. These sequences were therefore retained when constructing an antibody heavy chain variable domain library using human bone marrow cDNA. The phage displayed libraries were screened against 1,3-diketone **35** and 30-BSA in order to select antibodies that would tolerate β-aldol substrates such as **36–38** (**Figure 21**). The phage-selected clones were further screened by ELISA to identify soluble Fab capable of binding both **30** and **35**. The last stage of selection, which utilized fluorogenic substrates **36–38** (**Figure 21**), identified antibody Fab 28. Though the catalytic mechanism of parental antibodies appeared to be conserved in Fab 28-catalyzed reactions, the k_{cat} values of this antibody were superior relative to those of the parental antibodies for the same substrates, approximately 3–10-fold higher.

Mechanism-based inhibitors covalently react with the active site in target proteins and inhibit their activities; they therefore provide a wealth of information to guide the design of immunogens for immunization.[86–88] For example, penam sulfones have been shown to be potent mechanism-based inhibitors of β-lactamase by forming an acyl enzyme intermediate, which inspired the design of the sulfone hapten **39** targeting the hydrolysis of the lactam functionality built in the substrate **40**.[89] Immunoconjugate 39-KLH was used for immunization, and an scFv library was constructed using the spleen cells of immunized mice. Screening of the library afforded two scFv antibodies, FT6 and FT12; these antibodies catalyzed the hydrolysis of **40** with rate accelerations (k_{cat}/k_{uncat}) of 5200 and 320, respectively (**Figure 22**).

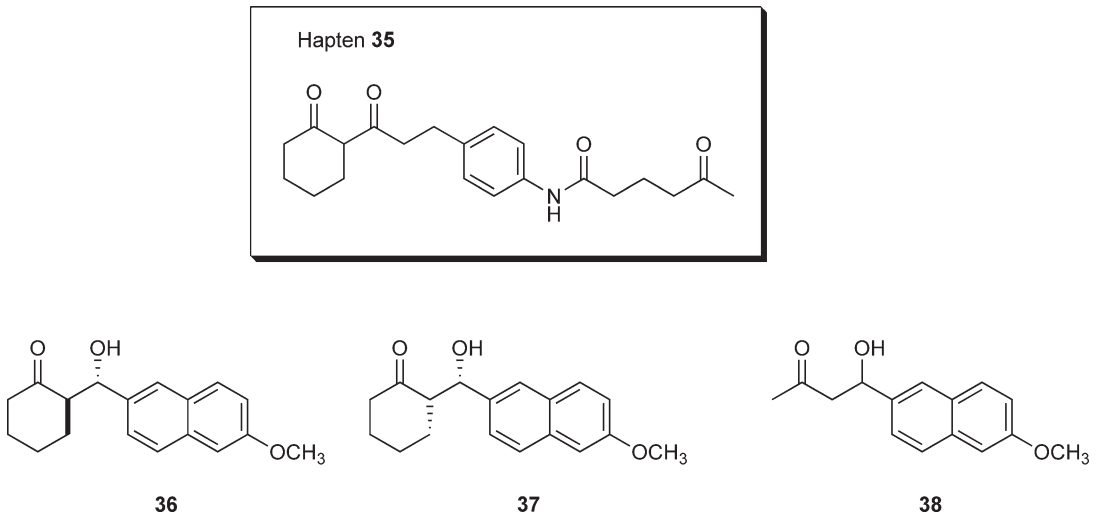


Figure 21 Hapten **35** and fluorogenic substrates **36–38** used to screen phage displayed Fabs with an aldolase activity.

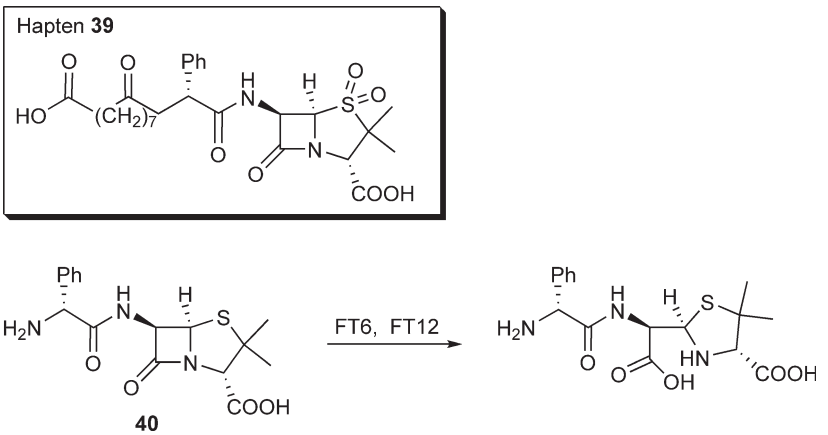


Figure 22 Penam sulfone **39**, a mechanism-based inhibitor of β-lactamase, used as a hapten to generate scFv antibodies, FT6 and FT12, with a β-lactamase activity.

5.12.6 Antibodies Using Cofactors

The addition of cofactors to antibodies is a sure means to confer a catalytic activity to them insofar as this cofactor is responsible for the activity. Indeed for many enzymes, the interaction with cofactors such as thiamins, flavins, pyridoxal phosphate, and ions or metal complexes is absolutely essential for the catalysis. It is thus a question there of building a new biocatalyst with two partners: the cofactor responsible for the catalytic activity, and the antibody which binds not only the cofactor but also the substrate that it positions in a specific way one with respect to the other, and can possibly take part in the catalysis thanks to some of its amino acids.

According to this strategy, antibodies are associated with inorganic or organic cofactors.[1–8] For example, external nucleophilic cofactors can be employed to improve the catalytic activity of the catalytic antibodies. Pyridoxal 50-phosphate (PLP) has been shown to be an effective cofactor for antibody-catalyzed aldol and retro-aldol reactions.[90] Aldolase antibody 10H2, elicited to hapten **41**, catalyzed the aldol reaction between glycine and aldehyde **42**, when combined with cofactor PLP, with a rate acceleration double that of the background reaction where no PLP was applied (**Figure 23(a)**).[91] This incorporation of PLP also improved the rates of the retro-aldol reactions of the threo- and erythro-isomers with rate enhancements of 4-fold and 2.5-fold, respectively (**Figure 23(b)**).

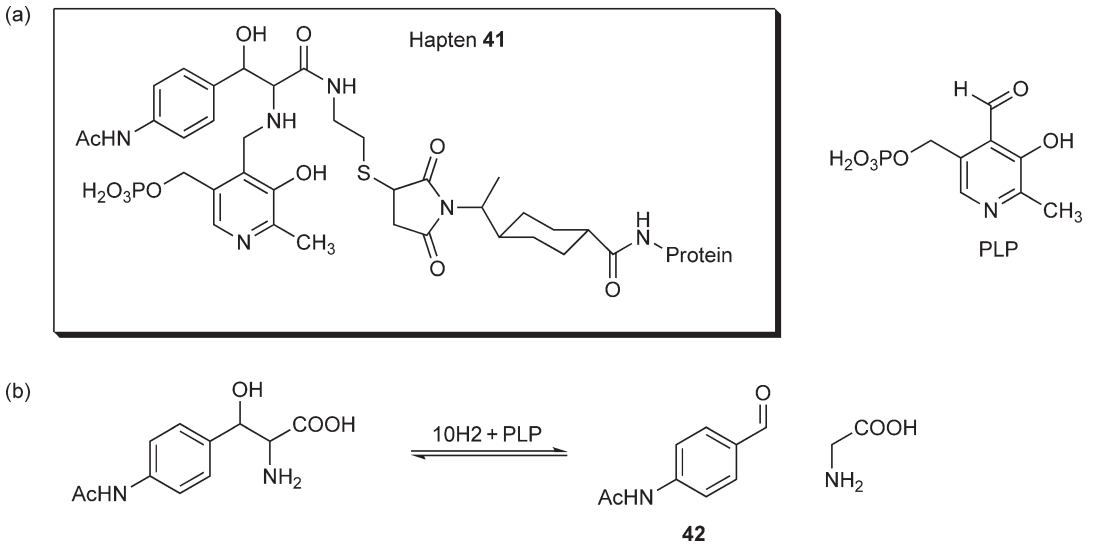


Figure 23 (a) Structure of PLP and hapten **41** conjugated to carrier protein, (b) aldol and retro-aldol reactions catalyzed by mAb 10H2 with PLP 62.

Metal-coordinated enzymes are ubiquitous in nature. A metallic species at the active site of an enzyme often enhances substrate selectivity and accelerates reaction rates. In the field of catalytic antibodies, substantial effort has been directed toward the development of improved antibody catalysts that recruit metals at the active site. Antibody aldolase 38C2 was chosen as the parent antibody by several groups for the development of novel catalytic antibodies by the recruitment of cofactors. Nicholas *et al.* have employed bis-imidazolyl ligand coordinated copper complexes as a cofactor, taking advantage of the high copper-binding affinity ($K_d = 10^{-12}$) of the ligand.[92,93] Insertion of this cofactor was achieved by the covalent attachment of the bis-imidazolyl ligand, which was equipped with a reactive succinic anhydride moiety, to the residue LysH93 and further incorporation of copper(II) by reaction with $CuCl_2$. This semisynthetic metalloantibody, 38C2-58-$CuCl_2$, catalyzed the hydrolysis of a picolic acid ester in aqueous buffer under physiological conditions. This study exemplified that modification of the active site by a metal-coordinated ligand could alter the catalytic nature of the parent antibody, affording a catalyst with very different catalytic activity. Incorporation of metallic cofactors demonstrated the potential to change not only the function of catalytic antibodies, but also to improve the parent antibody regarding substrate selectivity, turnover, and efficiency, while retaining the same catalytic mechanism. For example, when palladium(II), $Pd(en)Cl_2$, or Na_2PdCl_2 was added to aldolase antibodies 38C2 and 33F2, their reaction rates were accelerated.[94] Notably, $Pd(en)Cl_2^-$ antibody binding is reversible, and enantioselectivity was improved in the case of 38C2 by addition of Pd(II).

However, the target of choice remains without any doubt the realization of model systems of hemoproteins, such as cytochrome P-450 monooxygenases and peroxidases, by associating antibodies and natural or synthetic hemes, so as to produce new biocatalysts for selective oxidations. Thus, many antibodies were generated using as haptens palladium- or tin-porphyrins, *N*-substituted porphyrins, or tetraarylporphyrins substituted in meso by carboxyphenyl substituents.[95] The first monoclonal antibody, 7G12, was obtained by Cochran and Schultz[96] using *N*-methy1mesoporphyrin IX (*N*-CH_3–MPIX) (**Figure 24**), as a hapten. The corresponding Fe(III)–MPIX–7G12 complex was shown to catalyze the oxidation of several typical peroxidase cosubstrates, such as *o*-dianisidine, ABTS, and pyrogallol, by H_2O_2. The structure of a mesoporphyrin IX–7G12 complex was determined by X-ray diffraction studies[97] (**Figure 24**). It shows that approximately two-thirds of the porphyrin are interacting with the antibody pocket, three pyrrole rings being packed tightly against residues of the V_H domain and two pyrrole rings packed against tyrosine residues of the V_L domain. In addition, a methionine H100c interacts specifically with one pyrrole ring and forces it to adopt a tilted conformation which should be favorable for the insertion of metal ions in the porphyrin ring and could explain the ferrochelatase activity of antibody 7G12. To date, five antibodies showed in the presence of their Fe(III)-porphyrin cofactor an interesting peroxidase activity. Two of those antibodies, directed

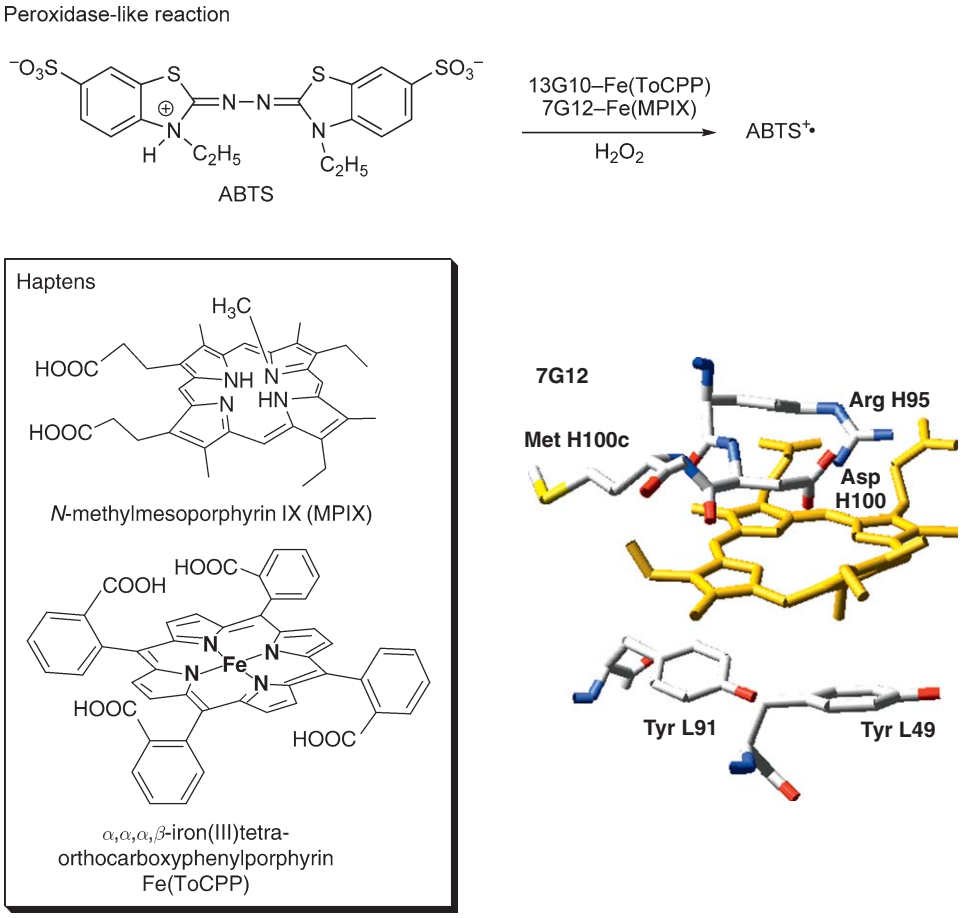


Figure 24 Peroxidase reaction catalyzed by 13G10–Fe(ToCPP) and 7G12–Fe(MPIX) porphyrin–antibody complexes and X-ray structure of the 7G12–MPIX complex.

against an $\alpha,\alpha,\alpha,\beta$-iron(III)tetra-orthocarboxyphenylporphyrin present, in the presence of this cofactor, a peroxidase activity characterized by a catalytic constant $k_{cat} = 560\ min^{-1}$ (**Figure 24**).[95] This constant is higher than that of the horseradish peroxidase itself, but the K_M constant for the H_2O_2 substrate (10 mmol l^{-1}) is much larger than that observed in the case of the enzyme (5 mmol l^{-1}), which makes the antibody much less effective than the enzyme.

Finally, the most convincing results in the field of antibodies with a heme cofactor were obtained independently by the team of E. Keinan[98] and that of J-P Mahy.[99,100] Using tin-tetraarylporphyrin carrying an axial a-naphthoxy ligand as a hapten to mimic the likely TS involved in the oxidation of aromatic sulphides by cytochrome P-450, Keinan *et al.* obtained an antibody that was able, in the presence of a ruthenium-porphyrin, to catalyze the enantioselective oxidation of thioanisole with 43% ee in favor of the *S* isomer (**Figure 25**). On the other hand, using as hapten microperoxidase 8 (MP8), a heme octapeptide produced by pepsic and trypsic digestion of cytochrome *c*, Ricoux *et al.*[99,100] obtained an antibody 3A3 that was able, in the presence of MP8, to catalyze the enantioselective oxidation of thioanisole with 45% ee in favor of the *R* isomer and the regioselective nitration of phenol into 2-nitrophenol by H_2O_2/NO_2^- (**Figure 25**).

5.12.7 Anti-Idiotypic Antibodies

The idiotype is constituted by the whole of the variable parts of the antibody, i.e. those parts which are involved in the recognition of the antigen. The 'anti-idiotype' strategy takes as a starting point the postulate suggested by Niels Jerne[101] according to which for each Ab1 antibody generated against an antigen Ag, there exists a

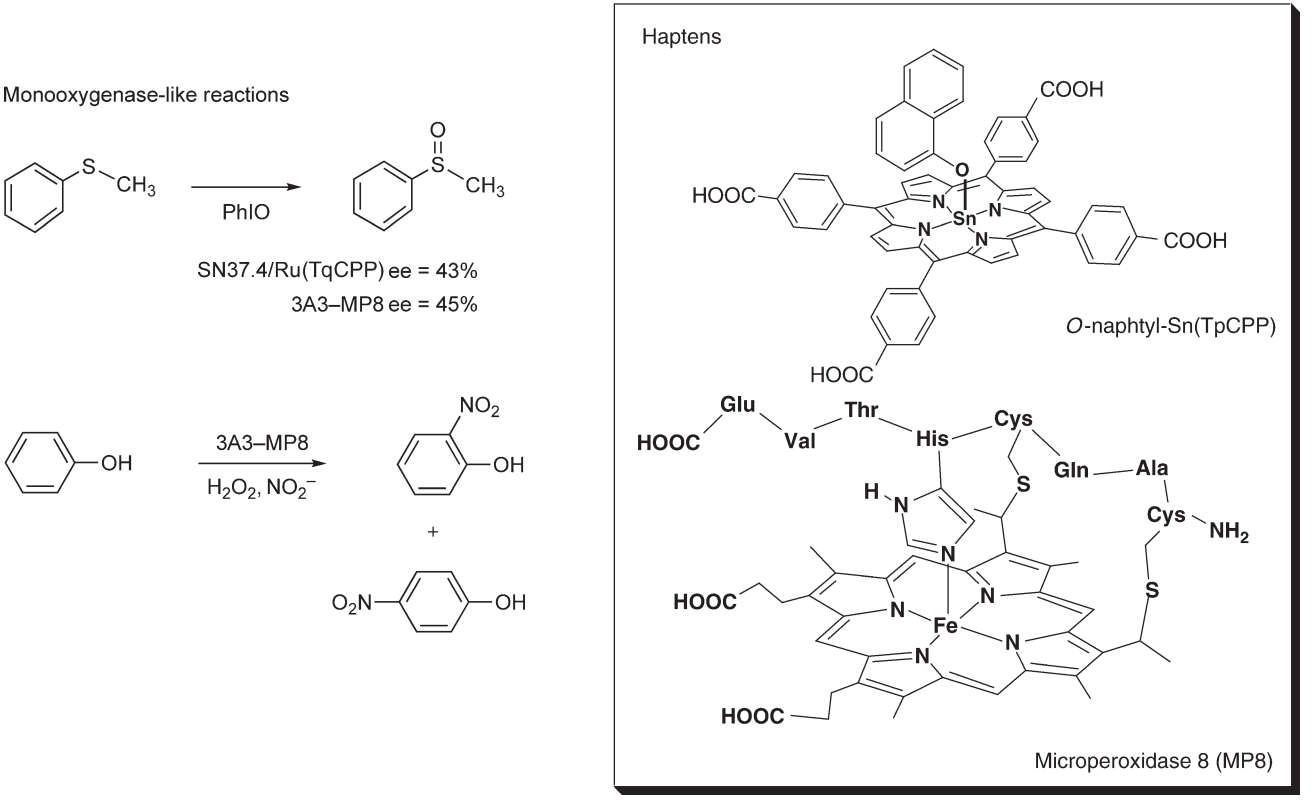


Figure 25 Stereoselective oxidation of sulfides catalyzed by SN37.4–Ru(ToCPP) and 3A3–MP8 complexes and regioselective nitration of phenol catalyzed by 3A3–MP8.

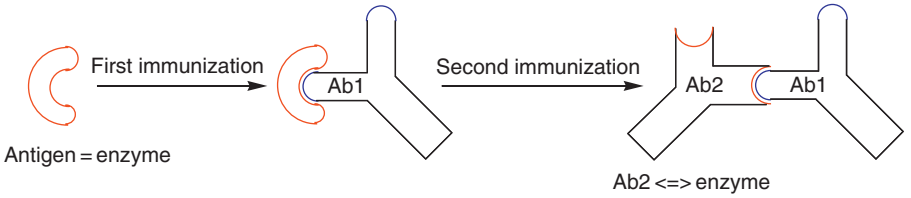


Figure 26 Production of abzymes using the anti-idiotypic antibodies method.

complementary antibody Ab2 directed against the idiotype of the first antibody. Consequently, some of the anti-idiotype antibodies Ab2 have in their recognition site a part mimicking the structure of the antigen; they thus constitute an internal image of this antigen. The strategy then consists in generating Ab1 antibodies directed against the active site of a given enzyme (**Figure 26**). At this stage, the choice of the best Ab1 antibody that is the most complementary as possible of the active site of the enzyme is paramount, one of the best means being to choose the Ab1 that is the best competitive inhibitor of the enzyme. Then, the Ab2 antibodies directed against the idiotype of Ab1 are produced and among them those having in their binding site an internal image of the active site of the antigen enzyme potentially possessed a catalytic activity (**Figure 26**).

The most outstanding example illustrating this strategy came from the team of Alain Friboulet and Daniel Thomas,[102] who produced anti-idiotype antibodies against a monoclonal antibody AE2 that was a competitive inhibitor of acetylcholine esterase. One of the selected antibodies, 9A8, catalyzes the hydrolysis of acetylthiocholine with a pseudo first-order rate constant $k_{cat} = 81\ s^{-1}$ and a factor of acceleration of 4.2×10^8. These remarkable parameters, which are only two orders of magnitude lower when compared to those of the enzyme, make abzyme 9A8 the most powerful abzyme known until now.

5.12.8 Other Approaches

To improve the substrate specificity of catalytic antibodies, a new modular assembly strategy has been developed, which relies on the assembly of a small peptide with enzyme-like qualities by combining a known catalytic peptide and fluorescein-binding peptide.[103] A 24 amino acid residue peptide (FT-YLK3: YKLLKELLAKLKWLLRKLLGPTCL) has been shown to catalyze the retro-aldol reaction, albeit with relatively poor substrate affinity ($K_M = 1.8\ mmol\ l^{-1}$).[104] In other work, a small 12 amino acid residue peptide (FluS303: YPNEFDWWDYYY) was identified as a binder for fluorescein.[105] These two peptides were covalently linked to afford peptide FluS303-FTYLK3, which catalyzed the retro-aldol reaction of a β-aldol substitued by a fluorescein-derived moiety, with a K_M of $8\ \mu mol\ l^{-1}$ and a k_{cat} of $2.3 \times 10^{-4}\ min^{-1}$. A similar hapten without the fluorophore moiety was also applied in order to examine the substrate recognition of the peptide FTYLK3. A fourfold higher rate acceleration of peptide-catalyzed retro-aldol reaction was observed for the fluorescein-derived substrate relative to that bearing no fluorophore, which clearly indicated an improved substrate specificity.

Immunopharmacotherapy has recently appeared as a viable treatment strategy for cocaine abuse using an animal model for relapse. Larsen *et al.*[106] were the first to construct a monoclonal antibody, GNC92H2, that was raised against hapten GNC (**Figure 27(a)**), and that was found to bind selectively to cocaine with respect to its metabolites with a K_d of $10^{-7}\ mol\ l^{-1}$ (**Figure 27(b)**).

In further work that aimed at getting antibodies that would be able to hydrolyze cocaine, Janda and co-workers[107] studied the importance of the linker moiety in the structure of haptens with regard to immunogenicity. In these studies, immunogen GNP72 and GNL73 were carefully evaluated (**Figure 27(a)**). GNP 72 employed a short linker for conjugation with carrier proteins. In comparison, the linker moiety of GNL 73 is extended by addition of a β-alanine. Surprisingly, this simple change in the linker structure imposed dramatically different immune responses during immunization. Conjugate 72-KLH afforded no antibodies with the desired catalytic activity, while conjugate 73-KLH successfully elicited several antibody catalysts, such as GNL3A6, GNL4D3, and GNL23A6. To fully investigate the importance of linker moiety, Janda *et al.* designed

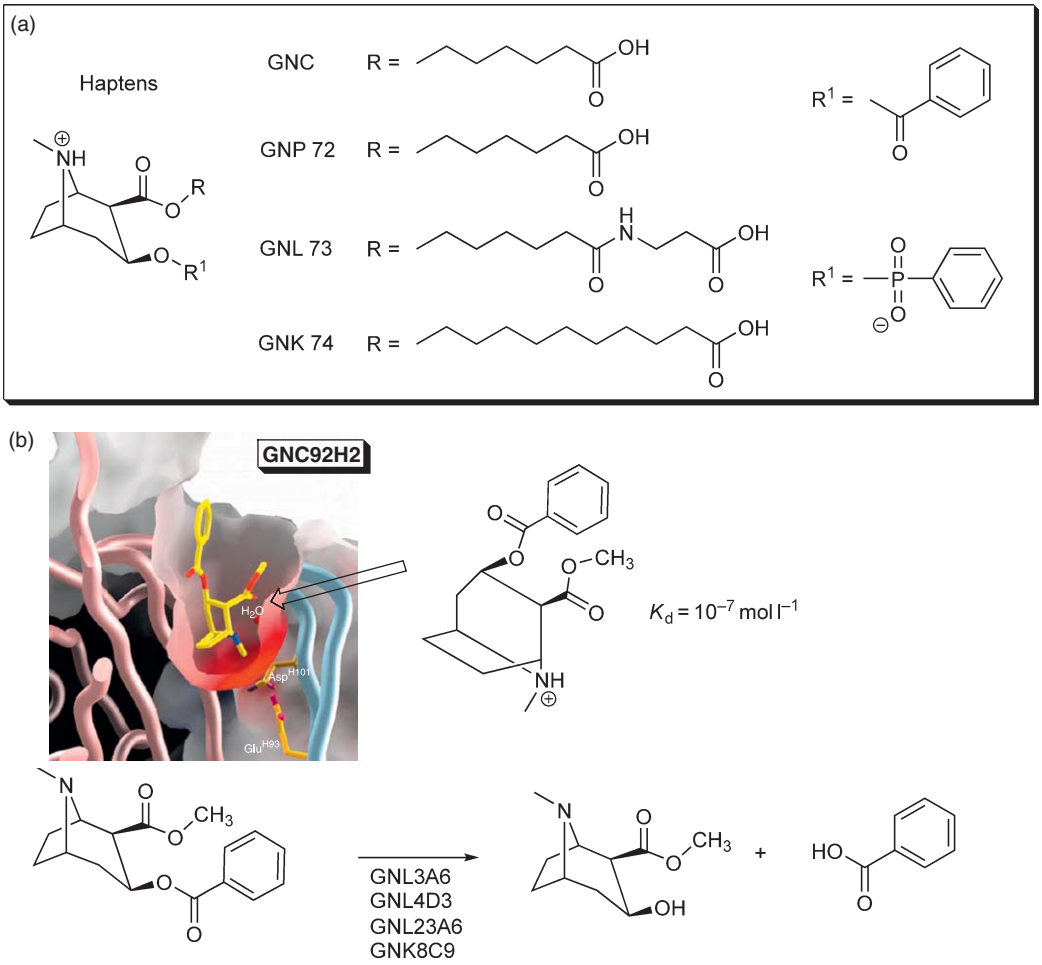


Figure 27 Various haptens used to elicit antibodies able to catalyze the hydrolysis of cocaine and X-ray structure of the GNC92H2 complex.

a third hapten, GNK 74, which is equipped with a linker of the same length as the one in 73 but without the amide functionality (**Figure 27(a)**). During immunization of 74-KLH, even though antibodies with catalytic capacity were obtained, they demonstrated a much lower activity. It was suggested that both the length of the linker and the precise position of the internal amide bond within the linker structure are required for a hapten to adequately trigger the defense mechanism of the adaptive immune system.

5.12.9 Conclusion

Since the first successful examples of tailored antibodies that catalyze chemical transformations, great attention has been drawn to this field of research. In the last two decades, a large number of antibodies that are able to catalyze a variety of chemical processes have been investigated and the mechanism of those abzymes has greatly been elucidated thanks to structural studies. Indeed, a good number of three-dimensional structures were solved during the past few years,[2] which made it possible to elucidate the mechanism whereby the catalytic antibodies function.

First, based on the concept of catalytic antibodies as first described by Jencks, the use of a TSA of the target reaction to elicit antibodies during an immune response has been developed and identified as a classic strategy in the search for catalytic antibodies. Structural studies have in particular shown that this strategy has led to

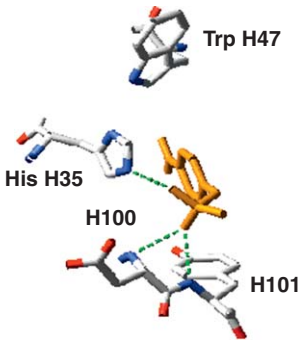


Figure 28 Stabilization of the phosphonate TSA for the hydrolysis of esters, by a network of H bonds involving H35 histidine and amide functions of the principal chain of antibody CNJ206.

antibodies that function by several typical mechanisms that are close to those displayed by enzymes but much less efficient. It has then appeared first that, in accordance with what was initially required, certain antibodies can stabilize the TS for a given reaction. It is the case of the antibodies catalyzing the hydrolysis of esters such as 43C9,[19,20] 48G7, 17E8,[23] and finally CNJ206,[108] which stabilizes, by a network of hydrogen bonds, implying the H35 histidine and the amide functions of the principal chain, the negative charge developed on the tetrahedral intermediate of the reaction and form, as in the esterases or the peptidases a real 'oxy-anion hole' (**Figure 28**).

The catalysis of bimolecular reactions such as the Diels–Alder reaction can be realized effectively and specifically by abzymes such as 1E9.[36] The antibody then acts as a template and maintains the two partners, diene and dienophile, in a position favorable to the formation of the TS that is in a boat conformation in this reaction (**Figure 8**).

Antibodies can also act as entropy traps by stabilizing a particular conformation of a substrate convenient for the formation of the TS. It is the case of 1F7 antibodies catalyzing the transformation of chorismate into prephenate[45] that stabilize, thanks to several hydrogen bonds and an ionic bond between an arginine (Arg H95) and a carboxylate of the substrate, the conformation of the chorismate that will give rise to the chair-like TS of this reaction (**Figure 11**).

Antibodies also well catalyze the reactions where a molecule should be forced to adopt a particular and reactive conformation, thanks to privileged interactions with the amino acids of the binding site. For example, abzymes with a ferrochelatase activity, 7G12, force the mesoporphyrin IX ring to adopt a distorted conformation favorable to the insertion of a Cu^{2+} ion in the center of the macrocycle,[97] thanks to an interaction with the H100c methionine which constrains one of the pyrrole rings to be left outside the plane of porphyrin (**Figure 24**).

The newer established 'bait and switch' and reactive immunization strategies allow for the recruitment of the desired amino acid functionalities in the catalytic machinery of the antibodies during immunization. However, the analysis of the known three-dimensional structures to date revealed that no abzyme had more than one catalytic residue in its active site. This has particularly been shown by the X-ray structure of antibody 4B2, that was obtained by the 'bait and switch' strategy, using a hapten carrying an amidinium function that allowed to generate a glutamate in the site of antibody 4B2 (**Figure 16**).[109]

Overall, from what we have learnt until now, it is clear that catalytic antibodies do not fulfill all the requirements for being the 'Taylor-made catalysts' that would catalyze any kind of reaction that a chemist would like to. It would be exaggerated however, to regard the abzymes as lures for chemists. It is simply necessary to ask them to do what they can do. One thus should not ask them to compete with the enzymes which, for comparable reactions, present much higher rate constants and factors of acceleration. The real hope for the chemist remains rather in the catalysis of reactions for which there exists no or little equivalent enzymes (we already mentioned the case of the Diels–Alder reaction[35–44] or in the induction of original selectivities for certain reactions of synthetic interest, as shown by two striking examples recently described in the literature: the use of an antibody (14D9) to catalyze the enantioselective cleavage of an enol ether in a key step of the total synthesis of the (−)-α-multistriatine pheromone[1] or the production by kinetic resolution reactions catalyzed by

the abzyme 38C2 of (+)-syn-β-hydroxyketones intermediates in the synthesis of the anticancer drugs epothilones A and C.[110] On the other hand, the lack of substrate selectivity of the antibody 38C2 confers on it an advantage when compared to aldolases since it can catalyze a much broader range of aldolic condensation reactions.[81] Finally, therapeutic applications can also be envisioned for abzymes. Thus, recent results have shown that 38C2 was able to activate a prodrug of the anticancer etoposide by retro-aldolization/retro-Michael reaction *in vivo*, to deliver this drug specifically to the tumoral cells of mice suffering from neuroblastoma.[111]

Even if the strategies used until now do not allow to obtain abzymes capable of competing with the enzymes, the mechanistic and structural studies provided a set of interesting informations on the enzymatic catalysis and its evolution, informations that are more difficult to obtain starting from natural enzymes optimized by directed evolution techniques. Such techniques should now be increasingly employed to improve the performances of the abzymes of first generation and to explore their limits and the chemist will have to rely probably on the biochemist for more powerful tools. However, the creativity of the chemist could still be put in profit for the design of more elaborate haptens, as testifies the recent appearance of the first antibody produced against a zwitterionic hapten in order to generate two catalytic residues in the active site of a single antibody.[112]

Other approaches could probably be envisioned in a near future. Some of them have been recently developed to improve the performances of antibodies. For example, improved hapten carrier linkages that can greatly enhance the immune response to otherwise poorly immunogenic haptens have been developed. The application of cofactors has also been shown to be a valuable strategy to yield antibodies with desired catalytic capabilities. Finally, the employment of chemosensor and phage display systems have greatly advanced the efficiency of the screening process for detection of antibody catalysts. Hopefully, these advancements, as well as future breakthroughs in this field of research, will facilitate the discovery of catalytic antibodies that demonstrate catalytic efficiencies comparable with natural enzymes. Furthermore, we anticipate the generation of more robust antibody catalysts that could provide synthetic chemists with novel tools to target challenging problems.

References

1. J. D. Stevenson; N. R. Thomas, *Nat. Prod. Rep.* **2000**, *17*, 535.
2. B. Golinelli-Pimpaneau, *Curr. Opin. Struct. Biol.* **2000**, *10*, 697.
3. D. Hilvert, *Annu. Rev. Biochem.* **2000**, *69*, 751.
4. J. L. Reymond, *Top. Curr. Chem.* **1999**, *200*, 59–93.
5. D. B. Smithrud; S. J. Benkovic, *Curr. Opin. Biotechnol.* **1997**, *8*, 459.
6. N. R. Thomas, *Nat. Prod. Rep.* **1996**, *13*, 479.
7. A. J. Kirby, *Acta Chem. Scand.* **1996**, *50*, 203.
8. P. G. Schultz; R. A. Lerner, *Science* **1995**, *269*, 1835.
9. L. Pauling, *Am. Sci.* **1948**, *36*, 51.
10. W. P. Jencks, *Catalysis in Chemistry and Enzymology*; McGraw-Hill: New York, 1969, p 288.
11. A. Tramontano; K. D. Janda; R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6736.
12. S. J. Pollack; J. W. Jacobs; P. G. Schultz, *Science* **1986**, *234*, 1570.
13. H. Eyring, *J. Chem. Phys.* **1935**, *3*, 107.
14. H. Eyring, *Chem. Rev.* **1935**, *17*, 65.
15. F. Tanaka, *Chem. Rev.* **2002**, *102*, 4885.
16. J. D. Stewart; S. J. Benkovic, *Nature* **1995**, *375*, 388.
17. M. M. Mader; P. A. Bartlett, *Chem. Rev.* **1997**, *97*, 1281.
18. K. Katayama; T. Kobayashi; H. Oikawa; M. Honma; A. Ichihara, *Biochim. Biophys. Acta* **1998**, *1384*, 387.
19. K. D. Janda; D. Schloeder; S. J. Benkovic; R. A. Lerner, *Science* **1988**, *241*, 1188.
20. J. D. Stewart; J. F. Krebs; G. Siuzdak; A. J. Berdis; D. B. Smithrud; S. J. Benkovic, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7404.
21. V. A. Roberts; J. Stewart; S. J. Benkovic; E. D. Getzoff, *Protein Eng.* **1993**, *6*, 85.
22. V. A. Roberts; J. Stewart; S. J. Benkovic; E. D. Getzoff, *J. Mol. Biol.* **1994**, *235*, 1098.
23. M. M. Thayer; E. H. Olender; A. S. Arvai; C. K. Koike; I. L. Canestrelli; J. D. Stewart; S. J. Benkovic; E. D. Getzoff; V. A. Roberts, *J. Mol. Biol.* **1999**, *291*, 329.
24. L. T. Chong; P. Bandyopadhyay; T. S. Scanlan; I. D. Kuntz; P. A. Kollman, *J. Comput. Chem.* **2003**, *24*, 1371.
25. R. Aggarwal; F. Benedetti; F. Berti; S. Buchini; A. Colombatti; F. Dinon; V. Galasso; S. Norbedo, *Chem. Eur. J.* **2003**, *9*, 3132.
26. J. Hasserodt; K. D. Janda; R. A. Lerner, *J. Am. Chem. Soc.* **1997**, *119*, 5993.
27. C. M. Paschall; J. Hasserodt; T. Jones; R. A. Lerner; K. D. Janda; D. W. Christianson, *Angew. Chem., Int. Ed.* **1999**, *38*, 1743.
28. T. Y. Li; K. D. Janda; J. A. Ashley; R. A. Lerner, *Science* **1994**, *264*, 1289.
29. X. Y. Zhu; A. Heine; F. Monnat; K. N. Houk; K. D. Janda; I. A. Wilson, *J. Mol. Biol.* **2003**, *329*, 69.
30. J. E. Baldwin, *J. Chem. Soc., Chem. Commun.* **1976**, (18), 734.
31. J. E. Baldwin, *J. Chem. Soc., Chem. Commun.* **1976**, (18), 738.

32. K. D. Janda; C. G. Shevlin; R. A. Lerner, *Science* **1993**, *259*, 490.
33. K. Gruber; B. Zhou; K. N. Houk; R. A. Lerner; C. G. Shevlin; I. A. Wilson, *Biochemistry* **1999**, *38*, 7062.
34. J. Sauer, *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 211.
35. D. Hilvert; K. W. Hill; K. D. Nared; M. T. M. Auditor, *J. Am. Chem. Soc.* **1989**, *111*, 9261.
36. J. G. Chen; Q. L. Deng; R. X. Wang; K. N. Houk; D. Hilvert, *Chembiochem* **2000**, *1*, 255.
37. S. P. Kim; A. G. Leach; K. N. Houk, *J. Org. Chem.* **2002**, *67*, 4250.
38. A. C. Braisted; P. G. Schultz, *J. Am. Chem. Soc.* **1990**, *112*, 7430.
39. H. D. Ulrich; P. A. Patten; P. L. Yang; F. E. Romesberg; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11907.
40. F. E. Romesberg; B. Spiller; P. G. Schultz; R. C. Stevens, *Science* **1998**, *279*, 1929.
41. X. Y. Zhang; Q. L. Deng; S. H. Yoo; K. N. Houk, *J. Org. Chem.* **2002**, *67*, 9043.
42. J. T. Ylikauhaluoma; J. A. Ashley; C. H. Lo; L. Tucker; M. M. Wolfe; K. D. Janda, *J. Am. Chem. Soc.* **1995**, *117*, 7041.
43. A. Heine; E. A. Stura; J. T. Yli-Kauhaluoma; C. S. Gao; Q. L. Deng; B. R. Beno; K. N. Houk; K. D. Janda; I. A. Wilson, *Science* **1998**, *279*, 1934.
44. C. E. Cannizzaro; J. A. Ashley; K. D. Janda; K. N. Houk, *J. Am. Chem. Soc.* **2003**, *125*, 2489.
45. M. R. Haynes; E. A. Stura; D. Hilvert; I. A. Wilson, *Science* **1994**, *263*, 646.
46. W. V. E. Doering; W. R. Roth, *Tetrahedron* **1962**, *18*, 67.
47. M. J. S. Dewar; L. E. Wade, *J. Am. Chem. Soc.* **1977**, *99*, 4417.
48. H. D. Ulrich; E. Mundroff; B. D. Santarsiero; E. M. Driggers; R. C. Stevens; P. G. Schultz, *Nature* **1997**, *389*, 271.
49. E. C. Mundorff; M. A. Hanson; A. Varvak; H. Ulrich; P. G. Schulz; R. C. Stevens, *Biochemistry* **2000**, *39*, 627.
50. T. Asada; H. Gouda; P. A. Kollman, *J. Am. Chem. Soc.* **2002**, *124*, 12535.
51. M. G. Rossman; A. Liljas; C. I. Branden; L. J. Banaszak, Evolutionary and Structural Relationship among Dehydrogenases. In *The Enzymes*; Academic Press: New York, 1975; Vol. 11.
52. J. Schroer; M. Sanner; J. L. Reymond; R. A. Lerner, *J. Org. Chem.* **1997**, *62*, 3220.
53. A. Radzicka; R. Wolfenden, *Science* **1995**, *267*, 90.
54. G. MacBeath; D. Hilvert, *Chem. Biol.* **1996**, *3*, 433.
55. A. Tramontano; G. Gololobov; S. Paul, Proteolytic Antibodies: Origins, Selection and Induction. In *Chemical Immunology, Catalytic Antibodies*; S. Paul, Ed.; S. Karger AG: Basel, Switzerland, 2000; Vol. 77; p 1.
56. J. D. Stewart; S. J. Benkovic, *Chem. Soc. Rev.* **1993**, *22*, 213.
57. P. A. Patten; N. S. Gray; P. L. Yang; C. B. Marks; G. J. Wedemayer; J. J. Boniface; R. C. Stevens; G. Schultz, *Science* **1996**, *271*, 1086.
58. A. Tramontano; B. Ivanov; G. Gololobov; S. Paul, *Appl. Biochem. Biotechnol.* **2000**, *83*, 233.
59. M. Barbany; H. Gutierrez-de-Teran; F. Sanz; J. Villa-Freixa; A. Warshel, *ChemBioChem* **2003**, *4*, 277.
60. J. E. Hanson; A. P. Kaplan; P. A. Bartlett, *Biochemistry* **1989**, *28*, 6249.
61. B. P. Morgan; J. M. Scholtz; M. Ballinger; I. Zipkin; P. A. Bartlett, *J. Am. Chem. Soc.* **1991**, *113*, 297.
62. O. Wiest; K. N. Houk, *J. Org. Chem.* **1994**, *59*, 7582.
63. D. J. Tantillo; K. N. Houk, *J. Org. Chem.* **1999**, *64*, 3066.
64. K. N. Houk; A. G. Leach; S. P. Kim; X. Y. Zhang, *Angew. Chem., Int. Ed.* **2003**, *42*, 4872.
65. G. M. Blackburn; A. Datta; H. Denham; P. Wentworth, Jr., *Adv. Phys. Org. Chem.* **1998**, *31*, 249.
66. A. Warshel, *Acc. Chem. Res.* **1981**, *14*, 284.
67. G. Yang; J. Chun; H. Arakawa-Uramoto; X. Wang; M. A. Gawinowicz; K. Zhao; D. W. Landry, *J. Am. Chem. Soc.* **1996**, *118*, 5881.
68. D. Hilvert, *Annu. Rev. Biochem.* **2000**, *69*, 751.
69. K. M. Shokat; C. J. Leumann; R. Sugasawara; P. G. Schultz, *Nature* **1989**, *338*, 269.
70. K. D. Janda; M. I. Weinhouse; D. M. Schloeder; R. A. Lerner; S. J. Benkovic, *J. Am. Chem. Soc.* **1990**, *112*, 1274.
71. P. Wentworth; Y. Q. Liu; A. D. Wentworth; P. Fan; M. J. Foley; K. D. Janda, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5971.
72. S. N. Thorn; R. G. Daniels; M. T. Auditor; D. Hilvert, *Nature* **1995**, *373*, 228.
73. O. Goncalves; T. Dintinger; J. Lebreton; D. Blanchard; C. Tellier, *Biochem. J.* **2000**, *346*, 691.
74. F. Tanaka; C. F. Barbas, *J. Immunol. Methods* **2002**, *269*, 67.
75. P. Wirsching; J. A. Ashley; C. H. L. Lo; K. D. Janda; R. A. Lerner, *Science* **1995**, *270*, 1775.
76. C. H. L. Lo; P. Wentworth; K. W. Jung; J. Yoon; J. A. Ashley; K. D. Janda, *J. Am. Chem. Soc.* **1997**, *119*, 10251.
77. A. Datta; P. Wentworth; J. P. Shaw; A. Simeonov; K. D. Janda, *J. Am. Chem. Soc.* **1999**, *121*, 10461.
78. D. W. Chen; R. J. Kubiak; J. A. Ashley; K. D. Janda, *J. Chem. Soc., Perkin Trans.* **2001**, *1*, 2796.
79. C. H. Lin; T. Z. Hoffman; Y. L. Xie; P. Wirsching; K. D. Janda, *Chem. Commun.* **1998**, (10), 1075.
80. J. Wagner; R. A. Lerner; C. F. Barbas, *Science* **1995**, *270*, 1797.
81. C. F. Barbas; A. Heine; G. F. Zhong; T. Hoffmann; S. Gramatikova; R. Bjornestedt; B. List; J. Anderson; E. A. Stura; I. A. Wilson; R. A. Lerner, *Science* **1997**, *278*, 2085.
82. S. C. Sinha; J. Sun; G. P. Miller; M. Wartmann; R. A. Lerner, *Chem. Eur. J.* **2001**, *7*, 1691.
83. C. Rader; J. M. Turner; A. Heine; D. Shabat; S. C. Sinha; I. A. Wilson; R. A. Lerner; C. F. Barbas, *J. Mol. Biol.* **2003**, *332*, 889.
84. G. F. Zhong; R. A. Lerner; C. F. Barbas, *Angew. Chem., Int. Ed.* **1999**, *38*, 3738.
85. F. Tanaka; R. A. Lerner; C. F. Barbas, *J. Am. Chem. Soc.* **2000**, *122*, 4835.
86. J. Fisher; R. L. Charnas; S. M. Bradley; J. R. Knowles, *Biochemistry* **1981**, *20*, 2726.
87. D. G. Brenner; J. R. Knowles, *Biochemistry* **1981**, *20*, 3680.
88. R. B. Silverman, *Mechanism-Based Enzyme Inactivation: Chemistry and Enzymology,* CRC Press: Boca Raton, FL, 1998; Vol. 1.
89. F. Tanaka; H. Almer; R. A. Lerner; C. F. Barbas, *Tetrahedron Lett.* **1999**, *40*, 8063.
90. F. Tanaka; M. Oda; I. Fujii, *Tetrahedron Lett.* **1998**, *39*, 5057.
91. T. Kimura; V. P. Vassilev; G. J. Shen; C. H. Wong, *J. Am. Chem. Soc.* **1997**, *119*, 11734.
92. K. M. Nicholas; P. Wentworth; C. W. Harwig; A. D. Wentworth; A. Shafton; K. D. Janda, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2648.

93. F. Tanaka; R. A. Lerner; C. F. Barbas, *Chem. Commun.* **1999**, (15), 1383.
94. M. G. Finn; R. A. Lerner; C. F. Barbas, *J. Am. Chem. Soc.* **1998**, *120*, 2963.
95. R. Quilez; S. de Lauzon; B. Desfosses; D. Mansuy; J. P. Mahy, *FEBS Lett.* **1996**, *395*, 73.
96. A. G. Cochran; P. G. Schultz, *J. Am. Chem. Soc.* **1990**, *112*, 9414.
97. F. E. Romesberg; B. D. Santarsiero; B. Spiller; J. Yin; D. Barnes; P. G. Schultz; R. C. Stevens, *Biochemistry* **1998**, *37*, 14404.
98. S. Nimri; E. Keinan, *J. Am. Chem. Soc.* **1999**, *121*, 8978.
99. R. Ricoux; E. Lukowska; F. Pezzotti; J. P. Mahy, *Eur. J. Biochem.* **2004**, *271*, 1277.
100. R. Ricoux; E. Girgenti; H. Sauriat-Dorizon; D. Blanchard; J. P. Mahy, *J. Protein Chem.* **2002**, *21*, 471.
101. N. K. Jerne, *Ann. Immunol.* **1974**, *125C*, 373.
102. L. Izadar; A. Friboulet; M. H. Remy; A. Roseto; D. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8876.
103. F. Tanaka; C. F. Barbas, *J. Am. Chem. Soc.* **2002**, *124*, 3510.
104. F. Tanaka; C. F. Barbas, *Chem. Commun.* **2001**, *9*, 769.
105. M. N. Rozinov; G. P. Nolan, *Chem. Biol.* **1998**, *5*, 713.
106. N. A. Larsen; B. Zhou; A. Heine; P. Wirsching; K. D. Janda; I. A. Wilson, *J. Mol. Biol.* **2001**, *311*, 9.
107. M. Matsushita; T. Z. Hoffman; J. A. Ashley; B. Zhou; P. Wirsching; K. D. Janda, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 87.
108. J. B. Charbonnier; E. Carpenter; B. Gigant; B. Golinelli-Pimpaneau; Z. Eshhar; B. S. Green; M. Knossow, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11721.
109. B. Golinelli-Pimpaneau; O. Goncalves; T. Dintinger; D. Blanchard; M. Knossow; C. Tellier, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9892.
110. C. F. Barbas; A. Heine; G. Zhong; T. Hoffmann; S. Gramatikova; R. Bjornestedt; B. List; J. Anderson; E. A. Stura; I. A. Wilson; R. A. Lerner, *Science* **1997**, *278*, 2085.
111. D. Shabat; H. N. Lode; U. Pertl; R. A. Reisfeld; C. Rader; R. A. Lerner; C. F. Barbas, III, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7528.
112. T. Tsumuraya; N. Takazawa; A. Tsunakawa; R. Fleck; S. Masamune, *Chemistry* **2001**, *7*, 3748.

Biographical Sketches

Rémy Ricoux, 39 years old, is currently Ingénieur d'Etudes CNRS in the Laboratory of Bioorganic and Bioinorganic Chemistry at the Institute of Molecular Chemistry and Materials at Orsay (ICMMO) at the University of Paris 11 (France).

Dr. Ricoux prepared his Ph.D. in the Laboratory of Pharmacological and Toxicological Chemistry and Biochemistry (Director Dr. D. Mansuy), at Paris 5, on a bioinorganic subject dealing with the production and reactivity studies of catalytic antibodies with a metalloporphyrin cofactor or 'hemoabzymes.' He received his Ph.D. from the University of Paris 12 in 2001.

Upon completion of his Ph.D., he was appointed in 2001 as a 'Ingenieur in the CNRS' in the Laboratory of Bioorganic and Bioinorganic Chemistry, at the Institute of Molecular Chemistry and Materials at Orsay (ICMMO) at the University of Paris 11 (France), where he first went on to develop the field of hemoabzymes.

He got a 'Habilitation à diriger les recherches' in 2008 and he is now developing his own project that consists of the elaboration of new hybrid metalloprotein catalysts for selective oxidation reactions, by insertion of metal cofactors into xylanases. He then studies their peroxidase, catalase, and monooxygenase activities, in particular in the selective oxidation of sulfides, alkanes, and alkenes.

He is author of 25 publications and has received an award from the CNRS (Cristal of the CNRS). He was also nominated for the FEBS Journal Prize in 2004.

Jean-Pierre Mahy, 52 years old, is currently a professor of Bioorganic and Bioinorganic Chemistry at the University of Paris 11. He is also vice-chairman of the Institute of Molecular Chemistry and Materials at Orsay (ICMMO) and head of the Laboratory of Bioorganic and Bioinorganic Chemistry at the University of Paris 11 (Orsay, France).

Professor Mahy received his Ph.D. in 1990 from the University of Paris 6 on a bioinorganic subject dealing with the importance of the complexes with an Fe–N bond in the chemistry of iron-porphyrin and hemoproteins.

After a postdoc in the US (Professor S. J. Benkovic, Pennsylvania State University) where he worked on catalytic antibodies, he was appointed in 1996 as 'Directeur de Recherches in the CNRS' in the Laboratory of Pharmacological and Toxicological Chemistry and Biochemistry (Director Dr. D. Mansuy), at Paris 5, where he developed the field of 'hemoabzymes: catalytic antibodies with an iron-porphyrin cofactor.'

In 1999, he moved to the University of Paris 11, where he was appointed as a professor and the head of LCBB. Since 2006 he is also the vice-chairman of ICMMO.

He teaches courses in bioorganic and bioinorganic chemistry (in particular on the role of metals in biology and on the iron-heme enzymes such as peroxidases and cytochromes P-450 monooxygenases).

His main fields of research are:

- Elaboration of new hybrid metalloproteins catalysts for selective oxidation reactions.
- Peroxidase, catalase, and monooxygenase activities of the hybrid metalloproteins.
- Selective oxidation of sulfides, alkanes, and alkenes using iron biocatalysts and mimics.

He is author of 70 publications and has received awards from the CNRS (bronze medal) and the French Academy of Sciences (Berthault prize); he has also been awarded 'Doctor Honoris Causa' from the University of Oradea (Romania).

5.13 Recent Progress on Understanding Ribosomal Protein Synthesis

Ute Kothe, University of Lethbridge, Lethbridge, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

5.13.1 Introduction

In each living cell, proteins are synthesized by the ribosome according to the information encoded in the genome. Thus, ribosomal protein synthesis constitutes the ultimate step in gene expression. The genetic information is delivered to the ribosome in the form of a messenger RNA (mRNA), a copy of the gene produced during transcription. This mRNA is then translated into a functional protein by the ribosome. Both transcription and translation are amplifying steps since several mRNAs are synthesized from one gene and several proteins are made from one mRNA. Thereby, the ribosome produces a large number of identical proteins based on a single gene. At the same time, the ribosome can synthesize many different proteins when programed by different mRNAs. Together with RNA-polymerases, the ribosome is one of the most important cellular machines as it is responsible for generating one of the largest classes of biomolecules, namely the proteins.

Being a fundamental cellular process, ribosomal protein synthesis is highly conserved. All living cells, from simple bacteria to plants and humans, contain ribosomes of similar structure that perform similar steps during protein synthesis. This allows us to study ribosome function in easy-to-handle organisms such as *Escherichia coli* or yeast while obtaining knowledge relevant to all forms of life. For this reason, we focus in this chapter on recent findings on prokaryotic protein synthesis, most of which also hold true for eukaryotic organisms. The ribosome consists mainly of RNA and might be a very ancient biomachine that enabled the transition from RNA-based life to today's cells in which proteins are the major players. Thus, ribosomal RNA is the most conserved component of the ribosome. This is used widely also in phylogenetic studies since the sequence of ribosomal RNA can be easily obtained even if the complete genome of an organism is not sequenced.[1]

Several distinct steps are required in addition to the actual formation of a peptide bond to allow for ribosomal protein synthesis directed by genetic information (**Figure 1**). In principle these steps are dictated by the nature of the mRNA. Most importantly, the mRNA contains the coding sequence where each codon (i.e., three nucleotides) encodes for one amino acid. Before and after the coding sequence, there are untranslated regions, which, among other functions, allow the ribosome to bind to the mRNA. These regions often also contain regulatory signals influencing the expression level of the specific mRNA. The ribosome first has to bind to the mRNA in a region preceding the coding sequence and has to identify the first codon, called the start codon, in a process termed as initiation. In this stage, the first amino acid is also bound to the ribosome in the form of an aminoacyl-tRNA. Typically, cellular proteins begin with methionine and a specialized initiator Met-$tRNA_i^{Met}$ is used as the starting tRNA that recognizes the AUG start codon on the mRNA. In prokaryotes, the start codon is defined by a nearby ribosome-binding site on the mRNA, the so-called Shine–Dalgarno sequence that can base pair with the ribosomal RNA. This process is facilitated by three initiation factors

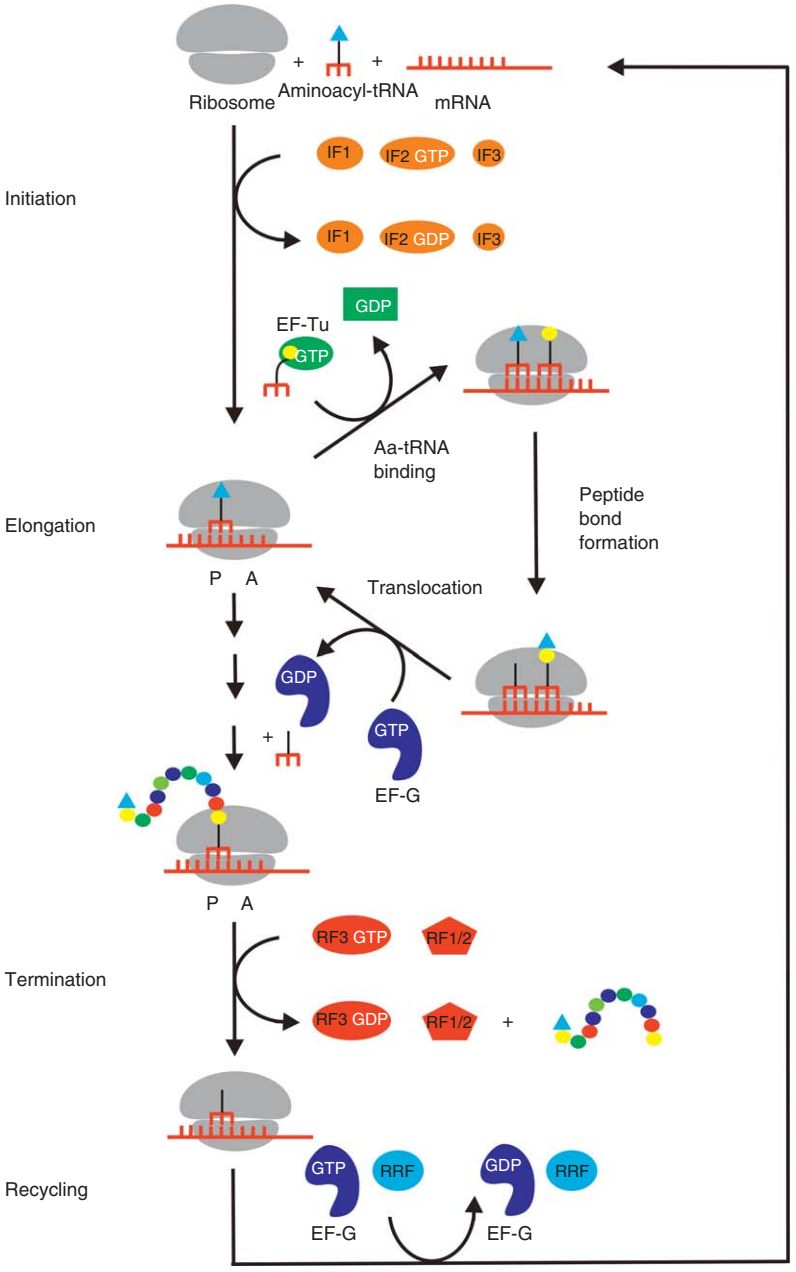


Figure 1 The four stages of ribosomal protein synthesis. During initiation, the mRNA and the first aminoacyl-tRNA are positioned on the ribosome with the help of three initiation factors (IFs). Next, one amino acid at a time is added to the growing peptide according to the genetic information contained in the mRNA during the elongation cycle, which is repeated many times. Therein, elongation factor Tu (EF-Tu) delivers the next aminoacyl-tRNA to the ribosome, which is followed by peptide bond formation and EF-G-catalyzed translocation of the mRNA–tRNA complex within the ribosome. The complete peptide is then set free during termination, which is facilitated by two release factors (RFs). Subsequently, mRNA and tRNA dissociation from the ribosome is catalyzed by elongation factor G and ribosome recycling factor (RRF) in the recycling phase.

(IF1–3). In contrast, eukaryotic initiation is a rather complex process involving a large number of initiation factors (eIFs, **Table 1**). This is also the stage of eukaryotic ribosomal protein synthesis, which is most highly regulated to achieve differential protein expression. Elaborating the details of eukaryotic initiation is beyond the scope of this chapter.

Table 1 Translation factors in prokaryotes and eukaryotes

Translation phase	*Prokaryotes*	*Eukaryotes*
Initiation	Initiation factor 1 (IF1) Initiation factor 2 (IF2) Initiation factor 3 (IF3)	~12 eIFs
Elongation	Elongation factor Tu (EF-Tu) Elongation factor Ts (EF-Ts) Elongation factor G (EF-G)	eEF1α eEF1β, γ, δ eEF2 eEF3 (in fungi)
Termination	Release factor 1 (RF1) Release factor 2 (RF2) Release factor 3 (RF3)	eRF1 eRF3
Recycling	EF-G Ribosome recycling factor (RRF)	eIF3, eIF1, eIF1A, eIF3j

Actual protein synthesis commences once the ribosome has bound an mRNA start codon and the initiator tRNA in its P site. It takes place in a cyclic fashion where one amino acid at a time is added to the C-terminus of the growing peptide. This stage is referred to as elongation (**Figure 1**). In this chapter, the main focus will be on the elongation stage since it is the most conserved and best-understood core process of ribosomal protein synthesis. In brief, the aminoacyl-tRNAs are delivered to the ribosomal A site by prokaryotic elongation factor Tu (EF-Tu) (see **Table 1** for eukaryotic translation factors). It is critical that only correct aminoacyl-tRNAs that match the codon presented in the ribosomal A site are accepted by the ribosome. Incorporation of wrong amino acids into the growing peptide ultimately leads to nonfunctional proteins, which can result in cell death. The high accuracy of the ribosome in discriminating against wrong aminoacyl-tRNAs is remarkable, and GTP hydrolysis by EF-Tu critically contributes to the fidelity of aminoacyl-tRNA selection. After the aminoacyl end of the tRNA is bound to the ribosome, peptide bond formation is catalyzed by the ribosome. Therein, the amino group of the incoming aminoacyl-tRNA undergoes a nucleophilic attack on the ester bond linking the peptide to the P-site tRNA. Thus, the already synthesized peptide is elongated by one amino acid and transferred onto the A-site tRNA. During the last few years, substantial insight has been gained into the details of how the ribosome catalyzes peptidyl transfer. As a result of peptide bond formation, the P-site-bound tRNA is deacylated while the A-site tRNA is connected to the growing peptide. In order to prepare the ribosome for a new elongation cycle, the tRNA–mRNA complex has to move relative to the ribosome to present the next codon in the A site. This translocation event is catalyzed by elongation factor G (EF-G), which also hydrolyzes GTP. During the large-scale movement of tRNAs and mRNA, the reading frame on the mRNA has to be maintained. Thus, it is important that exactly three nucleotides, that is, the next codon, move into the ribosomal A site and that the tRNAs do not slip on the mRNA. While the peptidyl-tRNA moves from the A site to the P site, the deacylated tRNA is first translocated from the P site to the E site and then dissociates from the ribosome. Ultimately, the elongation cycle is repeated many times until the peptide encoded by the mRNA is completely synthesized. Compared to the other stages of ribosomal protein synthesis, elongation takes the longest time. Thus, it is optimized for speed to ensure that protein synthesis takes place at a rate that is compatible with rapid cell division. For example, *E. coli* cells divide about once every 20 min under optimal conditions and have to double their protein content in this time.

Ribosomal protein synthesis is terminated once the ribosome reaches a specific stop codon on the mRNA (UAG, UGA, or UAA). The stop codon, which is followed by a 3′ untranslated region, signals to the ribosome to terminate its movement on the mRNA. Most importantly, the peptide is set free from the peptidyl-tRNA during termination. This step depends on another set of soluble proteins, the release factors (**Figure 1**). First, one release factor (RF1 or RF2), called class I release factor, binds into the ribosome, specifically recognizes the stop codon, and induces hydrolysis of the peptidyl-tRNA – a reaction we are just beginning to understand. Second, another release factor, RF3 or class II release factor, removes RF1 or RF2 from the ribosome while hydrolyzing GTP. Thus, the ribosome remains bound to an mRNA and a deacylated tRNA in its P site after termination. This step is followed by a recycling phase that is well understood only for prokaryotes. Therein,

EF-G and the ribosome recycling factor (RRF) catalyze dissociation of the tRNA and the mRNA from the ribosome as well as dissociation of the two ribosomal subunits. Consequently, the ribosome as well as the mRNA are prepared for a new round of protein synthesis.

In the following, the most significant recent findings on ribosomal protein synthesis will be discussed. We will focus on the elongation and termination phases as the most conserved stages of ribosomal protein synthesis. We will see that the last decade has dramatically improved our understanding of ribosome function on a molecular level. This knowledge could only be achieved by a combination of structural and functional studies using modern biophysical techniques.

5.13.2 Ribosome Structure

The ribosome is a unique cellular machine in that its main functional component is RNA whereas proteins seem to play only a structural role. For a long time, it has been debated whether RNA or proteins contribute most to the ribosome's function. With the determination of high-resolution crystal structures, this question could finally be answered. Clearly, these structures have revolutionized the field of ribosome studies. Already in the 1980s, Yonath and coworkers had grown crystals of active ribosomes that diffracted to about 0.6 nm (6 Å) (1 Å = 0.1 nm) resolution.[2] However, owing to the large size of the ribosome of about 2 500 000 Da (1 Da = 1 g mol^{-1}), the ribosome structure was not solved to atomic resolution until the year 2000.

Even before the ribosome structure had been determined, it was known that the ribosome is a rather complex biomolecular machine. Not only is the ribosome very large with a diameter of 25 nm, but it also consists of three different ribosomal RNAs (16S, 23S, and 5S rRNA) and about 50 different ribosomal proteins. The 70S ribosome is organized in two subunits called the small or 30S subunit, and the large or 50S subunit (S stands for Svedberg, the sedimentation coefficient used, e.g., to characterize ribosomal subunits and ribosomal RNA). The first ribosome structures that were solved at high resolution were the structures of isolated subunits.[3–5] Almost simultaneously, Ramakrishnan and coworkers reported the crystal structure of the small subunit from *Thermus thermophilus* at 0.3 nm (3 Å) resolution[5] while Steitz and coworkers solved the structure of the large subunit from *Haloarcula marismortui* at 0.24 nm (2.4 Å) resolution in 2000.[3] Shortly thereafter, Yonath and coworkers also published the small subunit structure from *T. thermophilus*[4] and then a large subunit structure from *Deinococcus radiodurans*.[6] At about the same time, Noller and coworkers determined the structure of the *T. thermophilus* 70S ribosome at 0.55 nm (5.5 Å) resolution.[7] These achievements constitute a major breakthrough as the ribosome structure was the largest crystal structure that had been solved at that time. However, the fact that ribosomes from different organisms had been used caused difficulties in evaluating subtle differences between the structures and in relating the new structural information to biochemical findings obtained with *E. coli*, despite the high conservation of most ribosomal features. This problem was overcome by the 70S crystal structure of the *E. coli* ribosome determined at 0.35 nm (3.5 Å) resolution by Cate and coworkers in 2005.[8] The most recent milestones in ribosome structure determination were a significant improvement in resolution when Ramakrishnan and coworkers published the *T. thermophilus* 70S structure at 0.28 nm (2.8 Å) resolution in 2006.[9] Also, Noller and coworkers achieved to solve the crystal structure of the 70S ribosome bound to RF1 at 0.32 nm (3.2 Å) resolution in 2008, which is the first 70S structure with bound factor at high resolution.[10]

The major finding from the ribosome crystal structures is that the ribosomal core consists mainly of RNA while the proteins are mostly found at the surface of the ribosome – this is true for both the small and the large subunits (**Figure 2**). The 30S ribosomal subunit is clearly structured in different domains that correspond to previously identified domains in 16S rRNA secondary structure. The 30S domains seem to be movable relative to each other during translation. In contrast, the 50S subunit folds into a large rigid body where different parts of the 23S rRNA interact with each other. Of the proteins on the surface of the 50S subunit, L10 and L7/12 could not be resolved within the subunit structure. Later, independent studies showed that the L7/12 stalk formed by these proteins is a highly dynamic extension to the large ribosomal subunit that reaches far into solution recruiting the translation factors to the ribosome.[11]

The interface between the small and large ribosomal subunits is built predominantly of RNA. Thus, the two subunits interact through various intersubunit bridges formed by RNA. Only the interactions at the outside of the

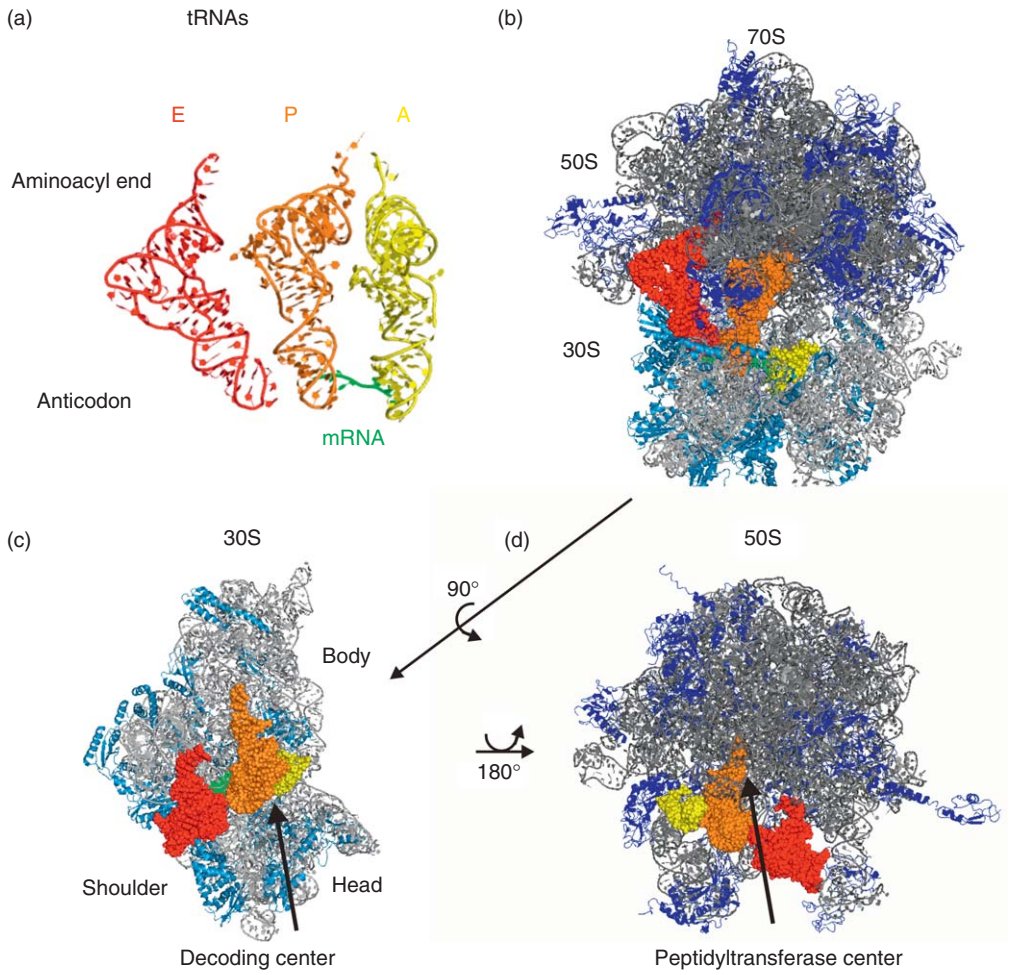


Figure 2 Three-dimensional structure of the ribosome. (a) The relative position of the three tRNAs bound to the E, P, and A site of the ribosome (not shown) (Protein Data Bank ID code 1GIX).[7] While the anticodons of the P- and A-site tRNA (orange and yellow) interact with the mRNA (green) in the decoding center of the 30S subunit, their aminoacyl ends are positioned closely to each other in the peptidyltransferase center of the 50S subunit. The E-site tRNA (red) no longer interacts with the mRNA and its aminoacyl end is bound to a distant site on the 50S subunit. (b) High-resolution crystal structure of the 70S ribosome (Protein Data Bank ID code 2J00 and 2J01).[9] The small 30S subunit is shown at the bottom with 16S rRNA in silver and proteins in sky blue. At the top, the large 50S subunit is seen with 23S and 5S rRNA in gray and proteins in blue. The E-, P-, and A-site tRNAs are shown in space filled in red, orange, and yellow, respectively (same orientation as in (a)). Only the anticodon stem, but not the aminoacyl end of the A-site tRNA is visible in this crystal structure. (c) The 30S subunit of the ribosome presented in (b) as seen from the interface between the subunits. (d) The 50S subunit as seen from the interface. For clarity the tRNAs are shown with both subunits. This figure was generated with Pymol (http://www.pymol.org).

interface involve proteins as well as RNAs. While the two subunits themselves are rather compact, there is significant space between the two subunits in the inside of the ribosome to accommodate the tRNAs in the A, P, and E sites. The anticodon stems of the tRNAs interact with the small ribosomal subunit while the elbow region and aminoacyl stem of the tRNAs contact the large ribosomal subunit (**Figure 2**). As the ribosome has to interact with about 40 different tRNAs that successively move through the ribosome during protein synthesis, the interactions are not sequence specific, but rather involve the backbone and the invariant 3′CCA end of the tRNAs.[12] The ribosome crystal structures also revealed the path of mRNA through the small ribosomal subunit at low resolution.[13] While two codons are presented in the A and P sites of the decoding center, the rest of the mRNA is buried in the small ribosomal subunit as the mRNA wraps around the neck of the 30S subunit in a channel. Overall, the ribosome binds only to about 30 nucleotides of the mRNA, which is progressively threaded through the ribosome during elongation.

The ribosome has two main functional centers. While the small subunit contains the decoding center where tRNA and mRNA interact, the large subunit forms the peptidyltransferase center that catalyzes peptide bond formation between aminoacyl- and peptidyl-tRNAs. The ribosome crystal structures have dramatically improved our understanding of ribosomal decoding and peptide bond formation on an atomic level. Here, only the overall structure of the functional centers will be described while their functional significance will be discussed in the following sections. The decoding center is located between the small subunit body and head domains at the interface to the large subunit (**Figure 2(c)**). It is formed by the tip of helix 44 that spans the entire body as well as helix 34 and the 530-loop of 16S rRNA, which is part of the head domain. In addition to 16S rRNA, a few residues of protein S12 are in proximity to the decoding center and take part in reading the codon–anticodon duplex.[14] The peptidyltransferase center is buried deep in the large subunit, which only the 3′ ends of the tRNAs can reach (**Figure 2(d)**). The tRNAs are held in place by direct base pairing between the conserved tRNA 3′CCA ends and the 23S rRNA in the so-called A and P loops. As the crystal structures unambiguously revealed, there is generally no ribosomal protein in proximity to the peptidyltransferase center. Of the ribosomal crystal structures obtained thus far, it is only in the *T. thermophilus* ribosome that we see an exception to this rule, with protein L27 reaching into the peptidyltransferase center. However, it has been shown by computational methods that this nonconserved part of L27 is not significantly involved in catalysis.[15] Thus, the ribosome seems to be a ribozyme that catalyzes peptide bond formation solely with the help of RNA.

In addition to providing insight into the atomic mechanism of protein synthesis, the ribosome structures also revealed the inhibitory mechanisms of several antibiotics targeting the ribosome.[16,17] These structures will help to design new antibiotics and to understand development of resistance. Most antibiotics target either the peptidyltransferase center including the peptide exit channel on the 50S subunit or important regions on the 30S ribosomal subunit. Thus, they interfere directly with the two central ribosomal processes, namely, decoding of the mRNA and extension of the growing peptide. Interestingly, antibiotics mainly interact with ribosomal RNA rather than with ribosomal proteins, again indicating that rRNA is the main functional player during protein synthesis. The macrolides, streptogramins, chloramphenicol, and the oxazolidinones bind close to the peptidyltransferase center in the ribosomal exit tunnel, a channel in the large ribosomal subunit through which the growing peptide leaves the ribosome. Structures of anisomycin, sparsomycin, blasticidine S, chloramphenicol, and streptogramin A,[18–20] and second-generation macrolides[21] bound to the 50S subunit have been solved. By binding to the exit channel, the macrolides such as erythromycin sterically block extension of the growing peptide chain and thus inhibit peptide bond formation indirectly. The 30S subunit is the main binding site for aminoglycosides and other antibiotics such as paramomycin and streptomycin. Structures with bound paramomycin, streptomycin,[22] tetracycline, pactamycin, hygromycin B,[23] and several other antibiotics targeting the small ribosomal subunit have now been determined. In general, these drugs seem to interfere with the dynamic conformational changes in the 30S subunit during decoding of the mRNA. Interestingly, a combination of structural and biochemical studies revealed that many of these antibiotics typically do not inhibit protein synthesis *per se*, but rather induce a high-error rate in aminoacyl-tRNA selection, which ultimately results in the synthesis of inactive proteins. Some examples of inhibitory mechanisms of antibiotics will be discussed below in the context of the ribosomal decoding mechanism.

In summary, the recent crystal structures have provided valuable insight into the atomic structure of the ribosome and its interactions with tRNAs, mRNA, and antibiotics. These structures are important snapshots toward a mechanistic understanding of ribosomal protein synthesis on an atomic level. The next challenge lies in determining structures of the ribosome interacting with the many translation factors. So far, most progress in this direction has been achieved using cryo-electron microscopy (cryo-EM). For example, this technique has revealed the interaction of elongation factors Tu[24,25] and G,[26,27] initiation factor 2 (IF2)[28,29] and release factor 3 (RF3)[30,31] with the ribosome. However, the maximum resolution that can currently be obtained by cryo-EM is about 10 nm (8–12 Å), far from the desired atomic resolution. Therefore, the crystal structures of the 30S subunit with initiation factors 1 and 3 (IF1[32]; IF3[33]) and of the 70S subunit with release factors 1 and 2 (RF1/2[10,34]) as well as RRF[35] have been important milestones toward understanding the interaction of the ribosome with protein factors.

5.13.3 Ribosomal Aminoacyl-tRNA Selection

The fidelity of aminoacyl-tRNA selection by the ribosome is very high with only one incorrect per 1000–10 000 correct amino acids incorporated into proteins in prokaryotes.[36] To achieve this accuracy, the ribosome has to discriminate efficiently and quickly between the various cognate (correct) codon–anticodon interactions and the very similar near-cognate interactions, which might vary in only a single base pair (**Figure 3**). However, from a theoretical viewpoint the three base pairs formed between the mRNA codon and the tRNA anticodon are not sufficient to explain the high ribosomal fidelity. The stability of a three base pair helix with correct Watson–Crick base pairs is very similar to that of a helix with a single mismatch.[37–39] In fact, an energetic difference of 2–3 kcal mol^{-1} between cognate and near-cognate tRNA typical of single mismatches could account only for an error rate of 1:100 in selecting the correct aminoacyl-tRNA by the ribosome. Thus, the thermodynamics of base pairing cannot be used to discriminate efficiently against mismatches in the codon–anticodon helix. This is known as the decoding problem. The situation is further complicated by the fact that in the cell concentrations of tRNAs that are specific for a particular codon (isoacceptor tRNAs) vary between 1 and 10% of total tRNA.[40] This implies that a large excess of incorrect ternary complexes compete with the single correct substrate for binding to the A site at a given codon. Furthermore, protein synthesis has to occur with a high speed of about 10–20 s^{-1} within a living cell.[41] Thus, the ribosome does not have enough time to allow codon–anticodon interaction to reach equilibrium, further diminishing the potential for discrimination based on energetic differences of cognate and near-cognate interactions.

The crystal structures of 30S ribosomal subunits with mRNA and tRNA fragments bound increased our understanding of how the ribosome selects for correct Watson–Crick base pairing between codon and anticodon.[14,42] In particular, critical interactions between the codon–anticodon duplex and the conserved bases A1492 and A1493 from helix 44 of 16S rRNA as well as G530 from the head have been identified (**Figure 4**) (*E. coli* numbering of ribosomal bases will be used throughout this chapter). Whereas in the structure of the small subunit alone, A1492 and A1493 are stacked in the interior of helix 44,[43] these bases are flipped out upon binding of tRNA to the A site. Also, binding of aminoacyl-tRNA induces a switch of G530 from a syn to anti conformation (**Figures 4(a)** and **4(b)**). Through these conformational changes, the three ribosomal bases monitor the correct geometry of a Watson–Crick base pair in the minor groove of the first two codon–anticodon positions. This is achieved by specific interactions with the 2′OH groups of the codon and anticodon riboses as the distance between these functional groups is characteristic of a Watson–Crick base pair (**Figures 4(d)** and **4(e)**). In the third position,

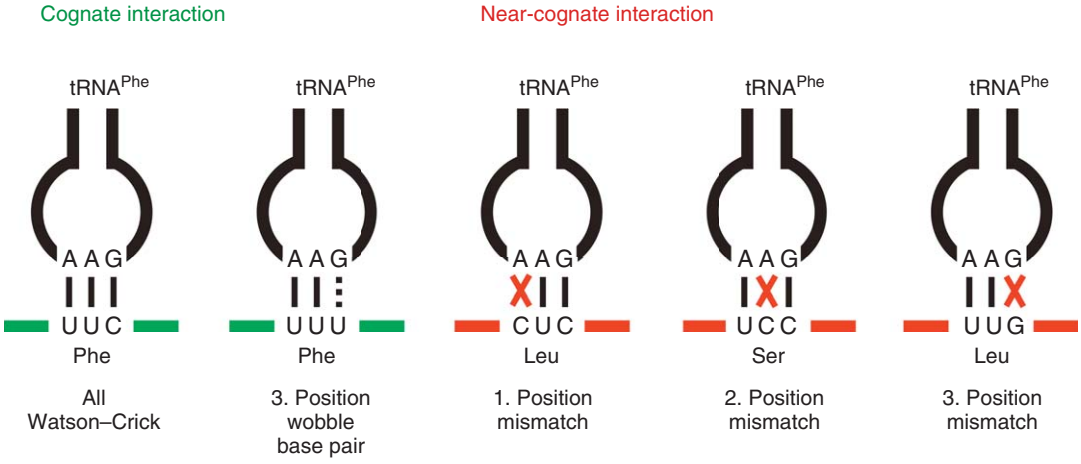


Figure 3 Cognate and near-cognate codon–anticodon interactions. The anticodon loop of $tRNA^{Phe}$ is shown as an example interacting with various codons on the mRNA. In correct, cognate codon–anticodon pairings, two Watson–Crick base pairs can be formed in the first two positions while the third position contains either a Watson–Crick or a wobble base pair. In incorrect, near-cognate interactions, one mismatch is found between codon and anticodon compared to the cognate pairings. The mismatch can reside in the first, second, or third position. Interactions with more than one mismatch are called noncognate (not shown).

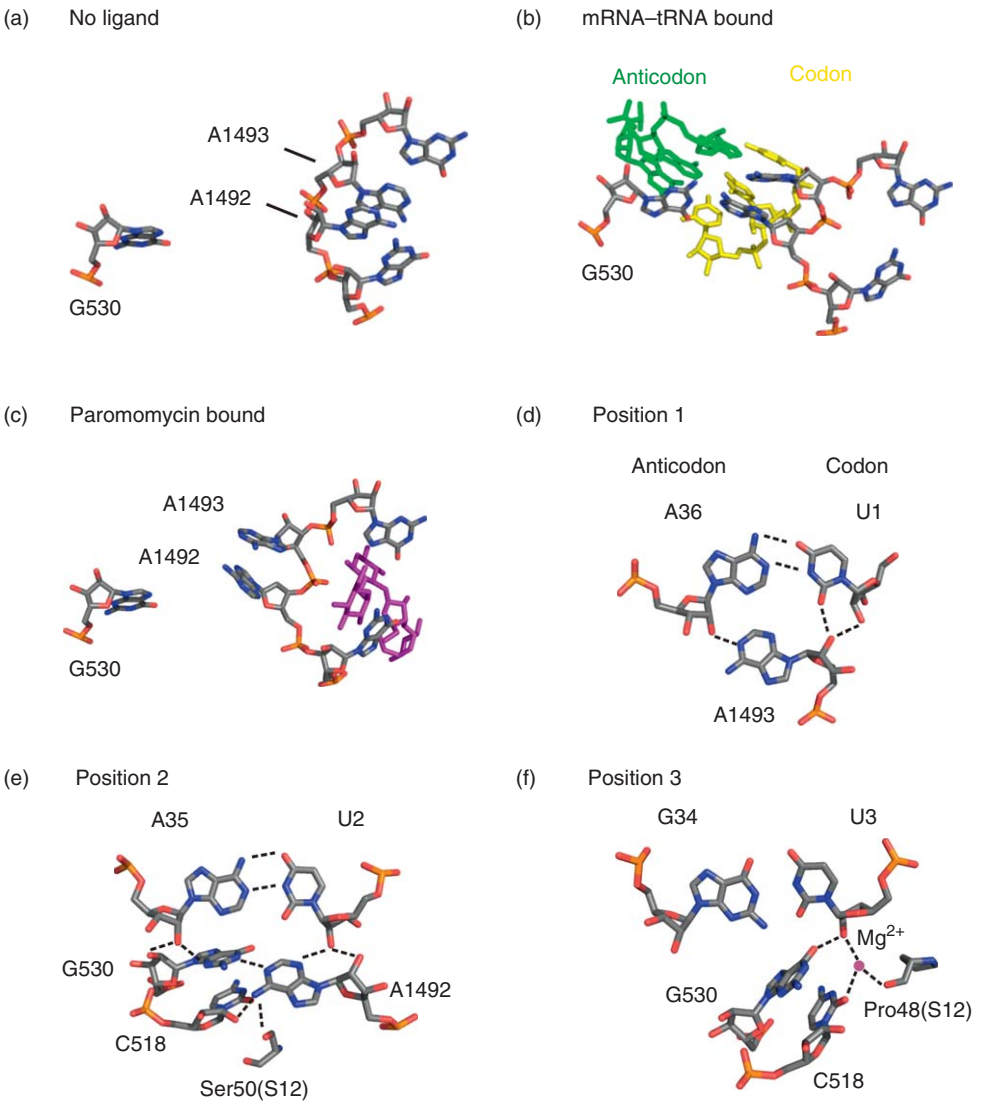


Figure 4 The decoding center of the 30S ribosomal subunit. (a–c) Conformational changes in the 16S rRNA bases A1492, A1493, and G530 upon formation of the cognate codon–anticodon complex and upon binding of the antibiotic paromomycin. (a) Without any ligand bound, A1492 and A1493 are located in the interior of helix 44 of the 16S rRNA, and G530 is in the syn conformation (Protein Data Bank ID code 1J5E).[5] (b) When a cognate tRNA is bound to the mRNA in the decoding center, A1492 and A1493 flip out and G530 switches into the anti conformation to interact with the codon–anticodon duplex (Protein Data Bank ID code 1IBM).[14] (c) Binding of the antibiotic paromomycin (purple) induces A1492 and A1493 to flip out, similar to the conformational changes observed upon binding of a cognate tRNA, while G530 stays in the syn conformation (Protein Data Bank ID code 1IBK).[14] (d–f) Monitoring of the three codon–anticodon positions by ribosomal bases (Protein Data Bank ID code 1IBM).[14] (d) In the first position, the distance between the 2′-hydroxyl groups of codon and anticodon is sensed by an A-minor interaction with A1493. (e) In the second codon–anticodon position, a network of hydrogen bonds is formed among A1492, G530, C518, and Ser50 of ribosomal protein S12, which monitors the position of the 2′-hydroxyl groups of codon and anticodon. (f) G530, C518, and Pro48 of ribosomal protein S12 interact directly or via an Mg^{2+} ion with the 2′-hydroxyl group of the third codon position. No interactions are formed between the ribosome and the third anticodon position. This figure was generated with Pymol (http://www.pymol.org).

monitoring of codon–anticodon base pairing is less stringent as wobble base pairs are allowed (**Figure 4(f)**). Thereby, the ribosome specifically controls formation of correct base pairs in a sequence-independent manner and can thus accept all different correct mRNA–tRNA pairings. Upon binding of near-cognate aminoacyl-tRNAs, only some of these structural changes seem to be induced as A1493 and A1492 are fully and partly flipped out,

respectively.[42] In addition to the described local changes, a global conformational change in the 30S subunit has been observed upon binding of a cognate tRNA. The 30S head and shoulder domain move to close around the A site. However, in the presence of a near-cognate aminoacyl-tRNA the 30S subunit seems to stay in an open conformation as the shoulder does not move and the head rotates into a different direction.[42]

While these structures are important snapshots of the ribosomal tRNA selection process, dynamic studies have been required to fully understand ribosomal decoding as the kinetics rather than the thermodynamics of tRNA–mRNA interaction are critical. In particular, rapid kinetic investigations have revealed how the ribosome overcomes the decoding problem by kinetic proofreading and induced fit (**Figure 5**).[44–46] The kinetic mechanism of aminoacyl-tRNA selection has also been confirmed by single-molecule studies[47]. Here, first a general overview will be provided of the steps during delivery of a new aminoacyl-tRNA by EF-Tu to the ribosomal A site and their contribution to ribosomal fidelity. Subsequently, the crucial steps will be discussed in detail. First, the ternary complex EF-Tu–GTP–aminoacyl-tRNA forms a labile initial binding complex with the ribosome, which is independent of the codon presented in the ribosomal A site.[48,49] Only subsequently does codon recognition take place when the anticodon stem of the tRNA moves into the ribosomal A site and base pairs with the A-site codon. Presumably, codon recognition actually occurs in several substeps.[47] Correct codon–anticodon interaction triggers GTPase activation of EF-Tu, which is rate limiting for GTP hydrolysis. EF-Tu is a classical G protein switch that changes its conformation upon release of inorganic phosphate.[50] Thereby, EF-Tu adopts its inactive GDP-bound conformation that has low affinity for the aminoacyl-tRNA and dissociates from the ribosome. Thus, the aminoacyl-tRNA is free to accommodate into the 50S A site. During accommodation, the aminoacyl end swings into the peptidyltransferase center on the large ribosomal subunit. Here, it takes part in rapid peptide bond formation with the peptidyl-tRNA in the ribosomal P site. Alternatively, the aminoacyl-tRNA may dissociate from the ribosome prior to accommodation.[46] It has now been confirmed that this mechanism is the same for different tRNAs.[51–53]

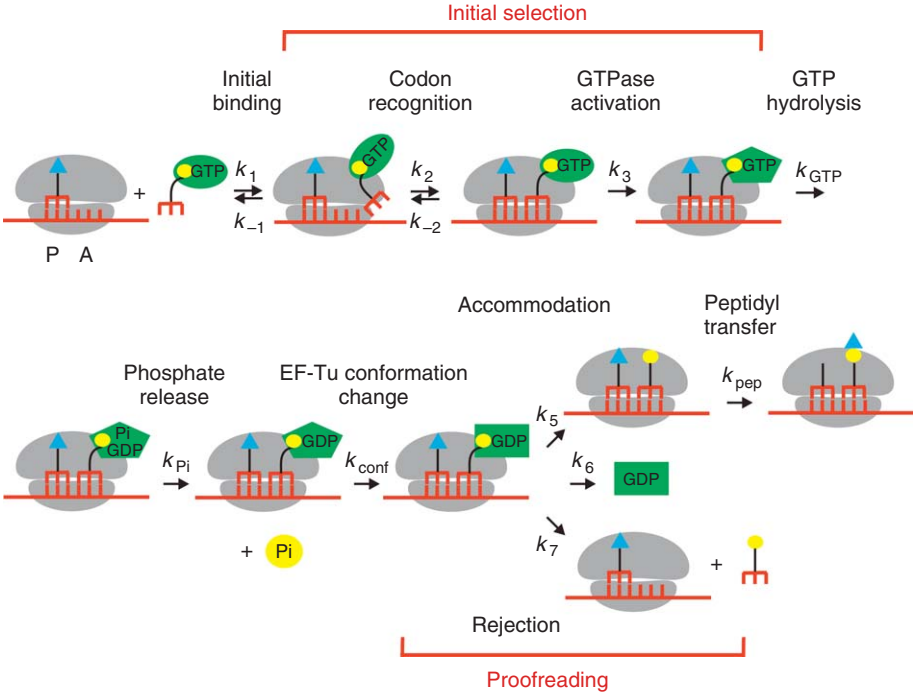


Figure 5 Kinetic mechanism of aminoacyl-tRNA selection by the ribosome. The aminoacyl-tRNAs are delivered to the ribosome in the form of a ternary complex with EF-Tu–GTP. Incorrect aminoacyl-tRNAs can either dissociate as a ternary complex in the initial selection phase or later as free aminoacyl-tRNA in the proofreading phase. The two selection phases are separated through the irreversible GTP hydrolysis by EF-Tu. Discrimination against incorrect tRNAs is achieved by increased dissociation rate constants (k_{-2} and k_7) as well as decreased forward rate constants (k_3 and k_5) compared to cognate tRNAs.

How does the described kinetic mechanism of A-site binding allow the ribosome to efficiently discriminate against near-cognate aminoacyl-tRNAs with only one mismatch to the codon presented in the decoding site? This question has been addressed by comparing the rate constants of the individual steps during delivery of aminoacyl-tRNA by EF-Tu to a cognate codon with those for near-cognate codons presented in the 30S A site. Already in the 1970s, it had been suggested that the ribosome uses kinetic proofreading to increase the discrimination against near-cognate mRNA–tRNA interactions.[54,55] Kinetic proofreading relies on two independent selection phases that are separated by an irreversible step. Thus, the energetic differences between cognate and near-cognate codon–anticodon pairings can be employed twice for selection. In the case of A-site binding, irreversible GTP hydrolysis by EF-Tu is the crucial event to separate the two selection phases. Thus, A-site binding can be divided into the initial selection phase and the proofreading phase (**Figure 5**). During initial selection, the ternary complex of EF-Tu–GTP–aminoacyl-tRNA can dissociate from the ribosome prior to GTP hydrolysis since both initial binding and codon recognition are reversible. The important rate constant in this selection phase is the dissociation rate of the codon recognition complex. For $tRNA^{Phe}$ it has been found that the dissociation rates of the codon recognition complex for near-cognate interactions are universally 1000-fold faster than for cognate pairings.[56] Proofreading occurs after GTP hydrolysis by EF-Tu and its dissociation in the GDP form. Thus it is the aminoacyl-tRNA alone that can dissociate in the proofreading phase rather than moving forward by accommodating into the 50S subunit. Again, the dissociation rates are significantly higher for near-cognate interactions than for cognate interactions.

Kinetic proofreading alone is not sufficient for high-fidelity protein synthesis in the cell since it relies on the thermodynamic differences between correct and incorrect codon–anticodon interactions as reflected in the different dissociation rates. However, A-site binding is rather fast occurring with an overall rate constant of 10–20 s^{-1} in bacteria.[41] Consequently, the critical binding steps such as codon recognition cannot reach equilibrium as the forward rate constants are comparably rapid. This problem was solved with the discovery of an additional induced-fit mechanism during aminoacyl-tRNA delivery to the ribosome by EF-Tu. As known for many enzymes, binding of the correct substrate can induce a conformational change in the enzyme that brings it into the active conformation. Thus, a forward step such as catalysis is selectively accelerated for correct but not for incorrect substrates where the enzyme remains in an inactive conformation. While this had been observed often for protein enzymes, it was a surprise that the ribosome, a large ribonucleoprotein machine, also uses an induced-fit mechanism. This was only discovered by using rapid kinetic techniques, such as stopped-flow and quench-flow, which allowed determination of all the individual rate constants during A-site binding (**Table 2**). Thereby, it has been demonstrated that the forward steps of GTPase activation and accommodation are significantly faster for cognate than for near-cognate codon–anticodon pairings.[45] In fact, under high-fidelity conditions the induced-fit mechanism is the major factor for achieving high discrimination and high speed in ribosomal protein synthesis. Importantly, GTPase activation and subsequent instantaneous GTP hydrolysis is 650-fold faster for cognate than near-cognate tRNAs under high-fidelity conditions.[46] Thus, cognate tRNAs

Table 2 Rate constants for aminoacyl-tRNA selection

	Cognate interaction	*Near-cognate interaction*
Initial binding[46]		
k_1	140 $\mu mol^{-1} l\, s^{-1}$	140 $\mu mol^{-1} l\, s^{-1}$
k_{-1}	85 s^{-1}	85 s^{-1}
Codon recognition[52,56]		
k_2	140–190 s^{-1}	100–280 s^{-1}
k_{-2}	0.1–0.2 s^{-1}	100–240 s^{-1}
GTPase activation, k_3[52,56]	60–260 s^{-1}	0.06–1.3 s^{-1}
GTP hydrolysis, k_{GTP}[44]	Rapid (limited by k_3)	Rapid
Pi release, k_{Pi}[50]	20 s^{-1}	20 s^{-1}
EF-Tu conformation change, k_{conf}[50]	Rapid (limited by k_{Pi})	Rapid
Accommodation, k_5[52,56]	7–12 s^{-1}	(0.02–0.3 s^{-1})[a]
Rejection, k_7[56]	<0.3 s^{-1}	(0.8–2.1 s^{-1})[a]

[a] Determined under low fidelity buffer conditions.

undergo very rapid A-site binding due to the fast forward rates. For near-cognate interactions, GTPase activation is sufficiently slow to allow for (at least some) dissociation of the codon recognition complex. These findings clearly demonstrate that aminoacyl-tRNA selection is kinetically rather than thermodynamically controlled.

In summary, the ribosome uses kinetic proofreading with two selection phases (initial selection and proofreading) in combination with a powerful induced-fit mechanism to achieve highly accurate protein synthesis. The relative contributions of initial selection and proofreading to overall selection *in vivo* are not known. *In vitro* kinetic and biochemical experiments indicated that during proofreading alone, about one amino acid out of 10–100 near-cognate aminoacyl-tRNAs was incorporated into the growing peptide.[41,45,46,57] This suggested that in order to achieve a high overall selectivity, two selection steps must contribute to the overall fidelity and the initial selection phase must be about as accurate as proofreading. Thus, not more than one out of 10–100 near-cognate ternary complexes should pass the initial selection screen. In fact, some of the near-cognate ternary complex can pass through initial selection, because the rate of GTPase activation can be relatively large compared to dissociation from the codon recognition complex. In contrast, most noncognate ternary complexes with less than two correct base pairs between codon and anticodon are efficiently rejected in the initial selection phase. Kinetic experiments carried out under high-fidelity conditions *in vitro*, which resemble closely the situation in the cell showed an overall incorporation of one incorrect per 450 correct amino acids, indicating an efficiency of initial selection of 30.[46] Thus, both selection steps, initial selection prior to and proofreading after GTP hydrolysis, are required for efficient tRNA discrimination under all experimental conditions studied. Therefore, it is very likely that the two steps operate also *in vivo*.

The kinetic mechanism of aminoacyl-tRNA selection demonstrates that the forward steps of EF-Tu GTPase activation and accommodation are crucial for high fidelity. How can this observation be explained on a structural level? Unfortunately, we have high-resolution structures only of the ribosome prior to A-site binding and of the tRNAs bound to the ribosome after accommodation. Of the intermediate states, only the low-resolution structure of a ternary complex EF-Tu–GTP–aminoacyl-tRNA stalled after GTP hydrolysis has been determined by cryo-EM (**Figure 6**).[24,25] These studies revealed that EF-Tu interacts with the so-called GTPase-activating center on the surface of the 50S subunit while holding the aminoacyl end of the

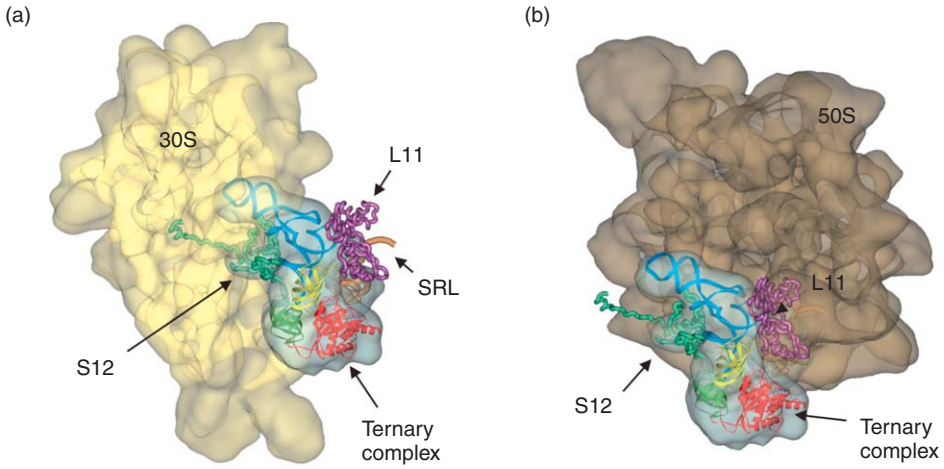


Figure 6 Cryo-electron microscopy reconstruction of the ternary complex EF-Tu–GTP–aminoacyl-tRNA bound to the ribosome.[24] A-site binding has been blocked by the antibiotic kirromycin following the release of inorganic phosphate and prior to the conformational change of EF-Tu. (a) Interactions of the ternary complex with the 30S subunit (yellow) and elements of the 50S subunit. The ternary complex is represented by a blue surface with the crystal structure of EF-Tu–GTP–aminoacyl-tRNA fitted into the density. The anticodon stem of the tRNA reaches into the decoding center while its aminoacyl end is bound to EF-Tu on the outside of the ribosome. Protein S12 (green) from the 30S and protein L11 (purple) from the 50S subunit contact the tRNA (blue). The Sarcin-Ricin loop (SRL, brown), from the 23S rRNA of the 50S subunit interacts with the G domain of EF-Tu (red). (b) Interactions of the ternary complex with the 50S subunit (brown) and protein S12 from the 30S subunit. L11 and the SRL are part of the GTPase-activating center on the 50S subunit. Reprinted by permission from Macmillan Publishers Ltd.: *Nature Structural Biology*, H. Stark; M. V. Rodnina; H. J. Wieden; F. Zemlin; W. Wintermeyer; M. van Heel, *Nat. Struct. Biol.* **2002**, *9* (11), 849–854, copyright 2002.

tRNA. Thus, the ternary complex EF-Tu–GTP–aminoacyl-tRNA spans from the decoding center in the middle of the 30S subunit to the GTPase-activating center on the outside of the 50S subunit. The active site of EF-Tu where the GTP is bound is about 7 nm away from the decoding site where codon–anticodon interaction takes place. This immediately raises the question of how the signal of correct mRNA–tRNA pairing is conveyed to EF-Tu during GTPase activation. A definitive answer on this question as well as on the exact mechanism of GTPase activation of EF-Tu has to await a crystal structure of the ternary complex on the ribosome. At the moment, several biochemical studies provide some hints on this crucial step during aminoacyl-tRNA selection.

The GTPase-activating signal obviously originates from the decoding site, and some conformational changes in the 30S subunit have been identified from the crystal structures of cognate and near-cognate tRNA fragments as outlined above.[14,42] Importantly, A1492, A1493, and G530 of 16S rRNA change their conformations upon correct codon–anticodon pairing, which induces a global conformational change in the 30S subunit (**Figure 4**). The importance of these interactions is further confirmed by mutational analysis of A1492, A1493, and G530, which revealed strong decreases for all mutations in the rate of GTPase activation as well as accommodation.[51] The conformational changes in the 30S subunit are altered by antibiotics such as streptomycin and paromomycin. Interestingly, both antibiotics have direct effects on the GTPase activation step: paromomycin, which causes A1492 and A1493 to flip out even in the absence of a tRNA[14,42] induces an active conformation of the ribosome and thus activates EF-Tu's GTPase for both cognate and near-cognate interactions (**Figure 4(c)**).[58,59] Streptomycin restricts the flexibility of the 30S subunit and decreases the rate of GTPase activation with cognate while increasing its rate with near-cognate ternary complexes resulting in an almost complete loss of selectivity.[58]

Following the conformational changes in the 30S subunit, the GTPase activation signal has to be submitted to EF-Tu on the 50S subunit. In principle, two signaling pathways can be envisioned: either conformational changes are propagated through intersubunit bridges to the 50S subunit and then to EF-Tu, or the signal is directly transferred through the tRNA body to EF-Tu. So far, not much is known about the first signaling pathway since only a few crystal structures of the ribosome with high resolution of the intersubunit bridges have been obtained.[9,12] To detect conformational changes at least one more such crystal structure in a different state, for example, with bound EF-Tu, would be required. However, it is known that elements of the 50S GTPase-activating center play a crucial role. Deletion and mutational studies combined with rapid kinetics revealed that the C-terminal domains of ribosomal protein L7/12 contribute significantly to EF-Tu GTPase activation by stabilizing its transition state.[11] Cryo-EM reconstructions also showed extensive interactions of the switch regions of the G domain with the Sarcin-Ricin loop (SRL) of 23S rRNA,[24,60] indicating that the SRL stabilizes the transition state conformation of the switch regions of EF-Tu (**Figure 6**). In fact, cleavage of the SRL impedes progression from the GTPase activated state;[47] as a result, GTP hydrolysis by EF-Tu is abolished. Recent single-molecule studies suggest that the interaction of EF-Tu with the GTPase-activating center may depend on thermal fluctuations as the ternary complex fluctuates reversibly between the codon recognition and the GTPase activated states repeatedly.[61]

The tRNA itself is clearly critically involved in transmitting the GTPase activation signal from the decoding site to EF-Tu as demonstrated by several biochemical studies. The cryo-EM structures of ternary complexes stalled on the ribosome show structural rearrangements in the tRNA. The anticodon arm of tRNA on the ribosome is reoriented by bending at the junction of the anticodon and D stems.[60] As the tRNA connects the decoding center with EF-Tu, the bending of aminoacyl-tRNA might be part of a signaling mechanism through the tRNA. This is supported by the finding that intact tRNA is required to trigger GTP hydrolysis in EF-Tu.[62] Further strong evidence for an active role for tRNA in GTPase activation was obtained by the kinetic analysis of the Hirsh suppressor, a G24A mutation within the D arm of Trp-tRNA^{Trp}, which promotes tryptophan incorporation on UGA stop codons, in addition to the Trp codon UGG.[63] In this case, the elevated levels of miscoding are achieved by acceleration of the forward selection steps, GTPase activation and accommodation, independently of codon–anticodon pairing. This mutation might facilitate bending of tRNA by rendering it more easily deformable such that no additional energy from correct codon–anticodon interaction is required to stabilize the distorted tRNA.[56] Interestingly, the Hirsh suppressor Trp-tRNA^{Trp} cannot overcome the effects on GTPase activation (and accommodation) of mutations of A1492, A1493, and G530 in the decoding center in contrast to the antibiotics paromomycin and streptomycin.[51] This led to the conclusion

that two independent transition state barriers have to be crossed for both GTPase activation and accommodation. It might be that these two transition state barriers are related to the conformational changes within the decoding center and the deformation of the tRNA, respectively.[51]

Rapid accommodation of cognate aminoacyl-tRNAs is the other important induced-fit step during A-site binding next to GTPase activation. During accommodation, the 3′ end of aminoacyl-tRNA has to move a distance of almost 7 nm (70 Å) within the ribosome from its binding site on EF-Tu into the peptidyltransferase center. Aminoacyl-tRNA accommodation in the peptidyltransferase center (PTC) limits the rate of peptide bond formation, which is intrinsically very rapid.[44,45,64] Accommodation of all cognate aminoacyl-tRNAs proceeds rapidly and efficiently, with essentially no dissociation of aminoacyl-tRNA at this step, yet it is the rate-limiting step during A-site binding of cognate aminoacyl-tRNA. In contrast, near-cognate aminoacyl-tRNAs are rejected at this step due to both low stability of aminoacyl-tRNA binding and slow accommodation in the A site. Currently, we know even less about the structural requirements for accommodation than for GTPase activation. The tRNA again plays a crucial role as the mutation in the D arm of $tRNA^{Trp}$ in the Hirsh suppressor also accelerates accommodation on near-cognate UGA stop codons.[63] More insight into accommodation has been derived from a molecular dynamics simulation of the path of the aminoacyl-tRNA through the ribosome into the peptidyltransferase center.[65] This study suggested four stages of aminoacyl-tRNA accommodation. From the state immediately after dissociation of EF-Tu, the tRNA body moves into a state where it is almost completely in the accommodated form. This could constitute a relaxation from the bent tRNA conformation observed in the cryo-EM reconstructions of ternary complex stalled on the ribosome after GTP hydrolysis. However, the CCA end of aminoacyl-tRNA is impeded from reaching the peptidyltransferase center by the A loop of 23S rRNA. In the next step, relaxation of the tRNA acceptor stem takes place by movement of the CCA end, which relies on the flexibility of both the acceptor stem and the A loop. Lastly, the CCA end slowly passes through a gate formed by the universally conserved bases U2492, C2556, and C2573. The molecular dynamics simulation provides a rationale for the importance of accommodation for aminoacyl-tRNA selection. As cognate tRNAs are tightly anchored with the anticodon stem in the decoding center, they remain bound to the ribosome after relaxation of the tRNA body although the CCA end is in a strained conformation. Thus, cognate tRNA can stay within the ribosome until the A loop moves and the 3-nucleotide gate opens such that the acceptor end can reach the peptidyltransferase center. In contrast, all less tightly bound near-cognate tRNAs would usually dissociate rapidly as soon as their CCA end clashes with the A loop upon relaxation of the tRNA body.

Owing to the progress in ribosomal crystal structure determination and in biophysical studies of the kinetic mechanism of A-site binding, we now understand fairly well how the ribosome selects the correct aminoacyl-tRNAs. In particular, the atomic details of codon–anticodon monitoring by the ribosome have been unraveled. Also, the kinetic determinants of aminoacyl-tRNA selection by kinetic proofreading and induced fit and the importance of EF-Tu GTPase activation and accommodation have been identified. The future challenge lies in understanding the interplay between structure and kinetics to explain on a molecular level the high fidelity of ribosomal protein synthesis.

5.13.4 Peptide Bond Formation

Ribosomal peptide bond formation takes place between the amino acid attached to the A-site tRNA and the growing peptide attached to the P-site tRNA. The α-amine of the A-site aminoacyl-tRNA attacks the carbonyl carbon of the ester bond linking the nascent peptide to the P-site tRNA (**Figure 7(a)**). Thus, ribosomal peptide bond formation is the aminolysis of an ester. By forming the new peptide bond, the growing peptide is transferred from the P-site tRNA to the A-site tRNA. The ribosome-independent formation of aminoacyl-tRNAs is a prerequisite for ribosome-catalyzed peptide bond formation. Aminoacylation is catalyzed by aminoacyl synthetases, which have an extremely high specificity for both the tRNA and the amino acid to ensure high accuracy in aminoacyl-tRNA formation. Aminoacyl synthetases activate an amino acid using ATP by forming an aminoacyl-adenylate in the first reaction step. Subsequently, the amino acid is transferred from the aminoacyl-adenylate to the tRNA. Both reaction steps are energetically favored by the hydrolysis of the liberated pyrophosphate. Thus, aminoacylation requires energy in form of ATP, and the

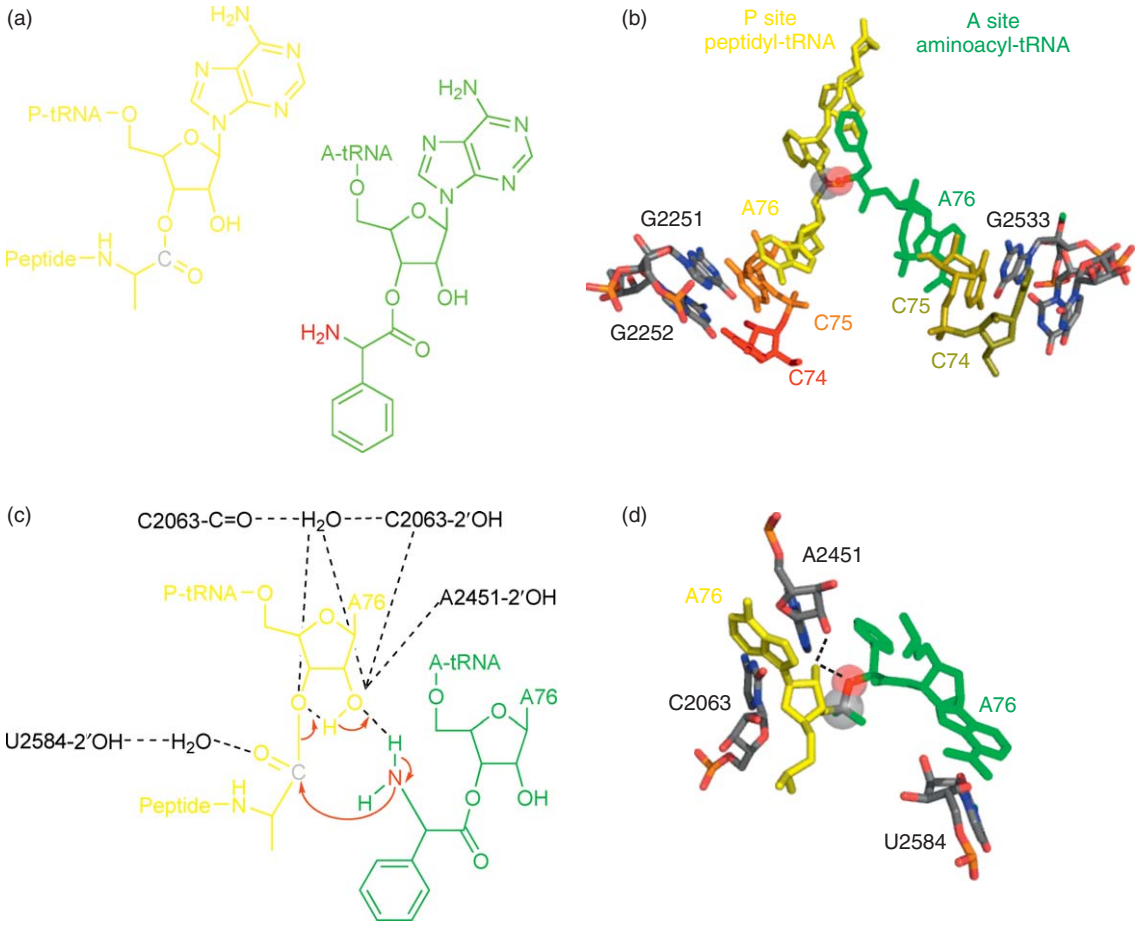


Figure 7 Ribosome-catalyzed peptide bond formation. (a) The α-amino group (red) of the aminoacyl-tRNA (green) attacks the carbonyl group (gray) of the peptidyl-tRNA (yellow). (b) Crystal structure of substrates bound to the ribosomal peptidyltransferase center (Protein Data Bank ID code 1VQN).[70] Peptide bond formation has been inhibited by the use of a hydroxyl group instead of the attacking α-amino group (red sphere). C74 and C75 of the P-site peptidyl-tRNA base pair with G2252 and G2251 of 23S rRNA, respectively. Similarly, C75 of the A-site aminoacyl-tRNA interacts with G2533. (c) Proposed model of ribosome-catalyzed peptide bond formation using a proton shuttle mechanism (red arrows). A six-membered ring is formed during the transition state, which includes the critical 2′ hydroxyl group of A76 of the P-site tRNA. During the nucleophilic attack of the α-amino group (red) on the carbonyl group (gray), A76 2′OH simultaneously abstracts a proton from the α-amine and donates a proton to the leaving group. The reaction is facilitated by an intricate hydrogen bonding network of the reactants with ribosomal bases (black). (d) Crystal structure of a transition state analogue (RAP) bound to the ribosomal peptidyltransferase center (Protein Data Bank ID code 1VQP).[69] Only the crucial hydrogen bonds of the attacking α-amino group to the 2′ hydroxyl group of A76 and the hydrogen bond between the 2′ hydroxyl group of A76 and A2451 are shown. This figure was generated with Pymol (http://www.pymol.org).

resulting aminoacyl-tRNA contains an activated amino acid. Using aminoacyl- and peptidyl-tRNAs as substrates, peptide bond formation is a spontaneous process. The reactivity of esters with amines is intrinsically relatively high with rates of the uncatalyzed reaction of $\sim 10^{-4}\,\mathrm{mol}^{-1}\,\mathrm{l}\,\mathrm{s}^{-1}$ at room temperature. Ribosomal peptide bond formation has been estimated to occur with a rate $>300\,\mathrm{s}^{-1}$.[64] Thus, the ribosome increases the rate of peptide bond formation by 10^6–10^7-fold. This rate enhancement is comparable to that of many protein enzymes, but far from the maximum rate enhancements by enzymes observed in nature.[66]

Peptide bond formation is the essential reaction catalyzed by the ribosome. Despite its importance, it was for a long time not the focus of ribosomal research, for several reasons. First, before the determination of the high-resolution ribosome crystal structures almost nothing was known about the active site. Second, under most experimental conditions accommodation of the incoming aminoacyl-tRNA is rate limiting for peptide bond

formation so that this step is difficult to be observed independently in biochemical experiments. The ribosome crystal structures and the development of new assays opened the door for mechanistic studies of ribosomal peptide bond formation. As mentioned above, the structure of the 50S subunit from *H. marismortui* determined at 0.24 nm (2.4 Å) resolution by Steitz and coworkers in 2000 showed the location of the peptidyltransferase center in a deep cleft of the large ribosomal subunit.[3] The residues forming the active site are all part of domain V of 23S rRNA, which is highly conserved. The A and P loops of 23S rRNA position the two tRNAs in the active site by interactions with the universal CCA ends of tRNA. In the A site, C74 of the tRNA stacks on U2555, C75 base pairs with G2553 and A76 is held in place by interactions with A2451 and G2583. The P-site tRNA forms two base pairs of C74 and C75 with G2252 and G2251, respectively, and the terminal residue A76 interacts with the same residues G2583 and A2450 as the A-site tRNA (**Figure 7(b)**). At the heart of the peptidyltransferase center, the conserved bases A2451, U2506, U2585, C2452, and A2602 are found. The *H. marismortui* ribosome structures showed no proteins close to the peptidyltransferase center as also seen in the ribosome structures from *D. radiodurans*.[6] Only in *T. thermophilus* ribosomes, the flexible N-terminal tail of ribosomal protein L27 is close to the P-site tRNA in the active site.[9] Also, it has been shown that deletion of the three N-terminal residues of L27 significantly reduces peptidyl transferase acitivity.[67] This raised again the question of protein involvement in ribosomal peptide bond formation. However, the N-terminus of protein L27 is not conserved, and further computational studies showed only a minor effect of L27 on catalysis.[15] Hence, all evidence indicates that ribosomal peptide bond formation is catalyzed by RNA alone, and that the ribosome is indeed a ribozyme.

In principle, several catalytic mechanisms for peptide bond formation can be envisioned. First of all, the ribosome will bind the two substrates, P-site peptidyl-tRNA and A-site aminoacyl-tRNA in close proximity to each other and orient them for peptide bond formation. Upon attack of the carbonyl group of P-site aminoacyl-tRNA by the A-site α-amino group, a tetrahedral transition state is formed. This state could be dipolar (zwitterionic) with a positive charge on the attacking nitrogen and a negative charge on the carbonyl oxygen. The formation of the transition state could be facilitated by acid–base catalysis where a general base could abstract a proton from the attacking amino group and/or a general acid could donate a proton to the leaving group. If a general base was involved, the transition state would be negatively charged. In both cases, the oxyanion of the transition state could be electrostatically stabilized as often observed in protein enzymes. This transition state stabilization could involve metal ions bound to 23S rRNA.

Since 2000, a combination of structural, mutational, kinetic, and computational studies have elucidated the catalytic strategies used by the ribosome to accelerate peptide bond formation. The crystal structures provide important insight into the structure of the active site. By now, 50S structures with substrates, different transition state analogues, and products have been solved.[20,68–70] However, this information is static and limited by the quality of the transition state analogue, which can never fully represent all features of the true transition state. Thus, these studies have to be complemented by other techniques to obtain information on the dynamic changes during catalysis. For a long time, peptide bond formation was not accessible for kinetic investigations as it is faster than the preceding accommodation of the incoming aminoacyl-tRNA. This could be overcome by the use of small, rapidly binding substrate analogues such as puromycin, which mimics the terminal adenine of tRNA with a linked tyrosyl residue.[71] In fact, puromycin is an antibiotic as the tyrosyl analogue is linked by a nonhydrolyzable amide bond to the adenosine causing premature termination of elongation as no further peptide bond can be formed. However, subsequent investigations also revealed that puromycin has slightly different properties than full-length tRNA, particularly since it is missing important interactions with the 23S rRNA A loop.[70,72] Thus, it is of limited use for mechanistic studies. Recent studies indicated that adjustment of buffer conditions might allow direct measurement of peptide bond formation with full-length tRNAs, but these protocols await further validation in other laboratories.[73] In addition to determination of ribosomal crystal structures, the ability to produce pure ribosome mutants for *in vitro* studies has significantly enhanced our understanding of ribosome function. Many mutations of ribosomal bases are lethal for the cell. Thus, mutant ribosomes can only be expressed in the cell in the presence of wild-type ribosomes. With the development of novel affinity tags engineered into ribosomal RNA, it was possible to purify mutant ribosomes from the cellular pool of wild-type ribosomes.[72,74,75] Thereby, it became feasible to study catalysis of peptide bond formation by ribosomes where the conserved bases in the peptidyltransferase center had been mutated in order to identify their precise roles in catalysis.[72,76,77] Lastly, these experiments

have been complemented by computational studies using molecular dynamics and free energy perturbation simulations of the peptidyltransferase center, which examined possible catalytic mechanisms.[78,79]

Since enzymatic reactions very often employ acid–base catalysis, it was suggested early on that the ribosome might use the same catalytic mechanism. In particular, A2451 was suggested as a general base since its N3 is located within 0.3–0.4 nm (3–4 Å) from the attacking α-amino group.[68] Mutation of A2451 reduces the rate of peptide bond formation with puromycin by a factor of 100.[71] Also, the rate of peptide bond formation with puromycin had been found to depend strongly on pH where protonation of a ribosomal group of pK_a 7.5 reduced the reaction rate.[71] This ribosomal group could be A2451, or it could reflect a pH-dependent conformational rearrangement required for catalysis. However, the pK_a of A2451 would have to be perturbed significantly to act as a general base at physiological pH, and the ribosome structures did not reveal how this might be achieved. Several subsequent studies provided evidence against A2451 or another ribosomal base acting as a general acid or base. Mutation of A2451 was shown to affect only the rate of peptide bond formation with puromycin, but not with full-length aminoacyl-tRNA.[72] At the same time, no effect of replacing the conserved U2506, U2585, and A2602 could be identified for reaction with aminoacyl-tRNA.[72] Similarly, mutation of A2451 in ribosomes from the Gram-positive bacterium *Mycobacterium smegmatis* did not abolish the pH dependence of the reaction with puromycin.[80] Lastly, Bieling *et al.* used a modified A-site substrate, Phelac-tRNA^{Phe} in which the nucleophilic amino group is replaced with a hydroxyl group. This substrate decreased the reaction rate of the chemical step so that accommodation was no longer rate limiting while the reaction mechanism should not be affected. Phelac-tRNA^{Phe} did not show any pH dependence in reacting with peptidyl-tRNA indicating that no general acid or base is involved in the reaction mechanism.[64] In summary, it has now clearly been shown that none of the conserved bases in the peptidyltransferase center acts as a single catalytic residue and that the ribosome does not use acid–base catalysis for peptide bond formation.

Comparisons of thermodynamic characteristics of the catalyzed and uncatalyzed reaction of peptide bond formation yielded crucial information about the mechanism of catalysis. Originally, the ribosome-catalyzed reaction with puromycin was compared with the aminolysis of the ethylene glycol ester of *N*-formylglycine by tris(hydroxymethyl)aminomethane.[81] By now, these results have been confirmed in a number of other systems.[73,80] First, these studies demonstrated a 2×10^7-fold rate enhancement of ribosomal peptide bond formation compared to the uncatalyzed reaction. Second, it was revealed that the enthalpy of activation is practically the same for the catalyzed and uncatalyzed reaction. This is another indication that the ribosome does not use conventional chemical catalysis such as acid–base catalysis. Rather, the ribosome catalyzes peptide bond formation by significantly lowering the entropy of activation leading to the conclusion that the ribosome is an 'entropy trap'.[81] In principle, a reduction in the entropy of activation might be the result of positioning the substrates within the catalytic site or of excluding water from the active site. Interestingly, the decrease in entropy of activation was also observed under saturating substrate conditions where substrate binding does not influence the first-order rate constant of ribosomal peptide bond formation. This could only be explained if substrate binding and positioning is not the major factor in lowering the entropy of activation. Indeed, computational studies confirmed that the catalytic effect is entirely of entropic origin and suggested that it is caused by reduction of solvent reorganization energy.[78]

Detailed knowledge about the transition state in ribosomal peptide bond formation helps develop an understanding of the catalytic mechanism. Toward this aim, 50S crystal structures with several different transition state analogues have been determined.[68–70] In particular, it was interesting to determine the binding pocket of the oxyanion in the transition state as well as the proximity of ribosomal groups facilitating catalysis. While an early transition state analogue was derived from puromycin (Yarus inhibitor),[68] the experimental results showing different behaviors of puromycin and full-length tRNA required larger transition state analogues to be used. To obtain an accurate picture of the ground state of peptide bond formation, Steitz and coworkers solved crystal structures with puromycin derivatives containing one or two additional cytidines to mimic the interaction of tRNA with the ribosomal A loop.[70] The presence of C74 stacking to U2590 is critical for correct positioning of the A-site substrate as it moves the attacking α-amino group by almost 0.1 nm (1 Å) in the active site. The importance of this interaction was also confirmed in biochemical experiments. Addition of a single cytidine residue to puromycin abolishes the pH dependence of peptide bond formation, which was only observed with puromycin, but not with full-length tRNA.[82] In addition, the interaction of C74 with the peptidyltransferase center results in a

90° rotation of U2541 and subtle conformational changes in the conserved core ribosomal bases G2618, U2619, and U2620, which seem to activate the catalytic center.[70] It has been proposed that these conformational changes constitute an induced-fit mechanism for peptide bond formation. However, it remains to be investigated if this conformational change is rate limiting for catalysis or if it is a rapid event due to the inherent flexibility of ribosomal bases as suggested by molecular dynamics simulations.[79] By using tetrahedral transition state analogues based on C–C–puromycin, it was subsequently determined that the ribosome-catalyzed peptidyl transferase reaction proceeds through a tetrahedral intermediate with S chirality.[69] In contrast to earlier suggestions,[68] the oxyanion is pointing away from A2451 providing even more evidence that A2451 cannot act as a general acid (**Figures 7(c)** and **7(d)**). Instead, the oxyanion seems to interact with a water molecule, which in turn is positioned by hydrogen bonds to A2602 and U2584. Interestingly, no metal ion could be found close to the active site ruling out the possibility of electrostatic transition state stabilization by a cation.[69] Lastly, the structures of different transition states bound to the peptidyltransferase center revealed a network of hydrogen bonds including the P-site A76 2′ hydroxyl group, which interacts with the α-amino group.[69]

The first functional group shown to be of direct and critical importance for ribosomal peptide bond formation was A76 2′OH of the P-site tRNA.[83] Replacement of this group with 2′H or 2′F reduces the rate of peptidyl transfer by a factor of 10^6. This indicates that the ribosome is using substrate-assisted catalysis where a substrate rather than the enzyme provides an essential functional group. Notably, the 2′F derivative could form a weak hydrogen bond to the attacking α-amino group. Thus, the large effect on peptide bond formation suggests a function of A76 2′OH beyond taking part in the hydrogen bonding network in the peptidyltransferase center. Another functional role for A76 2′OH would be to act as the general base abstracting a proton from the nearby α-amino group. However, as outlined above, there are no indications for acid–base catalysis, and this mechanism would require a substantial perturbation of the pK_a of this functional group. Instead, molecular dynamics studies suggested an essential function of A76 2′OH in a proton shuttle mechanism (**Figure 7(c)**).[78] Therein, a concerted movement of protons would take place upon attack of the P-site carbonyl by the α-amino group. The A76 2′OH could abstract a proton from the nucleophilic α-amino group while simultaneously donating a proton to the leaving 3′OH group. In fact, the structures of transition state analogues confirmed that A76 2′OH can form hydrogen bonds to both the attacking α-amino and the leaving 3′OH group.[69] Thus, a six-membered transition state could be formed. Computational investigations indicate that such a proton shuttle mechanism has a strong catalytic effect. It could be further supported by the observed intricate hydrogen bonding network in the active site, which stabilizes the position of the functional groups involved. Importantly, this network includes the conserved ribosomal bases C2063, A2451, and U2584 as well as some water molecules.[69] Recently, it has been shown that the hydrogen bonding capacity of the 2′OH of A2451's ribose is also essential for peptide bond formation.[84] Presumably, A2451 is a critical part of the hydrogen bonding network by directly interacting with P-site A76 2′OH, thus stabilizing the six-membered ring of the transition state (**Figures 7(c)** and **7(d)**). The ribosome seems to contribute an active site with several equally important bases perfectly positioned for the proposed proton shuttle without there being single specific catalytic residue. Interestingly, the proton shuttle mechanism depends on a preorganized hydrogen bond network such that no reorganization of water molecules is required as in the uncatalyzed reaction. This could explain the observed decrease in the entropy of activation of ribosome-catalyzed peptide bond formation compared to the uncatalyzed reaction.[81]

In summary, the past years have revealed the mechanism of ribosomal peptide bond formation. Upon correct substrate binding including interactions of all three terminal bases of the tRNAs with the A and P loops, respectively, the ribosome forms a preorganized hydrogen bonding network including the P-site A76 2′OH. Peptide bond formation is facilitated by a proton shuttle mechanism where A76 2′OH simultaneously receives a proton from the attacking α-amino group and donates a proton to the 3′ leaving group. Thus, ribosomal peptide bond formation is entropically driven and does not involve chemical catalysis such as acid–base or metal ion catalysis.

5.13.5 Translocation

For processive peptide polymerization, the ribosome has to move along the mRNA. Following peptide bond formation, the ribosomal A site is occupied by a peptidyl-tRNA whereas the P site contains a deacylated tRNA. During translocation, the complex of the two tRNAs with the mRNA has to move relative to the ribosome to

position the peptidyl-tRNA in the P site and the deacylated tRNA in the E site where it dissociates. This process is catalyzed by EF-G, which hydrolyzes GTP. Therein, the codon–anticodon interaction has to be maintained for the peptidyl-tRNA. The next codon then becomes positioned in an empty A site, and the ribosome is prepared for a new elongation cycle. Translocation is the least understood step of the elongation cycle since it involves large-scale movements of tRNAs, mRNA, the ribosome, and EF-G.

The ability to translocate the tRNA–mRNA complex by one codon might be an intrinsic property of the ribosome. It has been shown that the ribosome can slowly synthesize poly(Phe) based on an artificial poly(U) template in the absence of any elongation factor.[85,86] Thus, an important function of EF-G is to accelerate translocation. Interestingly, recent reports show that reverse translocation can also occur spontaneously on the ribosome.[87,88] For some tRNAs, direct reverse movement of tRNAs has been observed. Therein, a cognate deacylated tRNA in the E site translocates to the pretranslocation binding site in the P site while the peptidyl-tRNA in the P site shifts to the A site. This suggests that the ribosome has the ability to allow for spontaneous movement of the tRNA–mRNA complex, but does not confer directionality on this process. Without EF-G, the direction of movement depends on the relative affinities of the tRNAs to the different ribosome-binding sites, which seem to depend on the type of tRNA. Hence, the second function of EF-G is to impose directionality on the translocation process by preventing reverse translocation. Interestingly, a new translation factor, LepA, has been identified recently, which resembles EF-G and might catalyze reverse translocation.[89,90] It has been speculated that LepA might rescue the ribosome after a mistranslocation event. However, it remains to be shown what type of mistranslocation might occur.

The movement of the ribosome along the mRNA can be compared to the movement of cellular motor proteins along the cytoskeleton and to the movement of polymerases along DNA templates. As indicated above, EF-G ensures rapid translocation in a forward direction, a process that is driven by GTP hydrolysis by EF-G.[91] Thus, EF-G might function as a motor protein for the ribosome.[92] During translocation, ribosomes encounter secondary structure elements such as hairpins on the mRNA, and it has been shown that the ribosome can act as a helicase to resolve these structures.[93] Typically, helicases require energy and this might result from the GTP hydrolysis of EF-G during translocation. Recently, the stepwise movement of the ribosome one codon at a time along an mRNA containing a hairpin has been directly observed by single-molecule force spectroscopy.[94] These experiments show that translocation and mRNA unwinding occur simultaneously.

Translocation includes both the movement of the tRNA acceptor stems within the large ribosomal subunit as well as the shift of the tRNA anticodon stems together with the mRNA on the small ribosomal subunit. According to the hybrid state model of translocation, tRNA movement on the two ribosomal subunits can occur independently. Footprinting studies suggest that directly after peptide bond formation, the peptidyl end of A-site-bound tRNA can move into the 50S P site while the CCA end of deacylated P-site-bound tRNA shifts into the 50S E site.[95] These tRNA configurations are denoted A/P and P/E hybrid states (**Figure 8(a)**). Recent biochemical and single-molecule investigations indicate that the tRNAs can spontaneously fluctuate between a classical (A/A, P/P) and a hybrid state (A/P, P/E).[96,97] The P/E hybrid state of a tRNA could also be observed in a cryo-EM reconstruction of ribosome with bound EF-G–GDP.[98] In fact, this spontaneous movement into the hybrid state could facilitate translocation by reducing the affinity of the tRNAs for their binding sites and thus lowering the energy barrier.[99] Thereby, the energy released in peptide bond formation could at least partially be used for the large-scale movements during translocation. It is now clear that the hybrid states are a true intermediate of translocation[100] as binding of EF-G shifts the equilibrium between the classical and hybrid states toward the latter state.[101] Currently, it is under debate if the two tRNAs move simultaneously or sequentially into the hybrid state. There are some indications that an intermediate state can be observed where the P-site tRNA is in a hybrid P/E state while the A-site tRNA is still in the classical A/A state.[102–104] Obviously, the P-site tRNA has to shift first in order to free the 50S P site for the incoming A-site tRNA. However, it remains to be investigated whether the A-site tRNA movement follows instantaneously upon hybrid state formation in the P/E site or if there is a delay between A- and P-site tRNA translocation on the 50S subunit.

In order to allow for translocation of the tRNA–mRNA complex, the ribosome will have to undergo conformational changes as well. The contacts described above between the decoding center and the codon–anticodon helix as well as the base pairs between the 50S A and P loops and the tRNA acceptor stems will have

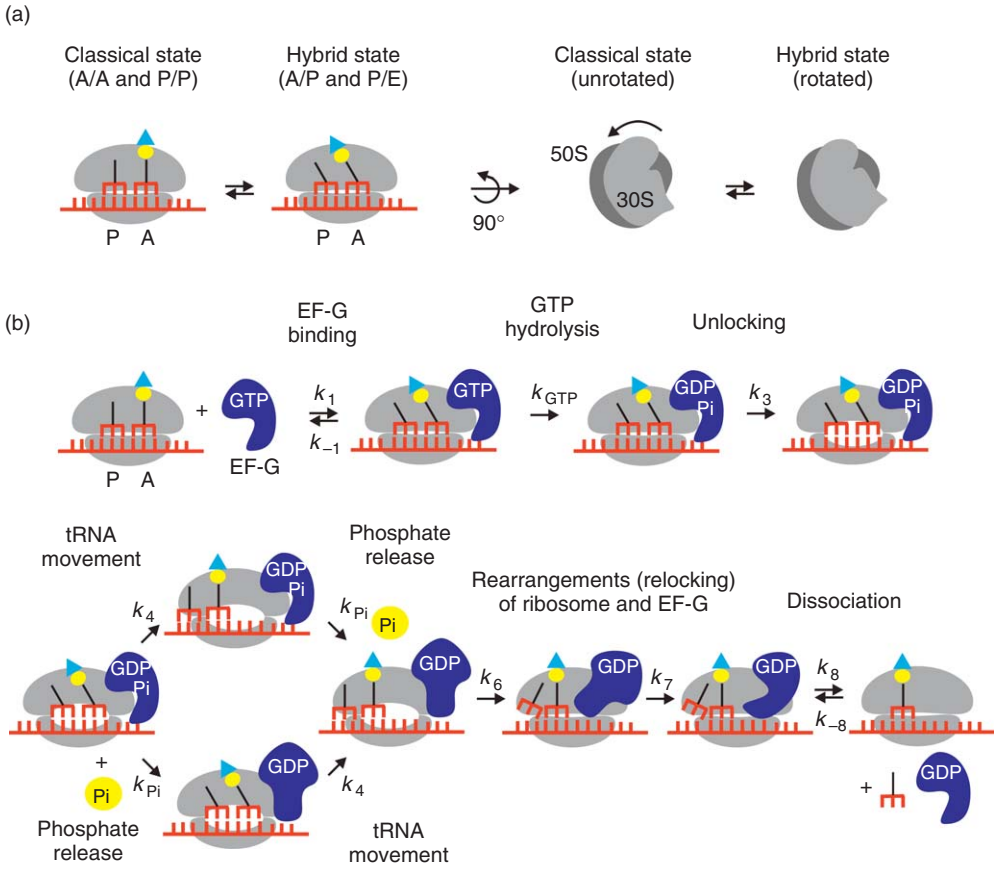


Figure 8 EF-G-catalyzed translocation of the tRNA–mRNA complex within the ribosome. (a) Hybrid state formation and intersubunit rotation. Upon peptide bond formation, the ribosome fluctuates between the classical state and a hybrid state. In the classical state, the tRNAs are bound to the A and P site on both the 30S and 50S subunit. In the hybrid state, the anticodons remain in the A and P site on the 30S subunit whereas the acceptor ends move into the P and E site on the 50S subunit, respectively. Simultaneously to hybrid state formation, the 30S subunit rotates relative to the 50S subunit as shown on the right site. (b) Kinetic mechanism of EF-G-catalyzed translocation. Upon GTP hydrolysis, unlocking occurs through a ribosomal rearrangement. Only subsequently, tRNA and mRNA movement as well as dissociation of the inorganic phosphate from EF-G take place.

to be resolved. To date, little is known about local conformational changes within the ribosome during translocation as high-resolution structures of translocation intermediates are missing. However, low-resolution cryo-EM reconstructions of pretranslocation and post-translocation ribosomes as well as ribosomes bound to EF-G have revealed global conformational changes within the ribosome.[26,27] Binding of EF-G induces a rotation of the small ribosomal subunit relative to the large ribosomal subunit by 6° in the direction of mRNA movement during translocation (**Figure 8(a)**).[105] At the same time, the L1 stalk of the 50S subunit, which is close to the E site undergoes a large-scale movement that might facilitate movement of the deacylated tRNA into the E site.[98,106] Independently of intersubunit movement, a rotation of the small subunit head has been observed that could further assist translocation of the tRNA–mRNA complex.[8,107,108] Lastly, the GTPase-activating center changes its conformation upon interaction with EF-G.[26,109,110]

The intersubunit rotation is required for translocation as ribosomes trapped in the nonrotated state by an engineered intersubunit disulfide bridge fail in tRNA–mRNA movement.[111] Real-time observation of intersubunit movement by fluorescence resonance energy transfer (FRET)[112] showed that intersubunit movement occurs concomitantly with hybrid state formation, and that the rotated state can be trapped by the antibiotic viomycin.[113] Similarly to the fluctuation of tRNAs between classical and hybrid states,[102] single-molecule studies have detected spontaneous intersubunit movement where the 30S subunit fluctuates between a rotated

and nonrotated state.[114] Thus, hybrid state formation and intersubunit rotation occur simultaneously and spontaneously following peptidyl transfer. It is the function of EF-G during translocation to induce a stable hybrid state of the tRNAs with rotated subunits.[101,114]

The most challenging question regarding translocation addresses the exact functional role of EF-G and how it relates to structural changes of the ribosome, the tRNA–mRNA complex, and EF-G itself. In particular, the concerted movement of tRNAs and mRNA on the 30S subunit remains poorly understood. More insight into EF-G's function has been derived from rapid kinetic experiments (**Figure 8(b)**). EF-G binds to the ribosome in the form of EF-G–GTP, and the interaction with the GTPase-activating center closes the nucleotide-binding pocket.[115] The GTPase-activating center seems to induce GTP hydrolysis by stabilizing the switch regions of EF-G.[116] In addition, ribosomal protein L7/12 contributes to GTPase activation.[11,117,118] The kinetic experiments clearly revealed that GTP hydrolysis by EF-G precedes the movement of tRNAs and mRNA[92] indicating that the energy from GTP hydrolysis is used to induce translocation. Thus, EF-G might act as a motor protein to drive tRNA and mRNA movement. It has been suggested that GTP hydrolysis by EF-G causes a rotation of domains 3–5 relative to domains 1 and 2 of EF-G, which positions the tip of domain 4 in the decoding center.[108] In fact, conformationally restricted EF-G can hydrolyze GTP, but is unable to promote translocation,[119] and the tip of domain 4 is likewise required for translocation.[120] Interestingly, release of inorganic phosphate from EF-G is significantly delayed relative to GTP hydrolysis indicating that EF-G remains bound to the ribosome in an activated EF-G–GDP–Pi complex.[121] Therein, EF-G resembles motor proteins such as myosin. It was shown that GTP hydrolysis by EF-G induces a rate-limiting ribosomal rearrangement called unlocking, which precedes both tRNA–mRNA movement and release of inorganic phosphate.[121] One hypothesis that remains to be proven is that this unlocking corresponds to stabilization of intersubunit rotation and hybrid state formation. Another hypothesis is that EF-G binding is sufficient to induce the hybrid states. Rather, ribosome unlocking might represent the movement of EF-G domain 4 into the decoding center. There, EF-G might disrupt the interactions of the codon–anticodon helix with the 30S ribosomal subunit allowing translocation on the 30S subunit. Following unlocking, complete tRNA–mRNA translocation takes place in parallel to release of inorganic phosphate from EF-G.[121] Therein, the codon–anticodon interaction of the peptidyl-tRNA has to be maintained during its movement from the A to the P site. Pi release is controlled by ribosomal protein L7/12, which might be important for maintaining the conformational coupling between EF-G and the ribosome.[122] Following translocation and Pi release, the ribosome and EF-G have to undergo further conformational changes, but these are not well characterized. It has been shown that the post-translocation state with a peptidyl-tRNA in the P site stabilizes the nonrotated conformation of the ribosome.[114] Thus, relocking might involve a reverse rotation of the 30S subunit relative to the 50S subunit and maybe also a rearrangement of the 30S head into its ground state position. Also, EF-G–GDP has to dissociate from the ribosome.

In summary, translocation involves large-scale conformational changes of the ribosome, the tRNA–mRNA complex, and EF-G. Translocation of the tRNA acceptor ends on the 50S subunit can occur spontaneously following peptide bond formation, but is not stable as the tRNAs fluctuate between different states (classical versus hybrid state). GTP hydrolysis by EF-G accelerates translocation by stabilizing a ribosome conformation that allows tRNA–mRNA movement. By positioning its domain 4 into the decoding center, EF-G might facilitate translocation on the small ribosomal subunit and sterically prevent reverse translocation by blocking the A site. Following translocation, the presence of the peptidyl moiety in the 50S P site locks the ribosome in the post-translocation state by preventing intersubunit rotation.

5.13.6 Termination

Upon encountering a stop codon on the mRNA, the ribosome will halt incorporation of further amino acids into the polypeptide as there is no tRNA complementary to a stop codon (UAG, UGA, UAA). In order to liberate the polypeptide, the ester bond between the peptide and the tRNA residing in the P site has to be hydrolyzed – a reaction that is also catalyzed in the peptidyltransferase center. It is critical for protein synthesis that peptide release is tightly coupled to the presence of a stop codon in the decoding center to avoid premature termination resulting in shortened, nonfunctional proteins. Both functions, recognizing the stop codon and triggering

peptide release, are performed by a class I release factor, RF1 or RF2 in prokaryotes. Additionally, a class II release factor (RF3) is involved in termination. The main questions regarding termination address (1) the specificity of stop codon recognition by protein factors, (2) the catalytic mechanism of ester bond hydrolysis in comparison to peptide bond formation, and (3) the functional role of RF3, a GTPase related to EF-Tu and EF-G, within the overall mechanism of termination.

Class I release factors resemble tRNAs as they bind to the ribosomal subunit interface and interact simultaneously with the 30S decoding center and the 50S peptidyltransferase center. The two class I release factors, RF1 and RF2, share high sequence identity and are supposed to be structurally and functionally very similar except for their interaction with the stop codon. Both RF1 and RF2 recognize the UAA stop codon. Additionally, RF1 recognizes the UAG stop codon, while RF2 recognizes UAG. Low-resolution cryo-EM and crystal structures of RF1 or RF2 bound to the ribosome identified the overall binding site of the class release factor. Domains 2/4 of RF1/2 interact with the decoding center whereas domain 3 reaches into the peptidyltransferase center.[34,123,124] Domain 1 is situated close to the GTPase-activating center where it might contact the GTPase RF3 during termination. The same binding geometry has very recently been visualized by high-resolution crystal structures as well.[10] The open conformation of RF2 observed on the ribosome has been confirmed by solution structures[125] in contrast to the closed structure determined by crystallography.[126] These findings indicate that RF1/2 could undergo some conformational changes during stop codon recognition and induction of peptide release.

Stop codon recognition by class I release factors is extremely accurate with a 10^3–10^6-fold discrimination against sense codons to prevent premature termination.[127] In contrast to aminoacyl-tRNA selection by EF-Tu, no function of the GTPase RF3 in promoting accuracy could be identified. Currently, it remains unknown if the kinetic mechanism of termination, which has not yet been fully characterized contributes to fidelity. More progress has been made toward understanding the interaction between RF1/2 with the stop codons. By mutagenesis and domain swaps, tripeptides in RF1 and RF2 have been identified, which determine the specificity of the two factors: a Pro–Ala–Thr (consensus sequence: PxT) motif in RF1 and a Ser–Pro–Phe (consensus sequence: SxP) motif in RF2 seem to be important for stop codon recognition.[128] These motifs can directly interact with the stop codon as they are located in a loop of domain 2/4 close to the ribosomal decoding site.[10,125,126] Interestingly, the recognition of stop codons by release factors and of sense codons by tRNAs seems to depend on different conformational changes in the decoding center as mutations in ribosomal RNA, removal of 2′OH groups in the codon, and addition of antibiotics affect the two steps differently.[129] In both cases, an induced-fit mechanism involving local structural rearrangements in the decoding center has been proposed. The very recent crystal structure of a ribosome–RF1 complex at 0.32 nm (3.2 Å) resolution in 2008 provides detailed insight into stop codon recognition and explains previous biochemical findings (**Figure 9**).[10] Indeed, the PxT motif of RF1 interacts with the stop codon, and the structural changes in the decoding center are quite different from the ones during sense codon recognition. Interestingly, A1492 and G530 of 16S rRNA take part in stop and sense codon recognition although by completely different interactions. A1492 flips out of helix 44 to allow its backbone to interact with the riboses at codon positions 1 and 2 while A1493 remains stacked within helix 44. There, A1493 interacts with the tip of helix 69 from the large subunit, which explains the role of this helix in peptide release.[130] G530 is in its anti conformation and stacks on the third codon base that is unstacked from the first two codon bases. These interactions agree with the proposed local induced-fit mechanism in the decoding site that differs from sense codon recognition.[129] A hydrogen bonding network between the codon bases and RF1 is responsible for the specificity of stop codon recognition.[10] These interactions include the tripeptide PxT as well as other conserved elements of RF1 (**Figure 9**). Thus, we now understand on a structural level how RF1 reads a stop codon. It remains to be investigated how the individual interactions contribute to the observed high fidelity of stop codon recognition.

In the peptidyltransferase center, the class I release factors trigger hydrolysis of the ester bond linking the polypeptide to the P-site tRNA. From a chemical perspective, this reaction is more challenging than peptide bond formation because the attacking water oxygen is significantly less nucleophilic than the α-amino group. Furthermore, hydrolysis has to be prevented during the normal elongation cycles as the peptidyl-tRNA has to be stable in the P site. This raises the question of how peptide release is regulated and whether the catalytic mechanism for peptide release is related to peptide bond formation. While hardly anything is known about the coupling of stop codon recognition to peptide release, several functional groups have been identified to be

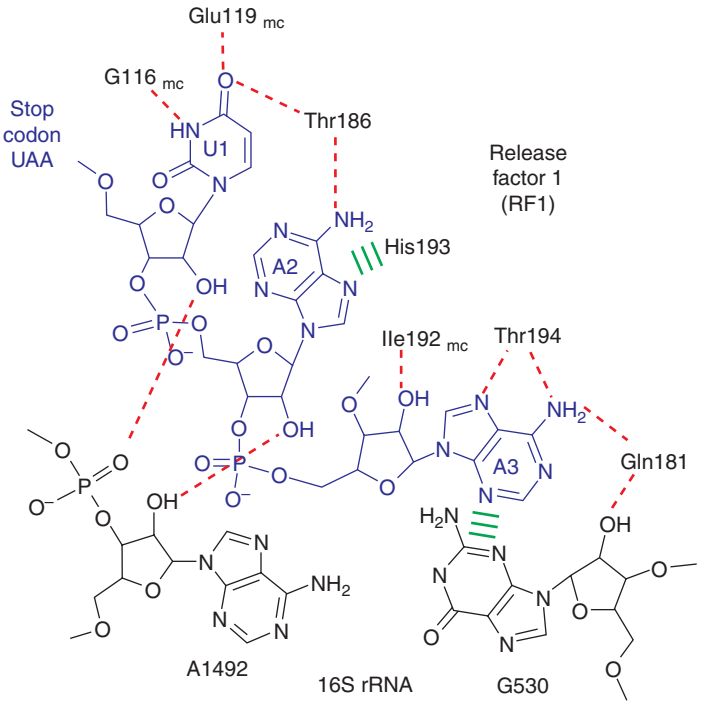


Figure 9 Schematic representation of stop codon recognition by release factor 1. Interactions shown are based on the crystal structure of the ribosome–RF1 complex.[10] The high specificity for the UAA stop codon is achieved through a hydrogen bonding network (red dashes) between the codon bases and RF1, which includes Thr186 of the conserved PxT tripeptide (interactions with RF1's main chain backbone are indicated by a lower case mc). Of the ribosomal RNA, the backbone of A1492 forms hydrogen bonds to the 2′ hydroxyl groups in the first and second codon position while G530 stacks upon the adenine in the third codon position (stacking interactions are shown by green parallel dashes).

crucial for catalysis. Of the ribosomal residues, A2602 and in particular its ribose is most important for peptide release.[72,131,132] Interestingly, mutations of the other conserved nucleotides in the peptidyltransferase center also impair peptide release without affecting peptidyl transfer.[72] Positioning of the P-site tRNA by the C74-G2252 base pair to the P loop is likewise critical for peptide release,[133] and again the 2′OH of the P-site A76 is absolutely required for catalysis.[134] In addition to ribosomal functional groups, the release factor could be directly involved in catalysis. Indeed a critical GGQ motif (Gly–Gly–Gln) has been identified in the class I release factors, which reaches into the peptidyltransferase center.[10,135–138] Nevertheless, the exact function of this motif, in particular of the glutamine, remains unclear. While mutation of the central glycine dramatically inhibits peptide release,[138] most substitutions of the glutamine affect the reaction only mildly.[137] Only mutations to asparagine or aspartate significantly decrease the rate of peptide hydrolysis. Interestingly, it has been shown that the release factor renders the reaction specific to water as other nucleophiles can attack the ester bond in the absence of a release factor, and it is the glutamine residue that is responsible for this specificity. In particular, upon mutation of this glutamine to a small residue, aminolysis of the ester bond by hydroxylamine becomes favorable over hydrolysis.[137] These results can be explained by a catalytic mechanism where the release factor on the one hand stimulates the active site without direct contribution of the conserved glutamine to allow for rapid peptide release. On the other hand, the glutamine of the release factor could form a defined cavity and interact with the nucleophilic water molecule thus restricting access of larger nucleophiles to the active site.[137] In fact, such a mechanism of peptide release in combination with a proton shuttle has been proposed based on computational studies.[139]

The new crystal structure of the ribosome–RF1 complex sheds more light into the interactions between the GGQ motif and the peptidyltransferase center.[10] This complex represents the product state of peptide release since a deacylated tRNA is bound to the P site. Importantly, the main chain amide of the conserved glutamine hydrogen bonds to the 3′OH of A76 in the P site, which is the leaving group of the hydrolysis

reaction. Thus, the main chain amide might facilitate catalysis by coordinating the leaving group and maybe stabilizing the oxyanion in the transition state. The exact role of the glutamine side chain will only be elucidated by a high-resolution structure of the substrate or transition state complex. In general, the ribosome–RF1 structure is compatible with the proposed mechanism wherein this glutamine interacts with the attacking water molecule. The 2′OH of A76 could be part of a proton shuttle mechanism similar to the catalytic mechanism of peptide bond formation. In contrast, the crystal structure does not support a catalytic role of a ribosomal residue.[10] The critical A2602 is clearly important for positioning of the GGQ motif as it is buried in a cavity of RF1, but cannot reach directly into the active site. Likewise, the conserved glycines of RF1 are required for positioning the glutamine as this loop of RF1 makes a tight turn within the peptidyltransferase center.

Regarding peptide release, a conserved glutamine of the class I release factor RF1 or RF2 and the A76 2′OH of the P-site tRNA seem to be intimately involved in catalyzing hydrolysis of the ester bond between peptide and tRNA during termination. The catalytic mechanism might consist of a proton shuttle comparable to the mechanism of peptide bond formation. Further experiments will be required to identify the exact catalytic mechanism of peptide release and how the ribosome modulates the function of its active center. Regarding the signaling of stop codon recognition to the peptidyltransferase center, a conformational change between domains 3 and 4 of RF1 has been suggested, which would be required for correct positioning of domain 3 in the peptidyltransferase center; however, this proposed rearrangement awaits further verification.[10]

While EF-Tu's GTPase activity is significantly contributing to the fidelity of aminoacyl-tRNA selection, GTPase activity of RF3 has been suggested to have a different role. Biochemical studies by Ehrenberg and coworkers indicate that RF3 will bind in its GDP-bound form to the ribosome-RF1/2 complexes.[140] Cryo-EM structures show that RF3 interacts with the GTPase-activating center on the 50S subunit similarly to EF-Tu and EF-G.[30] Only if peptide release has taken place will the ribosome complex facilitate exchange of GTP for GDP in RF3.[138] It might be that peptide release induces conformational changes in the ribosome such as P/E hybrid state formation of the tRNA and intersubunit rotation, which influence RF3.[30] Since RF3 is a GTPase, its conformation will presumably be different in the GTP and GDP state. Thus, binding of GTP might induce conformational changes in RF3, which tightens its interaction with the ribosome and ultimately results in dissociation of RF1/2 from the ribosome.[140] Removal of RF1/2 might be caused by disruption of its interactions with the decoding center and the GTPase-activating center. According to this model, GTP hydrolysis by RF3 will only occur at the end of termination and result in dissociation of RF3–GDP. The overall function of RF3 is thus to catalyze dissociation of the class I release factor from the ribosome after peptide release. Interestingly, eukaryotic release factors are not closely related to their prokaryotic counterparts and a different role for eRF3 during termination has been suggested.[141]

In conclusion, the last years have provided substantial insights into the molecular mechanism of termination. We are now beginning to understand the extremely accurate recognition of the stop codon by protein release factors. Also, significant progress has been made in elucidating the modulation of the peptidyltransferase center required to catalyze both peptide bond formation and peptide release depending on the ligand bound to the A site (aminoacyl-tRNA versus class I release factor, respectively). However, the exact interplay between stop codon recognition and activation of peptide release remains poorly understood. Future investigations are likely to enhance our knowledge of termination to a level similar to that of the elongation cycle.

5.13.7 Ribosomal Incorporation of Non-Natural Amino Acids

Based on our current understanding of ribosomal protein synthesis, several strategies have been developed to incorporate amino acids other than the 20 standard proteinogenic amino acids into a peptide using the ribosomal machinery[142–144]. This allows for the design of peptides with novel properties. On the one hand, such a system can be used to synthesize nonstandard peptides that are important pharmaceuticals. In nature, such peptides are produced by nonribosomal peptide synthetases, which operate in complex pathways. On the other hand, non-natural residues are a useful tool in biochemistry and biophysics to study proteins. For example, incorporation of non-natural residues by the ribosome allows for site-specific labeling of proteins with spin labels for electron paramagnetic resonance spectroscopy, with

fluorescence labels, or with crosslinkers. However, several challenges have to be overcome for ribosomal synthesis of non-natural peptides. (1) The genetic code, that is, the codon–anticodon interaction has to be altered to program incorporation of a non-natural amino acid at a specific site in a peptide. (2) Novel aminoacyl-tRNAs have to be created where the non-natural amino acid is attached to a tRNA, which can read the altered genetic code. (3) The non-natural amino acid has to be accepted in the peptidyltransferase center of the ribosome and take part in peptide bond formation. In this chapter, only a brief overview of how the different challenges have been addressed can be provided.

To date, several methods have been developed to assign a non-natural amino acid to a codon within an mRNA (**Figure 10**). First of all, nonsense suppression has been used where a stop codon, typically the stop codon UAG, can be read by a suppressor tRNA with the anticodon CUA, which has been charged with the non-natural amino acid.[145] This strategy is related to the ribosomal incorporation of selenocysteine or pyrrolysine in natural proteins.[146,147] However, in this system the suppressor tRNA is competing with release factor 1 (**Figure 10(a)**). Thus, even by using a high concentration of suppressor tRNA, premature release of peptides at the codon for the non-natural amino acids cannot be avoided *in vivo*. Recently, this problem has been solved by the development of a highly purified *in vitro* translation system (PURE system), which allows omitting RF1 if UAG is used only for the non-natural amino acid, but not as a stop codon.[148,149] In order to incorporate several non-natural amino acids, sense suppression has to be used where sense codons are reassigned to the non-natural amino acids.[150,151] This requires a purified translation system where the standard amino acids whose codons have been reassigned can be omitted to avoid competition between the suppressor tRNA and the natural tRNAs. Lastly, new codon–anticodon pairs have been designed to

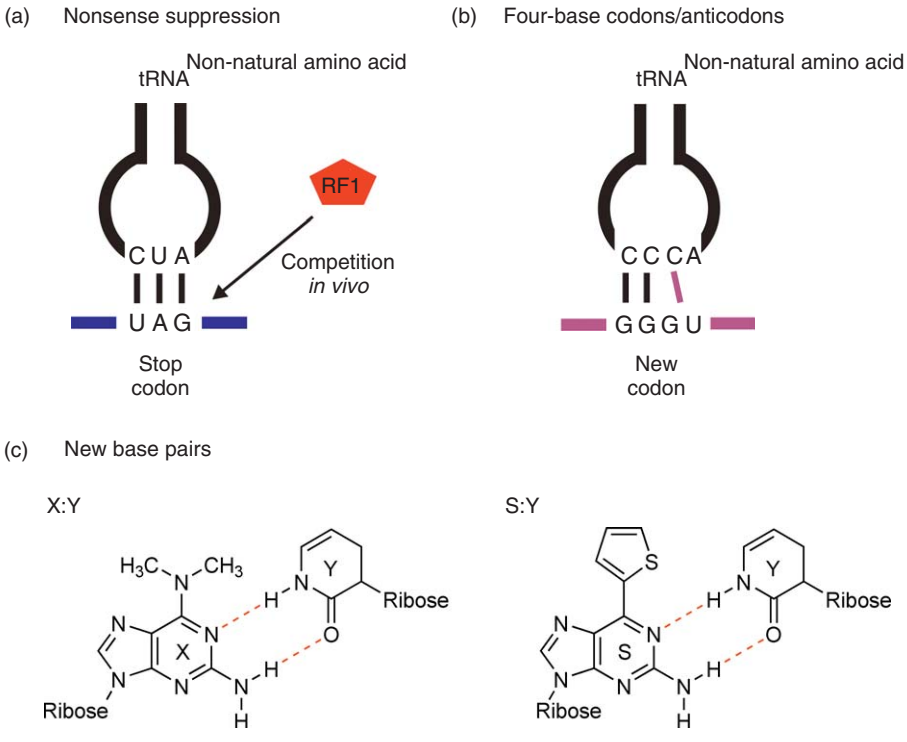


Figure 10 Alteration of the genetic code for incorporation of non-natural amino acids. (a) In nonsense suppression, the stop codon UAG is decoded by a non-natural tRNA with the anticodon CUA. *In vivo* decoding of the UAG codon by this tRNA is in competition with termination of protein synthesis by release factor 1 (RF1). Purified *in vitro* translation systems allow omission of RF1 from the reaction mixture. (b) A new codon–anticodon pair can be created using four-base codons such as GGGU.[153] Crystal structures of these codon–anticodon complexes in the ribosomal decoding center revealed that the C in the third anticodon position interacts with both the third and fourth codon position (purple line) while the extra A in the anticodon loop does not contact the codon.[156] (c) Non-natural base pairs also allow creation of new codon–anticodon pairs. Shown here is the interaction of the base Y with either base X or S[159,160] (hydrogen bonds are indicated by red dashes).

extend the genetic code to more than the 20 standard amino acids. For example, four-base and five-base codon–anticodon interactions have been identified that can be used for non-natural amino acids (**Figure 10(b)**).[152–154] In nature, four-base anticodons have been found in suppressor tRNAs, which induce a +1 frameshift on the mRNA by interacting with a four-base codon showing the flexibility of the ribosome in reading codon–anticodon interactions.[155,156] As an alternative, unnatural base pairs have been developed such as isoguanosine and isocytidine[157,158] or 2-amino-6-(*N*,*N*-dimethylamino)purine (X), pyridine-2-one (Y), and 2-amino-6-(2-thienyl)purine (S) (**Figure 10(c)**).[159,160] In general, it is now possible to assign multiple codons within an mRNA to novel non-natural amino acids.

Since aminoacyl-tRNAs are the adaptor molecules linking the mRNA sequence to the peptide sequence, the second challenge resides in the synthesis of novel aminoacyl-tRNAs. Usually, aminoacyl-tRNAs are generated by highly specific aminoacyl-tRNA synthetases with a powerful editing system to avoid misaminoacylation. Novel combinations of non-natural amino acids and tRNAs can be achieved by chemically attaching an amino acid to a short dinucleotide, which is subsequently connected to a truncated tRNA by RNA ligase.[161,162] However, this is a very labor-intensive process. Therefore, strategies have been developed to use aminoacyl-tRNA synthetases also for non-natural amino acids. These are so-called orthogonal tRNA synthetase pairs, which are inert to endogenous tRNAs and synthetases, that is, the suppressor tRNA cannot be charged by natural aminoacyl-tRNA synthetases and the new aminoacyl-tRNA synthetase cannot react with natural tRNAs. This can be achieved either by using a promiscuous synthetase from a different organism than the translation system, or by using mutant synthetases and tRNAs.[163,164] But again, each tRNA synthetase pair has to be generated individually and optimized, for example, by *in vitro* evolution. Interestingly, Szostak and coworkers have shown that natural aminoacyl-tRNA synthetases can charge natural tRNAs with a variety of non-natural amino acids when high concentrations are used (**Figure 11(a)**).[165] In combination with the pure *in vitro* translation system where standard amino acids can be omitted from the reaction mixture, this allows efficient incorporation of up to 13 different non-natural amino acids in one peptide.[166] Lastly, a powerful aminoacylation system using ribozymes has been developed by Suga and coworkers. These ribozymes called flexizymes can attach any amino acid to any tRNA (**Figure 11(b)**).[167] The flexizyme uses any activated amino acid that is esterified to a 3,5-dinitrobenzyl, a 4-chloro-benzyl group, or a cyanomethyl group. Since the flexizyme recognizes the leaving group, there is no limitation on the nature of the amino acid esterified to the leaving group. The tRNA is recognized by its conserved 3′CCA end[168] so that any tRNA can be used in this

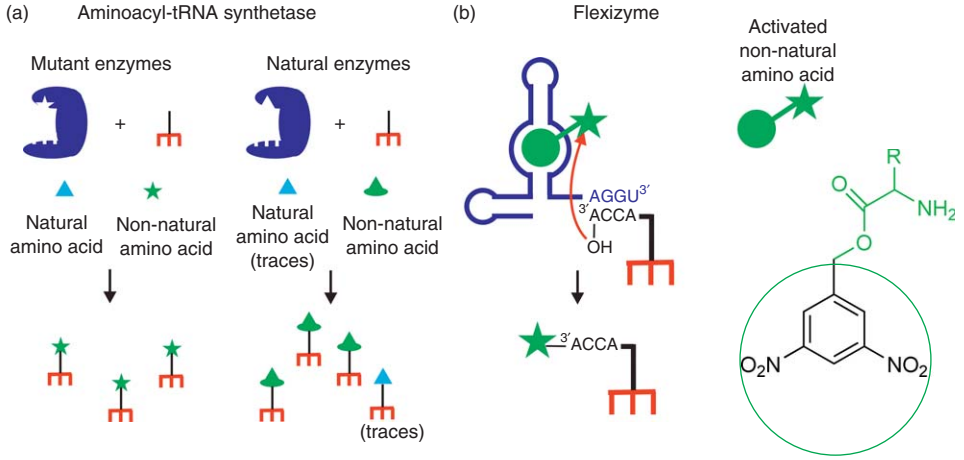


Figure 11 Aminoacylation of tRNAs with non-natural amino acids. (a) Mutant aminoacyl-tRNA synthetases can be engineered to selectively accept only the non-natural amino acid also in the presence of the natural amino acid (left site). Alternatively, naturally occurring aminoacyl-tRNA synthetases can accept non-natural amino acids in certain cases and attach them to a natural tRNA (right site). To ensure high efficiency, the natural amino acid typically used by this synthetase has to be efficiently removed from the translation system. (b) Flexizyme is a small ribozyme that recognizes any tRNA through its 3′CCA end and any activated amino acid through its ester group. The dFlexizyme shown here interacts with the 3,5-dinitrobenzylalcohol leaving group allowing aminoacylation of any tRNA with any non-natural amino acid.

in vitro aminoacylation system as long as no other tRNAs are present. By using orthogonal tRNAs, reaminoacylation of these tRNAs by natural aminoacyl-tRNA synthetases in the translation reaction can be avoided. In conclusion, multiple ways have been developed to generate tRNAs charged with non-natural amino acids for ribosomal peptide synthesis.

Lastly, the ribosome has to be able to catalyze peptide bond formation with these non-natural amino acids. Thus, the question is whether the ribosome itself discriminates against non-natural amino acids. In fact, many of the tRNAs charged with non-natural amino acids take part in ribosomal protein synthesis.[166] This demonstrates that the ribosome can accept most side chain derivatives. Impressively, the ribosomal P site basically has no limitations in accepting amino acid derivatives as it has been shown that large side chains such as fatty acids as well as D-amino acids can be used as substrates during initiation of protein synthesis.[169,170] In the A site, the ribosome is more stereospecific and typically does not accept D-amino acids. However, recent studies using mutant ribosomes indicate that it is possible to expand the ribosomal tolerance for D-amino acids.[171] Interestingly, the ribosome not only catalyzes peptide bond formation, but also accepts changes in the backbone. For example, *N*-methyl and *N*-ethyl substituted residues are readily incorporated into peptides by the ribosome.[172] Notably, the ribosome also synthesizes polyesters when α-hydroxy amino acids are used.[173] These findings demonstrate that a variety of different non-natural amino acids can be incorporated at multiple sites within a peptide through ribosomal synthesis. The main future task lies in optimization, in particular to increase the efficiency and the yield of non-natural peptide syntheses. The speed of translation is significantly enhanced if the non-natural aminoacyl-tRNAs can be delivered to the ribosome by EF-Tu. However, EF-Tu has some specificity for correctly aminoacylated, natural aminoacyl-tRNAs and discriminates against misaminoacylated tRNAs.[174] Accordingly, mutant EF-Tu variants are under development to facilitate interaction with large aromatic amino acids attached to the tRNA.[175]

The combination of all these powerful tools to manipulate ribosomal protein synthesis has allowed the synthesis of drug-like nonstandard peptides containing *N*-methylations on their backbones. As with natural product peptides, these ribosome-synthesized peptides can be cyclic. For example, a 2-chloroacyl group attached to the initiator amino acid spontaneously reacts with a single cysteine forming a stable thioether bond.[176] Thus, a large variety of different peptides containing non-natural amino acids can now be synthesized by the ribosome. In summary, the progress in understanding ribosomal protein synthesis in the last 10 years has made it possible to use the ribosome as a tool to produce artificial proteins and peptides for applications in basic research and in pharmacology.

5.13.8 Conclusion

Tremendous progress has been made in understanding ribosomal protein synthesis on a molecular to atomic level since determination of high-resolution ribosome crystal structures in 2000. In particular, the combination of X-ray crystallography of ribosomes bound to different ligands with state-of-the-art biophysical methods such as rapid kinetics provided new insight into the long-studied ribosomal machinery. Only by this combination, it has been possible to obtain knowledge on the dynamic nature of ribosomal protein synthesis on a molecular level. Today, it is well established that the ribosome is a ribozyme catalyzing peptide bond formation solely with the help of RNA and without involvement of any protein. We now also understand how the interaction of many translation factors with the ribosome increases the accuracy and speed of protein synthesis in the living cell. This knowledge is instrumental for the development of new antibiotics targeting the ribosome as well as for the construction of novel peptides containing non-natural amino acids for basic research and pharmacology. These successes promise to be continued in the future. In particular, many questions on the details of translation factor functioning remain to be investigated. For example, it still only poorly understood how the ribosome activates the GTPases IF2, EF-Tu, EF-G, and RF3 in a controlled fashion. In conclusion, ribosomal protein synthesis has been under investigation since the 1960s, and promises to remain an interesting research area in the future thanks to the new tools in ribosome research developed in the last decade.

References

1. W. Doolittle, *Science* **1999**, *284* (5423), 2124–2129.
2. A. Yonath; H. G. Wittmann, *Biophys. Chem.* **1988**, *29* (1–2), 17–29.
3. N. Ban; P. Nissen; J. Hansen; P. B. Moore; T. A. Steitz, *Science* **2000**, *289* (5481), 905–920.
4. F. Schluenzen; A. Tocilj; R. Zarivach; J. Harms; M. Gluehmann; D. Janell; A. Bashan; H. Bartels; I. Agmon; F. Franceschi; A. Yonath, *Cell* **2000**, *102* (5), 615–623.
5. B. T. Wimberly; D. E. Brodersen; W. M. Clemons, Jr.; R. J. Morgan-Warren; A. P. Carter; C. Vonrhein; T. Hartsch; V. Ramakrishnan, *Nature* **2000**, *407* (6802), 327–339.
6. J. Harms; F. Schluenzen; R. Zarivach; A. Bashan; S. Gat; I. Agmon; H. Bartels; F. Franceschi; A. Yonath, *Cell* **2001**, *107* (5), 679–688.
7. M. M. Yusupov; G. Z. Yusupova; A. Baucom; K. Lieberman; T. N. Earnest; J. H. Cate; H. F. Noller, *Science* **2001**, *292* (5518), 883–896.
8. B. S. Schuwirth; M. A. Borovinskaya; C. W. Hau; W. Zhang; A. Vila-Sanjurjo; J. M. Holton; J. H. Cate, *Science* **2005**, *310* (5749), 827–834.
9. M. Selmer; C. M. Dunham; F. V. Murphy IV; A. Weixlbaumer; S. Petry; A. C. Kelley; J. R. Weir; V. Ramakrishnan, *Science* **2006**, *313* (5795), 1935–1942.
10. M. Laurberg; H. Asahara; A. Korostelev; J. Zhu; S. Trakhanov; H. F. Noller, *Nature* **2008**, *454* (7206), 852–857.
11. M. Diaconu; U. Kothe; F. Schlunzen; N. Fischer; J. M. Harms; A. G. Tonevitsky; H. Stark; M. V. Rodnina; M. C. Wahl, *Cell* **2005**, *121* (7), 991–1004.
12. A. Korostelev; S. Trakhanov; M. Laurberg; H. F. Noller, *Cell* **2006**, *126* (6), 1065–1077.
13. G. Z. Yusupova; M. M. Yusupov; J. H. Cate; H. F. Noller, *Cell* **2001**, *106* (2), 233–241.
14. J. M. Ogle; D. E. Brodersen; W. M. Clemons, Jr.; M. J. Tarry; A. P. Carter; V. Ramakrishnan, *Science* **2001**, *292* (5518), 897–902.
15. S. Trobro; J. Aqvist, *Biochemistry* **2008**, *47* (17), 4898–4906.
16. T. Auerbach; A. Bashan; A. Yonath, *Trends Biotechnol.* **2004**, *22* (11), 570–576.
17. T. Hermann, *Curr. Opin. Struct. Biol.* **2005**, *15* (3), 355–366.
18. J. L. Hansen; J. A. Ippolito; N. Ban; P. Nissen; P. B. Moore; T. A. Steitz, *Mol. Cell* **2002**, *10* (1), 117–128.
19. J. L. Hansen; P. B. Moore; T. A. Steitz, *J. Mol. Biol.* **2003**, *330* (5), 1061–1075.
20. F. Schlunzen; R. Zarivach; J. Harms; A. Bashan; A. Tocilj; R. Albrecht; A. Yonath; F. Franceschi, *Nature* **2001**, *413* (6858), 814–821.
21. F. Schlunzen; J. M. Harms; F. Franceschi; H. A. Hansen; H. Bartels; R. Zarivach; A. Yonath, *Structure* **2003**, *11* (3), 329–338.
22. A. P. Carter; W. M. Clemons; D. E. Brodersen; R. J. Morgan-Warren; B. T. Wimberly; V. Ramakrishnan, *Nature* **2000**, *407* (6802), 340–348.
23. D. E. Brodersen; W. M. Clemons, Jr.; A. P. Carter; R. J. Morgan-Warren; B. T. Wimberly; V. Ramakrishnan, *Cell* **2000**, *103* (7), 1143–1154.
24. H. Stark; M. V. Rodnina; H. J. Wieden; F. Zemlin; W. Wintermeyer; M. van Heel, *Nat. Struct. Biol.* **2002**, *9* (11), 849–854.
25. M. Valle; J. Sengupta; N. K. Swami; R. A. Grassucci; N. Burkhardt; K. H. Nierhaus; R. K. Agrawal; J. Frank, *EMBO J.* **2002**, *21* (13), 3557–3567.
26. R. K. Agrawal; P. Penczek; R. A. Grassucci; J. Frank, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (11), 6134–6138.
27. H. Stark; M. V. Rodnina; H. J. Wieden; M. van Heel; W. Wintermeyer, *Cell* **2000**, *100* (3), 301–309.
28. G. S. Allen; A. Zavialov; R. Gursky; M. Ehrenberg; J. Frank, *Cell* **2005**, *121* (5), 703–712.
29. A. G. Myasnikov; S. Marzi; A. Simonetti; A. M. Giuliodori; C. O. Gualerzi; G. Yusupova; M. Yusupov; B. P. Klaholz, *Nat. Struct. Mol. Biol.* **2005**, *12* (12), 1145–1149.
30. H. Gao; Z. Zhou; U. Rawat; C. Huang; L. Bouakaz; C. Wang; Z. Cheng; Y. Liu; A. Zavialov; R. Gursky; S. Sanyal; M. Ehrenberg; J. Frank; H. Song, *Cell* **2007**, *129* (5), 929–941.
31. B. P. Klaholz; A. G. Myasnikov; M. van Heel, *Nature* **2004**, *427* (6977), 862–865.
32. A. P. Carter; W. M. Clemons, Jr.; D. E. Brodersen; R. J. Morgan-Warren; T. Hartsch; B. T. Wimberly; V. Ramakrishnan, *Science* **2001**, *291* (5503), 498–501.
33. M. Pioletti; F. Schlunzen; J. Harms; R. Zarivach; M. Gluhmann; H. Avila; A. Bashan; H. Bartels; T. Auerbach; C. Jacobi; T. Hartsch; A. Yonath; F. Franceschi, *EMBO J.* **2001**, *20* (8), 1829–1839.
34. S. Petry; D. E. Brodersen; F. V. Murphy IV; C. M. Dunham; M. Selmer; M. J. Tarry; A. C. Kelley; V. Ramakrishnan, *Cell* **2005**, *123* (7), 1255–1266.
35. A. Weixlbaumer; S. Petry; C. M. Dunham; M. Selmer; A. C. Kelley; V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14* (8), 733–737.
36. J. Parker, *Microbiol. Rev.* **1989**, *53* (3), 273–298.
37. H. J. Grosjean; S. de Henau; D. M. Crothers, *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75* (2), 610–614.
38. R. Kierzek; M. E. Burkard; D. H. Turner, *Biochemistry* **1999**, *38* (43), 14214–14223.
39. T. Xia; J. SantaLucia, Jr.; M. E. Burkard; R. Kierzek; S. J. Schroeder; X. Jiao; C. Cox; D. H. Turner, *Biochemistry* **1998**, *37* (42), 14719–14735.
40. H. Dong; L. Nilsson; C. G. Kurland, *J. Mol. Biol.* **1996**, *260* (5), 649–663.
41. R. C. Thompson; A. M. Karim, *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79* (16), 4922–4926.
42. J. M. Ogle; F. V. Murphy; M. J. Tarry; V. Ramakrishnan, *Cell* **2002**, *111* (5), 721–732.
43. B. T. Wimberly; D. E. Brodersen; W. M. Clemons, Jr.; R. J. Morgan-Warren; A. P. Carter; C. Vonrhein; T. Hartsch; V. Ramakrishnan, *Nature* **2000**, *407* (6802), 327–339.
44. T. Pape; W. Wintermeyer; M. V. Rodnina, *EMBO J.* **1998**, *17* (24), 7490–7497.
45. T. Pape; W. Wintermeyer; M. Rodnina, *EMBO J.* **1999**, *18* (13), 3800–3807.
46. K. B. Gromadski; M. V. Rodnina, *Mol. Cell* **2004**, *13* (2), 191–200.
47. S. C. Blanchard; R. L. Gonzalez; H. D. Kim; S. Chu; J. D. Puglisi, *Nat. Struct. Mol. Biol.* **2004**, *11* (10), 1008–1014.
48. U. Kothe; H. J. Wieden; D. Mohr; M. V. Rodnina, *J. Mol. Biol.* **2004**, *336* (5), 1011–1021.

49. M. V. Rodnina; T. Pape; R. Fricke; L. Kuhn; W. Wintermeyer, *J. Biol. Chem.* **1996**, *271* (2), 646–652.
50. U. Kothe; M. V. Rodnina, *Biochemistry* **2006**, *45* (42), 12767–12774.
51. L. Cochella; R. Green, *Science* **2005**, *308* (5725), 1178–1180.
52. U. Kothe; M. V. Rodnina, *Mol. Cell* **2007**, *25* (1), 167–174.
53. S. Ledoux; O. C. Uhlenbeck, *Mol. Cell* **2008**, *31* (1), 114–123.
54. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71* (10), 4135–4139.
55. J. Ninio, *Biochimie* **1975**, *57* (5), 587–595.
56. K. B. Gromadski; T. Daviter; M. V. Rodnina, *Mol. Cell* **2006**, *21* (3), 369–377.
57. T. Ruusala; M. Ehrenberg; C. G. Kurland, *EMBO J.* **1982**, *1* (6), 741–745.
58. K. B. Gromadski; M. V. Rodnina, *Nat. Struct. Mol. Biol.* **2004**, *11* (4), 316–322.
59. T. Pape; W. Wintermeyer; M. V. Rodnina, *Nat. Struct. Biol.* **2000**, *7* (2), 104–107.
60. M. Valle; A. Zavialov; J. Sengupta; U. Rawat; M. Ehrenberg; J. Frank, *Cell* **2003**, *114* (1), 123–134.
61. T. H. Lee; S. C. Blanchard; H. D. Kim; J. D. Puglisi; S. Chu, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (34), 13661–13665.
62. O. Piepenburg; T. Pape; J. A. Pleiss; W. Wintermeyer; O. C. Uhlenbeck; M. V. Rodnina, *Biochemistry* **2000**, *39* (7), 1734–1738.
63. L. Cochella; R. Green, *Science* **2005**, *308* (5725), 1178–1180.
64. P. Bieling; M. Beringer; S. Adio; M. V. Rodnina, *Nat. Struct. Mol. Biol.* **2006**, *13* (5), 423–428.
65. K. Y. Sanbonmatsu; S. Joseph; C. S. Tung, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (44), 15854–15859.
66. C. Lad; N. H. Williams; R. Wolfenden, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (10), 5607–5610.
67. J. Wower; S. S. Hixson; R. A. Zimmermann, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (14), 5232–5236.
68. P. Nissen; J. Hansen; N. Ban; P. B. Moore; T. A. Steitz, *Science* **2000**, *289* (5481), 920–930.
69. T. M. Schmeing; K. S. Huang; D. E. Kitchen; S. A. Strobel; T. A. Steitz, *Mol. Cell* **2005**, *20* (3), 437–448.
70. T. M. Schmeing; K. S. Huang; S. A. Strobel; T. A. Steitz, *Nature* **2005**, *438* (7067), 520–524.
71. V. I. Katunin; G. W. Muth; S. A. Strobel; W. Wintermeyer; M. V. Rodnina, *Mol. Cell* **2002**, *10* (2), 339–346.
72. E. M. Youngman; J. L. Brunelle; A. B. Kochaniak; R. Green, *Cell* **2004**, *117* (5), 589–599.
73. M. Johansson; E. Bouakaz; M. Lovmar; M. Ehrenberg, *Mol. Cell* **2008**, *30* (5), 589–598.
74. L. Lancaster; H. F. Noller, *Mol. Cell* **2005**, *20* (4), 623–632.
75. E. M. Youngman; R. Green, *Methods* **2005**, *36* (3), 305–312.
76. M. Beringer; S. Adio; W. Wintermeyer; M. Rodnina, *RNA* **2003**, *9* (8), 919–922.
77. A. E. Hesslein; V. I. Katunin; M. Beringer; A. B. Kosek; M. V. Rodnina; S. A. Strobel, *Nucleic Acids Res.* **2004**, *32* (12), 3760–3770.
78. S. Trobro; J. Aqvist, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (35), 12395–12400.
79. S. Trobro; J. Aqvist, *Biochemistry* **2006**, *45* (23), 7049–7056.
80. M. Beringer; C. Bruell; L. Xiong; P. Pfister; P. Bieling; V. I. Katunin; A. S. Mankin; E. C. Bottger; M. V. Rodnina, *J. Biol. Chem.* **2005**, *280* (43), 36065–36072.
81. A. Sievers; M. Beringer; M. V. Rodnina; R. Wolfenden, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (21), 7897–7901.
82. J. L. Brunelle; E. M. Youngman; D. Sharma; R. Green, *RNA* **2006**, *12* (1), 33–39.
83. J. S. Weinger; K. M. Parnell; S. Dorner; R. Green; S. A. Strobel, *Nat. Struct. Mol. Biol.* **2004**, *11* (11), 1101–1106.
84. K. Lang; M. Erlacher; D. N. Wilson; R. Micura; N. Polacek, *Chem. Biol.* **2008**, *15* (5), 485–492.
85. S. Pestka, *J. Biol. Chem.* **1968**, *243* (10), 2810–2820.
86. S. Pestka, *J. Biol. Chem.* **1969**, *244* (6), 1533–1539.
87. A. L. Konevega; N. Fischer; Y. P. Semenkov; H. Stark; W. Wintermeyer; M. V. Rodnina, *Nat. Struct. Mol. Biol.* **2007**, *14* (4), 318–324.
88. S. Shoji; S. E. Walker; K. Fredrick, *Mol. Cell* **2006**, *24* (6), 931–942.
89. R. N. Evans; G. Blaha; S. Bailey; T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (12), 4673–4678.
90. Y. Qin; N. Polacek; O. Vesper; E. Staub; E. Einfeldt; D. N. Wilson; K. H. Nierhaus, *Cell* **2006**, *127* (4), 721–733.
91. V. I. Katunin; A. Savelsbergh; M. V. Rodnina; W. Wintermeyer, *Biochemistry* **2002**, *41* (42), 12806–12812.
92. M. V. Rodnina; A. Savelsbergh; V. I. Katunin; W. Wintermeyer, *Nature* **1997**, *385* (6611), 37–41.
93. S. Takyar; R. P. Hickerson; H. F. Noller, *Cell* **2005**, *120* (1), 49–58.
94. J. D. Wen; L. Lancaster; C. Hodges; A. C. Zeri; S. H. Yoshimura; H. F. Noller; C. Bustamante; I. Tinoco, *Nature* **2008**, *452* (7187), 598–603.
95. D. Moazed; H. F. Noller, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88* (9), 3725–3728.
96. D. Sharma; D. R. Southworth; R. Green, *RNA* **2004**, *10* (1), 102–113.
97. S. C. Blanchard; H. D. Kim; R. L. Gonzalez, Jr.; J. D. Puglisi; S. Chu, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (35), 12893–12898.
98. M. Valle; A. Zavialov; J. Sengupta; U. Rawat; M. Ehrenberg; J. Frank, *Cell* **2003**, *114* (1), 123–134.
99. Y. P. Semenkov; M. V. Rodnina; W. Wintermeyer, *Nat. Struct. Biol.* **2000**, *7* (11), 1027–1031.
100. S. Dorner; J. L. Brunelle; D. Sharma; R. Green, *Nat. Struct. Mol. Biol.* **2006**, *13* (3), 234–241.
101. P. C. Spiegel; D. N. Ermolenko; H. F. Noller, *RNA* **2007**, *13* (9), 1473–1482.
102. J. B. Munro; R. B. Altman; N. O'Connor; S. C. Blanchard, *Mol. Cell* **2007**, *25* (4), 505–517.
103. D. Pan; S. Kirillov; C. M. Zhang; Y. M. Hou; B. S. Cooperman, *Nat. Struct. Mol. Biol.* **2006**, *13* (4), 354–359.
104. S. E. Walker; S. Shoji; D. Pan; B. S. Cooperman; K. Fredrick, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (27), 9192–9197.
105. J. Frank; R. K. Agrawal, *Nature* **2000**, *406* (6793), 318–322.
106. J. Fei; P. Kosuri; D. D. MacDougall; R. L. Gonzalez, Jr., *Mol. Cell* **2008**, *30* (3), 348–359.
107. C. M. Spahn; M. G. Gomez-Lorenzo; R. A. Grassucci; R. Jorgensen; G. R. Andersen; R. Beckmann; P. A. Penczek; J. P. Ballesta; J. Frank, *EMBO J.* **2004**, *23* (5), 1008–1019.
108. D. J. Taylor; J. Nilsson; A. R. Merrill; G. R. Andersen; P. Nissen; J. Frank, *EMBO J.* **2007**, *26* (9), 2421–2431.
109. P. P. Datta; M. R. Sharma; L. Qi; J. Frank; R. K. Agrawal, *Mol. Cell* **2005**, *20* (5), 723–731.
110. Y. Wang; H. Qin; R. D. Kudaravalli; S. V. Kirillov; G. T. Dempsey; D. Pan; B. S. Cooperman; Y. E. Goldman, *Biochemistry* **2007**, *46* (38), 10767–10775.
111. L. H. Horan; H. F. Noller, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (12), 4881–4885.
112. D. N. Ermolenko; Z. K. Majumdar; R. P. Hickerson; P. C. Spiegel; R. M. Clegg; H. F. Noller, *J. Mol. Biol.* **2007**, *370* (3), 530–540.

113. D. N. Ermolenko; P. C. Spiegel; Z. K. Majumdar; R. P. Hickerson; R. M. Clegg; H. F. Noller, *Nat. Struct. Mol. Biol.* **2007**, *14* (6), 493–497.
114. P. V. Cornish; D. N. Ermolenko; H. F. Noller; T. Ha, *Mol. Cell* **2008**, *30* (5), 578–588.
115. B. Wilden; A. Savelsbergh; M. V. Rodnina; W. Wintermeyer, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (37), 13670–13675.
116. S. R. Connell; C. Takemoto; D. N. Wilson; H. Wang; K. Murayama; T. Terada; M. Shirouzu; M. Rost; M. Schuler; J. Giesebrecht; M. Dabrowski; T. Mielke; P. Fucini; S. Yokoyama; C. M. Spahn, *Mol. Cell* **2007**, *25* (5), 751–764.
117. D. Mohr; W. Wintermeyer; M. V. Rodnina, *Biochemistry* **2002**, *41* (41), 12520–12528.
118. A. Savelsbergh; D. Mohr; B. Wilden; W. Wintermeyer; M. V. Rodnina, *J. Biol. Chem.* **2000**, *275* (2), 890–894.
119. F. Peske; N. B. Matassova; A. Savelsbergh; M. V. Rodnina; W. Wintermeyer, *Mol. Cell* **2000**, *6* (2), 501–505.
120. A. Savelsbergh; N. B. Matassova; M. V. Rodnina; W. Wintermeyer, *J. Mol. Biol.* **2000**, *300* (4), 951–961.
121. A. Savelsbergh; V. I. Katunin; D. Mohr; F. Peske; M. V. Rodnina; W. Wintermeyer, *Mol. Cell* **2003**, *11* (6), 1517–1523.
122. A. Savelsbergh; D. Mohr; U. Kothe; W. Wintermeyer; M. V. Rodnina, *EMBO J.* **2005**, *24* (24), 4316–4323.
123. B. P. Klaholz; T. Pape; A. V. Zavialov; A. G. Myasnikov; E. V. Orlova; B. Vestergaard; M. Ehrenberg; M. van Heel, *Nature* **2003**, *421* (6918), 90–94.
124. U. B. Rawat; A. V. Zavialov; J. Sengupta; M. Valle; R. A. Grassucci; J. Linde; B. Vestergaard; M. Ehrenberg; J. Frank, *Nature* **2003**, *421* (6918), 87–90.
125. B. Vestergaard; S. Sanyal; M. Roessle; L. Mora; R. H. Buckingham; J. S. Kastrup; M. Gajhede; D. I. Svergun; M. Ehrenberg, *Mol. Cell* **2005**, *20* (6), 929–938.
126. B. Vestergaard; L. B. Van; G. R. Andersen; J. Nyborg; R. H. Buckingham; M. Kjeldgaard, *Mol. Cell* **2001**, *8* (6), 1375–1382.
127. D. V. Freistroffer; M. Kwiatkowski; R. H. Buckingham; M. Ehrenberg, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (5), 2046–2051.
128. K. Ito; M. Uno; Y. Nakamura, *Nature* **2000**, *403* (6770), 680–684.
129. E. M. Youngman; S. L. He; L. J. Nikstad; R. Green, *Mol. Cell* **2007**, *28* (4), 533–543.
130. I. K. Ali; L. Lancaster; J. Feinberg; S. Joseph; H. F. Noller, *Mol. Cell* **2006**, *23* (6), 865–874.
131. M. Amort; B. Wotzel; K. Bakowska-Zywicka; M. D. Erlacher; R. Micura; N. Polacek, *Nucleic Acids Res.* **2007**, *35* (15), 5130–5140.
132. N. Polacek; M. J. Gomez; K. Ito; L. Xiong; Y. Nakamura; A. Mankin, *Mol. Cell* **2003**, *11* (1), 103–112.
133. J. S. Feinberg; S. Joseph, *J. Mol. Biol.* **2006**, *364* (5), 1010–1020.
134. J. L. Brunelle; J. J. Shaw; E. M. Youngman; R. Green, *RNA* **2008**, *14* (8), 1526–1531.
135. L. Y. Frolova; R. Y. Tsivkovskii; G. F. Sivolobova; N. Y. Oparina; O. I. Serpinsky; V. M. Blinov; S. I. Tatkov; L. L. Kisselev, *RNA* **1999**, *5* (8), 1014–1020.
136. A. Seit-Nebi; L. Frolova; J. Justesen; L. Kisselev, *Nucleic Acids Res.* **2001**, *29* (19), 3982–3987.
137. J. J. Shaw; R. Green, *Mol. Cell* **2007**, *28* (3), 458–467.
138. A. V. Zavialov; L. Mora; R. H. Buckingham; M. Ehrenberg, *Mol. Cell* **2002**, *10* (4), 789–798.
139. S. Trobro; J. Aqvist, *Mol. Cell* **2007**, *27* (5), 758–766.
140. A. V. Zavialov; R. H. Buckingham; M. Ehrenberg, *Cell* **2001**, *107* (1), 115–124.
141. E. Z. Alkalaeva; A. V. Pisarev; L. Y. Frolova; L. L. Kisselev; T. V. Pestova, *Cell* **2006**, *125* (6), 1125–1136.
142. A. Ohta; Y. Yamagishi; H. Suga, *Curr. Opin. Chem. Biol.* **2008**, *12* (2), 159–167.
143. L. Wang; J. Xie; P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
144. T. L. Hendrickson; V. de Crecy-Lagard; P. Schimmel, *Annu. Rev. Biochem.* **2004**, *73*, 147–176.
145. C. J. Noren; S. J. Anthony-Cahill; M. C. Griffith; P. G. Schultz, *Science* **1989**, *244* (4901), 182–188.
146. G. Srinivasan; C. M. James; J. A. Krzycki, *Science* **2002**, *296* (5572), 1459–1462.
147. S. Commans; A. Bock, *FEMS Microbiol. Rev.* **1999**, *23* (3), 335–351.
148. Y. Shimizu; A. Inoue; Y. Tomari; T. Suzuki; T. Yokogawa; K. Nishikawa; T. Ueda, *Nat. Biotechnol.* **2001**, *19* (8), 751–755.
149. Y. Shimizu; T. Kanamori; T. Ueda, *Methods* **2005**, *36* (3), 299–304.
150. A. C. Forster; Z. Tan; M. N. Nalam; H. Lin; H. Qu; V. W. Cornish; S. C. Blacklow, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (11), 6353–6357.
151. Z. Tan; S. C. Blacklow; V. W. Cornish; A. C. Forster, *Methods* **2005**, *36* (3), 279–290.
152. T. Hohsaka; Y. Ashizuka; H. Murakami; M. Sisido, *Nucleic Acids Res.* **2001**, *29* (17), 3646–3651.
153. T. Hohsaka; Y. Ashizuka; H. Taira; H. Murakami; M. Sisido, *Biochemistry* **2001**, *40* (37), 11060–11064.
154. M. Taki; T. Hohsaka; H. Murakami; K. Taira; M. Sisido, *J. Am. Chem. Soc.* **2002**, *124* (49), 14586–14590.
155. C. M. Cummins; T. F. Donahue; M. R. Culbertson, *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79* (11), 3565–3569.
156. C. M. Dunham; M. Selmer; S. S. Phelps; A. C. Kelley; T. Suzuki; S. Joseph; V. Ramakrishnan, *RNA* **2007**, *13* (6), 817–823.
157. J. D. Bain; C. Switzer; A. R. Chamberlin; S. A. Benner, *Nature* **1992**, *356* (6369), 537–539.
158. J. A. Piccirilli; T. Krauch; S. E. Moroney; S. A. Benner, *Nature* **1990**, *343* (6253), 33–37.
159. T. Ohtsuki; M. Kimoto; M. Ishikawa; T. Mitsui; I. Hirao; S. Yokoyama, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (9), 4922–4925.
160. I. Hirao; T. Ohtsuki; T. Fujiwara; T. Mitsui; T. Yokogawa; T. Okuni; H. Nakayama; K. Takio; T. Yabuki; T. Kigawa; K. Kodama; T. Yokogawa; K. Nishikawa; S. Yokoyama, *Nat. Biotechnol.* **2002**, *20* (2), 177–182.
161. S. M. Hecht; B. L. Alford; Y. Kuroda; S. Kitano, *J. Biol. Chem.* **1978**, *253* (13), 4517–4520.
162. S. A. Robertson; C. J. Noren; S. J. Anthony-Cahill; M. C. Griffith; P. G. Schultz, *Nucleic Acids Res.* **1989**, *17* (23), 9649–9660.
163. D. R. Liu; T. J. Magliery; M. Pastrnak; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (19), 10092–10097.
164. L. Wang; A. Brock; B. Herberich; P. G. Schultz, *Science* **2001**, *292* (5516), 498–500.
165. M. C. Hartman; K. Josephson; J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (12), 4356–4361.
166. M. C. Hartman; K. Josephson; C. W. Lin; J. W. Szostak, *PLoS ONE* **2007**, *2* (10), e972.
167. H. Murakami; A. Ohta; H. Ashigai; H. Suga, *Nat. Methods* **2006**, *3* (5), 357–359.
168. H. Xiao; H. Murakami; H. Suga; A. R. Ferre-D'Amare, *Nature* **2008**, *454* (7202), 358–361.
169. Y. Goto; H. Murakami; H. Suga, *RNA* **2008**, *14* (7), 1390–1398.
170. Y. Goto; A. Ohta; Y. Sako; Y. Yamagishi; H. Murakami; H. Suga, *ACS Chem. Biol.* **2008**, *3* (2), 120–129.
171. L. M. Dedkova; N. E. Fahmi; S. Y. Golovine; S. M. Hecht, *Biochemistry* **2006**, *45* (51), 15541–15551.
172. J. A. Ellman; D. Mendel; P. G. Schultz, *Science* **1992**, *255* (5041), 197–200.

173. A. Ohta; H. Murakami; E. Higashimura; H. Suga, *Chem. Biol.* **2007**, *14* (12), 1315–1322.
174. F. J. LaRiviere; A. D. Wolfson; O. C. Uhlenbeck, *Science* **2001**, *294* (5540), 165–168.
175. Y. Doi; T. Ohtsuki; Y. Shimizu; T. Ueda; M. Sisido, *J. Am. Chem. Soc.* **2007**, *129* (46), 14458–14462.
176. Y. Sako; Y. Goto; H. Murakami; H. Suga, *ACS Chem. Biol.* **2008**, *3* (4), 241–249.

Biographical Sketch

Ute Kothe studied biochemistry in Regensburg, Bochum, and Witten in Germany and completed her studies with a diploma in 2002. For her Ph.D. Ute Kothe worked under the supervision of Marina Rodnina at the Institute of Physical Biochemistry, University of Witten/Herdecke, Germany, where she studied the accuracy of ribosomal protein synthesis. Since 2006, Ute Kothe is an assistant professor of biochemistry at the University of Lethbridge, Alberta, Canada.

5.14 Glutaminyl-tRNA and Asparaginyl-tRNA Biosynthetic Pathways

Jonathan L. Huot, Jacques Lapointe, and Robert Chênevert, Université Laval, Québec, QC, Canada

Marc Bailly and Daniel Kern, Institut de Biologie Moléculaire et Cellulaire, Strasbourg, France

© 2010 Elsevier Ltd. All rights reserved.

5.14.1 Introduction

5.14.1.1 Generalities on tRNAs and Aminoacyl-tRNAs

Transfer ribonucleic acids (tRNAs) are a family of small stable RNAs generally containing 76 nucleotides[1] that play a central role in the translation of genetic information into amino acid sequence information (protein biosynthesis) in all living organisms.[2] The main relatively stable forms of activated amino acids in living cells are the aminoacyl-tRNAs (aa-tRNAs), where the α-COOH group of an amino acid is esterified to the 2′ or 3′ OH group of the ribose of an adenosine residue universally conserved at the 3′ end of tRNAs. Aminoacyl-tRNAs are formed by the reaction of tRNAs with a more reactive and unstable form of activated amino acid, the aminoacyl-adenylate (aa∼AMP), where the α-COOH group is linked to the phosphate group of AMP by an acid anhydride bond.[3] The three-letter abbreviations generally used for the free amino acids and their residues that are the focus of this chapter are Gln for glutamine, Asn for asparagine, Glu for glutamic acid, and Asp for aspartic acid. Glx denotes either Glu or Gln, and Asx either Asp or Asn. The upper index at the right of tRNA indicates its specificity; for instance, Asp-tRNA^{Asn} is a tRNA specific for Asn that has been esterified at its 3′ end with an Asp residue, as a result of a misaminoacylation reaction.

In extant organisms, aa-tRNAs are mostly used in the synthesis of polymers of amino acid residues (proteins) that takes place on ribosomes. In certain microorganisms, some aa-tRNAs used in protein biosynthesis are formed by the modification of the amino acid side chain of a different aa-tRNA, such as Gln-tRNA^{Gln} from Glu-tRNA^{Gln}, Asn-tRNA^{Asn} from Asp-tRNA^{Asn} (see **Figure 2**), and Cys-tRNA and selenocysteinyl-tRNA from *O*-phosphoseryl-tRNA formed by either direct aminoacylation or phosphorylation of Ser-tRNA.[4,5] The precursor–product relationship (in a metabolic sense) between the aminoacyl groups involved in these transformations supports the theory of coevolution of the genetic code and the amino acid biosynthetic pathways, which postulates that prebiotic synthesis was an inadequate source of all 20-protein amino acids, and therefore some of them had to be derived from the coevolving pathways of amino acid biosynthesis.[6–8] This theory is consistent with the synthesis of aspartic acid and glutamic acid but not of their amide derivatives Asn and Gln, following electric discharges or other energetic perturbations of gas mixtures proposed to mimic the atmosphere of the primitive earth.[9] The synthesis of some amino acids from precursors linked to a rigid holder, such as tRNA, is in line with models of prebiotic metabolism taking place at the surface of solid particles instead of within a prebiotic soup;[10,11] indeed, the size of tRNA (M_r about 25 000 Da) and its compact structure discovered at high resolution first for yeast tRNA specific for phenylalanine (tRNA^{Phe})[12,13] suggest that its ancestors could have played the role of rigid holders.

Aminoacyl-tRNAs are also precursors in the biosynthesis of heme and chlorophyll through the reduction of Glu-tRNA^{Glu} into glutamate-1-semialdehyde catalyzed by the Glu-tRNA reductase,[14] of the peptidoglycan of bacterial cell walls,[15,16] or of a lipid, lysylphosphatidylglycerol.[17] Some aa-tRNAs also have regulatory functions, for example, Asn-tRNA, which is involved in the transcriptional control of the *Lactobacillus bulgaricus* operon which encodes asparagine synthetase A and asparaginyl-tRNA synthetase (AsnRS).[18] Some aa∼AMPs

transfer their activated amino acids for other functions: for instance, the YadB protein of *Escherichia coli*, a paralogue of the glutamyl-tRNA synthetase (GluRS), is a tRNA modification enzyme that glutamylates the $tRNA^{Asp}$ anticodon and is inhibited by the Glu∼AMP analogue Glu-ol-AMP.[19]

5.14.1.2 aa-tRNA Formation by the Direct Pathway

For most amino acids, the ester linkage between the α-COOH group of the amino acid and the 3′-terminal adenosine of a cognate tRNA is formed in a two-step mechanism catalyzed by an aminoacyl-tRNA synthetase (aaRS).[20] In this so-called direct pathway, the aaRS first catalyzes the reaction of the amino acid with adenosine triphosphate (ATP), yielding the enzyme-bound high-energy intermediate aa∼AMP and PP_i; in the second step, this aaRS-bound intermediate reacts with tRNA to yield aa-tRNA and AMP (**Figure 1**).

5.14.1.3 aa-tRNA Formation by the Indirect Pathway

For glutamine and/or asparagine, the formation of the correctly charged cognate tRNA in some organisms takes place through the following indirect pathway catalyzed by two enzymes (**Figure 2**).[21] In the first reaction, an amino acid precursor (glutamate and aspartate, respectively) to the one corresponding to the tRNA substrate is used by an ND-aaRS to misacylate that tRNA. In the second reaction, the misacylated Glu-$tRNA^{Gln}$ or Asp-$tRNA^{Asn}$ is corrected by amidation of the side chains of these aa-tRNAs by an aminoacyl-tRNA amidotransferase (AdT), yielding the correctly charged Gln-$tRNA^{Gln}$ or Asn-$tRNA^{Asn}$ following the reaction mechanism described in **Figure 3**.

5.14.1.4 Some Differences Between the Metabolic and Structural Functions of Glutamine and Asparagine

Glutamine and asparagine are the only amino acids containing an amide group in their side chain, and they differ only by one methylene group. In spite of this structural similarity, their metabolic functions and structural roles in proteins are very different, probably because of the extra length and additional degree of freedom of the glutamine side chain.

Glutamine provides the major entry point for assimilation of reduced nitrogen in the form of NH_4^+, in particular for bacterial cells grown in ammonia-limited media.[22] Its amide group is a source of amino groups in a wide range of biosynthetic processes[23] including the amidation of aspartate to form asparagine. Its concentration or that of its precursor glutamate is much higher than that of the other amino acids in most types of cells. Consistent with these important metabolic functions, the regulation of glutamine synthetase biosynthesis[22] and activity[23] is very complex. This enzyme catalyzes glutamine biosynthesis in two steps: in the first step, it activates the γ-COOH group of glutamate using the energy of ATP hydrolysis to adenosine diphosphate (ADP) + P_i to form γ-glutamyl phosphate, and in the second step, it catalyzes the reaction of this intermediate with NH_4^+ to yield glutamine and P_i. In various transcription factors, glutaminyl residues in the recognition α-helices of helix–turn–helix motifs make bidentate H-bonds between their side chain amide group and the O6 and N7 positions of adenine involved in an A-T base pair in the major groove of double-stranded DNA.[24,25]

Asparagine is synthesized through the ammonia-dependent asparagine synthetase AS-A and/or through the glutamine-dependent asparagine synthetase AS-B, encoded respectively by the *asnA* and *asnB* genes. AS-A is homologous to aspartyl-tRNA synthetase.[26,27] In contrast to glutamine, asparagine is not involved in ammonia assimilation because it is not an amino group donor in any biosynthetic reaction. Little is known about the regulation of the expression of asparagine synthase genes, except that the transcription of *asnA* is activated by the protein AsnC[28] and is controlled by an antitermination mechanism.[18] In proteins, asparaginyl residues are frequently present at the N-terminal end of several α-helices, where they stabilize the helix through a hydrogen bond between the C=O group of their side chain and an unsatisfied NH of the main chain of the α-helix. Glutamine cannot play this role because its additional methylene group does not allow the positioning of its side chain C=O to make this H-bond.[29,30]

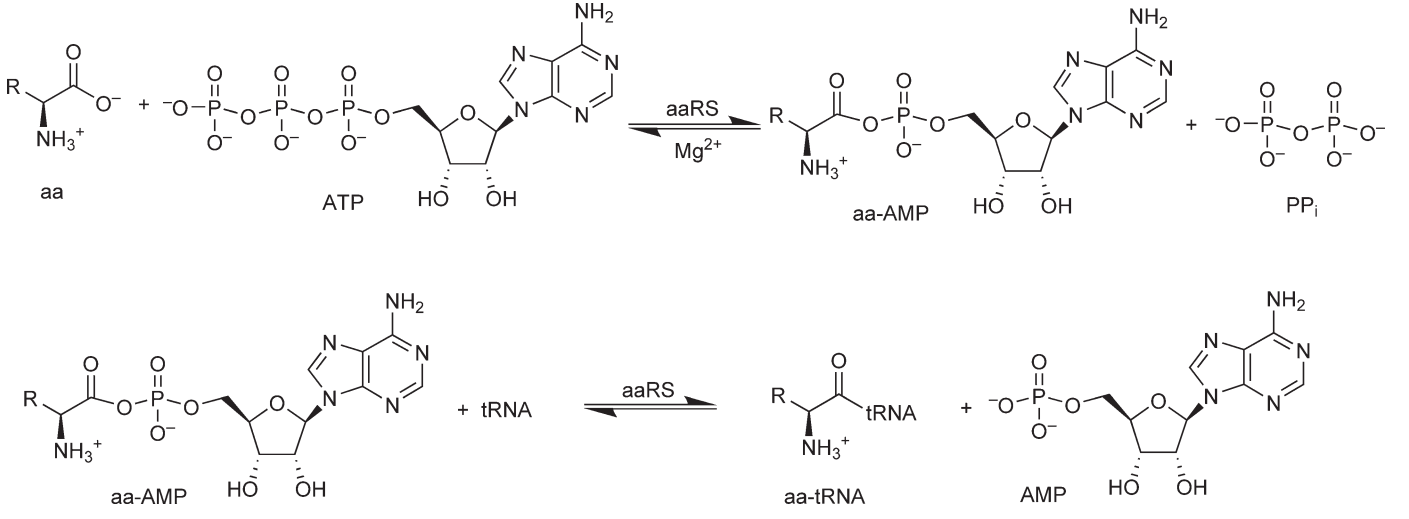


Figure 1 Direct pathway of tRNA aminoacylation. aaRS, aminoacyl-tRNA synthetase.

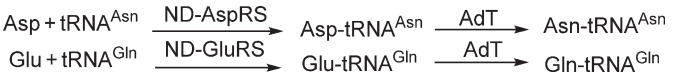


Figure 2 Indirect pathway for Asn-tRNA^{Asn} and Gln-tRNA^{Gln} biosynthesis. ND-AspRS, nondiscriminating aspartyl-tRNA synthetase; ND-GluRS, nondiscriminating glutamyl-tRNA systhetase; AdT, aminoacyl-tRNA amidotransferase.

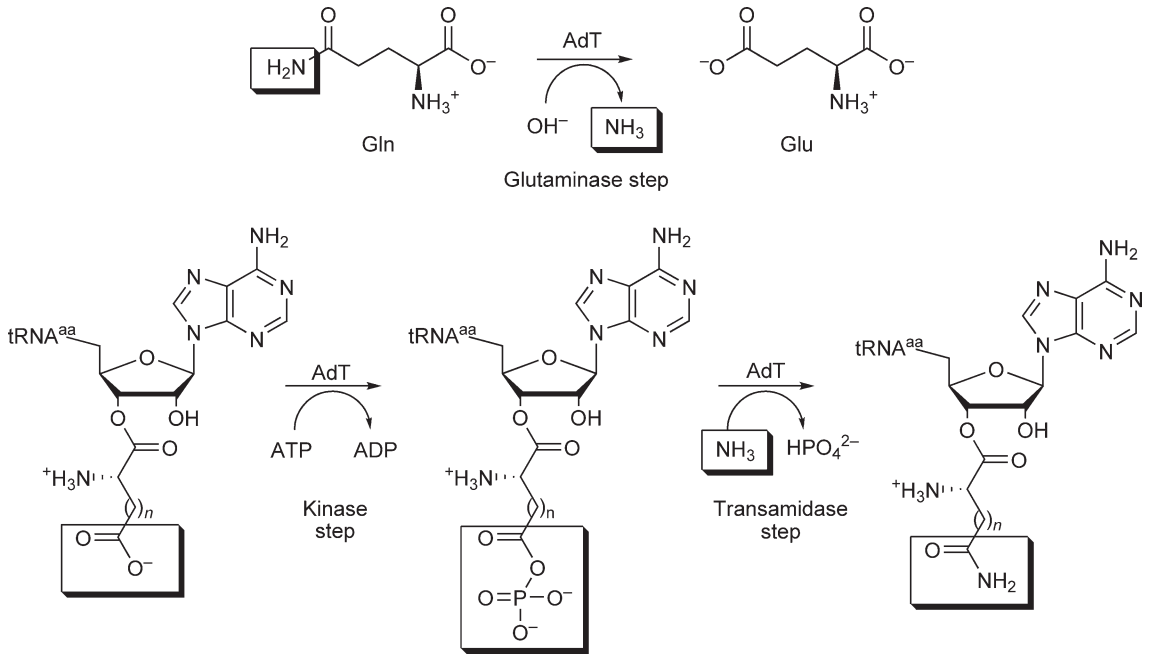


Figure 3 Reaction mechanism of aminoacyl-tRNA amidotransferase. For Asp-tRNA^{Asn} amidotransferase (AspAdT), $n = 1$, tRNA^{aa} = tRNA^{Asn}; for Glu-tRNA^{Gln} amidotransferase (GluAdT), $n = 2$, tRNA^{aa} = tRNA^{Gln}.

5.14.2 Nonenzymatic Deamidation of Glutaminyl and Asparaginyl Residues in Protein

One of the most common processes of chemical aging of proteins is the deamidation of asparagine residues, particularly those preceding glycine residues in the polypeptide sequence.[31] At alkaline pH, the Asn residue is much more labile than the Gln residue, and the reverse is true at acidic pH, but these reactions may also occur at neutral pH.[32] These deamidations introduce negative charges carried by the resulting aspartate or glutamate residues and lead to protein inactivation (through denaturation or conformational change) when such residues are located in their hydrophobic core.[33] Robinson[34] proposed that such deamidation reactions in peptides and proteins may serve as molecular clocks that control biological processes, and that these clocks may be set by rejection or accumulation during evolution of appropriate sequences of residues, including a glutaminyl or an asparaginyl residue. The role of deamidation as a molecular timer of *in vivo* protein turnover has been demonstrated using rabbit muscle aldolase.[35]

The rate of deamidation of a given residue depends upon its position in the primary structure (sequence of residues) of the protein, the three-dimensional (3D) structure of the protein, and the solution properties such as pH, temperature, ionic strength, and buffer ions. Using the deamidation rates at neutral pH of Asn residues in most of the 400 possible near-neighbor combinations in pentapeptide models, and of a representative group of Gln pentapeptides, combined with the 3D structures of proteins with well-characterized deamidation rates, a computerized technique for the quantitative estimation of the deamidation rates of any protein whose 3D structure is known has been developed and was shown to have good quantitative agreement with experimental values.[32,36,37] This approach showed that Asn deamidation is a biologically relevant phenomenon in many human proteins.[38]

Analyses of soluble proteins from complete genomes of 20 thermophilic or mesophilic microorganisms revealed that higher amounts of Glu and Asp residues and lower amounts of Gln and Asn residues are present in proteins of thermophiles compared to proteins of mesophiles.[39–41] This result suggests that Gln and Asn residues destabilize protein structure at high temperature, possibly through their deamidation.

5.14.3 Introduction of Glutamine to the Genetic Code

The molecular basis for the incorporation of Gln into proteins during biosynthesis has been revealed to be both surprisingly diverse and inextricably related to the incorporation of the related Glu. Early evidence suggested that only Gln could be added to the nascent polypeptide and that posttranslational hydrolysis of the Gln side chain amide group was responsible for the presence of Glu in protein.[42] However, further studies soon established the existence of two distinct enzymes, glutaminyl-tRNA synthetase (GlnRS) and GluRS, responsible for the formation of Gln-tRNA and Glu-tRNA, respectively.[43,44] Sequence homology studies suggest that GlnRS evolved from a duplicated *GlxRS* gene in early eukaryotes, which gradually acquired increased specificity for Gln and $tRNA^{Gln}$.[45,46] Prokaryotic organisms would have then acquired this early prototype of GlnRS through horizontal gene transfer from eukarya and retained or rejected it according to the suitability of their tRNA pool.[47] Contemporary GlnRS and GluRS are modular enzymes; composed of a core domain with major contacts to the acceptor stem of tRNA, ATP, and the cognate amino acid (aa), and of appended domains, which are mostly responsible for anticodon stem–loop recognition. The duplication event resulting in the appearance of a glutamine-specific aaRS is likely to have occurred before the acquisition of these added domains, which differ substantially between GluRS and GlnRS.[48]

5.14.3.1 Glutaminyl-tRNA Synthetase: Direct Formation of Gln-$tRNA^{Gln}$

GlnRS is a class I aaRS, with the corresponding Rossman fold dinucleotide-binding site, containing the conserved HIGH and KMSKS motifs involved in binding ATP and the aa substrate.[49–51] This class of aaRS catalyzes the esterification of an amino acid to the 2′ OH of the terminal adenosine of a cognate tRNA in a two-step process: activation of the aa by the formation of an aa∼AMP intermediate, and subsequent transfer of the aa to the terminal adenosine. GlnRS is one of four aaRSs for which the binding of the cognate tRNA is necessary for activating the aa.[52] The protein is organized into three main regions: with a core domain that is organized into five parallel β-strands, alternating with α-helices, with a sixth β-strand lying in a helical subdomain that is in contact with both the dinucleotide fold, and with the anticodon-binding domain.[53] A 110-aa domain, called the acceptor stem-binding domain, which is responsible for the binding and proper orientation of the acceptor end of tRNA, composed of both β-strands and α-helices, is inserted between the third β-strand and the third α-helix of the dinucleotide-binding domain. One of these α-helices, helix H, contains residues important for $tRNA^{Gln}$ specificity and also forms part of the dinucleotide-binding site of the Rossman fold. The anticodon-binding domain, composed of two β-barrels, interacts with the active site in at least two ways: both barrels are bridged by the second helix of a conserved α-helix-turn–β-strand–α-helix motif in the helical subdomain, which itself has significant contacts with the Rossman fold, and the more C-terminal barrel, which is closer to the core in the tertiary structure, interacts with the Rossman fold through an antiparallel β-ribbon reaching the helical subdomain near the acceptor stem-binding domain.[53] It has been suggested that this long loop plays a part in signal transmission between the anticodon-binding domain and the acceptor stem-binding domain, a mechanism that also appears to rely on the covalent continuity of $tRNA^{Gln}$.[54] This is partly due to the importance of a deformation of the tRNA, resulting in a shift of the acceptor stem and the anticodon loop relative to their usual conformation in most free tRNAs, caused by interactions between the helical subdomain and the phosphate backbone of the inner side of the L-shaped $tRNA^{Gln}$.[55,56] Further biochemical proof of signal transmission between the anticodon-binding domain and acceptor stem-binding domain was obtained with a revertant of a mutant GlnRS showing relaxed specificity for the third base of the anticodon: a secondary mutation in a site distant from the anticodon-binding domain allowed the enzyme to regain specificity for G36.[57] When the potential differences between signal transmission pathways and

specificity elements for efficient glutaminylation and that for $tRNA^{Gln}$ specificity were highlighted, correct G36 binding was found to be only minimally important for the cross-domain activation of glutaminylation, whereas the mutation of U35 diminished this activity considerably.[58]

5.14.3.1.1 Eukaryote-specific domains of GlnRS

GlnRS is composed of three domains in most organisms, or five, counting the helical subdomain and considering the two β-barrels of the anticodon-binding domain separately. Additional domains have been found in several cases. The GlnRS from *Deinococcus radiodurans*, a bacterium containing the GatCAB amidotransferase, bears a C-terminal extension to its primary sequence encoding a polypeptide from the putatively RNA-binding Yqey family. A mutant GlnRS from which this domain has been truncated shows a diminished affinity for $tRNA^{Gln}$, but any functions other than tRNA binding remain unknown.[59] It is also interesting to note that the Yqey domain is also found to be inserted in the GatB and GatE subunits of the GatCAB and GatDE amidotransferases, respectively.[60,61] In yeast, GlnRS possesses an N-terminal extension of over 200 aa, the removal or inactivation of which has little effect on homologous Gln-$tRNA^{Gln}$ formation.[62,63] *E. coli* GlnRS is only able to glutaminylate yeast $tRNA^{Gln}$ poorly *in vitro* or *in vivo*, but the fusion of the yeast N-terminal domain allowed an increase in specificity in both cases. Interestingly, the fusion to *E. coli* GlnRS of the unrelated Arclp protein, also from yeast and capable of nonspecific interactions with RNA, also resulted in an increase in specificity.[64] Mammalian GlnRS contains a considerable C-terminal extension, which is a docking platform for other components of the mammalian multienzyme synthetase complex.[65,66] Part of the dinucleotide-binding domain was later also found to be important for the formation of this complex,[67] as well as being capable of interacting and inhibiting the apoptosis signal-regulating kinase 1 (ASK1).[68]

5.14.3.1.2 Binding of the acceptor stem of $tRNA^{Gln}$ by GlnRS

Binding of the acceptor stem-binding domain to $tRNA^{Gln}$ induces the disruption of the U1–A72 base pair with Leu136 stacking between A72 and G2–C71, whereas Asp235, found in the same domain, forms hydrogen bonds with a water molecule held in the minor groove between base pairs G2–C71 and G3–C70. A water molecule also plays a role in allowing hydrogen bond formation between G10 and Glu323.[53] Sequence-specific interactions by Asp235 and other residues such as Arg130, Glu131, and Arg133 in the acceptor stem-binding domain with base pairs G2–C71 and G3–C70, and other sites in the acceptor stem, also define these base pairs as identity elements.[69–71] Terminal A76 binds with the ATP- and Gln-binding Rossman fold, whereas the first base of the CCA terminus of the tRNA, C74, is flipped out to interact with a binding pocket in the acceptor stem-binding domain. This allows a stacking of A76, C75, and G73, as well as formation of a hydrogen bond between A72 and discriminator base G73, which stabilizes the hairpin conformation adopted by the acceptor stem of $tRNA^{Gln}$ upon binding GlnRS (**Figure 4**).[53]

5.14.3.1.3 Binding of the anticodon loop of $tRNA^{Gln}$ by GlnRS

Many specificity elements for $tRNA^{Gln}$ are also the result of interactions with the three bases of the anticodon loop: C34 (and U34), U35, and G36 are each bound within separate pockets of the anticodon-binding domain.[72] Although it is likely that in free $tRNA^{Gln}$, like most other free tRNAs, anticodon bases are normally stacked with one another, binding by GlnRS disrupts this base stacking, allowing each base to be recognized by a distinct loop. Even though GlnRS must bind one of two $tRNA^{Gln}$ isoacceptors, $tRNA_1^{Gln}$ (UUG) and $tRNA_2^{Gln}$ (CUG), C34 is bound in its pocket by the concerted action of specific interactions with Arg412 and Trp458, as well as with the backbone of residues 411–414.[72] The U34 base of $tRNA_1^{Gln}$, which is sometimes modified into 5-methylaminomethyl-2-thiouridine (mnm^5s^2U), appears to be bound by a different mechanism dependent on the base modification.[73–75] U35 forms hydrogen bonds with Arg520, Gln517, and Arg341, which itself is positioned through interactions with the backbone of a nearby α-helix. The side chain of E519, which forms a bidentate hydrogen bond with the loop binding C34, also interacts with the U35 ring. G36 is held by its 5′ phosphate by Lys401, which also holds its ring through hydrophobic interactions. The G36 ring is also bound by Arg402 as well as by interactions with the polypeptide backbone (**Figure 4**).[72] The A-form helical conformation of the anticodon stem is maintained in spite of this sequestering of the anticodon bases through the formation of non-Watson–Crick base pairs, each comprising at least one modified nucleotide: 2′ *O*-methyl-

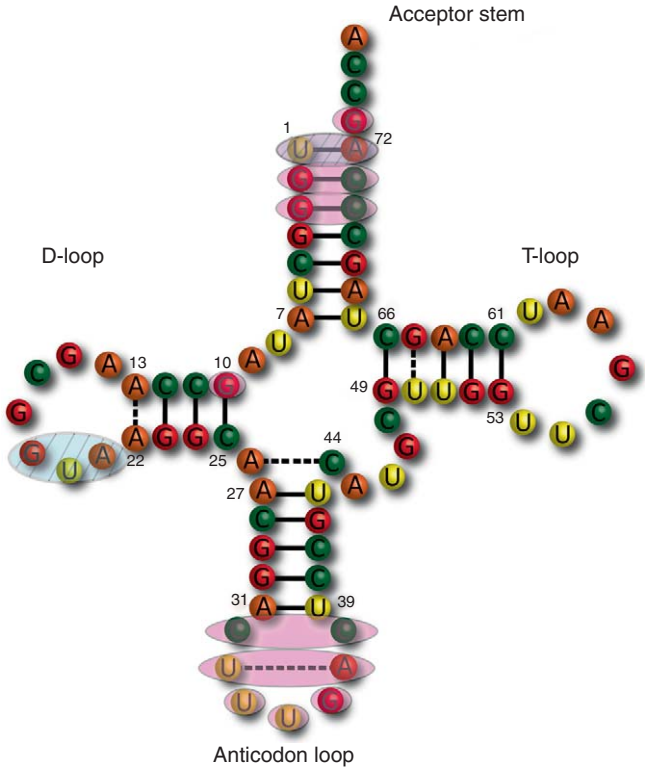


Figure 4 Recognition elements of $tRNA^{Gln}$. Highlighted in blue with bars are the recognition elements for Gln-$tRNA^{Gln}$ synthesis by GatCAB and GatDE. The U1–A72 base pair is a common identity element, while antideterminants for noncognate tRNAs are scanned for in the D-loop. Highlighted in magenta are the identity elements of $tRNA^{Gln}$ for GlnRS. $tRNA^{Gln(UUG)}$ from *Bacillus subtilis* is shown, but base modifications at positions 32–38 and 33–37 allow base pairing in *E. coli* $tRNA^{Gln(CUG)}$.

U32-pseudouridine 38 (Ψ) and U33-2-methyl-A37. Both base pairs are further strengthened by the formation of additional hydrogen bonds through a network of bound water molecules, whereas 2-methyl-A47 and Ψ38 form hydrogen bonds with Asn370. Modified bases can also serve as antideterminants against noncognate tRNAs bearing different modifications, often simply through steric hindrance.[72] The importance of antideterminants was highlighted in experiments where $tRNA^{Glu}$ was modified to hold all known identity elements of the acceptor stem and anticodon loop,[76] as well as the tertiary core of $tRNA^{Gln}$. These modified tRNAs were accepted by GlnRS only 25 times less efficiently than $tRNA^{Gln}$. The inability of GlnRS to use this heavily modified $tRNA^{Gln}$ as well as it uses $tRNA^{Gln}$ indicates that antideterminants must be found in $tRNA^{Glu}$, which hinder its use by GlnRS.[75]

5.14.3.1.4 Binding of ATP and glutamine in the GlnRS active site

The binding of tRNA works with the binding of the two other substrates of GlnRS: ATP and Gln. The binding site for ATP is mainly composed of conserved residues found in the Rossman fold: specificity for ATP in particular, rather than the other nucleotides known to be bound by dinucleotide folds in other proteins, is conferred by hydrogen bonds between the purine ring of ATP and Leu261 and Arg260. The HIGH signature interacts with the phosphates of ATP through His40 and His43. The ATP phosphates were found to interact with Lys270, of the KMSKS signature, along with several other residues as well as Mg^{2+}.[77] This orients the ATP molecule so that the α-phosphate, responsible for the activation of Gln, is near enough to the terminal A76 of tRNA for hydrogen bonds to be formed. Tyr211 also binds A76, allowing both to form parts of the glutamine-binding site.[78] Two interactions occur on the carboxamide moiety of Gln: the oxygen is thought to interact with the 3′ OH of ATP, whereas its hydrogens participate in the formation of a pentagonal array of

hydrogen bonds with Gln255, Tyr211, and two water molecules, said to be an antidromic circular hydrogen-bonding array. This is the first example of a synthetase recognizing its cognate aa through such ambiguous hydrogen bonds.[79,80] Curiously, Glu may also interact with the residues and water molecules involved in this array, albeit differently, even forming a supplemental hydrogen bond with Arg30.[79] This resulting shift in the position held by Glu relative to that held by Gln leaves it poorly placed to react with ATP and A76 and is also thought to result in conformational shifts inside the active site, which contribute greatly to a reduction in the affinity of GlnRS for $tRNA^{Gln}$.[81] The electrostatic environment of the active site was also found to be an important determinant in Gln specificity, in accordance with earlier modeling.[80,82]

5.14.3.2 The Indirect Pathway of Gln-$tRNA^{Gln}$ Biosynthesis

Soon after GlnRS was discovered, it was found that in some organisms, such as *Bacillus megaterium*, an enzyme existed that was capable of activating and charging Glu onto both $tRNA^{Gln}$ and $tRNA^{Gln}$. A novel enzyme, glutaminyl-tRNA amidotransferase (Glu-AdT), was found to be responsible for the amidation of Glu-$tRNA^{Gln}$ into Gln-$tRNA^{Gln}$.[83] The formation of Glu-$tRNA^{Gln}$ in *B. subtilis* was later identified as a function of its GluRS, which was said to be nondiscriminating (ND).[84] This indirect pathway for the formation of Gln-$tRNA^{Gln}$ was progressively found to be the norm in most bacteria, with a few exceptions, such as certain γ-proteobacteria[85] and the Deinococcus/Thermus group.[86,87] Eukaryotes have been found to use the direct pathway for Gln-$tRNA^{Gln}$ formation using GlnRS in the cytoplasm, whereas in organelles, the indirect pathway appears to be used in plants and may or may not be used in yeast and mammals.[88–91] In all these organisms, the indirect pathway for the incorporation of glutamine into protein is allowed by the heterotrimeric GatCAB amidotransferase.[92] In archaea, GlnRS is entirely absent, and the indirect pathway for Gln-$tRNA^{Gln}$ formation is catalyzed by a different heterodimeric amidotransferase called GatDE.[93] The GatA and GatD subunits play the same role, but they evolved from different protein families; GatA bears similarities to the amidase family, whereas GatD belongs to the type-I L-asparaginase family. GatB and GatE play the same role in their respective enzymes, binding the cognate tRNAs and catalyzing the amidation of an ester-linked dicarboxylic amino acid and are more related to one another than to any other protein family.[93] It was recently found that in archaea possessing the GatCAB enzyme as well as GatDE, the heterotrimeric amidotransferase is restricted to the synthesis of Asn-$tRNA^{Asn}$ through the action of domain-specific tRNA identity elements.[94] In eukarya and bacteria, GatCAB is able to catalyze the formation of both Gln-$tRNA^{Gln}$ and Asn-$tRNA^{Asn}$, being limited in function *in vivo* to the availability of the mischarged substrates and the ND-GluRS and ND-AspRS.

5.14.3.2.1 Structure of glutamyl-tRNA synthetase: Catalyzing the first step of the indirect pathway

The structure of the D-GluRS from *Thermus thermophilus* revealed that GluRS bears great similarity to GlnRS.[95] Like all other class I aaRSs it contains a core Rossman fold domain, which is responsible for the binding of Glu and ATP. Inserted between the two halves of this domain is the acceptor stem-binding domain, composed of α-helices and β-strands, which is responsible for the binding of the acceptor arm of cognate tRNAs. This domain is homologous to the acceptor stem-binding domain of GlnRS. The C-terminal extension from the second half of the Rossman fold is composed of two helical domains, forming the greater anticodon-binding region. These two domains bear the same function as the two C-terminal β-barrels of the GlnRS anticodon-binding domain, but are unrelated in both sequence and structure. As for GlnRS, tRNA must be bound to GluRS for activation of the cognate aa to occur.[96] GluRS is composed of these four (sometimes five, as explained below) domains in bacteria, archaea, and some eukaryotes such as yeast, whereas in other eukaryotes, it forms the N-terminal part of a polypeptide with both GluRS and ProRS moieties, called GluProRS.[97,98] When expressed by themselves, both of these moieties retain their activity, and sequence homology studies suggest that their fusion into a single polypeptide was a relatively recent event.[99] In yeast, GluRS and MetRS bind the Arclp protein, which plays a role in cellular localization, whereas in other eukaryotes, the GluProRS enzyme is a part of the multisynthetase complex.[98,100,101]

5.14.3.2.2 Acceptor stem recognition by GluRS

As for GlnRS, the specificity of GluRS for its cognate tRNAs is dependent on identity elements as well as on the overall tertiary structure of tRNA. Binding of $tRNA^{Glu}$ also induces conformational shifts in GluRS, inducing an 8° rotation between the C-terminal part of the dinucleotide fold, which has been called the stem-contact fold and is homologous to the helical subdomain of GlnRS,[96,102] and the N-terminal part of the dinucleotide fold, whereas an 8° rotation also occurs between both helix bundles of the anticodon-binding domain. The acceptor stem-binding domain also rotates after tRNA binding, shifting by 7° relative to the rossman fold, which aids in binding the acceptor stem. On the acceptor stem, C74 does not interact with the adjacent nucleotides, instead it forms hydrogen bonds with Glu107, Ser181, and Arg147, allowing the 3′ end of the tRNA to fold back into a hairpin conformation. A73, C75, and A76 stack with Trp209 and Tyr187, with C75 interacting with Asp44 and Arg47 on the Rossman fold. Other residues, as well as a bound zinc atom, have also been found to play a part in the stabilization of this hairpin.[103] Meanwhile, A76 forms hydrogen bonds with Lys180, Tyr187, and Thr43. The hairpin conformation of the single-strand end of the tRNA, as well as the base stacking of most single-strand nucleotides, is similar to the conformational changes GlnRS induces in $tRNA^{Gln}$, with the notable exception that the G1–C72 base pair of $tRNA^{Glu}$ is not disrupted.[95] A GluRS-specific insertion motif appears to interact with G10, U11, G12, and Ψ13 in the augmented D-loop of $tRNA^{Glu}$[104,105] through the residues in the stem-contact fold mentioned above. These interactions and others induce a shifting of the KISKR (the *T. thermophilus* GluRS KMSKS) loop toward the active site upon tRNA binding. Other identity elements, some allowing the formation of base triplets, were found in the D-loop of *E. coli* $tRNA^{Glu}$ and include U11–A24, C12–G23–C9, and U13–G22–A46.[105]

5.14.3.2.3 Binding of ATP and glutamic acid by GluRS

The binding of ATP occurs in the dinucleotide fold, with the adenine ring being bound in a hydrophobic pocket, while the phosphates interact with the loop containing the KISKR residues. Although ATP and Glu could bind GluRS in the absence of tRNA, ATP was bound in an unproductive manner preventing the activation of Glu, which is recognized through interactions with its γ-carboxyl, thereby discriminating against the similar Gln.[96] When tRNA is bound, ATP rotates by 37°, with the adenine ring remaining in the hydrophobic pocket, whereas the phosphates bind entirely new residues due to the tRNA-induced shifting away of the KISKR motif, bringing the α-phosphate into close proximity with A76 and Glu and allowing activation. Two other sets of interactions with the acceptor stem are also thought to contribute to the conformational shifts allowing productive binding of ATP.[106]

5.14.3.2.4 Anticodon recognition: Important differences between the D- and ND-GluRSs

Binding of the tRNA anticodon is where the fewest similarities lie between GluRS and GlnRS. In contrast to GlnRS, GluRS does not disrupt the base stacking of the anticodon, with both C-terminal helix domains making a binding pocket able to accommodate all three bases. In $tRNA_2^{Glu}$, C34 forms specific interactions with Leu427, Leu447, Phe448, and Arg435. When binding $tRNA_1^{Glu}$, the modified mnm^5s^2U forms hydrogen bonds with Arg435 and Leu426, while a part of the modified base is accommodated by an empty space. Thr444, Gln432, and Leu442 form specific interactions with U35, whereas C36 is bound by Arg358, which itself is held by Leu354, Pro445, and A37.[107]

The anticodon bases are important specificity elements for both GlnRS and GluRS, as only the third base differs between $tRNA^{Glu}$ and $tRNA^{Gln}$, with both likely possessing the same modification on the U34 of one of their isoacceptors. This binding strength was highlighted by the significant inhibition of *E. coli* GluRS by minihelices mimicking the anticodon stem–loop of $tRNA^{Glu}$.[108] It was, therefore, expected that recognition of the third anticodon base would differ between the D- and ND-GluRS. This less specific and possibly weaker interaction with the third anticodon base may be compensated by an increased affinity for U34, as suggested by the inability of *B. subtilis* GluRS to glutamylate the *E. coli* $tRNA_2^{Gln}$ and other C34 variants.[109] In bacteria from the *Bacillus/Clostridium* group, which rely on an ND-GluRS and GatCAB to form $Gln\text{-}tRNA^{Gln}$, Arg358 is replaced by a glutamine residue. Although structural evidence is still lacking to demonstrate this explicitly, it is thought that the smaller side chain of glutamine would avoid steric hindrance, which would be inevitable between arginine and the 2-amino and 5′ phosphate groups of G36. Mutational analysis of this hypothesis with the *T. thermophilus* enzyme revealed that Arg358Gln GluRS is able to glutamylate $tRNA^{Gln}$ as well as

$tRNA^{Glu}$.[107] The 3D structure of an ND-GluRS bound to $tRNA^{Gln}$ has not yet been revealed, but modeling of $tRNA^{Gln}$ binding using the crystal structure of the ND-GluRS from *Thermosynecholoccus elongatus* has given hints as to how this enzyme binds two significantly different tRNAs. This ND-GluRS bears 37% sequence identity with that of *T. thermophilus*, with conservation being greater for the N-terminal portion of the enzyme, particularly in the ATP- and Glu-binding regions. Several differences between them have been suggested to be of minimal importance to tRNA specificity, as they result in structural differences in areas of GluRS not thought to be involved in tRNA binding. While Arg358 appears to be an important contributor to discrimination against $tRNA^{Gln}$ in *T. thermophilus*, this residue is present in many ND-GluRS. It has been proposed that the presence of Glu443, a residue absent from ND-GluRS, which has Arg358, is involved in the discrimination mechanism. Glu443 forms a salt bridge with Arg358, which is broken upon binding of $tRNA^{Gln}$, allowing Glu443 to bind the tRNA backbone.[110] The ability of a single residue to restrict tRNA specificity from the ancestral nondiscriminating glutamylation of $tRNA^{Gln}$ and $tRNA^{Glu}$ to that of $tRNA^{Glu}$ alone reinforces the hypothesis that this event happened independently in individual organisms after the acquisition of eukaryotic GlxRS.[45] In the light of this, it is not surprising that, in some organisms, a GluRS has evolved to have greater or even exclusive specificity for $tRNA^{Gln}$ alone (see below).

5.14.3.2.5 The noncanonical GluRS2

Helicobacter pylori is a pathogenic bacterium frequently involved in the development of stomach ulcers.[111] This bacterium survives in the presence of stomach acids by producing large quantities of urease, which releases ammonia, neutralizing its environment.[112] The sequencing of the *H. pylori* genome allowed the identification of two open reading frames (ORFs) with sequence homology to GluRS, which were named GluRS1 and GluRS2.[113,114] Biochemical assays confirmed that both enzymes catalyzed the glutamylation of tRNA. GluRS1 was revealed to be of the discriminating type, whereas GluRS2 was shown to be a novel, discriminating enzyme that specifically misacylates $tRNA^{Gln}$ with Glu. GluRS2 is a paralogue of GluRS1, believed to have evolved from a duplicated *GluRS* gene. Indeed, the two enzymes show a sequence homology of 38% and a similarity of 53%, with GluRS2 sharing no significant homology to GlnRS. These results support the hypothesis that the acquisition of specificity for $tRNA^{Gln}$ or $tRNA^{Glu}$ is an event that occurred independently over time.[115] A similar scenario was found in *Acidithiobacillus ferrooxidans*, another bacterium that survives in an acidic environment and that contains an ND-GluRS1, as well as a GluRS2, which preferentially glutamylates $tRNA^{Gln}$. Both *H. pylori* and *A. ferrooxidans* lack AsnRS and GlnRS, and both rely on the GatCAB amidotransferase to form Asn-$tRNA^{Asn}$ and Gln-$tRNA^{Gln}$.[114,116] As mentioned previously, the presence of an augmented D-loop containing triple base pairings is an important identity element for GluRS.[105] This augmented D-loop may be an antideterminant for some ND-GluRS and GluRS2 enzymes against $tRNA^{Glu}$ as well as $tRNA^{Gln}$ species that contain this element.[109,115]

Sequence analysis has revealed that nearly all D-GluRS contain an arginine cognate to the Arg358 of *T. thermophilus*, while this residue is always absent from GluRS2 and variable in ND-GluRS. About 30 other organisms with a gene showing homology to GluRS2 were identified, and the putative GluRS2 gene in these organisms was always accompanied by a GluRS1 gene. The Thr444 residue of *T. thermophilus* GluRS is often present in other D-GluRSs or replaced by a similarly small residue, whereas in GluRS2, the equivalent position always holds Gly. The role of anticodon recognition in the $tRNA^{Gln}$ preference of GluRS2 was examined using mutagenesis. Using the *T. thermophilus* numbering, Arg358Glu GluRS1 and Glu358Arg GluRS2 were created; the mutated GluRS1 was still able to glutamylate $tRNA^{Glu}$ and did not gain the ability to glutamylate $tRNA^{Gln}$, whereas Glu358Arg GluRS2 gained no significant activity for $tRNA^{Gln}$ and lost activity for $tRNA^{Gln}$. These results indicate that Arg358 is an antideterminant for the aminoacylation of $tRNA^{Gln}$ for GluRS2. Two more GluRS2 mutants were used to discover how GluRS2 discriminates against $tRNA^{Glu}$: Gly444Thr GluRS2 and the double mutant Gly444Thr/Glu358Arg GluRS2 were created to study the role of the conserved GluRS2 Gly444. The Gly444Thr GluRS2 was able to glutamylate $tRNA^{Glu}$, but not $tRNA^{Gln}$, while Gly444Thr/Glu358Arg GluRS2 glutamylated $tRNA^{Glu}$ even more efficiently and did not possess a restored $tRNA^{Gln}$ activity. This increase is intriguing, as the single Glu358Arg mutant was unable to increase $tRNA^{Glu}$ activity on its own. Gly444 in GluRS2 is almost always followed by Pro, and it has been suggested that these two residues allow a turn in the polypeptide chain, modulating anticodon recognition in a still undetermined fashion.[117] This mechanism would be reminiscent of the role of Glu443 in *T. thermophilus* GluRS, altering anticodon recognition through a shift in the anticodon-binding domain.[110]

5.14.3.3 Divergence of the Indirect Pathways for the Formation of Gln-tRNA^{Gln}

The heterotrimeric GatCAB is found in bacteria and organelles in cases where either GlnRS or AsnRS or both enzymes are missing, whereas in archaea it is present only when AsnRS is absent.[27] As described earlier, the bacterial and eukaryotic GatCAB enzymes are capable of catalyzing the formation of both Gln- tRNA^{Gln} and Asn-tRNA^{Asn} *in vitro* with their homologous tRNAs, whether or not they hold both functions *in vivo*.[89,118] The GatCAB enzyme appears to use the same identity elements to recognize tRNA^{Asn} and tRNA^{Gln} and the same discrimination elements to exclude tRNA^{Asp} and tRNA^{Glu}.[60,119] The structure and mechanism of GatCAB for the formation of Asn-tRNA^{Asn} and Gln-tRNA^{Gln} are described in detail in Section 5.14.4.2.4.

In archaea, the reason behind the presence of GatCAB and the related absence of AsnRS could be that the archaeal GatCAB is unable to transamidate its homologous tRNA^{Gln}.[94] This divergence is the result of sequence and structure differences between the tRNA^{Asp} and tRNA^{Asn} of archaea and those of most bacteria and eukarya: tRNA^{Asn} has a G1–C72 base like tRNA^{Asp}, which means that this base pair cannot be an antideterminant for GatCAB as it is in bacteria. Discrimination of archaeal GatCAB against tRNA^{Asp} was found to be, in part, the function of an additional base in the region between the T-stem and the variable loop of tRNA^{Asp}, U49. Consequently, this GatCAB is still able to transamidate Glu-tRNA^{Gln} when the tRNA is of bacterial origin.[94] It has been suggested that the G1–C72 tRNA^{Asn} in archaea has evolved to hinder its binding by the Glu-tRNA^{Gln}-specific GatDE, subsequently forcing the archaeal GatCAB to alter its recognition elements for tRNA^{Asn}.[120] This would have also kept the archaea from acquiring GlnRS, which also depends on the absence of a G1–C72 base pair.

5.14.3.3.1 Evolution of the dimeric, monospecific GatDE enzyme from archaea

The GatE subunit is similar to GatB, except that it contains a major insertion domain.[93] Therefore, it might be expected that GatE is an evolutionary offshoot of GatB, specific to archaea. However, phylogeny studies have revealed that GatE sequences group together separately from GatB, unlike the case of GlnRS, which can be grouped with the eukaryotic GluRSs.[45,121] This indicates that GatE diverged from GatB before the separation of bacteria and archaea. In the light of these results, as well as others, a model was proposed for the evolution of GatDE and GatCAB: the ancestral GatB/E was duplicated in the last universal communal ancestor (LUCA), with GatB eventually associating with a protein of the amidase family as well as the precursor of GatC, and with GatE associating with the type I asparaginase ancestor of GatD. GatDE would have been then kept in the archaea, whereas GatCAB was kept in some organisms of all domains according to the coevolution of GatDE and tRNA^{Gln}, as well as differences in the general tRNA pool of these organisms.

5.14.3.3.2 The 3D crystal structure of the GatDE amidotransferase

Structures of GatDE with and without bound tRNA^{Gln} were obtained from *Pyrococcus abyssi* and *Methanothermobacter thermautotrophicus*, respectively.[61,122] GatDE was shown to form a tetramer through interactions between the GatD subunits of two GatDE proteins. GatE contains three structural modules. The N-terminal domain is folded around a twisted and rolled 10-stranded mixed β-sheet. The convex side of this sheet rests upon three α-helices and held to them through hydrophobic interactions between the side chains of this interface. This structure forms a half torus, which is extended by a mixed β-sheet and α-helix, associated with the α-helices that rest against the convex side of the 10-stranded β-sheet. Together, these elements form what has been called the cradle domain, which has not yet been found in any other protein except GatB.[60] Inserted into the primary structure between the cradle domain and its underlying α-helices is a domain homologous to the insertion domain found in bacterial-type AspRS,[123] which contains conserved residues (**Figure 5**). This suggested that it may have a role in acceptor arm binding, as this insertion domain does in bacterial AspRS.[124] The GatDE-tRNA^{Gln} cocrystal structure revealed that the AspRS-like domain does not interact with the tRNA, but that it may play a role in substrate channeling between GatDE and the ND-GluRS, which provides its Glu-tRNA^{Gln} substrate.[61] At the other end of the cradle domain's β-sheet lies a helical domain that has little contact with the core domain, whereas the C-terminal end of GatE bears a tail domain that could not be resolved entirely, but which is known to hold sequence similarity to the *B. subtilis* Yqey protein.[122]

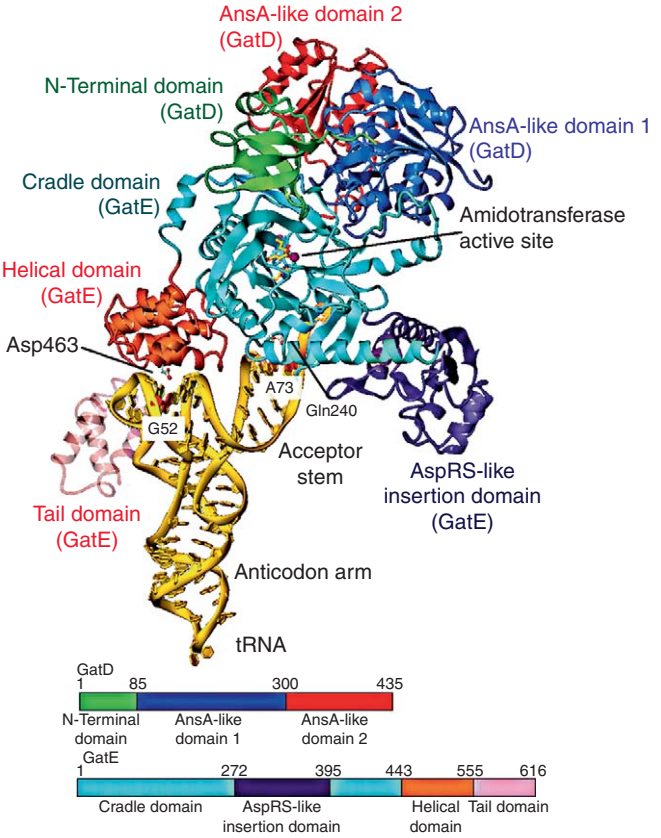


Figure 5 Structure of the archaeal GatDE aminoacyl-tRNA amidotransferase. GatDE forms a heterotetramer, with two GatDE proteins binding through their GatD subunits. Binding of $tRNA^{Gln}$ induces a conformational shift of the helical and tail domains, which scan the D-loop and the T-arm. This figure was reproduced from H. Oshikane; K. Sheppard; S. Fukai; Y. Nakamura; R. Ishitani; T. Numata; R. L. Sherrer; L. Feng; E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam; D. Söll; O. Nurehi, *Science* **2006**, *312*, 1950–1954.

5.14.3.3.3 Specificity of GatDE for archaeal $tRNA^{Gln}$

The cradle domain contains conserved residues in both GatE and GatB when the presence of the AspRS-like insertion domain of GatE is taken into account. These residues are mostly located on the concave surface of the cradle domain and are oriented toward the same region.[122] $tRNA^{Gln}$ has been shown to bind only to the GatE subunit, which does not contact the anticodon loop. Binding of the tRNA induces a shift in the orientation of both the helical domain, which contacts the minor groove of the T-arm, and the C-terminal tail domain, which seems to interact with the D-loop–T-loop assembly region. Upon binding GatDE, the last four nucleotides of the acceptor stem terminus pass through an opening between the β-strands of the cradle domain. The discriminator base A73 is then bound by Gln240 through a bidentate hydrogen bond, whereas the phosphate backbone of C74 and A73 also interacts with the helical domain Arg503 and the cradle domain Lys401, respectively. Mutation of either Gln240 or Arg503 to Ala resulted in a reduced transamidase activity.[61] The helical domain also recognizes several nucleotides of the T-stem; Arg 503 binds the phosphate backbone of C62, whereas Tyr496, Gln467, and Asp463 have specific interactions with G53, C62, and G52, respectively. Although mutation of some of their interacting GatE residues resulted in a decreased GatE activity, G52, G53, and C62 are conserved among $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Asn}$ and, therefore, cannot be identity elements for GatDE (**Figure 5**). It was proposed that these interactions serve to allow more robust interactions between the tail domain and the T-loop and D-loop. Although the residues involved are not known, these interactions seem to twist the C56–G18 base pair, whereas the tail domain binds the tRNA closely and adopts a concave surface preventing the binding of larger T-loops and D-loops, such as that of $tRNA^{Glu}$. In contrast, $tRNA^{Asn}$ has a

D-loop of a similar size as that of $tRNA^{Gln}$, which may require that a different mechanism be present to exclude it from GatDE activity. Gel-shift and transamidase activity assays using $tRNA^{Gln}$ with either a T-loop replaced by that of $tRNA^{Glu}$ or a D-loop replaced by that of $tRNA^{Glu}$ or $tRNA^{Asn}$ confirmed the existence of such a supplementary antideterminant: As predicted, the replacement by a $tRNA^{Glu}$ D-loop eliminated binding and activity, whereas substitutions with a $tRNA^{Asn}$ D-loop diminished binding and eliminated activity. This loss of activity is thought to be caused by unfavorable interactions with positions 19 and 20 of $tRNA^{Asn}$. The substitution with the $tRNA^{Glu}$ T-loop reduced the binding efficiency but still allowed some activity, establishing the size of the D-loop as the main discriminator for $tRNA^{Glu}$. In addition to the antideterminants identified in its D-loop, archaeal $tRNA^{Asn}$ is also discriminated against because of its G1–C72 base pair.[94] The replacement of the U1–A72 first base pair of $tRNA^{Gln}$ by G1–C72 was sufficient to reduce the amidotransferase activity by a factor of 50. With the wild-type $tRNA^{Gln}$, the position of A76 could not be determined, possibly because glutamylation is required for proper placement of A76. However, a pocket large enough to accommodate glutamyl-adenosine was located in the vicinity of C75.

5.14.3.3.4 The predicted mechanism for tRNA-dependent amidation by GatDE

The conserved residues His15, Glu157, and Glu184 appeared in the catalytic pocket as predicted, with a coordinated Mg^{2+}. In the absence of a structure with the active substrate bound, modeling was performed to confirm that the acceptor stem and glutamylated A76 could be held correctly through the cradle domain to allow the side chain carboxyl oxygen of Glu to take a position allowing transamidation with the help of the coordinated Mg^{2+}. This brought the α-carbonyl and α-amino groups of the glutamyl-adenosine, which would also be present on aspartyl-adenosine, near enough to the conserved Lys237 and Glu235 for hydrogen bonds to be formed. Modeling was also done to study the putative kinase site, using data from the ADP/aluminum fluoride (AlF)-bound GatB structure;[60] in this manner, the active site Mg^{2+} was predicted to coordinate octahedrally with His15, Glu157, Glu184, the γ-carboxyl group of glutamyl-A76, the γ-phosphate of ATP, and a water molecule to allow the formation of the γ-phosphoryl-glutamyl-$tRNA^{Gln}$ intermediate.[61] Both GatE and GatB are thought to activate Glu-$tRNA^{Gln}$ through a similar mechanism, but only GatE has been shown to be capable of activating Glu-$tRNA^{Gln}$ in the absence of an ammonia group donor.[125]

5.14.3.3.5 Dimerization of GatDE and the glutaminase activity

The GatD subunit was found to be structurally similar to the type I L-asparaginase, as expected. This subunit is composed of two domains: domain 1 contains a flavodoxin fold, whereas domain 2 is composed of four parallel β-strands and three flanking α-helices. From this core domain extends a linker that holds an additional domain – an open β-barrel formed by a five-strand β-sheet and an α-helix, which was shown to be important for the binding of GatE. Assembly of a core GatD dimer was shown to be mediated by extensive contacts between domains 1 and 2 of the two GatD subunits, forming a four-bridge link. This dimerization leaves the glutaminase active site of each GatD subunit free to bind its substrate and also forms a platform upon which two GatE subunits bind to form the heterotetramer. Conserved motifs in each GatE bind its own GatD through domain 1 and its N-terminal linked barrel as well as through contacts with domain 2 of the other GatD dimer subunit (**Figure 5**). Other GatE–GatD interactions occur, among which are contacts between the overhanging domain of the GatE cradle and the region of the glutaminase site.[122]

5.14.3.3.6 Bridging the 40-Å gap between glutamine hydrolysis and Gln-$tRNA^{Gln}$ synthesis

The glutaminase active site was confirmed to be on GatD in the region of Thr101, Thr177, Asp178, and Lys254, as was suggested by earlier studies where the mutation of these residues resulted in a GatDE unable to hydrolyze amide donors, but still able to transamidate using free ammonia.[125] This positions the glutaminase site 40 Å from the transamidase and kinase pocket on GatE. An opening was found on GatD, near the glutaminase site.[61] This opening allows ammonia molecules produced by the hydrolysis of Gln or Asn to access a tunnel, which passes through both domains of GatD, through the GatD/GatE interface and to the Mg^{2+} ion in the GatE catalytic pocket. Tyr373 is near the opening of the tunnel on GatD, and mutation of this residue to Phe reduced the transamidase and glutaminase rates of the enzyme considerably. This has led to the hypothesis that Tyr373 attracts ammonia ions released in the glutaminase site, allowing them to be taken up by

the transport mechanism, which guides them across the tunnel to GatE. This mechanism is often found with amidotransferase enzymes using glutamine as a donor and may also be similar to the proton-relay tunnel, which has been proposed for GatCAB. In both these examples, mechanisms have also been proposed for closing the ammonia channel under the condition of lack of the presence of all substrates; such a mechanism for the GatDE amidotransferase remains to be found.[60,126]

5.14.4 Partners of Direct and Indirect Pathways of Asn-tRNA^{Asn} Biosynthesis and their Reaction Mechanism

5.14.4.1 Direct Pathway of Asn-tRNA^{Asn} Formation

5.14.4.1.1 Characterization of asparaginyl-tRNA synthetase

AsnRS was first characterized in the early 1960s in *L. arabinosus* by its distinct chromatographic properties from aspartyl-tRNA synthetase and Asn synthetase[127,128] and, at the end of the decade in *E. coli* as a protein of M_r 90–100 kDa able to aminoacylate tRNA^{Asn} but not tRNA^{Asp}.[129,130] The enzyme was further characterized in mitochondria from *Neurospora crassa*[131] and in rat liver[132] where tRNA asparaginylation activity was found associated with proteins of M_r 35 and 90 kDa. The AsnRS purified from *B. stearothermophilus* of M_r 127 kDa was characterized as a homodimer α_2 ($\alpha = 51$ kDa).[133] AsnRSs isolated until now from various prokaryotic and eukaryotic organisms show conservation of the homodimeric structure. Investigation of *B. stearothermophilus* AsnRS allowed the determination of its physicochemical parameters: $S_{20,\omega} = 6.6 \times 10^{-13}$ s, $D_{20,\omega} = 5.1 \times 10^{-7}$ cm^2 s^{-1}, and Stokes radius $= 41.4 \times 10^{-8}$ cm.[134] The M_r of 127 kDa derived from centrifugation experiments equals that determined by gel filtration (120 kDa), in agreement with a globular shape of the protein. For the AsnRS of Chinese hamster ovary (CHO), cells with an $S_{20,\omega}$ value of 8.1×10^{-13} s has been characterized; its M_r of 152 kDa determined by centrifugation agrees with the value of 155 kDa found by gel filtration.[135]

5.14.4.1.2 Catalytic properties of AsnRS

Kinetic constants determined for *E. coli* and *T. thermophilus* AsnRSs are similar to those determined for mammalian AsnRSs and reveal strong affinities of AsnRS for the small substrates. *E. coli* AsnRS from exhibits K_m values for Asn and ATP of 15 and 500 μmol l^{-1}, respectively, in ATP–PP$_i$ exchange and 29 and 76 μmol l^{-1} in tRNA aminoacylation.[136] Rate constants of 3 and 2.8 s^{-1} were determined at 37 °C for these reactions, respectively. For AsnRS from *T. thermophilus*, a K_m of 48 nmol l^{-1} was determined for tRNA^{Asn} and a k_{cat} of 0.4 s^{-1} for tRNA aminoacylation at 70 °C. For the enzyme from CHO cells, K_m values of 30 μmol l^{-1} and 60 nmol l^{-1} were determined for Asn and tRNA^{Asn} respectively. These kinetic constants are in agreement with those determined for the aaRS family. Preliminary studies of the rat liver enzyme have suggested that like GluRS, GlnRS, and ArgRS, AsnRS requires the cognate tRNA to promote aa activation.[137] However, investigations of AsnRSs of various origins clearly showed that like AspRS, AsnRS activates the cognate aa in the absence of tRNA. Until now, the kinetic data concerning the asparaginylation system are poorly documented, probably because of the lack of commercial availability of labeled Asn. The sole mechanistic information of AsnRS concerns the rate-limiting step of the overall reaction. Comparison of the rate constants of ATP–PP$_i$ exchange (3 s^{-1}), which represents a minimal value of aa activation, and of tRNA charging (2.8 s^{-1}) at 37 °C suggests that the transfer step or the release of Asn-tRNA^{Asn} determines the steady-state rate of Asn-tRNA^{Asn} formation.[136] In contrast, cloning and sequencing of the genes of AsnRS of various origins and analysis of the sequenced genomes allowed a comparison of the protein sequences of a vast number of AsnRSs. The alignments of AsnRS sequences and resolution of the 3D structures of the enzymes from *T. thermophilus*[138] and *Pyrococcus horikoshii*[139] reinforced the knowledge of the structural properties of these aaRS species.

The first AsnRS sequenced was that of *E. coli* after the gene was cloned.[140] It was also one of the last aaRSs sequenced from this organism. Alignment with other aaRS sequences revealed important similarities with AspRSs and, to a lesser extent, with LysRSs, suggesting strong structural and phylogenetic interrelations between these aaRS species.

5.14.4.1.3 Structural properties of AsnRS and recognition of the small substrates

AsnRS belongs to class II aaRSs, characterized by three more or less well-conserved structural motifs. Motif I is involved in dimerization, and an invariant Pro residue plays a crucial role in the organization of the catalytic center. In GlyRSs, this Pro is substituted by a Ser or a Thr residue. Motifs II and III containing invariant Arg residues are involved in binding of the amino acid and ATP substrates and in catalysis. The catalytic center is formed by a nine-stranded antiparallel β-sheet surrounded by four α-helices. These aaRSs bind the ATP in the bent conformation in the presence of three Mg^{2+} ions and the tRNA by the major groove side. In the aaRS:tRNA complex, the 3′ CCA end of tRNA adopts an extended conformation. This class of aaRSs catalyzes the esterification of the amino acid to the 3′ OH ribose of the terminal adenosine of tRNA through a two-step process: activation of the amino acid into aa∼AMP and transfer of the activated amino acid onto the tRNA.[49–51]

5.14.4.1.3(i) The 3D structure of free* T. thermophilus *AsnRS The crystals obtained with *T. thermophilus* AsnRS diffracting at 2.6 Å of resolution[138] permitted the determination of its 3D structure,[141] which was the 12th representative of the 20 solved aaRS structures.

The modular organization of AsnRS resembles that of AspRS and LysRS. Each subunit of the dimeric AsnRS is formed by an N-terminal β-barrel connected to the C-terminal domain built upon an $\alpha\beta$-fold by a small hinge region. An N-terminal segment of variable length, which is significantly longer in eukaryotes than in prokaryotes, precedes the β-barrel. The N-terminal domain is formed by a five-stranded β-barrel in which two strands are separated by an α-helix. By analogy of the 3D structures of AspRS and LysRS complexed to the cognate RNA, this domain is involved in tRNA anticodon recognition. This β-barrel characterizes subclass IIb aaRSs and is functionally equivalent to the $\alpha\beta$-fold in the C-terminus of subclass IIa aaRSs.

The C-terminal domain is built around an $\alpha\beta$-fold including the three class-defining motifs. It contains the catalytic site formed by a six-stranded antiparallel β-sheet including motifs 2 and 3 and by a dimer interface, which includes motif 1. This β-sheet is interrupted between motifs 2 and 3 by a module formed by four helices and a β-strand.

The hinge region connecting the N-and C-terminal domains is formed by a turn followed by two α-helices. It resembles that of LysRSs but is longer in AspRSs.

AsnRS, LysRS, and AspRS, the partners of class IIb aaRSs, display strong resemblances in the modular organization of their catalytic domain, but superposition reveals angular shifts of the N-terminal domain, which can reach up to 18.8°.[142]

Association of the two subunits involves four types of interactions, which are mostly conserved in AspRSs and LysRSs. On the top of the interface, the contacts involve Asp, Glu, and Arg residues from motif 1 distributed along the interfaces of an α-helix and the following β-strand, which are reinforced by stacking of the His142 residue from each subunit. On the bottom of the dimer interface, two β-strands from each monomer form an antiparallel β-sheet. The internal parts of the interfaces are associated by hydrophobic interactions involving Ile, Leu, and Val residues, three β-strands, and Phe207 from motif 2. The association is strengthened by intersubunit contacts between the C-terminal domain from each subunit and the N-terminal and hinge domains from the other subunit.

5.14.4.1.3(ii) The 3D structure of* T. thermophilus *AsnRS bound to the small ligands Crystals of the enzyme bound to asparaginyl-adenylate (Asn∼AMP) diffracting at 3.2 Å of resolution were obtained by cocrystallization of the protein with ATP and $MgCl_2$. The structure was refined from crystals formed by the protein bound to the nonhydrolyzable 5′-*O*-(*N*-asparaginyl-sulfamoyl) adenosine analogue diffracting at 2.65 Å.[141]

Interactions between the subunits intimately link the dimer interface with the ATP-binding site. The adenine ring and the γ-phosphate of ATP interact, as in other class II aaRSs, with conserved residues from motifs 2 and 3. Three Mg^{2+} ions associate with the triphosphate group, one with the α- and β-phosphates and with Asp352 and Glu361 and the other two with the β-and γ-phosphates on each side of the phosphoester bond. These interactions stabilize ATP in the bent conformation, which is essential for the activation of the aa by class II aaRSs. The adenosine moiety interacts as in ATP: the adenine ring with residues Glu210, Leu218, Phe235, and Arg412 and the 2′ OH ribose with Ile362 and Gly409. Glu361 interacts with the 3′ OH and together with Asp352 through a Mg^{2+} ion with the O^- of the phosphoryl group. Arg208 interacts with the oxygen of both the

phosphoryl and the aa carbonyl groups. The aa α-amino group is positioned by Glu164, Ser185, and Gln187, whereas the carbonyl and the amido groups of the side chain make hydrogen bonds with Arg368 and Glu225, respectively.

Comparison of the 3D structures of free and Asn∼AMP-complexed AsnRSs shows that activation of Asn promotes domain movements in the protein and an increased ordering of the loops surrounding the catalytic site resulting in a more precise delimitation of the catalytic site. The motif 2 loop moves in a concerted way with the C-terminal peptide toward the active site, closing its access and sequestering the activated Asn. These contacts promote rotation by 2.5 Å of the N-terminal β-barrel relative to the catalytic domain. The novel conformation is stabilized by intersubunit interactions. Exposure of motif 2 loop on the enzyme surface agrees with cleavage by limited proteolysis of a peptide of M_r 26 kDa starting with the His243 included in this loop.[140] Involvement of this loop in the stabilization of the complex formed with the small ligands is supported by the temperature-dependent increase in K_m for Asn and ATP of the *E. coli* HO202 AsnRS mutant in which Pro231 from motif 2 is substituted by Leu without significant alteration of the k_{cat},[136] suggesting that this residue is involved in positioning of the loop and in sequestration of the small ligands in the catalytic site.

5.14.4.1.3(iii) Discrimination of asparagine against aspartate by AsnRS Comparison of the 3D structures of the catalytic domains of AsnRS and AspRS complexed with the cognate aa∼AMP shows conservation of most of the aa residues contacting the ligands and explains how aa discrimination occurs.[141] AspRSs and AsnRSs are characterized by three differences determining an appropriate structural context that directs the selection of the cognate aa: (1) a conserved Lys residue in AspRSs is replaced by a small or an uncharged residue in AsnRSs; (2) this residue is followed by a Gln residue in AspRSs and a Glu residue in AsnRSs; and (3) an invariant Asp in motif 2 of AspRSs is substituted by Glu in AsnRSs (**Figure 6**).

Binding of Asp to AspRS involves electrostatic interaction between the negatively charged β-carboxylate group of Asp and the positively charged side chains of Arg and Lys residues (Arg483 and Lys204 in *T. thermophilus* AspRS), which are positioned by the carboxylate groups of Glu241 and Asp239. This structural context prevents the interaction of the carboxamide group of Asn, which would lead to an unfavorable head-on interaction with the Lys and Arg residues. Conversely, AsnRS lacks the conserved Lys residue that is replaced by a small side chain residue (Ala190) but displays the conserved Arg (Arg368) in the same configuration. However, the unfavorable head-on interaction of the carboxamide of Asn with the Arg residue is avoided by a 60° rotation of the C–Cα bond, which turns the Asn side chain more into the catalytic site. As a consequence, the carbonyl and the amido groups of Asn are hydrogen bonded to Arg368 and the adjacent Glu225. This residue plays the dual role of recognition of Asn and rejection of the negatively charged β-carboxylate group of Asp. The rotation of the side chain of Asn implies a displacement of its α-amino group, which is precisely positioned by three hydrogen bonds with conserved Ser185, Gln187, and Glu164 residues. In AspRSs, the Lys204 fulfills the reverse dual role of Glu225 in AsnRS by interacting with the negatively charged β-carboxylate of Asp and discriminating against Asn. This residue is positioned by Asp239 at a nearer position than by its equivalent Glu225 in AsnRS and thus is too distant from the Asp side chain to provoke electrostatic repulsion.

The residues that orient the Asn side chain in AsnRS are conserved in AspRS (Ser199, Gln201, and Glu177) but subtle changes in the position of their side chains induce a different orientation of the α-amino group of Asp. In conclusion, a restricted number of residues differ in the catalytic sites of AspRSs and AsnRSs and are directly involved in the selection of the homologous aa, but additional residues participate in correct orientation of the substrate side chains and thus are involved indirectly in their selection.

Comparison of the 3D structure of the free AsnRS of *P. horikoshii* solved at 1.45 Å of resolution[139] and that of the enzyme complexed with either the Asn∼AMP or its analogue asparaginyl-sulfamoyl adenosine solved at 1.98 and 1.8 Å of resolution, respectively, reveals a conformational change upon aa activation, since as the electron density map corresponding to residues 165–172 and 214–219 that are missing in the apo AsnRS becomes ordered after Asn∼AMP formation. This conformational change was also observed in *T. thermophilus* AsnRS and in other class II aaRSs, but it could be better characterized in the *P. horikoshii* complex because of the high resolution of the crystals. The high resolution facilitated the identification of many water molecules that form a hydrogen-bonding network covering the entire AsnRS molecule, including the catalytic center, and two

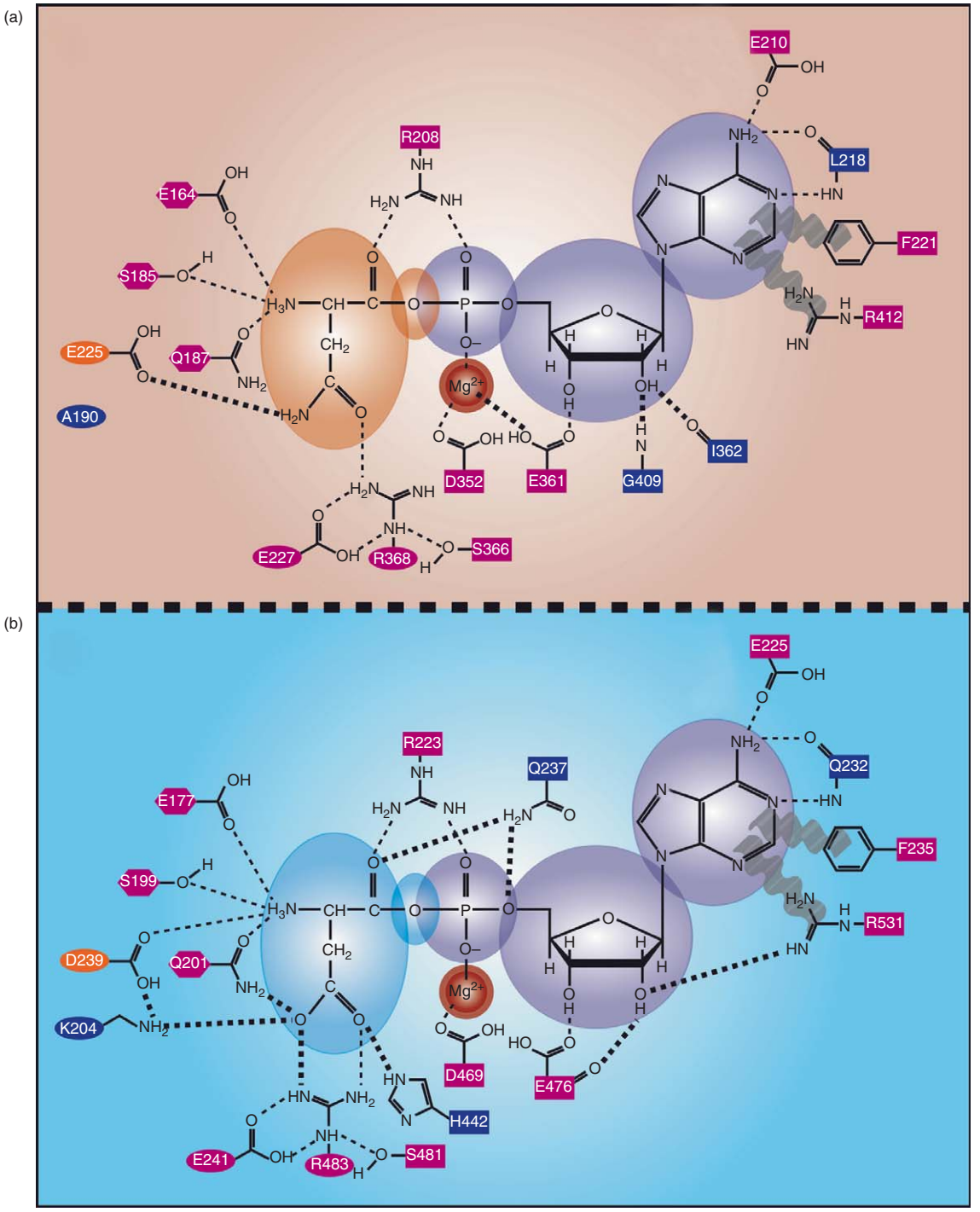


Figure 6 Recognition of the aminacyl adenylate by aspartyl- and asparaginyl-tRNA synthetases. (a) Recognition of Asn ~ AMP by asparaginyl-tRNA synthetase. (b) Recognition of Asp~AMP by aspartyl-tRNA synthetase. Residues on dark blue, orange, and red backgrounds, are respectively, not conserved or of the same chemical nature, or strictly conserved in the active center of both enzymes. Residues in circles and hexagons are involved in the selection of the amino acid; circled residues interact with the carboxamide and the carboxylate groups of Asp and Asn, respectively; residues in hexagons determine the orientation of the side chains. Hydrogen bonds between enzyme residues and aa ~ AMP groups are in dotted or bold-dotted lines when they are respectively conserved or differ in the two enzymes. Stacking interactions are represented by gray waves. The AMP, Asn, and Asp moieties of aminoacyl-adenylates are respectively on violet, orange, and blue backgrounds, the O involved in the mixed anhydride bond is circled and on orange or blue background, and the Mg^{2+} ion is on a brown background.

water molecules that contribute to the precise recognition of Asn. In the complex, the Asn∼AMP is fixed on the center of the antiparallel β-sheet of the catalytic core. One side of the molecule is wrapped by a unique α-helix β-hairpin loop and a curved β-sheet with a long connecting loop, which ensures specific recognition of the product. Some portions of these regions become ordered after the formation of Asp∼AMP, indicating an induced fit of the mobile loops upon formation of Asn∼AMP. The left side of Asn∼AMP, which has few protein interactions with Glu357, Ile358, and Arg408, is exposed to the solvent.

The structure reveals a layer of well-ordered solvent molecules that covers half of the surface of Asn∼AMP. The pocket is filled with water molecules, which form a network of water–water, water–protein, and water–Asn∼AMP hydrogen-bonding interactions. The Asn side chain is docked to the binding pocket that consists of Glu228, Arg364, Leu229, Glu230, Tyr333, Phe402, and Gly403. Residues Glu228 and Arg364 play a major role in defining Asn specificity specificity, as they form hydrogen bonds with the Asn amide and carbonyl groups. However, this Asn pocket is loose in contrast to the highly complementary Asp-binding pocket in AspRS composed exclusively of protein residues. Two water molecules fill the void between the Asn side chain and the protein residues. One forms hydrogen bonds with the carbonyl groups of Leu229 and Glu230 and with the amide group of Asn. The second water molecule forms hydrogen bonds with the Tyr333 hydroxyl group and the Asn β-carbonyl group, and accepts hydrogen bonds from Arg364 and a third water molecule, thereby filling the space between these partners. Likewise, one side of the bound Asn-sulfamoyl adenosine is completely covered by the solvent molecules that complement the binding site. Two of the water molecules interact directly with the Asn amide and carbonyl groups and contribute to the formation of a pocket highly complementary to the Asn side chain. It has been proposed that these water molecules play a key role in the structural recognition of Asn and in the discrimination against Asp by AsnRS. The water-assisted Asn recognition by AsnRS contrasts with the recognition of Asp by AspRS, which is achieved exclusively through interactions with protein residues.

5.14.4.1.4 Structure of tRNA^{Asn} and recognition by AsnRS

5.14.4.1.4(i) Structural and functional peculiarities of tRNA^{Asn} All organisms contain a unique tRNA^{Asn} species with a GUU anticodon that recognizes the AAU and AAC codons. Codons possess a small variable loop containing five, or exceptionally four, nucleotides. tRNA^{Asn} from prokaryotes and eukaryotes are formed by 76 and 77 nucleotides, respectively, because their D-loops contain eight and nine nucleotides. Exceptions in the length of the D-loops were reported for archaebacterial and mitochondrial tRNA^{Asn}. The D-loops of eubacterial tRNA^{Asn} contain two 2-dihydro-uridine (D) residues. In contrast, the D-loops of eukaryotic tRNA^{Asn} contain four modified uridine residues: yeast tRNA^{Asn} contains four D residues and mammalian tRNA^{Asn} contains three D residues and the acp3 (3-amino-3-carboxypropyl)-uridine. D is also found at position 47 of the variable loop of eukaryotic tRNA^{Asn}, and pseudouridine (ψ) is present at positions 27 and 28 of the anticodon stem in mammalian tRNA^{Asn}. The D-loop of mitochondrial tRNA^{Asn} displays 3–11 nucleotides. In contrast to other tRNA species, no atypical secondary structures were found in the mitochondrial tRNA^{Asn}. The D-loop of *Methanobacterium thermautotrophicus* tRNA^{Asn} contains 10 nucleotides; it is deprived of D residues like other archaeal tRNA^{Asn} but contains the modified acp3-uridine.[1]

Two unusual modifications are found in tRNA^{Asn}: N^6-threonylcarbamoyladenosine (t^6A) is present at position 37 of almost all tRNA^{Asn} and queuine at position 34 of eukaryotic and most eubacterial tRNA^{Asn} resulting from exchange of the guanine residue catalyzed by the tRNA-guanine transglycosylase.[143] *In vitro* protein synthesis experiments conducted with yeast tRNA^{Asn} with or without queuosine at position 34 have shown that absence of the modified base promotes efficient frameshifting in eukaryotes.[144] Certain eubacteria such as *T. thermophilus* and all archaea are unable to synthesize queuine and thus contain G34.[145] Archaeal tRNA^{Asn} contains archaeosine at position 15, an unknown modification at position 34, m^5C at position 48 and/or 49, 1-methylpseudouridine at position 54, pseudouridine at positions 54 and/or 55, Cm at position 56, and 1-*O*-methylinosine at position 57.[146,147] Like all other tRNAs sequenced, *T. thermophilus* tRNA^{Asn} contains three invariant posttranscriptional modifications: Gm18, s^2T54, and m^1 A58.[145,148–150] Combination of these modifications is not found in other prokaryotic tRNAs. It has been shown that they reinforce the thermostability of the tRNAs of *T. thermophilus* by increasing the T_m by about 3 °C.[148,151] *T. thermophilus* tRNA^{Asn} contains m^2G at positions 6 and 10. m^2G10 and m^2_2G26 are found in eukaryotic tRNA^{Asn} and m^1G9 in mammalian tRNA^{Asn}.[1]

5.14.4.1.4(ii) Specificity of tRNA asparaginylation by AsnRS Cross-reactions using $tRNA^{Asn}$ and AsnRSs of distinct origins reveal species specificity of tRNA asparaginylation. Human and *Brugia malayi* AsnRSs efficiently charge the eukaryotic $tRNA^{Asn}$ from mammals and yeast but only poorly the prokaryotic $tRNA^{Asn}$ from *E. coli*.[152–154] In contrast, *T. thermophilus* AsnRS aspartylates efficiently prokaryotic $tRNA^{Asn}$ from *T. thermophilus* and *E. coli* but only poorly the eukaryotic $tRNA^{Asn}$ from yeast.[155] It has been proposed that the species specificity of the AsnRSs is related to the supernumerary nucleotide inserted at position 21 of the D-loop of eukaryotic $tRNA^{Asn}$.[154]

5.14.4.1.4(iii) Identify of $tRNA^{Asn}$ Because bacterial $tRNA^{Asn}$ starts with 5′ pU, which prevents efficient transcription of the gene with T_7 RNA polymerase, the identity of $tRNA^{Asn}$ could not be investigated easily by analysis of *in vitro*-synthesized mutated transcripts. Therefore, the identity determinants were characterized by analysis of *E. coli* $tRNA^{Asn}$ variants expressed *in vivo* and of variants of *E. coli* $tRNA^{Lys(UUU)}$ transcripts synthesized *in vitro*.[156,157] More recent kinetic analysis of the *T. thermophilus* $tRNA^{Asn}$ transcript starting with the 5′U, formed by self-cleavage catalyzed by the hammerhead ribozyme attached at the 5′ end, showed that posttranscriptional modifications exert only a minor effect on the accepting capacity, as the asparaginylation efficiency of the transcript is decreased only 3.5-fold compared to that of the native $tRNA^{Asn}$ (k_{cat} and K_m 0.4 s^{-1} and 48 nmol l^{-1} for the modified tRNA and 0.7 s^{-1} and 290 nmol l^{-1} for the transcript).[155]

In vivo investigations of the charging capacity of tRNA variants showed that the anticodon and the discriminator G73 are essential in $tRNA^{Asn}$ identity. Random mutagenesis of yeast $tRNA^{Asn(CUA)}$ creates variants charged by GlnRS.[158] Conversion of $tRNA^{Asn}$ into amber and opal suppressors by change of the anticodon inactivates the tRNA.[159] However, the $tRNA^{Asn}$ amber suppressor already containing the Gln identity elements acquired Gln specificity by disruption of the first base pair from the acceptor stem, which is known to be essential for tRNA glutaminylation.[158] The essential role of G34 was revealed by a switch of the specificity of $tRNA^{Lys}$ transplanted with both the anticodon and the discriminator base of $tRNA^{Asn}$. The $tRNA^{Lys(GUU),A73\rightarrow G}$ variant is well charged by AsnRS.[160] Because $tRNA^{Lys(UUU),A73\rightarrow G}$ does not accept Asn, G34 plays a crucial role in Asn identity. Furthermore, substitution in $tRNA^{Lys(UUU),G73}$ of G73 by any other nucleotide decreases asparaginylation efficiency, indicating that the discriminator base is also important. *In vivo* analysis of the charging capacity of $tRNA^{Asn}$ variants showed that U35→C and U36→C mutations abolish the charging capacity of $tRNA^{Asn}$, whereas the G34→C substitution confers Lys-accepting capacity to the mutated tRNA.[156] The chimeric $tRNA^{Asn(CUU),A73}$ is deprived of asparaginylation capacity and switches *in vivo* to lysylation specificity. However, the role of these nucleotides also seems to depend on the structural context of the tRNA because $tRNA^{fMet}$ containing the Asn anticodon and G73 is unable to initiate polypeptide chain synthesis starting with an Asn codon.[156]

5.14.4.1.4(iv) Elements of AsnRS involved in the recognition of $tRNA^{Asn}$ No crystal structure of an AsnRS complexed to $tRNA^{Asn}$ is currently available. Alignments of AsnRS and AspRS polypeptide chains show conservation in AsnRS of the residues of AspRS contacting U35 and G73 of $tRNA^{Asp}$. In the 3D structure of the yeast AspRS·$tRNA^{Asp}$ complex, U35 contacts Arg119, Phe127, and Gln138, and G73 contacts Glu327.[161,162] Implication of these residues in tRNA recognition was confirmed by site-directed mutagenesis.[163] Because U35 and G73 are conserved in $tRNA^{Asn}$ and contribute to efficient tRNA asparaginylation,[164] these nucleotides are probably contacted by the conserved aa residues from AsnRS. The docking model of *P. horikoshii* AsnRS and $tRNA^{Asn}$ shows that Glu89, Phe36, and Gln47 interact with G34 and U35 from $tRNA^{Asn}$, suggesting that the recognition pattern of the first and the second nucleotides of the anticodon is conserved between AsnRS and AspRS. The third nucleotide, U36, an essential identity element that distinguishes $tRNA^{Asn}$ from $tRNA^{Asp}$, is contacted on its N3 and O4 groups by the carbonyl and guanidine groups of Arg83 by hydrogen bonds. Substitution of Arg63 by Ala disrupting the hydrogen bonds with U36 drastically decreases aminoacylation efficiency.[139]

5.14.4.1.4(v) Discrimination by AsnRS of $tRNA^{Asn}$ against $tRNA^{Asp}$ and $tRNA^{Lys}$ Because the prevalent identity elements in $tRNA^{Asn}$, that is, the anticodon and the discriminator bases, are partly conserved in $tRNA^{Asp}$ and $tRNA^{Lys}$, one may wonder how AsnRS discriminates $tRNA^{Asn}$ against $tRNA^{Asp}$ and $tRNA^{Lys}$. Among the nucleotides determining aspartate and asparagines identities, only nucleotide 36 differs (C in

$tRNA^{Asn}$ and U in $tRNA^{Asp}$). Because substitution in *E. coli* $tRNA^{Asn}$ of U36 by C drastically decreases the efficiency of asparaginylation,[160] C36 prevents recognition of $tRNA^{Asp}$ by AsnRS. In contrast, substitution in $tRNA^{Asp}$ of C36 by U affects tRNA aspartylation only moderately, suggesting that U36 in $tRNA^{Asn}$ does not efficiently prevent aspartylation by AspRS.[87] Specific asparaginylation of $tRNA^{Asn}$ by AsnRS may be reinforced by the posttranscriptional modification t6A that should prevent aspartylation. Among the nucleotides involved in the identity of $tRNA^{Asn(GUU)}$ and $tRNA^{Lys(UUU)}$, only nucleotide 34 (G and hypermodified U respectively) differs in cytosolic eukaryotic tRNAs whereas both nucleotide 34 and the discriminator base (G73 and A73 respectively) differ in prokaryotic tRNAs.[157,165–167] Thus, in eukaryotic $tRNA^{Lys(UUU)}$, the modified U34 probably solely prevents asparaginylation, whereas in prokaryotic $tRNA^{Lys(UUU)}$, in addition to the modified U34 the discriminator base A73 may also prevent asparaginylation. Finally, it has been suggested that the strong contribution of U35 and U36 to Asn identity determines the phylogenetic distribution of the two distantly related class I LysRSs that recognize U35 and U36 in organisms possessing AsnRS such as *Borrelia burgdorferi* or U36 alone in organisms lacking AsnRS such as *Methanococcus maripaludis*.[168]

5.14.4.2 The Indirect Pathway of tRNA Asparaginylation

5.14.4.2.1 Discovery of the indirect pathway of tRNA asparaginylation and phylogenetic distribution

Biochemical investigations and analysis of prokaryotic genomes revealed the absence of GlnRS and AsnRS in various bacteria and archaebacteria.[142,169] All archaea are deprived of GlnRS and nearly 50% of AsnRS, whereas 80% of the bacteria do not contain GlnRS and 50% lack AsnRS, although their proteins contain Gln and Asn. It was shown that in these cases Gln-$tRNA^{Gln}$ and Asn-$tRNA^{Asn}$ are formed by a two-step process involving conversion of the aa mischarged on the orphan tRNA. The indirect pathway of Asn-$tRNA^{Asn}$ formation was first suggested in the archaea *Haloferax volcanii* deprived of AsnRS where biochemical investigations have shown that Asn-$tRNA^{Asn}$ is formed by conversion of Asp mischarged on $tRNA^{Asn}$.[170] The nonconventional pathway of Asn-$tRNA^{Asn}$ formation was deciphered in *T. thermophilus*.[87] This thermophilic bacterium, deprived of the Asn synthetase, is unable to form free Asn and thus depends on exogenous Asn to form Asn-$tRNA^{Asn}$ by AsnRS. It was demonstrated that this organism forms Asn-$tRNA^{Asn}$ indirectly through a pathway involving two peculiar enzymes: an ND-AspRS that aspartylates $tRNA^{Asn}$ and a tRNA-dependent amidotransferase (AdT) that converts the Asp mischarged on $tRNA^{Asn}$ into Asn by amidation to form the homologous Asn-$tRNA^{Asn}$ [118,171] (**Figure 7**). This pathway resembles the indirect pathway of Gln-$tRNA^{Gln}$ formation used by archaea and bacteria to compensate the absence of GlnRS[83,84] (**Figure 8**). In the eukaryotic cytosol, both Asn-$tRNA^{Asn}$ and Gln-$tRNA^{Gln}$ are formed by direct charging of the tRNA with the cognate aaRSs, AsnRS and GlnRS, respectively. In mitochondria and chloroplasts, Asn-$tRNA^{Asn}$ is also formed directly by AsnRS, whereas the formation of Gln-$tRNA^{Gln}$ occurs through the indirect pathway.[88,89] Indirect tRNA asparaginylation and glutaminylation involve the formation of Asp-$tRNA^{Asn}$ and Glu-$tRNA^{Gln}$ intermediates whose use for protein synthesis would be lethal. It has been demonstrated that the heterologous aa-tRNAs do not bind the elongation factor EF-Tu and thus are not carried to the ribosomes.[87,172]

In organisms that use the indirect pathway to form Asn-$tRNA^{Asn}$, the AspRS exhibits a dual specificity by aspartylating $tRNA^{Asn}$ as efficiently as the cognate $tRNA^{Asp}$ (3.3 and 1.5 $s^{-1}\,\mu mol^{-1}\,l$, respectively). Alignments show that the AspRSs can be grouped into three structurally distinct families that superimpose with their phylogenetic distribution.[145,173] The AspRSs of the bacterial type of the largest size (subunit M_r, about 65 kDa) contain additional domains not found in other AspRSs, in particular the ferredoxinlike or GAD domain inserted in the catalytic core between consensus motifs 2 and 3 and a C-terminal extension. The eukaryotic AspRSs (subunit M_r 60–65 kDa) contain an N-terminal extension upstream of the anticodon-binding domain, which, in the yeast enzyme, contacts the tRNA by the minor groove side of the anticodon stem and increases the stability of the complex with tRNA and the global aminoacylation efficiency.[174,175] Finally, the archaeal-type AspRSs of a minimalist structure are formed by the catalytic and the anticodon-binding domains deprived of any additional domain and belong to the smallest AspRSs (subunit M_r, about 50 kDa).

AspRSs of dual specificity involved in tRNA asparaginylation have first been characterized in archaea and in the Deinococcales (*T. thermophilus* and *D. radiodurans*). Interestingly, these eubacteria contain two AspRSs: the

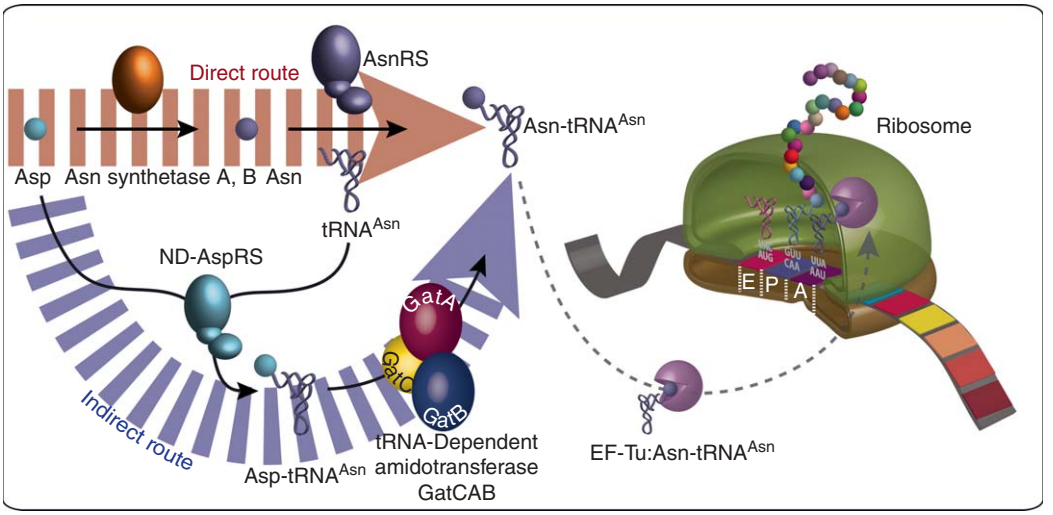


Figure 7 The direct and indirect pathways of tRNA asparaginylation. The direct pathway consists of charging by AsnRS on $tRNA^{Asn}$ of free Asn formed with asparagine synthetase A or B. The Asn-$tRNA^{Asn}$ binds the EF-Tu factor in bacteria (or EF-1A in eukaryotes and archaea) to be carried to the ribosome. In the indirect pathway, a nondiscriminating AspRS (ND-AspRS) charges Asp on $tRNA^{Asn}$; Asp-$tRNA^{Asn}$ does not bind the elongation factor but is converted by the tRNA-dependent trimeric amidotransferase GatCAB into Asn-$tRNA^{Asn}$, which binds the EF-Tu factor and is carried to the ribosome where it is used for polypeptide chain elongation.

D-AspRS (AspRS1) of bacterial type and of strict specificity for $tRNA^{Asp}$ charging and the ND-AspRS (AspRS2) of archaeal type and of relaxed specificity for $tRNA^{Asp}$ and $tRNA^{Asn}$ charging.

The first biochemical investigations have shown that archaea are deprived of AsnRS, whereas bacteria contain AsnRS. The analysis of the first genomes sequenced was in agreement with this observation, as the archaeal genomes lacked the gene encoding AsnRS whereas the bacterial genomes contained it. Thus, it was believed that Asn-$tRNA^{Asn}$ is formed in archaea by the indirect pathway and in bacteria, as in eukaryotes, by the direct pathway. As a consequence, archaeal-type AspRSs should exhibit relaxed specificity and aspartylate $tRNA^{Asp}$ and $tRNA^{Asn}$, and bacterial AspRSs, like the eukaryotic AspRSs, should have a strict specificity and aspartylate only $tRNA^{Asp}$. Comparison of the 3D structures of the AspRSs from yeast and the archaea *P. kodakaraensis* suggested the structural bases determining the strict and the relaxed specificities of the AspRSs. Indeed, the 3D structure of yeast AspRS complexed to $tRNA^{Asp}$ shows implication of the L1 loop from the anticodon-binding domain in the recognition of C36 from $tRNA^{Asp}$ anticodon through backbone contacts.[161,162] In contrast, modelization of the complex of *P. kodakaraensis* AspRS with $tRNA^{Asp}$ revealed absence of contact of the L1 loop with nucleotide 36 from tRNA.[176] Interestingly, archaeal-type AspRSs differ from other AspRSs by the size of this loop (residues 173–186 from alignment), which consists of 14–16 residues in bacterial and eukaryal AspRSs but only 5–9 in archaeal AspRSs.[145] Because $tRNA^{Asp}$ contains C and $tRNA^{Asn}$ contains U at position 36, it has been proposed that the strict and the relaxed specificities of bacterial and archaeal AspRSs are determined by the size of the L1 loop surrounding tRNA nucleotide 36 in the complex. The large loop of eukaryal and bacterial AspRSs would determine strict specificity by contacting C36 of $tRNA^{Asp}$, whereas the small loop of archaea AspRSs able to recognize C36 and U36 of $tRNA^{Asp}$ and $tRNA^{Asn}$ equally would determine relaxed specificity. However, this assumption was refuted by functional investigations of a large variety of AspRSs of bacterial and archaeal origins that showed that strict and relaxed specificities are found in both types. The bacterial AspRSs of *H. pylori*[177] and *Chlamydia trachomatis*[178] display relaxed specificity despite their large L1 loop, as both organisms are deprived of AsnRS and use the indirect pathway to form Asn-$tRNA^{Asn}$. In contrast, archaeal AspRSs from *P. kodakaraensis* and *Ferroplasma acidarmanus*[179] exhibit strict specificity and charge only $tRNA^{Asp}$ despite the presence of a small L1 loop, as both organisms contain an AsnRS that is able to form Asn-$tRNA^{Asn}$ directly.

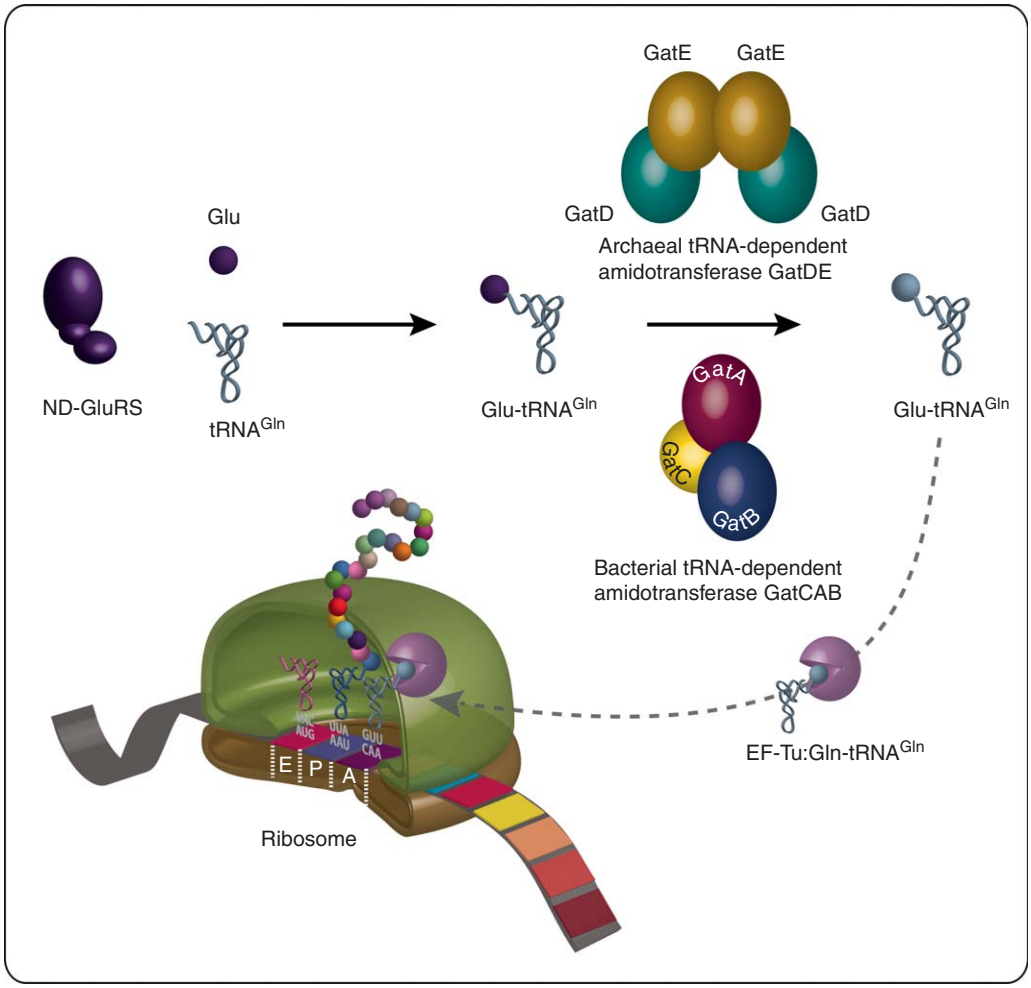


Figure 8 The indirect pathway of tRNA glutaminylation. The nondiscriminating GluRS (ND-GluRS) charges free Glu on $tRNA^{Gln}$. The mischarged Glu-$tRNA^{Gln}$ does not bind the elongation factor, but is converted into Gln-$tRNA^{Gln}$ by a tRNA-dependent amidotransferase, the trimeric GatCAB in bacteria or the dimeric GatDE in archaea. Gln-$tRNA^{Gln}$ binds the elongation factor EF-Tu in bacteria and EF-1A in archaea, and is carried to the ribosome where it is used for polypeptide chain elongation.

5.14.4.2.2 Structure–function relationship of ND- and D-AspRSs

More precise information concerning the structural basis determining the strict and the relaxed specificities of AspRSs was obtained after resolution of the 3D structure of *T. thermophilus* ND-AspRS (AspRS2).[180] This AspRS presents nonconventional features in the OB-fold of its anticodon-binding domain, that is, the absence of an Hα-helix, which in canonical OB-folds is inserted between the S3 and S4 β-strands. This insertion, without peculiar amino acid composition, is conserved in almost all known AspRSs. In the ND-AspRS2, it forms the Lα loop of eight residues. Interestingly, the ND-AspRS2 from *D. radiodurans* also lacks the Hα-helix. In contrast, the archaeal D-AspRS from *P. kodakaraensis* contains an 18-residue-long insertion that forms an α-helix. Comparison of the 3D structures of *T. thermophilus* AspRS2 and *P. kodakaraensis* AspRS reveals that the two AspRSs differ in the orientation and the sequence of their L1 loop joining the S4 and S5 strands[180] and provide an insight into a possible role of this loop in tRNA recognition. The L1 loop is located opposite the Lα loop or the Hα-helix in the OB-fold, and both subdomains are linked through strand S4. The fact that the Lα loop is short compared to the Hα-helix introduces a structural constraint that may mediate different orientations of the L1 loop in *T. thermophilus* AspRS2 and *P. kodakaraensis* AspRS through the S4 strand. However, the

absence of the Hα-helix does not obligatorily imply a relaxed specificity of AspRS, as since sequence alignments suggest the presence of this helix in some archaeal ND-AspRSs.[180]

The crystal structures of D-AspRS·tRNA^{Asp} complexes show interaction of the L1 loop with determinant C36 of tRNA^{Asp}.[124,162,181,182] Three residues of the extended loop of the *T. thermophilus* D-AspRS1 (Asn82, Arg78, Glu80) contact three different atoms from C36 of tRNA^{Asp}.[181] Analysis of the Asp identity elements of tRNAs of various origins has shown that C36 is a universal identity determinant,[164,183,184] and mutation in yeast AspRS of the aa contacting this residue strongly affects tRNA^{Asp} recognition.[163] The distinct discrimination properties of *T. thermophilus* AspRS2 and *P. kodakaraensis* AspRS may thus be related to distinct conformations of the L1 loops of both AspRSs which do not superimpose,[180] determining distinct recognition patterns of nucleotide 36. Interestingly, the L1 loop of *T. thermophilus* AspRS2 superimposes perfectly with that of *T. thermophilus* AsnRS, but strong sequence and conformation variations occur with the extended loops of the D-AspRSs from yeast, *E. coli*, and *T. thermophilus*, which are very similar and superimpose.[180]

A closer analysis of the sequences of L1 loops has shown that nondiscrimination in tRNA recognition is accompanied by conservation of a Pro residue in both the short and the long L1 loops, whereas specific tRNA recognition relies on the absence of such residue in short loops.[180] For instance, Pro72 in *T. thermophilus* AspRS2 corresponds to a Lys residue in *P. kodakaraensis* AspRS and in other discriminating archaeal AspRSs. The presence or the absence of a Pro residue at position 72 imposes different stiffnesses to the L1 loop. In *P. kadakaraensis* AspRS, the difference may even be enhanced by the presence of a Pro residue of the opposite site of the loop at position 68. Interestingly, *in vivo* selection of variants of *D. radiodurans* ND-AspRS2 that conserved the capacity to aspartylate tRNA^{Asn} showed that substitution of Pro77 by Cys, Ile, Leu, Lys, Phe, Ser, or Val results in the loss of the capacity to aspartylate tRNA^{Asp} but does not affect the aspartylation capacity of tRNA^{Asn}.[169] Furthermore, the His28Gln substitution increased threefold the specificity for tRNA^{Asp} over that of tRNA^{Asn}. These results strongly suggest that residues Pro77 and His28 determine the ND-tRNA recognition properties of this archaeal-type AspRS. A similar conclusion derives from mutational investigation of the ND-AspRS from *Pseudomonas aeruginosa*. It was shown that residue His31 from conserved motif RRRDH/L contained in the first β-barrel and residue Gly83 conserved in the L1 loop of almost all ND-AspRSs determine relaxed specificity, as the His31Leu and Gly83Lys variants of *P. aeruginosa* AspRS aspartylate tRNA^{Asp} respectively five- and sixfold more efficiently than tRNA^{Asn}, whereas the wild type charges both tRNAs with similar efficiencies.[185] These observations agree with the capability of the mutated ND-AspRSs from *P. aeruginosa* to complement the *E. coli* strain expressing a thermosensitive mutant of AspRS, whereas expression of the wild-type enzyme was lethal because of the depletion of free tRNA^{Asn} mischarged with Asp.[185] Likewise, the toxicity of the wild-type ND-AspRS from *H. pylori* in *E. coli* is higher than that of its Leu81Asn and Leu86Met variants, which are 1.5- and 2-fold more specific, respectively, than the wild-type AspRS for tRNA^{Asp} than for tRNA^{Asn}.[177]

The conservation of the conformation of the L1 loop in the ND-AspRS2 and in AsnRS, and the fact that the anticodon residues C36 in tRNA^{Asp} and U36 in tRNA^{Asn} are aspartate and asparagine identity determinants, respectively, suggests a functional role of the L1 loop in AspRS2. This view is supported by the fact that the U36 determinant for tRNA asparaginylation contains two chemical groups, O2 and N3, also present in C36 recognized in tRNA^{Asp} by the D-AspRSs[162] and by the acquisition by the *P. kodakaraensis* AspRS of the capacity to charge tRNA^{Asp} and tRNA^{Asn} with equivalent efficiencies when transplanted with the L1 loop of *T. thermophilus* AspRS2.[180] However, the slightly decreased activity for tRNA^{Asn} charging suggests that the discrimination process is not solely mediated by the L1 loop and may be tuned through indirect effects by additional regions of AspRS such as the Hα/Lα domain. This is in agreement with the fact that the variant of *T. thermophilus* AspRS2 transplanted with the L1 loop of *P. kodakaraensis* did not acquire the strict specificity.[180]

5.14.4.2.3 Recognition of tRNA^{Asp} and tRNA^{Asn} by the ND-AspRS

Because D-AspRS and AsnRS select tRNA^{Asp} and tRNA^{Asn}, respectively, they recognize distinct elements in each tRNA whereas ND-AspRS that aspartylates both tRNAs recognizes elements common to the two tRNAs. Most tRNA^{Asp} identity elements are conserved in the various species[186]; recognition of tRNA^{Asp} by AspRS occurs by a similar pattern.[164] Study of the tRNA elements determining aspartylation by the D- and ND-AspRSs from *T. thermophilus*, AspRS1 and AspRS2 respectively, has shown that with only one exception the same nucleotides determine aspartylation by the two AspRSs.[142] Specificity of each AspRS is related to distinct

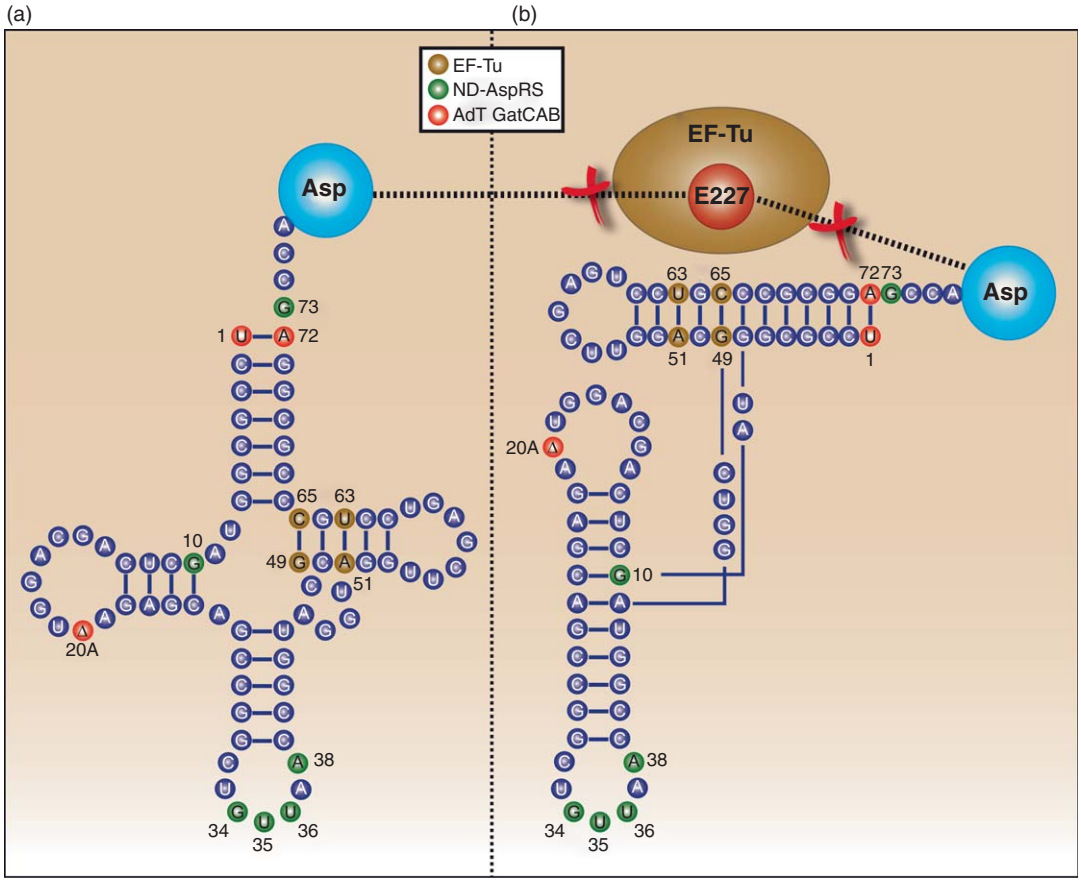


Figure 9 The elements of tRNA^{Asn} determining recognition by the ND-AspRS, the amidotransferase GatCAB, and the EF-Tu factor. (a) Cloverleaf structure of tRNA^{Asn}. (b) The 3D structure of tRNA^{Asn}. The elements of the tRNA^{Asn} determining recognition by the ND-AspRS are on a green background. The elements determining recognition of Asp-tRNA^{Asn} by GatCAB are on a red background; the U1–A72 pair constitutes the identity element with a prevalent contribution of the U1 residue; recognition requires the lack of nucleotide U20A (Δ) that is present in tRNA^{Asp} to prevent recognition of Asp-tRNA^{Asp}. The tRNA elements determining the binding capacity on the elongation factor EF-Tu, the G49–C65, and A51–U63 base pairs are in brown. Base pairs G49–U65 and G51–C63 in tRNA^{Asp} prevent binding of EF-Tu. Residue Glu227 of the protein determines the poor affinity of the Asp moiety of aa-tRNA as indicated by the crosses.

contributions of the tRNA elements to aspartylation efficiency. The three major elements that determine charging by the D-AspRS are U35 and C36 from the anticodon followed by the discriminator base G73. This triad is followed by the G2–C71 pair from the acceptor stem, the first anticodon base G34, and finally C38. The hierarchy of contribution of these elements differs for aspartylation by the ND-AspRS. Contribution of U35 immediately precedes that of G73, which is followed by G34, C38, and finally C36 (**Figure 9**). Thus, specific aminoacylation of tRNA^{Asp} by the D-AspRS is likely promoted by the nucleotide C36 from the anticodon and by the G2–C71 pair. In tRNA^{Asn}, C36 is substituted by U36 and the G2–C71 pair by C1–G71. Thus, the relaxed specificity of the ND-AspRS is due to the minimized contribution of nucleotide 36 in tRNA recognition and to the absence of recognition of the G2–C71 pair.

5.14.4.2.4 The tRNA-dependent bacterial amidotransferase GatCAB

5.14.4.2.4(i) Functional properties of the bacterial GatCAB The tRNA-dependent Gln synthetase activity was discovered in *B. megaterium* and *B. subtilis* where it was shown that the tRNA cognate with Gln codons is charged with Glu by the GluRS and that the bound Glu is then converted into Gln to form Gln-tRNA^{Gln}.[83] This pathway of Gln-tRNA^{Gln} formation is however not confined to prokaryotes. It was shown 20 years later

that a similar route forms Gln-tRNA^{Gln} in chloroplasts.[88] This pathway probably occurs also in *Saccharomyces cerevisiae* mitochondria, as the tRNA that reads Gln codons is charged directly with Glu.[187] The activity that was able to convert Glu mischarged on tRNA^{Gln} to Gln was first studied in *B. subtilis*,[83] and the enzyme was partially purified later from *Euglena gracilis* chloroplasts.[88] The genes encoding the heterotrimeric enzyme in *B. subtilis* were cloned and sequenced. The enzyme, overexpressed in *E. coli*, was purified and structurally and functionally characterized. This enzyme, formed by the association of the GatA, GatB, and GatC polypeptide chains, was called GatCAB.[86]

After the discovery of the transamidation pathway forming Asn-tRNA^{Asn} in *T. thermophilus*, [87] ORFs encoding the GatA, GatB, and GatC polypeptide chains were sought in the *T. thermophilus* HB 8 strain and identified by alignments with the chains from *B. subtilis*. Polypeptide chains presenting, respectively, 46.2, 45.4, and 27.1% of identity with GatA, GatB, and GatC from *B. subtilis* were characterized.[142] The three genes were cloned into an artificial operon. The trimeric protein expressed in *E. coli* catalyzed tRNA-dependent conversion of Asp into Asn. The important sequence conservation of the enzymes from *B. subtilis* and *T. thermophilus*, despite their distinct activities, prompted the investigation of their functional properties *in vitro*. It was shown that both enzymes exhibit dual transamidation specificity. The enzyme from *B. subtilis*, which forms Gln-tRNA^{Gln} *in vivo*, catalyzes the formation of both Gln-tRNA^{Gln} and Asn-tRNA^{Asn} *in vitro* and GatCAB from *T. thermophilus*, which is dedicated to Asn-tRNA^{Asn} formation *in vivo*, catalyzes the formation of the two aa-tRNAs *in vitro*.[118] Dual specificity of bacterial GatCAB *in vitro* was generalized. Indeed, GatCAB from *D. radiodurans*[92] and *P. aeruginosa*,[188] which forms Asn-tRNA^{Asn} *in vivo*, forms both aa-tRNAs *in vitro*. Likewise, the mitochondrial GatCAB from *Arabidopsis thaliana*,[89] which forms Gln-tRNA^{Gln} *in vivo*, catalyzes the formation of the two aa-tRNAs *in vitro*. Finally, the enzymes from *A. ferrooxidans*,[116] *C. trachomatis*,[178] and *H. pylori*,[189] which are involved in the *in vivo* synthesis of Asn-tRNA^{Asn} and Gln-tRNA^{Gln}, also catalyze the *in vitro* formation of the two aa-tRNAs. Thus, a unique GatCAB is involved in the formation of Asn-tRNA^{Asn} and Gln-tRNA^{Gln}. Its cellular function is dictated by the biochemical and genetic backgrounds of the host. In the absence of AsnRS, GatCAB supplies the organism with Asn-tRNA^{Asn} by amidation of the Asp-tRNA^{Asn} formed by the ND-AspRS; in the absence of GlnRS, the enzyme forms Gln-tRNA^{Gln} by amidation of Glu-tRNA^{Gln} formed by the ND-GluRS; finally, in the absence of both GlnRS and AsnRS, GatCAB, together with ND-AspRS and -GluRS, provides the organism with both aa-tRNAs.[142]

Despite identical functional properties, important divergences are observed in the genomic organization of the genes encoding GatA, GatB, and GatC. In *B. subtilis*, *A. ferrooxidans,* and *C. trachomatis* they are organized in an operon, whereas in *T. thermophilus* and *D. radiodurans* they are dispersed in the genome.[142]

Alignment of the GatCAB polypeptide chains shows two interesting features in GatA. First, the polypeptide chain contains the signature sequence of amidases rich in Gly, Ser, and Ala residues involved in the metabolic pathways of various amino acids (Arg, Pro, Phe, Trp) in prokaryotes such as *Rhodococcus*, *Brevibacterium*, and *Pseudomonas*.[86,93] Interestingly, the Asp and Ser residues of these amidases are present within the active site sequences of aspartic proteinases, suggesting an evolutionary relationship between the two enzymes.[190] Analogues of glutamyl-γ-boronate, an analogue of Gln that reacts with Ser residues, inhibit the growth of various bacteria that use the transamidation pathway to form Gln-tRNA^{Gln}, such as *Streptococcus pyogenes*, *S. pneumoniae*, and *Enterobacter faecalis*,[191] whereas these analogues are without effect on the growth of bacteria deprived of GatCAB that use the direct pathway to form Gln-tRNA^{Gln} and Asn-tRNA^{Asn}. Furthermore, *in vitro* analysis shows inhibition of the glutaminase and transamidase activities of GatCAB by these analogues. The crucial presence of a Ser residue in Gln deamidation by GatCAB is confirmed by the fact that Ser176Ala substitution in *S. pyogenes* GatA decreases the glutaminase- and Gln-dependent transamidase activities more than 300-fold but retains NH_3-dependent transamidase activity.[192] Second, GatA, when coexpressed with GatC, promotes ATP-independent deamidation of Gln and also contains a P-loop (residues 142–148), a structural motif involved in ATP and GTP binding, for which a function in GatCAB has yet to be identified;[86] GatB, structurally related to Pet112, an essential mitochondrial protein, may be involved in tRNA binding. Mechanistic studies have shown that in the absence of Glu-tRNA^{Gln}, GatCAB hydrolyzes Gln to Glu, while it uses the released NH_3 to amidate tRNA^{Gln}-bound Glu in concert with ATP hydrolysis when it is present.[193] In the absence of Glu-tRNA^{Gln}, the enzyme exerts a basal glutaminase activity that is unaffected by ATP ($k_{cat} = 0.019$ and $0.017\,s^{-1}$). Glu-tRNA^{Gln} activates the glutaminase activity about 10-fold ($k_{cat} = 0.019$ and $0.14\,s^{-1}$), and ATP increases this activation sevenfold (0.14 and $0.96\,s^{-1}$). The increased activities result mainly

from increased k_{cat} without significant effects on the K_m for Gln. Analogues of ATP are unable to stimulate the glutaminase activity, except ATP-γS, which stimulates it at the same level as ATP, but without significant hydrolysis and without concomitant transamidase activity. Glu-tRNA^{Gln} slightly stimulates basal ATP hydrolysis, whereas full ATP hydrolysis activity is observed only in the presence of Gln and Glu-tRNA^{Gln}. These results agree with those obtained by sedimentation of the GatCAB from *Chlamydomonas reinhardtii* on glycerol gradient in the presence or absence of the ligands, which show that the enzyme forms a stable complex with Glu-tRNA^{Gln} only when ATP is present.[194] This suggests that stimulation of the glutaminase activity by ATP and Glu-tRNA^{Gln} results either from an allosteric effect promoted by binding of the ligands or from a structural change that accompanies ATP hydrolysis.

5.14.4.2.4(ii) Structural properties of the bacterial GatCAB The 3D structure of GatCAB from *Staphylococcus aureus*, either in free form or complexed with Gln or Asn, and a nonhydrolyzable ATP analogue and Mn^{2+} ions, has brought new insights into the structure–function relationship of GatCAB.[60] The structure reveals that distinct catalytic centers for glutaminase and transamidase activities are markedly distant but are connected by a 30-Å-long channel. GatA consists of a central, mixed 11-stranded β-sheet core, covered on the top and the bottom by double layers of α-helices. This subunit faces its loop-rich plane to, and caps a loop-rich side of, the cradle domain of GatB, burying 7% of the total surface area in the subunit interface. GatC wraps around the interface region as a belt and makes extensive interactions with GatA and GatB. Two amphiphilic N-terminal α-helices of GatC form a helical bundle with the hydrophobic core of GatA, relieving local hydrophobicity, whereas the internal loop region crosses over the loop-rich side of GatB stabilized through a hydrogen-bond network involving the invariant Arg64 and Asp66 residues. The association of GatA and GatB is further stabilized by hydrophobic interactions of an α-helix of GatC with GatA and GatB and the two C-terminal β-strands that form an antiparallel four-stranded sheet with a β-hairpin of GatB. A putative ammonia channel runs through the middle of the interface.

In the cocrystal, Gln is found in the catalytic center of GatA that contains the amidase signature sequence, suggesting that GatA uses the same mechanism of hydrolysis as the amidases. Gln is bound by its amide group to Asp425 and by the carboxyl group to Arg358, whereas its side chain is located close to the conserved Ser178–*cis*-Ser154–Lys79 catalytic scissor. The catalytic Ser178 makes a tight covalent bond with the amide group, which suggests the formation of the tetrahedral covalent intermediate stabilized by the oxyanion hole formed by the nitrogen atoms of the backbone of Thr175, Gly176, Gly177, and Ser178 residues.

GatB is comprised of two domains connected by a 60-Å linker loop. The globular cradle N-terminal domain is topologically unique. The C-terminal domain is built with seven helices and an additional three-helix bundle similar to the Yqey protein. Interestingly, this protein of unknown function is found in free form in *B. subtilis* and appended to the anticodon-binding domain of *D. radiodurans* GlnRS where it significantly increases the affinity for tRNA^{Gln}.[59] The cocrystal of GatCAB with the ATP analogue ADP $-$ A1F_4^- reveals that the adenosine moiety dips into a hydrophobic pocket formed by Va16, Phe205, and Pro155. N1 and N6 interact with the invariant Ser196, the ribose OH groups with the backbone of a loop inserted between an α-helix and a β-strand, whereas the O4 ribose is recognized by the conserved Asn194. The β-phosphate is hydrogen bonded to the conserved Glu10 and an anchored water molecule. The cocrystal with $MnCl_2$ reveals two Mn^{2+} ions in the active site of GatB at a distance of 6.3 Å. One is bound as Mg^{2+} in the native enzyme, whereas the second is coordinated to a cluster of three conserved acidic residues, Glu10, Asp192, and Glu210, at an appropriate distance to interact with the β- and γ-phosphates of ATP. These observations suggest that GatB uses a two-metal-ion mechanism of catalysis resembling that used by Gln synthetases. It has been proposed that in the functional complex, one Mg^{2+} is coordinated to the γ-carboxyl group of Glu acylating tRNA^{Gln}, whereas the second Mg^{2+} interacts with the β-and γ-phosphates of ATP and, by polarizing the γ-phosphate, promotes the nucleophilic attack of the γ-Glu carboxylate to form a γ-phosphoryl-Glu-tRNA^{Gln}.[60] Activation of the aa carboxyl side chain by phosphorylation with ATP before amidation has been demonstrated for Gln-tRNA^{Gln} formation by *B. megaterium* GatCAB.[195]

In the 3D structure of GatCAB solved at 2.3 Å of resolution, the glutaminase active center and the binding site of the CCA end of tRNA are 30 Å apart. The two sites are connected by a channel, mostly hydrophobic in the outside, but lined inside with a succession of alternating strictly conserved positive and negative residues. The structure suggests a 'proton relay' mechanism that carries ammonia from one site to the other by repeated

protonations and deprotonations promoted by these residues. The proton donor and acceptor residues are Thr175 at the entrance and Lys79 at the exit of the channel. The crystalline structure reveals a continuous electron density sprouting out from the amide group of the Gln side chain, probably reflecting the ammonia released from the substrate. Because Gln hydrolysis depends stringently on the binding of glutamylated but not free or glutaminylated $tRNA^{Gln}$, it has been suggested that Glu-$tRNA^{Gln}$ induces a conformational change displacing residue Glu125, whose open–close motion constitutes a gate susceptible to tRNA binding. Two residues, Asp126 and Arg190, located at the entrance of the active site where the CCA end of tRNA approaches, form a salt bridge. It has been proposed that the CCA end of the incoming tRNA stimulates the movement of Arg190, which disrupts the salt bridge, resulting in the relaxation of an adjacent distorted β-strand and in the opening of the Glu125 gate.[60]

5.14.4.2.4(iii) tRNA recognition by the bacterial GatCAB The nucleotides of potential $tRNA^{Asn}$ determinants for amidation of the bound Asp were sought by characterizing the nucleotides conserved in $tRNA^{Asn}$ but absent in $tRNA^{Asp}$ in organisms using the transmidation pathway to form Asn-$tRNA^{Asn}$. Elements found in $tRNA^{Asn}$ were transplanted into $tRNA^{Asp}$ and those found in $tRNA^{Asp}$ into $tRNA^{Asn}$ before analysis of the capability of the tRNA variants to promote tRNA-dependent amidation of Asp.[119] It was shown that the first base pair U1–A72 of $tRNA^{Asn}$ and the absence of a surnumerary nucleotide in the D-loop determine amidation of the tRNA-bound Asp into Asn. Nucleotide U1 plays a prevalent role, as the $tRNA^{Asn}$ variant containing the U1–G72 base pair promotes transamidation. Amidation of Asp bound on $tRNA^{Asp}$ is prevented by the presence of the G1–C72 pair replacing the U1–A72 pair of $tRNA^{Asn}$ and by the surnumerary nucleotide U20A in the D-loop playing the role of an antideterminant. Substitution in $tRNA^{Asp}$ of the G1–C72 pair by U1–A72 and deletion of nucleotide U20A create a variant able to promote conversion by GatCAB of the bound Asp into Asn as efficiently as $tRNA^{Asn}$. In contrast, substitution in $tRNA^{Asn}$ of the U1–A72 base pair by G–C and introduction of U at position 20A prevent amidation of the bound Asp by GatCAB (**Figure 9**).

Analysis of the $tRNA^{Gln}$ and $tRNA^{Glu}$ sequences of organisms that form Gln-$tRNA^{Gln}$ by the indirect pathway shows conservation respectively of the elements that in $tRNA^{Asn}$ promote amidation of Asp and those that in $tRNA^{Asp}$ amidation of Asp, respectively, suggesting that the same tRNA elements determine specific amidation of bound Asp and Glu.[119] This prediction has been confirmed by mutational analysis of $tRNA^{Gln}$ and GatCAB from *S. aureus*.[60] Gel-shift assays show that GatCAB does not bind the anticodon of $tRNA^{Gln}$ and that exchange of the upper part of the acceptor stem of $tRNA^{Gln}$ with that of $tRNA^{Glu}$ drastically decreases its binding capacity for GatCAB, whereas exchange of the lower part preserved significant binding capacity. Substitution of only the first base pair U1–A72 by G1–C72 of $tRNA^{Glu}$ extinguishes the binding capacity of $tRNA^{Gln}$ to GatCAB. Insertion of U20B of $tRNA^{Glu}$ into $tRNA^{Gln}$ is lethal for tRNA binding. Thus, as for the tRNA-dependent Asp transamidation, the U1–A72 base pair constitutes a positive determinant for GatCAB to discriminate $tRNA^{Gln}$ from $tRNA^{Glu}$, whereas the U20B in the D-loop constitutes an antideterminant that promotes rejection of $tRNA^{Glu}$.

It is unclear how GatCAB discriminates the U1–A72 pair in $tRNA^{Gln}$ and $tRNA^{Asn}$ from the G1–C72 pair in $tRNA^{Glu}$ and $tRNA^{Asp}$. Because the 3D structure of the *E. coli* GlnRS·$tRNA^{Gln}$ complex shows that the base pair U1–A72 of $tRNA^{Gln}$ is disrupted by Leu136 located at the tip of the β turn in the acceptor-binding domain of GlnRS, use of a similar mechanism by GatCAB has been investigated. Interestingly, two-turn loops are good candidates to destabilize the A–U pair with a smaller free energy cost than the G–C pair. Furthermore, the C-terminal helical domain of GatB essential for $tRNA^{Gln}$ binding contains the Leu472 residue at the fourth position from the C-terminus. Deletion of this residue causes a loss of tRNA-binding capacity, whereas the C-terminus deletion mutants harboring the Leu472 residue did not impair binding capacity. It has been proposed that the absence of this Leu residue disrupts the configuration of the last three helices connected by a long flexible linker to a bundle of helices expected to bind the D-loop of $tRNA^{Gln}$. This helical bundle may discriminate $tRNA^{Gln}$ by probing the size and the configuration of the D-loop.[60]

5.14.4.2.5 The archaeal tRNA-dependent amidotransferase GatCAB

The first functional investigation of a GatCAB of archaeal origin suggested that this enzyme, like the bacterial one, possesses a dual specificity. Indeed, it has been shown that GatCAB from *M. thermautotrophicus* amidates *in vitro* Glu-$tRNA^{Gln}$ from *H. pylori* and *B. subtilis* and Asp-$tRNA^{Asn}$ from *M. thermautotrophicus*.[93] However, analysis of the transamidation reaction using $tRNA^{Asn}$ and $tRNA^{Gln}$, both originating from *M. thermautotrophicus*,

revealed that the archaeal enzyme forms Asn-tRNA^{Asn} but not Gln-tRNA^{Gln}.[94] This agrees with the observation that in archaea GatCAB is present only when AsnRS is absent, whereas in bacteria GatCAB is present whenever AsnRS or GlnRS or both are absent. Thus, the archaeal GatCAB is confined *in vivo* to Asn-tRNA^{Asn} formation, whereas Gln-tRNA^{Gln} is formed by the dimeric GatDE amidotransferase found exclusively in archaea.[93,169] Because the enzyme conserves the ability to transamidate Glu-tRNA^{Gln} from *H. pylori*, recognition of glutamylated archaeal tRNA^{Gln} is not prevented by the Glu residue acylating tRNA^{Gln} but by the antideterminants present in the archaeal tRNA^{Gln}.

In *Methanosarcina barkeri* and *M. thermautotrophicus*, tRNA^{Asn}, like tRNA^{Asp}, contains a G1–C72 pair in the first position. Thus, this base pair cannot promote specific transamidation of Asp-tRNA^{Asn} as in bacteria; in contrast to the bacterial GatCAB, the archaeal enzyme cannot discriminate aspartylated bacterial tRNA^{Asn} from tRNA^{Asp}. Indeed, cross-species transamidation experiments showed that GatCAB from *Neisseria meningitidis* is unable to promote amidation of Asp bound to archaeal tRNA^{Asn}, whereas GatCAB from *M. barkeri* amidates Asp bound to tRNA^{Asp} and tRNA^{Asn} from *N. meningitidis*, confirming that the tRNA elements determining the conversion of Asp-tRNA^{Asn} differ in bacterial and archaeal systems.[119] Alignments of archaeal tRNA^{Asp} and tRNA^{Asn} suggest that nucleotides G46 and U47 of the variable region of tRNA^{Asn} are involved in archaeal transamidation because nucleotide 46 is lacking in tRNA^{Asp}, and U47 is substituted by A. Site-directed mutagesis showed that the aspartylated yeast tRNA^{Asp} displaying the variable region of archaeal tRNA^{Asp} (absence of nucleotides 47 and A46) is not a substrate of archaeal GatCAB, whereas the aspartylated yeast tRNA$^{Phe\rightarrow Asp}$ variant displaying the Asp identity and nucleotides G46 and U47 in the variable region is a substrate.[119] These observations were confirmed by binding and transamidation experiments involving aspartylated tRNA^{Asn} and tRNA^{Asp} variants and GatCAB from *M. thermautotrophicus*[120] that showed that the archaeal GatCAB discriminates Asp-tRNA^{Asn} against Asp-tRNA^{Asp} by the use of U49, the D-loop, and, to a lower extent, the variable loop. Anticodon-swapping mutants had nearly the same binding affinity as the wild-type Asp-tRNA^{Asn} for GatCAB (0.6 μ mol l^{-1}), whereas T-stem-loop mutants had decreased affinity. Decreased transamidation efficiencies were observed with aspartylated tRNA^{Asn} variants in which positions in the T-arm and D-loop were replaced by the corresponding ones of tRNA^{Asp}. It was shown that A9 located between the acceptor and the T-stems, U47 in the variable loop and C56 in the T-loop, contribute to transamidation efficiency. Furthermore, nucleotide U49 of tRNA^{Asp} constitutes a major antideterminant for GatCAB recognition. In most archaea encoding a GatCAB, position 49 differs in tRNA^{Asp} and tRNA^{Asn}: when a purine is present in tRNA^{Asn}, tRNA^{Asp} has a pyrimidine and vice versa. Interestingly, position 49 is invariant between tRNA^{Asp} and tRNA^{Asn} in archaea encoding AsnRS and deprived of GatCAB. However, *Nanoarchaeum equitans* and *Methanopyrus kandleri* are exceptions, as this nucleotide is not conserved as an antideterminant in archaeal tRNA^{Asp}.[120]

5.14.4.2.6 Assembly of the partners of the indirect pathway of tRNA asparaginylation

Gel filtration, dynamic light scattering, and polyacrylamide gel shift experiments have shown that the partners of the transamidation pathway for tRNA asparaginylation from *T. thermophilus* assemble into a complex of a 1/2/2 (dimeric AspRS/GatCAB/tRNA^{Asn}) stoichiometry.[196] This complex of M_r 380 000 Da called transamidosome converts free Asp into tRNA-bound Asn in the presence of ATP and an amido group donor and is able to promote channeling of the charged aa acceptor end of tRNA^{Asn} from AspRS to the GatCAB active site. The complex is stable enough to be isolated under nonequilibrium conditions by gel filtration. Comparison of the properties of the partners in free form or inside the complex revealed the advantages conferred by their assembly (**Figure 10**). GatCAB binds tRNA^{Asn} bound on the AspRS 60-fold stronger than it binds free tRNA^{Asn} (K_D 0.6 and >10 μm, respectively). Kinetic investigations showed that when bound on the AspRS·tRNA^{Asn} complex, GatCAB increases the k_{cat} of tRNA aspartylation by AspRS eightfold (0.012 and 0.094 s^{-1} in the absence and presence of GatCAB, respectively). In contrast, AspRS decreases the rate of transamidation by GatCAB about 30-fold (3.4 and 0.11 s^{-1} in the absence and presence of AspRS, respectively). The pre–steady-state rate of Asn-tRNA^{Asn} formation inside the complex is determined by tRNA aspartylation (0.094 s^{-1}), whereas the steady-state rate is determined by the release of the Asn-tRNA^{Asn} triggered by dissociation of the ternary complex. The transamidosome promotes the formation of Asn-tRNA^{Asn} fourfold faster than the free partners (0.048 and 0.012 s^{-1}, respectively) (**Figure 10**). *In vivo* formation of the transamidosome has been investigated in an *E. coli* strain auxotrophic for Asn coexpressing the three partners of the indirect asparaginylation pathway of distinct origins, that is, the ND-AspRS of *D. radiodurans*, the GatCAB of *N. meningitidis*, and the

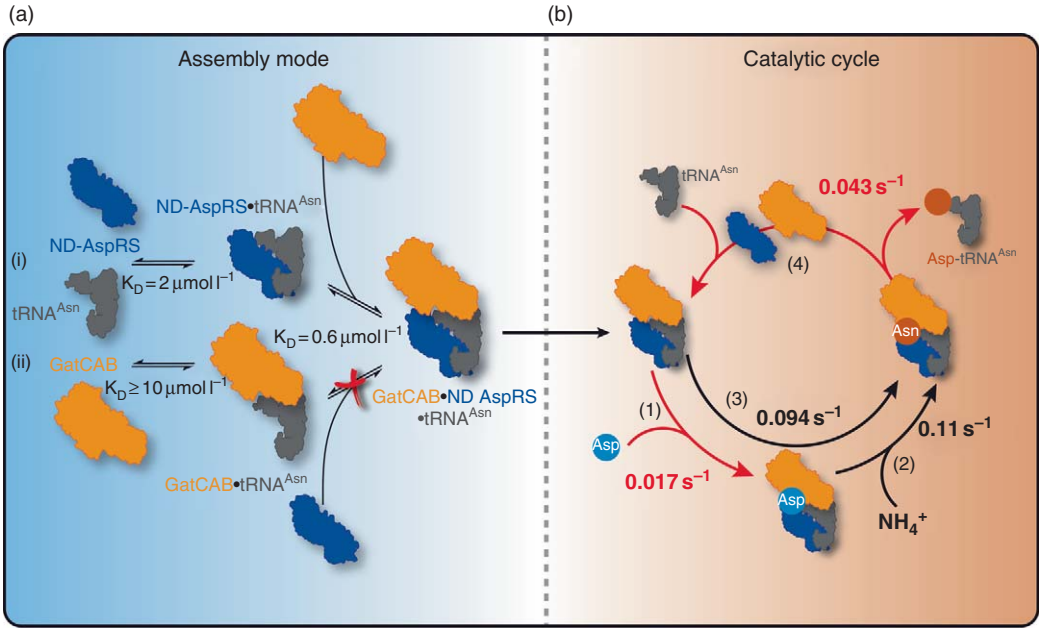


Figure 10 The transamidosome of *Thermus thermophilus* catalyzing tRNA asparaginylation. (a) Formation of the transamidosome (i) The AspRS (in blue) binds $tRNA^{Asn}$ (in gray, $K_D = 2\,\mu mol\,l^{-1}$) before association of the amidotransferase GatCAB (in orange, $K_D = 0.6\,\mu mol\,l^{-1}$) to form the ternary complex. (ii) The free GatCAB binds $tRNA^{Asn}$ with a poor affinity ($K_D \geq 10\,\mu mol\,l^{-1}$) before association of $tRNA^{Asn}$; thus pathway (i) is preferred for the formation of the transamidosome. (b) The catalytic cycle of the transamidosome. In the absence of free $tRNA^{Asn}$, the transamidosome aminoacylates $tRNA^{Asn}$ with a first-order rate constant of $0.017\,s^{-1}$ (1) and amidates the $tRNA^{Asn}$-bound Asp into Asn with a rate constant of $0.11\,s^{-1}$ (2). In the presence of an excess of free $tRNA^{Asn}$, the first Asn-$tRNA^{Asn}$ is formed with a rate constant of $0.094\,s^{-1}$ (3), whereas the following catalytic cycles occur with a rate constant of $0.043\,s^{-1}$ (4), indicating that dissociation of the newly formed Asn-$tRNA^{Asn}$ accompanied by the disruption of the complex is rate-limiting at the steady state.

$tRNA^{Asn}$ of *T. thermophilus*. Expression of the tRNA-dependent pathway of asparagine formation complemented the Asn auxotrophy of the strain, which acquired the ability to grow in the absence of free Asn. Gel filtration of the protein extract of this strain showed coelution of the three partners in a complex of increased M_r compared to those of the binary AspRS·$tRNA^{Asn}$ and GatCAB·$tRNA^{Asn}$ complexes. When incubated in the presence of Asp, ATP, and an amido group donor, this complex forms Asn-$tRNA^{Asn}$.[197]

Kinetic analysis of the partners of the transamidation pathway of Asn-$tRNA^{Asn}$ formation of *P. aeruginosa* and *H. pylori* also suggests the formation of a ternary complex. Indeed, the ND-AspRSs from *P. aeruginosa* and *H. pylori* decrease the K_m of *H. pylori* GatCAB for Asp-$tRNA^{Asn}$ ($2.24\,\mu mol\,l^{-1}$) 1.4- and 2-fold respectively without affecting significantly the k_{cat} of transamidation.[198] These observations suggest that the three partners interact at least transiently.

Further investigations revealed additional advantages conferred by the association of the partners in the *T. thermophilus* transamidosome. The kinetics of the hydrolysis of the aa-tRNAs in free form or bound in the complex show that the transamidosome stabilizes the ester bond of the Asp-$tRNA^{Asn}$ intermediate by increasing its half-life about twofold (half-lives 315 and 204 min) and that of the end product Asn-$tRNA^{Asn}$ fourfold (half-lives 18 and 66 min). Finally, measurements of the thermostability of the protein partners reveal a significant increase in the stability of GatCAB and the ND-AspRS at 85 °C, the optimal growth temperature of *T. thermophilus*, when associated in the transamidosome.[199]

5.14.4.2.7 Mechanisms that prevent the use of the mischarged Asp-$tRNA^{Asn}$ and Glu-$tRNA^{Gln}$ for protein synthesis

Upon being formed by the aaRSs, aa-tRNAs are trapped by the elongation factor EF-Tu in prokaryotes and EF-1A in eukaryotes and archaea and carried to the ribosome where they are used for elongation of the

polypeptide chain. The EF-Tu factor binds all species of homologous aa-tRNA, except fMet-tRNA^{Meti}, formed in prokaryotes, which binds the initiation factor IF2 to form the initiation complex, and Sec-tRNA^{Sec}, which binds the SelB protein in prokaryotes and EF-Sec in eukaryotes and archaea. These protein factors promote recognition of the selenocysteine insertion sequence (SECIS) element on the mRNA, which determines the incorporation of Sec on the UGA stop codon.[200,201] However, when bound on EF-Tu, mischarged tRNAs promote misincorporation of aa into proteins. Indeed, Chapeville *et al.*[202] showed that conversion of Cys acylating tRNA^{Cys} by Raney platinum into Ala results in misincorporation of Ala into the polypeptide chains. Thus, binding of Asp-tRNA^{Asn} and Glu-tRNA^{Gln} on EF-Tu would result in misincorporation of Asp and Glu into the polypeptide chain on Asn and Gln codons. However, the poor affinity of Asp-tRNA^{Asn} and Glu-tRNA^{Gln} for EF-Tu prevents their use for protein synthesis. Furthermore, sequestration of Asp-tRNA^{Asn} by the transamidosome constitutes an additional barrier preventing the use of this aa-tRNA for polypeptide chain elongation.

Nitrocellulose disc filtration experiments of guanine triphosphate (GTP)-activated EF-Tu•aa-tRNA complexes revealed that the noncognate Asp-tRNA^{Asn} does not bind *T. thermophilus* EF-Tu under conditions where both the cognate Asp-tRNA^{Asp} and Asn-tRNA^{Asn} are retained.[172] These results agree with the fact that His-tagged EF-Tu from *T. thermophilus* linked to a nickel-containing matrix retains the aspartylated tRNA^{Asp} and the asparaginylated tRNA^{Asn} but not the aspartylated tRNA^{Asn}.[87] Furthermore, it has also been shown that Glu-tRNA^{Gln} formed by the ND-GluRS from *Pisum sativum* chloroplasts does not bind the organellar EF-Tu factor.[203] These observations agree with those of protection experiments with *H. pylori* and *E. coli* EF-Tu against hydrolysis of aa-tRNA by RNase A or against alkaline hydrolysis of the ester bond, which show that EF-Tu from both organisms binds Asp-tRNA^{Asn} and Glu-tRNA^{Gln} much less strongly than the homologous Asp-tRNA^{Asp} and Glu-tRNA^{Glu}.[204] Absence of binding of Asp-tRNA^{Asn} and Glu-tRNA^{Gln} on the EF-Tu factor is rationalized by the general rule defining the binding properties of aa-tRNAs on EF-Tu described by Uhlenbeck and coworkers. This group demonstrated that to be released from the ribosome at a similar rate, the various homologous aa-tRNAs bind the EF-Tu factor with similar affinities, as a consequence of thermodynamic compensation between the aa and the tRNA moieties, due to the strong affinity of one partner compensating for the poor affinity of the second partner.[205,206] Absence of binding of Asp-tRNA^{Asn} on EF-Tu is provoked by the combination of the poor affinity of both the aa and the tRNA moieties because the mischarged aa-tRNA acquires the binding capacity only after conversion of the Asp moiety into Asn (**Figure 7**). The poor affinity of the Asp residue is related to a steric hindrance provoked by its side chain and the Glu227 residue of the protein, whereas the poor affinity of the tRNA^{Asn} moiety is related to base pairs G49–C65 and A51–U63 from the T-arm. In Asp-tRNA^{Asp}, the poor affinity of Asp for the protein is compensated by the stronger affinity of tRNA^{Asp} promoted by base pairs G49–U65 and G51–C63, whereas in Asn-tRNA^{Asn}, the poor affinity of tRNA^{Asn} is compensated by the stronger affinity of Asn. Substituting Glu227 in EF-Tu by Ala, and base pairs G49–C65 and A51–U63 in tRNA^{Asn} by base pairs G49–U65 and G51–C63 present in tRNA^{Asp}, promotes the binding of the aspartylated tRNA^{Asn} variant as strongly as aspartylated tRNA^{Asp} (**Figure 9**).[172]

Similar structural characteristics prevent binding on EF-Tu of Glu-tRNA^{Gln} formed by the indirect pathway of tRNA glutaminylation. According to investigations by the Uhlenbeck group, Glu and Gln constitute low and strong affinity partners of aa-tRNA, respectively. Furthermore, analysis of tRNA^{Glu} and tRNA^{Gln} sequences of organisms using the indirect pathway of tRNA glutaminylation reveals conservation in tRNA^{Glu} of base pairs G49–C65 and A51–U63, decreasing tRNA affinity, and in tRNA^{Gln} of base pairs G49–U65 and G51–C63, increasing tRNA affinity. Therefore, the structural elements determining the absence of binding of Asp-tRNA^{Asn} on EF-Tu also prevent binding of Glu-tRNA^{Gln}.[172] Measurements of the interaction between EF-Tu and aa-tRNA have shown that the discrimination property of EF-Tu against Asp-tRNA^{Asn} and Glu-tRNA^{Gln} is conserved even in bacteria using the direct pathway for tRNA asparaginylation and glutaminylation, which do not form these misacylated tRNAs.[204] Finally, a detailed analysis of *E. coli* tRNA sequences established that the structural elements that determine the poor affinity of tRNA^{Asn} and tRNA^{Gln} for EF-Tu and those determining the strong affinity of tRNA^{Asp} and tRNA^{Glu} for the factor are conserved in other tRNA species and contribute with the aa moiety to the thermodynamic compensation for appropriate binding of aa-tRNAs on EF-Tu.[172]

Interestingly, discrimination of the mischarged aa-tRNAs by EF-Tu occurs efficiently only when the partners are expressed at the physiological concentrations found in the wild-type cells. Overexpression of an ND-AspRS or -GluRS or high local concentrations of the EF-Tu factor induce binding of Asp-tRNA^{Asn} and Glu-tRNA^{Gln} on EF-Tu. Overexpression of the ND-GluRS from *B. subtilis* in *E. coli* provokes lethality of the host strain because of misincorporation of Glu into proteins provoked by the accumulation of Glu-tRNA^{Gln}.[207] Growth inhibition was relieved by coexpression of GlnRS preventing misacylation of tRNA^{Gln}, whereas coexpression of *B. subtilis* GatCAB converting tRNA^{Gln}-bound Glu into Gln also rescued the cells from the toxic effects of Glu-tRNA^{Gln}.[208] The Gln358Arg mutant of the ND-GluRS from *B. subtilis* rescued the thermosensitive GluRS mutation of the *E. coli* JP1449 strain at the nonpermissive temperature, as a consequence of the decreased efficiency of the enzyme to glutamylate tRNA^{Gln}.[209] Furthermore, expression in the *E. coli* trpA49 strain auxotrophic for Trp of the GluRS2 from *A. ferrooxidans* or *H. pylori* that specifically glutamylates tRNA^{Gln} restores growth of the strain in the absence of Trp. Because Trp synthetase activity is abolished by substitution of the Glu49 codon in the TrpA protein by a Gln codon in this strain, the acquisition of auxotrophy of the transformed strain results from incorporating Glu at position 49 by Glu-tRNA^{Gln} formed by GluRS2.[209] These results demonstrate that Glu-tRNA^{Gln} formed *in vivo* by *A. ferrooxidans* and *H. pylori* GluRS2 can be used for protein synthesis.

Similar results were reported for the mischarged Asp-tRNA^{Asn}. Although large amounts of Asp-tRNA^{Asn} are detrimental for *E. coli* growth, smaller amounts support protein synthesis and allow missense suppression. Overexpression of *H. pylori* ND-AspRS in *E. coli* is toxic, but the toxicity is rescued either by coexpression of the *H. pylori* GatCAB, which converts Asp-tRNA^{Asn} formed by the AspRS into Asn-tRNA^{Asn}, or by the Leu81Asn, Leu86Met, and Leu81Asn/Leu86Met mutations in the anticodon-binding domain, which, by increasing the specificity of the AspRS for tRNA^{Asp}, decrease its toxicity.[177] Expression of the ND-AspRS from *H. pylori* in the *E. coli* trpA34 strain auxotrophic for Trp restores growth of the strain in the absence of Trp. Because Trp synthetase activity is abolished in this strain by substituting the Asp34 codon in the TrpA protein with an Asn codon, the acquisition of auxotrophy results from incorporation of Asp at position 34 by Asp-tRNA^{Asn} formed by the ND-AspRS.[177,210] However, the highest level of growth of the transformed auxotrophic strain reaches only 38% that of the wild-type strain, probably because an increase in the level of Asp-tRNA^{Asn} formation causes lethality due to misincorporation of Asp into other proteins.[210]

5.14.5 Evolution of the Enzymes Involved in Gln-tRNA and in Asn-tRNA Biosynthesis

5.14.5.1 Evolution of aaRSs from Two Unlinked Ancestors

About 20 species of amino acids are incorporated into polypeptides. The aaRSs that catalyze the formation of the corresponding aa-tRNAs are derived from two precursors that have no evolutionary linkage, as evidenced mostly by the fact that their structures have completely different topologies.[51,211–213] and by phylogenetic analyses of the sequences of their amino acid residues.[214] This classification is also consistent with the differences observed in the interactions of these enzymes with modified substrates and in their reactions with reactive groups mounted on substrate analogues.[215]

Each of these two classes of aaRSs has about 10 amino acid substrates with the same spectrum of physicochemical properties (charged, polar, hydrophobic); for instance, glutamate, glutamine, and tyrosine are substrates of class I aaRSs, whereas aspartate, asparagines, and phenylalanine are substrates of class II aaRSs. These facts suggest the possibility that each aaRS class could have once supported an independent system for protein biosynthesis. As each aaRS class interacts with the opposite sides of their tRNA substrates (class I aaRSs approach the acceptor stem of tRNA from the minor groove side, whereas class II aaRSs approach it from the major groove side[161]), the possibility of cross-class pairings of ancestral aaRS on tRNA was indicated by modeling of ternary complexes such as TyrRS/tRNA/PheRS and was suggested to have protected the tRNA acceptor stem in an environment deleterious to RNA in progenotes using a more degenerate and ambiguous genetic code than that of extant organisms.[216,217] The specific

recognition of $tRNA^{Asp}$ by the GluRS paralogue YadB in *E. coli*[19,218] is another hint of the existence of such interclass aaRSI/tRNA/aaRSII ternary complexes.

5.14.5.2 Evolution of the tRNA Glutaminylation and Asparaginylation Pathways, and the Late Emergence of the Direct Pathways

According to the theory of the coevolution of the genetic code and pathways for amino acid biosynthesis,[6] glutamine and asparagine have been incorporated into proteins later than most other amino acids. The aaRSs specific for them, GlnRS and AsnRS, are missing in many organisms where Gln-$tRNA^{Gln}$ and/or Asn-$tRNA^{Asn}$ are synthesized through indirect pathways (see Sections 5.14.3 and 5.14.4).

GlnRS was found in the cytoplasm of all eukaryotes where it was searched for, but is absent in many Gram-positive bacteria as originally observed by Wilcox and Nirenberg;[83] they discovered in 1968 the indirect pathway of Gln-tRNA biosynthesis in *B. subtilis*,[83] and Wilcox did the first characterization of a Glu-$tRNA^{Gln}$ amidotransferase.[195] Only about three decades later was the *gatCAB* operon encoding the trimeric AdT of *B. subtilis* cloned and its trimeric product purified and characterized.[86] Although GlnRS was found in *E. coli* and in several Gram-negative bacteria, its absence and the presence of the indirect pathway were reported for the Gram-negative *Rhizobium meliloti*.[85] Later, GlnRS was shown to have a very sparse distribution among bacteria and to be absent in archaea.[219] Phylogenetic analyses of GluRSs and GlnRSs indicated that GlnRS evolved from a GluRS, probably an ND-GluRS,[46] in early eukaryotes, and that its gene was transferred horizontally[220] from them to a few bacterial species.[45] In this evolutionary process, the ND-GluRS changed its specificity for its amino acid substrate (from glutamate to glutamine) and for its tRNA substrates (keeping $tRNA^{Gln}$ but losing affinity for $tRNA^{Glu}$). This transition of ND-GluRS to GlnRS in early eukaryotes, and that of ND-GluRS to D-GluRS in bacterial cells that received a eukaryotic GlnRS by horizontal gene transfer, probably required few mutations as suggested by the following two observations of reverse evolution: first, human GlnRS modified by the incorporation of a few GluRS-specific features preferentially aminoacylates tRNA with glutamate instead of glutamine,[221] although this variant has a much reduced activity (k_{cat}); second, the D-GluRS of *T. thermophilus* was made nondiscriminating by a single-residue substitution, Arg358 to Gln, in the anticodon-binding domain 4.[107] A probable intermediate in the evolution of ND-GluRS toward GlnRS was found in *H. pylori* and *A. ferrooxidans*.[114,115] These bacteria contain two GluRSs: one specific for $tRNA^{Glu}$ named GluRS1 and the other that charges Glu on $tRNA^{Gln}$ but not on $tRNA^{Glu}$. The latter, a putative link in the evolution of GlnRS, was named GluGlnRS.

AsnRS, as GlnRS, was found in the cytoplasm of all eukaryotes where it was searched for, but in contrast to GlnRS, it is present in a wide range of bacteria and in two archaea.[219] Phylogenetic studies show that although AsnRS and GlnRS belong to evolutionarily unlinked aaRS classes, AsnRS (class II) and GlnRS (class I) evolved, respectively, within the clusters of the synthetases for their corresponding diacid, AspRS and GluRS; in the case of AsnRS, the origin is localized only to the archaeal genre of AspRS in general,[27,219] which is nondiscriminating. For both archaeal and bacterial ND-AspRSs, substitutions of only a few residues in the anticodon-binding region were sufficient to give variants endowed with more discriminating properties.[169,185]

Two different types of AdTs participate in the indirect pathways presented above: a trimeric GatCAB AdT, which generally transamidates either Glu-$tRNA^{Gln}$ or Asp-$tRNA^{Asn}$, and thus is said to have both Glu-AdT and Asp-AdT activities,[52] and a dimeric GatDE Glu-AdT, which transamidates only Glu-$tRNA^{Gln}$. Phylogenetic analysis of the GatB and GatE subunits, which catalyze the transamidation reaction, revealed that they evolved from a common ancestor and that the split between them took place prior to the phylogenetic divide between bacteria and archaea;[121] archaea retained both AdTs and bacteria retained only GatCAB. From this, it would appear that archaea have two types of Glu-AdT activity, but this is probably not the case, at least for the archaea *M. thermautotrophicus*, as the structure of its Glu-$tRNA^{Gln}$ does not allow its transmidation by this archaeal GatCAB.[94] In contrast to bacterial GatCAB, this enzyme uses Asn almost as well as Gln as an amide donor.

The repartition in the universal phylogenetic tree of life, of the direct and indirect pathways for Gln-tRNA and Asn-tRNA biosynthesis in bacterial lineages, is shown in **Figure 11**.[222]

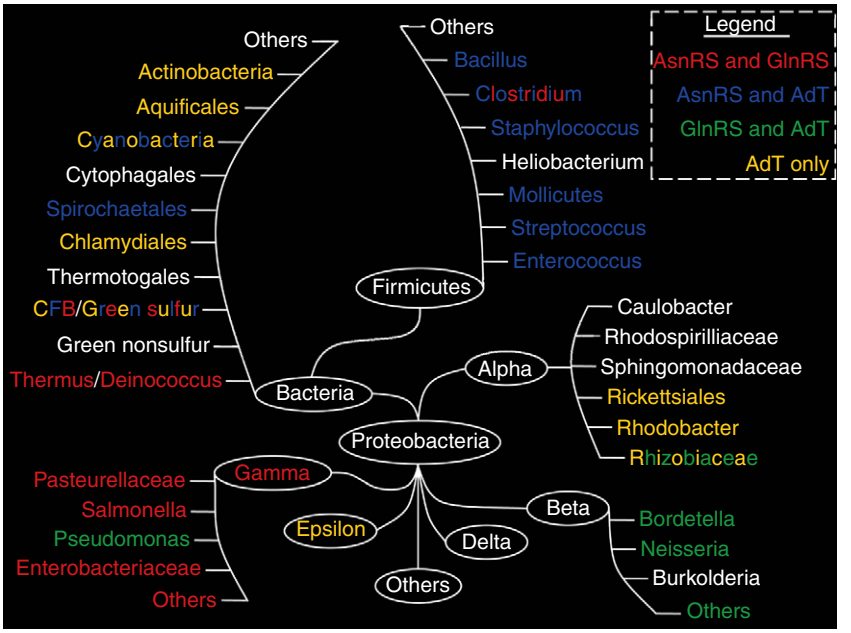


Figure 11 Repartition of the direct and indirect pathways of Asn-tRNA^{Asn} and Gln-tRNA^{Gln} biosynthesis throughout the bacterial phylogenetic tree (Reproduced from M. Ibba; D. Söll, *Genes Dev.* **2004**, *18*, 731–738). The organisms containing both AsnRS and GlnRS are in red; AsnRS and AdT in blue; GlnRS and AdT in green; AdT only in yellow. Letters in alternated colors indicate that the group contains organisms in several of the above-mentioned categories.

5.14.6 Inhibitors of Enzymes Involved in Gln-tRNA and Asn-tRNA Biosynthesis as Tools for Structural and Mechanistic Studies and Leads for Therapeutic Applications

Rationally designed synthetic inhibitors of aaRSs are typically stable analogues of aa∼AMPs (**Figure 12**, also see Section 5.14.1). Stability is achieved by replacing the labile mixed anhydride function with nonhydrolyzable bioisosteres. Several aminoalkyl adenylates, aminoacylsulfamoyl adenosines (sulfamates), aminoacylsulfamide adenosines, or β-ketophosphonates have been synthesized and shown to be inhibitors of corresponding aaRS[223] (**Figure 12**). Several natural products with various chemical structures have been identified as inhibitors of aaRS[223]. Inhibitors were also discovered using high-throughput screening[224] or virtual screening of compound libraries.[225]

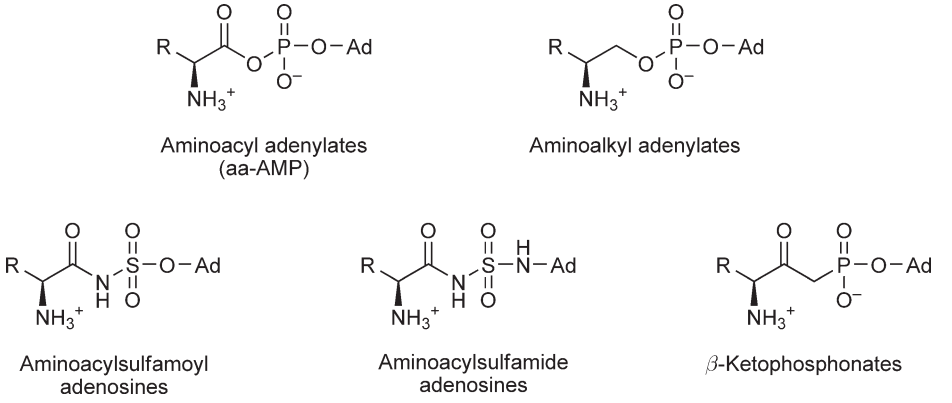


Figure 12 Aminoacyl-adenylates and stable bioisosteres. Ad, adenosine; R, amino acid side chain.

5.14.6.1 Direct Pathway: Inhibitors of Glutaminyl-tRNA Synthetase and Asparaginyl-tRNA Synthetase

Glutaminol adenylate (**1**) (**Table 1**) is a competitive inhibitor of GlnRS with respect to glutamine ($K_i = 0.28\ \mu mol\ l^{-1}$) and ATP ($K_i = 0.86\ \mu mol\ l^{-1}$).[226] The corresponding methyl phosphate ester (**2**) is a weaker inhibitor ($K_i = 10\ \mu mol\ l^{-1}$) with respect to glutamine. The 50-fold increase of K_i due to this phosphate methylation indicates that the negative charge of the phosphate group is important for its interaction with GlnRS.

5′-*O*-[*N*-(L-Glutaminyl)sulfamoyl] adenosine (**3**) inhibits GlnRS with a K_i of $1.3\ \mu mol\ l^{-1}$, which is significantly smaller than the K_m values of 110 and $111\ \mu mol\ l^{-1}$ for Gln and ATP,[78] respectively, but is larger than the expected nanomolar range K_i of most sulfamates for their corresponding aaRS.[223]

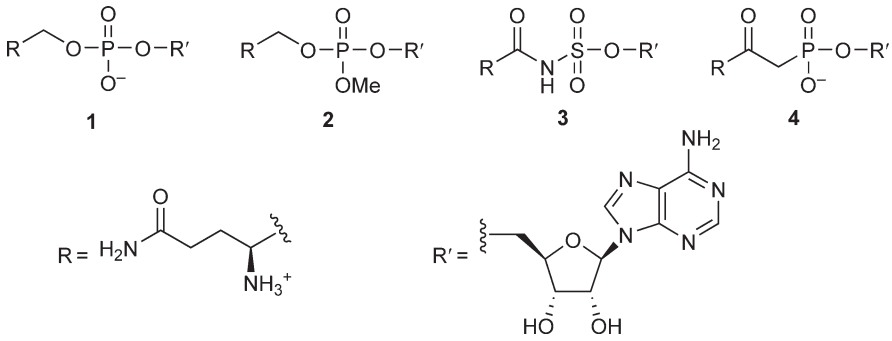
Glutaminyl-β-ketophosphonate (**4**) inhibits *E. coli* GlnRS by binding competitively but weakly at two distinct sites on this enzyme (average $K_i = 650\ \mu mol\ l^{-1}$).[227] The kinetic results indicate that the glutamine and the ATP moieties of **4**, connected by the β-ketophosphonate linker, cannot bind GlnRS simultaneously and that one molecule binds the AMP-binding site of GlnRS through its AMP module, whereas another molecule binds the Gln-binding site through its glutamine module.

Studies on AsnRS ligands are rare. Three asparaginyl adenylate mimics that target AsnRS have been reported:[228] sulfamate (**5**), cycloadenosine sulfamate (**6**), and L-aspartate-β-hydroxamate (**7**) (**Table 2**). The IC_{50} values of compounds **5** and **7** were quite similar (4.5 and $4\ \mu mol\ l^{-1}$, respectively) for the filarial roundworm *B. malayi* AsnRS, whereas the value for the cycloadenosine derivative **6** was approximately 20 times higher ($IC_{50} = 76\ \mu mol\ l^{-1}$). Compounds **5** and **6** have no significant selectivity for *B. malayi* relative to human AsnRS. Compound **5** has previously been shown to bind in the crystal structure of *T. thermophilus* AsnRS and was used to determine the mechanism of discrimination between asparagine and aspartic acid.[141]

5.14.6.2 Indirect Pathway: Inhibitors of Glutamyl-tRNA Synthetase, Aspartyl-tRNA Synthetase, and Aminoacyl-tRNA Amidotransferase

In the initial step of the so-called indirect (transamidation) pathway, an ND-GluRS aminoacylates $tRNA^{Gln}$ with glutamate or an ND-AspRS aminoacylates $tRNA^{Asn}$ with aspartate; in the second step, the incorrectly

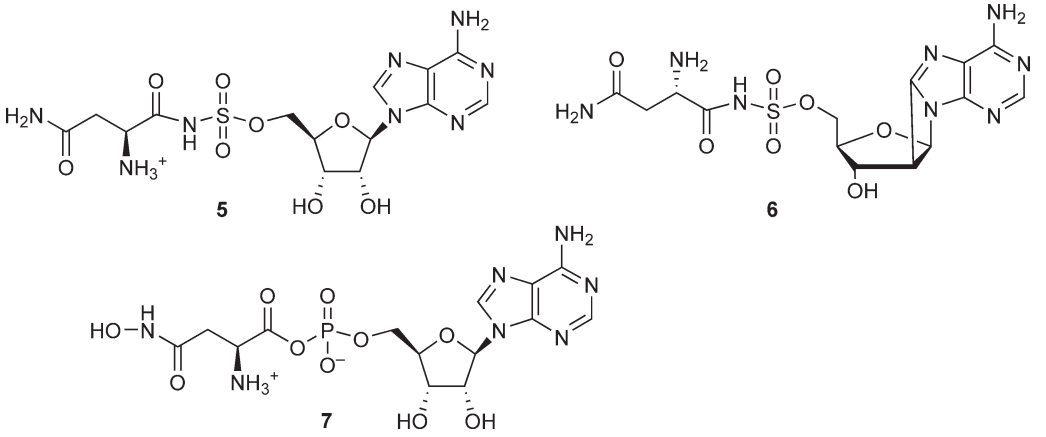
Table 1 Inhibitors of glutaminyl-tRNA synthetase (GlnRS)



Source	*Compound*	*Inhibition*[a]	*Reference*
E. coli	**1**	$K_i = 0.28\ \mu mol\ l^{-1}$ (vs Gln) $K_i = 0.86\ \mu mol\ l^{-1}$ (vs ATP)	226
	2	$K_i = 10\ \mu mol\ l^{-1}$ (vs Gln)	
E. coli	**3**	$K_i = 1.3\ \mu mol\ l^{-1}$	78
E. coli	**4**	$K_i = 650\ \mu mol\ l^{-1}$ (two sites)	227

[a] Competitive inhibition; measured in the aminoacylation reaction.
E. coli, *Escherichia coli*.

Table 2 Inhibition of asparaginly-tRNA synthetase (AsnRS)



	IC_{50} (μmol l^{-1})[a]	
Inhibitor	*Brugia malayi*	*Human*
5	4.5	1.7
6	76	90
7	4	–

[a] Measured in the ATP–PP_i exchange reaction.

aminoacylated tRNAs are transformed, respectively, into Gln-$tRNA^{Gln}$ and Asn-$tRNA^{Asn}$ by amidotransferases (AdT), which transform the side chain carboxyl function of Glu or Asp on tRNA into an amide group (see Section 5.14.1). Because GluRS and AspRS are involved in this pathway, their inhibition is reviewed in this section.

Glutamol adenylate (**8**) (**Table 3**) is a competitive inhibitor ($K_i = 3\ \mu mol\ l^{-1}$) of GluRS from *E. coli*.[229] The N^6-benzoyl adenine derivative is also an inhibitor ($K_i = 60\ \mu mol\ l^{-1}$). Replacing adenine with other bases (purine, cytosine, dihydrocytosine, uridine) resulted in a more than 1000-fold loss of activity, indicating the importance of the contribution of the adenine ring to enzyme binding.

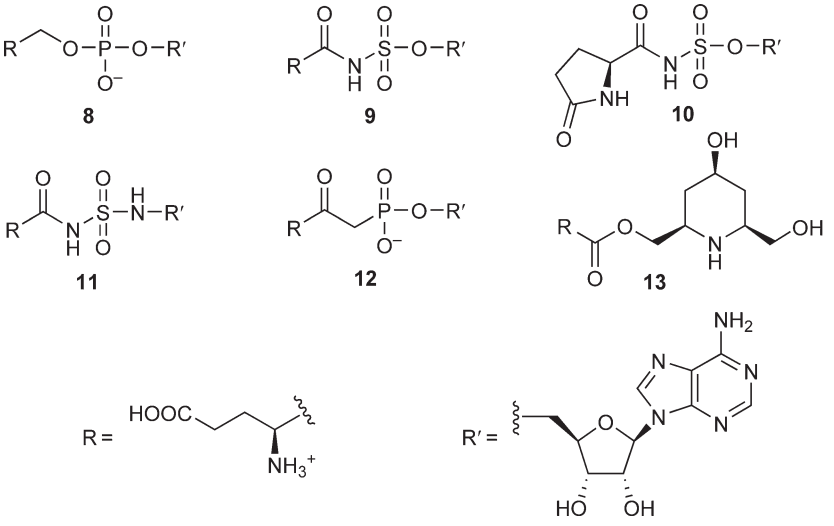
5′-*O*-[*N*-(L-Glutamyl)sulfamoyl] adenosine (**9**) is a potent competitive inhibitor of *E. coli* GluRS with respect to glutamic acid; with its $K_i = 2.8\ nmol\ l^{-1}$, compound **9** is by far the best inhibitor of this class I aaRS.[230] It is a weaker inhibitor ($K_i = 70\ nmol\ l^{-1}$) of mammalian (murine) GluRS. 5′-*O*-[*N*-(L-Pyroglutamyl)sulfamoyl] adenosine (**10**), which differs from **9** only by the cyclization of the γ-carboxylic acid side chain with the α-amino group of glutamic acid, is a weaker inhibitor ($K_i = 15\ \mu mol\ l^{-1}$).

Sulfamide (**11**) inhibits *S. aureus* GluRS with a $K_i = 150\ nmol\ l^{-1}$ and has only a marginal twofold selectivity between bacterial and human (HeLa GluRS, $K_i = 300\ nmol\ l^{-1}$) enzymes.[231]

Glutamyl-β-ketophosphonate (**12**) is a competitive inhibitor of *E. coli* GluRS with a K_i of $18\ \mu mol\ l^{-1}$ with respect to its substrate glutamate, and binds at one site on this monomeric enzyme.[227] Compound **12** inhibits bovine liver GluRS 145-fold less efficiently than *E. coli* GluRS because of competitive weak binding at two distinct sites (average $K_i = 2.6\ mmol\ l^{-1}$). Various glutamic acid esters in which the alcohol moiety is ribose, prolinol, or substituted piperidines are inhibitors of *E. coli* GluRS; the best inhibitor is obtained with piperidine derivative (**13**) ($K_i = 20\ \mu mol\ l^{-1}$).[232] This ester resulted from the esterification of a meso diol and is a mixture of diastereoisomers.

GluRS has the characteristic, which is also common to Gln, Arg, and an unusual LysRS, of requiring the presence of its cognate tRNA to catalyze the activation of its amino acid substrate (**Figure 1**, Section 5.14.1). The crystal structures of complexes of *T. thermophilus* GluRS with ligand **8** or **9** have been solved, providing an

Table 3 Inhibitors of glutamyl-tRNA synthetase (GluRS)



Source	*Compound*	*Inhibition*	*Reference*
E. coli	**8**	$K_i = 3\ \mu mol\ l^{-1}$	229
E. coli	**9**	$K_i = 2.8\ nmol\ l^{-1}$	230
Mammalian, murine	**9**	$K_i = 70\ nmol\ l^{-1}$	
E. coli	**10**	$K_i = 15\ \mu mol\ l^{-1}$	
S. aureus	**11**	$K_i = 150\ nmol\ l^{-1}$	231
Mammalian, HeLa	**11**	$K_i = 300\ nmol\ l^{-1}$	
E. coli	**12**	$K_i = 18\ \mu mol\ l^{-1}$	227
Mammalian, bovine	**12**	$K_i = 2.6\ mmol\ l^{-1}$	
E. coli	**13**	$K_i = 20\ \mu mol\ l^{-1}$	232

E. coli, Escherichia coli; S. aureus, Staphylococcus aureus.

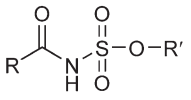
insight into the induced fit mechanism in which the binding of the cognate tRNA is required in order to form a productive active site conformation and trigger amino acid activation.[96,106]

Nonhydrolyzable aspartyl adenylate analogues (**14**–**16**) (**Table 4**) have been prepared and tested as inhibitors of *E. coli* AspRS.[233] Sulfamate (**14**) is a potent competitive inhibitor ($K_i = 15\ nmol\ l^{-1}$), whereas L-aspartol adenylate (**15**) is a weaker inibitor ($K_i = 45\ \mu mol\ l^{-1}$) with respect to aspartic acid. The corresponding β-ketophosphonate (**16**) is also a strong inhibitor ($K_i = 123\ nmol\ l^{-1}$). Replacing the 5′-oxygen on the ribose with an NH group (compound **17**) resulted in an equally potent inhibitor of the *S. aureus* enzyme ($K_i = 15\ nmol\ l^{-1}$).[231]

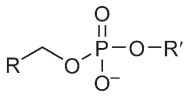
Asn-tRNA formation in *P. aeruginosa* involves an ND-AspRS, which forms both Asp-tRNA^{Asp} (direct pathway for Asp) and Asp-tRNA^{Asn} (indirect pathway for Asn). L-Aspartol adenylate (**15**) inhibits aspartylation of tRNA^{Asp} ($K_i = 41\ \mu mol\ l^{-1}$) more efficiently than that of tRNA^{Asn} ($K_i = 215\ \mu mol\ l^{-1}$), the other natural tRNA substrate of this enzyme.[234]

Microcin C (**18**) (**Figure 13**) is a peptide–nucleotide antibiotic produced by *E. coli*. The heptapeptide Met-Arg-Thr-Gly-Asn-Ala-Asp is substituted at the N-terminus by a formyl group and the C-terminal aspartic acid is linked to the phosphodiester of adenosine and *n*-aminopropanol through an amide group (*N*-acyl phosphoramidate linkage).[235] Microcin C is introduced into sensitive cells by peptide transporters[236] and processed by peptidases[237] to release the stable aspartyl adenylate (**19**) that inhibits aspartyl-tRNA synthetase (AspRS), leading to cessation of translation and bacterial cell growth.[238]

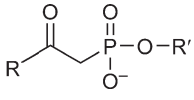
Table 4 Inhibitors of aspartyl-tRNA synthetase (AspRS)



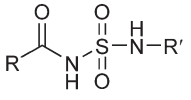
14



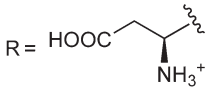
15

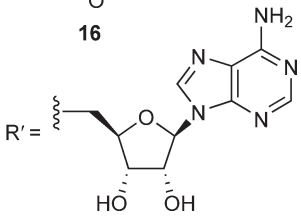


16



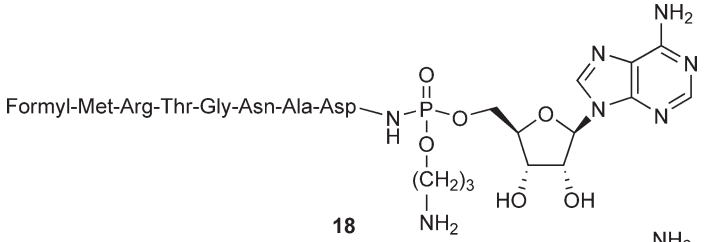
17



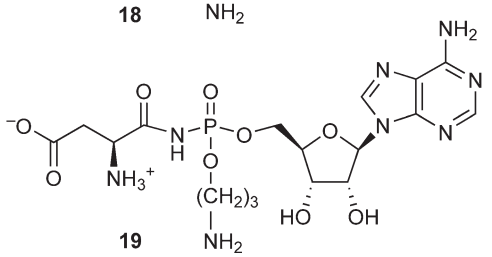


Source	*Compound*	*Inhibition*	*Reference*
E. coli	**14**	$K_i = 15\ \text{nmol l}^{-1}$	233
	15	$K_i = 45\ \mu\text{mol l}^{-1}$	
	16	$K_i = 123\ \text{nmol l}^{-1}$	
S. aureus	**17**	$K_i = 15\ \text{nmol l}^{-1}$	231
P. aeruginosa ND-AspRS	**15**	$K_i = 41\ \mu\text{mol l}^{-1}$	234
		$K_i = 215\ \mu\text{mol l}^{-1}$	

E. coli, Escherichia coli; S. aureus, Staphylococcus aureus; P. aeruginosa ND-AspRS, *Pseudomonas aeruginosa* nondiscriminating AspRS.



18

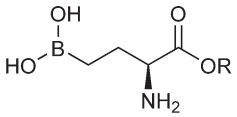


19

Figure 13 Structures of microcin C (**18**) and processed microcin C (**19**), an inhibitor of AspRS.

Glutamyl-γ-boronate and analogues are mechanism-based inhibitors of *S. pyogenes* Glu-tRNA^{Gln} amidotransferase.[191] Compounds **20** and **21** (**Figure 14**) provide potent inhibition of glutaminase (absence of ATP and tRNA substrates) and Gln-dependent transferase activities of the enzyme. Structure–activity studies revealed a narrow range of tolerated chemical changes that maintained activity. The compounds were designed to engage the putative catalytic serine nucleophile required for the glutaminase reaction. Boronic acids have been shown to form reversible covalent bonds with the active site serine or threonine residues of diverse enzymes leading to a tetrahedral adduct (transition state analogue).

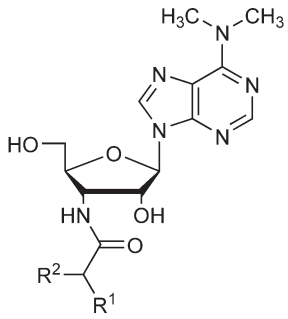
Glutamine and ATP analogues[192,193] were useful to probe the reaction mechanism, but these inhibitors are likely to interfere with many other enzymes acting on the same substrates. More recently, analogues of puromycin (**22**) (**Table 5**) were synthesized and evaluated as mechanism-based selective inhibitors of *H. pylori*

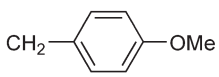
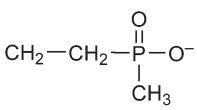


		IC_{50} (μmol l^{-1})	
		Glutaminase	Transferase
R = OH	**20**	1.6	0.10
R = OCH_3	**21**	1.3	0.05

Figure 14 Inhibition of *Streptococcus pyogenes* Glu-AdT by glutamyl-γ-boronates.

Table 5 Inhibitors of *Helicobacter pylori* GatCAB amidotransferase derived from puromycin



	R^1	R^2	K_i (μmol l^{-1})
22		(L)-NH_2	4100
23	CH_2-COO^-	(L)-NH_2	134
24	$CH_2-CH_2-COO^-$	(L)-NH_2	105
25	$CH_2-CH_2-COO^-$	(L)-OH	130
26	$CH_2-CH_2-CONH_2$	(L)-NH_2	45
27		(D,L)-NH_2	33
28	$CH_2-CH_2-SOCH_3$	(L)-NH_2	11
29	$CH_2-CH_2-SO_2CH_3$	(L)-NH_2	4

GatCAB amidotransferase.[198,239] This natural product mimics the charged 3′-terminus of aa-tRNA and has been used as a tool for the study of protein biosynthesis. The parent compound **22** is a very weak inhibitor of AdT. The amino acid chain is related to tyrosine and differs from the glutamic and aspartic side chains transformed in the kinase or the transamidase steps. Replacement of the methoxyphenyl moiety of puromycin by carboxylic acid derivatives (**23**–**26**) improved the ability to inhibit this AdT. Stable analogues of the transition state in the last step of the transamidation process (**27**–**29**) where the carbonyl to be attacked by NH_3 is replaced by tetrahedral sulfur or phosphorus atom with a methyl group mimicking ammonia exhibited the highest activity.

5.14.6.3 Inhibitors as Leads for Therapeutic Applications

aa-tRNA synthetases and aa-tRNA amidotransferases are essential enzymes of the translation apparatus, which transforms the information stored in nucleic acids into proteins. Their inhibition results in inhibition of protein biosynthesis, which in turn leads to cell growth arrest. Consequently, these enzymes have been acknowledged as rational targets for antiproliferative or anti-infective drug development.[223,231,240–244] Selective inhibition of bacterial aaRS has proved to be a successful strategy for the production of antibacterial agents. Pseudomonic acid (generic name: mupirocin) isolated from *P. fluorescens* is a highly potent ($K_i = 6\,\text{nmol}\,l^{-1}$ for *E. coli* IleRS) and selective (8000-fold selectivity for pathogens vs. mammalian) inhibitor of bacterial IleRS.[245] Mupirocin is the sole aaRS inhibitor currently in clinical use.

The high *in vitro* activity of the inhibitors often does not translate into valuable antibiotic activity because of the combination of poor cell penetration and lack of selectivity. 5′-*O*-[*N*-(Aminoacyl)sulfamoyl] adenosines are potent inhibitors of aaRS, but their whole-cell antibacterial activity is very limited. 5′-*O*-[*N*-(Dipeptidyl)sulfamoyl] adenosines (**Figure 15**) showed improved antibacterial activity.[246] These compounds can be regarded as synthetic analogues of the natural product microcin C (**18**) (**Figure 13**).

Kuhn and coworkers applied structure-based computational ligand screening and design to develop inhibitors of *B. malayi* AsnRS.[228] *B. malayi* is a nematode worm that causes lymphatic filariasis (elephantiasis), a debilitating disease that afflicts more than 200 million people in tropical countries. For instance, compound **30** (**Figure 16**) showed significant inhibition of AsnRS ($IC_{50} = 51\,\mu\text{mol}\,l^{-1}$), whereas compounds **31** and **32** showed weaker IC_{50} values (123 and 175 $\mu\text{mol}\,l^{-1}$), with a three- to eightfold selectivity for *B. malayi* over human AsnRS.

Glu-AdT inhibitors (**20** and **21**) (**Figure 14**) inhibited growth of several bacteria that use the transamidation pathway (*S. pyogenes, S. pneumoniae, E. faecalis, H. pylori*).[191] Notably, these compounds did not inhibit growth of two *E. coli* strains, consistent with the absence of Glu-AdT in this microorganism.

5′-*O*-[*N*-(L-Aminoacyl)sulfamoyl] adenosines containing all 20 proteinogenic amino acids and analogues are potent and synergistic immunosuppressants.[247] Standard amino acid derivatives show activity in the $IC_{50} = 0.3–5.6\,\mu\text{mol}\,l^{-1}$ range when tested in a cellular mixed lymphocyte reaction assay, a test clinically performed to predict transplant rejection. Compounds related to Asn and Gln are among the most potent with $IC_{50} = 0.3$ and $0.8\,\mu\text{mol}\,l^{-1}$, respectively. These immunosuppressive effects may hamper the use of these sulfamates as systemic antibiotics.

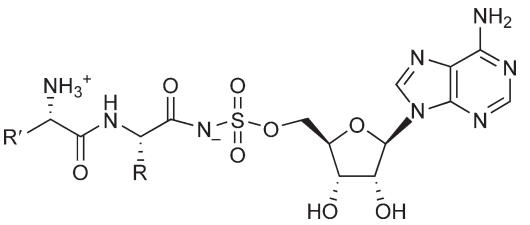
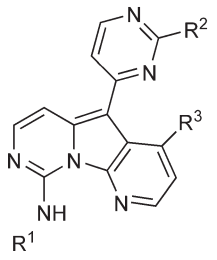


Figure 15 5′-*O*-[*N*-(Dipeptidyl)sulfamoyl] adenosine.



				IC_{50} ($\mu\text{mol}\,l^{-1}$)	
				Brugia malayi AsnRS	Human AsnRS
30	$R^1 = H$	$R^2 = SMe$	$R^3 = H$	51	101
31	$R^1 = (CH_2)_2–NMe_2$	$R^2 = NH_2$	$R^3 = OMe$	173	>500
32	$R^1 = (CH_2)_3–NMe_2$	$R^2 = NH_2$	$R^3 = OMe$	123	>1000

Figure 16 Inhibitors of *Brugia malayi* AsnRS from computational screening.

5.14.7 Conclusions

The proteins of the diverse community of primitive cells present on earth during the early steps of life[248] probably did not contain glutamine or asparagine residues, as proposed by the theory of the coevolution of the genetic code and the amino acid biosynthetic pathways[6] and as suggested by their low levels compared to those of glutamic acid and aspartic acid in the proteins of extant thermophilic and hyperthermophilic organisms[39] living at high temperatures postulated to be similar to those at the surface of the primitive earth.[249] The sensitivity of glutamine and asparagine to deamidation may have been a factor for their late selection for incorporation into proteins, compared to that of their precursors glutamate and aspartate and most other amino acids present in extant proteins. Free Asn and AsnRS probably appeared late in evolution, as both AsnRS and Asn synthetase evolved from an ancestral AspRS by gene duplication.[27] The exceptional sensitivity of Gln and Asn residues under physiological conditions is used to control protein turnover and many other biological processes.[32]

The synthesis in many extant organisms of these two amide residues from their respective precursors glutamate and aspartate esterified to tRNA (the indirect aminoacylation pathways described in Sections 5.14.3 and 5.14.4) and that of other amino acid residues, such as selenocysteine (which is also synthesized from a precursor esterified on a tRNA[250]) support the model of prebiotic metabolism taking place at the surface of solid particles,[10] analogous to ancestral RNAs.

More organisms kept the ancestral indirect pathway for Gln-tRNA biosynthesis than that for Asn-tRNA biosynthesis, as evidenced by the presence of AsnRS in many more species than those carrying a GlnRS (see Section 5.14.5). This preferential retention of the indirect pathway for Gln-tRNA biosynthesis may be related to the predominant importance of Gln compared to Asn in the nitrogen metabolism (see Section 5.14.1.4). A comparative study of nitrogen status and metabolism in species that kept this pathway and others that lost it, using the tools of metabolomics, might reveal the physiological function of this pathway. An alternative explanation for the poor distribution of GlnRS in prokaryotes is the incompatibility between GlnRS that appeared in eukaryotes and prokaryotic $tRNA^{Gln}$ identity (see Section 5.14.3). The presence of AsnRS in about 50% of bacteria and archaebacteria may result from more frequent successful horizontal gene transfer from eukaryal or archaeal ancestors[219,251] allowed by a more permissive tRNA identity barrier for AsnRS.

So far, several inhibitors of enzymes involved in the formation of Gln-tRNA and Asn-tRNA were designed to serve as probes for mechanistic studies or ligands for X-ray crystal structure determination (see Section 5.14.6), but selective and high whole-cell activity against microorganisms remained elusive. The availability of these enzymes for high-throughput screening provides an opportunity for developing new inhibitors. The widespread use of the transamidation pathway among bacteria and its absence in the mammalin cytoplasm identifies the aa-tRNA amidotransferases as targets for antibiotic therapy; the presence of these types of enzymes in human mitochondria, however, calls for caution.

Abbreviations

aa	amino acid
aa∼AMP	aminoacyl-adenylate
aaRS	aminoacyl-tRNA synthetase
aa-tRNA	aminoacyl-tRNA
Acp3	3-amino-3-carboxypropyl
ADP/AlF	adenosine diphosphate/Aluminum Fluoride
AdT	aminoacyl-tRNA amidotransferase
ASK1	apoptosis signal-regulating kinase 1
Asn	asparagine
Asn∼AMP	asparaginyl-adenylate
AsnRS	asparaginyl-tRNA synthetase
Asp	aspartic acid
Asp∼AMP	aspartyl-adenylate
Asp-AdT	aspartyl-tRNA amidotransferase

AspRS	aspartyl-tRNA synthetase
AspRS1	bacterial-type AspRS
AspRS2	archaeal-type AspRS
Asx	asparagine and/or aspartic acid
ATP	adenosine triphosphate
CHO	Chinese hamster ovary
D	2-dihydro-uridine
D-aaRS	discriminating aminoacyl-tRNA synthetase
Gln	glutamine
Gln∼AMP	glutaminyl-adenylate
GlnRS	glutaminyl-tRNA synthetase
Glu	glutamic acid
Glu∼AMP	glutamyl-adenylate
Glu-AdT	glutaminyl-tRNA amidotransferase
GluRS	glutamyl-tRNA synthetase
GluRS1	canonical $tRNA^{Glu}$-specific GluRS
GluRS2	noncanonical $tRNA^{Gln}$-specific GluRS
Glx	glutamine and/or glutamic acid
GTP	guanine triphosphate
LUCA	last universal communal ancestor
mnm^5s^2U	5-methylaminomethyl-2-thiouridine
ND	nondiscriminating (AspRS or GluRS)
ND-aaRS	nondiscriminating aminoacyl-tRNA synthetase
ND-AspRS	nondiscriminating Asp-tRNA synthetase
ND-GluRS	nondiscriminating GluRS
ORFs	open reading frames
SECIS	selenocysteine insertion sequence
t^6A	N^6-threonylcarbamoyladenosine
tRNA	transfer ribonucleic acid

Nomenclature

k_{cat}	catalytic constant
K_D	dissociation constant
K_i	inhibition constant
K_m	Michaelis constant
M_r	molecular radius
Ψ	pseudouridine

References

1. F. Jühling; M. Mörl; R. K. Hartmann; M. Sprinzl; P. F. Stadler; J. Pütz, *Nucleic Acids Res.* **2009**, *37*, D159–D162.
2. R. Giegé; M. Frugier, Transfert RNA Structure and Identity. In *Translation Mechanisms*; J. Lapointe, L. Brakier-Gingras, Eds.; Eurekah.com/Landes Bioscience: Georgetown: TX, 2003; pp 1–24.
3. P. Berg, *J. Biol. Chem.* **1958**, *233*, 601–607.
4. T. Cathopoulis; P. Chuawong; T. L. Hendrickson, *Mol. Biosyst.* **2007**, *3*, 408–418.

5. K. Sheppard; J. Yuan; M. J. Hohn; B. Jester; K. M. Devine; D. Söll, *Nucleic Acids Res.* **2008**, *36*, 1813–1825.
6. J. T. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1909–1912.
7. M. Di Giulio, *Phys. Life Rev*. **2004**, *1*, 128–137.
8. J. T. Wong, *Bioessays* **2005**, *27*, 416–425.
9. B. M. Rode; D. Fitz; T. Jakschitz, *Chem. Biodivers.* **2007**, *4*, 2674–2702.
10. G. Wächtershäuser, *Microbiol. Rev.* **1988**, *52*, 452–484.
11. A. Danchin, *Prog. Biophys. Mol. Biol.* **1989**, *54*, 81–86.
12. S. H. Kim; J. L. Sussman; F. L. Suddath; G. J. Quigley; A. McPherson; A. H. Wang; N. C. Seeman; A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 4970–4974.
13. W. Saenger, tRNA – A Treasury of Stereochemical Information. In *Principles of Nucleic Acid Structure*; C. R. Cantor, Ed.; Springer-Verlag: New York, 1984; pp 331–349.
14. E. Verkamp; A. M. Kumar; A. Lloyd; O. Martins; N. Stange-Thomann; D. Söll, Glutamyl-tRNA as an Intermediate in Glutamate Conversions. In *tRNA: Structure, Biosynthesis, and Function*; D. Söll, U. L. RajBhandary, Eds.; American Society for Microbiology: Washington, DC, 1995; pp 545–550.
15. R. J. Roberts, *J. Biol. Chem.* **1974**, *249*, 4787–4796.
16. A. J. Lloyd; A. M. Gilbey; A. M. Blewett; G. De Pascale; A. El Zoeiby; R. C. Levesque; A. C. Catherwood; A. Tomasz; T. D. Bugg; D. I. Roper; C. G. Dowson, *J. Biol. Chem.* **2008**, *283*, 6402–6417.
17. J. A. Nesbitt, III; W. J. Lennarz, *J. Biol. Chem.* **1968**, *243*, 3088–3095.
18. S. I. Kim; J. E. Germond; D. Pridmore; D. Söll, *J. Bacteriol.* **1996**, *178*, 2459–2461.
19. M. Blaise; H. D. Becker; J. Lapointe; C. Cambillau; R. Giegé; D. Kern, *Biochimie* **2005**, *87*, 847–861.
20. E. A. First, Catalysis of the tRNA Aminoacylation Reaction. In *The Aminoacyl-tRNA Synthetases*; M. Ibba, C. Francklyn, S. Cusack, Eds.; Eurekah.com/Landes Bioscience: Georgetown: TX, 2005; pp 328–352.
21. L. Feng; D. Tumbula-Hansen; B. Min; S. Namgoong; J. Salazar; O. Orellana; D. Söll, Transfer RNA-Dependent Amidotransferases: Key Enzyme for Asn-tRNA and Gln-tRNA Synthesis in Nature. In *The Aminoacyl-tRNA Synthetases*; M. Ibba, C. Francklyn, S. Cusack, Eds.; Eurekah.com/Landes Bioscience: Georgetown: TX, 2005; pp 314–319.
22. L. J. Reitzer, Ammonia Assimilation and the Biosynthesis of Glutamine, Glutamate, Aspartate, Asparagine, L-alanine, and D-alanine. In Escherichia coli *and* Salmonella, *Cellular and Molecular Biology*; F. C. Neidhardt, Ed.; ASM Press: Washington, DC, 1996; Vol. 1, pp 391–407.
23. S. G. Rhee; P. B. Chock; E. R. Stadtman, *Methods Enzymol.* **1985**, *113*, 213–241.
24. J. E. Anderson; M. Ptashne; S. C. Harrison, *Nature* **1987**, *326*, 846–852.
25. T. A. Steitz, *Q. Rev. Biophys.* **1990**, *23*, 205–280.
26. S. K. Hinchman; S. Henikoff; S. M. Schuster, *J. Biol. Chem.* **1992**, *267*, 144–149.
27. H. Roy; H. D. Becker; J. Reinbolt; D. Kern, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9837–9842.
28. R. Kölling; H. Lother, *J. Bacteriol.* **1985**, *164*, 310–315.
29. J. S. Richardson; D. C. Richardson, *Science* **1988**, *240*, 1648–1652.
30. J. S. Richardson; D. C. Richardson, Principles and Patterns of Protein Conformation. In *Prediction of Protein Structure and the Priciples of Protein Conformation*; G. D. Fasman, Ed.; Plenum Press: New York, 1989; pp 1–98.
31. T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman and Company: New York, 1993.
32. N. E. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5283–5288.
33. D. B. Volkin; A. M. Klibanov, Minimising Protein Inactivation. In *Protein Function: A Practical Approach*; T. E. Creighton, Ed.; IRL Press at Oxford University Press: Oxford, 1989; pp 1–24.
34. A. B. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 885–888.
35. J. H. McKerrow; A. B. Robinson, *Science* **1974**, *183*, 85.
36. N. E. Robinson; A. B. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4367–4372.
37. N. E. Robinson; Z. W. Robinson; B. R. Robinson; A. L. Robinson; J. A. Robinson; M. L. Robinson; A. B. Robinson, *J. Pept. Res.* **2004**, *63*, 426–436.
38. N. E. Robinson; A. B. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12409–12413.
39. S. Chakravarty; R. Varadarajan, *FEBS Lett.* **2000**, *470*, 65–69.
40. S. Kumar; R. Nussinov, *Cell. Mol. Life Sci.* **2001**, *58*, 1216–1233.
41. D. Charlier; L. Droogmans, *Cell. Mol. Life Sci.* **2005**, *62*, 2974–2984.
42. G. Zubay, *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 894–896.
43. J. M. Ravel; S. F. Wang; C. Heinemeyer; W. Shive, *J. Biol. Chem.* **1965**, *240*, 432–438.
44. M. P. Deutscher, *J. Biol. Chem.* **1967**, *242*, 1123–1131.
45. V. Lamour; S. Quevillon; S. Diriong; V. C. N'Guyen; M. Lipinski; M. Mirande, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8670–8674.
46. J. R. Brown; W. F. Doolittle, *J. Mol. Evol.* **1999**, *49*, 485–495.
47. K. C. Rogers; D. Söll, *J. Mol. Evol.* **1995**, *40*, 476–481.
48. M. Siatecka; M. Rozek; J. Barciszewski; M. Mirande, *Eur. J. Biochem.* **1998**, *256*, 80–87.
49. T. Webster; H. Tsai; M. Kula; G. A. Mackie; P. Schimmel, *Science* **1984**, *226*, 1315–1317.
50. J. J. Burbaum; R. M. Starzyk; P. Schimmel, *Proteins: Struct. Funct. Genet.* **1990**, *7*, 99–111.
51. G. Eriani; M. Delarue; O. Poch; J. Gangloff; D. Moras, *Nature* **1990**, *347*, 203–206.
52. M. Ibba; H. D. Becker; C. Stathopoulos; D. L. Tumbula; D. Söll, *Trends Biochem. Sci.* **2000**, *25*, 311–316.
53. M. A. Rould; J. J. Perona; D. Söll; T. A. Steitz, *Science* **1989**, *246*, 1135–1142.
54. D. J. Wright; S. A. Martinis; M. Jahn; D. Söll; P. Schimmel, *Biochimie* **1993**, *75*, 1041–1049.
55. W. H. McClain; J. Schneider; K. Gabriel, *Biochimie* **1993**, *75*, 1125–1136.
56. M. J. Rogers; T. Adachi; H. Inokuchi; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 291–295.
57. M. Kitabatake; M. Ibba; K. W. Hong; D. Söll; H. Inokuchi, *Mol. Gen. Genet.* **1996**, *252*, 717–722.
58. N. T. Uter; J. J. Perona, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14396–14401.
59. M. Deniziak; C. Sauter; H. D. Becker; C. A. Paulus; R. Giegé; D. Kern, *Nucleic Acids Res.* **2007**, *35*, 1421–1431.
60. A. Nakamura; M. Yao; S. Chimnaronk; N. Sakai; I. Tanaka, *Science* **2006**, *312*, 1954–1958.

61. H. Oshikane; K. Sheppard; S. Fukai; Y. Nakamura; R. Ishitani; T. Numata; R. L. Sherrer; L. Feng; E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam; D. Söll; O. Nureki, *Science* **2006**, *312*, 1950–1954.
62. S. W. Ludmerer; P. Schimmel, *J. Biol. Chem.* **1987**, *262*, 10807–10813.
63. S. W. Ludmerer; D. J. Wright; P. Schimmel, *J. Biol. Chem.* **1993**, *268*, 5519–5523.
64. C. C. Wang; P. Schimmel, *J. Biol. Chem.* **1999**, *274*, 16508–16512.
65. R. Fett; R. Knippers, *J. Biol. Chem.* **1991**, *266*, 1448–1455.
66. B. Schray; P. Thömmes; R. Knippers, *Biochim. Biophys. Acta* **1990**, *1087*, 226–234.
67. T. Kim; S. G. Park; J. E. Kim; W. Seol; Y. G. Ko; S. Kim, *J. Biol. Chem.* **2000**, *275*, 21768–21772.
68. Y. G. Ko; E. Y. Kim; T. Kim; H. Park; H. S. Park; E. J. Choi; S. Kim, *J. Biol. Chem.* **2001**, *276*, 6030–6036.
69. T. A. Steitz; M. A. Rould; J. J. Perona, *Mol. Biol. Rep.* **1990**, *14*, 213–214.
70. J. G. Arnez; T. A. Steitz, *Biochemistry* **1996**, *35*, 14725–14733.
71. I. Weygand-Durasević; E. Schwob; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2010–2014.
72. M. A. Rould; J. J. Perona; T. A. Steitz, *Nature* **1991**, *352*, 213–218.
73. T. Seno; P. F. Agris; D. Söll, *Biochim. Biophys. Acta* **1974**, *349*, 328–338.
74. K. C. Rogers; A. T. Crescenzo; D. Söll, *Biochimie* **1995**, *77*, 66–74.
75. K. C. Rogers; D. Söll, *Biochemistry* **1993**, *32*, 14210–14219.
76. M. J. Rogers; I. Weygand-Durasević; E. Schwob; J. M. Sherman; K. C. Rogers; T. Adachi; H. Inokuchi; D. Söll, *Biochimie* **1993**, *75*, 1083–1090.
77. J. J. Perona; M. A. Rould; T. A. Steitz, *Biochemistry* **1993**, *32*, 8758–8771.
78. V. L. Rath; L. F. Silvian; B. Beijer; B. S. Sproat; T. A. Steitz, *Structure* **1998**, *6*, 439–449.
79. T. L. Bullock; N. Uter; T. A. Nissan; J. J. Perona, *J. Mol. Biol.* **2003**, *328*, 395–408.
80. E. M. Corigliano; J. J. Perona, *Biochemistry* **2009**, *48*, 676–687.
81. N. T. Uter; I. Gruic-Sovulj; J. J. Perona, *J. Biol. Chem.* **2005**, *280*, 23966–23977.
82. D. F. Green; B. Tidor, *J. Mol. Biol.* **2004**, *342*, 435–452.
83. M. Wilcox; M. Nirenberg, *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *61*, 229–236.
84. J. Lapointe; L. Duplain; M. Proulx, *J. Bacteriol.* **1986**, *165*, 88–93.
85. Y. Gagnon; L. Lacoste; N. Champagne; J. Lapointe, *J. Biol. Chem.* **1996**, *271*, 14856–14863.
86. A. W. Curnow; K. Hong; R. Yuan; S. Kim; O. Martins; W. Winkler; T. M. Henkin; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11819–11826.
87. H. D. Becker; D. Kern, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12832–12837.
88. A. Schön; C. G. Kannangara; S. Gough; D. Söll, *Nature* **1988**, *331*, 187–190.
89. C. Pujol; M. Bailly; D. Kern; L. Maréchal-Drouard; H. Becker; A. M. Duchêne, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6481–6485.
90. J. Rinehart; B. Krett; M. A. Rubio; J. D. Alfonzo; D. Söll, *Genes Dev.* **2005**, *19*, 583–592.
91. M. Frechin; A. M. Duchene; H. D. Becker, *RNA Biol.* **2009**, *6*, 31–34.
92. A. W. Curnow; D. L. Tumbula; J. T. Pelaschier; B. Min; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12838–12843.
93. D. L. Tumbula; H. D. Becker; W. Z. Chang; D. Söll, *Nature* **2000**, *407*, 106–110.
94. K. Sheppard; R. L. Sherrer; D. Söll, *J. Mol. Biol.* **2008**, *377*, 845–853.
95. O. Nureki; D. G. Vassylyev; K. Katayanagi; T. Shimizu; S. Sekine; T. Kigawa; T. Miyazawa; S. Yokoyama; K. Morikawa, *Science* **1995**, *267*, 1958–1965.
96. S.-i. Sekine; O. Nureki; D. Y. Dubois; S. Bernier; R. Chênevert; J. Lapointe; D. G. Vassylyev; S. Yokoyama, *EMBO J.* **2003**, *22*, 676–688.
97. C. Cerini; P. Kerjan; M. Astier; D. Gratecos; M. Mirande; M. Semeriva, *EMBO J.* **1991**, *10*, 4267–4277.
98. S. Quevillon; J. C. Robinson; E. Berthonneau; M. Siatecka; M. Mirande, *J. Mol. Biol.* **1999**, *285*, 183–195.
99. E. Berthonneau; M. Mirande, *FEBS Lett.* **2000**, *470*, 300–304.
100. H. Simader; M. Hothorn; C. Kohler; J. Basquin; G. Simos; D. Suck, *Nucleic Acids Res.* **2006**, *34*, 3968–3979.
101. M. P. Golinelli-Cohen; M. Mirande, *Mol. Cell. Biochem.* **2007**, *300*, 47–59.
102. I. Sugiura; O. Nureki; Y. Ugaji-Yoshikawa; S. Kuwabara; A. Shimada; M. Tateno; B. Lorber; R. Giegé; D. Moras; S. Yokoyama; M. Konno, *Struct. Folding Des.* **2000**, *8*, 197–208.
103. E. Madore; R. S. Lipman; Y. M. Hou; J. Lapointe, *Biochemistry* **2000**, *39*, 6791–6798.
104. J. E. Ladner; A. Jack; J. D. Robertus; R. S. Brown; D. Rhodes; B. F. Clark; A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4414–4418.
105. S.-i. Sekine; O. Nureki; M. Tateno; S. Yokoyama, *Eur. J. Biochem.* **1999**, *261*, 354–360.
106. S.-i. Sekine; M. Shichiri; S. Bernier; R. Chênevert; J. Lapointe; S. Yokoyama, *Structure* **2006**, *14*, 1791–1799.
107. S.-i. Sekine; O. Nureki; A. Shimada; D. G. Vassylyev; S. Yokoyama, *Nat. Struct. Biol.* **2001**, *8*, 203–206.
108. E. M. Gustilo; D. Y. Dubois; J. Lapointe; P. F. Agris, *RNA Biol.* **2007**, *4*, 85–92.
109. S. I. Kim; D. Söll, *Mol. Cells* **1998**, *8*, 459–465.
110. J. O. Schulze; A. Masoumi; D. Nickel; M. Jahn; D. Jahn; W. D. Schubert; D. W. Heinz, *J. Mol. Biol.* **2006**, *361*, 888–897.
111. R. L. Sidebotham; J. H. Baron, *Lancet* **1990**, *335*, 193–195.
112. B. J. Marshall; L. J. Barrett; C. Prakash; R. W. McCallum; R. L. Guerrant, *Gastroenterology* **1990**, *99*, 697–702.
113. J. F. Tomb; O. White; A. R. Kerlavage; R. A. Clayton; G. G. Sutton; R. D. Fleischmann; K. A. Ketchum; H. P. Klenk; S. Gill; B. A. Dougherty; K. Nelson; J. Quackenbush; L. Zhou; E. F. Kirkness; S. Peterson; B. Loftus; D. Richardson; R. Dodson; H. G. Khalak; A. Glodek; K. McKenney; L. M. Fitzegerald; N. Lee; M. D. Adams; E. K. Hickey; D. E. Berg; J. D. Gocayne; T. R. Utterback; J. D. Peterson; J. M. Kelley; M. D. Cotton; J. M. Weidman; C. Fujii; C. Bowman; L. Watthey; E. Wallin; W. S. Hayes; M. Borodovsky; P. D. Karp; H. O. Smith; C. M. Fraser; J. C. Venter, *Nature* **1997**, *388*, 539–547.
114. S. Skouloubris; L. Ribas de Pouplana; H. De Reuse; T. L. Hendrickson, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11297–11302.
115. J. C. Salazar; I. Ahel; O. Orellana; D. Tumbula-Hansen; R. Krieger; L. Daniels; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13863–13868.
116. J. C. Salazar; R. Zúñiga; G. Raczniak; H. Becker; D. Söll; O. Orellana, *FEBS Lett.* **2001**, *500*, 129–131.

117. J. Lee; T. L. Hendrickson, *J. Mol. Biol.* **2004**, *344*, 1167–1174.
118. H. D. Becker; B. Min; C. Jacobi; G. Raczniak; J. Pelaschier; H. Roy; S. Klein; D. Kern; D. Söll, *FEBS Lett.* **2000**, *476*, 140–144.
119. M. Bailly; S. Giannouli; M. Blaise; C. Stathopoulos; D. Kern; H. D. Becker, *Nucleic Acids Res.* **2006**, *34*, 6083–6094.
120. S. Namgoong; K. Sheppard; R. L. Sherrer; D. Söll, *FEBS Lett.* **2007**, *581*, 309–314.
121. K. Sheppard; D. Söll, *J. Mol. Biol.* **2008**, *377*, 831–844.
122. E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam, *Structure* **2005**, *13*, 1421–1433.
123. M. Delarue; A. Poterszman; S. Nikonov; M. Garber; D. Moras; J. C. Thierry, *EMBO J.* **1994**, *13*, 3219–3229.
124. S. Eiler; A.-C. Dock-Bregeon; L. Moulinier; J.-C. Thierry; D. Moras, *EMBO J.* **1999**, *18*, 6532–6541.
125. L. Feng; K. Sheppard; D. Tumbula-Hansen; D. Söll, *J. Biol. Chem.* **2005**, *280*, 8150–8155.
126. S. Mouilleron; B. Golinelli-Pimpaneau, *Curr. Opin. Struct. Biol.* **2007**, *17*, 653–664.
127. C. Hedgcoth; J. Ravel; W. Shive, *Biochem. Biophys. Res. Commun.* **1963**, *13*, 495–499.
128. S. J. Norton; J. M. Ravel; C. Lee; W. Shive, *J. Biol. Chem.* **1963**, *238*, 269–274.
129. G. Nass; G. Stöffler, *Mol. Gen. Genet.* **1967**, *100*, 378–382.
130. K. H. Muench; P. A. Safille, *Biochemistry* **1968**, *7*, 2799–2808.
131. W. E. Barnett; D. H. Brown; J. L. Epler, *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1775–1781.
132. M. R. Davies; R. D. Marshall, *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1386–1395.
133. H. Grojean; J. Charlier; C. Darte; G. Dirheimer; R. Giegé; S. de Henau; G. Keith; R. Parfait; C. Takada, *Experientia Suppl.* **1976**, *26*, 347–362.
134. T. Samuelsson; L. Lundvik, *J. Biol. Chem.* **1978**, *253*, 7033–7039.
135. I. L. Andrulis; C. S. Chiang; S. M. Arfin; T. A. Miner; G. W. Hatfield, *J. Biol. Chem.* **1978**, *253*, 58–62.
136. D. Madern; J. Anselme; M. Härtlein, *FEBS Lett.* **1992**, *299*, 85–89.
137. M. R. Davies; R. D. Marshall, *Biochim. Biophys. Acta* **1975**, *390*, 94–104.
138. L. Seignovert; M. Hartlein; R. Leberman, *Eur. J. Biochem.* **1996**, *239*, 501–508.
139. W. Iwasaki; S. Sekine; C. Kuroishi; S. Kuramitsu; M. Shirouzu; S. Yokoyama, *J. Mol. Biol.* **2006**, *360*, 329–342.
140. J. Anselme; M. Härtlein, *Gene* **1989**, *84*, 481–485.
141. C. Berthet-Colominas; L. Seignovert; M. Hartlein; M. Grotli; S. Cusack; R. Leberman, *EMBO J.* **1998**, *17*, 2947–2960.
142. D. Kern; H. Roy; H. D. Becker, Asparaginyl-tRNA Synthetases. In *The Aminoacyl-tRNA Synthetases*; M. Ibba, C. Francklyn, S. Cusack, Eds.; Eurekah.com/Landes Bioscience: Georgetown: TX, 2005; pp 193–209.
143. U. Grädler; H. D. Gerber; D. M. Goodenough-Lashua; G. A. Garcia; R. Ficner; K. Reuter; M. T. Stubbs; G. Klebe, *J. Mol. Biol.* **2001**, *306*, 455–467.
144. B. A. Carlson; S. Y. Kwon; B. J. Lee; D. Hatfield, *Mol. Cells* **2000**, *10*, 113–118.
145. H. D. Becker; H. Roy; L. Moulinier; M.-H. Mazauric; G. Keith; D. Kern, *Biochemistry* **2000**, *39*, 3216–3230.
146. X.-R. Gu; K. Nicoghosian; R. J. Cedergren, *FEBS Lett.* **1984**, *176*, 462–466.
147. K. Nicoghosian; X.-R. Gu; R. Cedergren, *FEBS Lett.* **1985**, *193*, 255–260.
148. N. Horie; M. Hara-Yokoyama; S. Yokoyama; K. Watanabe; Y. Kuchino; S. Nishimura; T. Miyazawa, *Biochemistry* **1985**, *24*, 5711–5715.
149. B. S. Choi; A. G. Redfield, *Biochemistry* **1986**, *25*, 1529–1534.
150. G. Keith; M. Yusupov; C. Briand; D. Moras; D. Kern; C. Brion, *Nucleic Acids Res.* **1993**, *21*, 4399.
151. M.-H. Mazauric, Le système de la glycylation de l'ARNt de *Thermus thermophilus.* Études structurales et fonctionnelles et interrelation avec d'autres systèmes de glycylation. Ph.D. Thesis, Université Louis Pasteur, Strasbourg, France, 1997.
152. M. Kron; K. Marquard; M. Härtlein; S. Price; R. Leberman, *FEBS Lett.* **1995**, *374*, 122–124.
153. I. Landrieu; M. Vandenbol; R. Leberman; D. Portetelle; M. Härtlein, *Yeast* **1998**, *14*, 527–533.
154. M. Beaulande; N. Tarbouriech; M. Härtlein, *Nucleic Acids Res.* **1998**, *26*, 521–524.
155. H. Roy, Évolution des voies de formation d'asparagine et d'asparaginylation de l'ARNt. Ph.D. Thesis, Université Louis Pasteur, Strasbourg, France, 2002.
156. S. Li; H. Pelka; L. H. Schulman, *J. Biol. Chem.* **1993**, *268*, 18335–18339.
157. K. Tamura; H. Himeno; H. Asahara; T. Hasegawa; M. Shimizu, *Nucleic Acids Res.* **1992**, *20*, 2335–2339.
158. F. Martin; G. Eriani; J. Reinbolt; G. Dirheimer; J. Gangloff, *Nucleic Acids Res.* **1995**, *23*, 779–784.
159. J. Normanly; J. Abelson, *Annu. Rev. Biochem.* **1989**, *58*, 1029–1049.
160. M. Shimizu; H. Asahara; K. Tamura; T. Hasegawa; H. Himeno, *J. Mol. Evol.* **1992**, *35*, 436–443.
161. M. Ruff; S. Krishnaswamy; M. Boeglin; A. Poterszman; A. Mitschler; A. Podjarny; B. Rees; J. C. Thierry; D. Moras, *Science* **1991**, *252*, 1682–1689.
162. J. Cavarelli; B. Rees; M. Ruff; J. C. Thierry; D. Moras, *Nature* **1993**, *362*, 181–184.
163. G. Eriani; J. Gangloff, *J. Mol. Biol.* **1999**, *291*, 761–773.
164. H. D. Becker; R. Giegé; D. Kern, *Biochemistry* **1996**, *35*, 7447–7458.
165. W. H. McClain; K. Foss; R. A. Jenkins; J. Schneider, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9260–9264.
166. K. Shiba; T. Stello; H. Motegi; T. Noda; K. Musier-Forsyth; P. Schimmel, *J. Biol. Chem.* **1997**, *272*, 22809–22816.
167. M. Francin; M. Mirande, *Biochemistry* **2006**, *45*, 10153–10160.
168. D. Söll; H. D. Becker; P. Plateau; S. Blanquet; M. Ibba, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14224–14228.
169. L. Feng; J. Yuan; H. Toogood; D. Tumbula-Hansen; D. Söll, *J. Biol. Chem.* **2005**, *280*, 20638–20641.
170. A. W. Curnow; M. Ibba; D. Söll, *Nature* **1996**, *382*, 589–590.
171. H. D. Becker, Mise en évidence et propriétés de deux aspartyl-ARNt synthétases dans *Thermus thermophilus.* Un exemple de coexistence d'une voie ancestrale et d'une voie moderne d'aminoacylation des ARNt. Ph.D. Thesis, Université Louis Pasteur, Strasbourg, France, 1998.
172. H. Roy; H. D. Becker; M.-H. Mazauric; D. Kern, *Nucleic Acids Res.* **2007**, *35*, 3420–3430.
173. R. Giegé; B. Rees, Aspartyl-tRNA Synthetases. In *The Aminoacyl-tRNA Synthetases*; M. Ibba, C. Francklyn, S. Cusack, Eds.; Eurekah.com/Landes Bioscience: Georgetown: TX, 2005; pp 210–226.
174. M. Frugier; L. Moulinier; R. Giegé, *EMBO J.* **2000**, *19*, 2371–2380.
175. M. Ryckelynck; R. Giegé; M. Frugier, *J. Biol. Chem.* **2003**, *278*, 9683–9690.

176. E. Schmitt; L. Moulinier; S. Fujiwara; T. Imanaka; J.-C. Thierry; D. Moras, *EMBO J.* **1998**, *17*, 5227–5237.
177. P. Chuawong; T. L. Hendrickson, *Biochemistry* **2006**, *45*, 8079–8087.
178. G. Raczniak; H. D. Becker; B. Min; D. Söll, *J. Biol. Chem.* **2001**, *276*, 45862–45867.
179. D. Tumbula-Hansen; L. Feng; H. Toogood; K. O. Stetter; D. Söll, *J. Biol. Chem.* **2002**, *277*, 37184–37190.
180. C. Charron; H. Roy; M. Blaise; R. Giegé; D. Kern, *EMBO J.* **2003**, *22*, 1632–1643.
181. C. Briand; A. Poterszman; S. Eiler; G. Webster; J. Thierry; D. Moras, *J. Mol. Biol.* **2000**, *299*, 1051–1060.
182. L. Moulinier; S. Eiler; G. Eriani; J. Gangloff; J.-C. Thierry; K. Gabriel; W. H. McClain; D. Moras, *EMBO J.* **2001**, *20*, 5290–5301.
183. J. Pütz; J. D. Puglisi; C. Florentz; R. Giegé, *Science* **1991**, *252*, 1696–1699.
184. R. Giegé; M. Sissler; C. Florentz, *Nucleic Acids Res.* **1998**, *26*, 5017–5035.
185. D. Bernard; P. M. Akochy; D. Beaulieu; J. Lapointe; P. H. Roy, *J. Bacteriol.* **2006**, *188*, 269–274.
186. R. Giegé; C. Florentz; D. Kern; J. Gangloff; G. Eriani; D. Moras, *Biochimie* **1996**, *78*, 605–623.
187. G. Dirheimer; G. Keith; A.-P. Sibler; R. P. Martin, The Primary Structure of tRNAs and Their Rare Nucleosides. In *Transfer RNA: Structure, Properties, and Recognition*; P. Schimmel, D. Söll, J. N. Abelson, Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1979; Vol. 9A, pp 19–41.
188. P. M. Akochy; D. Bernard; P. H. Roy; J. Lapointe, *J. Bacteriol.* **2004**, *186*, 767–776.
189. K. Sheppard; P. M. Akochy; J. C. Salazar; D. Söll, *J. Biol. Chem.* **2007**, *282*, 11866–11873.
190. M. Kobayashi; Y. Fujiwara; M. Goda; H. Komeda; S. Shimizu, *Proc. Natl. Acad. Sci. U. S.A.* **1997**, *94*, 11986–11991.
191. C. P. Decicco; D. J. Nelson; Y. Luo; L. Shen; K. Y. Horiuchi; K. M. Amsler; L. A. Foster; S. M. Spitz; J. J. Merrill; C. F. Sizemore; K. C. Rogers; R. A. Copeland; M. R. Harpel, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2561–2564.
192. M. R. Harpel; K. Y. Horiuchi; Y. Luo; L. Shen; W. Jiang; D. J. Nelson; K. C. Rogers; C. P. Decicco; R. A. Copeland, *Biochemistry* **2002**, *41*, 6398–6407.
193. K. Y. Horiuchi; M. R. Harpel; L. Shen; Y. Luo; K. C. Rogers; R. A. Copeland, *Biochemistry* **2001**, *40*, 6450–6457.
194. D. Jahn; Y. C. Kim; Y. Ishino; M. W. Chen; D. Söll, *J. Biol. Chem.* **1990**, *265*, 8059–8064.
195. M. Wilcox, *Eur. J. Biochem.* **1969**, *11*, 405–412.
196. M. Bailly; M. Blaise; B. Lorber; H. D. Becker; D. Kern, *Mol. Cell* **2007**, *28*, 228–239.
197. M. Bailly; M. Blaise; H. Roy; M. Deniziak; B. Lorber; C. Birck; H. D. Becker; D. Kern, *Methods* **2008**, *44*, 146–163.
198. J. L. Huot; C. Balg; D. Jahn; J. Moser; A. Émond; S. P. Blais; R. Chênevert; J. Lapointe, *Biochemistry* **2007**, *46*, 13190–13198.
199. M. Bailly, Nouveaux aspect des voies indirectes de formation des aa-ARNt amidés. Ph.D. Thesis, Université Louis Pasteur, Strasbourg, France, 2008.
200. A. Böck; M. Thanbichler; M. Rother; A. Resch, Selenocysteine. In *The Aminoacyl-tRNA Synthetases*; M. Ibba, C. Francklyn, S. Cusack, Eds.; Eurekah.com/Landes Bioscience: Georgetown: TX, 2005; pp 320–327.
201. C. Allmang; A. Krol, *Biochimie* **2006**, *88*, 1561–1571.
202. F. Chapeville; F. Lipmann; G. Von Ehrenstein; B. Weisblum; W. J. Ray, Jr.; S. Benzer, *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 1086–1092.
203. M. Stanzel; A. Schön; M. Sprinzl, *Eur. J. Biochem.* **1994**, *219*, 435–439.
204. T. J. Cathopoulis; P. Chuawong; T. L. Hendrickson, *Biochemistry* **2008**, *47*, 7610–7616.
205. F. J. LaRiviere; A. D. Wolfson; O. C. Uhlenbeck, *Science* **2001**, *294*, 165–168.
206. H. Asahara; O. C. Uhlenbeck, *Biochemistry* **2005**, *44*, 11254–11261.
207. M. Pelchat; L. Lacoste; F. Yang; J. Lapointe, *Can. J. Microbiol.* **1998**, *44*, 378–381.
208. J. W. Baick; J. H. Yoon; S. Namgoong; D. Söll; S. I. Kim; S. H. Eom; K. W. Hong, *J. Microbiol.* **2004**, *42*, 111–116.
209. H. Núñez; C. Lefimil; B. Min; D. Söll; O. Orellana, *FEBS Lett.* **2004**, *557*, 133–135.
210. B. Min; M. Kitabatake; C. Polycarpo; J. Pelaschier; G. Raczniak; B. Ruan; H. Kobayashi; S. Namgoong; D. Söll, *J. Bacteriol.* **2003**, *185*, 3524–3526.
211. S. Cusack; C. Berthet-Colominas; M. Hartlein; N. Nassar; R. Leberman, *Nature* **1990**, *347*, 249–255.
212. C. W. Carter, Jr., *Annu. Rev. Biochem.* **1993**, *62*, 715–748.
213. J. G. Arnez; D. Moras, *Trends Biochem. Sci.* **1997**, *22*, 211–216.
214. G. M. Nagel; R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8121–8125.
215. F. Cramer; W. Freist, *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 190–200.
216. P. Schimmel; L. Ribas de Pouplana, *Cold Spring Harb. Symp. Quant. Biol.* **2001**, *66*, 161–166.
217. L. Ribas de Pouplana; P. Schimmel, *Trends Biochem. Sci.* **2001**, *26*, 591–596.
218. D. Y. Dubois; M. Blaise; H. D. Becker; V. Campanacci; G. Keith; R. Giegé; C. Cambillau; J. Lapointe; D. Kern, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7530–7535.
219. C. R. Woese; G. J. Olsen; M. Ibba; D. Söll, *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 202–236.
220. I. K. Jordan; E. V. Koonin, Horizontal Gene Transfer and Prokaryotic Genome Evolution. In *Microbial Evolution; Gene Establishment, Survival, and Exchange*; R. V. Millner, M. J. Day, Eds.; ASM Press: Washington, DC, 2004; pp 319–338.
221. F. Agou; S. Quevillon; P. Kerjan; M. Mirande, *Biochemistry* **1998**, *37*, 11309–11314.
222. M. Ibba; D. Söll, *Genes Dev.* **2004**, *18*, 731–738.
223. R. Chênevert; S. Bernier; J. Lapointe, Inhibitors of Aminoacyl-tRNA Synthetases as Antibiotics and Tools for Structural and Mechanistic Studies. In *Translation Mechanisms*; J. Lapointe, L. Brakier-Gingras, Eds.; Eurekah.com/Landes Bioscience and Kluwer Academic/Plenum Publishers: Georgetown: TX, 2003; pp 416–428.
224. D. J. Payne; M. N. Gwynn; D. J. Holmes; D. L. Pompliano, *Nat. Rev. Drug Discov.* **2007**, *6*, 29–40.
225. J. Finn; M. Stidham; M. Hilgers; G. C. Kedar, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3932–3937.
226. S. Bernier; D. Y. Dubois; M. Therrien; J. Lapointe; R. Chênevert, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2441–2444.
227. C. Balg; S. P. Blais; S. Bernier; J. L. Huot; M. Couture; J. Lapointe; R. Chênevert, *Bioorg. Med. Chem.* **2007**, *15*, 295–304.
228. S. C. Sukuru; T. Crepin; Y. Milev; L. C. Marsh; J. B. Hill; R. J. Anderson; J. C. Morris; A. Rohatgi; G. O'Mahony; M. Grøtli; F. Danel; M. G. Page; M. Härtlein; S. Cusack; M. A. Kron; L. A. Kuhn, *J. Comput. Aided Mol. Des.* **2006**, *20*, 159–178.
229. M. Desjardins; S. Garneau; J. Desgagnés; L. Lacoste; F. Yang; J. Lapointe; R. Chênevert, *Bioorg. Chem.* **1998**, *26*, 1–13.
230. S. Bernier; D. Y. Dubois; C. Habegger-Polomat; L.-P. Gagnon; J. Lapointe; R. Chênevert, *J. Enzyme Inhib. Med. Chem.* **2005**, *20*, 61–67.

231. J. Tao; P. Schimmel, *Expert Opin. Investig. Drugs* **2000**, *9*, 1767–1775.
232. M. Desjardins; J. Desgagnés; L. Lacoste; F. Yang; M.-P. Morin; J. Lapointe; R. Chênevert, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2363–2366.
233. S. Bernier; P.-M. Akochy; J. Lapointe; R. Chênevert, *Bioorg. Med. Chem.* **2005**, *13*, 69–75.
234. D. Bernard; P.-M. Akochy; S. Bernier; O. Fisette; O. Côté-Brousseau; R. Chênevert; P. H. Roy; J. Lapointe, *J. Enzyme Inhib. Med. Chem.* **2007**, *22*, 77–82.
235. J. I. Guijarro; J. E. González-Pastor; F. Baleux; J. L. San Millán; M. A. Castilla; M. Rico; F. Moreno; M. Delepierre, *J. Biol. Chem.* **1995**, *270*, 23520–23532.
236. M. Novikova; A. Metlitskaya; K. Datsenko; T. Kazakov; A. Kazakov; B. Wanner; K. Severinov, *J. Bacteriol.* **2007**, *189*, 8361–8365.
237. T. Kazakov; G. H. Vondenhoff; K. A. Datsenko; M. Novikova; A. Metlitskaya; B. L. Wanner; K. Severinov, *J. Bacteriol.* **2008**, *190*, 2607–2610.
238. A. Metlitskaya; T. Kazakov; A. Kommer; O. Pavlova; M. Praetorius-Ibba; M. Ibba; I. Krasheninnikov; V. Kolb; I. Khmel; K. Severinov, *J. Biol. Chem.* **2006**, *281*, 18033–18042.
239. C. Balg; J. L. Huot; J. Lapointe; R. Chênevert, *J. Am. Chem. Soc.* **2008**, *130*, 3264–3265.
240. U. A. Ochsner; X. Sun; T. Jarvis; I. Critchley; N. Janjic, *Expert Opin. Investig. Drugs* **2007**, *16*, 573–593.
241. J. Finn; J. Tao, Aminoacyl-tRNA Synthetases as Anti-infective Drug Targets. In *The Aminoacyl-tRNA Synthetases*; M. Ibba, C. Francklyn, S. Cusack, Eds.; Eurekah.com/Landes Bioscience: Georgetown, TX, 2005; pp 405–413.
242. J. G. Hurdle; A. J. O'Neill; I. Chopra, *Antimicrob. Agents Chemother.* **2005**, *49*, 4821–4833.
243. J. Pohlmann; H. Brötz-Oesterhelt, *Curr. Drug Targets Infect. Disord.* **2004**, *4*, 261–272.
244. S. Kim; S. W. Lee; E.-C. Choi; S. Y. Choi, *Appl. Microbiol. Biotechnol.* **2003**, *61*, 278–288.
245. J. Hughes; G. Mellows, *Biochem. J.* **1980**, *191*, 209–219.
246. P. Van de Vijver; G. H. Vondenhoff; S. Denivelle; J. Rozenski; J. Verhaegen; A. Van Aerschot; P. Herdewijn, *Bioorg. Med. Chem.* **2009**, *17*, 260–269.
247. P. Van de Vijver; T. Ostrowski; B. Sproat; J. Goebels; O. Rutgeerts; A. Van Aerschot; M. Waer; P. Herdewijn, *J. Med. Chem.* **2008**, *51*, 3020–3029.
248. C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8742–8747.
249. C. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6854–6859.
250. J. Yuan; S. Palioura; J. C. Salazar; D. Su; P. O'Donoghue; M. J. Hohn; A. M. Cardoso; W. B. Whitman; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18923–18927.
251. Y. I. Wolf; L. Aravind; N. V. Grishin; E. V. Koonin, *Genome Res.* **1999**, *9*, 689–710.

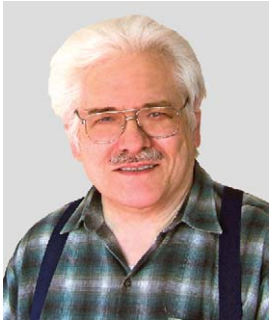
Biographical Sketches

Jonathan L. Huot earned a B.Sc. in microbiology from Université Laval in Quebec City, Canada, in 2005. He is currently a doctoral fellow of the Fonds québécois de la recherche sur la nature et les technologies working with Professor Jacques Lapointe at Université Laval, where he is investigating the pathways catalyzed in bacteria by the aminoacyl-tRNA amidotransferase (GatCAB) and its substrate-forming enzymes, the nondiscriminating aminoacyl-tRNA synthetases.

Jacques Lapointe is Professor of Biochemistry at Université Laval, Québec, Canada. He did his undergraduate studies in physics at the Université de Montréal, his graduate studies (Ph.D. 1971) in the field of protein biosynthesis in the Department of Molecular Biophysics and Biochemistry of Yale University (New Haven, CT, USA), and his postdoctoral studies in genetics at the Université Libre de Bruxelles (Belgium). His research group studies protein biosynthesis, mostly by characterizing the reaction mechanisms of the multidomain enzymes that catalyze the aminoacylation of tRNAs in bacteria: the discriminating and the nondiscriminating aminoacyl-tRNA synthetases, and the aminoacyl-tRNA amidotransferases. In collaboration with several colleagues in Québec, France, Japan, the USA, and China, they study these essential enzymes and their interactions with their substrates and with inhibitors analogous to substrates or reaction intermediates, using structure/function, kinetic, optical, and microcalorimetric approaches. In 1986, his laboratory published the first report on a nondiscriminating aminoacyl-tRNA synthetase (*J. Bacteriol.166*, 88–93).

Robert Chênevert is Professor of Organic Chemistry at Université Laval, Québec, Canada. He studied chemistry (B.Sc. and M.Sc.) at the Université de Montréal. After receiving his Ph.D. in organic chemistry in 1975 at the Université de Sherbrooke under the supervision of Professor Pierre Deslongchamps, he spent a postdoctoral year at Harvard (R. B. Woodward's group). His main research interest is the application of biocatalysts in asymmetric synthesis. He is also interested in the design of inhibitors of enzymes involved in the aminoacylation of tRNA (aminoacyl-tRNA synthetases and aminoacyl-tRNA amidotransferases).

Marc Bailly holds a Ph.D. in molecular and cellular biology. He did his undergraduate studies in biochemistry at the University Louis Pasteur in Strasbourg (France). He did his graduate studies and obtained Ph.D. in 2008 in the team 'Evolution of the translation machinery and of the genetic code' headed by Professor Daniel Kern in the UPR 9002 (Architecture et Réactivité de l' ARN) at the Institut de Biologie Moléculaire et Cellulaire (IBMC) of the CNRS. During the preparation of his thesis, he was a teaching assistant in biochemistry. Now, he is a postdoctoral fellow in the Microbiology and Cell Science Department of Florida University (Gainesville, USA), where he is working with Professor Valérie De Crecy Lagard's group, on predicting functions of genes by using genetic and bioinformatics tools.



Daniel Kern is Professor of Biochemistry at the Université de Strasbourg (France) where he manages the Master of Biochemistry, Molecular and Cellular Biology. He did his undergraduate studies in biochemistry and biology at the Université Louis Pasteur in Strasbourg and obtained the grade of Docteur d'État és Sciences in 1981 in the field of biochemistry at the Institut de Biologie Moléculaire et Cellulaire (IBMC) du CNRS in Jean-Pierre Ebel's laboratory. His thesis was devoted to the mechanisms of tRNA aminoacylation and selection of the cognate substrates by the aminoacyl-tRNA synthetases. He is leading the research group 'Evolution of the translation machinery and of the genetic code' in the UPR 9002 (Architecture et Réactivité de l' ARN) of the CNRS (IBMC). The team investigates the atypical pathways of tRNA aminoacylation and the paralogues of the partners of the translation machinery, and studies functional interrelations between the translation machinery and other biological processes. In collaboration with other groups in France, Québec, Japan, Greece, and the USA it analyzes the structural and functional properties of the atypical partners involved in translation – nondiscriminating aminoacyl-tRNA synthetases, paralogues of aminoacyl-tRNA synthetases, tRNA-dependent amidotransferases, and so on, – and investigates their interaction with other cellular partners using biophysical, biochemical, and *in vivo* approaches.

5.15 Posttranslational Modification of Proteins

Keith D. Green and Sylvie Garneau-Tsodikova, University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

5.15.1 Introduction

Proteins are the workhorses of the cell, performing most of the reactions for the cell to carry out its daily function. The size of a protein can range from 70 amino acids to subunits as large as 15 000 amino acids roughly ranging in molecular weight from 7 kD to 1.7 MDa. Since the genome contains 30 000 genes, there exists at least an identical number of proteins. Covalent posttranslational modifications are the methods that the cells utilize to diversify the limited number of proteins that can be generated, altering the activity and functions of the proteins. A posttranslational modification is, as suggested, a modification to a protein after its translation from messenger ribonucleic acid (mRNA) to its specific sequence, transforming the 20 natural amino acids found in the body into a much more diversified collective. Shown in **Table 1** is a list of posttranslational modifications and the residues onto which they occur. It is important to point out that there are no known modifications to alanine, isoleucine, leucine, phenylalanine, and valine side chains.

The most common posttranslational modifications, discussed in the following sections, include phosphorylation, sulfation, disulfide formation, N-methylation, O-methylation, S-methylation, N-acetylation, hydroxylation, glycosylation, ADP-ribosylation, prenylation, biotinylation, lipoylation, and phosphopantetheine tethering. Many of the posttranslational modifications are proven to be cross talks.[1,2] Other modifications exist in a smaller extent and include oxidation of methionine, C-methylation, ubiquitylation, carboxylation, and amidation. These topics will not be covered in this chapter which is meant to focus primarily on the recent literature (2005–08). For a more complete coverage of all posttranslational modifications and earlier literature (up to 2005), the reader is referred to Professor Christopher T. Walsh's book *Posttranslational Modification of Proteins: Expanding Nature's Inventory*.[3,4]

Table 1 Posttranslational modifications by amino acid side chain modified

Amino acid	*Reaction*	*Amino acid*	*Reaction*
Arg	N-methylation	His	Phosphorylation
	N-ADP ribosylation		Aminocarboxypropylation
Asn	N-glycosylation		N-methylation
	N-ADP ribosylation	Lys	N-methylation
	Protein splicing		N-acylation by acetyl, biotinyl, lipoyl, ubiquitinyl groups
Asp	Phosphorylation		C-hydroxylation
	Isomerization to isoAsp	Met	Oxidation to sulfoxide
Cys	S-hydroxylation (S-OH)	Pro	C-hydroxylation
	Disulfide bond formation	Ser	Phosphorylation
	Phosphorylation		O-glycosylation
	S-acylation		Phosphopantetheinylation
	S-prenylation		Autocleavages
	Protein splicing	Thr	Phosphorylation
Gln	Transglutamination	Trp	C-mannosylation
Glu	Methylation	Tyr	Phosphorylation
	Carboxylation		Sulfation
	Polyglycination		Ortho-nitration
	Polyglutamylation		TOPA quinone
Gly	C-hydroxylation		

No known modifications on Ala, Ile, Leu, Phe, and Val side chains.

Posttranslational modifications can be broken down into two main classes: those that are reversible and those that are irreversible. Included in the large group of reversible posttranslational modifications are phosphorylation, acetylation, and disulfide formation. Irreversible posttranslational modifications include peptide bond cleavage as in intein splicing; also irreversible is the introduction of a phosphopantetheinyl group during fatty acid, polyketide, and nonribosomal peptide biosyntheses. The current debate is whether to classify lysine N-methylation as reversible or irreversible. Recently, there have been reports of lysine demethylases.[5]

Posttranslational modifications would not be possible without small molecule high-energy cosubstrates to serve as a source of electrophiles for the nucleophilic proteins. Depicted in **Figure 1** are the structures of several cosubstrates, including adenosine 5-triphosphate (ATP) used for phosphorylation, 3′-phosphoadenosine-5′-phosphosulfate (PAPS) used in sulfation, acetyl-coenzyme A (AcCoA) specifically used in acetylation, but other derivatives are used in acylations, 5′-nicotinamide 5′-adenosine diphosphate (NAD^+) used in oxidations, reductions and ADP-ribosylations, and *S*-adenosylmethionine (SAM) used in methylation. Other molecules shown are (2-*N*-acetyl)-glucosamine uracil 5′-diphosphate (UDP-GlcNAc) and glucose uracil 5′-diphosphate (UDP–glucose) used for the glycosylation of serine.

It is important to note that the enzymes responsible for posttranslational modifications have been proven to be promiscuous with regard to their consumed cosubstrates. Kinases, the enzymes responsible for phosphorylation, have displayed remarkable promiscuity in accepting γ-phosphate-modified ATP derivative in place of ATP.[6–8] In a similar fashion, galactosyltransferase, which normally transfers galactose to a terminal *N*-acetyl glucosamine (GlcNAc), has been shown to accept a galactose derivative with a reactive ketone.[9] Another modification to protein cosubstrates that has been explored using *N*-acetyltransferases is the transfer of a 2-chloroacetyl group to amino functionalities instead of the normal acetyl moiety.[10] Yet another example of enzymes exploiting cosubstrate promiscuity are methyltransferases, using derivatized *S*-adenosylmethionine.[11] Cosubstrate promiscuity has also been reported for farnesyltransferase (FTase) using propargyl ethers appended to the end of the farnesyl group.[12] 4′-phosphopantetheinyltransferase (PPTase), using AcCoA, can also use a maleimide-CoA.[13] The maleimide was derivatized with fluorophores, biotin, or a sugar to identify the phosphopantetheinylated proteins either directly in the case of fluorophores or through a secondary protein interaction using streptavidin (biotin) or concanavalin that binds mannose. The promiscuity of these enzymes has served as a valuable tool for tracking the abundant posttranslational modifications.

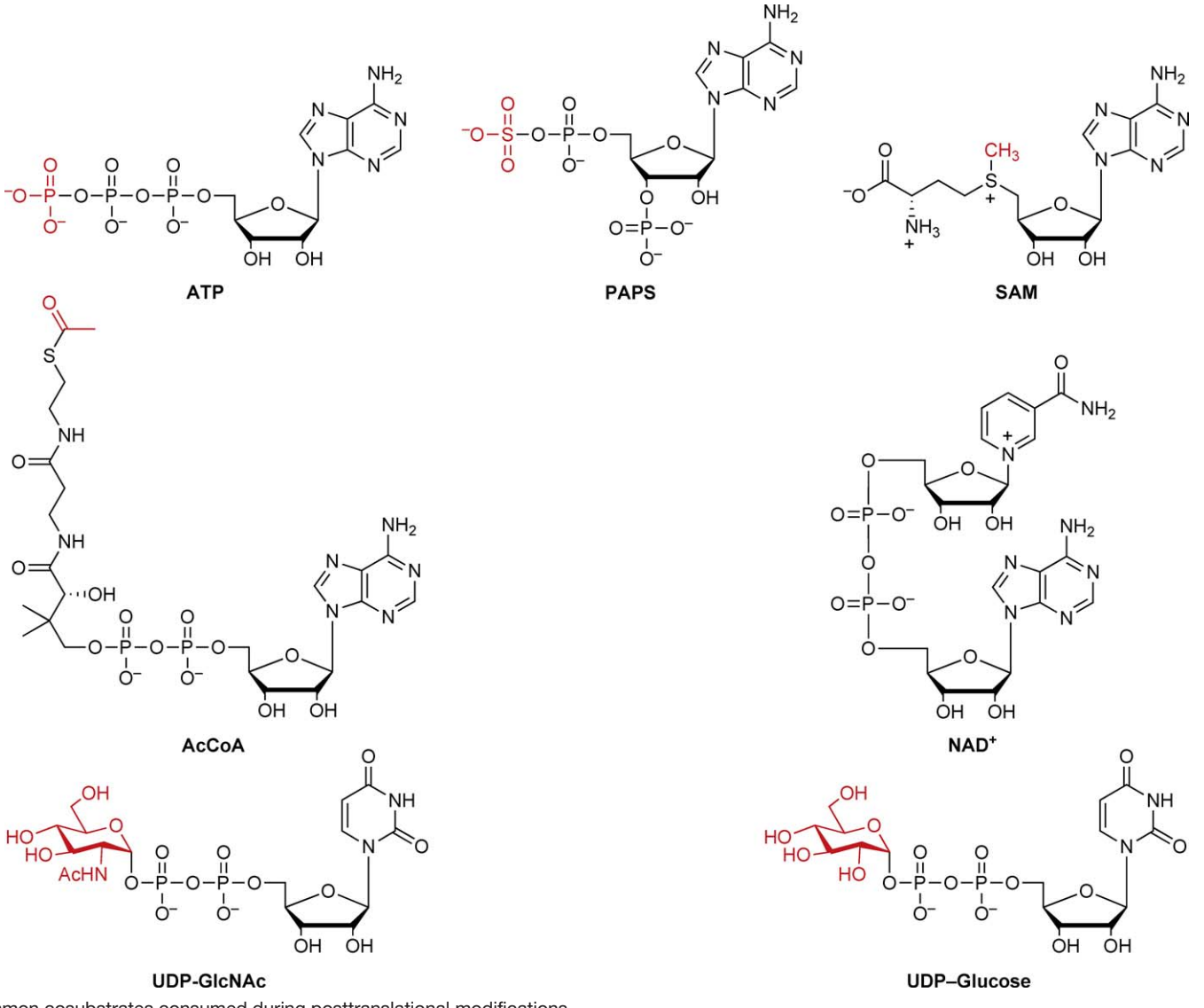


Figure 1 Common cosubstrates consumed during posttranslational modifications.

5.15.2 Phosphorylation and Sulfation

5.15.2.1 Phosphorylation

Phosphorylation in cells involves the transfer of phosphate groups from a high-energy molecule, usually ATP, but in some instances guanine 5′-triphosphate (GTP), to specific amino acid residues or small molecules. The phosphorylation reaction is catalyzed by a class of enzymes known as kinases. Proteins are phosphorylated at the hydroxyl-containing amino acid residues serine, threonine, and tyrosine, as well as at other amino acids in much smaller amounts (**Scheme 1(a)**).[14] The two mechanisms (associative[15] or dissociative[16]) by which the phosphorylation reactions occur are Mg^{2+}-dependent and proceed in three steps: (1) substrate binding to the kinase–ATP complex, (2) transfer of γ-phosphate from ATP to the hydroxyl moiety of the amino acid substrate, and (3) release of the phosphorylated substrate from the kinase–ADP complex (**Scheme 1(b)**).[17] At any time during the functional existence of a protein, one or more of these amino acid residues can be phosphorylated. While some proteins are phosphorylated immediately after translation and for their entire lifetime, others may be transiently phosphorylated and in some cases only up to 0.5% of the protein is phosphorylated.[18] Phosphorylation is an important posttranslational modification responsible for many biological functions. With an estimated 500 protein kinases[19] and 100 000 phosphorylation consensus sites in the genome,[20] discovery of phosphorylation sites and the study of their formation are widely valued.

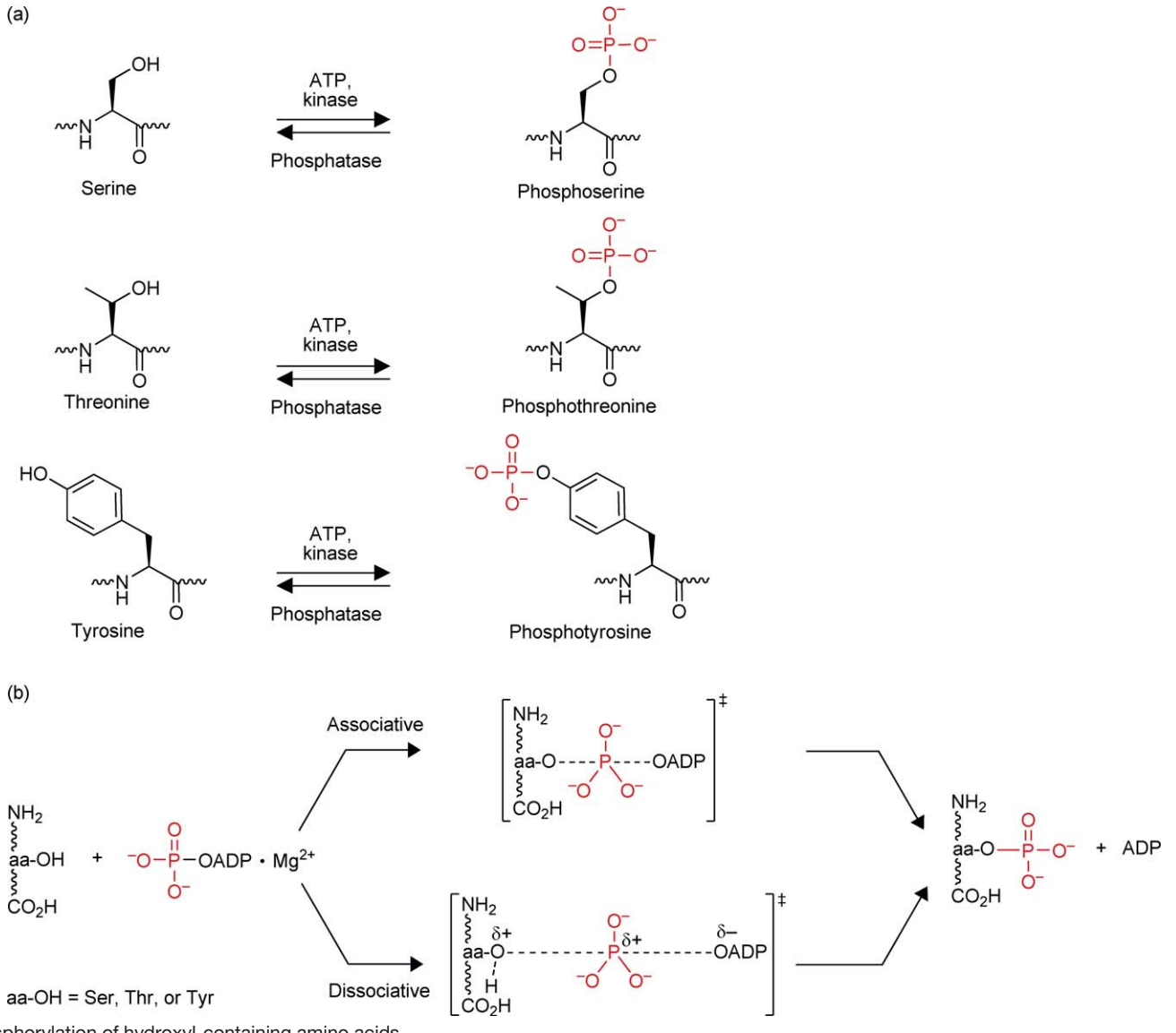
Phosphorylation, like most biological processes, is dynamic. The enzymes responsible for the removal of phosphate moieties are known as phosphatases. There are remarkably fewer phosphatases than kinases (about half), and unsurprisingly phosphatases are less specific than kinases.[21] The dynamic nature of phosphorylation comes into effect when considering signal transduction *in vivo*. Kinases and phosphatases are complementary to each other, where one enzyme can turn a signal 'on' and the complementary enzyme will turn the signal 'off'.

Based on their substrate specificity, kinases have been divided into two broad classes: serine/threonine- and tyrosine-specific kinases. Of the estimated 500 kinases, roughly one-fifth of these are tyrosine kinases. The two classes are further broken down and grouped by function, required cofactors, and substrates.[21]

Despite the various functions and biological requirements of protein kinases, the primary structures of the catalytic site are very similar, with short amino acid stretches unique to each class.[22] Nine amino acids are conserved throughout protein kinase active sites.[23] These amino acids include Gly52, Lys72, Glu91, Asp166, Asn171, Asp184, Gly186, Glu208, and Arg280 (numbering corresponds to protein kinase A (PKA)). Six of these amino acids are responsible for stabilizing the triphosphate (**Figure 2(a)**). The remaining amino acids are unsurprisingly responsible for the binding of the adenosine portion of ATP. In addition, throughout the family of kinases, the core structure is highly conserved.[20] There are two domains: an α-helical portion where substrate binding occurs and a smaller portion comprised of β-sheets. These two domains are joined by a linker segment that produces an ATP-binding pocket (**Figure 2(b)**).[24] ATP is nestled deep in the pocket of the highly conserved catalytic site (**Figure 2(a)**).[24] Upon binding of ATP and substrate, the two domains close together, causing both the protein substrate and ATP to bind tighter to the kinase.[25] The catalytic site employs multiple charged residues to chelate two magnesium ions that aid in the stabilization of the multiple charged triphosphate.

The protein kinase mechanism uses aspartate as a general base, weakening the hydrogen–oxygen bond of the hydroxyl-containing residues. This allows the hydroxyl nucleophile of the substrate to attack the γ-phosphate of ATP, producing a phosphorylated peptide product and ADP (**Scheme 2**).

It has recently been shown that some phosphorylations may be useful only to activate a protein and once in its active form or complexed, phosphorylation is no longer needed (**Scheme 3**).[26] Karwowska-Desaulniers *et al.*[26] recently hypothesized that phosphorylation of histone deacetylase 1 (HDAC1) at two key residues is important for complex formation of active protein *in vivo*. However, when the phosphorylation sites were mutated to alanine, no active protein was expressed, yet complex formation was still observed. Even when native protein was expressed and dephosphorylated, complex formation and activity were still observed. The complex formation is still not fully understood; the complexes are thought to place HDAC in the correct position. No active noncomplexed HDAC has been generated in order to study the effects of complexation. Interestingly, when HDAC1 was expressed in bacteria and subsequently phosphorylated there was no



Scheme 1 Phosphorylation of hydroxyl-containing amino acids.

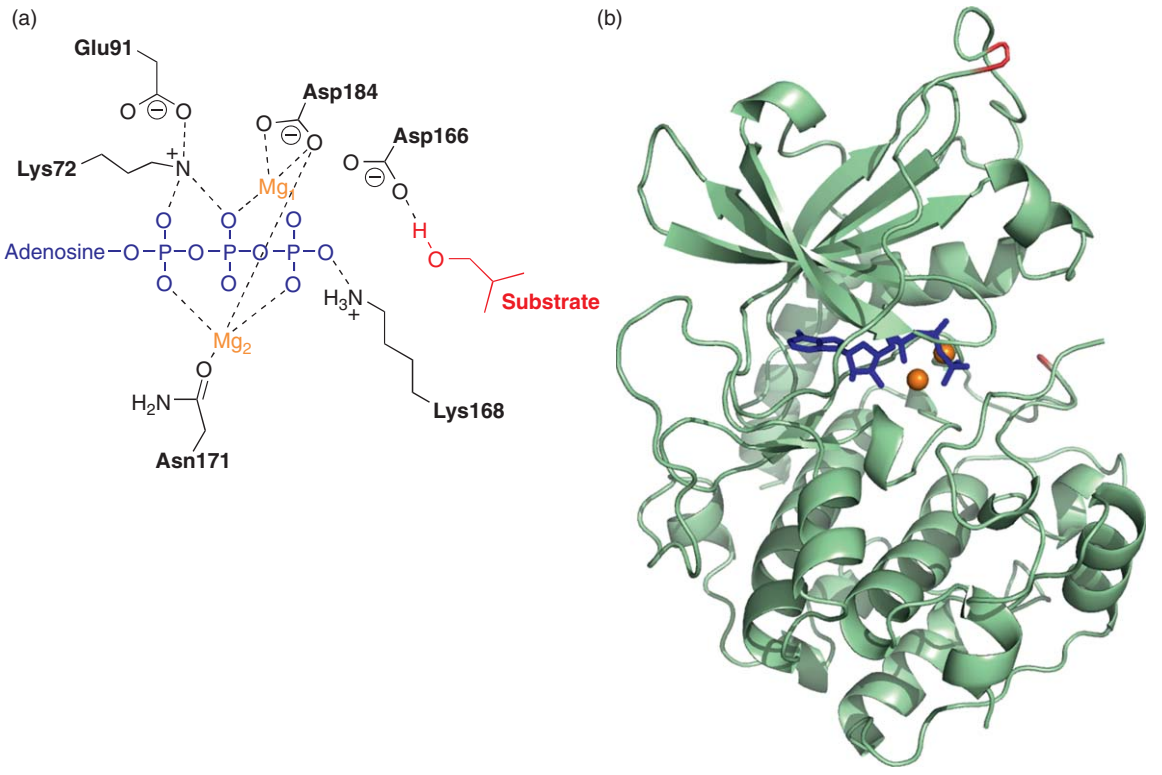
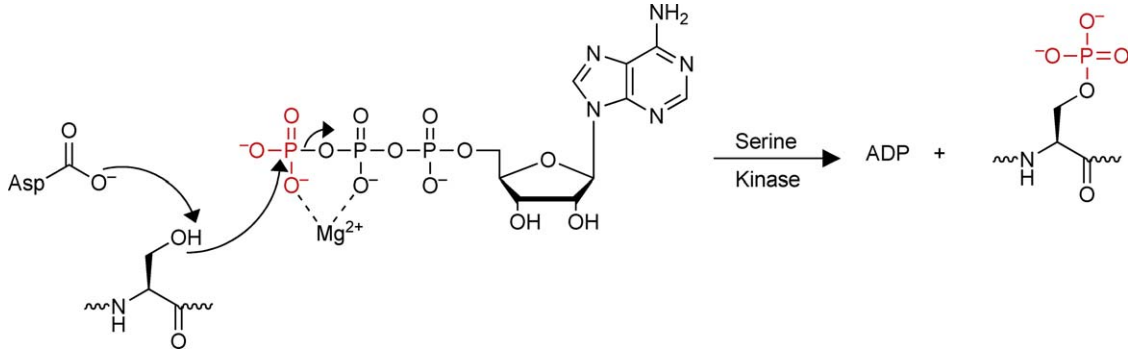
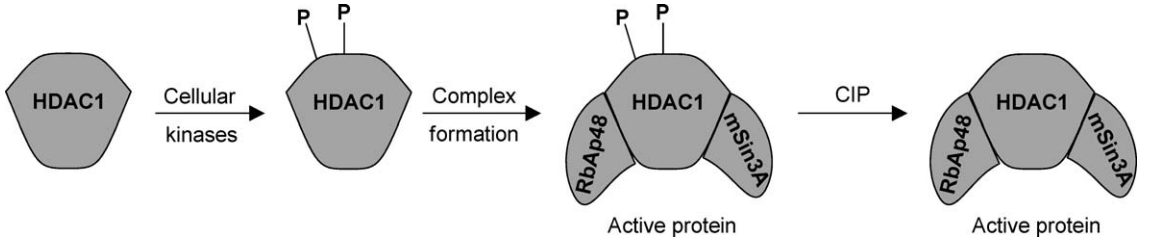


Figure 2 Crystal structure of PKA and conserved active site. (a) Crystal structure image of PKA (green) in complex with peptide substrate inhibitor. ATP is displayed as a ball and stick structure (blue) and Mg^{2+} ions (orange). (b) ATP binding residues of the kinase with ATP (blue), Mg^{2+} ions (orange), and substrate (red) are shown.



Scheme 2 The catalytic mechanism of phosphorylation. Aspartate acts as a general base to deprotonate the hydroxyl group for nucleophilic attack at the γ-phosphate of ATP.



Scheme 3 Dispensable phosphorylation of HDAC1. HDAC1 is posttranslationally phosphorylated. Upon phosphorylation, HDAC1 forms complexes with RbAp48 and mSin3A, which generate active HDAC1. When the protein complex is treated with calf intestinal phosphatase (CIP), the phosphates are removed, but activity and complex formation remain.

deacetlyase activity. These experiments show that phosphorylation can be required for initial activity and complex formation, but not required to maintain activity as once thought.

Phosphorylation has also been implicated in viral growth. A recent study showed that a lithium-sensitive kinase (e.g., GSK-3β) phosphorylated serine 54 of the human respiratory syncytial virus (HRSV) structural P protein, which is a cofactor of the viral RNA polymerase.[27] The removal of the phosphate is controlled by phosphatase 2A (PP2A). When lithium-sensitive kinases were inhibited, the virus was unable to replicate and did not survive (**Figure 3**).

Phosphorylation also plays an evident part in cell division. Specifically, cytoskeleton-associated protein 2 (CKAP2) has a phosphorylation site at threonine 596. This threonine has been tracked and gets phosphorylated between prophase and metaphase, and dephosphorylated a short time later before anaphase, suggesting that the protein is phosphorylated during the formation of mitotic spindles.[28]

Playing a major role in signal transduction,[29] phosphorylation, as mentioned above, is in practice a switch. The change of a signaling cascade is heavily dependent on phosphorylation. Another protein using phosphorylation as a switch is the general splicing regulator p38 (SRp38). Upon phosphorylation, the protein switches from a splicing repressor to a sequence-specific activator.[30] This study further displays the importance and versatility of phosphorylation.

Other functions, like carbohydrate metabolism, also utilize kinase-signaling cascades for regulation. The insulin pathway is very important in the uptake and maintenance of the body's glucose levels and glycogen storage. The pathway starts with the release of insulin from the pancreas upon an increase in blood sugar levels. The insulin makes its way through the blood stream to insulin receptors (IRs), where the process of storing glucose in the form of glycogen begins.

Upon binding of insulin to the β-subunits of IR (IRβ), the receptor autophosphorylates. This phosphorylation event can result in one of the two overall signaling pathways (**Figure 4**). In one pathway, the β-subunit of IR phosphorylates its insulin receptor substrate, IRS-1. Upon phosphorylation, IRS-1 becomes active and activates phosphoinositide 3-kinase (PI3K) through phosphorylation. PI3K then activates pyruvate dehydrogenase kinase 1 (PDK1), which activates protein kinase B (PKB), a serine kinase. PKB deactivates glycogen synthase kinase 3 (GSK3)[31] by phosphorylation, leading to the activation of glycogen synthase and glycogen synthesis. Activation of PKB also results in migration of glucose transporter (GLUT4) vesicles from their intracellular storage to the cell surface, where they allow the uptake of glucose into the cell.[32]

The second signaling transduction pathway allows for the expression of genes to influence glucose uptake. IRβ can also tyrosine phosphorylate Src homologues and collagen protein (Shc), which allows Shc to complex with the growth receptor-bound protein 2 (Grb2) and binds to son of sevenless 1 (Sos1). Sos1 initiates the Ras kinase (RAS)/Raf kinase (RAF)/mitogen-activated or extracellular signal-regulated protein kinase (MEK)

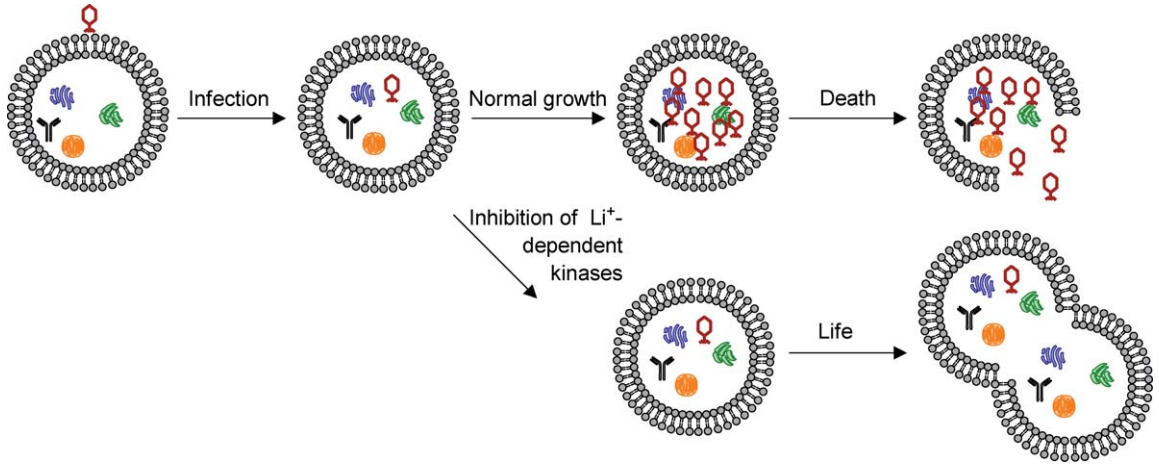


Figure 3 Phosphorylation-dependent viral growth. In the presence of Li^+-dependent kinases, viruses are replicating and ultimately lead to cell death (top). When the Li^+-dependent kinases are inhibited, the virus no longer replicates and the cell undergoes normal functions like mitosis.

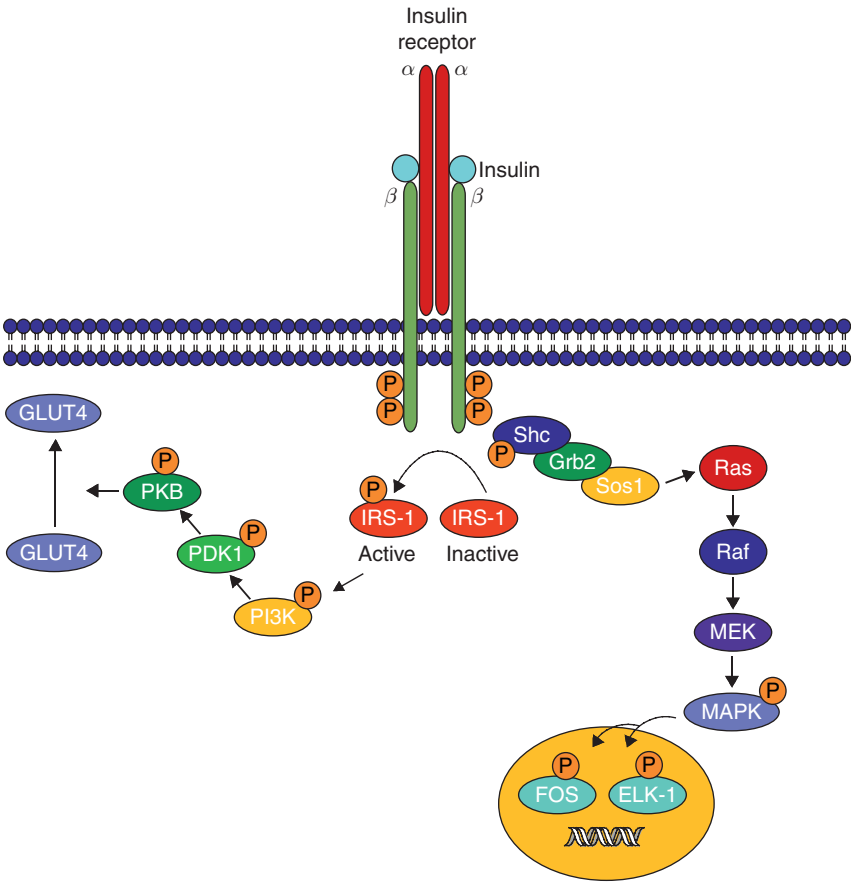


Figure 4 Two insulin signaling pathways regulating glycogen synthesis/glucose uptake and gene expression. Upon binding of insulin to the receptor IRβ, it auto-phosphorylates. IRβ then phosphorylates IRS-1 or Shc. IRS-1 phosphorylates PI3K, which phosphorylates PDK1, which phosphorylates PKB. PKB activates the translocation of the glucose transporter GLUT4 from intracellular vesicles to the cell surface to commence the transport of glucose into the cells. Shc triggers the Ras/Raf/MEK activation of MAPK that activates gene transcription through FOS and ELK-1 to increase the expression of genes involved in glucose uptake. Orange circles represent phosphorylations.

pathway, leading to activation of mitogen-activated protein kinase (MAPK) and mitogenic responses of gene transcription through c-fos protein (FOS) and E-26-like protein 1 (ELK-1).[32] It is well established that the insulin pathway is involved in diabetes,[31] demonstrating again the importance of phosphorylation and its role in diseases.

The culmination of gene transcription and cell metabolism is cell growth. Kinase phosphorylation cascades also control cell growth, growth arrest, and can even lead to apoptosis. Shown in **Figure 5** is a phosphorylation-dependent signaling pathway to commence neuronal cell growth. A family of receptors known as Trk proto-oncogene contain four members TrkA, TrkB, TrkC, and TrkE, which are expressed variably throughout the nervous system. The pathway shown in **Figure 5** involves the TrkA neuronal growth factor receptor (NGFR).

When neuronal growth factor (NGF) is present in the extracellular matrix, it binds to TrkA and stimulates its tyrosine autophosphorylation. The phosphotyrosine can then associate with Shc[33] and commence the Ras/Raf/MEK signaling pathway, thereby activating MAPK. MAPK phosphorylates ELK-1 that can recruit the general transcriptional machinery and turn on the genes for cell growth. In addition to stimulating cell growth, the activation of Ras also enhances the activity of RhoA that inhibits the MAPK stimulation of $p21^{CIP}$, a protein responsible for induction of growth arrest.[33] The Trk pathways are very similar to other growth receptors; however NGFR has an additional binding site for Shc that is thought to promote differentiation in the neuronal cells.[33]

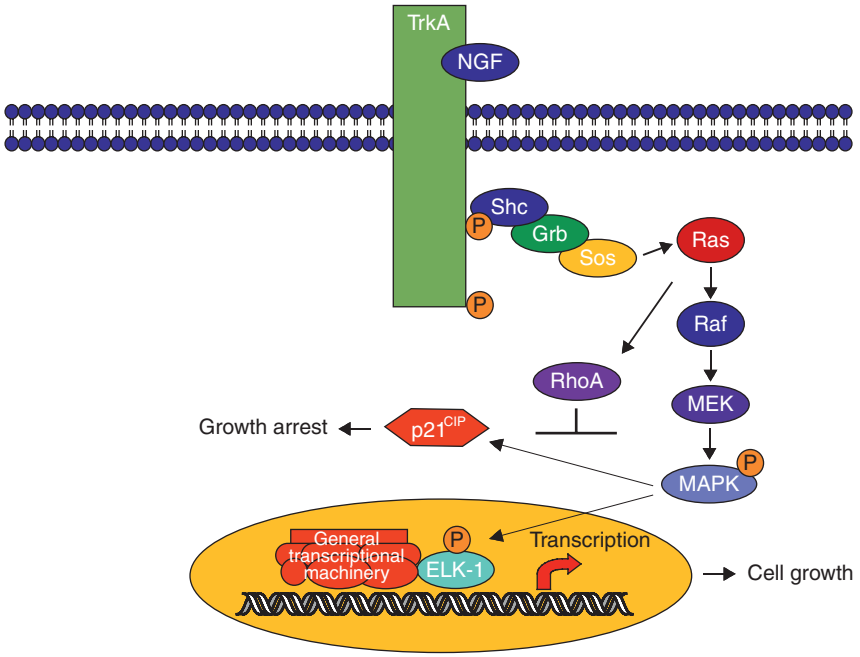


Figure 5 Signaling of cell growth through NGF and NGFR TrkA. Binding of NGF to TrkA results in TrkA auto-phosphorylation. The phosphorylated receptor can interact with Shc and triggers the Ras/Raf/MEK pathway. Phosphorylation of ELK-1 initiates gene transcription and cell growth. Ras also activates RhoA, which inhibits the growth arrest protein $p21^{CIP}$.

Phosphorylation can enhance protein binding. One example involves the proteins cyclic adenosine 3′,5′-monophosphate (cAMP) responsive element-binding protein (CREB) and CREB-binding protein (CBP), two transcription factors.[34,35] The phosphorylation of CREB at Ser133 transforms the residue from a hydrogen bond donor into a strong hydrogen bond acceptor. The neutral serine is simultaneously transformed into a charged phosphoserine. The new phosphoserine at position 133 can now hydrogen bond with Tyr658 of CBP, accounting for the initial interaction of the two proteins.[34] The phosphorylation at Ser133 of CREB is critical for the interaction of the two enzymes, leading to the activation of gene expression.[36] The binding of CBP greatly increases the activity of CREB in transcriptional activation.[37,38]

As a second example of phosphorylation-dependent interactions, p53 is greatly stabilized when phosphorylated.[39] This stabilization of structure causes p53 to dissociate from Mdm3, and bind to DNA. Once bound to DNA, p53 enhances gene transcription.[39] In both cases (CREB and p53), it is clearly seen that any error in phosphorylation levels has the potential to cause serious effects on gene expression. If a signal causes overexpression of a gene, the results could be devastating to a cell or an organism.

In addition to its influence on protein–protein interactions, phosphorylation also affects protein structure and activity. One case involves a protein termed dematin headpiece (DHP), an actin-binding protein found in a variety of tissues including heart, brain, skeletal muscle, kidney, and lung.[40] DHP is known to interact with Ras-guanine nucleotide exchange factor (Ras-GRF2) and this interaction can modulate MAPK pathways, which can link the cytoskeleton and signaling pathways.[41]

Phosphorylation of DHP causes its N-terminal domain to migrate closer to the phosphorylation site.[42] The change in conformation not only affects the overall structure of the protein, but also its activity in bundling actin. Recent models where Ser74 is mutated to a glutamate, to mimic phosphorylation, have led to elucidation of the effects of phosphorylation on DHP. *In vitro* studies comparing phosphorylated DHP and DHP S74E show similar activity.[41] The NMR studies comparing wild-type DHP and DHP S74E show a 25° shift in the C-terminal helices of the protein toward the N-terminal domain. The shift in structure drastically changes the activity of the protein. Upon mutation, DHP no longer bundles actin, which can cause spherocytosis in leukocytes, displaying the importance of phosphorylation regulation on cellular structure.

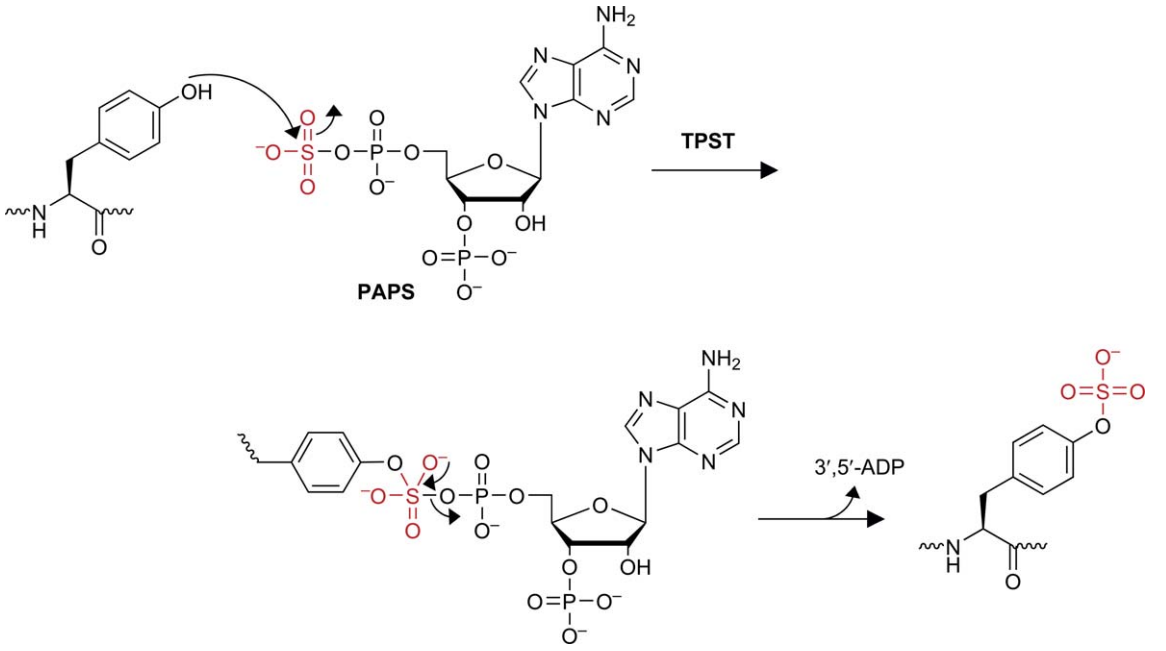
A second phosphorylation-dependent protein is the first-identified and best-studied tyrosine kinase, c-Src.[43] c-Src is regulated by phosphorylation of a C-terminal tyrosine residue. This residue is phosphorylated by a second tyrosine kinase labeled C-terminal Src kinase (Csk). When unphosphorylated, Src is in its open form and catalytically active. Upon phosphorylation, the conformation of the kinase is altered and it becomes significantly less active.[44] The inactivation is a result of the src homologue 2 (SH2) domain of Src binding the C-terminal phosphotyrosine. This binding allows an appreciable interaction between the linker sequence connecting the SH2 domain and the SH1 or kinase domain, and the SH3 domain, normally a type II polyproline helix (PPII) binder. The interaction between the two portions of the enzyme was unexpected due to the lack of PPII character possessed by the linker region. Because the interaction is intramolecular, it is believed to be enhanced. The series of events changes the overall structure of the enzyme and turns off catalytic activity. Upon dephosphorylation of the tyrosine residue, catalytic activity is restored,[44] making phosphorylation a reversible switch.

It is important to highlight that c-Src or cellular Src behaves as described above; however there is a second form of Src termed v-Src or viral Src. v-Src was originally discovered as a component of the Rous Sarcoma virus, which causes cancer in chickens. The sequences of c-Src and v-Src are nearly identical. The major difference in the two proteins occurs in the C-terminal tail. While c-Src is regulated through phosphorylation of the C-terminus, v-Src has no C-terminal phosphorylation site and therefore is constitutively active and unregulated.

5.15.2.2 Sulfation

Protein sulfation occurs exclusively at tyrosine residues.[45] It has been suggested that up to 1% of the tyrosine protein content becomes sulfated, which is the most abundant posttranslational modification for tyrosine, with phosphorylation occurring only on 0.5% of tyrosine protein content.[46] Sulfation occurs mostly on excreted proteins or trans-membrane proteins. Sulfation is catalyzed by tyrosylprotein sulfotransferase (TPST), with PAPS as a cosubstrate (**Scheme 4**). Like kinases, sulfotransferases have a biological inverse known as sulfatases.[47]

Sulfation in most aspects is very similar to phosphorylation, except that sulfation is not involved in intracellular signal transduction, but in other forms of signaling. The mechanism of sulfation is similar to that of phosphorylation as a general base from the enzyme active site that deprotonates the hydroxyl groups of tyrosine residues. The nucleophilic oxygen then attacks the β-position, in contrast to the γ-position in phosphorylation, and releases adenosine 3′,5′-diphosphate.



Scheme 4 Catalytic mechanism of protein sulfation. Tyrosine attacks electrophilic sulfur generating 3′,5′-ADP as a leaving group.

Since most of tyrosine sulfation occurs on proteins that are secreted or membrane bound, it is no surprise that chemokine receptors have the potential to be sulfated. Chemokines are a class of proteins secreted by cells, generally characterized by their small size (roughly 8–10 kDa) and the conserved placement of four cysteines that play a key role in the structure of the protein.[48] Sulfation is so important in the interaction of chemokines and chemokine receptors that in some chemokine receptors it affects their affinity for the targeted chemokine. These changes in affinity have an effect on chronic or acute events of cellular immunity, which can include asthma and chronic obstructive pulmonary disease (COPD).[49] These events are in part due to the leukocyte trafficking and airway inflammation associated with cellular immunity. The binding of chemokines is thought to trigger cellular signaling events.

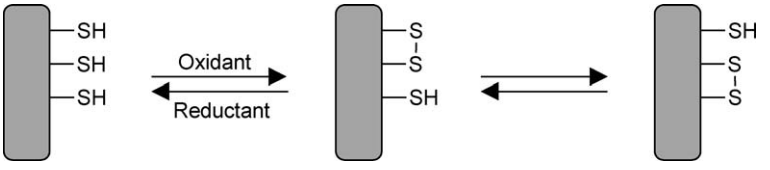
Sulfation has also been documented in salivary proteins, specifically statherin.[50] The enzymes in saliva are the first wave of the human digestion process and statherin prevents the precipitation of calcium phosphate in the salivary gland and saliva.[51] Since TPST is secreted along with statherin, sulfation plays a role in digestion by binding hydroxylapatite and preventing its precipitation.

5.15.3 Cysteine Disulfide Formation

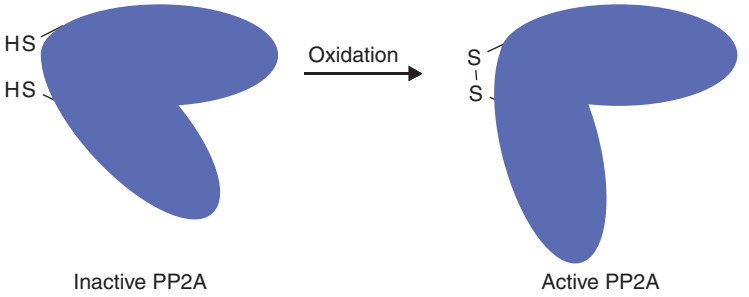
Cysteine disulfide formation is one of the most important posttranslational modifications involved in protein structure. Disulfides play a crucial role in maintaining the structure of many proteins including insulin, keratin, and many other structurally important proteins. While the cytoplasm and nucleus are reducing microenvironments, the Golgi and other organelles can have oxidizing environments and process proteins to contain disulfide bonds (**Scheme 5**).

Disulfide formation has elucidated the properties of PP2A in the brain. PP2A was isolated from the soluble portion of rat brain after the enzyme was reacted with a biotin-linked sulfide. The biotin-linked PP2A was purified on a streptavidin resin and eluted with a thiol-reducing agent. PP2A was confirmed to be a dithiol containing protein. In addition, the activity of PP2A was reduced 2.5-fold when the disulfide was not intact.[52] Remarkably, when PP2A was eluted from the column with a reducing agent the activity went up to sevenfold greater than the soluble fraction. These data indicate that disulfides play an important role in enzyme activity and can alter the dynamics of phosphorylation, that is, the function of PP2A (**Scheme 6**). PP2A was also discovered to be oxidation sensitive.[52]

Disulfides have also removed the activity of certain proteins in bacteria. The bacterium *Paracoccus pantotrophus* has a cluster of enzymes responsible for the oxidation of sulfides to sulfate known as the sulfur-oxidizing



Scheme 5 General disulfide formation equilibrium.



Scheme 6 Protein phosphatase 2A disulfide dependence. PP2A requires disulfide bonds for optimal activity.

enzyme system (Sox). On in-depth study of the SoxY portion of the system, it was noted that a disulfide could form at Cys138.[53] Upon disulfide formation, the heterodimer, SoxYZ, would become a heterotetramer $SoxYZ_2$. Once the disulfide was formed, the catalytic activity of the system was diminished. The cysteine involved with disulfide formation is the cysteine of the catalytic site. The bacteria utilizing this system would be reducing agent sensitive in the absence of environmental sulfate.

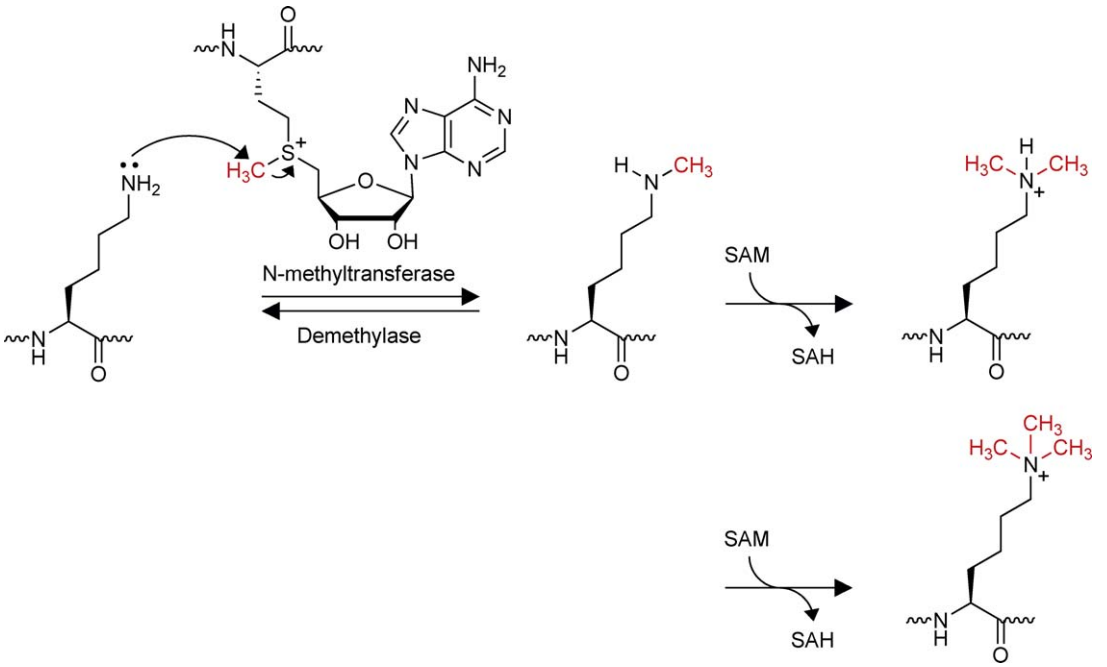
5.15.4 Methylation

5.15.4.1 N-Methylation

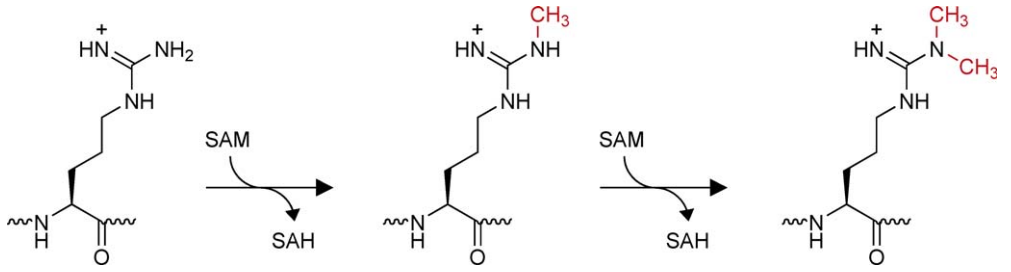
Lysine is the most common site for N-methylation, but methylation can also occur on arginine, histidine, glutamine, and asparagine. The enzymes responsible for N-methylation are known as N-methyltransferases aided with SAM as a cosubstrate. All forms of methylation share the same mechanism: the nucleophilic amino acid side chain attacks the electrophilic methyl group of SAM and releases *S*-adenosylhomocysteine (SAH) (**Scheme 7**).

Methylation plays an important role in transcriptional regulation and a lesser role in signal transduction.[54] Histones are heavily methylated proteins. Single, double, or triple methylated lysines play an important role on histones.[55,56] Lysine methylation is a more subtle transcriptional control than acetylation. Lysine methylation has come to light in another protein known as p53.[57] p53 is a protein expressed in low levels in the cell and stabilized by posttranslational modifications including phosphorylation, acetylation, and now N-methylation. There are several C-terminal lysines on p53 that increase its stability.[57] The addition of the methylation modifications adds complexity to p53 and fine-tunes its activity and ultimately suppresses tumor formation.[57]

Like lysines, there are arginine methyltransferases, some of which create the monomethyl-arginine and others that generate dimethyl-arginine with the consumption of two SAM molecules (**Scheme 8**). Also like methylation of lysine, arginine methylation helps regulate gene transcription. Recent work has shown methylation of arginine to regulate the coupling of transcription and the processing of mRNA.[58] The arginine



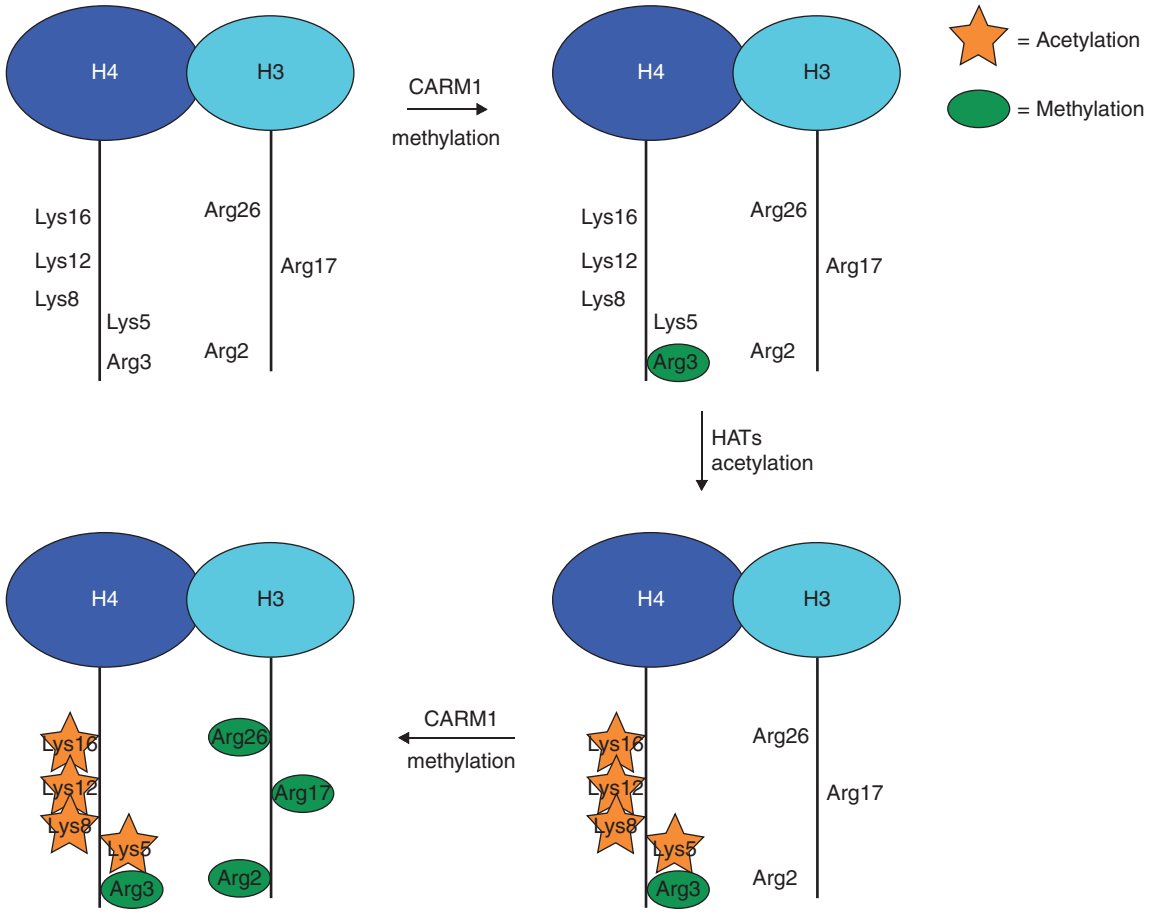
Scheme 7 Catalytic mechanism of methylation. The nucleophilic side chain (N, O, S) attacks the methyl group attached to the electrophilic sulfur generating methylated protein and *S*-adenosylhomocysteine. Lysine can be methylated up to three times under an identical mechanism.



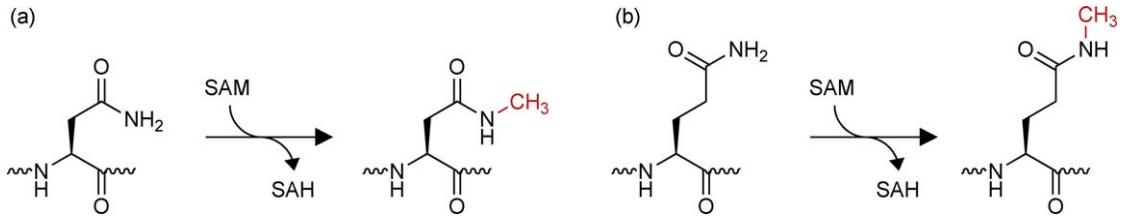
Scheme 8 Mono- and di-N-methylation of arginine residues.

methyltransferase named coactivator-associated arginine methyltransferase (CARM1) is recruited by multiple transcription factors and is used as a positive regulator. On investigation of the substrates for CARM1, it was found that several of the proteins were involved in splicing, implicating the involvement of CARM1 in alternative splicing. It has also been found that methylation of Arg3 on histone 4 (H4) induces the acetylation of Lys5, Lys8, Lys12, and Lys16 on H4 (**Scheme 9**).[59] This series of acetylations then spurs the methylation of Arg2, Arg17, and Arg26 on histone 3 (H3) through the coactivator CARM1.[60]

Asparagine and glutamine methylation are certainly unexpected. The carboxamido functional group is not known for its nucleophilicity, but nature has created pathways that can overcome low nucleophilicity of



Scheme 9 Methylation-induced modification cascade. The methylation of Arg3 on H4 causes the acetylation of Lys5, Lys8, Lys12, and Lys16. These acetylated residues then trigger the methylation of Arg2, Arg17, and Arg26 on H3.



Scheme 10 N-methylation of (a) asparagine and (b) glutamine residues.

molecules. Methylation of asparagine has been noted in bacteria, and methylation of glutamine residues has been shown to play a role in the release of proteins after ribosomal synthesis (**Scheme 10**).[61] No additional work has been investigated at this time.

N-terminal methylation of proteins is possible and can play a major role in mitosis.[62] Chen *et al.*[63] discovered that the N-terminal serine or proline of regulator of chromatin condensation 1 (RCC1) is methylated. RCC1 associates with chromatin, binding to histones H2A and H2B, regulated by Ran, a GTPase.[64,65] For methylation to occur, the initiating methionine must be removed and a proline and lysine must be at positions 3 and 4, respectively.[63] If RCC1 is N-terminal methylation deficient, its binding to histones loses efficiency, which can cause spindle pole defects and ultimately cause problems in cell division. This work suggests the first known function for N-terminal protein methylation.[63]

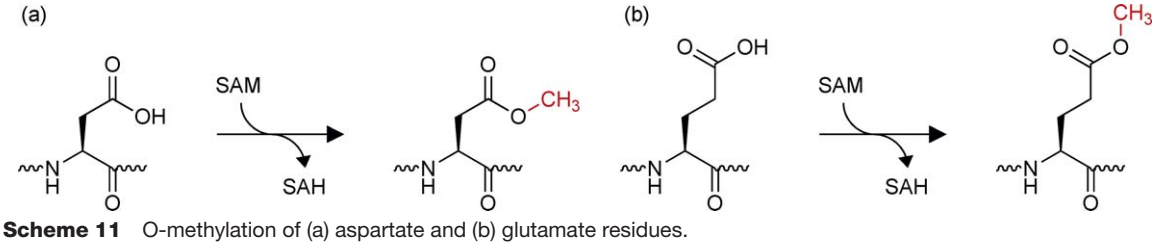
5.15.4.2 O-Methylation

Aspartate and glutamate O-methylation has not been studied as extensively as lysine and arginine N-methylation (**Scheme 11**). Glutamate O-methylation is known to play a role in modulating chemotactic responses in *Escherichia coli*, but the methyltransferases have not been found in eukaryotes.[66] Aspartate O-methylation has been observed in both prokaryotic and eukaryotic organisms. In prokaryotes it has been implicated as an enzyme repair mechanism.[66] Aspartate O-methylation was found to be abundant in about 2% of eukaryotic cells, suggesting that aspartate O-methylation is an important posttranslational modification and could rival lysine and arginine N-methylation.[66]

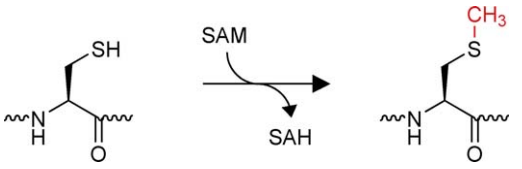
In addition to the O-methylation of aspartate and glutamate residues, the C-terminal carboxyl group of GTPases are methylated.[67] In the G-protein family C-terminal cysteines are prenylated at the sequence CZZX, where Z is a hydrophobic amino acid and X represents any residue.[67,68] Once the ZZX sequence is cleaved by a special protease, the isoprenylcysteine carboxymethyltransferase (Icmt) methylates the C-terminal carboxyl group and effectively creates a more hydrophobic enzyme.[67]

5.15.4.3 S-Methylation

S-methylation involves the methylation of cysteine (**Scheme 12**). While there are no current studies on the S-methylation of cysteine, it is thought to be involved in the aging process of enzymes.[69]



Scheme 11 O-methylation of (a) aspartate and (b) glutamate residues.



Scheme 12 S-methylation of cysteine.

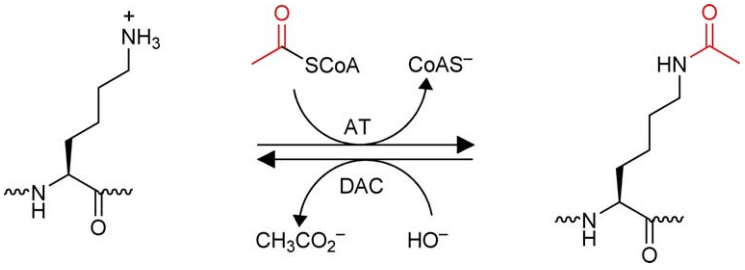
5.15.5 N-Acetylation

While we will focus here on N-acetylation, it is important to note that the acetylation of serine and threonine has also been discovered.[70] No machinery is known for the modification, and research on this acetylation reaction is in its infancy. N-acetylation has played an important role in gene regulation.[71,72] During acetylation, lysine attacks the acetyl group of AcCoA to generate acetylated lysine and CoA (**Scheme 13**). In addition to being methylated, the polylysine tails of histones have been shown to be acetylated (**Figure 6**). Acetylation is achieved in this case by a class of enzymes known as histone acetyltransferases (HATs). The acetyl group can be removed with HDACs (**Scheme 13**). As in methylation, the acetylation of a lysine transforms the positively charged amine into a neutral acetylamine. The change from positive to neutral no longer allows the lysines to bind the DNA, opening the histone complex to DNA transcription or replication.[73] Acetylation of histones has the ability to cross talk with other posttranslational modification on histones. For example, the phoshorylation of Ser10 on H3 prevents subsequent methylation of Lys9.[74] The methylation at Lys9 would also block acetylation of the same residue.[75]

Recent studies on asthma and COPD have shown variations in the acetylation levels of inflammatory proteins and histones that are altered compared to those without the diseases. In asthma, HAT activity is increased, while HDAC activity remains the same, while those with COPD show the opposite.[76] These studies implicate protein hyperacetylation in inflammatory response and hypoacetylation in the lack of inflammatory response.

In addition to the role played in histone acetylation and transcription, acetylation can also have an effect on cellular trafficking. One such mechanism is the increased acetylation of tubulin. In a recent study, increased acetylation of tubulin due to inhibition of tubulin deacetylase or histone deacetylase 6 (HDAC6) led to an increase in cell movement.[77] A plasmid microinjected into cells lacking HDAC6 took only 5 min to migrate to the nucleus, less than half the time for cells containing HDAC6. These experiments implicate that acetylation not only imparts activity to transcription, but also intracellular movement of proteins, plasmids, and organelles.

Acetylation also plays a role in mitochondrial function and basal levels of ATP. Mouse studies have shown that in mice lacking a deacetylase (Sirt3), basal ATP levels were lower than that in normal mice. In organs such as the heart, kidneys, and liver, the ATP levels were decreased more than 50%.[78] In concurrence with lower ATP levels, higher acetylation levels of mitochondrial proteins were also noted. When Sirt3 was added, ATP levels returned to normal. This work leads to the conclusion that acetylation controls the levels of ATP in the cell.



Scheme 13 N-acetylation of lysine. Lysine attacks the acetyl group of acetyl-CoA generating an acetylated protein and CoA.

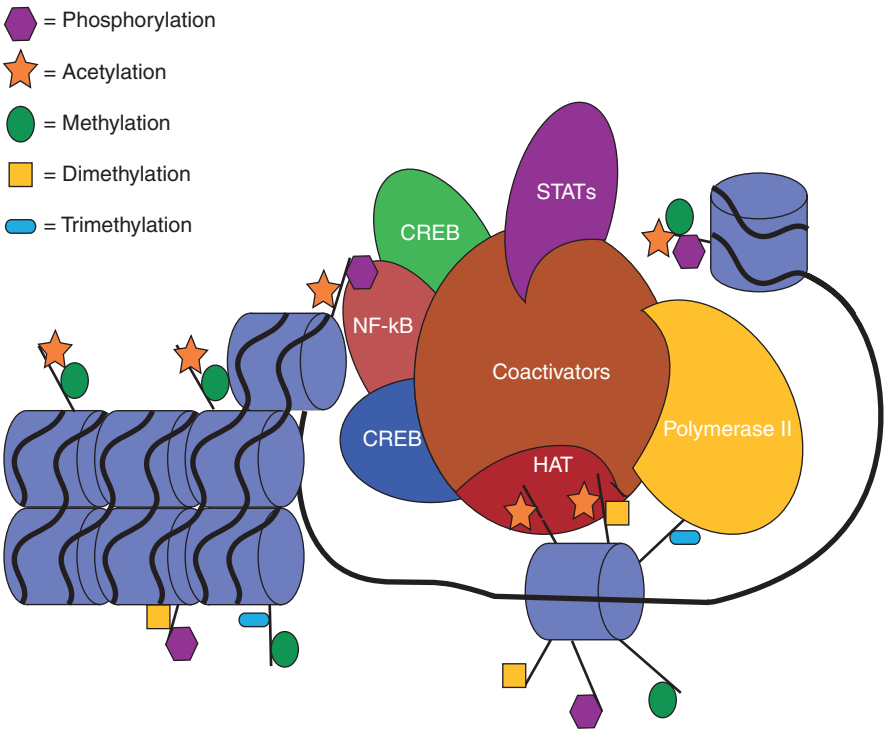


Figure 6 Cartoon of histone posttranslational modifications. Histone proteins are known for having multiple numbers of the same posttranslational modification and multiple types of posttranslational modifications on their lysine tails responsible for binding DNA. The histone code as it is called relies on the dynamic nature of phosphorylation, acetylation, and mono-, di-, and trimethylation.

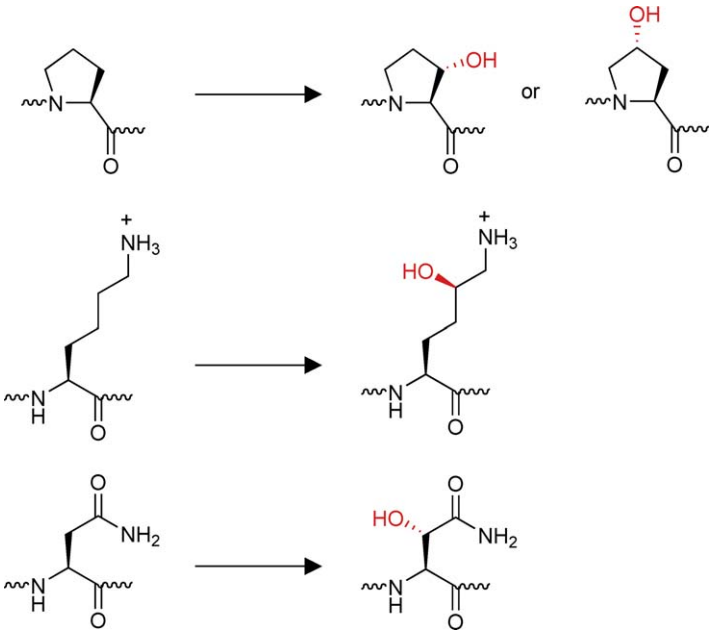
5.15.6 Hydroxylation

Hydroxylation of proteins occurs on three residues, most commonly proline at either the 3- or 4-position, lysine at the 5-position, and asparagine at the 3-position (**Scheme 14**). Enzymes catalyzing protein hydroxylation are iron-dependent hydroxylases, which consume oxygen and α-ketoglutarate as cosubstrates. Hydroxyproline and hydroxylysine play important roles in the maturation of collagen fibers,[79,80] while hydroxyasparagines can be found in antifungals and antibiotic compounds.[81]

Conus snails employ hydroxyproline in certain conotoxins: μ-, ω-, and α-conotoxins.[82] μ-GIIIA naturally contains three hydroxyproline residues. When the toxin was synthesized with prolines replacing their hydroxylated counterparts, the ability of GIIIA to block sodium channels was reduced, while the folding remained intact.[82] ω-MVIIC had decreased folding when the hydroxyproline was replaced with proline, but little effect was seen on biological activity.[82] The α-conotoxins ImI and GI showed results similar to that of GIIIA, where the folding was increased, but biological activity decreased in the presence of hydroxyproline.[82]

Hydroxyproline-rich glycoproteins (HRGPs) are an important family of self-assembling proteins that are vital to the cell walls of plants.[83,84] *Chlamydomonas* has been studied to look at the evolutionary role of the HRGPs.[85] The findings from studying *Chlamydomonas reinhardtii* and *Chlamydomonas incerta* suggest that the misreading of the HRGP hydroxylases may cause changes in fundamental structures of the plants, which create an evolutionary step in the species. The two species of *Chlamydomonas* share a common ancestry, the difference in the HRGPs could be a source of specification.

Hypoxia-inducible factor (HIF) is a transcriptional regulating protein whose stability is regulated by proline hydroxylation.[86–88] Under oxygen-poor conditions, or hypoxia, HIF functions normally expressing HIF targeted genes. When oxygen levels increase, HIFα subunits become hydroxylated at specific residues and are tagged for destruction.[89] A secondary path controlling gene transcription by HIF is hydroxylation of the



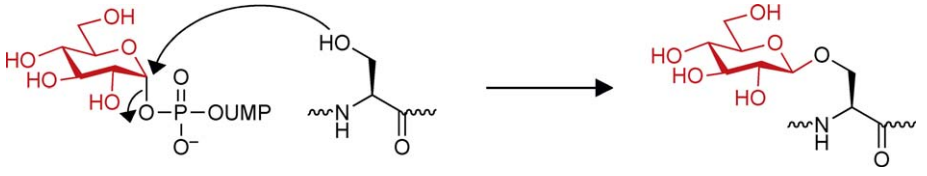
Scheme 14 Protein hydroxylation. Hydroxylations occur at the 3- or 4-position of proline, the 5-position of lysine, and the 3-position of asparagine.

factor inhibiting HIF (FIH), a hydroxylase that modifies Asn803 of HIF.[90] This hydroxylation event prevents the recruitment of transcriptional coactivators p300 and CBP and gene transcription does not occur.[89] A nitric oxide donor, *S*-nitroso-*N*-acetylpenicillamine (SNAP), stabilized HIF and prevented the hydroxylation events from degrading the protein or preventing protein interactions.[91] The hydroxylation of proline was not inhibited, but the downstream interaction with degradative enzymes was inhibited, while asparagine hydroxylation was completely inhibited.[91]

Adipopectin is one of many small peptides secreted by adipose tissues, which alters the sensitivity of insulin in the bloodstream by stimulating fatty acid oxidation.[92] In addition to altering insulin sensitivity, adipopectin also decreases plasma triglyceride levels and improves glucose metabolism.[92] In the collagen domain of adipopectin, which allows the formation of high molecular weight (HMW) oligomeric complex, four conserved lysines must be hydroxylated and, in some cases, glycosylated.[93] Prevention of the posttranslational modifications resulted in reduced activity and formation of the HMW oligomeric complexes. Hydroxylation is required for the formation of the oligomer and contributes to the insulin-sensitizing activity of adipopectin in hepatocytes.

5.15.7 Glycosylation

Protein glycosylation occurs mainly on serine and asparagine residues,[94,95] but can also occur on hydroxylysine and hydroxyproline (**Scheme 15**).[96] Glycosylation is very important in the endoplasmic reticulum and Golgi apparatus and can be involved in cell signaling.[97] Many of the membrane-bound proteins and excreted proteins are glycosylated. Protein glycosylation is important in all forms of eukaryotes.[98,99]



Scheme 15 General mechanism of glycosylation. The hydroxyl group of serine attacks the activated anomeric carbon of UDP-activated sugars.

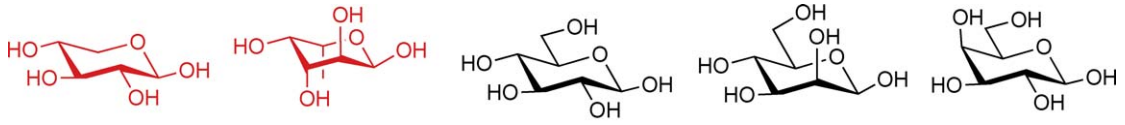


Figure 7 Common sugars used in glycosylation reactions. β-Xylose and β-fucose (red) are used primarily by plant and insects, while β-glucose, β-mannose, and β-galactose are used in mammalian systems.

There are several glycoproteins that contain fucose (insects) or fucose and xylose (plants). These glycoproteins are known to cause allergic reactions in humans.[100] Most mammalian glycoproteins contain glucose, galactose, mannose, and their derivatives (**Figure 7**). Some individuals are so sensitive to these fucose- and xylose-containing glycoproteins that they develop allergic reactions to certain foods or insect bites. These glycosides happen to be linked through an asparagine residue, which is not abundantly seen in human tissues. These glycoproteins can lead to an increased amount of antibodies (IgE), targeting the fucose and xylose derivatives, displaying the importance of glycosylation in the immune system.

Glycosylation has many effectors in the cell; diabetes can affect the amount of glycosylation in individuals carrying the disease.[101] Lately, a turmeric extract, curcumin, has been used in investigations as an anticancer drug and has also been used in diabetic studies.[101] In studies using endothelial cells, fed with large amount of glucose to simulate diabetes, it was found that curcumin lowered the levels of glycosylated proteins.[102] Not only was the amount of glycosylated proteins reduced, but also the amounts of oxidized lipids and oxygen radicals.[102] This work displays that over-glycosylation can be hazardous to one's health. Glycosylation has been hypothesized to have an effect on the initiation and development of atherosclerotic lesions in diabetics.[103] When smooth muscle cells were treated with high quantities of glucose, an increase in thrimbospondin-1 (TSP-1) was observed. TSP-1 is a potent anti-angiogenic and pro-atherogenic protein[104,105] found in smooth muscle cells and provides a link between diabetes and vascular complications.

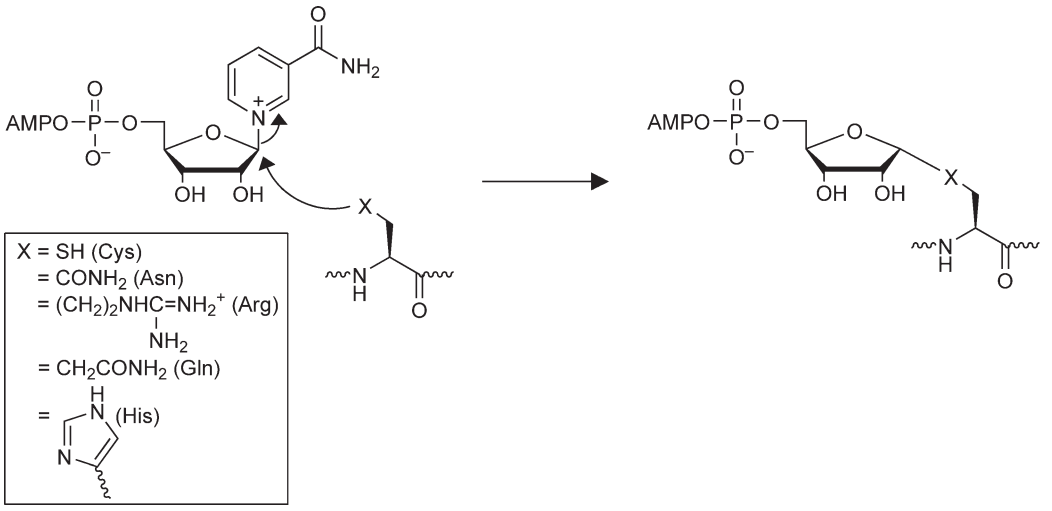
Cell surface glycoproteins can become sulfated similarly to tyrosine sulfation.[106,107] In recent studies, a heparin sulfate endosulfatase, Sulf1, which is involved in growth factor signaling through regulation of 6-O-sulfation states of cell surface and matrix heparin sulfate proteoglycans, was shown to be glycosylated at an asparagine residue.[107] The glycosylation, which occurs in the active site, is vital to the Srf1 activity, membrane targeting, and secretion. Glycosylation has now been implicated to have a key role in the control of sulfatase enzymatic activity and indirectly growth factor signaling.[107]

In filamentous fungi, the secretory proteins are mannosylated by protein O-mannosyltransferase (PMT) starting in the endoplasmic reticulum (ER).[98] After leaving the ER the proteins are subsequently glycosylated in the Golgi apparatus to diversify the structure of the O-glycans. These glycosylations improve the solubility of secreted proteins and cause proteins to become less pervious to proteolysis. In addition, the fungal morphology, development, and differentiation are also determined by glycosylation.[108,109]

5.15.8 ADP-Ribosylation

Many of the ADP-ribose transfer enzymes are secreted from bacteria, as means of protection, virulence, or aggression, and the protein is taken up into the host cells. Recently the mammalian counterparts have been termed ADP-ribosyltransferases (ARTs) and ADP-ribosylhydrolases (ARHs).[110] The ART enzymes use the metabolite NAD^+ as a cosubstrate and transfer the ADP-ribose onto cysteine, arginine, asparagine, glutamine, or histidine releasing nicotinamide.[111] ARHs hydrolyze the sugar releasing ADP-ribose from the protein. ADP-ribosylation has been found to affect mitosis, differentiation and proliferation, telomere length and longevity, and cell death processes (**Scheme 16**).[112]

The bacterium *Pseudomonas aeruginosa* produces several virulence factors. In particular ExoS has an ADP-ribosylation domain that affects arginine residues. When pleotypic cells are infected with ExoS, the phosphorylation of ezrin/radixin/moesin (ERM) proteins no longer occurs.[113] It happens that ERM proteins are high-affinity substrates for the ADP-ribosylation domain of ExoS. When the sequence of moesin was



Scheme 16 Catalytic mechanism of ADP-ribosylation. The nucleophilic side chain of the amino acid attacks the anomeric carbon of ribose carrying nicotinamide to create ADP-ribosylated proteins.

analyzed, three arginines were found to be ADP-ribosylated: Arg553, Arg560, and Arg563 (**Scheme 17**). These three arginines are clustered around Thr558, which is normally phosphorylated by protein kinase C (PKC) and Rho kinase. Due to the lack of phosphorylation of this residue, cells were unable to maintain their shape and became spherical. Here ADP-ribosylation is used by the bacteria to inhibit the phosphorylation of proteins.

In addition to ADP-ribosylation, protein can become polyADP-ribosylated.[114] Enzymes known as polyADP-ribosylation polymerases (PARPs) catalyze multiple ADP-ribosylations to create polyADP-ribosylated proteins (**Scheme 18**).[115] PolyADP-ribosylated proteins play an important role in many biological functions including maintenance of genomic stability, transcriptional regulation,[116] energy metabolism, and cell death.[117,118] How polyADP-ribosylation achieves these effects is still under investigation. One theory is through steric inhibition/activation.[117]

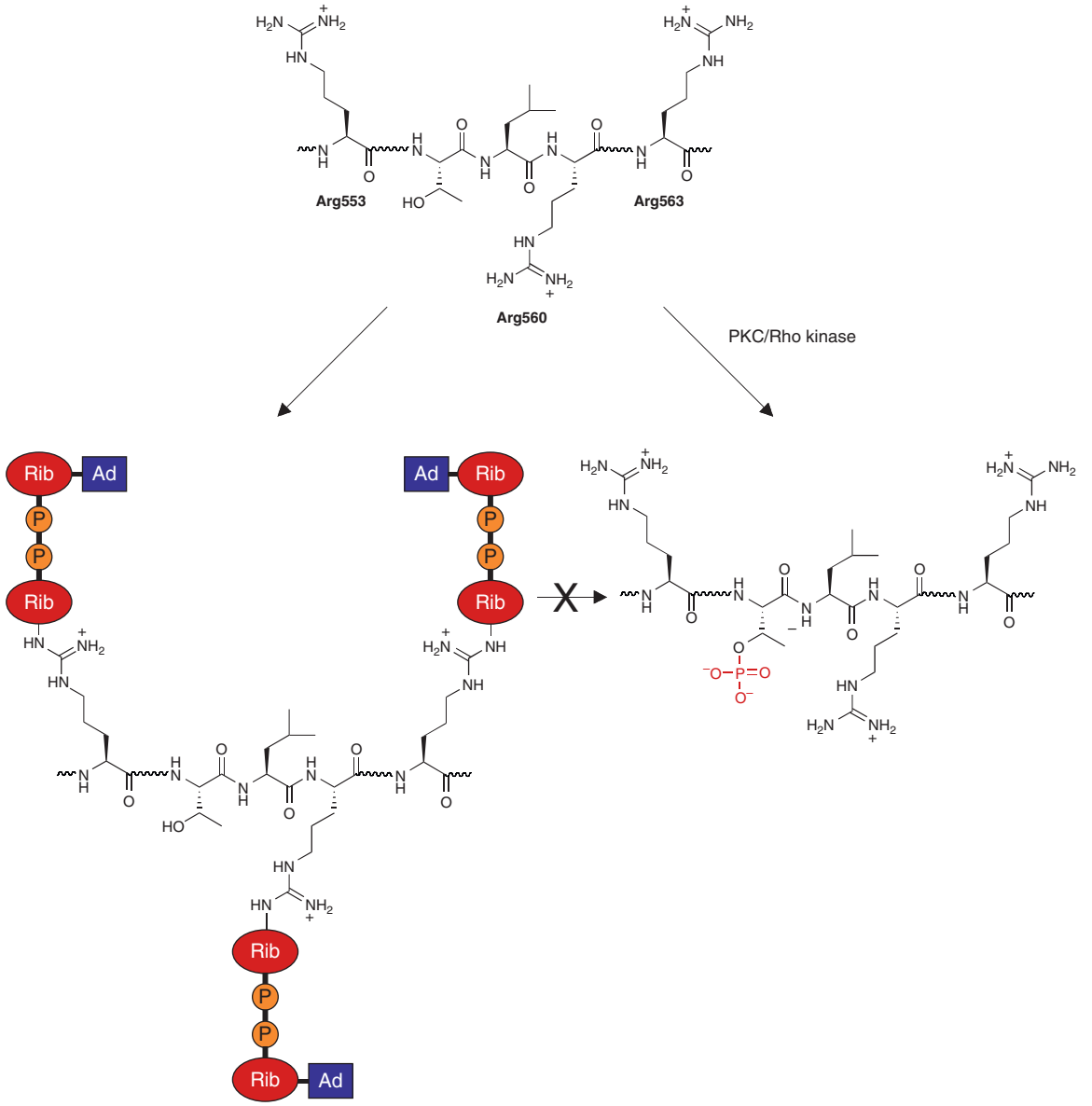
PolyADP-ribosylation has been reported to play a role in traumatic brain injury (TBI), excitotoxic, and oxidative injury.[119] In the mitochondria after TBI, PARPs are activated and polyADP-ribosylate multiple proteins involved in electron transfer. Since the ribosylation of these proteins shuts down electron transport, cells are sent into an apoptotic state. This gives insight into mitochondrial-based brain injuries and diseases.

5.15.9 Prenylation

Protein prenylation involves the transfer of isoprenyl groups such as farnesyl and geranylgeranyl lipid groups (**Scheme 19**). The respective transferases utilize farnesyl pyrophosphate and geranylgeranyl pyrophosphate as cosubstrates. Farnesylation and geranylgeranylation are very important in C-terminal anchoring of proteins to the cell membrane.[120,121] Among the proteins that become prenylated are small GTPases such as Ras, Rho, and Rab family proteins, which require prenylation for activity.

The C-terminal cysteine of GTPases is commonly prenylated and is thought to be involved in anchoring G-proteins to the membranes of various organelles and the cell membrane so that they may effectively interact with their effector proteins.[122–126] Several studies have suggested that inhibition of prenylation leads to accumulation of G-proteins in the cytosol, due to decreased hydrophobicity of the unmodified protein.[68] Prenylation levels of G-proteins in pancreatic β-cells is linked closely with extracellular glucose concentrations,[127] giving prenylation an important role in sugar metabolism.

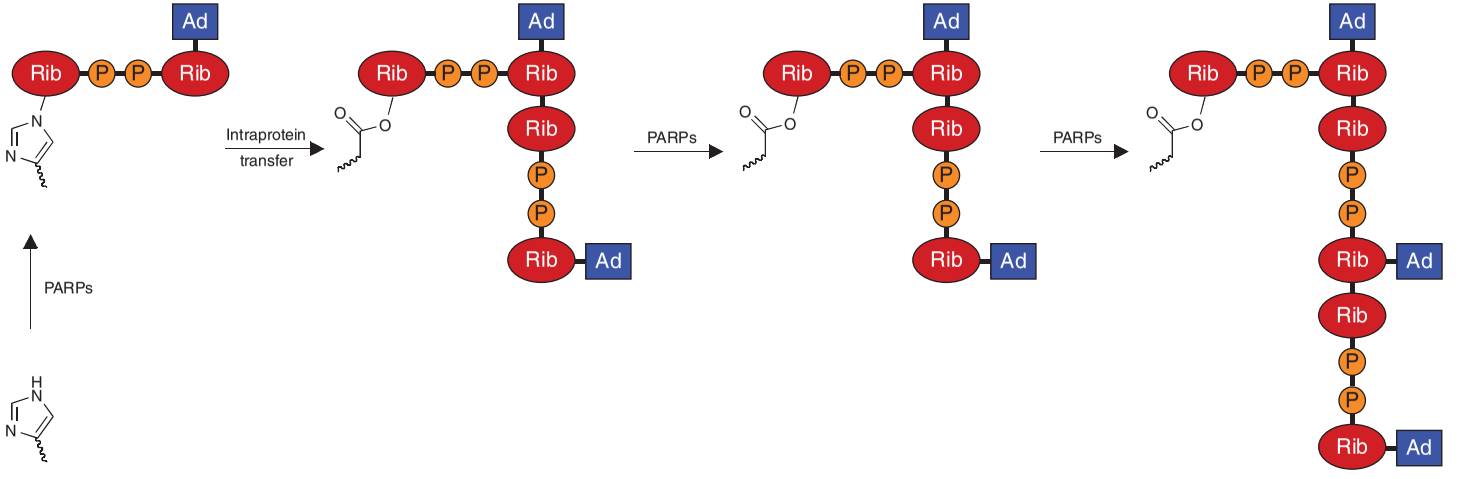
Prenylation has been implicated in the prevention of HIV infection.[128] Statins have been used to inhibit HIV infection by interacting with Rho GTPases[129,130] and suppress the intercellular adhesion required for viral entry.[131] In another study with statins and HIV the mechanism of action was elucidated. The lypophilic statins



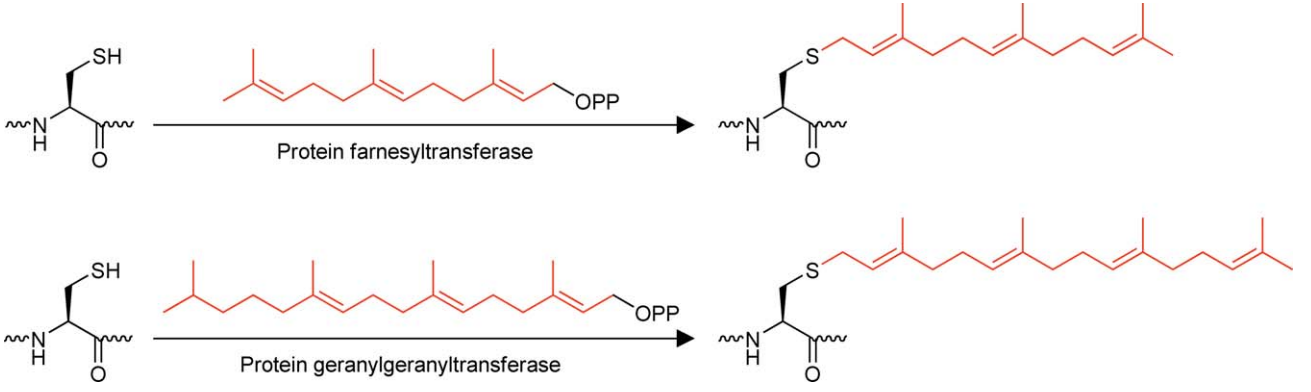
Scheme 17 ADP-ribosylation of moesin inhibits phosphorylation. Moesin is phosphorylated at Thr558, but when cells are infected with *Pseudomonas aeruginosa*, moesin becomes ADP-ribosylated at Arg553, Arg560, and Arg563, which effectively blocks phosphorylation.

suppress HIV release from tumor necrosis factor α, through inhibition of geranylgeranylation. The process did not deplete cholesterol.[128] When geranylgeranyl pyrophosphate was added the effects were reversed demonstrating that geranylgeranylation is required for the virility of HIV-1, and opens gernanylgeranyltransferase inhibitors as HIV drugs.

Prenylation, specifically farnesylation, plays an important part in rheumatoid arthritis.[132] Arthritic fibroblasts secret matrix metalloproteinases (MMPs) into the intercellular matrix and upon their release MMPs commence the digestion of cartilage tissue.[133,134] When arthritic fibroblasts were treated with a farnesyltransferase inhibitor (FTI), the amount of certain MMPs secreted from the cells decreased.[132] In contrast, inhibition with GGTases gave mixed results. GGTI-298, a GGTase inhibitor, enhanced, while geranylgeranylpyrophosphate (GGPP) inhibited, MMP-1 secretion. This indicates that protein farnesylation, but not



Scheme 18 Polyribosylation of proteins. In some cases proteins become polyADP-ribosylated by PARPs creating long chains of diribose diphosphate.



Scheme 19 Prenylation of cysteine. Cysteine can become prenylated with farnesyl or geranylgeranyl groups to anchor proteins to membranes.

geranylgeranylation, is required for expression and secretion of MMP-1 from arthritic fibroblasts. Geranylgeranylation actually decreases the amount of MMP-1 expressed and secreted in arthritic fibroblasts.

5.15.10 Biotin, Lipoate, and Phosphopantetheine Tethering

Some enzymes are nonfunctional until posttranslationally modified. Examples of these enzymes include the acyl- and carboxyltransferases. While lipoate and phosphopantetheine are necessary for acyl transfer chemistry, tethered biotin is used in carboxyl transfer chemistry. Biotin and lipoate tethering occur under a similar mechanism; the natural small molecule is activated with ATP to form biotinyl-AMP or lipoyl-AMP (**Scheme 20**). A lysine from the target protein then attacks the activated acid and transfers the group to the protein. The phosphopantetheine moiety is transferred using its own enzyme, the phosphopantetheinyltransferase (PPTase). The PPTase uses a nucleophilic hydroxy-containing amino acid, serine, to attach the phosphopantetheinyl (Ppant) arm found in coenzyme A to convert the apo (inactive) carrier protein to its holo (active) form. The reaction is Mg^{2+}-dependent.

5.15.10.1 Biotinylation

Protein biotinylation is catalyzed by biotin protein ligase (BPL). In the active site of the enzyme, biotin is activated at the expense of ATP to form AMP-biotin; the activated biotin can then react with a nucleophile on the targeted protein. BPL transfers the biotin to a special lysine on biotin carboxyl carrier protein (BCCP), a subunit of AcCoA carboxylase (**Scheme 21**).[135] Biotinylation of BCCP is very important in fatty acid biosynthesis, starting the growth of the fatty acid with AcCoA carboxylase to generate malonyl-CoA. Recently the crystal structures of mutated BPL and BCCP have been solved together with biotin and ATP to get a better idea of how the transfer functions.[135]

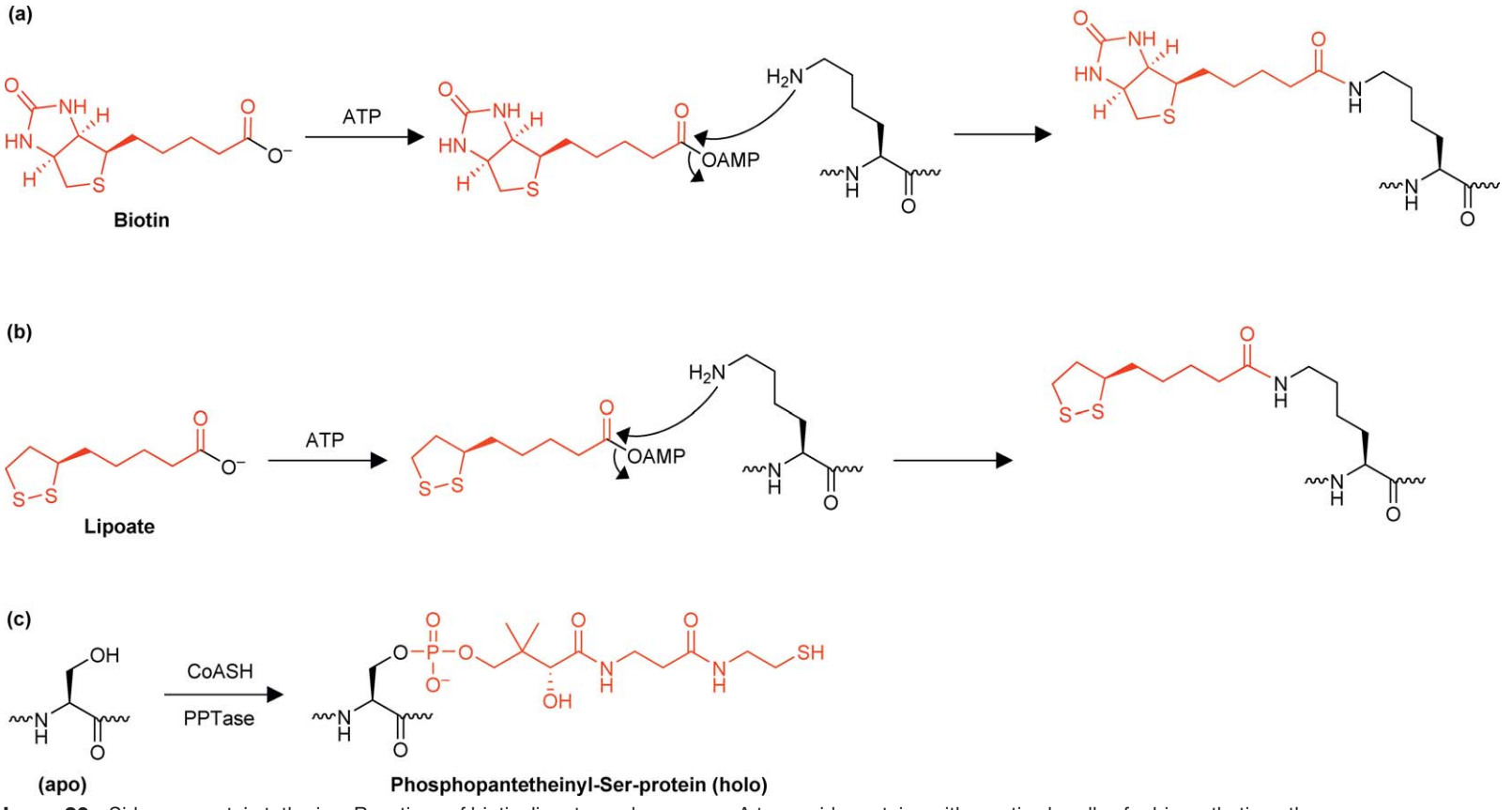
5.15.10.2 Lipoylation

While lipoate becomes activated similarly to biotin, the mechanism of acyl transfer is different. Lipoamide is used to ferry acyl groups between functional domains of multi-domain enzymatic assembly lines. Lipoate becomes acetylated by reacting with thiamine pyrophosphate (TPP) (**Scheme 22**). After several rearrangements TPP is consumed and tethered lipoate is loaded with an acetyl group ready for biosynthesis. Lipoate has been suggested to play roles in sulfane sulfur metabolism.[136] Lipoylation is used in many metabolic enzymes including pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase, branched chain 2-oxoacid dehydrogenase, and glycine cleavage.[137]

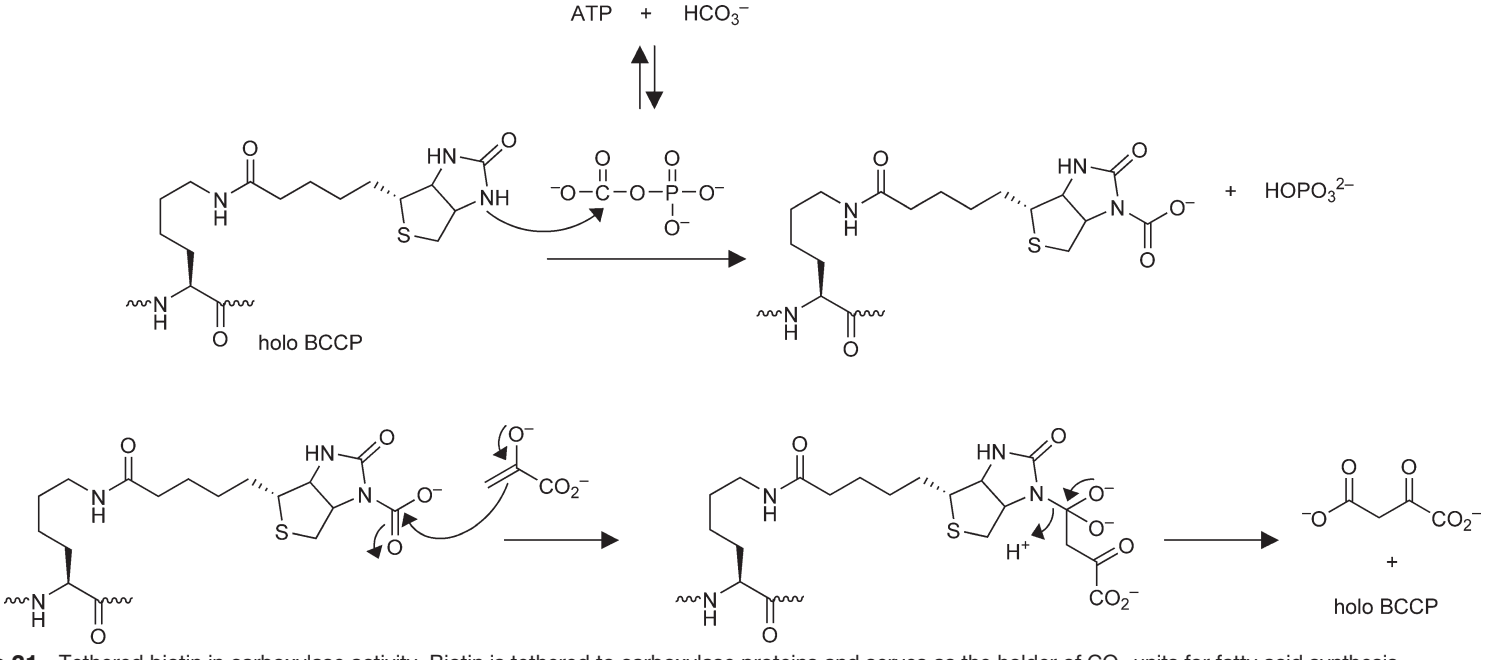
5.15.10.3 Phosphopantetheine Attachment

5.15.10.3.1 Carrier proteins

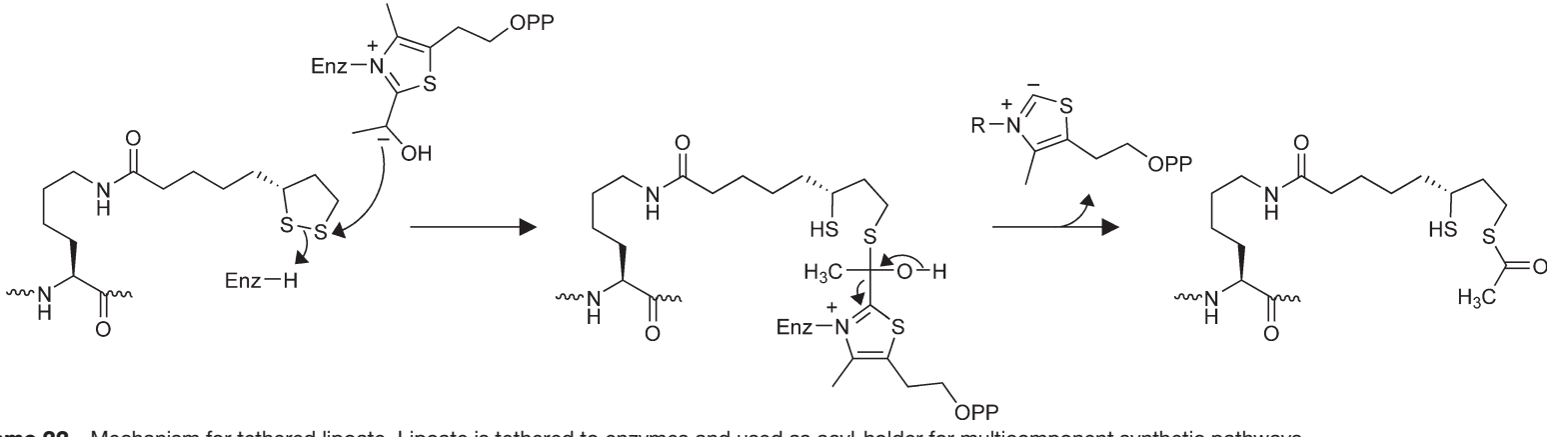
Phosphopantetheine tethering is a posttranslational modification that takes place on the active site serine of carrier proteins[138] – 'acyl carrier proteins (ACPs) and peptidyl carrier proteins (PCPs), also termed thiolation (T) domains' – during the biosynthesis of fatty acids (FAs) (use ACPs) (**Scheme 23**), polyketides (PKs) (use ACPs) (**Scheme 24**), and nonribosomal peptides (NRPs) (use T domain) (**Scheme 25**). It is only after the covalent attachment of the ~20-Å Ppant arm, required for facile transfer of the various building block constituents of the molecules to be formed, that the carrier proteins can interact with the other components of the different multi-modular assembly lines (fatty acid synthases (FASs), polyketide synthases (PKSs), and nonribosomal peptide synthetases (NRPSs)) on which the compounds of interest are assembled. The structural organizations of FASs, PKSs, and NRPSs are analogous and can be divided into three broad classes: the types I, II, and III systems. Even though the role of the carrier proteins is the same in all systems, their mode of action differs from one system to another. In the type I systems the carrier proteins usually only interact in cis with domains to which they are physically attached, with the exception of the PPTases and external type II thioesterase (TEII) domains that act in trans. In the type II systems the carrier proteins selectively interact



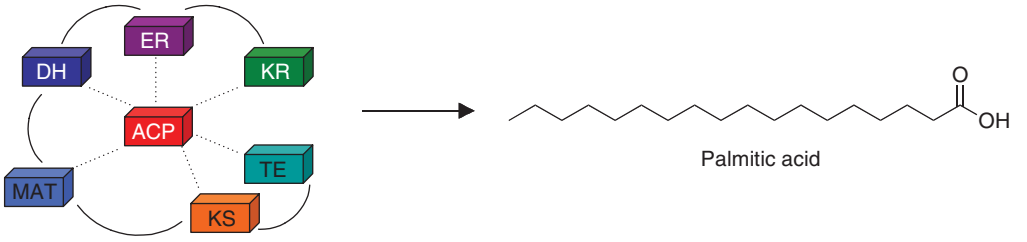
Scheme 20 Side-arm protein tethering. Reactions of biotin, lipoate, and coenzyme A to provide proteins with reactive handles for biosynthetic pathways.



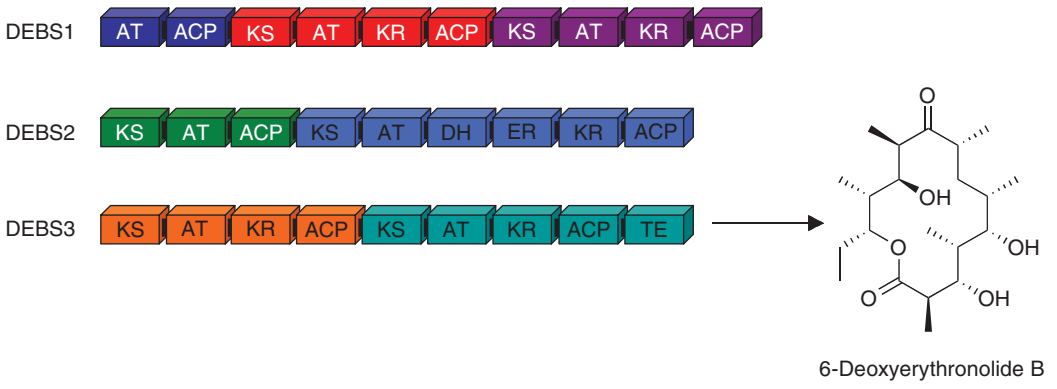
Scheme 21 Tethered biotin in carboxylase activity. Biotin is tethered to carboxylase proteins and serves as the holder of CO_2 units for fatty acid synthesis.



Scheme 22 Mechanism for tethered lipoate. Lipoate is tethered to enzymes and used as acyl-holder for multicomponent synthetic pathways.



Scheme 23 Example of an acyl carrier protein (ACP in red) in a type I FAS. The palmitic acid is depicted as a representative fatty acid. During its biosynthesis, the ACP (red) interacts iteratively with each domain (DH, dehydrogenase; ER, enoyl reductase; KR, ketoreductase; KS, ketosynthase; TE, thioesterase) until the palmitic acid has reached its proper length.



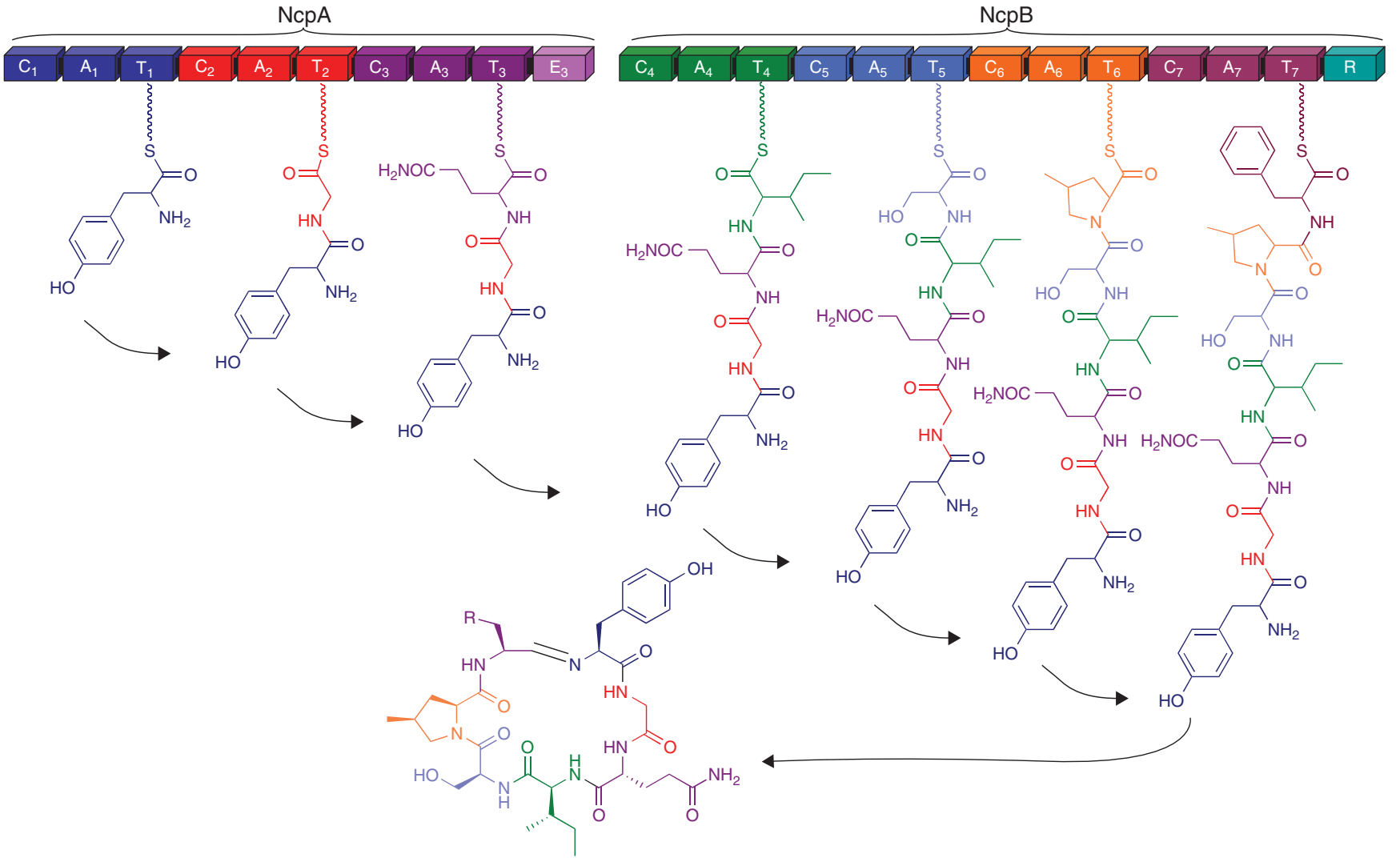
Scheme 24 Example of acyl carrier proteins (ACPs) in type I PKS. The 6-deoxyerythronolide B is shown as a representative polyketide. Seven ACPs are found in the seven modules, each depicted in a different color, responsible for the incorporation of the seven building block constituents of 6-deoxyerythronolide B.

in trans, in an iterative manner, and in a specific order with the tailoring and extending enzymes. The type III systems do not utilize carrier proteins.[139]

The first refined structure of a carrier protein was obtained for the FAS ACP from *E. coli* about 20 years ago.[140] All carrier proteins display a four α-helical structure with a conserved active site serine residue (contained in the motif DSX, where X = lysine or arginine) in a loop close to the beginning of helix II. Since then, several NMR solution structures and X-ray crystal structures of various apo and holo ACPs and PCPs have been determined, which provide insight into the mode of interaction of PPTases with carrier proteins.[141–149] However, until 2 years ago the mechanism of interaction between the carrier proteins and other domains was mostly unknown. Invaluable insight into carrier protein–other domain interaction during nonribosomal peptide biosynthesis was very recently obtained by determination of the crystal structure of a thiolation-condensation (T-C) di-domain from the nonribosomal tyrocidine synthetase,[150] the dynamic solution structure of a T-TEI di-domain from the *E. coli* enterobactin synthetase,[151] and the NMR structure of a T-TEII complex from *B. subtilis*.[152] The most recent report of a 2.6-Å crystal structure of the C-A-T-TE SrfA-C termination module from *B. subtilis* showed that during nonribosomal peptide biosynthesis the adenylation (A) and condensation domains get in close proximity to associate and form a catalytic platform where they display their active sites on the same side (**Figure 8**).[153] The flexible T domain with its substrate-loaded-Ppant arm is attached to this platform and is able to move from the active site of the A domain to the donor site of the C domain.

5.15.10.3.2 PPTases

PPTases are the first proteins that carrier proteins interact with during the biosynthesis of FAs, PKs, and NRPs. Many PPTases have been identified prior to 2006[154–162] and can be categorized into three main classes: (1) AcpS[163]-like PPTases involved in primary metabolism acting on carrier proteins from FAS systems



Scheme 25 Example of thiolation (T) domains in type I NRPS. Nostocyclopeptide is shown as a representative nonribosomal peptide. Seven T domains are found in the seven modules, each depicted in a different color, responsible for the incorporation of the seven amino acid building block constituents of nostocyclopeptide. A reductase (R) domain replaces the more common TE domain for the formation of the unusual imine functionality.

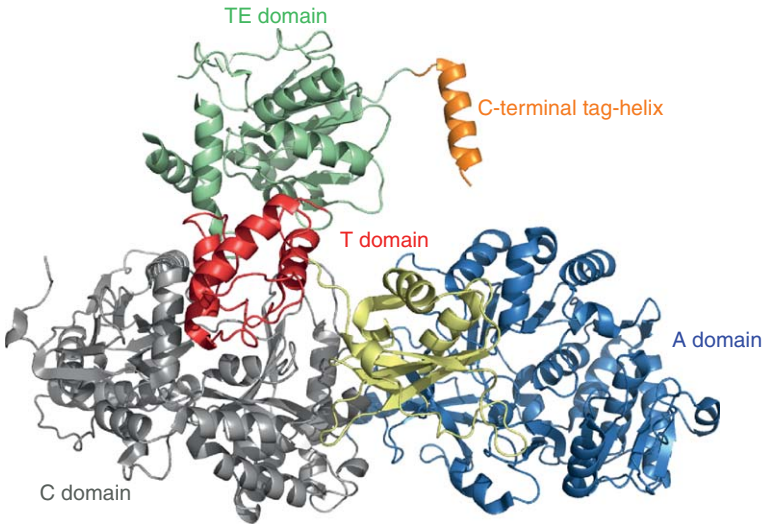


Figure 8 Crystal structure of the C-A-T-TE SrfA-C termination module from *Bacillus subtilis*. The C domain is shown in gray, the A domain in yellow and blue, the T domain can be seen in red, the TE domain in green, and the C-terminal tag is shown in orange.

and displaying little activity with carrier proteins from NRPS and PKS systems, (2) Sfp[164]-like PPTases with broad substrate specificity towards various ACPs and PCPs, and (3) PPTases contained in fungal and yeast FASs.[165,166]

More recently two functionally redundant Sfp-type PPTase, MxPpt1 and MxPpt2, have been found to differentially activate biosynthetic pathways in *Myxococcus xanthus*.[167] MxPpt1 and MxPpt2 exhibit broad substrate specificity as supported by the fact that the complex PKS-NRPS hydrids epothilone and myxothiazol from *Sorangium cellulosum* and *Stigmatella aurantiaca*, respectively, could be expressed in *M. xanthus* without the need for an external PPTase.

In all *Mycobacterium* spp. two conserved PPTases are found: an AcpS that activates two FASs and another PPTase that acts on NRPSs and PKSs that are involved in the biosynthesis of virulence factors.[168] The essential role of both PPTases was demonstrated for the survival of *Mycobacterium smegmatis*, making them promising targets for the development of new antituberculosis drugs.

The characterization of the first functional cyanobacterial PPTase[169] from *Nodularia spumigena* NSOR10 was reported and its ability to modify carrier proteins from heterocyst glycolipid synthesis and nodularin toxin synthesis was demonstrated.[170] This now opens the door to harness the biotechnological potential of cyanobacterial natural products formation.

In the bacterium *Penicillium chrysogenum* the genomic PPTase is similar to others. When the gene was knocked out, the bacteria needed lysine to survive and no longer produced penicillin or pigmentation. However the bacterium still completed the syntheses of roquefortine, another mycotoxin, and fatty acids.[171] This work suggests that the *pptase* gene is not needed for the synthesis of fatty acids and roquefortine in this species of bacteria.

Based on studies on the neocarzinostatin (NCS) and C-1027 biosynthesis (**Figure 9**), it has come to light that phosphopantetheine tethering is required for the biosynthesis of members of the family of some of the most potent anticancer drugs known to date, the enediynes.[172] Until this time, it was unsure if the enediyne structures had common pathways, but with the discovery of a C-terminal PPTase domain on the gene cluster and the production of linear fragments of the enediyne core structure, the production of these potent anticancer drugs is becoming more relevant. A novel type of PPTase with distinct pseudo-trimeric structure was also recently discovered in the calicheamicin gene cluster, another member of the family of enediynes (**Figure 9**).[173]

The filamentous fungi *Aspergillus nidulans* are also dependent on phosphopantetheine tethering. In fungi lacking the *cfwA* gene and the *A. nidulans* PPTase, several polyketides were missing including emericellin,

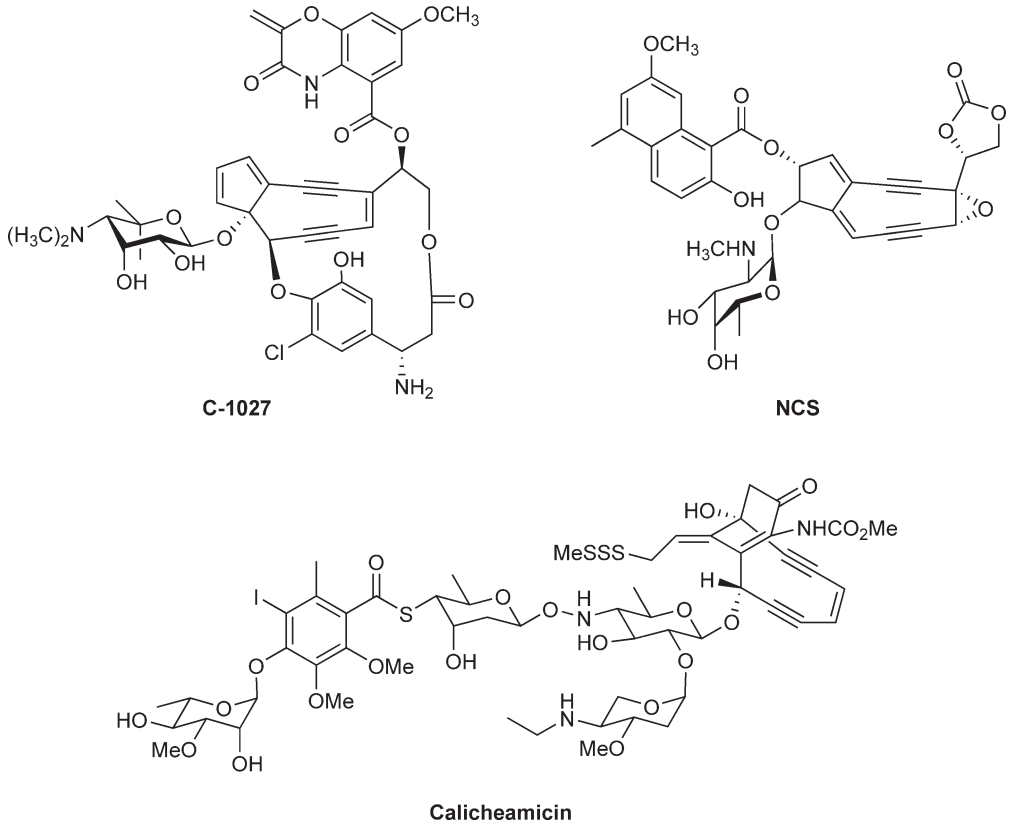


Figure 9 Structures of representative members of the enediyne family. C-1027 and NCS are two enediynes that require phosphopantetheinylation of proteins for their biosynthesis.

dekydroaustinol, peroxyergosterol, and shamizanthone. However, fatty acids, ergosterol, and cervisterol were still synthesized.[174] The deficient fungi were also not able to reproduce asexually without *cfwA* in the presence of the missing compounds. Phosphopantetheine tethering must be involved in some mechanism of the asexual reproduction of the filamentous fungi.

In mammals a single PPTase is used for the posttranslational modification of three different apo-proteins: the carrier proteins of mitochondrial and cytosolic FASs and the aminoadipate semialdehyde reductase implicated in lysine degradation. The crystal structure of human PPTase has been determined and found to be most closely related to the class II Sfp-like enzymes.[175] Architectural and mechanistic differences between the type II human PPTase and the type I bacterial PPTases include a divalent cation coordinated by the α-phosphate of CoA, a Glu and an Asp residue, and three water ligands in type I PPTases versus a divalent cation coordinated by α- and β-phosphates of CoA, two to three protein side chains, and a water molecule in the human PPTase.

5.15.11 Conclusions

Posttranslational modifications are a sundry set of transformations that help to diversify the limited genome of organisms. The modifications discussed in this chapter have been shown to modify a wide variety of proteins whose functions vary from cell division to metabolism and regulation. While a large selection of posttranslational modifications has been discussed, the presentation is not all-inclusive of all modifications. Emphasis has been placed on the discoveries made since 2005 and on the more common modifications. The importance of posttranslational modifications on protein structure and function and cellular function has been emphasized.

Acknowledgments

We would like to acknowledge the work on posttranslational modification of proteins of those not cited in this chapter either due to the scope of the chapter or our ignorance. We thank Dr. Oleg V. Tsodikov for critical reading of the manuscript and help in the preparation of **Figures 2**, **6**, and **9**. SGT would also like to thank Professors John C. Vederas and Christopher T. Walsh for their invaluable support and mentorship.

Abbreviations

A	adenylation
AcCoA	acetyl coenzyme A
ACP	acyl carrier protein
ADP	adenosine 5′-diphosphate
AMP	adenosine 3′,5′-monophoshpate
ARH	ADP-ribosylhydrolase
ART	ADP-ribosyltransferase
ATP	adenosine 5′-triphosphate
BCCP	biotin carboxyl carrier protein
BPL	biotin protein ligase
C	condensation
cAMP	3′,5′ cyclic adenosine-monophosphate
CARM1	coactivator-associated arginine methyltransferase
CBP	CREB binding protein
CIP	calf intestinal phosphatase
CKAP2	cytoskeleton associated protein 2
CoA	coenzyme A
COPD	chronic obstructive pulmonary disease
CREB	cAMP responsive element binding protein
Csk	C-terminal Src kinase
DHP	dematin head-piece
DNA	deoxyribonucleic acid
ELK-1	E-26 like protein 1
ER	endoplasmic reticulum
ERM	ezrin/radixin/moesin protein
FA	fatty acid
FAS	fatty acid synthase
FIH	factor inhibiting HIF
FOS	c-fos protein
FTase	farnesyltransferase
FTI	FTase inhibitor
GGTase	geranylgeranyltransferase
GGTI-298	geranylgeranyltransferase inhibitor 298
GlcNAc	*N*-acetyl glucosamine
GLUT4	glucose transporter 4
Grb2	growth receptor-bound protein 2
GSK-3b	glycogen synthase kinase 3, b isoform
GTP	guanine 5′-triphosphate
GTPases	a family of membrane-bound proteins that hydrolyze GTP, also known as G-proteins
H3	histone 3
H4	histone 4
HAT	histone acetyltransferase

HDAC	histone deacetylase
HIF	hypoxia-inducible factor
HMW	high molecular weight
HRGP	hydroxyproline-rich glycoprotein
HRSV	human respiratory syncytial virus
Icmt	isoprenylcysteine carboxylmethyltransferase
IR	insulin receptor
IRS-1	insulin receptor substrate 1
MAPK	mitogen-activated protein kinase
MEK	mitogen-activated or extracellular signal-regulated protein kinase
MMP	matrix metalloproteinase
mRNA	messenger RNA
NAD^+	nicotinamide adenine dinucleotide
NCS	neocarzinostatin
NGF	nerve growth factor
NGFR	nerve growth factor receptor
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthetase
PAPS	3′-phosphoadenosine-5′-phosphosulfate
PARP	poly-ADP-ribosyl polymerase
PCP	peptidyl carrier protein
PDK1	pyruvate dehydrogenase kinase 1
PI3K	phosphoinositide 3 kinase
PKB	protein kinase B
PKC	protein kinase C
PK	polyketide
PKS	polyketide synthase
PMT	protein O-mannosyltransferase
PP2A	protein phosphatase 2A
Ppant	phosphopantetheinyl
PPII	polyproline helix 2
PPTase	4′-phosphopantetheinyltransferase
Raf	protein kinase produced by *raf* gene
Ras	protein kinase produced by *ras* gene
RCC1	regulator of chromatin condensation 1
RhoA	aplysia Ras-related homologue A
RNA	ribonucleic acid
SAH	*S*-adenosylhomocysteine
SAM	*S*-adenosylmethionine
SH2	src homologue 2
Shc	src homologues and collagen protein
Sirt3	silent mating type information regulation 2 homologue protein 3
SNAP	*S*-nitrosyl-*N*-acetylpenicillamine
Sox	sulfur-oxidizing enzyme system
Src	kinase produced by *src* gene
SRp38	splicing regulator p38
Sulf1	sulfotransferase 1
T	thiolation
TBI	traumatic brain injury
TEI	type I thioesterase
TEII	type II thioesterase

TPP	thiamine pyrophosphate
TPST	tyrosylprotein sulfotransferase
TrkA	*trk* proto-oncogene A
TSP-1	thrimbospondin-1
UDP-GlcNAc	uracil 5′-diphosphate 2-*N*-acetyl glucosamine
UDP–glucose	uracil 5′-diphosphate glucose

References

1. T. Hunter, *Mol. Cell* **2007**, *28*, 730–738.
2. P. Chiarugi; F. Buricchi, *Antioxid. Redox Signal.* **2007**, *9*, 1–24.
3. C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, 1st ed.; Roberts and Company Publishers: Englewood, CO, 2005.
4. C. T. Walsh; S. Garneau-Tsodikova; G. J. Gatto, Jr., *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 7342–7372.
5. P. Trojer; D. Reinberg, *Cell* **2006**, *125*, 213–217.
6. Z. Wang; J. Lee; A. R. Cossins; M. Brust, *Anal. Chem.* **2005**, *77*, 5770–5774.
7. Z. Wang; R. Levy; D. G. Fernig; M. Brust, *J. Am. Chem. Soc.* **2006**, *128*, 2214–2215.
8. K. D. Green; M. K. Pflum, *J. Am. Chem. Soc.* **2007**, *129*, 10–11.
9. N. Khidekel; S. Arndt; N. Lamarre-Vincent; A. Lippert; K. G. Poulin-Kerstien; B. Ramakrishnan; P. K. Qasba; L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2003**, *125*, 16162–16163.
10. M. Yu; L. P. de Carvalho; G. Sun; J. S. Blanchard, *J. Am. Chem. Soc.* **2006**, *128*, 15356–15357.
11. C. Dalhoff; G. Lukinavicius; S. Klimasauskas; E. Weinhold, *Nat. Protoc.* **2006**, *1*, 1879–1886.
12. B. P. Duckworth; Z. Zhang; A. Hosokawa; M. D. Distefano, *Chembiochem* **2007**, *8*, 98–105.
13. J. J. La Clair; T. L. Foley; T. R. Schegg; C. M. Regan; M. D. Burkart, *Chem. Biol.* **2004**, *11*, 195–201.
14. A. Sickmann; H. E. Meyer, *Proteomics* **2001**, *1*, 200–206.
15. J. Aqvist; K. Kolmodin; J. Florian; A. Warshel, *Chem. Biol.* **1999**, *6*, R71–R80.
16. P. A. Cole; D. Sondhi; K. Kim, *Pharmacol. Ther.* **1999**, *82*, 219–229.
17. S. Garneau-Tsodikova; V. J. Thannickal, *Curr. Med. Chem.* **2008**, *15*, 2632–2640.
18. J. Reinders; A. Sickmann, *Proteomics* **2005**, *5*, 4052–4061.
19. G. Manning; D. B. Whyte; R. Martinez; T. Hunter; S. Sudarsanam, *Science* **2002**, *298*, 1912–1934.
20. J. A. Adams, *Chem. Rev.* **2001**, *101*, 2271–2290.
21. T. Hunter, *Cell* **1995**, *80*, 225–236.
22. S. K. Hanks, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 388–392.
23. S. K. Hanks; A. M. Quinn; T. Hunter, *Science* **1988**, *241*, 42–52.
24. K. Niefind; M. Putter; B. Guerra; O. G. Issinger; D. Schomburg, *Nat. Struct. Biol.* **1999**, *6*, 1100–1103.
25. V. Helms; J. A. McCammon, *Protein Sci.* **1997**, *6*, 2336–2343.
26. P. Karwowska-Desaulniers; A. Ketko; N. Kamath; M. K. H. Pflum, *Biochem. Biophys. Res. Commun.* **2007**, *361*, 349–355.
27. A. Asenjo; J. C. Gonzalez-Armas; N. Villanueva, *Virology* **2008**, *380* (1), 26–33.
28. K. U. Hong; Y.-B. Choi; J.-H. Lee; H.-J. Kim; H.-R. Kwon; Y.-S. Seong; H. T. Kim; J. Park; C.-D. Bae; K.-M. Hong, *Exp. Mol. Med.* **2008**, *40*, 377–386.
29. C. Jorgensen; R. Linding, *Brief Funct. Genomic Proteomic* **2008**, *7*, 17–26.
30. Y. Feng; M. Chen; J. L. Manley, *Nat. Struct. Mol. Biol.* **2008**, *15* (10), 1040–1048.
31. J. Zdychova; J. Vesela; L. Kazdova; R. Komers, *Physiol. Res.* **2007**, *54*, 1–16.
32. P. Bevan, *J. Cell Sci.* **2001**, *114*, 1429–1430.
33. S. O. Yoon; S. P. Soltoff; M. V. Chao, *J. Biol. Chem.* **1997**, *272*, 23231–23238.
34. J. C. Chrivia; R. P. Kwok; N. Lamb; M. Hagiwara; M. R. Montminy; R. H. Goodman, *Nature* **1993**, *365*, 855–859.
35. D. Parker; K. Ferreri; T. Nakajima; V. J. LaMorte; R. Evans; S. C. Koerber; C. Hoeger; M. R. Montminy, *Mol. Cell Biol.* **1996**, *16*, 694–703.
36. I. Radhakrishnan; G. C. Perez-Alvarado; D. Parker; H. J. Dyson; M. R. Montminy; P. E. Wright, *Cell* **1997**, *91*, 741–752.
37. B. Mayr; M. Montminy, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 599–609.
38. M. Johannessen; M. P. Delghandi; U. Moens, *Cell Signal.* **2004**, *16*, 1211–1227.
39. A. M. Bode; Z. Dong, *Nat. Rev. Cancer* **2004**, *4*, 793–805.
40. A. C. Kim; A. C. Azim; A. H. Chishti, *Biochim. Biophys. Acta* **1998**, *1398*, 382–386.
41. M. Lutchman; A. C. Kim; L. Cheng; I. P. Whitehead; S. S. Oh; M. Hanspal; A. A. Boukharov; T. Hanada; A. H. Chishti, *Eur. J. Biochem.* **2002**, *269*, 638–649.
42. Z. G. Jiang; C. J. McKnight, *Structure* **2006**, *14*, 379–387.
43. T. Hunter; B. M. Sefton, *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1311–1315.
44. P. A. Cole; K. Shen; Y. Qiao; D. Wang, *Curr. Opin. Chem. Biol.* **2003**, *7*, 580–585.
45. F. Monigatti; B. Hekking; H. Steen, *Biochim. Biophys. Acta* **2006**, *1764*, 1904–1913.
46. T. Pawson; J. D. Scott, *Science* **1997**, *278*, 2075–2080.
47. G. Diez-Roux; A. Ballabio, *Annu. Rev. Genomics Hum. Genet* **2005**, *6*, 355–379.
48. E. J. Fernandez; E. Lolis, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 469–499.

49. J. Liu; S. Louie; W. Hsu; M. K. Yu; H. B. Nicholas, Jr.; G. L. Rosenquist, *Am. J. Respir. Cell Mol. Biol.* **2008**, *38*, 738–743.
50. C. Kasinathan; N. Gandhi; P. Ramaprasad; P. Sundaram; N. Ramasubbu, *Int. J. Biol. Sci.* **2007**, *3*, 237–241.
51. D. I. Hay; D. J. Smith; S. K. Schluckebier; E. C. Moreno, *J. Dent. Res.* **1984**, *63*.
52. T. D. Foley; M. E. Kintner, *Biochem. Biophys. Res. Commun.* **2005**, *330*, 1224–1229.
53. A. Quentmeier; P. Janning; P. Hellwig; C. G. Friedrich, *Biochemistry* **2007**, *46*, 10990–10998.
54. W. K. Paik; D. C. Paik; S. Kim, *Trends Biochem. Sci.* **2007**, *32*, 146–152.
55. K. Orford; P. Kharchenko; W. Lai; M. C. Dao; D. J. Worhunsky; A. Ferro; V. Janzen; P. J. Park; D. T. Scadden, *Dev. Cell* **2008**, *14*, 798–809.
56. L. P. O'Neill; H. T. Spotswood; M. Fernando; B. M. Turner, *Biol. Chem.* **2008**, *389*, 365–370.
57. A. Scoumanne; X. Chen, *Cell. Mol. Biol.* **2008**, *23*, 1143–1149.
58. D. Cheng; J. Cote; S. Shaaban; M. T. Bedford, *Mol. Cell* **2007**, *25*, 71–83.
59. S. Daujat; U. M. Bauer; V. Shah; B. Turner; S. Berger; T. Kouzarides, *Curr. Biol.* **2002**, *12*, 2090–2097.
60. S. Huang; M. Litt; G. Felsenfeld, *Genes Dev.* **2005**, *19*, 1885–1893.
61. C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, 1st ed.; Roberts and Company Publishers: Englewood, CO, 2006.
62. M. Hetzer; O. J. Gruss; I. W. Mattaj, *Nat. Cell Biol.* **2002**, *4*, E177–E184.
63. T. Chen; T. L. Muratore; C. E. Schaner-Tooley; J. Shabanowitz; D. F. Hunt; I. G. Macara, *Nat. Cell Biol.* **2007**, *9*, 596–603.
64. M. E. Nemergut; C. A. Mizzen; T. Stukenberg; C. D. Allis; I. G. Macara, *Science* **2001**, *292*, 1540–1543.
65. H. Y. Li; D. Wirtz; Y. Zheng, *J. Cell Biol.* **2003**, *160*, 635–644.
66. R. Sprung; Y. Chen; K. Zhang; D. Cheng; T. Zhang; J. Peng; Y. Zhao, *J. Proteome Res.* **2008**, *7*, 1001–1006.
67. K. F. Leung; R. Baron; B. R. Ali; A. I. Magee; M. C. Seabra, *J. Biol. Chem.* **2007**, *282*, 1487–1497.
68. A. Kowluru, *J. Cell. Mol. Med.* **2008**, *12*, 164–173.
69. L. L. Cook; A. Gafni, *J. Biol. Chem.* **1988**, *263*, 13991–13993.
70. S. Mukherjee; Y. H. Hao; K. Orth, *Trends Biochem. Sci.* **2007**, *32*, 210–216.
71. X. J. Yang; E. Seto, *Mol. Cell* **2008**, *31*, 449–461.
72. H. Fukuda; N. Sano; S. Muto; M. Horikoshi, *Brief Funct. Genomic Proteomic* **2006**, *5*, 190–208.
73. M. G. Rosenfeld; V. V. Lunyak; C. K. Glass, *Genes Dev.* **2006**, *20*, 1405–1428.
74. T. Hirota; J. J. Lipp; B. H. Toh; J. M. Peters, *Nature* **2005**, *438*, 1176–1180.
75. E. J. Richards; S. C. Elgin, *Cell* **2002**, *108*, 489–500.
76. K. Ito; C. E. Charron; I. M. Adcock, *Pharmacol. Ther.* **2007**, *116*, 249–265.
77. E. E. Vaughan; R. C. Geiger; A. M. Miller; P. L. Loh-Marley; T. Suzuki; N. Miyata; D. A. Dean, *Mol. Ther.* **2008**, *16* (11), 1841–1847.
78. B.-H. Ahn; H.-S. Kim; S. Song; I. H. Lee; J. Liu; A. Vassilopoulos; C.-X. Deng, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14447–14452.
79. V. S. Ananthanarayanan, *J. Biomol. Struct. Dyn.* **1983**, *1*, 843–855.
80. C. L. Jenkins; R. T. Raines, *Nat. Prod. Rep.* **2002**, *19*, 49–59.
81. Y. Lim; J. W. Suh; S. Kim; B. Hyun; C. Kim; C. H. Lee, *J. Antibiot. (Tokyo)* **1994**, *47*, 1406–1416.
82. E. Lopez-Vera; A. Walewska; J. J. Skalicky; B. M. Olivera; G. Bulaj, *Biochemistry* **2008**, *47*, 1741–1751.
83. G. I. Cassab, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 281–309.
84. A. M. Showalter, *Cell. Mol. Life Sci.* **2001**, *58*, 1361–1362.
85. J. H. Lee; S. Waffenschmidt; L. Small; U. Goodenough, *Plant Physiol.* **2007**, *144*, 1813–1826.
86. M. Ivan; K. Kondo; H. Yang; W. Kim; J. Valiando; M. Ohh; A. Salic; J. M. Asara; W. S. Lane; W. G. Kaelin, Jr., *Science* **2001**, *292*, 464–468.
87. P. Jaakkola; D. R. Mole; Y. M. Tian; M. I. Wilson; J. Gielbert; S. J. Gaskell; A. Kriegsheim; H. F. Hebestreit; M. Mukherji; C. J. Schofield; P. H. Maxwell; C. W. Pugh; P. J. Ratcliffe, *Science* **2001**, *292*, 468–472.
88. F. Yu; S. B. White; Q. Zhao; F. S. Lee, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9630–9635.
89. W. G. Kaelin, *Annu. Rev. Biochem.* **2005**, *74*, 115–128.
90. D. Lando; D. J. Peet; J. J. Gorman; D. A. Whelan; M. L. Whitelaw; R. K. Bruick, *Genes Dev.* **2002**, *16*, 1466–1471.
91. Y. K. Park; D. R. Ahn; M. Oh; T. Lee; E. G. Yang; M. Son; H. Park, *Mol. Pharmacol.* **2008**, *74*, 236–245.
92. J. Nedvidkova; K. Smitka; V. Kopsky; V. Hainer, *Physiol. Res.* **2005**, *54*, 133–140.
93. Y. Wang; K. S. Lam; L. Chan; K. W. Chan; J. B. Lam; M. C. Lam; R. C. Hoo; W. W. Mak; G. J. Cooper; A. Xu, *J. Biol. Chem.* **2006**, *281*, 16391–16400.
94. E. Weerapana; B. Imperiali, *Glycobiology* **2006**, *16*, 91R–101R.
95. F. Blanchet; B. T. Schurter; O. Acuto, *Curr. Opin. Immunol.* **2006**, *18*, 321–328.
96. M. Kieliszewski; R. de Zacks; J. F. Leykam; D. T. Lamport, *Plant Physiol.* **1992**, *98*, 919–926.
97. N. E. Zachara; G. W. Hart, *Biochim. Biophys. Acta* **2006**, *1761*, 599–617.
98. M. Goto, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 1415–1427.
99. W. B. Dias; G. W. Hart, *Mol. Biosyst.* **2007**, *3*, 766–772.
100. F. Altmann, *Int. Arch. Allergy Immunol.* **2006**, *142*, 99–115.
101. B. Aggarwal; A. Kumar; A. Bharti, *Anticancer Res.* **2003**, *23*, 363–398.
102. S. K. Jain; J. Rains; K. Jones, *Free Radical Biol. Med.* **2006**, *41*, 92–96.
103. P. Raman; I. Krukovets; T. E. Marinic; P. Bornstein; O. I. Stenina, *J. Biol. Chem.* **2007**, *282*, 5704–5714.
104. J. S. Isenberg; F. Hyodo; L. A. Ridnour; C. S. Shannon; D. A. Wink; M. C. Krishna; D. D. Roberts, *Neoplasia* **2008**, *10*, 886–896.
105. G. O. Elpek; G. A. Gokhan; S. Bozova, *World J. Gastrol.* **2008**, *14*, 2213–2217.
106. M. Ueno; S. Yamada; M. Zako; M. Bernfield; K. Sugahara, *J. Biol. Chem.* **2001**, *276*, 29134–29140.
107. R. K. Ambasta; X. Ai; C. P. Emerson, Jr., *J. Biol. Chem.* **2007**, *282*, 34492–34499.
108. S. K. Prill; B. Klinkert; C. Timpel; C. A. Gale; K. Schroppel; J. F. Ernst, *Mol. Microbiol.* **2005**, *55*, 546–560.
109. C. Bai; X. L. Xu; F. Y. Chan; R. T. Lee; Y. Wang, *Eukaryotic Cell* **2006**, *5*, 238–247.
110. F. Koch-Nolte; S. Kernstock; C. Mueller-Dieckmann; M. S. Weiss; F. Haag, *Front. Biosci.* **2008**, *13*, 6716–6729.
111. H. Lin, *Org. Biomol. Chem.* **2007**, *5*, 2541–2554.

112. P. O. Hassa; S. S. Haenni; M. Elser; M. O. Hottiger, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 789–829.
113. A. W. Maresso; Q. Deng; M. S. Pereckas; B. T. Wakim; J. T. Barbieri, *Cell. Microbiol.* **2007**, *9*, 97–105.
114. A. I. Scovassi, *Riv. Biol.* **2007**, *100*, 351–360.
115. A. Burkle, *FEBS J.* **2005**, *272*, 4576–4589.
116. M. Cohen-Armon, *Trends Pharmacol. Sci.* **2007**, *28*, 556–560.
117. P. O. Hassa; M. O. Hottiger, *Front. Biosci.* **2008**, *13*, 3046–3082.
118. S. Beneke; A. Burkle, *Nucleic Acids Res.* **2007**, *35*, 7456–7465.
119. Y. Lai; Y. Chen; S. C. Watkins; P. D. Nathaniel; F. Guo; P. M. Kochanek; L. W. Jenkins; C. Szabo; R. S. Clark, *J. Neurochem.* **2008**, *104*, 1700–1711.
120. K. T. Lane; L. S. Beese, *J. Lipid Res.* **2006**, *47*, 681–699.
121. K. F. Leung; R. Baron; M. C. Seabra, *J. Lipid Res.* **2006**, *47*, 467–475.
122. Y. Takai; T. Sasaki; T. Matozaki, *Physiol. Rev.* **2001**, *81*, 153–208.
123. P. J. Casey; M. C. Seabra, *J. Biol. Chem.* **1996**, *271*, 5289–5292.
124. S. Maurer-Stroh; S. Washietl; F. Eisenhaber, *Genome Biol.* **2003**, *4*, 212.
125. M. C. Seabra; Y. Reiss; P. J. Casey; M. S. Brown; J. L. Goldstein, *Cell* **1991**, *65*, 429–434.
126. H. W. Fu; P. J. Casey, *Recent Prog. Horm. Res.* **1999**, *54*, 315–342; Discussion 342–343.
127. A. Kowluru; G. Li; S. A. Metz, *J. Clin. Invest.* **1997**, *100*, 1596–1610.
128. T. Amet; M. Nonaka; M. Z. Dewan; Y. Saitoh; X. Qi; S. Ichinose; N. Yamamoto; S. Yamaoka, *Microbes Infect.* **2008**, *10*, 471–480.
129. G. del Real; S. Jimenez-Baranda; E. Mira; R. A. Lacalle; P. Lucas; C. Gomez-Mouton; M. Alegret; J. M. Pena; M. Rodriguez-Zapata; M. Alvarez-Mon; A. C. Martinez; S. Manes, *J. Exp. Med.* **2004**, *200*, 541–547.
130. C. Gilbert; M. Bergeron; S. Methot; J. F. Giguere; M. J. Tremblay, *Viral Immunol.* **2005**, *18*, 474–489.
131. J. F. Giguere; M. J. Tremblay, *J. Virol.* **2004**, *78*, 12062–12065.
132. A. M. Abeles; N. Marjanovic; J. Park; M. Attur; E. S. Chan; H. E. Al-Mussawir; M. Dave; M. C. Fisher; S. A. Stuchin; S. B. Abramson; M. H. Pillinger, *Arthritis Rheum.* **2007**, *56*, 2840–2853.
133. C. Ritchlin, *Arthritis Res.* **2000**, *2*, 356–360.
134. U. Muller-Ladner; S. Gay, *Ann. Rheum. Dis.* **2002**, *61*, 957–959.
135. B. Bagautdinov; Y. Matsuura; S. Bagautdinova; N. Kunishima, *J. Biol. Chem.* **2008**, *283*, 14739–14750.
136. A. Bilska; M. Dudek; M. Iciek; I. Kwiecien; M. Sokolowska-Jezewicz; B. Filipek; L. Wlodek, *Pharmacol. Rep.* **2008**, *60*, 225–232.
137. R. N. Perham, *Annu. Rev. Biochem.* **2000**, *69*, 961–1004.
138. A. C. Mercer; M. D. Burkart, *Nat. Prod. Rep.* **2007**, *24*, 750–773.
139. M. B. Austin; J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110.
140. Y. Kim; J. H. Prestegard, *Proteins* **1990**, *8*, 377–385.
141. M. P. Crump; J. Crosby; C. E. Dempsey; J. A. Parkinson; M. Murray; D. A. Hopwood; T. J. Simpson, *Biochemistry* **1997**, *36*, 6000–6008.
142. K. D. Parris; L. Lin; A. Tam; R. Mathew; J. Hixon; M. Stahl; C. C. Fritz; J. Seehra; W. S. Somers, *Structure* **2000**, *8*, 883–895.
143. T. Weber; R. Baumgartner; C. Renner; M. A. Marahiel; T. A. Holak, *Structure* **2000**, *8*, 407–418.
144. G. Y. Xu; A. Tam; L. Lin; J. Hixon; C. C. Fritz; R. Powers, *Structure* **2001**, *9*, 277–287.
145. A. Roujeinikova; C. Baldock; W. J. Simon; J. Gilroy; P. J. Baker; A. R. Stuitje; D. W. Rice; A. R. Slabas; J. B. Rafferty, *Structure* **2002**, *10*, 825–835.
146. H. C. Wong; G. Liu; Y. M. Zhang; C. O. Rock; J. Zheng, *J. Biol. Chem.* **2002**, *277*, 15874–15880.
147. Q. Li; C. Khosla; J. D. Puglisi; C. W. Liu, *Biochemistry* **2003**, *42*, 4648–4657.
148. G. A. Zornetzer; B. G. Fox; J. L. Markley, *Biochemistry* **2006**, *45*, 5217–5227.
149. V. Y. Alekseyev; C. W. Liu; D. E. Cane; J. D. Puglisi; C. Khosla, *Protein Sci.* **2007**, *16*, 2093–2107.
150. S. A. Samel; G. Schoenafinger; T. A. Knappe; M. A. Marahiel; L. O. Essen, *Structure* **2007**, *15*, 781–792.
151. D. P. Frueh; H. Arthanari; A. Koglin; D. A. Vosburg; A. E. Bennett; C. T. Walsh; G. Wagner, *Nature* **2008**, *454*, 903–906.
152. A. Koglin; F. Lohr; F. Bernhard; V. V. Rogov; D. P. Frueh; E. R. Strieter; M. R. Mofid; P. Guntert; G. Wagner; C. T. Walsh; M. A. Marahiel; V. Dotsch, *Nature* **2008**, *454*, 907–911.
153. A. Tanovic; S. A. Samel; L. O. Essen; M. A. Marahiel, *Science* **2008**, *321*, 659–663.
154. L. Wang; J. McVey; L. C. Vining, *Microbiology* **2001**, *147*, 1535–1545.
155. A. G. Bobrov; V. A. Geoffroy; R. D. Perry, *Infect. Immun.* **2002**, *70*, 4204–4214.
156. C. Sanchez; L. Du; D. J. Edwards; M. D. Toney; B. Shen, *Chem. Biol.* **2001**, *8*, 725–738.
157. H. Oberegger; M. Eisendle; M. Schrettl; S. Graessle; H. Haas, *Curr. Genet.* **2003**, *44*, 211–215.
158. A. K. Joshi; L. Zhang; V. S. Rangan; S. Smith, *J. Biol. Chem.* **2003**, *278*, 33142–33149.
159. K. J. Weissman; H. Hong; M. Oliynyk; A. P. Siskos; P. F. Leadlay, *Chembiochem* **2004**, *5*, 116–125.
160. S. Sunaga; H. Li; Y. Sato; Y. Nakagawa; T. Matsuyama, *Microbiol. Immunol.* **2004**, *48*, 723–728.
161. Q. Liu; Y. Ma; L. Zhou; Y. Zhang, *Arch. Microbiol.* **2005**, *183*, 37–44.
162. Y. Huang; E. Wendt-Pienkowski; B. Shen, *J. Biol. Chem.* **2006**, *281*, 29660–29668.
163. R. H. Lambalot; C. T. Walsh, *J. Biol. Chem.* **1995**, *270*, 24658–24661.
164. L. E. Quadri; P. H. Weinreb; M. Lei; M. M. Nakano; P. Zuber; C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595.
165. I. B. Lomakin; Y. Xiong; T. A. Steitz, *Cell* **2007**, *129*, 319–332.
166. F. Fichtlscherer; C. Wellein; M. Mittag; E. Schweizer, *Eur. J. Biochem.* **2000**, *267*, 2666–2671.
167. P. Meiser; R. Muller, *Chembiochem* **2008**, *9*, 1549–1553.
168. C. Chalut; L. Botella; C. de Sousa-D'Auria; C. Houssin; C. Guilhot, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8511–8516.
169. J. N. Copp; B. A. Neilan, *Appl. Environ. Microbiol.* **2006**, *72*, 2298–2305.
170. J. N. Copp; A. A. Roberts; M. A. Marahiel; B. A. Neilan, *J. Bacteriol.* **2007**, *189*, 3133–3139.
171. C. Garcia-Estrada; R. V. Ullan; T. Velasco-Conde; R. P. Godio; F. Teijeira; I. Vaca; R. Feltrer; K. Kosalkova; E. Mauriz; J. F. Martin, *Biochem. J.* **2008**, *415*, 317–324.
172. J. Zhang; S. G. Van Lanen; J. Ju; W. Liu; P. C. Dorrestein; W. Li; N. L. Kelleher; B. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1460–1465.

173. E. Murugan; Z. X. Liang, *FEBS Lett.* **2008**, *582*, 1097–1103.
174. O. Marquez-Fernandez; A. Trigos; J. L. Ramos-Balderas; G. Viniegra-Gonzalez; H. B. Deising; J. Aguirre, *Eukaryotic Cell* **2007**, *6*, 710–720.
175. G. Bunkoczi; S. Pasta; A. Joshi; X. Wu; K. L. Kavanagh; S. Smith; U. Oppermann, *Chem. Biol.* **2007**, *14*, 1243–1253.

Biographical Sketches

Keith D. Green was born in Michigan, USA, in 1980. He majored in chemistry at Michigan Technological University (BS 2003), where he attended summer research programs at the North Carolina State University (Summer 2001) and the University of Alabama at Tuscaloosa (Summer 2002) where he worked with Professors Michael Overcash and Robin Rogers, respectively. He obtained his Ph.D. in chemistry in 2008 at Wayne State University where he explored the cosubstrate promiscuity of protein kinases in the laboratory of Professor Mary Kay Pflum. Now a postdoctoral fellow in Professor Sylvie Garneau-Tsodikova's laboratory, he is studying new tools for combinatorial biosynthesis and the production of novel aminoglycoside antibiotics.

Sylvie Garneau-Tsodikova was born in Québec, Canada, in 1973. She majored in chemistry at the Université Laval (BSc 1995, MSc 1997), where she worked with Professors Robert Chênevert and Perséphone Canonne. She obtained her Ph.D. in chemistry in 2003 at the University of Alberta where she studied new antimicrobial agents acting on bacterial cell walls in the laboratory of Professor John C. Vederas. As a postdoctoral fellow in Professor Christopher T. Walsh laboratory at Harvard Medical School, she studied halogenation and pyrrole formation during the biosynthesis of various natural products. In 2006, she joined the University of Michigan as the John G. Searle Assistant Professor of Medicinal Chemistry at the College of Pharmacy and the Life Sciences Institute. Her research interests lie in the areas of mechanistic enzymology, in the development of new tools for combinatorial biosynthesis of novel nonribosomal peptides, and in the chemoenzymatic formation of new aminoglycoside antibiotics.

5.16 Collagen Formation and Structure

Hans Peter Bächinger, Kazunori Mizuno, Janice A. Vranka, and Sergei P. Boudko, Oregon Health & Science University, Portland, OR, USA

© 2010 Elsevier Ltd. All rights reserved.

5.16.1 Introduction

This chapter aims to provide an extensive coverage of the collagen proteins. Collagens are the major components of the extracellular matrix of multicellular animals, and facilitate the formation and maintenance of a multicellular system, particularly in tissue or organ skeletal structures in large vertebrates. It has become evident that collagens serve as solid-state regulators of cellular functions. The primary sequence of polypeptides required for collagen triple helix formation has a glycine at every third residue, that is, Gly-Xaa-Yaa-. This primary sequence allows for the formation of the characteristic triple helix. Glycine is the only amino acid that can be packed tightly at the center of the triple-stranded collagen monomer, where Gly provides an NH group for hydrogen bonding to the O=C– group of the X position residue of another chain (see Section 5.16.6.2 for details). The triple helical structure of collagen (**Figure 1**) shows the unique characteristics of collagens: (1) a highly elongated shape, (2) all side chains of the Xaa and Yaa position residues exposed on the molecular surface, and (3) the surface of the molecule consisting of three staggered polypeptide chains. In this chapter, recent progress in collagen research during the past decade is reviewed, with the main focus on the triple helical structure, the biological function of different collagens, collagens in relation to human diseases, and mouse models for collagens and collagen-modifying proteins.

5.16.2 Molecular Structure and Biological Function of Collagen Types

Collagen is the most abundant extracellular matrix protein family in vertebrates. Proteins in the collagen superfamily all have three polypeptide chains with the required -Gly-Xaa-Yaa- repeated sequence, where Xaa and Yaa are frequently proline and 4-hydroxyproline, respectively. At present, more than 30 molecular species of vertebrate proteins called 'collagen' are classified into 28 types as type I, II, III,. . ., XXVIII. They are typically called 'type N collagen', or 'collagen N'. In addition, there are many more collagen-like proteins that

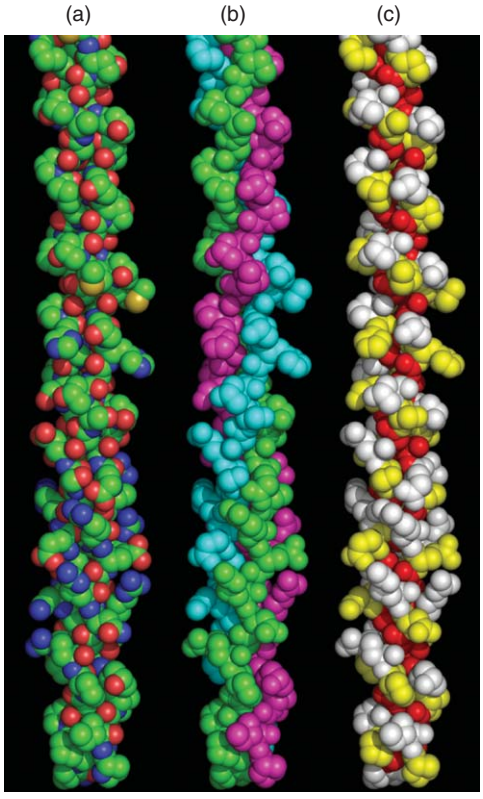


Figure 1 The collagen triple helix. (a) Colored by element C (green), O (red), N (blue), S (yellow); (b) colored by chains; and (c) colored by Gly (red), Xaa (yellow), and Yaa (white) positions. The models were built using PyMOL[1] with the type III collagen model peptide (PDB accession number 3DMW).[2] Glycine residues in every third position are located on the inside of the triple helix. Each chain is staggered by one residue to accommodate the Gly in the center of the triple helix.

are not called 'collagen', most of which are soluble proteins related to innate immunity, such as adiponectin[3] CTRP9,[4] macrophage scavenger receptor,[5] surfactant proteins,[6] hibernation proteins,[7] and C1q.[8–10] In invertebrates, which constitute more than 95% of all known species of animals, there is a much wider variety of collagenous proteins.[11–13] In this chapter, we focus only on the extracellular and membrane collagen molecules from vertebrates that are called 'collagen'. Each vertebrate collagen gene is termed COL*m*A*n*, where *m* indicates the type of collagen and *n* the number of α chains. Some types of collagen have more than one gene, that is, they form heterotrimeric molecules (e.g., type I, IV, or V collagen) or two different homotrimers (e.g., type VIII collagen). The human gene name is represented by uppercase letters as in human COL1A1. In other species, the genes are indicated by lowercase letters as in mouse collal. The polypeptide chain of each collagen gene product is called the 'α chain'. Cross-linked α chains are called β for two covalently bonded chains, 'γ' for three, 'δ' for four, and so on. These are relics from the original biochemical analyses of collagen in which the unit of the type I collagen molecule was unknown. Collagens are multidomain proteins. Some of these domains are enzymatically processed in order to form higher ordered structures that are found in tissues. The polypeptide chains that are processed in tissue form are called pro-α chains. The proteolytically cleaved parts of the portion of the procollagen molecule are called the propeptides. Many of the collagen family members contain more than one domain. In such cases, the domains are usually called NC1, NC2, and NC3, or COL1 and COL2 domains, where NC and COL indicate noncollagenous and collagenous domains, respectively. Domain 1 is usually at the C-terminal domain. However, some researchers use a different notation and refer to the N-terminal domain as NC1 domain.

Collagen was previously thought of as a matrix scaffolding protein without any significant biological roles, but more recent work has established that collagen plays a vital role in various physiological functions of all

tissues and organs in the body. Recent studies have demonstrated essential roles for the collagen molecules in neural development.[14,15] In this chapter, we will briefly summarize the structure and function of the currently known vertebrate collagen molecules. Vertebrate collagens are classified into several groups by structure, supramolecular assembly, or localization in tissues. The human collagen gene products are schematically illustrated in **Figure 2**.

We classify the 28 types of vertebrate collagens as classical fibrillar collagens (I, II, III, V, and XI), basement membrane collagens (IV), basement membrane zone collagens (XV and XVIII), type VI collagen, type VII collagen, short-chain collagens (VIII and X), FACIT (fibril-associated collagen with interrupted triple helices) (IX, XII, and XIV), FACIT-like collagens (XVI, XIX, XX, XXI, and XXII), transmembrane collagens (XIII, XVII, XXIII, and XXV), new fibrillar collagens (XXIV and XXVII), type XXVI collagen, and type XXVIII collagen. The localization of different types of collagen molecules is schematically shown in **Figure 3**.

5.16.2.1 Classical Fibrillar Collagens (Types I, II, III, V, and XI)

Fibrillar collagen, or fibril-forming collagen, has more than 1000 residues of an uninterrupted repeat of -Gly-Xaa-Yaa- sequences. Type I, II, III, V, and XI collagens belong to the classical fibrillar collagen. Recently, two more fibrillar collagens have been found: types XXIV[16] and XXVII.[17,18] The genes of these two new types are less than 1000 residues with two interruptions of -Gly-Xaa-Yaa- repeated sequences. Classical fibrillar collagens (I, II, III, V, and XI) are the major components of collagen fibrils in the body. These five types of fibrillar collagen can be further classified based on sequence similarity as type I–III and V/XI collagens. Type I and II collagens are the major components of the collagen fibril. They do not have large NC domains after processing. Type V and XI collagens are the minor components of collagen fibrils. Fibrillar collagens are synthesized as procollagen molecules with the N- and the C-propeptides. During deposition into tissues, these propeptides are cleaved by enzymes. The C-propeptide of fibrillar collagen is the association domain that selects the appropriate three pro-α chains and initiates the folding of the triple helix.[19]

5.16.2.2 Fibrillar Collagens (Types I, II, and III)

5.16.2.2.1 Type I collagen

Type I collagen is the most abundant collagen in vertebrates. It is suggested that evolution of this type of collagen made it possible for larger vertebrate animals to live outside water and to resist the gravitational forces that are present outside of an aquatic environment. It is estimated that more than 90% of all collagen molecules in the body are type I collagen. This molecule is a major component of collagen fibrils in bone, tendon, ligament, skin, and all major organs such as heart, kidney, liver, lung, and spleen. In the human body, type II collagen exists as the major component of collagen fibrils only in hyaline cartilage tissue.

Type I collagen molecules are heterotrimeric molecules with an α chain composition of two α1 (I) chains and one α2(I) chain from the COL1A1 and COL1A2 genes, respectively. This α chain composition is written as $[\alpha1(I)]_2\alpha2(I)$. A type I collagen molecule is synthesized as a procollagen type I molecule. The N- and C-terminal propeptides are cleaved as part of the processing in tissues and are called N- and C-propeptides, respectively. The human pro-α(I) and pro-α2(I) have 246 and 247 residues in their C-propeptides, respectively. The C-propeptide contains interchain disulfide bonds, which are important for the stabilization of the three α chains in the endoplasmic reticulum (ER).

However, the mechanism by which the 2:1 heterotrimer ratio is formed is still unknown. Folding of the collagen triple helix starts at the C-terminus and proceeds in a zipper-like fashion toward the amino terminus. *In vitro* experiments have demonstrated that the α2(I) homotrimer is less stable compared with α1(I) homotrimer and $[\alpha1(I)]_2\alpha2(I)$ heterotrimer.[20] Under specific conditions, a homotrimer of α1(I) can be synthesized, for example, in the *oim* mouse model, where the α2(I) chain is inactivated, and this makes the type I collagen molecule an α1(I) homotrimer.[21] Since the synthesized procollagen molecules are soluble in cultured media, one of the predicted functions of the N-propeptide is to prevent unexpected aggregation of newly synthesized procollagen chains in the rough endoplasmic reticulum (rER) and before deposition into the extracellular space. Another function is to help in the transport of the procollagen molecules from the ER to Golgi. A final predicted function of the N-propeptide is to interact with other matrix molecules. Recently, many processing

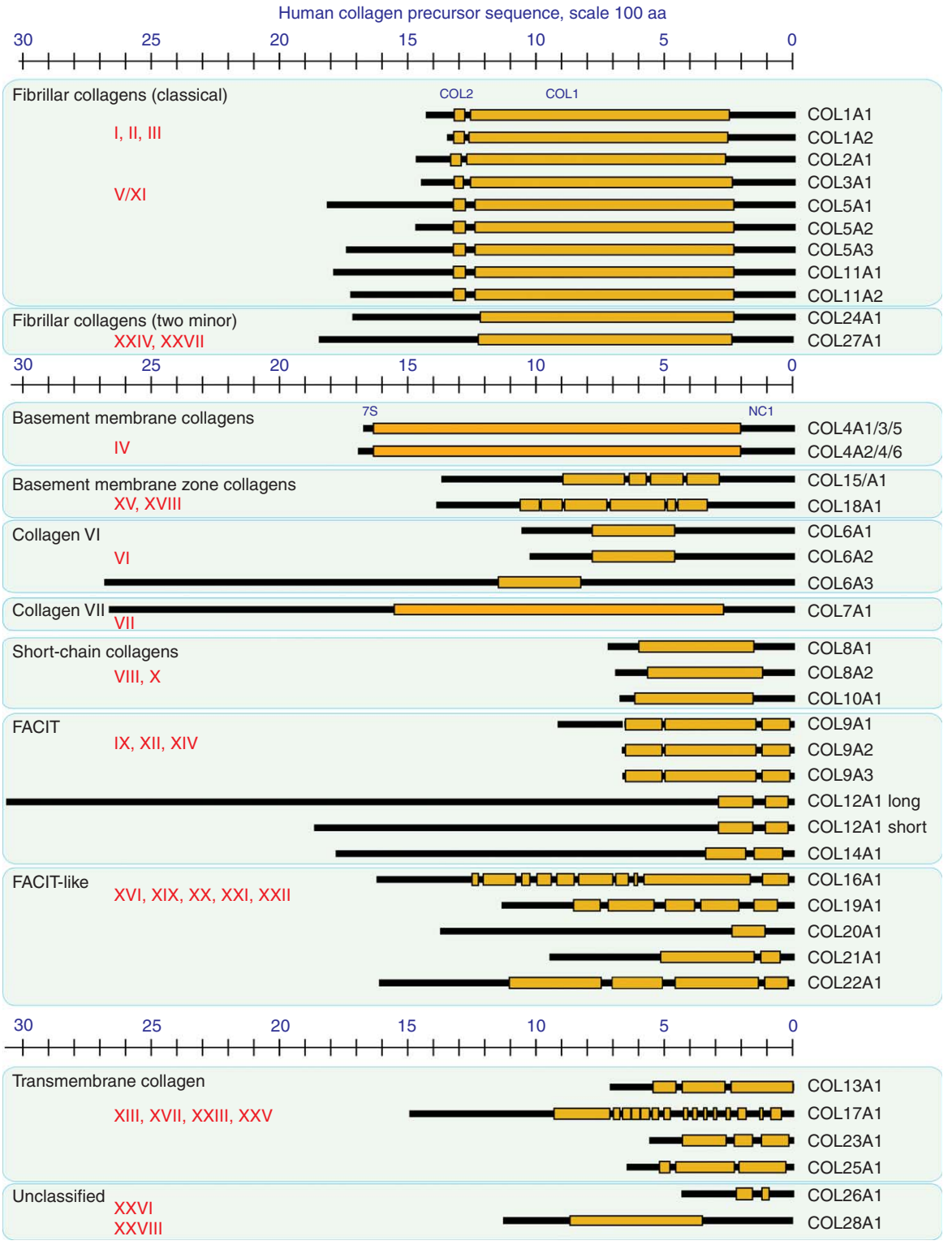


Figure 2 Schematic illustration of the varying lengths of the collagenous domains in human collagen types. The collagenous domain is indicated by a yellow-colored box. The noncollagenous domains are indicated as black bars. Some collagen types are processed before integration in their final location within the extracellular matrix.

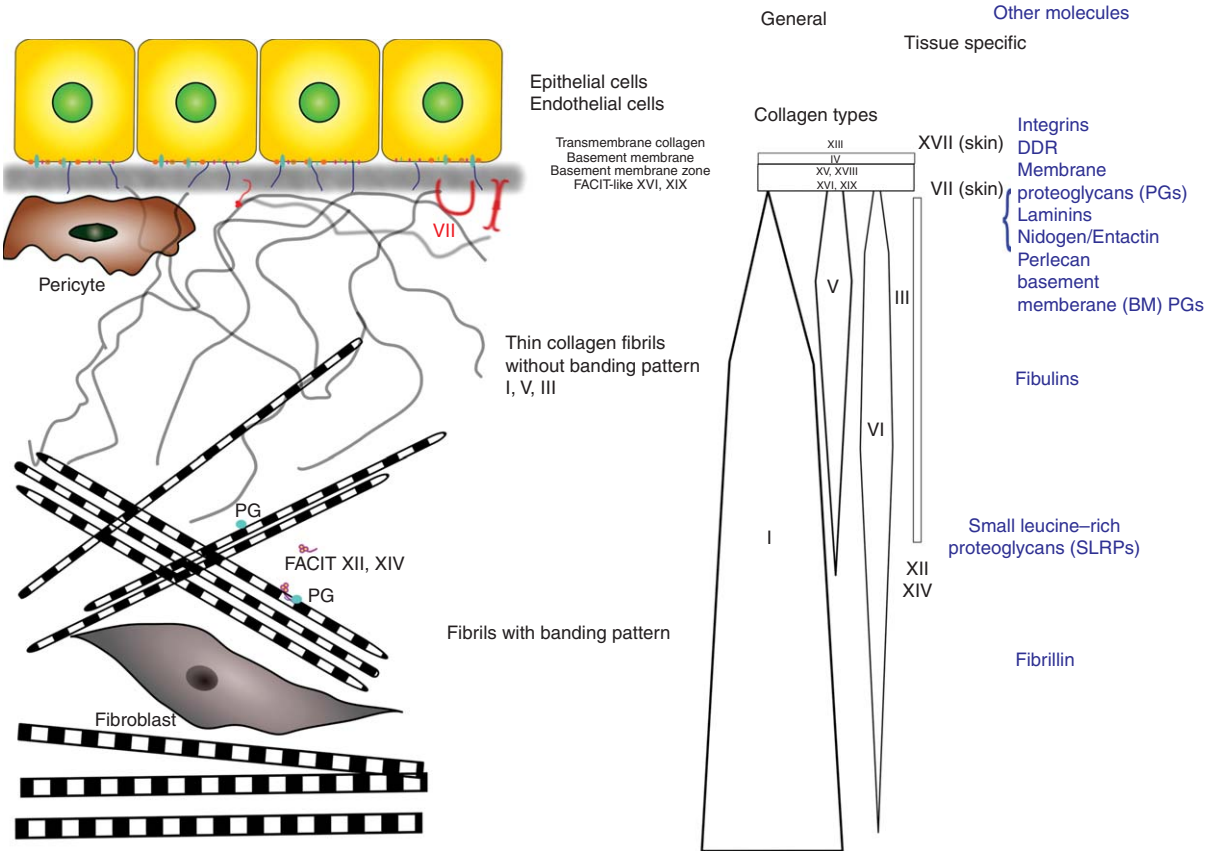


Figure 3 Schematic illustration of the distribution of different types of collagen in tissues. On the epithelial cell surface, transmembrane collagen, type XIII collagen, is expressed as a membrane-bound molecule. At the vicinity of the epithelial cells or endothelial cells, type IV collagen is a major component of basement membrane. Next, basement zone collagens, such as types XV and XVIII, localize in the basement membrane and vicinity. Type V and VI collagens also exist close to the basement membrane. Type VI collagen forms 100 nm periodic beaded microfibrils and may also exist as different supramolecular structures. Type III collagen is distributed especially in skin and artery extracellular matrix (ECM). Type I collagen is the most abundant type of collagen and is found at a little distance from the epithelial cells. Type V collagen is a minor component of collagen fibrils and one of its functions is to regulate their diameter. FACIT collagens, type XII and XIV collagens, are found in tissues containing a high content of type I collagen. The interactions of type XII and XIV collagens with collagen fibril might not be direct but rather mediated by proteoglycans. Collagen fibrils form 65 nm period banding fibrils. Some types of collagen are tissue specific, for example, type VII collagen is found as the major component of anchoring fibrils of epidermal–dermal junctions. Type XVII collagen is the membrane-bound collagen of keratinocytes of skin. The suggested ratio of the relative content of various types of collagen is schematically indicated by the boxes on the right side. Note that the size of the cells and ECM components are not proportional in size.

enzymes for collagen have been characterized. The N-terminal propeptide of the type I procollagen molecule is cleaved by an enzyme called pN-protype I collagen proteinase (EC 3.4.24.14; ADAMTS-2 (ADAMTS: a disintegrin and metalloproteinase with thrombospondin (TSP) motifs)) (see Section 5.16.3.7). ADAMTS-3 and -14 also may cleave the N-propeptide of the pro-α1(I) and the pro-α2(I) chains.[22,23] In tissues, the TIMP (tissue inhibitor of metalloproteinase) family is the major inhibitor of metalloproteinases. Among the four TIMPs, the procollagen N-proteinase ADAMTS-2 activity can only be inhibited by TIMP-3 but not TIMP-1, -2, or -4.[24] At the C-terminal end, pC-proteinase (EC 3.4.23.5; bone morphogenetic protein-1 (BMP-1); tolloid) cleaves the C-propeptide of the α1(I) and the α2(I) chains.[25] Mammalian tolloid (mTLD), mammalian tolloid-like 1 (mTLL-1), and mammalian tolloid-like 2 (mTLL-2) are the family of BMP-1-like proteinases.[25]

mTLD is an alternative splice variant of BMP-1. The BMP-1/mTLD double null mouse still has procollagen C-proteinase activity derived from mTLL-1,[26] indicating overlapping enzymatic activities of the different

family members. These enzymes are also known as the astacin family of zinc endopeptidases,[27] named after the freshwater crayfish *Astacus astacus*.[28] The concentration and diffusion gradient of these enzymes may regulate the formation and the organization of the extracellular matrix.[25,29] BMP-1-like proteinases also activate the transforming growth factor β superfamily of proteins.

The majority of the proline residues in the Yaa position in the -Gly-Xaa-Yaa-repeated sequences of collagens are posttranslationally modified to 4(*R*)-hydroxyproline by prolyl 4-hydroxylase (P4H) (see Section 5.16.3.1). One proline in the Xaa position (Pro986, counted from the first residue of the major triple helical domain) of the α1(I) chain is posttranslationally modified to 3(*S*)-hydroxyproline (3Hyp) by prolyl 3-hydroxylase (P3H) (see Section 5.16.3.2). A single proline residue of the α2(I) chain also has one 3Hyp but the site has not yet been identified. Approximately, one-third of all the lysine residues in the Yaa position are posttranslationally modified to 5-hydroxylysine by lysyl hydroxylase (LH) (see Section 5.16.3.3). At least one lysine residue in the major collagenous domain of the α1(I) chain, K87 in the sequence GMKGHR, is posttranslationally modified to a disaccharide with a unique connection, 2-*O*-α-D-glycopyranosyl-*O*-D-galactopyranosyl hydroxylysine.[30] Normal type I collagen molecules may also have another glycosylated hydroxylysine (Hyl) at K930 in the sequence GIKGHR in the α1(I) chain or at K933 in the sequence GFKGHN in the α2(I) chain in the triple helical domain. Lys87 is the position of 0.4D, and K930 is the position of 4D (see Section 5.16.2.4 about D periodicity). K87 may form an intermolecular cross-link with the C-telopeptide of the adjacent molecule, and K930 of the α1(I) chain and the K933 of the α1(I) chain may interact with the N-telopeptide of the adjacent molecule in the fibril. The function of glycosylation in the triple helical domain is unclear, but it may regulate the alignment of the molecular surface and pack the molecules in a directional manner to maximize the strength in some tissues. It may also add flexibility to the fibrils in some tissues. In addition, the specific three-dimensional organization may raise the frequency of chemical reaction of intermolecular cross-linking that stabilizes the fibrils. The cross-links of collagen have been reviewed in detail.[31–37]

The *oim* mouse that does not synthesize the α2(I) chain[21] can be used as an animal model to study the function of type I collagen, especially the role of the α2(I) chain. This mouse produces homotrimeric type I collagen, $[\alpha 1(I)]_3$. Surprisingly, the defect of the α2(I) chain does not result in lethality, although there are many abnormalities in many different tissues.[38–40] This indicates the functional importance of the heterotrimeric type I collagen. However, the basic information about the formation of heterotrimeric type I collagen molecule is still needed. In particular, the order of the three chains required to form the triple helix in a heterotrimer is not known. The sequence of the α2(I) chain is similar to that of the α1(I) chain but has a significant difference in the triple helical domain and telopeptides. When the 2:1 heterotrimer forms a triple helix with one-residue stagger, there are three possibilities of the chain alignments. In the case of type I collagen, both the α1(I) and the α2(I) chains have the same length collagenous domain of 1014 residues (338 tripeptide repeats). If the three α chains align with a one-residue stagger with respect to each other, the location of the α2(I) chain might be one of the following arrangements from the N-terminus: '1-1-2', '1-2-1', or '2-1-1'. This staggered order affects the side-chain alignment on the molecular surface (**Figure 4**), which affects the intermolecular interactions. It is still unknown whether the α chain staggered order of each heterotrimeric collagen molecule is specifically regulated to only one set of the chain stagger order or not. The synthesis of short stagger specific collagen heterotrimeric peptides has proven difficult.[43] In order to analyze heterotrimeric type I collagen in comparison with other molecules, the correct staggered order needs to be identified.

Type I collagen can interact with various other molecules in the extracellular matrix and on the cell surface. Cell receptors of type I collagen have been identified and include the heterodimeric integrin family receptors, α1, α2, α10, and α11, especially $\alpha 2\beta 1$, the receptor tyrosine kinase discoidin domain receptor (DDR), and the heparan sulfate membrane proteoglycan syndecan family.[44,45]

Synthetic homotrimeric collagen model peptides interact with the MIDAS (metal ion-dependent association site) in the I domain of the integrin α2 subunit.[42] These type I collagen model peptides have a GFOGER sequence (where O = 4-hydroxyproline), which is located at amino acids 501–506 of the major triple helical domain. If this GFOGER-binding motif is also true *in vivo*, and the chain stagger order is strictly determined, the α chain order of the type I collagen molecule may be '1-1-2' or '2-1-1', and cannot be a '1-2-1' chain stagger order because the homotrimeric collagen peptides interact with the integrin I domain with two adjacent chains. On the other hand, the '1-2-1' staggered heterotrimeric collagen model was successfully used as a substrate for

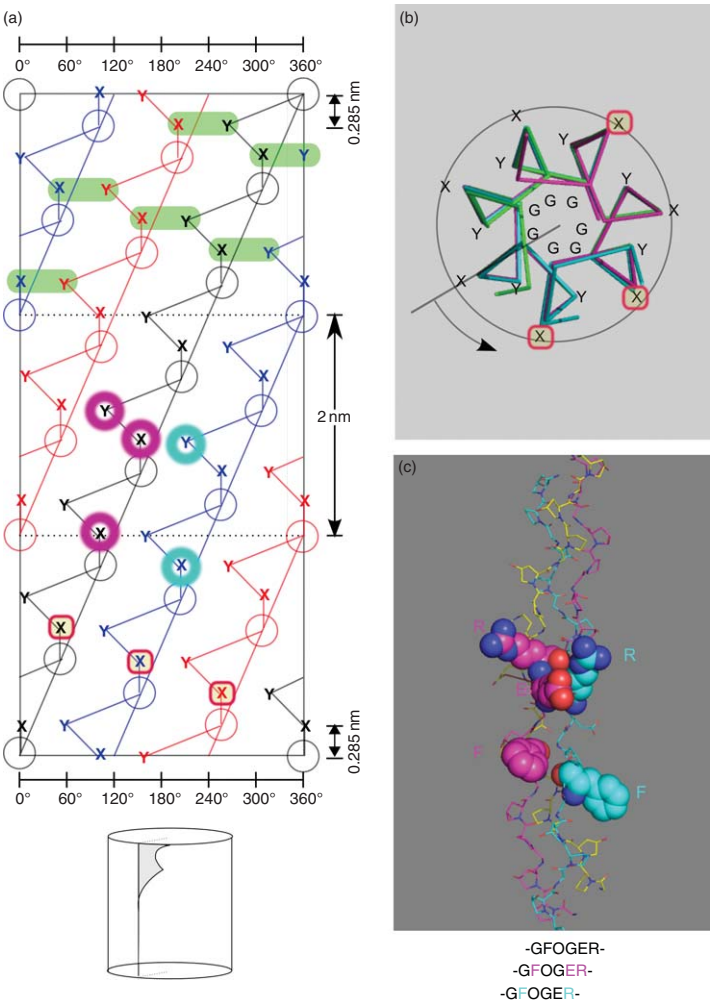


Figure 4 Schematic illustration of the three-dimensional side-chain distribution on the collagen triple helix. This model assumes that the triple helix is a cylinder. (a) The outer side of the collagen helix is developed as a two-dimensional rectangle. The base corresponds to the circumference and the height corresponds to the helix height. The three polypeptide chains are indicated by black-, blue-, and red-colored lines. Small circles indicate the Cα atom of Gly in three chains. In this figure, the helical symmetry is a the left-handed 7/2 helix, which is found in Pro/Hyp-rich collagen model peptide crystals. In regions with less Pro/Hyp, the helix in crystal structures is more loosely packed. Every 0.28–0.29 nm, two side chains from the two α chains shown in light green face the outer surface of the triple helix to create potential intermolecular interaction sites. (b) The top view of the Cα connections of the triple helix made from the $(Pro\text{-}Pro\text{-}Gly)_9$ structure (PDB accession number 2CUO).[41] Each chain forms a left-handed polyproline II helix, and the three chains form a loose right-handed superhelix. (c) The crystal structure of the collagen model peptide side chains that interact with the integrin I domain (PDB accession number 1DZI). Parts (b) and (c) were made by PyMOL' with PDB accession number 1DZI.[42] Note that the side chains that are close in the primary sequence are apart in azimuthal angles and the five residues from the two chains shown in cyan and magenta are interacting with the ligand. The colored circles in (a) show the side-chain distribution of these five residues.

the matrix metalloproteinases (MMPs) in enzyme assays.[46,47] Typically, a series of short collagen-like fragments are used to find the binding sites within the collagen molecule.[48] In globular proteins, hydrophobic residues are usually folded into the inside of the protein but in collagen, all of the side chains in the Xaa and the Yaa positions are exposed to the molecular surface. The total content of hydrophobic residues in collagen is much less than what is typically found in globular proteins. For example, the content of F(Phe) in the triple helical domains of the α1(I) and the α2(I) are 12 and 11 in 1014 residues, respectively. These hydrophobic residues might be important to form specific binding sites on the surface of collagen. It should be noted that if

'1-2-1' is the only natural type I collagen chain stagger, some of the binding sites formed by two adjacent different α chains cannot be detected using homotrimeric model peptides. Also, in tissues, virtually all type I collagen molecules are packed as fibrils. On the fibril surface, for example, multiple side chains from two adjacent molecules could be arranged as a binding site. Unfortunately, the molecular packing of type I collagen in fibrils is not understood in detail. This is one of the difficulties in studying collagen–protein interactions.

Mouse models have been developed to analyze the biological function and molecular mechanisms of collagen proteins. In **Table 1**, the phenotypes of the mouse models of collagen genes and collagen biosynthesis-related proteins are briefly summarized.

Mutations in the type I collagen genes are known to cause osteogenesis imperfecta (OI). Most forms of OI are autosomal dominant with mutations in one of the two genes that code for type I collagen α chains, COL1A1 and COL1A2. More than 300 mutations in the α1(I) and the α2(I) chains have been reported.[90,91] A single point mutation of Gly is the most common event causing the disease. OI has been classified into groups I–VIII based on differences in phenotype and disease severity ranging from mild to lethal (**Table 2**). The mechanism of how mutations in the collagen genes ultimately lead to OI is still not well understood. Mineralization of the bone collagen matrix, the molecular packing of bone collagen fibrils, and the cross-linking of bone collagens are not well-characterized processes, but are likely to be involved in the development of the disease. Most of the mutations that cause OI occur in the major triple helical domain; however, some have also been found in the C-propeptide.[93,94] Collagen molecules isolated from cultured fibroblasts from OI patients have a little lower transition temperature than the control type I collagen molecules, indicating that they are slightly less stable than normal type I collagen. Cells from OI patients were also shown to synthesize overmodified type I collagen molecules in some cases as indicated by their electrophoretic mobility and amino acid analysis.[95,96] More recently, it was found that mutations in prolyl 3-hydroxylase 1 (P3H1) and cartilage-associated protein (CRTAP) cause a recessive form of OI that ranges in severity from severe to lethal.[88,97] Type I collagen secreted from some of these recessive OI patient fibroblasts was shown to have a higher transition temperature,[97] suggesting that the stability of the type I collagen molecule itself is most likely not the cause of the bone fractures in OI patients. Overhydroxylation and overglycosylation during biosynthesis are reasonable reactions of the cell to make a more stable type I collagen molecule, but for bone tissues, this might cause strain in molecular packing causing improper assembly of collagen fibrils within the bone matrix. If this turns out to be true, it may be appropriate to try to treat OI patients by enhancing the quality control mechanisms during type I collagen biosynthesis, folding, and secretion, which would limit the amount of overmodification that could occur.

5.16.2.2.2 Type II collagen

Type II collagen is a major component of hyaline cartilage. Miller and Matukas[98] identified the new collagen polypeptide chain from chick cartilage in 1969. Type II collagen is also expressed in noncartilaginous tissues in development. The mRNA of col2a1 is widely and transiently expressed in developmental noncartilage tissues such as notochord, eye, heart, and brain.[99,100] The molecular structure is similar to type I collagen in general, but the chain composition of the type II collagen molecule is a homotrimer with the chain composition of $[\alpha 1(II)]_3$. The major triple helical domain with 1014 residues of the repeated -Gly-Xaa-Yaa- sequence is flanked by N- and C-propeptides as in type I collagen. ADAMTS-2 and ADAMTS-3 cleave the N-propeptide of type II collagen.[23] In hyaline cartilage, type II collagen molecules assemble with type XI and IX collagens as well as with chondroitin and the keratan sulfate proteoglycan aggrecan. Eyre *et al.*[32] estimated the amount of the different types of collagen in the cartilage matrix as type II (75% fetal, >90% adult), type III (>10% in adult human articular), type IX (10% fetal, 1% adult), type X (hypertrophic cartilage only), type XI (10% fetal, 3% adult), type VI (chondron basket, microfilaments <1%), type XII/XIV, and type XIII.

From amino acid analysis and the primary structure, almost all the Yaa position lysine residues of hyaline cartilage type II collagen are hydroxylated. In addition, nearly half of the Hyl was glycosylated as determined by Fmoc-labeled amino acid analysis.[101] While type I collagen has only one or two glycosylated Hyl per α chain, type II collagen has about 10 glycosylated Hyl per α chain. Recombinant type II collagen, with low levels of Hyl and glycosylated Hyl, forms thicker fibrils than the recombinant type II collagen with high Hyl and glycosylated Hyl levels.[102] One of the biological functions of glycosylation of the collagen helix may be in the regulation of the lateral growth of collagen fibrils.

Table 1 Examples of mouse models of collagen and related protein genes

Gene	*Mouse phenotype*	*References*
col1a1	Mov13 mice carry a provirus that prevents transcription initiation of the α1(I) collagen gene. Mutant mice homozygous for the null mutation produce no type I collagen and die at E12–14	49–51
col1a2	Mice without the α2(I) chain form type I collagen homotrimer [α1(I)]$_3$ molecules. Not lethal	21,40, 52–56
col2a1	Col2a1$^{-/-}$ mice die either just before or shortly after birth. Their cartilage consists of highly disorganized chondrocytes with a complete lack of extracellular fibrils by electron microscopy. There is no endochondrial bone or epiphyseal growth plate in long bones; however, many skeletal structures such as the cranium and ribs are normally developed and mineralized	57
col3a1	About 10% of the homozygous mutants survive to adult but have shorter lifespans. Major cause of death was rupture of the blood vessels, similar to EDS type IV	58
col4a1	Mice with deletion of exon 40 (17 aa) in triple helical domain have high ratio of cerebral hemorrhage; mutant mice are smaller and have multiple pleiotropic phenotypes including ocular abnormalities, mild renal abnormalities, and reduced fertility. Homozygous mutant mice are not viable after mid-embryogenesis, and 50% of heterozygous mice die within a day of birth	59–61
col4a3	Col4a3 null mice display a renal phenotype strikingly similar to Alport syndrome: decreased glomerular filtration (leading to uremia), compromised glomerular integrity (leading to proteinuria), structural changes in glomerular BL, and glomerulonephritis	62, 63
col4a5	A mouse model of X-linked Alport syndrome (XLAS) was generated by targeting a human nonsense mutation, G5X, to the mouse Col4a5 gene. Hemizygous mutant male mice are null and heterozygous carrier female mice are mosaic for α5(IV) chain expression. Mutant male mice and carrier female mice are viable and fertile. Mutant male mice die spontaneously at 6–34 weeks of age, and carrier female mice die at 8–45 weeks of age, manifesting proteinuria, azotemia, and progressive and manifold histologic abnormalities of the kidney glomerulus and tubulointerstitium. This mouse model recapitulates findings of human XLAS showing ultrastructural abnormalities of the glomerular basement membrane	64
col4a6	col4a6$^{-/Y}$ mice are healthy, fertile, and display no known defects[a]	
col4a1 and col4a2	Col4a1/a2 null mice are embryonic lethal at E10.5-11.5; basement membrane is not formed	59
col4a3 & col4a4	Lack of col4a3 and col4a4 in mice results in an earlier onset Alport syndrome-like phenotype than in mice lacking α3 only	65
col5a1	The col5a1$^{-/-}$ animals are embryonic lethal at approximately E10. Mice demonstrate a virtual lack of collagen fibril formation. Col5a1$^{+/-}$ mice are viable but have poorly organized abnormal collagen fibrils	66, 67
col5a2	Homozygous mice with short N-terminal noncollagenous domain (exon 6 deletion) poorly survive possibly because of severe spinal deformation, kypholordosis. Collagen fibrils are disorganized	68
col6a1	Col6a1$^{-/-}$ mice show histological features of myopathy such as fiber necrosis, phagocytosis, and a pronounced variation in muscle fiber diameter. Muscles also show signs of stimulated regeneration of fibers. Necrotic fibers are frequent in the diaphragm. Milder alterations are detected in heterozygous mice, indicating haploinsufficiency	69, 70

col7a1	Conditional inactivation of Col7a1 expression results in col7a1 hypomorphic animals expressing 10% of normal col7a1 levels. Homozygous mice appear normal at birth, but develop blisters on the paws by 24–48 h after birth	71
col8a1, col8a2	No major structural defects in most organs, but anterior segment abnormalities are found in the eye. Type VIII collagen is required for normal anterior eye development, particularly the formation of a corneal stroma with the appropriate number of fibroblastic cell layers and Descemet's membrane of appropriate thickness	72
col10a1	Type X collagen-deficient mice have phenotypes resembling SMCD, such as abnormal trabecular bone architecture. Mice with inactivated Col10a1 manifest variable phenotypes reflecting skeletohematopoietic defects. A subset of the knockout mice (~11%) died 3 weeks after birth, and others continued to exhibit defects with age including erythrocyte predominance in marrow, reduced spleen and thymus size, altered B and T lymphocyte development, aberrant endochondral ossification, and dwarfism	73, 74
col11a1	Autosomal recessive mutations in the col11a1 gene (*cho* mice) result in newborn limb bones that are wider at the metaphyses and only about half the normal length. Type XI is essential for normal formation of cartilage collagen fibrils and the cohesive properties of cartilage	
col11a2	Mice with a targeted disruption of Col11a2 show hearing loss. Electron microscopy of the tectorial membrane reveals loss of organization of the collagen fibrils. The findings reveal a unique ultrastructural malformation of the inner ear architecture associated with nonsyndromic hearing loss	75
col13a1	Overexpressing type XIII collagen enhances bone formation. Lack of type XIII collagen causes myopathy	76, 77
col14a1	col14a1$^{-/-}$ mice show slightly different fiber organization at postnatal day 4. No significant differences are observed in the adult mice	78
col15a1	Col15a1 null mice develop and reproduce normally. However, null mice show progressive histologic changes characteristic of muscular disease after 3 months of age, and are more vulnerable to exercise-induced muscle injury. The development of the vasculature appeared normal; however, null mice had collapsed capillaries and endothelial cell degeneration in heart and skeletal muscle by EM analysis	79
col18a1	Col18a1$^{-/-}$ mice have abnormalities in the iris and ciliary body. Mice exhibit dilation of the brain ventricles with 20% of mice developing hydrocephalus. There is significant broadening of the epithelial basement membrane of the choroid plexuses. Cells from Col18a1 null mice contain more vacuoles, suggestive of alterations in CSF production. Markedly broadened basement membranes are found in the atrioventricular valves of the heart and in the kidney tubules, whereas the glomerular mesangial matrix of kidneys is expanded. Also serum creatinine levels are elevated, indicating alterations in kidney filtration capacity	80–82
P4Ha(I)	P4ha1$^{-/-}$ mice are embryonic lethal at E10.5. Basement membrane and collagen fibrils are abnormal	83
LH1	PLOD1$^{-/-}$ mice show frequent aorta rupture; decrease of LH activity in skin	84
LH3	LH3 null mice are embryonic lethal at E9.5–14.5, show fragmentation of basement membrane and possibly some type IV collagen abnormalities	85–87
CRTAP	Loss of Crtap in mice causes an osteochondrodysplasia characterized by severe osteoporosis and decreased osteoid production	88
P3H1	P3h1$^{-/-}$ mice do not die prematurely but have abnormal collagen fibril morphology in tendon, decreased bone density, and abnormalities in skin and developing limbs[b]	
HSP47	lethal at E11.5; basement membrane is not formed well by electron microscopy	89

[a] Professor Yoshifumi Ninomiya, Okayama University, Japan, personal communication.
[b] J. A. Vranka *et al.*, unpublished results.

Table 2 Classification of osteogenesis imperfecta

	Symptoms	*Cause of gene mutations*
I	The mildest and most common form; about 50% of total OI, often blue sclerae; dentinogenesis imperfecta is often absent	Autosomal dominant mutations in COL1A1 or COL1A2
II	Lethal perinatal, OI with radiographically crumpled femora; the most severe type; blue sclerae	Autosomal dominant mutations in COL1A1 or COL1A2
III	Progressively deforming OI with normal sclerae; the most severe type among children who survive the neonatal period. The degree of bone fragility and the fracture rate vary widely	The majority result from dominant mutations in COL1A1 or COL1A2
IV	Mild-to-moderate bone deformity and variable short stature; hearing loss occurs in some families; white or blue sclerae; the child might not fracture until he or she is walking; dentinogenesis imperfecta may be present or absent	Autosomal dominant mutations in COL1A1 or COL1A2
V	Phenotypically indistinguishable from type IV OI. Distinctive histology of irregular arrangement or mesh-like appearance of lamellae. Also have triad of hypertrophic callus formation, dense metaphyseal bands, and ossification of the interosseus membranes of the forearm. Normal type I collagen; no mutations detected. People with OI type IV have moderate-to-severe growth retardation, which is one factor that distinguishes them clinically from people with type I OI	Dominantly inherited, gene unknown
VI	Extremely rare. Phenotypically indistinguishable from type IV OI Diagnosed on the basis of unique histological features. Elevated alkaline phosphatase activity. 'Fish-scale' appearance of bone under the microscope	Probably recessive, the gene has not yet been identified
VII	Rhizomelia and coxa vara in all affected patients; decreased cortical width and trabecular number, increased bone turnover, and preservation of the birefringent pattern of lamellar bone	Recessive inheritance of a mutation in the cartilage-associated protein (CRTAP)
VIII	The phenotype of the probands overlapped Sillence lethal type II/severe type III osteogenesis imperfecta; severe osteoporosis, shortened long bones, and a soft skull with wide open fontanel; white sclerae, severe growth deficiency, extreme skeletal undermineralization, and bulbous metaphyses	Prolyl 3-hydroxylase 1 mutation

Types I–IV are Sillence classification by clinical symptom.[92] Types V–VIII were added. Types VII and VIII are diagnosed by the mutated genes not from the clinical symptom.

In the type II collagen null mouse, there was no endochondral bone or epiphyseal growth plate in the long bones. However, surprisingly, many skeletal structures such as the cranium and ribs were normally developed and mineralized.[57] The cartilaginous elements have more severe phenotypes in the type II null mice. They do not have growth plates in the long bones.[103] This study also shows that cartilage-specific proteins such as type IX and XI collagens, aggrecan, fibromodulin, COMP (cartilage oligomeric matrix protein), CMP (cartilage matrix protein), and chondroadherin are expressed even in the absence of type II collagen. Polypeptides from type XI collagen, in which the gene of the $\alpha 3$(XI) chain is identical to that of the $\alpha 1$(II) chain (Col2a1), are pepsin labile in the type II collagen null mice, indicating that the $\alpha 3$(XI) chain is required to form a type XI collagen molecule.[103]

More than 90 mutations of COL2A1 gene have been reported and cause a variety of disorders including spondyloepiphyseal dysplasia (SED), Stickler syndrome, Kniest dysplasia, rhegmatogenous retinal detachment, osteochondrodysplasia, hypochondrogenesis, achondrogenesis 2, and osteoarthritis.[104,105] These autosomal dominant disorders range in severity from achondrogenesis type II, which is lethal at or before birth, to late-onset SED, the major feature of which is precocious osteoarthritis. Disorders of intermediate severity include SED congenita, SED Strudwick, and Kniest dysplasia. Stickler syndrome is an autosomal dominant disorder with characteristic ophthalmological and orofacial features, deafness, and arthritis.[106] The mutations are thought to act through a dominant-negative mechanism to reduce the number of overall collagen fibrils in the cartilage matrix.

5.16.2.2.3 Type III collagen

Type III collagen belongs to the fibrillar collagen group. Its molecular structure is similar to that of type I collagen. The molecule is a homotrimer with the chain composition of $[\alpha 1(\text{III})]_3$. The major tissue localization of type III collagen is in dermis and aorta. In tendon and bone, the ratio of type III collagen to type I collagen is very low. Recent studies have shown that type III collagen represents almost 10% of the total collagen components in cartilage.[32] Type II and III collagens exist in single fibrils.[107] Some of the type III collagen molecules are cross-linked with type II collagen.[32] Type III procollagen (pN-type III collagen) can be purified from tissue for biophysical analysis. Type III collagen contains unique covalent cross-links at both the N- and C-terminal ends adjacent to the triple helical domain and is a good source for refolding experiments.[105] The C-terminal end of the major collagenous domain has the sequence -GPPGAPGPCCG-. The six cysteine residues from the three chains form three pairs of disulfide bonds (see Section 5.16.4.1.1). Also, the major triple helical domain, 1029 residues in humans, is a little longer than the same domain in type I and II collagen molecules containing 1014 residues. Type III collagen is synthesized as procollagen III in a similar manner to type I collagen. However, while most of the N-propeptide of the procollagen type I molecule is cleaved and difficult to extract from tissue, the pN-type III collagen is processed partially,[109,110] and is known to be deposited in the skin at significant levels.[111] Immunoelectron histochemistry using a monoclonal antibody made against the triple helical domain suggested that type III collagen is localized in skin collagen fibrils regardless of fibril diameter.[112,113] The type III collagen null mouse[58] was shown to have ruptures in major blood vessels, indicating that this molecule is essential for the structural integrity of the circulatory system. However, the molecular function of type III collagen and pN-type III collagen is still completely unknown. The tyrosine residues in a pepsin-susceptible domain of type III collagen are sulfated by tyrosylprotein sulfotransferase (TPST, EC 2.8.2.20).[114] This modification was shown to occur in the trans-Golgi during secretion of the procollagen molecule.[115] The biological role of sulfation might be related to growth regulation of the collagen fibrils although details of this mechanism are unknown.[116]

A series of short peptide fragments of type III collagen have been used to find binding sites with other molecules. DDR-2 has been shown to interact with a type III collagen model peptide with the minimal sequence of GVMGFO (where O is 4-hydroxyproline).[117] It is interesting to note that this same sequence is also involved with the collagen interaction of SPARC.[118]

More than 100 mutations of the COL3A1 gene can cause Ehlers–Danlos syndrome (EDS) type IV, an autosomal dominant disorder characterized by joint and dermal manifestations similar to the other forms of the syndrome, but in addition these individuals are prone to spontaneous ruptures of bowel and large arteries.

5.16.2.3 Fibrillar Collagens (Type V and XI)

5.16.2.3.1 Type V collagen

Type V and XI collagens are highly homologous and are often classified as type V/XI collagen. In fact, in some cases, the type V and XI collagen chains form a heterotypic molecule. Type V collagen is a fibrillar collagen molecule, which is a minor component of collagen fibrils with type I collagen as the major component. Garrone and coworkers have hypothesized that collagens from primitive marine animals such as sponge, sea pen, and jellyfish could represent ancestral forms of the vertebrate types V/XI.[119,120] There are three genes in type V collagen, COL5A1–COL5A3. At least two chain compositions of type V collagen molecules have been isolated and these are $[\alpha1(V)]_2\alpha2(V)$ and $\alpha1(V)\alpha2(V)\alpha3(V)$ in normal tissues.[121–124] The subtype with the chain composition of $[\alpha1(V)]_2\alpha2(V)$ is ubiquitously distributed, but the subtype of $\alpha1(V)\alpha2(V)\alpha3(V)$ is found in placenta,[123,125,126] skin,[127] synovial membrane,[128] and uterus.[129,130] In cornea, no $\alpha3(V)$ chains have been detected, suggesting that the $\alpha3(V)$ chain-containing subtype might exist only in the tissues associated with the vascular system. Immunohistochemical studies have shown that type V collagen localizes both near the basement membrane and in the interstitial connective tissue.[131] Biochemical and immunohistochemical data have shown that near the basement membrane, the subtype of $\alpha1(V)\alpha2(V)\alpha3(V)$ might be more prominent, whereas near the type I collagen-rich area, the subtype of $[\alpha1(V)]_2\alpha2(V)$ might be more populated. The $[\alpha1(V)]_3$ homotrimer is also expressed in hamster cell culture media[132,133] but further details of the function of subtypes are still unknown. The tissue form of type V collagen molecules may retain a larger N-terminal NC domain compared to that of type I collagen.[134] The N-terminal propeptide of the $\alpha(V)$ chain consists of the signal peptide, a proline- and arginine-rich protein (PARP) domain, a globular 'variable region', and a small collagenous subdomain (COL2).

Type V collagen has interesting biological effects on cells that cannot be seen with type I collagen in culture. Type V collagen molecules act as antiadhesive agents against cultured cells.[135] Vascular endothelial cells in particular tend to detach from dishes coated with type V collagen molecules,[136] suggesting that type V collagen prevents cell adhesion to the culture dish. In contrast to these conclusions, a recent study showed that the strong adhesion of type V collagen to cells is the cause of their apparent detachment from the dish.[137] The fibrils made of pepsinized type V collagen interact strongly with the cells and form cell clumps with type V collagen fibrils.[138] When kidney glomerular endothelial cells are cultured on type V collagen fibrils, the cells dynamically move with extending long filopodial protrusions and wide lamellipodia compared with cultures on type I collagen fibrils.[139] The extracellular signaling is mediated through $\beta1$ integrin-activated phosphorylation of paxillin and of focal adhesion kinase. Also the attached type V collagen activates FAKpY397 via neuroglycan 2.[139] The triple helical domain of type V collagen interacts with various molecules with much higher affinity than type I collagen. TSP,[140] heparin,[141,142] and heparan sulfate[143] interact with type V collagen much stronger than with type I collagen.

Type V collagen has a function to limit the lateral growth of collagen fibrils.[131,144] In the cornea, the ratio of type V collagen to type I collagen is higher than in other tissues, and conversely the fibril diameter of cornea is smaller than in other tissues. One possible explanation for the limited growth of fibrils would be that tissue form of type V collagen has the bulky NC domain, but the collagenous domain itself retains this function. Pepsin-treated type V collagen without a bulky globular domain limits the lateral growth of reconstituted pepsin-treated type I collagen *in vitro*.[145] Pepsin-treated type V collagen forms thin, highly homogeneous fibrils without merging and branching as compared with type I collagen fibrils.[146]

The presence of type V collagen in the vicinity of basement membranes and in collagen fibers suggests that it can act as a linker and can also contribute to fibril structure.[131,147] Fibril thickness, length, and flexibility can affect cell behaviors such as differentiation, motility, and growth. Compared with type I collagen, the triple helical domain of a type V collagen molecule has much more glycosylated Hyl. Almost all the Yaa position matrix-assisted laser desorption ionization mass spectrometry residues in a type V collagen molecule are modified to glucosylgalactosyl hydroxylysine as shown by MALDI-MS analysis.[148,149]

Processing of a newly synthesized type V procollagen molecule is also different from that of type I collagen. BMP-1, which is the C-propeptidase of type I, II, and III collagens, cleaves the N-terminal propeptide of the pro-$\alpha1(V)$ chain, and a furin-like proteinase cleaves the C-terminal propeptide.[150,151] The tyrosine residue in the N-terminal NC domain of type V collagen is sulfated.[152,153] This modification might be related to

protein–protein interactions. The processing of the N-terminal domain of type V procollagen to the appropriate tissue sizes consists of more than one step involving multiple enzyme activities.[150,154–156] Both the complicated processing and the posttranslational modifications of type V collagen suggest that this molecule could play key roles in the regulation of extracellular matrix organization in development, differentiation, and tissue repair.

Mutations of COL5A1 and COL5A2 genes cause EDS. Mice that are heterozygous for a targeted inactivating mutation in col5a1 causing a 50% reduction in col5a1 mRNA levels showed EDS-like histology with decreased aortic stiffness and tensile strength and hyperextensible skin with decreased tensile strength in both normal and wounded skin.[66] Col5a1 null mice died at E10.[67] This is earlier than for the type I and III collagen null mice, suggesting that some functions of type V collagen are essential for early stages in development. This idea is supported by the evidence that the fibrillar collagen of primitive invertebrates is more similar to type V collagen, indicating the essential biological role of this collagen.

For more information on type V collagen, see Adachi *et al.*[131], Birk[144], Fessler and Fessler[147], and Fichard *et al.*[157]

5.16.2.3.2 Type XI collagen

Type XI collagen is a minor component of hyaline cartilage collagen fibrils. The chain composition of type XI collagen is $\alpha1(XI)\alpha2(XI)\alpha3(XI)$. The gene of the $\alpha3(XI)$ chain is the splicing variant of collagen IIB exon 2.[121,158–160] Type XI collagen is highly homologous to type V collagen. In some sources, such as rhabdomyosarcoma[161] and bovine vitreous tissues,[162] collagen molecules with a chain composition of $[\alpha1(XI)]_2\alpha2(V)$ are synthesized. The $\alpha1(XI)$ and the $\alpha2(XI)$ chains are more similar to the $\alpha1(V)$ chain.

Since the gene of the $\alpha3(XI)$ chain is COL2A1, all type II collagen gene mutations in COL2A1 (except exon 2 mutations) should also affect the type XI collagen molecule structure. Mutations of COL11A1 cause Stickler dysplasia type III and Marshall syndrome (OMIM #154780),[163] and mutations of COL11A2 cause nonocular Stickler syndrome type II.[164] Type XI collagen is highly expressed in the nucleus pulposus of the intervertebral disc. A single-nucleotide polymorphism of COL11A1 has been shown to cause an increase in the genetic risk of developing a lumbar disc herniation in the Japanese population.[165]

In the col2a1 (or the col11a3) null mouse, unstable type XI collagen is detected in cartilage in the absence of the $\alpha3(XI)$ (or the $\alpha1(II)$) chain.[103] The chain composition of the formed type XI collagen without the $\alpha3(XI)$ or the $\alpha1(II)$ chain is not clear, but the molecule is more susceptible to pepsin digestion suggesting that the correct chain combination is important to have a functional type XI collagen molecule.

5.16.2.4 Collagen Fibril

In tissues, banded fibrils with about 65 nm periodicity are observed by transmission electron microscopy (TEM). Purified type I collagen molecules isolated from tissues also from collagen fibrils with about 65 nm periodicity without any other proteins. Due to dehydration of the sample preparation for TEM, up to 25–30% shrinkage can occur in the collagen fibril estimated from the comparison with the X-ray analysis.[166,167] A similar 65 nm periodic banding pattern is observable not only by TEM but also by scanning electron microscopy and atomic force microscopy. The details of the mechanism of molecular packing in collagen fibrils are still unclear. At least one-dimensionally, the molecules are staggered along the fibril at about 65 nm in distance, the so-called D periodicity, or 2–4 times the D. The D is derived from the name of the clearest band observed on collagen fibrils by TEM.[168] When the primary structures of the α chains of type I collagen are aligned, hydrophobic periodicity can be matched with a 234-residue stagger. One major triple helical domain of type I collagen can be interpreted as 4.4D, that is, 1014 (residues) = 234 × 4.33. From the analysis of cross-links within the type I collagen fibril, 0D and 4D staggered cross-links are found. Synthetic peptides of the C-telopeptide of type I collagen have been found to inhibit the self-assembly of type I collagen.[169] The synthetic C-telopeptide interacts near the collagenase cleavage site Gly775-I1e776, which corresponds to 3.4D stagger.[169] The three-dimensional alignment of the molecules within a fibril affects the interaction between collagen fibrils and the other molecules. The other types of fibrillar collagen do not have this periodic hydrophobic residue distribution. There are many models that attempt to explain the substructure of collagen fibrils. Collagen molecules are not a stiff rod, and it is reasonable to hypothesize that collagen molecules in a fibril are twisted as well as the

collagen fibrils themselves. Electron microscopy analyses have shown the helical subfibril organization in collagen fibrils.[170,171] Thus far, no models can explain the molecular packing of collagen molecules in a fibril. Molecular packing, especially on the surface of the fibril, is important for interactions with other extracellular matrix components, proteins in the extracellular space, and molecules on the cell surface. For more details on collagen fibrils, refer the recent reviews.[170,172–177]

5.16.2.5 Basement Membrane Collagen (Type IV)

Type IV collagen is a major component of the basement membrane, which is the specialized sheet-like extracellular matrix of multicellular tissues that exists around certain cell types (e.g., skeletal muscle cells, smooth muscle cells, heart muscle cells, and adipocytes). Basement membranes also exist between connective tissues and epithelial cells, endothelial cells, or Schwann cells. Type IV collagen also localizes at liver sinusoid. Type IV collagen has six α chains, α1(IV)–α6(IV), which are classified into two groups by their gene structure, 1, 3, and 5 are one set, and 2, 4, and 6 are another set. The chain composition of type IV collagens is estimated from immunohistochemical analysis using chain-specific monoclonal antibodies. From the staining patterns of the six antibodies in many different tissues, the current combinations of the type IV collagen α chains were determined to be $[\alpha1(IV)]_2\alpha2(IV)$, $\alpha3(IV)\alpha4(IV)\alpha5(IV)$, and $[\alpha5(IV)]_2\alpha6(IV)$.[178] Almost all basement membranes contain the subtype with the chain composition of $[\alpha1(IV)]_2\alpha2(IV)$. The $\alpha3(IV)\alpha4(IV)\alpha5(IV)$ subtype is localized in the basement membrane of kidney glomerulus, the alveoli of the lung,[179,180] and the neuromuscular synapse.[181] The $[\alpha5(IV)]_2\alpha6(IV)$ subtype is thought to localize to basement membranes of the epidermis, smooth muscle cells, mammary glands, and epithelium of the alimentary tract.[180,182,183]

The type IV collagen molecule is comprised of the N-terminal collagenous 7S domain, which has a sedimentation coefficient of 7S,[184,185] the central major collagenous domain with more than 20 interruptions of repeated -Gly-Xaa-Yaa- sequences, and the C-terminal NC1 domain. The major collagenous domain is longer than that of type I collagen. Additionally, 3-hydroxyproline and glycosylated Hyl contents are higher in type IV than in type I collagen. Since the collagenous domain of type IV collagen has more than 20 interruptions in the -Gly-Xaa-Yaa- repeats, the collagenous domain is susceptible to pepsin treatment, which is a convenient method commonly used to extract the collagenous domain from tissues.

Type IV collagen subtype $[\alpha1(IV)]_2\alpha2(IV)$ from bovine lens capsule forms a fine meshwork structure containing many branches with about 20 nm between the branching points.[186] There are three types of supramolecular assembly of type IV collagen molecules: dimer formation (with six polypeptide chains) at the C-terminal NC1 domain,[187,188] the association of four molecules (12 chains) in a 30 nm tetramer formation at the N-terminal 7S domain,[189] and lateral interaction at the collagenous domain (**Figure 5(a)**).[190–192]

Recently, the crystal structure of the NC1 domains of two molecules, in hexamer form, was analyzed by two different groups using either human placenta[193] or bovine lens capsule.[194] A unique nonreducible but labile cross-link between Met and Lys/Hyl was found (**Figure 5(b)**).[193,195,196]

The NC1 domains of the α1(IV), α2(IV), and α3(IV) chains are named arresten, canstatin, and tumstatin, respectively.[197] These NC1 fragments are reported to have several functions: suppression of tumor growth, inhibition of endothelial cell migration, or induction of apoptosis (reviewed in Mundel and Kalluri[197]). Under certain circumstances, secreted monomeric nontriple helical type IV collagen shows inhibition of collagenase activity similar to that of the TIMPs.[198] The NC1 domain of type IV collagen has sequence homology with TIMP, suggesting that the potential inhibitory activities might be activated by proteases.

Mutations of COL4A3, COL4A5, and COL4A6 genes cause Alport syndrome, a hereditary form of progressive renal disease. Goodpasture syndrome, an autoimmune disease affecting the lungs and the kidneys, has been shown to be caused by antibodies to the α3(IV) chain. Mice that do not express col4a1/col4a2 are lethal at embryonic stage E9.5.[59]

5.16.2.6 Basement Membrane Zone Collagens, Multiplexins (Types XV and XVIII)

Type XV and XVIII collagens are classified as basement membrane zone collagens or multiplexins (multiple triple helix domains and interruptions) by Olsen and coworkers.[199] From a biological aspect, basement

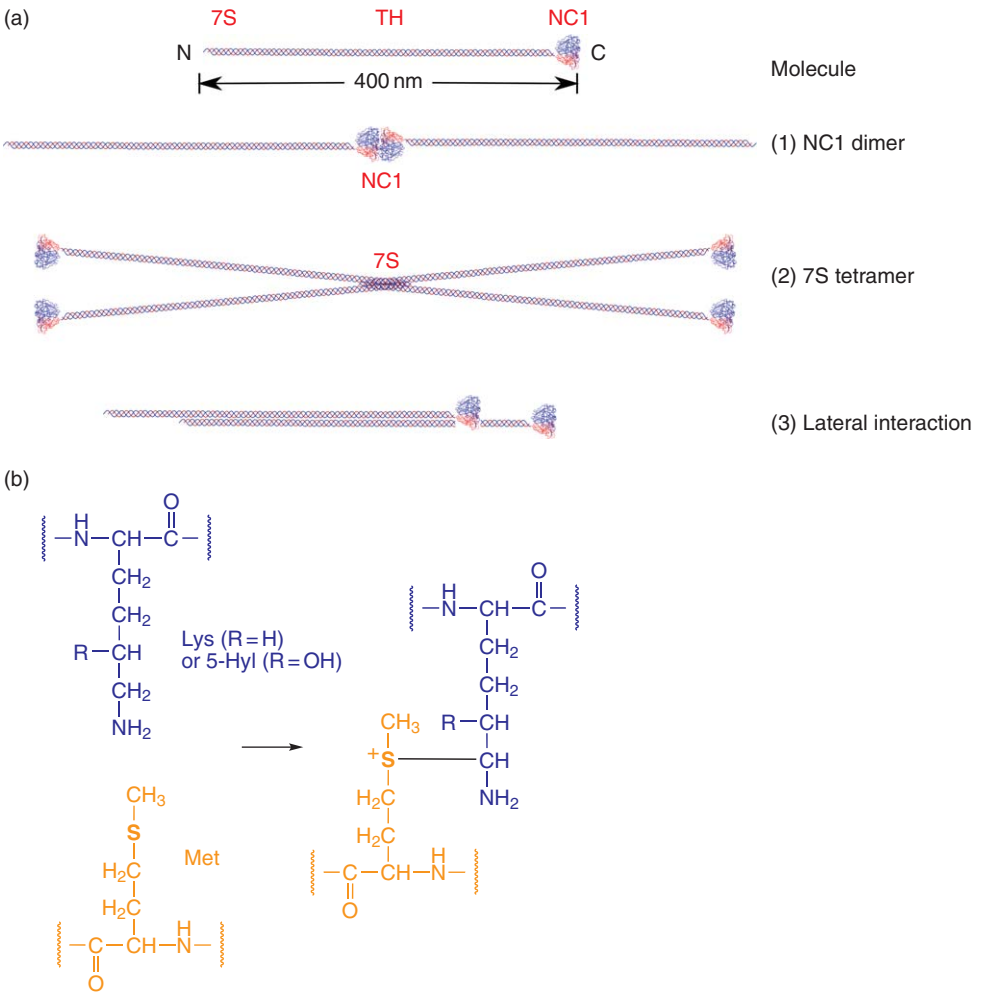


Figure 5 (a) Supramolecular assembly of type IV collagen molecules: (1) a dimer formation at the C-terminal NC1 domain, (2) the association of four molecules at the N-terminal 7S domain, and (3) lateral interaction at the collagenous domain. (b) *S*-(hydroxy)lysylmethionine found in the NC1 domain dimer of two type IV collagen molecules (hexamer of the α chains).

membrane zone collagen is more appropriate than multiplexin. Transmembrane collagens and type VII collagen are also sometimes called basement membrane zone collagens. The structural and physiological name of these two types of collagen would be basement membrane collagens with glycosaminoglycan chains. Type XV collagen is highly homologous to type XVIII collagen.

5.16.2.6.1 Type XV collagen

Type XV collagen has a collagenous domain of 577 residues containing 13 interruptions of the repeated -Gly-Xaa-Yaa- sequence.[200] The N-terminal domain consists of 530 residues and the C-terminal domain consists of 256 residues. Two Cys in the middle part of the molecule are involved in interchain disulfide linkages.[201] Chondroitin sulfate chains are attached at the N-terminal NC domain.[201] The umbilical cord is the major source of type XV collagen with only about $1–2 \times 10^{-4}$ percent of total protein. Thus, type XV collagen is the second most rare collagen protein isolated from tissue to date,[202] exceeded only by type XIX collagen.[203] The type XV collagen multimer structure has a crisscross shape by electron microscopy.[202] Type XV collagen localizes to basement membranes and may connect collagen fibrils with the basement membrane.[204,205] Type XV collagen might also function as a biological spring to stabilize and enhance resistance to compression or expansion forces.[202]

The C-terminal NC1 domain, called type XV endostatin or restin, has antiangiogenic activity similar to endostatin, the NC1 domain of type XVIII collagen (see below).[206,207] The NC1 domain has a trimerization domain, a hinge region that is more sensitive to proteolysis in type XVIII collagen, and the endostatin domain. Lack of type XV collagen in mice causes a skeletal myopathy and cardiovascular defects.[79]

5.16.2.6.2 Type XVIII collagen

Type XVIII collagen is a heparan sulfate proteoglycan consisting of three α1(XVIII) chains and localized in basement membranes.[208–210] Type XVIII collagen is expressed as three splice variants differing in their N-terminal NC domains. The longest variant has a cysteine-rich domain homologous to the frizzled receptors and this frizzled domain was shown to inhibit wnt/β-catenin signaling after its proteolytic release.[211] Endostatin is a 20 kDa fragment derived from the C-terminal NC domain of type XVIII collagen and has been reported to inhibit angiogenesis and tumor growh.[212,213] The endostatin domain also has heparin affinity.[207,214] The therapeutical potential of endostatin for tumor suppression was proven to be low.[215,216] The mutations in human collagen XVIII (COL18A1) were identified in patients with Knobloch syndrome, characterized by high myopia, vitroretinal degeneration, and occipital encephalocele.[217] Analyses of Col18al null mice revealed that collagen XVIII is critical for normal blood vessel formation in eye[80] and those mice have defects not only in the posterior but also the anterior parts of eye.[218] Age-dependent loss of vision in those mice is associated with abnormalities of the retinal pigment epithelium and formation of basement membrane deposits.[219]

5.16.2.7 Type VI Collagen

Type VI collagen does not belong to any subgroup of collagen. It is thought to have a unique molecular structure and set of characteristics. In general, the chain composition of the molecule is α1(VI)α2(VI)α3(VI). However, recently, three new α chains of type VI were found.[220,221] All of these three chains are very similar to the α3(VI) chain. The α3–α6(VI) chains consist of seven von Willebrand factor A (VWA) domains followed by a collagenous domain, two C-terminal VWA domains, and a unique domain. In humans, the COL6A4 gene is separated into two parts, and both are suggested to be pseudogenes. As of now, most of type VI collagen is believed to form an α1(VI)α2(VI)α3(VI) heterotrimer. If the new three α chains α4(VI), α5(VI), and α6(VI) were to form a similar molecular structure, the heterotrimer chain compositions would be α1(VI)α2(VI)α4(VI), α1(VI)α2(VI)α5(VI), or α1(VI)α2(VI)α6(VI). Therefore, some of the earlier studies should be reviewed from the perspective of potentially more complex chain compositions with the newer chain forms now identified.

Type VI collagen has a collagenous domain (100 nm) that is about one-third the size of that in type I collagen (300 nm). The collagenous domain is flanked by two large N- and C-terminal globular domains, which are homologous to VWA domains. The ratio of the NC domain with respect to the collagenase domain is much larger in type VI collagen compared with type I–V collagens. In particular, the α3(VI)–α6(VI) chains are by far longer than the other two α chains in type VI collagen. In the NC domain, there are 3, 3, and 12 VWA domains in the α1(VI), α2(VI), and α3–α6(VI) chains, respectively. The N-terminal globular domain of the α3(VI)–α6(VI) chain is much larger than the others, and contains up to 10 N-terminal A subdomains (N10–N1) and 2 C-terminal A subdomains (C1 and C2).[222]

A unique feature of the collagenous domain of type VI collagen α chains is a much lower frequency of the stable repeat -GPP-, where the second Pro is expected to be 4-hydroxylated. For example, the collagen domain of COL1A1 has 12% of the repeating tripeptide units (42 GPP of 338), whereas the total of COL6A1–A3 has only 4.7% (15, 7, and 4 in α1(VI), α2(VI), and α3(VI), respectively, of 320 tripeptide units). Each α chain has one Cys in the collagenous domain. The collagenous domain of type VI collagen has a high content of glucosylgalactosyl hydroxylysine,[223] which may stabilize the triple helical structure. Interestingly, lysyl hydroxylase 3 (also a glycosvltransferase) null mice displayed abnormalities in the supramolecular assembly of type VI collagen beaded fibrils.[85] Additionally, the intracellular tetramerization of type VI collagen molecules was affected indicating a role for lysyl hydroxylation and glycosylation during the secretion of type VI collagen.[85]

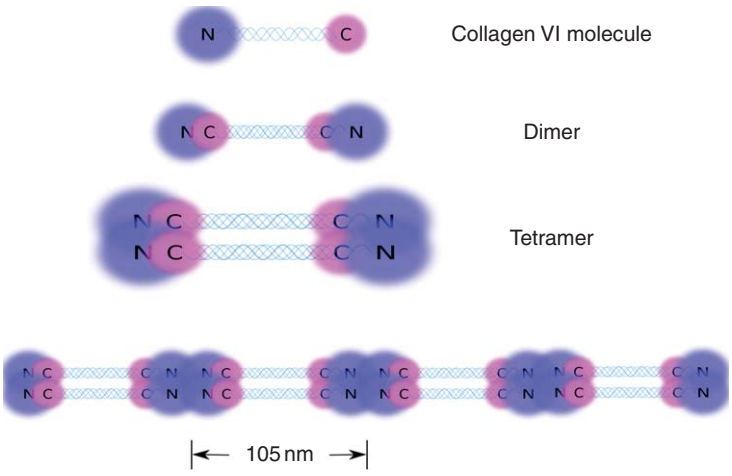


Figure 6 Model of type VI collagen assembly. Two type VI collagen molecules assemble with 30 nm overlap[224] with two pairs of disulfide bonds between cysteines, one in collagenous domain and another in the C-terminal globular domain.[225] Two dimers form a tetramer with disulfide bonds presumably in the α3(VI) chains. The tetramers assemble into the long beaded filamentous structure with 105 nm periodicity.

A basic supramolecular assembly of type VI collagen molecules is schematically illustrated in **Figure 6**. Type VI collagen forms a beaded filament with 105 nm periodicity by TEM. From biochemical and electron microscopy analyses, the current model is that the two molecules form a dimer with antiparallel interaction. Then, two dimers form a tetramer and the tetramers overlap to make the so-called banded fibrils. Although the amount of type VI collagen is abundant in various tissues, its function is still unclear. The type VI collagen meshwork is fine and difficult to detect by electron microscopy. A monoclonal antibody or ruthenium staining is required to see the ultrastructure of the type VI collagen meshwork.[226] Electron microscopy showed that the C-terminal subdomain C5 of the α3(VI) chain is critical for microfibril formation.[227] The three recently found α chains might influence the organization of the other type of supramolecular assembly of type VI collagen.

Type VI collagen exists ubiquitously throughout the body. Whereas most types of collagen localize in either hyaline cartilage or noncartilage tissue, type VI collagen localizes in both. Type VI collagen is abundant in bone marrow and interacts with hematopoietic cells via its collagenous domain.[228] Type VI collagen may interact with basement membrane type IV collagen, and also with collagen fibrils consisting of type I collagen. One of the functions of type VI collagen might be to anchor the basement membranes by interacting with type IV collagen.[229] In skeletal muscle, type IV and VI collagens localize near the basement membrane zone. Type VI collagen is also found in areas far from basement membranes.

Type VI collagen interacts with many proteins such as small leucine-rich repeat proteoglycans (SLRPs), decorin,[230] biglycan,[231] and the cell surface proteoglycan NG2.[232] Recently, interesting biological roles of type VI collagen were reported. The α3(VI) chain of native type VI collagen is cleaved by MMP-11, and this processing might be related to the differentiation of adipocytes.[233] Macrophages synthesize type VI collagen, suggesting that it has a significant role in tissue repair.[234] Type VI collagen is increased in brain tissues from Alzheimer's disease patients, and is produced by neurons. The addition of type VI collagen to cultured neuron cells blocks the association of amyloid β peptide oligomers with neurons, and prevents neurotoxicity. Type VI collagen is an important component of the neuronal injury response and has neuron protective activity.[235]

Mutations in both the NC and collagenous domains of COL6A1, COL6A2, or COL6A3 cause Bethlem myopathy,[236,237] a dominantly inherited disorder. Ulrich congenital muscular dystrophy can be caused by recessive mutations in any of COL6A1–A3 genes encoding the subunits of type VI collagen.[238,239]

Col6a1-deficient mice display an early onset of myopathy,[69] and dysfunction of mitochondria.[70] Details of the relation between the α1(VI) chain and pathogenesis of mitochondria are still unknown.

5.16.2.8 Type VII Collagen

Procollagen VII is a homotrimeric molecule consisting of a central 145 kDa triple helical collagenous domain flanked by the N-terminal NC1 domain of 140 kDa and the C-terminal NC2 domain of 30 kDa.[240–243] The collagenous domain has 1530 residues, which is longer than that in fibrillar collagen, and has more than 20 interruptions of the Gly-Xaa-Yaa- repeat sequence. Type VII collagen localizes beneath stratified squamous epithelium, and is a major component of anchoring fibrils at the epidermal–dermal junction in skin. Originally, the collagenous domain of type VII collagen was isolated from pepsin-treated human skin. During the initial stages of anchoring fibril formation, procollagen VII molecules form a dimer, which overlaps in an antiparallel manner between the C-terminal NC (NC2) and collagenous domains. The dimer is stabilized by an interchain disulfide bond formed between the Cys2634 residue located in the triple helical region close to the C-terminus and either of the two Cys residues (Cys2802 or Cys2804) located in the NC2 domain of the adjacent molecule.[244,245] *In vitro* cleavage assays have confirmed that the NC2 domain of procollagen VII is processed by the BMP-1 family of enzymes, BMP-1, mTLL-1, and mTLL-2.[246] This process is essential for the interaction between the NC2 domain and the collagenous domain of type VII collagen.[247] There is evidence that type VII collagen is fully processed in the skin of mouse embryos null for the Bmp-1 gene, which encodes both BMP-1 and mTLD.[246] This suggests that mTLL-1 and mTLL-2 may compensate for the loss of the Bmp-1 gene in the mouse especially with regard to the processing of the NC2 domain of procollagen VII.

The NC1 domain of type VII collagen binds to the $\beta 3$ chain of laminin,[248] laminin-5 ($\alpha 3\beta 3\gamma 2$), and type IV collagen.[249] The triple helical domain of type VII collagen functions to promote the migration of human keratinocytes.[250]

Mutations of COL7A1 cause a skin disease called dystrophic epidermolysis bullosa. It is a blistering disease caused by the separation of the dermal–epidermal junction.[251,252] Biochemical analysis revealed the single amino acid substitution in type VII collagen that affects the supramolecular assembly of type VII collagen.[250,253]

5.16.2.9 Short-Chain Collagens (Types VIII and X)

Type VIII and X collagens are called short collagens or meshwork-forming collagens. For example, the collagenous domain of human COL8A1 is 454 residues long, which is less than half the size of the major triple helical domain of COL1A1. The short-chain collagens contain a short triple helical domain flanked by a short N-terminal and a large C-terminal NC domain.

5.16.2.9.1 Type VIII collagen

Type VIII collagen was originally called EC collagen because it was characterized from cultured endothelial cells.[254] Type VIII collagen is localized in subepithelial or subendothelial matrices.[255–257] Descemet's membrane of bovine eye shows hexagonal-like meshwork structure with thin type VIII collagen fibrils. Immunohistochemical analysis and electron microscopic observation have shown that type VIII collagen is a major component of the skeleton of this structure.[258] Additionally, type VIII collagen may form a tetrahedron-like supramolecular structure observed by rotary shadowing analysis.[259]

There are two similar α chains in type VIII collagen, $\alpha 1$(VIII) and $\alpha 2$(VIII). Two homotrimeric subtypes, $[\alpha 1(\text{VIII})]_3$ and $[\alpha 2(\text{VIII})]_3$, are suggested to be the major molecular species,[260,261] but a heterotrimer can also exist.[262,263] Stephan *et al.*[259] analyzed the purified type VIII collagen by atomic force microscopy and electron microscopy. They confirmed the tetrahedron model from biochemical analysis and ultrastructural studies.

The NC1 domain of type VIII collagen is important for trimerization. ACRP30/adiponectin, a member of the complement C1q family of proteins,[264] and the type X collagen NC1 domain[265] have similar structures as the NC1 domain of type VIII collagen; however, the type VIII collagen NC1 trimer lacks the buried calcium cluster found in the type X collagen NC1 trimer.[265] The crystal structure of this domain has similarity to the TNF (tumor necrosis factor) family of proteins (PDB accession number 1O91).[266]

Double knockout mouse of two type VIII collagen genes, Col8a1$^{-/-}$/Col8a2$^{-/-}$, develop a distinct phenotype of dysgenesis of the anterior segment of the eye with a globoid, keratoglobus-like protrusion of the anterior chamber. The corneal stroma is diffusely thin, similar to what is seen in human keratoglobus. Descemet's membrane is markedly thinned and lacks the anterior banded zone. The corneal endothelial cells

are enlarged and reduced in number. Finally, mutant corneal endothelial cells show a decreased ability to proliferate in response to different growth factors *in vitro*, suggesting that type VIII collagen may function as an enhancer of growth factor-induced proliferation of cells.[72]

In humans, mutations in the $\alpha 2$(VIII) gene result in corneal endothelial dystrophies, Fuchs endothelial dystrophy (FECD), and posterior polymorphous corneal dystrophy (PPCD).[267] Recently, it was reported that the mutation L450W in COL8A2 causes an early-onset subtype of Fuchs corneal dystrophy.[268]

5.16.2.9.2 Type X collagen

Type X collagen was first reported in 1983.[269] It localizes to the hypertrophic zone of mineralizing cartilage during bone growth, and within the growth plate.[270] The chain composition of type X collagen is $[\alpha 1(X)]_3$. Type X collagen has a short triple helical collagen domain flanked by the N-terminal NC2 and the C-terminal NC1 domains. The C-terminal NC1 domain has complement C1q-like structure. The NC1 trimeric assembly is strengthened by a buried cluster of calcium ions (PDB accession number 1GR3).[265] Observations from rotary shadowing electron microscopy of isolated chicken cartilage type X collagen have shown that the initial oligomerization of type X collagen begins at the C-terminal NC1 domain, and it can form a regular hexagonal packing with an intermodular length of 100 nm *in vitro*.[73]

Mutations of COL10A1 cause Schmid metaphyseal chondrodysplasia (SMCD), an osteochondrodysplasia characterized by short-limbed short stature with normal face structure, but with generalized metaphyseal dysplasias of the long and short tubular bones.[271,272] From the crystal structure of the NC1 domain of type X collagen, SMCD mutations in most internal residues most likely prevent protein folding, whereas mutations of surface residues may affect type X collagen oligomerization in a dominant-negative manner.[265]

From recent reports of the mouse model studies, type X collagen might have an important role not only in growth plate but also in bone marrow functions.[74,273]

5.16.2.10 FACIT and FACIT-Like Collagens (Types IX, XII, XIV, XVI, XX, XXI, and XXII)

Originally, FACIT was named for type IX collagen,[274] but later was extended to a number of other collagens. So far, type IX collagen remains the only collagen that has actually been shown to interact directly with fibrillar collagen molecules.[275] Although some reports suggest the interaction of other FACIT collagens with collagen fibrils,[276] the molecular mechanism and binding sites are unknown. The newer FACIT collagens have been characterized very little by biochemical analysis. Some FACITs might not actually interact with collagen fibrils consisting of type I collagen. All FACIT genes have about 250 residues of a LamG domain (TSP module) in the N-terminus, conserved positions of the interrupted repeats of Gly-Xaa-Yaa- sequence in the COL1 domain, and conserved positions of NC2 domain (**Figure 7**). All of the NC1 domains have conserved spacing and location of cysteines in a CXXXXC sequence. The NC2 domain of type XIX collagen has recently been shown to be a stable trimerization domain but it is not known if this domain has the same function in other FACIT collagens.[278] FACIT collagens, except type IX, XII, and XIV collagens, are classified based on the primary sequence, although the name of FACIT might not be the best choice in the classification of these types of collagen.

5.16.2.10.1 Type IX collagen

Type IX collagen belongs to the group of FACIT collagens and is a component of collagen fibrils in hyaline cartilage. It forms a heterotrimer with three different chains, $\alpha 1$(IX)$\alpha 2$(IX)$\alpha 3$(IX). The domains of type IX collagen are three triple helical domains (COL1, COL2, and COL3) separated and flanked by nontriple helical (NC) domains, NC1–NC4. There is a splicing variant in the NC4 domain of the $\alpha 1$(IX) chain. The NC3 domain has one chondroitin sulfate chain attachment site.[279] The length of the chondroitin sulfate chain may vary depending on the tissue type and developmental stage. The short form of type IX collagen in chick embryo vitreous humor has a long chondroitin sulfate chain in the NC3 domain of the $\alpha 2$(IX) chain.[280] The $\alpha 1$(IX) chain in bovine nucleus pulposus, the central zone of the intervertebral disc, has only a short form of the $\alpha 1$(IX) chain and lacks the NC4 domain, while the $\alpha 1$(IX) chain in hyaline cartilage has the long form of the $\alpha 1$(IX) chain.[281]

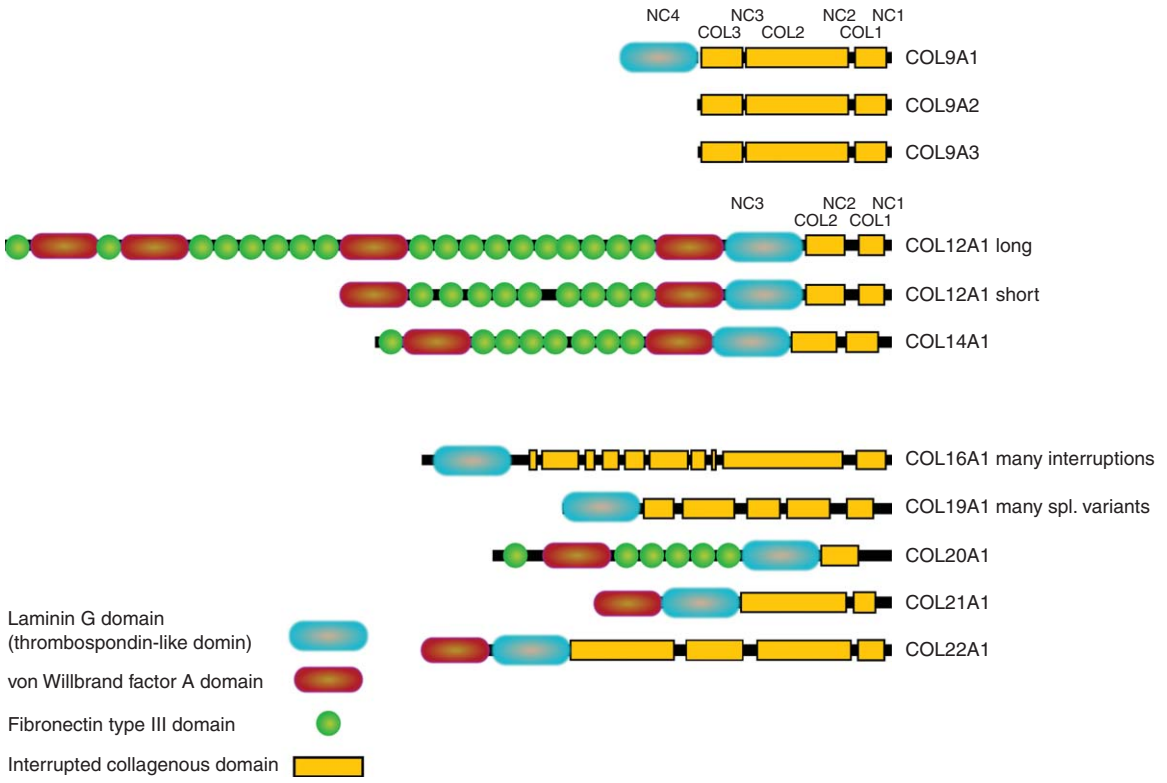


Figure 7 Domain structure of FACIT/FACIT-like collagens. The domains are named from the C-terminus as NC1, COL1, NC2 domains. The structures of the C-terminal domains are relatively homologous but the rest of molecules are various among the molecules. The domain annotation is dependent on the reference.[277]

Type IX collagen molecules assemble with type II and XI collagens to form the fibrils in hyaline cartilage. The interaction of types II and IX is via the triple helical COL1 domain of type IX.[282] Synthetic peptides of NC1 domains of type IX α1, α2, and α3 chains can form a 1:1:1 heterotrimer with disulfide bonds *in vitro*.[283] The yield was only 10% but more than 95% of the formed oligomer peak was the $\alpha1\alpha2\alpha3$ trimer.[283] The crystal structure of the col9a1 N-terminal NC4 domain was analyzed[284] and found to contain a TSP-1 domain, which also exists in at least 17 other collagen polypeptides in the N-terminal region. The structure of the NC4 domain of col9a1 is similar to the previously characterized laminin–neurexin–sex hormone-binding globulin (LNS) structure,[285] which is dominated by an antiparallel β-sheet sandwich. In addition, a zinc ion was found in a position similar to that of the metal-binding site of other LNS domains. From the determined K_d constant of 11.5 m mol l^{-1} by nuclear magnetic resonance (NMR) titration, the zinc ion is thought to work as a regulator rather than playing a structural role.[284]

The mouse model with inactivated col9a1 gene does not synthesize type IX collagen.[286] Col9a1 knockout mice develop normally without detectable abnormalities, but severe joint disease develops by approximately 9 months of age.[287] While knockout of either of the two cartilage matrix components, TSP-3 and TSP-5 (also known as COMP), does not show a significant phenotype in the knockout mouse, the triple knockout of the TSP-3, TSP-5, and type IX collagen results in severe skeletal abnormalities with 20% reduction in limb length.[288]

Mutations in COL9A1, COL9A2, and COL9A3 cause multiple epiphyseal dysplasia (MED), an autosomal dominant chondrodysplasia.[289–291] Also, mutations in COMP, which interacts with type IX collagen, cause MED.

5.16.2.10.2 Type XII collagen

Type XII collagen consists of two short collagenous domains COL1 and COL2 along with three NC domains (NC1–NC3) (**Figure 7**).[292,293] The N-terminal NC3 domain makes up close to 90% of the total molecular mass. The NC3 domain contains 2–4 VWA domains, several FNIII repeats, and a TSPN domain. There are

splicing variants called the long form, XIIA-1, XIIA-2, and short form, XIIB-1, XIIB-2. Three chains are covalently cross-linked with disulfide bonds.

Type XII and XIV collagens may play a role in the organization of the collagen fibrils. When a small aliquot of type XII and XIV collagens is added to the collagen gel with dermal fibroblasts, the gel contracts in a dose-dependent manner.[294] The N-terminal NC3 domain is responsible for this activity.

5.16.2.10.3 Type XIV collagen

Type XIV collagen is very similar to type XII collagen with 92.5% similarity. Type XIV collagen was originally isolated from chicken embryos. The same protein from human placenta was identified and named undulin.[295] Type XIV collagen is a FACIT, but its interaction with collagen types I–V is weak.[296] It interacts with heparin at more than one site, and has a significant affinity for the triple helical domain of type VI collagen[296] Decorin may bind to the N-terminal fibronectin type III repeat of NC3 of type XIV collagen.[297] Therefore, a recent report suggested that the major interaction of type XIV collagen with type I collagen molecules is indirect.

Col14a1 null mice have fibril diameter abnormalities in the early postnatal stage but the adult showed no significant differences.[78] Type XIV collagen might function early in development regulating the entry of fibril intermediates into lateral fibril growth.[78]

5.16.2.10.4 Type XVI collagen

Type XVI collagen is composed of 10 collagenous domains (COL1–COL10) flanked by 11 NC domains.[298] Type XVI collagen localizes near the dermal–epidermal junction. From immunoelectron microscopy of the papillary dermis, type XVI collagen associates with a fibrillin-1-containing matrix but not on collagen fibrils.[299]

5.16.2.10.5 Type XIX collagen

Type XIX collagen was initially found from human rhabdomyosarcoma cell line cDNA library.[300] From the primary structure, it is classified as one of FACIT collagens, but the experimental data suggest that type XIX collagen is one of the basement membrane zone collagen family like type XV and XVIII collagens.[301,302] Type XIX collagen was purified from human umbilical cord and biochemically analyzed.[203] Type XIX collagen is by far the least abundant collagen so far purified. The amount is only 10^{-5} percent of the dry weight of umbilical cord.[203] This collagen is easily solubilized with salt from the tissue, indicating that it is not tightly associated with the extracellular matrix. Rotary shadowing electron microscopy observation of the purified molecule showed about 240 nm rope-like structure with a globular end, which must be the N-terminal largest NC domain.[203] The positions of sharp kinks corresponded to the estimated position of NC domains.[203] Electron microscopic analysis suggests that type XIX collagen self-assembles into oligomers via the N-terminal NC domain.[203] Embryonic expression of the type XIX collagen gene is transient and confined almost exclusively to differentiating muscles, and transiently expressed in brain by *in situ* hybridization analysis.[303] Like type XIII collagen, type XIX collagen has many splicing variants.[304] The col9a1 gene knockout mouse showed an abnormality in esophageal structure and function.[302]

5.16.2.10.6 Type XX collagen

Type XX collagen was reported as chick cDNA clone.[305] No further details are known at present.

5.16.2.11 Transmembrane Collagens (Types XIII, XVII, XXIII, and XXV)

Most types of collagen are extracellular matrix proteins but four of the vertebrate collagens (XIII, XVII, XXIII, and XXV) are classified as type II membrane proteins. They contain an N-terminal cytoplasmic domain, a transmembrane domain, and an extracellular collagenous domain. Also, a membrane-bound protein, ectodysplasin-A, is a membrane protein with 19 -Gly-Xaa-Yaa- repeat.[306] Macrophage scavenger receptor,[5,307] macrophage receptor with collagenous structure (MARCO),[308,309] scavenger receptor with C-type lectin (SRCL),[310,311] and collomin[312] are also in the transmembrane collagen-like family. All collagenous domains of these membrane proteins are on the extracellular side.

At the amino acid level, type XXIII collagen is similar to type XIII and XXV collagens with 54 and 56% identity, respectively. There is high identity across the 20-amino-acid C-terminal NC domain. In addition, the

COL1 domain of type XXIII collagen contains multiple copies of a conserved KGD motif used for integrin-mediated cell adhesion by collagen type XVII.[313]

5.16.2.11.1 Type XIII collagen

Type XIII collagen is the first collagen to be identified as the transmembrane collagen.[314] It exists in the plasma membrane of cells with a short N-terminal cytosolic domain and a large collagenous ectodomain with many interruptions. It forms a homotrimer. The interchain disulfide bonds and prolyl 4-hydroxylation of Pro stabilize the triple helical structure.[315] In contrast to most other collagens, the domains of type XIII collagen are numbered from the N-terminus. Type XIII collagen consists of three collagenous domains (COL1 – COL3) flanked by four NC domains (NC1 – NC4). One of the unique features of type XIII collagen is the number of splicing variants. Type XIII collagen is widely distributed in tissues, such as in myotendinous junction, and at cell–cell interaction sites such as the intercalated disc of heart muscle. In fibroblast cell culture, type XIII collagen is found at the focal adhesion site.[316]

Type XIII collagen interacts with the $\alpha 1\beta 1$ integrin receptor.[317] The mouse without the transmembrane domain of col13a1 showed progressive myopathy-like histology,[76] implying that one of the functions of type XIII collagen is working as a component of muscle. The ectodomain of type XIII collagen interacts with many molecules including nidogen-2, fibronectin, and perlecan,[318] suggesting that type XIII collagen is one of the key molecules of cell adhesion.

5.16.2.11.2 Type XVII collagen

Type XVII collagen is known as 180 kDa bullous pemphigoid antigen (BP180, BPAG2) on epithelial cells.[319] Type XVII collagen is located in hemidesmosomes and spans into the basement membrane of the epidermal–dermal junction. Also central nervous system neurons express type XVII collagen in retina.[320] Type XVII collagen can be cleaved at the NC16A domain adjacent to the cell membrane and released from the cell surface.[321] From rotary shadowing electron microscopy analysis, the length of the ectodomain is 160–200 nm. Type XVII collagen is cleaved by a membrane-associated metalloprotease expressed in epithelial cells.[322] TNF-α converting enzyme (TACE), ADAM-10, and ADAM-9 may cleave the ectodomain.[323] These enzymes are expressed by keratinocytes.[323] From immunoelectron microscopy analysis with three different monoclonal antibodies, the ectodomain of type XVII collagen is suggested to make a loop structure.[324] Genetic defects of COL17A1 cause epidermal detachment in junctional epidermolysis bullosa, a heritable skin fragility syndrome.

5.16.2.11.3 Type XXIII collagen

Type XXIII collagen contains about 540 residues.[325] The NC1 domain contains a transmembrane sequence and is followed by COL1, NC2, COL2, NC3, COL3, and NC4 at the C-terminus. Furin proteinases may cleave the NC1 domain, and the cleaved secreted portion of type XXIII collagen may also form multimers and display low-affinity binding to heparin *in vitro*.[325]

5.16.2.11.4 Type XXV collagen

Type XXV collagen was found from Alzheimer's disease brain as a precursor of CLAC (collagenous Alzheimer amyloid plaque component), one of the senile plaque components.[326] Type XXV collagen is similar to type XIII collagen.

5.16.2.12 New Fibrillar Collagens (Types XXIV and XXVII)

Recently, two types of new collagens similar to fibrillar collagen were found. These are types XXIV and XXVII. Type XXIV collagen has some similarity to type V/XI collagen.[16] The triple helical domain is shorter than that of the classical fibrillar collagens. One of the differences of type XXIV and XXVII collagens compared with classical fibrillar collagen (types I, II, III, V, and XI) is the interruption of the -Gly-Xaa-Yaa- repeats in the collagenous domain. The short collagenous domain (COL2) at the N-terminus is missing. NC domain sequences are more similar to type V/XI collagen. Type XXIV collagen is developmentally expressed in cornea and bone.[16] The gene structure of type XXVII collagen, col27a1, is similar to col24a1, and also to a few of type

V/XI collagen genes,[17,18] col5a1, col5a3, and col11a2. Immunohistochemical analysis suggested that type XXVII collagen localizes at the ossification site of hypertrophic cartilage.[327]

5.16.2.13 The other types of collagens (types XXVI and XXVIII)

Type XXVI collagen has two collagenous domains (COL1 and COL2) flanked by three NC domains (NC1–NC3).[328] The N-terminal domain has a cysteine-rich region. This protein is identical to Emu2, which has a cysteine-rich EMI domain as found in emillin and multimerin.[329] Since Emu1 and Emu2 are highly homologous, Emu1 and Col26a1 (Emu2) might form a heterotrimeric molecule. Emu1 is not called a 'collagen'.

Type XXVIII collagen belongs to the class of VWA domain-containing proteins.[330] The primary structure is similar to type VI collagen. It is mainly a component of the basement membranes around Schwann cells in the peripheral nervous system.[330]

5.16.3 Biosynthesis of Collagen Molecules

One of the unique structural properties of the collagen domain is the complicated multistep process of posttranslational modifications. It is likely that collagen receives the highest number of enzymatic reactions for the lifetime of the molecule, especially relative to most other proteins. To synthesize one type I collagen molecule, several hundreds of reactions are catalyzed by enzymes such as prolyl hydroxylases (EC 1.14.11.2), lysyl hydroxylases (LHs), peptidyl prolyl cis–trans isomerases, protein disulfide isomerase (PDI; EC 5.3.4.1), glycosyltransferases, and specific proteinases to cleave the N- and C-propeptides from the procollagen molecules.[331–335] In addition to enzymes, many other molecules assist in folding, solubilizing, sorting, and transporting the collagen molecules through the cell. Type I collagen biosynthesis is schematically illustrated in **Figure 8**. In this section, representative examples of vertebrate posttranslational modification enzymes and chaperones are summarized. **Figure 9** shows the posttranslational modifications of collagenous proteins, and **Figure 10** shows the domain structure of the enzymes.

5.16.3.1 Prolyl 4-Hydroxylation

Prolyl 4-hydroxylation is the most abundant posttranslational modification of collagens. 4-Hydroxylation of proline residues increases the stability of the triple helix and is a key element in the folding of the collagen triple helix.[337–340] In vertebrates, almost all the Yaa position prolines of the Gly-Xaa-Yaa repeat are modified to 4(*R*)-hydroxylproline by the enzyme P4H (EC 1.14.11.2), a member of Fe(II)- and 2-oxoglutarate-dependent dioxygenases. This enzyme is an $\alpha 2\beta$b2-type heterotetramer in which the β subunit is PDI (EC 5.3.4.1), which is a ubiquitous disulfide bond catalyst. The P4H α subunit needs the β subunit for solubility; however, the β subunit, PDI, is soluble by itself and is present in excess in the ER. Three isoforms of the α subunit have been identified and shown to combine with PDI to form $[\alpha(\mathrm{I})]_2\beta_2$, $[\alpha(\mathrm{II})]_2\beta_2$, or $[\alpha(\mathrm{III})]_2\beta_2$ tetramers, called the type I, II, or III enzymes, respectively, and all possess similar catalytic properties.[334,341–343] The type I enzyme is the main form. Type II is the major form in chondrocytes, osteoblasts, endothelial cells, and some other cell types, whereas type III is expressed in many tissues but at much lower levels than types I and II. Collagen P4Hs catalyze the formation of 4-hydroxyproline by the hydroxylation of proline in -X-Pro-Gly- triplets in collagens and other proteins with collagen-like sequences. P4H- requires unfolded chains as a substrate. If 4-hydroxylation is prevented, unfolded procollagens remain bound to P4H and are retained within the ER.[344,345]

The crystal structure of the peptide substrate-binding domain (140–245 of 517 residues of human α1 subunit) of the human type I enzyme forms 2.5 tetratricopeptide (TPR) repeat domains with five α helices (PDB accession number 1TJC).[346,347] The organization of tyrosine residues is suggested to be key to its interaction with the substrate peptide in a polyproline II helix.[347] The TPR motif is composed of a 34 amino acid repeated α helical motif, and is typically involved in protein–protein interactions. The tandem repeats of TPR motifs are found in many proteins related to chaperone, cell cycle, transcription, and protein transport complexes.[348,349] P3H and CRTAP also have multiple TPR domains, which may be important in mediating their interactions with each other as well as other proteins.

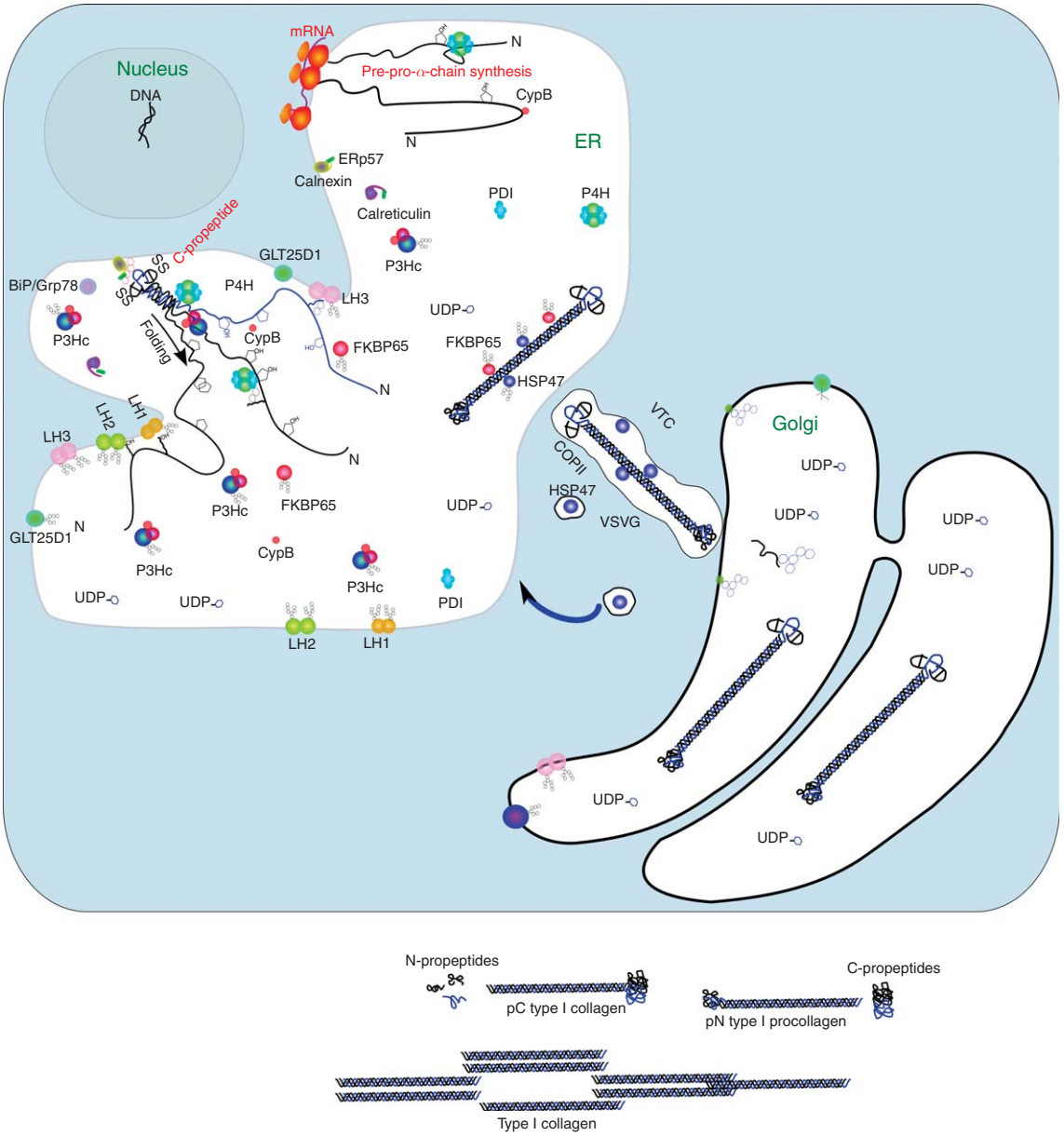


Figure 8 Schematic illustration of type I collagen biosynthesis. After the transcription of procollagen α1(I) and α2(I) genes, pre-pro-α chains are synthesized at the ribosome, and those with a signal peptide enter the ER. The two α1(I) and one α2(I) chains associate via their C-terminal propeptides where disulfide bonds are catalyzed by protein disulfide isomerase (PDI). Prolines in the Yaa position are 4-hydroxylated by the prolyl 4-hydroxylase/PDI tetramer. One Pro in the 986 position of the α1(I) chain is modified to 3-hydroxyproline by prolyl 3-hydroxylase 1, which forms a complex (P3H1c) with cartilage-associated protein (CRTAP) and cyclophilin B (CypB), a peptidyl prolyl cis–trans isomerase. About one-third of Lys residues in the Yaa position are modified to 5-hydroxylysine by lysyl hydroxylase LH1–LH3. The Lys residues in the telopeptides are modified to hydroxylysine by LH2. Some of the Asp residues in the noncollagenous domains are N-glycosylated. Lys87 of the α1(I) and the α2(I) chains, Lys930 of the α1(I), and the Lys933 of the α2(I) can be modified to glucosylgalactosyl hydroxylysine. Calnexin and calreticulin may interact with the N-linked carbohydrate of the noncollagenous prodomains. BiP/Grp78 is also a typical ER chaperone protein thought to be present during collagen biosynthesis. FKBP65 is a peptidyl prolyl cis–trans isomerase present during collagen biosynthesis and one of the most abundant ER proteins of chick embryos. The folding of the long triple helical domain proceeds in the C-terminal to N-terminal direction. HSP47 interacts with triple helical collagen and prevents premature aggregation. After the folding of a triple helix, the procollagen molecule is transferred to the Golgi apparatus via COP (coat protein complex) I and II, VSVG, and other proteins known to be involved in ER-to-Golgi transport. Intracellular vesicles of developing chick and mouse embryo tendon fibroblasts contain D-periodic banded collagen fibrils with hexagonal packing as seen in the extracellular space.[336]

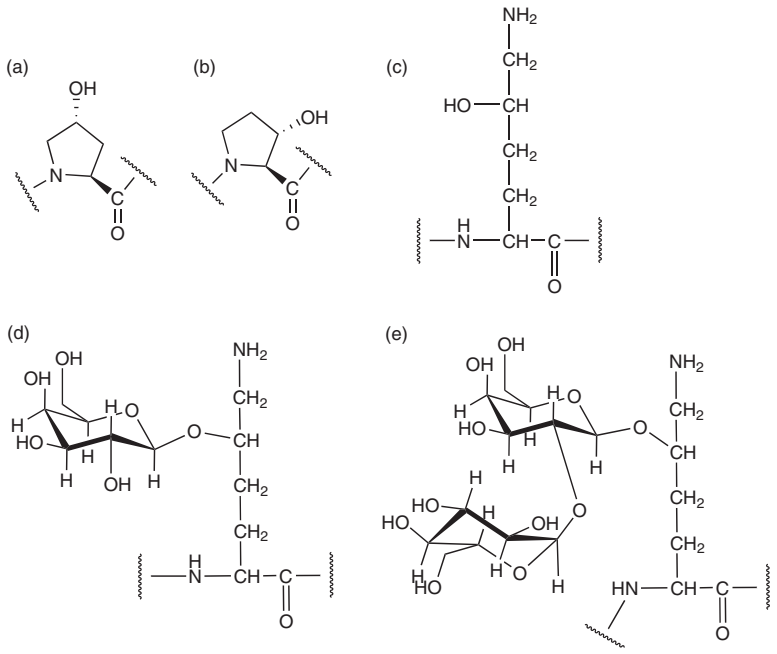


Figure 9 Posttranslationally modified residues found in collagen. (a) 4-(*R*)-hydroxyproline, (b) 3-(*S*)-hydroxyproline, (c) 5-hydroxylysine, (d) β-D-galactosyl-5-hydroxylysine, and (e) α-D-glucopyranosyl-(1 $\rightarrow$ 2)-β-*O*-D-galactopyranosyl hydroxylysine.

The P4Hα(I) (αI subunit) null mouse had abnormal type IV collagen deposition in the basement membranes and a significant decrease of P4H activity in cultured fibroblasts.[83] The null mice were embryonic lethal at day E10.5; however, collagen fibrils were formed with D-periodic banding patterns suggesting that P4Hα(II) and P4Hα(III) might compensate for the lack of P4Hα(I) in fibrillar collagens at least until the E10.5 stage. Importantly, inability to form basement membranes suggested that adequate formation of type IV collagen requires the P4Hα(I) subunit.

5.16.3.2 Prolyl 3-Hydroxylation

Proline in the Xaa position of the -Gly-Xaa-Yaa- repeated sequence is sometimes modified to 3Hyp by (P3H) (EC 1.14.11.7). In vertebrates, three isoforms P3H1, P3H2, and P3H3 are reported.[350] P3H1 was shown to catalyze the formation of 3Hyp on an unfolded type I procollagen substrate *in vitro*.[350] The 3-hydroxylation of proline has been found only in the Xaa position of the -Gly-Xaa-4Hyp- sequences. In contrast to 4-hydroxylation, 3-hydroxylation occurs much less frequently. Studies with synthetic model peptides showed that the 3Hyp in the Yaa position destabilizes the collagen helix.[351,352] However, the insertion of 1–9 3Hyp residues in the Xaa position of the H-(Gly-Pro-4(*R*)Hyp)$_9$-OH stabilizes the triple helix by about 0.5 °C per 3Hyp residue.[353] The crystal structure of the triple helical collagen model peptide with the sequence H-(Gly-Pro-4(*R*)Hyp)$_3$-(Gly-3(*S*)Hyp-4(*R*)Hyp)$_2$-(Gly-Pro-4(*R*)Hyp)$_4$-OH shows almost identical main-chain structure when compared to that of the -Gly-Pro-Pro- or -Gly-Pro-4(*R*)Hyp- repeated peptide with 7/2 helix symmetry,[354] suggesting that the modification from Pro to 3Hyp has very little effect on the triple helical structure of collagen molecules. The physiological role of the 3Hyp residues is still unknown. The 3Hyp modification might affect the interaction with other molecules such as SLRPs.

P3H1 exists in a stable complex with CRTAP and cyclophilin B (CypB).[88,350] Mutations in CypB cause hereditary equine regional dermal asthenia (HERDA), a degenerative skin disease that affects the Quarter Horse breed.[355] This suggests that one of the functions of the P3H/CRTAP/CypB complex might be disturbed by a CypB mutation. The function of 3-hydroxyproline in collagen is still uncertain; however, mutations in

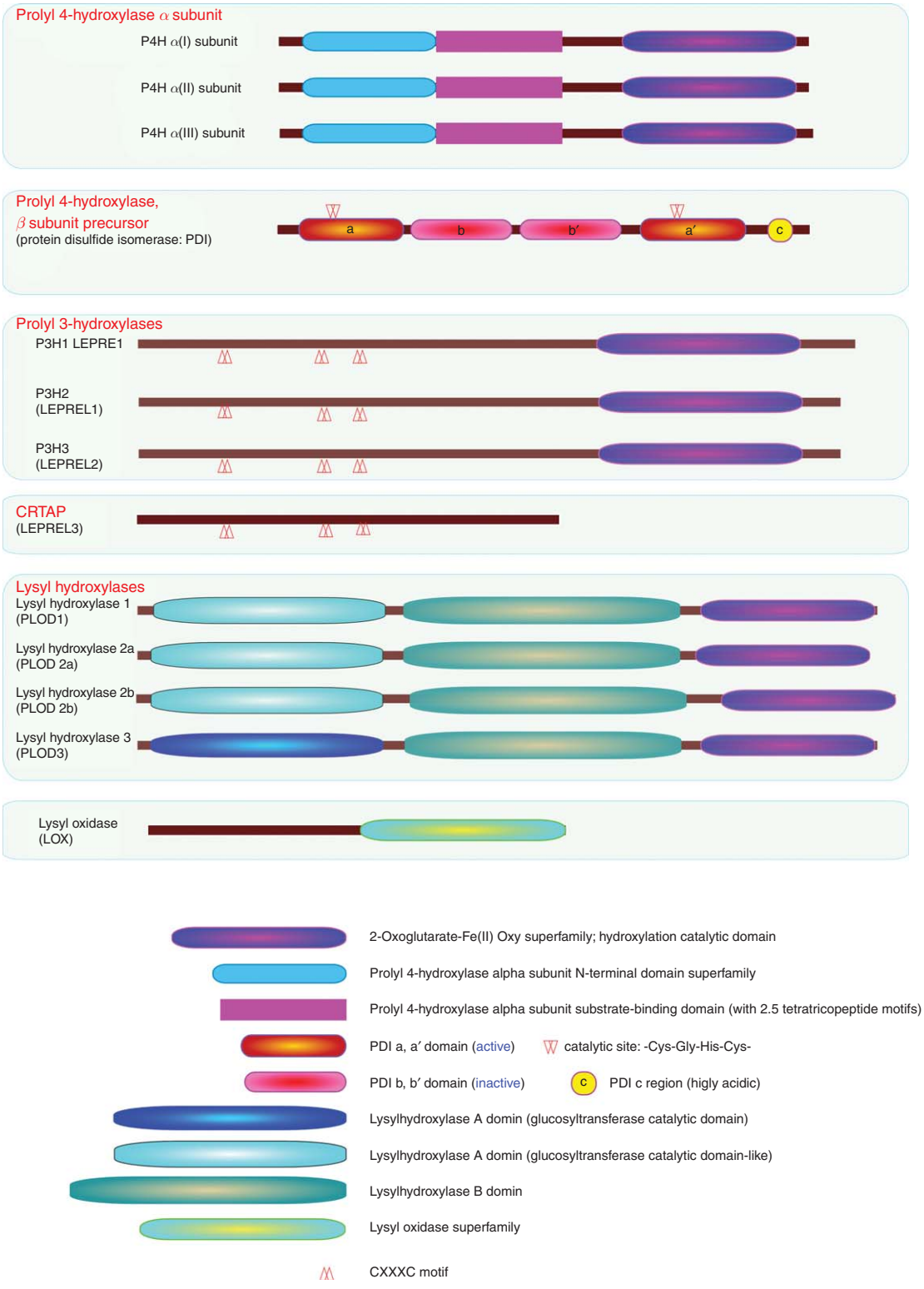


Figure 10 Schematic illustration of the posttranslational enzymes and related proteins. P4H α subunits have three isoforms. Each has three domains. The substrate-binding domain is in the middle. The catalytic domain is at the C-terminal end. Lysyl hydroxylase-3 (LH3) has two different catalytic activities. The N-terminal domain has the glucosyltransferase activity and the C-terminal domain has the hydroxylase activity. LH1 and LH2 also have similar domain structures but the glucosyltransferase activities are not detected *in vitro*.

CRTAP or P3H1 in human patients have recently been shown to lead to a recessive form of OI indicating a possible role in collagen fibril formation and development.[88,97,356] P3H2 has recently been shown to preferentially catalyze the 3-hydroxylation of proline residues *in vitro* using peptide substrates corresponding to type IV collagen sequences.[357] These results indicate that different isoforms of the P3H family may have different substrate specificities within the collagen family. Additionally, *in situ* hybridization and real-time PCR data suggest some tissue- and developmental stage-specific expression patterns of the P3H family members.[358]

Mutations in P3H1 cause OI type VIII, but the molecular mechanism of this type OI is still unknown. **Table 3** lists the occurrences of 3Hyp in collagens.

5.16.3.3 Lysyl 5-Hydroxylation

Lysine residues in the Yaa position of the -Gly-Xaa-Yaa- sequence are partially hydroxylated by LHs (EC 1.14.11.4) as a co- and posttranslational event during collagen biosynthesis.[335,375,376] Also in the NC telopeptides, -Hyl-Ser- or -Hyl-Ala- sequences are found. In type I collagen, human COL1A1 has both N- and C-(PPQEKAHDG) telopeptides containing lysyl hydroxylation sites, whereas COL1A2 has only one potential Hyl site in the N-telopeptide. The type IV collagen NC1 domain also has Hyl in STLKAGEL (COL4A1) and SADTLKAGLIR (COL4A2) sequences. Additionally, Hyl is found in collagenous proteins such as complement C1q, acetylcholinesterase, and adiponectin.[377] Prolyl 4-hydroxylation of the Yaa position of -Gly-Xaa-Yaa- sequences is close to 100%. In contrast, lysyl hydroxylation occurs to a much lower extent, especially in type I collagen. From amino acid analysis data, lysyl hydroxylation occurs in about one-third of the Yaa position lysines in the type I collagen triple helical domain. The LH does not hydroxylate free lysine[378,379] or the tripeptide Lys-Gly-Pro-,[379] but a single tripeptide of X-Lys-Gly fulfills the minimal requirement for

Table 3 Contents of 3-hydroxyproline in collagens

	Residues	*Reference*
Type I collagen		
Bone, 3-week-old chick	1/1000	359
Skin, human α2 chain	0.5/316 (CB6)	360
Skin, human α1 chain	0.5/204 (CB6)	361
Skin, human α2 chain	0.5/314 (CB5)	361
Type II collagen		
Sternal cartilage, chick embryo	2.2/1000	362
Cartilage, human/bovine	2/116 α1(II)–CB(9,7)	363
Type IV collagen		
Human placental tissue peptide	1/1000	364
Capsule, bovine	6–8/1000	365
EHS tumor, mouse	3/1000	366
Lens capsule, bovine	15/1000	367
Human placental peptides	3–8/1000	368
Glomerular basement membrane, human	11/1000	369
Type V/XI collagen		
Placenta, human	5/1000	370
Placenta	3/1000	370
Calvaria, human infant	5/1000	370
Epiphyseal cartilage, human infant	5/1000	370
Type X collagen		
Porcine cartilage	1/463	371
Unknown types	–	
Human aortic skin	7–10/1000	372
Swine kidney cortex	16/1000	373

Values are expressed as the 3-hydroxyproline content per analyzed collagenous polypeptides. CB means the CNBr fragments. Numbering of the CB peptides is written in Henkel and Dreisewerd[149], and Seyer *et al.*[374]

recognition by the enzyme.[379] From the experimental data of various peptides,[379,380] both the amino acid sequence around lysine and the length of the peptide chain are important determinants in the synthesis of Hyl by the enzymes. From the amino acid analysis data, the number of Hyl residues in type I collagen is about 20–25 residues. This is about one-third of the lysine in the Yaa position. The Hyl content varies in developmental and pathological states. In contrast with type I collagen, most Lys residues in the Yaa positions are hydroxylated in type IV,[368] V[149], and VI[381] collagens.

In vertebrates, three isoforms of the LH enzymes, LH1, LH2, and LH3, exist and all are expressed in the same cells.[382,383] LH2 has alternative splicing variants, a longer LH2b, and a shorter LH2a. All three catalyze the lysyl hydroxylation reaction of the collagenous domain, but only LH2 catalyzes the telopeptide Lys residues, which form intermolecular cross-links.[384] LHs contain about 700 amino acid residues and are composed of three domains, A, B, and C. The C domain contains the catalytic site. However, LH3 has recently been shown to be a multifunctional enzyme that possesses not only LH activity but also hydroxylysyl galactosyltransferase (GT) and galactosyl hydroxylysyl glucosyltransferase (GGT) activities *in vitro*.[86,383] LH3 has been shown to be localized not only within the ER but also in the extracellular matrix.[385] The active LH forms a homodimer. LHs do not have a typical four-residue ER retention signal but rather they have a much longer C-terminal sequence, which is essential for ER localization and which also contains the catalytic domain.[386,387] There are four potential N-linked glycosylation sites in LH. The glycosylation is required for maximum enzyme activity.[386,387]

The LH1 (PLOD1) knockout mouse shows aortic ruptures, abnormal gait, and early death.[84] The content of cross-links between collagen molecules is decreased in the null mouse, and the diameter of collagen fibrils in the aorta and skin is irregular and the fibrils are thicker as determined by electron microscopy. Mutations in LH2 cause an autosomal recessive disease called Bruck syndrome.[388] Biochemical analysis of the patient bone revealed that the Lys residues of the collagenous domain are hydroxylated normally, but the telopeptide Lys residues are not hydroxylated.[389] The LH3 (PLOD3) null mice are embryonic lethal. This is partially because LH3 is also the glucosyltransferase of Hyl (see the next section).

5.16.3.4 Glycosylation of Collagen

Some of the Hyl residues are modified to β-galactosyl hydroxylysine, and sometimes to 2-α-glucosyl-1-β-galactosyl hydroxylysine. The reaction is shown in **Figure 11**. This unique $1 \rightarrow 2$ type glycosidic linkage is only found in collagens. For example, type I collagen has about 25 residues of Hyl in the major collagenous domain. However, only two sites are glycosylated in normal tissues. In contrast, for type IV, V,[148,149] VI, and X[371] collagens, most of the lysine residues in the Yaa position are glycosylated. Lys residues of the collagenous domain of complement C1q are 83% posttranslationally modified to glycosyl hydroxylysine.[390]

One of the LHs, LH3, is a multifunctional enzyme with two catalytic domains. *In vitro*, this protein has three different enzyme activities, that is, procollagen-lysine 5-dioxygenase (lysyl hydroxylase) (EC 1.14.11.4), collagen galactosyltransferase (EC 2.4.1.50), and collagen glucosyltransferase (EC 2.4.1.66).[391,392] The observed *in vitro* glycosyltransferase activities are very low, suggesting that the enzyme needs some other chaperone molecules and/or strict structural requirements of the substrates. While the hydroxylation catalytic domain resides in the C-terminal C domain, the glucosyltransferase activity is at the N-terminal A domain.[391] The A domain is actually not required for the lysyl hydroxylation activity.[391] LH1 and LH2 have homologous A domains, but their glucosyltransferase activities have not been detected *in vitro*. LH3 null mice show an embryonic lethal phenotype due to abnormal basement membrane formation, especially nonspecific aggregation of type IV collagen molecules. The enzyme activity of glucosyltransferase was reduced by almost 85%, but the galactosyltransferase activity was not changed.[87]

Myllylä and coworkers produced three different mouse lines with a manipulated LH3 gene.[86] The mouse line (LH mutant) with a point mutation that specifically destroys the LH activity of LH3 without affecting glycosyltransferase activities was viable.[392] A complete loss of LH3 causes early embryonic lethality in mice.[86,87]

Recently, two galactosyltransferases, GLT25D1 (glucosyltransferase 25 domain containing 1) and GLT25D2, were identified as collagen hydroxylysine galactosyltransferase.[393] GLT25D2 is a type II transmembrane protein with four potential N-glycosylation sites. GLT25D1 is highly homologous to GLT25D2. GLT25D1 gene is widely expressed while GLT25D2 is expressed at low levels in the nervous system.[393] Since

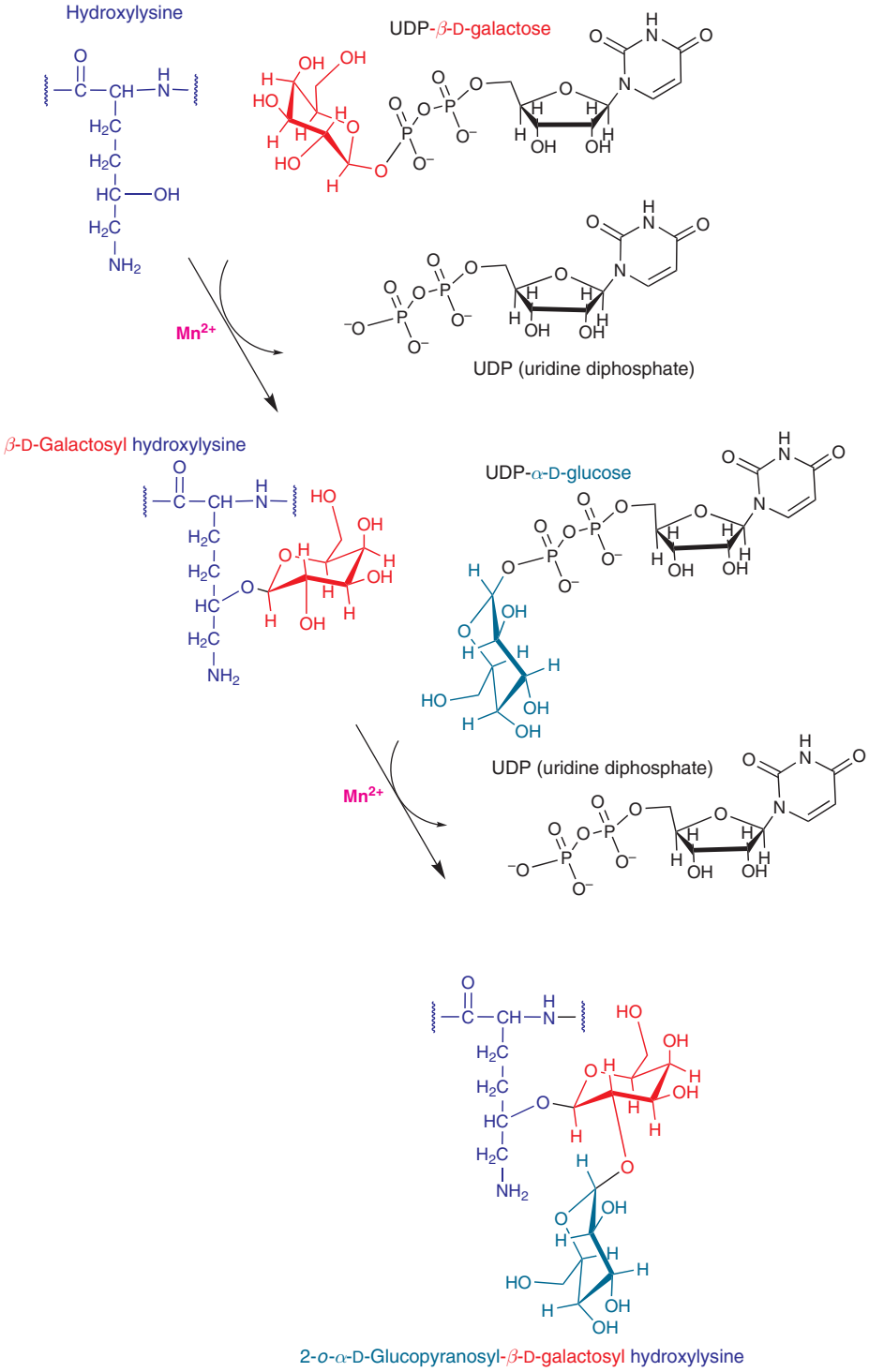


Figure 11 Reaction scheme of glycosylation of hydroxylysine.

the LH3 showed little glucosyltransferase activity,[392,399] it is suggested that the major enzymes of collagen galactosyltransferase might be GLT25D1/2 and the glucosyltransferase might be LH3.

Extensive analysis of cyanogen bromide fragmented peptides of type I, III, and V collagens by MALDI-MS has shown that type V collagen from bovine skin is highly glycosylated, while type I and III collagens contain

Table 4 Content of posttranslationally modified lysine residue in collagenous peptides

	Tissue	*Lys*	*Total Hyl*	*GH*	*GGH*	*Reference*
α1(I)	Placenta	26	9	1	1	394
α2(I)	Placenta	18	12	1	2	394
α1(V)	Skin	14	36	5	29	372
α1(V)	Skin		41	7	34	149
α2(V)	Skin	13	23	3	5	395
α2(V)	Skin		22	10	12	149
α3(V)	Placenta	15	43	6.5	17	125
Type VI	Placenta	17	62	7	48	223
α1(VII)	Skin	13	41	ND	ND	396
α1(X)	Cartilage	20	33	23	29	397
α1(XI)	Cartilage	19	37	2	25	398
α2(XI)	Cartilage	17	40	3.5	26	398
α3(XI)	Cartilage	15	21	4	11	398

Hyl, hydroxylysine; GH, galactosylhydroxylysine; GGH, glucosylgalactosyl hydroxylysine; ND, not determined.
All samples were prepared by pepsin treatment.
Data shown are the values per 1000 residues of pepsin-resistant collagenous domain.

only a few glycosylated residues[149] The MALDI-MS analyses have shown that almost all the Hyl residues in the α1(V) and the α2(V) chains from bovine skin are glycosylated.[148] Compared with the α1(V) chain, the ratio of the disaccharide units to monosaccharide is significantly lower in the α2(V) chain.[149] While there are many common sequences in type I and V collagen α chains, it is still unclear how glycosylation is regulated.

Table 4 summarizes the contents of galactosyl- and glycosylgalactosyl hydroxylysine residues in several types of collagen.

So far, the function of lysyl hydroxylation and glycosylation of Hyls is not clear. In some diseases such as OI, an overmodification of lysine residues occurs. It has been postulated that the triple helix formation is slower in these cases and therefore the unfolded chains are exposed for a longer time to the modifying enzymes.

Adiponectin is a solubule multimeric peptide hormone with demonstrated antidiabetic, antiatherogenic, and anti-inflammatory properties.[400] Serum adiponectin forms trimer, hexamer, and higher molecular weight multimers. Adiponectin has four glucosylgalactosyl hydroxylysine residues in its collagenous domain. The mutation of these lysine residues to arginine is the cause of the inability of the adiponectin to form multimers, which are essential for its physiological role.[401]

5.16.3.5 Disulfide Bond Formation and Chaperone Activities

Disulfide bond formation within the individual propeptides precedes folding and trimers are then formed by association of the C-terminal propeptides.[402] Disulfide bonds between the chains are then formed and this formation is most likely catalyzed by PDI.[403] As triple helix formation proceeds, the rate-limiting step in this process is the cis–trans isomerization of peptidyl-Pro bonds. This process can be catalyzed by peptidyl-prolyl cis–trans isomerases (cyclophilins and FKBPs). This activity is required to convert the proline residues to the trans form required for triple helix formation.[404,405]

Since procollagen molecules are only marginally stable, it has been proposed that folding of procollagen molecules inside cells requires special chaperones,[406] with HSP47 and FKBP65[407] as potential candidates. The collagen-specific chaperone HSP47 is known to be required during collagen biosynthesis.[408,409] Its functional importance as a collagen chaperone is shown by the embryonic lethality of HSP47 knockout mice.[89] Almost all proteins in ER are glycosylated. While type I collagen is thermally unstable at body temperature,[20] carbohydrates may stabilize the collagen helix. It has been well known that polyols stabilize the collagen triple helix and also inhibit the lateral aggregations.[410,411] The reason for the thermal stability and the prevention of the lateral aggregates of collagen in the ER may be partially due to the highly glycosylated environment. Additionally, collagen molecules have a strong tendency to aggregate and ultimately form insoluble complexes that stabilize

connective tissues such as tendon, bone, and skin. This process must be inhibited inside the cells and allowed only after secretion into the extracellular space. HSP47 travels from the ER to the cis-Golgi with procollagens whereby it dissociates and is recycled to the ER.[412] Additional predicted function of PDI is the inhibition of the aggregation of the collagen molecules in cells. Type X collagen interacts transiently with PDI in cells.[413] Also, type I procollagen interacts with PDI in ER.[414,415] Accumulation of BiP/GRP78 has been shown in cells with mutations in the collagen chains.[416] Premature association between procollagen chains is thought to be prevented by chaperones such as PDI, BiP/GRP78, GRP94, FKBP65, and HSP47, and collagen-modifying enzymes until the biosynthesis of the individual chain is completed. Proteins cannot exit the ER until they have achieved their correctly folded conformation.[417]

The chain selection and association of monomeric procollagen chains during folding is determined by the C-terminal propeptides in fibrillar collagens. Nucleation of the procollagen monomers occurs at the C-propeptide and triple helix formation proceeds in the C-to-N direction for type I collagen and most likely the other fibrillar collagens as well.[418,419] The interaction of the three procollagen polypeptides and initial folding of type I collagen might occur close to the ER membrane.[19,420,421] Some membrane-bound molecules might keep the procollagen α chains close to the ER membrane.

5.16.3.6 Procollagen Transport

Transport of procollagen molecules from the ER to the Golgi was shown to occur in transport complexes and it requires the action of both COPI and COPII.[422–424] Vesicular tubular clusters (VTCs) containing folded procollagen molecules are thought to bud directly from the ER membrane en route to the cis-Golgi.[425,426] Transport of procollagen occurs through the Golgi stacks without ever leaving its lumen and the Golgi mediates the lateral association of procollagen into bundles.[422,427] Golgi distensions occur during transport due to the long length of the procollagen molecules and changes in their supramolecular organization.[428,429] As bundles of procollagen molecules are released from the trans-Golgi they form secretory vacuoles known as Golgi-to-plasma membrane carriers (GPCs).[428,430]

5.16.3.7 Processing of Collagen Molecules

Finally, all of the classical fibrillar collagen molecules (I, II, III, V, and XI) are cleaved by enzymes to make mature tissue forms. Among the other types of collagens, types IV, VII, XV, and XVIII, and membrane-bound collagens (XIII, XVII, and XXV) have some processed forms in tissues. Type IV collagen from bovine lens capsule has a significant amount of the 20 kDa short α1(IV) chain IV collagen with an intact NC1 domain.[431–433] The type VI collagen C5 domain of the α3(VI) chain in cartilage is immediately processed after secretion,[434] and type VII collagen was reported to be cleaved by BMP-1.[246] These processing events allow more options for the proteins to organize three-dimensionally and also allow for time-dependent modifications of the extracellular matrix. In addition, the cleaved short fragments of collagen molecules can also work as signals to cells. Whole families of proteinases in the extracellular space or on cell membrane may work as regulators of extracellular matrix during development, growth, and injury response.

Removal of the N- and C-terminal propeptides from fully folded procollagens occurs only after transport of procollagens across the Golgi stacks and results in collagen molecules that are then able to assemble into fibrils.[177,423] C-proteinase activity is possessed by members of the tolloid family of zinc metalloproteinases, BMP-1, mTLD, and tolloid-like 1 (TLL-1).[435–438] Members of the ADAMTS family are responsible for the N-terminal processing of procollagens.[439] ADAMTS-2 acts on procollagens I, II, and III, whereas ADAMTS-3 and ADAMTS-14 have been linked only with procollagen II and I processing, respectively.[440–442]

Most collagen molecules are cleaved by processing enzymes to form a final tissue form for deposition into the matrix. These processing steps can regulate the solubility and the affinity of the molecules. Mutations in ADAMTS-2, a procollagen N-proteinase of type I, II, and III collagens, can cause the recessive disease EDS VIIC.[443] The skin of the ADAMTS-2 gene knockout mouse is fragile, and male mice are sterile although females are fertile.[444] Thus, neither ADAMTS-3 nor ADAMTS-14 can compensate for the function of ADAMTS-2. In bovine, this mutation causes a disease called dermatosparaxis. For more details, see the recent review by Greenspan *et al.*[445]

5.16.3.8 Lysyl Oxidation

Some lysine and Hyl residues are modified to allysine. This reaction is catalyzed by lysyl oxidase, also known as LOX (EC 1.4.3.13). At least four more proteins, LOX-like 1–4, are also suggested to have a similar function to LOX. LOX needs two cofactors, Cu^{2+} and a unique covalently integrated organic cofactor identified as lysine tyrosylquinone (LTQ).[446] LTQ is autocatalytically derived from a specific tyrosine (Y349) and a specific lysine (K314) residue within the nascent enzyme (**Figure 12**). This modification is believed to occur extracellularly. The ε-amino group of Lys/Hyl residues is catalyzed by oxidative deamination (**Figure 12**).

The modified aldehyde group can spontaneously condense with vicinal peptidyl aldehydes or with ε-amino groups of peptdidyl lysine. An example of such a cross-link found in collagen is shown in **Figure 12**.

5.16.4 Collagen Peptides as a Model of the Triple Helix

Synthetic peptides have been extensively used to study the thermal stability and folding of the triple helix. These peptides can be synthesized as either single chains or cross-linked peptides. Early on, such peptides were synthesized by polycondensation of tri- or hexapeptides, which led to a broad mass distribution that was difficult to separate. With the advances of solid-phase synthesis methods, peptides with defined chain length became available. The most studied collagen-like peptides are $(Gly\text{-}Pro\text{-}Pro)_n$ or $(Pro\text{-}Pro\text{-}Gly)_n$ and $(Gly\text{-}Pro\text{-}4(R)Hyp)_n$ or $(Pro\text{-}4(R)Hyp\text{-}Gly)_n$ with n varying from 5 to 15. Sutoh and Noda[447] introduced the concept

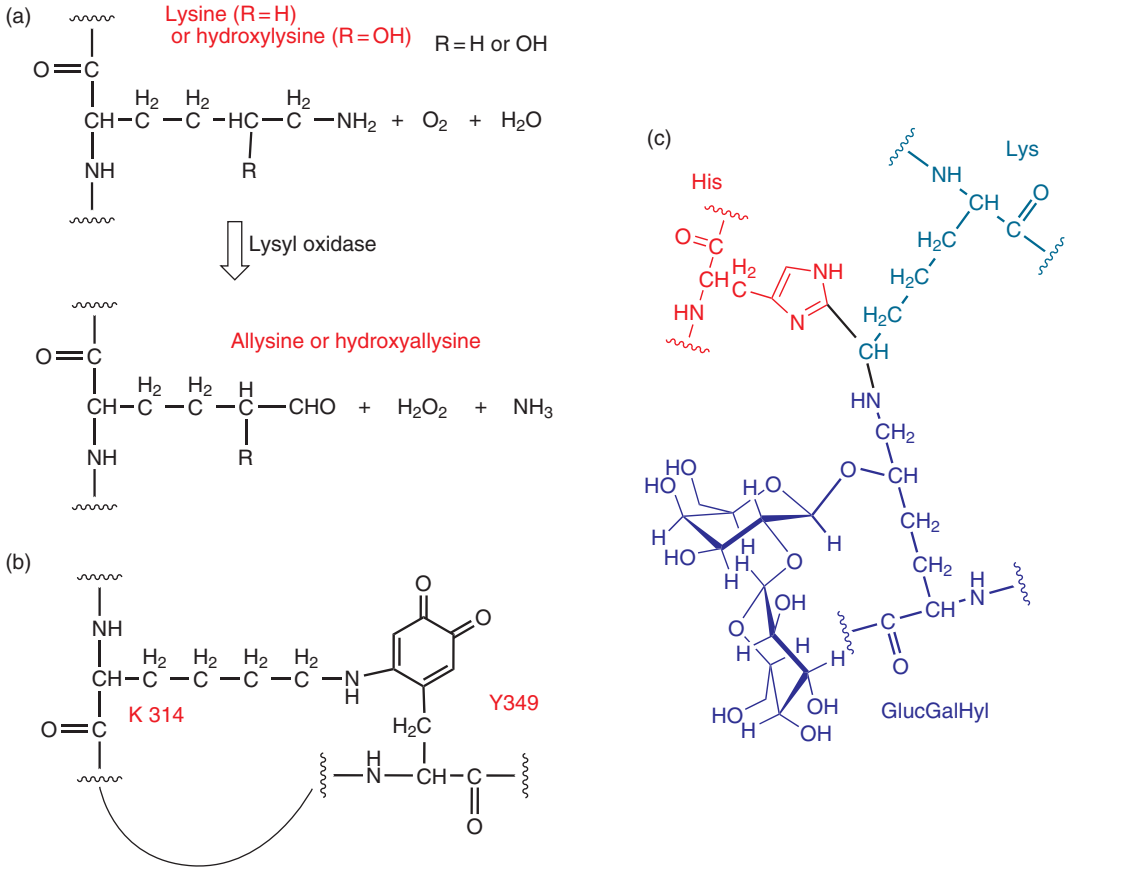


Figure 12 (a) Lysyl oxidation reaction, (b) cofactor of lysyl oxidase, lysyl tyrosine quinone (LTQ), and (c) an example of the intermolecular cross-link, histidinyl glucosylgalactosyl hydroxylysinorleucine.

of block copolymers, where two blocks of $(Pro\text{-}Pro\text{-}Gly)_n$, where $n = 5$, 6, or 7, were separated by a block of $(Ala\text{-}Pro\text{-}Gly)_m$, where $m = 5$, 3, and 1, respectively. This concept was later extended to include all amino acids and was called host–guest peptide system. The most studied host–guest system uses the sequence acetyl-(Gly-Pro-4(*R*)Hyp)$_3$-Gly-Xaa-Yaa-(Gly-Pro-4(*R*)Hyp)$_4$-Gly-Gly-NH_2.[448,449] Such peptides established the stabilizing role of 4(*R*)Hyp, the contribution of individual amino acids to the stability of the triple helix, and a thermodynamic and kinetic analysis of the triple helix–coil transition. These aspects have been recently reviewed[450,451] and in this chapter we will focus on the synthesis and analysis of cross-linked collagen model peptides.

5.16.4.1 Natural Cross-Linking

5.16.4.1.1 Collagen type III C-terminal cystine knot

Native collagens are molecules that are difficult to handle because of their high tendency toward aggregation and denaturation. If tissue-extracted collagens are denatured *in vitro*, the presence of interchain disulfide bonds is crucial for efficient and fast refolding into the triple helical structure.[108,452] In the absence of these interchain crosslinks or other trimerization sites, refolding of denatured molecules is very slow and incomplete, and mismatched triple helices differing in length and stability are formed. Following this observation, early studies of the kinetics and mechanisms of triple helical folding have been performed with the COL1–3 fragment of procollagen type III and the mature collagen type III. The COL1–3 fragment comprises the entire N-terminal precursor-specific region of procollagen III and contains a cystine knot resulting from interchain cross-linking of two cysteine residues separated by a dipeptide sequence. Mature collagen type III, on the other hand, contains a C-terminal cystine knot derived from two vicinal cysteine residues located at the C-terminus of the triple helix.[453] Using these natural forms of collagen, the C-to-N propagation of the triple helix in a zipper-like mode as well as the effects of the cis-to-trans isomerization of the aminoacyl-proline/hydroxyproline bonds on the folding kinetics were characterized.[108, 452, 454] The natural way to covalently cross-link three chains in the triple helix was successfully used to markedly reduce the unfavorable entropy of self-association of single-stranded collagenous peptides and thus to stabilize the triple helix.[455–460] These were all done with the use of the C-terminal cystine knot of collagen type III with the sequence GPCCGG. The interchain cystine knot of the N-terminal propeptide COL1–3 of collagen type III has so far not been applied to the oxidative assembly of synthetic collagen trimers. Attempts to oxidatively form the cystine knot in a recombinant peptide that corresponds to the native C-terminal sequence of collagen type III, but with the hydroxyproline residues replaced by proline (i.e., $GS(GXY)_{13}GPCCGGG$) failed because of the low triple helix propensity of the natural sequence.[459] However, using the obligatory trimeric molecule minifibritin as a fusion protein, which dictates trimerization of this collagen fragment, formation of the cystine knot was readily achieved at low temperatures in yields over 50%. Since at room temperature mainly cross-linked dimers and monomers were obtained, the necessity of the triple helix conformation with properly aligned chains prior to the oxidation of the cysteine residues was assessed. This observation fully agrees with the results of another group by analyzing the synthetic collagen molecules Ac-$(POG)_n$PCCGGG-NH_2, where n varied from 3 to 7.[461] With $n = 3$, self-association into a triple helical structure was not achieved according to the circular dichroism (CD) spectra, even after incubation at low temperatures and at 1 m mol l^{-1} concentration in order to entropically control the process. As a consequence, oxidation was found to generate a mixture of products, in which the desired trimer could not be detected by MS. On the other hand, preequilibration of the peptide with $n = 5$ at low temperature (7–8 °C) and at 1 m mol l^{-1} concentration yielded a CD spectrum that reflects a high content of triple helix ($T_m = 20.3$ °C). From the oxidation experiments at this concentration, at pH 8, and at a temperature far below the T_m value (i.e., at 7−8 °C), air oxidation was found to generate a product distribution consisting on an average ≈70% trimer. The remaining product was essentially oxidized monomer. Two case-limiting mechanisms for oxidative folding that allow for structural interpretation (i.e., the folded precursor and the quasistochastic mechanism) have emerged from extensive studies on cysteine-rich peptides and proteins.[462] In the folded precursor mechanism, local regions of the polypeptide chain adopt native-like structures, which are locked in by disulfide bonds; a piecemeal accumulation of such local microfolding events leads to global folding. In the quasistochastic mechanism on the other hand, a set of disulfide bonds form initially following the proximity rule for loop formation in early fast oxidation and reshuffling steps. This conformation is then locked

in and protected from further rearrangements by local or global conformational folding. While this second mechanism applies well to most of the folding pathways of globular proteins, formation of the type III collagen disulfide knot, at least in model peptides, follows the folded precursor mechanism as it requires a proper spatial alignment of all the cysteine residues in the triple helix.

Liquid chromatography (LC)–MS analysis of the product mixture of oxidized Ac-$(POG)_n$PCCGGG-NH_2 revealed the presence of a single trimer and even the NMR spectra were consistent with a single set of disulfide connectivities; prefolding of the collagenous monomeric peptide, followed by oxidation, leads to the formation of a well-defined cystine knot isomer.[461] However, it was not possible to derive the exact cysteine pairings from NMR structural analysis.[461] There are eight possible ways to connect the six cysteines to form the cystine knot. Two models were previously suggested based on steric compatibility.[452,461] The three collagen chains are designated A, B, and C, where chain A has a one-residue stagger toward the N-terminus, followed by the B chain and, finally, the C chain (**Figure 13**). Based on this type of stagger, the first model proposed by Bruckner *et al.*[452] has the connectivity A1–B1/A2–C1/B2–C2. The second model proposed by Barth *et al.*[461] has the connectivity A1–B2/A2–C1/B1–C2. The two models share the A2–C1 disulfide bridge yet differ in the other disulfide connectivities. Recent advances in solving the crystal structure of human type III collagen fragment G991–G1032 containing the cystine knot allowed to unambiguously establish only one disulfide connectivity, that is, A1–B2,[2] which is consistent with the second model (**Figure 13**). Another possibility is a model with the connectivity A1–B2/A2–C2/B1–C1 (**Figure 13**). Further analysis is required to establish which one is correct. The poor electron density that was observed in the cystine knot region might be due to the flexibility of the last residues of the polypeptide chain. Indeed, there are no crystal contacts to the C-terminal region, which could stabilize these residues.[278] Extending the C-terminal region by a few more residues that belong to the telopeptide might solve this problem although it might affect crystallization.

Although the collagen type III C-terminal cystine knot naturally localizes on the very C-terminal end of the triple helix, its use on the N-terminal end of short model peptides also appeared to be successful.[458, 461] This approach was initially applied for recombinant constructs containing collagen peptides such as $GS(GPP)_{10}$GPPGPCCGGG[457] and GSYGPPGPCCSGPP$(GPP)_{10}$.[458] The oxidation was performed in the presence of GSH/GSSG (9:1) at 20 °C (i.e., at a temperature slightly below the triple helix melting point of $(GPP)_{10}$ ($T_m = 25$ °C). The trimers were formed in good yields and both showed a significant increase of the thermal stability, with the C-terminal knot being more efficient ($T_m = 82$ °C) than the N-terminal disulfide cross-link ($T_m = 67$ °C). Since the refolding rates of both peptides ($0.000\,12\ s^{-1}$ for the C-terminal and $0.000\,37\ s^{-1}$ for the N-terminal cystine knot peptide) are very similar, triple helix formation can be nucleated at both ends. The rate-limiting step is still represented by the cis-to-trans isomerization as supported by the very similar activation energy, which is $>50\ kJ\ mol^{-1}$ for both trimers.[457, 458] These results confirmed that N-terminal triple helix nucleation with N-to-C propagation can occur as efficiently as from the C-terminus. It also explains the efficient *in vivo* folding of membrane-associated collagens where nucleation and triple helix folding were proposed to start at the N-terminus.[463,464]

In native collagen, all Gly-Pro and Xaa-Hyp peptide bonds are in the trans conformation, whereas in the unfolded state, a significant fraction of cis isomers populates at each Gly-Pro and Xaa-Hyp peptide bond. cis-to-trans isomerization reactions at prolyl peptide bonds are the origin for the observed slow kinetics of triple helix formation[450] as shown by their high activation energy ($\approx 72\ kJ\ mol^{-1}$)[454] and their acceleration by prolyl

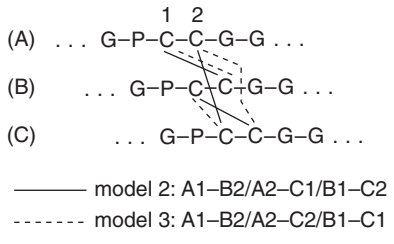


Figure 13 Possible collagen type III C-terminal cystine knot conformations.

isomerases.[404] Another rate-determining step in collagen triple helix formation is to bring the individual chains into correct register.[457, 465, 466] To gain information on the dynamics of triple helix formation in the absence of these slow reactions, stopped-flow double-jump experiments were performed on cross-linked trimeric fragments of type III collagen containing the natural cystine knot.[460] Cross-linked collagens were rapidly unfolded at high guanidinium chloride concentrations in a stopped-flow instrument and the refolding was initiated by diluting out denaturant before prolyl isomerization could occur (double-jump experiments). The refolding from collagen molecules with all Xaa-Pro peptide bonds in their native trans isomerization state was studied. The results showed that triple helix formation occurs with a rate constant of $113 \pm 20\,s^{-1}$ at 3.7 °C and is virtually independent of temperature, indicating a purely entropic barrier. Comparison of the effect of guanidinium chloride on folding kinetics and stability revealed that the rate-limiting step is represented by bringing 10 consecutive tripeptide units (3.3 per strand) into a triple helical conformation. The following addition of tripeptide units occurs on a much faster timescale and cannot be observed experimentally. These results supported an entropy-controlled zipper-like nucleation/growth mechanism for collagen triple helix formation.

5.16.4.1.2 Cystine knots of FACIT collagens

FACIT collagens contain short triple helical domains interdispersed with nontriple helical domains. Collagen type IX is composed of three different α chains, while all others are homotrimers. Unlike the fibril-forming collagens, the FACITs have significantly shorter C-terminal NC domains, but share a remarkable sequence homology in the first collagenous domain (COL1) with two imperfections within the triple helix. Moreover, the FACITs contain in their COL1/NC1 junctions two strictly conserved cysteine residues separated by four residues: $(GXY)_nGxCxxxxC$. In the folded molecules, these cysteine residues form an interchain disulfide knot of unknown disulfide connectivities.[467] The most detailed investigation of the intrinsic propensity of the highly conserved cystine knot sequences of FACITs for oxidative trimerization *in vitro* was done by van der Rest and coworkers.[468] Two synthetic model peptides derived from the COL1/NC1 junction of collagen type XIV were used: GYCDPSSCAG (peptide I) and $(GPO)_3$ GYCDPSSCAG (peptide II). Peptide II was N-terminally extended with three (GPO) repeats to replace the GSQGPAGPO sequence of type XIV collagen. All oxidative experiments performed with peptide I failed to produce the homotrimer in satisfactory yields. The majority of the oxidized product was the intrachain disulfide-bridged monomer. Conversely, oxidative refolding of peptide II was found to produce a variety of species: monomers, dimers, and trimers at ratios strongly depending upon the experimental conditions. At room temperature, the intrachain disulfide-bonded monomer was the predominant product, whereas the trimer formation was negligible. The significant yield of trimeric species (up to 55% of the product mixture) was achieved at a temperature of −20 °C and with 30% MeOH in the buffer (pH 8.5). Low temperature and presence of alcohols are known factors for stabilizing the collagen triple helix. Dilution of the peptide below $1\,mg\,ml^{-1}$ had a negative effect on trimer formation (i.e., by increasing the entropic penalty of self-association of the three peptide chains). Therefore, trimerization into the triple helix register via the $(GPO)_3$ repeats is apparently required for oxidative formation of the disulfide knot. The triple helical structure of the trimer as isolated product was confirmed by the CD spectrum and NMR structural analysis. [468] NMR conformational analysis of the trimer also confirmed the formation of a cystine knot although an unambiguous differentiation between two models could not be made (**Figure 14**).

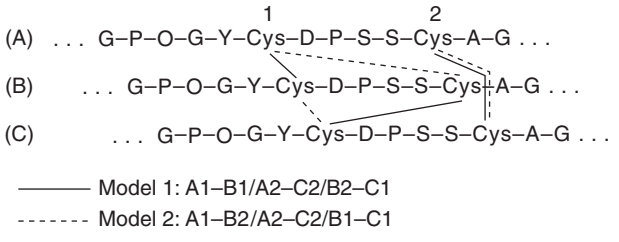


Figure 14 Possible FACIT cystine knot conformations based on NMR data of type XIV collagen cystine knot-containing peptide.

Another successful example of exploiting FACIT cystine knot was made on type XII collagen. A minichain consisting of COL1 and the N-terminal portion of the NC1 domain has been successfully self-assembled into a triple helical homotrimer stabilized by the cystine knot.[469]

The most interesting and practically important way would be to form a heterotrimeric FACIT cystine knot found in type IX collagen. It plays a dual function by stabilizing the triple helix of three different chains and dictating the stagger. A naturally derived low-molecular-weight pepsin-resistant fragment of the heterotrimeric type IX collagen was used as a model for oxidative *in vitro* experiments.[470] The fragment consisted of COL1 and the N-terminal part of NC1. Upon reduction and isolation of the three truncated chains, they were individually analyzed for their *in vitro* oxidative self-association into homotrimers, and as stoichiometric mixture of all three chains into a heterotrimer. All three chains are able to form triple helical homotrimeric reassociation products. A mixture of the α1, α2, and α3 chain in the presence of reduced and oxidized glutathione leads to an oxidized product distribution consisting of monomers, homodimers and homotrimers of the α1 and α2 chains, homodimers and homotrimers of the α3 chain, which are formed to much lesser extents, and heterotrimer $\alpha1\alpha2\alpha3$. The total amount of disulfide-bonded trimers did not exceed 25%. The low yield of cross-linked trimers and their heterogeneity indicated that COL1/NC1 junction is insufficient to guide the quantitative heterotrimer formation. Very recent results demonstrated that full-length COL1 and NC1 domains are not required for trimerization of type IX collagen.[471]

5.16.4.2 Artificial Cross-Linking

To synthesize three collagen peptide chains together, several chemical knots were developed. Moroder and colleagues adopted a cystine knot with a stagger with three different chains.[46,47] This method can be applied to make a heterotrimeric type I collagen peptide consisting of residues 772–784/785 of $[\alpha1(I)]_2\alpha2(I)$ for the substrate of MMP-8[46] and MMP-1.[472] Fields *et al.*[473] and Lauer-Fields *et al.*[474] used a Lys-linked knot at the C-terminal end to synthesize type I collagen α1(I) 772–786 and type II collagen α1(II) 772–783 sequences. Some of the other knot/nucleation domains used are summarized in **Figure 15**. The requirements of the ideal knot are the stability in the wide range of pH, solubility in aqueous solvents, little strains of the three staggered peptides, and the simple and quick reaction to connect three different sequence peptides. So far, all methods have some difficulties. Each C-α of Gly should form an isosceles triangle 4 Å apart with 112° angle to align the staggered collagen helix.

5.16.5 Collagen Chain Selection, Trimerization, and Triple Helix Formation

Although short collagen model peptides are able to form the triple helix without any additional trimerization domain,[480] it becomes increasingly clear that full-length collagen molecules require specialized domains for chain selection, trimerization, and subsequent triple helix formation. Only 30 years ago, it was still believed that collagens (at that time only few types of collagen were known) consist of a triple helix and short nontriple helical peptides located on both N- and C-terminal ends, the so-called N- and C-telopeptides.[451] Early experiments on collagen refolding suffered from poor yield, chain misalignment, and formation of wrong products, simply due to the use of collagen extracted from tissues (mainly type I collagen). This form of collagen represents a processed species ready for fibril formation. It lacks the so-called N- and C-propeptides, NC domains situated on both N- and C-termini, which are important in chain selection, alignment, and initiation of the triple helix formation. The presence of propeptides also prevents fibril formation. Several steps might be observed during the folding of collagen: synthesis of a procollagen molecule and modifications of residues (proline hydroxylation, lysine hydroxylation); chain selection, alignment, and trimerization by a specialized domain (or domains); triple helix formation; formation of intra- and interchain disulfides; cleavage of propeptides; and fibril formation. In this chapter, we review advances in studying specialized domains involved in chain selection and trimerization as well as triple helix formation, steps required to build a folded molecule that is not yet involved in the formation of the higher order structures.

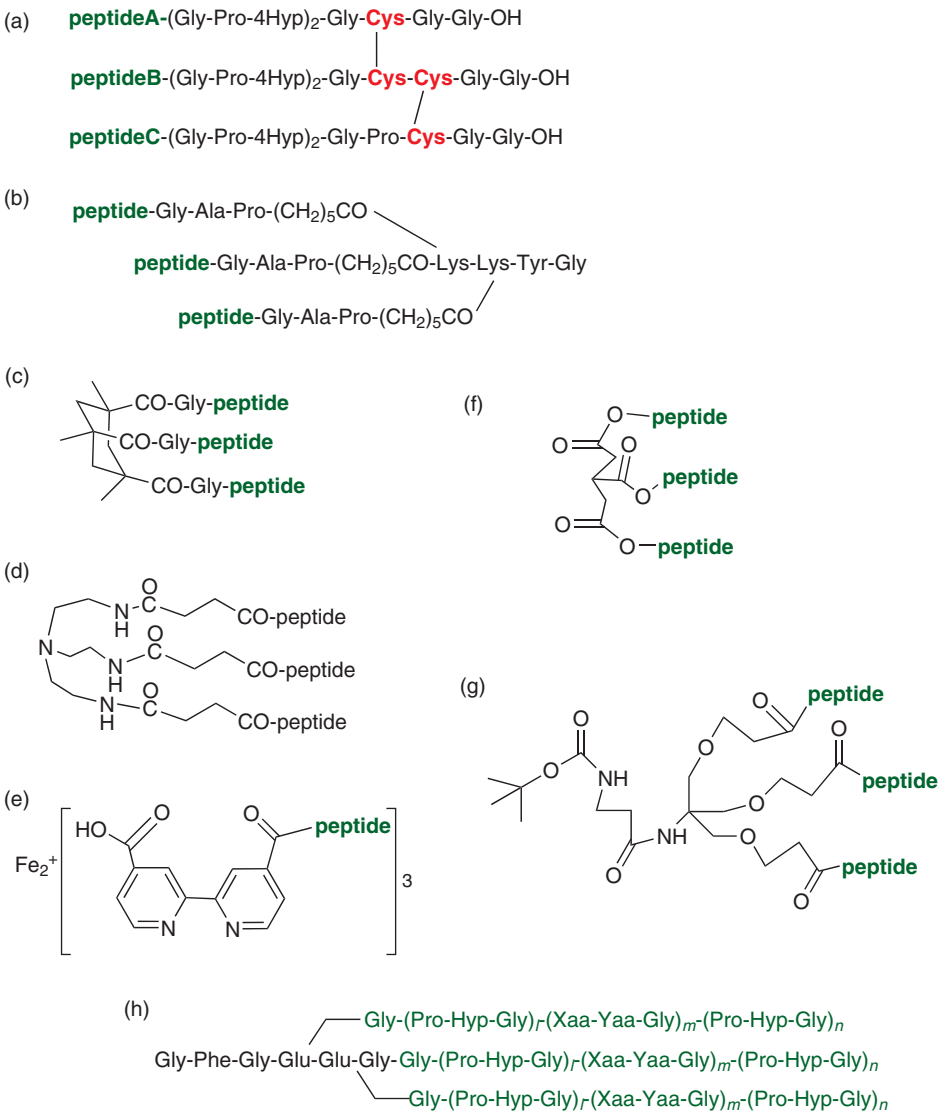


Figure 15 Examples of artificial collagen model peptide knots. (a) Regioselective artificial cystine knot,[46] (b) dilysine scaffold,[473] (c) *cis,cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid 'Kemp tri acid',[475] (d) TREN-(suc-OH)3, (Tris(2-aminoethyl)amine) succinic acid,[476] (e) Fe^{2+} [2,2′-bipyridyl-peptide]3,[477] (f) monoalkyl chains,[478] (g) Boc-β-Ala-TRIS-$(OH)_3$ knot, and (h) N-terminal GFGEEG link.[479]

5.16.5.1 Chain Selection and Trimerization

For many collagens, trimerization domains are found at the C-terminus, leading to the idea that triple helices may only or preferentially fold from the C-terminus. It was also suggested that a three-stranded coiled coil domain at the N-terminus of type XIII collagen can induce collagen folding from the N-terminus.[315] Similar suggestions have been made for type XVII,[463] XXIII,[325] and XXV[326] collagens. FACIT collagens are the most intriguing in deciphering the folding mechanism since they have a very short C-terminal domain, which is not sufficient for trimerization.[468,481] A potential role of the NC2 domain (the second NC domain counted from the C-terminal end) in the trimerization was suggested based on an α-helical coiled coil sequence prediction[482] and the experimental proof came very recently for type XIX collagen.[278]

5.16.5.1.1 Fibril-forming collagens *(types, I, II, III, V, XI, XXIV, and XXVII)*

All of the 12 fibril-forming α chains share a long uninterrupted collagenous domain flanked by N- and C-terminal NC propeptides. The α chains assemble into at least 12 type-specific protomers, characterized as homo- and heterotrimers. The chains of fibrillar collagens associate first through a series of noncovalent interactions between the C-terminal NC domains (NC1), which provide correct alignment and registration for the nucleation of triple helix.[108,189,483,484] The terminal propeptides are further cleaved by specific proteases that promote oligomerization and fibril formation.

5.16.5.1.2 Network-forming collagens *(types IV, VIII, and X)*

In contrast to the fibril-forming collagens, the C-terminal NC domains responsible for trimerization are retained in their suprastructures. Importantly, crystal structures of the trimerization domains are known (see Section 5.16.6.5). Type IV collagen has provided an opportunity to gain insights into chain recognition mechanisms. The six homologous α chains assemble into three specific protomers out of 76 possible combinations, reflecting a remarkable specificity for chain selection *in vivo*.[485] The monomeric C-terminal NC domains play a critical role in trimerization *in vivo*[486] and could also recognize each other and reassemble *in vitro* into their original trimeric and hexameric compositions.[487] The crystal structure of the C-terminal NC hexamer is known[193,194] and reveals the nature of intermolecular interactions and their role in chain selectivity.

Type VIII collagen is composed of highly conserved α1(VIII) and α2(VIII) chains, which contain a short collagenous domain of 454 residues flanked by an N-terminal NC domain of 117 residues and a C-terminal NC domain of 173 residues.[488] The chains assemble into two distinct homotrimers that assemble into hexagonal lattices.[259,260] The crystal structure of the α1 C-terminal NC homotrimerization domain was determined[266] (see Section 5.16.6.5).

Type X collagen has a single α1 chain, which contains a short collagenous domain of 460 residues, flanked by an N-terminal NC domain of 37 residues and C-terminal NC domain of 161 residues.[489] Its protomer is a homotrimer. The C-terminal NC domain is responsible for trimerization and multimer formation, which is based on experiments with recombinant domains.[490] The crystal structure of type X collagen trimerization domain[265] is similar to that of type VIII collagen[266] (see also Section 5.16.6.5).

5.16.5.1.3 FACIT collagens (types IX, XII, XIV, XVI, XIX, XX, XXI, and XXII)

In contrast to the fibril-forming collagens, the FACITs have significantly shorter C-terminal NC domains: 75 residues for type XII collagen and fewer than 30 residues for collagen IX, whereas those of fibrillar collagen are about 260 residues.[491] The FACITs share a remarkable sequence homology at their COL1/NC1 junctions by having two strictly conserved cysteine residues separated by four residues at the NC1 domain. Those cysteines form interchain disulfide bonds and thus covalently cross-link all three chains. Experiments on folding of COL1/NC1 regions of type XII and XIV collagens showed that the correct interchain cystine knot formation depends on the presence and ability to form the triple helix of COL1 domain or artificial collagen-like sequence.[468,469] Moreover, only the N-terminal part (about seven residues) is required for the disulfide knot formation.[469] Very recently, it was shown that trimerization of type IX collagen α chains does not require the presence of COL1 and NC1 domains at all.[471] The trimerization and the collagen triple helix stabilization properties of the NC2 domain of type XIX collagen were demonstrated recently.[278] Given that the NC2 domain plays the same role in other FACITs, especially in type IX (which is a heterotrimer), we can conveniently produce any recombinant heterotrimeric collagen fragment with desired chain selectivity.

5.16.5.1.4 Beaded filament and anchoring fibril collagens (types VI and VII)

The α1 and α2 chains of type VI collagen are similar in size and contain one N-terminal NC domain and a C-terminal NC domain with two subdomains, C1 and C2. In contrast, the α3 chain is much longer, and the N-terminal NC domain contains 10 subdomains, N1–N10, and 5 C-terminal NC subdomains, C1–C5.[227] The protomer is an α1α2α3 heterotrimer. Deletion studies have demonstrated that whereas the C5 subdomain of the α3 chain is required for the extracellular microfibril formation, the C1 subdomain, in all chains, is sufficient for chain recognition and protomer assembly.[227,492]

Type VII collagen is composed of three identical α chains, each consisting of a 145 kDa central collagenous triple helical segment, flanked by a large 145 kDa N-terminal NC domain (NC1) and a smaller 34 kDa

C-terminal NC domain (NC2).[493] The deletion of the N-terminal NC domain (NC1) did not abolish the triple helix formation and subsequent association of two molecules into disulfide-bonded hexameric aggregates,[493] an intermediate of anchoring fibril formation driven by the NC1 domain.[493] Thus, most probably, the C-terminal NC domain (NC2) contains the trimerization properties.

5.16.5.1.5 Transmembrane collagens *(types XIII, XVII, XXIII, and XXV)*

The α chain of each type contains an N-terminal intracellular region, a transmembrane region, and a C-terminal extracellular region. The extracellular region is composed of multiple collagenous domains interrupted by NC domains. The protomer of each α chain is a homotrimer. The extracellular region contains NC domains with a high probability to form an α-helical coiled coil, the conformation of which is thought to prompt trimerization and subsequent zipper-like folding of the triple helical domain. Studies using deletion constructs have shown that such N-terminal NC domains of collagens XIII and XVII are necessary for triple helix formation,[463,464] suggesting that nucleation of the triple helix occurs at the N-terminal region and proceeds in the N- to C-terminal direction, which is opposite to that of all other classes of collagens. Later, for type XIII collagen, which has four NC domains, numbered 1–4 starting from the N-terminal end, it was shown that the potential α-helical coiled coil region in NC1 is necessary only for the folding and the stabilization of the membrane-proximal collagenous domain COL1, whereas the COL2 and COL3 domains are correctly folded and stabilized due to the NC3 domain, another potential α-helical coiled coil.[494] This is also predicted for the related type XXIII and XXV collagens.[494] It was recently suggested that the NC2 domain might also form an α-helical coiled coil and thus also participate in the folding and stabilization of the triple helix together with potential coiled coils in the NC1 and NC3 domains.[312] Unfortunately, no biophysical and structural experiments were done to prove the formation of the α-helical coiled coil and characterize its trimerizing and stabilizing properties for any transmembrane collagen so far.

5.16.5.1.6 Multiplexins *(types XV and XVIII)*

Collagens XV and XVIII are composed of a single α chain that contains a collagenous domain, with frequent interruptions, flanked by the N-terminal and the C-terminal (NC1) NC domains. The NC1 domain of both XV and XVIII forms trimers.[207,495] The trimerization properties of the NC1 domain were also demonstrated for an analogue of type XVIII collagen in *Caenorhabditis elegans* and *Drosophila melanogaster*.[496] It was shown that the NC1 domain is prone to proteolysis, which leads to a release of a monomeric form of the endostatin domain and the stable trimeric peptide.[495] The trimeric peptide was mapped to the N-terminal end of the NC1 domain and its length was estimated to be about 50 residues.[495]

5.16.5.2 Stabilization of Short Collagen-Like Triple Helices by Exogenous Trimerization Domains

The collagen-like peptide $(GlyProPro)_{10}$ could be substantially stabilized by the recombinant fusion to the foldon domain. Single $(GlyProPro)_{10}$ chains are able to form a collagen triple helix of low stability (melting temperature 24 °C at 20 μmol l^{-1} chain concentration) and of high dependence on concentration. The stability of the collagen peptide fused to the N-terminus of foldon dramatically increased to 75 °C and became concentration independent. Such high stability is achieved by the high intrinsic concentration of the C-terminal ends of collagen chains, which are held very close to each other by the foldon domain (at ~1 mol l^{-1} concentration).[497] This stabilization could also be explained in terms of entropy. The foldon domain substantially restricts the freedom of the unfolded collagen chains and aligns them. However, the foldon domain cannot adopt the staggered register of the collagen triple helix, which leads to a substantial kink between the helix and the foldon seen in the crystal structure[498] (see also Section 5.16.6.5.3).

5.16.5.3 Triple Helix Formation

For most collagens, the folding of the triple helical domain proceeds from the carboxyl end toward the amino end of the trimeric molecule in a zipper-like fashion with a rate that is limited by cis–trans isomerization of peptidyl prolyl bonds.[499] The fast propagation of the triple helix formation is followed by a slower folding

determined by cis peptide bond isomerization at proline and hydroxyproline residues. These peptide bonds need to be isomerized into trans conformation to allow triple helix formation to continue. The process can be facilitated by adding the enzyme peptidyl prolyl cis–trans isomerase (PPIase) as was shown *in vitro*.[404] An accelerating effect of PPIases on the folding of type III[404] and IV collagens[405] was only close to 2, which is very little compared to an acceleration factor of 100 or more for other proteins. The relatively small accelerations may originate from the use of enzymes that are not specific for collagens.

More detailed analysis of the triple helix folding was made using short collagen peptides. The very early studies were made on single-chain peptides that are able to form the triple helix by themself without any additional trimerization domain.[500,501] The folding of such peptides has a very complex mechanism, which varies from third- to first-order reaction depending on the amino acid composition and the peptide concentration.[457] A nucleation of the triple helix in such peptides makes it difficult to analyze the triple helix propagation. The invention of covalently cross-linked peptides and peptides carrying NC trimerization domains allowed to overcome the problem of the triple helix initiation and focus on the triple helix propagation.[457,458,460,461]

Pioneering kinetic work resembling the natural way of the collagen triple helix formation was performed on the long central triple helical domain of type III collagen, which consists of three identical chains and is terminated by a disulfide knot.[108] The C-terminal propetide provides a noncovalent link between the chains in the nonprocessed state but even after its removal the chains remain linked by the disulfide knot. The unfolding–refolding of this trimerized, purely triple helical domain was completely reversible and end products of refolding at 25 °C were identical to the native molecules as judged by their melting profiles, molecular weights, and sedimentation behavior. The growth of the triple helix was found to proceed from the disulfide knot at the C-terminus at a rather uniform rate in a zipper-like fashion.[108] From the temperature dependence of refolding, activation energies of 85 kJ mol^{-1} were derived, supporting that the propagation of the zipper is rate limited by cis–trans isomerization steps.[454]

5.16.6 Atomic Structures of the Collagen Triple Helix and Collagen Trimerization Domains

5.16.6.1 Collagen Triple Helix

Although there were early reports of crystallization of a cyanogen bromide fragment of collagen,[502] the collagen molecule itself has not proved amenable to investigations at the molecular level. The path to the molecular details of the collagen triple helix has been through collagen model peptides, which have yielded high-resolution X-ray structures and allowed NMR characterization of dynamic and conformational features. The fiber diffraction models of collagen left unresolved controversies about the nature of interchain hydrogen bonding, including the possibility of hydrogen bonds involving α carbon atoms, and the precise geometrical parameters of both the basic helix and the supercoil. The high-resolution crystal structures of peptides confirmed the conformation derived from fiber diffraction data of collagen, and resolved a number of long-standing controversies about hydrogen bonding and hydration. The molecular details also raise new issues for consideration, including variability in helix twist and the mechanism of hydroxyproline stabilization. Crystallization and structure determination of collagen peptides remain a challenging task. All available crystal structures are of either artificial mimics (like $(GPP)_n$ or $(GPO)_n$, where O = 4(*R*)-hydroxyproline) or host–guest peptides, where a short stretch of 1–3 native tripeptide units is flanked by 3–5 GPO repeats. Unfortunately, these structures give only limited insight into how side chains other than imino acid residues in the Xaa and Yaa positions contribute to collagen structure and stability. Nevertheless, an impressive progress was made in resolving the atomic structure of the collagen triple helix since 1950s when two structural models for collagen with different fiber periods and helical symmetries were proposed based on the fiber diffraction pattern of native collagen. These were the 7/2-helical model with a 20 Å axial repeat[503,504] and the 10/3-helical model with a 30 Å axial repeat.[503] These single-helical models were discarded after the proposal of a triple helical structure with 10/3-helical symmetry and a 28.6 Å axial repeat by Ramachandran and Kartha.[505] The first crystal structure of $(Pro\text{-}Pro\text{-}Gly)_{10}$ showed a triple helix with 7/2-helical symmetry with a 20 Å axial repeat.[506] Recently, the fiber diffraction analysis of native collagen was performed based on the advanced diffraction data acquisition techniques and revealed that the X-ray diffraction data can be explained not only by

Table 5 Crystal structures of collagen triple helices

Peptide	*Sequence*	*Number of residues*	*PDB accession number*	*References*
PPG9	$(PPG)_9$	27	ITT, 2CUO	41, 508
PPG10	$(PPG)_{10}$	30	1A3I, 1A3J, 1G9W, 1K6F	509–511
POG10	$(POG)_{10}$	30	1V7H	512
POG11	$(POG)_{11}$	33	1V4F, 1V6Q	512
$(GPP)_{10}$ foldon	GS $(GPP)_{10}$ foldon	30	1NAY	498
PPG9-POG	$(PPG)_4$ POG $(PPG)_4$	27	2D3F	514
PPG9-PaOG	$(PPG)_4$ P alloO G $(PPG)_4$	27	1X1K	513
PPG9-OOG	$(PPG)_4$ OOG $(PPG)_4$	27	2D3H	514
POG10-LOG1	$(POG)_4$ LOG $(POG)_5$	30	2DRT	514
POG10-LOG2	$(POG)_4$ LOGLOG $(POG)_4$	30	2DRX	514
GPO9-$G^{3(S)}$OO2	$(GPO)_3$ $(G3OO)_2$ $(GPO)_4$	27	2G66	354
GOO9	$(GOO)_9$	27	1YM8	515
G → A	$(POG)_4$ POA $(POG)_5$	30	1CAG, 1CGD	516, 517
Hyp-	$(POG)_4$ PG $(POG)_5$	29	1EI8	518
T3-785	$(POG)_3$ ITGARGLAG $(POG)_4$	30	1BKV	519
EKG	$(POG)_4$ EKG $(POG)_5$	30	1QSU	519
IBP	$(GPO)_2$ GFOGER $(GPO)_3$	21	1Q7D	520
IBP in complex	In complex with integrin	21	1DZI	42
Hug peptide	$(GPO)_4$ GPRGRT $(GPO)_4$ in complex with adhesin	30	2F6A	521
G982–G1023	Collagen III fragment G982–G1023 with disulfide knot	42	3DMW	2

the prevailing 10/3-helical model, but also by the 7/2-helical model.[507] At this time, the crystal structures of 20 different triple helical peptides have been reported (**Table 5**). Two of them are in complex: one with a bound integrin domain and another with an adhesin (**Table 5**). Almost all high-resolution structures of model peptides adopt a 7/2-helical symmtery and the conformation close to the 10/3-helix appears only in the guest region of host–guest peptides, like the T3-785 peptide and the integrin-binding collagen peptide in complex with integrin.[507] These two peptides can be viewed as having three zones: N-terminal Gly-Pro-Hyp repeats; a central collagen sequence; and C-terminal Gly-Pro-Hyp repeats. In T3-785, the two terminal Gly-Pro-Hyp regions show 7/2 symmetry, while the central Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala region is closer to 10/3 symmetry. In IBP, the terminal Gly-Pro-Hyp repeats have 7/2 symmetry and the central Gly-Phe-Hyp-Gly-Glu-Arg sequence is intermediate between 7/2 and 10/3. In each case, these three zones are slightly bent and twisted with respect to each other. It appears that the 7/2 symmetry is generated by the steric restrictions of repeating Gly–imino acid–imino acid units and is maintained when only one Gly-X-Y triplet is introduced (e.g., EKG peptide). The presence of two or three tripeptide units, where X and Y are not imino acids, starts to change the helix twist toward 10/3 symmetry. A very recent crystal structure of a native type III collagen peptide G991–G1032 with a cystine knot further confirms this finding. The peptide contains four consequent tripeptide units without imino acids, which adopt 10/3 symmetry, and the rest of the peptide has 7/2 symmetry. This is the first structure containing only the natural sequence (with one exception, one glutamine is replaced by selenomethionine for crystallographic phasing). Altogether, the results suggest that the collagen molecule, with its varied Gly-X-Y sequence, is likely to have a nonuniform helical twist along its length. It is likely that sequences poor in imino acids will have a symmetry close or equal to 10/3, while stretches of Gly-Pro-Hyp units may have 7/2 symmetry.

5.16.6.2 Hydrogen Bonding

Hydrogen bonding is a critical part of triple helix stabilization. The triple helix has repetitive backbone hydrogen-bonding networks, but differs from β sheets or α helices in that the repeating tripeptide unit consists

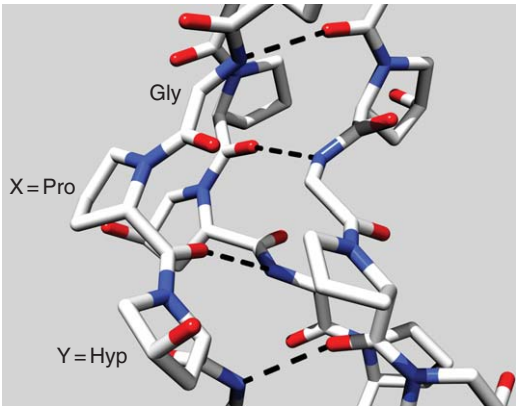


Figure 16 Interchain hydrogen bond network in collagen triple helix. The figure was generated using the UCSF Chimera package[524] and coordinates from PDB structure 1QSU.

of three nonequivalent peptide groups, and not all backbone peptide groups participate in hydrogen bonding. It was clear from earliest models that one strong interchain hydrogen bond NH. . .CO could be formed per Gly-X-Y tripeptide unit.[522] Ramachandran and Kartha[505] originally argued that a second NH. . .CO bond was possible when the residue in the X position was not Pro, but because of distortion and steric problems, this view was modified to suggest an interaction mediated by water.[523] All crystal structures show a hydrogen bond between the NH of Gly in one chain and the C=O of the residue in the X position of the neighboring chain, as predicted (**Figure 16**).

5.16.6.3 Collagen Hydration

The importance of water in collagen structure has been known from many techniques, and extensive hydration networks are seen in all crystal structures of collagen triple helical peptides.[516,517] These hydrogen-bonded networks often adopt pentagonal arrangements of water and are anchored to the peptide chain through backbone carbonyls and hydroxyl groups of Hyp. The observation of Hyp involvement in a hydrogen-bonded water network is consistent with that proposed by Ramachandran's group on the basis of modeling and Privalov on the basis of calorimetry studies on collagens.[525,526] The mechanism of stabilization of the triple helix by Hyp has been controversial, and studies on peptides with fluoroproline suggest that stereoelectronic effects may lead to an *exo*-puckering of the imino acid ring, which is favorable for the triple helix.[527] The crystal structures of peptides clarified a long-standing debate regarding hydrogen bonding-within the triple helix. The direct interchain [(Gly) NH. . .CO (X position)] has been a basic feature of all models, but there had been proposal of a second H bond between NH (X position) and the CO of a Gly in a neighboring chain, either directly or mediated by water.[523] The determination of several crystal structures of peptides with amino acids, rather than Pro, in the X position showed a second interchain backbone hydrogen bond mediated by a water NH(X). . .W. . .CO(Gly) (**Figure 17**).[2,514,519,528]

5.16.6.4 Side-Chain Interactions

The availability of high-resolution structures of peptides EKG, T3-785, IBP, and G991–G1032, which include residues other than Pro and Hyp in the X and Y positions, offers the opportunity to investigate the conformation and interactions of side chains from residues typically found within the collagen triple helix. In the peptide with an EKG tripeptide sequence, the Lys and Glu residues did not form direct intermolecular or intramolecular ion pairs, even though such pairs are sterically feasible.[528]

Instead, the Lys side chains bond to Y position carbonyl groups of an adjacent chain, while one Glu directly interacts with a Hyp hydroxyl group. There was also a range of water-mediated interactions involving the polar side chains. In peptide T3-785, with the central region Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala, the Arg side

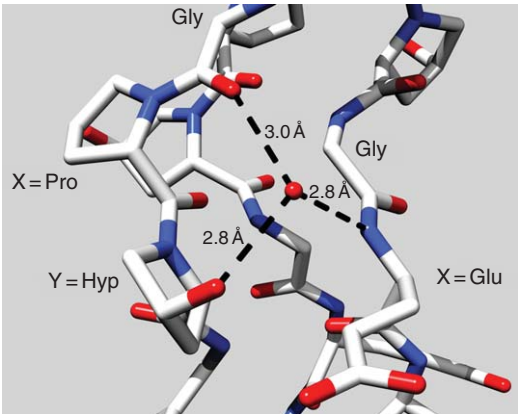


Figure 17 Water-mediated interchain hydrogen bond. The figure was created with the UCSF Chimera package[553] using coordinates from PDB structure 1QSU.

chains make direct contacts with backbone carbonyl groups on an adjacent chain, confirming predictions of this interaction and clarifying the high stability of Arg in the Y position.[519,529] In T3-785, the Arg side chains also make hydrophobic interactions with Leu and Ile side chains from the same or neighboring molecules, forming nonpolar clusters that minimize exposure to solvent. The Thr side chains are involved in bonding with the water, which mediates hydrogen bonds between the amide groups from the X position and Gly C=O groups. The participation of Thr in the Y position in the water network, much as Hyp does when it is in the Y position, suggests that Thr could play a similar stabilizing role in invertebrates and bacteria. The IBP structure, with the central Gly-Phe-Hyp-Gly-Glu-Arg sequence, shows a direct and a water-mediated electrostatic interaction between Glu and Arg side chains from adjacent strands, but no intrastrand interactions (**Figure 18**). The interactions between IBP Glu and Arg side chains are disrupted upon integrin binding (**Figure 19**).[520] A comparison of IBP and IBP-complex main-chain conformation reveals the flexible nature of the triple helix backbone in the imino-poor GFOGER region. This flexibility could be important for the integrin–collagen interaction and provides a possible explanation for the unique orientation of the three GFOGER strands observed in the integrin–IBPc complex crystal structure. In the G991–G1032 structure, there are numerous intra- and interchain ion pair and hydrogen-bonding interactions observed between charged and polar residues. There are one intrachain ionic interaction between the side chains of Arg15(A) and Glu18(A)

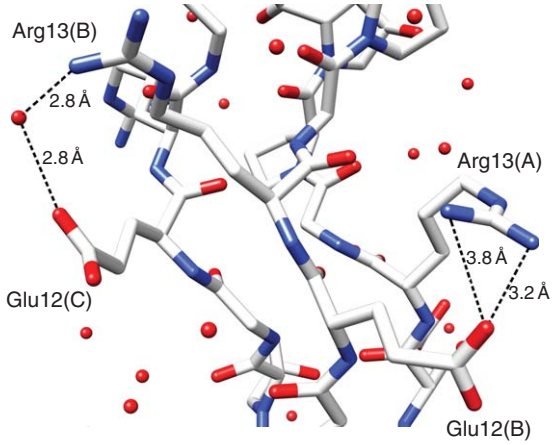


Figure 18 Molecular structure of IBP collagen (PDB accession number 1Q7D) showing a direct intrahelical ion bridge and a water-mediated interaction between Glu and Arg of different chains. The figure was generated using the UCSF Chimera package.[524]

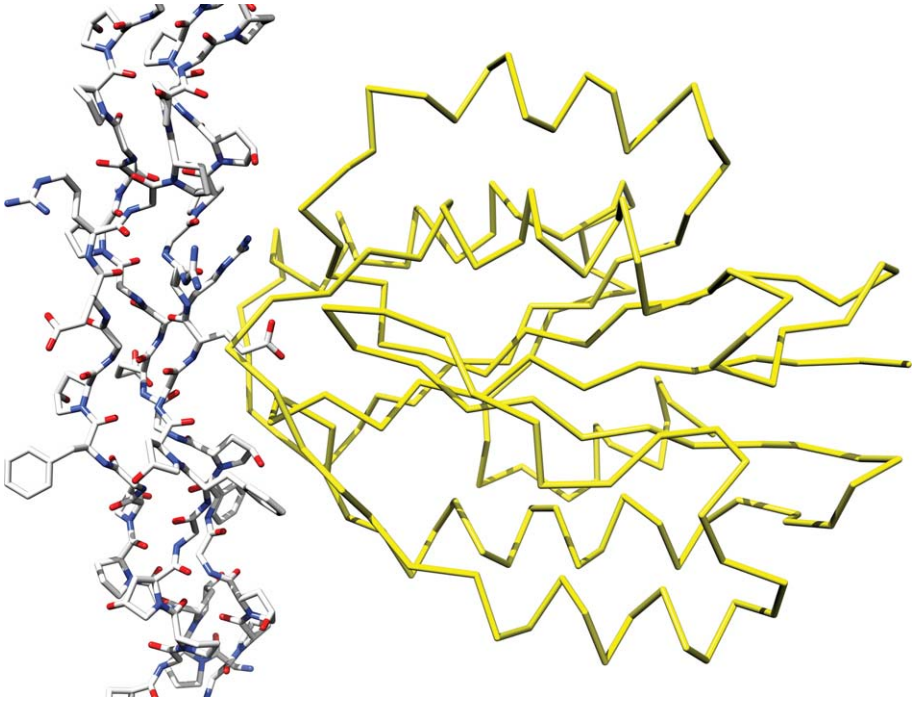


Figure 19 Complex structure of IBP and integrin (PDB accession number 1DZI). The figure was generated using the UCSF Chimera package.[524]

and two interchain ion pairs between Arg9(C) and Glu14(A) and Arg12(A) and Glu14(B). In addition to ion-pairing interactions, there are numerous interchain hydrogen bonds. Interestingly, many of the contacts involve serine residues. Interestingly, there is a continuous stretch of ionic and hydrogen-bonding interactions involving seven residues that belong to four different tripeptide units. This is the first observation of such extended interactions in collagen triple helix. Thus, theories on the thermal stability of short triple helical sequences that are based only on local ($\pm$1 tripeptide unit) interactions[530] must be reconsidered.

5.16.6.5 Collagen Trimerization Domains

5.16.6.5.1 Type IV collagen

Mammalian type IV collagen comprises a family of six homologous α chains, designated α1–α6.[531] Each chain is characterized by a long collagenous domain of approximately 1400 residues of Gly-Xaa-Yaa repeats interrupted by approximately 20 short NC sequences, an NC1 domain of approximately 230 residues at the C-terminus, and a small NC sequence at the N-terminus.[187] The six chains assemble into three distinct protomers, differing in chain composition, that are assembled by the association of the NC1 domains followed by triple helix formation of the collagenous domains. These protomers are α1α2, α3α4α5, and α5α5α6.[531] The protomers self-associate to form three distinct networks: α1α2, α3α4α5, and α1α2–α5α6 networks.[531] The networks are assembled by the dimerization at the C-terminus through NC1 domain interactions and by the tetramerization of the 7S domain at the N-terminus. The specificity of both the protomer and the network assembly is governed by molecular recognition sequences encoded within the NC1 domains.[487,532] The NC1 domains of the six α chains can be divided as α1-like (α1, α3, and α5) and α2-like (α2, α4, and α6) subfamilies based on sequence identity.[533,534] Furthermore, each NC1 domain consists of two homologous subdomains with approximately 35% sequence identity.[194] The ubiquitous [(α1)2.α2]2 NC1 hexamer was isolated from bovine lens capsule basement membrane and its crystal structure was solved at 2.0 Å resolution.[194] Independently, [(α1)2.α2]2 NC1 hexamer from human placenta was crystallized and its 1.9 Å structure was solved.[193] The NC1 monomer folds into a novel tertiary structure with predominantly β- strands. The two α1 chains in the

trimer are identical, and the $\alpha 2$ chain has a similar overall structure (**Figure 20**). The Cα atoms of 214 matching residues in one of the $\alpha 1$ chains and the $\alpha 2$ chain superimpose with an r.m.s. deviation of 0.9 Å. Each chain can be divided into two homologous subdomains, the N- and C-subdomains. The two subdomains fold in a similar topology, and the Cα atoms of 96 matching residues of two subdomains of the $\alpha 1$ chain superimpose with an r.m.s. deviation of 1.0 Å. The 12 invariant cysteine residues form six disulfides, 3 in each subdomain, at conserved positions.

Two chains of $\alpha 1$ and one chain of $\alpha 2$ form the trimer structure with a pseudo-three-fold molecular symmetry (**Figure 20**). Because each chain is made up of topologically similar subdomains, there is even a pseudo six-fold symmetry. The trimer structure is approximately cone-shaped with a base diameter of about 65 Å and a hollow core of about 12–14.0 Å inner diameter. This is about the same as the diameter of the collagen triple helix, with the N-termini of all three chains coming together at the vertex of the cone where the triple helical collagenous domain links the NC1 domain. The trimer is tightly packed through several interchain hydrophobic and hydrogen-bonding interactions.[194]

The football-shaped hexamer is made up of two identical trimers. The two NC1 trimers are related by a two-fold noncrystallographic symmetry axis lying in the equatorial plane and perpendicular to the pseudo-three-fold axis (**Figure 20**). The interface, which covers about 4400 Å^2 of solvent-accessible area per trimer, is formed by the nearly flat surfaces of the two trimers. The polar (45.5%) and nonpolar atoms (54.5%) in the interface are in almost equal proportions, underscoring the importance of both types of interactions in the hexamer stabilization.[194]

5.16.6.5.2 Type VIII and X collagens

Type VIII collagen is a nonfibrillar short-chain collagen comprised of $\alpha 1$(VIII) and $\alpha 2$(VIII) chains, and it was originally identified as a biosynthetic product of endothelial cells from bovine aorta[254] and rabbit cornea.[255] Type VIII collagen is a major component of Descemet's membrane,[535] a specialized basement membrane separating the corneal endothelium and stroma, and is also present in vascular subendothelial matrices, heart, liver, kidney, and lung, as well as in malignant tumors.[536] Collagen X is a short-chain, homotrimeric collagen and it is expressed specifically by hypertrophic chondrocytes during endochondral ossification in the developing vertebrate embryo.[269,270,537,538] In the adult animal, collagen X expression is reactivated during fracture repair[539] and in osteoarthritis.[540] Both type VIII and X collagens consist of a short N-terminal NC2 domain followed by approximately 150 collagen triple helix-forming repeats and a C-terminal NC1 domain of

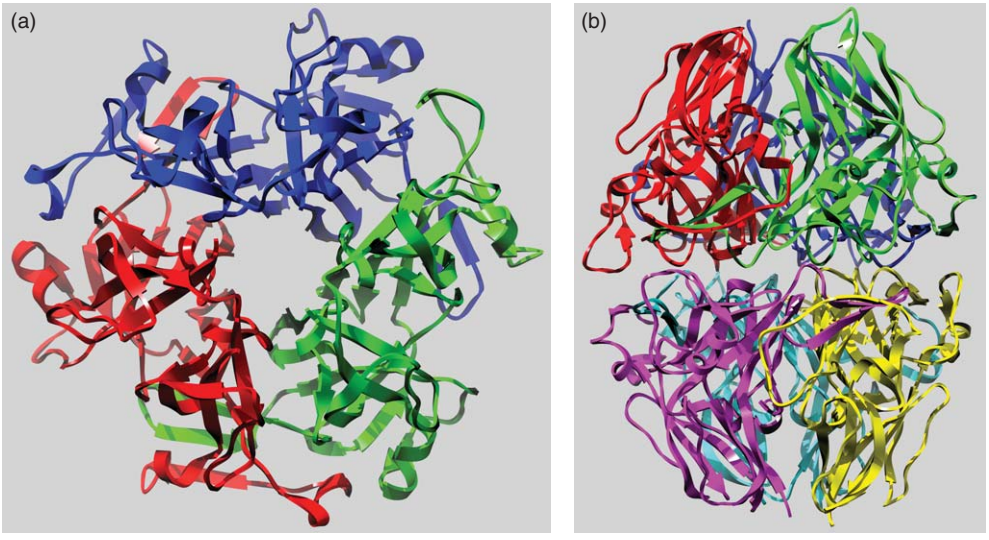


Figure 20 Ribbon representations of structures of the type IV collagen NC1 trimer (PDB accession number 1M3D) (a) down the pseudo-three-fold axis and hexamer (b) down the two-fold NCS axis. The figure was generated using the UCSF Chimera package.[524]

approximately 160 residues, which is homologous to the globular domain of the complement protein C1q and to C1q-like domains in other proteins.[541–543]

The crystal structures of the mouse collagen α1 (VIII) NC1 trimer[266] (PDB accession number 1GR3) and the human collagen X NC1 trimer[265] (PDB accession number 1O91) were solved (**Figure 21** and **22**). The sequence identity of NC1 domains of type VIII and X collagens is 61.5%. The NC1 domains of type VIII and X collagens are very similar, with an r.m.s. deviation of only 0.51 Å (99 Cα atoms) when single subunits are superimposed and 0.61 Å when the entire trimers are superimposed (297 Cα atoms).[266] The first C1q-like domain to have its structure determined was the globular domain of ACRP30.[264] The sequence identity between type VIII collagen NC1 and the globular domain of ACRP30 is 40%, resulting in higher r.m.s. deviations of 0.92 Å for superimposing single subunits (ACRP30 subunit A; 94 Cα atoms) and 2.3 Å for the entire trimers (273 Cα atoms).[266] The most striking feature of the NC1 surfaces of type VIII and X collagens is three strips of partially exposed hydrophobic residues. The strips extend across each subunit interface and

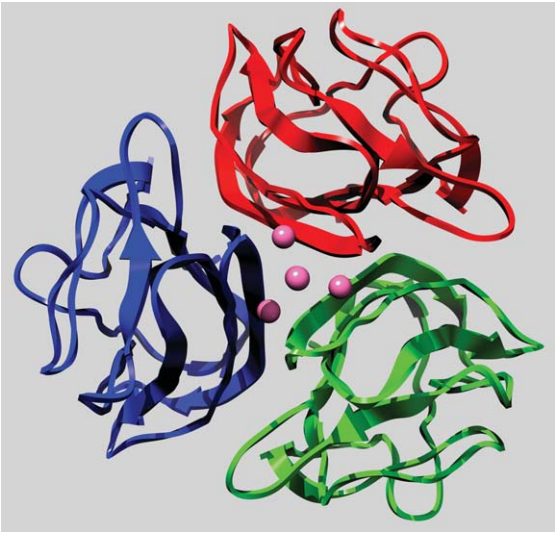


Figure 21 Structure of the collagen X NC1 trimer (PDB accession number 1GR3). Ribbon representation of the NC1 trimer viewed down the crystallographic three-fold axis is given. Ca^{2+} ions are represented as pink spheres. The figure was generated using the UCSF Chimera package.[524]

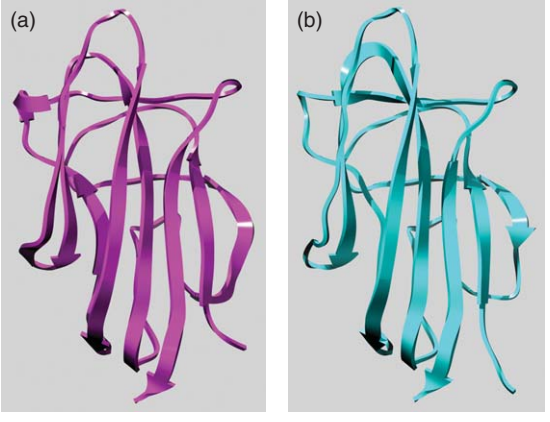


Figure 22 Structures of type VIII (a) and X (b) collagens NC1 single chains (PDB accession numbers 1O91 and 1GR3, respectively) viewed perpendicular to the three-fold axis. The figure was generated using the UCSF Chimera package.[524]

include the side chains of F625, W674, F678, Y686, Y688, F714, Y726, and Y730 for type VIII collagen NC1[266] and the side chains of Y562, W611, Y615, Y623, Y625, W651, Y663, and Y667 for collagen X NC1.[265] The apolar nature of the aromatic strips is underscored by the presence of a partially ordered CHAPS detergent molecule, which is bound in the cleft between subunits. Type VIII collagen is expressed during angiogenesis and in response to vascular injury and may form a provisional matrix during remodeling of the blood vessel wall.[536] Collagen X forms a transient matrix during endochondral ossification in growing long bones.[544] Both type VIII and X collagens have been observed to form higher order structures *in vitro*, the assembly of which is thought to involve the NC1 domains.[73,258] A possible mechanism for polygonal lattice formation in type VIII and X collagens has been suggested in which three hydrophobic strips on the surface of the NC1 trimer initiate the supramolecular assembly.[265,266]

Although the overall structures of the NC1 domains of type VIII and X collagens are very similar, there is a striking and unexpected difference: the collagen X NC1 trimer contains four buried calcium ions, which are absent from the type VIII collagen NC1 structure. The difference is due to a single amino acid replacement in the solvent channel, from T629 in collagen X to K692 in type VIII collagen, which allows the side chains of K692 to replace the calcium ions while maintaining a similarly dense network of interactions near the apex of the NC1 trimer.[266]

5.16.6.5.3 Crystal structure of $(GPP)_{10}$ foldon

Despite the requirement of NC domains for correct oligomerization of the collagen triple helix, the NC domain structures are divergent among all collagens. The structure of the NC1 domain of type IV collagen[193,194] is unrelated to that of type VIII and X collagens.[265,266] However, common to all the different domains is a trimerization potential, which leads to a high local concentration of collagen chains, which, in turn, promotes triple helix nucleation by entropic reasons.[457,497,545] In agreement with this hypothesis, a trimeric foldon domain of bacteriophage T4 fibritin stabilizes the collagen model peptide $(GPP)_{10}$ if linked to its C-terminus.[497] Crystal structure analysis at 2.6 Å resolution of this artificial fusion revealed conformational changes within the interface of both domains compared with the structure of the isolated molecules.[498] A striking feature is an angle of 62.5° between the symmetry axis of the foldon domain and the axis of the triple helix (**Figure 23**). Although there are two linker residues (GS) in the interface, the last three residues of the triple helix and the first three residues of the foldon domain are disobeying the geometry of the stand-alone molecules. Unfortunately, there are no atomic structures of natural interfaces between a collagen triple helix and a trimerization domain available so far.

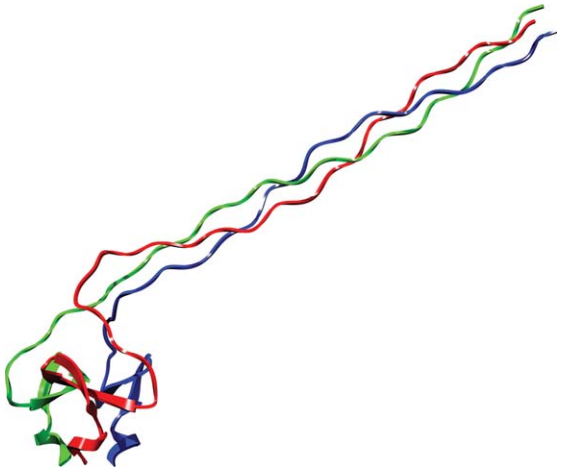


Figure 23 The overall structure of the $(GlyProPro)_{10}$ foldon (PDB accession number 1NAY). The symmetry axis and symmetry elements are shown by a line and symbolic representations, respectively. The backbones of different chains are shown in red, blue, and green.[498] The figure was generated using the UCSF Chimera package.[524]

5.16.7 Other Collagens

In addition to classical collagens, the collagen triple helix is present as a motif in a variety of proteins. The kinked triple helix domain in C1q binds to the serine proteases C1r and C1s, and mediates the self-association of six trimer molecules that generates the bouquet-type structure important for activity.[546] The family of collectins, which includes the mannose-binding lectin SP-A and SP-D, all contain a collagenous domain, a coiled coil α-helical domain, and a terminal carbohydrate-recognition domain. Ficolins have a triple helix domain with a terminal carbohydrate-binding fibrinogen domain. Both collectins and ficolins are host defense molecules, which bind to the carbohydrate groups of microbes, leading to complement activation and phagocytosis.[547] The triple helix domain of the macrophage scavenger receptor is responsible for ligand recognition,[548, 549] while the collagenous tail of the asymmetric form of acetylcholinesterase binds to heparan sulfate, localizing the enzyme to the neuromuscular junction.[550]

Orthologues of fibril-forming collagens and type IV collagen are found in a number of invertebrates, in addition to specialized collagens such as the cuticle collagens of *C. elegans* and the hydra nematocyst minicollagens.[11] It was thought that collagen was a defining feature of multicellular animals, but recent observations have shown triple helix domains in bacteria and viruses. Collagen-like the presence of domains Scl1 and Scl2 were observed in proteins expressed on the cell surface of group A streptococcus, and were shown to adopt a triple helix conformation.[551] A highly repetitive collagen-like domain was also identified in the filament protein of the exosporium of anthrax spores, and the length of this collagenous domain appears to determine the filament length.[552] A search of the genomes of various bacteria and viruses indicated the presence of at least 100 novel proteins containing collagen-related structural motifs.[553] These findings confirm that the collagen triple helix motif can be part of many different kinds of proteins and can fill a wider than expected set of biological niches.

References

1. W. L. DeLano, *The PyMOL Molecular Graphic System*; DeLano Scientific LLC: Palo Alto, CA, 2006.
2. S. P. Boudko; J. Engel; K. Okuyama; K. Mizuno; H. P. Bächinger; M. A. Schumacher, *J. Biol. Chem.* **2008**, *283*, 32580–32589.
3. E. Hu; P. Liang; B. M. Spiegelman, *J. Biol. Chem.* **1996**, *271*, 10697–10703.
4. G. W. Wong; S. A. Krawczyk; C. Kitidis-Mitrokostas; G. Ge; E. Spooner; C. Hug; R. Gimeno; H. F. Lodish, *FASEB J.* **2009**, *23*, 241–258.
5. T. Kodama; M. Freeman; L. Rohrer; J. Zabrecky; P. Matsudaira; M. Krieger, *Nature* **1990**, *343*, 531–535.
6. R. T. White; D. Damm; J. Miller; K. Spratt; J. Schilling; S. Hawgood; B. Benson; B. Cordell, *Nature* **1985**, *317*, 361–363.
7. N. Kondo; J. Kondo, *J. Biol. Chem.* **1992**, *267*, 473–478.
8. K. B. Reid, *Biochem. J.* **1974**, *141*, 189–203.
9. S. S. Bohlson; D. A. Fraser; A. J. Tenner, *Mol. Immunol.* **2007**, *44*, 33–43.
10. I. L. Johnstone, *Trends Genet.* **2000**, *16*, 21–27.
11. J. Exposito; C. Cluzel; R. Garrone; C. Lethias, *Anat. Rec.* **2002**, *268*, 302–316.
12. J. Engel, *Science* **1997**, *277*, 1785–1786.
13. A. P. Page; I. L. Johnstone, *WormBook* **2007**, 1–15.
14. B. Stevens; N. J. Allen; L. E. Vazquez; G. R. Howell; K. S. Christopherson; N. Nouri; K. D. Micheva; A. K. Mehalow; A. D. Huberman; B. Stafford; A. Sher; A. M. Litke; J. D. Lambris; S. J. Smith; S. W. M. John; B. A. Barres, *Cell* **2007**, *131*, 1164–1178.
15. T. Xiao; H. Baier, *Nat. Neurosci.* **2007**, *10*, 1529–1537.
16. M. Koch; F. Laub; P. Zhou; R. A. Hahn; S. Tanaka; R. E. Burgeson; D. R. Gerecke; F. Ramirez; M. K. Gordon, *J. Biol. Chem.* **2003**, *278*, 43236–43244.
17. R. P. Boot-Handford; D. S. Tuckwell; D. A. Plumb; C. F. Rock; R. Poulsom, *J. Biol. Chem.* **2003**, *278*, 31067–31077.
18. J. M. Pace; M. Corrado; C. Missero; P. H. Byers, *Matrix Biol.* **2003**, *22*, 3–14.
19. K. Beck; B. A. Boswell; C. C. Ridgway; H. P. Bächinger, *J. Biol. Chem.* **1996**, *271*, 21566–21573.
20. E. Leikina; M. V. Mertts; N. Kuznetsova; S. Leikin, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1314–1318.
21. S. D. Chipman; H. O. Sweet; D. J. J. McBride; M. T. Davisson; S. C. J. Marks; A. R. Shuldiner; R. J. Wenstrup; D. W. Rowe; J. R. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1701–1705.
22. A. Colige; I. Vandenberghe; M. Thiry; C. A. Lambert; J. Van Beeumen; S. Li; D. J. Prockop; C. M. Lapiere; B. V. Nusgens, *J. Biol. Chem.* **2002**, *277*, 5756–5766.
23. R. J. Fernandes; S. Hirohata; J. M. Engle; A. Colige; D. H. Cohn; D. R. Eyre; S. S. Apte, *J. Biol. Chem.* **2001**, *276*, 31502–31509.
24. W. Wang; G. Ge; N. H. Lim; H. Nagase; D. S. Greenspan, *Biochem. J.* **2006**, *398*, 515–519.
25. D. R. Hopkins; S. Keles; D. S. Greenspan, *Matrix Biol.* **2007**, *26*, 508–523.
26. W. N. Pappano; B. M. Steiglitz; I. C. Scott; D. R. Keene; D. S. Greenspan, *Mol. Cell. Biol.* **2003**, *23*, 4428–4438.
27. J. S. Bond; R. J. Beynon, *Protein Sci.* **1995**, *4*, 1247–1261.
28. W. Stöcker; R. Zwilling, *Meth. Enzymol.* **1995**, *248*, 305–325.

29. G. Ge; D. S. Greenspan, *Birth Defects Res. C Embryo Today* **2006**, *78*, 47–68.
30. R. G. Spiro, *J. Biol. Chem.* **1967**, *242*, 4813–4823.
31. D. R. Eyre; M. A. Weis; J. Wu, *Methods* **2008**, *45*, 65–74.
32. D. R. Eyre; M. A. Weis; J. Wu, *Eur. Cell Mater.* **2006**, *12*, 57–63.
33. D. R. Eyre; M. A. Paz; P. M. Gallop, *Annu. Rev. Biochem.* **1984**, *53*, 717–748.
34. D. Eyre, *Meth. Enzymol.* **1987**, *144*, 115–139.
35. N. C. Avery; A. J. Bailey, *Scand. J. Med. Sci. Sports* **2005**, *15*, 231–240.
36. L. Knott; J. F. Tarlton; A. J. Bailey, *Biochem. J.* **1997**, *322* (Pt. 2), 535–542.
37. A. J. Bailey; R. G. Paul; L. Knott, *Mech. Ageing Dev.* **1998**, *106*, 1–56.
38. S. M. Carleton; D. J. McBride; W. L. Carson; C. E. Huntington; K. L. Twenter; K. M. Rolwes; C. T. Winkelmann; J. S. Morris; J. F. Taylor; C. L. Phillips, *Bone* **2008**, *42*, 681–694.
39. G. E. Lopez Franco; A. Huang; N. Pleshko Camacho; D. S. Stone; R. D. Blank, *J. Dent. Res.* **2006**, *85*, 1032–1036.
40. C. A. Miles; T. J. Sims; N. P. Camacho; A. J. Bailey, *J. Mol. Biol.* **2002**, *321*, 797–805.
41. C. Hongo; K. Noguchi; K. Okuyama; Y. Tanaka; N. Nishino, *J. Biochem.* **2005**, *138*, 135–144.
42. J. Emsley; C. G. Knight; R. W. Farndale; M. J. Barnes; R. C. Liddington, *Cell* **2000**, *101*, 47–56.
43. C. Renner; B. Saccà; L. Moroder, *Biopolymers* **2004**, *76*, 34–47.
44. G. A. Di Lullo; S. M. Sweeney; J. Korkko; L. Ala-Kokko; J. D. San Antonio, *J. Biol. Chem.* **2002**, *277*, 4223–4231.
45. S. M. Sweeney; J. P. Orgel; A. Fertala; J. D. McAuliffe; K. R. Turner; G. A. Di Lullo; S. Chen; O. Antipova; S. Perumal; L. Ala-Kokko; A. Forlino; W. A. Cabral; A. M. Barnes; J. C. Marini; J. D. San Antonio, *J. Biol. Chem.* **2008**, *283*, 21187–21197.
46. J. Ottl; R. Battistuta; M. Pieper; H. Tschesche; W. Bode; K. Kühn; L. Moroder, *FEBS Lett.* **1996**, *398*, 31–36.
47. J. Ottl; L. Moroder, *J. Am. Chem. Soc.* **1999**, *121*, 653–661.
48. C. G. Knight; L. F. Morton; A. R. Peachey; D. S. Tuckwell; R. W. Farndale; M. J. Barnes, *J. Biol. Chem.* **2000**, *275*, 35–40.
49. H. Wu; M. H. Byrne; A. Stacey; M. B. Goldring; J. R. Birkhead; R. Jaenisch; S. M. Krane, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 5888–5892.
50. J. Löhler; R. Timpl; R. Jaenisch, *Cell* **1984**, *38*, 597–607.
51. S. Hartung; R. Jaenisch; M. Breindl, *Nature* **1986**, *320*, 365–367.
52. D. J. J. McBride; K. E. Kadler; Y. Hojima; D. J. Prockop, *Matrix* **1992**, *12*, 256–263.
53. J. Saban; M. A. Zussman; R. Havey; A. G. Patwardhan; G. B. Schneider; D. King, *Bone* **1996**, *19*, 575–579.
54. P. Fratzl; O. Paris; K. Klaushofer; W. J. Landis, *J. Clin. Invest.* **1996**, *97*, 396–402.
55. N. V. Kuznetsova; D. J. McBride; S. Leikin, *J. Mol. Biol.* **2003**, *331*, 191–200.
56. D. J. J. McBride; V. Choe; J. R. Shapiro; B. Brodsky, *J. Mol. Biol.* **1997**, *270*, 275–284.
57. S. W. Li; D. J. Prockop; H. Helminen; R. Fässler; T. Lapveteläinen; K. Kiraly; A. Peltarri; J. Arokoski; H. Lui; M. Arita; J. S. Khillan, *Genes Dev.* **1995**, *9*, 2821–2830.
58. X. Liu; H. Wu; M. Byrne; S. Krane; R. Jaenisch, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1852–1856.
59. E. Pöschl; U. Schlötzer-Schrehardt; B. Brachvogel; K. Saito; Y. Ninomiya; U. Mayer, *Development* **2004**, *131*, 1619–1628.
60. D. B. Gould; F. C. Phalan; G. J. Breedveld; S. E. van Mil; R. S. Smith; J. C. Schimenti; U. Aguglia; M. S. van der Knaap; P. Heutink; S. W. M. John, *Science* **2005**, *308*, 1167–1171.
61. D. B. Gould; F. C. Phalan; S. E. van Mil; J. P. Sundberg; K. Vahedi; P. Massin; M. G. Bousser; P. Heutink; J. H. Miner; E. Tournier-Lasserve; S. W. M. John, *N. Engl. J. Med.* **2006**, *354*, 1489–1496.
62. J. H. Miner; J. R. Sanes, *J. Cell Biol.* **1996**, *135*, 1403–1413.
63. Y. Hamano; J. A. Grunkemeyer; A. Sudhakar; M. Zeisberg; D. Cosgrove; R. Morello; B. Lee; H. Sugimoto; R. Kalluri, *J. Biol. Chem.* **2002**, *277*, 31154–31162.
64. M. N. Rheault; S. M. Kren; B. K. Thielen; H. A. Mesa; J. T. Crosson; W. Thomas; Y. Sado; C. E. Kashtan; Y. Segal, *J. Am. Soc. Nephrol.* **2004**, *15*, 1466–1474.
65. W. Lu; C. L. Phillips; P. D. Killen; T. Hlaing; W. R. Harrison; F. F. Elder; J. H. Miner; P. A. Overbeek; M. H. Meisler, *Genomics* **1999**, *61*, 113–124.
66. R. J. Wenstrup; J. B. Florer; J. M. Davidson; C. L. Phillips; B. J. Pfeiffer; D. W. Menezes; I. Chervoneva; D. E. Birk, *J. Biol. Chem.* **2006**, *281*, 12888–12895.
67. R. J. Wenstrup; J. B. Florer; E. W. Brunskill; S. M. Bell; I. Chervoneva; D. E. Birk, *J. Biol. Chem.* **2004**, *279*, 53331–53337.
68. K. Andrikopoulos; X. Liu; D. R. Keene; R. Jaenisch; F. Ramirez, *Nat. Genet.* **1995**, *9*, 31–36.
69. P. Bonaldo; P. Braghetta; M. Zanetti; S. Piccolo; D. Volpin; G. M. Bressan, *Hum. Mol. Genet.* **1998**, *7*, 2135–2140.
70. W. A. Irwin; N. Bergamin; P. Sabatelli; C. Reggiani; A. Megighian; L. Merlini; P. Braghetta; M. Columbaro; D. Volpin; G. M. Bressan; P. Bernardi; P. Bonaldo, *Nat. Genet.* **2003**, *35*, 367–371.
71. A. Fritsch; S. Loeckermann; J. S. Kern; A. Braun; M. R. Bösl; T. A. Bley; H. Schumann; D. von Elverfeldt; D. Paul; M. Erlacher; D. Berens von Rautenfeld; I. Hausser; R. Fässler; L. Bruckner-Tuderman, *J. Clin. Invest.* **2008**, *118*, 1669–1679.
72. U. Hopfer; N. Fukai; H. Hopfer; G. Wolf; N. Joyce; E. Li; B. R. Olsen, *FASEB J.* **2005**, *19*, 1232–1244.
73. T. M. Schmid; T. F. Linsenmayer, *J. Cell Biol.* **1985**, *100*, 598–605.
74. C. J. Gress; O. Jacenko, *J. Cell Biol.* **2000**, *149*, 983–993.
75. W. T. McGuirt; S. D. Prasad; A. J. Griffith; H. P. Kunst; G. E. Green; K. B. Shpargel; C. Runge; C. Huybrechts; R. F. Mueller; E. Lynch; M. C. King; H. G. Brunner; C. W. Cremers; M. Takanosu; S. W. Li; M. Arita; R. Mayne; D. J. Prockop; G. Van Camp; R. J. Smith, *Nat. Genet.* **1999**, *23*, 413–419.
76. A. P. Kvist; A. Latvanlehto; M. Sund; L. Eklund; T. Väisänen; P. Hägg; R. Sormunen; J. Komulainen; R. Fässler; T. Pihlajaniemi, *Am. J. Pathol.* **2001**, *159*, 1581–1592.
77. R. Ylönen; T. Kyrönlahti; M. Sund; M. Ilves; P. Lehenkari; J. Tuukkanen; T. Pihlajaniemi, *J. Bone Miner. Res.* **2005**, *20*, 1381–1393.
78. H. L. Ansorge; X. Meng; G. Zhang; G. Veit; M. Sun; J. F. Klement; D. P. Beason; L. J. Soslowsky; M. Koch; D. E. Birk, *J. Biol. Chem.* **2009**, *284*, 8427–8438.
79. L. Eklund; J. Piuhola; J. Komulainen; R. Sormunen; C. Ongvarrasopone; R. Fässler; A. Muona; M. Ilves; H. Ruskoaho; T. E. Takala; T. Pihlajaniemi, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1194–1199.

80. N. Fukai; L. Eklund; A. G. Marneros; S. P. Oh; D. R. Keene; L. Tamarkin; M. Niemelä; M. Ilves; E. Li; T. Pihlajaniemi; B. R. Olsen, *EMBO J.* **2002**, *21*, 1535–1544.
81. A. Utriainen; R. Sormunen; M. Kettunen; L. S. Carvalhaes; E. Sajanti; L. Eklund; R. Kauppinen; G. T. Kitten; T. Pihlajaniemi, *Hum. Mol. Genet.* **2004**, *13*, 2089–2099.
82. A. G. Marneros; B. R. Olsen, *Invest. Ophthalmol. Vis. Sci.* **2003**, *44*, 2367–2372.
83. T. Holster; O. Pakkanen; R. Soininen; R. Sormunen; M. Nokelainen; K. I. Kivirikko; J. Myllyharju, *J. Biol. Chem.* **2007**, *282*, 2512–2519.
84. K. Takaluoma; M. Hyry; J. Lantto; R. Sormunen; R. A. Bank; K. I. Kivirikko; J. Myllyharju; R. Soininen, *J. Biol. Chem.* **2007**, *282*, 6588–6596.
85. L. Sipilä; H. Ruotsalainen; R. Sormunen; N. L. Baker; S. R. Lamandé; M. Vapola; C. Wang; Y. Sado; A. Aszodi; R. Myllylä, *J. Biol. Chem.* **2007**, *282*, 33381–33388.
86. H. Ruotsalainen; L. Sipilä; M. Vapola; R. Sormunen; A. M. Salo; L. Uitto; D. K. Mercer; S. P. Robins; M. Risteli; A. Aszodi; R. Fässler; R. Myllylä, *J. Cell Sci.* **2006**, *119*, 625–635.
87. K. Rautavuoma; K. Takaluoma; R. Sormunen; J. Myllyharju; K. I. Kivirikko; R. Soininen, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14120–14125.
88. R. Morello; T. K. Bertin; Y. Chen; J. Hicks; L. Tonachini; M. Monticone; P. Castagnola; F. Rauch; F. H. Glorieux; J. Vranka; H. P. Bächinger; J. M. Pace; U. Schwarze; P. H. Byers; M. Weis; R. J. Fernandes; D. R. Eyre; Z. Yao; B. F. Boyce; B. Lee, *Cell* **2006**, *127*, 291–304.
89. N. Nagai; M. Hosokawa; S. Itohara; E. Adachi; T. Matsushita; N. Hosokawa; K. Nagata, *J. Cell. Biol.* **2000**, *150*, 1499–1506.
90. R. Dalgleish, *Nucleic Acids Res.* **1998**, *26*, 253–255.
91. R. Dalgleish, *Nucleic Acids Res.* **1997**, *25*, 181–187.
92. D. O. Sillence; A. Senn; D. M. Danks, *J. Med. Genet.* **1979**, *16*, 101–116.
93. J. M. Pace; M. Wiese; A. S. Drenguis; N. Kuznetsova; S. Leikin; U. Schwarze; D. Chen; S. H. Mooney; S. Unger; P. H. Byers, *J. Biol. Chem.* **2008**, *283*, 16061–16067.
94. S. R. Lamandé; S. D. Chessler; S. B. Golub; P. H. Byers; D. Chan; W. G. Cole; D. O. Sillence; J. F. Bateman, *J. Biol. Chem.* **1995**, *270*, 8642–8649.
95. M. Raghunath; P. Bruckner; B. Steinmann, *J. Mol. Biol.* **1994**, *236*, 940–949.
96. R. Tenni; M. Valli; A. Rossi; G. Cetta, *Am. J. Med. Genet.* **1993**, *45*, 252–256.
97. W. A. Cabral; W. Chang; A. M. Barnes; M. Weis; M. A. Scott; S. Leikin; E. Makareeva; N. V. Kuznetsova; K. N. Rosenbaum; C. J. Tifft; D. I. Bulas; C. Kozma; P. A. Smith; D. R. Eyre; J. C. Marini, *Nat. Genet.* **2007**, *39*, 359–365.
98. E. J. Miller; V. J. Matukas, *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *64*, 1264–1268.
99. L. J. Ng; P. P. Tam; K. S. Cheah, *Dev. Biol.* **1993**, *159*, 403–417.
100. K. S. Cheah; E. T. Lau; P. K. Au; P. P. Tam, *Development* **1991**, *111*, 945–953.
101. R. A. Bank; B. Beekman; R. Tenni; J. M. TeKoppele, *J. Chromatogr. B Biomed. Sci. Appl.* **1997**, *703*, 267–272.
102. H. Notbohm; M. Nokelainen; J. Myllyharju; P. P. Fietzek; P. K. Müller; K. I. Kivirikko, *J. Biol. Chem.* **1999**, *274*, 8988–8992.
103. A. Aszódi; D. Chan; E. Hunziker; J. F. Bateman; R. Fässler, *J. Cell Biol.* **1998**, *143*, 1399–1412.
104. W. A. Horton; J. T. Hecht, Chondrodysplasias: Disorders of Cartilage Matrix Proteins. In *Connective Tissue and Its Heritable Disorders*; P. M. Royce, B. Steinmann, Eds.; Wiley-Liss: New York, 2002; pp 909–937.
105. W. A. Horton; J. T. Hecht, Chondrodysplasias: General Concepts and Diagnostic and Management Considerations. In *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects*; P. M. Royce, B. Steinmann, Eds.; Wiley-Liss: New York, 2002; pp 901–908.
106. M. P. Snead; J. R. Yates, *J. Med. Genet.* **1999**, *36*, 353–359.
107. R. D. Young; P. A. Lawrence; V. C. Duance; T. Aigner; P. Monaghan, *J. Histochem. Cytochem.* **2000**, *48*, 423–432.
108. H. P. Bächinger; P. Bruckner; R. Timpl; D. J. Prockop; J. Engel, *Eur. J. Biochem.* **1980**, *106*, 619–632.
109. L. I. Fessler; J. H. Fessler, *J. Biol. Chem.* **1979**, *254*, 233–239.
110. L. I. Fessler; R. Timpl; J. H. Fessler, *J. Biol. Chem.* **1981**, *256*, 2531–2537.
111. P. H. Byers; K. H. McKenney; J. R. Lichtenstein; G. R. Martin, *Biochemistry* **1974**, *13*, 5243–5248.
112. D. R. Keene; L. Y. Sakai; R. E. Burgeson; H. P. Bächinger, *J. Histochem. Cytochem.* **1987**, *35*, 311–318.
113. D. R. Keene; L. Y. Sakai; H. P. Bächinger; R. E. Burgeson, *J. Cell Biol.* **1987**, *105*, 2393–2402.
114. A. Jukkola; J. Risteli; O. Niemelä; L. Risteli, *Eur. J. Biochem.* **1986**, *154*, 219–224.
115. P. A. Baeuerle; W. B. Huttner, *J. Cell Biol.* **1987**, *105*, 2655–2664.
116. K. L. Moore, *J. Biol. Chem.* **2003**, *278*, 24243–24246.
117. A. D. Konitsiotis; N. Raynal; D. Bihan; E. Hohenester; R. W. Farndale; B. Leitinger, *J. Biol. Chem.* **2008**, *283*, 6861–6868.
118. E. Hohenester; T. Sasaki; C. Giudici; R. W. Farndale; H. P. Bächinger, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18273–18277.
119. E. Tillet; J. M. Franc; S. Franc; R. Garrone, *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **1996**, *113*, 239–246.
120. J. Exposito; C. Larroux; C. Cluzel; U. Valcourt; C. Lethias; B. M. Degnan, *J. Biol. Chem.* **2008**, *283*, 28226–28235.
121. R. E. Burgeson; D. W. Hollister, *Biochem. Biophys. Res. Commun.* **1979**, *87*, 1124–1131.
122. R. K. Rhodes; E. J. Miller, *Biochemistry* **1978**, *17*, 3442–3448.
123. R. K. Rhodes; E. J. Miller, *Coll. Relat. Res.* **1981**, *1*, 337–343.
124. H. Sage; P. Pritzl; P. Bornstein, *Biochemistry* **1981**, *20*, 3778–3784.
125. H. Sage; P. Bornstein, *Biochemistry* **1979**, *18*, 3815–3822.
126. C. Niyibizi; P. P. Fietzek; M. van der Rest, *J. Biol. Chem.* **1984**, *259*, 14170–14174.
127. R. A. Brown; C. A. Shuttleworth; J. B. Weiss, *Biochem. Biophys. Res. Commun.* **1978**, *80*, 866–872.
128. R. A. Brown; J. B. Weiss, *FEBS Lett.* **1979**, *106*, 71–75.
129. M. Z. Abedin; S. Ayad; J. B. Weiss, *Biosci. Rep.* **1982**, *2*, 493–502.
130. M. Z. Abedin; S. Ayad; J. B. Weiss, *Biochem. Biophys. Res. Commun.* **1981**, *102*, 1237–1245.
131. E. Adachi; I. Hopkinson; T. Hayashi, *Int. Rev. Cytol.* **1997**, *173*, 73–156.
132. M. A. Haralson; W. M. Mitchell; R. K. Rhodes; T. F. Kresina; R. Gay; E. J. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 5206–5210.

133. M. A. Haralson; W. M. Mitchell; R. K. Rhodes; E. J. Miller, *Arch. Biochem. Biophys.* **1984**, *229*, 509–518.
134. H. P. Bächinger; K. J. Doege; J. P. Petschek; L. I. Fessler; J. H. Fessler, *J. Biol. Chem.* **1982**, *257*, 14590–14592.
135. C. Luparello; R. Schillaci; I. Pucci-Minafra; S. Minafra, *Eur. J. Cancer* **1990**, *26*, 231–240.
136. K. Yamamoto; M. Yamamoto; T. Noumura, *Exp. Cell Res.* **1992**, *201*, 55–63.
137. T. Kihara; Y. Takemura; Y. Imamura; K. Mizuno; T. Hayashi, *Cell Tissue Res.* **2004**, *318*, 343–352.
138. T. Kihara; Y. Imamura; Y. Takemura; K. Mizuno; E. Adachi; T. Hayashi, *J. Biochem.* **2008**, *144*, 625–633.
139. Y. Murasawa; T. Hayashi; P. Wang, *Exp. Cell Res.* **2008**, *314*, 3638–3653.
140. S. M. Mumby; G. J. Raugi; P. Bornstein, *J. Cell Biol.* **1984**, *98*, 646–652.
141. Y. Yaoi; K. Hashimoto; H. Koitabashi; K. Takahara; M. Ito; I. Kato, *Biochim. Biophys. Acta* **1990**, *1035*, 139–145.
142. K. Mizuno; T. Hayashi, *J. Biochem.* **1996**, *120*, 934–939.
143. J. D. San Antonio; M. J. Karnovsky; S. Gay; R. D. Sanderson; A. D. Lander, *Glycobiology* **1994**, *4*, 327–332.
144. D. E. Birk, *Micron* **2001**, *32*, 223–237.
145. E. Adachi; T. Hayashi, *Connect. Tissue Res.* **1986**, *14*, 257–266.
146. K. Mizuno; E. Adachi; Y. Imamura; O. Katsumata; T. Hayashi, *Micron* **2001**, *32*, 317–323.
147. J. Fessler; L. Fessler, Type V Collagen. In *Structure and Function of Collagen Types*; R. Mayne, R. Burgeson, Eds.; Academic Press: Orlando, FL, 1987; pp 81–103.
148. K. Dreisewerd; A. Rohlfing; B. Spottke; C. Urbanke; W. Henkel, *Anal. Chem.* **2004**, *76*, 3482–3491.
149. W. Henkel; K. Dreisewerd, *J. Proteome Res.* **2007**, *6*, 4269–4289.
150. Y. Imamura; B. M. Steiglitz; D. S. Greenspan, *J. Biol. Chem.* **1998**, *273*, 27511–27517.
151. B. Gopalakrishnan; W. Wang; D. S. Greenspan, *J. Biol. Chem.* **2004**, *279*, 30904–30912.
152. L. I. Fessler; S. Brosh; S. Chapin; J. H. Fessler, *J. Biol. Chem.* **1986**, *261*, 5034–5040.
153. L. I. Fessler; S. Chapin; S. Brosh; J. H. Fessler, *Eur. J. Biochem.* **1986**, *158*, 511–518.
154. L. I. Fessler; N. Shigaki; J. H. Fessler, *J. Biol. Chem.* **1985**, *260*, 13286–13293.
155. C. Bonod-Bidaud; M. Beraud; E. Vaganay; F. Delacoux; B. Font; D. J. S. Hulmes; F. Ruggiero, *Biochem. J.* **2007**, *405*, 299–306.
156. A. Colige; F. Ruggiero; I. Vandenberghe; J. Dubail; F. Kesteloot; J. Van Beeumen; A. Beschin; L. Brys; C. M. Lapière; B. Nusgens, *J. Biol. Chem.* **2005**, *280*, 34397–34408.
157. A. Fichard; J. P. Kleman; F. Ruggiero, *Matrix Biol.* **1995**, *14*, 515–531.
158. N. P. Morris; H. P. Bächinger, *J. Biol. Chem.* **1987**, *262*, 11345–11350.
159. J. J. Wu; D. R. Eyre, *J. Biol. Chem.* **1995**, *270*, 18865–18870.
160. M. C. Ryan; L. J. Sandell, *J. Biol. Chem.* **1990**, *265*, 10334–10339.
161. J. P. Kleman; D. J. Hartmann; F. Ramirez; M. van der Rest, *Eur. J. Biochem.* **1992**, *210*, 329–335.
162. R. Mayne; R. G. Brewton; P. M. Mayne; J. R. Baker, *J. Biol. Chem.* **1993**, *268*, 9381–9386.
163. M. Majava; K. P. Hoornaert; D. Bartholdi; M. C. Bouma; K. Bouman; M. Carrera; K. Devriendt; J. Hurst; G. Kitsos; D. Niedrist; M. B. Petersen; D. Shears; I. Stolte-Dijkstra; J. M. Van Hagen; L. Ala-Kokko; M. Männikkö; G. R. Mortier, *Am. J. Med. Genet. A* **2007**, *143*, 258–264.
164. T. Pihlajamaa; D. J. Prockop; J. Faber; A. Winterpacht; B. Zabel; A. Giedion; P. Wiesbauer; J. Spranger; L. Ala-Kokko, *Am. J. Med. Genet.* **1998**, *80*, 115–120.
165. F. Mio; K. Chiba; Y. Hirose; Y. Kawaguchi; Y. Mikami; T. Oya; M. Mori; M. Kamata; M. Matsumoto; K. Ozaki; T. Tanaka; A. Takahashi; T. Kubo; T. Kimura; Y. Toyama; S. Ikegawa, *Am. J. Hum. Genet.* **2007**, *81*, 1271–1277.
166. E. F. Eikenberry; B. B. Brodsky; A. S. Craig; D. A. D. Parry, *Int. J. Biol. Macromol.* **1982**, *4*, 393–398.
167. E. F. Eikenberry; B. Brodsky; D. A. D. Parry, *Int. J. Biol. Macromol.* **1982**, *4*, 322–328.
168. J. Gross; F. O. Schmitt, *J. Exp. Med.* **1948**, *88*, 555–568.
169. D. J. Prockop; A. Fertala, *J. Biol. Chem.* **1998**, *273*, 15598–15604.
170. V. Ottani; D. Martini; M. Franchi; A. Ruggeri; M. Raspanti, *Micron* **2002**, *33*, 587–596.
171. J. H. Lillie; D. K. MacCallum; L. J. Scaletta; J. C. Occhino, *J. Ultrastruct. Res.* **1977**, *58*, 134–143.
172. D. R. Eyre; J. J. Wu; R. J. Fernandes; T. A. Pietka; M. A. Weis, *Biochem. Soc. Trans.* **2002**, *30*, 893–899.
173. K. E. Kadler; A. Hill; E. G. Canty-Laird, *Curr. Opin. Cell Biol.* **2008**, *20*, 495–501.
174. T. Starborg; Y. Lu; K. E. Kadler; D. F. Holmes, *Methods Cell Biol.* **2008**, *88*, 319–345.
175. T. J. Wess, *Adv. Protein Chem.* **2005**, *70*, 341–374.
176. G. Zhang; B. B. Young; Y. Ezura; M. Favata; L. J. Soslowsky; S. Chakravarti; D. E. Birk, *J. Musculoskelet. Neuronal Interact.* **2005**, *5*, 5–21.
177. E. G. Canty; K. E. Kadler, *J. Cell Sci.* **2005**, *118*, 1341–1353.
178. Y. Sado; M. Kagawa; I. Naito; Y. Ueki; T. Seki; R. Momota; T. Oohashi; Y. Ninomiya, *J. Biochem.* **1998**, *123*, 767–776.
179. M. Mariyama; A. Leinonen; T. Mochizuki; K. Tryggvason; S. T. Reeders, *J. Biol. Chem.* **1994**, *269*, 23013–23017.
180. Y. Ninomiya; M. Kagawa; K. Iyama; I. Naito; Y. Kishiro; J. M. Seyer; M. Sugimoto; T. Oohashi; Y. Sado, *J. Cell Biol.* **1995**, *130*, 1219–1229.
181. J. H. Miner; J. R. Sanes, *J. Cell Biol.* **1994**, *127*, 879–891.
182. S. Nakano; K. Iyama; M. Ogawa; H. Yoshioka; Y. Sado; T. Oohashi; Y. Ninomiya, *Lab. Invest.* **1999**, *79*, 281–292.
183. A. Simoneau; F. E. Herring-Gillam; P. H. Vachon; N. Perreault; N. Basora; Y. Bouatrouss; L. P. Pageot; J. Zhou; J. F. Beaulieu, *Dev. Dyn.* **1998**, *212*, 437–447.
184. R. Timpl; J. Risteli; H. P. Bächinger, *FEBS Lett.* **1979**, *101*, 265–268.
185. J. Risteli; H. P. Bächinger; J. Engel; H. Furthmayr; R. Timpl, *Eur. J. Biochem.* **1980**, *108*, 239–250.
186. E. Adachi; Y. Takeda; K. Nakazato; M. Muraoka; M. Iwata; T. Sasaki; Y. Imamura; I. Hopkinson; T. Hayashi, *J. Electron Microsc. (Tokyo)* **1997**, *46*, 233–241.
187. R. Timpl; H. Wiedemann; V. van Delden; H. Furthmayr; K. Kühn, *Eur. J. Biochem.* **1981**, *120*, 203–211.
188. R. Dölz; J. Engel; K. Kühn, *Eur. J. Biochem.* **1988**, *178*, 357–366.
189. K. Kühn; H. Wiedemann; R. Timpl; J. Risteli; H. Dieringer; T. Voss; R. W. Glanville, *FEBS Lett.* **1981**, *125*, 123–128.
190. P. D. Yurchenco; H. Furthmayr, *Biochemistry* **1984**, *23*, 1839–1850.
191. P. D. Yurchenco; G. C. Ruben, *J. Cell Biol.* **1987**, *105*, 2559–2568.

192. P. D. Yurchenco; G. C. Ruben, *Am. J. Pathol.* **1988**, *132*, 278–291.
193. M. E. Than; S. Henrich; R. Huber; A. Ries; K. Mann; K. Kühn; R. Timpl; G. P. Bourenkov; H. D. Bartunik; W. Bode, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6607–6612.
194. M. Sundaramoorthy; M. Meiyappan; P. Todd; B. G. Hudson, *J. Biol. Chem.* **2002**, *277*, 31142–31153.
195. M. E. Than; G. P. Bourenkov; S. Henrich; K. Mann; W. Bode, *Biol. Chem.* **2005**, *386*, 759–766.
196. R. M. Vanacore; D. B. Friedman; A. L. Ham; M. Sundaramoorthy; B. G. Hudson, *J. Biol. Chem.* **2005**, *280*, 29300–29310.
197. T. M. Mundel; R. Kalluri, *Microvasc. Res.* **2007**, *74*, 85–89.
198. D. Kajimura; S. Takahashi; K. Yoshikawa; S. Hattori; Y. Sado; Y. Imamura; T. Hayashi, *Biochem. Biophys. Res. Commun.* **2004**, *314*, 11–16.
199. S. P. Oh; M. L. Warman; M. F. Seldin; S. D. Cheng; J. H. Knoll; S. Timmons; B. R. Olsen, *Genomics* **1994**, *19*, 494–499.
200. J. C. Myers; S. Kivirikko; M. K. Gordon; A. S. Dion; T. Pihlajaniemi, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10144–10148.
201. D. Li; C. C. Clark; J. C. Myers, *J. Biol. Chem.* **2000**, *275*, 22339–22347.
202. J. C. Myers; P. S. Amenta; A. S. Dion; J. P. Sciancalepore; C. Nagaswami; J. W. Weisel; P. D. Yurchenco, *Biochem. J.* **2007**, *404*, 535–544.
203. J. C. Myers; D. Li; P. S. Amenta; C. C. Clark; C. Nagaswami; J. W. Weisel, *J. Biol. Chem.* **2003**, *278*, 32047–32057.
204. P. S. Amenta; N. A. Scivoletti; M. D. Newman; J. P. Sciancalepore; D. Li; J. C. Myers, *J. Histochem. Cytochem.* **2005**, *53*, 165–176.
205. J. C. Myers; A. S. Dion; V. Abraham; P. S. Amenta, *Cell Tissue Res.* **1996**, *286*, 493–505.
206. R. Ramchandran; M. Dhanabal; R. Volk; M. J. Waterman; M. Segal; H. Lu; B. Knebelmann; V. P. Sukhatme, *Biochem. Biophys. Res. Commun.* **1999**, *255*, 735–739.
207. T. Sasaki; H. Larsson; D. Tisi; L. Claesson-Welsh; E. Hohenester; R. Timpl, *J. Mol. Biol.* **2000**, *301*, 1179–1190.
208. Y. Muragaki; S. Timmons; C. M. Griffith; S. P. Oh; B. Fadel; T. Quertermous; B. R. Olsen, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8763–8767.
209. M. Rehn; T. Pihlajaniemi, *J. Biol. Chem.* **1995**, *270*, 4705–4711.
210. A. G. Marneros; B. R. Olsen, *FASEB J.* **2005**, *19*, 716–728.
211. D. Quélard; E. Lavergne; I. Hendaoui; H. Elamaa; U. Tiirola; R. Heljasvaara; T. Pihlajaniemi; B. Clément; O. Musso, *PLoS ONE* **2008**, *3*, e1878.
212. M. S. O'Reilly; T. Boehm; Y. Shing; N. Fukai; G. Vasios; W. S. Lane; E. Flynn; J. R. Birkhead; B. R. Olsen; J. Folkman, *Cell* **1997**, *88*, 277–285.
213. T. Sasaki; E. Hohenester; R. Timpl, *IUBMB Life* **2002**, *53*, 77–84.
214. A. Olsson; I. Johansson; H. Akerud; B. Einarsson; R. Christofferson; T. Sasaki; R. Timpl; L. Claesson-Welsh, *Cancer Res.* **2004**, *64*, 9012–9017.
215. J. P. J. Eder; J. G. Supko; J. W. Clark; T. A. Puchalski; R. Garcia-Carbonero; D. P. Ryan; L. N. Shulman; J. Proper; M. Kirvan; B. Rattner; S. Connors; M. T. Keogan; M. J. Janicek; W. E. Fogler; L. Schnipper; N. Kinchla; C. Sidor; E. Phillips; J. Folkman; D. W. Kufe, *J. Clin. Oncol.* **2002**, *20*, 3772–3784.
216. J. H. M. Schellens; M. J. Ratain, *J. Clin. Oncol.* **2002**, *20*, 3758–3760.
217. A. L. Sertié; V. Sossi; A. A. Camargo; M. Zatz; C. Brahe; M. R. Passos-Bueno, *Hum. Mol. Genet.* **2000**, *9*, 2051–2058.
218. R. Ylikärppä; L. Eklund; R. Sormunen; A. I. Kontiola; A. Utriainen; M. Määttä; N. Fukai; B. R. Olsen; T. Pihlajaniemi, *FASEB J.* **2003**, *17*, 2257–2259.
219. A. G. Marneros; D. R. Keene; U. Hansen; N. Fukai; K. Moulton; P. L. Goletz; G. Moiseyev; B. S. Pawlyk; W. Halfter; S. Dong; M. Shibata; T. Li; R. K. Crouch; P. Bruckner; B. R. Olsen, *EMBO J.* **2004**, *23*, 89–99.
220. J. Fitzgerald; C. Rich; F. H. Zhou; U. Hansen, *J. Biol. Chem.* **2008**, *283*, 20170–20180.
221. S. K. Gara; P. Grumati; A. Urciuolo; P. Bonaldo; B. Kobbe; M. Koch; M. Paulsson; R. Wagener, *J. Biol. Chem.* **2008**, *283*, 10658–10670.
222. M. L. Chu; R. Z. Zhang; T. C. Pan; D. Stokes; D. Conway; H. J. Kuo; R. Glanville; U. Mayer; K. Mann; R. Deutzmann; R. Timpl, *EMBO J.* **1990**, *9*, 385–393.
223. R. Jander; J. Rauterberg; R. W. Glanville, *Eur. J. Biochem.* **1983**, *133*, 39–46.
224. H. Furthmayr; H. Wiedemann; R. Timpl; E. Odermatt; J. Engel, *Biochem. J.* **1983**, *211*, 303–311.
225. E. Odermatt; J. Risteli; V. van Delden; R. Timpl, *Biochem. J.* **1983**, *211*, 295–302.
226. D. R. Keene; C. C. Ridgway; R. V. Iozzo, *J. Histochem. Cytochem.* **1998**, *46*, 215–220.
227. S. R. Lamandé; M. Mörgelin; N. E. Adams; C. Selan; J. M. Allen, *J. Biol. Chem.* **2006**, *281*, 16607–16614.
228. G. Klein; C. A. Müller; E. Tillet; M. L. Chu; R. Timpl, *Blood* **1995**, *86*, 1740–1748.
229. H. J. Kuo; C. L. Maslen; D. R. Keene; R. W. Glanville, *J. Biol. Chem.* **1997**, *272*, 26522–26529.
230. D. J. Bidanset; C. Guidry; L. C. Rosenberg; H. U. Choi; R. Timpl; M. Hook, *J. Biol. Chem.* **1992**, *267*, 5250–5256.
231. C. Wiberg; E. Hedbom; A. Khairullina; S. R. Lamandé; A. Oldberg; R. Timpl; M. Mörgelin; D. Heinegård, *J. Biol. Chem.* **2001**, *276*, 18947–18952.
232. M. A. Burg; E. Tillet; R. Timpl; W. B. Stallcup, *J. Biol. Chem.* **1996**, *271*, 26110–26116.
233. E. R. Motrescu; S. Blaise; N. Etique; N. Messaddeq; M. Chenard; I. Stoll; C. Tomasetto; M. Rio, *Oncogene* **2008**, *27*, 6347–6355.
234. M. Schnoor; P. Cullen; J. Lorkowski; K. Stolle; H. Robenek; D. Troyer; J. Rauterberg; S. Lorkowski, *J. Immunol.* **2008**, *180*, 5707–5719.
235. J. S. Cheng; D. B. Dubal; D. H. Kim; J. Legleiter; I. H. Cheng; G. Yu; I. Tesseur; T. Wyss-Coray; P. Bonaldo; L. Mucke, *Nat. Neurosci.* **2009**, *12*, 119–121.
236. G. J. Jobsis; P. A. Bolhuis; J. M. Boers; F. Baas; R. A. Wolterman; G. W. Hensels; M. de Visser, *Neurology* **1996**, *46*, 779–782.
237. S. R. Lamandé; K. A. Shields; A. J. Kornberg; L. K. Shield; J. F. Bateman, *J. Biol. Chem.* **1999**, *274*, 21817–21822.
238. R. A. Pace; R. A. Peat; N. L. Baker; L. Zamurs; M. Mörgelin; M. Irving; N. E. Adams; J. F. Bateman; D. Mowat; N. J. C. Smith; P. J. Lamont; S. A. Moore; K. D. Mathews; K. N. North; S. R. Lamandé, *Ann. Neurol.* **2008**, *64*, 294–303.
239. E. Mercuri; Y. Yuva; S. C. Brown; M. Brockington; M. Kinali; H. Jungbluth; L. Feng; C. A. Sewry; F. Muntoni, *Neurology* **2002**, *58*, 1354–1359.

240. H. Bentz; N. P. Morris; L. W. Murray; L. Y. Sakai; D. W. Hollister; R. E. Burgeson, *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3168–3172.
241. R. E. Burgeson; A. M. Christiano, *Curr. Opin. Cell Biol.* **1997**, *9*, 651–658.
242. G. P. Lunstrum; L. Y. Sakai; D. R. Keene; N. P. Morris; R. E. Burgeson, *J. Biol. Chem.* **1986**, *261*, 9042–9048.
243. G. P. Lunstrum; H. J. Kuo; L. M. Rosenbaum; D. R. Keene; R. W. Glanville; L. Y. Sakai; R. E. Burgeson, *J. Biol. Chem.* **1987**, *262*, 13706–13712.
244. N. P. Morris; D. R. Keene; R. W. Glanville; H. Bentz; R. E. Burgeson, *J. Biol. Chem.* **1986**, *261*, 5638–5644.
245. H. P. Bächinger; N. P. Morris; G. P. Lunstrum; D. R. Keene; L. M. Rosenbaum; L. A. Compton; R. E. Burgeson, *J. Biol. Chem.* **1990**, *265*, 10095–10101.
246. A. Rattenholl; W. N. Pappano; M. Koch; D. R. Keene; K. E. Kadler; T. Sasaki; R. Timpl; R. E. Burgeson; D. S. Greenspan; L. Bruckner-Tuderman, *J. Biol. Chem.* **2002**, *277*, 26372–26378.
247. M. Colombo; R. J. Brittingham; J. F. Klement; I. Majsterek; D. E. Birk; J. Uitto; A. Fertala, *Biochemistry* **2003**, *42*, 11434–11442.
248. M. Chen; M. P. Marinkovich; J. C. Jones; E. A. O'Toole; Y. Y. Li; D. T. Woodley, *J. Invest. Dermatol.* **1999**, *112*, 177–183.
249. R. Brittingham; J. Uitto; A. Fertala, *Biochem. Biophys. Res. Commun.* **2006**, *343*, 692–699.
250. D. T. Woodley; Y. Hou; S. Martin; W. Li; M. Chen, *J. Biol. Chem.* **2008**, *283*, 17838–17845.
251. R. Varki; S. Sadowski; J. Uitto; E. Pfendner, *J. Med. Genet.* **2007**, *44*, 181–192.
252. A. M. Christiano; I. Anton-Lamprecht; S. Amano; U. Ebschner; R. E. Burgeson; J. Uitto, *Am. J. Hum. Genet.* **1996**, *58*, 682–693.
253. R. Brittingham; M. Colombo; H. Ito; A. Steplewski; D. E. Birk; J. Uitto; A. Fertala, *J. Biol. Chem.* **2005**, *280*, 191–198.
254. H. Sage; P. Pritzl; P. Bornstein, *Biochemistry* **1980**, *19*, 5747–5755.
255. P. D. Benya, *Ren. Physiol.* **1980**, *3*, 30–35.
256. Y. Muragaki; C. Shiota; M. Inoue; A. Ooshima; B. R. Olsen; Y. Ninomiya, *Eur. J. Biochem.* **1992**, *207*, 895–902.
257. H. Sage; P. Pritzl; P. Bornstein, *Coll. Relat. Res.* **1982**, *2*, 465–479.
258. H. Sawada; H. Konomi; K. Hirosawa, *J. Cell Biol.* **1990**, *110*, 219–227.
259. S. Stephan; M. J. Sherratt; N. Hodson; C. A. Shuttleworth; C. M. Kielty, *J. Biol. Chem.* **2004**, *279*, 21469–21477.
260. N. S. Greenhill; B. M. Rüger; Q. Hasan; P. F. Davis, *Matrix Biol.* **2000**, *19*, 19–28.
261. C. Illidge; C. Kielty; A. Shuttleworth, *J. Biol. Chem.* **1998**, *273*, 22091–22095.
262. K. Mann; R. Jander; E. Korsching; K. Kühn; J. Rauterberg, *FEBS Lett.* **1990**, *273*, 168–172.
263. C. Illidge; C. Kielty; A. Shuttleworth, *Int. J. Biochem. Cell Biol.* **2001**, *33*, 521–529.
264. L. Shapiro; P. E. Scherer, *Curr. Biol.* **1998**, *8*, 335–338.
265. O. Bogin; M. Kvansakul; E. Rom; J. Singer; A. Yayon; E. Hohenester, *Structure* **2002**, *10*, 165–173.
266. M. Kvansakul; O. Bogin; E. Hohenester; A. Yayon, *Matrix Biol.* **2003**, *22*, 145–152.
267. S. Biswas; F. L. Munier; J. Yardley; N. Hart-Holden; R. Perveen; P. Cousin; J. E. Sutphin; B. Noble; M. Batterbury; C. Kielty; A. Hackett; R. Bonshek; A. Ridgway; D. McLeod; V. C. Sheffield; E. M. Stone; D. F. Schorderet; G. C. Black, *Hum. Mol. Genet.* **2001**, *10*, 2415–2423.
268. J. D. Gottsch; O. H. Sundin; S. H. Liu; A. S. Jun; K. W. Broman; W. J. Stark; E. C. L. Vito; A. K. Narang; J. M. Thompson; M. Magovern, *Invest. Ophthalmol. Vis. Sci.* **2005**, *46*, 1934–1939.
269. T. M. Schmid; T. F. Linsenmayer, *J. Biol. Chem.* **1983**, *258*, 9504–9509.
270. A. P. Kwan; C. E. Cummings; J. A. Chapman; M. E. Grant, *J. Cell Biol.* **1991**, *114*, 597–604.
271. J. F. Bateman; S. Freddi; R. McNeil; E. Thompson; P. Hermanns; R. Savarirayan; S. R. Lamandé, *Hum. Mutat.* **2004**, *23*, 396.
272. S. Ikegawa; K. Nakamura; A. Nagano; N. Haga; Y. Nakamura, *Hum. Mutat.* **1997**, *9*, 131–135.
273. O. Jacenko; D. W. Roberts; M. R. Campbell; P. M. McManus; C. J. Gress; Z. Tao, *Am. J. Pathol.* **2002**, *160*, 2019–2034.
274. L. M. Shaw; B. R. Olsen, *Trends Biochem. Sci.* **1991**, *16*, 191–194.
275. L. Vaughan; M. Mendler; S. Huber; P. Bruckner; K. H. Winterhalter; M. I. Irwin; R. Mayne, *J. Cell Biol.* **1998**, *106*, 991–997.
276. M. Koch; B. Bohrmann; M. Matthison; C. Hagios; B. Trueb; M. Chiquet, *J. Cell Biol.* **1995**, *130*, 1005–1014.
277. A. Marchler-Bauer; J. B. Anderson; M. K. Derbyshire; C. DeWeese-Scott; N. R. Gonzales; M. Gwadz; L. Hao; S. He; D. I. Hurwitz; J. D. Jackson; Z. Ke; D. Krylov; C. J. Lanczycki; C. A. Liebert; C. Liu; F. Lu; S. Lu; G. H. Marchler; M. Mullokandov; J. S. Song; N. Thanki; R. A. Yamashita; J. J. Yin; D. Zhang; S. H. Bryant, *Nucleic Acids Res.* **2007**, *35*, D237–D240.
278. S. P. Boudko; J. Engel; H. P. Bächinger, *J. Biol. Chem.* **2008**, *283*, 34345–34351.
279. D. McCormick; M. van der Rest; J. Goodship; G. Lozano; Y. Ninomiya; B. R. Olsen, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4044–4048.
280. T. Yada; S. Suzuki; K. Kobayashi; M. Kobayashi; T. Hoshino; K. Horie; K. Kimata, *J. Biol. Chem.* **1990**, *265*, 6992–6999.
281. J. Wu; D. R. Eyre, *J. Biol. Chem.* **2003**, *278*, 24521–24525.
282. D. R. Eyre; S. Apon; J. J. Wu; L. H. Ericsson; K. A. Walsh, *FEBS Lett.* **1987**, *220*, 337–341.
283. D. E. Mechling; J. E. Gambee; N. P. Morris; L. Y. Sakai; D. R. Keene; R. Mayne; H. P. Bächinger, *J. Biol. Chem.* **1996**, *271*, 13781–13785.
284. V. Leppänen; H. Tossavainen; P. Permi; L. Lehtiö; G. Rönnholm; A. Goldman; I. Kilpelaïnen; T. Pihlajamaa, *J. Biol. Chem.* **2007**, *282*, 23219–23230.
285. G. Rudenko; E. Hohenester; Y. A. Muller, *Trends Biochem. Sci.* **2001**, *26*, 363–368.
286. R. Hagg; E. Hedbom; U. Möllers; A. Aszódi; R. Fässler; P. Bruckner, *J. Biol. Chem.* **1997**, *272*, 20650–20654.
287. R. Fässler; P. N. Schnegelsberg; J. Dausman; T. Shinya; Y. Muragaki; M. T. McCarthy; B. R. Olsen; R. Jaenisch, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5070–5074.
288. K. L. Posey; K. Hankenson; A. C. Veerisetty; P. Bornstein; J. Lawler; J. T. Hecht, *Am. J. Pathol.* **2008**, *172*, 1664–1674.
289. M. Czarny-Ratajczak; J. Lohiniva; P. Rogala; K. Kozlowski; M. Perälä; L. Carter; T. D. Spector; L. Kolodziej; U. Seppänen; R. Glazar; J. Królewski; A. Latos-Bielenska; L. Ala-Kokko, *Am. J. Hum. Genet.* **2001**, *69*, 969–980.
290. P. Holden; E. G. Canty; G. R. Mortier; B. Zabel; J. Spranger; A. Carr; M. E. Grant; J. A. Loughlin; M. D. Briggs, *Am. J. Hum. Genet.* **1999**, *65*, 31–38.
291. J. Fiedler; J. Stöve; F. Heber; R. E. Brenner, *Am. J. Med. Genet.* **2002**, *112*, 144–153.
292. M. van der Rest; B. Dublet, *Semin. Cell Dev. Biol.* **1996**, *7*, 639–648.
293. B. Dublet; M. van der Rest, *J. Biol. Chem.* **1991**, *266*, 6853–6858.

294. T. Nishiyama; A. M. McDonough; R. R. Bruns; R. E. Burgeson, *J. Biol. Chem.* **1994**, *269*, 28193–28199.
295. D. Schuppan; M. C. Cantaluppi; J. Becker; A. Veit; T. Bunte; D. Troyer; F. Schuppan; M. Schmid; R. Ackermann; E. G. Hahn, *J. Biol. Chem.* **1990**, *265*, 8823–8832.
296. J. C. Brown; K. Mann; H. Wiedemann; R. Timpl, *J. Cell Biol.* **1993**, *120*, 557–567.
297. T. Ehnis; W. Dieterich; M. Bauer; H. Kresse; D. Schuppan, *J. Biol. Chem.* **1997**, *272*, 20414–20419.
298. T. C. Pan; R. Z. Zhang; M. G. Mattei; R. Timpl; M. L. Chu, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6565–6569.
299. A. Kassner; U. Hansen; N. Miosge; D. P. Reinhardt; T. Aigner; L. Bruckner-Tuderman; P. Bruckner; S. Grässel, *Matrix Biol.* **2003**, *22*, 131–143.
300. J. C. Myers; M. J. Sun; J. A. D'Ippolito; E. W. Jabs; E. G. Neilson; A. S. Dion, *Gene* **1993**, *123*, 211–217.
301. H. Sumiyoshi; K. Inoguchi; M. Khaleduzzaman; Y. Ninomiya; H. Yoshioka, *J. Biol. Chem.* **1997**, *272*, 17104–17111.
302. H. Sumiyoshi; N. Mor; S. Y. Lee; S. Doty; S. Henderson; S. Tanaka; H. Yoshioka; S. Rattan; F. Ramirez, *J. Cell Biol.* **2004**, *166*, 591–600.
303. H. Sumiyoshi; F. Laub; H. Yoshioka; F. Ramirez, *Dev. Dyn.* **2001**, *220*, 155–162.
304. K. Inoguchi; H. Yoshioka; M. Khaleduzzaman; Y. Ninomiya, *J. Biochem.* **1995**, *117*, 137–146.
305. M. Koch; J. E. Foley; R. Hahn; P. Zhou; R. E. Burgeson; D. R. Gerecke; M. K. Gordon, *J. Biol. Chem.* **2001**, *276*, 23120–23126.
306. M. Bayés; A. J. Hartung; S. Ezer; J. Pispa; I. Thesleff; A. K. Srivastava; J. Kere, *Hum. Mol. Genet.* **1998**, *7*, 1661–1669.
307. M. Penman; A. Lux; N. J. Freedman; L. Rohrer; Y. Ekkel; H. McKinstry; D. Resnick; M. Krieger, *J. Biol. Chem.* **1991**, *266*, 23985–23993.
308. J. R. M. Ojala; T. Pikkarainen; A. Tuuttila; T. Sandalova; K. Tryggvason, *J. Biol. Chem.* **2007**, *282*, 16654–16666.
309. O. Elomaa; M. Kangas; C. Sahlberg; J. Tuukkanen; R. Sormunen; A. Liakka; I. Thesleff; G. Kraal; K. Tryggvason, *Cell* **1995**, *80*, 603–609.
310. K. Nakamura; H. Funakoshi; K. Miyamoto; F. Tokunaga; T. Nakamura, *Biochem. Biophys. Res. Commun.* **2001**, *280*, 1028–1035.
311. K. Ohtani; Y. Suzuki; S. Eda; T. Kawai; T. Kase; H. Keshi; Y. Sakai; A. Fukuoh; T. Sakamoto; H. Itabe; T. Suzutani; M. Ogasawara; I. Yoshida; N. Wakamiya, *J. Biol. Chem.* **2001**, *276*, 44222–44228.
312. A. Snellman; A. Tuomisto; A. Koski; A. Latvanlehto; T. Pihlajaniemi, *J. Biol. Chem.* **2007**, *282*, 14898–14905.
313. P. Nykvist; K. Tasanen; T. Viitasalo; J. Kapyla; J. Jokinen; L. Bruckner-Tuderman; J. Heino, *J. Biol. Chem.* **2001**, *276*, 38673–38679.
314. P. Hägg; M. Rehn; P. Huhtala; T. Väisänen; M. Tamminen; T. Pihlajaniemi, *J. Biol. Chem.* **1998**, *273*, 15590–15597.
315. A. Snellman; M. R. Keränen; P. O. Hägg; A. Lamberg; J. K. Hiltunen; K. I. Kivirikko; T. Pihlajaniemi, *J. Biol. Chem.* **2000**, *275*, 8936–8944.
316. P. Hägg; T. Väisänen; A. Tuomisto; M. Rehn; H. Tu; P. Huhtala; S. Eskelinen; T. Pihlajaniemi, *Matrix Biol.* **2001**, *19*, 727–742.
317. P. Nykvist; H. Tu; J. Ivaska; J. Käpylä; T. Pihlajaniemi; J. Heino, *J. Biol. Chem.* **2000**, *275*, 8255–8261.
318. H. Tu; T. Sasaki; A. Snellman; W. Göhring; P. Pirilä; R. Timpl; T. Pihlajaniemi, *J. Biol. Chem.* **2002**, *277*, 23092–23099.
319. G. J. Giudice; D. J. Emery; L. A. Diaz, *J. Invest. Dermatol.* **1992**, *99*, 243–250.
320. T. Claudepierre; M. K. Manglapus; N. Marengi; S. Radner; M. Champliaud; K. Tasanen; L. Bruckner-Tuderman; D. D. Hunter; W. J. Brunken, *J. Comp. Neurol.* **2005**, *487*, 190–203.
321. Y. Hirako; J. Usukura; J. Uematsu; T. Hashimoto; Y. Kitajima; K. Owaribe, *J. Biol. Chem.* **1998**, *273*, 9711–9717.
322. Y. Hirako; K. Yoshino; D. Zillikens; K. Owaribe, *J. Biochem.* **2003**, *133*, 197–206.
323. C. Franzke; K. Tasanen; H. Schäcke; Z. Zhou; K. Tryggvason; C. Mauch; P. Zigrino; S. Sunnarborg; D. C. Lee; F. Fahrenholz; L. Bruckner-Tuderman, *EMBO J.* **2002**, *21*, 5026–5035.
324. S. Nonaka; A. Ishiko; T. Masunaga; M. Akiyama; K. Owaribe; H. Shimizu; T. Nishikawa, *J. Invest. Dermatol.* **2000**, *115*, 889–892.
325. J. Banyard; L. Bao; B. R. Zetter, *J. Biol. Chem.* **2003**, *278*, 20989–20994.
326. T. Hashimoto; T. Wakabayashi; A. Watanabe; H. Kowa; R. Hosoda; A. Nakamura; I. Kanazawa; T. Arai; K. Takio; D. M. A. Mann; T. Iwatsubo, *EMBO J.* **2002**, *21*, 1524–1534.
327. R. Hjorten; U. Hansen; R. A. Underwood; H. E. Telfer; R. J. Fernandes; D. Krakow; E. Sebald; S. Wachsmann-Hogiu; P. Bruckner; R. Jacquet; W. J. Landis; P. H. Byers; J. M. Pace, *Bone* **2007**, *41*, 535–542.
328. K. Sato; K. Yomogida; T. Wada; T. Yorihuzi; Y. Nishimune; N. Hosokawa; K. Nagata, *J. Biol. Chem.* **2002**, *277*, 37678–37684.
329. C. Leimeister; C. Steidl; N. Schumacher; S. Erhard; M. Gessler, *Dev. Biol.* **2002**, *249*, 204–218.
330. G. Veit; B. Kobbe; D. R. Keene; M. Paulsson; M. Koch; R. Wagener, *J. Biol. Chem.* **2006**, *281*, 3494–3504.
331. K. I. Kivirikko; T. Pihlajaniemi, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, *72*, 325–398.
332. S. R. Lamandé; J. F. Bateman, *Semin. Cell Dev. Biol.* **1999**, *10*, 455–464.
333. L. M. Hendershot; N. J. Bulleid, *Curr. Biol.* **2000**, *10*, R912–R915.
334. J. Myllyharju, *Matrix Biol.* **2003**, *22*, 15–24.
335. J. Myllyharju; K. I. Kivirikko, *Trends Genet.* **2004**, *20*, 33–43.
336. E. G. Canty; Y. Lu; R. S. Meadows; M. K. Shaw; D. F. Holmes; K. E. Kadler, *J. Cell Biol.* **2004**, *165*, 553–563.
337. R. A. Berg; D. J. Prockop, *Biochem. Biophys. Res. Commun.* **1973**, *52*, 115–120.
338. S. Jimenez; M. Harsch; J. Rosenbloom, *Biochem. Biophys. Res. Commun.* **1973**, *52*, 106–114.
339. J. Rosenbloom; M. Harsch; S. Jimenez, *Arch. Biochem. Biophys.* **1973**, *158*, 478–484.
340. J. Myllyharju, *Ann. Med.* **2008**, *40*, 402–417.
341. T. Helaakoski; P. Annunen; K. Vuori; I. A. MacNeil; T. Pihlajaniemi; K. I. Kivirikko, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 4427–4431.
342. P. Annunen; T. Helaakoski; J. Myllyharju; J. Veijola; T. Pihlajaniemi; K. I. Kivirikko, *J. Biol. Chem.* **1997**, *272*, 17342–17348.
343. L. Kukkola; R. Hieta; K. I. Kivirikko; J. Myllyharju, *J. Biol. Chem.* **2003**, *278*, 47685–47693.
344. K. Juva; D. J. Prockop; G. W. Cooper; J. W. Lash, *Science* **1966**, *152*, 92–94.
345. A. R. Walmsley; M. R. Batten; U. Lad; N. J. Bulleid, *J. Biol. Chem.* **1999**, *274*, 14884–14892.
346. M. Pekkala; R. Hieta; P. Kursula; K. I. Kivirikko; R. K. Wierenga; J. Myllyharju, *Acta Crystallogr. D. Biol. Crystallogr.* **2003**, *59*, 940–942.
347. M. Pekkala; R. Hieta; U. Bergmann; K. I. Kivirikko; R. K. Wierenga; J. Myllyharju, *J. Biol. Chem.* **2004**, *279*, 52255–52261.

348. G. L. Blatch; M. Lässle, *BioEssays* **1999**, *21*, 932–939.
349. A. K. Das; P. W. Cohen; D. Barford, *EMBO J.* **1998**, *17*, 1192–1199.
350. J. A. Vranka; L. Y. Sakai; H. P. Bächinger, *J. Biol. Chem.* **2004**, *279*, 23615–23621.
351. C. L. Jenkins; L. E. Bretscher; I. A. Guzei; R. T. Raines, *J. Am. Chem. Soc.* **2003**, *125*, 6422–6427.
352. K. Mizuno; T. Hayashi; D. H. Peyton; H. P. Bächinger, *J. Biol. Chem.* **2004**, *279*, 282–287.
353. K. Mizuno; T. Hayashi; D. H. Peyton; J. Engel; H. P. Bächinger, *FEBS J.* **2008**, *275*, 5830–5840.
354. M. A. Schumacher; K. Mizuno; H. P. Bächinger, *J. Biol. Chem.* **2006**, *281*, 27566–27574.
355. R. C. Tryon; S. D. White; D. L. Bannasch, *Genomics* **2007**, *90*, 93–102.
356. A. M. Barnes; W. Chang; R. Morello; W. A. Cabral; M. Weis; D. R. Eyre; S. Leikin; E. Makareeva; N. Kuznetsova; T. E. Uveges; A. Ashok; A. W. Flor; J. J. Mulvihill; P. L. Wilson; U. T. Sundaram; B. Lee; J. C. Marini, *N. Engl. J. Med.* **2006**, *355*, 2757–2764.
357. P. Tiainen; A. Pasanen; R. Sormunen; J. Myllyharju, *J. Biol. Chem.* **2008**, *283*, 19432–19439.
358. J. A. Vranka; H. Stadler; H. P. Bächinger, *Cell Struct. Funct.* **2009**, *34*, 97–104.
359. E. J. Miller; G. R. Martin; K. A. Piez; M. J. Powers, *J. Biol. Chem.* **1967**, *242*, 5481–5489.
360. A. H. Kang; S. Igarashi; J. Gross, *Biochemistry* **1969**, *8*, 3200–3204.
361. E. M. Click; P. Bornstein, *Biochemistry* **1970**, *9*, 4699–4706.
362. E. J. Miller, *Biochemistry* **1971**, *10*, 3030–3035.
363. E. J. Miller; L. G. Lunde, *Biochemistry* **1973**, *12*, 3153–3159.
364. T. F. Kresina; E. J. Miller, *Biochemistry* **1979**, *18*, 3089–3097.
365. S. Gay; E. J. Miller, *Arch. Biochem. Biophys.* **1979**, *198*, 370–378.
366. H. K. Kleinman; M. L. McGarvey; L. A. Liotta; P. G. Robey; K. Tryggvason; G. R. Martin, *Biochemistry* **1982**, *21*, 6188–6193.
367. N. A. Kefalides, *Int. Rev. Connect. Tissue Res.* **1973**, *6*, 63–104.
368. H. Sage; R. G. Woodbury; P. Bornstein, *J. Biol. Chem.* **1979**, *254*, 9893–9900.
369. N. A. Kefalides, *Biochem. Biophys. Res. Commun.* **1971**, *45*, 226–234.
370. R. K. Rhodes; E. J. Miller, *J. Biol. Chem.* **1979**, *254*, 12084–12087.
371. K. J. Bos; G. J. Rucklidge; B. Dunbar; S. P. Robins, *Matrix Biol.* **1999**, *18*, 149–153.
372. E. Chung; K. Rhodes; E. J. Miller, *Biochem. Biophys. Res. Commun.* **1976**, *71*, 1167–1174.
373. R. M. Gryder; M. Lamon; E. Adams, *J. Biol. Chem.* **1975**, *250*, 2470–2474.
374. J. M. Seyer; K. A. Hasty; A. H. Kang, *Eur. J. Biochem.* **1989**, *181*, 159–173.
375. A. G. Brownell; A. Veis, *J. Biol. Chem.* **1976**, *251*, 7137–7143.
376. K. I. Kivirikko; R. Myllylä, *Meth. Enzymol.* **1982**, *82* (Pt. A), 245–304.
377. Y. Wang; K. S. L. Lam; L. Chan; K. W. Chan; J. B. B. Lam; M. C. Lam; R. C. L. Hoo; W. W. N. Mak; G. J. S. Cooper; A. Xu, *J. Biol. Chem.* **2006**, *281*, 16391–16400.
378. K. I. Kivrikko; D. J. Prockop, *Arch. Biochem. Biophys.* **1967**, *118*, 611–618.
379. K. I. Kivirikko; K. Shudo; S. Sakakibara; D. J. Prockop, *Biochemistry* **1972**, *11*, 122–129.
380. K. I. Kivirikko; L. Ryhänen; H. Anttinen; P. Bornstein; D. J. Prockop, *Biochemistry* **1973**, *12*, 4966–4971.
381. R. Timpl; J. Engel, Type VI Collagen. In *Structure and Function of Collagen Types*; R. Mayne, R. E. Burgeson, Eds.; Academic Press: Orlando, FL, 1987; pp 105–143.
382. A. M. Salo; L. Sipilä; R. Sormunen; H. Ruotsalainen; S. Vainio; R. Myllylä, *Matrix Biol.* **2006**, *25*, 475–483.
383. R. Myllylä; C. Wang; J. Heikkinen; A. Juffer; O. Lampela; M. Risteli; H. Ruotsalainen; A. Salo; L. Sipilä, *J. Cell. Physiol.* **2007**, *212*, 323–329.
384. K. Takaluoma; J. Lantto; J. Myllyharju, *Matrix Biol.* **2007**, *26*, 396–403.
385. A. M. Salo; C. Wang; L. Sipilä; R. Sormunen; M. Vapola; P. Kervinen; H. Ruotsalainen; J. Heikkinen; R. Myllylä, *J. Cell Physiol.* **2006**, *207*, 644–653.
386. M. Suokas; O. Lampela; A. H. Juffer; R. Myllylä; S. Kellokumpu, *Biochem. J.* **2003**, *370*, 913–920.
387. M. Suokas; R. Myllyla; S. Kellokumpu, *J. Biol. Chem.* **2000**, *275*, 17863–17868.
388. R. Ha-Vinh; Y. Alanay; R. A. Bank; A. B. Campos-Xavier; A. Zankl; A. Superti-Furga; L. Bonafé, *Am. J. Med. Genet. A* **2004**, *131*, 115–120.
389. R. A. Bank; S. P. Robins; C. Wijmenga; L. J. Breslau-Siderius; A. F. Bardoel; H. A. van der Sluijs; H. E. Pruijs; J. M. TeKoppele, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1054–1058.
390. H. Shinkai; K. Yonemasu, *Biochem. J.* **1979**, *177*, 847–852.
391. K. Rautavuoma; K. Takaluoma; K. Passoja; A. Pirskanen; A. Kvist; K. I. Kivirikko; J. Myllyharju, *J. Biol. Chem.* **2002**, *277*, 23084–23091.
392. J. Heikkinen; M. Risteli; C. Wang; J. Latvala; M. Rossi; M. Valtavaara; R. Myllylä, *J. Biol. Chem.* **2000**, *275*, 36158–36163.
393. B. Schegg; A. J. Hülsmeier; C. Rutschmann; C. Maag; T. Hennet, *Mol. Cell. Biol.* **2009**, *29*, 943–952.
394. E. J. Miller; S. Gay, *Meth. Enzymol.* **1982**, *82* (Pt. A), 3–32.
395. R. E. Burgeson; P. A. Hebda; N. P. Morris; D. W. Hollister, *J. Biol. Chem.* **1982**, *257*, 7852–7856.
396. R. E. Burgeson, Type VII Collagen. In *Structure and Function of Collagen Types*; R. Mayne, R. E. Burgeson, Eds.; Academic Press: Orlando, FL, 1987; pp 145–172.
397. N. Quarto; R. Cancedda; F. Descalzi-Cancedda, *Eur. J. Biochem.* **1985**, *147*, 397–400.
398. D. Eyre; J. J. Wu, Type XI or $1\alpha2\alpha3\alpha$ Collagen. In *Structure and Function of Collagen Types*; R. Mayne, R. E. Burgeson, Eds.; Academic Press: Orlando, FL, 1987; pp 261–281.
399. C. Wang; H. Luosujärvi; J. Heikkinen; M. Risteli; L. Uitto; R. Myllylä, *Matrix Biol.* **2002**, *21*, 559–566.
400. M. E. Trujillo; P. E. Scherer, *J. Intern. Med.* **2005**, *257*, 167–175.
401. A. A. Richards; T. Stephens; H. K. Charlton; A. Jones; G. A. Macdonald; J. B. Prins; J. P. Whitehead, *Mol. Endocrinol.* **2006**, *20*, 1673–1687.
402. K. J. Doege; J. H. Fessler, *J. Biol. Chem.* **1986**, *261*, 8924–8935.
403. J. Koivu, *FEBS Lett.* **1987**, *212*, 229–232.
404. H. P. Bächinger, *J. Biol. Chem.* **1987**, *262*, 17144–17148.
405. J. M. Davis; B. A. Boswell; H. P. Bächinger, *J. Biol. Chem.* **1989**, *264*, 8956–8962.

406. E. Makareeva; S. Leikin, *PLoS ONE* **2007**, *2*, e1029.
407. Y. Ishikawa; J. Vranka; J. Wirz; K. Nagata; H. P. Bächinger, *J. Biol. Chem.* **2008**, *283*, 31584–31590.
408. K. Nagata, *Trends Biochem. Sci.* **1996**, *21*, 22–26.
409. K. Nagata, *Matrix Biol.* **1998**, *16*, 379–386.
410. T. Hayashi; Y. Nagai, *J. Biochem.* **1972**, *72*, 749–758.
411. K. Gekko; S. Koga, *J. Biochem.* **1983**, *94*, 199–205.
412. M. Satoh; K. Hirayoshi; S. Yokota; N. Hosokawa; K. Nagata, *J. Cell Biol.* **1996**, *133*, 469–483.
413. S. H. McLaughlin; N. J. Bulleid, *Biochem. J.* **1998**, *331* (Pt. 3), 793–800.
414. R. A. Lumb; N. J. Bulleid, *EMBO J.* **2002**, *21*, 6763–6770.
415. R. Wilson; J. F. Lees; N. J. Bulleid, *J. Biol. Chem.* **1998**, *273*, 9637–9643.
416. S. D. Chessler; P. H. Byers, *J. Biol. Chem.* **1993**, *268*, 18226–18233.
417. C. Hammond; A. Helenius, *Curr. Opin. Cell Biol.* **1995**, *7*, 523–529.
418. G. Bellamy; P. Bornstein, *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1138–1142.
419. J. D. Schofield; D. J. Prockop, *Clin. Orthop. Relat. Res.* **1973**, *97*, 175–195.
420. A. Veis; A. G. Brownell, *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 902–905.
421. N. J. Bulleid; J. A. Dalley; J. F. Lees, *EMBO J.* **1997**, *16*, 6694–6701.
422. L. Bonfanti; A. A. J. Mironov; J. A. Martínez-Menárguez; O. Martella; A. Fusella; M. Baldassarre; R. Buccione; H. J. Geuze; A. A. Mironov; A. Luini, *Cell* **1998**, *95*, 993–1003.
423. D. J. Stephens; R. Pepperkok, *J. Cell Sci.* **2002**, *115*, 1149–1160.
424. A. A. Mironov; A. A. J. Mironov; G. V. Beznoussenko; A. Trucco; P. Lupetti; J. D. Smith; W. J. C. Geerts; A. J. Koster; K. N. J. Burger; M. E. Martone; T. J. Deerinck; M. H. Ellisman; A. Luini, *Dev. Cell* **2003**, *5*, 583–594.
425. S. I. Bannykh; W. E. Balch, *J. Cell Biol.* **1997**, *138*, 1–4.
426. J. A. Martínez-Menárguez; H. J. Geuze; J. W. Slot; J. Klumperman, *Cell* **1999**, *98*, 81–90.
427. M. I. Cho; P. R. Garant, *Anat. Rec.* **1981**, *199*, 459–471.
428. C. P. Leblond, *Anat. Rec.* **1989**, *224*, 123–138.
429. H. R. Pelham, *J. Cell Biol.* **2001**, *155*, 1099–1101.
430. R. L. Trelstad; K. Hayashi, *Dev. Biol.* **1979**, *71*, 228–242.
431. M. Iwata; Y. Imamura; T. Sasaki; T. Hayashi, *J. Biochem.* **1995**, *117*, 1298–1304.
432. M. Iwata; Y. Sado; T. Sasaki; Y. Imamura; Y. Ninomiya; T. Hayashi, *J. Biochem.* **1996**, *120*, 133–137.
433. M. Muraoka; T. Hayashi, *J. Biochem.* **1993**, *114*, 358–362.
434. T. Aigner; L. Hambach; S. Söder; U. Schlötzer-Schrehardt; E. Pöschl, *Biochem. Biophys. Res. Commun.* **2002**, *290*, 743–748.
435. E. Kessler; K. Takahara; L. Biniaminov; M. Brusel; D. S. Greenspan, *Science* **1996**, *271*, 360–362.
436. S. W. Li; A. L. Sieron; A. Fertala; Y. Hojima; W. V. Arnold; D. J. Prockop, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5127–5130.
437. I. C. Scott; I. L. Blitz; W. N. Pappano; Y. Imamura; T. G. Clark; B. M. Steiglitz; C. L. Thomas; S. A. Maas; K. Takahara; K. W. Cho; D. S. Greenspan, *Dev. Biol.* **1999**, *213*, 283–300.
438. N. Hartigan; L. Garrigue-Antar; K. E. Kadler, *J. Biol. Chem.* **2003**, *278*, 18045–18049.
439. S. Porter; I. M. Clark; L. Kevorkian; D. R. Edwards, *Biochem. J.* **2005**, *386*, 15–27.
440. A. Colige; A. Beschin; B. Samyn; Y. Goebels; J. Van Beeumen; B. V. Nusgens; C. M. Lapière, *J. Biol. Chem.* **1995**, *270*, 16724–16730.
441. A. Colige; S. W. Li; A. L. Sieron; B. V. Nusgens; D. J. Prockop; C. M. Lapière, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2374–2379.
442. W. Wang; S. Lee; B. M. Steiglitz; I. C. Scott; C. C. Lebares; M. L. Allen; M. C. Brenner; K. Takahara; D. S. Greenspan, *J. Biol. Chem.* **2003**, *278*, 19549–19557.
443. A. Colige; A. L. Sieron; S. W. Li; U. Schwarze; E. Petty; W. Wertelecki; W. Wilcox; D. Krakow; D. H. Cohn; W. Reardon; P. H. Byers; C. M. Lapière; D. J. Prockop; B. V. Nusgens, *Am. J. Hum. Genet.* **1999**, *65*, 308–317.
444. S. W. Li; M. Arita; A. Fertala; Y. Bao; G. C. Kopen; T. K. Långsjö; M. M. Hyttinen; H. J. Helminen; D. J. Prockop, *Biochem. J.* **2001**, *355*, 271–278.
445. D. Greenspan, Biosynthetic Processing of Collagen Molecules. In *Collagen Primer in Structure, Processing and Assembly Topics in Current Chemistry*; J. Brinckmann, H. Notbohm, P. K. Müller, Eds.; Springer Berlin Heidelberg: New York, 2005; pp 149–183.
446. S. X. Wang; M. Mure; K. F. Medzihradszky; A. L. Burlingame; D. E. Brown; D. M. Dooley; A. J. Smith; H. M. Kagan; J. P. Klinman, *Science* **1996**, *273*, 1078–1084.
447. K. Suto; H. Noda, *Biopolymers* **1974**, *13*, 2461–2475.
448. A. V. Persikov; J. A. Ramshaw; A. Kirkpatrick; B. Brodsky, *Biochemistry* **2000**, *39*, 14960–14967.
449. N. K. Shah; J. A. Ramshaw; A. Kirkpatrick; C. Shah; B. Brodsky, *Biochemistry* **1996**, *35*, 10262–10268.
450. H. P. Bächinger; J. Engel, The Thermodynamics and Kinetics of Collagen Folding. In *Protein Folding Handbook*; J. Buchner, T. Kiefhaber, Eds.; Wiley-VCH: Weinheim, 2005; pp 1059–1110.
451. J. Engel; H. P. Bächinger, Structure, Stability and Folding of the Collagen Triple Helix. In *Collagen Primer in Structure, Processing and Assembly Topics in Current Chemistry*; J. Brinckmann, H. Notbohm, P. K. Müller, Eds.; Springer Berlin Heidelberg: New York , 2005; pp 7–33.
452. P. Bruckner; H. P. Bächinger; R. Timpl; J. Engel, *Eur. J. Biochem.* **1978**, *90*, 595–603.
453. R. W. Glanville; H. Allmann; P. P. Fietzek, *Hoppe-Seyler's Z. Physiol. Chem.* **1976**, *357*, 1663–1665.
454. H. P. Bächinger; P. Bruckner; R. Timpl; J. Engel, *Eur. J. Biochem.* **1978**, *90*, 605–613.
455. K. Mann; D. E. Mechling; H. P. Bächinger; C. Eckerskorn; F. Gaill; R. Timpl, *J. Mol. Biol.* **1996**, *261*, 255–266.
456. D. E. Mechling; H. P. Bächinger, *J. Biol. Chem.* **2000**, *275*, 14532–14536.
457. S. Boudko; S. Frank; R. A. Kammerer; J. Stetefeld; T. Schulthess; R. Landwehr; A. Lustig; H. P. Bächinger; J. Engel, *J. Mol. Biol.* **2002**, *317*, 459–470.
458. S. Frank; S. Boudko; K. Mizuno; T. Schulthess; J. Engel; H. P. Bächinger, *J. Biol. Chem.* **2003**, *278*, 7747–7750.
459. S. P. Boudko; J. Engel, *J. Mol. Biol.* **2004**, *335*, 1289–1297.
460. A. Bachmann; T. Kiefhaber; S. Boudko; J. Engel; H. P. Bächinger, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13897–13902.

461. D. Barth; O. Kyrieleis; S. Frank; C. Renner; L. Moroder, *Chemistry* **2003**, *9*, 3703–3714.
462. E. Welker; W. J. Wedemeyer; M. Narayan; H. A. Scheraga, *Biochemistry* **2001**, *40*, 9059–9064.
463. S. K. Areida; D. P. Reinhardt; P. K. Muller; P. P. Fietzek; J. Kowitz; M. P. Marinkovich; H. Notbohm, *J. Biol. Chem.* **2001**, *276*, 1594–1601.
464. A. Snellman; H. Tu; T. Väisänen; A. P. Kvist; P. Huhtala; T. Pihlajaniemi, *EMBO J.* **2000**, *19*, 5051–5059.
465. J. Engel; D. J. Prockop, *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 137–152.
466. J. Baum; B. Brodsky, *Curr. Opin. Struct. Biol.* **1999**, *9*, 122–128.
467. C. Boulègue; H. Musiol; M. G. Götz; C. Renner; L. Moroder, *Antioxid. Redox Signal.* **2008**, *10*, 113–125.
468. A. Lesage; F. Penin; C. Geourjon; D. Marion; M. van der Rest, *Biochemistry* **1996**, *35*, 9647–9660.
469. M. Mazzorana; C. Giry-Lozinguez; M. van der Rest, *Matrix Biol.* **1995**, *14*, 583–588.
470. L. Labourdette; M. van der Rest, *FEBS Lett.* **1993**, *320*, 211–214.
471. J. Jäälinoja; J. Ylöstalo; W. Beckett; D. J. S. Hulmes; L. Ala-Kokko, *Biochem. J.* **2008**, *409*, 545–554.
472. J. Ottl; D. Gabriel; G. Murphy; V. Knäuper; Y. Tominaga; H. Nagase; M. Kröger; H. Tschesche; W. Bode; L. Moroder, *Chem. Biol.* **2000**, *7*, 119–132.
473. C. G. Fields; D. J. Mickelson; S. L. Drake; J. B. McCarthy; G. B. Fields, *J. Biol. Chem.* **1993**, *268*, 14153–14160.
474. J. L. Lauer-Fields; K. A. Tuzinski; K. I. Shimokawa; H. Nagase; G. B. Fields, *J. Biol. Chem.* **2000**, *275*, 13282–13290.
475. D. S. Kemp; K. S. Petrakis, *J. Org. Chem.* **1981**, *46*, 5140–5143.
476. J. Kwak; A. De Capua; E. Locardi; M. Goodman, *J. Am. Chem. Soc.* **2002**, *124*, 14085–14091.
477. T. Koide; M. Yuguchi; M. Kawakita; H. Konno, *J. Am. Chem. Soc.* **2002**, *124*, 9388–9389.
478. Y. Yu; M. Tirrell; G. Fields, *J. Am. Chem. Soc.* **1998**, *120*, 9979–9987.
479. S. T. Khew; Y. W. Tong, *Biochemistry* **2008**, *47*, 585–596.
480. B. Brodsky; G. Thiagarajan; B. Madhan; K. Kar, *Biopolymers* **2008**, *89*, 345–353.
481. M. Mazzorana; H. Gruffat; A. Sergeant; M. van der Rest, *J. Biol. Chem.* **1993**, *268*, 3029–3032.
482. A. McAlinden; T. A. Smith; L. J. Sandell; D. Ficheux; D. A. D. Parry; D. J. S. Hulmes, *J. Biol. Chem.* **2003**, *278*, 42200–42207.
483. H. P. Bächinger; L. I. Fessler; R. Timpl; J. H. Fessler, *J. Biol. Chem.* **1981**, *256*, 13193–13199.
484. S. H. McLaughlin; N. J. Bulleid, *Matrix Biol.* **1998**, *16*, 369–377.
485. B. G. Hudson; K. Tryggvason; M. Sundaramoorthy; E. G. Neilson, *N. Engl. J. Med.* **2003**, *348*, 2543–2556.
486. S. Söder; E. Pöschl, *Biochem. Biophys. Res. Commun.* **2004**, *325*, 276–280.
487. A. Boutaud; D. B. Borza; O. Bondar; S. Gunwar; K. O. Netzer; N. Singh; Y. Ninomiya; Y. Sado; M. E. Noelken; B. G. Hudson, *J. Biol. Chem.* **2000**, *275*, 30716–30724.
488. Y. Muragaki; O. Jacenko; S. Apte; M. G. Mattei; Y. Ninomiya; B. R. Olsen, *J. Biol. Chem.* **1991**, *266*, 7721–7727.
489. Y. Ninomiya; M. Gordon; M. van der Rest; T. Schmid; T. Linsenmayer; B. R. Olsen, *J. Biol. Chem.* **1986**, *261*, 5041–5050.
490. Y. Zhang; Q. Chen, *J. Biol. Chem.* **1999**, *274*, 22409–22413.
491. J. Khoshnoodi; K. Sigmundsson; J. Cartailler; O. Bondar; M. Sundaramoorthy; B. G. Hudson, *J. Biol. Chem.* **2006**, *281*, 6058–6069.
492. S. G. Ball; C. Baldock; C. M. Kielty; C. A. Shuttleworth, *J. Biol. Chem.* **2001**, *276*, 7422–7430.
493. M. Chen; D. R. Keene; F. K. Costa; S. H. Tahk; D. T. Woodley, *J. Biol. Chem.* **2001**, *276*, 21649–21655.
494. A. Latvanlehto; A. Snellman; H. Tu; T. Pihlajaniemi, *J. Biol. Chem.* **2003**, *278*, 37590–37599.
495. T. Sasaki; N. Fukai; K. Mann; W. Göhring; B. R. Olsen; R. Timpl, *EMBO J.* **1998**, *17*, 4249–4256.
496. C. J. Kuo; K. R. J. LaMontagne; G. Garcia-Cardeña; B. D. Ackley; D. Kalman; S. Park; R. Christofferson; J. Kamihara; Y. H. Ding; K. M. Lo; S. Gillies; J. Folkman; R. C. Mulligan; K. Javaherian, *J. Cell Biol.* **2001**, *152*, 1233–1246.
497. S. Frank; R. A. Kammerer; D. Mechling; T. Schulthess; R. Landwehr; J. Bann; Y. Guo; A. Lustig; H. P. Bächinger; J. Engel, *J. Mol. Biol.* **2001**, *308*, 1081–1089.
498. J. Stetefeld; S. Frank; M. Jenny; T. Schulthess; R. A. Kammerer; S. Boudko; R. Landwehr; K. Okuyama; J. Engel, *Structure* **2003**, *11*, 339–346.
499. B. R. Olsen, *Curr. Opin. Cell Biol.* **1995**, *7*, 720–727.
500. K. Suto; H. Noda, *Biopolymers* **1974**, *13*, 2477–2488.
501. K. A. Piez; M. R. Sherman, *Biochemistry* **1970**, *9*, 4134–4140.
502. A. Yonath; W. Traub, *FEBS Lett.* **1975**, *57*, 93–95.
503. P. M. Cowan; S. McGavin; A. C. North, *Nature* **1955**, *176*, 1062–1064.
504. H. Cohen, *Br. J. Phys. Med.* **1953**, *16*, 169–170.
505. G. N. Ramachandran; G. Kartha, *Nature* **1955**, *176*, 593–595.
506. K. Okuyama; N. Tanaka; T. Ashida; M. Kakudo; S. Sakakibara, *J. Mol. Biol.* **1972**, *72*, 571–576.
507. K. Okuyama; X. Xu; M. Iguchi; K. Noguchi, *Biopolymers* **2006**, *84*, 181–191.
508. C. Hongo; V. Nagarajan; K. Noguchi; S. Kamitori; Y. Tanaka; N. Nishino; K. Okuyama, *Polym. J.* **2001**, *33*, 812–818.
509. R. Berisio; L. Vitagliano; L. Mazzarella; A. Zagari, *Protein Sci.* **2002**, *11*, 262–270.
510. R. Z. Kramer; L. Vitagliano; J. Bella; R. Berisio; L. Mazzarella; B. Brodsky; A. Zagari; H. M. Berman, *J. Mol. Biol.* **1998**, *280*, 623–638.
511. V. Nagarajan; S. Kamitori; K. Okuyama, *J. Biochem.* **1998**, *124*, 1117–1123.
512. K. Okuyama; C. Hongo; R. Fukushima; G. Wu; H. Narita; K. Noguchi; Y. Tanaka; N. Nishino, *Biopolymers* **2004**, *76*, 367–377.
513. N. Jiravanichanun; C. Hongo; G. Wu; K. Noguchi; K. Okuyama; N. Nishino; T. Silva, *ChemBioChem* **2005**, *6*, 1184–1187.
514. K. Okuyama; H. Narita; T. Kawaguchi; K. Noguchi; Y. Tanaka; N. Nishino, *Biopolymers* **2007**, *86*, 212–221.
515. M. Schumacher; K. Mizuno; H. P. Bächinger, *J. Biol. Chem.* **2005**, *280*, 20397–20403.
516. J. Bella; M. Eaton; B. Brodsky; H. M. Berman, *Science* **1994**, *266*, 75–81.
517. J. Bella; B. Brodsky; H. M. Berman, *Structure* **1995**, *3*, 893–906.
518. J. Bella; J. Liu; R. Kramer; B. Brodsky; H. M. Berman, *J. Mol. Biol.* **2006**, *362*, 298–311.
519. R. Z. Kramer; J. Bella; B. Brodsky; H. M. Berman, *J. Mol. Biol.* **2001**, *311*, 131–147.
520. J. Emsley; C. G. Knight; R. W. Farndale; M. J. Barnes, *J. Mol. Biol.* **2004**, *335*, 1019–1028.
521. Y. Zong; Y. Xu; X. Liang; D. R. Keene; A. Höök; S. Gurusiddappa; M. Höök; S. V. L. Narayana, *EMBO J.* **2005**, *24*, 4224–4236.

522. A. Rich; F. H. Crick, *J. Mol. Biol.* **1961**, *3*, 483–506.
523. G. N. Ramachandran; R. Chandrasekharan, *Biopolymers* **1968**, *6*, 1649–1658.
524. E. F. Pettersen; T. D. Goddard; C. C. Huang; G. S. Couch; D. M. Greenblatt; E. C. Meng; T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
525. G. N. Ramachandran; M. Bansal; R. S. Bhatnagar, *Biochim. Biophys. Acta* **1973**, *322*, 166–171.
526. P. L. Privalov, *Adv. Protein Chem.* **1982**, *35*, 1–104.
527. C. L. Jenkins; R. T. Raines, *Nat. Prod. Rep.* **2002**, *19*, 49–59.
528. R. Z. Kramer; M. G. Venugopal; J. Bella; P. Mayville; B. Brodsky; H. M. Berman, *J. Mol. Biol.* **2000**, *301*, 1191–1205.
529. L. Vitagliano; G. Némethy; A. Zagari; H. A. Scheraga, *Biochemistry* **1993**, *32*, 7354–7359.
530. A. V. Persikov; J. A. M. Ramshaw; B. Brodsky, *J. Biol. Chem.* **2005**, *280*, 19343–19349.
531. B. G. Hudson; S. T. Reeders; K. Tryggvason, *J. Biol. Chem.* **1993**, *268*, 26033–26036.
532. D. B. Borza; O. Bondar; Y. Ninomiya; Y. Sado; I. Naito; P. Todd; B. G. Hudson, *J. Biol. Chem.* **2001**, *276*, 28532–28540.
533. J. Zhou; M. Ding; Z. Zhao; S. T. Reeders, *J. Biol. Chem.* **1994**, *269*, 13193–13199.
534. K. O. Netzer; K. Suzuki; Y. Itoh; B. G. Hudson; R. G. Khalifah, *Protein Sci.* **1998**, *7*, 1340–1351.
535. U. Labermeier; T. A. Demlow; M. C. Kenney, *Exp. Eye Res.* **1983**, *37*, 225–237.
536. C. A. Shuttleworth, *Int. J. Biochem. Cell Biol.* **1997**, *29*, 1145–1148.
537. C. M. Kielty; A. P. Kwan; D. F. Holmes; S. L. Schor; M. E. Grant, *Biochem. J.* **1985**, *227*, 545–554.
538. G. J. Gibson; M. H. Flint, *J. Cell Biol.* **1985**, *101*, 277–284.
539. W. T. Grant; G. J. Wang; G. Balian, *J. Biol. Chem.* **1987**, *262*, 9844–9849.
540. K. von der Mark; T. Kirsch; A. Nerlich; A. Kuss; G. Weseloh; K. Glückert; H. Stöss, *Arthritis Rheum.* **1992**, *35*, 806–811.
541. J. T. Thomas; C. J. Cresswell; B. Rash; H. Nicolai; T. Jones; E. Solomon; M. E. Grant; R. P. Boot-Handford, *Biochem. J.* **1991**, *280* (Pt. 3), 617–623.
542. E. Reichenberger; F. Beier; P. LuValle; B. R. Olsen; K. von der Mark; W. M. Bertling, *FEBS Lett.* **1992**, *311*, 305–310.
543. A. Brass; K. E. Kadler; J. T. Thomas; M. E. Grant; R. P. Boot-Handford, *FEBS Lett.* **1992**, *303*, 126–128.
544. D. Chan; O. Jacenko, *Matrix Biol.* **1998**, *17*, 169–184.
545. J. Engel; R. A. Kammerer, *Matrix Biol.* **2000**, *19*, 283–288.
546. U. Kishore; K. B. Reid, *Immunopharmacology* **1999**, *42*, 15–21.
547. J. Lu; C. Teh; U. Kishore; K. B. M. Reid, *Biochim. Biophys. Acta* **2002**, *1572*, 387–400.
548. T. Doi; K. Higashino; Y. Kurihara; Y. Wada; T. Miyazaki; H. Nakamura; S. Uesugi; T. Imanishi; Y. Kawabe; H. Itakura; Y. Yazaki; A. Matsumoto; T. Kodama, *J. Biol. Chem.* **1993**, *268*, 2126–2133.
549. H. Shirai; T. Murakami; Y. Yamada; T. Doi; T. Hamakubo; T. Kodama, *Mech. Ageing Dev.* **1999**, *111*, 107–121.
550. P. Deprez; E. Doss-Pepe; B. Brodsky; N. C. Inestrosa, *Biochem. J.* **2000**, *350* (Pt. 1), 283–290.
551. Y. Xu; M. Bhate; B. Brodsky, *Biochemistry* **2002**, *41*, 8143–8151.
552. P. Sylvestre; E. Couture-Tosi; M. Mock, *J. Bacteriol.* **2003**, *185*, 1555–1563.
553. M. Rasmussen; M. Jacobsson; L. Björck, *J. Biol. Chem.* **2003**, *278*, 32313–32316.

Biographical Sketches

Hans Peter Bächinger received his Ph.D. degree from the University of Basel, Switzerland in 1979 under Professor Jürgen Engel. He worked on the stability and folding of the collagen triple helix at the Biocenter in Basel. He then worked as a postdoctoral fellow with Professor John Fessler at UCLA, studying the biosynthesis of collagens. He was appointed assistant professor of medicine and biochemistry at the Medical University of South Carolina in Charleston and a few years later moved to the Shriners Hospital for Children and Oregon Health Sciences University in Portland, OR. He is now a senior investigator at Shriners Hospital for Children and Professor in the Department of Biochemistry and Molecular Biology at Oregon Health & Science University. His research interests are the

biophysical–chemical characterization, structure, function, and biosynthesis of extracellular matrix molecules, with a major focus on collagens.

Kazunori Mizuno is a postdoctoral research associate in the laboratory of Hans Peter Bächinger at the Shriners Hospitals for Children in Portland, OR. He joined the laboratory of Professor Toshihiko Hayashi at the University of Tokyo in 1989. After 5 years of graduate school course, he was appointed as an assistant professor at the Department of Life Science (Sports and Health Sciences) in the University of Tokyo, Komaba, from 1993. He continued the collagen research with Professor Hayashi until 2001, and received his Ph.D. degree in 2001. He had an opportunity for one sabbatical year at the laboratory of Hans Peter Bächinger in 2001. After the additional 2 years of the research leave, Dr. Mizuno officially moved to Bächinger's lab. His research interests are broad ranging from the atomic molecular structure of collagen to macroscopic physiology and sports sciences related to collagen proteins.

Dr. Vranka is currently a postdoctoral research associate in the laboratory of Hans Peter Bächinger at the Shriners Hospitals for Children in Portland, OR. Dr. Vranka received her Ph.D. at Oregon Health and Science University in the Department of Biochemistry and Molecular Biology in 1997. After completing her Ph.D. in the field of MMPs and TIMPs involved in eye-related diseases, Dr. Vranka has pursued postdoctoral studies in the laboratory of Hans Peter Bächinger in the field of collagen biophysical chemistry and the study of key enzymes involved in collagen biosynthesis. She has been involved in the characterization of the family of prolyl 3-hydroxylase (P3H) enzymes and the identification of the complex of proteins consisting of P3H1 and cartilage-associated protein (CRTAP) in which mutations have recently been shown to cause osteogenesis imperfecta (OI). Recently, Dr. Vranka has characterized P3H-deficient mice in Bächinger's lab as a model system to study the underlying mechanism of a recessive form of Dr. Vranka's research is focused on understanding

how mutations in P3H1 or CRTAP in humans lead to a severe OI-like bone disorder involving the collagen architecture where there are no primary defects in collagen genes.

Sergei P. Boudko received his Ph.D. degree from the Biozentrum, the University of Basel, Switzerland, in 2003 under Professor P. D. Jürgen Engel. At the Biozentrum, he worked on the structure, stability, and folding of trimeric fibrous proteins, such as collagen and T4 phage fibritin. He continued with Professor P. D. Jürgen Engel for another year as a postdoctoral fellow. In 2004, he joined Professor Michael G. Rossmann at Purdue University, where he focused on structural studies of different viral proteins. Since 2006 he is a postdoctoral research associate in the laboratory of Professor Hans Peter Bächinger at the Shriners Hospitals for Children in Portland, OR. His research interests focus on structural and biophysical studies of collagen triple helix and collagen trimerization domains.

5.17 Lipidation of Peptides and Proteins

Kristina Görmer and Herbert Waldmann, Max Planck Institute of Molecular Physiology and Technische Universität Dortmund, Dortmund, Germany

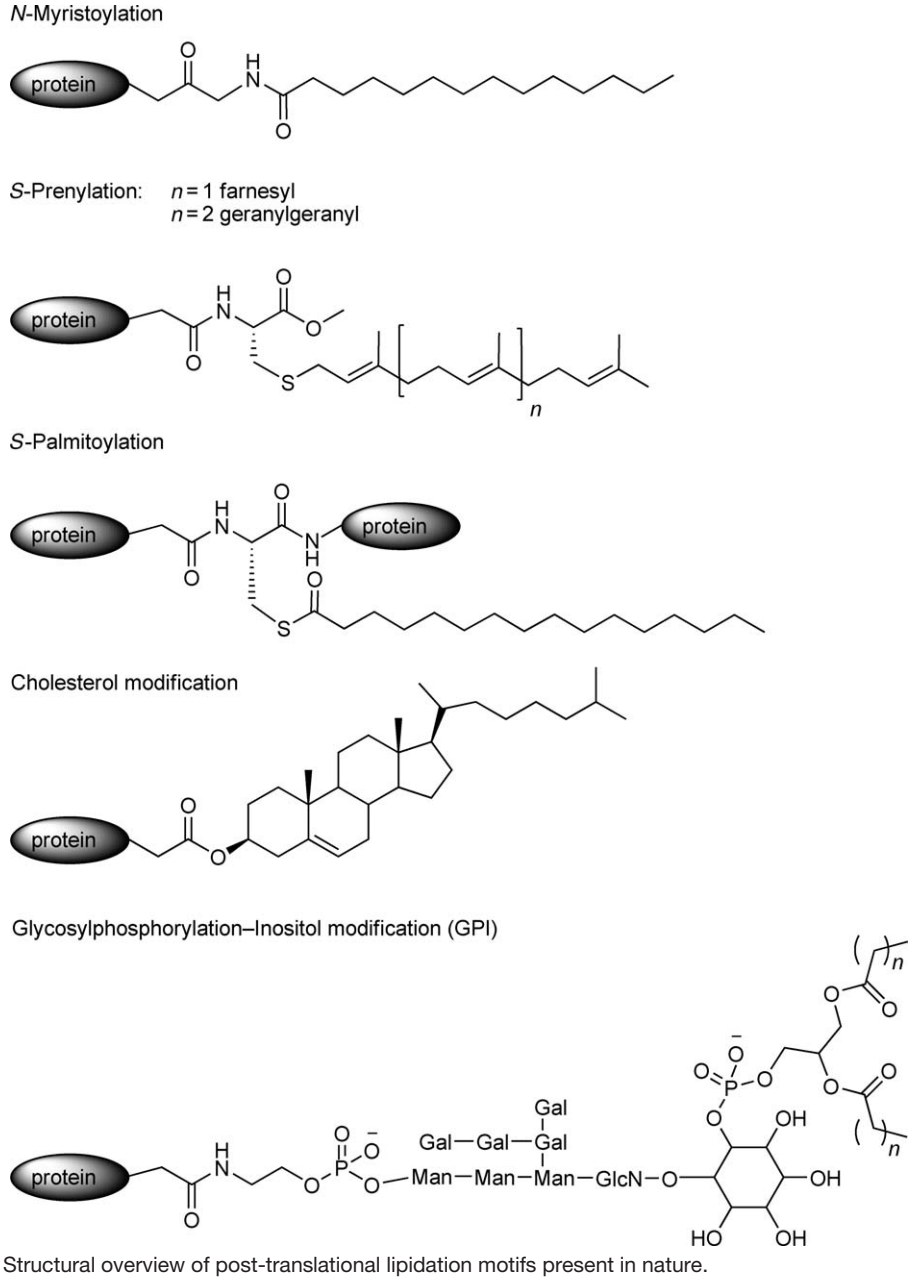
Luc Brunsveld, Max Planck Institute of Molecular Physiology and Chemical Genomics Centre of the Max Planck Society, Dortmund, Germany

© 2010 Elsevier Ltd. All rights reserved.

5.17.1 Introduction

Lipid modifications of proteins are on the forefront of scientific interest. These special co- or post-translational modifications are widely occurring in nature,[1] and it has become increasingly clear that they play an essential role in many important cellular processes such as signaling and mass transport.[2] This recognition has resulted in protein lipidation becoming a target of therapeutic interest.[3–7]

The attachment of lipid moieties to proteins leads to crucial changes in structures and physicochemical properties of the proteins, thus essentially affecting their biological activity, stability, and cellular localization. Folch discovered the first proteins with lipid modifications by extraction of rat brain in 1951.[8] However, it took 20 years to prove that a fatty acid was covalently attached to the protein. In 1978, the first *S*-farnesylated peptide, the mating factor from fungus *Rhodospiridium toruloides*, was found.[9] Ever since then the number of known intra- and extracellular proteins and peptides modified with lipid moieties is increasing. The number of known lipids that are attached co- or post-translationally to proteins is limited, especially considering, for example, the large variety of lipid molecules found in cellular membranes. The protein lipid functionalities can be divided into cytoplasm-oriented lipidations, such as N-myristoylation, S-prenylation, and S-acylation, and extracellularly oriented modifications, such as the attachment of glycosylphosphatidylinositol anchors or cholesterol (**Scheme 1**).



Scheme 1 Structural overview of post-translational lipidation motifs present in nature.

This introductory section will give an overview of the protein lipid modifications observed in nature. This section is then followed by Sections 5.17.2 and 5.17.3.

5.17.1.1 Post-Translational Lipidation of Cytoplasmic Proteins

5.17.1.1.1 N-Myristoylation

Protein myristoylation is the irreversible attachment of a myristate, a C14 saturated carboxylic acid, to an N-terminal glycine.[10] The formation of a stable amide bond occurs co-translational in eukaryotic and viral organisms. This lipidation follows the removal of an initiator methionine by cellular methionylaminopeptidases.[11] In the proapoptotic protein BH3 interacting domain death agonist (BID) myristoylation takes place post-translationally. In this case the recognition sequence for myristoylation is released upon proteolysis by caspase 8.[12] Myristoylated N-termini are found on several proteins of diverse protein families, such as tyrosine kinases of the Src family, the alanine-rich C-kinase substrate (MARCKS), the HIV Nef phosphoprotein, the α-subunit of heterotrimeric G proteins, etc.[10] The transfer of myristate is catalyzed by monomeric *N*-myristoyl transferases (NMT), which show high specificity toward the acyl donor, only accepting myristoyl-CoA. The NMT of *Saccharomyces cerevisiae* was purified and studied in 1987.[13] It recognizes a sequence motif of a substrate protein and attaches the lipid chain to the N-terminal glycine.

Myristoyl groups, being slightly shorter than, for example, palmitoyl groups, only promote weak and reversible membrane–protein and protein–protein interactions. Generally, myristoylation acts in combination with further regulation mechanisms. For example, in MARCKS and Src proteins, myristoylation provides membrane association in concert with electrostatic interactions between positively charged side chains and negatively charged membrane phospholipids via a 'myristoyl-electrostatic switch'.[14,15] 'Myristoyl-conformational switches' are observed in Arfs[16] or HIV-1 $Pr55^{gag}$/P17MA[17] in which conformational changes due to ligand binding or proteolytic cleavage lead to exposition of the acyl chain. Furthermore, myristoylation in combination with palmitoylation is encountered in the Src family of tyrosine kinases (Fyn, Lck), in the α-subunit of heterotrimeric G proteins, endothelial nitric oxide synthase, and the yeast vacuolar protein Vac8p.[18]

5.17.1.1.2 S-Prenylation

Prenylation is the covalent attachment of a lipid consisting of either three (farnesyl) or four (geranylgeranyl) isoprene units to a free thiol of a cysteine side chain at or near the C-terminus of a protein. The number of S-prenylated proteins is large including fungal mating factors, nuclear lamins, Ras and Ras-related GTP-binding proteins, the subunits of trimeric G proteins, protein kinases, and viral proteins.[19] In the early 1970s and 1980s the first evidence was found that mating factors of fungi can be prenylated.[5] The discovery that mammalian proteins can be prenylated as well took longer, for which the nuclear envelope protein Lamin B was the first example.[20–22] At about the same time, yeast genetics revealed that Ras proteins are subject to farnesylation and that this modification is essential for the ability of oncogenic Ras forms to transform cells.[23–26]

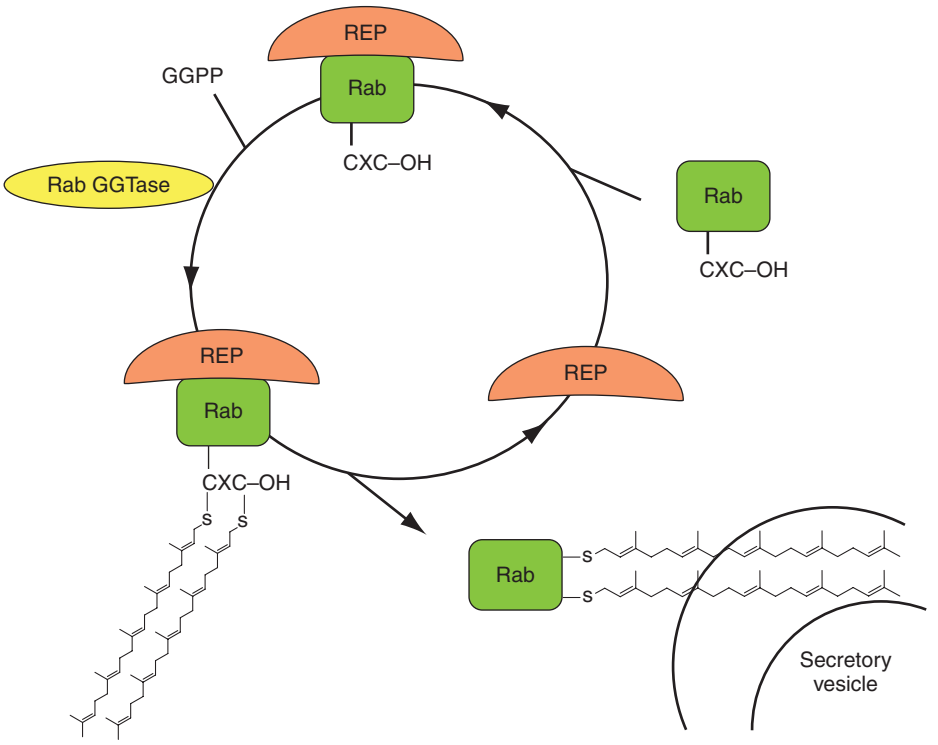
Studying the sequences of farnesylated proteins indicated that all lipidated proteins bear a cysteine residue near the C-terminus revealing the CAAX-motif, where C is a cysteine, A stands for an aliphatic amino acid, and X can be any amino acid. Database searches resulted in more prenylated proteins, all bearing the CAAX-motif, in systems from lower eukaryotes to mammals. A closer look at the mature proteins revealed that prenylation was only the first step of processing of the CAAX-motif-encoded proteins. After transfer of the isoprene unit, the last three amino acids are cleaved proteolytically by an endoprotease and the C-terminal cysteine is carboxymethylated by a methyltransferase.[27]

During the studies on farnesylation it turned out that farnesyl was not the only prenyl moiety to be attached to proteins bearing the CAAX-motif. In fact, the other isoprene residue found, the geranylgeranyl residue, is predominant in cellular proteins.[28,29] Accompanied with this finding was the discovery of geranylgeranylated proteins lacking the CAAX-motif. These proteins belong to the Rab family, a subgroup of the Ras-related G proteins. The Rab proteins are processed in a different way and possess two cysteines at the C-terminus.[30,31] Accordingly, prenylated proteins can be divided into two classes: those bearing the CAAX-motif and those that do not contain this motif, the so-called CC or CXC-containing proteins.

The CAAX-containing proteins are a diverse class of proteins for which the attachment of the farnesyl or geranylgeranyl moiety is catalyzed by either protein farnesyl transferase (FTase) or protein geranylgeranyltransferase type I (GGTase-I). These two proteins act in a similar manner, transferring the prenyl group from farnesyldiphosphate (FPP) and geranylgeranyldiphosphate (GGPP) to the thiol of the CAAX-box cysteine. The C-terminal amino acid of the CAAX-motif, the 'X', determines which isoprene unit will be incorporated. FTase recognizes serine, methionine, and glutamine, whereas leucine at the C-terminus leads to geranylgeranylated proteins.[19]

The second class of S-prenylated proteins consists only of the Rab family of the small GTP-binding proteins, which play important roles in intracellular membrane trafficking. The enzyme, catalyzing the transfer of this geranylgeranyl moiety, protein geranylgeranyltransferase type II (GGTase-II), has a different mode of action compared to FTase or GGTase-I. GGTase-II does not recognize the C-terminal short CC/CXC sequences alone as substrates. The substrate of GGTase-II is a complex of the unprenylated Rab protein and its accessory protein, termed Rab escort protein (REP). REP recognizes prenylated as well as unprenylated Rab proteins. It forms a complex with unlipidated Rab and presents the Rab protein to the catalytic dimer of GGTase-II. The complex of Rab and REP remains present until the (doubly) lipidated Rab has been successfully inserted into the target membrane. REP is then released and can act in another cycle of geranylgeranylation (**Scheme 2**).[32–34] It is important to mention the fact that most Rab proteins undergo double geranylgeranylation and this double lipidation is absolutely required for correct membrane targeting. Singly geranylgeranylated Rab mistargets to the membrane of the endoplasmic reticulum (ER) and it remains localized within the ER.[35]

Protein prenylation leads to an increased hydrophobicity of proteins, typically resulting in an increased affinity for membranes. In 2004 studies on the cellular location of prenylated RhoB proteins showed that RhoB can undergo farnesylation (RhoB-F) as well as geranylgeranylation (RhoB-GG). With the aid of specific prenyl transferase inhibitors, it was revealed that RhoB-GG is localized to multivesicular late endosomes.



Scheme 2 Intracellular processing of Rab proteins. Rab proteins form a complex with REP (Rab escort protein) that is subsequently recognized by Rab GGTase-II and leads to geranylgeranylation of Rab.

Furthermore, sorting of epidermal growth factor (EGF) receptor to the lysosome was reduced and its recycling to the membrane increased. Hence this study exemplarily indicates the importance of each prenyl moiety as a determinant for the subcellular targeting.[36]

5.17.1.1.3 S-Acylation

In contrast to N-acylation and S-alkylation, S-acylation is special as it is a reversible post-translational modification. The attachment of a fatty acid via a thioester to the thiol side chain of cysteines combines membrane targeting with a fundamental role in cellular signaling, regulating both the localization and function of proteins.[37] So far there is no structural motif known that is essential for palmitoylation apart from the presence of a cysteine. Nevertheless, palmitoylated proteins can be categorized according to their sequence context.[38] On the one hand there are proteins that are synthesized on free ribosomes associating peripherally with membranes, either solely lipidated or dually lipidated. On the other hand there is the group of palmitoylated proteins with transmembrane domains (TMD). For the lipidation site, palmitoylation is again different to myristoylation or prenylation as this modification is not limited to the C-terminus or N-terminus but is observed at different sites of the protein, frequently at pairs or longer stretches of cysteine residues. TMD proteins are found to be acylated at the interface of the cytoplasm and membrane or in the cytoplasmic tail.

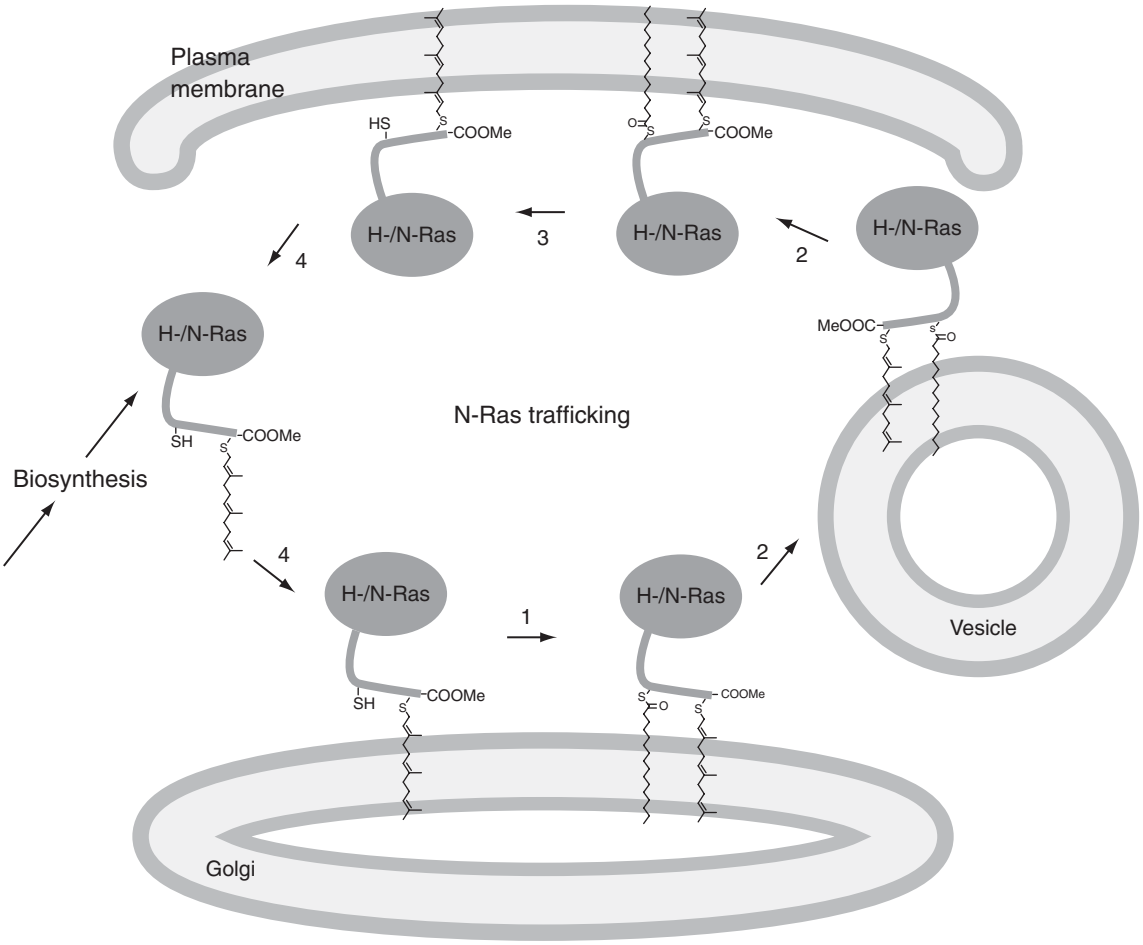
Acylation is used by the cell for diverse purposes, the most common being membrane association. The membrane affinity of soluble proteins is increased via acylation, which affects the localization and function of proteins. This is found for singly and dually lipidated proteins. The localization of doubly lipidated proteins is determined by dynamic palmitoylation and depalmitoylation. Palmitoylation is therefore a tool for regulating protein trafficking in the cell. The most prominent examples are the small Ras GTPases.

Another function of acylation is modulation of protein stability, serving as a quality check. For instance, the yeast chitin synthase Chs3, which has six to eight TMD, is only palmitoylated if the protein is folded properly. If misfolding has appeared, palmitoylation does not occur thus leading to aggregation of Chs3 and retainment in the ER.[39] Another example for this check point function is the SNARE protein Tlg1, which is located at the Golgi. The interaction with the E3 ubiquitin ligase of the protein is prevented if the protein is palmitoylated.[40] Thus, palmitoylation can be seen as a tool which the cell uses for modulating proteins' stability.

Proteins can undergo different rounds of palmitoylation and depalmitoylation, either constitutively or as a response to signals.[41,42] Here the Ras proteins are the most commonly discussed examples. As described above, all Ras proteins are expressed with the CAAX-box and are subject to post-translational modifications. First, they get farnesylated and after proteolysis and methylation of the C-terminus, H-/N-Ras as well as K-Ras 4A get further palmitoylated at additional cysteines present in their C-terminus. Palmitoylation occurs in the Golgi apparatus and via vesicular transport the farnesylated and palmitoylated proteins are directed to the plasma membrane (PM). The palmitoyl thioester is hydrolyzed at multiple cellular sites and the protein is transported back to the Golgi via a nonvesicular pathway (**Scheme 3**).[41]

Deacylation is an enzyme-mediated process; however, deacylating enzymes have not unambiguously been proven.[43,44] Similarly, an enzymatic S-palmitoylation has not been conclusively proven yet. Evidence has been found for both, an enzyme-mediated[45,46] and a nonenzymatic process.[47,48] The spontaneous acylation model shows no involvement of any proteins in the transfer of palmitate. The thiol group is believed to be reactive enough to react with acyl-CoA.[49] The high local level of acyl-CoA in mitochondria seems to be responsible for the acylation of several mitochondrial enzymes.[50] Kummel and coworkers proposed self-palmitoylation of the yeast transport protein Bet3. Palmitoylation is crucial for the stability of the protein. Bet3 was found to incorporate palmitate stoichiometrically when incubated with palmitoyl-CoA at physiological concentration and pH.[51] Furthermore, palmitoylated Bet3 can still be observed in yeast strains in which the palmitoylating enzymes have been deleted.[52]

According to the second hypothesis the incorporation of palmitate is mediated by protein *S*-acyl transferases (PAT). Biochemical approaches have demonstrated PAT activity in membranes but until the recent work of Linder and Deschenes,[53] the molecular identification of these biocatalysts having PAT activity and thus proving the PAT hypothesis by both genetic and biochemical criteria was not possible. In 1994, Deschenes and coworkers showed a Ras mutant (Ras2), which cannot get farnesylated due to a C-terminal extension with basic amino acids.[54] However, Ras2 still gets palmitoylated at the cysteine residue like the wild type, and its palmitoylation is essential for the viability of the yeast. Having this mutant in hand, screenings for Ras2 PAT candidates could be performed,



Scheme 3 Cycle of N-Ras trafficking. Farnesylated N-Ras is palmitoylated in the Golgi (1), via vesicular transport it targets the PM (2), and after deacylation (3) it trafficks back to the Golgi (4).

which led to the identification of Erf2 and Erf4 (effect on Ras2 function). Erf2 and Erf4 form a protein complex, which is localized at the ER where Ras2 is further processed before trafficking to the PM. Deletion of one of these proteins led to reduction of palmitoylation, thus leading to relocalization of Ras2 to internal membranes. The PAT activity of the Erf2/Erf4 complex could be proven by incubation of purified Erf2/Erf4 complex with Ras2 and palmitoyl-CoA.[55,56] Meanwhile Davis and coworkers identified a PAT candidate for casein kinase 2, AKR1, a yeast protein with two additional TMDs and a series of N-terminal ankyrin repeats.[57] Notable is that both Erf2 and Ark1 possess a DHHC (Asp–His–His–Cys)-motif embedded in a cysteine-rich domain (CRD), which is critical for PAT activity on Ras *in vitro* and for Ras function *in vivo*. Furthermore remarkable is the fact that Erf2 is only active in complex with Erf4 whereas Akr1 does not require any accessory protein for PAT activity. Owing to the similarity between yeast and mammalian cells in terms of lipidation, mammalian orthologues with PAT activity were sought. By iterative basic local alignment search tool (BLAST) searches DHHC9 and GCP16 (Golgi complex protein of 16 kDa) could be found as orthologues to Erf2 and Erf4, localizing as a complex at the Golgi. By kinetic studies Swarthout *et al.* revealed the catalytic nature of the palmitate transfer.[58]

Erf2 and Akr1 can thus be seen as the founding members of a family of putative PATs. In databases, over 120 DHHC-CRD genes have been found so far, with over 20 in humans.[59] The DHHC-CRD-family of proteins contains 7 members in yeast and 23 in humans. The localization of the PATs determines where soluble proteins get stably associated to the membrane.[60] In yeast, the DHHC proteins are widely distributed on membranes. In 2006, a global analysis of palmitoylation in yeast was performed, showing the great impact of DHHC proteins on the cellular palmitoylation events.

So far the debate is still hot, but it can be stated that most likely a protein activity is required for palmitoylation. However, it still has to be elucidated if this protein activity is required in a chaperone-like way that is not involved in the catalytic process or if the palmitate transfer is enzyme mediated.[61]

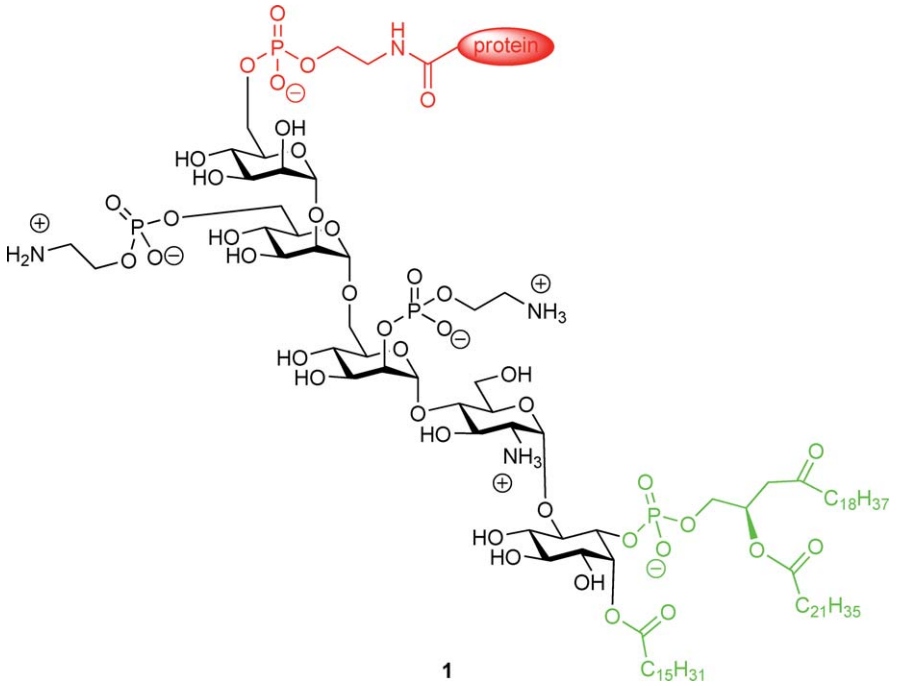
5.17.1.2 Post-Translational Lipidation of Extracellularly Oriented Proteins

5.17.1.2.1 GPI-Anchored proteins

About 10–20% of all transmembrane proteins that are targeted to the ER and subsequently enter the secretory pathway are subject to post-translational modification with glycosylphosphatidyl-inositol (GPI).[62] Proteins bearing the GPI anchor are involved in signal transduction, immune response, cancer cell invasion, and metastasis and the pathobiology of trypanosomal parasites.[63] The structure of the GPI anchor has been analyzed for mammals, protozoa, and yeast.[64–69] The general structure of the glycolipid structure is shown in Scheme 4.

The C-terminus of the modified protein is linked to a phosphoethanolamine unit that is positioned at the primary hydroxyl group of a mannose belonging to the conserved glycan core (Manα(1,2)Manα(1,6)Manα(1,4)GlcNα(1,6)-*myo*-inositoyl). The core structure is further modified during the biosynthesis in yeasts and mammals. Acylation takes place at the 2-hydroxyl group of the inositol. This palmitoyl moiety leads to stability toward the cleavage by a phospholipase C. Furthermore, various ethanolamine phosphates can be attached to the core mannoses. In the Golgi, further incorporation of various carbohydrates can take place. After GPI is transferred to the protein, the palmitoyl moiety attached to the inositol is removed in the ER, and remodeling of the diacylglycerol unit present on the GPI precursor is initiated. The lipid chains that are attached to the glycerol part may also vary, being either saturated or unsaturated and differing in length.

It is obvious that the biosynthesis of such a complex structure involves many proteins and steps. In fact, the anchor is synthesized in the ER, requiring a membrane-bound multistep pathway in which more than 20 gene products, mainly polytopic membrane proteins, take part.[44] The first two steps of the biosynthesis occur on the cytoplasmic site of the ER, and after flipping to the lumen of the ER the biosynthesis is completed. The GPI



1

Scheme 4 Structure of the native GPI anchor of human erythrocyte acetylcholineesterase, divided into three main parts: the phosphoethanolamine linker (red), the glycan core (black), and the phospholipid tail (green).

precursor is then transferred to proteins that have been translocated across the ER membrane displaying a signal sequence at the C-terminus. After secretion, GPI proteins target the cell surface; in yeast a covalent attachment to the cell wall was detected.

So far GPI biosynthesis raises many questions as its correct regulation is crucial for viability in yeast and for embryonic development in mammals. Many steps remain unclear, also due to the fact that most of the proteins involved still have to be characterized. The same is true for the biological function of the GPI anchor apart from membrane insertion. Several suggestions have been made; according to which GPI anchors are targeting lipid rafts, specific intracellular compartments, or the apical membrane of polarized epithelial cells.[70–78]

5.17.1.2.2 Post-Translational lipidation of secreted proteins

In metazoans, pattern formation is controlled by secreted protein signals. Acting as morphogens, these secreted proteins elicit concentration-dependent responses in cells surrounding a localized site of signal production and release. Two important families of signaling proteins, the Hedgehog and Wnt proteins, get post-translationally lipidated. Lipidation has great impact on their distribution, thus on their signaling activity. Hedgehog and Wnt morphogens are important for specifying the pattern of proliferation and differentiation in many tissues and structures during embryogenesis.[79]

Hedgehog proteins play an important role in the development of metazoans as they control patterning, growth, and cell migration.[80] Their correct regulation is required to prevent tumorigenesis in the adults too.[81] Hedgehog proteins are expressed as proproteins, which need to undergo post-translational modification to reach their mature state. When the proprotein enters the secretory pathway the N-terminal signal sequence is cleaved. This is followed by the autocatalytic cleavage of the 45 kDa proprotein, which leads to a 19 kDa N-terminal fragment. In an intein-like mechanism the C-terminus is cleaved off and replaced by cholesterol.[82,83] The cholesteryl moiety inserts into the lipid bilayer of the cell membrane, thereby restricting the spatial spread of the mature signal, which subsequently affects the cellular response of tissue. Cholesterol is also believed to have an impact on intracellular and extracellular trafficking and localization of the signal.

The C-terminus mediates the autocatalytic cleavage. In two nucleophilic displacements the active morphogen is generated. In an N, S-shift, the thiol of a cysteine attacks the carbonyl, leading to a thioester, which is subsequently attacked by the hydroxyl group of the cholesterol.[79] The N-terminal fragment, which is then covalently linked to cholesterol at its C-terminus, undergoes further lipidation. A palmitate is added to a cysteine residue near the N-terminus by an ER transmembrane protein.[84] The Hedgehog protein needs to be fully lipidated to gain complete signaling activity.[85–89] If it is palmitoylated and covalently linked to cholesterol, the Hedgehog protein shows high affinity to cell membranes. The signal transduction pathway of the Hedgehog protein as well as the complete function of the lipid residues, which are attached to it, still have to be elucidated.

The second family of secreted proteins that is covalently lipidated is the family of Wnt proteins. They are also involved in numerous processes like proliferation of stem cells,[90] specification of the neural crest,[91] and the expanding of specific cell types. The correct regulation of this pathway is important for animal development.[92] Willert and coworkers were the first to isolate an active Wnt molecule. Mass spectroscopy studies carried out with the isolated protein revealed that cysteine 93 is palmitoylated. Mutating this amino acid to alanine led to almost complete loss of the signaling activity.[93] Later in 2006, a second lipidation was found on a serine in Wnt3a.[94] In this case, the hydroxyl side chain is acylated with palmitoleic acid. This unsaturated fatty acid seems to be crucial for the progression of the protein through the secretory pathway. The attachment of two different lipid chains may therefore serve different functions.[95]

5.17.2 Synthesis of Lipidated Peptides

5.17.2.1 General Considerations

Lipidated peptides can be synthesized following different strategies, such as solid- or solution-phase techniques, or the use of lipidated building blocks versus lipidation of the peptides; also, the choice of the protecting group strategy is important whereas for the temporary protecting group there is the *tert*-butoxycarbonyl (Boc) strategy versus 9-fluorenylmethoxycarbonyl (Fmoc) strategy and for permanent protecting groups enzymatic or noble metal-sensitive groups can be used. The type and number of lipid groups, their position in the peptide,

and the length of the peptide sequence determine the strategy to be followed. Furthermore, C-terminal functionalization has to be taken into account for the choice as well as the optional presence of additional functional groups such as fluorescent markers, spin labels, and of course the purification strategy. With regard to the lipids and functional groups the synthetic strategy has to be fully adapted to the reactivities of the modifications. When planning a solid-phase synthesis of lipidated peptides, choosing a suitable linker to the resin is an additional criterion. Therefore, some general guidelines and compound characteristics have to be considered for the synthesis of lipidated peptides:

- Prenyl groups, such as farnesyl or geranylgeranyl, cannot be combined with strong acid-labile protecting groups or linker systems because acids attack the double bonds and lead to isomerization of addition products.[96] The geranylgeranyl group is in this respect the more sensitive prenyl group.
- Similarly, protecting groups susceptible to hydrogenolytical cleavage cannot be used in combination with prenyl moieties, as the unsaturated carbon–carbon bonds do not survive the cleavage conditions.
- Thioesters, such as palmitoylated cysteines, are susceptible to cleavage by nucleophiles.[97]
- In line with the above-mentioned reactivity of cysteine thioesters is the occurrence of a S, N-acyl shift of the palmitoyl group from the thiol side chain to the α-amino group, when this amino group is present as a free amine.
- Several prenylated dipeptide methyl esters undergo rapid diketopiperazine formation upon N-terminal deprotection.
- Finally, it should be considered that additional non-natural functional groups, which are often incorporated in the lipidated peptides for biological studies such as fluorophores or photoactive groups, typically lead to additional restrictions for the synthesis protocols.

The purification of conventional peptides typically is performed using preparative high-performance liquid chromatography (HPLC) on reversed-phase silica. Lipidated peptides, however, feature significantly different physicochemical properties in comparison with nonmodified peptides. Depending on the number and type of lipid functionalities, the water solubility is significant reduced. These peptides often display detergent-like characteristics. Therefore the purification of lipidated peptides is a demanding task, for which typically no general protocol exists. Fortunately, the often strongly dissimilar nature of the coupling partners, concerning, for example, hydrophobicity, can frequently be used to facilitate the separation of product and starting compounds or, for example, the separation of double lipidated from single lipidated peptide.

Considering all these features and limitations, it is understandable that currently there is no general procedure available, such as the Fmoc or Boc protocols that are available for nonfunctionalized peptides. However, several approaches have been developed in the last two decades to make lipidated peptides accessible. This chapter describes both the corresponding solution-phase approaches and solid-support approaches for the synthesis of lipidated peptides. In line with the framework sketched above, both the different protecting groups and solid-phase linker systems that have been developed will be reviewed.

5.17.2.2 Solution-Phase Approach for the Synthesis of Lipopeptides

The flexibility of the peptide synthesis is often determined by the number of different types of lipids that need to be incorporated in the peptide. An increasing number of lipids of one type also narrow down the number of approaches that can be followed, but the uniform nature of the lipids in this case typically allows for less demanding strategies than when different types of lipids are to be incorporated. In general, there are two general strategies to incorporate lipid chains in peptides. The lipid moiety can be introduced using prelipidated amino acid building blocks or the attachment of the lipid(s) is carried out on the peptide backbone, either after the complete synthesis of the peptide or stepwise during the synthesis.

5.17.2.2.1 Synthesis of lipopeptides containing one type of lipid group

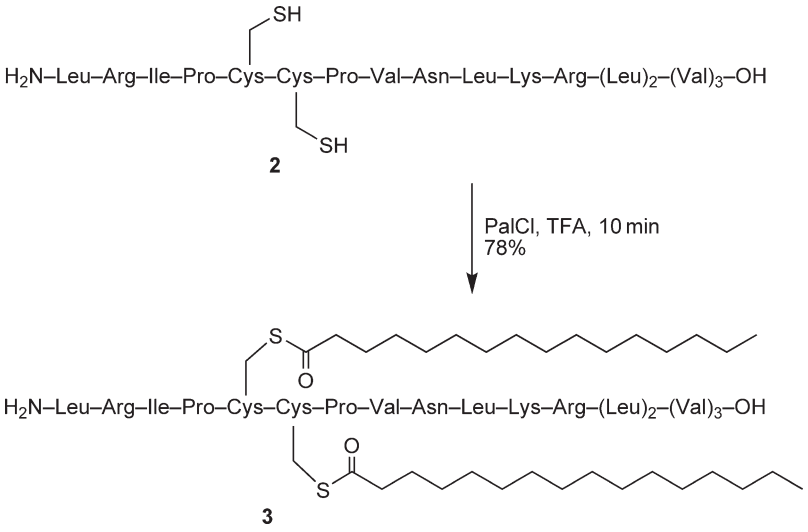
If only one type of lipid modification is incorporated, the synthesis can be performed using standard protecting group strategies according to the particular sensitivity of the lipid moiety. Palmitoyl thioesters are acid-stable, therefore the Boc strategy can be used for the buildup of the peptide. For the acid-labile prenyl residues the Fmoc strategy can typically be used. In principle, however, other protecting groups, such as highly acid-labile,

enzymatic-labile, or noble metal-labile protecting groups, can also be used. Myristoyl amides are generally stable and tolerate basic as well as acidic reaction conditions and therefore typically do not limit the protecting group strategies.

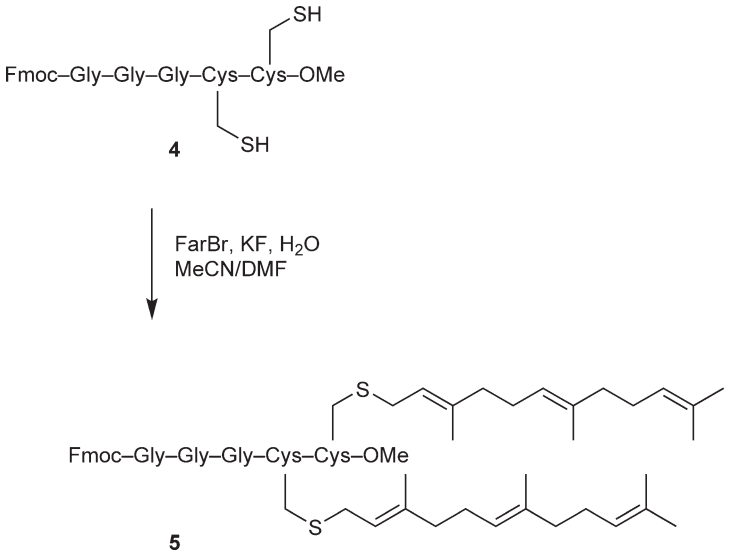
5.17.2.2.1(i) Synthesis of N-myristoylated peptides The synthesis of myristoylated peptides can be seen as the least complex of the lipopeptide syntheses because myristoylation occurs at the N-terminal glycine of signal transducing proteins and both the lipid chain and the amide bond are neither basic nor acid labile. As such the incorporation of a myristoyl functionality does not impose limitations on the protecting group strategy. The lipid modification can be incorporated either via use of prelipidated building blocks or by direct acylation of the peptide N-terminus. Both approaches are compatible with solid-phase synthesis, which is typically favored for this type of peptides.[98–100]

5.17.2.2.1(ii) Synthesis of S- or O-palmitoylated peptides The synthesis of palmitoylated peptides is limited due to a possible shift of the palmitoyl group either from O to N on serines, or the even faster S to N shift on cysteines, as soon as the amine group is not protected.[101] This characteristic implies that a sequential N-terminal elongation synthesis of palmitoylated peptides with prelipidated building blocks is challenging. The number of reports on the synthesis of such peptides is rare, especially considering their significant occurrence in nature. One way to generate S- or O-palmitoylated peptides is via introduction of the palmitoyl functionality after the assembly of the peptide backbone. Typically, the acylation is performed on a selectively deprotected peptide, in which the cysteine side chains are unmasked, but all other reactive side chain functionalities are protected, thus allowing the use of a large excess of acylation reagent without acylation of noncysteine side chains.[102] In 1999 Strömberg *et al.* published the S- and O-acylation of unprotected peptides in trifluoroacetic acid (TFA). The acidic conditions temper the reactivity of the amine functionalities. To obtain the double acylated peptide in a yield of 78%, the palmitoylation reagent was used in 20-fold excess, without detecting N-acylation (**Scheme 5**).[103]

5.17.2.2.1(iii) Synthesis of S-prenylated peptides Various syntheses of S-prenylated peptides[104] have relied on the assembly of the peptide backbone on the solid phase, using standard protocols, and subsequent S-prenylation in solution. As a consequence, any final deprotection steps of other amino acids after the S-prenylation cannot be performed under strongly acidic conditions due to the acid lability of the prenyl group. The prenylation reactions themselves can be carried out under basic or mildly acidic conditions. Typical synthesis problems that arise during the S-alkylation are: (1) incomplete conversion because of solubility problems, (2) oxidation of the thiol group to disulfides under basic conditions, (3) formation of the sulfonium



Scheme 5 Palmitoylation under acidic conditions using 20-fold excess of palmitoyl chloride.



Scheme 6 Double farnesylation under basic conditions using potassium fluoride as a base.

ion when the prenyl halide is used in excess, (4) prenylation of other functional groups, and (5) hydrolysis of the alkylation reagent.

When using basic conditions the major concern is the formation of disulfides. These however can typically be ruled out by working under inert gas atmosphere. The prenylation of free amines can be excluded by using only one equivalent of alkylation reagent. To overcome solubility problems and thereby achieving a rapid and complete conversion, solvent systems like liquid ammonia in methanol,[105] pure *N*,*N*-dimethylformamide (DMF), or dimethyl sulfoxide (DMSO)/DMF/acetonitrile (MeCN) mixtures have been used.[106] For example, the double farnesylation of peptide **4** could be achieved in an MeCN/DMF solvent mixture by treatment with an excess of farnesyl bromide in the presence of potassium fluoride dehydrate (**Scheme 6**).[107] By adding potassium fluoride as a base the yield and quality of prenylation reactions can be increased. The neutral catalyst leads to faster and cleaner reactions in organic solvents even if the alkylating reagent shows low reactivity.[108] This protocol can generally be applied for the alkylation of cysteine thiol groups.[109]

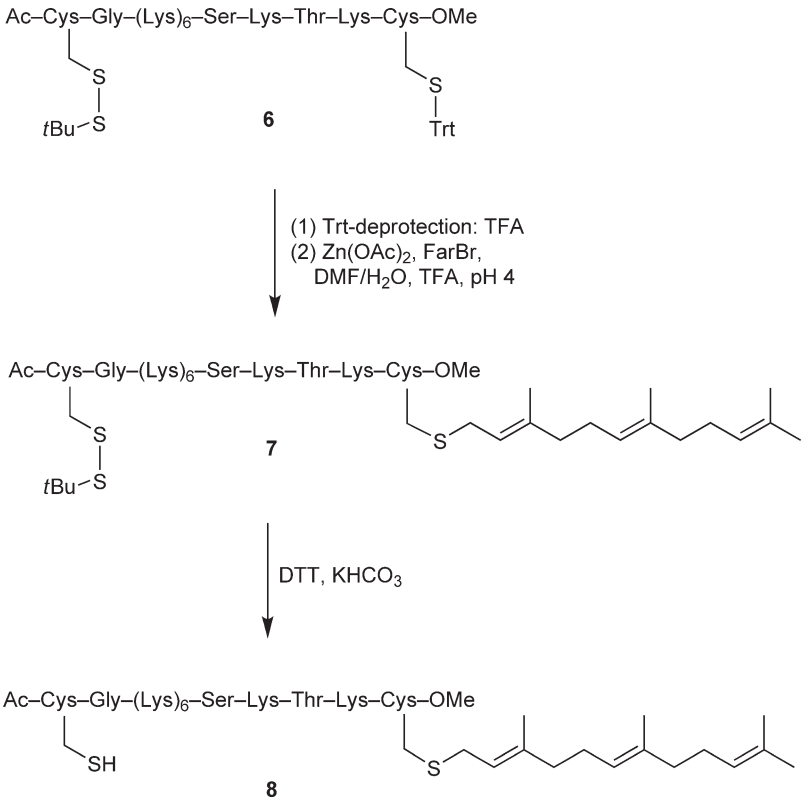
The synthesis of peptide **7** used an improved solubility in acidic medium. Using zinc acetate, efficient alkylation was achieved in aqueous/organic solvent mixtures containing TFA or AcOH at mild pH to give access to the C-terminus of K-Ras B.[110] Since two cysteines are present in the sequence two different protecting groups were chosen to allow selective deprotection. The trityl group can be removed with TFA, in contrast to the disulfide, which is stable under this condition but after alkylation can be removed reductively using dithiothreitol (DTT) (**Scheme 7**).

Alternatively to the incorporation of prenyl groups in peptides using organic chemistry, prenyl functionalities can also be introduced in peptides and proteins, using the corresponding enzymes such as FTase or GGTase. These prenyltransferases show great substrate tolerance allowing also the introduction of modified farnesyl and geranylgeranyl moieties.[111–113] This enzymatic approach will be discussed in detail in Section 5.17.3.2.

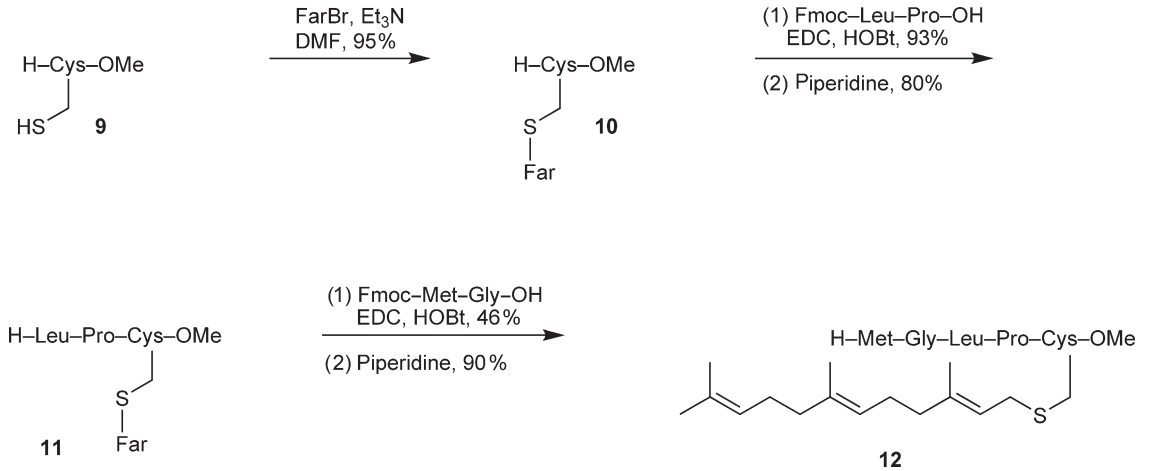
As an alternative strategy, prelipidated building blocks can be used to generate prenylated peptides (**Scheme 8**). Thereby advantage is taken of the stability of the thioether linkage that offers the possibility of applying, for example, the Fmoc strategy for such peptides. Short N-Ras C-termini, for example, are readily available via this method.[96]

5.17.2.2.2 Synthesis of peptides containing different lipid groups

For the incorporation of different types of lipid moieties, the synthetic strategies become more complex; thus a more sophisticated protecting group assembly is required. In **Figure 1** an overview of different options is given. The combination of prenyl and palmitoyl groups significantly limits the possible protecting group strategies.



Scheme 7 Selective deprotection of the trityl group for farnesylation under acidic conditions and subsequent deprotection of the second cysteine under reductive conditions.



Scheme 8 Synthesis of the farnesylated N-Ras C-terminus; the cysteine methyl ester is first farnesylated, subsequently the sequence is assembled.

Therefore the development of protecting groups that can be cleaved under mild conditions in the presence of the acid-labile prenyl moiety as well as of the nucleophile-labile palmitoyl group is required. Below, two powerful strategies will be described that allowed the successful synthesis of peptides bearing more than one type of lipid chain: the application of enzyme-labile protecting groups and a noble metal-sensitive blocking group.[114]

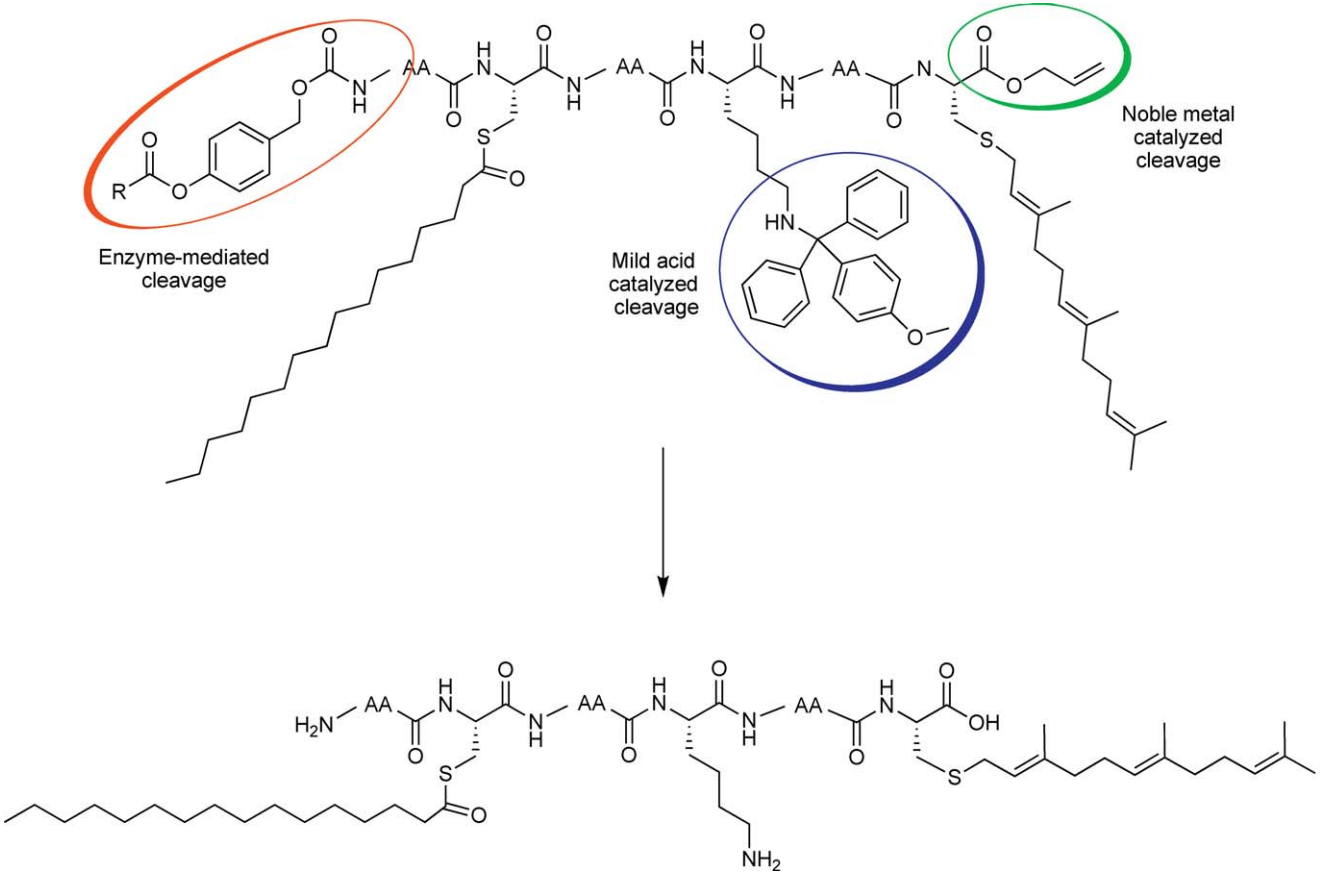
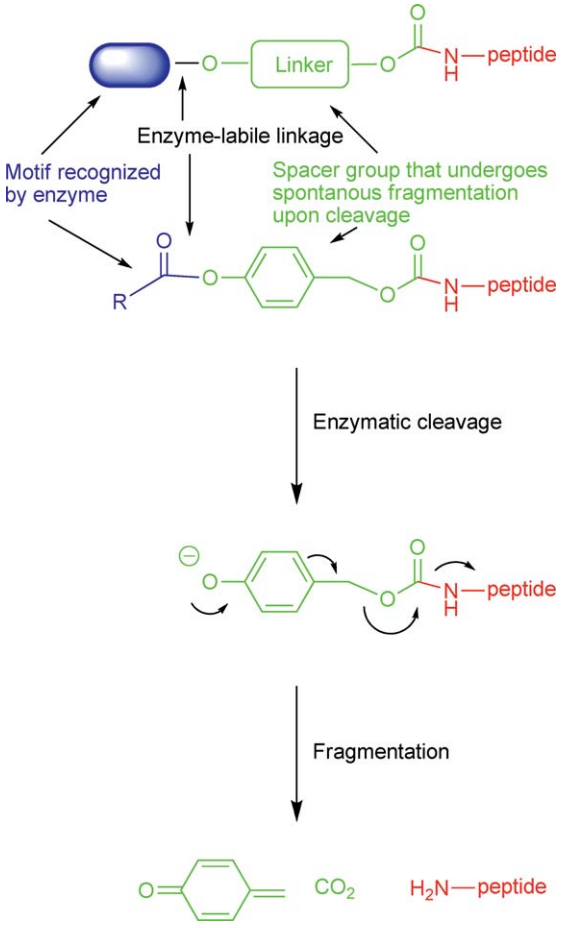


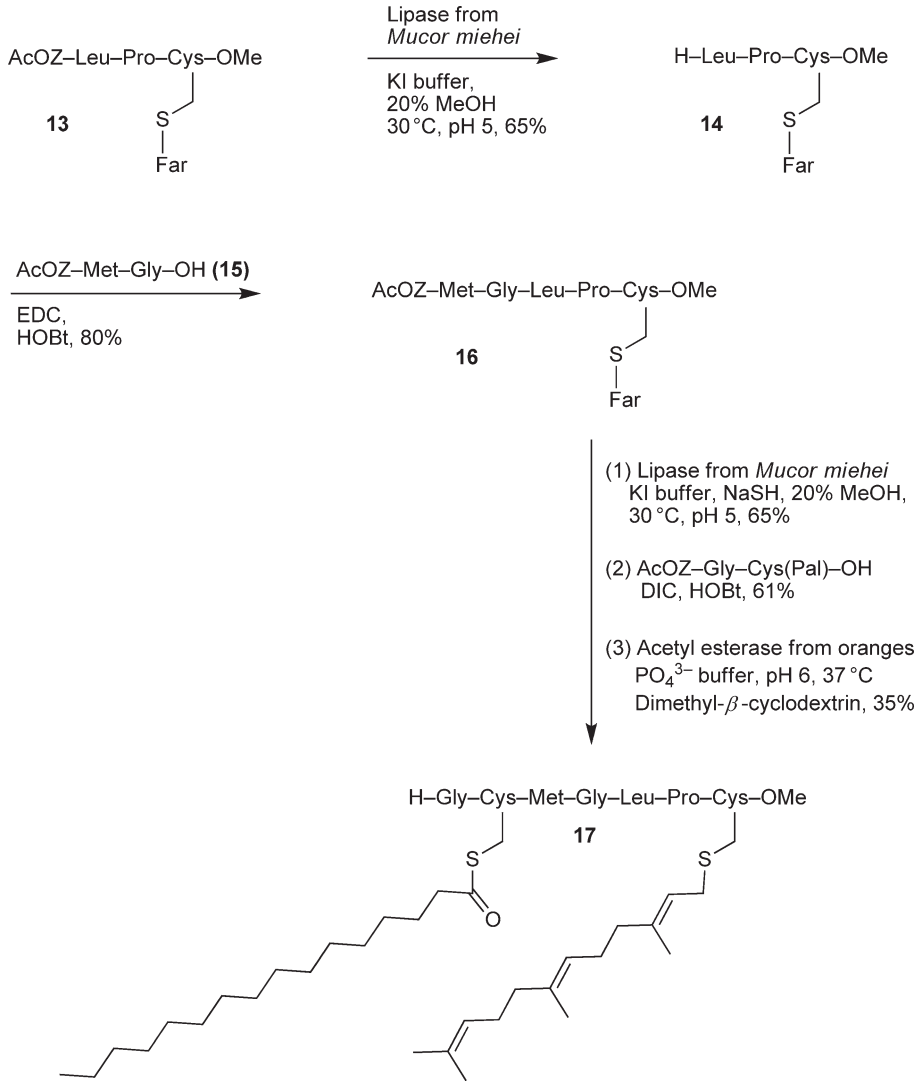
Figure 1 Overview of different protecting groups that have been used in solution-phase synthesis to access multiply lipidated peptides.

5.17.2.2.2(i) Enzyme-labile amine protecting groups The development of an enzyme-labile amine protecting group, which meets all restrictions of farnesylated as well as palmitoylated peptides, was challenging because this blocking group would need to be cleavable under mild conditions to avoid, on the one hand addition to the prenyl double bonds in acidic environment, and on the other hand base-mediated hydrolysis of the palmitoyl thioester. Furthermore, the cleaving enzyme should not harm other functional groups present in the peptide. The protecting group should embody a specific functionality recognized by a biocatalyst and furthermore, to avoid racemization upon amino acid activation, it should include a urethane motif. As most of the available enzymes do not attack urethane structures, a biocatalyst was chosen, which attacks *O*-alkyl or ester bonds, which are linked to the urethane unit via a spacer. The enzyme-labile *p*-acetoxybenzyloxycarbonyl (AcOZ) urethane protecting group was successfully applied to the synthesis of doubly lipidated peptides.[115] In **Scheme 9** the general setup and cleavage reaction steps of such blocking groups is illustrated. The enzyme recognizes a specific motif, which leads to cleavage of the enzyme-labile ester bond. This leads to spontaneous fragmentation, in which a carbamic acid derivative is liberated. Subsequent decarboxylation releases the desired peptide.

The AcOZ group can be cleaved enzymatically under very mild conditions in a pH range between 5 and 6. Appropriate enzymes to remove the group and initiate the subsequent fragmentation are the lipase from *Mucor miehei* or an acetyl esterase of the flavedo of oranges.[116] The lipase allows the use of substantial amounts of methanol as a cosolvent. The acetyl esterase attacks acetyl groups but not longer or branched acyl chains, which allows its use in the presence of palmitoyl thioesters; furthermore, it does not show any amidase activity. In **Scheme 10** the successful application of the AcOZ group for the synthesis of a palmitoylated and farnesylated N-Ras C-terminus is shown. The AcOZ group of the tripeptide **13** was removed using 20% MeOH as



Scheme 9 General setup of the enzymatic cleavage and subsequent fragmentation of the enzymatic cleavable protecting group AcOZ (R = Me), respectively PhAcOZ (R = Bn, Benzyl).



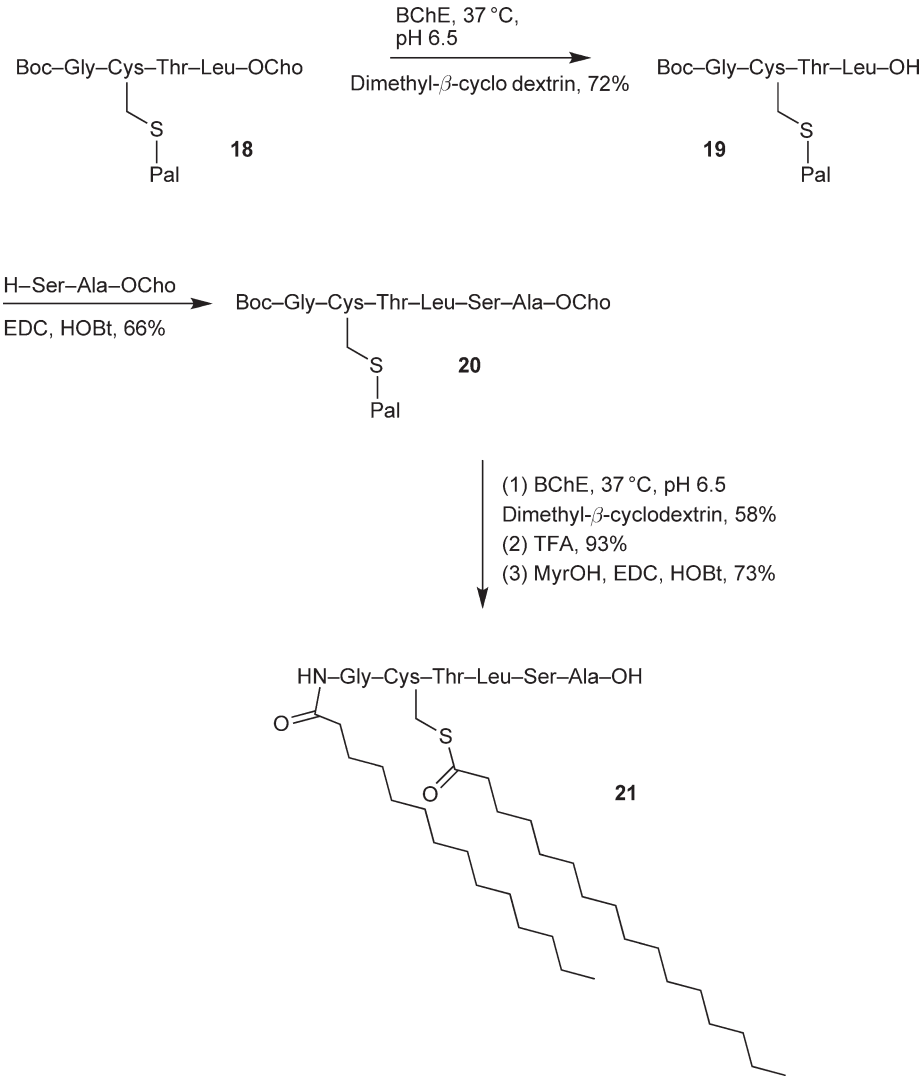
Scheme 10 AcOZ strategy for the synthesis of the palmitoylated and farnesylated C-terminus of N-Ras.

solubilizing cosolvent.[117] After recognition and cleavage of the acetyl bond by the esterase, the resulting *p*-hydroxybenzyl urethane is not stable but undergoes spontaneous fragmentation to give, via decarboxylation, quinone methide and the desired peptide with a free N-terminus. To trap the reactive quinone methide, which is formed during fragmentation, an excess of potassium iodide is added. Using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and *N*-hydroxybenzotriazole (HOBt) the deprotected tripeptide **14** subsequently was coupled to the dipeptide **15** also bearing an N-terminal AcOZ group. After enzymatic deprotection the pentapeptide **16** is elongated by coupling to a palmitoylated dipeptide also N-terminally protected with the AcOZ group. For the removal of the amino protecting group of the palmitoylated and farnesylated heptapeptide, acetyl esterase from oranges is used as this enzyme can distinguish between the palmitoyl thioester and the AcOZ group. To increase the solubility of the peptides cyclodextrins are added, finally giving the farnesylated and palmitoylated peptide **17**. During all these deprotection and elongation steps, the cysteine methyl ester was not harmed.

The scope of this kind of protecting group is quite high as the enzymatic deprotection is general. If another acyl group is chosen, the fragmentation of the *p*-hydroxy-benzyl urethane is initiated upon hydrolysis using a different enzyme. For instance, the *p*-phenylacetoxybenzyloxycarbonyl group (PhAcOZ) can be cleaved by

penicillin G acylase.[118] Using this enzymatic masking for amine groups, a 29-mer peptide corresponding to the N-terminus of endothelial NO synthase, which bears two *S*-palmitoyl thioesters and a myristic acid amide, was successfully built up.[119]

5.17.2.2.2(ii) Enzyme-labile carboxyl protecting groups The choline ester group has been applied as an enzymatically labile blocking group for C-termini of simple peptides as well as for sensitive conjugated peptides like glycosylated, phosphorylated peptides,[120] nucleopeptides,[121] and lipidated peptides.[97,122,123] The removal of the choline ester is carried out under neutral conditions using acetyl choline esterase (AChE) or butyryl choline esterase (BChE). Generally, the deprotection with BChE proceeds faster and in higher yields. Both enzymes are highly selective without harming any other ester bonds in the molecule, and the deprotection is chemoselective. As an example, the synthesis of the myristoylated and palmitoylated hexapeptide **21** corresponding to the N-terminus of the $G_{\alpha O}$ protein was achieved using this protecting group (**Scheme 11**).[123] To this end, the palmitoylated tetrapeptide **18** bearing an N-terminal Boc masking group and a C-terminal choline ester was enzymatically deprotected with BChE. The hydrolyzed peptide fragment was then coupled to a dipeptide also C-terminally protected with the choline ester. To obtain the

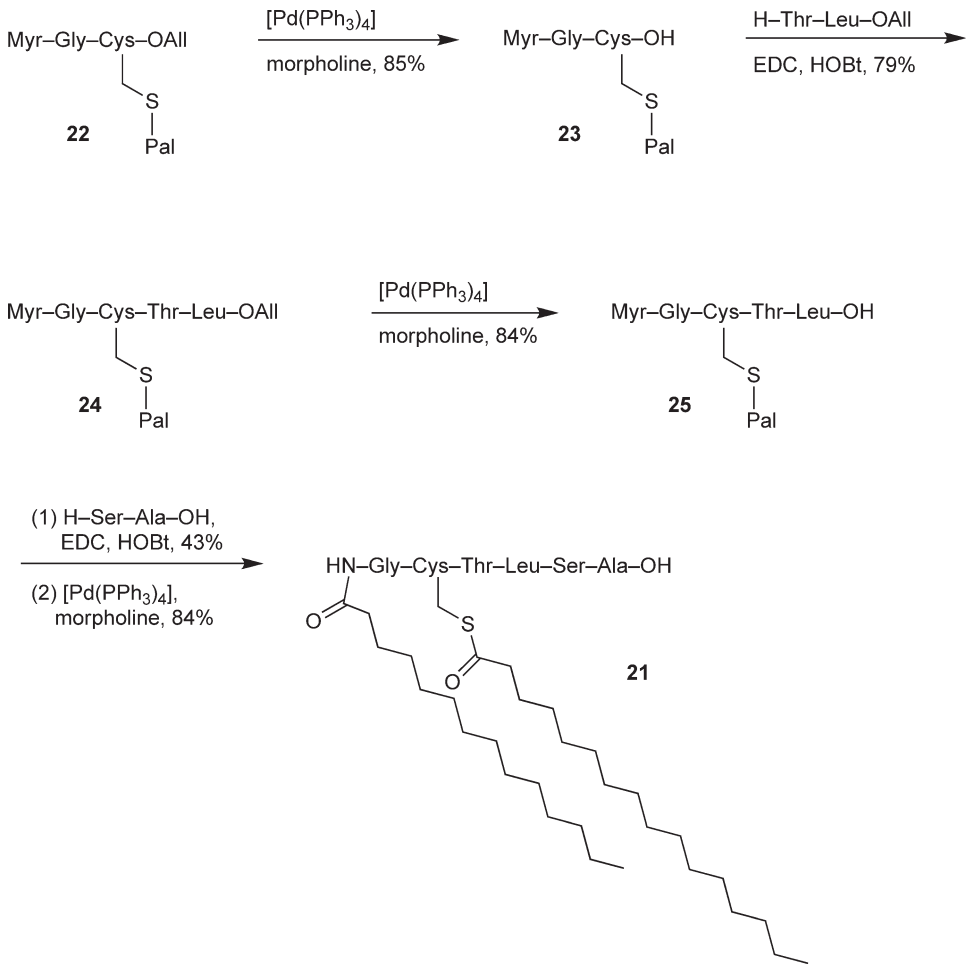


Scheme 11 Choline ester strategy for the synthesis of the myristoylated and palmitoylated hexapeptide (**21**) corresponding to the $G_{\alpha O}$ protein N-terminus.

unprotected C-terminus **21** BChE was used again and cyclodextrin was added to increase the solubility of the peptide. As there was no prenyl group present in this sequence, the use of the Boc strategy for the amine function was possible, allowing deprotection with TFA and subsequent myristoylation of the N-terminus with EDC and HOBt.

5.17.2.2.2(iii) Noble metal-sensitive protecting groups for amines and carboxyl groups An alternative and orthogonal approach to enzymatically removable protecting groups are Pd(0)-sensitive blocking groups like the allyl ester (All) and the allyloxycarbonyl urethane (Aloc). These masking groups have initially found application in glycopeptide chemistry[124,125] and peptide conjugate chemistry,[126,127] where advantage is taken of the high selectivity and the mildness of the deprotection. As such, these protecting groups also allow the application to the synthesis of lipopeptides containing acid-labile prenyl moieties as well as base-labile thioesters. The deprotection reaction is normally performed by adding a catalytic amount of $[Pd(PPh_3)_4]$ to a solution containing the protected peptide and a nucleophile as a scavenger. This nucleophile needs to be chosen carefully, taking into account the stability of the peptide and the intended purification method. Morpholine and dimethylbarbituric acid (DMB) are suitable and can be extracted with aqueous buffer. Furthermore, if a water-soluble ligand, such as triphenylphosphanyltrisulfonate (TPPTS) in combination with $Pd(OAc)_2$, is used, very pure crude products may be obtained.[128–130] If problems of incomplete conversion are encountered the use of phenylsilane can help. Although it cannot be extracted easily, its application led to complete conversion for the synthesis of lipopeptides.

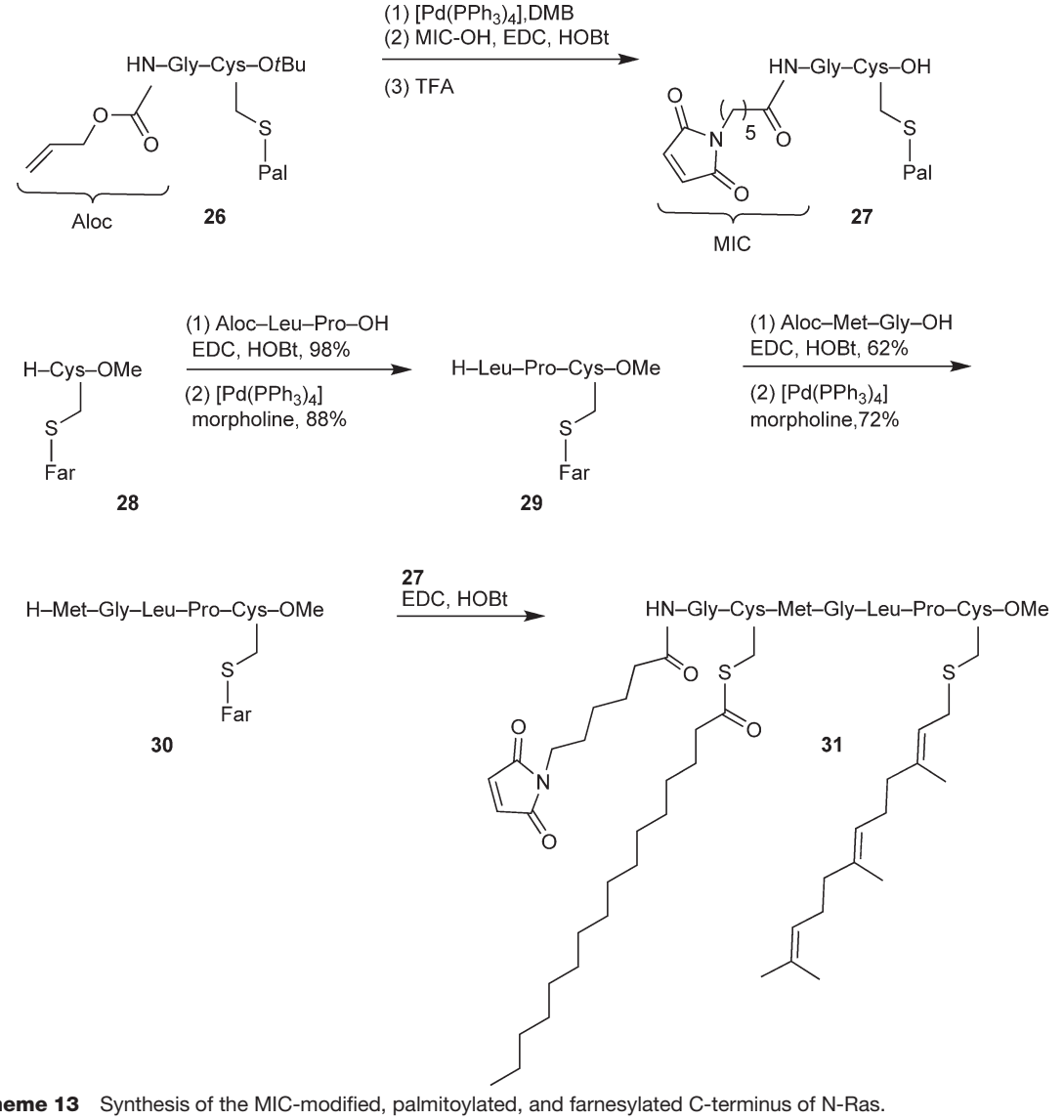
Peptide **21**, synthesized using the C-terminal enzymatic-cleavable choline ester (**Scheme 11**), was also synthesized using the C-terminal allyl ester protecting group (**Scheme 12**). The selective Pd(0)-catalyzed



Scheme 12 Allyl ester strategy for the synthesis of the myristoylated and palmitoylated hexapeptide (**21**) corresponding to the $G_{\alpha O}$ protein N-terminus.

C-terminal deprotection was performed using morpholine as an allyl trap, resulting in high yields. The most significant advantage of the allyl group in this peptide synthesis in comparison with the choline ester is that it overcomes the solubility problem encountered in the enzymatic deprotection approach. The deprotection of the allyl ester is performed in organic solvents, in which these lipidated peptide are readily soluble.[123,131]

The All ester, as well as the analogous Aloc group, have been successfully used for the synthesis of lipidated Ras peptides.[122,132–134] Furthermore, application of the Aloc group allowed the incorporation of additional functional groups.[135] As an example the synthesis of the maleimido (MIC)-modified, S-palmitoylated, and S-farnesylated N-Ras C-terminus **31** is shown in **Scheme 13**. The *N*-Aloc-protected dipeptide **26** bearing a palmitoyl moiety at the cysteine side chain and a *tert*-butyl-protected carboxylic acid was deprotected using Pd(0). After attachment of the MIC group, the ester was hydrolyzed using TFA. This modified dipeptide **27** was then coupled to the pentapeptide **30**, which was built up starting with a farnesylated cysteine methyl ester **28**. The methyl ester was subsequently coupled to a dipeptide bearing an N-terminal Aloc group that was cleaved under neutral and mild conditions using Pd(0). Elongation with a second N-terminally protected Aloc dipeptide and subsequent deprotection resulted in the pentapeptide **30**, which was finally coupled to the palmitoylated and MIC-modified dipeptide **27** to give peptide **31**.



Scheme 13 Synthesis of the MIC-modified, palmitoylated, and farnesylated C-terminus of N-Ras.

5.17.2.2.3 Selected lipidated peptide syntheses in solution

5.17.2.2.3(i) Small lipidated model peptides for biophysical investigations Small cysteine-containing peptides, similar to sequences often found in naturally occurring *S*-acylated proteins, were predominantly synthesized in solution. The tetrapeptide Bimane-SC(S*t*Bu)RC(Far)OMe **32** representative for the C-terminus of H-Ras and featuring the Bimane fluorophore was prepared in solution using Fmoc chemistry, introducing the farnesyl group at the stage of the Fmoc-protected dipeptide.[102]

The acylated peptides (Myr)GCX-Bimane **31 a–e** (X = G, L, R, T, V), which are found in certain nonreceptor tyrosine kinases and α-subunits of several heterotrimeric G-proteins, were synthesized in solution using common solution-phase peptide synthesis with *N*-myristoylglycine as a building block. These model peptides were used for acylation studies with palmitoyl-CoA in phospholipid vesicles at physiological pH. For such uncatalyzed spontaneous reactions only a modest molar excess of acyl donor species (2.5:1) was necessary. Unprotected side chains of threonine or serine are not interfering with this S-acylation (**Scheme 14**).

Double lipidated peptides incorporating a C(GerGer)XC(GerGer)-OMe **33** motif that is found in several Rab and homologous proteins were also synthesized in solution via Fmoc chemistry following cysteine deprotection and geranylgeranylation.[107]

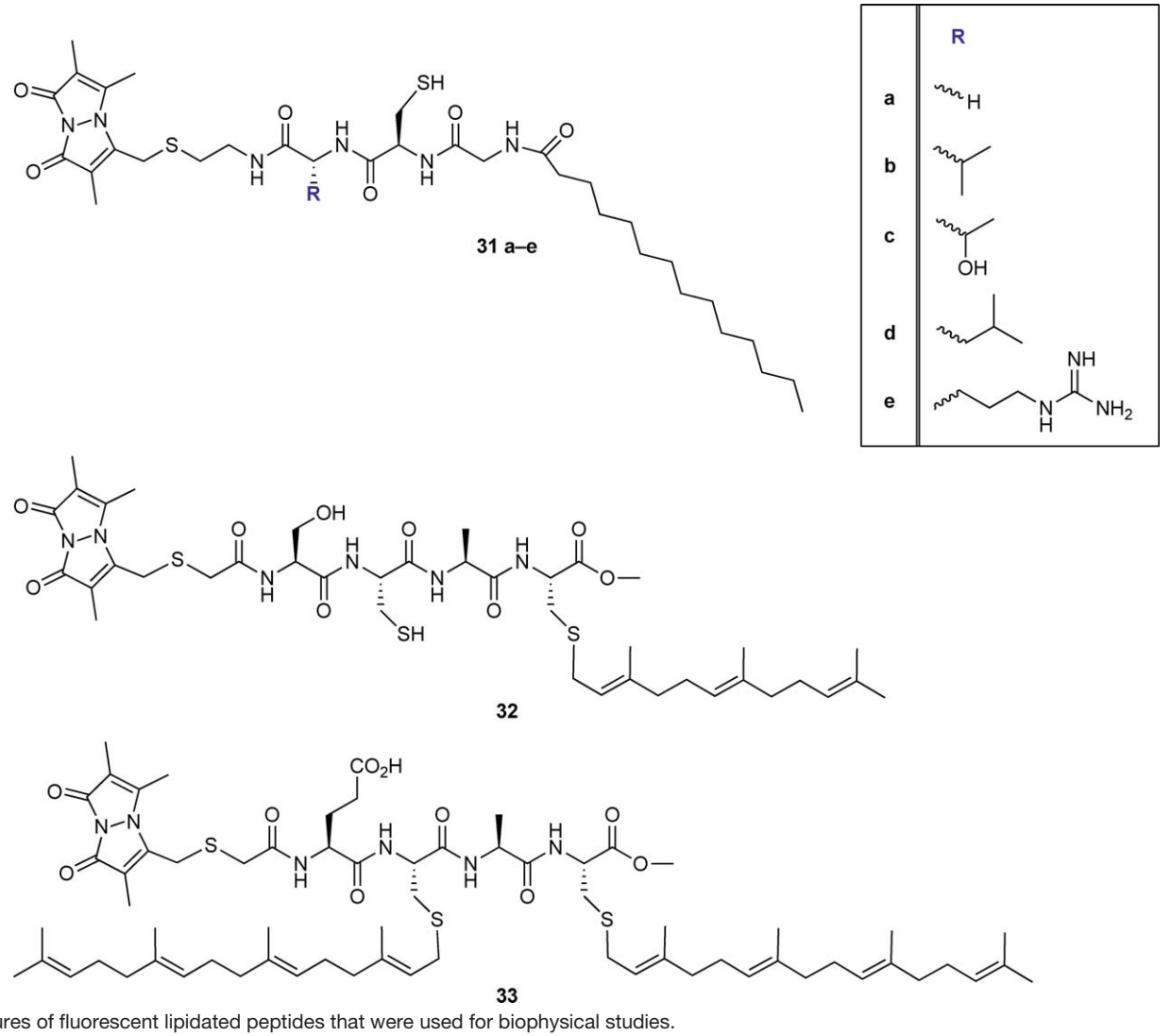
5.17.2.2.3(ii) C-terminally lipidated peptide of the influenza virus hemagglutinin A The influenza virus hemagglutinin A contains a lipidated peptide fragment at the C-terminus featuring two palmitoylated cysteines and two amino acids with a polar side chain. This peptide **37** was synthesized in solution and the synthesis strategy was based on the fragment condensation of the lipidated tetrapeptide TIC(Pal)I **34**, which was coupled to the palmitoylated protected dipeptide RC(Pal) **35** and after deprotection the resulting peptide was N-terminally elongated with NBD-aca-labeled methionine (**Scheme 15**).[136] For this block coupling strategy, a set of three orthogonal protecting groups was required in which the use of base/nucleophile-labile and hydrogenolytically removable protecting groups was not possible. The use of the acid-labile Boc group for the N-terminus, the Pd(0)-sensitive allyl ester for the C-terminus, and the Pd(0)-cleavable Aloc group for the arginine side chain function turned out to be a successful combination.

5.17.2.2.3(iii) C-terminally lipidated H- and N-Ras peptides A large body of work has been devoted to the synthesis of the C-terminally lipidated peptides of the small GTPases H- and N-Ras. These peptides have been synthesized in solution, via combined solution- and solid-phase approaches as well as completely on solid phase.

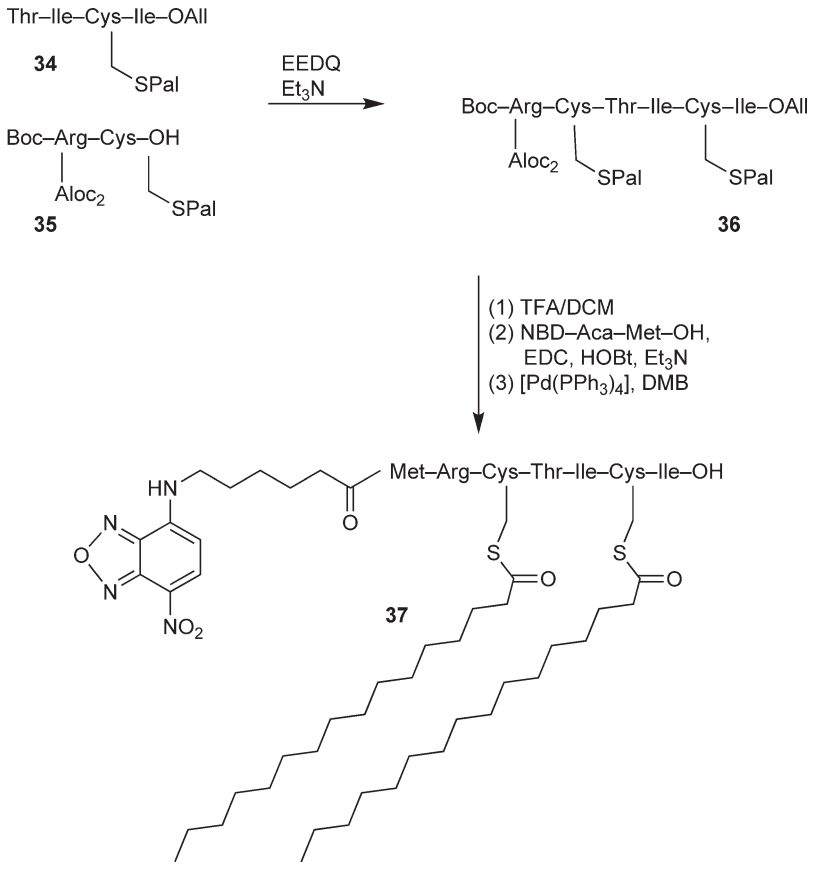
For the synthesis of a small library of palmitoylated and isoprenylated N-Ras peptides in solution, a modular strategy was adopted, with the tetrapeptide MGLP **38** as a key intermediate. This tetrapeptide allowed further elongation at its C-terminus with lipidated or nonlipidated cysteine methyl esters, as well as the addition of various N-terminally MIC-labeled dipeptides, consisting of different GC lipidated units **39–41** (**Scheme 16**).[135,137] The synthesis was performed under common conditions using the Fmoc-, Boc-, and Aloc-protecting group strategy. Utilizing this pathway a number of N-Ras derivatives containing natural and non-natural lipid residues were produced and the technique was extended also allowing the synthesis of fluorescent derivatives.

In 2005 Schmidt and coworkers published a reversed approach for synthesizing palmitoylated and farnesylated peptides in solution via an S_N2 displacement of a bromide, which was embedded in a bromoalanine-containing sequence. Thiol-bearing lipids can act as nucleophiles, for example, thiopalmitic acid or farnesylmercaptane.[138] The method gives access to farnesylated, palmitoylated, and doubly lipidated peptides (**Scheme 17**).

5.17.2.2.3(iv) Synthesis of polybasic lipidated peptides in solution The synthesis of polybasic isoprenylated lipidated peptides such as the C-termini of K-Ras 4B, D-Ral, and Rho A was most successful on solid support. Solution strategies to the peptides, generally through fragment condensation, were difficult due to different polarities and solubilities of the fragments. The solution synthesis of the polybasic C-terminus of Rho A required coupling of the protected KKK-containing peptide **47**, which previously had been connected to the fluorescent-labeled glycine **45**, with the SGC(GerGer) tripeptide containing the geranylgeranylated cysteine



Scheme 14 Structures of fluorescent lipidated peptides that were used for biophysical studies.



Scheme 15 Synthesis of the doubly palmitoylated heptapeptide corresponding to the C-terminus of hemagglutinin A bearing an N-terminal NBD function using Boc, Aloc, and the allyl ester as protecting groups.

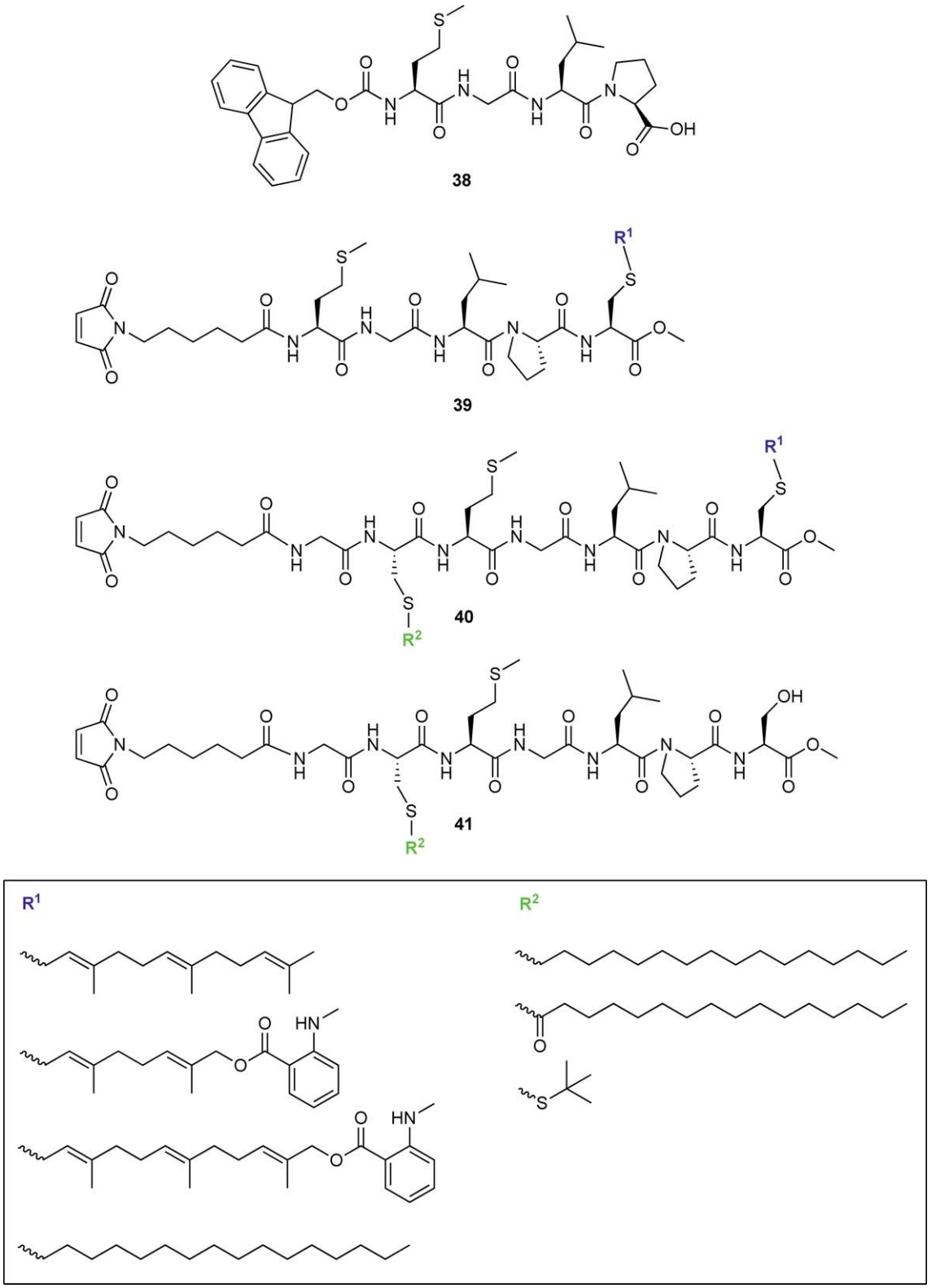
methylester (**Scheme 18**).[139,140] For this approach the orthogonality of the Fmoc-, Aloc-, and O*t*Bu-protecting groups was explored. However, it was not possible to apply the same methodology to the synthesis of the polybasic C-terminus of K-Ras 4B due to low coupling yields resulting from the different solubility properties of the fragments.

5.17.2.3 Synthesis of Lipopeptides on Solid Support

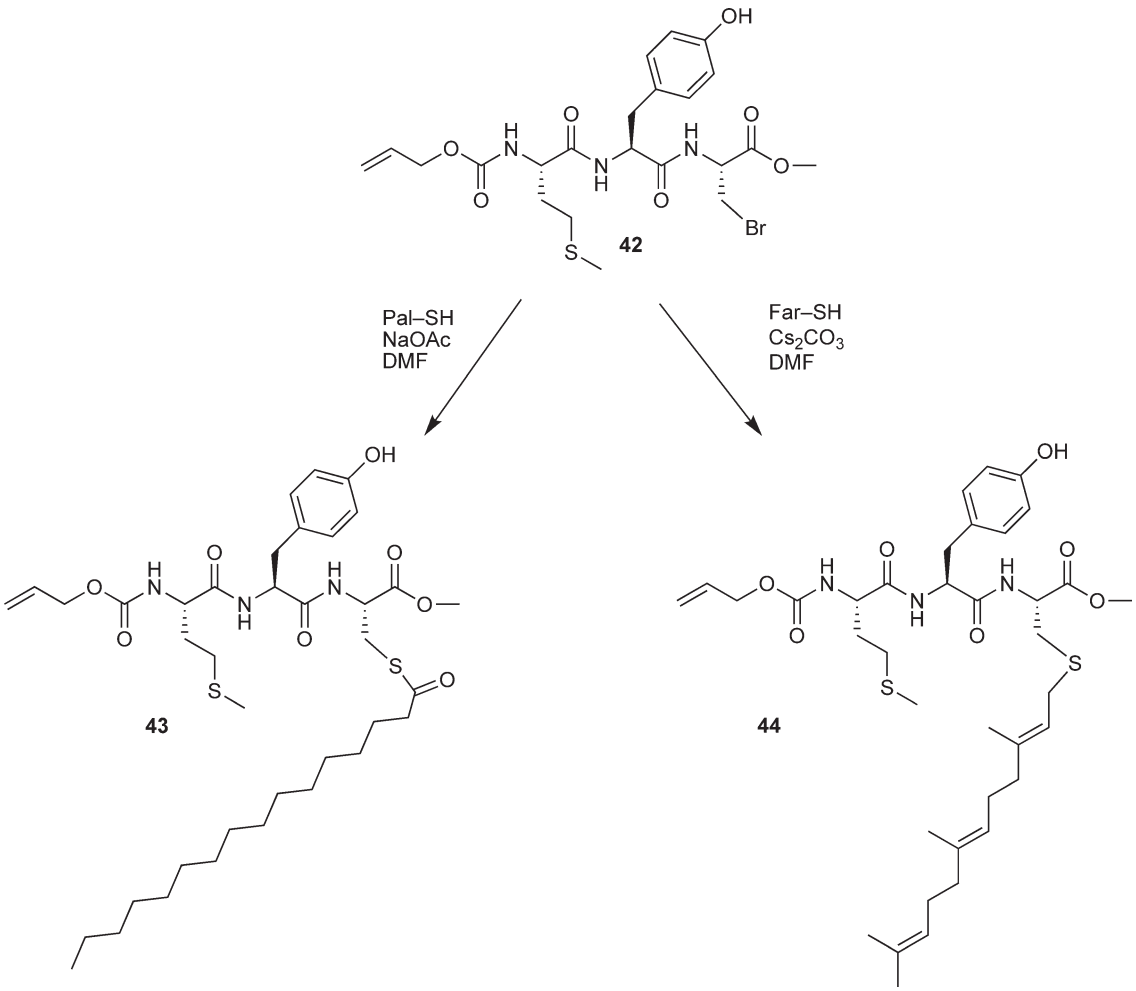
By analogy to the solution-phase approaches, the introduction of lipid functionalities on peptides on the solid support can also follow two general approaches, either using prelipidated building blocks in the standard solid-phase peptide synthesis or via selective lipidation on resin.[139,141–148]

5.17.2.3.1 Lipidation on resin

If lipidation is carried out on solid support, an orthogonal protecting group strategy is required that allows the introduction of different types of lipids, such as farnesyl and palmitoyl groups, on the same peptide in a stepwise manner but also avoiding undesired lipidation on side chains. To obtain this goal, generally a large excess of the lipidating species is used. This might, for instance, cause problems if isotopically labeled or modified lipids for biological studies need to be introduced. To overcome these problems, dehydroalanine can be placed in the sequence which subsequently can be attacked by thiolate nucleophiles leading to the desired lipidated peptide.[149] For stereoselective control of the conjugate addition, aziridine-2-carboxylic acids can be used in the solid-phase synthesis (**Scheme 19**).[150]



Scheme 16 Library of N-Ras sequences modified with palmitoyl, farnesyl, hexadecyl, and prenylated fluorophores whose syntheses were based on the key intermediate (**38**).



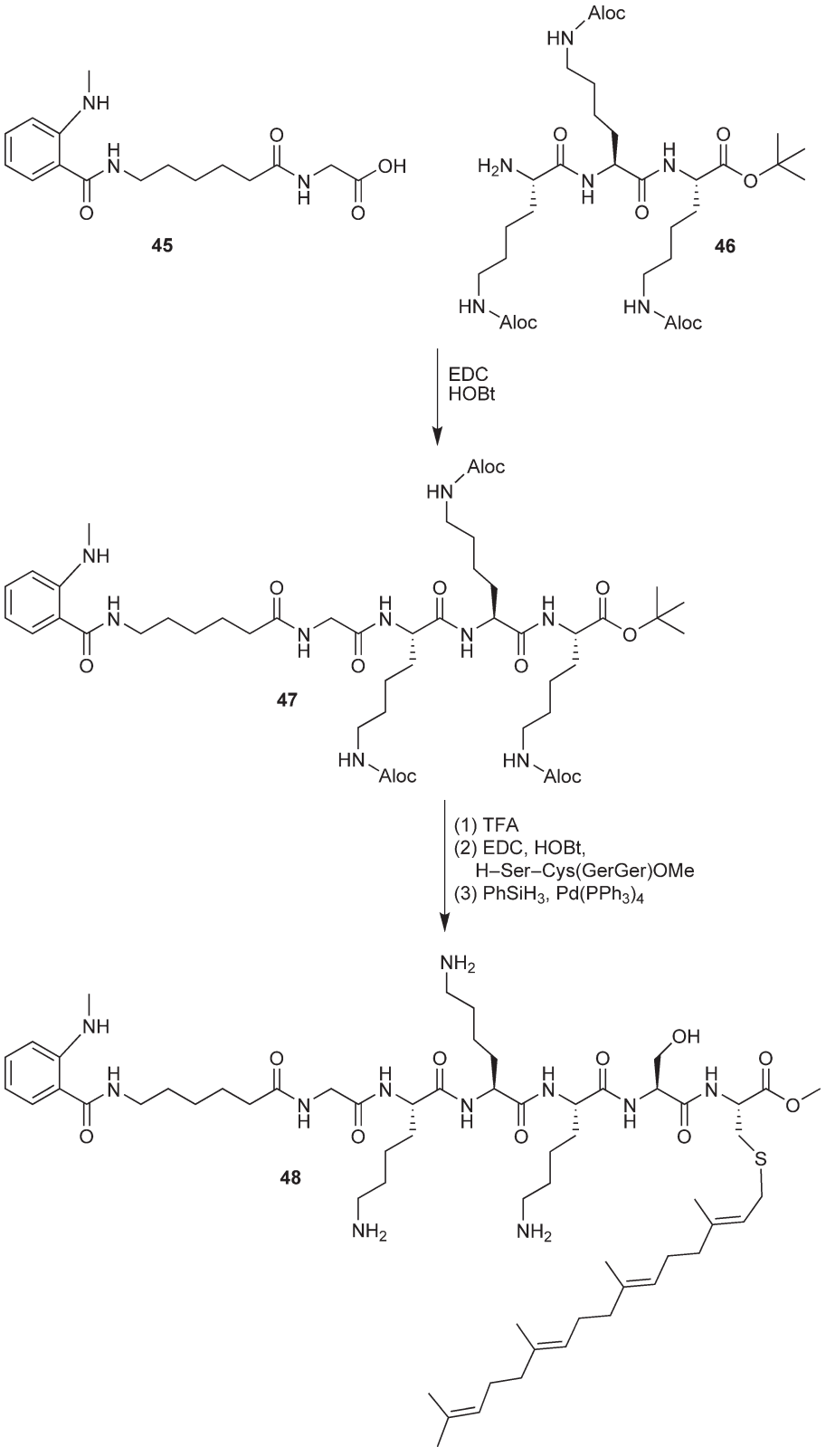
Scheme 17 Reverse approach for the synthesis of lipidated peptides by nucleophilic attack of thiol nucleophiles to substitute a bromide incorporated in the sequence.

5.17.2.3.2 Use of prelipidated building blocks

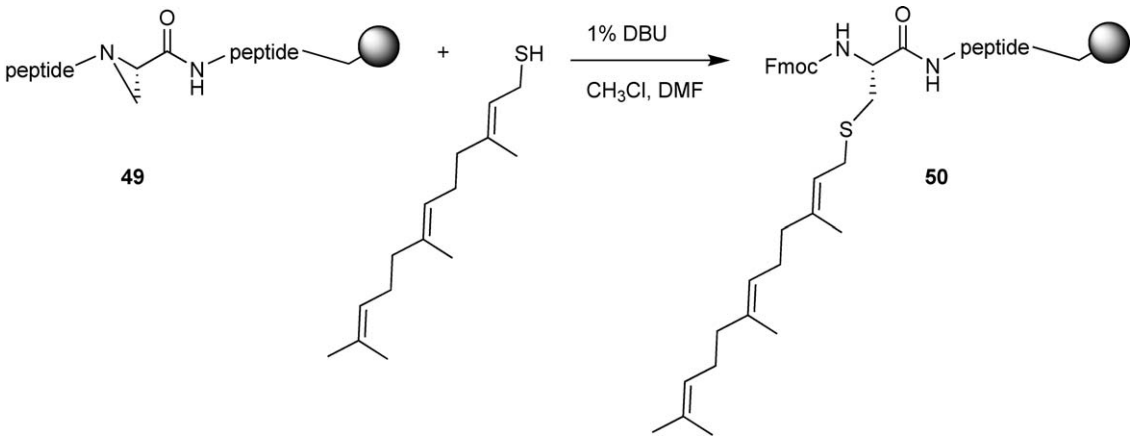
The alternative to resin lipidation is the use of prelipidated building blocks. **Scheme 20** shows a variety of lipidated cysteines, which were synthesized and incorporated into peptide sequences also bearing reporter groups for biological studies.[151–153] For the preparation of alkylated cysteines, a new method was recently published.[154] These Fmoc-protected building blocks, bearing the desired lipid moiety, are used in solid-phase peptide synthesis like normal amino acids. Since the protecting group strategy does not need to be modified, this method gives access to a flexible and generic synthesis of lipidated peptides.

5.17.2.3.3 Linker systems for the synthesis of lipidated peptides on solid support

Recently, the main focus for the preparation of lipopeptides has shifted to the synthesis on solid support. Various linker systems and protecting group strategies are under investigation to find the ideal lipopeptide assembly system. This implies a protecting group strategy, which is orthogonal to the different lipid moieties and other functional groups in the peptide; the cleavage of the lipidated product from the resin should additionally in itself be orthogonal again to all the functionalities in the peptide, lipid repertoire, and protecting groups. The cleavage from the resin should furthermore allow the introduction of functional groups at the C-terminus of the peptide, since many lipidated proteins are post-translationally processed in this position;



Scheme 18 Successful synthesis of the polybasic C-terminus of Rho A via solution-phase approach.



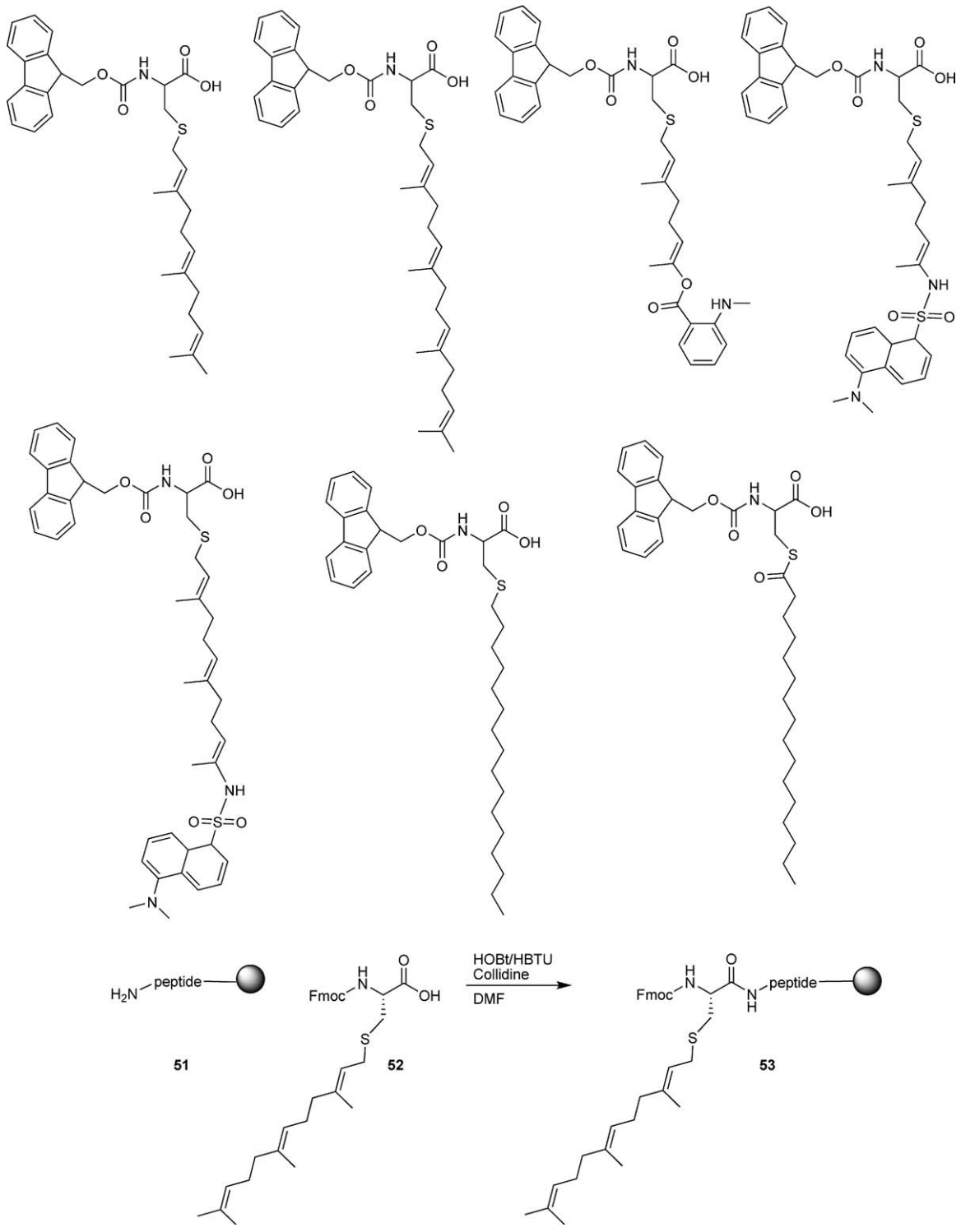
Scheme 19 Use of aziridine-2-carboylic acids in the peptide backbone to synthesize farnesylated peptides in a stereoselective way.

typically, a methyl ester is required. In addition, the C-terminus provides an ideal handle to insert molecular tags, such as fluorescent groups, for biophysical studies (**Scheme 21**).

In the following sections different solid-phase approaches will be reviewed according to their linker system. As previously mentioned, the major challenge for the synthesis of lipopeptides is the orthogonality between the protecting groups, the lipid residues, and the linker. The commonly used solid-phase linkers for unmodified peptides, like the Wang- or Rink-amide linker are not suitable for lipopeptide synthesis because they require, for example, strong acidic cleavage, not compatible with prenyl moieties.

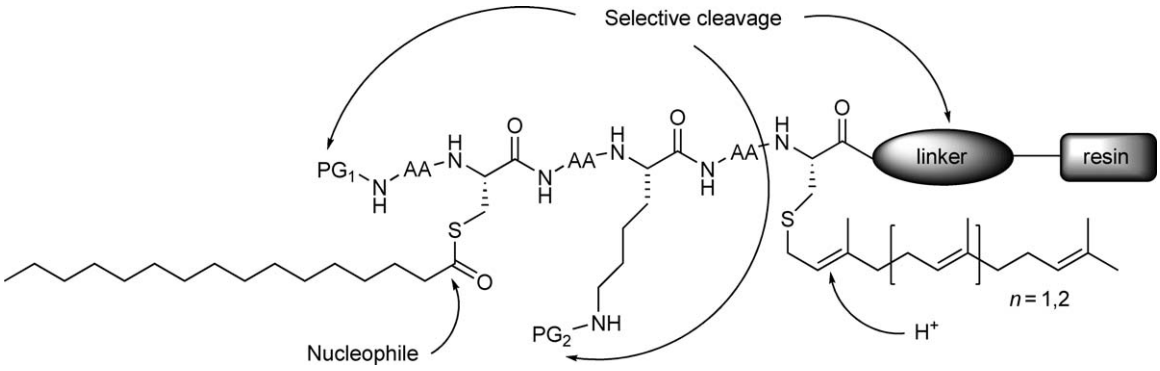
5.17.2.3.3(i) Kaiser benzophenone oxime resin This resin was first published by Degrado and Kaiser in 1980[155,156] and has been used for the synthesis of tetrapeptides bearing farnesylated cysteines at the N-terminus[156] as well as for the synthesis of biotinylated and radiolabeled peptides.[157] The *N*-Boc-protected amino acids are converted into active esters via reaction with *N*,*N*-dicyclohexylcarbodiimide (DCC) and HOBt in tetrahydrofuran (THF), and by treatment of the resin with TFA in CH_2Cl_2 the Boc-protecting group is removed. The desired product is released from the resin with L-amino acid benzyl ester tosylates under mildly acidic conditions. Since the resin is used with the Boc strategy, acid-labile functionalities can only be incorporated at the N-terminus, which limits the application of this resin (**Scheme 22**).

5.17.2.3.3(ii) Trityl linker The trityl linker system allows cleavage under very mildly acidic conditions, making it compatible with prenyl groups. Although the linker can be attached to the C-terminus of the peptide, cleavage typically results in the generation of a free carboxylic acid only and does not allow the introduction of variety at this position. Alternatively, the peptide can be anchored to the resin via a side chain, for example, the ε-amino group of a lysine, which offers possibilities for the synthesis of lipopeptides with functionalized C-termini. This strategy has been successfully applied to polybasic sequences, such as the C-termini of Rho A, K-Ras 4B, and D-Ral.[140,158] In all of these syntheses the ε-amino group of a lysine close to the C-terminal end was anchored to the resin **61**. The lysine's carboxylic function was protected with an allyl ester, which could be selectively removed with Pd(0) and subsequently coupled to a prelipidated cysteine methyl ester (**Scheme 23**). The resulting dipeptide **62** attached to the solid support was then elongated using standard Fmoc-SPPS protocol. For the cleavage of the peptide sequence, 1% TFA was used in combination with a scavenger. Due to the high dilution double-bond isomerization of the prenyl functionalities could be avoided.[159] The use of the Aloc-protecting group for the amino-groups of the side chain limits the compatibility of this approach with nucleophile sensitive palmitoyl thioesters, since allyl trapping scavengers are required for successful cleavage of multiple amino groups. Similarly, the maleimido functionality,[137] incorporated into



Scheme 20 Selection of differently modified lipidated Fmoc-cysteines and their coupling as normal building blocks on solid support.

lipopeptides for ligation purposes, is not compatible with the phosphine ligands of the Pd(0) catalyst used in the Aloc deprotection. Nevertheless, the trityl linker is one of the most suitable linker systems to generate, especially, polybasic lipopeptides in a very efficient way.



Scheme 21 Demands of an ideal strategy for the synthesis of lipidated peptides on solid support.

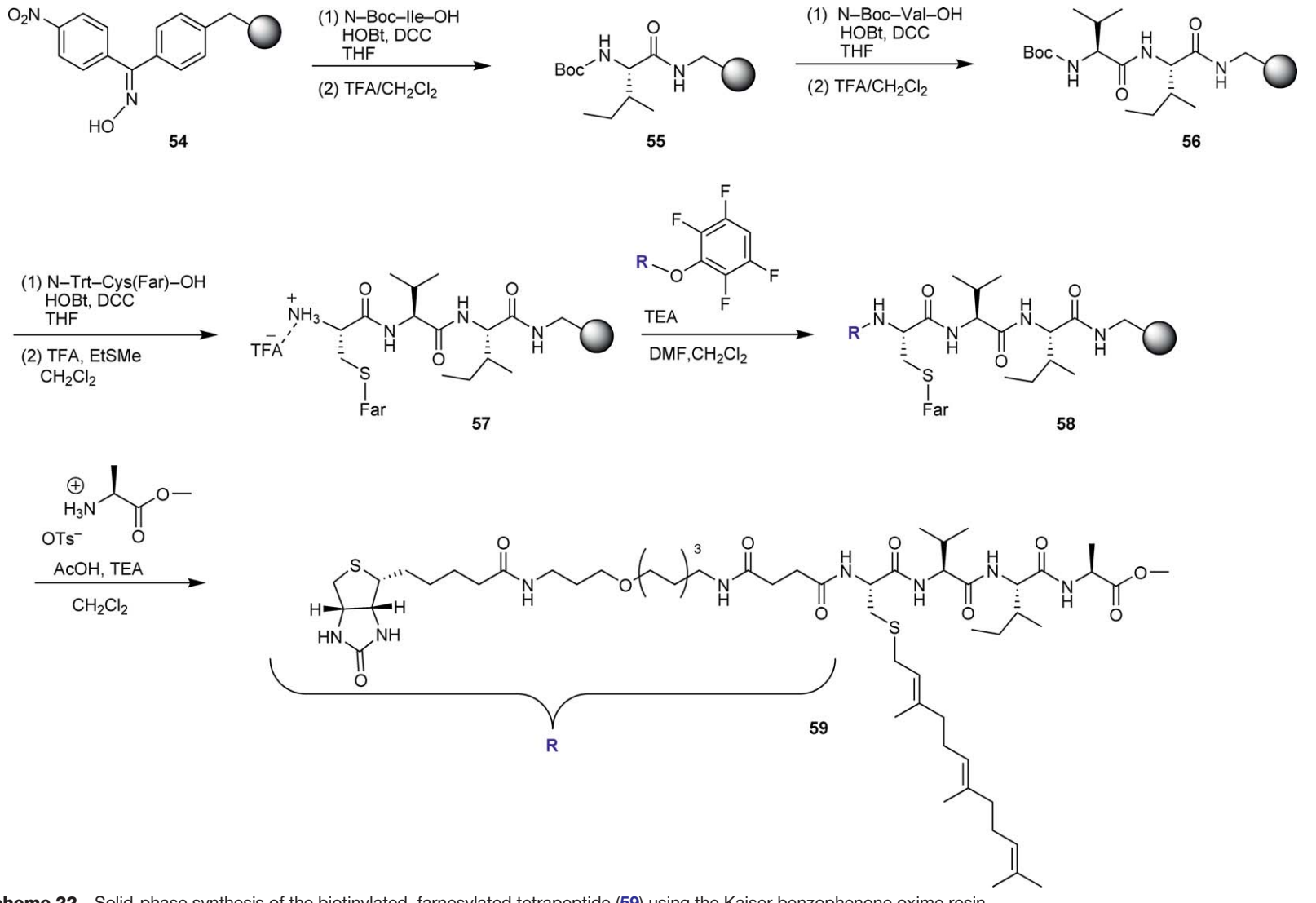
5.17.2.3.3(iii) Hydrazine linker The aryl hydrazine linker has been subject to intense investigation for the application of the synthesis of lipopeptides. The oxidation-labile group was described in 1970[160] and since 1995 it has been applied in the solid-phase synthesis for small peptide esters and amines[161,162] as well as for the synthesis of cyclic peptides.[163] The peptide cleavage reaction occurs via oxidation of the linker to an acyldiazene. This is subsequently attacked by a suitable nucleophile, leading to the release of the peptide with the respective nucleophile attached to the C-terminus (**Scheme 24**).

Two methods are frequently applied for the oxidation of the linker. One method uses $Cu(OAc)_2$, O_2, and a nucleophile for *in situ* oxidation and cleavage. The other method is a two-step procedure involving the oxidation via *N*-bromosuccinimide (NBS) followed by a nucleophilic cleavage with the desired nucleophile. Typically, nucleophiles like water, amines, or alcohols are used.[164] The linker is orthogonal to classical urethane-protecting groups such as Boc, Fmoc, and Aloc, and racemization does not occur upon cleavage. The oxidation sensitivity of the linker does require the coupling reactions and especially Fmoc deprotection reactions to be performed under exclusion of oxygen, because of possible oxidation of the linker and cleavage with nucleophilic reagents present.

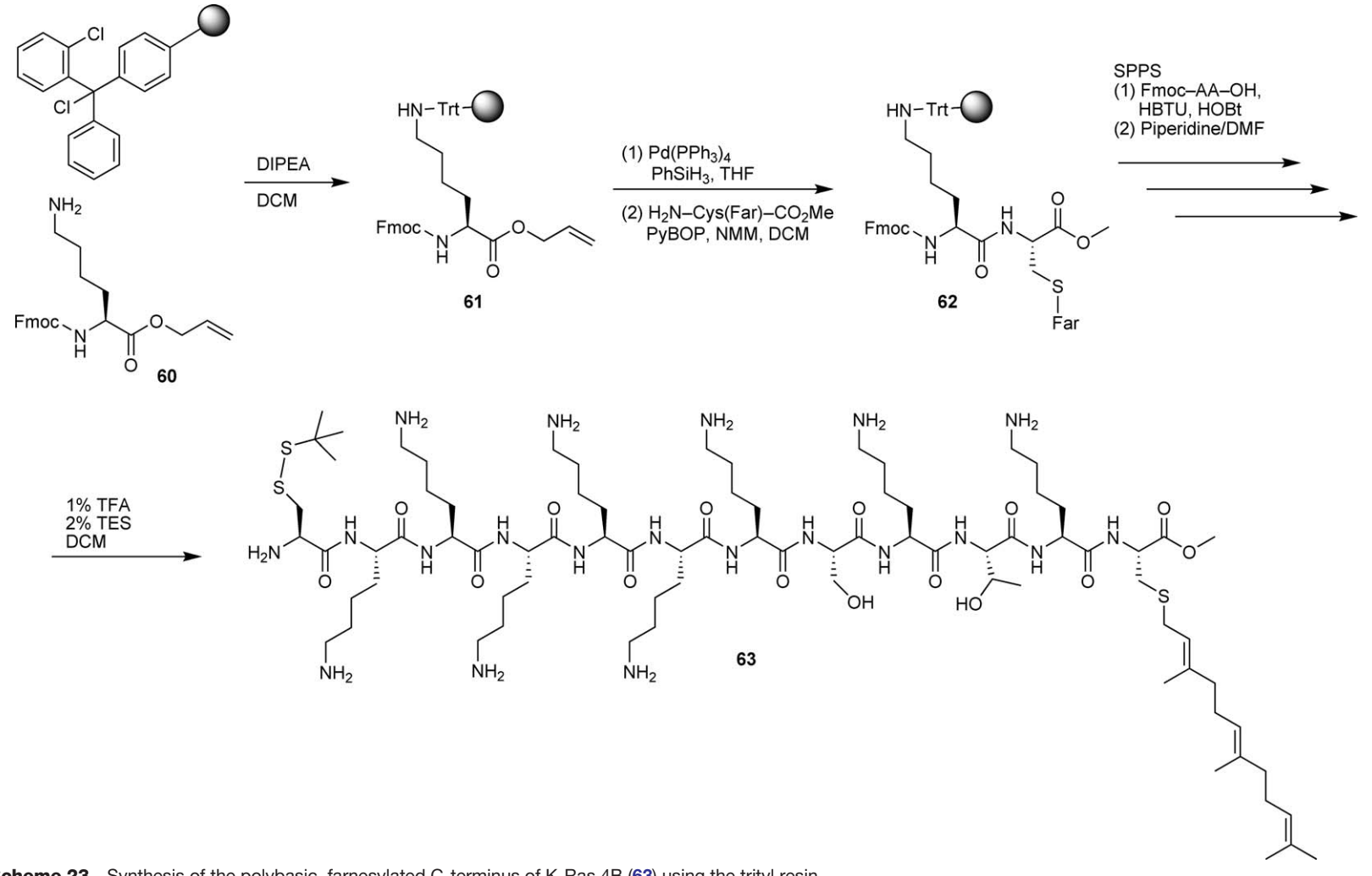
Although used previously for peptide synthesis, the special utility of the hydrazide linker really came to light when applied to the synthesis of lipidated peptides.[148,152,153,165] A typical sequence synthesized on the hydrazide linker is shown in **Scheme 25**. The first amino acid attached is a farnesylated cysteine **65**. To avoid racemization equimolar amounts of collidine are used as base instead of *N,N*-diisopropylethylamine (DIPEA) and the coupling reaction to the resin is extended to 12 h. Afterward, normal coupling conditions using HBTU, HOBt, and DIPEA are used. A remarkable feature in this sequence is the palmitoyl moiety that was introduced as a prepalmitoylated building block using collidine as a base. To avoid the rapid S, N-shift of the palmitoyl group of **67** during the Fmoc-deprotection step, a solution of 1% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in DMF was used for 2×30 s. The next coupling was then immediately performed with the preactivated amino acid and HATU as coupling reagent.[153] After assembly of the palmitoylated and farnesylated peptide **68** on solid support, the cleavage occurs, as mentioned above, upon oxidation with $Cu(OAc)_2$. Subsequent nucleophilic attack of methanol finally gives the desired double lipidated peptide with the C-terminal methyl ester **69**.

In **Scheme 26** a selection of lipidated peptide sequences synthesized via this approach is shown. A variety of N- and H-Ras-derived peptide sequences was obtained bearing different types of lipids like farnesyl, palmitoyl, and geranylgeranyl and a fluorescent-labeled geranyl group. Furthermore, fluorescent markers like nitrobenzoxadiazole and photoactivatable groups were introduced.

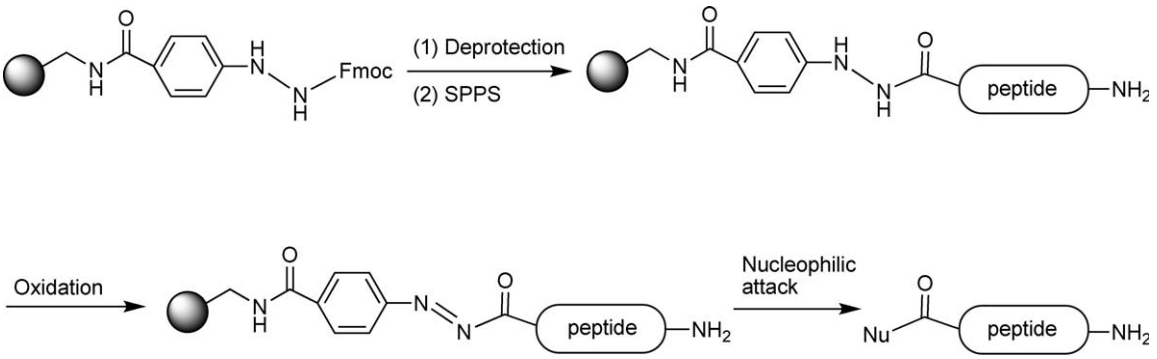
Alternatively to using prelipidated building blocks palmitoylation on resin is possible with the hydrazine linker. In **Scheme 27** the synthesis route for the palmitoylated and farnesylated N-Ras peptide **78** is shown. Here the initial loading of trityl-protected cysteine to the hydrazine linker was mediated by *N,N*-diisopropylcarbodiimide (DIC) and HOBt. After Fmoc removal the proline was coupled using HBTU and HOBt. The trityl-protected dipeptide **75** was subsequently S-deprotected using TFA with triethylsilane (TES) as a scavenger. Farnesylation of the free thiol was achieved with an excess of farnesyl bromide.



Scheme 22 Solid-phase synthesis of the biotinylated, farnesylated tetrapeptide (**59**) using the Kaiser benzophenone oxime resin.



Scheme 23 Synthesis of the polybasic, farnesylated C-terminus of K-Ras 4B (**63**) using the trityl resin.



Scheme 24 Cleavage of the hydrazine linker by oxidation and nucleophilic attack of a nucleophile.

Afterward, the peptide chain was elongated following the standard Fmoc-based protocol. Before cleavage of the peptide the incorporated Mmt-protected cysteine was deprotected using 1% TFA. Under these very mild conditions the farnesyl moiety was not harmed. Palmitoylation could be achieved using an excess of palmitoyl chloride. Cleavage with copper acetate and methanol as a nucleophile gave the farnesylated and palmitoylated N-Ras sequence with the C-terminal methyl ester **78**.[152]

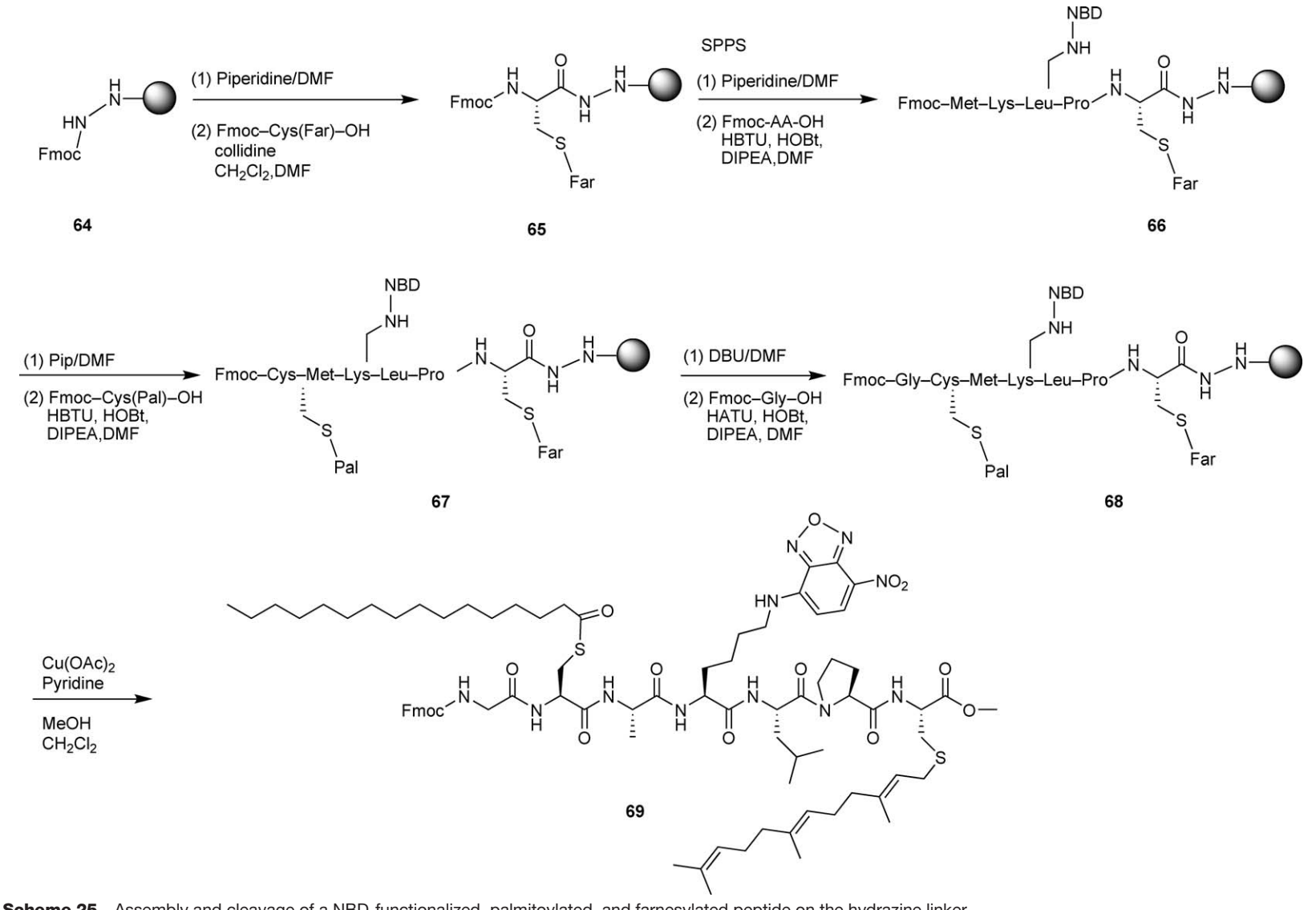
In summary, the hydrazine linker is currently the most suitable linker system for the synthesis of C-terminally methylated lipidated peptides. Allowing the incorporation of different types of lipid chain in combination with the MIC group at the N-terminus the hydrazine linker gives access to a variety of lipidated peptides of the Ras superfamily.

5.17.2.3.3(iv) Ellman sulfonamide linker The Ellman sulfonamide linker offers another option to synthesize lipidated peptides on solid support. This linker system is stable toward acids and bases. It is activated upon N-alkylation of the *N*-acyl sulfonamide **80** with selectivity over the methionine thioether. The target sequence is released from the resin via nucleophilic attack (**Scheme 28**).[166]

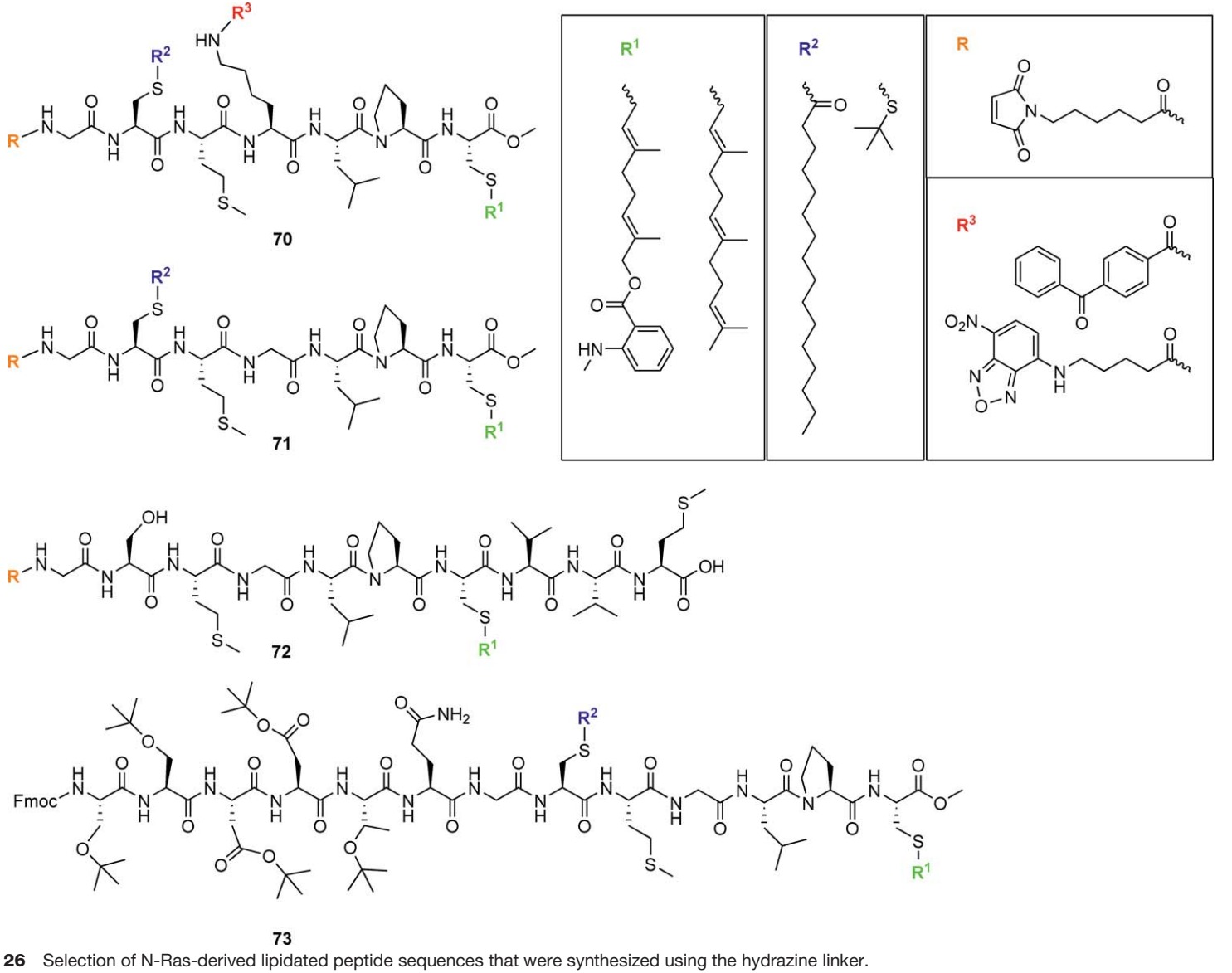
Having been used already for the synthesis of different peptide derivatives and other compound classes,[167] this linker was successfully applied to the synthesis of lipidated peptides of the Ras superfamily, since it meets all requirements of these sequences. The Ellman linker turned out to be also suitable for the synthesis of larger peptides containing more than 20 amino acids with multiple lipidation sites. As an example, the synthesis of the NO synthase-derived hexacosapeptide **85** that is triply lipidated is shown in **Scheme 29**. The 26-mer peptide was built up using the Ellman linker. Normal solid-phase peptide synthesis was used except for the couplings after the palmitoylated cysteines. Here, the protocol was changed to the rapid deprotection with 1% DBU in DMF and fast coupling using HATU. After assembly of the doubly palmitoylated 26-mer sequence **84** on solid support, the N-terminal Fmoc group was removed using 1% DBU in DMF and the resulting free amine was myristoylated. After alkylation with iodoacetonitrile and nucleophilic attack of water, the peptide was released from the resin. In a final deprotection step the acid-labile permanent protecting groups were cleaved off with TFA and triisopropyl silane (TIS) as a scavenger, giving the desired doubly palmitoylated and myristoylated 26-mer **85** in a yield of 24%.[168]

5.17.2.4 Approaches Combining Solution- and Solid-Phase Peptide Chemistry

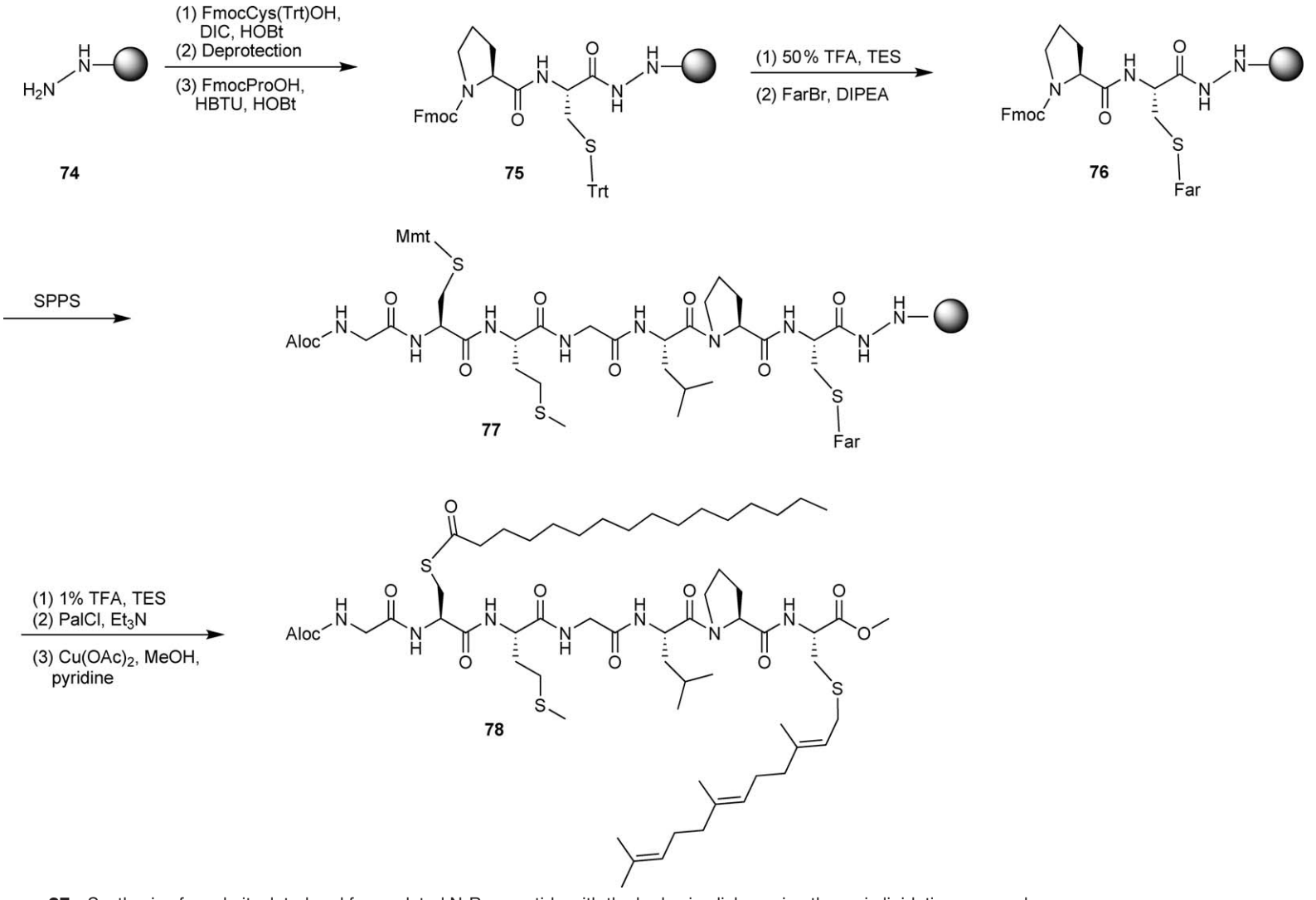
A possibility to overcome the time-consuming solution-phase peptide synthesis, but avoiding specific synthetic problems that cannot be addressed on the solid support, is to follow a combined solution-/solid-phase approach. In such an approach the majority of the peptide sequence is typically assembled on solid support, but critical steps are performed in solution. Below, some exemplary peptides are discussed.



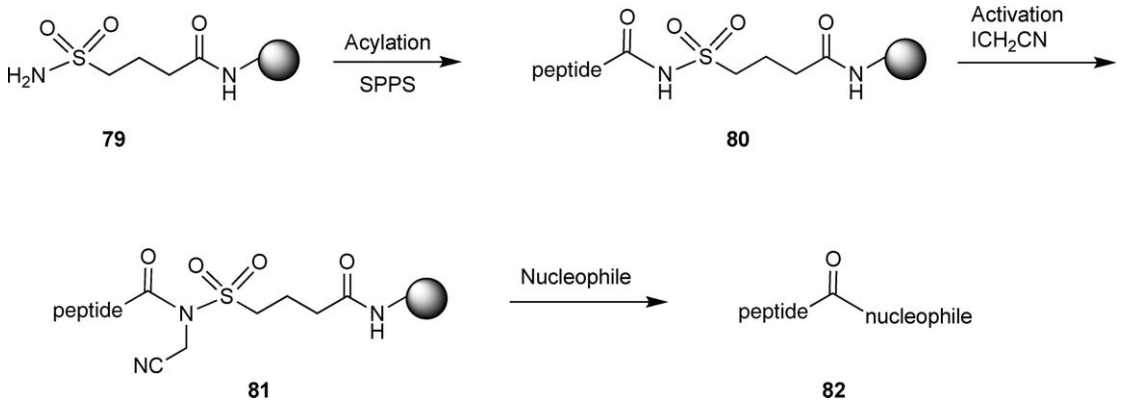
Scheme 25 Assembly and cleavage of a NBD-functionalized, palmitoylated, and farnesylated peptide on the hydrazine linker.



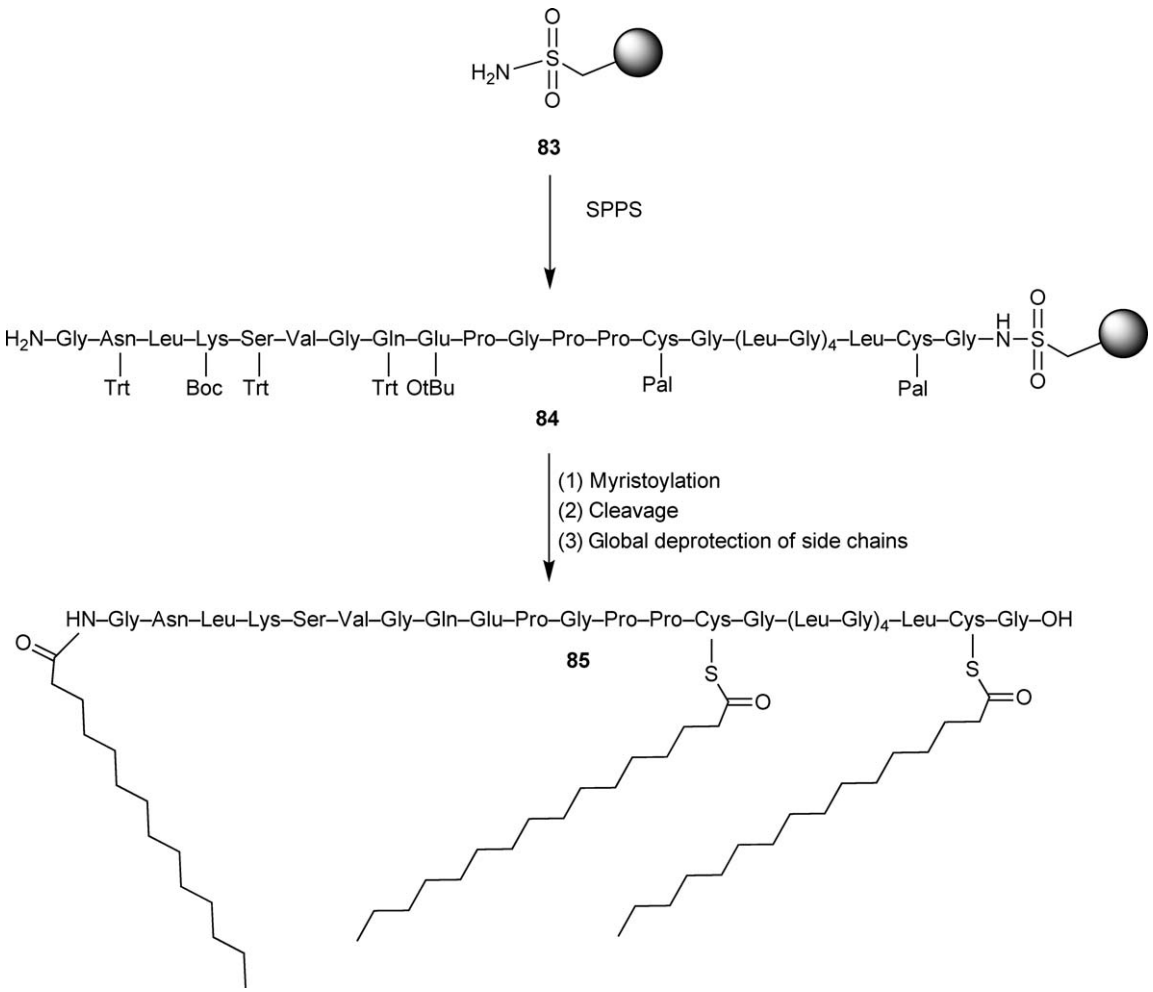
Scheme 26 Selection of N-Ras-derived lipidated peptide sequences that were synthesized using the hydrazine linker.



Scheme 27 Synthesis of a palmitoylated and farnesylated N-Ras peptide with the hydrazine linker using the resin lipidation approach.



Scheme 28 Activation and cleavage of the Ellman sulfonamide linker.



Scheme 29 Synthesis of the NO synthase-derived 26-mer bearing two palmitoyl and one myristoyl moiety.

5.17.2.4.1 Synthesis of O-palmitoylated peptides on solid support

The 44-mer peptide PLTX II from the venom of the spider *Plectreurys* is O-palmitoylated at a C-terminal threonine residue. Synthesis of this peptide using the Fmoc protocol gave unsatisfying results as an O, N-shift was

observed in the presence of base (piperidine) during the Fmoc deprotection.[169] By switching to the Boc strategy the shift could be avoided. The 44-mer was divided into six segments that were each synthesized using the Boc strategy. After coupling of these segments in solution the desired palmitoylated sequence could be obtained.[170]

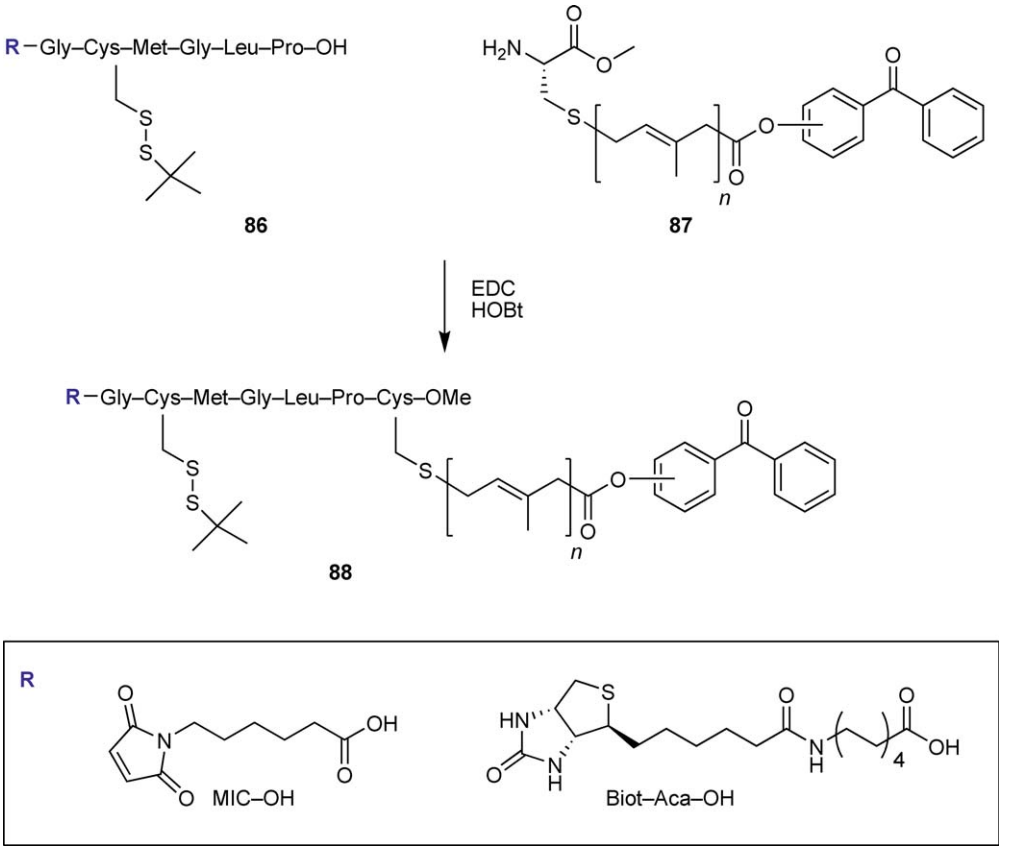
5.17.2.4.2 Synthesis of differently modified N-Ras peptides

In a combined solution- and solid-phase approach lipidated peptides, such as **88**, corresponding to the C-terminus of N-Ras were synthesized by fragment condensation of N-terminal hexapeptides **86** including N-terminal modifications like maleimidocaproic acid or biotin to C-terminally lipidated cysteine methyl esters **87**, which feature, for example, a farnesyl functionality incorporating benzophenone or further reporter groups.[171] This allowed the hexapeptide to be assembled on solid support using the trityl linker in a very fast and efficient way. The second cysteine was protected as a *tert*-butyl disulfide, thus allowing reductive cleavage under physiological conditions. The N-terminal amine of the peptide was connected to an MIC group for protein ligation purposes.

Furthermore, these two fragments are of similar complexity, which is desirable for a high-yielding convergent synthesis. The use of the protecting and functional groups made the final segment condensation a relatively straightforward task. An important issue that has to be kept under scrutiny when performing such segment condensations is the possibility of racemization. Therefore the selection of validated coupling techniques, which avoid the use of basic conditions and polar solvents, is critical. However, this has usually proved sufficient to avoid or minimize racemization (**Scheme 30**).[171]

5.17.2.4.3 Lipidated peptides of Hedgehog proteins

The C-terminal steroid-modified heptapeptides, corresponding to the C-terminus of Hedgehog proteins, were synthesized in a combined solid-/solution-phase approach.[172] The strategy was based on a dipeptide,



Scheme 30 Combination of solution- and solid-phase approach to synthesize modified N-Ras sequences.

Fmoc–Ser–Gly–OAll, which was synthesized in solution. Subsequently, it was connected to the trityl resin via the free hydroxyl functionality of the serine. The C-terminal allyl ester was cleaved by noble metal-mediated allyl transfer to phenylsilane. Using PyBOP as a coupling reagent, the carboxylic acid was connected to glycine-steroid esters. Afterward, the N-terminal peptide chain elongation was achieved by means of standard Fmoc solid-phase peptide chemistry to yield differently fluorescent-labeled peptides carrying an NBD group at a lysine side chain or at the N-terminus or MIC-modified peptides. The peptides were cleaved from the resin under mild conditions with 5% TFA, furnishing the desired products in high yields without any side reactions.

5.17.3 Synthesis of Lipidated Proteins

5.17.3.1 Synthesis of Lipidated Proteins via Biological Methods and Their Limitation

Increasingly it has become clear that biochemical studies need to be performed on the protein featuring its post-translational modifications. This not only allows the study of several processes that cannot be investigated without these post-translational modifications, such as protein–protein and protein–membrane interactions, but it also allows a more valid comparison between cellular and biochemical data. Although there has been substantial progress in generating authentic and engineered proteins, post-translationally modified proteins are not generally accessible by genetic methods. This imbalance is primarily due to the intricacy of protein modification pathways as well as their ways of manipulation.[173]

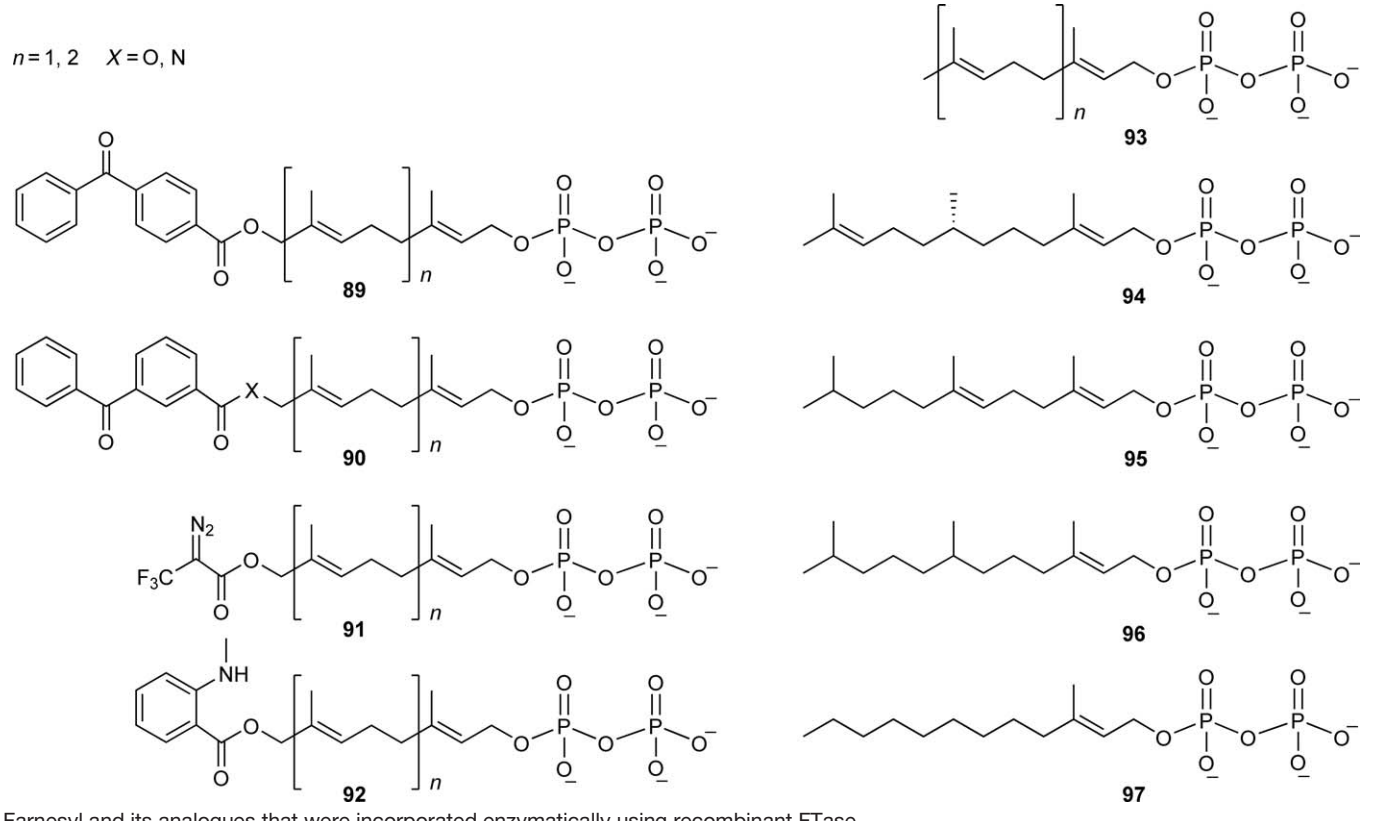
Post-translational modifications, such as phosphorylation, complex glycosylation, and lipidation, typically occur in eukaryotic organisms. Therefore, their expression in prokaryotic systems like *Escherichia coli* is difficult. However, it should be noted that via clever engineering and coexpression of specific enzymes, access can be granted to specific lipidated proteins via expression in bacteria, for example, via the expression of *N*-myristoyltransferase in *E. coli*.[174] Eukaryotic systems that can be used for the expression of post-translationally modified proteins are yeast[175] and *Dictyostelium discoidum*.[176,177] Furthermore, lipidated proteins, such as the Rab proteins, can be obtained via purification from tissue sources or from membrane fractions of insect cells that had been infected with baculovirus bearing a Rab gene.[178,179]

However, these methods are in most cases time consuming with low yields and high costs. For expression of proteins with lipid residues in multiple sites or the incorporation of various lipid moieties and non-natural groups for biological studies, the situation gets even more complicated.[61] Currently, there are two approaches that are used to access lipidated proteins: (1) via the incorporation of the lipid with the aid of lipidating enzymes and (2) via the synthesis of lipidated and modified peptides featuring post-translational modification and their ligation to protein cores accessible via molecular biology techniques.

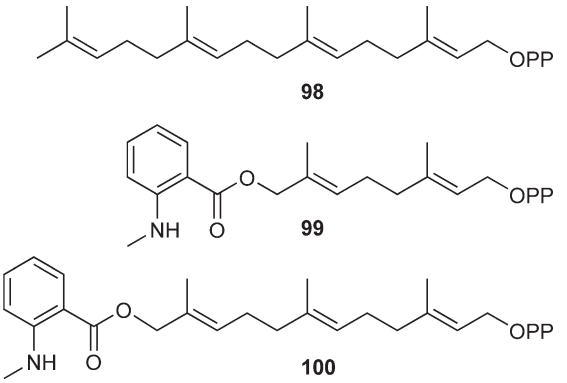
5.17.3.2 Synthesis of Lipidated Proteins Using Lipidating Enzymes

Alternatively to expression of lipidated proteins, there is the option of expressing the biocatalyst and using it for the lipidation of peptides and even entire proteins.[19,180] The natural substrates of FTase and GGTase-I and GGTase-II are prenylpyrophosphates, such as farnesyl or geranylgeranyl pyrophosphate. Since these transferases show broad tolerance regarding the protein substrate as well as the pyrophosphate, they can be used for prenylation of a broad range of entire proteins. Distefano and coworkers have explored the scope of the substrate tolerance of the enzymes with the aid of a large collection of non-natural pyrophosphates.[113,181–184] Following this approach, Ras proteins could be successfully prenylated with different types of prenylpyrophosphates. In **Scheme 31** a selection of pyrophosphates **89–92** is shown bearing photoaffinity labels in the prenyl part that served as tools for elucidating the mechanism of farnesylation by FTase.[113,181,184] **Scheme 31** also provides an overview of further non-natural farnesyl analogues **93–97**, which could be incorporated into inactive H-Ras derived from bacteria to study their activity toward the activation of MAP kinase in a cell-free system. These findings are in line with results obtained by Gelb and coworkers who had shown that inactive H-Ras can be converted into active, farnesylated H-Ras via *in vitro* incubation with FPP and recombinant FTase.[185,186]

In 2001 Alexandrov and coworkers developed a new two-plasmid expression system that allowed protein expression of GGTase-II at high levels and easy purification.[187] Through this method the proteins were expressed as heterodimers with α-subunits bearing a cleavable 6His-glutathione *S*-transferase (GST) tag,



Scheme 31 Farnesyl and its analogues that were incorporated enzymatically using recombinant FTase.



Scheme 32 Structure of geranylgeranyldiphosphate and two fluorescent derivatives.

which facilitated the purification in a two-step procedure. The GST-tagged GGTase-II can be used for preparative *in vitro* prenylation. Owing to this easy and efficient purification method, prenylation of Rab proteins could be achieved. Despite the different mode of action, GGTase-II also shows a broad substrate tolerance, thereby allowing incorporation of differently modified prenyl moieties bearing reporter groups.

Two fluorescent derivatives **99** and **100** of geranylgeranyldiphosphate **98** could be efficiently attached to the Rab7 protein via an enzymatic transfer reaction that required Rab GGTase as well as its accessory protein REP (**Scheme 32**).[188] These non-natural substrates are efficient tools to elucidate the mechanism of prenylation. Furthermore, the establishment of a highly sensitive fluorescence assay was possible, allowing investigations on the discrimination of GGTase-II between its natural substrate and other phosphoisoprenoids via monitoring the interaction between the enzyme and the lipid moieties.[189]

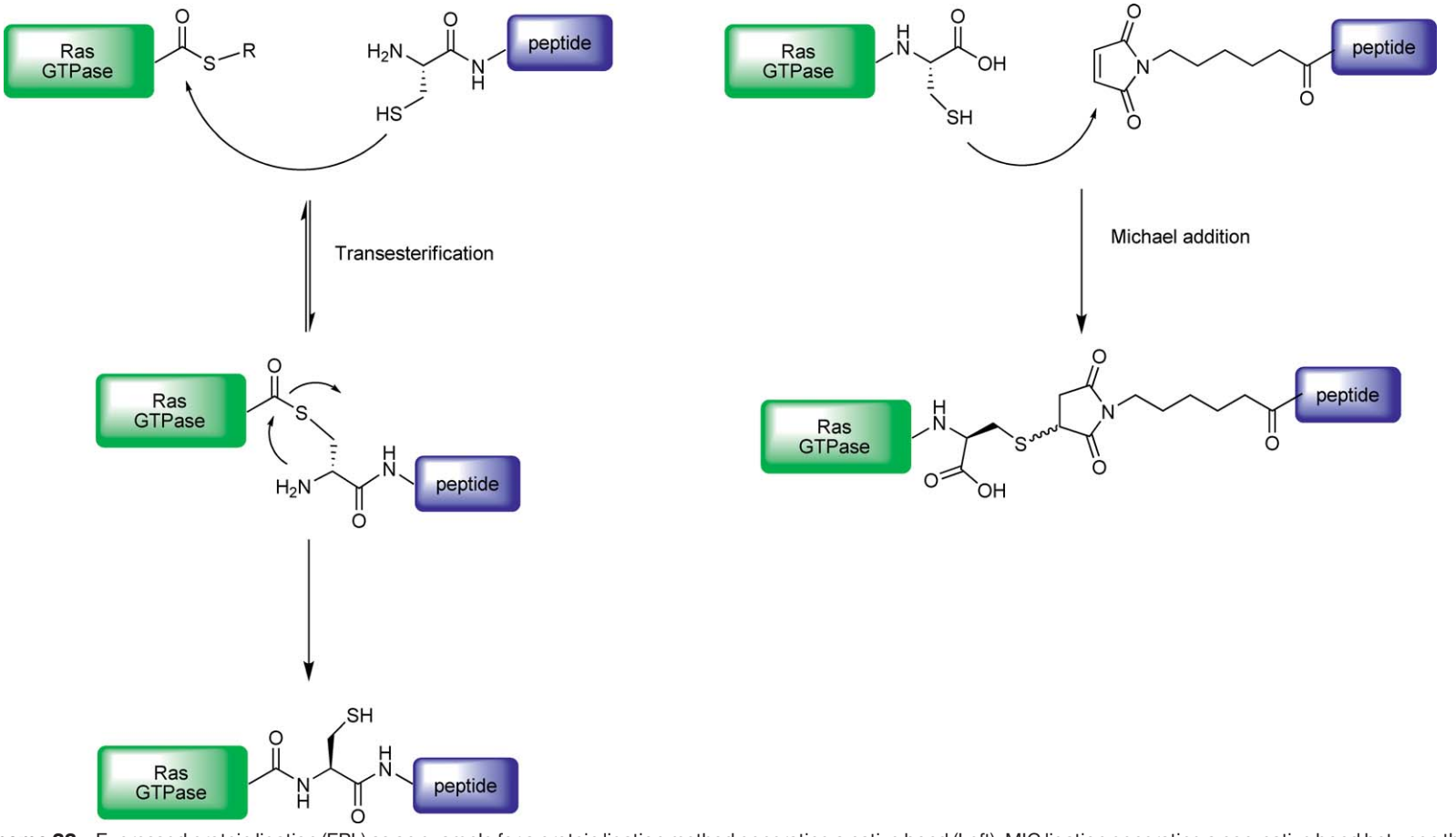
Recently, the purification of recombinant protein acyl transferases was published.[190] Future investigations will show whether these biocatalysts may also serve as tools for acylating proteins. Depending on their substrate tolerance the incorporation of non-natural acyl analogues could also be possible.

5.17.3.3 Synthesis of Lipidated Proteins Using Ligation Methods

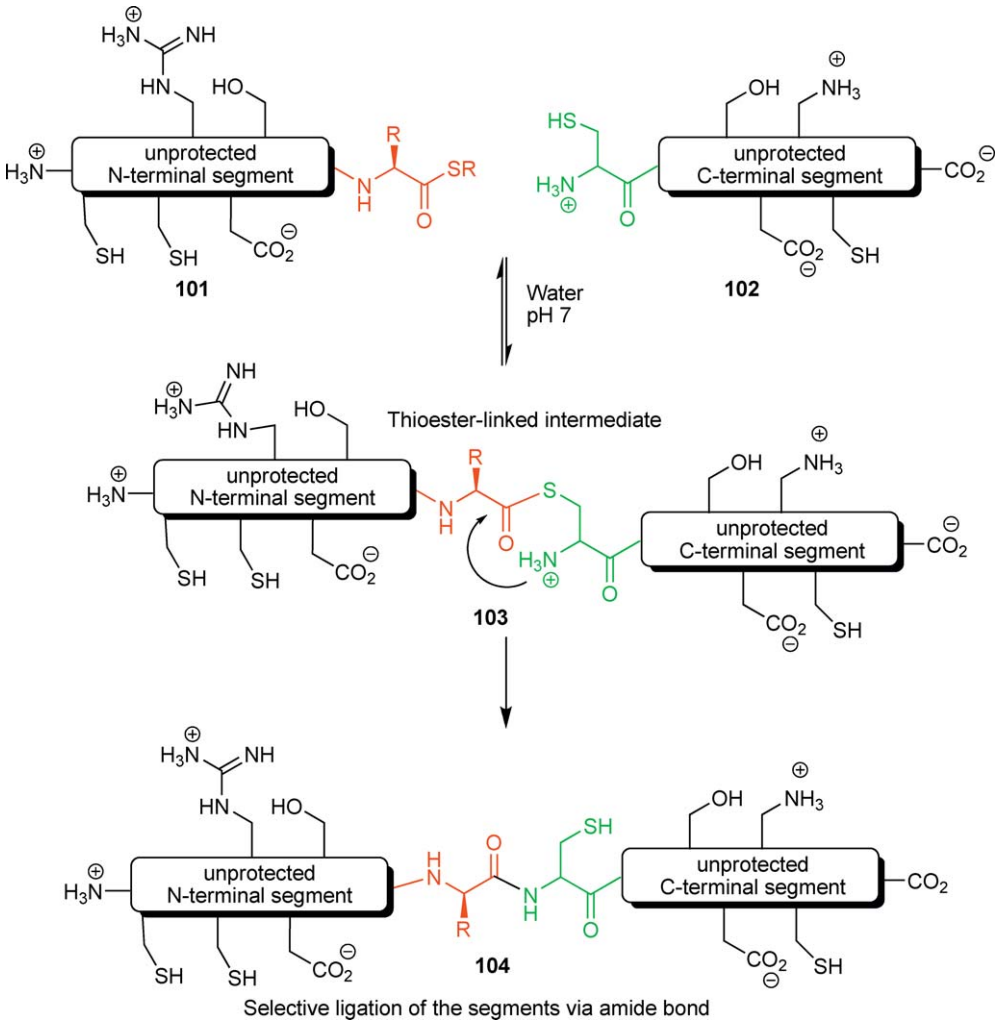
Enzymatic prenylation of protein, which is in specific cases highly successful, is limited to the confined changes in the prenyl moiety. For example, the introduction of modifications in the protein at sites other than the predetermined site of attachment for the prenyl functionality is not feasible. For studying the biological roles of post-translational modifications and the role of the modified proteins in signal transduction, preparative amounts of lipidated proteins are required that also bear non-natural groups like fluorophores, photoreactive and spin-labeled groups, and differently modified lipids. The recent progress in protein ligation methods offers a platform for combining large recombinant protein scaffolds with peptides derived from organic synthesis.[191–195] For the generation of lipidated proteins via ligation methods two different classes of ligation methods can be used. The first type of method generates a native bond as connection site between the two individual parts (e.g., prior thiol capture, native chemical ligation, expressed protein ligation). The second type of ligation method generates a non-native bond between the expressed protein and the synthesized peptide (e.g., imine capture ligation, oxime ligation, maleimidocaproic acid ligation, Diels–Alder ligation) (**Scheme 33**).

5.17.3.3.1 Native chemical ligation

Published in 1994 by Kent and coworkers, the native chemical ligation (NCL) has become an often used and powerful method for connecting peptide fragments via a native amide bond.[196] Two unprotected peptide segments, one with a C-terminal thioester **101** and the other bearing an N-terminal cysteine **102**, react to generate a transient thioester **103** which is not isolated but is destined to undergo a rearrangement via intramolecular nucleophilic attack of the αN-amine (see **Scheme 34**). Owing to the two-step reaction, in which the first is reversible and the second step is irreversible, only one product **104** is formed.[197] For this ligation method the fragments do not need to be protected and no racemization occurs during the coupling. Thus, NCL allows the chemical synthesis of large proteins.[198]



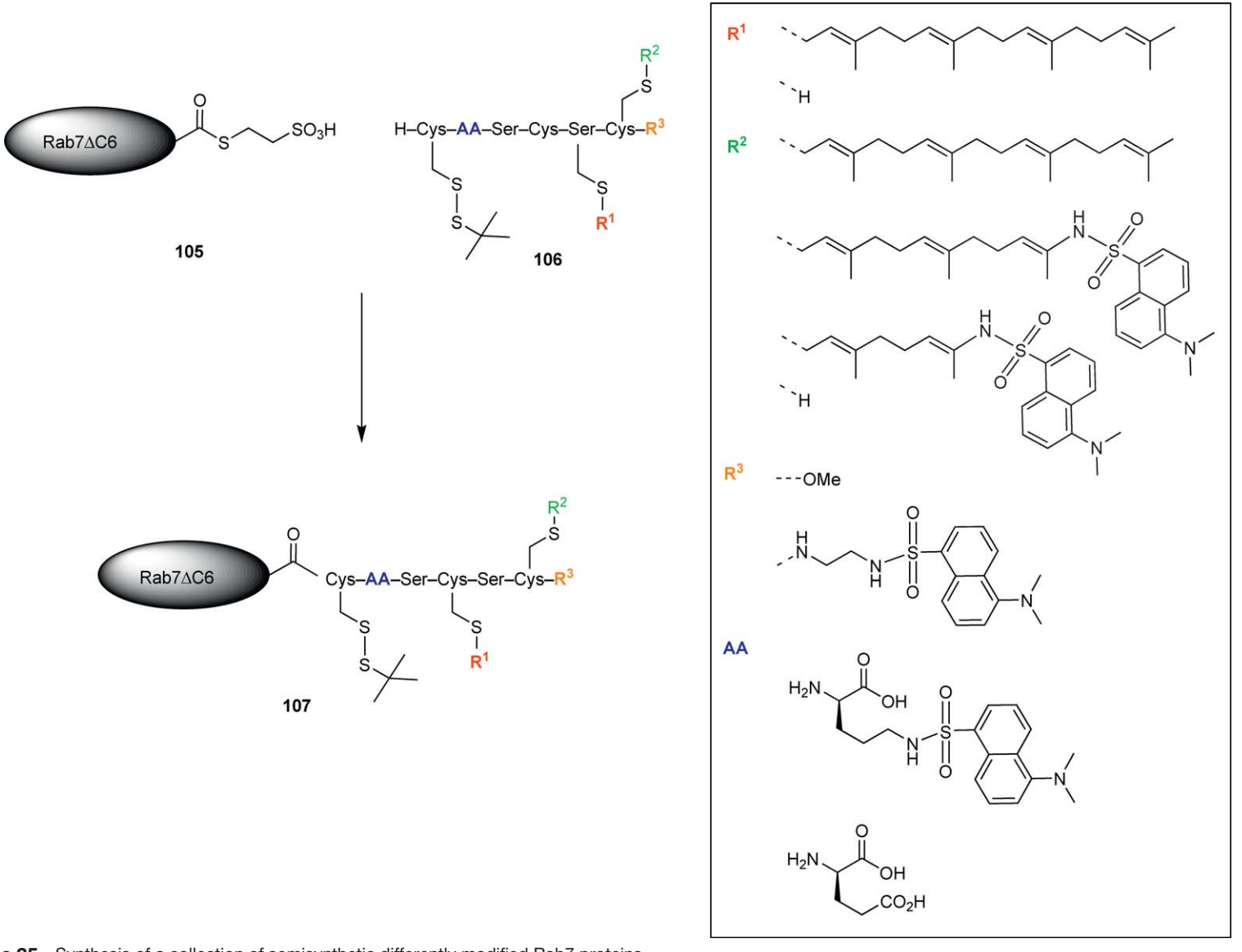
Scheme 33 Expressed protein ligation (EPL) as an example for a protein ligation method generating a native bond (Left). MIC ligation generating a non-native bond between the two coupled segments (right).



Scheme 34 Overview of native chemical ligation (NCL). Two unprotected segments react in a reversible thiol/thioester reaction; only the thioester product between the C-terminal thioester and the N-terminal cysteine can react further to form the desired amide bond via nucleophilic attack of the cysteine amine group.

5.17.3.3.2 Expressed protein ligation

In 1998 Muir and coworkers published a powerful expansion of the NCL; recombinantly expressed proteins bearing a C-terminal thioester were coupled to synthetic peptide sequences, thus leading to a method for protein engineering that allows the chemoselective addition of a peptide to a recombinant protein.[191] The incorporation of non-natural amino acids, biophysical probes, as well as post-translational modifications became feasible with this method allowing more detailed analysis and studies of proteins.[199–203] This method was successfully applied to the synthesis of lipoproteins. For instance, Rab proteins can be expressed in *E. coli* with a C-terminal truncation and fused to an engineered intein. This construct is additionally followed by a chitin-binding domain (CBD), allowing the purification and isolation of the corresponding C-terminal thioesters of Rab proteins in their C-terminally truncated form **105**.[204] By ligating differently modified Rab C-termini **106** to the truncated Rab7 protein core, the establishment of a fluorescence-based assay for *in vitro* prenylation was possible, elucidating mechanisms like the Rab7 double geranylgeranylation, which is catalyzed by Rab GGTase-II. **Scheme 35** shows a collection of semisynthetic Rab7 proteins **107**, into which lipid moieties as well as fluorophores were incorporated.[151,205,206]



Scheme 35 Synthesis of a collection of semisynthetic differently modified Rab7 proteins.

Furthermore, the interaction of prenylated Rab GTPase with REP and GDP dissociation inhibitor (GDI) protein could be studied and analyzed.[207]

These ligation reactions proceed overnight with yields around 85% of the coupled lipoprotein, which typically precipitates after ligation. After extraction with organic solvents an insoluble pellet remains that can be dissolved in 6 mol l^{-1} guanidinium chloride and then refolded by stepwise dilution into 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS)-containing buffer solution. Subsequently, REP-1 or GDI protein is added as solubilizing chaperone and after further purification via dialysis and gel filtration, correctly folded Rab GTPases can be obtained.

For Ras proteins, there is no chaperone protein available or known to solubilize the lipidated protein; thus the refolding of, for example, lipidated N- and H-Ras is not feasible. However, lipidated K-Ras **110** can be generated via expressed protein ligation (EPL), since the polybasic C-terminal region of this protein helps to solubilize the protein even when farnesylated and facilitates the purification (**Scheme 36**).[158]

Another example of lipidated proteins that have been succesfully generated via EPL are GPI-anchored proteins. Various proteins are post-translationally modified by the attachment of glycosylphosphatidylinositol anchors. These proteins are widely involved in signal transduction, immune response, cancer cell invasion, metastasis, and the pathobiology of trypanosomal parasites.[63] The membrane anchor usually consists of the glycolipid structure, also shown in **Scheme 4**. Through a phosphoethanolamine linker the C-terminus of the protein is linked to the glycan core (Manα(1,2)Manα(1,6)Manα(1,4)GlcNα(1,6)-*myo*-inositoyl) and the phospholipid chains are attached to the last carbohydrate. The glycan core is highly variable; side chains such as phosphoethanolamines and other carbohydrates can be incorporated. Furthermore, the lipid chains vary as well; they can be either saturated or unsaturated and differ in length.

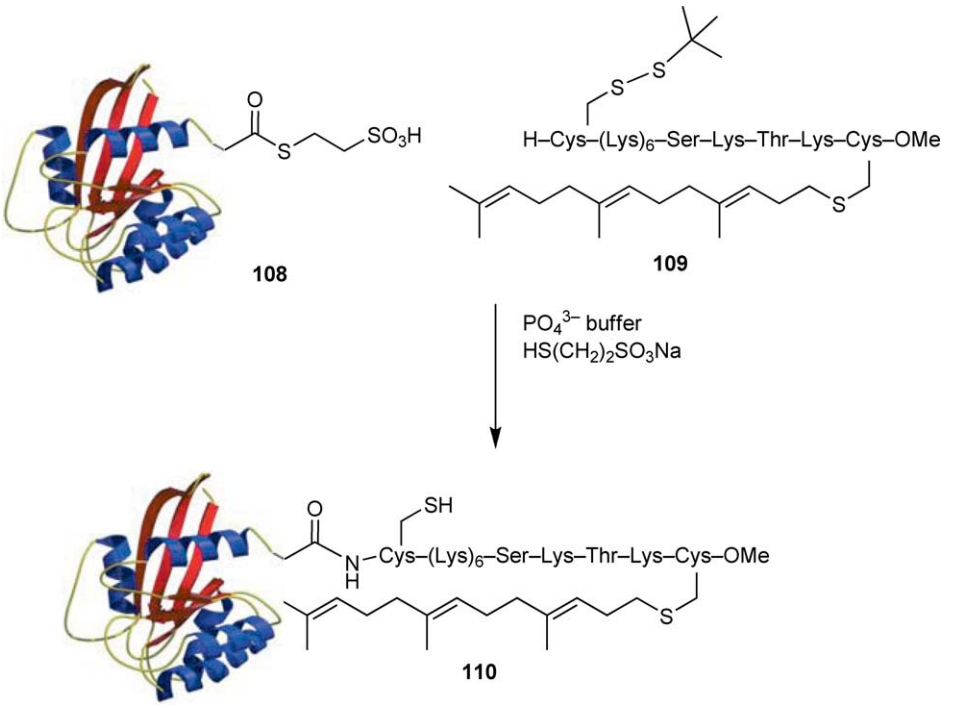
The biological function of the GPI anchor still remains unknown besides the membrane insertion. Several suggestions have been made; like GPI anchors may be targeting lipid rafts, and specific compartments.[70–74,76–78,208] Furthermore, its involvement in transmembrane signaling is discussed.[209–211] The difficulties with elucidating its biological function can be explained by the heterogeneity of GPI anchors and their limited quantities, which makes it even harder to study this complex system. Nevertheless, recently, Bertozzi and coworkers described a chemical approach for studying GPI-anchored proteins.[212,213] GPI protein analogues **111–113** were investigated in living cells with respect to their diffusion behavior (**Scheme 37**). Their insertion into membranes of various mammalian cells and the direction to the recycling endosomes was similar to the natural GPI-anchored protein and decreased diffusion could be seen when incorporating less monosaccharide units.

In this approach, GPI analogues were synthesized for which the glycan core was modified systematically with unnatural linkers; phospholipid tails and a phosphoethanolamine linker were attached. Subsequently, the phosphoethanolamine linker was coupled to a cysteine, offering the option of EPL of the GPI analogue to green fluorescent protein (GFP) bearing a thioester on the C-terminus. For the ligation of GFP and the GPI analogues **115** the GFP was overexpressed in *E. coli* as an intein fusion and a C-terminal thioester **114** was generated. This construct could then be ligated to a diverse set of membrane anchors (**Scheme 38**). By dividing the GPI proteins into two parts, the GPI analogues containing a cysteine moiety and protein thioesters, a wide range of variously modified C-terminally anchored GPI proteins is accessible.

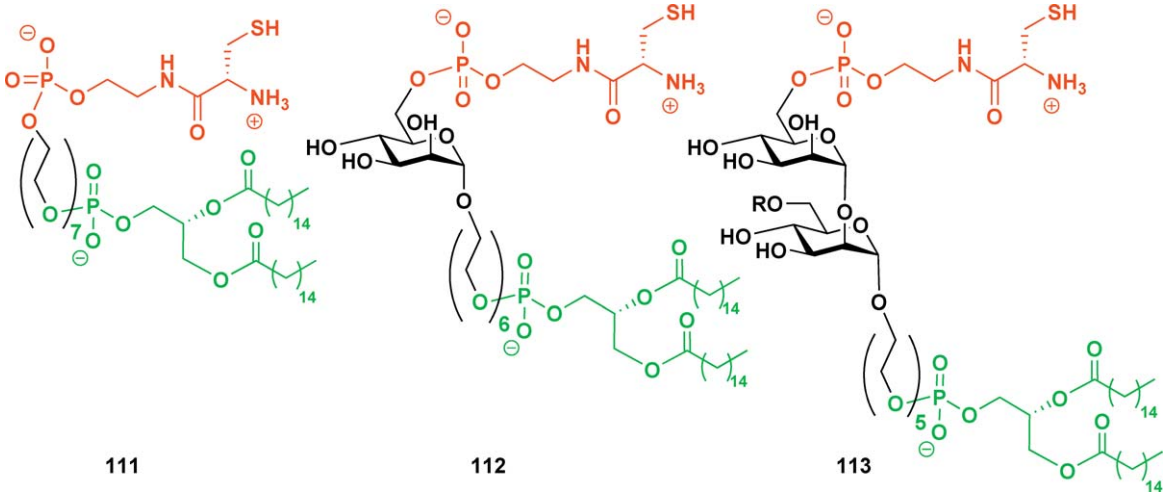
5.17.3.3.3 Maleimide ligation

The maleimide ligation (MIC ligation) method consists of the ligation of a C-terminally truncated non-lipidated protein **117** bearing a C-terminal cysteine to a N-terminal maleimido-modified lipidated peptide **118**.[135] This ligation method has frequently been applied for the generation of biologically active Ras proteins **119**. The C-terminus of the final protein is derived from organic chemistry, which allows the incorporation of various types of lipids as well as reporter groups required for biological studies, such as fluorophores, photo-activatable groups, and nonhydrolyzable palmitoyl thioester analogues. The modular nature of this approach also offers the opportunity of introducing additional non-natural building blocks (see Section 5.17.2.2).

For MIC ligations truncated H-Ras protein with a C-terminal cysteine at the position 181 was recombinantly expressed in *E. coli*. The removal of the last eight C-terminal amino acids (MSCKCVLS) had no effect on



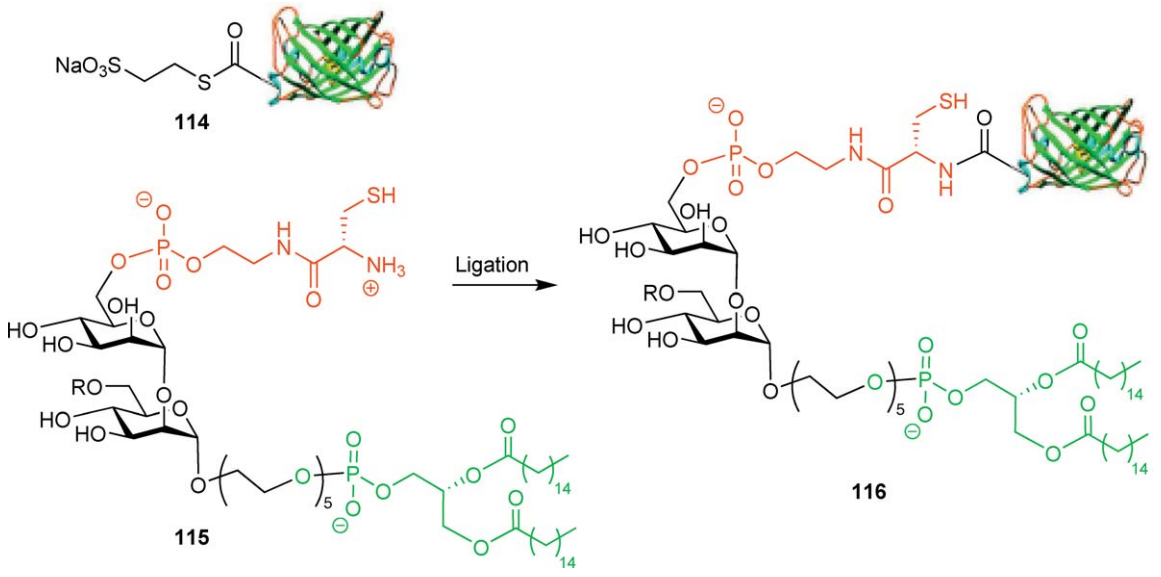
Scheme 36 Expressed protein ligation of the farnesylated K-Ras 4B C-terminus.



Scheme 37 Structures of GPI analogues that were synthesized by the Bertozzi group in order to study the diffusion behaviour in living cells.

protein expression in *E. coli* CK600K, nucleotide binding and interaction with the exchange factor Cdc25 or the Ras-binding domain of Raf-kinase.

Having the expressed truncated protein in hand, the MIC-modified lipidated peptide sequences corresponding to the C-terminus were coupled to the protein.[214] The C-terminal cysteine is highly exposed to the solvent because of the flexibility of the C-terminus. This makes the ligation reaction fast and selective. Further cysteines in the protein sequence are buried and, therefore, less accessible, limiting the extent of side reactions.[137] The ratio of peptide to protein has to be limited and generally should not exceed 3:1 to prevent additional MIC ligations. Although the hydrophobicity of the protein is increased by attachment of lipopeptides, lipidated Ras proteins still retain their high solubility in aqueous buffer, offering the option to be purified



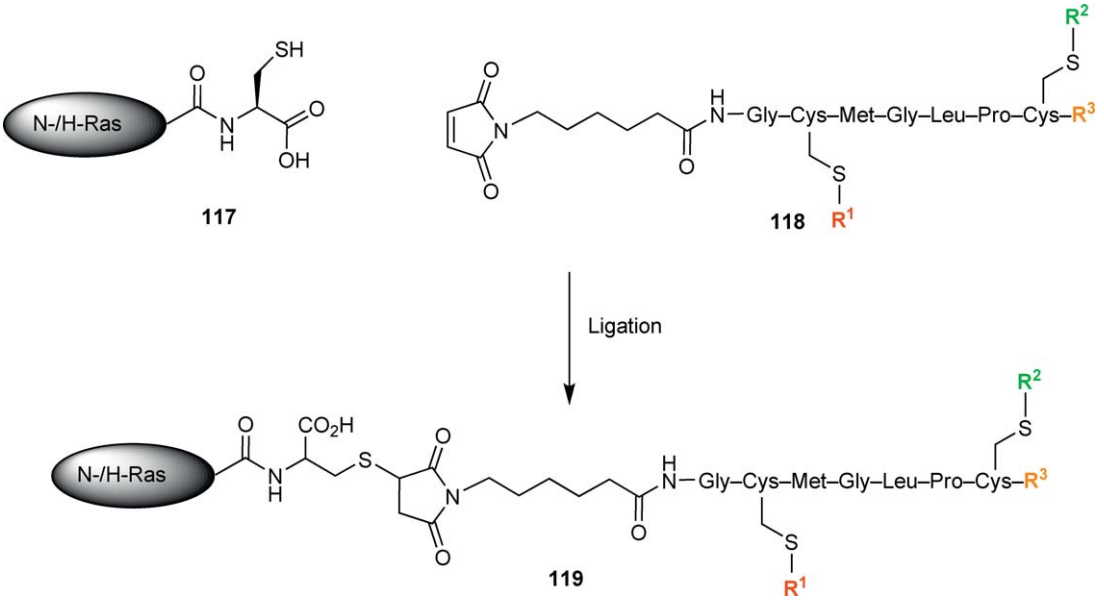
Scheme 38 Synthesis of GFP–GPI analogues: GPI analogues are coupled to GFP thioesters that were previously generated via EPL.

via extraction with a Triton X-114 saturated solution.[215] At a temperature above 30 °C the ligated protein can be separated from the unligated protein, subsequently followed by an ion exchange chromatography. A selection of semisynthetic proteins generated via this technology is shown in **Scheme 39**.

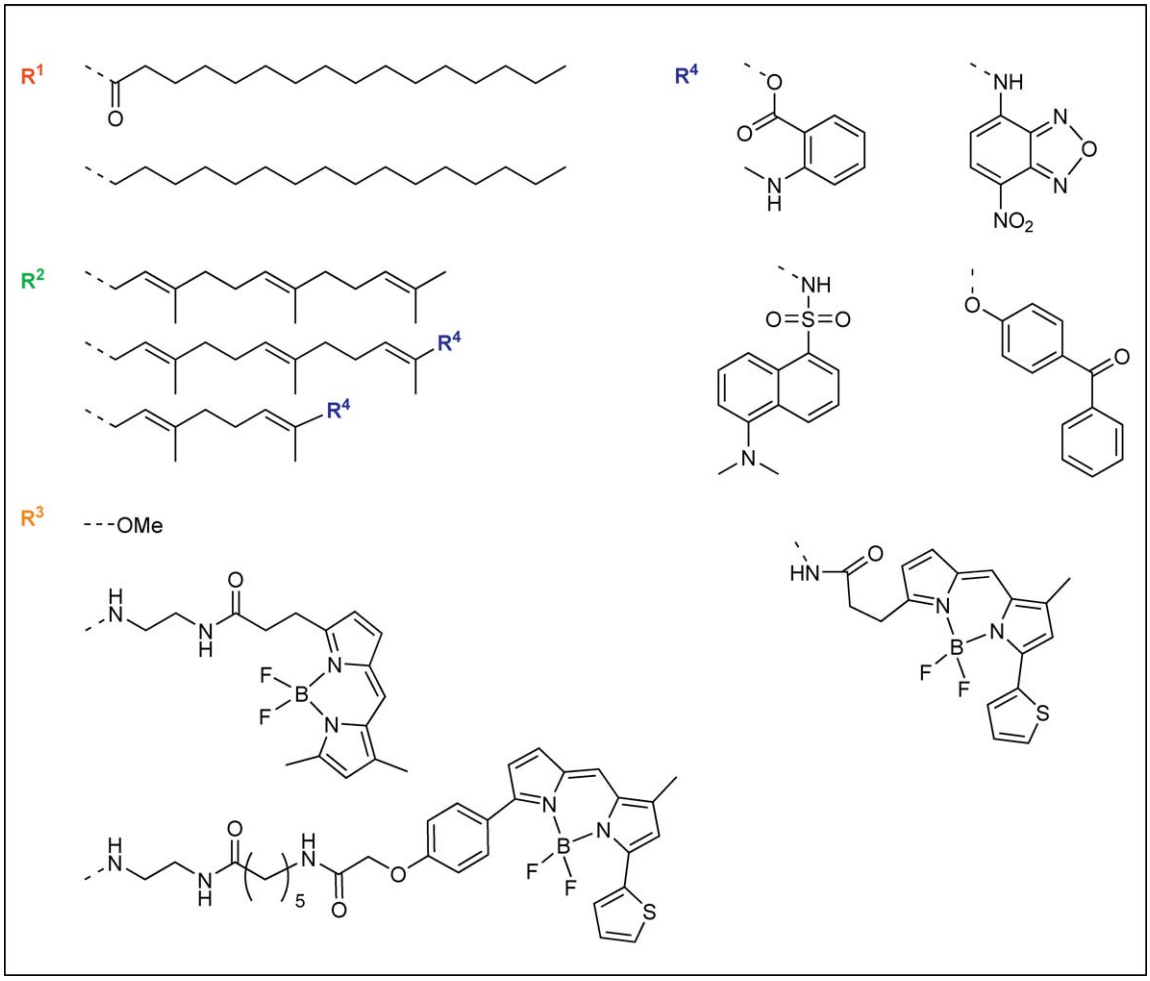
These semisynthetic proteins have served as useful tools to investigate and study the role of Ras proteins in the cell, for instance, new insights in the so-called Ras acylation cycle[41] could be obtained as well as solid-state nuclear magnetic resonance (NMR) spectroscopic analysis of the lipidated membrane anchor and proteins became possible.[216–219]

5.17.3.3.4 Diels–Alder ligation

In 2006 an alternative protein ligation was published employing the highly selective and fast Diels–Alder reaction to the synthesis of lipoproteins.[220] This reaction has already been used for bioconjugation



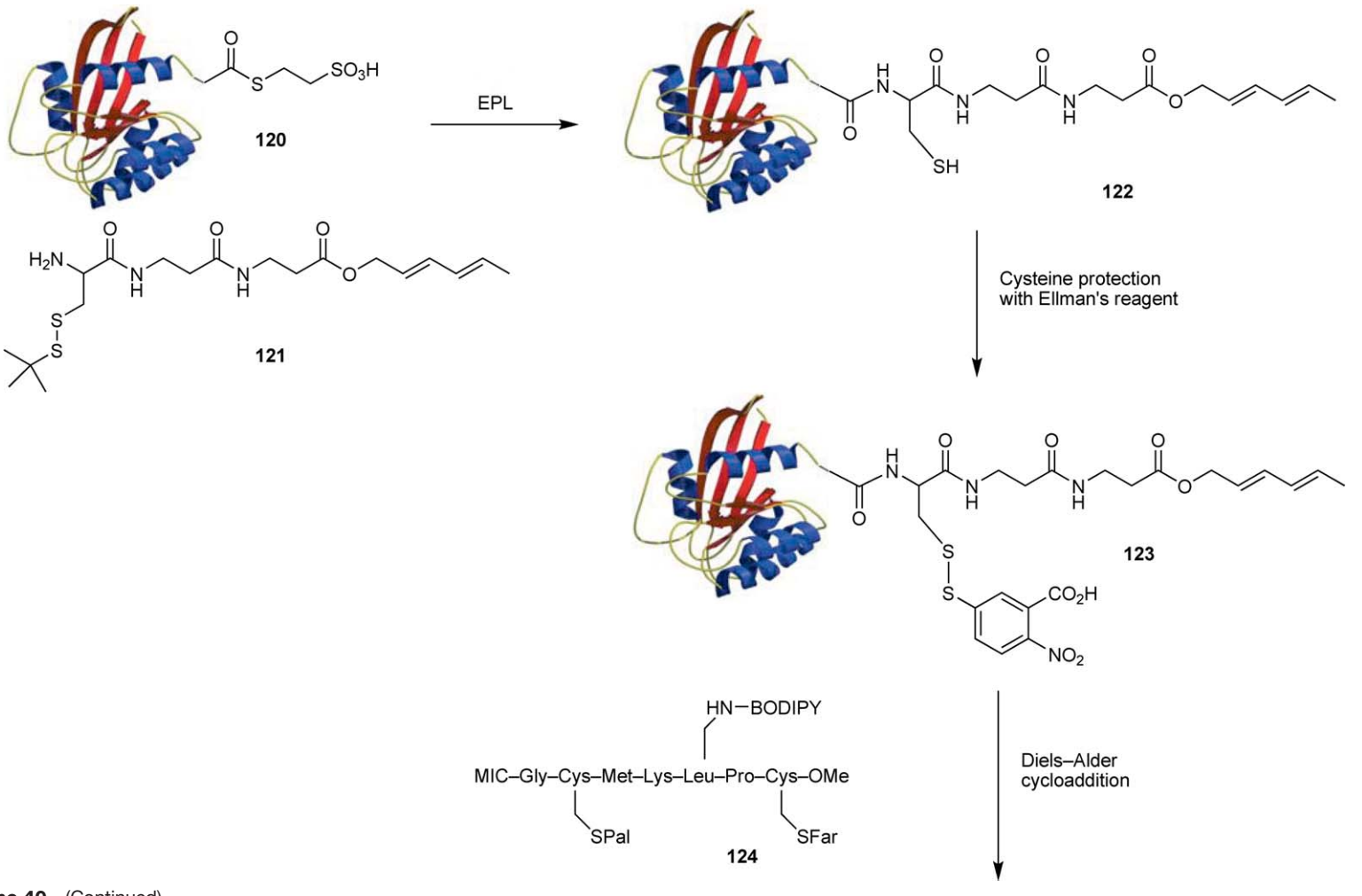
Scheme 39 (Continued)



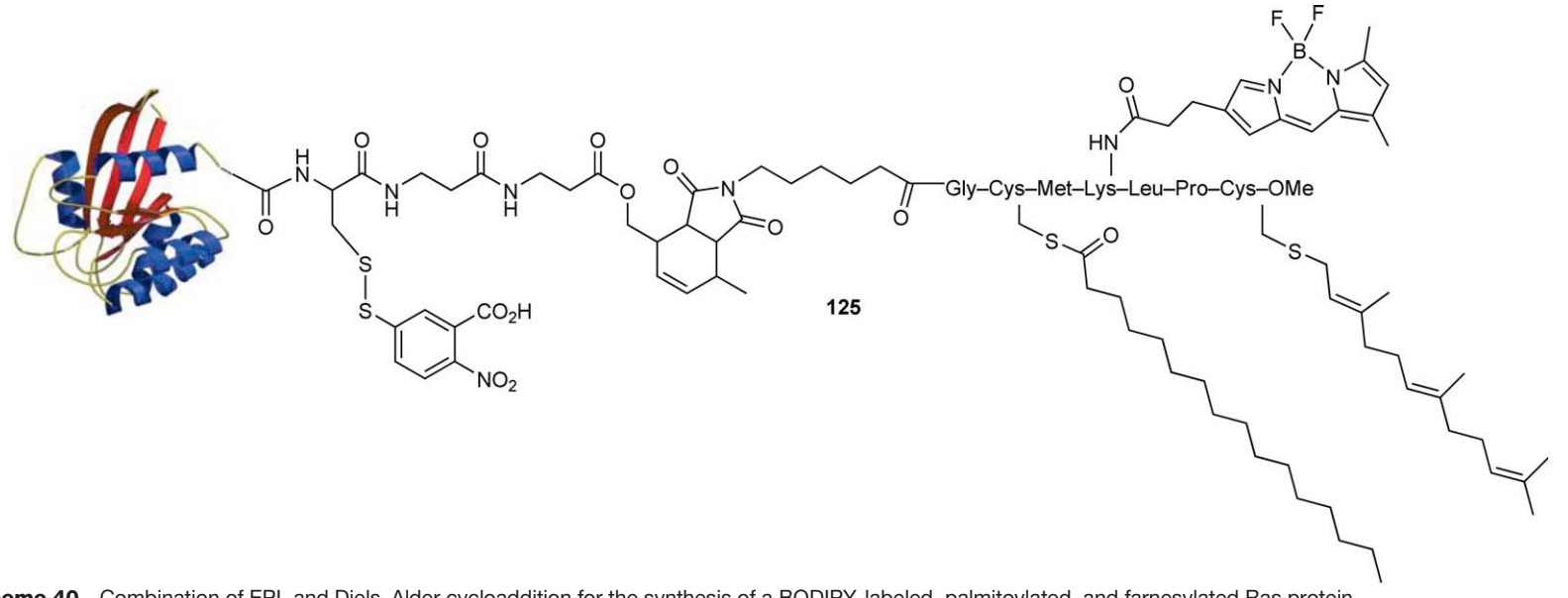
Scheme 39 Generation of semisynthetic N-/H-Ras sequences and an overwiew of proteins that could be obtained via this method.

and/or immobilization of oligonucleotides and other biomolecules.[221–229] Now, its use could be extended to ligation reactions of lipidated peptides and entire functional proteins. The cycloaddition proceeds under very mild condition at 25 °C in water within 48 h with almost complete conversion. The N-terminal segment is functionalized with an 2,4-hexadienyl ester at its C-terminus and the coupling partner requires a N-terminal maleimido group. The conjugation is compatible with all amino acid side chain functional groups except with the presence of freely accessible cysteine thiols, as these can cross react with the MIC functionality.

This method allows the combination with further conjugation and ligation techniques. For instance, expressed protein ligation can serve as a tool to introduce the diene functionality at the C-terminus. Therefore, a recombinantly expressed thioester-tagged protein is required. Since EPL generates a nucleophilic cysteine at the ligation site, a protection strategy is required for this thiol group and further cysteine side chains present in the sequence. Ellman's reagent,[230] 5,5-dithiobis(2-nitrobenzoic acid) (DTNB), was found to be suitable to block all cysteine side chains by forming stable disulfide bonds. The protocol applied for EPL included a genetically engineered intein and a chitin-binding domain as fusion partners to express and purify the desired Rab thioester.[204,231] Ligation of the thioester **120** to the peptide hexadienyl ester **121** was carried out under reducing conditions overnight at 16 °C, in the presence of GDP and $MgCl_2$ for stabilization. After masking all incorporated free cysteine side chains the Diels–Alder reaction was performed. Remarkably, this reaction allows the incorporation of reporter groups, which are not stable under the conditions of EPL. For example, 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BODIPY) lost its fluorescence during ligation via EPL (**Scheme 40**).[232,233]



Scheme 40 (Continued)



125

Scheme 40 Combination of EPL and Diels–Alder cycloaddition for the synthesis of a BODIPY-labeled, palmitoylated, and farnesylated Ras protein.

5.17.4 Conclusions

Lipidated peptides and proteins are a fascinating class of natural products. The combination of both the peptide characteristics and the lipid features gives these molecules specific properties, not shared by other types of natural products. The dynamic interaction of lipidated proteins with membranes and with specific protein chaperones are only two examples highlighting this fact. The specific and unique characteristics of lipidated peptides and proteins have, however, also limited their synthetic and biological access. Neither biological nor chemical approaches alone turned out to be sufficient to generate quantitative amounts of such proteins. With the advent of chemical biology, the generation of such systems has become possible in the last decade. The synthesis of lipidated peptides and their application to the synthesis of lipidated proteins now allows lipidated proteins to be generated, including desired chemical modifications for the study of these proteins. With these proteins biological phenomena previously not accessible via chemical or biological approaches alone can now be studied.

For such an integrated research activity, differently modified peptides and proteins that carry modifications whose structure can be changed at will through synthesis are invaluable tools. Therefore, the synthesis of the lipidated peptides is an important theme. Lipidated peptides can typically not be accessed via standardized peptide synthesis methods. However, employing the synthetic tools developed and presented here, most types of lipidated peptides can now be synthesized and obtained in pure form. Even though solution-phase approaches still play a significant role in the synthesis of lipidated peptides, the recently developed solid-phase synthesis methods delineate the preferred strategy to access the majority of the required lipidated peptides.

For the generation of lipidated proteins, protein semisynthesis approaches have been developed. The main focus in this field has been on the family of Ras proteins. These proteins feature diverse sets of lipidation patterns and can be generated via a variety of ligation methods. The intrinsic properties of the lipid motifs significantly alter the physicochemical characteristics of the proteins. Therefore, protein purification and handling methodologies were developed in parallel. This has resulted in a diverse set of protein semisynthesis approaches to generate lipidated proteins. From a chemical and protein semisynthesis point of view, the methods developed for the synthesis of Ras lipopeptides and proteins can now be applied for the investigation of other lipidated peptides and proteins with diverse and not readily accessible lipidation motifs. Recent progress in the field of GPI-anchored proteins has shown that the general methodology to generate and study lipidated proteins indeed exists and awaits application to the broad spectrum of lipidated proteins present.

Acknowledgments

Our research was generously supported by the Max Planck Gesellschaft, the Deutsche Forschungsgesellschaft, the Volkswagen Stiftung, and the Fonds der Chemischen Industrie. Luc Brunsveld would like to thank the Alexander von Humboldt foundation and the BMBF for support via the Sofja Kovalevskaja Award.

Abbreviations

Ac	acetate
AChE	acetyl choline esterase
AcOH	acetic acid
AcOZ	*p*-acetoxybenzyloxycarbonyl
All	allyl
Aloc	allyloxycarbonyl
BChE	butyryl choline esterase
Boc	*tert*-butoxycarbonyl
CBD	chitin-binding domain

Cho	choline
CRD	cysteine-rich domain
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCC	*N*,*N*′-dicyclohexylcarbodiimide
DIC	*N*,*N*′-diisopropylcarbodiimide
DIPEA	*N*,*N*-diisopropylethylamine
DMB	dimethylbarbituric acid
DMF	*N*,*N*-dimethylformamide
DMSO	dimethyl sulfoxide
DTNB	5,5′-dithiobis(2-nitrobenzoic acid)
DTT	dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EGF	epidermal growth factor
EPL	expressed protein ligation
ER	endoplasmic reticulum
Far	farnesyl
Fmoc	9-fluorenylmethoxycarbonyl
FPP	farnesyldiphosphate
FTase	farnesyl transferase
GDI	GDP dissociation inhibitor
GDP	guanosine-5′-diphosphate
GGPP	geranylgeranyldiphosphate
GGTase-I	geranylgeranyltransferase type I
GGTase-II	geranylgeranyltransferase type II
GPI	glycosylphosphatidylinositol
GST	glutathione *S*-transferase
GTP	guanosine-5′-triphosphate
HATU	2-(1H-9-azabenzotriazole-1-yl)-1,1,3,3-tetramethyl-aminium hexafluorophosphate
HBTU	2-(1H-9-benzotriazole-1-yl)-1,1,3,3-tetramethyl-aminium hexafluorophosphate
HOBt	*N*-hydroxybenzotriazole
HPLC	high-performance liquid chromatography
kDa	kiloDalton
MeCN	acetonitrile
MIC	maleimido caproic acid
Mmt	4-methoxytrityl
Myr	myristoyl
NBS	*N*-bromosuccinimide
NCL	native chemical ligation
NMT	*N*-myristoyl transferase
Pal	palmitoyl
PAT	protein *S*-acyl transferases
PhAcOZ	*p*-phenylacetoxybenzyloxycarbonyl
PM	plasma membrane
PPh_3	triphenylphosphine
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
REP	Rab escort protein
SPPS	solid-phase peptide synthesis
***t*Bu**	*tert*-butyl
TEA, Et_3N	triethylamine
TES	triethylsilane
TFA	trifluoroacetic acid

THF	tetrahydrofuran
TMD	transmembrane domains
TPPTS	triphenylphosphanyltrisulfonate
Trt	trityl

References

1. M. D. Resh, *Nat. Chem. Biol.* **2006**, *2*, 584–590.
2. Y. Takai; T. Sasaki; T. Matozaki, *Physiol. Rev.* **2001**, *81*, 153–208.
3. P. W. Bowyer; E. W. Tate; R. J. Leatherbarrow; A. A. Holder; D. F. Smith; K. A. Brown, *ChemMedChem* **2008**, *3*, 402–408.
4. R. T. Eastman; F. S. Buckner; K. Yokoyama; M. H. Gelb; W. C. Van Voorhis, *J. Lipid Res.* **2006**, *47*, 233–240.
5. M. H. Gelb; L. Brunsveld; C. A. Hrycyna; S. Michaelis; F. Tamanoi; W. C. Van Voorhis; H. Waldmann, *Nat. Chem. Biol.* **2006**, *2*, 518–528.
6. P. Konstantinopoulos; M. V. Karamouzis; A. G. Papavassiliou, *Nat. Rev. Drug Discov.* **2007**, *6*, 541–555.
7. S. M. Sebti, *Cancer Cell* **2005**, *7*, 297–300.
8. J. Folch; I. Ascoli; M. Lees; J. A. Meath; F. N. Lebaron, *J. Biol. Chem.* **1951**, *191*, 833–841.
9. Y. Kamiya; A. Sakurai; S. Tamura; N. Takahashi; K. Abe; E. Tsuchiya; S. Fukui; C. Kitada; M. Fujino, *Biochem. Biophys. Res. Commun.* **1978**, *83*, 1077–1083.
10. T. A. Farazi; G. Waksman; J. I. Gordon, *J. Biol. Chem.* **2001**, *276*, 39501–39504.
11. A. Wolven; H. Okamura; Y. Rosenblatt; M. D. Resh, *Mol. Biol. Cell* **1997**, *8*, 1159–1173.
12. J. P. Zha; S. Weiler; K. J. Oh; M. C. Wei; S. J. Korsmeyer, *Science* **2000**, *290*, 1761–1765.
13. D. A. Towler; S. P. Adams; S. R. Eubanks; D. S. Towery; E. Jacksonmachelski; L. Glaser; J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 2708–2712.
14. M. D. Resh, *Biochim. Biophys. Acta* **1999**, *1451*, 1–16.
15. S. Mclaughlin; A. Aderem, *Trends Biochem. Sci.* **1995**, *20*, 272–276.
16. J. Goldberg, *Cell* **1998**, *95*, 237–248.
17. L. Hermida-Matsumoto; M. D. Resh, *J. Virol.* **1999**, *73*, 1902–1908.
18. J. T. Dunphy; M. E. Linder, *Biochim. Biophys. Acta* **1998**, *1436*, 245–261.
19. F. L. Zhang; P. J. Casey, *Annu. Rev. Biochem.* **1996**, *65*, 241–269.
20. C. C. Farnsworth; S. L. Wolda; M. H. Gelb; J. A. Glomset, *J. Biol. Chem.* **1989**, *264*, 20422–20429.
21. S. L. Wolda; J. A. Glomset, *J. Biol. Chem.* **1988**, *263*, 5997–6000.
22. L. A. Beck; T. J. Hosick; M. Sinensky, *J. Cell Biol.* **1988**, *107*, 1307–1316.
23. J. F. Hancock; A. I. Magee; J. E. Childs; C. J. Marshall, *Cell* **1989**, *57*, 1167–1177.
24. W. R. Schafer; R. Kim; R. Sterne; J. Thorner; S. H. Kim; J. Rine, *Science* **1989**, *245*, 379–385.
25. P. J. Casey; P. A. Solski; C. J. Der; J. E. Buss, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8323–8327.
26. M. Barbacid, *Annu. Rev. Biochem.* **1987**, *56*, 779–827.
27. S. Clarke, *Annu. Rev. Biochem.* **1992**, *61*, 355–386.
28. C. C. Farnsworth; M. H. Gelb; J. A. Glomset, *Science* **1990**, *247*, 320–322.
29. H. C. Rilling; E. Breunger; W. W. Epstein; P. F. Crain, *Science* **1990**, *247*, 318–320.
30. B. T. Kinsella; W. A. Maltese, *J. Biol. Chem.* **1991**, *266*, 8540–8544.
31. R. Khosravifar; G. J. Clark; K. Abe; A. D. Cox; T. Mclain; R. J. Lutz; M. Sinensky; C. J. Der, *J. Biol. Chem.* **1992**, *267*, 24363–24368.
32. D. A. Andres; M. C. Seabra; M. S. Brown; S. A. Armstrong; T. E. Smeland; M. F. P. Cremers; J. L. Goldstein, *Cell* **1993**, *73*, 1091–1099.
33. S. Araki; K. Kaibuchi; T. Sasaki; Y. Hata; Y. Takai, *Mol. Cell. Biol.* **1991**, *11*, 1438–1447.
34. K. Alexandrov; H. Horiuchi; O. Steelemortimer; M. C. Seabra; M. Zerial, *EMBO J.* **1994**, *13*, 5262–5273.
35. M. Calero; C. Z. Chen; W. Y. Zhu; N. Winand; K. A. Havas; P. M. Gilbert; C. G. Burd; R. N. Collins, *Mol. Biol. Cell* **2003**, *14*, 1852–1867.
36. M. Wherlock; A. Gampel; C. Futter; H. Mellor, *J. Cell Sci.* **2004**, *117*, 3221–3231.
37. G. Milligan; M. Parenti; A. I. Magee, *Trends Biochem. Sci.* **1995**, *20*, 181–186.
38. M. J. Nadolski; M. E. Linder, *FEBS J.* **2007**, *274*, 5202–5210.
39. K. K. Y. Lam; M. Davey; B. M. Sun; A. F. Roth; N. G. Davis; E. Conibear, *J. Cell Biol.* **2006**, *174*, 19–25.
40. J. Valdez-Taubas; H. Pelham, *EMBO J.* **2005**, *24*, 2524–2532.
41. O. Rocks; A. Peyker; M. Khams; P. J. Verveer; C. Koerner; M. Lumbierres; J. Kuhlmann; H. Waldmann; A. Wittinghofer; P. I. H. Bastiaens, *Science* **2005**, *307*, 1746–1752.
42. I. R. Vetter; A. Wittinghofer, *Science* **2001**, *294*, 1299–1304.
43. J. A. Duncan; A. G. Gilman, *J. Biol. Chem.* **1998**, *273*, 15830–15837.
44. L. X. Liu; F. Margottin; S. LeGall; O. Schwartz; L. Selig; R. Benarous; S. Benichou, *J. Biol. Chem.* **1997**, *272*, 13779–13785.
45. J. T. Dunphy; W. K. Greentree; M. E. Linder, *J. Biol. Chem.* **2001**, *276*, 43300–43304.
46. A. S. Varner; C. E. Ducker; Z. P. Xia; Y. Zhuang; M. L. De Vos; C. D. Smith, *Biochem. J.* **2003**, *373*, 91–99.
47. M. C. Bano; C. S. Jackson; A. I. Magee, *Biochem. J.* **1998**, *330*, 723–731.
48. M. Veit, *Biochem. J.* **2000**, *345*, 145–151.

49. L. Berthiaume; I. Deichaite; S. Peseckis; M. D. Resh, *J. Biol. Chem.* **1994**, *269*, 6498–6505.
50. M. M. Corvi; C. L. M. Soltys; L. G. Berthiaume, *J. Biol. Chem.* **2001**, *276*, 45704–45712.
51. D. Kummel; U. Heinemann; M. Veit, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12701–12706.
52. A. F. Roth; J. M. Wan; A. O. Bailey; B. M. Sun; J. A. Kuchar; W. N. Green; B. S. Phinney; J. R. Yates; N. G. Davis, *Cell* **2006**, *125*, 1003–1013.
53. S. Lobo; W. K. Greentree; M. E. Linder; R. J. Deschenes, *J. Biol. Chem.* **2002**, *277*, 41268–41273.
54. D. A. Mitchell; L. Farh; T. K. Marshall; R. J. Deschenes, *J. Biol. Chem.* **1994**, *269*, 21540–21546.
55. L. H. Zhao; S. Lobo; X. W. Dong; A. D. Ault; R. J. Deschenes, *J. Biol. Chem.* **2002**, *277*, 49352–49359.
56. D. J. Bartels; D. A. Mitchell; X. W. Dong; R. J. Deschenes, *Mol. Cell. Biol.* **1999**, *19*, 6775–6787.
57. A. F. Roth; Y. Feng; L. Y. Chen; N. G. Davis, *J. Cell Biol.* **2002**, *159*, 23–28.
58. J. T. Swarthout; S. Lobo; L. Farh; M. R. Croke; W. K. Greentree; R. J. Deschenes; M. E. Linder, *J. Biol. Chem.* **2005**, *280*, 31141–31148.
59. M. E. Linder; R. J. Deschenes, *J. Cell Sci.* **2004**, *117*, 521–526.
60. D. A. Mitchell; A. Vasudevan; M. E. Linder; R. J. Deschenes, *J. Lipid Res.* **2006**, *47*, 1118–1127.
61. T. Magee; M. C. Seabra, *Curr. Opin. Cell Biol.* **2005**, *17*, 190–196.
62. P. Orlean; A. K. Menon, *J. Lipid Res.* **2007**, *48*, 993–1011.
63. O. Nosjean; A. Briolay; B. Roux, *Biochim. Biophys. Acta* **1997**, *1331*, 153–186.
64. W. L. Roberts; S. Santikarn; V. N. Reinhold; T. L. Rosenberry, *J. Biol. Chem.* **1988**, *263*, 18776–18784.
65. I. A. Brewis; M. A. J. Ferguson; A. Mehlert; A. J. Turner; N. M. Hooper, *J. Biol. Chem.* **1995**, *270*, 22946–22956.
66. C. Fankhauser; S. W. Homans; J. E. Thomasoates; M. J. Mcconville; C. Desponds; A. Conzelmann; M. A. J. Ferguson, *J. Biol. Chem.* **1993**, *268*, 26365–26374.
67. S. W. Homans; M. A. J. Ferguson; R. A. Dwek; T. W. Rademacher; R. Anand; A. F. Williams, *Nature* **1988**, *333*, 269–272.
68. M. A. J. Ferguson; S. W. Homans; R. A. Dwek; T. W. Rademacher, *Science* **1988**, *239*, 753–759.
69. W. L. Roberts; J. J. Myher; A. Kuksis; M. G. Low; T. L. Rosenberry, *J. Biol. Chem.* **1988**, *263*, 18766–18775.
70. P. Sharma; R. Varma; R. C. Sarasij; Ira; K. Gousset; G. Krishnamoorthy; M. Rao; S. Mayor, *Cell* **2004**, *116*, 577–589.
71. R. Varma; S. Mayor, *Nature* **1998**, *394*, 798–801.
72. T. Friedrichson; T. V. Kurzchalia, *Nature* **1998**, *394*, 802–805.
73. S. Sabharanjak; P. Sharma; R. G. Parton; S. Mayor, *Dev. Cell* **2002**, *2*, 411–423.
74. B. J. Nichols; A. K. Kenworthy; R. S. Polishchuk; R. Lodge; T. H. Roberts; K. Hirschberg; R. D. Phair; J. Lippincott-Schwartz, *J. Cell Biol.* **2001**, *153*, 529–541.
75. S. Chatterjee; E. R. Smith; K. Hanada; V. L. Stevens; S. Mayor, *EMBO J.* **2001**, *20*, 1583–1592.
76. M. Fivaz; F. Vilbois; S. Thurnheer; C. Pasquali; L. Abrami; P. E. Bickel; R. G. Parton; F. G. van der Goot, *EMBO J.* **2002**, *21*, 3989–4000.
77. S. Schuck; K. Simons, *J. Cell Biol.* **2006**, *172*, 963–965.
78. S. Paladino; T. Pocard; M. A. Catino; C. Zurzolo, *J. Cell Biol.* **2006**, *172*, 1023–1034.
79. R. K. Mann; P. A. Beachy, *Annu. Rev. Biochem.* **2004**, *73*, 891–923.
80. S. Eaton, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 437–445.
81. D. M. Berman; S. S. Karhadkar; A. Maitra; R. M. de Oca; M. R. Gerstenblith; K. Briggs; A. R. Parker; Y. Shimada; J. R. Eshleman; D. N. Watkins; P. A. Beachy, *Nature* **2003**, *425*, 846–851.
82. J. A. Porter; K. E. Young; P. A. Beachy, *Science* **1996**, *274*, 255–259.
83. J. A. Porter; S. C. Ekker; W. J. Park; D. P. vonKessler; K. E. Young; C. H. Chen; Y. Ma; A. S. Woods; R. J. Cotter; E. V. Koonin; P. A. Beachy, *Cell* **1996**, *86*, 21–34.
84. Z. Chamoun; R. K. Mann; D. Nellen; D. P. von Kessler; M. Bellotto; P. A. Beachy; K. Basler, *Science* **2001**, *293*, 2080–2084.
85. R. B. Pepinsky; L. L. Chen; W. Meier; B. P. Wallner, *J. Biol. Chem.* **1991**, *266*, 18244–18249.
86. A. Gallet; R. Rodriguez; L. Ruel; P. P. Therond, *Dev. Cell* **2003**, *4*, 191–204.
87. A. Gallet; L. Ruel; L. Staccini-Lavenant; P. R. Therond, *Development* **2006**, *133*, 407–418.
88. J. D. Lee; P. Kraus; N. Gaiano; S. Nery; J. Kohtz; G. Fishell; C. A. Loomis; J. E. Treisman, *Dev. Biol.* **2001**, *233*, 122–136.
89. M. H. Chen; Y. J. Li; T. Kawakami; S. M. Xu; P. T. Chuang, *Genes Dev.* **2004**, *18*, 641–659.
90. M. van de Wetering; E. Sancho; C. Verweij; W. de Lau; I. Oving; A. Hurlstone; K. van der Horn; E. Batlle; D. Coudreuse; A. P. Haramis; M. Tion-Pon-Fong; P. Moerer; M. van den Born; G. Soete; S. Pals; M. Eilers; R. Medema; H. Clevers, *Cell* **2002**, *111*, 241–250.
91. M. I. Garcia-Castro; C. Marcelle; M. Bronner-Fraser, *Science* **2002**, *297*, 848–851.
92. K. M. Cadigan; R. Nusse, *Genes Dev.* **1997**, *11*, 3286–3305.
93. K. Willert; J. D. Brown; E. Danenberg; A. W. Duncan; I. L. Weissman; T. Reya; J. R. Yates; R. Nusse, *Nature* **2003**, *423*, 448–452.
94. R. Takada; Y. Satomi; T. Kurata; N. Ueno; S. Norioka; H. Kondoh; T. Takao; S. Takada, *Dev. Cell* **2006**, *11*, 791–801.
95. X. Franch-Marro; F. Wendler; J. Griffith; M. M. Maurice; J. P. Vincent, *J. Cell Sci.* **2008**, *121*, 1587–1592.
96. P. Stober; M. Schelhaas; E. Nagele; P. Hagenbuch; J. Retey; H. Waldmann, *Bioorg. Med. Chem.* **1997**, *5*, 75–83.
97. M. Schelhaas; S. Glomsda; M. Hansler; H. D. Jakubke; H. Waldmann, *Angew. Chem. Int. Ed.* **1996**, *35*, 106–109.
98. K. H. Wiesmuller; W. G. Bessler; G. Jung, *Int. J. Pept. Protein Res.* **1992**, *40*, 255–260.
99. H. J. Schramm; J. Boetzel; J. Buttner; E. Fritsche; W. Gohring; E. Jaeger; S. Konig; O. Thumfart; T. Wenger; N. E. Nagel; W. Schramm, *Antiviral Res.* **1996**, *30*, 155–170.
100. B. Deprez; J. P. Sauzet; C. Boutillon; F. Martinon; A. Tartar; C. Sergheraert; J. G. Guillet; E. Gomard; H. GrasMasse, *Vaccine* **1996**, *14*, 375–382.
101. R. G. Hiskey; T. Mizoguch; T. Inui, *J. Org. Chem.* **1966**, *31*, 1192–1195.
102. S. Quesnel; J. R. Silvius, *Biochemistry* **1994**, *33*, 13340–13348.
103. E. Yousefi-Salakdeh; J. Johansson; R. Stromberg, *Biochem. J.* **1999**, *343*, 557–562.
104. F. R. Naider; J. M. Becker, *Biopolymers* **1997**, *43*, 3–14.
105. Y. S. Or; R. F. Clark; J. R. Luly, *J. Org. Chem.* **1991**, *56*, 3146–3149.
106. C. C. Yang; C. K. Marlowe; R. Kania, *J. Am. Chem. Soc.* **1991**, *113*, 3177–3178.

107. S. Shahinian; J. R. Silvius, *Biochemistry* **1995**, *34*, 3813–3822.
108. C. B. Xue; J. M. Becker; F. Naider, *Int. J. Pept. Protein Res.* **1991**, *37*, 476–486.
109. M. Koppitz; T. Spellig; R. Kahmann; H. Kessler, *Int. J. Pept. Protein Res.* **1996**, *48*, 377–390.
110. F. Ghomashchi; X. H. Zhang; L. Liu; M. H. Gelb, *Biochemistry* **1995**, *34*, 11910–11918.
111. M. W. Rose; J. H. Xu; T. A. Kale; G. O'Doherty; G. Barany; M. D. Distefano, *Biopolymers* **2005**, *80*, 164–171.
112. T. C. Turek; I. Gaon; M. D. Distefano; C. L. Strickland, *J. Org. Chem.* **2001**, *66*, 3253–3264.
113. R. L. Edelstein; M. D. Distefano, *Biochem. Biophys. Res. Commun.* **1997**, *235*, 377–382.
114. D. Kadereit; J. Kuhlmann; H. Waldmann, *ChemBioChem* **2000**, *1*, 144–169.
115. T. Kappes; H. Waldmann, *Carbohydr. Res.* **1997**, *305*, 341–349.
116. H. Waldmann; A. Heuser, *Bioorg. Med. Chem.* **1994**, *6*, 477–482.
117. E. Nagele; M. Schelhaas; N. Kuder; H. Waldmann, *J. Am. Chem. Soc.* **1998**, *120*, 6889–6902.
118. T. Pohl; H. Waldmann, *J. Am. Chem. Soc.* **1997**, *119*, 6702–6710.
119. R. Machauer; H. Waldmann, *Angew. Chem. Int. Ed.* **2000**, *39*, 1449–1453.
120. J. Sander; H. Waldmann, *Angew. Chem. Int. Ed.* **1999**, *38*, 1250–1252.
121. S. Flohr; V. Jungmann; H. Waldmann, *Chem. A Eur. J.* **1999**, *5*, 669–681.
122. M. Schelhaas; E. Nagele; N. Kuder; B. Bader; J. Kuhlmann; A. Wittinghofer; H. Waldmann, *Chem. A Eur. J.* **1999**, *5*, 1239–1252.
123. A. Cotte; B. Bader; J. Kuhlmann; H. Waldmann, *Chem. A Eur. J.* **1999**, *5*, 922–936.
124. H. Kunz; H. Waldmann, *Angew. Chem. Int. Ed.* **1984**, *23*, 71–72.
125. H. Kunz; C. Unverzagt, *Angew. Chem. Int. Ed.* **1984**, *23*, 436–437.
126. F. Guibe, *Tetrahedron* **1998**, *54*, 2967–3042.
127. F. Guibe, *Tetrahedron* **1997**, *53*, 13509–13556.
128. J. P. Genet; E. Blart; M. Savignac; S. Lemeune; S. Lemaireaudoire; J. M. Paris; J. M. Bernard, *Tetrahedron* **1994**, *50*, 497–503.
129. S. LemaireAudoire; M. Savignac; E. Blart; J. M. Bernard; J. P. Genet, *Tetrahedron Lett.* **1997**, *38*, 2955–2958.
130. S. LemaireAudoire; M. Savignac; G. Pourcelot; J. P. Genet; J. M. Bernard, *J. Mol. Catal. A Chem.* **1997**, *116*, 247–258.
131. T. Schmittberger; A. Cotte; H. Waldmann, *Chem. Commun.* **1998**, 937–938.
132. H. Waldmann; M. Schelhaas; E. Nagele; J. Kuhlmann; A. Wittinghofer; H. Schroeder; J. R. Silvius, *Angew. Chem. Int. Ed.* **1997**, *36*, 2238–2241.
133. T. Schmittberger; H. Waldmann, *Biorg. Med. Chem.* **1999**, *7*, 749–762.
134. I. Rubiol; U. Wittig; C. Meyer; R. Heinze; D. Kadereit; H. Waldmann; L. Downward; R. Wetzker, *Eur. J. Biochem.* **1999**, *266*, 70–82.
135. K. Kuhn; D. J. Owen; B. Bader; A. Wittinghofer; J. Kuhlmann; H. Waldmann, *J. Am. Chem. Soc.* **2001**, *123*, 1023–1035.
136. F. Eisele; J. Kuhlmann; H. Waldmann, *Chem. A Eur. J.* **2002**, *8*, 3362–3376.
137. B. Bader; K. Kuhn; D. J. Owen; H. Waldmann; A. Wittinghofer; J. Kuhlmann, *Nature* **2000**, *403*, 223–226.
138. K. Pachamuthu; X. M. Zhu; R. R. Schmidt, *J. Org. Chem.* **2005**, *70*, 3720–3723.
139. B. Ludolph; F. Eisele; H. Waldmann, *J. Am. Chem. Soc.* **2002**, *124*, 5954–5955.
140. B. Ludolph; F. Eisele; H. Waldmann, *ChemBioChem* **2002**, *3*, 901–904.
141. M. Joseph; R. Nagaraj, *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1025–1028.
142. N. Pfender; E. Guenin; J. M. Greer; E. Trifilieff, *Lett. Pept. Sci.* **2003**, *10*, 581–588.
143. B. Denis; E. Trifilieff, *J. Pept. Sci.* **2000**, *6*, 372–377.
144. D. T. S. Rijkers; J. A. W. Kruijtzer; J. A. Killian; R. M. J. Liskamp, *Tetrahedron Lett.* **2005**, *46*, 3341–3345.
145. B. Pallavi; R. Nagaraj, *J. Biol. Chem.* **2003**, *278*, 12737–12744.
146. A. S. Varner; M. L. De Vos; S. P. Creaser; B. R. Peterson; C. D. Smith, *Analyt. Biochem.* **2002**, *308*, 160–167.
147. P. Mayer-Fligge; J. Volz; U. Kruger; E. Sturm; W. Gernandt; K. P. Schafer; M. Przybylski, *J. Pept. Sci.* **1998**, *4*, 355–363.
148. B. Ludolph; H. Waldmann, *Chem. A Eur. J.* **2003**, *9*, 3683–3691.
149. Y. Zhu; W. van der Donk, *Org. Lett.* **2001**, *3*, 1189–1192.
150. D. P. Galonic; N. D. Ide; W. A. van der Donk; D. Y. Gin, *J. Am. Chem. Soc.* **2005**, *127*, 7359–7369.
151. T. Durek; K. Alexandrov; R. S. Goody; A. Hildebrand; I. Heinemann; H. Waldmann, *J. Am. Chem. Soc.* **2004**, *126*, 16368–16378.
152. G. Kragol; M. Lumbierres; J. Palomo; H. Waldmann, *Angew. Chem. Int. Ed.* **2004**, 5839–5842.
153. L. Brunsveld; A. Watzke; T. Durek; K. Alexandrov; R. S. Goody; H. Waldmann, *Chem. A Eur. J.* **2005**, *11*, 2756–2772.
154. G. Triola; L. Brunsveld; H. Waldmann, *J. Org. Chem.* **2008**, 3646–3649.
155. W. F. Degrado; E. T. Kaiser, *J. Org. Chem.* **1980**, *45*, 1295–1300.
156. E. K. Dolence; J. M. Dolence; C. D. Poulter, *J. Comb. Chem.* **2000**, *2*, 522–536.
157. E. K. Dolence; J. M. Dolence; C. D. Poulter, *Bioconjugate Chem.* **2001**, *12*, 35–43.
158. D. Gottlieb; C. Grunwald; C. Nowak; J. Kuhlmann; H. Waldmann, *Chem. Commun.* **2006**, 260–262.
159. D. Kadereit; P. Deck; I. Heinemann; H. Waldmann, *Chem. A Eur. J.* **2001**, *7*, 1184–1193.
160. T. Wieland; J. Lewalter; C. Birr, *Annalen Der Chemie-Justus Liebig* **1970**, *740*, 31–47.
161. A. N. Semenov; K. Y. Gordeev, *Int. J. Pept. Protein Res.* **1995**, *45*, 303–304.
162. C. R. Millington; R. Quarrell; G. Lowe, *Tetrahedron Lett.* **1998**, *39*, 7201–7204.
163. C. Rosenbaum; H. Waldmann, *Tetrahedron Lett.* **2001**, *42*, 5677–5680.
164. C. Peters; H. Waldmann, *J. Org. Chem.* **2003**, *68*, 6053–6055.
165. B. Ludolph; F. Eisele; H. Waldmann, *J. Am. Chem. Soc.* **2002**, *124*, 5954–5955.
166. B. J. Backes; J. A. Ellman, *J. Org. Chem.* **1999**, *64*, 2322–2330.
167. P. Heidler; A. Link, *Bioorg. Med. Chem.* **2005**, *13*, 585–599.
168. J. M. Palomo; M. Lumbierres; H. Waldmann, *Angew. Chem. Int. Ed.* **2006**, *45*, 477–481.
169. W. D. Branton; C. G. Fields; V. L. Vandrisse; G. B. Fields, *Tetrahedron Lett.* **1993**, *34*, 4885–4888.
170. J. Bodi; H. Nishio; Y. Zhou; W. D. Branton; T. Kimura; S. Sakakibara, *Pept. Res.* **1995**, *8*, 228–235.
171. M. Volkert; K. Uwai; A. Tebbe; B. Popkirova; M. Wagner; J. Kuhlmann; H. Waldmann, *J. Am. Chem. Soc.* **2003**, *125*, 12749–12758.
172. C. Peters; A. Wolf; M. Wagner; J. Kuhlmann; H. Waldmann, *Proc. Natl Acad. Sci. U.S.A.* **2004**, *101*, 8531–8536.

173. R. S. Goody; K. Alexandrov; M. Engelhard, *ChemBioChem* **2002**, *3*, 399–403.
174. L. J. Knoll; J. I. Gordon, *J. Biol. Chem.* **1993**, *268*, 4281–4290.
175. J. L. Cereghino; J. M. Cregg, *FEMS Microbiol. Rev.* **2000**, *24*, 45–66.
176. M. B. Slade; K. R. Emslie; K. L. Williams, *Biotechnol. Genet. Eng. Rev.* **1997**, *14*, 1–35.
177. D. J. Manstein; K. M. Ruppel; L. Kubalek; J. A. Spudich, *J. Cell Sci.* **1991**, 63–65.
178. A. Kikuchi; T. Yamashita; M. Kawata; K. Yamamoto; K. Ikeda; T. Tanimoto; Y. Takai, *J. Biol. Chem.* **1988**, *263*, 2897–2904.
179. H. Horiuchi; O. Ullrich; C. Bucci; M. Zerial, *Methods Enzymol.* **1995**, *257*, 9–15.
180. H. W. Park; L. S. Beese, *Curr. Opin. Struct. Biol.* **1997**, *7*, 873–880.
181. I. Gaon; T. C. Turek; V. A. Weller; R. L. Edelstein; S. K. Singh; M. D. Distefano, *J. Org. Chem.* **1996**, *61*, 7738–7745.
182. T. A. Kale; S. J. Hsieh; M. W. Rose; M. D. Distefano, *Curr. Top. Med. Chem.* **2003**, *3*, 1043–1074.
183. M. W. Rose; N. D. Rose; J. Boggs; S. Lenevich; J. Xu; G. Barany; M. D. Distefano, *J. Pept. Res.* **2005**, *65*, 529–537.
184. T. C. Turek; I. Gaon; D. Gamache; M. D. Distefano, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2125–2130.
185. T. Dudler; M. H. Gelb, *Biochemistry* **1997**, *36*, 12434–12441.
186. P. Mcgeady; S. Kuroda; K. Shimizu; Y. Takai; M. H. Gelb, *J. Biol. Chem.* **1995**, *270*, 26347–26351.
187. A. Kalinin; N. H. Thoma; A. Iakovenko; I. Heinemann; E. Rostkova; A. T. Constantinescu; K. Alexandrov, *Protein Expr. Purif.* **2001**, *22*, 84–91.
188. D. J. Owen; K. Alexandrov; E. Rostkova; A. J. Scheidig; R. S. Goody; H. Waldmann, *Angew. Chem. Int. Ed.* **1999**, *38*, 509–512.
189. N. H. Thoma; A. Iakovenko; D. Owen; A. S. Scheidig; H. Waldmann; R. S. Goody; K. Alexandrov, *Biochemistry* **2000**, *39*, 12043–12052.
190. C. Budde; M. J. Schoenfish; M. E. Linder; R. J. Deschenes, *Methods* **2006**, *40*, 143–150.
191. T. W. Muir; D. Sondhi; P. A. Cole, *Proc. Natl Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710.
192. R. M. Hofmann; T. W. Muir, *Curr. Opin. Biotechnol.* **2002**, *13*, 297–303.
193. T. J. Tolbert; C. H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 5421–5428.
194. J. P. Tam; J. X. Xu; K. D. Eom, *Biopolymers* **2001**, *60*, 194–205.
195. G. G. Kochendoerfer; S. B. H. Kent, *Curr. Opin. Chem. Biol.* **1999**, *3*, 665–671.
196. P. E. Dawson; T. W. Muir; I. Clarklewis; S. B. H. Kent, *Science* **1994**, *266*, 776–779.
197. P. E. Dawson; S. B. H. Kent, *Annu. Rev. Biochem* **2000**, *69*, 923–960.
198. T. M. Hackeng; J. H. Griffin; P. E. Dawson, *Proc. Natl Acad. Sci. U.S.A.* **1999**, *96*, 10068–10073.
199. M. E. Hahn; J. P. Pellois; M. Vila-Perello; T. W. Muir, *ChemBioChem* **2007**, *8*, 2100–2105.
200. V. Muralidharan; T. W. Muir, *Nat. Methods* **2006**, *3*, 429–438.
201. T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
202. Z. Machova; A. Beck-Sickinger, *Methods Mol. Biol.* **2005**, *298*, 115–130.
203. T. W. Muir, *Synlett* **2001**, 733–740.
204. T. Durek; R. S. Goody; K. Alexandrov, *In vitro* Semisynthesis and Applications of C-Terminally Modified Rab Proteins. In *Bioconjugation Protocols: Strategies and Methods*; C. Niemeyer, Ed.; Methods in Molecular Biology 283; Humana Press: Totowa, 2004; pp 233–244.
205. K. Alexandrov; I. Heinemann; T. Durek; V. Sidorovitch; R. S. Goody; H. Waldmann, *J. Am. Chem. Soc.* **2002**, *124*, 5648–5649.
206. A. Rak; O. Pylypenko; T. Durek; A. Watzke; S. Kushnir; L. Brunsveld; H. Waldmann; R. S. Goody; K. Alexandrov, *Science* **2003**, *302*, 646–650.
207. Y. W. Wu; K. T. Tan; H. Waldmann; R. S. Goody; K. Alexandrov, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12294–12299.
208. S. Chatterjee; S. Mayor, *Cell. Mol. Life Sci.* **2001**, *58*, 1969–1987.
209. K. Simons; D. Toomre, *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 31–39.
210. V. Horejsi; K. Drbal; M. Cebecauer; J. Cerny; T. Brdicka; P. Angelisova; H. Stockinger, *Immunol. Today* **1999**, *20*, 356–361.
211. D. R. Jones; I. Varela-Nieto, *Int. J. Biochem. Cell Biol.* **1998**, *30*, 313–326.
212. M. G. Paulick; M. B. Forstner; J. T. Groves; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20332–20337.
213. M. G. Paulick; A. R. Wise; M. B. Forstner; J. T. Groves; C. R. Bertozzi, *J. Am. Chem. Soc.* **2007**, *129*, 11543–11550.
214. M. Wagner; J. Kuhlmann, Protein Lipidation. In *Bioconjugation Protocols: Strategies and Methods*; C. Niemeyer, Ed.; Methods in Molecular Biology 283; Humana Press: Totowa, 2004; pp 245–255.
215. C. Bordier, *J. Biol. Chem.* **1981**, *256*, 1604–1607.
216. G. Reuther; K. Tan; J. Köhler; C. Nowak; A. Pampel; K. Arnold; J. Kuhlmann; H. Waldmann; D. Huster, *Angew. Chem. Int. Ed.* **2006**, *45*, 5387–5390.
217. A. Vogel; G. Reuther; K. T. Tan; H. Waldmann; J. Kuhlmann; K. Arnold; S. E. Feller; M. F. Brown; D. Huster, *Biophys. J.* **2007**, *93*, 2697–2712.
218. G. Reuther; K. T. Tan; J. Kohler; C. Nowak; A. Pampel; K. Arnold; J. Kuhlmann; H. Waldmann; D. Huster, *Angew. Chem. Int. Ed.* **2006**, *45*, 5387–5390.
219. A. Vogel; C. P. Katzka; H. Waldmann; K. Arnold; M. F. Brown; D. Huster, *J. Am. Chem. Soc.* **2005**, *127*, 12263–12272.
220. A. D. de Araujo; J. M. Palomo; J. Cramer; M. Kohn; H. Schroder; R. Wacker; C. Niemeyer; K. Alexandrov; H. Waldmann, *Angew. Chem. Int. Ed.* **2006**, *45*, 296–301.
221. K. W. Hill; J. Taunton-Rigby; J. D. Carter; E. Kropp; K. Vagle; W. Pieken; D. P. C. McGee; G. M. Husar; M. Leuck; D. J. Anziano; D. P. Sebesta, *J. Org. Chem.* **2001**, *66*, 5352–5358.
222. G. M. Husar; D. J. Anziano; M. Leuck; D. P. Sebesta, *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 559–566.
223. H. A. Latham-Timmons; A. Wolter; J. S. Roach; R. Giare; M. Leuck, *Nucleosides Nucleotides Nucleic Acids* **2003**, *22*, 1495–1497.
224. B. T. Houseman; J. H. Huh; S. J. Kron; M. Mrksich, *Nat. Biotechnol.* **2002**, *20*, 270–274.
225. W. S. Dillmore; M. N. Yousaf; M. Mrksich, *Langmuir* **2004**, *20*, 7223–7231.
226. B. T. Houseman; M. Mrksich, *Chem. Biol.* **2002**, *9*, 443–454.
227. A. Berkin; B. Coxon; V. Pozsgay, *Chem. A Eur. J.* **2002**, *8*, 4424–4433.
228. V. Pozsgay; N. E. Vieira; A. Yergey, *Org. Lett.* **2002**, *4*, 3191–3194.
229. W. S. Yeo; M. N. Yousaf; M. Mrksich, *J. Am. Chem. Soc.* **2003**, *125*, 14994–14995.

230. G. L. Ellman, *Arch. Biochem. Biophys.* **1959**, *82*, 70–77.
231. O. Pylypenko; A. Rak; T. Durek; S. Kushnir; B. E. Dursina; N. H. Thomae; A. T. Constantinescu; L. Brunsveld; A. Watzke; H. Waldmann; R. S. Goody; K. Alexandrov, *EMBO J.* **2006**, *25*, 13–23.
232. M. Lumbierres; J. Palomo; G. Kragol; S. Roehrs; O. Müller; H. Waldmann, *Chem. Eur. J.* **2005**, *11*, 7405–7415.
233. A. D. de Araujo; J. M. Palomo; J. Cramer; O. Seitz; K. Alexandrov; H. Waldmann, *Chem. A Eur. J.* **2006**, *12*, 6095–6109.

Biographical Sketches

Kristina Görmer studied chemistry at the University of Bayreuth, Germany and at the University of Granada, Spain. Her diploma thesis under the supervision of Professor C. Unverzagt in the Department of Bioorganic Chemistry at the University of Bayreuth, Germany, dealt with the synthesis of glycosyl-triazoles. After an internship at the Institute for Glycomics at the Griffith University of Brisbane, Australia, where she focused on the synthesis of sialic acids, she joined the group of Professor H. Waldmann in Dortmund, Germany as a Ph.D. student in 2007. She is currently working on the synthesis of lipopeptides and their application for biological research.

Herbert Waldmann received his Ph.D. in 1985 from the University of Mainz under the guidance of Professor Kunz in organic chemistry after which he completed a postdoctoral appointment with Professor G. Whitesides at Harvard University. He was appointed as professor of organic chemistry at the University of Bonn (1991), full professor of organic chemistry at the University of Karlsruhe (1993), and director at the MPI of Molecular Physiology Dortmund and professor of organic chemistry at the University of Dortmund (1999). His research interests lie in chemical biology research employing small molecule and protein probes and microarray technology.

Luc Brunsveld was born in 1975 in Culemborg, the Netherlands. He received his Ph.D. in supramolecular chemistry at the Eindhoven University of Technology in 2001 under the guidance of Bert Meijer. Subsequently he was postdoc at the MPI in Dortmund, Germany and group leader at Organon in Oss, the Netherlands. From 2005 to 2008 he was group leader at the Max Planck Institute of Molecular Physiology in Dortmund, Germany. In 2008 he was appointed as full professor of chemical biology at the Eindhoven University of Technology, the Netherlands. His research interests lie in supramolecular chemistry, protein semisynthesis, and chemical biology.

5.18 Genetic Incorporation of Unnatural Amino Acids into Proteins

Angela R. Parrish and Lei Wang, The Salk Institute for Biological Studies, La Jolla, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

5.18.1 Introduction

With few exceptions, the genetic code is universal. Codons in different organisms have identical meanings, specifying the insertion of one of the canonical 20 amino acids or translational termination. Although amino acids different from the canonical 20 have been discovered in some organisms, only two amino acids,

selenocysteine[1] and pyrrolysine,[2] are genetically incorporated into proteins in response to a stop codon in specific mRNA locations. Built by the canonical 20 amino acids, proteins carry out a variety of functions for life needs. Nonetheless, proteins are somewhat limited in the reactions they can participate in alone, requiring cofactors and posttranslational modifications to fulfill many natural functions.

The introduction of new functional groups apart from those found in the common 20 amino acids into proteins can provide specific and malleable tools to researchers for probing protein structure and function, as well as eventually generating novel protein activities. There are a number of methods to introduce unnatural amino acids into proteins, which can be broadly cataloged as chemical approaches and biosynthetic approaches.[3] Chemical approaches rely on chemistry to prepare or modify the protein. For instance, reactive side chains of exposed amino acids in proteins can be chemically modified for derivatization.[4] Complete chemical synthesis has been used to introduce unnatural amino acids into peptides and small proteins.[5] Semisynthetic protein ligation methods, in which two or more protein fragments are chemically ligated to make the full-length protein, further enable unnatural amino acids to be introduced into large proteins.[6] On the other hand, biosynthetic approaches harness the endogenous cellular machinery to translate protein and to introduce unnatural amino acids. For instance, using bacterial strains auxotrophic for a particular amino acid, close structural analogues of this amino acid can globally replace the amino acid in proteins.[7] An *in vitro* biosynthetic method in which a suppressor tRNA is chemically acylated with an unnatural amino acid allows for the site-specific incorporation of unnatural amino acids into proteins in cell extracts supporting translation.[8,9] All of these methods have been proven valuable but are limited by the lack of site selectivity, the *in vitro* nature of the method, or the low incorporation efficiency.

A great challenge is to genetically incorporate unnatural amino acids at any specified site in the proteome with translational fidelity and efficiency parallel to that of common amino acids – in essence, to expand the genetic code to include unnatural amino acids *in vivo*. This chapter will focus on the methodology of genetically encoding unnatural amino acids *in vivo* in different organisms, as well as discuss a number of applications for which unnatural amino acids have been used and could be used in the future.

5.18.2 Methodology

5.18.2.1 Basic Requirements

There are four requirements that must be fulfilled for the genetic incorporation of unnatural amino acids *in vivo* to be both specific and effective (**Figure 1**):[10] (1) an orthogonal tRNA that no endogenous synthetases will charge with natural amino acids and that decodes a unique codon; (2) a unique codon to signal insertion of the unnatural amino acid in the mRNA, such that no endogenous tRNAs will decode the unique codon with a natural amino acid; (3) an orthogonal aminoacyl-tRNA synthetase that is specific for the unnatural amino acid and charges only the orthogonal tRNA; and (4) unnatural amino acid availability in the cytoplasm of the cells of interest, so that the orthogonal synthetase has access to the molecule for acylation. Once these criteria have been met, the orthogonal synthetase will charge the orthogonal tRNA with the desired unnatural amino acid only, and the unnatural amino acid-acylated tRNA will incorporate the unnatural amino acid in response to the unique codon into proteins by cooperating with the endogenous protein translational machinery. The introduced components function in a similar manner as the counterparts that genetically encode common amino acids. This method is particularly powerful because it enables exploitation of novel properties of unnatural amino acids *in vivo* and has the potential to be applied to almost all genetically tractable organisms.

5.18.2.2 Encoding Unnatural Amino Acids in Prokaryotes

Efforts to develop a general method for genetically encoding unnatural amino acids in live cells first focused on *Escherichia coli* because it is easily manipulated genetically and its translational machinery had been extensively studied. *Escherichia coli* has high transformation efficiency for introducing large DNA libraries, and many genetic selections or screens have been successfully applied in this organism to identify distinct phenotypes. These features of *E. coli* facilitated the generation of orthogonal tRNA–codon and synthetase, critical components for incorporating unnatural amino acids *in vivo*.

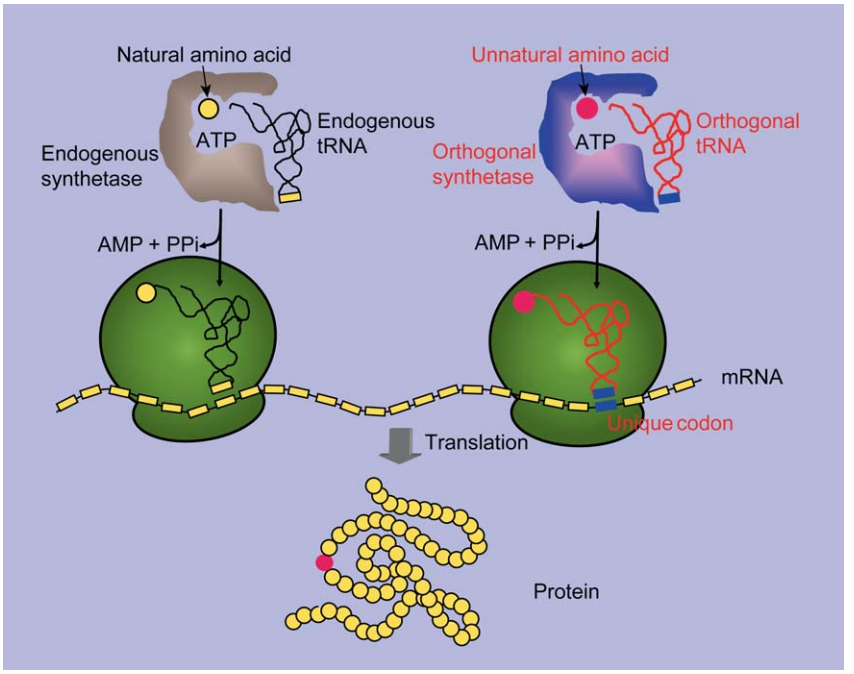


Figure 1 A general method for genetically encoding unnatural amino acids in live cells.

5.18.2.2.1 Generation of orthogonal codon–tRNA–synthetase set

A unique codon in mRNA is required to specify the unnatural amino acid, and an orthogonal tRNA able to recognize this unique codon is needed for decoding. The sense codons in the genetic code cannot be used for this purpose because they already code for an amino acid. Using a 'blank' stop codon to encode the unnatural amino acid is an attractive strategy, because the orthogonal tRNA introducing the unnatural amino acid does not need to compete with an endogenous tRNA. However, the tRNA will still compete with the release factors that terminate protein translation. Initially, the amber stop codon (UAG) was selected to specify the unnatural amino acid. Among the three stop signals, the amber codon is the least used stop codon in *E. coli* and *Saccharomyces cerevisiae*. Amber suppressor tRNAs, tRNAs that can read through the amber stop codon by incorporating a natural amino acid, have been identified in *E. coli*, and these suppressor tRNAs do not significantly affect the growth rates of the host *E. coli* cells.[11] Moreover, tRNAs have also been engineered to suppress the amber codon for amino acid mutagenesis in proteins.[12] To make a tRNA recognize the amber stop codon is also straightforward; the anticodon of the tRNA simply needs to be changed to CUA to be complementary to the amber codon UAG.

There are multiple tRNAs and synthetases inside cells, and the synthetase recognizes its cognate tRNAs through specific interactions at various regions of the tRNA.[13] An orthogonal tRNA for incorporating the unnatural amino acid should not interact with any endogenous synthetases, and an orthogonal synthetase should not interact with any endogenous tRNAs. Meanwhile, both the orthogonal tRNA and the synthetase should be compatible with all other components in the translational machinery. These stringent criteria and the multiple component interactions inside cells make it very challenging to generate an orthogonal tRNA–synthetase pair. Rational engineering of an *E. coli* glutaminyl tRNA–synthetase pair produced an orthogonal tRNA, but no orthogonal synthetase could be generated to selectively recognize this tRNA.[14] A successful strategy is to import a tRNA–synthetase pair from species of a different kingdom of life. This strategy is based on the observation that *in vitro* cross-aminoacylation using tRNA–synthetase pairs from different kingdoms is often low. Attention was focused on archaea as a source of orthogonal tRNA–synthetase pairs for use in *E. coli*. Archaeal aminoacyl-tRNA synthetases are more similar to their eukaryotic than prokaryotic counterparts.[13] Moreover, early work indicated that most tRNAs from the halophile *Halobacterium cutirebrum* cannot be charged by *E. coli* aminoacyl-tRNA synthetases.[15]

The first orthogonal *E. coli* tRNA–synthetase pair generated from archaeal bacteria was derived from the tyrosyl pair taken from *Methanococcus jannaschii*.[16] *In vitro* experiments showed that the major recognition elements of *M. jannaschii* $tRNA^{Tyr}$ include the discriminator base A73 and the first base pair, C1-G72, in the acceptor stem (**Figure 2(a)**). The anticodon triplet participates only weakly in identity determination. By contrast, *E. coli* $tRNA^{Tyr}$ uses A73, G1-C72, a long variable arm, and the anticodon as identity elements. The *M. jannaschii* tyrosyl-tRNA synthetase (*Mj* TyrRS) also has a minimalist anticodon loop binding domain,[17] making it possible to change the anticodon loop of its cognate tRNA to CUA with little loss in affinity by the synthetase. In addition, the TyrRS does not have an editing mechanism, which removes amino acids misacylated onto the cognate $tRNA^{Tyr}$. This lack of the editing function can prevent unnatural amino acid from being deacylated from the orthogonal tRNA. Indeed, an amber suppressor *M. jannaschii* $tRNA^{Tyr}_{CUA}$ $\left(MjtRNA^{Tyr}_{CUA}\right)$ and its cognate *Mj* TyrRS were shown to function efficiently in translation in *E. coli*, but some degree of aminoacylation of this $MjtRNA^{Tyr}_{CUA}$ by endogenous *E. coli* synthetases was observed.[18]

A general strategy was thus developed to evolve orthogonal tRNAs in *E. coli* from heterologous precursors.[18] In this method, a combination of negative and positive selections is applied to a library of tRNA mutants derived from a heterologous suppressor tRNA in the absence and presence of the cognate synthetase, respectively (**Figure 2(b)**). The tRNA library is first introduced into *E. coli* along with a mutant barnase gene in the negative selection. Amber nonsense codons are introduced in the barnase gene at sites permissive to substitution by other amino acids. When a member of the suppressor tRNA library is aminoacylated by any endogenous *E. coli* synthetase (i.e., it is not orthogonal to the *E. coli* synthetases), the amber codons are suppressed and the ribonuclease barnase is produced, resulting in cell death. Only cells harboring orthogonal or nonfunctional tRNAs can survive. All tRNAs from surviving clones are then subjected to a positive selection in the presence of the cognate heterologous synthetase and a β-lactamase gene with an amber codon at a permissive site. For a cell to survive the selection pressure from ampicillin, tRNAs must be good substrates for the cognate heterologous synthetase and function in translation to suppress the amber codon and produce active β-lactamase, which hydrolyzes ampicillin. Therefore, only tRNAs that (1) are not substrates for endogenous *E. coli* synthetases, (2) can be aminoacylated by the heterologous synthetase of interest, and (3) function in translation will survive both selections.

This approach was applied to the $MjtRNA^{Tyr}_{CUA}$ to further reduce recognition of this tRNA by endogenous *E. coli* synthetases, while preserving activity with both the cognate *Mj* TyrRS and translational machinery. Eleven nucleotides of $MjtRNA^{Tyr}_{CUA}$ that do not interact directly with the *Mj* TyrRS were randomly mutated to generate a suppressor tRNA library (**Figure 2(c)**). This tRNA library was passed through rounds of negative and positive selections to afford a functional, orthogonal tRNA ($mutRNA^{Tyr}_{CUA}$) that functions efficiently with *Mj* TyrRS to translate the amber codon.[18]

The orthogonal $mutRNA^{Tyr}_{CUA}$ has high activity for its cognate *Mj* TyrRS, which is critical for the success of changing the substrate specificity of the synthetase to be specific for an unnatural amino acid. Because unnatural amino acid-specific synthetase mutants are generally less active than the wild-type synthetase, these mutants will be difficult to identify if the starting wild-type synthetase–tRNA pair is not highly active. The $mutRNA^{Tyr}_{CUA}$–*Mj* TyrRS pair has been evolved to incorporate more than 40 unnatural amino acids in *E. coli* to date, and is currently the most widely used orthogonal tRNA–synthetase pair.[19] Additional orthogonal tRNA–synthetase pairs have since been generated and include a $tRNA^{Asp}_{CUA}$–AspRS pair derived from yeast[20] and an *E. coli* initiator $tRNA^{fMet}_{CUA}$–yeast TyrRS pair.[21] Orthogonal suppressor tRNAs can also be derived from consensus sequences of multiple archaeal tRNAs and then improved with the above selections.[22] This approach has been used to evolve an orthogonal *Methanococcus thermoautotrophicum* $tRNA^{Leu}_{CUA}$–LeuRS pair,[23] an orthogonal *Methanosarcina mazei* $tRNA^{Glu}_{CUA}$–GluRS pair,[24] and an orthogonal *Pyrococcus horikoshii* $tRNA^{Lys}_{CUA}$–LysRS pair.[22] These pairs hold the potential for further expanding the range of unnatural amino acids to be incorporated, but have yet to be exploited.

5.18.2.2.2 Evolving unnatural amino acid-specific synthetases

Aminoacyl-tRNA synthetases charge the appropriate tRNA with the correct amino acid, which is important in maintaining the fidelity of protein translation. To genetically encode an unnatural amino acid, the substrate specificity of the orthogonal synthetase needs to be altered to charge the orthogonal tRNA with only the desired unnatural amino acid and none of the common 20 amino acids. A general scheme was developed for

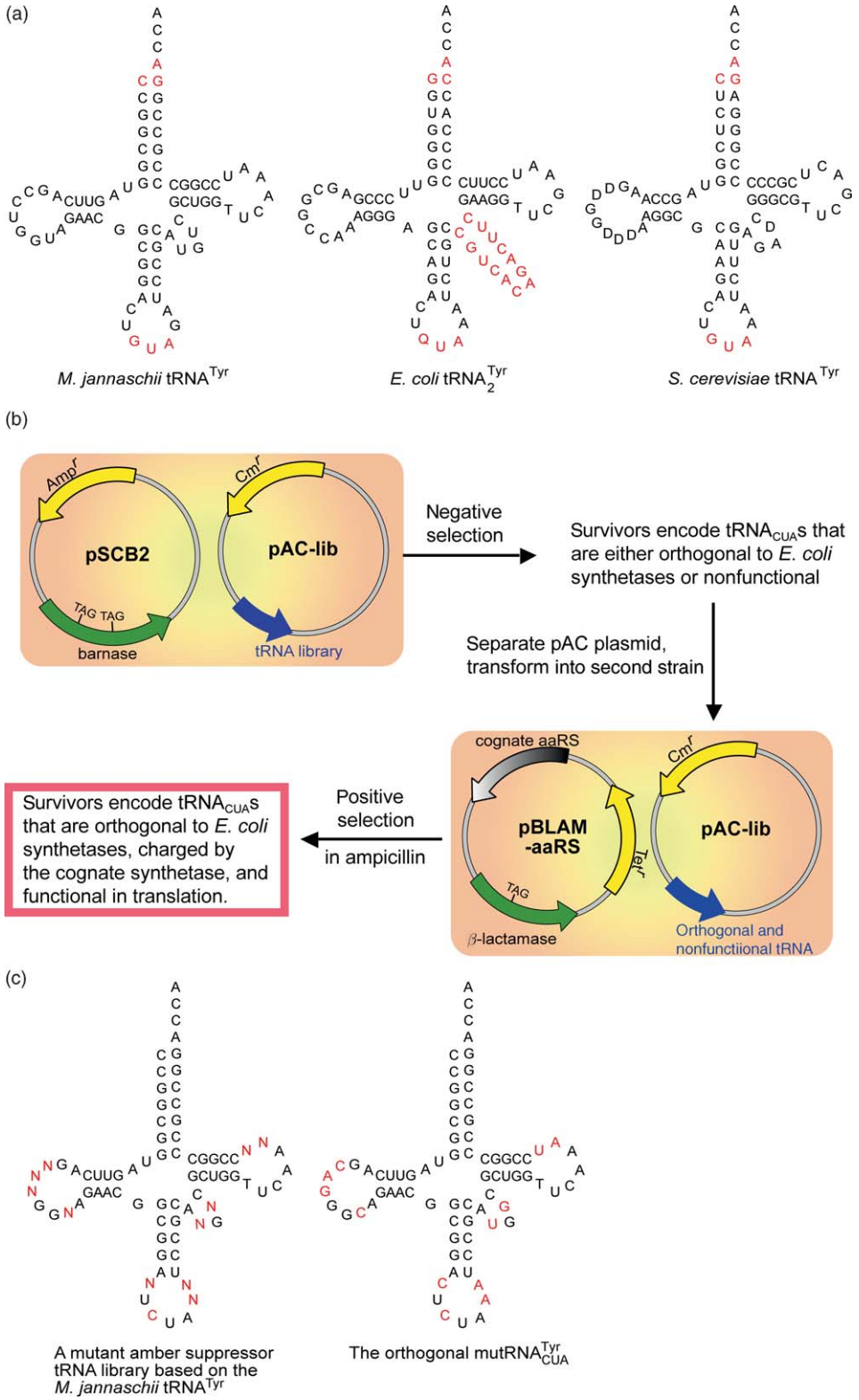


Figure 2 (a) Sequences of $tRNA^{Tyr}$ from three different species. Major identity elements for recognition by the cognate synthetase are highlighted in red. (b) A general method for evolving orthogonal tRNA using negative and positive selections. (c) A library of amber suppressor $tRNA^{Tyr}$ generated by randomizing 11 nucleotides in red (left), and the evolved orthogonal tRNA that has been used to incorporate a variety of unnatural amino acids in *Escherichia coli* (right).

evolving the specificity of aminoacyl-tRNA synthetases that is independent of the structure of the amino acid of interest. In this approach, a library of synthetase active-site mutants is generated and then subjected to a combination of positive and negative selections to evolve synthetases that aminoacylate the tRNA with the unnatural and not endogenous amino acids.[3,25] Active synthetases charging unnatural or natural amino acids are identified using positive selection, and variants charging natural amino acids are subsequently eliminated from the sample pool using negative selection.

By analyzing the X-ray crystal structure of the synthetase (or a homologue) complexed with its cognate amino acid or aminoacyl adenylate, residues in the substrate binding pocket of the synthetase are identified. Codons for these residues are then randomized using degenerate oligonucleotides to make a DNA library, which is transformed and expressed in *E. coli* cells to generate the mutant synthetase library. *Escherichia coli* cells harboring the synthetase library are selected using alternating rounds of positive and negative selections (**Figure 3**). The positive selection is based on resistance to chloramphenicol conferred by suppression of an amber mutation at a permissive site in the chloramphenicol acetyl transferase (CAT) gene; the negative selection uses the barnase gene with amber mutations at permissive sites. When the library of synthetase mutants is passed through the positive selection in the presence of the unnatural amino acid, cells with mutant synthetases that can acylate the orthogonal tRNA with either the unnatural amino acid or an endogenous amino acid will be able to make full-length CAT, which enables the cell to survive and propagate in the antibiotic chloramphenicol. Plasmids encoding active mutant synthetases are then transformed into the negative selection strain, and selections are carried out in the absence of the unnatural amino acid. Cells containing mutant

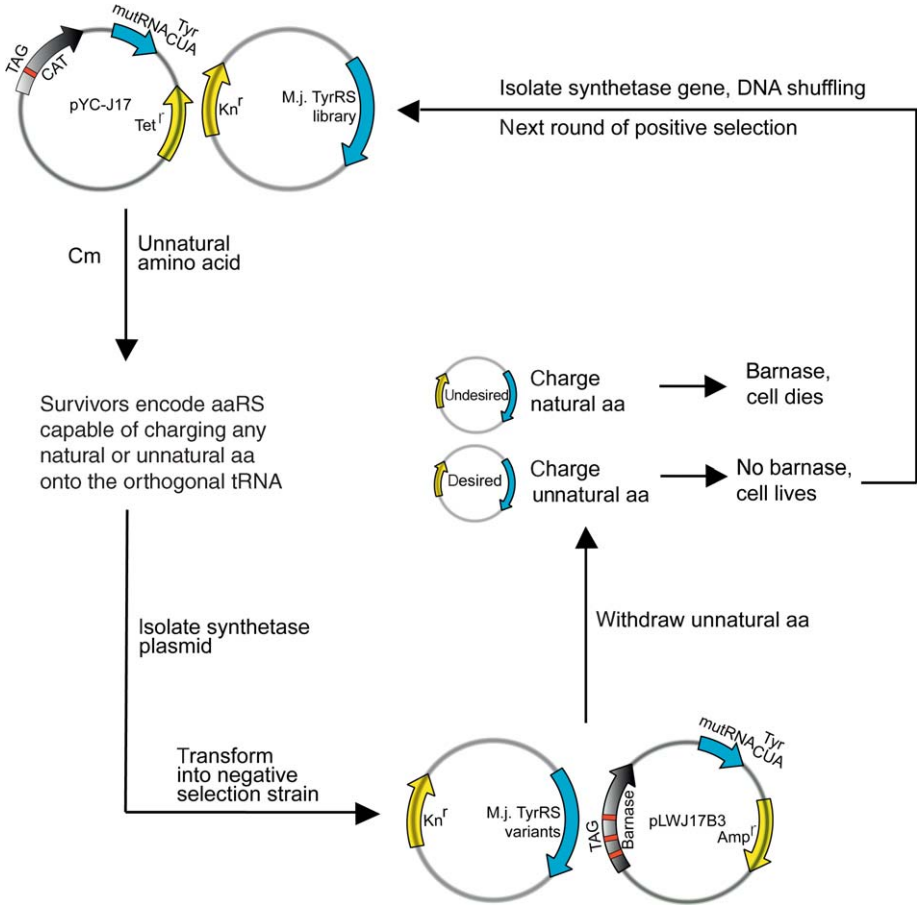


Figure 3 A general positive and negative selection strategy for evolving aminoacyl-tRNA synthetase variants specific for an unnatural amino acid.

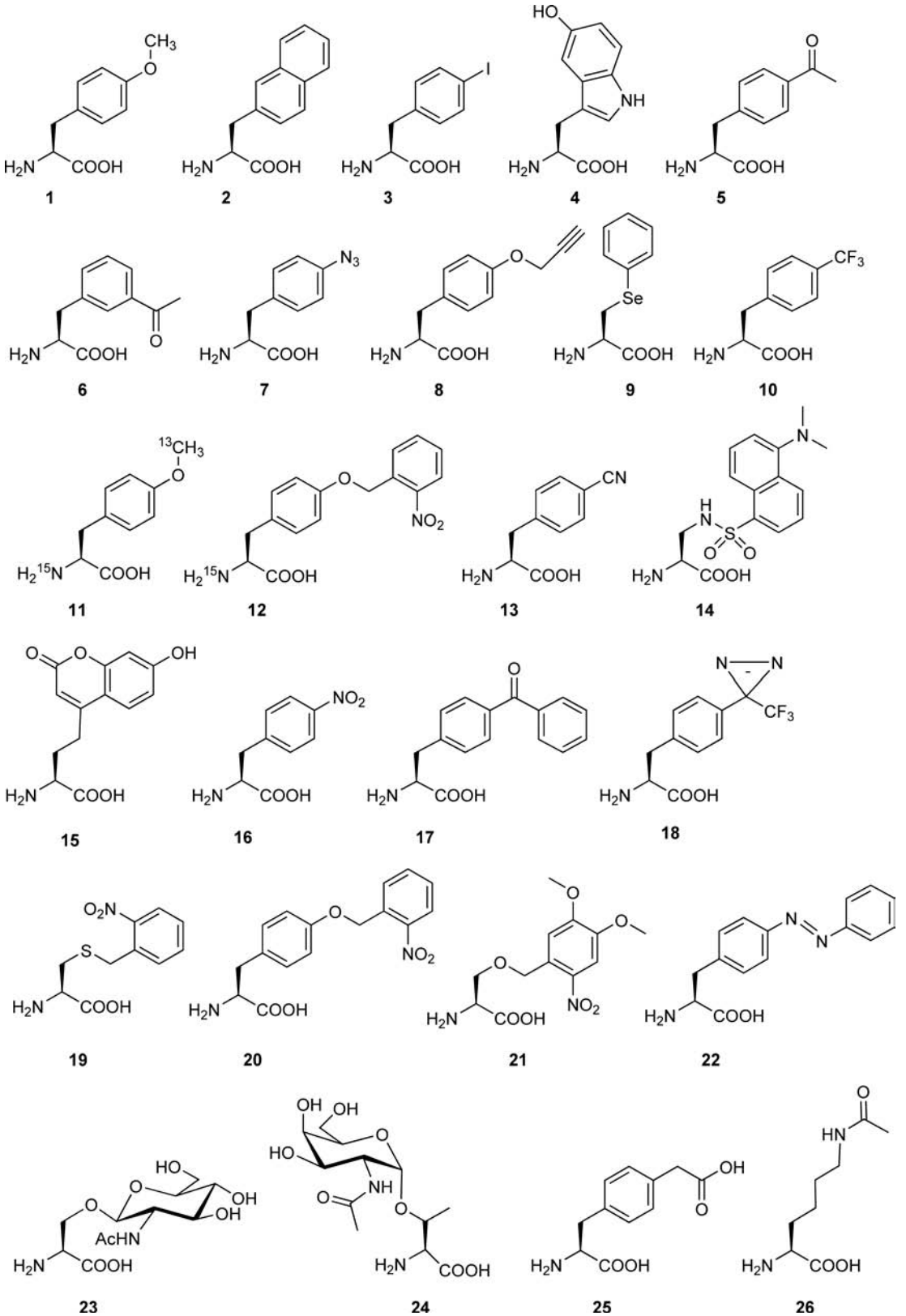
synthetases that recognize endogenous amino acids incorporate the latter in response to the amber codons in the barnase gene, and the produced barnase kills the cell. Therefore, only synthetase variants specific for the unnatural amino acid can survive both selections. For further optimization, more mutations can be introduced into the synthetase gene of the initial hits using random mutagenesis or DNA shuffling. Repeated rounds of positive and negative selections can be carried out with increased selection pressure (increasing the concentration of chloramphenicol in the positive selection and decreasing the number of amber codons in the barnase gene in the negative selection), until mutant synthetases that can specifically incorporate the unnatural amino acid in response to the amber codon are isolated.

An alternative selection scheme makes use of an amber-T7-GFP (green fluorescent protein) instead of an amber-barnase reporter in the negative selection.[26] Suppression of amber codons introduced at permissive sites in T7 RNA polymerase (Pol) produces full-length T7 RNA Pol, which drives the expression of GFP under the control of the T7 promoter. In this approach, both the amber-CAT reporter and amber-T7-GFP reporter are encoded in a single plasmid. After positive selection in chloramphenicol, surviving cells are grown in the absence of both the unnatural amino acid and chloramphenicol. Cells containing mutant synthetases that can acylate the tRNA with any of the 20 common amino acids express GFP, whereas cells containing mutant synthetases that can acylate the tRNA with only the unnatural amino acid do not. These nonfluorescent cells are separated from the fluorescent cells using fluorescence activated cell sorting. One advantage of this latter method is that both reporters are contained within a single genetic construct, eliminating the need for plasmid shuttling between positive and negative selections. However, fluorescence intensity is not as clear-cut as survival/death of bacteria, and effective sorting of tiny bacteria cells is technically challenging. Other selection schemes have also been pursued, including cell surface and phage display systems, but these are less general (i.e., require capture reagents specific for the amino acid of interest) or not as efficient.[27] The CAT-barnase selection strategy has been most efficient in generating a large number of unnatural amino acid-specific synthetases.

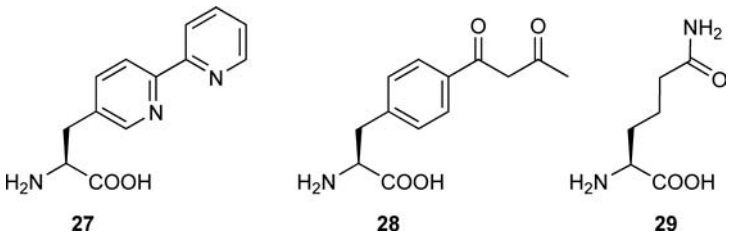
5.18.2.2.3 Expansion of the genetic code

The first unnatural amino acid genetically encoded into proteins in *E. coli* is *O*-methyl-L-tyrosine (OMeTyr, **1**) (**Scheme 1**).[25] A library of synthetase mutants (10^9 in size) was generated by randomizing the five active-site residues (Tyr32, Glu107, Asp158, Ile159, and Leu162) in the *Mj* TyrRS based on the crystal structure of the homologous *Bacillus stearothermophilus* TyrRS–tyrosyl adenylate complex (**Figure 4**). This library was then subjected to two rounds of positive selection and negative screening to afford clones that survived at high chloramphenicol concentrations in the presence of OMeTyr and at low chloramphenicol concentrations in its absence. To verify the incorporation of OmeTyr in response to the amber stop codon by the evolved mutRNA$^{Tyr}_{CUA}$–mutant TyrRS pair, the third codon of the *E. coli* dihydrofolate reductase (DHFR) gene was mutated to TAG and a His6 tag was added to the C-terminus to facilitate protein purification and separation from endogenous DHFR. Full-length DHFR was produced only when the mutant TyrRS, mutRNA$^{Tyr}_{CUA}$, and OMeTyr were each present. In the absence of any one component, no DHFR protein could be detected by silver stain or Western blot analysis. Insertion of OMeTyr in response to the TAG codon was confirmed by mass spectrometric analysis of both the intact protein and tryptic fragments. No incorporation of tyrosine or other amino acids at the TAG position was observed, and OMeTyr was incorporated only in response to TAG and not any other sites in DHFR. These results demonstrated that one could rationally engineer bacteria that genetically encode an unnatural amino acid with high efficiency and translational fidelity.

OMeTyr is structurally similar to tyrosine and phenylalanine and provided an excellent case to demonstrate the high translational fidelity that can be achieved by this approach. Meanwhile, it was important to test whether the method could be used to genetically encode unnatural amino acids that significantly deviate from common amino acids in structure. Efforts were then attempted to genetically encode the second unnatural amino acid, L-3-(naphthyl)-alanine (**2**), which represents a significant structural perturbation relative to tyrosine.[28] A slightly different mutant *Mj* TyrRS library (in which Tyr32, Asp158, Ile159, Leu162, and Ala167 were randomized) was constructed, from which four mutant synthetases were identified that have high activity for this unnatural amino acid. One round of DNA shuffling of these four synthetase genes followed by selection resulted in a mutant synthetase (SS12) with enhanced activity for L-3-(naphthyl)alanine and



Scheme 1 (Continued)



Scheme 1

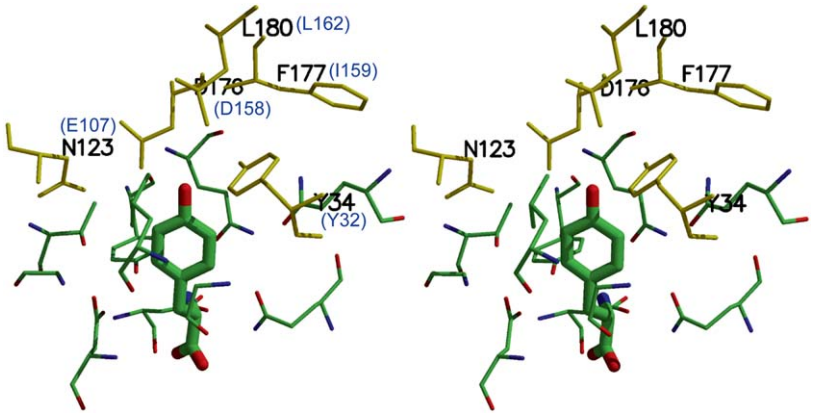


Figure 4 A library of *Mj* TyrRS mutants was generated by randomizing five residues in the tyrosine binding site. The crystal structure of *Bacillus stearothermophilus* TyrRS is shown with residues labeled in black. Corresponding residues in the *Methanococcus jannaschii* TyrRS are labeled in blue. Reproduced from L. Wang; A. Brock; B. Herberich; P. G. Schultz, *Science* **2001**, *292*, 498–500, with permission from AAAS.

greatly reduced the activity for endogenous amino acids. Protein expression and mass spectrometric analysis confirmed that the mutRNA$^{Tyr}_{CUA}$/SS12 mutant synthetase selectively incorporates L-3-(naphthyl)alanine into proteins with an efficiency and fidelity rivaling that of the common 20 amino acids. This result suggested that the above methodology may be applicable to a large variety of unnatural amino acids.

Using the above methodology, more than 40 unnatural amino acids have been genetically incorporated into proteins in *E. coli*.[19] In general, for most unnatural amino acids, suppression efficiencies range from 25 to 75% of wild-type protein and translational fidelity is >99%. Yields of unnatural amino acid-containing proteins are in the range of several milligrams to tens of milligrams per liter of cell culture. In optimized expression systems, the yield could reach about $1\,g\,l^{-1}$, sufficient for large-scale protein production. Other than structural analogues of natural amino acids, which can probe very fine aspects of amino acid structure as it relates to protein function, many unnatural amino acids have very interesting chemical properties that make them particularly useful for biological research. Applications of representative unnatural amino acids will be discussed in Section 5.18.3.

5.18.2.3 Encoding Unnatural Amino Acids in Eukaryotes

The general strategy developed to genetically encode unnatural amino acids in bacteria should also be applicable to higher organisms. Orthogonal tRNA–codon–synthetase sets have to be generated for each organism of interest by following the same principle.

5.18.2.3.1 Technical issues and solutions of the transition to yeast

A number of orthogonal tRNA–synthetase pairs have been generated for use in eukaryotic organisms, and most of them are derived from bacterial pairs, again due to the low cross-kingdom aminoacylation. For instance, it

has been shown that the *E. coli* tyrosyl amber suppressor ($tRNA^{Tyr}_{CUA}$) is not aminoacylated by eukaryotic synthetases and acts as an amber suppressor in *S. cerevisiae* in the presence of *E. coli* TyrRS.[29,30] In addition, *E. coli* TyrRS does not aminoacylate yeast tRNAs.[15] Thus, *E. coli* $tRNA^{Tyr}_{CUA}$–TyrRS functions as an orthogonal pair in yeast. A human initiator tRNA-derived amber suppressor and *E. coli* GlnRS form another orthogonal pair for use in yeast.[21] In addition, the *E. coli* $tRNA^{Gln}_{CUA}$–GlnRS pair has also been used in mammalian cells for efficient suppression of the amber codon.[31] To selectively introduce an unnatural amino acid into proteins in eukaryotes, Yokoyama and coworkers[32] screened a small collection of designed active-site variants of *E. coli* TyrRS in a wheat germ translation system and discovered a mutant synthetase that utilizes 3-iodotyrosine (**3**) more effectively than tyrosine. This mutant synthetase was used with the *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$ to incorporate 3-iodotyrosine into proteins in mammalian cells.[33]

To set up a general selection scheme in yeast analogous to that used in *E. coli* for evolving synthetases specific for unnatural amino acids, two amber stop codons were introduced into the permissive sites of the transcriptional activator protein GAL4.[34] The GAL4(2TAG) mutant was expressed in yeast MaV203, a commercially available yeast strain that has the endogenous GAL4 deleted and contains three GAL4-inducible reporter genes, HIS3, URA3, and LACZ. Suppression of these amber codons leads to the production of full-length GAL4, which drives transcription of the reporters. Expression of HIS3 and URA3 complements the histidine and uracil auxotrophy in this strain and provides a positive selection for clones expressing active synthetase mutants. On the contrary, addition of 5-fluoroorotic acid (5-FOA), which is converted into a toxic product by URA3, results in the death of cells expressing active synthetases. In the absence of the unnatural amino acid, this serves as a negative selection to remove synthetases specific for endogenous amino acids. Like GFP, the lacZ reporter can serve as another marker for colorimetrically distinguishing active synthetases from inactive ones.

This selection scheme was used to evolve the orthogonal *E. coli* $tRNA^{Tyr}_{CUA}$–TyrRS pair in yeast.[34] A synthetase library (10^8 in size) was similarly constructed by randomizing five active-site residues in *E. coli* TyrRS corresponding to the five residues randomized in the *Mj* TyrRS. Mutant synthetases were identified after several rounds of positive and negative selection that incorporate a number of unnatural amino acids into proteins, albeit with rather low protein yields (about 0.05 mg l^{-1}). A similar approach has been used to evolve orthogonal *E. coli* leucyl $tRNA_{CUA}$–LeuRS pairs that selectively incorporate photochromic and fluorescent amino acids into proteins in yeast.[35,36]

Low incorporation efficiency creates difficulties in making use of the genetically encoded unnatural amino acids in yeast. To address this problem, new methods have been developed to efficiently express orthogonal prokaryotic tRNA and to improve the target mRNA stability in yeast. Prokaryotes and eukaryotes differ significantly in tRNA transcription and processing (**Figure 5(a)**). *Escherichia coli* tRNAs are transcribed by the sole RNA Pol through promoters upstream of the tRNA gene. However, the transcription of eukaryotic tRNAs by Pol III depends principally on promoter elements within the tRNA known as the A- and B-box.[37] The A- and B-box identity elements are conserved among eukaryotic tRNAs but are lacking in many *E. coli* tRNAs. Creating the consensus A- and B-box sequences in *E. coli* tRNAs through mutation could cripple the tRNA.[33] In addition, all *E. coli* tRNA genes encode full tRNA sequences, whereas eukaryotic tRNAs have the 3′-CCA trinucleotide enzymatically added after transcription.[38] The *E. coli* $tRNA^{Tyr}_{CUA}$ has the B-box but no fully matched A-box. In the previously described system,[34] only the structural gene of *E. coli* $tRNA^{Tyr}_{CUA}$ was inserted in the plasmid. The resultant basal level of tRNA expression is likely driven by a cryptic promoter elsewhere on the plasmid and should contribute to the low incorporation efficiency of unnatural amino acids.

One method to increase the tRNA expression uses the 5′- and 3′-flanking sequences of an endogenous yeast suppressor $tRNA^{Tyr}_{CUA}$ (SUP4) (**Figure 5(b)**).[39] The *E. coli* $tRNA^{Tyr}_{CUA}$ without the 3′-CCA trinucleotide was flanked by these sequences, and this arrangement was repeated 3–6 times in tandem. A strong phosphoglycerate kinase 1 (PGK1) RNA Pol II promoter was additionally placed upstream of the tandem arrangements. This combination gave an overall >50-fold increase in tRNA levels compared to the previous tRNA-alone scheme. Together with the optimized expression of synthetase, this approach improved the yield of unnatural amino acid-containing proteins to 6–8 mg l^{-1} of culture.

A new method to efficiently express prokaryotic tRNAs in yeast involves an external Pol III promoter containing the consensus A- and B-box sequences (**Figure 5(c)**).[40] When placed upstream of the *E. coli* tRNA (without the 3′-CCA trinucleotide), the promoter drives transcription of a primary RNA transcript consisting

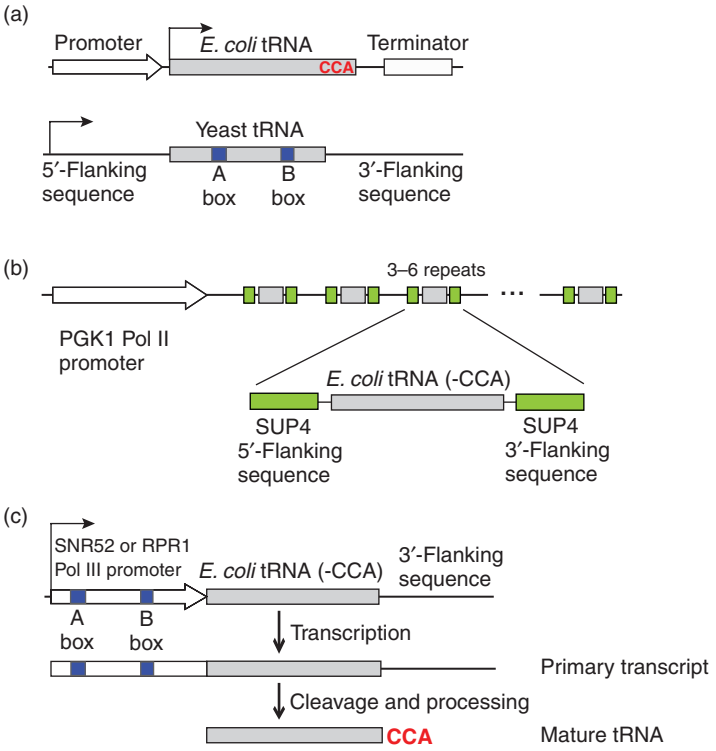


Figure 5 (a) Gene elements for tRNA transcription in *Escherichia coli* and yeast. (b) Increasing tRNA expression in yeast. A tandemly repeated (3–6 times) tRNA cassette containing 5′ and 3′ yeast tRNA flanking sequences around an *Escherichia coli* tRNA and driven by an upstream RNA Pol II promoter is shown. (c) A general method for expressing prokaryotic tRNAs in yeast using an external Pol III promoter that contains the consensus A- and B-box sequences and that is cleaved from the primary transcript.

of the promoter and the tRNA. The promoter is then cleaved posttranscriptionally to yield the mature tRNA. Two yeast Pol III promoters, the SNR52 promoter and the RPR1 promoter, have been shown to drive the expression of the *E. coli* $tRNA^{Tyr}_{CUA}$ efficiently in yeast.[40] The expressed *E. coli* $tRNA^{Tyr}_{CUA}$ are six- to ninefold more active in protein translation than the tRNA transcribed using the SUP4 5′-flanking sequence. The increased activity is not through an increase in tRNA transcription level based on Northern blot results, underscoring the importance of proper tRNA processing. The cleavage of the SNR52 or RPR1 promoter from the primary RNA transcript could directly generate the correct 5′ end of the tRNA. When a 5′-flanking sequence is used, RNaseP is required to generate the 5′ tRNA end. It may be difficult for yeast RNaseP to process prokaryotic tRNA efficiently. This external promoter method has also been used to express *E. coli* $tRNA^{Leu}_{CUA}$ in yeast, and the resultant $tRNA^{Leu}_{CUA}$ functions efficiently in protein translation as well, suggesting that it can be a general method for functionally expressing different prokaryotic tRNAs in yeast.

The stability of mRNA is an important issue for unnatural amino acid incorporation in eukaryotes. Eukaryotic cells have an mRNA surveillance mechanism, nonsense-mediated mRNA decay (NMD), to identify mRNA with premature stop codons and target the mRNA for rapid degradation. This mechanism is intended to protect cells from generating undesired truncated proteins. When stop codons are used to encode unnatural amino acids, NMD could result in a shorter lifetime for the target mRNA and thus a lower protein yield. Inactivation of NMD would preserve the stability of the UAG-containing target mRNA and thus enhance the incorporation efficiency of unnatural amino acids. An NMD-deficient yeast strain was generated by knocking out the UPF1 gene, an essential component for NMD in yeast. The unnatural amino acid incorporation efficiency was indeed increased more than twofold in the upf1Δ strain compared to the wild-type yeast.[40] NMD in yeast shows a polar effect of nonsense codon position.[41,42] The reduction in steady-state

mRNA levels is larger when the nonsense codon is closer to the 5′ end than to the 3′ end. The increase of unnatural amino acid incorporation efficiency in the upf1Δ strain similarly correlates with the position of the UAG codon, with >2-fold increase when the amber codon is within the N-terminal two-thirds of the gene and no significant increase when within the C-terminal fourth of the coding region (Q. Wang and L. Wang, unpublished results).

By using the external SNR52 promoter and the NMD-deficient upf1Δ strain, the overall purified yields of unnatural amino acid-containing proteins have reached ~15 mg l^{-1} of yeast cells, ~300-fold higher than the previous system and comparable to the yield in *E. coli*.

5.18.2.3.2 Incorporating unnatural amino acids in mammalian cells

Genetic incorporation of unnatural amino acids into proteins in mammalian cells is a big leap, as mammalian cells are biologically complex, and therefore all of the hurdles to incorporation in yeast will be amplified. Specifically, one challenge is the efficient expression of prokaryotic orthogonal tRNAs that are functional in translation in mammalian cells, again due to the aforementioned difference in tRNA transcription and processing between prokaryotes and eukaryotes. In addition, the transfection efficiency of mammalian cells is much lower than the transformation efficiency of *E. coli* and yeast, making it impractical to generate large synthetase mutant libraries inside mammalian cells. Survival–death selection in mammalian cells is also not as efficient as in *E. coli* and yeast. These factors lead to a second challenge: how to evolve mutant orthogonal synthetases to be specific for unnatural amino acids for use in mammalian cells.

Initial attempts at introducing unnatural amino acids into proteins in mammalian cells involved the transfection of an amber suppressor tRNA that is chemically acylated with an unnatural amino acid *in vitro*.[43,44] This approach limits the amount of overall protein that can be produced since the acylated suppressor tRNA is consumed stoichiometrically and cannot be regenerated inside cells. It is also technically demanding to prepare the chemically acylated tRNA.

Since the *E. coli* $tRNA^{Tyr}_{CUA}$–TyrRS pair is orthogonal in mammalian cells, it should be possible to evolve this pair to incorporate unnatural amino acids in mammalian cells. However, the *E. coli* $tRNA^{Tyr}_{CUA}$ does not express well in mammalian cells, presumably due to the lack of a matched A-box inside the tRNA structural gene. Another bacterial $tRNA^{Tyr}_{CUA}$ derived from *B. stearothermophilus* happens to contain the consensus A- and B-box sequences in eukaryotic tRNAs. This tRNA is not aminoacylated by any endogenous synthetase in mammalian cells, and functions with the *E. coli* TyrRS for suppressing the TAG codon with tyrosine. The *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$ (lacking the 3′-CCA) was linked to the 5′-flanking sequence of the human $tRNA^{Tyr}$, and nine tandem repeats of this gene arrangement were used for tRNA expression.[33] A small collection of designed active-site variants of the *E. coli* TyrRS was screened using *in vitro* biochemical assays, and a mutant synthetase that uses 3-iodotyrosine more effectively than Tyr was identified. This mutant synthetase was used with nine tandem *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$ to incorporate 3-iodotyrosine (**3**) into proteins in Chinese hamster ovary (CHO) cells and HEK293 cells with approximately 95% fidelity. In another attempt, mutations were introduced into the A-box region of the *Bacillus subtilis* tryptophan opal suppressor tRNA $\left(tRNA^{Trp}_{CUA}\right)$ to create a consensus A-box.[45] The mutant tRNA (lacking the 3′-CCA) was expressed using the 5′- and 3′-flanking sequences from the *Arabidopsis* $tRNA^{Trp}$. Together with a rationally designed *B. subtilis* TrpRS mutant, this tRNA incorporated 5-hydroxytryptophan (**4**) into the foldon protein in HEK293 cells in response to the UGA opal codon with ~97% fidelity.

The above methods may not be generally applied to other tRNA–synthetase pairs and various unnatural amino acids. Most bacterial tRNAs do not have the consensus A- and B-box sequences, and heterologous pairing of the tRNA and synthetase often results in low activity. Mutations to create such consensus sequences in tRNA, although they did not dramatically decrease the activity of the *B. subtilis* $tRNA^{Trp}_{UCA}$,[45] greatly impaired the suppression ability of *E. coli* $tRNA^{Tyr}_{CUA}$,[33] since the A- and B-box lie at nucleotides involved in tertiary interactions that support the L-shaped structure of the tRNA. For generating synthetase variants that are specific for unnatural amino acids, it is difficult to predict *a priori* which active-site residues need to be mutated. As mutations are often required at multiple sites to achieve high substrate specificity, small collections of synthetase mutants will likely fall short. Mutants thus generated may still recognize common amino acids, as is the case with the synthetase used to incorporate 3-iodotyrosine.[33] Therefore, it remained a challenge to incorporate different unnatural amino acids in mammalian cells with high efficiency and fidelity.

For efficient expression of functional prokaryotic tRNAs in mammalian cells, a novel method was developed that involves the use of an external type-3 Pol III promoter.[46] A type-3 Pol III promoter does not require intragenic elements (such as the A- and B-box) for transcription; its promoter sequence resides exclusively upstream of the coding sequence. Thus tRNAs with or without the consensus A- and B-box sequences can be transcribed by this type of promoter in mammalian cells. In addition, the transcription initiation site of some type-3 Pol III promoters, such as the H1 promoter[47] and the U6 small nuclear RNA promoter,[48] is well defined, which could be used to generate the correct 5′ end of the tRNA without further posttranscriptional processing. A fluorescence-based translation assay was set up to systematically identify expression elements that are required to efficiently drive the transcription of *E. coli* tRNAs and to generate tRNAs functional in protein translation in mammalian cells. The H1 promoter, which drives the expression of human H1 RNA (the RNA component of the human nuclear RNase P), the 3′-CCA trinucleotide, and the 5′- and 3′-flanking sequence from the human $tRNA^{fMet}$ are linked to the *E. coli* $tRNA^{Tyr}_{CUA}$ in different combinations for tRNA expression (**Figure 6(a)**). The gene arrangement that yields tRNAs with the highest translational activity is the H1 promoter, followed by the *E. coli* $tRNA^{Tyr}_{CUA}$ (without 3′-CCA) and by the 3′-flanking sequence of $tRNA^{fMet}$. The *E. coli* $tRNA^{Tyr}_{CUA}$ expressed using this method is ~70-fold more active in translation than the one expressed with the 5′- and 3′-flanking sequences of human $tRNA^{fMet}$ (**Figure 6(b)**). This method has also been used to express different *E. coli* tRNAs successfully, and it is effective in various mammalian cells such as HEK293, HeLa, and primary neurons.

A transfer strategy was developed to address the challenge of evolving unnatural amino acid-specific synthetases directly in mammalian cells. Since the *E. coli* $tRNA^{Tyr}_{CUA}$–TyrRS pair is orthogonal in mammalian cells (in addition to yeast), and the translational machinery of yeast is homologous to that of higher eukaryotes, it should be possible to evolve the specificity of this synthetase in yeast (which is well suited for genetic selections with large libraries) and transfer the optimized tRNA–synthetase pairs directly to mammalian cells. In one report, mutant synthetases that were evolved in yeast from the *E. coli* TyrRS specific for different unnatural amino acids, together with the *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$, have been successfully used in CHO cells and HEK293 cells for incorporating the corresponding unnatural amino acid.[49] Mass spectrometry confirmed a high translational fidelity for the unnatural amino acid, and proteins could be produced with efficiencies up to 1 μg per 2×10^7 cells. In another report, mutant synthetases evolved in yeast from the *E. coli* TyrRS were used together with the cognate *E. coli* $tRNA^{Tyr}_{CUA}$ expressed by type-3 Pol III H1 promoter in mammalian cells.[46] Because the H1 promoter enabled the functional expression of the *E. coli* $tRNA^{Leu}_{CUA}$, mutant

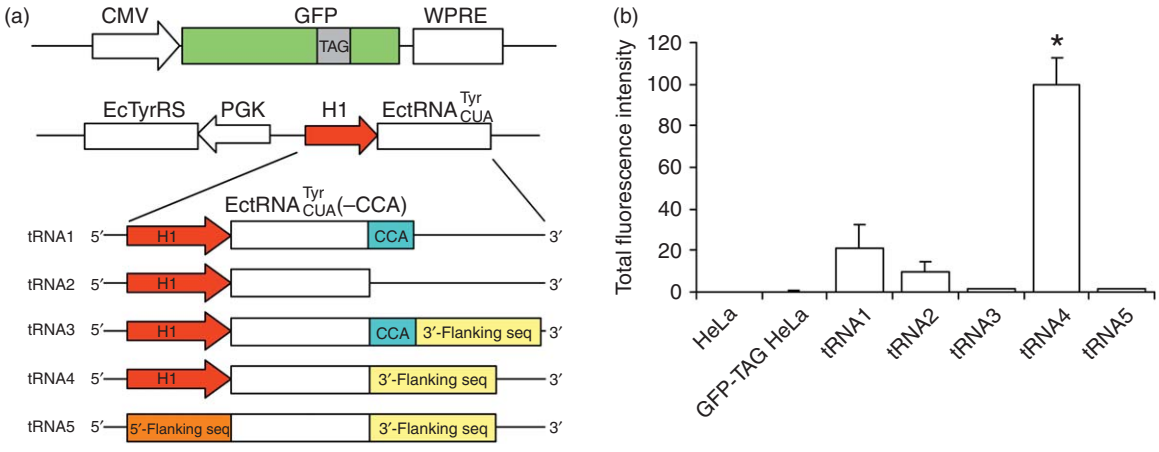


Figure 6 A general method for expressing prokaryotic tRNAs in mammalian cells. (a) A type-3 Pol III promoter, the H1 promoter, was combined with other gene elements in different ways to express the *Escherichia coli* amber suppressor $tRNA^{Tyr}$. The GFP gene containing the TAG mutation at a permissive site serves as a fluorescence reporter for amber suppression. Translation of full-length GFP generates green fluorescence and indicates the expression of functional *Escherichia coli* amber suppressor tRNA in cells. (b) Total fluorescence intensity of cells transfected with constructs shown in (a). The combination in tRNA4 yields the most efficient tRNA expression.

synthetases evolved in yeast from the *E. coli* LeuRS were also successfully transferred in mammalian cells to incorporate the fluorescent unnatural amino acid dansylalanine (**14**). The incorporation efficiencies of the unnatural amino acids were in the range of 13–41%, depending on the activity of the evolved synthetase, and the incorporation fidelity was also high judging by a sensitive fluorescence assay. Unnatural amino acids have been incorporated in HEK293 cells, HeLa cells, and even in primary neurons. Finally, unnatural amino acids were then incorporated into a voltage-sensitive K^+ channel, Kv1.4, in HEK293 cells, which allowed for a mechanistic analysis of the fast inactivation of the ion channel.

5.18.2.4 Further Considerations

5.18.2.4.1 Additional codons for encoding unnatural amino acids

The total number of unnatural amino acids that can be encoded in any organism is limited by the number of noncoding codons. It should be possible to use quadruplet codons and cognate suppressor tRNAs with expanded anticodon loops to encode additional amino acids. There are many examples of naturally occurring +1 frameshift suppressors including UAGN (N = A, G, C, or T) suppressors derived from Su7, which encodes glutamine,[50] *suf7*-derived suppressors of ACCN codons encoding threonine,[51] and CAAA suppressors derived from $tRNA^{Lys}$ and $tRNA^{Gln}$.[52] Moreover, genetic selections have been used to identify efficient four- and five-base codon suppressor tRNAs from libraries of mutant tRNAs.[53,54] Frameshift suppressor tRNAs can efficiently incorporate a number of the common 20 amino acids into proteins *in vivo*,[55] and chemically aminoacylated frameshift suppressors have also been used to incorporate unnatural amino acids into proteins using *in vitro* translation systems with four- and five-base codons.[56,57] An orthogonal four-base suppressor tRNA–synthetase pair was generated from the $tRNA^{Lys}$–LysRS of archaebacteria *P. horikoshii*, which efficiently incorporates the unnatural amino acid homoglutamine (**29**) into proteins in *E. coli* in response to the quadruplet codon AGGA.[22] Frameshift suppression with homoglutamine does not significantly affect protein yields or cell growth rates, and is mutually orthogonal with amber suppression. This approach allows the simultaneous incorporation of two unnatural amino acids at distinct sites within proteins. The orthogonal AGGA-specific $tRNA^{Lys}_{UCCU}$ and the UAG-specific $mutRNA^{Tyr}_{CUA}$ were expressed with a mutant myoglobin gene containing the Gly24(AGGA) and Ala75(TAG) mutations. In the presence of both of the corresponding orthogonal homoglutamine-specific and OMeTyr-specific synthetases and the two unnatural amino acids, 1.7 mg l^{-1} of mutant myoglobin was produced with an overall suppression efficiency of ~25%. Electrospray mass spectrometric analysis of the full-length protein confirmed that myoglobin contained both unnatural amino acids.

Another solution for generating unique codons is to develop additional unnatural base pairs, that is, to expand the existing genetic alphabet. One extra base pair would increase the number of triplet codons from 64 to 125. Essential requirements for third base pair candidates include stable and selective base pairing, efficient enzymatic incorporation into DNA with high fidelity by a Pol, and the efficient continued primer extension after synthesis of the nascent unnatural base pair. For *in vivo* usage, the unnatural nucleoside must be membrane permeable and be phosphorylated to form the corresponding triphosphate. In addition, the increased genetic information must be stable and not destroyed by cellular enzymes. Previous efforts took advantage of hydrogen bonding patterns that are different from those in canonical Watson–Crick pairs, the most noteworthy example of which is the *iso*-C:*iso*-G pair.[58–60] These bases generally mispair to some degree with natural bases and cannot be enzymatically replicated. Hydrophobic packing interactions between bases have emerged as a promising strategy that can replace hydrogen bonding to drive the formation of base pair, as well as disfavoring mispairing with the natural nucleobases.[61,62] Hydrophobic interactions are also sufficient for the enzymatic synthesis of an unnatural base pair by incorporation of an unnatural nucleoside triphosphate using an unnatural nucleotide as the template,[63–66] but extension beyond the unnatural hydrophobic base pair tends to be inefficient to date. In an effort to develop an unnatural base pair satisfying all the above requirements, a series of unnatural hydrophobic bases have been systematically synthesized and studied, from which fundamental principles underlying the genetic alphabet are being revealed and promising base pair candidates may be developed in the near future.[67]

As our ability to synthesize large DNA molecules keeps improving, it may be possible to generate a synthetic *E. coli* variant in which rare or degenerate codon–tRNA pairs are eliminated from the wild-type genome by replacing them with sense codons specifying the same amino acid. The liberated codons can instead be used to

encode unnatural amino acids. All these possibilities suggest that neither the number of available triplet codons nor the translational machinery itself represents a significant barrier to further expansion of the code.

5.18.2.4.2 Intracellular availability of unnatural amino acids

Unnatural amino acids are added to the growth medium in most experiments. There are a large number of amino acid and amine transporters that are relatively nonspecific and which may help to transport the unnatural amino acids into cells. From measurements of cytoplasmic levels of amino acids, it is found that a large number of unnatural amino acids are efficiently transported to the *E. coli* cytoplasm in millimolar concentrations. Highly charged or hydrophilic amino acids may require derivatization (e.g., esterification, acylation) with groups that are hydrolyzed in the cytoplasm. Metabolically labile amino acids or analogues (e.g., α-hydroxy acids, *N*-methyl amino acids) may require strains in which specific metabolic enzymes are deleted.

An alternative to adding exogenous amino acids to the growth media involves engineering a pathway for the biosynthesis of the unnatural amino acid directly in the host organism. For example, genes for the biosynthesis of *p*-amino-L-phenylalanine (*p*AF) from simple carbon sources can be heterologously expressed in *E. coli*.[68] *p*AF was biosynthesized from the metabolic intermediate chorismic acid using the *papA*, *papB*, and *papC* genes from *Streptomyces venezuelae* in combination with a nonspecific *E. coli* transaminase (**Figure 7**). *Escherichia coli* containing these genes produced *p*AF at levels comparable to those of other aromatic amino acids and had normal growth rates. In the presence of a *p*AF-specific, orthogonal mutRNA$^{Tyr}_{CUA}$–synthetase pair, *E. coli* transformed with *papA–C* produced mutant proteins containing *p*AF at sites encoded by the amber codon with excellent yield and fidelity. In addition to *p*AF, it should be possible to biosynthesize and genetically encode other amino acids *in vivo* as well.

5.18.3 Applications

Unnatural amino acids have been used to study various biological problems involving proteins. As virtually every cellular process involves proteins, this technology has very broad applicability. This section will focus on applications of unnatural amino acids that are genetically encoded using the above methodology. For applications of unnatural amino acids introduced using other methods, readers are suggested to refer to the many excellent reviews written on the subject.[3,69–74] Given the wide variety of functional groups that have been incorporated into proteins using genetically encoded unnatural amino acids, this section has been split into several subsections. Each subsection will group unnatural amino acids based on their desirable properties or on common problems that they can address. As the selected examples below illustrate, the novel chemical, physical, and biological properties embodied by unnatural amino acids could benefit research in a diverse set of disciplines.

5.18.3.1 Chemical Handles for Protein Labeling and Modification

Cysteine is the most frequently used residue for selective chemical modification of proteins due to its relatively low abundance in proteins and the increased nucleophilicity of the thiol group relative to other natural amino acid side chains. The intrinsic selectivity is low unless no cysteine is present or unless all unwanted cysteines

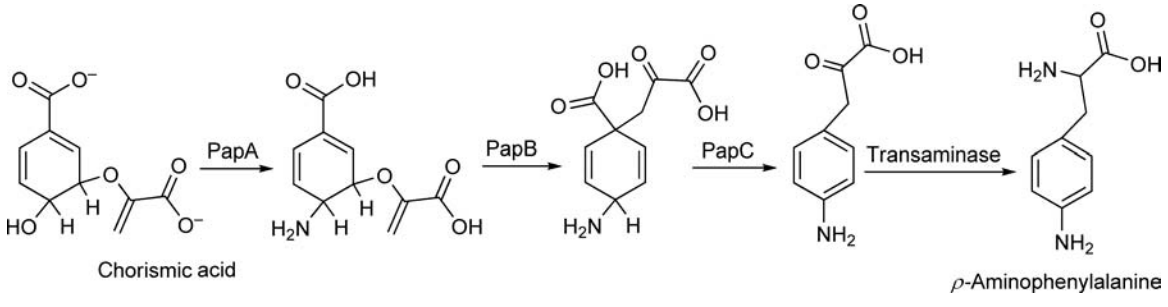


Figure 7 The biosynthesis of unnatural amino acid *p*-amino-L-phenylalanine in *Escherichia coli*.

can be mutated to other amino acids in the target protein. A unique and nonproteinogenic chemical group on an unnatural amino acid can serve as a chemical handle for bio-orthogonal reactions, occurring only on the handle and not on any endogenous protein residues. When introduced into proteins site-specifically, the chemical handle will allow a variety of reagents to be selectively appended to proteins *in vitro* or directly in cells. These reagents can be biophysical probes, tags, posttranslational modifications, and groups that can modify protein stability or activity. The diversity of molecules that can be attached to proteins through different chemical handles allows a variety of modified proteins to be generated for many purposes.

5.18.3.1.1 The keto group

The keto group undergoes a variety of reactions including addition, aldol, transamination, and isomerization reactions. It reacts with hydrazides, alkoxyamines, and semicarbazides under aqueous, mild conditions to produce hydrazone, oxime, and semicarbazone linkages that are stable under physiological conditions. The keto group is not present in any of the canonical amino acids, making it an attractive chemical handle for site-specific labeling using bio-orthogonal reactions.

The keto functional group was genetically encoded in *E. coli* in the form of *p*-acetyl-L-phenylalanine (**5**) and was specifically labeled *in vitro* with fluorescein hydrazide and biotin hydrazide with greater than 90% yield (**Figure 8(a)**).[75] In another application of generating homogenous glycoprotein mimetics, sugars were prepared in the aminooxy form and attached to *p*-acetyl-L-phenylalanine incorporated at defined sites in proteins (**Figure 8(b)**).[76] The attached sugar was subsequently elaborated by adding additional saccharides with glycosyltransferases to generate oligosaccharides with defined structures. Alternatively, an aminooxy-derivatized glycan can also be covalently coupled to the *p*-acetyl-L-phenylalanine-containing protein in one step. Such glycoprotein mimetics with pure and known glycan structures will be valuable for elucidating the functional role of sugars in glycoproteins. Similarly, fluorescent dyes were selectively labeled with the *m*-acetyl-L-phenylalanine (**6**) introduced into the membrane protein LamB in *E. coli*.[77] Rhodopsin, a visual photoreceptor and a transmembrane G-protein-coupled receptor (GPCR), was expressed in mammalian cells to contain *p*-acetyl-L-phenylalanine or *p*-benzoyl-L-phenylalanine.[78] After purification, the mutant rhodopsin was labeled *in vitro* using the keto-containing unnatural amino acid with fluorescein hydrazide, which did not perturb its function. These results are promising, and ideally it will be possible to attach fluorophores onto proteins that are not easily tagged by fluorescent proteins directly in live mammalian cells in the future.

Another useful application of the chemical handles is the addition of polyethylene glycol (PEG) specifically to the incorporated unnatural amino acids. PEG is capable of shielding proteins from proteases and increases the apparent size of the protein, which reduces clearance rates. PEGylation thus increases protein stability as well as solubility. Compared with PEGylation by modifying common amino acids, site-specific PEGylation using a unique chemical handle also increases the homogeneity of the final protein product and provides flexibility in choosing PEGylation sites to optimize protein activity. PEG was attached to *p*-acetyl-L-phenylalanine introduced in the human growth hormone to afford a protein that retained wild-type activity but had a considerably improved half-life in serum (H. Cho and T. Daniel, unpublished results). This approach can be extended to other therapeutic proteins to generate purer and more stable protein drugs.

Another unnatural amino acid (**28**), the β-diketone, has also been genetically incorporated and used *in vitro* to label proteins with hydroxylamine derivatives of biotin or fluorescent dyes,[79] thereby increasing the number of unique chemical handles that can be genetically incorporated into proteins *in vivo*.

5.18.3.1.2 Azide and acetylene

The azide and acetylene functional groups can react with each other through a copper(I)-catalyzed [2 + 3] cycloaddition reaction. Unnatural amino acid *p*-azidophenylalanine (**7**) containing the azide group and *p*-propargyloxyphenylalanine (**8**) containing the acetylene group were genetically incorporated into human superoxide dismutase-1 (SOD) in *E. coli* and yeast.[80,81] Purified mutant proteins were then labeled *in vitro* with fluorescent dyes or PEG derivatized with the complementary functional group (**Figure 9(a)**). The requirement of Cu^{+} as catalyst may make the use of this reaction inside cells difficult.

The azide can also react with phosphine derivatives through the Staudinger ligation. *p*-Azidophenylalanine was incorporated into the Z-domain protein in *E. coli* or into peptides displayed on phage, and was labeled with fluorescein-derived phosphines in phosphate buffer at room temperature (**Figure 9(b)**).[82]

p-Azidophenylalanine can also be used for photocrosslinking, which will be discussed in Section 5.18.3.3.1.

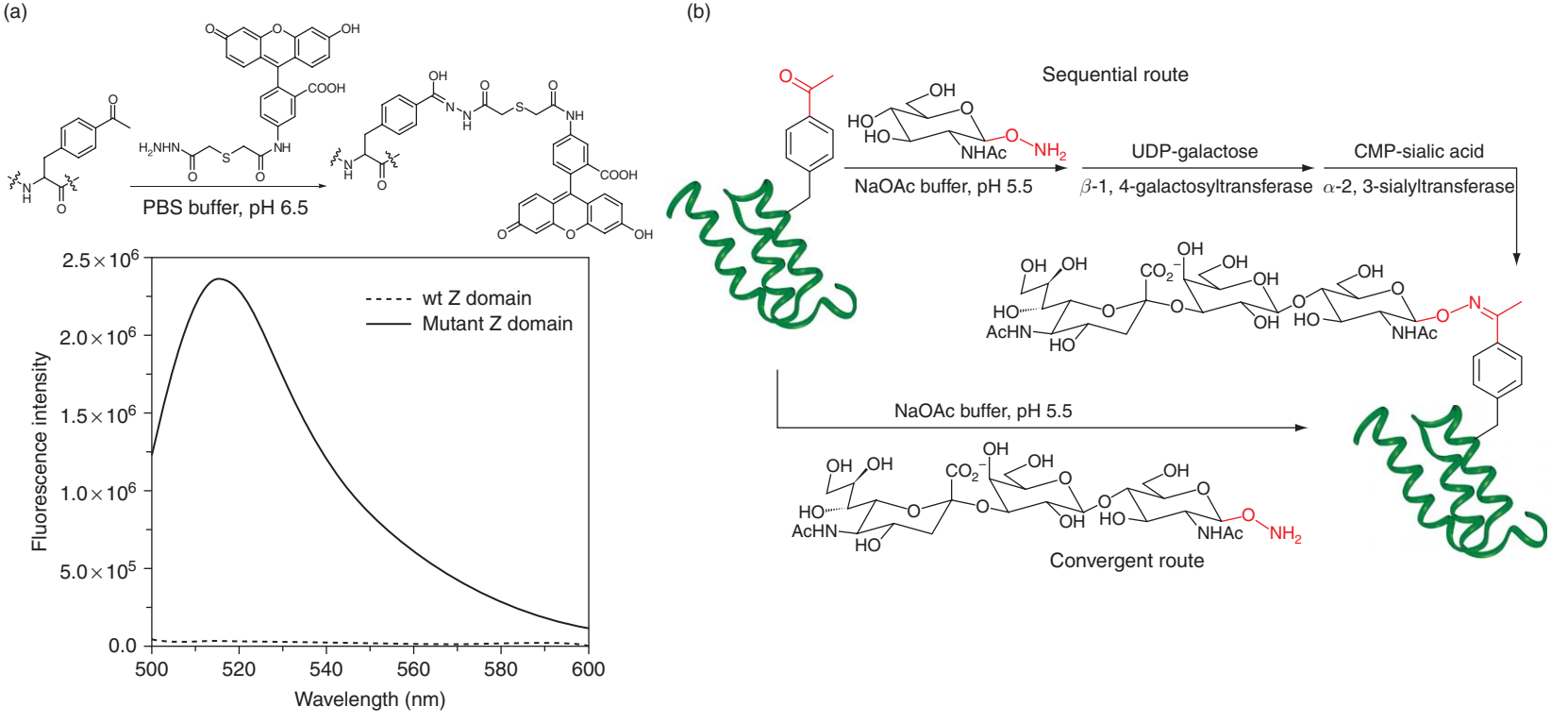


Figure 8 Selective protein modification using a keto amino acid, *p*-acetyl-L-phenylalanine. (a) Labeling of fluorescein hydrazide to the Z domain protein. Only the mutant protein containing *p*-acetyl-L-phenylalanine was labeled and became fluorescent. (b) A general method for preparing glycoprotein mimetics with defined glycan structure.

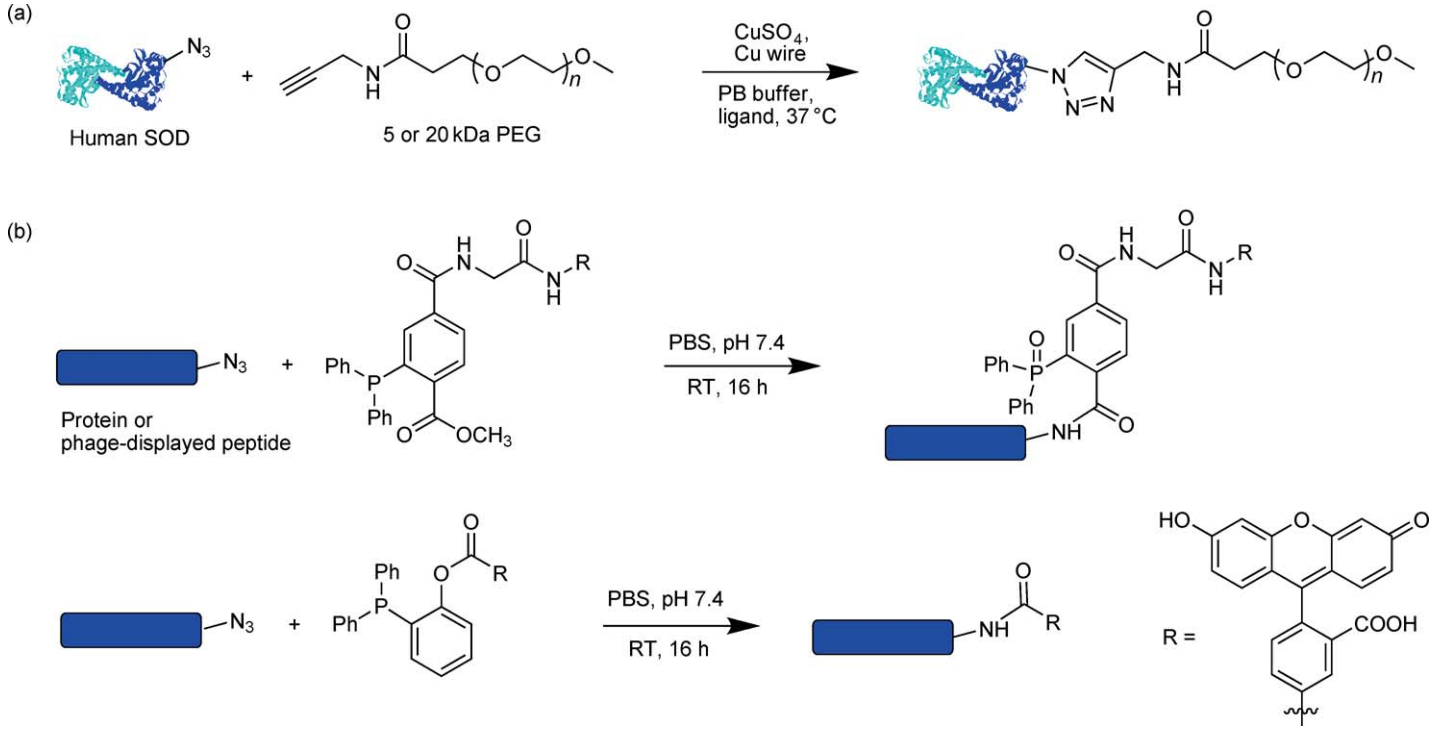


Figure 9 (a) Protein modification with PEG through a copper-catalyzed cycloaddition reaction. (b) Protein modification with fluorescein through Staudinger ligation.

5.18.3.1.3 Phenylselenide

Phenylselenide, after oxidation and β-elimination in mild conditions, undergoes Michael addition reactions with thiols. One unnatural amino acid, dehydroalanine, is naturally found in a number of proteins and normally arises through posttranslational modification of serine or cysteine. This α,β-unsaturated side chain can alter the conformation and rigidity of the peptide backbone. Dehydroalanine is unstable as a free amino acid, however, and was genetically incorporated as a protected precursor unnatural amino acid, phenylselenocysteine (**9**), into GFP (**Figure 10**).[83] After deprotection with hydrogen peroxide, the resultant dehydroalanine in purified proteins was selectively labeled with thiol-containing mannopyranose to generate a glycosylated GFP. A similar approach was taken to generate a palmitoylated protein with a nonhydrolyzable analogue, demonstrating the generality of this method for different posttranslational modifications.

The phenylselenocysteine has also been used successfully to chemically append analogues of methyl- or acetyl-lysine, important histone modifications that can contribute to chromatin structure and accessibility of transcriptional machinery in eukaryotes. By introducing phenylselenocysteine into the *Xenopus* histone H3, both acetyl-lysine and mono-, di-, and trimethyl-lysine analogues were appended to the purified unnatural amino acid-containing H3 protein (**Figure 10**).[84] Additionally, the H3 protein with a modification mimicking acetylation of lysine 9 can be deactylated by a histone deacetylation complex and is also a substrate for phosphorylation by Aurora B kinase. Such purified and chemically labeled histones are likely functional in nucleosomes, and preparation of specifically modified histones for comprehensive analysis of chromatin structure and accessibility is particularly suited to this chemical labeling technique.

5.18.3.2 Biophysical Probes for Studying Protein Structure and Function

Site-specific incorporation of biophysical probes into proteins allows the structure and function of proteins to be probed with greater precision and accuracy using various biophysical means. Genetically encoded biophysical probes in the form of unnatural amino acids further extend the potential of these studies into live cells, the

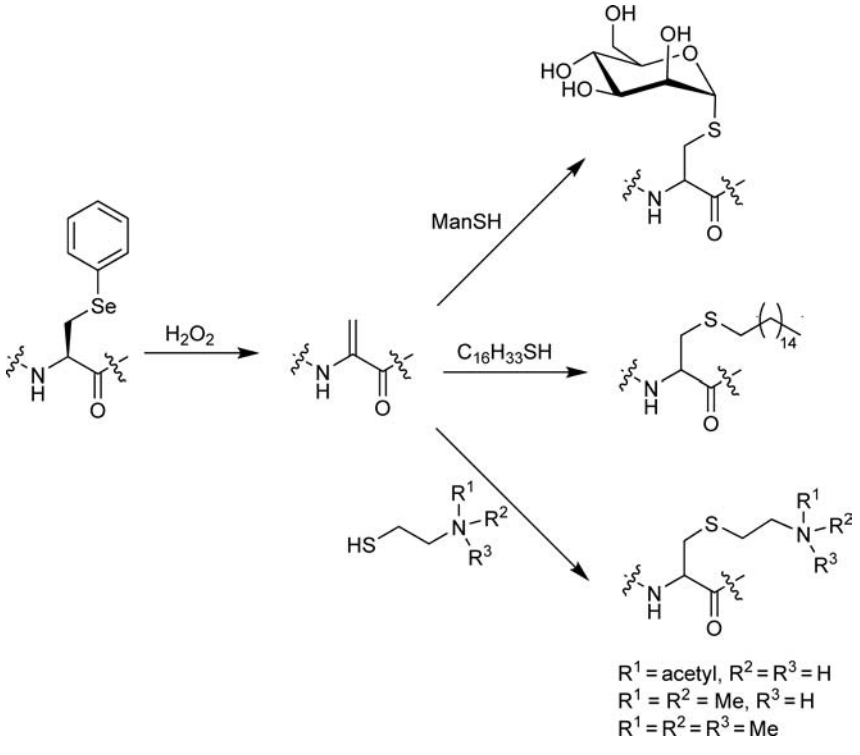


Figure 10 Protein modification through the genetically incorporated phenylselenocysteine.

native environment of proteins. A handful of unnatural amino acids that can serve as biophysical probes have been genetically incorporated into proteins in *E. coli*, yeast, and mammalian cells, and representative applications of some probes are discussed here.

5.18.3.2.1 Heavy atom-containing amino acids for X-ray crystallography

Standard techniques to introduce heavy atoms into protein crystals for phase determination use telluromethionine or selenomethionine to replace methionine during protein translation, or soak the protein crystal in solutions containing ions for heavy atoms. The soaking method can produce a number of low-occupancy sites whose positions must be determined before the phase can be derived, while quantitative replacement of methionine is difficult. A novel method to introduce heavy atoms into proteins is to genetically incorporate *p*-iodo-L-phenylalanine (**3**) into proteins in *E. coli* or yeast cells. This approach ensures that the heavy atom iodine is quantitatively introduced at a specific site of the target protein, and a large quantity of proteins can be conveniently prepared for single-wavelength anomalous dispersion (SAD) phasing. *p*-Iodo-L-phenylalanine was incorporated into T4 lysozyme, and the mutant protein was crystallized.[85] Diffraction data were collected with a laboratory CuKα X-ray source, and the structure was solved using SAD. A single iodinated amino acid among 164 residues resulted in a strong anomalous signal, about 3% of the total intensities, which compares favorably with the level achieved with selenomethionine using synchrotron beams. In addition, this amino acid caused little structural perturbation when substituted for Phe in the core of T4 lysozyme. The strong anomalous signal, the possibility of incorporation at multiple sites and in different cell types, and the use of an in-house X-ray source should facilitate solving novel protein structures in a high-throughput manner, providing more structural data to the scientific community.

5.18.3.2.2 Nuclear magnetic resonance probes

Nuclear magnetic resonance (NMR) can be used to determine the structure and dynamics of a protein or complex of proteins in solution. Transient binding or conformational changes associated with binding to small molecules or other proteins could be measured, in principle, if NMR spectra from large proteins or complexes were less complicated to decode. The ability to introduce one or several specific NMR labels at defined locations in a protein would greatly reduce the complexity of its NMR spectra and simplify the signal assignment. Genetic incorporation of unnatural amino acids containing isotopic labels or fluorine can allow such NMR labels to be introduced site-selectively into proteins with high fidelity and efficiency.

Trifluoromethyl-L-phenylalanine (tfm-Phe, **10**) was placed into the binding interface of two well-studied obligate dimers, nitroreductase and histidinol dehydrogenase, both of which contain active sites at the dimer interface.[86] Possible decreases in enzyme activity from unnatural amino acid incorporation, as well as substrate- or inhibitor-induced conformational changes, could be measured by NMR, and there were clear conformation-based changes in signal output. In another report, the binding of a small molecule ligand to the thioesterase domain of fatty acid synthase was studied by NMR using three different NMR-active unnatural amino acids, tfm-Phe, ^{13}C- and ^{15}N-labeled OMeTyr (**11**), and ^{15}N-labeled *o*-nitrobenzyl tyrosine (**12**), at 11 different sites (most solvent exposed), none of which seriously perturbed the structure of the protein.[87] By comparing the spectra of a number of mutants and the changes in conformation upon addition of a small molecule, the binding site of this molecule was mapped on the protein. In addition, photodecaging of the ^{15}N-labeled *o*-nitrobenzyl tyrosine to regenerate ^{15}N-labeled tyrosine provides a new method of isotopic labeling of individual residues without altering the protein sequence. The utility of these methods for the identification of active small molecules and the exclusion of those binding to nondesired regions of the protein, as well as their effects on conformation, are likely to be used in the future for NMR structural studies.

5.18.3.2.3 Infrared probes

Other potential unnatural amino acids that can be used to finely probe the structure of proteins include the reported *p*-cyano-L-phenylalanine (*p*CNPhe, **13**), a probe for infrared (IR) spectroscopy.[88] The stretching vibration of the nitrile group of *p*CNPhe has strong absorption and a frequency (ν_{CN}) at $\sim$2200 cm^{-1}, which falls in the transparent window of protein IR spectra. *p*CNPhe was genetically incorporated into myoglobin in *E. coli* and used to examine different ligand-bound states of the heme group of myoglobin. A substitution of *p*CNPhe was made for His64, which is at the distal face and close to the iron center of the heme group in

myoglobin. In the ferric myoglobin, when the Fe(III) ligand was changed from water to cyanide, ν_{CN} shifted from 2248 to 2236 cm^{-1}, which indicates a less polar active site. In the ferrous myoglobin, a ν_{CN} absorption at 2239 cm^{-1} was observed for the linear Fe(II)CO complex, and the bent Fe(II)NO and Fe(II)O_2 complexes showed a ν_{CN} absorption at 2230 cm^{-1}. These results demonstrate that the nitrile group is a sensitive probe for ligand binding and for local electronic environment.

5.18.3.2.4 Bulk probes

The side chain size of unnatural amino acids can be conveniently altered at atomic precision to probe the effects of amino acid bulk on protein structure and function. For instance, the fast inactivation mechanism of the voltage-gated potassium channel Kv1.4 was examined using unnatural amino acids in HEK293 cells.[46] The classic ball-and-chain model for channel inactivation suggests that the N-terminal inactivation peptide forms a ball-like domain to occlude the channel exit for ions. In contrast, a new model hypothesizes that the inactivation peptide threads through a side portal and extends into the inner pore of the channel to block ion flow. To experimentally test the new model, tyrosine 19, a residue in the inactivation peptide, was initially mutated to phenylalanine or tryptophan, which showed no difference in channel inactivation. However, mutation of this residue to OMeTyr (**1**) resulted in a markedly slower inactivation, as did the incorporation of dansylalanine (**14**) at this site (**Figure 11**). Modeling suggested that the diameter of the inactivation peptide, which is unchanged for each of the canonical amino acids incorporated at this site but is larger for OMeTyr and dansylalanine, is important for channel inactivation. This is likely due to the narrow width of the side portal in

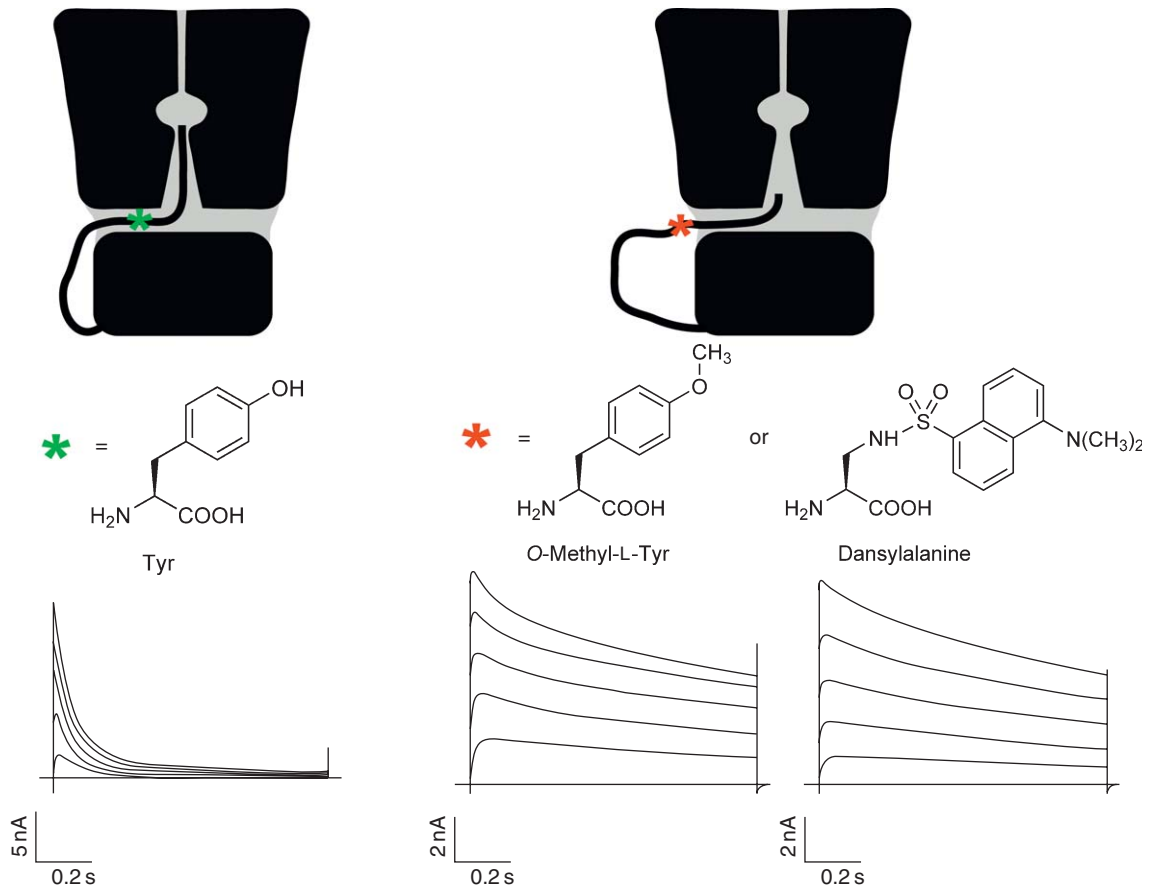


Figure 11 Unnatural amino acids to probe the inactivation mechanism of ion channel Kv1.4. OMeTyr or dansylalanine extends the side chain length of tyrosine, which impedes the inactivation peptide from threading through the side portal of the ion channel and abolishes the fast inactivation, as shown in the current–time curve in the bottom panel.

the channel, supporting the new model for channel inactivation. Elucidation of structure–function relationships involving specific residues in other proteins can also benefit from unnatural amino acid incorporation, as in this case, because conventional mutagenesis is unable to provide the desired increase or decrease residue bulk to clearly delineate the effect.

5.18.3.2.5 Fluorescent probes and quenchers

The ability to tag a protein with fluorescence has been a boon to molecular and cell biology, allowing *in vivo* imaging studies to detect protein expression, localization, or trafficking. Fluorescent amino acids would provide a similar level of genetic encodability, but without the bulkiness and potential perturbations of normal protein function and interaction that fluorescent proteins may introduce. Additionally, these unnatural amino acids can be engineered to be sensitive to microenvironments, lending themselves to potential use as sensors of posttranslational modification, dimerization, protein folding, pH, and a number of other physical and biochemical properties. Two fluorescent amino acids, 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide) propanoic acid (dansylalanine, **14**) and L-(7-hydroxycoumarin-4-yl) ethylglycine (**15**), have been genetically encoded into proteins. Each of these unnatural amino acids possesses a fluorescent moiety that is excited in the UV spectrum and emits blue or green light. The emission intensity or wavelength of these unnatural amino acids is dependent on the microenvironment where the unnatural amino acid is located. Dansylalanine is brighter in hydrophobic environments than hydrophilic ones, making it a feasible probe of dimerization or protein folding, while coumarin is sensitive to the pH in addition to the polarity of the surroundings.

Dansylalanine has been genetically encoded as a protein folding reporter in human SOD, which was expressed in *S. cerevisiae*, purified, and treated with guanidinium chloride.[36] Placing the unnatural amino acid in a site on the surface of a β-barrel in the protein showed little change in fluorescence after protein denaturation. When the unnatural amino acid was placed in an internal site of the β-barrel, there was a redshift in the emission wavelength of the unnatural amino acid and a large decrease in fluorescence intensity upon denaturation. The change from a buried to exposed environment was reported by changes in wavelength and intensity of the fluorescence from the incorporated dansylalanine. The potential to use dansylalanine or similar unnatural amino acids as sensors for protein–protein dimerization by incorporation at a binding interface, as a fluorescent reporter of receptor–ligand binding, or to identify membrane-associated proteins, and in a number of other possible applications, makes this unnatural amino acid attractive for further study.

The coumarin-based fluorescent unnatural amino acid **15** has properties similar to those of dansylalanine, and has also been used to monitor protein folding due to its solvent polarity-dependent fluorescence emission.[89] The coumarin unnatural amino acid was incorporated into sperm whale myoglobin at two different helices, A or C. An increase in fluorescence correlates to local unfolding of the region containing the coumarin due to increased exposure of the unnatural amino acid to solvent. At $5\,\mathrm{mol\,l^{-1}}$ urea, both sites show an equal fluorescence increase, indicating unfolding of the protein. At $2\,\mathrm{mol\,l^{-1}}$ urea, fluorescence increases only when the unnatural amino acid is incorporated in helix A but not in helix C. However, at $3\,\mathrm{mol\,l^{-1}}$ urea, coumarin placed in helix C shows a similar increase in fluorescence. NMR data indicate that at urea concentrations above $2.2\,\mathrm{mol\,l^{-1}}$, helices A and B are largely unfolded, while the remaining helices (including C) are not destabilized until urea concentrations are above $3\,\mathrm{mol\,l^{-1}}$, suggesting that the fluorescent emissions of coumarin incorporated in different locations are an accurate reporter of local protein destabilization. Coumarin can be used as a site-specific probe of local protein folding and conformational changes, and may be useful for other fluorescence-based protein assays sensitive to polarity.

The unnatural amino acid *p*-nitrophenylalanine (**16**) is capable of quenching intrinsic tryptophan fluorescence, provided a short enough distance.[90] It was incorporated in a leucine zipper protein, GCN4, in a dimeric region. The tryptophanyl fluorescence of GCN4 was quenched in a distance-dependent manner, making *p*-nitrophenylalanine a useful distance probe to monitor protein folding or conformational changes.

Though still in its infancy, the use of genetically encoded fluorescent amino acids as biosensors is promising, and it is likely that these and similar unnatural amino acids will be used as optical reporters directly in live cells in the future for studying protein trafficking and function in live cells and in real time.

5.18.3.3 Photoreactive Groups for Identifying and Regulating Protein Activity

It would be advantageous to control protein interaction or activity with an easy to administer, noninvasive, and selective method. Application of light of a specific wavelength would be an ideal way to control protein activities directly *in vivo* without disrupting the cellular environment. Many functional groups can respond to light by crosslinking with nearby molecules, removing protecting groups, or changing molecular conformation. A number of unnatural amino acids bearing such functional groups as side chains have been genetically incorporated into proteins. They have the potential to be very useful in a wide variety of applications, including the identification of protein–protein or protein–nucleic acid interactions and the optical control of protein structure or function *in vivo*.

5.18.3.3.1 Photocrosslinkers

Crosslinking has been an important technique for uncovering protein interactions in cells. Using specific antibodies and formaldehyde, interactions between proteins or proteins and DNA can be examined. However, formaldehyde crosslinking lacks both spatial and temporal resolution. In general, it is difficult to pinpoint transient or weak interactions between proteins, to identify which domains of the protein are involved in the interaction, and to determine whether the interaction is direct or involves an unknown protein acting as a scaffold in protein complexes. Site-specific crosslinking enabled by genetically encoded unnatural amino acids would allow subregions of a protein to be examined for binding of proteins and would distinguish direct interactions from indirect ones. In addition, transient or weak interactions can be enriched by a prolonged crosslinking time, as only the proteins containing the unnatural amino acid will be covalently linked, rather than capturing a snapshot of the whole cell at once.

Three unnatural amino acids, *p*-azidophenylalanine (**7**), *p*-benzoylphenylalanine (*p*BpA, **17**), and *p*-(3-trifluoromethyl-3*H*-diazirin-3-yl)-phenylalanine (TfmdPhe, **18**), with photocrosslinking side chains have been genetically incorporated into proteins. Upon excitation by UV light, *p*-azidophenylalanine crosslinks to C–H and N–H bonds, *p*BpA to C–H bonds, and TfmdPhe to C–H or O–H bonds. A mutant homodimeric glutathione *S*-transferase with *p*-azidophenylalanine or *p*BpA substituted site-specifically at the dimer interface could be crosslinked *in vitro*,[91,92] as well as in the cytoplasm of *E. coli* in case of *p*Bpa.[93]

Photocrosslinkers have been incorporated in *E. coli* to confirm close contact between specific residues of a protein and its substrate. ClpB, a heat shock protein that aids in the disaggregation and refolding of proteins during the heat shock response, has a conserved aromatic residue (Tyr251) in the central pore of its AAA+ domain, considered to be the main substrate recognition residue.[94] After Tyr251 in ClpB was replaced with *p*BpA, biotinylated substrate peptides were shown to be crosslinked upon UV light exposure, but not if *p*BpA was incorporated elsewhere in the AAA+ domain of this protein.

Photocrosslinkers have also been used in *S. cerevisiae* to specifically crosslink a ligand to its receptor. Ste2p, a GPCR, is known to bind a short peptide pheromone, α-factor, which affects downstream signaling and leads to growth arrest. Several extracellular loops of Ste2p, based on available topology, were mutagenized to the amber stop codon for *p*BpA incorporation and α-factor photocapture.[95] Two sites were able to specifically crosslink biotinylated α-factor and were identified by Western blot. It is possible in the future that GPCRs or other cell membrane receptors can be mutagenized to incorporate a photocrosslinker and identify novel ligands through mass spectrometry.

In mammalian cells, *p*BpA has been incorporated into Grb2 (growth factor receptor-bound protein 2), in the ligand binding pocket of the SH2 domain, which binds to specific, phosphorylated tyrosine residues. It was shown that *p*BpA incorporation into a site on the SH2 domain did not interfere with Grb2 binding to the epidermal growth factor receptor (EGFR) by co-immunoprecipitation, and that a larger molecular weight band can be seen when activated EGFR and Grb2 containing *p*BpA were coexpressed and treated with UV light.[96] Several positions on the SH2 domain were chosen to place *p*BpA, with differing crosslinking capacities based on their distance from the ligand. Therefore, closely interacting versus father interacting residues in ligand binding may be distinguished using photocrosslinkers placed in various locations on the binding domain.

These studies suggest that fine structural and mechanistic information can be obtained from site-specific photocrosslinking using genetically incorporated unnatural amino acids. The potential of photocrosslinkers to identify biomolecular interactions more specifically than general formaldehyde crosslinking, combined with its

compatibility with a wide range of biological processes, suggests that this *in vivo* photocrosslinking strategy will be used more extensively in the future to very specifically probe protein complexes and the interactions within.

5.18.3.3.2 Photocaged amino acids

To regulate protein activity using light, one strategy is to attach a photoremovable protecting group to the suitable amino acid in the target protein, which masks the amino acid and renders the protein inactive. Photolysis releases the caging group and converts the amino acid to an active form, which generates abrupt or localized changes to the target protein. The nitrobenzyl derivative is the most prevalent form for caged compounds. The side chain hydroxy or thiol groups of Cys, Ser, and Tyr are blocked by substituted nitrobenzyl groups that can be cleaved on irradiation with 365-nm light. These photocaged Cys, Ser, and Tyr have been genetically incorporated into different proteins.

In one example, mutation of the active-site cysteine residue in the proapoptotic protease caspase 3 to *o*-nitrobenzylcysteine (**19**) in yeast led to a catalytically inactive enzyme. UV illumination of the cell lysate converted ~40% of the caged caspase to the active enzyme.[35] In another report, *o*-nitrobenzyltyrosine (**20**) was incorporated into β-galactosidase to activate its enzymatic activity by using light both *in vitro* and in *E. coli*.[97] β-Galactosidase activity is dependent on a critical tyrosine residue (Tyr503), and placing *o*-nitrobenzyltyrosine in this site effectively reduced the activity of this enzyme to 5% of the wild-type form. After a 30 min exposure of bacterial cells to long-wavelength UV light, the enzyme regained activity to around 70% of wild-type levels. In a third example, protein phosphorylation in yeast is controlled by a photocaged serine, 4,5-dimethoxy-2-nitrobenzylserine (DMNB-Ser, **21**).[98] DMNB can be removed by light with a wavelength of 405 nm, reducing the side effect of UV light on cells. This DMNB-Ser blocks phosphorylation and is incorporated at different phosphoserine sites in the transcription factor Pho4. Upon photodecaging, serine is regenerated and subsequently phosphorylated, triggering the nuclear export of Pho4 (**Figure 12**).

This strategy will likely be used to light-activate other enzymes of interest in living cells provided that enzymatic function is dependent on just one or a few residues of the protein. Another useful application could be the decaging of phosphorylated, ubiquitinated, or otherwise posttranslationally modified residues with light at specific time points for examination of signal transduction pathways.

5.18.3.3.3 Photoisomerizable amino acids

Photolysis of a caged amino acid residue is an irreversible process. Reversible modulation can be achieved with the photochromic azobenzene compounds. Azobenzene undergoes a reversible *cis–trans* isomerization: The more stable *trans* isomer can be converted to the *cis* isomer upon illumination at 320–340 nm, and the *cis* form can revert to *trans* form either thermally or by irradiation at >420 nm (**Figure 13**). The resultant change in the geometry and/or dipole of the compound can be used for regulating protein activity in a reversible manner. The azobenzene group has been genetically encoded in the form of *p*-azophenyl-phenylalanine (AzoPhe, **22**). AzoPhe was incorporated at the Ile71 site of the *E. coli* catabolite activator protein (CAP), a transcriptional activator.[99] Its binding affinity for the promoter sequence decreased fourfold after irradiation at 334 nm, which converts the predominant *trans* AzoPhe to the *cis* form. The isomerized *cis* AzoPhe was then switched back to

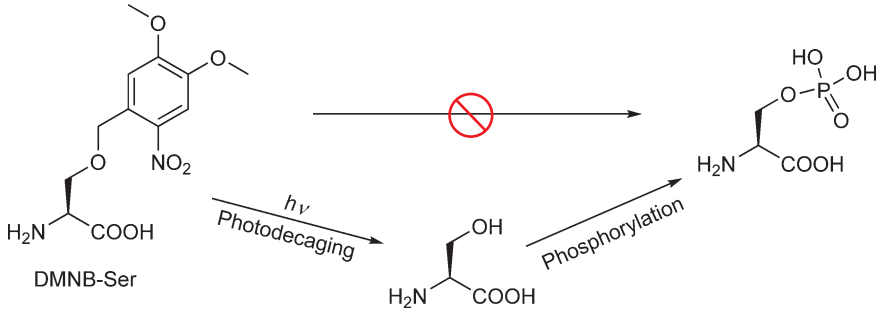


Figure 12 Photocaged serine is used to control when the target serine is regenerated and phosphorylated.

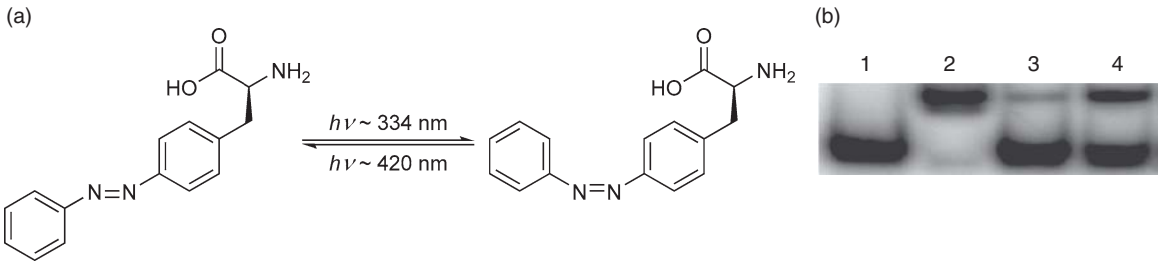


Figure 13 (a) The photoisomerizable unnatural amino acid *p*-azophenyl-phenylalanine (AzoPhe). (b) Gel mobility shift assay to determine the binding affinity of the catabolite activator protein (CAP) to the lactose promoter DNA fragment. Lane 1, DNA only; lane 2, DNA + wild-type CAP; lane 3, DNA + CAP with AzoPhe incorporated at residue 71 (after irradiation at 334 nm); lane 4, DNA + CAP with AzoPhe incorporated at residue 71 (before irradiation at 334 nm). Reprinted with permission from M. Bose; D. Groff; J. Xie; E. Brustad; P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 388–389. Copyright 2006 American Chemical Society.

the *trans* state by irradiation at >420 nm, after which the affinity of the protein for the promoter was completely recovered. Similarly, it should be possible to photomodulate the activity of other enzymes, receptors, and transcription factors.

5.18.3.4 Unnatural Amino Acids for Mimicking Posttranslational Modifications

One big challenge with studying protein function by expression and purification is the extremely large number of posttranslational modifications found on proteins, especially those from eukaryotes. These modifications often play important roles in protein stability, localization, and function, but cannot be easily mimicked or produced in bacterial cells. It is also difficult to prepare homogenously modified proteins from eukaryotic cells. Unnatural amino acids with the properties of posttranslational modifications, or chemical handles for the addition of these modifications, may allow the generation of homogenous proteins with the desired modifications to uncover the functional roles of posttranslational modifications and may enable us to control protein function by controlling the modifications appended to the protein of interest. Many posttranslational tags, like glycosylation, lipidation, acetylation, methylation, or phosphorylation, can be incorporated into proteins using genetically encoded unnatural amino acids.

5.18.3.4.1 Glycosylation

Glycosylation is a frequent modification of eukaryotic proteins and can have stability and functional consequences. Many proteins traverse the Golgi apparatus to be modified on specific residues with carbohydrate chains. Natural glycoproteins are often present as a population of many different glycoforms, which makes analysis of glycan structure and the study of glycosylation effects on protein structure and function complicated. Therefore, methods for the synthesis of natural and unnatural homogeonously glycosylated proteins are needed for the systematic understanding of glycan function and for the development of improved glycoprotein therapeutics.

One strategy to prepare homogenous glycoprotein mimetics is to genetically incorporate a chemical handle at the desired glycosylation site in the target protein and then label the chemical handle with sugar using bioorthogonal chemistry. Further elaboration with additional sugars will generate oligosaccharides with defined structure covalently linked to the protein. The keto- and phenylselenide-containing unnatural amino acids have been used in this way (see Section 5.18.3.1). This method will leave a small unnatural linkage between the protein and the first sugar.

To prepare a native glycosylated protein, mutant synthetases were evolved that site-specifically incorporate *β*-*N*-acetylglucosamine-*O*-serine (*β*-GlcNAc-Ser, **23**) into proteins in *E. coli*.[100] An *O*-GlcNAc-Ser mutant myoglobin generated by this method could be elaborated further to form more complex saccharides with galactosyltransferases. A similar approach was used to selectively introduce *N*-acetylgalactosamine-*α*-*O*-threonine (*α*-GalNAc-Thr, **24**) into proteins in *E. coli*.[101] Since there are no endogenous glycosyl modifications

in *E. coli* to confound the results, the ability to add a glycosylation tag onto proteins in *E. coli* can lead to improvement in the purification of many types of proteins that require glycosylation for proper folding and function.

5.18.3.4.2 Phosphorylation

Phosphorylation is a nearly ubiquitous posttranslational modification that can control protein function, conformation, and stability. Phosphorylation of serine, threonine, and tyrosine residues is the key switch in many signal transduction cascades. The ability to precisely control which residues are phosphorylated would enable dissection of which signaling events are tied to modifications on specific residues. Using photocaged serines, threonines, or tyrosines at critical sites, phosphorylation of specific residues could be prevented until a desired moment, which is a more attractive and versatile method than replacing these residues with natural amino acids that cannot be phosphorylated. On the contrary, it can be difficult to generate a signaling molecule that is constitutively active due to the reversible nature of phosphorylation and the large number of phosphatases in the cell. Creating a mimetic for a phosphorylated residue that cannot be cleaved by endogenous phosphatases can be used to identify the role of phosphorylation of a single residue, and can create more stable activated proteins for a variety of functional assays.

The unnatural amino acid *p*-carboxymethyl-L-phenylalanine (*p*CMF, **25**) is a nonhydrolyzable analogue of phosphotyrosine and was found to be capable of mimicking the phosphorylated state of Tyr.[102] This capability was demonstrated in a model phosphoprotein, the human signal transducer and activator of transcription-1 (STAT1). STAT1 has only a weak affinity for DNA, but during phosphorylation of Tyr701, STAT1 forms a homodimer and strongly binds a DNA duplex that contains M67 sites. The mutant STAT1 with Tyr701 substituted with *p*CMF also bound the M67-containing DNA duplex tightly, which suggests that *p*CMF could replace phosphotyrosine in the generation of constitutively active phosphoproteins. Nonhydrolyzable analogues of phosphorylated residues can be similarly used to study the functional consequences of constitutive phosphorylation of a number of proteins, and to study phosphorylation-induced changes in binding partners or binding affinity more easily.

5.18.3.4.3 Acetylation

Acetylation is a reversible modification on proteins that can also contribute to protein localization and function. Acetylation of lysine residues in histone proteins can control the secondary structure of chromatin as well as gene expression levels from certain loci, and chromatin remodeling and its consequences in a variety of molecular and cell biological questions are intensely researched. Many other proteins undergo reversible acetylation, and the functional consequences of these modifications are poorly understood in many cases.

One method for generating acetyl-lysine and methyl-lysine analogue-tagged histones was described in Section 5.18.3.1.3, using phenylselenocysteine as a caged chemical handle to make a large quantity of modified histone H3.[84] These synthetically generated histones retained the capability of being deacetylated and phosphorylated, suggesting that normal behavior in nucleosomes could be expected.

Unnatural amino acid N^{ε}-acetyllysine (**26**) was genetically incorporated into proteins in *E. coli*.[103] In the presence of a deacetylase inhibitor, manganese superoxide dismutase (MnSOD) containing N^{ε}-acetyllysine at residue Lys44 was prepared from *E. coli*. Acetylation of Lys44 in MnSOD, based on experiments comparing recombinant MnSOD with or without N^{ε}-acetyllysine at lysine 44, did not appear to affect the activity of the enzyme, however. N^{ε}-acetyllysine and other lysine analogues were also incorporated into GRB2 in HEK293 cells.[104]

5.18.3.5 Unnatural Amino Acids for Generating New Protein Function

One interesting aspect of incorporating unnatural amino acids into proteins is the ability to generate completely new protein functions. Though countless proteins have been generated through both natural and directed evolution and rational engineering, these strategies are all limited by the 20 canonical amino acids in the genetic code. Encoding additional functional groups can enable totally novel and previously impossible protein functions to be generated and evolved for various purposes.

5.18.3.5.1 Metal-chelating amino acids

Metal ions participate in a large number of catalytic and electron transfer reactions in cells, primarily as cofactors. These ions tend to be coordinated by a number of amino acid residues in proteins. It is difficult to predict and control the primary and secondary amino acid shells surrounding the metal ion, complicating protein engineering efforts. A potential solution is to genetically incorporate multidentate metal-chelating amino acids, which have coordinating atoms preorientated in the correct configuration.

Some unnatural amino acids have been designed with this metal-chelating property in mind. For instance, bipyridylalanine (BpyAla, **27**) has the bipyridyl group that chelates most transition metal ions and has been successfully incorporated into proteins in *E. coli*.[105] BpyAla was shown to reversibly bind copper ions when incorporated into T4 lysozyme, but a tyrosine in the same location was unable to bind copper, indicating that BpyAla is useful to coordinate copper ions to a protein of interest.

The Cu(II) binding property should make oxidative cleavage of the phosphosugar moiety of the nucleic acid backbone possible. Indeed, the BpyAla has been incorporated into the CAP and converts this bacterial transcription factor into a sequence-specific DNA-cleaving protein.[106] BpyAla was incorporated into CAP at Lys26, a site close to the protein–DNA interface. In the presence of Cu(II) and 3-mercaptopropionic acid, this CAP mutant cleaves double-stranded DNA at its consensus sequence with high specificity. This method may be generally applied to other DNA binding proteins to map their preferred DNA sequences.

5.18.3.5.2 Other possibilities

Unnatural amino acids can possess such diverse structures that there are a number of them that may be used for purposes previously unidentified. For instance, *p*-nitrophenylalanine (**16**), originally used as a distance probe to quench the fluorescence of tryptophan, has recently been found to stimulate potent immune responses for novel immunogenic applications.

In general, an animal will not mount a substantial immune response against a protein that is normally present in the body. It was shown that the incorporation of *p*-nitrophenylalanine into antigens is capable of breaking immune self-tolerance.[107] Incorporation of this unnatural amino acid into tumor necrosis factor α (TNF-α) led to an immune response directed at both the unnatural amino acid mutant and the natural version of the protein, even in the absence of adjuvant. Incorporation of the nitro group created a T-cell epitope that was not previously recognized, and mobilized a robust immune response against TNF-α. The immune response was retained in mice after exposure, because after lipopolysaccharide challenge, which leads to a TNF-α-mediated septic shock, animals immunized with the mutant TNF-α had improved survival rates compared with mock-immunized animals or those immunized with wild-type TNF-α. These striking observations have the potential for generating new therapies for a number of human diseases. This work further shows the immense diversity that unnatural amino acids can play in a number of disciplines and suggests that there are many more unknown applications to be identified.

5.18.4 Future Developments

Since 2001, when the first unnatural amino acid was genetically encoded in live cells,[25] more than 40 new unnatural amino acids have been incorporated into proteins in prokaryotic and eukaryotic cells. The number of novel experiments that can be performed increases each day another functional group is made available. Proof-of-principle experiments have demonstrated the utility of many of these unnatural amino acids, and their application in addressing challenging biological questions and uncovering unknowns should emerge in the coming years.

The use of this technology in mammalian cells is a burgeoning area, in which a variety of cell biology questions can be answered in the native cellular environment. Proteins containing fluorescent unnatural amino acids that behave as sensors of small molecules and protein–protein interactions, or proteins whose activity can be photoregulated *in vivo*, are particularly attractive. Mimics of posttranslational modifications or photocaged amino acids may be used to control signal transduction components. Genetically encoded unnatural amino acids can aid in the identification of proteins that are critical in the development of different tissue and cell lineages, can tease apart at the molecular level the process of cellular differentiation, and may generate new

therapies for a number of human diseases. The ability to control signal transduction events directly in live cells would enable a better understanding of diseases in which signal transduction has gone awry, including cancer.

Another future direction for unnatural amino acid incorporation would be the introduction of this technology into multicellular organisms. The ability to genetically incorporate unnatural amino acids in a multicellular organism would open the field to biological problems concerning how cells interact with one another in a variety of settings, allowing aspects of development, neural connectivity, and cellular signaling to be studied, to name only a few.

The genetically encoded unnatural amino acids, as new building blocks for proteins, should allow the design or evolution of proteins with novel properties. For instance, glycosylated or PEGylated therapeutic proteins may be rationally designed to improve pharmacological properties; various proteins can be engineered with light responsiveness to work as noninvasive molecular tools inside cells; protein properties never required and that have never existed in nature may be evolved from random unnatural amino acid mutagenesis. It may even be possible to experimentally test whether there is an evolutionary advantage for organisms with more than the 20 genetically encoded amino acids.

Glossary

aminoacyl-tRNA synthetase An enzyme that catalyzes the esterification of an amino acid to its cognate tRNAs.
bio-orthogonal reaction A chemical reaction that will not occur with the functional groups naturally present in cells, but only with the group introduced exogenously.
biophysical probe A chemical group whose physical property can be detected with a biophysical method or instrument such as NMR and optical microscope.
chemical handle A chemical group that can react with certain functional groups and can be used to attach molecules.
codon A nucleotide triplet that has a specific meaning during translation (one of the 20 canonical amino acids or a stop signal).
DNA shuffling A method to exchange portions of DNA among similar genes, in order to select for a certain trait.
genetic code The set of rules that decodes nucleotide triplets (codons) into amino acids, which is identical in nearly every organism.
heterologous expression The expression of a gene from one organism in a different organism. Green fluorescent protein is often heterologously expressed.
photocage A protecting group that is used to mask other molecules and can be removed upon illumination with a specific wavelength of light.
photocrosslink A method for crosslinking organic molecules by using chemical groups that covalently bind to adjacent molecules when illuminated with a certain wavelength of light.
posttranslational modification A covalent and often reversible modification occurring on a protein after it has been translated, and can include phosphorylation, glycosylation, acetylation, methylation, ubiquitination, sumoylation, and so on.
suppressor tRNA A tRNA that recognizes a stop codon (UAG, UGA, or UAA). When charged with an amino acid, it effectively prevents the stop codon from being recognized as such and the mRNA continues to be translated.
unnatural amino acid An amino acid with the same basic structure as a natural amino acid so that it can be incorporated into a polypeptide chain, but with a side chain containing a functional group different from the canonical 20 amino acids.

References

1. A. Bock; K. Forchhammer; J. Heider; W. Leinfelder; G. Sawers; B. Veprek; F. Zinoni, *Mol. Microbiol.* **1991**, *5*, 515–520.
2. G. Srinivasan; C. M. James; J. A. Krzycki, *Science* **2002**, *296*, 1459–1462.
3. L. Wang; P. G. Schultz; *Angew. Chem. Int. Ed. Engl.* **2004**, *44*, 34–66.
4. A. F. Carne, *Methods Mol. Biol.* **1994**, *32*, 311–320.
5. S. B. Kent, *Annu. Rev. Biochem.* **1988**, *57*, 957–989.

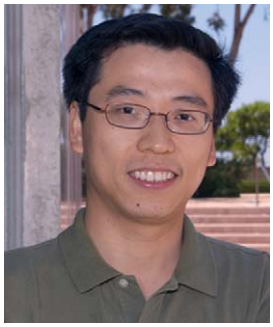
6. T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
7. G. Hortin; I. Boime, *Methods Enzymol.* **1983**, *96*, 777–784.
8. C. J. Noren; S. J. Anthony-Cahill; M. C. Griffith; P. G. Schultz, *Science* **1989**, *244*, 182–188.
9. J. D. Bain; C. G. Glabe; T. A. Dix; A. R. Chamberlin; E. S. Diala, *J. Am. Chem. Soc.* **1989**, *111*, 8013–8014.
10. L. Wang; P. G. Schultz, *Chem. Commun.* **2002**, *1*, 1–11.
11. S. Benzer; S. P. Champe, *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 1114–1121.
12. L. Bossi; J. R. Roth, *Nature* **1980**, *286*, 123–127.
13. M. Ibba; D. Soll, *Annu. Rev. Biochem.* **2000**, *69*, 617–650.
14. D. R. Liu; T. J. Magliery; P. G. Schultz, *Chem. Biol.* **1997**, *4*, 685–691.
15. Y. Kwok; J. T. Wong, *Can. J. Biochem.* **1980**, *58*, 213–218.
16. L. Wang; T. J. Magliery; D. R. Liu; P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 5010–5011.
17. B. Steer; P. Schimmel, *J. Biol. Chem.* **1999**, *274*, 35601–35606.
18. L. Wang; P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883–890.
19. L. Wang; J. Xie; P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
20. M. Pastrnak; T. J. Magliery; P. G. Schultz, *Helv. Chim. Acta* **2000**, *83*, 2277–2286.
21. A. K. Kowal; C. Kohrer; U. L. RajBhandary, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2268–2273.
22. J. C. Anderson; N. Wu; S. W. Santoro; V. Lakshman; D. S. King; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566–7571.
23. J. C. Anderson; P. G. Schultz, *Biochemistry* **2003**, *42*, 9598–9608.
24. S. W. Santoro; J. C. Anderson; V. Lakshman; P. G. Schultz, *Nucleic Acids Res.* **2003**, *31*, 6700–6709.
25. L. Wang; A. Brock; B. Herberich; P. G. Schultz, *Science* **2001**, *292*, 498–500.
26. S. W. Santoro; L. Wang; B. Herberich; D. S. King; P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044–1048.
27. M. Pastrnak; P. G. Schultz, *Bioorg. Med. Chem.* **2001**, *9*, 2373–2379.
28. L. Wang; A. Brock; P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 1836–1837.
29. H. Edwards; P. Schimmel, *Mol. Cell. Biol.* **1990**, *10*, 1633–1641.
30. H. Edwards; V. Trezeguet; P. Schimmel, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1153–1156.
31. H. J. Drabkin; H. J. Park; U. L. Rajbhandary, *Mol. Cell. Biol.* **1996**, *16*, 907–913.
32. D. Kiga; K. Sakamoto; K. Kodama; T. Kigawa; T. Matsuda; T. Yabuki; M. Shirouzu; Y. Harada; H. Nakayama; K. Takio; Y. Hasegawa; Y. Endo; I. Hirao; S. Yokoyama, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9715–9720. Epub 2002 Jul 9713.
33. K. Sakamoto; A. Hayashi; A. Sakamoto; D. Kiga; H. Nakayama; A. Soma; T. Kobayashi; M. Kitabatake; K. Takio; K. Saito; M. Shirouzu; I. Hirao; S. Yokoyama, *Nucleic Acids Res.* **2002**, *30*, 4692–4699.
34. J. W. Chin; T. A. Cropp; J. C. Anderson; M. Mukherji; Z. Zhang; P. G. Schultz, *Science* **2003**, *301*, 964–967.
35. N. Wu; A. Deiters; T. A. Cropp; D. King; P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 14306–14307.
36. D. Summerer; S. Chen; N. Wu; A. Deiters; J. W. Chin; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9785–9789.
37. G. Galli; H. Hofstetter; M. L. Birnstiel, *Nature* **1981**, *294*, 626–631.
38. F. Li; Y. Xiong; J. Wang; H. D. Cho; K. Tomita; A. M. Weiner; T. A. Steitz, *Cell* **2002**, *111*, 815–824.
39. S. Chen; P. G. Schultz; A. Brock, *J. Mol. Biol.* **2007**, *371*, 112–122.
40. Q. Wang; L. Wang, *J. Am. Chem. Soc.* **2008**, *130*, 6066–6067.
41. D. Cao; R. Parker, *Cell* **2003**, *113*, 533–545.
42. S. W. Peltz; A. H. Brown; A. Jacobson, *Genes Dev.* **1993**, *7*, 1737–1754.
43. C. Köhrer; L. Xie; S. Kellerer; U. Varshney; U. L. RajBhandary, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14310–14315.
44. S. L. Monahan; H. A. Lester; D. A. Dougherty, *Chem. Biol.* **2003**, *10*, 573–580.
45. Z. Zhang; L. Alfonta; F. Tian; B. Bursulaya; S. Uryu; D. S. King; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8882–8887.
46. W. Wang; J. Takimoto; G. V. Louie; T. J. Baiga; J. P. Noel; K. F. Lee; P. A. Slesinger; L. Wang, *Nat. Neurosci.* **2007**, *10*, 1063–1072.
47. E. Myslinski; J. C. Ame; A. Krol; P. Carbon, *Nucleic Acids Res.* **2001**, *29*, 2502–2509.
48. M. R. Paule; R. J. White, *Nucleic Acids Res.* **2000**, *28*, 1283–1298.
49. W. Liu; A. Brock; S. Chen; S. Chen; P. Schultz, *Nat. Methods* **2007**, *4*, 239–244.
50. J. F. Curran; M. Yarus, *Science* **1987**, *238*, 1545–1550.
51. L. Bossi; J. R. Roth, *Cell* **1981**, *25*, 489–496.
52. M. O'Connor, *Nucleic Acids Res.* **2002**, *30*, 1985–1990.
53. J. C. Anderson; T. J. Magliery; P. G. Schultz, *Chem. Biol.* **2002**, *9*, 237–244.
54. T. J. Magliery; J. C. Anderson; P. G. Schultz, *J. Mol. Biol.* **2001**, *307*, 755–769.
55. J. F. Atkins; R. B. Weiss; S. Thompson; R. F. Gesteland, *Annu. Rev. Genet.* **1991**, *25*, 201–228.
56. T. Hohsaka; Y. Ashizuka; H. Murakami; M. Sisido, *Nucleic Acids Res.* **2001**, *29*, 3646–3651.
57. T. Hohsaka; Y. Ashizuka; H. Taira; H. Murakami; M. Sisido, *Biochemistry* **2001**, *40*, 11060–11064.
58. C. Switzer; S. E. Moroney; S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.
59. J. A. Piccirilli; T. Krauch; S. E. Moroney; S. A. Benner, *Nature* **1990**, *343*, 33–37.
60. E. T. Kool, *Curr. Opin. Chem. Biol.* **2000**, *4*, 602–608.
61. A. A. Henry; F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
62. I. Hirao, *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627.
63. D. L. McMinn; A. K. Ogawa; Y. Q. Wu; J. Q. Liu; P. G. Schultz; F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11586–11587.
64. T. J. Matray; E. T. Kool, *J. Am. Chem. Soc.* **1998**, *120*, 6191–6192.
65. M. Chiaramonte; C. L. Moore; K. Kincaid; R. D. Kuchta, *Biochemistry* **2003**, *42*, 10472–10481.
66. I. Hirao; M. Kimoto; T. Mitsui; T. Fujiwara; R. Kawai; A. Sato; Y. Harada; S. Yokoyama, *Nat. Methods* **2006**, *3*, 729–735.
67. A. M. Leconte; G. T. Hwang; S. Matsuda; P. Capek; Y. Hari; F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343.
68. R. A. Mehl; J. C. Anderson; S. W. Santoro; L. Wang; A. B. Martin; D. S. King; D. M. Horn; P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935–939.
69. V. W. Cornish; D. Mendel; P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 621–633.

70. D. L. Beene; D. A. Dougherty; H. A. Lester, *Curr. Opin. Neurobiol.* **2003**, *13*, 264–270.
71. D. Schwarzer; P. A. Cole, *Curr. Opin. Chem. Biol.* **2005**, *9*, 561–569.
72. J. P. Pellois; T. W. Muir, *Curr. Opin. Chem. Biol.* **2006**, *10*, 487–491.
73. A. J. Link; M. L. Mock; D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.
74. T. Hohsaka; M. Sisido, *Curr. Opin. Chem. Biol.* **2002**, *6*, 809–815.
75. L. Wang; Z. Zhang; A. Brock; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56–61.
76. H. Liu; L. Wang; A. Brock; C. H. Wong; P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 1702–1703.
77. Z. Zhang; B. A. Smith; L. Wang; A. Brock; C. Cho; P. G. Schultz, *Biochemistry* **2003**, *42*, 6735–6746.
78. S. Ye; C. Köhrer; T. Huber; M. Kazmi; P. Sachdev; E. C. Yan; A. Bhagat; U. L. RajBhandary; T. P. Sakmar, *J. Biol. Chem.* **2008**, *283*, 1525–1533.
79. H. Zeng; J. Xie; P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5356–5359.
80. A. Deiters; T. A. Cropp; M. Mukherji; J. W. Chin; J. C. Anderson; P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783.
81. A. Deiters; T. A. Cropp; D. Summerer; M. Mukherji; P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5743–5745.
82. M. L. Tsao; F. Tian; P. G. Schultz, *ChemBioChem* **2005**, *6*, 2147–2149.
83. J. Wang; S. M. Schiller; P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 6849–6851.
84. J. Guo; J. Wang; J. S. Lee; P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 6399–6401.
85. J. Xie; L. Wang; N. Wu; A. Brock; G. Spraggon; P. G. Schultz, *Nat. Biotechnol.* **2004**, *22*, 1297–1301.
86. J. C. Jackson; J. T. Hammill; R. A. Mehl, *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166.
87. S. E. Cellitti; D. H. Jones; L. Lagpacan; X. Hao; Q. Zhang; H. Hu; S. M. Brittain; A. Brinker; J. Caldwell; B. Bursulaya; G. Spraggon; A. Brock; Y. Ryu; T. Uno; P. G. Schultz; B. H. Geierstanger, *J. Am. Chem. Soc.* **2008**, *130*, 9268–9281.
88. K. C. Schultz; L. Supekova; Y. Ryu; J. Xie; R. Perera; P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 13984–13985.
89. J. Wang; J. Xie; P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 8738–8739.
90. M. L. Tsao; D. Summerer; Y. Ryu; P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 4572–4573.
91. J. W. Chin; S. W. Santoro; A. B. Martin; D. S. King; L. Wang; P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027.
92. J. W. Chin; A. B. Martin; D. S. King; L. Wang; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020–11024.
93. J. W. Chin; P. G. Schultz, *ChemBioChem* **2002**, *3*, 1135–1137.
94. C. Schlieker; J. Weibezahn; H. Patzelt; P. Tessarz; C. Strub; K. Zeth; A. Erbse; J. Schneider-Mergener; J. W. Chin; P. G. Schultz; B. Bukau; A. Mogk, *Nat. Struct. Mol. Biol.* **2004**, *11*, 607–615.
95. L. Y. Huang; G. Umanah; M. Hauser; C. Son; B. Arshava; F. Naider; J. M. Becker, *Biochemistry* **2008**, *47*, 5638–5648.
96. N. Hino; Y. Okazaki; T. Kobayashi; A. Hayashi; K. Sakamoto; S. Yokoyama, *Nat. Methods* **2005**, *2*, 201–206.
97. A. Deiters; D. Groff; Y. Ryu; J. Xie; P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2728–2731.
98. E. A. Lemke; D. Summerer; B. H. Geierstanger; S. M. Brittain; P. G. Schultz, *Nat. Chem. Biol.* **2007**, *3*, 769–772.
99. M. Bose; D. Groff; J. Xie; E. Brustad; P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 388–389.
100. Z. Zhang; J. Gildersleeve; Y. Y. Yang; R. Xu; J. A. Loo; S. Uryu; C. H. Wong; P. G. Schultz, *Science* **2004**, *303*, 371–373.
101. R. Xu; S. R. Hanson; Z. Zhang; Y. Y. Yang; P. G. Schultz; C. H. Wong, *J. Am. Chem. Soc.* **2004**, *126*, 15654–15655.
102. J. Xie; L. Supekova; P. G. Schultz, *ACS Chem. Biol.* **2007**, *2*, 474–478.
103. H. Neumann; S. Y. Peak-Chew; J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234.
104. T. Mukai; T. Kobayashi; N. Hino; T. Yanagisawa; K. Sakamoto; S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822.
105. J. Xie; W. Liu; P. G. Schultz; *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 9239–9242.
106. H. S. Lee; P. G. Schultz, *J. Am. Chem. Soc.* **2008**, *130*, 13194–13195.
107. J. Grünewald; M. L. Tsao; R. Perera; L. Dong; F. Niessen; B. G. Wen; D. M. Kubitz; V. V. Smider; W. Ruf; M. Nasoff; R. A. Lerner; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11276–11280.

Biographical Sketches

Angela Parrish received her BS in Biology, magna cum laude with distinction in research, from Cornell University in 2003, where she was a Cornell Presidential Research Scholar and a Hughes Scholar. She has participated in biological research at the USDA, the Wellcome Trust Center at Cambridge University, and Georgetown University. She is currently a

graduate student with the Biology Department at the University of California, San Diego, where she is partially supported by the National Institutes of Health Cellular and Molecular Genetics Training Grant and the Chancellor's Fellowship. She is doing graduate research in Lei Wang's lab at the Salk Institute on new applications of unnatural amino acids.



Lei Wang received his BS from Peking University and Ph.D. from the University of California at Berkeley. He joined the Salk faculty in 2005, where he is currently the Frederick B. Rentschler assistant professor in Chemical Biology and Proteomics. His research interests focus on using genetically encoded unnatural amino acids to investigate biological processes *in vivo* and studying the evolution of the genetic code. He is a Beckman Young Investigator, a Searle Scholar, and a recipient of the NIH Director's New Innovator Award.

5.19 Nonribosomal Peptide Synthetases

Timothy J. Montavon and Steven D. Bruner, University of Florida, Gainesville, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

5.19.1 Introduction: Peptide-Based Natural Products

Peptidic natural products constitute a major class of secondary metabolites produced in nature. The family represents a vast array of biologically active compounds produced by diverse organisms ranging from bacteria to humans (**Figure 1**). Peptidic natural products are constructed by combinatorial arrangement of amino acids resulting in numerous chemical architectures. There are two general classes of peptide-based natural products. The first is synthesized by cellular ribosome machinery. The ribosome is a key player in the central dogma of

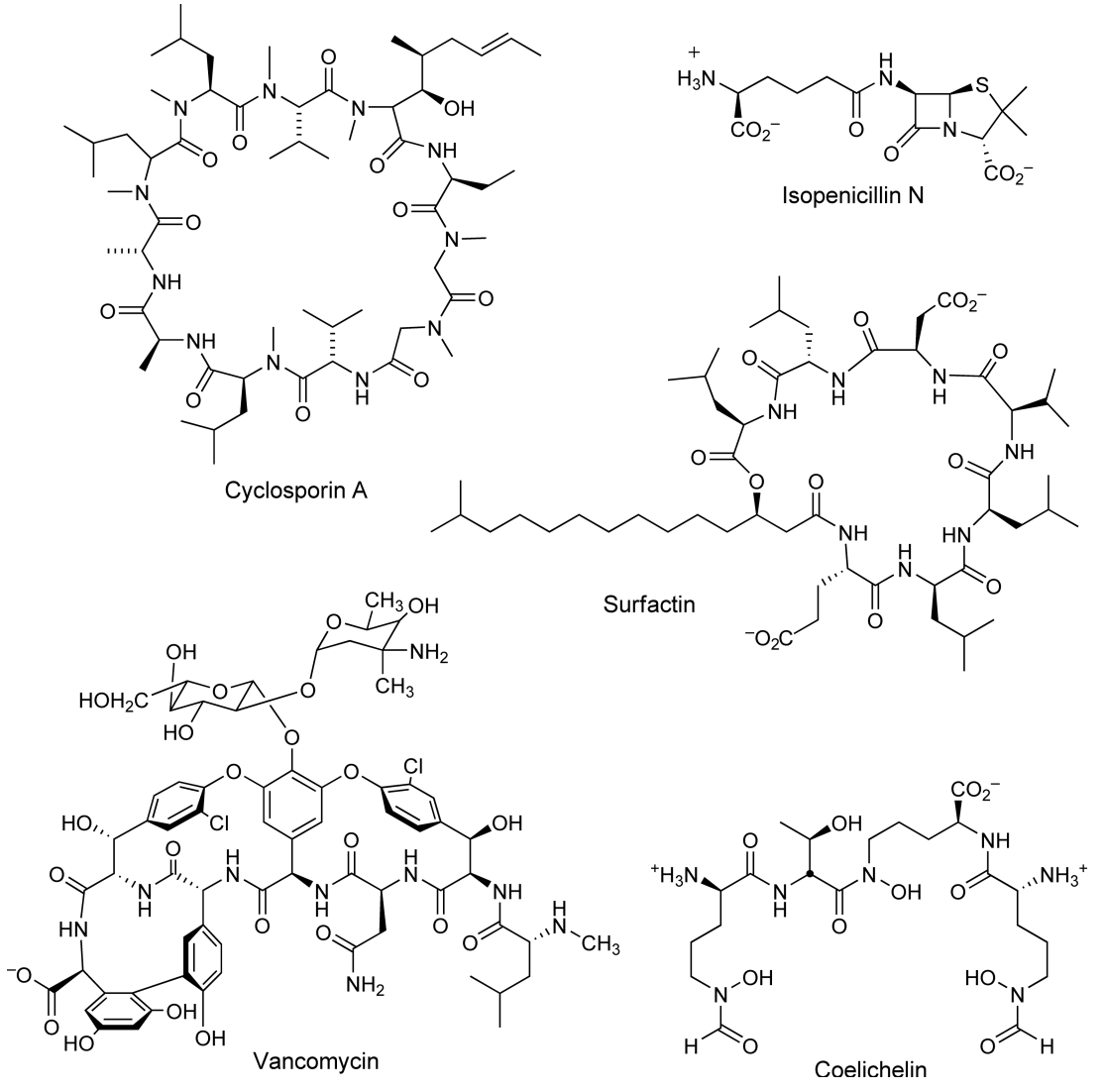


Figure 1 Examples of NRP natural products.

biology and has evolved complex mechanisms to produce amino acid polymers (proteins) with high fidelity.[1,2] Proteins generally contain on the order of 10^2–10^3 amino acids and the building block composition is limited to the 20 proteinogenic amino acids. Ribosomally synthesized secondary metabolites are commonly processed from larger protein precursors to yield the smaller biologically active peptides. These natural products include diverse molecules such as mammalian hormones and antibiotic peptides, and bacterial antibiotics including nisins and lantibiotics.[3] Although these metabolites can display complex architectures, their diversity is restricted by the limited building blocks available. The second class of peptide-based natural products is produced by a pathway that is distinct from RNA-templated ribosomal enzymology. In general, this class (nonribosomal peptide, NRP) displays greater structural diversity resulting from the use of nonproteinogenic amino acids, alternate linkages, and complex postsynthetic tailoring reactions.

5.19.2 Nonribosomal Peptide Natural Products

NRP natural products are short peptides produced by large multimodular enzymes.[4] The catalytic nonribosomal peptide synthetases (NRPSs) utilize a thiol-template mechanism that is functionally similar to the enzymology used by fatty acid and polyketide synthases.[5,6] The phosphopantetheine prosthetic group provides the NRPS with a flexible arm, about 20 Å in length, terminating with a thiol.[7] The prosthetic group, which is derived from coenzyme A, is added to small carrier domains of NRPSs by a phosphopantetheinyl transferase (PPTase).[8] Unlike the ribosome, which uses mRNA as a template for synthesis, the multidomain enzyme machinery itself acts as the template in NRP biosynthesis. The use of such a large template (~100 kDa of protein machinery per incorporated amino acid) puts a significant limit on the size of the final peptide product. Indeed, NRPS-derived peptides are, in general, <10 amino acids in length.

5.19.3 The Canonical Enzymology of NRPS Modules

The pathway to the clinically important antibiotic vancomycin provides a representative illustration of the overall strategy used to construct NRPs.[9] This pathway demonstrates how NRPS biosynthetic machinery is uniquely suited for the synthesis of architecturally complex natural products (**Figure 2**). The genes responsible for the biosynthesis of vancomycin (and other bacterial-derived NRPs) are clustered in a contiguous region of the genome. Included in the clusters are genes responsible for not only chemical synthesis but also regulation and export of the natural product, and self-resistance. Frequently, the first stage of NRP biosynthesis involves the production of nonproteinogenic amino acid building blocks. The vancomycin gene cluster, for example, includes genes for the biosynthesis of phenylglycines incorporated into the natural product. After the requisite building blocks have been assembled, they are linked together (along with proteinogenic amino acids) through amide bonds by the vancomycin NRPS to form a heptapeptide. This intermediate undergoes postsynthetic chemistry catalyzed by tailoring enzymes, which form biaryl and biaryl-ether cross-links and add carbohydrate moieties. For vancomycin, these tailoring steps give the natural product its cup-shaped structure, a key feature that allows it to bind the cell-wall biosynthetic precursor D-Ala-D-Ala leading to potent antibiotic activity.

NRPSs contain three core domains: adenylation (A), condensation (C), and peptidyl carrier protein (PCP). Together, these domains constitute the repeating module responsible for activation and incorporation of a single amino acid into the growing peptide chain (**Figure 3**). The basic chemical logic of NRPSs has been reviewed extensively in recent literature and will be summarized here.[4,10–12] The first step in oligomerization is the selection of a specific amino acid by the A domain. This enzyme plays a key role in determining specificity by activating an amino acid with ATP and loading the building block onto the assembly line. Using a chemical activation pathway that is shared with the aminoacyl-tRNA synthetases of ribosomal enzymology, A domains use ATP to convert the carboxylate of the amino acid into a mixed anhydride-AMP ester with concomitant release of pyrophosphate (PP_i).[13] The A domain then loads the activated amino acid onto the thiol-terminated phosphopantetheine arm attached to the PCP domain resulting in a covalent thioester adduct. After amino acids are loaded onto PCP domains, the C domain of

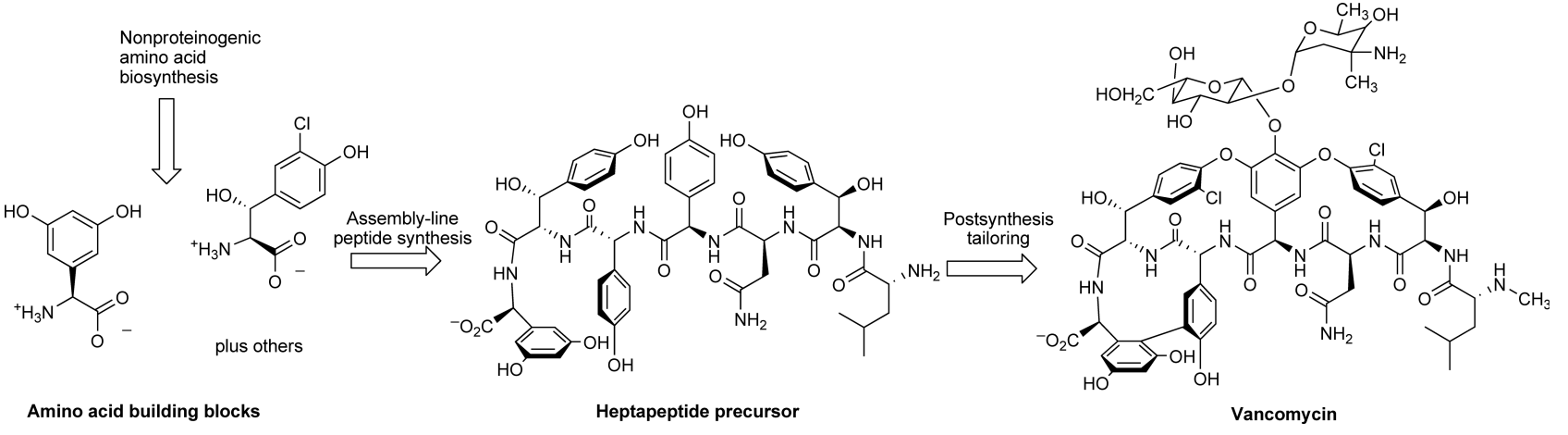


Figure 2 Summary of NRP natural product biosynthesis, exemplified by vancomycin.

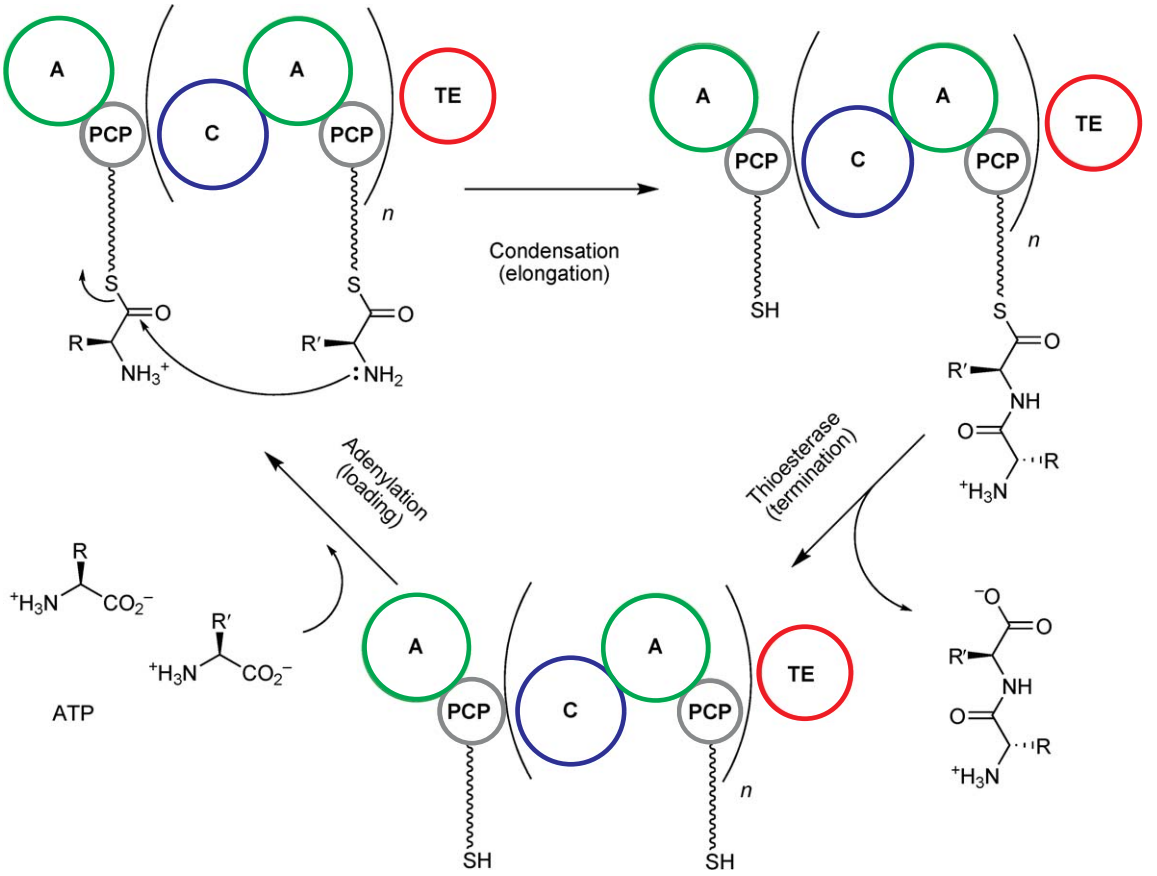


Figure 3 Summary of the steps in an NRPS assembly line. The example of the construction of a dipeptide is illustrated. R and R′ represent amino acid side chains.

the downstream module will catalyze the coupling of the amino group of the donor substrate with the upstream thioester, forming an amide bond and releasing a free thiol-terminated phosphopantetheine arm. At downstream modules, the oligomer is further lengthened according to a synthetic order that is colinear with the modules in the synthetase. At the conclusion of the peptide elongation process, the mature peptide remains tethered to the final module as a thioester. Release from the PCP domain is frequently catalyzed by a C-terminal thioesterase (TE) domain, hydrolyzing the bound intermediate to a carboxylic acid and regenerating the phosphopantetheine-functionalized synthetase. Owing to the modular, colinear organization of the synthesis, the overall process is frequently compared to an assembly line. The diverse structural architectures evident in NRPS natural products are an indication that departures from this general scheme are possible. Based on the structures of known secondary metabolites, diversity is obtained by combining a large precursor pool of building blocks with alternate linkages and enzymatic tailoring (Figure 1).

5.19.4 Classes of Nonribosomal Peptide Synthetases

NRPSs can be classified as linear (Type A), iterative (Type B), or nonlinear (Type C).[14] This classification is based upon the relationship between the amino acid sequence of the natural product and the modular organization of the synthetase.

5.19.4.1 Linear Synthetases

In linear NRPSs a product consisting of *n* amino acids is biosynthesized in an N- to C-terminal manner by the multidomain assembly line with a domain organization of A-PCP-(C-A-PCP)$_{n-1}$-TE. The initiation module of a linear NRPS lacks a C domain, while the following modules may include any required additional domains. After formation of the full-length peptide, the product is released from the assembly line by a termination domain. Thus, the number and order of amino acids in the peptide directly coincides with the number and order of synthetase modules. Many NRPs are biosynthesized in this manner, and characterized examples include the penicillin tripeptide precursor δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV, **Figure 4** (**a**)),[15] complestatin,[16] cyclosporin,[17] fengycin,[18] surfactin,[19] and tyrocidine.[20]

5.19.4.2 Iterative Synthetases

Similarly, iterative NRPSs operate in a linear fashion but utilize at least one domain or module multiple times for the synthesis of a single NRP product. Thus, peptides assembled by iterative synthetases contain short, repeating units of peptide building blocks. In such systems, the terminal PCP-TE (or infrequently PCP-C)[21] didomain is responsible for both condensation of the repeating peptide units and chain release from the assembly line. NRPs biosynthesized in this manner include enniatin,[22] enterobactin,[23] bacillibactin,[24] gramicidin S,[25] and the depsipeptides valinomycin[26] and cereulide.[27] Of these examples, condensation of the precursor peptides for both enterobactin and gramicidin S has been extensively studied and will be discussed in detail.

Enterobactin, a siderophore produced by *Escherichia coli*, is constructed by iterative condensations of 2,3-dihydroxybenzyl-serine followed by macrocyclization of the resulting trimer (**Figure 4(b)**). This natural product is biosynthesized by the enterobactin synthetase, which contains two NRPS modules.[23] The initiation module is composed of EntB, a free-standing 2,3-dihydroxybenzoic acid (Dhb) A domain, and the PCP domain EntE. The second NRPS module (EntF) contains a C domain, a serine-specific A domain, a PCP domain, and a C-terminal TE domain. After Dhb and serine are loaded onto their respective holo-PCP domains, the C domain of EntF catalyzes amide bond formation resulting in a Dhb-Ser-*S*-PCP thioester, which is transferred to the active site serine of the TE domain. The TE domain then catalyzes two successive oligomerizations of Dhb-Ser units and macrocyclization of the resulting TE-bound trimer to yield enterobactin (**Figure 1**). Protein-bound intermediates in enterobactin biosynthesis have been characterized by high-resolution ESI-FT mass spectral analysis confirming the pathway.[28]

In iterative NRPSs, there are two possible mechanisms for the oligomerization of repeating precursor units (see **Figure 5**). In a 'forward' mechanism, a heteroatom of the TE-bound peptide acts as a nucleophile, attacking the thioester carbon of the upstream peptidyl-*S*-PCP intermediate. The work of Walsh and coworkers has provided evidence that enterobactin is made in this manner as only TE-bound oligomers were observed under *in vitro* turnover conditions.[28] Alternatively, in a 'backward' mechanism, the PCP-bound peptide acts as a nucleophile by adding to the ester carbon of a TE-bound peptide. The resulting oligomer must be transferred back to the TE domain before elongation continues or chain termination occurs. Recent biochemical experiments performed by Marahiel and coworkers suggest that this mechanism is operative in the biosynthesis of gramicidin S.[29] In these studies, a recombinant PCP-TE didomain obtained from the GrsB module of gramicidin S synthetase was incubated with a pentapeptidyl-*S*-*N*-acetylcysteamine (-SNAC) substrate mimic. Analysis of the resulting product mixture by mass spectrometry showed the formation of gramicidin S along with a 15-membered macrocycle, a hydrolyzed pentamer, and a linear decapeptidyl-SNAC. The formation of this decapeptidyl-SNAC suggests that the TE domain of GrsB may act in a 'backward' fashion as this product could not be formed if the amino group of a TE-bound pentapeptide served as the nucleophile in the dimerization. However, it is unclear whether the decapeptidyl-SNAC is an intermediate or shunt product of the reaction catalyzed by the GrsB TE domain.

5.19.4.3 Nonlinear Synthetases

A number of NRPs have chemical structures that cannot be predicted from direct, linear (or iterative) interpretation of the synthetase sequence. These metabolites are produced by nonlinear synthetases, which

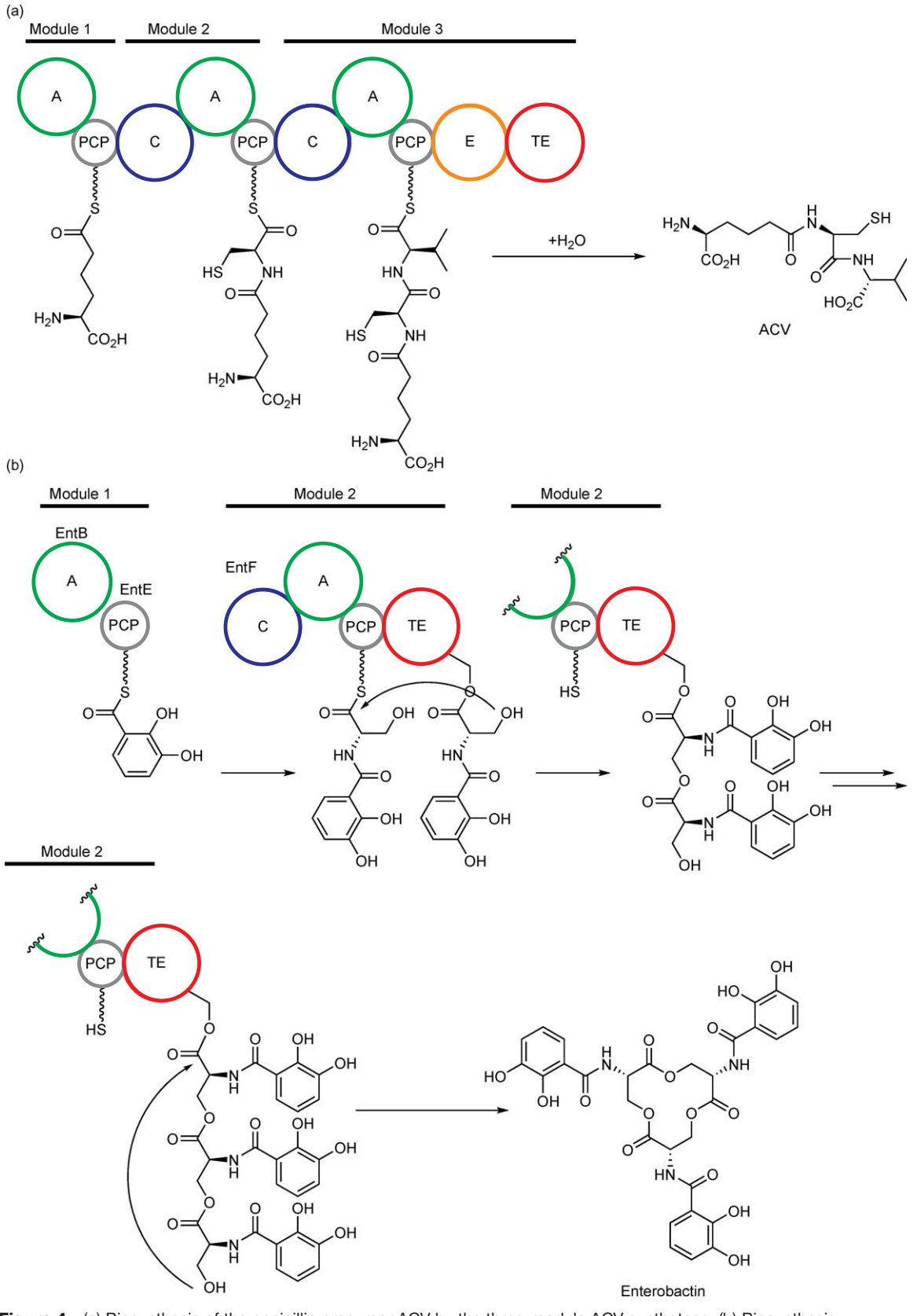


Figure 4 (a) Biosynthesis of the penicillin precursor ACV by the three-module ACV synthetase. (b) Biosynthesis of enterobactin by the iterative enterobactin synthetase.

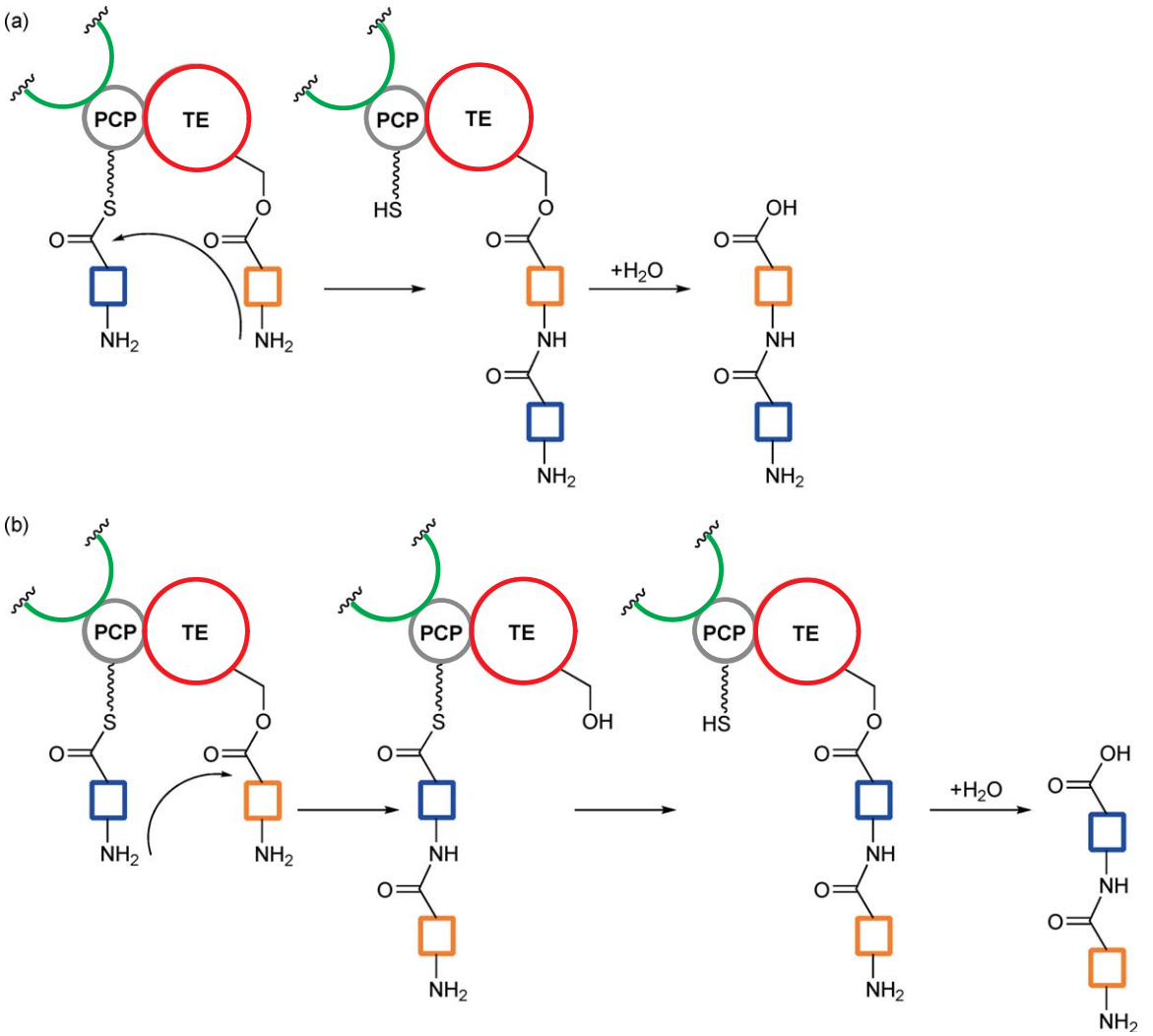


Figure 5 Iterative synthetases. (a) 'Forward' mechanism for oligomerization of peptides by an iterative synthetase. (b) 'Backward' mechanism for oligomerization of peptides by an iterative synthetase. Boxes represent peptidyl chains.

have at least one permutation from the canonical $(C\text{-}A\text{-}PCP)_n$ domain organization. Such systems often display internal cyclizations of peptide building blocks, domain skipping, or the incorporation of nontethered small molecules into the peptide backbone of the natural product. Three examples are described below to illustrate nonlinear NRPS biosynthetic pathways.

5.19.4.3.1 Domain skipping: biosynthesis of myxochromides

Myxochromides are lipopeptides produced by various myxobacterial species. In 2005, Müller and coworkers reported the structural characterization of the hybrid PKS/NRPS cyclic pentapeptides myxochromides S_{1-3}, which differ from each other in the lengths of their polyketide starter units.[30] Sequence analysis of the gene cluster responsible for the biosynthesis of these compounds in *Stigmatella aurantiaca* revealed one iterative polyketide synthase, $MchA_s$, and two enzymes $MchB_s$ and $MchC_s$ which are responsible for the NRP core of the natural products. Together, they form a six-module synthetase in which one module is skipped during the production of myxochromides S_{1-3} (**Figure 6(a)**).

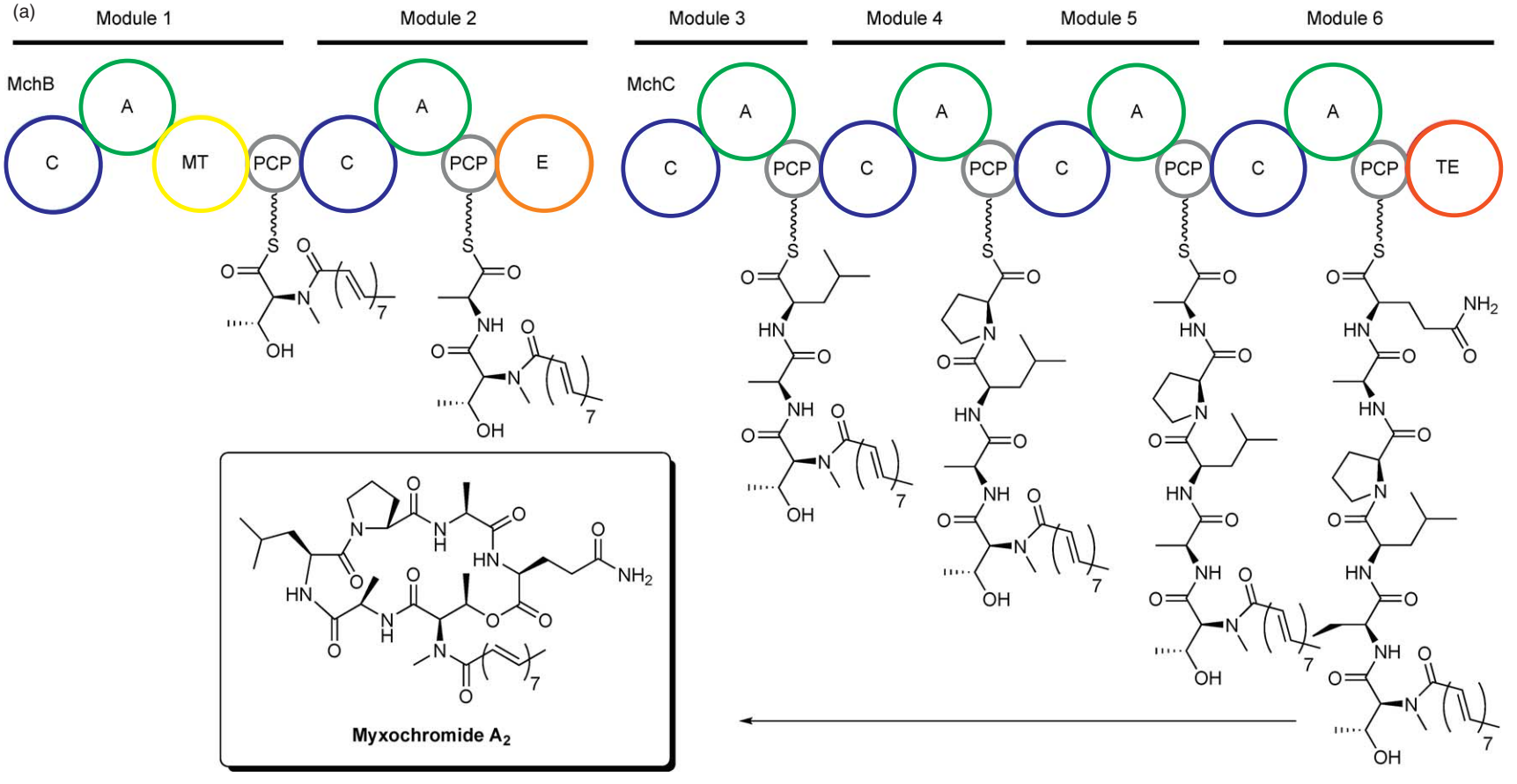


Figure 6 (Continued)

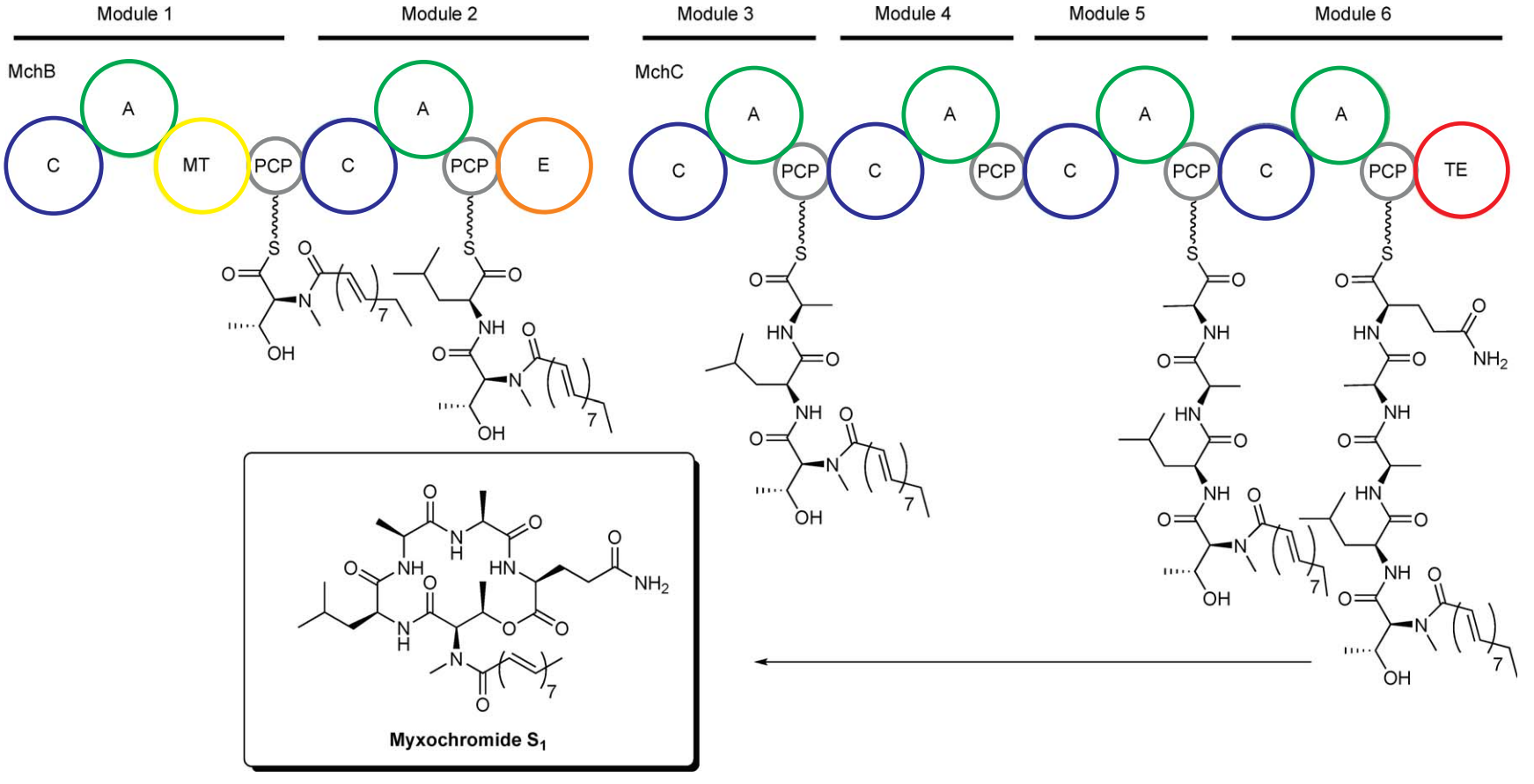


Figure 6 (Continued)

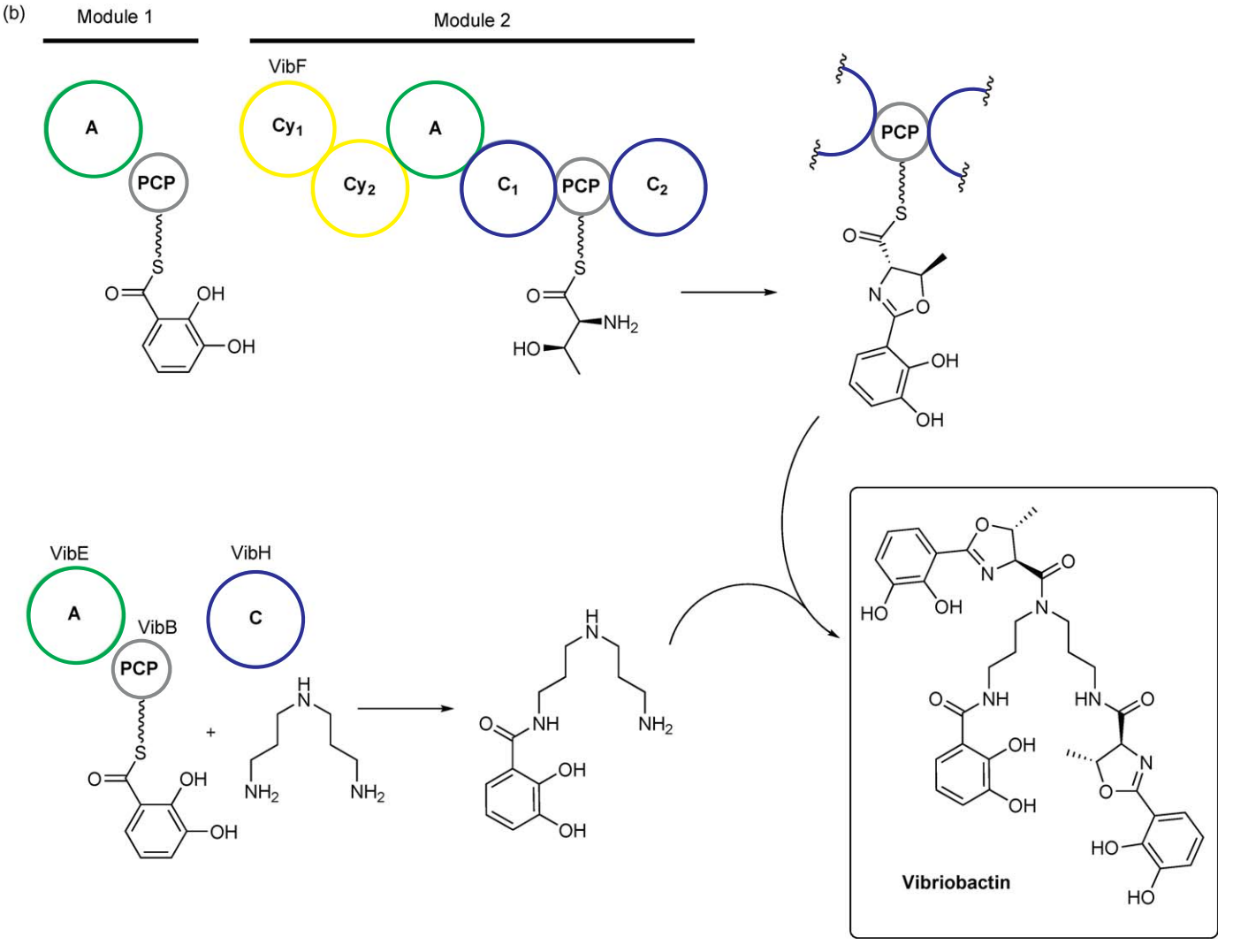


Figure 6 Nonlinear synthetases. (a) Organization of the NRPS responsible for myxochromide A_2 and myxochromide S_1 biosynthesis. The myxochromide S_1 synthetase contains an inactive module 4. (b) Biosynthesis of vibriobactin by the two-module vibriobactin synthetase.

The structurally related myxochromides A_{1-3} are cyclic hexapeptides produced by several *Myxococcus* species.[31] These examples contain a proline residue, which is not present in myxochromides S_{1-3}, as the fourth amino acid in their peptide core. The NRPSs responsible for myxochromides A and S biosynthesis have exactly the same module and domain organization; thus, the fourth module of the myxochromide S synthetase must be skipped to account for the natural product. Biochemical experiments revealed that the A domain of this module activates L-proline, but the adjacent PCP domain cannot be phosphopantetheinylated by a PPTase.[31] These results suggest that the C domain of module 5 reacts directly with the tripeptide intermediate bound to the PCP domain of module 3 in myxochromide S biosynthesis. A similar example of domain skipping has been noted in the biosynthesis of the mannopeptimycins.[32]

5.19.4.3.2 Vibriobactin biosynthesis

Another example of a nonlinear assembly line is the two-module, four-component vibriobactin synthetase from *Vibrio cholerae* (**Figure 6(b)**).[33,34] The siderophore vibriobactin consists of a central norspermidine core linked to a Dhb unit and two 2-dihydroxyphenyl-5-methyloxazolinyl-4-carbonyl (Dhp-Oxa) groups through amide bonds. Several interesting features of vibriobactin synthetase include the use of the free, nontethered small-molecule norspermidine in the condensation reactions, the utilization of the stand-alone C domain VibH, and the presence of a catalytically inactive C domain in the VibF module.

The initiation module of vibriobactin synthetase consists of the Dhb-specific A domain VibE, the PCP domain VibB, and the stand-alone C domain VibH, which catalyzes amide bond formation between VibB-bound Dhb and norspermidine. The second module of this system is VibF, which contains two cyclization domains (Cy), a threonine-specific A domain, a PCP domain, and two condensation domains in the order Cy_1-Cy_2-A-C_1-PCP-C_2. The Cy_2 domain of VibF catalyzes amide bond formation between VibB-bound Dhb and PCP-bound threonine. Cyclization and dehydration of the resulting intermediate by Cy_1 of VibF yields PCP-bound Dhp-Oxa.[35] Next, the C_2 domain of VibF catalyzes amide bond formation between DHB-norspermidine and DHP-Oxa, and the domains of VibF are reused in an iterative fashion to complete the synthesis of vibriobactin. The C_1 domain of VibF is not catalytically active, but the multidomain enzyme does form a catalytically important homodimer in solution. Although this condensation domain is not needed for vibriobactin production, dimer formation likely reinforces active confirmations between Cy_1-Cy_2-A and PCP-C_2 of VibF.[36]

5.19.4.3.3 Iterative module use: coelichelin biosynthesis

Several recently characterized siderophores including coelichelin[37] and the fuscachelins[38] are biosynthesized by nonlinear synthetases that utilize modules in an iterative fashion and display domain skipping. The gene cluster for coelichelin biosynthesis in *Streptomyces coelicolor* only encodes for a single, three-module NRPS termed CchH.[39] This NRPS is unusual because it lacks a C-terminal TE domain and utilizes both domain skipping and the iterative action of a C domain. Sequence analysis of CchH reveals that the A domains of modules 1, 2, and 3 should activate L-*δ*-*N*-formyl-*δ*-*N*-hydroxyornithine (L-hfOrn), L-threonine, and L-*δ*-*N*-hydroxyornithine (L-hOrn), respectively.[40]

In the model proposed for the biosynthesis of coelichelin,[37] modules 1 and 2 form an hfOrn-Thr-*S*-PCP thioester (**Figure 7**). Subsequent peptide bond formation between the *δ*-amino group of hOrn and this dipeptide yields D-hfOrn-D-*allo*-Thr-L-hOrn-*S*-PCP. At this stage, module 1 of CchH forms D-hfOrn-*S*-PCP, which undergoes condensation with the *α*-amino group of hOrn in the above tripeptide. The resulting tetrapeptide is then released from the NRPS by the action of the external TE CchJ, which likely displays high specificity for tetrapeptide hydrolysis to prevent premature chain release. The factors that control CchH's iterative use of module 1 and subsequent domain skipping to form only a single tetrapeptide product are not known.

An alternative biosynthetic model involving initial formation of a D-*allo*-Thr-L-hOrn-*S*-PCP dipeptide has also been put forth.[41] After the formation of this dipeptide, CchH would then catalyze the consecutive addition of two D-hfOrn residues to both amino groups of the PCP-bound dipeptide followed by hydrolysis of the resulting tetrapeptide to yield coelichelin.

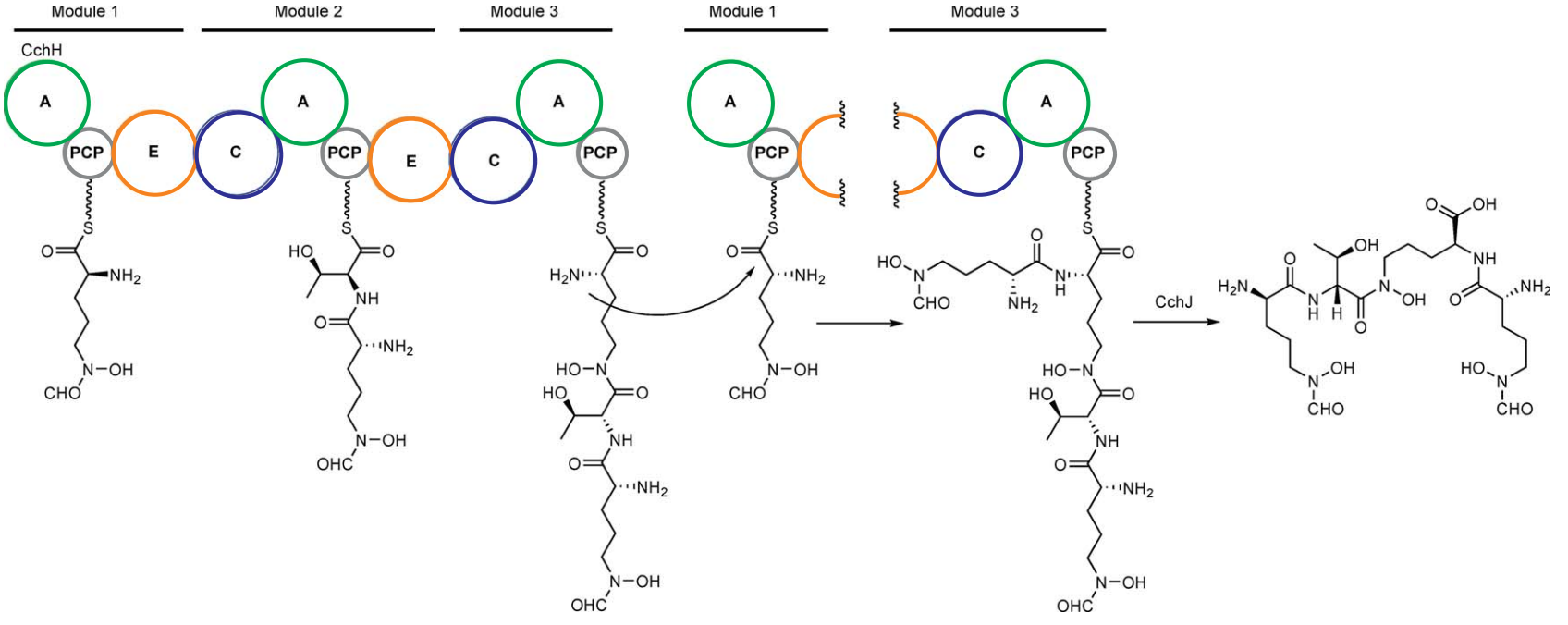


Figure 7 Biosynthesis of coelichelin by the nonlinear coelichelin synthetase. Chain release is catalyzed by the *trans*-acting thioesterase, CchJ.

5.19.5 Noncanonical Features of NRPSs

5.19.5.1 Initiation Modules

Linear NRPSs frequently contain an A-PCP initiation module linked *in cis* to the first elongation module as in the vancomycin[8] and the penicillin precursor ACV synthetases.[15] When D-amino acids are used as starter units, two initiation strategies can be employed. In the first method, a stand-alone A-PCP-epimerization (E) module is used for initiation as in tyrocidine biosynthesis.[20] Alternatively, if the D-amino acid is produced by an external racemase, it is directly activated by a stand-alone A domain. This A domain and the corresponding stand-alone PCP domain constitute an initiation module, as seen in leinamycin biosynthesis.[42] When N-acylated amino acids are incorporated into NRPs, the initiation module begins with a starter C domain, as described in detail below, and has a C-A-PCP organization. An initiation module of this type is used in fengycin[18] and surfactin[19] synthetases. For these and related NRPS lipopeptides, the source of the acyl donor is not always obvious and may come from a loaded carrier domain from fatty acid biosynthesis. Similarly, synthetases which incorporate N-formylated amino acid residues as starter units, such as anabaenopeptilides[43] and linear gramicidin,[44] contain formylation (F) domains incorporated into an F-A-PCP initiation module. In addition, if aryl acids are used as starter units, as in enterobactin biosynthesis,[23] the initiation module contains an acyl-CoA-ligase homologous to NRPS A domains. As a final example, bacterial NRPSs utilizing α-hydroxyacids as starter units often have initiation modules with an A-KR (ketoreductase)-PCP domain organization. In these examples, the A domain activates an α-ketoacid that is subsequently reduced to a PCP-bound α-hydroxythioester. Modules of this type are found in the cereulide[27] and valinomycin[26] synthetases. An alternative method for initiation module classification based on the organizational pattern of domains has been suggested by Shen and coworkers.[42]

5.19.5.2 Coupling of Building Blocks in NRPSs

C domains typically catalyze amide bond formation between the α-amine of a downstream PCP-bound amino acid acceptor and the thioester carbon of an upstream PCP-bound amino-, peptidyl-, or β-hydroxy acid donor. A recent phylogenetic analysis of condensation domains found in the genomes of bacterial species containing NRPSs revealed that there are several subtypes of condensation domains.[45] $^{L}C_{L}$ and $^{D}C_{L}$ condensation domains form peptide bonds between L-aminoacyl-*S*-PCP acceptors and L or D (respectively) aminoacyl/peptidyl-*S*-PCP donors.[46] Dual E/C domains are also known to exist in NRPS assembly lines. C domains of this type directly follow a C-A-PCP module that activates an L-amino acid. After the formation of the aminoacyl-*S*-PCP acceptor on the module containing the E/C domain has occurred, this dual domain catalyzes the epimerization of the C_{α}-carbon on the upstream donor followed by amide bond formation.[47] Thus, such domains function as $^{D}C_{L}$ catalysts with epimerase activity. Domains of this type are utilized in the biosynthesis of arthrofactin,[48] ramoplanin,[49] syringomycin,[50] and syringopeptin.[51] Starter C domains differ from the above subtypes in that they utilize β-hydroxy acids as PCP-bound donor substrates. Such domains are always the first C domain present in NRPSs that use this class of substrates, and they are found in the synthetases responsible for the construction of enterobactin,[23] surfactin,[19] fengycin,[18] arthrofactin,[48] pristinamycin,[52] and other NRP-containing natural products.

5.19.5.2.1 Formation of ester linkages

C domains can display functions that deviate from typical amide bond formation. Several C domains are postulated to act as ester synthases, catalyzing ester formation instead of amide formation. NRPS modules containing C domains that display this activity are present in the biosynthetic pathways for the kutznerides,[53] cryptophycins,[54] cereulide,[27] valinomycin,[26] hectochlorin,[55] and beauvericin.[56] Each of these C domains likely utilizes a PCP-bound α-hydroxyl acceptor in the condensation reaction. Another NRPS C domain that catalyzes ester bond formation is involved in the biosynthesis of the polyketide-derived mycotoxins known as the fumonisins.[57] Du and coworkers have shown that a recombinant PCP-C didomain of an NRPS involved in the biosynthetic pathway of the fumonisins can catalyze ester bond formation between hydroxyfumonisins and the *N*-acetylcysteamine thioester of tricarballylic acid, even though PCP-bound tricarballylic acid is not

the predicted substrate for this C domain.[58] This system differs from the ester synthases described above in that it utilizes the soluble (non-PCP-bound) hydroxyfumonisins as acceptor substrates in the condensation reaction.

5.19.5.2.2 C domains that catalyzed the formation of multiple amide bonds

Another set of unusual C domains include those that catalyze the formation of more than one amide bond on an acceptor substrate containing multiple amine moieties. For example, the C domain of the NRPS module FscI in fuscachelin biosynthesis likely catalyzes amide bond formation at both the α- and δ-amines of a PCP-bound L-hOrn intermediate.[38] Other domains displaying similar activity include the condensation domain of MxcG, the third C domain of CchH, and the second C domain of VibF from the biosynthetic pathways for myxochelin,[59] coelichelin,[37] and vibriobactin,[35] respectively.

5.19.5.2.3 Transglutaminases

Transglutaminases (TGases) are a class of enzymes that catalyzes cross-linking of peptide chains through isopeptide bond formation between amine donors and the γ-carboxyamide group of glutamine side chains.[60] This class of enzymes uses the Cys of a Cys-His-Asp catalytic triad to form acyl-enzyme intermediates prior to amide-bond formation. Sequence analysis of the gene cluster encoding the biosynthesis of the hybrid polyketide-NRP andrimid suggests that the stand-alone TGase AdmF is utilized for isopeptide bond formation between the β-phenylalanine and octa-2,4,6-trienoic acid subunits of this natural product.[61] Biochemical experiments have revealed that the TGase AdmF indeed catalyzes formation of an acyl-enzyme intermediate with the polyketide donor followed by isopeptide bond formation with the β-phenylalanine acceptor present in andrimid.[62] This example represents a novel mode of peptide bond formation in NRPS systems.

5.19.5.3 Chain Termination Strategies

Once assembly of a mature peptide has been completed by the NRPS, the product remains covalently linked to the PCP domain of the last module as a thioester. Release into solution from the assembly line is accomplished by a variety of enzyme-catalyzed reactions as described below (**Figure 8**).

5.19.5.3.1 Thioesterase-catalyzed chain release

Direct nucleophilic cleavage of the thioester is a common mechanism employed to generate the free peptide. TE domains present at the C-terminus of the last NRPS module catalyze this reaction. Such domains usually contain a Ser-His-Asp catalytic triad common to the hydrolase class of enzymes. However, known exceptions include the TE domains utilized in mycobactin[63] and pyochelin[64] biosynthesis, where the Ser of the triad is replaced with a Cys. In the TE-catalyzed reaction, the full-length peptidyl precursor is transferred to the catalytic serine resulting in an acyl-*O*-TE intermediate. The product can then be released from the assembly line through hydrolysis, as in ACV[15] and vancomycin[9] biosynthesis, or by intramolecular macrocyclization resulting in the formation of a lactone, lactam, or thiolactone product. Macrocyclization is the manner of chain release in most characterized NRPS systems.[65] In iterative systems such as thiocoraline,[66] gramicidin S,[25] and enterobactin[23] synthetases, the TE domain is responsible for both oligomerization of the requisite precursor peptides and macrocyclization of the mature peptide.

5.19.5.3.2 Alternative chain release through reduction

Several characterized NRPSs utilize alternative methods for chain termination. In some synthetases, the TE domain of the final module is replaced by an NAD(P)H-dependent reductase domain. Reduction of a peptidyl-*S*-PCP substrate through a two-electron reaction leads to the formation of a transient aldehyde, which is subsequently converted into a cyclic imine or hemiaminal through intramolecular cyclization. This two-electron reaction is utilized in the biosynthesis of nostocyclopeptides,[67,68] the saframycins,[69,70] and anthramycin.[71] Alternatively, a four-electron reduction to the primary alcohol is observed in the biosynthesis of mycobacterial peptidolipids,[72] linear gramicidin,[44] the myxalamides,[73] lyngbyatoxin,[74] and myxochelin A.[75,76] An alternative four-electron reduction pathway involving aldehyde formation, transamination, and reduction to a primary amine occurs in the biosynthesis of myxochelin B.[76]

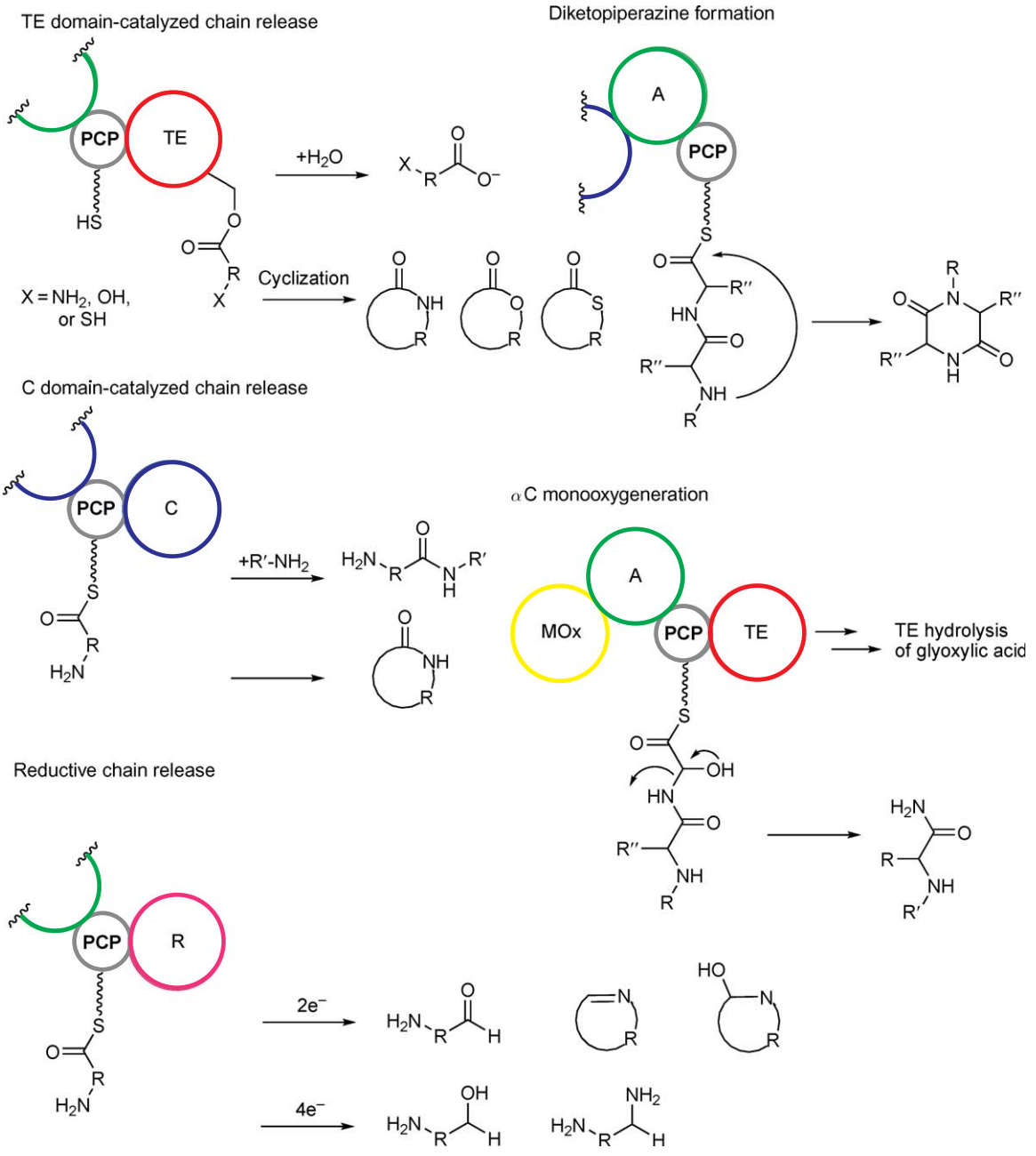


Figure 8 Summary of chain release strategies utilized by NRPS assemblies. The terminal domains involved in the chemistry are illustrated. The domain abbreviations are defined in the text. In general, for chemical structures, R and R′ represent peptide components and R″ represents various amino acid side chains.

5.19.5.3.3 Condensation domains as chain termination catalysts

In a limited number of NRPSs, the final module terminates in a specialized C domain that catalyzes chain release through amide bond formation. Modules of this type are found in the synthetases involved in the biosynthesis of enniatin,[22] vibriobactin,[33] cyclosporin,[17] HC-toxin,[77] and PF1032A.[78] Unlike TE termination, this method of chain release does not utilize an acyl-ester intermediate. Most likely, the chain termination precursor is presented to the C domain as an aminoacyl-*S*-PCP substrate. Most of these specialized C domains

catalyze intramolecular amide bond formation between the α-amino group of the N-terminal amino acid and the carbonyl carbon of the C-terminal amino acid in the requisite peptide precursor. As an example, the final C domain of the cyclosporin synthetase in *Tolypocladium niveum* likely catalyzes bond formation between the amino group of the first D-Ala and the carbonyl carbon of the terminal L-Ala in the peptide precursor, although no direct biochemical evidence for this reaction exists. Conversely, it has been shown that the C_2 domain of VibF in the vibriobactin synthetase catalyzes chain release of vibriobactin in an intermolecular manner (see **Figure 6(b)**). In such a reaction, peptide bond formation occurs between the PCP-bound Dhp-Oxa and the soluble norspermidine intermediate, resulting in vibriobactin production.

5.19.5.3.4 Diketopiperazine formation

Chain termination can also be realized through intramolecular diketopiperazine formation. This strategy is used in the biosynthesis of the ergot alkaloid ergotamine.[79,80] The ergotamine synthetase in *Claviceps purpurea* contains two subunits, LPS1: which activates Ala, Phe, and Pro, and LPS2, which activates D-lysergic acid. This synthetase lacks a TE domain, suggesting that diketopiperazine formation is a spontaneous process.[81] A majority of the NRP-derived diketopiperazines characterized to date contain a proline in the second position. Conformational constraints of the precursor induced by proline residues are likely important for the cyclization of such compounds. The formation of phenylalanylprolyl diketopiperazine by truncated constructs of both gramicidin S and tyrocidine synthetases corroborates this suggestion.[82,83] Conversely, the diketopiperazines gliotoxin and cyclomarazine do not contain Pro residues. Gliotoxin is produced by a three-module A_1-PCP_1-C_1-A_2-PCP_2-C_2-PCP_3 synthetase.[84] Biochemical studies have revealed that the unusual C_2-PCP_3 didomain is not necessary for diketopiperazine production.[85] Conversely, the cyclomarazines from *Salinispora arenicola* are produced as possible shunt products in the biosynthesis of the cyclic heptapeptide cyclomarins.[86] The factors which control diketopiperazine formation in these systems are not well understood.

5.19.5.3.5 Oxidative chain termination

An unusual method for chain termination is utilized in the biosynthesis of myxothiazol and melithiazol. The domains in the terminal modules of these synthetases are arranged in the following order: C-A-MOx-A-PCP-TE (MOx = monooxygenase domain).[87] In the termination modules of these synthetases, the MOx domain likely catalyzes hydroxylation of the α-carbon of the C-terminal glycine residue of the requisite peptidyl-*S*-PCP. The resulting intermediate then undergoes spontaneous conversion into PCP-bound glyoxylic acid and a terminal amide. This sequence yields the final product in the case of myxothiazol, whereas the terminal amide is transformed to a methyl ester to complete melithiazol biosynthesis.

5.19.6 Additional Enzyme Domains that Function in the NRPS Assembly Line

Multidomain synthetases can include specialized domains to modify the amino acids of peptide intermediates during chain elongation. The chemistry carried out by these domains introduces specific structural motifs, which are often important for biological activity, into the peptide natural product. A summary of additional NRPS enzyme chemistry is described below and illustrated in **Figure 9**.

5.19.6.1 Methyl Transferases

Many NRPs such as cyclosporin,[17] complestatin,[16] actinomycin,[88] and chondramide[89] contain *N*-methyl amides. *N*-Methyl transferase (*N*-MT) domains utilize *S*-adenosylmethionine (SAM) as a cofactor to catalyze the transfer of the methyl group from SAM to the α-amine of an aminoacyl-*S*-PCP substrate. The presence of *N*-methylamides in NRPs is believed to protect the peptide from proteolysis. Interestingly, *N*-MT domains are incorporated into the A domains of C-A-MT-PCP modules, between two of the core motifs (A8 and A9). MT domains contain three sequence motifs important for catalysis.[90,91] *O*-Methyl transferase domains are also found in NRPSs and likewise use the SAM cofactor. For instance, cryptophycin[54] and anabaenopeptilide[43] synthetases contain *O*-MT domains for the methylation of tyrosine side chains. These *O*-MT domains lack one of the three core motifs described for *N*-MT domains.[92]

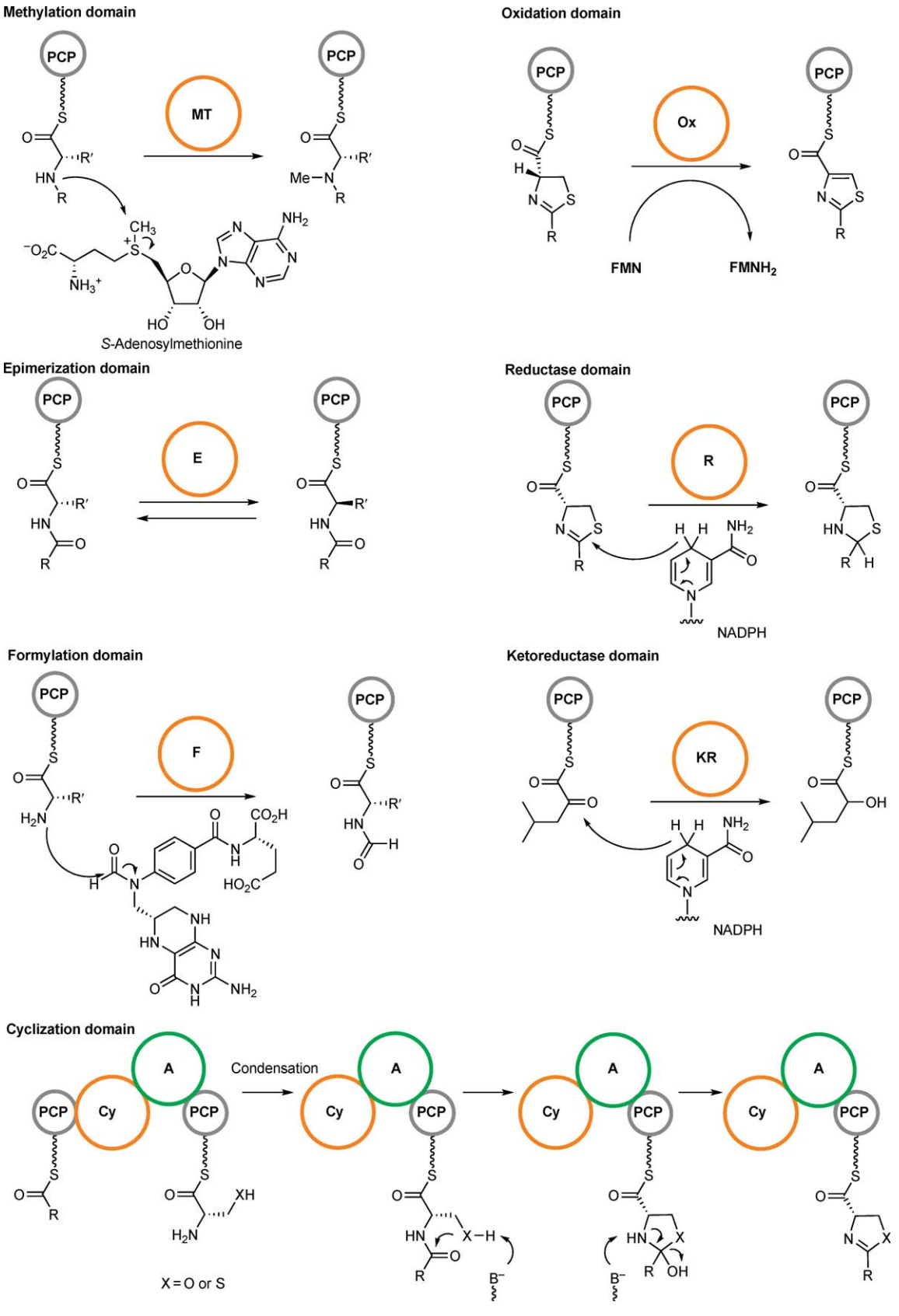


Figure 9 Summary of enzyme domains (orange) embedded in an NRPS module that act on intermediates in the assembly line peptide oligomerization. The variable group R represents a peptide component and R′ represents side chains.

Two additional types of methyltransferase domains have been identified in NRPSs. Yersiniabactin synthetase contains a carbon-MT domain within a Cy-MT-PCP-TE module.[93] This domain methylates at the α-carbon of an intermediate thiazoline ring. The melithiazol synthetase utilizes an MT domain *in trans* to form a methyl ester at the C-terminus of the natural product.[87] Recent *in silico* analysis of MT domains from secondary metabolite biosynthetic pathways has revealed the boundaries of these domains and suggests that they are typically approximately 200 amino acids in length, much shorter than previously thought. The results of this study by Mohanty and coworkers allow for the accurate prediction of N-, C-, or *O*-MT activity through sequence analysis.[94]

5.19.6.2 Epimerization Domains

A significant percentage of NRPs studied to date contain D-amino acids within their structure. Occasionally, as in cyclosporin[95] and leinamycin,[42] D-amino acids are formed using external PLP-dependent racemases before being utilized by NRPS assembly lines. However, in most cases, D-amino acids are formed on NRPS assembly lines through the action of an epimerization domain within a C-A-PCP-E module. E domains catalyze the epimerization of the α-carbon of aminoacyl/peptidyl-*S*-PCPs; they are approximately 470 amino acids in length and contain the HHXXXDG motif conserved among C domains.[96] Rapid-quench kinetic experiments have revealed that E domains form an equilibrated mixture of D and L PCP-bound intermediates.[97] The donor site of the downstream C domain displays enantioselectivity, allowing only the D-enantiomer of the upstream aminoacyl or peptidyl-*S*-PCP to react with the downstream aminoacyl-*S*-PCP acceptor.[98] As noted above, some synthetases utilize dual E/C domains. In addition to the first conserved His sequence shared with C domains, these domains contain an HH[I/L]XXXXGD motif at the N-terminus of the domain[47] and act as dual epimerization and $^{D}C_{L}$ catalysts.

5.19.6.3 Heterocyclization Domains

Several NRPs such as bacitracins,[99] vibriobactin,[33] and yersiniabactin[100] contain oxazoline or thiazoline rings, which are important for biological activity. Heterocyclization (Cy) domains are evolutionarily specialized C domains capable of catalyzing both amide bond formation and heterocyclization of Cys, Ser, or Thr side chains onto the growing peptide backbone. In such domains, the conserved HHXXXXDG motif of normal condensation domains is replaced with a DXXXXD sequence, in which the aspartate residues are important for both peptide bond formation and heterocyclization.[101] Recent mutational work on the Cy domain of bacitracin synthetase has revealed that peptide bond formation between the aminoacyl PCP-bound donor and the upstream acceptor is the first catalytic step.[102] Next, the Cy domain catalyzes ring formation between the C-terminal carbonyl carbon of the donor and the thiol or hydroxyl side chain of the acceptor followed by dehydration yielding thiazolines or oxazolines, respectively. As discussed earlier, vibriobactin synthetase utilizes an alternative logic and requires two heterocyclization domains for full activity.

5.19.6.4 Oxidative/Reductive Tailoring

The oxidation state of thiazolines and oxazolines can be adjusted by additional tailoring enzymes.[103] For instance, oxidation domains (Ox) composed of approximately 250 amino acids utilize the cofactor FMN (flavin mononucleotide) to form aromatic oxazoles and thiazoles from oxazolines and thiazolines, respectively. Such domains are likely utilized in the biosynthesis of the disorazoles,[104] diazonimides,[105] bleomycin,[106] and epothiolone.[107] The typical domain organization for a synthetase containing an oxidation domain is Cy-A-PCP-Ox; however, in myxothiazol biosynthesis one oxidation domain is incorporated into an A domain.[108] Alternatively, NRPSs can utilize NAD(P)H reductase domains to convert thiazolines and oxazolines into thiazolidines and oxazolidines, respectively. For instance, PchG is a reductase domain from the pyochelin biosynthetic pathway that acts *in trans* to reduce a thiazolyinyl-*S*-PCP-bound intermediate to the corresponding thiazolidynyl-*S*-PCP.[109]

5.19.6.5 Ketoreductase Domains

Recently, bacterial NRPS modules with the organization of A-KR-PCP have been discovered in the valinomycin and cereulide synthetases.[26,27] The A domains of these modules selectively activate α-keto acids. After the resulting adenylate is transferred to the PCP domain, the α-ketoacyl-*S*-PCP intermediate is reduced to a PCP-bound, α-hydroxythioester by the KR domain. These domains use NAD(P)H as a cofactor and are inserted into A domains between two conserved core motifs analogous to MT domains. Their substrate specificity differs from that of polyketide synthase KR domains, which reduce β-ketoacyl substrates. Similar fungal NRPSs, such as beauvericin synthetase,[56] utilize A domains that selectively activate α-hydroxy acids. These molecules are thought to be obtained using an *in trans* KR domain, which directly reduces the necessary, soluble α-keto acid.

5.19.6.6 Formylation Domains

As noted previously, linear gramicidin[44] and anabaenopeptilid[43] synthetases begin with modules containing formylation (F) domains in an F-A-PCP initiation module. *In vitro* experiments have revealed that formylation domains act upon aminoacyl-*S*-PCP intermediates before amide bond formation occurs within the downstream C domain. These F domains likely utilize N^{10}-formyltetrahydrofolate as a cofactor.[110]

5.19.7 Structure and Chemistry of NRPS Domains

Structural characterization of NRPSs has yielded significant insight into the enzymology of these complex biosynthetic machines and has provided a framework for engineering these systems toward novel function. As summarized in this section, X-ray crystal and NMR structures have been determined for both individual NRPS domains and multidomain constructs. Overall, these studies support a monomeric structure for NRPS assembly line units where significant domain motion is necessary to allow participation of the various active sites in the chemistry leading to peptide products.

5.19.7.1 Structure and Function of Peptidyl Carrier Protein Domains

The NRPS PCP domain is a small subunit that acts as the scaffold for the attachment of the phosphopantetheinyl prosthetic group and thus the site of covalent linkage for amino acids and the growing peptide chain. The PCP is the smallest of the NRPS domains, typically under 100 amino acids in length, and contributes no catalytic function. The domain is functionally homologous to acyl carrier protein (ACP) domains integral to fatty acid and polyketide biosynthesis. A conserved serine residue in the motif (I/L)GG(D/H)SL is the site of attachment for the pantetheine prosthetic group as a seryl phosphodiester. The first structural information for a PCP domain came from the NMR structure of an excised PCP domain from the tyrocidine NRPS of *Bacillus brevis*.[111] The structure revealed that approximately 75 residues of the domain form an ordered four-helix bundle. The functional serine residue is at the end of one of the helices on a disordered loop region. Although the PCPs have little sequence homology to ACPs, they share in common a helical bundle structure and the relative position of the serine residue (**Figure 10**).

The PCP domain plays the central role in the trafficking of substrates through the assembly line of enzyme domains and active sites. The PCP domains must deliver substrates to multiple protein partners presumably through the formation of specific protein–protein interactions. The small size of the PCP domain and the attached phosphopantetheinyl arm in relation to that of the other enzyme players in the pathway suggests that a significant amount of domain reorganization is necessary for the NRPS assembly line to function. Indeed, structural evidence suggests that the carrier domains of fatty acid and polyketide synthases are mobile elements in assembly lines, shuttling covalently tethered substrates to other catalytic domains.[112,113] Consistent with these predictions, distinct conformations of PCP domains have been observed by changing the state of the PCP (holo vs. apo) or by adding protein-binding partners.[114] As illustrated in **Figure 10**, reexamination of the tyrocidine PCP NMR spectra revealed two conformations that were in slow exchange on the timescale of the

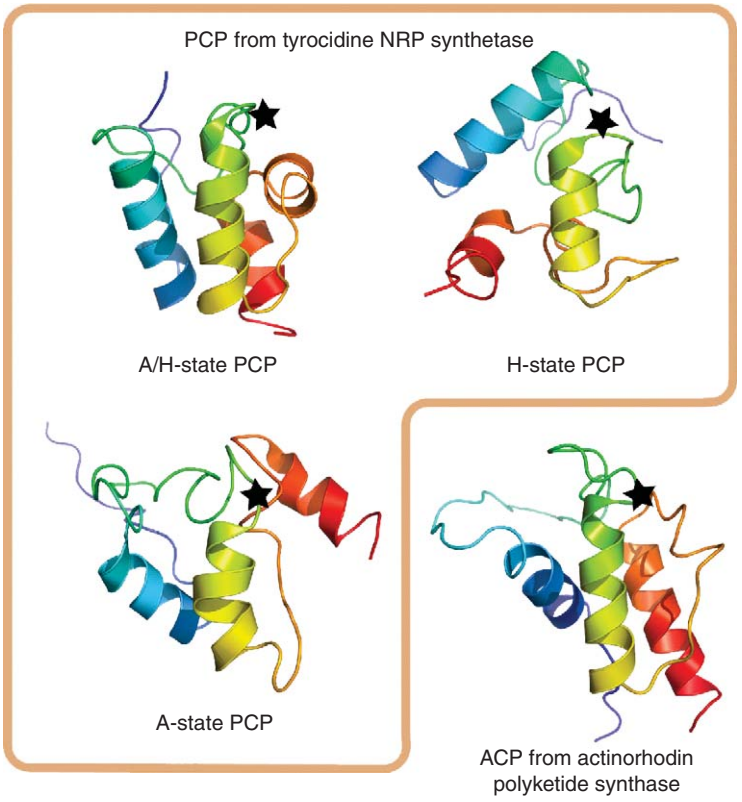


Figure 10 NMR structural analysis of carrier domains. Three conformations of the PCP domain from tyrocidine synthetase (brown box) and the NMR structure of the related ACP domain from a polyketide synthase. The star symbol signifies the position of the conserved phosphopantetheinylated serine residue. The protein ribbon representations are 'rainbow' colored from red (N-terminus) to violet. PDB codes: A/H state, 2GDW; H-state, 2GDX; A-state, 2GDY; ACP, 2AF8.

experiment. Both the apo and holo (with phosphopantetheine loaded) forms of the PCP domain each adopt two conformations. The A/H state is shared between the two, and a distinct state (A-state or H-state) is unique to each form of the carrier domain. Each of the structural conformations represents a significant reorientation of the helical bundle and the position of the prosthetic group. Importantly, these states showed differential interaction with partner enzymes. Using two domains that act with PCP-bound phosphopantetheinyl substrates *in trans*, it was shown that distinct conformational states interact with the PPTase, Sfp, or the housekeeping TE, TEII.

Biochemical experiments on the PCP domain from *E. coli* EntF support the described structural evidence for distinct PCP-interacting conformations.[115] Using combinatorial mutagenesis, specific regions of the PCP were demonstrated to be important for interacting with Sfp *in trans* and the EntF TE domain *in cis*. Related experiments on an aryl carrier domain and a polyketide acyl carrier domain[116] agree with these observations. Overall, the structural and biochemical work provides evidence that the carrier domain is not a simple scaffold for the attachment of the prosthetic arm. Instead, the domain specifically interacts with multiple partners integral to the function of the NRPS machinery.

5.19.7.2 Structure and Function of Adenylation Domains

An adenylation domain from the gramicidin S synthetase provided the first atomic resolution structural information for an NRPS domain (**Figure 11** (**a**)).[117] The excised protein is part of the first module of the synthetase. The domain is responsible for both activating the amino acid phenylalanine and loading the building block onto the adjacent PCP domain. The structure demonstrated that A domains share a common

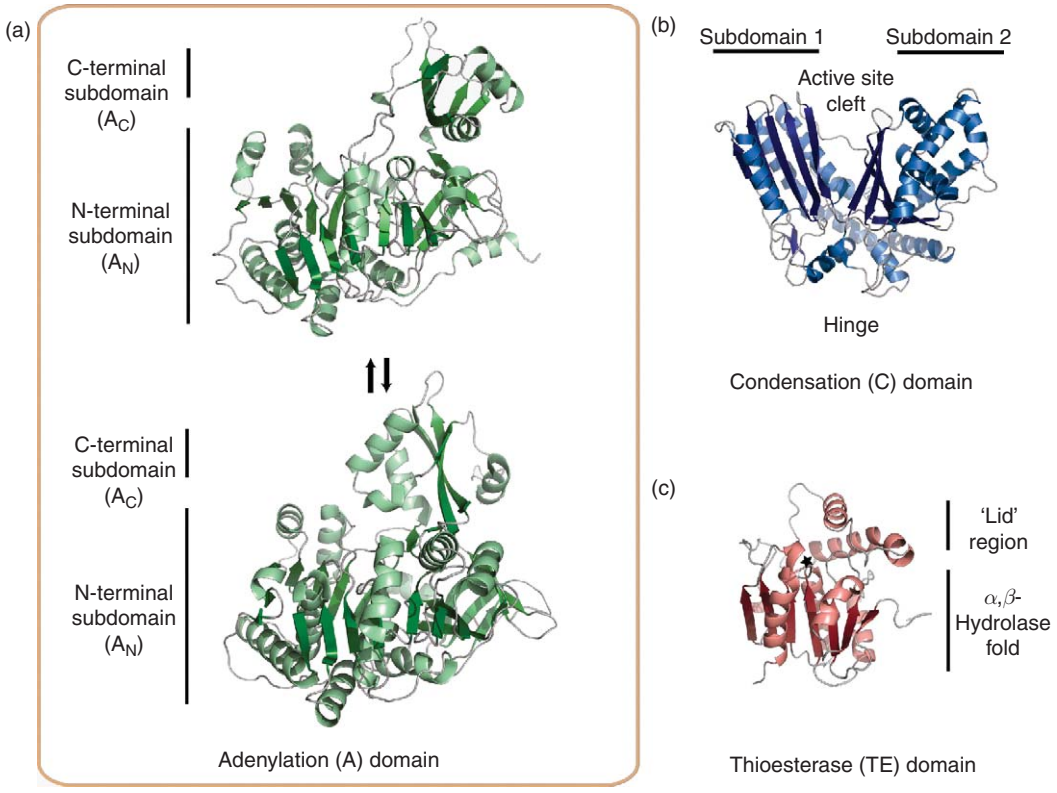


Figure 11 Structural representatives of the core NRPS domains from X-ray crystallographic analysis. (a) Two conformations (brown box) of A domains differing in the orientation of the subdomains. The top structure (PDB code, 1AMU) is postulated to be the conformation responsible for activating the amino acid and the lower (PDB code, 3CW9) for loading the amino acid onto the phosphopantetheine arm. (b) X-ray structure of the VibH condensation domain (PDB code, 1L5A) and (c) the TE domain from surfactin synthetase (PDB code, 1JMK) are also illustrated in ribbon format.

fold with adenylation enzymes such as firefly luciferase and acyl-CoA ligases.[118] Overall, A domains contain a large N-terminal region (A_N) with a majority of the catalytic functionality and a smaller C-terminal subdomain (A_C). The chemistry of A domains involves two steps: (1) ATP-dependent activation of the amino acid by adenylation of the carboxylic acid forming an aminoacyl adenylate and (2) reaction of the activated substrate with the phosphopantetheine arm of a PCP to generate an aminoacyl thioester. The structure of the gramicidin NRPS A domain allowed identification of core residues responsible for substrate recognition and led to the formulation of a code for the assignment of amino acid specificity based on primary sequence.[40,119,120] Combining these results with the colinearity of NRPS modules to the natural product allows for the prediction of the amino acid sequence of an NRP. This bioinformatic approach has led to the identification and structural assignment of novel NRPS-derived natural products through 'genome-mining' approaches.[121] In addition to the A domain from gramicidin S synthetase, the X-ray structure of an aryl acid-activating adenylation domain from the bacillibactin biosynthetic pathway has been determined.[122] The enzyme (DhbE) acts *in trans* with a second module of the bacillibactin NRPS to incorporate Dhb, a common building block in NRPS-derived siderophore natural products. The structure of DhbE revealed that A domains share a similar overall fold and confirmed the structural basis for the NRPS specificity code.

Structural analysis of several non-NRPS adenylation domains has provided significant insight into the basis for the multistep chemistry of NRPS A domains. Of note, the X-ray structures of 4-chlorobenzoate-CoA ligase bound to reaction intermediates showed two dramatically different orientations between the large and small domains.[123,124] The enzyme bound to a substrate analogue was in a similar conformation as the described NRPS A-domain structures. In contrast, the structure of the enzyme bound to a product analogue revealed that

the N- and C-terminal regions were rotated by approximately 140° relative to the previous structures, an orientation that has been observed in other adenylate ligases. This result suggests that the two half reactions catalyzed by A domains are controlled by the mobility of the two subdomains of the enzyme. Indeed, kinetic and mutagenic analysis of chlorobenzoate-CoA ligase supports this conclusion.[125] Structural information about additional amino acid-activating adenylation enzymes will provide direct insight into the chemistry of NRPSs and, in particular, refine the specificity code used to predict natural product structures. Toward this end, structures of the D-alanine-activating ligase DltA, an enzyme involved in the cell-wall biosynthetic pathway, provide insight into the basis for selective binding and activation of a nonproteinogenic amino acid.[126,127]

5.19.7.3 Structure and Function of Condensation Domains

NRPS C domains carry out the key catalytic function of forming amide bonds between amino acid building blocks. C domains catalyze amide formation by coupling the amine of a PCP-bound phosphopantetheinyl amino acid to the upstream peptidyl thioester, transferring the growing chain to the downstream PCP domain. NRPS C domains share sequence homology with acyltransferases, exemplified by chloramphenicol acetyltransferase.[128] Common among the family is a conserved HHXXXDG motif, which is believed to play a role in the chemistry of amide bond formation. In addition to C domains, this enzyme family includes E and Cy domains found in NRPS assemblies as described. Biochemical experiments support the involvement of the HHXXXDG motif in condensation chemistry and the second histidine is postulated to act as a general base.[83,129,130]

The first X-ray structure determined for an NRPS C domain was VibH, a free-standing domain in the vibriobactin biosynthetic pathway in *V. cholerae* (see **Figure 6(b)**).[96] The structure shows a pseudodimeric fold with a deep active site cleft formed between the two distinct subdomains of the C domain (**Figure 11(b)**). Chloramphenicol acetyltransferase and other structurally characterized homologues are trimers, and the two subdomains of VibH are structurally similar to a chloramphenicol acetyltransferase monomer. The structure of VibH supports assignment of the active site residues; however, as VibH is not an integral NRPS domain, details of condensation chemistry and substrate recognition in the context of an NRPS assembly are not readily apparent.

5.19.7.4 Structure and Function of Thioesterase Domains

TE domains are frequently found at the C-terminus of an NRPS and are responsible for release of the mature peptide from the assembly line into solution.[10] Release is accomplished either through hydrolysis of the phosphopantetheinyl thioester to form a free acid or by cyclization using an internal nucleophile in the peptide to cleave the thioester bond. As a significant percentage of biologically active peptide natural products are cyclic, cyclizing TE domains are commonly found in NRPS biosynthetic pathways.[131] TE domains are homologous to the large family of α,β-hydrolases.[132] The family commonly contains an active site catalytic triad with serine acting as a nucleophile to cleave the thioester, initially forming an acyl-enzyme intermediate. The X-ray structure of the TE domain from the last module of surfactin synthetase confirmed an α,β-hydrolase fold with a catalytic triad consisting of Ser-His-Asp (**Figure 11(c)**).[133] The active site forms a large hydrophobic cleft to accommodate the lipopeptide substrate in a conformation that promotes head-to-tail cyclization. The structure also reveals a flexible 'lid' region covering the active site reminiscent of lipase structures.[134–136] Additionally, the structure of the TE domain from fengycin biosynthesis has been determined.[137] This TE domain shares overall structural features with the surfactin TE including a large active site cleft and a 'lid' region with conformational flexibility. Structural and biochemical studies suggest that the large, protected active site plays a key role in sequestering the reactive intermediate and controlling the timing of hydrolytic bond cleavage.

NRPS-based gene clusters frequently contain a second TE gene (TEII) that is expressed as a free-standing enzyme.[10] The function of this domain is to act *in trans* with NRPS in a housekeeping function to repair misprimed phosphopantetheinyl groups through thioester bond hydrolysis.[109,138–140] *In vivo,* the nucleophilic thiol group of PCP-bound phosphopantetheine can react with various electrophiles. Moroever, mispriming of apo-PCP domains with acyl-CoAs by promiscuous PPTases can occur *in vivo.* Both scenarios will lead to a functionally stalled NRPS. The NMR structure of the TEII from surfactin biosynthesis was recently

determined.[141] Similar to the previously described NRPS TE domain of surfactin synthetase,[133] the structure has a α,β-hydrolase fold and a flexible 'lid' region around the active site. However, TEII has a shallow, hydrophobic active site supporting the predicted substrate specificity for short acyl chains while excluding phosphopantetheinyl peptides. NMR titrations with various PCP substrates showed that TEII preferentially recognizes the H state of the PCP domain but other binding conformations are also tolerated.

5.19.7.5 Multidomain NRPS Structural Information

The study of individual NRPS domain structures provides important information regarding the specificity and enzymology of individual steps in NRP biosynthesis. However, structural analysis of larger NRPS constructs is necessary to gain insight into aspects related to domain/domain interactions and the overall structure of the synthetase machinery. This information is particularly important for understanding the details of substrate trafficking and will assist efforts toward the rational manipulation of NRPSs.

5.19.7.5.1 PCP-C didomain structure

The X-ray crystal structure of an excised PCP-C didomain construct from the tyrocidine synthetase provided the first structural information for a multidomain NRPS construct, in addition to the first structure of a C domain that acts *in cis* within an assembly line.[142] The folding of the PCP domain in the crystal structure resembles the A/H state described for the NMR structure of the free-standing PCP domain.[111] However, the carrier domain in the PCP-C structure is not in an orientation that would allow it to interact with the active site of the C domain (**Figure 12**). Instead, the conformation may represent a domain organization, which occurs prior to amide bond formation. The C domain of this construct is structurally similar to the stand-alone VibH,[96] with a chloramphenicol acetyltransferase-like fold containing pseudodimeric subdomains and a large active site cleft. Using this structure as a guide, a mechanism of catalysis based on electrostatic stabilization of reaction intermediates, as opposed to the previously suggested acid–base catalysis by the HHXXXXD motif, has been proposed.[111]

5.19.7.5.2 PCP-TE didomain structure

The NMR structure of the 37 kDa PCP-TE didomain construct from the *E. coli* enterobactin NRPS synthetase provides a detailed picture of the functional interactions between these two domains.[143] As described in Section 5.19.5, the enterobactin TE works with the adjacent PCP domain to catalyze a trimerization/cyclization of 2,3-dihydroxybenzyl serine to generate the siderophore natural product (see **Figure 4(b)**). NMR analysis

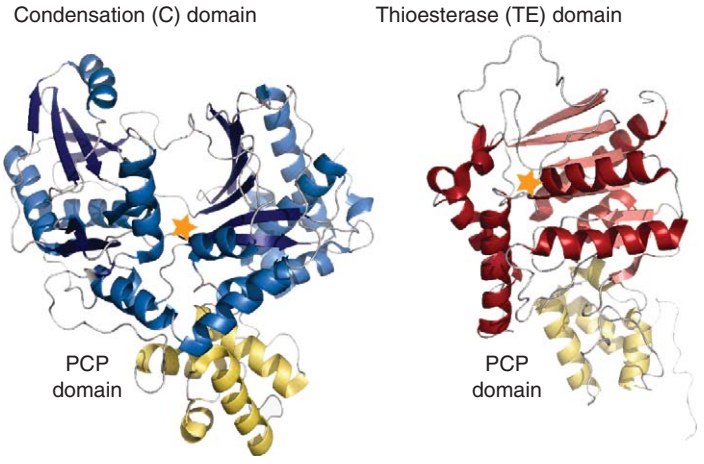


Figure 12 Structures of didomain NRPS constructs. The X-ray structure of a PCP-C (PDB code, 2JGP) and NMR structure of a PCP-TE (PDB code, 2ROQ) fragment. The location of the active site is indicated by the orange star and the relative orientation of the C and TE domains are consistent with the top multidomain structure in **Figure 13**.

shows that the two domains form a compact structure with an extensive hydrophobic interface between the two domains. Based on the orientation between the active site serine of each domain (**Figure 12**), the observed conformation appears to be catalytically relevant for PCP delivery of the substrate into the active site of the TE. As with the previously described structures of TEs,[133,137] the enterobactin TE contains a dynamic 'lid' region consisting of two α-helices. Overall, the structure displays plasticity that allows the PCP domain to interact with protein partners in addition to the TE. Based on NMR measurements, a second conformation of the didomain construct was proposed with a more open structure. In addition, NMR titrations with the interacting partners (PPtase and the upstream C domain) showed that the presence of these domains modulated the PCP-TE didomain interactions.

5.19.7.5.3 Structure of a C-A-PCP-TE termination module

The termination module of surfactin synthetase is a 144 kDa four-domain enzyme responsible for the incorporation of the final amino acid (L-Leu) into the surfactin peptide and subsequent cyclization of the resulting product. The structure of the TE domain of this construct was previously solved.[133] In the recently determined 2.6 Å X-ray structure of the C-A-PCP-TE construct, the entire protein chain is evident in the electron density maps.[144] The structural folds of the individual domains in this module are similar to structures of monomeric domains (**Figure 13**). The deviations observed in this multidomain structure include a slight difference in the hinge region of C domain subdomains and an orientation of the subdomains of the A domain that is not consistent with the open or closed conformations previously described. The A domain contains

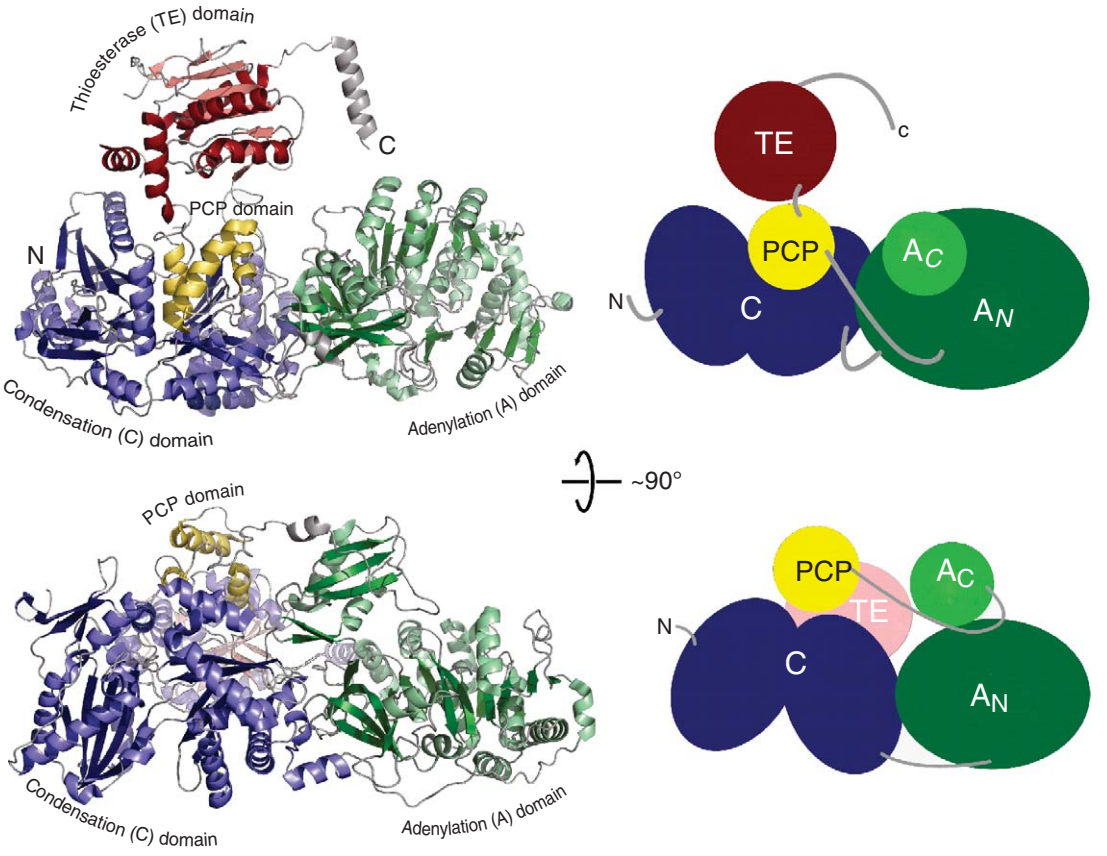


Figure 13 X-ray structure of the four-domain termination module of surfactin synthetase (PDB code, 2VSQ). The coloring and representation of the domains is the same as in **Figures 11** and **12**. A cartoon diagram of the relative domain structure is illustrated at the right of the two views. A_C and A_N signify the C-terminal and N-terminal subdomains of the A domain.

bound L-Leu and is proposed to occupy a confirmation that occurs before ATP binding. The TE domain is very similar to the previously described 'open' state observed in the structure of this individual domain.[133] The overall module architecture is dominated by the C domain and the A_N subdomain of the A domain. These two domains form an extensive interface that is likely invariant during the catalytic process. Both the PCP domain and the C-terminal subdomain of the A domain are adjacent to the C–A platform and appear to be more flexible. The observed distance between the C and A domain active sites (~63 Å) necessitates mobility of the smaller domains during the multiple steps of catalysis. The orientation of the PCP domain with respect to other domains suggests that the termination module crystallized in an orientation where the PCP domain interacts with the acceptor site of the C domain. The linker regions between domains also suggest that the C/A interaction is conformationally stable based on the well-ordered intervening segment. In contrast, the linkers between the A/PCP and PCP/TE domains, though shorter, appear less structured, suggesting flexibility. The X-ray structure also contains a serendipitous observation relevant to NRPS module/module interactions. A C-terminal helix that was not part of the synthetase, but that originated from the cloning vector, forms an interaction with the C domain of an adjacent synthetase in the crystal lattice. Though not biologically relevant, this helix is homologous to predicted linkers present in nonterminal modules. Thus, the helical interaction could mimic the natural structural elements that link NRPS modules *in trans*.

5.19.8 Pathways to Nonproteinogenic Amino Acids Incorporated into NRP Natural Products

Although a significant percentage of building blocks found in NRPs are included among the 20 canonical amino acids, they can also incorporate a variety of nonproteinogenic amino acids. The increased availability of building blocks enhances the structural diversity found in NRP natural products. As previously discussed, one mechanism to alter amino acids and form new chemical entities is through enzyme domains imbedded in the NRPS assembly line. With this methodology, modifications to amino acids can be introduced 'on the fly' as the peptide is assembled on the synthetase. This chemistry commonly includes epimerization and methylation, and to a lesser extent cyclization and reduction/oxidation reactions.

The following section explores the biosynthesis and utilization of nonproteinogenic amino acids by NRPSs. These building blocks are either biosynthesized *de novo*, often with genes included in the natural product gene cluster, or obtained from other cellular pathways present in the host. In these systems, the NRPS modules contain adenylation domains that specifically recognize and activate the uncommon amino acid. There are three general pathways to these building blocks as summarized below.

5.19.8.1 Nonproteinogenic Amino Acids Present as Cellular Metabolites

NRPSs can incorporate amino acids that are natural metabolic intermediates of the producing organism but not included among the 20 proteinogenic amino acids (**Figure 14** (**a**)). Frequently in these cases, the biosynthetic gene products for the amino acid do not reside in the NRP gene cluster. For instance, β-alanine is available from decarboxylation of aspartate, and is an intermediate in the biosynthesis of coA.[145,146] D-Ala is a common cellular metabolite in bacteria, where it is incorporated into cell walls. In such organisms, dedicated alanine racemases are used to epimerize L-Ala by a PLP-dependent mechanism.[147,148] The remaining examples are intermediates in the metabolic pathways of proteinogenic amino acids. L-Ornithine and L-α-aminoadipic acid are intermediates in the pathways to proline/arginine and lysine, respectively.[149] L-Ornithine is a common component of NRPs, in particular siderophores,[150] and L-α-aminoadipic acid is incorporated into penicillins.[151] L-Kynurenic acid is the first intermediate in the oxidative degradation of tryptophan and is incorporated into the daptomycin antibiotics among others.[152]

5.19.8.2 Modification of Proteinogenic Amino Acids

Enzymatic elaboration of proteinogenic amino acids is the most common pathway to nonproteinogenic amino acids. Six representative examples are illustrated in **Figure 14(b)**. Amino acids that have undergone oxidation

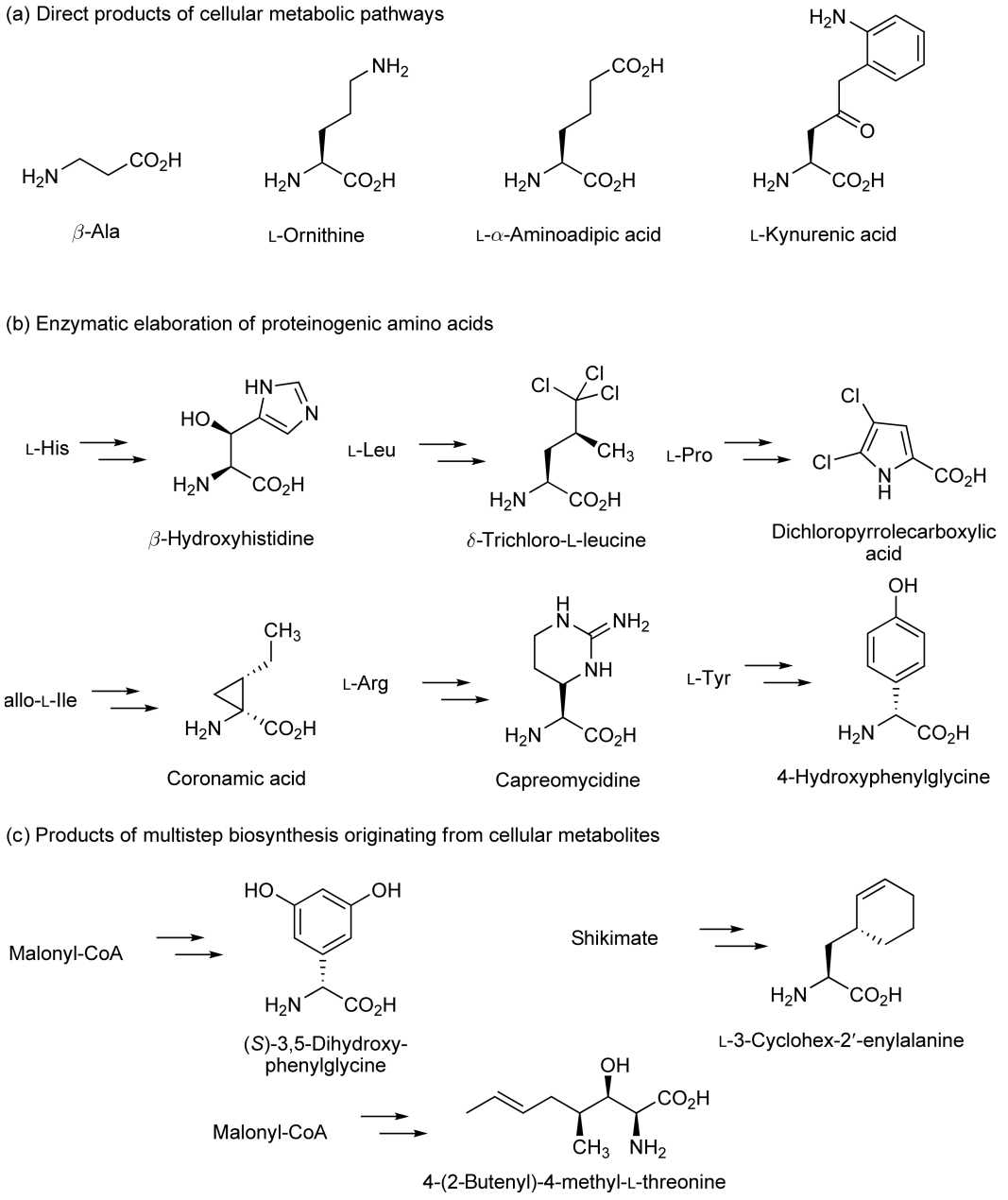


Figure 14 Summary of pathways (a–c) to nonproteinogenic amino acids found in NRP natural products.

at the β-position are frequently incorporated in NRPs. The biosynthesis of β-hydroxyhistidine from the nikkomycin pathway is a biochemically characterized example of this transformation.[153–155] The iron–heme-containing enzyme NikQ monohydroxylates L-His stereospecifically at the β-position using a mechanism common to the cytochrome P-450 enzyme family.[156] Another common oxidative modification is site-specific halogenation to incorporate chlorine and bromine atoms. Recent work has elucidated two general pathways to incorporate halogens into NRP natural products.[157,158] Flavin-dependent chlorination enzymes have been demonstrated to use a hypochlorous species to react with aromatic substrates through an electrophilic aromatic substitution mechanism.[159,160] This halogenation strategy is used in the production of the dichloropyrrolyl moiety found in the mixed NRP/polyketide pyoluteorin (see **Figure 14(b)**).[161] In addition, the chemically more

challenging halogenation of unactivated C–H groups is also observed in NRPS pathways. For instance, the marine cyanobacterium *Lyngbya majuscula* incorporates δ-trichloro-L-leucine into the NRP natural product barbamide.[47,162] In this pathway two genes encoding for nonheme, iron-dependent halogenases install the three chlorine atoms using α-ketoglutarate-dependent chemistry.

The three additional examples shown in **Figure 14(b)** illustrate the array of structurally complex amino acids possible by enzymatic elaboration of proteinogenic amino acids. The cyclopropyl-containing amino acid coronamic acid is derived from *allo*-L-Ile through chlorination of the aliphatic chain followed by an intramolecular ring-closing reaction that eliminates chloride ion.[163] Coronamic acid is subsequently incorporated into the natural product coronatine.[164] The cyclic amino acid capreomycidine is found in NRP natural products including the antibiotics viomycin and capreomycin.[165–167] This amino acid is derived from L-Arg in two enzymatic steps: β-hydroxylation, analogous to that described previously for β-hydroxyhistidine, followed by a PLP-dependent cyclization.[168,169]

Phenylglycines are important components of the vancomycin/teicoplanin antibiotics, and the conformationally restricted amino acids contribute to the unique architecture and biological function of these clinically important NRPs.[9] 4-Hydroxyphenylglycine is produced from L-tyrosine in a pathway that involves three enzymes.[170,171] In the key step, a nonheme iron oxidase catalyzes the oxidative decarboxylation of the α-keto acid derivative of L-tyrosine resulting in loss of carbon dioxide and generation of the phenylglycine carbon framework.[172]

In addition to β-alanine, several NRPs contain other β-amino acids derived from their α-amino acid counterparts. Aminomutases catalyzed a direct 1,2-amino shift producing the β-amino acid in a single step. There are two general strategies described for enzymes to promote this chemically challenging 1,2-amino shift. Aliphatic α-amino acid substrates are converted into β-amino acids through a mechanism involving the production of a radical intermediate at the β-carbon.[173] For example, lysine-2,3-aminomutase contains the cofactor PLP and uses an iron–sulfur cluster and SAM to generate the reactive radical intermediate.[174,175] β-Lysine is found in the antibiotics viomycin and capreomycin.[165–167] A second class of aminomutases act on aromatic amino acids and are dependent on the rare prosthetic group 4-methylideneimidazole-5-one.[176,177] The L-tyrosine aminomutase in the biosynthetic pathway to the mixed NRP/polyketide enediyne antitumor/antibiotics is a well-characterized example, and biochemical studies have demonstrated that the enzyme uses an electrophilic mechanism to catalyze the chemistry.[178,179]

5.19.8.3 Nonproteinogenic Amino Acids Derived from Multistep Pathways

A limited number of nonproteinogenic amino acids incorporated into NRP natural products are not derived from available amino acids. In these examples, the natural product biosynthetic gene clusters encode enzymes that prepare the nonproteinogenic amino acids from common cellular precursors. Three examples of therapeutically important natural products are briefly described. In addition to 4-hydroxyphenylglycine, the vancomycin/teicoplanin family of antibiotics also contain the (*S*)-3,5-dihydroxyphenylglycine (Dpg) building blocks.[9] Dpg is biosynthesized from the common cellular, two-carbon metabolite malonyl-CoA. Five genes are responsible for the production of Dpg, with the carbon skeleton constructed by three enzymes using Type III polyketide synthase methodology to generate the intermediate 3,5-dihydroxyphenylacetyl-CoA.[180,181] Benzylic oxidation by an unusual cofactor-independent dioxygenase followed by the action of an aminotransferase yields the amino acid, Dpg.[182,183] The biosynthesis of the clinically important cyclosporin by the fungus *T. niveum* provides another example.[17] This cyclic peptide contains multiple nonproteinogenic amino acids, including the unique 4-(2-butenyl)-4-methyl-L-threonine. As with Dpg, labeling studies have shown that 4-(2-butenyl)-4-methyl-L-threonine is prepared by a polyketide synthase mechanism, although the specific enzyme pathway has not been fully characterized.[184,185] Another recently described example of this category is the pathway to L-3-cyclohex-2′-enylalanine from salinosporamide biosynthesis.[186] Prephenate, a common intermediate in the primary metabolic pathway to several proteinogenic aromatic amino acids, is the starting material in this pathway.[187]

Frequently, NRPS domain enzymology is utilized in the biosynthetic pathways to nonproteinogenic amino acids. In these examples, there exists a dedicated A/PCP domain pair (either as two stand-alone enzymes or as a didomain construct) that loads the starting amino acids onto a PCP domain as a phosphopantenthienyl thioester (**Figure 15**). Various enzyme-catalyzed transformations are then employed to generate the nonproteinogenic

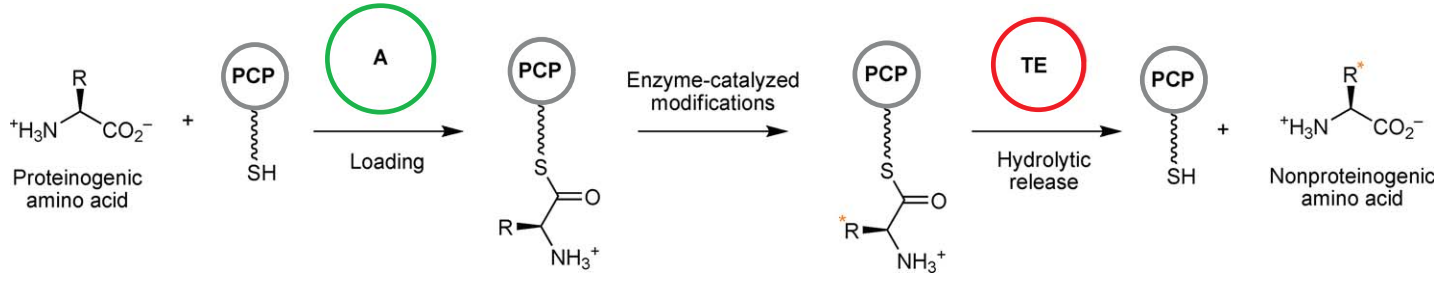


Figure 15 Use of NRPS enzymology in the biosynthetic pathways to nonproteinogenic amino acids. R^* signifies a nonproteinogenic side chain.

amino acid bound to the PCP domain. In order to produce the soluble nonproteinogenic amino acid, a dedicated stand-alone TE domain is necessary to cleave the product from the PCP domain. Using this approach, proteinogenic amino acids are diverted from primary metabolism to a secondary metabolic pathway. This use of NRPS methodology is distinct from the assembly-line NRPS producing the peptide, and the nonproteinogenic amino acid is activated by an assembly-line A domain specific for the modified amino acid. A well-characterized example of this approach is the pathway to δ-trichloro-L-leucine from barbamide biosynthesis.[188,189] A stand-alone A domain (BarD) loads L-Leu onto the PCP domain BarA, and trichlorination occurs on the PCP-bound amino acid. The stand-alone TE domain (BarC) present in the cluster is predicted to free the amino acid from the PCP domain. δ-Trichloro-L-leucine is then activated and incorporated into barbamide by the NRPS A domain of BarE.

5.19.8.4 Tailoring Enzymology in NRP Natural Products

Frequently, after assembly of the peptide natural product by the NRPS, enzyme-catalyzed reactions impart additional functionality and structural elements. The reactions are very diverse and a thorough survey of the known enzymology is beyond the scope of this chapter; however, several recent review articles describe the various chemistries that are known.[7,9,12,190]

5.19.9 Chemical Approaches Toward Mechanistic Probes and Inhibitors of NRPS Enzymes

A variety of chemical tools have been developed to probe the activity and specificity determinants of NRPS domains (**Figure 16**).

5.19.9.1 Adenylation Domain Chemistry and Specificity

Typically, A domain specificity toward amino acid substrates is determined using an ATP/PP_i exchange assay.[191,192] Alternatively, the adenylation of substrates and their transfer to phosphopantetheinyl arms can be studied using radioactive assays[156] with labeled substrates or by mass spectrometry.[193] Recently, Yin and coworkers have shown the utility of using alkyne-functionalized substrate analogues to profile adenylation domain activity (**Figure 16(b)**).[194] In this study, various alkyne-containing 3-amino-5-hydroxybenzoic acid analogues were incubated with the A-PCP didomain RifA from the rifamycin synthetase. The adenylation domain of this construct was able to activate these analogues and load them onto the phosphopantetheinyl arm of the RifA-PCP. The alkyne-containing aminoacyl-*S*-PCP was then conjugated to an azide-containing biotin tag through 'click chemistry' and could be visualized using standard ELISA or western blot analysis. The use of nonradioactive substrates and ease of analysis for such assays provide benefits over previous, widely used methodologies.

5.19.9.1.1 Inhibitors of adenylation domain function

In 2003, Marahiel and coworkers reported the use of 5′-*O*-[*N*-(aminoacyl)-sulfamoyl] adenosine (AMS) analogues to inhibit NRPS A domain activity.[195] These compounds were initially developed as inhibitors for tRNA synthetases, which catalyze the adenylation of amino acids and subsequent esterification with a hydroxyl group of tRNA.[196,197] The fact that AMSs inhibit both classes of enzymes is somewhat surprising as NRPS A domains and tRNA synthetases are not structurally homologous and bind vastly different conformations of aminoacyl adenylate intermediates in their respective active sites. Recent work has suggested that the A domain inhibitors may be viable drug candidates as they can be used to stop the production of NRPS-derived virulence factors produced by pathogenic bacteria.[198,199] The major challenge in developing such drugs is to selectively halt A domain activity, while avoiding inhibition of tRNA synthetases. One way of overcoming this obstacle is to target A domains that use nonproteinogenic amino acids or aryl acids as substrates. Following this hypothesis, Quadri and coworkers used salicyl-AMS to inhibit the growth of *Mycobacterium tuberculosis* and *Yersinia pestis*, which produce siderophore virulence factors containing salicylic acid monomers.[200] Synthetic routes to a

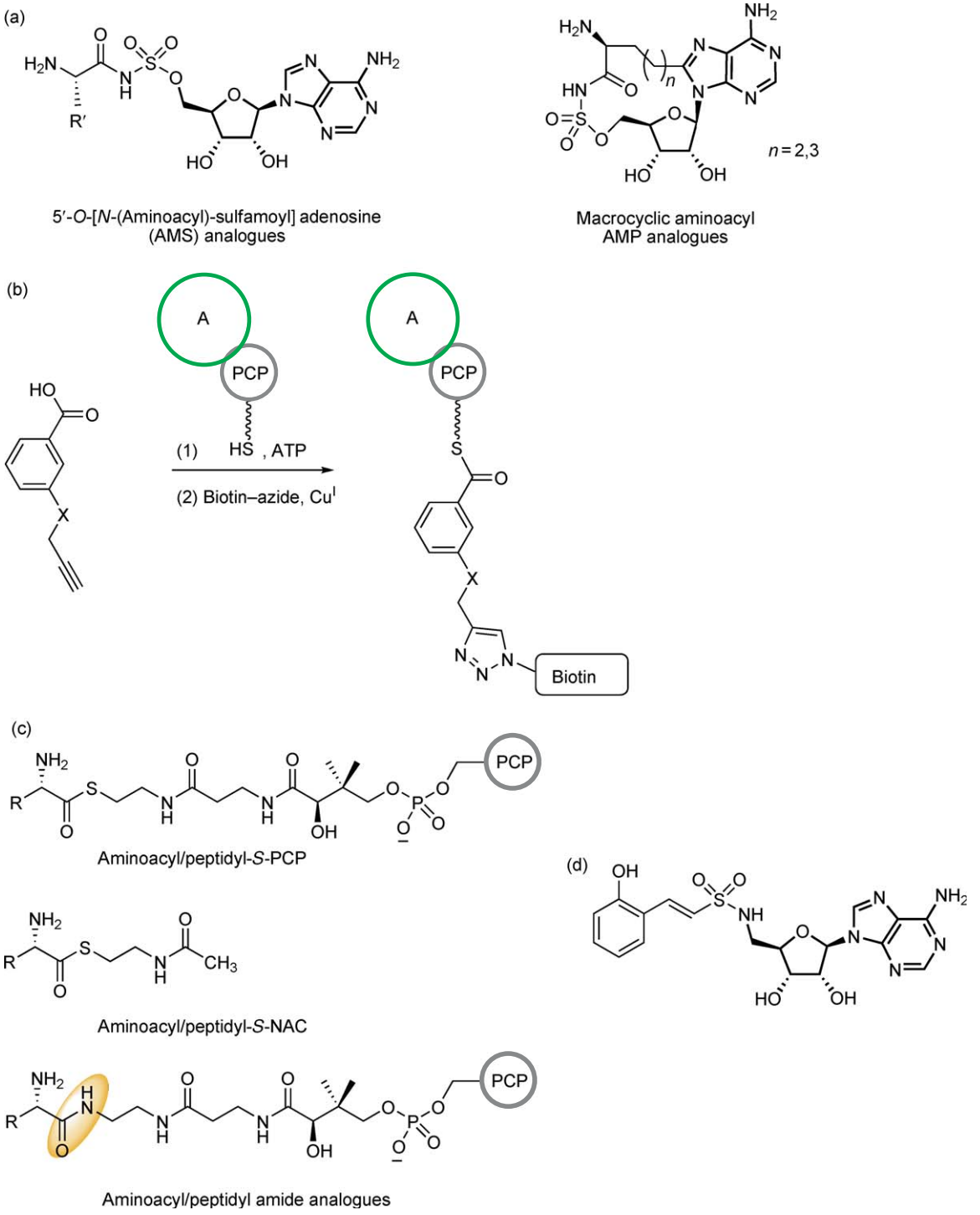


Figure 16 (a) Structures of adenylation domain intermediates and inhibitors: aminoacyl-sulfamoyl adenosine (AMS) and 'cisoid'-like macrocyclic inhibitor. (b) Alkyne-functionalized chemical probe for NRPS A and PCP domains. (c) Structure of aminoacyl PCP, SNAC substrate analogue, and hydrolytically stable phosphopantetheinyl analogue. (d) Structure of vinylsulfonamide probe. R represents a peptide component and R′ an amino acid side chain.

variety of salicyl-AMS analogues with variations in the aryl acid, sulfamate, and glycosyl moieties have been reported by Aldrich and coworkers yielding a number of potential antibiotic leads.[201,202] Additionally, the use of a fluorescently labeled AMS probe in competitive binding assays of aryl acid activating A domains has been demonstrated.[203] This assay was used for the high-throughput screening of chemical libraries, which revealed that several molecules that do not contain a nucleoside moiety can effectively inhibit aryl acid adenylation domains. Recently, macrocyclic aminoacyl AMP analogues have been used to selectively target A domains that use proteinogenic amino acids as substrates.[204] These compounds structurally mimic the 'cisoid' conformation of aminoacyl adenylates formed by A domains and should not bind to tRNA synthetases, which form adenylates adopting a 'transoid' geometry. Biochemical experiments have verified that these macrocyclic compounds are potent inhibitors of A domains and do not affect tRNA synthetase activity in *in vitro* translation assays.

5.19.9.2 Substrate Analogues for *In Vitro* Biochemical Analysis of NRPS Domain Function

Probing the reactions catalyzed by other NRPS domains can be difficult, as the reactants are frequently aminoacyl/peptidyl-*S*-PCP conjugates. In 2000, Walsh and coworkers reported that aminoacyl-*N*-acetylcysteamine thioesters (aminoacyl-SNACs), which mimic PCP-bound substrates, could be used to probe the reactivity and specificity of C domains.[205] Peptidyl and aminoacyl thioesters of NAC can also serve as substrate analogues for other NRPS domains, including TE, Cy, and MTs.[35,206,207]

In some cases, chemically synthesized peptidyl-SNACs do not serve as substrates for NRPS domains. For instance, the TE domains from fengycin, syringomycin, and mycosubtilin synthetases do not catalyze heterocyclization of the requisite peptidyl-SNACs.[208] When such a problem is encountered, apo-PCPs in multidomain constructs can be loaded with aminoacyl or peptidyl-CoA analogues by promiscuous phosphopantetheinyl-transferases such as Sfp.[209] This strategy has been used to study the activity of various domains including R, C, and TE domains.[210,211] However, this method is limited by the single-turnover nature of the reaction, which generates a holo-PCP domain that cannot be reloaded with the desired peptide. For TE domains, this limitation can sometimes be circumvented by using peptidyl thiophenol thioesters as substrates.[208]

5.19.9.3 Chemical Approaches Toward Understanding Domain Architecture

Using chemical modifications to restrict the movement of individual domains within a module could prove useful for structural and mechanistic studies of synthetases. Aldrich and coworkers have shown that vinylsulfonamide analogues of aminoacyl adenylates can modulate protein–protein interactions between NRPS domains.[212] These mechanism-based inhibitors can bind to A domains, which catalyze the 1,4 addition of the thiol group of a phosphopantetheine arm of a partner PCP to the vinyl group of the inhibitor. Gel-shift assays revealed that interactions between *in trans* acting PCP and A domains were strengthened upon covalent modification of the PCP with a vinylsulfonamide inhibitor. Additionally, hydrolytically stable, aminoacyl-CoA analogues have been designed to restrict carrier domain geometry and manipulate interdomain interactions.[213] Aminoacyl-CoA analogues (see **Figure 16(c)**) are transferred to a PCP domain by promiscuous phosphopantetheinyl-transferases. Biochemical experiments revealed that aminoacyl-PCP conjugates selectively inhibit A domain activity in a multidomain construct. Moreover, an α-chloro-CoA analogue was used to covalently link the active site Ser of EntF's TE domain to the PCP-bound phosphopantetheinyl analogue. Using this methodology to restrict PCP domain conformations will likely be useful in structural and mechanistic studies of multidomain constructs.[214]

5.19.10 Conclusions

NRPSs are responsible for the biosynthesis of a wide range of structurally diverse natural products. Included in this large family of small molecules are several examples of clinically important therapeutics. The basic enzymology of these systems represents a fascinating example of complex multicomponent

enzyme function. Peptide synthesis is carried out on large, multidomain proteins where the individual domains are organized into modules responsible for amino acid incorporation. The amino acid building blocks are covalently tethered to the synthetase as thioesters and are oligomerized in a manner that is colinear with the modules of the assembly line. Frequently, pathways toward NRPs have evolved to deviate from the simple oligomerization of amino acids. Use of these noncanonical mechanisms significantly expands the peptide architectures possible, resulting in the production of complex biologically active molecules. In addition, the use of nonproteinogenic amino acid building blocks and postsynthetic modification adds to the observed structural complexity.

The colinearity and predictability of substrate specificity of NRPSs has spurred proposals and efforts toward rational manipulation of pathways to generate novel biologically active compounds.[215–220] Recent advances in understanding the chemistry and structure of NRPSs have provided significant insight into the complex enzymology of these systems. However, several significant questions including the specificity determinants of condensation domains, how modules interact with one another, and the manner in which intermediates are shuttled through assembly lines remain unanswered. Further insight into these questions, along with improvement in the bioinformatics-based characterization of synthetases, will aid in efforts focused on the rational manipulation of synthetases that can be used to construct novel natural product analogues.

References

1. T. A. Steitz, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253.
2. H. S. Zaher; R. Green, *Nature* **2009**, *457*, 161–166.
3. J. M. Willey; W. A. van der Donk, *Annu. Rev. Microbiol.* **2007**, *61*, 477–501.
4. M. A. Fischbach; C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468–3496.
5. F. Lipmann; W. Gevers; H. Kleinkauf; R. Roskoski, Jr., *Adv. Enzyme Mol. Biol.* **1971**, *35*, 1–34.
6. F. Lipmann, *Adv. Microbiol. Phys.* **1980**, *21*, 227–266.
7. C. T. Walsh; A. M. Gehring; P. H. Weinreb; L. E. Quadri; R. S. Flugel, *Curr. Opin. Chem. Biol.* **1997**, *1*, 309–315.
8. R. H. Lambalot; A. M. Gehring; R. S. Flugel; P. Zuber; M. LaCelle; M. A. Marahiel; R. Reid; C. Khosla; C. T. Walsh, *Chem. Biol.* **1996**, *3*, 923–936.
9. B. K. Hubbard; C. T. Walsh, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 730–765.
10. R. Finking; M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.
11. C. T. Walsh, *Science* **2004**, *303*, 1805–1810.
12. C. T. Walsh, *Acc. Chem. Res.* **2007**, *41*, 4–10.
13. W. Freist; F. Cramer, *Angew. Chem. Int. Ed. Engl.* **2003**, *32*, 190–200.
14. H. D. Mootz; D. Schwarzer; M. A. Marahiel, *ChemBioChem* **2002**, *3*, 490–504.
15. D. J. Smith; A. J. Earl; G. Turner, *EMBO J.* **1990**, *9*, 2743–2750.
16. H. T. Chiu; B. K. Hubbard; A. N. Shah; J. Eide; R. A. Fredenburg; C. T. Walsh; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8548–8553.
17. G. Weber; K. Schorgendorfer; E. Schneiderscherzer; E. Leitner, *Curr. Genet.* **1994**, *26*, 120–125.
18. T. P. Lin; C. L. Chen; L. K. Chang; J. S. M. Tschen; S. T. Liu, *J. Bacteriol.* **1999**, *181*, 5060–5067.
19. P. Cosmina; F. Rodriguez; F. Deferra; G. Grandi; M. Perego; G. Venema; D. Vansinderen, *Mol. Microbiol.* **1993**, *8*, 821–831.
20. H. D. Mootz; M. A. Marahiel, *J. Bacteriol.* **1997**, *179*, 6843–6850.
21. M. Glinski; C. Urbanke; T. Hornbogen; R. Zocher, *Arch. Microbiol.* **2002**, *178*, 267–273.
22. A. Haese; M. Schubert; M. Herrmann; R. Zocher, *Mol. Microbiol.* **1993**, *7*, 905–914.
23. A. M. Gehring; I. Mori; C. T. Walsh, *Biochemistry* **1998**, *37*, 2648–2659.
24. J. J. May; T. M. Wendrich; M. A. Marahiel, *J. Biol. Chem.* **2001**, *276*, 7209–7217.
25. J. Kratzschmar; M. Krause; M. A. Marahiel, *J. Bacteriol.* **1989**, *171*, 5422–5429.
26. Y. Q. Cheng, *ChemBioChem* **2006**, *7*, 471–477.
27. N. A. Magarvey; M. Ehling-Schulz; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 10698–10699.
28. C. A. Shaw-Reid; N. L. Kelleher; H. C. Losey; A. M. Gehring; C. Berg; C. T. Walsh, *Chem. Biol.* **1999**, *6*, 385–400.
29. K. M. Hoyer; C. Mahlert; M. A. Marahiel, *Chem. Biol.* **2007**, *14*, 13–22.
30. S. C. Wenzel; B. Kunze; G. Hofle; B. Silakowski; M. Scharfe; H. Blocker; R. Muller, *Chembiochem* **2005**, *6*, 375–385.
31. S. C. Wenzel; P. Meiser; T. M. Binz; T. Mahmud; R. Muller, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301.
32. N. A. Magarvey; B. Haltli; M. He; M. Greenstein; J. A. Hucul, *Antimicrob. Agents Chemother.* **2006**, *50*, 2167–2177.
33. T. A. Keating; C. G. Marshall; C. T. Walsh, *Biochemistry* **2000**, *39*, 15513–15521.
34. T. A. Keating; C. G. Marshall; C. T. Walsh, *Biochemistry* **2000**, *39*, 15522–15530.
35. C. G. Marshall; M. D. Burkart; T. A. Keating; C. T. Walsh, *Biochemistry* **2001**, *40*, 10655–10663.
36. N. J. Hillson; C. J. Balibar; C. T. Walsh, *Biochemistry* **2004**, *43*, 11344–11351.
37. S. Lautru; R. J. Deeth; L. M. Bailey; G. L. Challis, *Nat. Chem. Biol.* **2005**, *1*, 265–269.
38. E. J. Dimise; P. F. Widboom; S. D. Bruner, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15311–15316.
39. G. L. Challis; J. Ravel, *FEMS Microbiol. Lett.* **2000**, *187*, 111–114.

40. G. L. Challis; J. Ravel; C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211–224.
41. S. W. Haynes; G. L. Challis, *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 203–218.
42. G. L. Tang; Y. Q. Cheng; B. Shen, *J. Biol. Chem.* **2007**, *282*, 20273–20282.
43. L. Rouhiainen; L. Paulin; S. Suomalainen; H. Hyytiainen; W. Buikema; R. Haselkorn; K. Sivonen, *Mol. Microbiol.* **2000**, *37*, 156–167.
44. N. Schracke; U. Linne; C. Mahlert; M. A. Marahiel, *Biochemistry* **2005**, *44*, 8507–8513.
45. C. Rausch; I. Hoof; T. Weber; W. Wohlleben; D. H. Huson, *BMC Evol. Biol.* **2007**, *7*, 78.
46. S. L. Clugston; S. A. Sieber; M. A. Marahiel; C. T. Walsh, *Biochemistry* **2003**, *42*, 12095–12104.
47. C. J. Balibar; F. H. Vaillancourt; C. T. Walsh, *Chem. Biol.* **2005**, *12*, 1189–1200.
48. N. Roongsawang; K. Hase; M. Haruki; T. Imanaka; M. Morikawa; S. Kanaya, *Chem. Biol.* **2003**, *10*, 869–880.
49. C. M. Farnet, E. Zazopoulos, A. Staffa, Gene Encoding a Nonribosomal Peptide Synthetase for the Production of Ramoplanin. U.S. Patent 7,078,185, 15 October 2001.
50. E. Guenzi; G. Galli; I. Grgurina; D. C. Gross; G. Grandi, *J. Biol. Chem.* **1998**, *273*, 32857–32863.
51. B. K. Scholz-Schroeder; J. D. Soule; D. C. Gross, *Mol. Plant-Microbe Interact.* **2003**, *16*, 271–280.
52. V. de Crecy-Lagard; V. Blanc; P. Gil; L. Naudin; S. Lorenzon; A. Famechon; J. Bamas; J. Crouzet; D. Thibaut, *J. Bacteriol.* **1997**, *179*, 705–713.
53. D. G. Fujimori; S. Hrvatin; C. S. Neumann; M. Strieker; M. A. Marahiel; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16498–16503.
54. N. A. Magarvey; Z. Q. Beck; T. Golakoti; Y. S. Ding; U. Huber; T. K. Hemscheidt; D. Abelson; R. E. Moore; D. H. Sherman, *ACS Chem. Biol.* **2006**, *1*, 766–779.
55. A. V. Ramaswamy; C. M. Sorrels; W. H. Gerwick, *J. Nat. Prod.* **2007**, *70*, 1977–1986.
56. Y. Q. Xu; R. Orozco; E. M. K. Wijeratne; A. A. L. Gunatilaka; S. P. Stock; I. Molnar, *Chem. Biol.* **2008**, *15*, 898–907.
57. R. H. Proctor; D. W. Brown; R. D. Plattner; A. E. Desjardins, *Fungal Genet. Biol.* **2003**, *38*, 237–249.
58. K. Zaleta-Rivera; C. P. Xu; F. G. Yu; R. A. E. Butchko; R. H. Proctor; M. E. Hidalgo-Lara; A. Raza; P. H. Dussault; L. C. Du, *Biochemistry* **2006**, *45*, 2561–2569.
59. B. Silakowski; B. Kunze; G. Nordsiek; H. Blocker; G. Hofle; R. Muller, *Eur. J. Biochem.* **2000**, *267*, 6476–6485.
60. C. T. Walsh, *Posttranslational Modification of Proteins, Expanding Nature's Inventory*; Roberts and Company Publishers: Englewood, CO, 2006.
61. M. Jin; M. A. Fischbach; J. Clardy, *J. Am. Chem. Soc.* **2006**, *128*, 10660–10661.
62. P. D. Fortin; C. T. Walsh; N. A. Magarvey, *Nature* **2007**, *448*, 824–828.
63. S. T. Cole; R. Brosch; J. Parkhill; T. Garnier; C. Churcher; D. Harris; S. V. Gordon; K. Eiglmeier; S. Gas; C. E. Barry; F. Tekaia; K. Badcock; D. Basham; D. Brown; T. Chillingworth; R. Connor; R. Davies; K. Devlin; T. Feltwell; S. Gentles; N. Hamlin; S. Holroyd; T. Hornby; K. Jagels; A. Krogh; J. McLean; S. Moule; L. Murphy; K. Oliver; J. Osborne; M. A. Quail; M. A. Rajandream; J. Rogers; S. Rutter; K. Seeger; J. Skelton; R. Squares; S. Squares; J. E. Sulston; K. Taylor; S. Whitehead; B. G. Barrell, *Nature* **1998**, *393*, 537–544.
64. C. Reimmann; L. Serino; M. Beyeler; D. Haas, *Microbiology* **1998**, *144*, 3135–3148.
65. F. Kopp; M. A. Marahiel, *Nat. Prod. Rep.* **2007**, *24*, 735–749.
66. F. Lombo; A. Velasco; A. Castro; F. de la Calle; A. F. Brana; J. M. Sanchez-Puelles; C. Mendez; J. A. Salas, *ChemBioChem* **2006**, *7*, 366–376.
67. J. E. Becker; R. E. Moore; B. S. Moore, *Gene* **2004**, *325*, 35–42.
68. F. Kopp; C. Mahlert; J. Grunewald; M. A. Marahiel, *J. Am. Chem. Soc.* **2006**, *128*, 16478–16479.
69. A. Pospiech; J. Bietenhader; T. Schupp, *Microbiology-UK* **1996**, *142*, 741–746.
70. L. Li; W. Deng; J. Song; W. Ding; Q. F. Zhao; C. Peng; W. W. Song; G. L. Tang; W. Liu *J. Bacteriol.* **2008**, *190*, 251–263.
71. Y. Hu; V. Phelan; I. Ntai; C. M. Farnet; E. Zazopoulos; B. O. Bachmann, *Chem. Biol.* **2007**, *14*, 691–701.
72. H. Billman-Jacobe; M. J. McConville; R. E. Haites; S. Kovacevic; R. L. Coppel, *Mol. Microbiol.* **1999**, *33*, 1244–1253.
73. B. Silakowski; G. Nordsiek; B. Kunze; H. Blocker; R. Muller, *Chem. Biol.* **2001**, *8*, 59–69.
74. J. A. Read; C. T. Walsh, *J. Am. Chem. Soc.* **2007**, *129*, 15762–15763.
75. B. Silakowski; B. Kunze; G. Nordsiek; H. Blocker; G. Hofle; R. Muller, *Eur. J. Biochem.* **2000**, *267*, 6476–6485.
76. N. Gaitatzis; B. Kunze; R. Muller, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11136–11141.
77. J. S. Scottcraig; D. G. Panaccione; J. A. Pocard; J. D. Walton, *J. Biol. Chem.* **1992**, *267*, 26044–26049.
78. W. Weckwerth; K. Miyamoto; K. Iinuma; M. Krause; M. Glinski; T. Storm; G. Bonse; H. Kleinkauf; R. Zocher, *J. Biol. Chem.* **2000**, *275*, 17909–17915.
79. P. Tudzynski; K. Holter; T. Correia; C. Arntz; N. Grammel; U. Keller, *Mol. Gen. Genet.* **1999**, *261*, 133–141.
80. B. Riederer; M. Han; U. Keller, *J. Biol. Chem.* **1996**, *271*, 27524–27530.
81. B. Walzel; B. Riederer; U. Keller, *Chem. Biol.* **1997**, *4*, 223–230.
82. T. Stachelhaus; H. D. Mootz; V. Bergendahl; M. A. Marahiel, *J. Biol. Chem.* **1998**, *273*, 22773–22781.
83. T. Stachelhaus; C. T. Walsh, *Biochemistry* **2000**, *39*, 5775–5787.
84. D. M. Gardiner; B. J. Howlett, *FEMS Microbiol. Lett.* **2005**, *248*, 241–248.
85. C. J. Balibar; C. T. Walsh, *Biochemistry* **2006**, *45*, 15029–15038.
86. A. W. Schultz; D. C. Oh; J. R. Carney; R. T. Williamson; D. W. Udwary; P. R. Jensen; S. J. Gould; W. Fenical; B. S. Moore, *J. Am. Chem. Soc.* **2008**, *130*, 4507–4516.
87. S. Weinig; H. J. Hecht; T. Mahmud; R. Muller, *Chem. Biol.* **2003**, *10*, 939–952.
88. F. Schauwecker; F. Pfennig; N. Grammel; U. Keller, *Chem. Biol.* **2000**, *7*, 287–297.
89. S. Rachid; D. Krug; B. Kunze; I. Kochems; M. Scharfe; T. M. Zabriskie; H. Blocker; R. Mueller, *Chem. Biol.* **2006**, *13*, 667–681.
90. D. Schwarzer; R. Finking; M. A. Marahiel, *Nat. Prod. Rep.* **2003**, *20*, 275–287.
91. C. Hacker; M. Glinski; T. Hornbogen; A. Doller; R. Zocher, *J. Biol. Chem.* **2000**, *275*, 30826–30832.
92. R. M. Kagan; S. Clarke, *Arch. Biochem. Biophys.* **1994**, *310*, 417–427.
93. D. A. Miller; C. T. Walsh; L. S. Luo, *J. Am. Chem. Soc.* **2001**, *123*, 8434–8435.

94. M. Z. Ansari; J. Sharma; R. S. Gokhale; D. Mohanty, *BMC Bioinf.* **2008**, *9*, 454.
95. K. Hoffmann; E. Schneiderscherzer; H. Kleinkauf; R. Zocher, *J. Biol. Chem.* **1994**, *269*, 12710–12714.
96. T. A. Keating; C. G. Marshall; C. T. Walsh; A. E. Keating, *Nat. Struct. Biol.* **2002**, *9*, 522–526.
97. L. S. Luo; C. T. Walsh, *Biochemistry* **2001**, *40*, 5329–5337.
98. P. J. Belshaw; C. T. Walsh; T. Stachelhaus, *Science* **1999**, *284*, 486–489.
99. H. Rieder; G. Heinrich; E. Breuker; M. M. Simlot; P. Pfaender, *Methods Enzymol.* **1975**, *43*, 548–559.
100. D. A. Miller; L. S. Luo; N. Hillson; T. A. Keating; C. T. Walsh, *Chem. Biol.* **2002**, *9*, 333–344.
101. T. A. Keating; D. A. Miller; C. T. Walsh, *Biochemistry* **2000**, *39*, 4729–4739.
102. T. Duerfahrt; K. Eppelmann; R. Muller; M. A. Marahiel, *Chem. Biol.* **2004**, *11*, 261–271.
103. R. S. Roy; A. M. Gehring; J. C. Milne; P. J. Belshaw; C. T. Walsh, *Nat. Prod. Rep.* **1999**, *16*, 249–263.
104. R. Jansen; H. Irschik; H. Reichenbach; V. Wray; G. Höfle, *Liebigs Ann. Chem.* **1994**, *8*, 759.
105. N. Lindquist; W. Fenical; G. D. Vanduyne; J. Clardy, *J. Am. Chem. Soc.* **1991**, *113*, 2303–2304.
106. L. C. Du; M. Chen; C. Sanchez; B. Shen, *FEMS Microbiol. Lett.* **2000**, *189*, 171–175.
107. T. L. Schneider; B. Shen; C. T. Walsh, *Biochemistry* **2003**, *42*, 9722–9730.
108. B. Silakowski; H. U. Schairer; H. Ehret; B. Kunze; S. Weinig; G. Nordsiek; P. Brandt; H. Blocker; G. Hofle; S. Beyer; R. Muller, *J. Biol. Chem.* **1999**, *274*, 37391–37399.
109. C. Reimmann; H. M. Patel; L. Serino; M. Barone; C. T. Walsh; D. Haas, *J. Bacteriol.* **2001**, *183*, 813–820.
110. G. Schoenafinger; N. Schracke; U. Linne; M. A. Marahiel, *J. Am. Chem. Soc.* **2006**, *128*, 7406–7407.
111. T. Weber; R. Baumgartner; C. Renner; M. A. Marahiel; T. A. Holak, *Structure* **2000**, *8*, 407–418.
112. S. Jenni; M. Leibundgut; T. Maier; N. Ban, *Science* **2006**, *311*, 1263–1267.
113. K. J. Weissman; R. Muller, *ChemBioChem* **2008**, *9*, 826–848.
114. A. Koglin; M. R. Mofid; F. Lohr; B. Schafer; V. V. Rogov; M. M. Blum; T. Mittag; M. A. Marahiel; F. Bernhard; V. Dotsch, *Science* **2006**, *312*, 273–276.
115. Z. Zhou; J. R. Lai; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11621–11626.
116. S. E. Evans; C. Williams; C. J. Arthur; S. G. Burston; T. J. Simpson; J. Crosby; M. P. Crump, *ChemBioChem* **2008**, *9*, 2424–2432.
117. E. Conti; T. Stachelhaus; M. A. Marahiel; P. Brick, *EMBO J.* **1997**, *16*, 4174–4183.
118. E. Conti; N. P. Franks; P. Brick, *Structure* **1996**, *4*, 287–298.
119. C. Rausch; T. Weber; O. Kohlbacher; W. Wohlleben; D. H. Huson, *Nucleic Acids Res.* **2005**, *33*, 5799–57808.
120. T. Stachelhaus; H. D. Mootz; M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493–505.
121. G. L. Challis, *J. Med. Chem.* **2008**, *51*, 2618–2628.
122. J. J. May; N. Kessler; M. A. Marahiel; M. T. Stubbs, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12120 –12115.
123. A. S. Reger; R. Wu; D. Dunaway-Mariano; A. M. Gulick, *Biochemistry* **2008**, *47*, 8016–8025.
124. A. M. Gulick; X. Lu; D. Dunaway-Mariano, *Biochemistry* **2004**, *43*, 8670–8679.
125. R. Wu; J. Cao; X. Lu; A. S. Reger; A. M. Gulick; D. Dunaway-Mariano, *Biochemistry* **2008**, *47*, 8026–8039.
126. L. Du; Y. He; Y. Luo, *Biochemistry* **2008**, *47*, 11473–11480.
127. H. Yonus; P. Neumann; S. Zimmermann; J. J. May; M. A. Marahiel; M. T. Stubbs, *J. Biol. Chem.* **2008**, *283*, 32484–32491.
128. A. G. Leslie, *J. Mol. Biol.* **1990**, *213*, 167–186.
129. A. Lewendon; I. A. Murray; C. Kleanthous; P. M. Cullis; W. V. Shaw, *Biochemistry* **1988**, *27*, 7385–7390.
130. C. G. Marshall; N. J. Hillson; C. T. Walsh, *Biochemistry* **2002**, *41*, 244–250.
131. R. M. Kohli; C. T. Walsh, *Chem. Commun.* **2003**, *3*, 297–307.
132. M. Nardini; B. W. Dijkstra, *Curr. Opin. Struct. Biol.* **1999**, *9*, 732–737.
133. S. D. Bruner; T. Weber; R. M. Kohli; D. Schwarzer; M. A. Marahiel; C. T. Walsh; M. T. Stubbs, *Structure* **2002**, *10*, 301–310.
134. J. J. Bellizzi III; J. Widom; C. Kemp; J. Y. Lu; A. K. Das; S. L. Hofmann; J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4573–4578.
135. Y. Devedjiev; Z. Dauter; S. R. Kuznetsov; T. L. Jones; Z. S. Derewenda, *Structure* **2000**, *8*, 1137–1146.
136. J. D. Schrag; F. K. Winkler; M. Cygler, *J. Biol. Chem.* **1992**, *267*, 4300–4303.
137. S. A. Samel; B. Wagner; M. A. Marahiel; L. O. Essen, *J. Mol. Biol.* **2006**, *359*, 876–889.
138. D. Chen; R. Wu; T. L. Bryan; D. Dunaway-Mariano, *Biochemistry* **2009**, *48*, 511–513.
139. D. Schwarzer; H. D. Mootz; U. Linne; M. A. Marahiel, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14083–14088.
140. E. Yeh; R. M. Kohli; S. D. Bruner; C. T. Walsh, *ChemBioChem* **2004**, *5*, 1290–1293.
141. A. Koglin; F. Lohr; F. Bernhard; V. V. Rogov; D. P. Frueh; E. R. Strieter; M. R. Mofid; P. Guntert; G. Wagner; C. T. Walsh; M. A. Marahiel; V. Dotsch, *Nature* **2008**, *454*, 907–911.
142. S. A. Samel; G. Schoenafinger; T. A. Knappe; M. A. Marahiel; L. O. Essen, *Structure* **2007**, *15*, 781–792.
143. D. P. Frueh; H. Arthanari; A. Koglin; D. A. Vosburg; A. E. Bennett; C. T. Walsh; G. Wagner, *Nature* **2008**, *454*, 903–906.
144. A. Tanovic; S. A. Samel; L. O. Essen; M. A. Marahiel, *Science* **2008**, *321*, 659–663.
145. J. E. Cronan, Jr., *J. Bacteriol.* **1980**, *141*, 1291–1297.
146. T. P. Begley; C. Kinsland; E. Strauss, *Vitam. Horm.* **2001**, *61*, 157–171.
147. E. Adams, *Adv. Enzyme Mol. Biol.* **1976**, *44*, 69–138.
148. R. A. John, *Biochim. Biophys. Acta* **1995**, *1248*, 81–96.
149. J. M. Berg; J. L. Tymoczko; L. Stryer, *Biochemistry*, 6th ed.; W. H. Freeman and Co.: New York, 2006.
150. M. Miethke; M. A. Marahiel, *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 413–451.
151. J. E. Baldwin; E. Abraham, *Nat. Prod. Rep.* **1988**, *5*, 129–145.
152. V. Miao; M. F. Coeffet-Legal; P. Brian; R. Brost; J. Penn; A. Whiting; S. Martin; R. Ford; I. Parr; M. Bouchard; C. J. Silva; S. K. Wrigley; R. H. Baltz, *Microbiology* **2005**, *151* (Pt. 5), 1507–1523.
153. C. Bormann; V. Mohrle; C. Bruntner, *J. Bacteriol.* **1996**, *178*, 1216–1218.
154. C. Bruntner; B. Lauer; W. Schwarz; V. Mohrle; C. Bormann, *Mol. Gen. Genet.* **1999**, *262*, 102–114.
155. B. Lauer; R. Russwurm; C. Bormann, *Eur. J. Biochem.* **2000**, *267*, 1698–1706.
156. H. Chen; B. K. Hubbard; S. E. O'Connor; C. T. Walsh, *Chem. Biol.* **2002**, *9*, 103–112.

157. D. G. Fujimori; C. T. Walsh, *Curr. Opin. Chem. Biol.* **2007**, *11*, 553–560.
158. C. S. Neumann; D. G. Fujimori; C. T. Walsh, *Chem. Biol.* **2008**, *15*, 99–109.
159. C. Dong; S. Flecks; S. Unversucht; C. Haupt; K. H. van Pee; J. H. Naismith, *Science* **2005**, *309*, 2216–2219.
160. E. Yeh; L. C. Blasiak; A. Koglin; C. L. Drennan; C. T. Walsh, *Biochemistry* **2007**, *46*, 1284–1292.
161. P. C. Dorrestein; E. Yeh; S. Garneau-Tsodikova; N. L. Kelleher; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13843–13848.
162. L. C. Blasiak; F. H. Vaillancourt; C. T. Walsh; C. L. Drennan, *Nature* **2006**, *440*, 368–371.
163. F. H. Vaillancourt; E. Yeh; D. A. Vosburg; S. E. O'Connor; C. T. Walsh, *Nature* **2005**, *436*, 1191–1194.
164. C. L. Bender; F. Alarcon-Chaidez; D. C. Gross, *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 266–292.
165. E. A. Felnagle; M. R. Rondon; A. D. Berti; H. A. Crosby; M. G. Thomas, *Appl. Environ. Microbiol.* **2007**, *73*, 4162–4170.
166. M. G. Thomas; Y. A. Chan; S. G. Ozanick, *Antimicrob. Agents Chemother.* **2003**, *47*, 2823–2830.
167. X. Yin; T. O'Hare; S. J. Gould; T. M. Zabriskie, *Gene* **2003**, *312*, 215–224.
168. X. Fei; X. Yin; L. Zhang; T. M. Zabriskie, *J. Nat. Prod.* **2007**, *70*, 618–622.
169. J. Ju; S. G. Ozanick; B. Shen; M. G. Thomas, *ChemBioChem* **2004**, *5*, 1281–1285.
170. B. K. Hubbard; M. G. Thomas; C. T. Walsh, *Chem. Biol.* **2000**, *7*, 931–942.
171. T. L. Li; O. W. Choroba; E. H. Charles; A. M. Sandercock; D. H. Williams; J. B. Spencer, *Chem. Commun.* **2001**, *18*, 1752–1753.
172. J. Brownlee; P. He; G. R. Moran; D. H. Harrison, *Biochemistry* **2008**, *47*, 2002–2013.
173. P. A. Frey; G. H. Reed, *Arch. Biochem. Biophys.* **2000**, *382*, 6–14.
174. F. Berkovitch; E. Behshad; K. H. Tang; E. A. Enns; P. A. Frey; C. L. Drennan, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15870–15875.
175. B. W. Lepore; F. J. Ruzicka; P. A. Frey; D. Ringe, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13819–13824.
176. L. Poppe; J. Retey, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 3668–3688.
177. L. Poppe, *Curr. Opin. Chem. Biol.* **2001**, *5*, 512–524.
178. C. V. Christianson; T. J. Montavon; G. M. Festin; H. A. Cooke; B. Shen; S. D. Bruner, *J. Am. Chem. Soc.* **2007**, *129*, 15744–15745.
179. C. V. Christianson; T. J. Montavon; S. G. Van Lanen; B. Shen; S. D. Bruner, *Biochemistry* **2007**, *46*, 7205–7214.
180. H. Chen; C. C. Tseng; B. K. Hubbard; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14901–14906.
181. V. Pfeifer; G. J. Nicholson; J. Ries; J. Recktenwald; A. B. Schefer; R. M. Shawky; J. Schroder; W. Wohlleben; S. Pelzer, *J. Biol. Chem.* **2001**, *276*, 38370–38377.
182. C. C. Tseng; F. H. Vaillancourt; S. D. Bruner; C. T. Walsh, *Chem. Biol.* **2004**, *11*, 1195–1203.
183. P. F. Widboom; E. N. Fielding; Y. Liu; S. D. Bruner, *Nature* **2007**, *447*, 342–345.
184. M. Offenzeller; G. Santer; K. Totschnig; Z. Su; H. Moser; R. Traber; E. Schneider-Scherzer, *Biochemistry* **1996**, *35*, 8401–8412.
185. M. Offenzeller; Z. Su; G. Santer; H. Moser; R. Traber; K. Memmert; E. Schneider-Scherzer, *J. Biol. Chem.* **1993**, *268*, 26127–26134.
186. D. W. Udwary; L. Zeigler; R. N. Asolkar; V. Singan; A. Lapidus; W. Fenical; P. R. Jensen; B. S. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10376–10381.
187. L. L. Beer; B. S. Moore, *Org. Lett.* **2007**, *9*, 845–848.
188. P. M. Flatt; S. J. O'Connell; K. L. McPhail; G. Zeller; C. L. Willis; D. H. Sherman; W. H. Gerwick, *J. Nat. Prod.* **2006**, *69*, 938–944.
189. D. P. Galonic; F. H. Vaillancourt; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 3900–3901.
190. E. S. Sattely; M. A. Fischbach; C. T. Walsh, *Nat. Prod. Rep.* **2008**, *25*, 757–793.
191. D. V. Santi; R. W. Webster; W. W. Cleland, *Methods Enzymol.* **1974**, *29*, 620–627.
192. L. G. Otten; M. L. Schaffer; B. R. M. Villiers; T. Stachelhaus; F. Hollfelder, *Biotechnol. J.* **2007**, *2*, 232–240.
193. P. C. Dorrestein; N. L. Kelleher, *Nat. Prod. Rep.* **2006**, *23*, 893–918.
194. Y. K. Zou; J. Yin, *ChemBioChem* **2008**, *9*, 2804–2810.
195. R. Finking; A. Neumuller; J. Solsbacher; D. Konz; G. Kretzschmar; M. Schweitzer; T. Krumm; M. A. Marahiel, *ChemBioChem* **2003**, *4*, 903–906.
196. H. Ueda; Y. Shoku; N. Hayashi; J. Mitsunaga; Y. In; M. Doi; M. Inoue; T. Ishida, *Biochim. Biophys. Acta* **1991**, *1080*, 126–134.
197. A. K. Forrest; R. L. Jarvest; L. M. Mensah; P. J. O'Hanlon; A. J. Pope; R. Sheppard, *J. Bioorg. Med. Chem. Lett.* **2000**, *10*, 1871–1874.
198. S. W. Bearden; J. D. Fetherston; R. D. Perry, *Infect. Immun.* **1997**, *65*, 1659–1668.
199. I. Smith, *Clin. Microbiol. Rev.* **2003**, *16*, 463–496.
200. J. A. Ferreras; J. S. Ryu; F. Di Lello; D. S. Tan; L. E. N. Quadri, *Nat. Chem. Biol.* **2005**, *1*, 29–32.
201. J. Vannada; E. M. Bennett; D. J. Wilson; H. I. Boshoff; C. E. Barry; C. C. Aldrich, *Org. Lett.* **2006**, *8*, 4707–4710.
202. J. Neres; N. P. Labello; R. V. Somu; H. I. Boshoff; D. J. Wilson; J. Vannada; L. Chen; C. E. Barry; E. M. Bennett; C. C. Aldrich, *J. Med. Chem.* **2008**, *51*, 5349–5370.
203. J. Neres; D. J. Wilson; L. Celia; B. J. Beck; C. C. Aldrich, *Biochemistry* **2008**, *47*, 11735–11749.
204. J. S. Cisar; J. A. Ferreras; R. K. Soni; L. E. N. Quadri; D. S. Tan, *J. Am. Chem. Soc.* **2007**, *129*, 7752–7753.
205. D. E. Ehmann; J. W. Trauger; T. Stachelhaus; C. T. Walsh, *Chem. Biol.* **2000**, *7*, 765–772.
206. J. W. Trauger; R. M. Kohli; H. D. Mootz; M. A. Marahiel; C. T. Walsh, *Nature* **2000**, *407*, 215–218.
207. T. Hornbogen; S. P. Riechers; B. Prinz; J. Schultchen; C. Lang; S. Schmidt; C. Mugge; S. Turkanovic; R. D. Sussmuth; E. Tauberger; R. Zocher, *ChemBioChem* **2007**, *8*, 1048–1054.
208. S. A. Sieber; J. H. Tao; C. T. Walsh; M. A. Marahiel, *Angew. Chem. Int. Ed.* **2004**, *43*, 493–498.
209. L. E. N. Quadri; P. H. Weinreb; M. Lei; M. M. Nakano; P. Zuber; C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595.
210. S. L. Clugston; S. A. Sieber; M. A. Marahiel; C. T. Walsh, *Biochemistry* **2003**, *42*, 12095–12104.
211. N. Kessler; H. Schuhmann; S. Morneweg; U. Linne; M. A. Marahiel, *J. Biol. Chem.* **2004**, *279*, 7413–7419.
212. C. H. Qiao; D. J. Wilson; E. M. Bennett; C. C. Aldrich, *J. Am. Chem. Soc.* **2007**, *129*, 6350–6351.
213. Y. Liu; S. D. Bruner, *ChemBioChem* **2007**, *8*, 617–621.

214. A. S. Worthington; H. Rivera; J. W. Torpey; M. D. Alexander; M. D. Burkart, *ACS Chem. Biol.* **2006**, *1*, 687–691.
215. H. D. Mootz; M. A. Marahiel, *Curr. Opin. Biotechnol.* **1999**, *10*, 341–348.
216. J. Staunton; B. Wilkinson, *Curr. Opin. Chem. Biol.* **2001**, *5*, 159–164.
217. C. D. Reeves, *Crit. Rev. Biotechnol.* **2003**, *23*, 95–147.
218. J. L. Fortman; D. H. Sherman, *ChemBioChem* **2005**, *6*, 960–978.
219. R. H. Baltz, *Nat. Biotechnol.* **2006**, *24*, 1533–1540.
220. H. G. Menzella; C. D. Reeves, *Curr. Opin. Microbiol.* **2007**, *10*, 238–245.

Biographical Sketches

Timothy J. Montavon was born in Wheelersburg, OH, in 1982. He attended Xavier's University in Cincinnati, OH, where he earned a B.S. in chemistry in 2004. While at Xavier's, he worked in the laboratory of Professor Aaron Baba developing novel fluoroionophores. In 2004, he began attending graduate school at Boston College working in the laboratory of Professor Steven Bruner. His current research is focused on the structure and mechanism of aminomutase enzymes used in the biosynthesis of enediyne natural products.

Steven D. Bruner received his B.S. degree from Boston College working in the laboratory of Marc Snapper. He received a Ph.D. degree in chemistry in 2000 from Harvard University under the direction of Professor Gregory Verdine. From 2000 to 2003, he was a Damon Runyon Cancer Research postdoctoral fellow in the laboratory of Professor Christopher Walsh at Harvard Medical School. In 2003, he joined the faculty in the Department of Chemistry at Boston College as an assistant professor. Research in his laboratory focuses on the use of structural biology and synthetic chemistry to characterize natural product biosynthetic enzymes.

5.20 The Properties, Formation, and Biological Activity of 2,5-Diketopiperazines

P. J. Milne and G. Kilian, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa

© 2010 Elsevier Ltd. All rights reserved.

5.20.1 Introduction

In modern times, it is a known fact that very few new medicines are being developed to the detriment of poorer countries and communities. The fall in the number of new drugs is reaching a crisis point, especially as the number of new drug applications worldwide has also declined dramatically.[1]

The battle between research (science) and disease has been constant throughout human history and for all the success we have witnessed in the twentieth century, only about one-third of the 30 000 known diseases can be treated effectively.[2–4] Several diseases remain incurable such as cancer, autoimmune disorders, viral diseases (e.g., influenza), and central nervous system (CNS) disorders (e.g., Alzheimer's disease).[4] The emergence of multidrug-resistant microorganisms have added to the urgency in discovering new novel drugs or agents, a process that can take many years before a product is regarded as *safe and efficacious* and reaches the market place.[5] The costs associated with the development of new drugs (especially clinical evaluation) have increased dramatically and there is a great need for the development of more active and selective drugs with minimal side effects, agents useful in prophylaxis and drugs that will cause less harmful contamination in an already polluted environment.[4,6]

Traditionally, medicinal lead structures have originated from plant, microbial fermentation, and animal sources.[7] Most new drug candidates today are identified through four approaches:

1. Chemical modification of a lead compound;
2. Random screening of natural products;
3. Rational drug design;
4. Biotechnology and cloning.[8]

Pharmaceutical development relies heavily on rational drug design, where therapies are engineered to block or to interfere with a disease's mechanism and synthetic peptides gain increasing importance as both versatile intermediates and active materials.[9] Proteins and peptides control numerous biochemical reactions in the body and may represent an untapped source of new agents for treating a variety of diseases.[10,11] As early as 1984, Vida and Gordon predicted that peptides and synthetic analogues would be among the major drugs of the future[12] owing to the fact that they generally possess properties ideal for optimal receptor interaction, selective mode of action, and clearance from the site of action.

The discovery of a number of naturally occurring peptides with potent and specific biological activities has further increased their importance.[13] Peptide ligands can act as either agonists or antagonists at cell-surface receptors or acceptors that modulate cell function and animal behavior. This area encompasses approximately 50% of current drugs and will become even more important in the future. Biological and behavioral activities that are controlled or modulated by these interactions include response to stress, pain, inflammation, addiction, learning, memory, feeding behavior, sexual behavior, reproduction, the immune response, thermal control, cardiovascular function/regulation, and kidney function. The development of peptide or peptidomimetic ligands that can target receptors or acceptors that modulate or control these biological activities is one of the challenges of pharmaceutical scientists.[14–16]

A form of peptidomimetics (a peptidomimetic is a compound that acts as a substitute for peptides in their interaction with receptors) is the formation of DKPs where amino acid residues are cyclized to form a stable ring structure. The design and development of peptidomimetics allows for the generation of new pharmacotherapeutic agents with the potential of exerting a more desirable or favorable physiological response than its prototype.[17]

Ideally, a peptidomimetic agent should have one or more of the following properties: (1) high specificity/selectivity for a single receptor, (2) metabolic stability, (3) prolonged duration of action, (4) high receptor affinity and thus potency, (5) minimal adverse effects, and (6) good bioavailability.[14,18]

Biologically active peptides range in size from small molecules, containing only two or three amino acids, to large molecules containing numerous amino acids.[13] Peptides are used by most tissues for cell-to-cell communication and play important roles in both autonomic and CNS function. It has been suggested that short antibiotic peptides may serve as a blueprint for the design of novel chemotherapeutic agents, partly due to their prevalence in nature as part of the intrinsic defense mechanisms of most organisms.[19] Fernandez-Lopez *et al.*[20]

has shown that certain peptides with a particular sequence and cyclic D,L-α-peptide structure is effective in killing Gram-positive and Gram-negative bacteria.

The use of peptides as drugs is limited by the following factors:

1. Their low metabolic stability toward proteolysis in the gastrointestinal tract and in serum.
2. Their poor absorption after oral ingestion, in particular due to their relatively high molecular mass or the lack of specific transport systems or both.
3. Their undesired effect caused by interaction of the conformationally flexible peptides with various receptors.
4. Their rapid excretion through liver and kidneys.
5. They are quickly hydrolyzed by plasma and tissue peptidases, and exhibit a short biological half-life, although there are exceptions to this.
6. They fail to cross or penetrate the blood–brain barrier.
7. They are expensive to manufacture.[14,21,22]

Therefore, the direction of research chemists in peptide design and synthesis is to modify the peptide structure into bioactive compounds with improved properties. Structure variations of the native peptides can range from changing or modifying the amino acid side chains or the peptide main chain into peptide analogues, wherein only the electronic and steric aspects of the parent peptide, which proved to be essential for activity, are retained. Cyclization or simple chemical modification of the peptide bond itself, or completely replacing it with a surrogate, could protect the peptide against enzymatic degradation, increasing the peptide's half-life and biological activity.[14,23] The rate of degradation by peptidases and other enzymes are reduced by isosteric structures,[24,25] retro-inverso peptides,[26] cyclic peptides, and nonpeptidomimetics.[27]

In the following sections, a relatively unexplored class of drug, commonly referred to as cyclic dipeptides are discussed. These heterocyclic compounds are 'modified' peptide drugs, having promising biological activity and favorable pharmacokinetic profiles. Cyclic peptides are often prepared as stepping stones in developing pharmaceutical compounds from linear peptide lead molecules or precursors.

5.20.2 What Are Cyclic Dipeptides?

Cyclic dipeptides are heterocyclic compounds comprising of two amino acid residues linked to a central diketopiperazine (DPK) ring structure. The general structure for DKPs can be seen in **Figure 1**.[27–30]

Substitution of the amino side-chain groups (R^1 and R^2) with any of the 20 endogenous L-α amino acids results in numerous potential chemical structures with varying degrees of biological activities. In addition, there are no limitations to the use of the D-enantiomers of the respective amino acids, thus adding to the number or multitude of permutations possible.

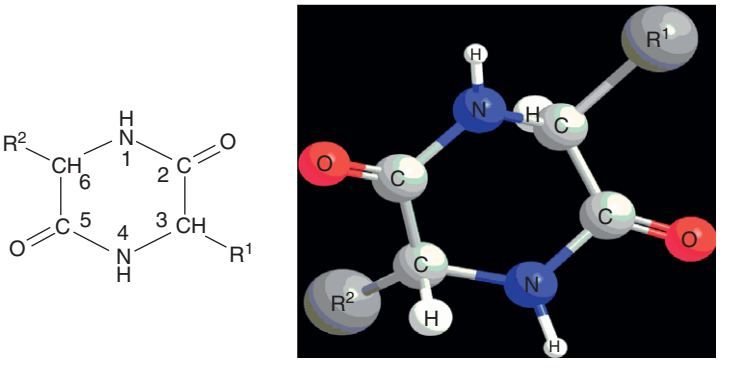


Figure 1 General structure of a 2,5-diketopiperazine, where R^1 and R^2 represent the substituting amino acid residues.

DKPs also comprise out of two other isomers, all possessing a piperazine core. The isomers, 2,3-DKPs, 2,5-DKPs, and 2,6-DKPs can be distinguished on the basis of intramolecular C–N cyclization, tandem reactions, and synthetic methods used. 2,3-DKPs have been used in medicinal chemistry and are found in natural products such as antibiotics (piperacillin),[31] cefoperazone,[32] and bicyclomycin.[33] 2,5-DKPs are common naturally occurring peptide derivatives and are frequently generated as unwanted by-products in the synthesis of oligopeptides.[34] 2,6-DKPs have been investigated as antiproliferative agents through the inhibition of DNA topoisomerase 11.[35]

5.20.3 Historical Aspects and Naturally Occurring Diketopiperazines

Cyclic dipeptides also known as cyclo dipeptides, DKPs, 2,5-dioxopiperazines (DOPs), or dipeptide anhydrides, are among the most common peptide derivatives found in nature.[36] Most cyclic dipeptides found to date appear to have emerged as by-products of fermentation and food processing. Other resulted from nonenzymatic cyclization of dipeptides and their amides during chemical and thermal manipulations[27,37–42] and during storage of proteins and peptides.[43] However, many are endogenous to the members of the animal and the plant kingdoms including marine sponges.[44–46] DKPs have been detected in proteins and polypeptide hydrolysates as well as fermentation broths and cultures of lichens, fungi, and yeast.[27,39,47–52] Cyclo(His-Pro) is the only DKP that has been conclusively shown to be endogenous to mammals (see **Table 1** for specific examples).[36,53]

DKPs have also been detected in a variety of processed foods and beverages, including various cereal grains,[64] cocoa,[65] Comte cheese,[66] hydrolyzed vegetable protein,[67] aged saki,[68] dried squid,[69] beer,[70] and roasted coffee.[71] Several DKPs were found to contribute to the perceived bitterness among some of the processed foods and beverages mentioned above. Gautschi and Schmid[70] studied a market place cross section of five commercial beers, manufactured in different countries, and reported cyclo(Phe-Pro) to be the most prominent DKP, present at concentrations up to 0.25 mmol l^{-1}.

5.20.4 Structural Types and Classes of Cyclic Dipeptides

The diversity of structural types within this group of compounds is well recognized. Sammes[27] arbitrarily divided DKPs into five main structural types or groups: (1) simple dioxypiperazines, (2) the echinulins and related derivatives, (3) hydroxypyrazine derivatives, (4) sulfur-bridged derivatives, and (5) bicyclomycin and dibromophakellin.

Table 1 Some naturally occurring simple cyclic dipeptides in the protist and plant kingdoms

Diketopiperazine (all amino acids are in the L-configuration)	*Species*	*Common name*
Cyclo(Pro-Leu)	*Rosellinia necatrix*	Fungus[54]
	Aspergillus fumigatus	Fungus[50]
Cyclo(Pro-Val)	*Rosellinia necatrix*	Fungus[54]
	Aspergillus ochraceus	Fungus[55]
	Metarrhizum ansiopha	Fungus (mold)[55]
Cyclo(Pro-Phe)	*Rosellinia necatrix*	Fungus[54]
Cyclo(Pro-Tyr)	*Alternaria alternate*	Fungus[56]
Cyclo(Pro-Trp)	*Penicillium brevicompactum*	Fungus (mold)[57]
Cyclo(Phe-Phe)	*Penicillium nigricans*	Fungus (mold)[58]
	Streptomyces noursei	49
Cyclo(Ala-Leu)	*Aspergillus niger*	Fungus (black mold)[59]
3,6-bis-2-(5-chloropiperazine)-2,5-dioxopiperazine	*Streptomyces griseoluteus*	60
Cycloserine dimer	*Streptomyces orchidaceus*	Actinomycetes[61]
Picroroccellin	*Rocella fuciformis*	Lichen[51]
Rhodoturulic acid	*Rhodotorula pilimanae*	Fungus (yeast)[48]
Albonoursin	*Streptomyces noursei*	Actinomycetes[62]
Amphomycin	*Streptomyces canus*	Actinomycetes[47,63]

Source: Adapted from C. Prasad, *Peptides* **1995**, *16*, 151–164.

Cyclic peptides are divided into two broad classes: those in which the ring is built entirely of amino acids, called homomeric cyclic peptides, and when one or more of the building units of the phenyl-containing homomeric ring are not amino acids, they are called heteromeric cyclic peptides. Homomeric cyclic peptides in which the amino acids of the ring are joined together through amide bonds are referred to as homodetic cyclic peptides. In this group many of the amide bonds are typical peptide bonds formed between carboxyl groups and α-amino groups such as that found in gramicidin S and tyrocidins. In other members of this group such as the polymixins, some of the amide bonds of the ring are formed with amino groups other than those alpha to the carboxyl function.[72,73]

In heterodetic peptides, the ring system is formed from amides and other heteroatom linkages, for example, a disulfide bridge, a thioether or a lactone linkage.[72] The most common nonamide links in heterodetic peptides are the disulfide bonds (found in posterior pituitary hormones such as oxytocin and vasopressin) and the ester (lactone) function.[74]

Cyclic dipeptides, bearing the DKP ring, in which the amide linkages are *cis*, are known to represent the simplest form of homodetic cyclic peptides.[75]

5.20.5 Properties of Diketopiperazines

5.20.5.1 Scope of Interest

The properties of DKPs differ from those of ordinary peptides in that:

1. They do not exist as zwitterions and are often neutral compounds.[27]
2. The less structurally complex members of this group are water soluble, although derivatives of phenylalanine (Phe) are only sparingly soluble.[27]
3. They are of interest in studies on the thermodynamic behavior of nonionic compounds in aqueous media because they have the ability of forming hydrogen bonds with the solvent (through the two *cis*-amide groups in the cyclic dipeptide ring), and give rise to hydrophobic interactions.[29]
4. DKPs are important metabolic intermediates and provoke the destruction of the secondary globular protein structure[36,76] (especially nonionic polar DKPs exhibit the requisites to act as denaturing agents).[29]
5. They are relatively simple compounds and aromatic- and/or Pro-containing DKPs serve as excellent models for theoretical studies.[30,77]
6. Their physicochemical and biological properties are dependent on the spatial arrangement of their constituent atoms.[78,79]
7. They are well known for their biological importance as hormones, antibiotics, toxins, ion-transport regulators, antivirals, phytotoxic compounds, and so on.[27,80]
8. DKP drugs, which were developed as analgesics and anti-Parkinson's agents, have entered the clinical trial phase.[81–83]
9. They do not have backbone terminal charged groups.[84]
10. Being conformationally constrained their popularity to determine three-dimensional structure and biological relationships has increased.[84]
11. Arylmethyl-containing DKPs prefer a folded conformation with the aromatic ring face-to-face with the six-membered heterocyclic DKP ring.[85]
12. They can serve as templates or useful scaffolds from which peptidomimetic drugs with enhanced activity can be designed.[86,87]
13. They are sensitive to oxidation, especially when *imino* acid residues are the building blocks.[88]
14. They may act as powerful hydrolytic catalysts.[89,90]
15. They are chirally enriched, easy accessible heterocycles, and owing to their constrained nature, do not exhibit unwanted physical and metabolic properties.[87]
16. Their reduced mobility allows them to be used in the study and mimicry of protein folding[91] and serve as valuable models for gaining conformational insight into the properties of larger peptides and proteins.[28]
17. The stereochemistry of the DKP moiety seems to play an important role in the inhibition of platelet-activating factor (PAF) activity.[92]

Conformationally restricted cyclic dipeptides have several advantages over their linear counterparts:

1. Cyclization of peptides may induce higher binding affinities and a greater degree of specificity in peptide–receptor interactions;[93]
2. Increased *in vivo* stability due to resistance to protease degradation.[94,95]

The reasons for DKPs being more stable than their linear dipeptides are as follows: (1) breakdown of a peptide chain proceeds most easily starting at C- or N-end groups. Therefore, the lack of a free terminal COOH or NH_2 will increase stability; (2) rigidity of the peptide backbone; and (3) reduction of conformational space. Thus in DKPs, degradation is slower and the biological activity is more sustained.[6]

It appears that cyclic dipeptides are indeed an unexplored class of bioactive agents that may hold great promise for the future.[36]

5.20.5.2 Physicochemical Properties of Diketopiperazines

5.20.5.2.1 Solubility

As a result of the intramolecular aminolysis reaction, DKPs are not capable of existing as zwitterions. Simpler members have been shown to exist exclusively as neutral compounds. DKPs do share the capability of establishing hydrogen bonds with the solvents through the two *cis*-amide groups in the DKP ring. In contrast, they may give rise to hydrophobic interactions, the extent of which is determined by the R^1 and R^2 substituents.[29] To date, the solubility of nonionic DKPs in water has been shown to be limited to those containing short alkyl groups, such as cyclo(Ala-Ala).[29,96]

5.20.5.2.2 Physical stability

According to literature, DKPs exist as a fairly stable entity; however, studies have shown that either of the two carbonyl groups of the DKP ring of cyclo(Phe-Pro) can undergo hydrolysis, thereby producing two different dipeptides, that is, Phe-Pro-OH and Pro-Phe-OH. This suggests their possible use as prodrugs. The results would suggest that contrary to the statements made with respect to the relative stability of DKPs, these entities are susceptible to degradation if conditions permit. Studies have shown that between the pH values of 3 and 8, cyclo(Phe-Pro) exists as a relatively stable entity, as previously proposed in the literature. In contrast, at pH values <3 and >8, the DKP nucleus of cyclo(Phe-Pro) undergoes hydrolysis to the dipeptide Phe-Pro-OH.[97]

Short-term stability studies were conducted on cyclo(Gly-Gly), cyclo(Gly-Trp), cyclo(Trp-Trp), cyclo(Gly-Tyr), and cyclo(Gly-Phe) to determine the stability of the dipeptides in solution under various storage conditions. Effects of temperature, pH differences, and buffer catalysis on the hydrolysis of the dipeptides were determined. The hydrolytic products expected were the respective linear dipeptide counterparts, which were obtained commercially. Assessment of the degradation products was undertaken using high-performance liquid chromatography (HPLC).

Varying concentrations of the respective cyclic dipeptides were exposed to a range of temperatures and pH values over a 6- to 24-week period.

Degradation of the dipeptides in aqueous buffer solution was found to be pH dependent and observed to follow second-order kinetics through acid- and base-catalyzed amide hydrolysis reactions. Maximum stability was observed between pH 4.6 and 7.0. The half-life of cyclo(Gly-Gly), cyclo(Gly-Trp), and cyclo(Trp-Trp) revealed that degradation, although significant, would not influence the majority of biological assays (pH values ± 7.4) with regard to loss of drug concentration, resultant of physicochemical degradation.

Buffers at acidic pH values influenced hydrolysis reactions of cyclo(Gly-Phe). Furthermore, exposure of cyclo(Gly-Phe) and cyclo(Gly-Tyr) to 60 °C resulted in hydrolysis of the DKP ring in both strong acids and bases.

Major degradation products were indeed found to be the respective linear counterparts.[98]

5.20.5.2.3 The isomerization factor

Naturally *cis*-fused DKPs are thermodynamically more stable, compared with their *trans*-fused counterparts. This seems logical considering their biosynthetic origin, usually from two proteinogenic L-α-amino acids. Some other *cis*- and *trans*-functional DKPs are derived from nonproteinogenic D-α-amino acids. Naturally

occurring DKPs containing D-α-amino acids other than Pro are not common, but do occur. Cyclo(D-Val-L-Trp) has been isolated from *Aspergillus chevalieri* and cyclo(*N*′-carboxy-D-Trp-D-Ile) has been isolated from the marine sponge *Rhaphisa pallida*.[99,100]

Biological activities of different isomers may differ in many ways:

1. All isomers may be as active as the next one, without exhibiting any stereochemistry of interaction.
2. The isomers may differ in their biological activities: one isomer may not show any biological activities, while other isomers may show unique biological activities.[101]

Substitution of the L-amino acids with their D-amino acid configurational analogue, will enhance resistance to enzymatic cleavage of peptide bonds. The LL isomer of cyclo(Leu-Leu) hydrolyzes 3.5 times faster than the DL isomer in the presence of 0.5 mol l^{-1} HCl. This difference was explained by both steric shielding (LD-*trans* isomer) and steric strain (LL-*cis* isomer).[102]

Peptide bonds involving α-amino acid residues have been shown to be in the *trans* conformation but for Pro-containing and *N*-methylglycine peptides the bond may be *cis* or *trans*. When isomerization of Pro-Pro peptide bond occurs, the α and δ protons of the pyrrolidine ring change positions with respect to the carbonyl group.[103]

5.20.5.2.4 Absorption, transportation, and metabolic stability

Peptide and protein drugs must be transported without metabolic degradation to the systemic circulation in order to exhibit or exert their pharmacological action. Although active transport of linear peptides and oligopeptides by intestinal oligopeptide transporters has been reported, overall intestinal absorption of peptides is very poor because of metabolic degradation by peptidases.[104–107]

The intestinal transport and metabolic stability parameters of various cyclic dipeptides have been investigated by the everted small intestine of a rat.[22,104] DKPs, when compared with their corresponding linear dipeptides, have been shown to be more stable and more absorbable in all parts of the intestine. Cyclo(Gly-Phe) was stable enough to be transported in any region of the small intestine although linear glycylphenylalanine and linear phenylalanylglycine were highly unstable to be transported.[104] Studies on the intestinal transport and metabolism of cyclo(Ser-Tyr), cyclo(Gly-Phe), cyclo(Asp-Phe), cyclo(His-Phe), and cyclo(His-pro) have been done and compared with linear derivatives.[22,104] Constant stability was found with regard to the DKPs being stable in the intestine. A study involving the DKPs, cyclo(Ser-Tyr), cyclo(Asp-Phe), and cyclo(Gly-Phe) showed energy-dependent, saturable transport. In addition, the transport was reduced by the presence of glycylsarcosine or cephalexin, an indication that the transport was due to the oligopeptide transporter.[105–108]

Peptide transporters, for example, PEPT1 (SLC15A1) and PEPT2 (SLC15A2) are integral plasma proteins, which actively transports dipeptides and tripeptides, including peptidomimetics, across cell membranes in animals, microbes, and plants.[109–111] PEPT1 is a low-affinity/high-capacity transporter that is found primarily on the brush-border membrane of the small intestine, and to a lesser extent in the proximal tubule of the kidney. PEPT2 is a high-affinity/low-capacity transporter expressed predominantly in the membrane of epithelial cells of the kidney.[112,113] Both transporters function electrogenically, as they are driven electrochemically through a gradient of protons across the intestinal or renal brush-border membrane in the lumen-to-cytoplasm direction.[114] In addition to amino acid transporters, peptide transporters contribute to the homeostasis of amino acids.[115] Mizuma *et al.*[104] showed that cyclo(Gly-Phe) is much more stable than linear Gly-Phe, and can be transported by PEPT1 in isolated rat small intestine without degradation by intestinal peptidases.

Mizuma *et al.*[116] found that some DKPs are actively transported in Caco-2 cell monolayers by PEPT1. The substrate-binding site of peptide transporters can accommodate a wide range of molecules of differing size, hydrophobicity, thus they represent excellent targets for the delivery of pharmacologically active compounds.[111]

Mizuma *et al.*[116] also studied the uptake of DKPs through the PEPT1 peptide transporters. They noted that cellular uptake of the DKPs was pH dependent, and that their uptake was inhibited by the addition of PEPT1 substrates, which indicates PEPT1-mediated transport of the DKPs. It was concluded that tyrosine (Tyr) had a high affinity for PEPT1 and that the phenolic hydroxyl group may enhance this affinity for PEPT1.

Certain cyclic dipeptides have the ability to be transported by additional paracellular mechanisms, thereby enhancing their transport.[104] Not only absorptive transport but also excretive transport are observed for certain cyclic dipeptides. The intestinal absorption of certain cyclic dipeptides consists of carrier-mediated absorptive

transport and carrier-mediated excretive transport in addition to passive transport. Absorptive transport is mediated by PEPT1 whereas excretive transport is mediated by a transporter that is closely linked to adenosine-5′-triphosphate (ATP). The concentration-dependent preference of these excretive and absorptive transport results in an atypical intestinal absorption of certain cyclic dipeptides.[117] As far as metabolism is concerned, possible recognition of peptide derivatives by hepatic cytochrome P-450 3A has been suggested by binding and metabolism of several pseudopeptidic compounds, for example, cyclosporine and ergot derivatives. Natural linear or cyclic dipeptides containing hydrophobic amino acids produced by microorganisms and present in mammals are able to interact with cytochrome P-450 3A. Some cyclic dipeptides are rapidly transformed by cytochrome P-450 3A to mono- or dihydroxylated metabolites. This metabolism of cyclic dipeptides occurs in eight species including humans.[118]

The DKPs studied to date have shown to be stable in the intestine when compared with their linear derivatives. These results suggest the possibility that DKPs may be orally active.

5.20.5.2.5 Organoleptic properties

A varying degree of bitterness has been reported with regard to the taste of DKPs containing L-leucine and L-tryptophan (Trp).[119–121]

5.20.5.2.6 IR spectroscopy: Characteristic properties

IR spectroscopy permits reliable discrimination between *cis* and *trans*-secondary amide bonds and the presence of 1550 cm^{-1} bands is indicative of *trans*-peptide bonds (**Table 2**).

5.20.5.2.7 Mass spectrometry: Characteristic mass spectral fragmentation patterns

Mass spectrometry (MS) has become a very important and rapid method of determining the structure of DKPs and the dominant features in the MS spectra are fragments corresponding to cleavages of side chains and the rupture of the six-membered DKP ring.[125] The parent or molecular ion is prominent and is followed by other principal fragmentations including the following: (1) loss of CO or CHO, (2) amine fragmentation, and (3) elimination of HNCO.[27,126]

The existence of the DKP ring can be confirmed by the ions m/e 85, m/e 113, and m/e 114, while the structure of the side chains may be determined from the presence of the ion m/e (113 + R) or from the ions $[M-R]^+$. The ions m/e 107, 91, and 128 are characteristic of Tyr, Phe, and alanyl residues, respectively.[125]

Table 2 Frequencies/absorption bands (cm^{-1}) of DKPs

Description of band	*Cis-amide absorption bands*	*Trans-amide absorption bands*
Amide 1 band (CO stretch)	1670–1690[27]	1650
Amide 11 band (NH-in plane vibration)	1440–1450[27]1420–1460[80]	1550[80,124]
Amide 111 (*cis*-CONH)	1300–1350[122]	Not present[122]
NH bending	1450[123]	1450
CN stretching	1350[123]	1350
NH stretching	3180–3195[27]	3350
Combination band of CO stretching and NH bending vibrations of the *cis*-CONH group	3100–3200[123]	3100–3300
CH_2 twisting	1249[123]	
CH_2 bending	1468[123]	
CH_2 wagging	1340[123]	
CH_2 rocking	998[123]	
CO in-plane bending	806[2]	
Skeletal stretching (NMe stretching)	1075[123]	
C-Me stretching	910[123]	
Free NH groups	3420–3480[80]	
Hydrogen-bonded NH groups	3300–3380[80]	

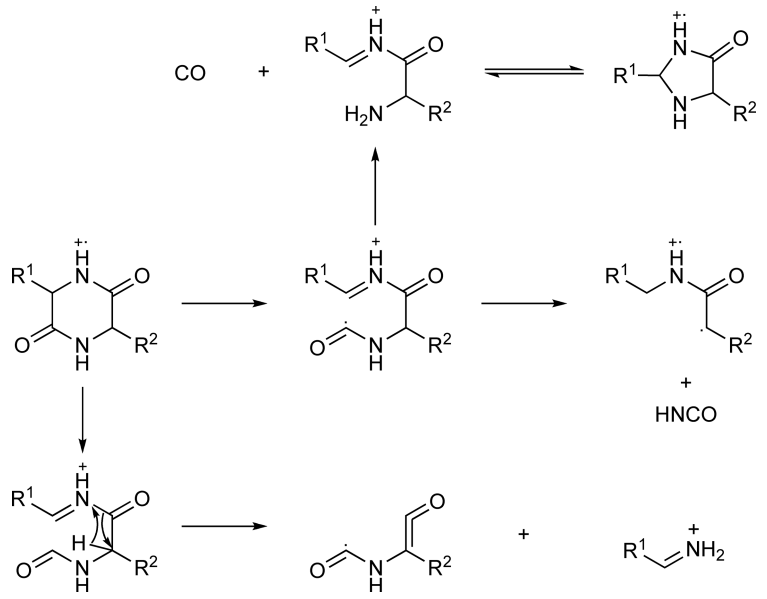


Figure 2 Principal fragmentations of the parent ion of DKPs. Adapted from P. G. Sammes, *Chem. Org. Naturstoffe (Wien)* **1975**, *32*, 51–118.

Studies by Svec and Junk on the mass spectra of dipeptides and DKPs lead to the conclusion that ionization occurs at the nitrogen atom resulting in charge localization, followed by the main fragmentation due to bond ruptures β to the charged nitrogen.

The peaks of some Pro-containing DKPs shift two mass units lower when compared with other DKPs (**Figure 2**).[126]

There are six main decomposition pathways for protonated DKPs and elimination of one of the amino acid side chains R^1 or R^2 produces an ion which is the origin of five other fragmentation pathways. Owing to the symmetric structure of the DKP ring with regard to R^1 and R^2, only half of the fragment ions are shown in **Figure 3**.[86]

5.20.6 Structural Relevance

5.20.6.1 Conformational Features of Diketopiperazines

The DKP backbone is an important pharmacore in peptide research, drug discovery and design, and for the development of peptidomimetics.[127] The DKP backbone is conformationally restrained by a six-membered ring with side chains that are oriented in a spatially defined manner. The DKP ring contains two hydrogen-bond-accepting centers and two hydrogen-bond-donating sites, which are important for potential interactions between the lead compound and receptor sites.[95]

The conformations and molecular structures of cyclic dipeptides containing the DKP ring have been studied by a number of physical techniques, including X-ray crystallography,[128,129] infrared (IR) spectroscopy,[122,130] nuclear magnetic resonance (NMR),[131–133] optical rotatory dispersion (ORD),[134] circular dichroism (CD) spectroscopy,[135] quantum mechanical calculations,[136] and reversed-phase liquid chromatography.[137] From these studies general rules relating to the conformations of DKPs emerged and it can be applied particularly well to aromatic-containing DKPs.

Three conformations of the DKP ring have been identified in the solid state, that is, chairs, boats, and twists (**Figure 4**).[138]

For the planar DKP ring, the β-carbon atom of an L-amino acid is located on the left-hand side, as shown in **Figure 5(a)**. For the boat conformation, the β-carbon atom of an L-amino acid is in the axial position (**Figure 5(b)**). For the bowsprit-boat form, the β-carbon atom of an L-amino acid is in the equatorial position (**Figure 5(c)**). Finally, for the twist-boat form to exist, the β-carbon atom of an L-amino acid must adopt a quasi-equatorial position (**Figure 5(d)**).

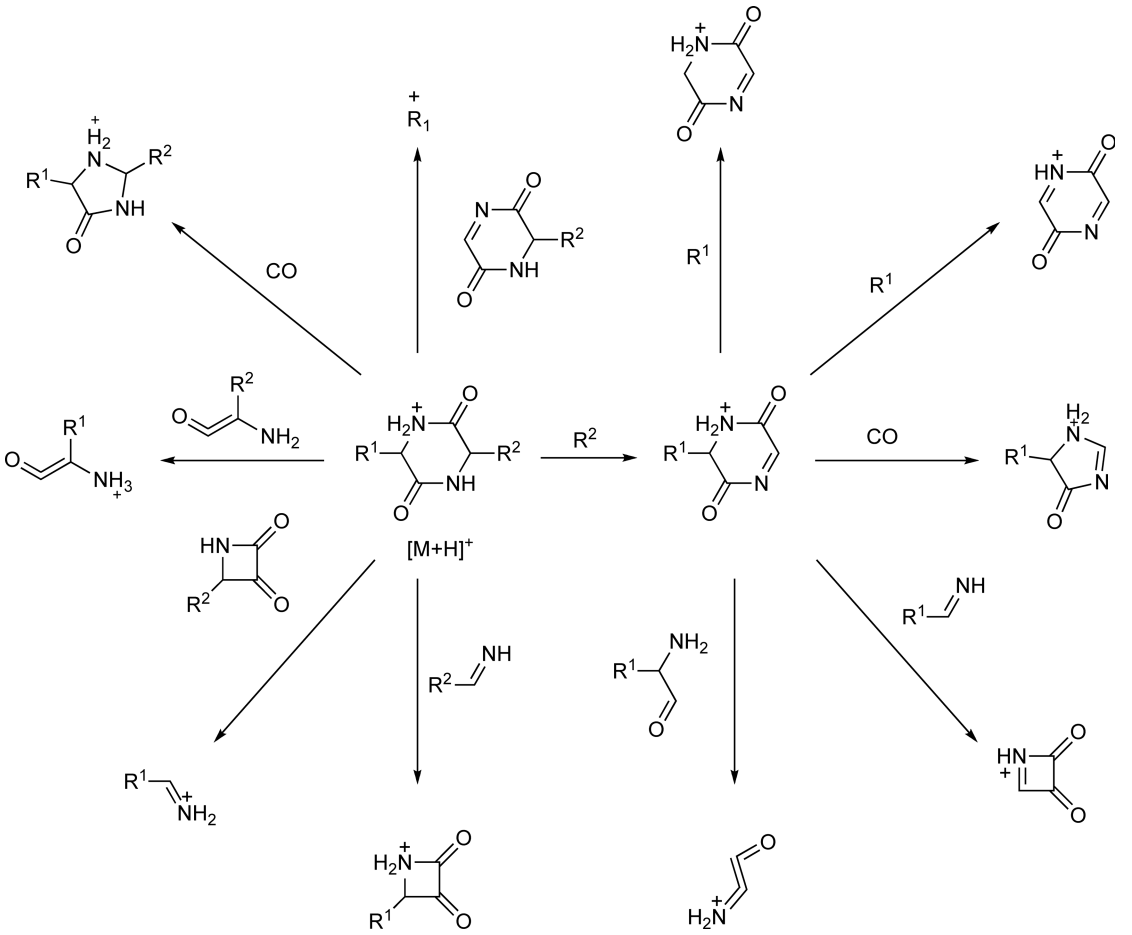


Figure 3 Possible fragmentation pathways for protonated DKPs. Adapted from M. Henczi; D. F. Weaver, *Rapid Commun. Mass Spectrom.* **1995**, *9*, 800–803.

Factors affecting the conformation of the DKP ring are as follows:

1. near planarity of the *cis*-peptide units – the ring will thus be flat or in a boat conformation, but not a chair conformation;
2. chirality;
3. substitutions at the α-carbons atoms – if the substitutions differ and are of the same chirality, the DKP ring is buckled to a boat conformation, which is significant when the side chain contains Pro;[30,132,133,140]
4. when the constituting subunits are N-substituted and have opposite configuration, the chair conformation will result;[84]
5. when a cyclic dipeptide contains aromatic side chains, maximal overlap will occur between the DKP ring and aromatic ring, such as cyclo(Gly-X), where X is an aromatic amino acid.[128] This preferred conformation is called 'flagpole' in which the aromatic ring stacks over the DKP ring;[84]
6. when one considers a molecule such as cyclo(L-Y-L-X), an unfolded, buckled ring will result if Y is bulky. In this context, X refers to aromatic residues, while Y refers to nonaromatic residues. X will interact with the DKP ring without interference between Y and X;[128]
7. if, however, both X and Y are nonaromatic, steric hinderance between X and Y will occur. This will result in the DKP ring taking on a boat conformation.[141]

Young *et al.*[77] have shown that the aromatic side chain of the two Phe-containing rotamers [cyclo(Phe-Pro); cyclo(Phe-D-Pro)] can adopt one of the three predominant conformations: a folded conformation, extended

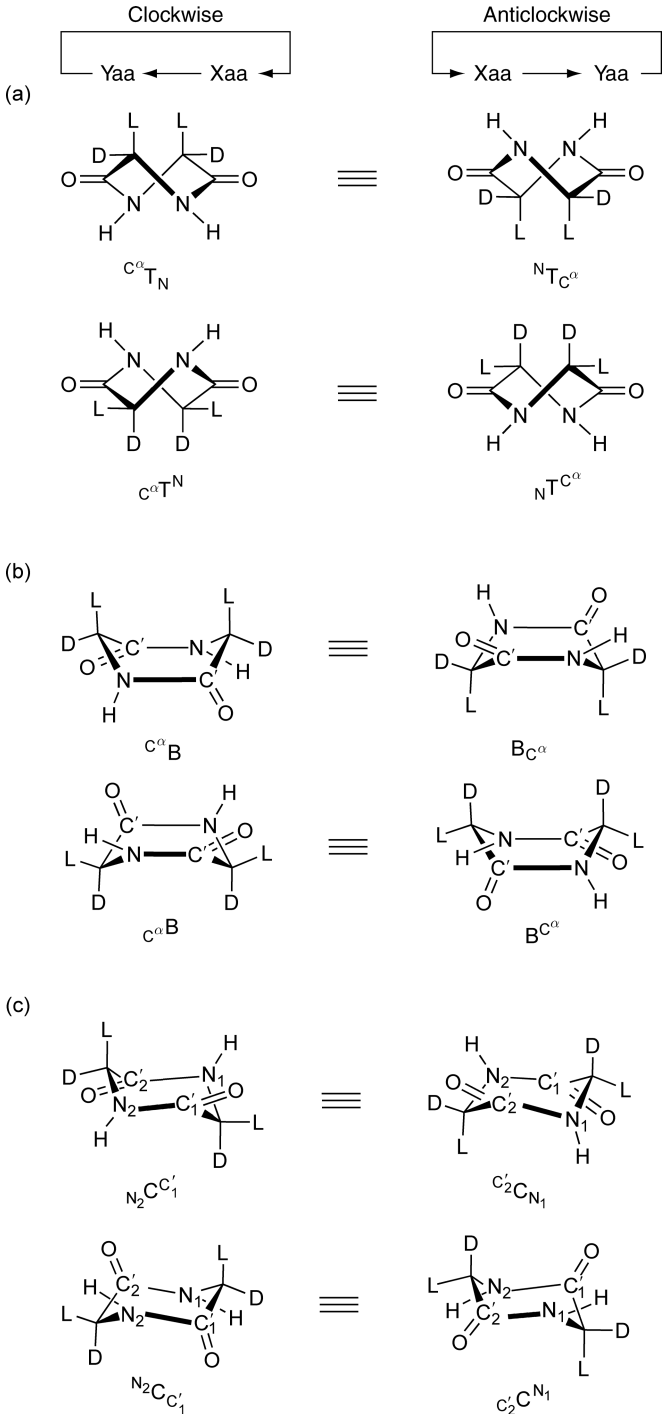


Figure 4 Major conformations of the six-membered DKP ring in the solid state (A = twists, B = boats, C = chairs). Adapted from R. Jankowska; J. Ciarkowski, *Int. J. Peptide Protein Res.* **1987**, *30*, 61–78.

toward the nitrogen (E_N), or extended toward the oxygen. Most cyclic dipeptide derivatives, bearing a single aromatic or Trp side chain, have shown to exist predominantly in the folded form, with only a few of these dipeptides existing as equimolar mixtures of unfolded and folded conformers.[74,77,137,142]

Increasing temperature favors the open or unfolded form.[133] The various forces resulting in a folded conformation of the DKP ring include dipole-induced dipole interactions, van der Waals forces, and interaction

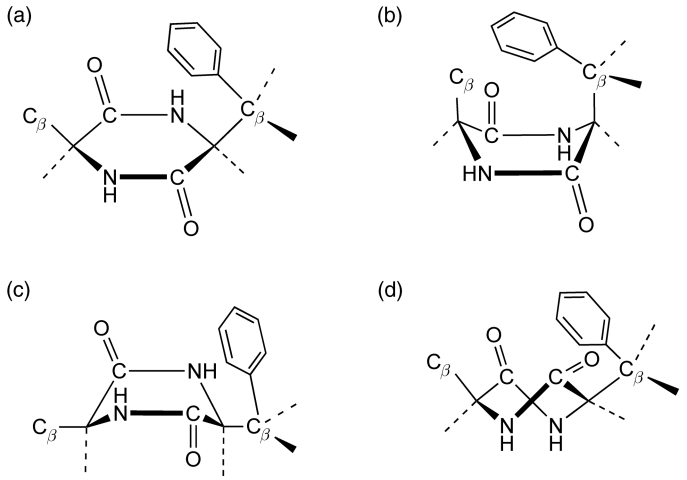


Figure 5 In solution the DKP ring adopts four main conformations: (a) planar, (b) boat, (c) bowsprit-boat, and (d) twist-boat forms.[137,139,140] Adapted from N. Funasaki; S. Hada; S. Neya, *Anal. Chem*. **1993**, *65*, 1861–1867.

between aromatic π electrons, and the polarized π system of the two amide groups.[132,133,140] Derivatives of DKPs containing an aromatic side chain, for example, cyclo(Gly-Phe) and cyclo(Gly-Trp) exhibit shielding of the *cis*-disposed glycyl hydrogen atom. This is explained by the side chain's aromatic ring that assumes a folded conformation over the DKP ring (**Figure 6**).[137,141,143]

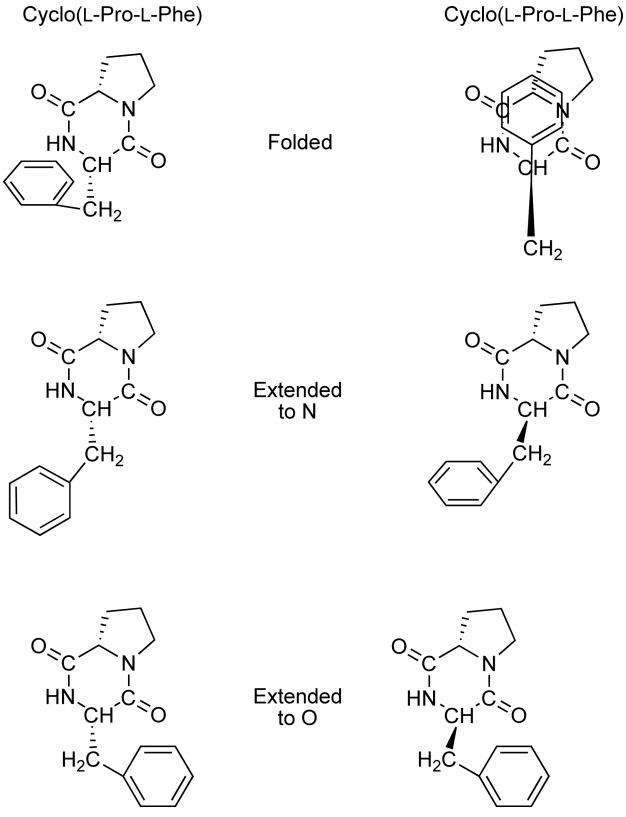


Figure 6 Side-chain conformations of cyclo(Phe-Pro) and cyclo(Phe-D-Pro). Adapted from P. E. Young; V. Madison; E. R. Blout, *J. Am. Chem. Soc*. **1976**, *98* (17), 5365–5371.

5.20.6.2 Side Chain of Proline

The Pro side chain (R group), is unique in that it incorporates the terminal amino group in the side chain. The incorporation of this amine group into a five-membered pyrrolidine ring, constrains the rotational freedom around the N–Cα bond in Pro to a specific rotational angle. This has important consequences for the protein structures in which Pro participates, for example, collagen.[144,145]

The pyrrolidine ring, with its secondary amine also confers unusual chemical and biological properties, compared with the primary amino acids commonly found in proteins (**Figure 7**).[146]

5.20.6.3 Puckering of the Pyrrolidine Ring of Proline

It has been found that C^{γ} and/or C^{β} are the atoms that are usually puckered from the plane of the pyrrolidine ring of Pro. The above phenomenon can be explained as follows: **Figure 8** shows the different structures of the pyrrolidine ring viewed along the bisectors of the C^{α}–N–C^{δ} angles. Results showed that almost half of the structures have a C_2 (half-chair) symmetry or twist,[84,149] with the twofold axis passing through the N atom and the midpoint of the C^{β}–C^{γ} bond.[148] In these structures the $NC^{\alpha}C^{\beta}C^{\delta}$ or $NC^{\alpha}C^{\gamma}C^{\delta}$ groups are not planar and they may be characterized as C_2–C^{γ}-exo (C^{β}-endo) or C_2–C^{γ}-endo (C^{β}-exo) depending on whether C^{γ} is displaced on the same side as C′ (i.e., C_2–C^{γ}-endo) or on the opposite side of (below) the $NC^{\alpha}C^{\delta}$ plane (i.e., C_2–C^{γ}-exo).[148]

The other structures have a Cs (envelope) symmetry, with the mirror plane passing through either the C^{γ} or C^{β} atoms. If the mirror plane passes through C^{β}, χ^4 is close to zero and the atoms $C^{\alpha}C^{\gamma}C^{\delta}N$ are coplanar, the structure may be designated as Cs–C^{β}-endo or Cs–C^{β}-exo depending on the direction of displacement of the C^{β} atom from the $NC^{\alpha}C^{\gamma}C^{\delta}$ plane (i.e., if C^{β} is above the plane of the ring, it is designated as Cs–C^{β}-endo and vice versa). If the C^{γ} atom is on the mirror plane, then θ is close to zero and the four atoms $C^{\alpha}C^{\beta}C^{\delta}N$ are coplanar, the structure may be designated by Cs–C^{γ}-exo or Cs–C^{γ}-endo.[148]

Balasubramanian *et al.*[147] concluded that the Pro ring has two typical puckered conformations: (1) conformation A characterized by negative values of $\chi 1$ (Cγ-exo) (found in residues having α-helix-type torsion angles) and (2) conformation B characterized by positive $\chi 1$ (Cγ-endo) (found in residues having collagen-type torsion angles).

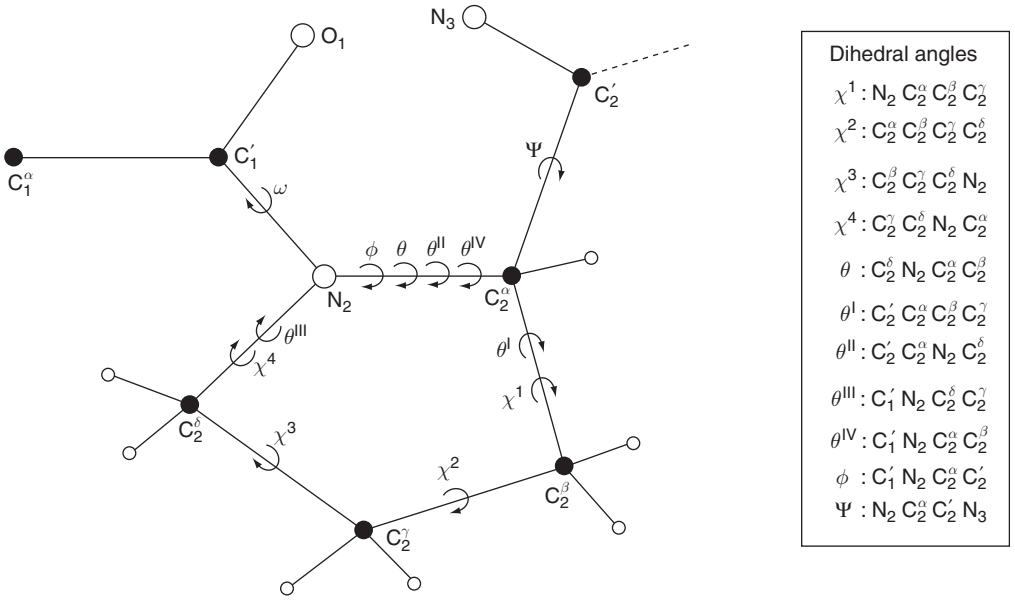


Figure 7 Dihedral angles, atoms, and bonds of a proline ring. Adapted from R. Balasubramanian; A. V. Lakshminarayanan; M. N. Sabesan; G. Tegoni; K. Venkatesan; G. N. Ramachandran, *Int. J. Protein Res.* **1971**, *3*, 25–33.

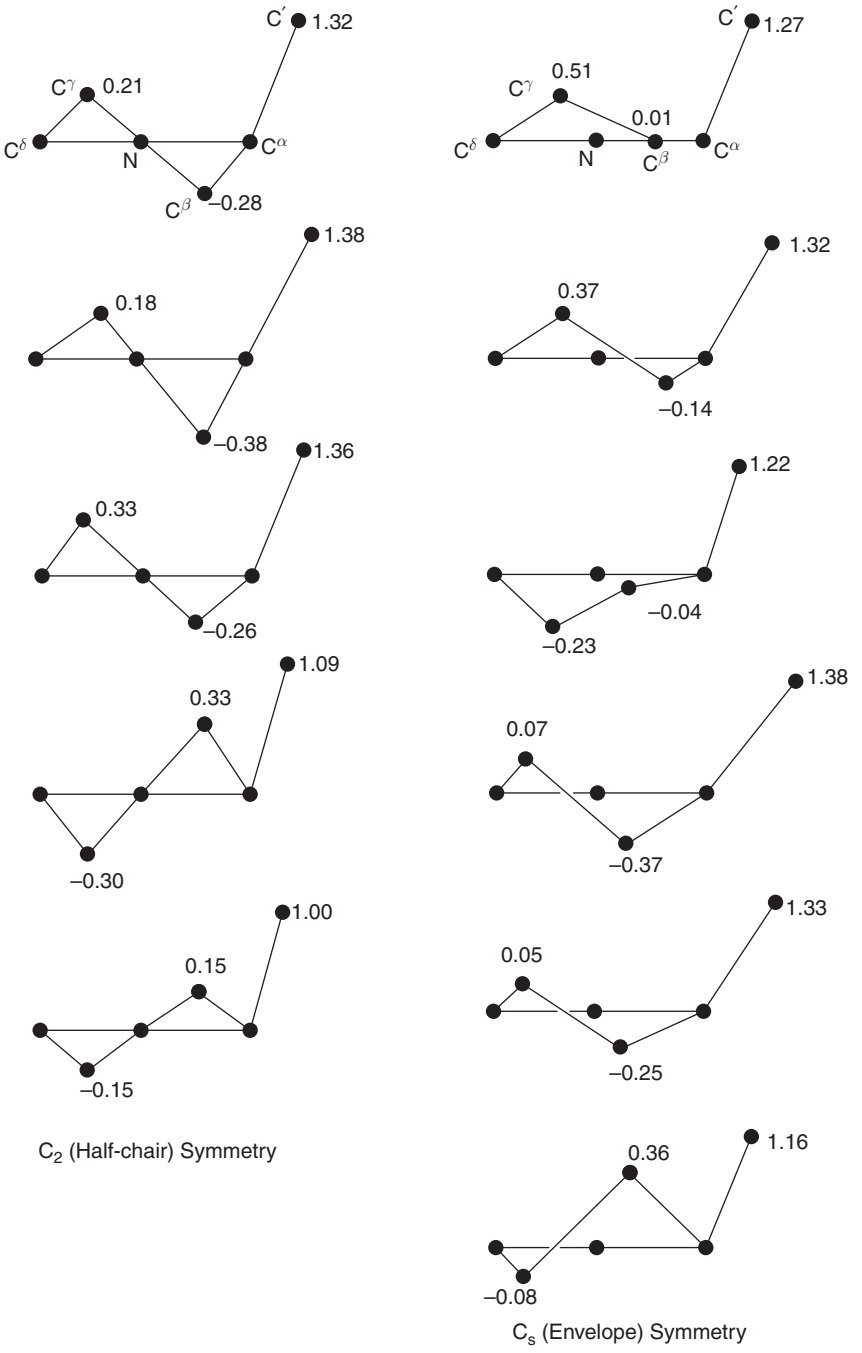


Figure 8 Puckering of the pyrrolidine ring. Adapted from T. Ashida; M. Kakudo, *Bull. Chem. Soc. Jpn.* **1974**, *47* (5), 1129–1133.

5.20.7 Relevance of Amino Acids

5.20.7.1 What Is an Amino Acid?

Amino acids are bifunctional compounds and the basic structural units/building blocks of proteins. In chemistry, an α-amino acid consists of an amino group, a carboxyl group, an R-group, and a hydrogen atom which is bonded to the α-carbon. The R-group represents a side chain specific to each amino acid and

20 different kinds of side chains (varying in shape, size, charge, hydrogen bonding capacity, and chemical reactivity) are commonly found in proteins. Although amino acids exist in either the L- or D-stereoisomer most of the amino acids found in nature are of the L-type and are isolated from proteins that contain only L-amino acids. L- and D- refer to the absolute configuration of optically active compounds.[11,150]

For all other amino acids, with the exception of glycine (Gly), the α-carbon is bonded to four different groups, and the two stereoisomers are mirror images that cannot be superimposed. Eukaryotic proteins are always composed of L-amino acids although D-amino acids are found in certain peptide antibiotics and some peptides of bacterial cell walls.[10,150] The physical properties of amino acids are influenced by the degree of ionization at different pH values.

Amino acids are not only important for protein synthesis but also serve as precursors for hormones, coenzymes, alkaloids, cell wall polymers, porphyrins, antibiotics, nucleotides, pigments, and neurotransmitters.[150] A deficiency of one amino acid may result in a negative nitrogen balance (**Figure 9**).[11]

5.20.7.2 Classification of Amino Acids

There are several possible ways of classifying amino acids. From their isoelectric points, their basic, acidic, or neutral character can be distinguished. Alternatively, we can consider in turn those with aromatic or aliphatic side chains.[151] Therefore, based on the properties of their R-groups (polarity and charge) amino acids are grouped into five main classes: (1) nonpolar (hydrophobic), aliphatic; (2) aromatic; (3) polar (hydrophilic), uncharged; (4) positively charged (basic); and (5) negatively charged (acidic).[150] On the basis of the properties of their side chains, Stryer[11] divided them into seven groups: (1) aliphatic side chains, (2) hydroxyl aliphatic side chains, (3) aromatic side chains, (4) basic side chains, (5) acidic side chains, (6) amide side chains, and (7) sulfur side chains. In biochemical terms they may be divided into either glucogenic (catabolized to pyruvate, α-ketoglutarate, succinyl CoA, fumarate, or oxaloacetate) or ketogenic (gives rise to ketone bodies) amino acids depending on the metabolism of the carbon chains.[11] Alanine (Ala), cysteine (Cys), Gly, serine (Ser), threonine (Thr), and Trp are all glucogenic and are converted into either acetyl-CoA or oxaloacetate.[150] Five amino acids (Phe, Tyr, isoleucine (Ile), Thr, and Trp) are both glucogenic and ketogenic. Only leucine (Leu) and lysine (Lys) are exclusively ketogenic. Both asparagine (Asn) and aspartate are degraded or converted into oxaloacetate whereas the three branched-chain amino acids (Leu, Ile, and valine (Val)) are oxidized as fuels only in extrahepatic tissues, for example, brain, muscle, kidney, and adipose.

Finally, they may be classified on the basis of whether they are nonessential, essential (exogenous), and conditionally essential amino acids.[150–152] Food and tissue proteins contain 20 amino acids of nutritional importance. Nine of these amino acids (histidine (His), Ile, Leu, Lys, methionine (Met), Phe, Thr, Trp, and Val) cannot be synthesized by the body and they are therefore essential or indispensable nutrients that must be

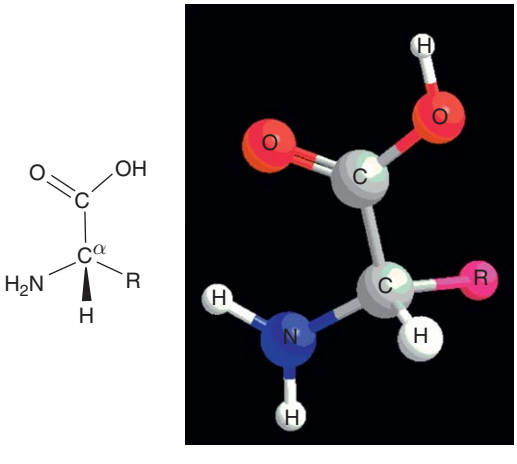


Figure 9 An α-amino acid.

obtained from the diet. The other 11 amino acids (Ala, arginine (Arg), aspartic acid (Asp), Asn, Cys, glutamic acid (Glu), glutamine (Gln), Gly, Pro, Ser, and Tyr) are also ordinarily obtained from the diet, but the body can synthesize them. They are therefore nonessential nutrients or nutritionally dispensable, but are equally as important as the indispensable amino acids for the nutrition of cells and for normal cell and organ function.[153] For the conditionally essential amino acids, there is either an increased demand of certain amino acids, as seen in specific diseases or a decreased endogenous synthesis. There are only a few diseases that may cause isolated amino acid deficiencies, for example, in catabolic diseases, some of the amino acids (Arg, Cys, Gln, Gly, Pro, histidine, Ser, and Tyr) can become conditionally essential.[150,152,154]

5.20.7.3 Choice of Amino Acid

Knowledge of how each individual amino acid and the manner in which they are combined affects important biochemical/pharmacological systems is important when considering their potential as novel agents. Understanding of the conformation, which an amino acid residue adopts during its interaction with either receptors or enzymes, is equally important.[155] Thus, the choice of an amino acid plays a very important role in biological and structural relevance.

Alanine (abbreviated Ala or A) ((*S*)-2-aminopropanoic acid; α-aminopropionic acid) is a nonpolar, neutral, aliphatic amino acid with the formula $HOOCCH(NH_2)CH_3$. Ala plays a major role in the transport of nitrogen from skeletal muscles to the liver.[150]

Arginine (Arg or R) (2-amino-5-(diaminomethylidene amino)-pentanoic acid) is a polar, basic, positively charged amino acid with the formula $HOOCCH(NH_2)(CH_2)_3NH(C{=}N^+H_2)NH_2$ and consists of a four-carbon aliphatic straight chain. The outer part of the side chain consisting of three nitrogens bonded to a carbon atom is called a guanidinium group. Owing to the conjugation between the double bond and the nitrogen lone pairs, the side chain of Arg can serve as hydrogen bond donor only.[11] Arg is the precursor of nitric oxide (NO), an very important biological messenger.[150] Arg-containing peptides can result in the formation of piperidones.[73] The side chains of Arg and Trp can serve as hydrogen donors only, while the side chains of Ser, Thr, Asn, and Gln can serve as hydrogen bond acceptors and donors.[11]

Asparagine (Asn or N) ((2*S*)-2-amino-3-carbamoyl-propanoic acid) is a polar, uncharged amino acid with the formula $HOOCCH(NH_2)CH_2CONH_2$. It has a carboxamide as the side chain's functional group. Asx or B represent either Asn or Asp. Asn are often found near the beginning and end of alpha-helices, and in turn motifs in beta sheets.[11,150]

Aspartic acid (Asp or D) ((2*S*)-2-aminobutanedioic acid) is a polar, acidic, negatively charged amino acid with the formula $HOOCCH(NH_2)CH_2COO^-$. The side chains of Asp and Glu are nearly always negatively charged at physiological pH; therefore, these amino acids are usually called aspartate (Asp) and glutamate (Glu). Asp and Glu play important roles as general acids in enzyme active centers, as well as in maintaining the solubility and ionic character of proteins.[11,150]

Cysteine (Cys or C) ((2*R*)-2-amino-3-sulfanyl-propanoic acid) is a polar, uncharged amino acid with the formula $HOOCCH(NH_2)CH_2SH$. It is an important thiol-containing (sulfur) amino acid and is classified as hydrophilic. The sulfhydryl of Cys is highly reactive and plays a crucial role in shaping proteins by forming disulfide links.[11] The Cys thiol group is nucleophilic and easily oxidized to the disulfide cystine. Owing to the ability of thiols to undergo redox reactions, Cys has antioxidant properties. Cys residues play a valuable role by crosslinking proteins. Insulin is an example of a protein with cystine crosslinking. Cys is useful to detoxify the body from harmful toxins.[11,150]

Glutamic acid (Glu or E) ((2*S*)-2-aminopentanedioic acid) is a polar, amino acid with the formula $HOOCCH(NH_2)CH_2CH_2COOH$. It has an acidic side chain. Glutamate is the precursor of Gln, Pro, and Arg, while Ser is the precursor of Gly and Cys.[11]

Glutamine (Gln or Q) ((2*S*)-2-amino-4-carbamoyl-butanoic acid) is a polar, uncharged amino acid with the formula $HOOCCH(NH_2)(CH_2)_2CONH_2$. The abbreviation Glx or Z represents either Gln or Glu. The side chain contains an amide group in place of the carboxylate and can be considered the amide of the acidic amino acid Glu. Gln is the most abundant naturally occurring, nonessential amino acid in the human body. It becomes conditionally essential in states of illness or injury. Both glutamate and Gln play key roles in nitrogen metabolism.[150]

Glycine (Gly or G) (aminoacetic acid, aminoethanoic acid) is a nonpolar, neutral, aliphatic amino acid with the formula $HOOCCH(NH_2)H$.[151] Gly is the simplest amino acid and plays important roles in peptide and protein chains. It does not contain a side chain and can thus fit into secondary structures where larger amino acids cannot.[156] Gly acts as a transmitter in the CNS where it accomplishes several functions.[157] Gly is a precursor of porphyrins.[150] Gly, Pro, aspartate, Ser, and Asn enable reverse turns.[150] The acylated amino group of Gly can accept a second acyl group to give rise to a diacylamide.[72]

Histidine (His or H) (2-amino-3-(3*H*-imidazol-4-yl) propanoic acid) is a polar, basic amino acid with the formula $HOOCCH(NH_2)CH_2$ ($C{=}CHNHCH{=}N^+H$).[151] His can be uncharged or positively charged, and is often found in active sites of enzymes where its imidazole ring can readily switch between these states resulting in the making and breaking of bonds.[11] The unprotonated imidazole is nucleophilic and can serve as a general base, while the protonated form can serve as a general acid. The imidazole moiety can play a role in stabilizing the folded structures of proteins and is useful as coordinating ligand in metalloproteins. Intramolecular nucleophilic attack by the imidazole nitrogen can cause lactam formation[158] as well as fission of peptide bonds.[159]

Isoleucine (Ile or I) ((2*S*,3*S*)-2-amino-3-methylpentanoic acid) is a nonpolar, neutral, aliphatic amino acid with the formula $HOOCCH(NH_2)CH(CH_3)CH_2CH_3$. Having a hydrocarbon side chain, Ile is classified as a hydrophobic amino acid. Together with Thr, Ile is one of the two common amino acids that have a chiral center. Although four stereoisomers of Ile are possible, Ile present in nature exists in only one enantiomeric form, that is, (2*S*,3*S*)-2-amino-3-methylpentanoic acid.[11,150]

Leucine (Leu or L) ((*S*)-2-amino-4-methyl-pentanoic acid) is a neutral, aliphatic amino acid with the formula $HOOCCH(NH_2)CH_2CH(CH_3)_2$ and with a hydrocarbon side chain, Leu is classified as a hydrophobic (nonpolar) amino acid.[11,150]

Lysine (Lys or K) (2,6-diaminohexanoic acid) is positively charged, amino acid with the formula $HOOCCH(NH_2)(CH_2)_4N^+H_3$. Lys, Arg, and His have *basic side chains*. The ε-amino group often participates in hydrogen bonding and as a general base in catalysis. Lys plays a role in the formation of collagen and is important for proper growth and bone development.[11]

Methionine (Met or M) ((*S*)-2-amino-4-(methylsulfanyl)-butanoic acid) is a nonpolar, neutral, amino acid with the formula $HOOCCH(NH_2)CH_2CH_2SCH_3$. Together with Cys, Met is one of the two sulfur-containing proteinogenic amino acids and a great antioxidant. Its derivative *S*-adenosyl methionine (SAM) serves as a methyl donor.[11,150]

Phenylalanine (Phe or F) (2-amino-3-phenyl-propanoic acid) is a neutral, aromatic amino acid with the formula $HOOCCH(NH_2)CH_2C_6H_5$. It is classified as nonpolar because of the hydrophobic nature of the benzyl side chain.[16,160] Tyr and Phe play a significant role not only in protein structure but also as important precursors for thyroid and adrenocortical hormones as well as in the synthesis of neurotransmitters such as dopamine and noradrenaline.[16,160] The genetic disorder phenylketonuria (PKU) is the inability to metabolize Phe. This is caused by a deficiency of phenylalanine hydroxylase with the result that there is an accumulation of Phe in body fluids. Individuals with this disorder are known as 'phenylketonurics' and must abstain from consumption of Phe. A nonfood source of Phe is the artificial sweetener aspartame (L-aspartyl-L-phenylalanine methyl ester), which is metabolized by the body into several by-products including Phe. The side chain of Phe is immune from side reactions, but during catalytic hydrogenations the aromatic ring can be saturated and converted into a hexahydrophenylalanine residue.[72]

Proline (Pro or P) ((*S*)-pyrrolidine-2-carboxylic acid) is a nonpolar, neutral amino acid and is regarded as a helix breaker. It has an aliphatic side chain but it differs from other amino acids in that its side chain is bonded to both the α-carbon and nitrogen. Pro contains a secondary rather than a primary amino group, which makes it an *imino* acid.[11] Because Pro lacks a hydrogen on the amide group, it cannot act as a hydrogen bond donor, only as a hydrogen bond acceptor. Pro is the only residue, which leads to an *N*-alkyl amide bond when incorporated into a peptide. Pro is therefore unique in having no amide proton to participate in hydrogen bonding and in having a cyclic side chain, which establishes conformational restrictions (the pyrrolidine ring decreases the conformational mobility of the DKP moiety, so that the side-chain rotamers for the nonprolyl residue is well defined).[77,161] Pro can cause reversal of direction of peptide chains in globular proteins[162] and acts as a conformational determinant in structural proteins, for example, collagen.[144] Brandl and Deber[163] suggested that *cis–trans* isomerism of Pro residues might play a role in transduction of transmembrane proteins. Inclusion

of Pro in DKPs increases the solubility of the compounds in chloroform ($CHCl_3$).[77] Futhermore, studies have showed that Pro has the ability to isomerize and undergoes a transition from all-*cis* form (polyproline 1) to the all-*trans* form (polyproline 11).[164] Pro is found in naturally occurring biologically active peptides, including peptide hormones, for example, angiotensin, bradykinin, oxytocin, vasopressin, melanocyte stimulating hormone (MSH) thyroid-releasing factor (TRF); gramicidin S and actinomycin; and antamanide.[165] Dipeptides containing a Pro or a hydroxyproline residue, exhibit a marked propensity or strong tendency for intramolecular cyclization and it is not surprising that several Pro-based DKPs are detected in fermented and thermally treated foods.[126,166,167] In addition, the thermal treatment of food appears to favor the preservation of some amino acids compared with others. For example, during the roasting of coffee beans there is a reduction in the amount of amino acids by 20–40%. The content of reactive amino acids such as Lys is strongly decreased, whereas others such as Pro or Phe remain nearly unchanged.[71] Both Pro- and Gly-containing peptides have a higher probability of cyclization than peptides containing other amino acids.[168]

Serine (Ser or S) ((*S*)-2-amino-3-hydroxypropanoic acid) is a polar, neutral, uncharged amino acid with the formula $HOOCCH(NH_2)CH_2OH$. It has an aliphatic hydroxyl side chain and can be seen as a hydroxylated version of Ala.[11] Ser participates in the biosynthesis of purines and pyrimidines and is also the precursor to several amino acids including Gly, Cys, and Trp (in bacteria). In addition, it is the precursor to numerous other metabolites, including sphingolipids[11] and is present in enzymes such as α-chymotrypsin.[169] Ser, Asn, and aspartate disrupt α helices.[150]

Threonine (Thr or T) ((2*S*,3*R*)-2-amino-3-hydroxybutanoic acid) has an aliphatic hydroxyl side chain and is classified as a polar, uncharged amino acid with the formula $HOOCCH(NH_2)CHOHCH_3$. Together with Ser and Tyr, Thr is one of the three proteinogenic amino acids bearing an alcohol group. Thr can be seen as a hydroxylated version of Val.[11] With two chiral centers, Thr can exist in four possible stereoisomers, or two possible diastereomers of L-Thr. However, the name L-Thr is used for one single enantiomer, (2*S*,3*R*)-2-amino-3-hydroxybutanoic acid. The second diastereomer (2*S*,3*S*), which is rarely present in nature, is called L-*allo*-Thr.

Both Ser and Thr are commonly considered to be hydrophilic due to the hydrogen bonding capacity of the hydroxyl group and play a dynamic role in the functioning of cellular processes. Thr has been observed to have an effect on the eating patterns of mammals.[170]

Tryptophan (Trp or W) ((*S*)-2-amino-3-(1*H*-indol-3-yl)-propionic acid, α-amino-β-[3-indolyl]propionic acid, 1-β-indolylalanine, 2-amino-3-indolylpropionic acid) is a nonpolar, neutral, aromatic amino acid and has a indole ring attached to the methylene group. Compounds containing an indole ring are found in many pharmacologically active compounds available in the market today. Their therapeutic uses range from antiemetics and anti-inflammatories to the treatment of hypertension, migraine, and Parkinson's disease.[9] Trp is also the precursor to 5-hydroxytryptamine (5-HT), also known as serotonin, which is an important neurotransmitter in the brain.[171] Trp is very sensitive under acidic conditions and its side chain is susceptible to oxidative degradation, dimerization, alkylation, and substitution with sulfenyl chlorides.[72]

Tyrosine (Tyr or Y) (4-hydroxyphenylalanine ((*S*)-2-amino-3-(4-hydroxyphenyl)-propanoic acid)) is a polar, neutral, aromatic amino acid with the formula $HOOCCH(NH_2)CH_2C_6H_5OH$ and is the precursor of thyroxin, dopamine, norepinephrine (noradrenaline), epinephrine (adrenaline), and the pigment melanin.[11] Being the precursor amino acid for the thyroid gland hormone thyroxin, a defect in this may result in hypothyroidism. Tyr is extremely soluble in water, a property that has proven useful in isolating this amino acid from protein hydrolysates. The occurrence of tyrosine-*O*-sulfate as a constituent of human urine and fibrinogen has been reported.[172]

Although Phe, Trp, and Tyr are nonpolar, Trp and Tyr are relatively more 'polar' than Phe because of the nitrogen indole ring and the Tyr hydroxyl group.[150]

Valine (Val or V) ((*S*)-2-amino-3-methyl-butanoic acid) is a nonpolar, neutral, aliphatic amino acid with the formula $HOOCCH(NH_2)CH(CH3)_2$. Along with Leu and Ile, Val is a branched-chain amino acid and is found in high concentrations in the muscles. Val is needed for muscle metabolism and coordination, tissue repair, and for the maintenance of proper nitrogen balance in the body.[150] The steric hindrance present in Val and Ile (caused by branching) lowers the rate of coupling reactions, resulting in an increase in side reactions.[72]

α-Helix formation is enhanced by glutamate, Leu, and Met whereas the formation of a β-sheet is favored by Val, Phe, and Ile.[150]

In conclusion, certain residues such as Gly, Pro, Val, and Ile enhance the tendency for cyclization. Aspartyl and Asn residues can result in the formation of aminosuccinimide derivatives. Alkylation of Tyr, Trp, and Met side chains are possible and can happen very quickly.[72]

5.20.8 Formation of 2,5-Diketopiperazines

5.20.8.1 History and Background

DKPs are simple and easy to obtain and are quite common by-products of synthetic, spontaneous, and biological formation pathways. DKP formation has been well documented as side reactions of solid-phase[173] and solution-phase peptide synthesis.[91] In addition, DKPs have been shown to be decomposition products of various peptides, proteins, and other commercial pharmaceuticals.[174,175] Cyclic dipeptides were found to be present in solutions of human growth hormone,[176] bradykinin,[177] histerlin,[178] and solutions of agents within the classes of penicillins and cephalosporins.[174] DKPs are also enzymatically synthesized in several protists and in members of the plant kingdom. Hydrolysates of proteins and polypeptides often contain these compounds and they are commonly isolated from yeasts, lichens, and fungi.[27]

Cyclo(Gly-Gly) was the first cyclic dipeptide synthesized and elucidated, since then a wide variety of members of the cyclic dipeptide family have been synthesized by various methodologies.[27,179] The cyclic dipeptides were initially prepared by the action of ammonia on the free dipeptide esters, liberated from the corresponding amine salts.[180,181] The long duration of exposure to ammonia, required by the free dipeptide esters to effect cyclization, lead to extensive racemization when employing optically pure linear dipeptide precursors.[182]

Early cyclic dipeptide synthetic methods have involved the following:

1. Cyclization of free dipeptide methyl esters, liberated from the corresponding amine salts, by the action of excess ammonia.[180,181]
2. Heating unprotected dipeptides with β-napthol.[183]
3. Heating of corresponding amino acids in refluxing ethylene glycol.[184]

A major breakthrough in solid-phase synthesis (SPPS) was provided by Merrifield in 1962. The method was based on a stepwise strategy and involved the following:

1. Synthesizing a peptide while keeping it attached at one end to a solid support.
2. The support is an insoluble polymer contained with a column.
3. By using a standard set of reactions the peptide is built on this support one amino acid at a time.
4. Protective groups blocking unwanted reactions at each successive step in the chemical cycle.[150]

Most pertinent and current methods are as follows:

1. Heating of the unprotected dipeptide or its hydrobromide salt in phenol.[185]
2. Cyclization of the linear dipeptide in a neutral medium.[182]
3. Hydrogenolysis of benzyloxycarbonyl dipeptide methyl esters in methanol over palladium or charcoal catalyst.[102,186]
4. Activated esters.[187]
5. Carboxy-catalyzed intramolecular aminolysis of the dipeptide ester.[173]
6. Refluxing dipeptide esters in 2-butanol containing 0.1 mol l^{-1} acetic acid for 3 h.[188]
7. Synthesizing the corresponding protected linear dipeptides by utilizing diethylphosphoryl cyanide (DEPC), triethylamine (Et_3N), and 1,2-dimethoxyethane at 0 °C. The *t*-Boc group was removed by using trifluoroacetic acid (TFA); followed by cyclization in saturated $NaHCO_3$ and extraction with $CHCl_3$.[189]
8. Solid-phase peptide synthesis.[190]

   It has been reported by Patel and Gordon[190] that SPPS is limited by poor yields with hindered amines, deactivated aromatic aldehydes and slight over alkylation with aliphatic aldehydes. Johnson *et al.*[94] argued that because of the speed and convenience of automated SPPS, support-bound cyclization protocols are ideal for the preparation of numerous cyclo peptide analogues.

9. A resin-bound amino acid is alkylated using an appropriate aldehyde and the resulting secondary amine is then acylated to yield an *N*-alkylated dipeptide. Deprotection followed by cleavage using toluene–ethanol solvent mixtures under basic or acidic conditions induces cyclization to yield the corresponding DKP. An advantage of this method is that potential side products that might be formed during the synthesis, for example, nonalkylated or nonacylated products, remain tethered to the solid support.[191]
10. Multicomponent Ugi reaction: a resin-bound amino acid is reacted with an aldehyde, isocyanide, and a second amino acid in a one-pot reaction to form the *N*-alkylated cyclic precursor dimer.[191]
11. Activated ester method in which formic acid (98%) and anisole (0.2 ml), a carbonium ion scavenger, were used to remove the *t*-Boc group. Haywood[192] investigated four different methods of cyclization of the unprotected linear dipeptide esters: (1) constant stirring for 5 days with saturated $NaHCO_3$, followed by extraction with $CHCl_3$ (85% yield); (2) boiling in 2-butanol and toluene (4:1) under reflux for 4 h (72% yield); (3) 2-butanol (15 ml) containing 1 mol l^{-1} acetic acid and 1 mmol *N*-methyl morpholine (NMM) added to the crude linear dipeptide ester and boiled under reflux for 4 h (44% yield); and (4) boiling in methanol for 18 h (56% yield, yellowish product).

Although DKPs are formed through head-to-tail cyclization, a strategically different approach to their synthesis is required due to their constrained nature.[91]

5.20.8.2 Problems Encountered with the Early Methods of Synthesis

Early methods of synthesis were riddled with problems such as racemization, poor percentage yield, long reaction times, and degradation of aromatic amino acid side chains.[27,180–182,193]

5.20.8.3 Racemization

Examination of the reaction mixtures when using the Fischer method, always showed the presence of some racemic material (5–40%).[182] Peptide synthesis therefore has to be directed in such a way that racemization cannot occur. Racemization is particularly favored in slightly alkaline solution, where the formation of an intermediate oxazolinone is responsible for the inversion of configuration. Racemization means the loss of biological activity, and amino acids having electronegative substituents in the β-position, for example, Cys, Ser, Thr, Asp, the aromatic amino acids (Phe, Tyr, Trp), and His are the most readily racemized. Racemization can occur during introduction and removal of protecting groups. During the activation of the carboxyl group, direct abstraction of the α-proton (carbanion formation) or oxazolinone formation can occur, resulting in racemization. Other factors that can cause racemization are as follows: (1) high temperatures; (2) coupling methods; (3) excess amounts of base; and (4) reversible β-elimination.[9,72,194]

Racemization can also be reduced by condensing Gly or Pro at the C-termini of the first amino acid.[72]

5.20.8.4 The Importance of Protecting Groups

In the preparation of even the smallest peptide, it is crucial that certain functional groups must be blocked. In order to avoid racemization, amino- and carboxylic protection is recommended. Two types of protecting groups are needed: (1) easily removing groups for α-amino functions and (2) semipermanent groups for all other blockings. In principle, three protective groups are necessary for efficient peptide synthesis: *t*-butyloxycarbonyl (*t*-Boc), 9-fluorenylmethyloxycarbonyl (Fmoc), and the triphenylmethyl (trityl) group.[9] The Boc protecting group is easily introduced by the reaction of the amino acid with di-*tert*-butyl dicarbonate in a nucleophylic acyl substitution reaction and is removed by brief treatment with a strong acid such as TFA[195] or can even be rapidly cleaved by mild acidic conditions.[196] However, TFA, alone or in mixtures with other solvents, is the most widely used reagent for the removal of the Boc group. The Boc group is stable to catalytic hydrogen action, sodium in liquid ammonia, alkali, and hydrazine.[196] The advantage of the Fmoc group is that it is cleaved under very mild basic conditions (e.g., piperidine), but stable under acidic conditions. Trityl substituents are removable by catalytic hydrogenation, TFA, dilute acetic acid, and hydrochloric acid in organic solvents.[196]

The carboxylic acid functional group of the amino acid can be preferable protected by converting it into the methyl ester amino acid or ethyl ester amino acid, whereas benzyl and *p*-nitrobenzyl esters are also convenient.[9] *t*-Butyl esters are better nucleophiles than methyl or ethyl esters.[9,196] This esterification can be achieved by dropping thionyl chloride into cold methanol, followed by the addition of the amino acid, which soon dissolves. Standing at room temperature or refluxing completes the reaction. Removal of the methyl ester group from the C-terminal, can usually be done by mild alkaline hydrolysis in acetone, methanol, or dioxane, preferably at or below room temperature.[196] *t*-Butyl esters are often used as permanent protecting groups, and will only be cleaved at the end of the whole synthesis with TFA.[9]

For total selectivity in lactam formation during SPPS, all other amino acid functional groups, other than those destined to form the lactam bridge, must be protected. Johnson *et al.*[94] found an unanticipated series of side products formed during the SPPS of cyclic peptide structures when using the orthogonal 4-{*n*-1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl]amino}benzyl ester (Dmab) carboxylic acid protecting group. It was shown that the Dmab group, when used for temporary protection of the glutamyl side-chain acid functionality, is prone to two side reactions. The first generates *N*-pyroglutamyl peptides through intramolecular cyclization either during N-Fmoc removal or subsequent acylation (coupling) reactions. The second side reaction is a result of the slow 1,6-elimination reaction of the 4-aminobenzyl group derived from hydrazine treatment of Glu(ODmab).

5.20.8.5 The Importance of Coupling Reagents

The creation of a peptide bond is a very important reaction in medicinal chemistry and there are many ways to synthesize DKPs. It is very important to prevent the formation of the wrong peptide bond. When two different amino acids are reacted with each other, it can give rise to four different products. It is therefore vital to protect the amino terminal of one and the carboxyl group of the other amino acid.

General coupling methods are as follows: (1) azide procedure, (2) anhydrides, (3) active esters, and (4) coupling reagents.[72]

5.20.8.5.1 Most commonly used coupling reagents

The following coupling reagents are useful and some can be used in solid-phase and liquid-phase peptide synthesis:

- Dicyclohexylcarbodiimide (DCC) is by far the most popular coupling agent, especially in Merrifield's SPPS. It is very reactive and gives high yield. DCC reacts rapidly with the amino acid or peptide carboxylic acid group to give an O-acylated isourea. The by-product formed is *N*-acylurea. This reaction has a big disadvantage: it can lead to isomerization and must be used in combination with auxiliary nucleophiles or additives, for example, 1-hydroxybenzotriazole, 3-hydroxy-3,4-dihydro-1,2,3-benzotriazin-4-one, or *N*-hydroxysuccinimide. A wide variety of solvents can be employed for DCC couplings including dioxane, DMF, CH_3CN, THF, $CHCl_3$, and CH_2Cl_2. DCC is allergenic and must be handled with great care.
- Mixed anhydrides. Ethyl chloroformate and isobutyl chloroformate are the reagents most often used.
- EEDQ (1-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline) and IIDQ (2-isobutyloxy-1-isobutyloxycarbonyl-1,2-dihydroquinoline) are agents which produce mixed anhydrides. EEDQ is a selective and efficient coupling reagent.
- Carbonyldiimidazole. This reagent reacts with acids in DMF, THF, or CH_3CN to form acylimidazoles.
- BOP(1-benzotriazolyl-tri-dimethylaminophosphonium hexafluorophosphate). Suitable solvents are DMF, CH_2Cl_2, or CH_3CN.
- HBTU(*O*-benzotriazolyl-tetramethylisouronium hexafluorophosphate).
- PPA (propane phosphonic acid anhydride) has a low tendency to racemization. By-products that are formed are water-soluble and thus easily removed.[9,72]
- DEPC. 1,2-Dimethoxyethane is a suitable solvent.[189]

5.20.8.6 The Four Basic Steps to Obtain Pure Diketopiperazines

The main steps to obtain pure DKPs are as follows:

1. the synthesis of the protected dipeptide ester precursor;
2. N-deprotecting;
3. cyclization;
4. crystallization.

5.20.8.6.1 Synthesis of the protected dipeptide precursor

The procedure requires the initial synthesis of the linear dipeptide precursor, bearing specific chemical blocking groups, capable of being selectively removed from the N- and C-terminals.[168]

The procedure for the condensation of two amino acids into a dipeptide essentially involves (1) the interaction of the carboxylic acid function of one amino acid with the amino function of the next in a coupling reaction; (2) loss of a water molecule; and (3) the formation of an amide bond, to give covalently linked amino acid residues. For the amino acid $H_2N–CHR^1–COOH$ (X) to be coupled with amino acid $H_2N–CHR^2–COOH$ (Y) to produce the dipeptide $H_2N–CHR^1–CHO–NH–CHR^2–COOH$ (XY), the hydroxyl group of the carboxyl-component (X) must react with the amino component (Y). This may be achieved either by the use of an activating agent or catalyst to drive the reaction. The inherent problem lies in the fact that the carboxyl group of (X) is not limited in acylating the amino component (Y) but may react with unreacted or unprotected functional groups of (X), yielding unwanted derivatives from the reaction (e.g., cyclic XX, tripeptide XXY, and tetrapeptide XXXY). Therefore, in the formation of the desired peptide, the carboxyl group (X) must be prevented from acylating its own α-amino group. This is achieved by temporarily attaching the α-amino group to chemical blocking groups such as *t*-Boc. The presence of diverse side-chain functional groups of the amino acids, and the need to maintain chiral integrity of the α-carbon stereocenter during coupling, complicates the synthesis process. Suitable side-chain protection or the judicious selection of reaction conditions for less problematic groups, and careful choice of activating groups or coupling reagents to assure high yields and efficacy while avoiding potential side reactions that generate unwanted by-products is crucial in peptide synthesis.[92]

5.20.8.6.2 N-deprotection

Besides the potential problems encountered during synthesis of racemization, poor percentage yield, and long exposure to the reagent,[182] butylation of aromatic residues can also occur.[197] Alakhov *et al.*[197] showed that *t*-Boc and *t*-butyl groups should not be used in the synthesis of Trp-containing peptides, since a high degree of butylation of Trp takes place during the removal of these groups with TFA. This side reaction can lead to a high percentage of impurities in the synthesis of complex natural peptides and their analogues by the solid-phase method, where repeated removal of *t*-Boc groups will inevitably lead to the accumulation of butyl-tryptophan-containing peptides. In addition, the use of elevated temperatures for the aminolysis reaction of amino acid precursors are generally not considered ideal due to the oxidative sensitivity of many of the amino acid side chains, such as Trp. In order to avoid or minimize such decomposition, acidolytic removal of protecting groups from Trp-containing peptides should be carried out at 0 °C under nitrogen.[72]

Tyr residues are also subject to O- and C-butylation, although to a lesser extent than Trp, during the removal of *t*-Boc group.[197] Carbonium ion scavengers, such as anisole or Met, have to be used to prevent butylation occurring during removal of the protective Boc group.[196] It is recommended that removal of protecting groups involving oxidation steps should be avoided when the chain contains sensitive residues such as Tyr, Trp, Cys, Ser, Met, and Thr.[196]

Nitecki *et al.*[182] reported on the successful deblocking of *t*-Boc peptides with formic acid (98%) at room temperature for 2 h resulting in the formation of the respective formate salts of the peptide esters. The action of the heat when boiling in a neutral medium (*sec*-butanol–toluene (4:1)) resulted in the facile conversion of the formates of dipeptide methyl esters to the corresponding DKPs. The synthetic approach, utilizing sterically pure peptide derivatives proved to be advantageous in delivering favorable yields rapidly while overcoming the shortfalls of racemization, previously reported as a limitation of earlier peptide cyclization procedures. It was reported that the advantages of the formic acid procedure include (1) the retention of steric homogeneity of the peptide and (2) improved product yield whenever acid-sensitive amino acids, such as Trp, were present in the precursor peptide.

5.20.8.6.3 Cyclization

The formation of the DKP ring, like any other cyclization procedure, requires the generation of mutually reactive chain ends and the reaction of these ends under conditions favoring intramolecular processes. Ring closure of dipeptide esters can take place spontaneously because the thermodynamic stability of the DKP ring overcomes the energy barrier in the formation of a *cis*-peptide bond. This reaction is accelerated by bases, for example, ammonia. Certain residues such as Val, Ile, *N*-methylamino acids, Pro, and Gly enhance the tendency for cyclization. If one of the residues has the L-configuration, while the other has the D-configuration, cyclization is accelerated. The reason being the amino acid side chains will lie on opposite sides of the plane of the DKP ring. Methyl, ethyl, or benzyl esters are more sensitive to nucleophilic attacks than dipeptide *tert* butyl esters.[72]

DKPs may be produced without any difficulties and may form without particular activation of the carboxyl function of the dipeptide precursor. However, the cyclization reaction may be a slow process, justifying the need for specific activation of the carboxyl moiety.[72] Kopple[168] stated that there is no single perfect method for the cyclization of all peptides, only guidelines aiding in the judicious choice of combinations of procedures to limit the generation of unwanted by-products of peptide synthesis.

According to Dinsmore and Beshore[87] there are five synthetic strategies that exist to synthesize 2,5-DKPs:

1. Intramolecular cyclization: DKP formation through intramolecular cyclization of the N_1–C_2 bond is an efficient route to ring closure and the construction of these head-to-tail dipeptides involves the coupling of an N-protected α-amino acid to an α-amino ester, followed by N-deprotection and cyclization.[194]
2. The second method of intramolecular cyclization to 2,5-DKPs involves formation of the N_1–C_6 bond. This reaction entails acylation of glycinamide with an α-halo acid halide, followed by ring closure under basic conditions.[198]
3. The third method involves the tandem formation of the N_1–C_2 and C_3–N_4 bonds. This strategy combines some of the methods used in the intramolecular cyclization approaches outlined in the first and second methods and is accomplished by the reaction of a bivalent electrophile and an α-amino acid.[199]
4. The fourth method also involves a tandem bond-forming strategy of N_1–C_2 and N_4–C_5, in which simultaneous formation of both amide bonds of a 2,5-DKP occurs in a one-pot reaction from an α-amino ester derivative.[200]
5. The fifth method involves the simultaneous formation of the N_1–C_2 and N_1–C_6 bonds, thus enabling the facile introduction of N_1-substituents by the use of primary amines as reactants. An example of this approach is the tandem acylation–alkylation of an amine with a suitable halo-ester[201] (see **Figure 1** for numbering sequence).

The most commonly used cyclization methods include the following:

1. Boiling in a neutral medium (*sec*-butanol–toluene (4:1))[182] (see Section 5.20.8.6.2).
2. Heating in phenol.[185]

An efficient method for the formation of cyclic dipeptides was reported by Kopple and Ghazarian.[185] They described the preparation of optically pure cyclic dipeptides by heating the unprotected dipeptide or its hydrobromide salt in phenol. Lichtenstein[183] initially reported the cyclization procedure for peptides utilizing molten β-naphthol at 135–150 °C. On removal of β-naphthol by ether extraction a sterically pure DKP was observed. Despite the apparent efficiency of the method outlined by Lichtenstein[183] the high temperatures, long reaction times, and the use of β-naphthol itself were identified as shortfalls of the procedure, enhancing the probability of generation of unwanted by-products by the increased risk of oxidation of sensitive chains during the cyclization procedure. The cyclization procedure therefore ideally required the substitution of β-naphthol. Derived from the aforementioned procedure, phenol was reported to be a more favorable solvent for cyclization. Phenol exhibited a lower melting point than β-naphthol therefore, making it readily removable by sublimation. The increased water solubility of phenol when compared with β-naphthol was proposed to be more beneficial due to evidence suggesting that many DPKs are more readily crystallized from water.[185] Additionally, the reduced reaction time, when utilizing phenol, further limits the chance for oxidation to occur. Cyclization of the free dipeptide was achieved on heating in phenol, in an oil bath held at 140–150 °C. When carried out under nitrogen, Kopple and Ghazarian[185] reported no discoloration of the reaction, with only the

detection of the required DPKs by thin-layer chromatography. The reports of consistent high yields, formation of sterically pure cyclic products, and commercial availability of free dipeptides suggest that the procedure utilizing phenol for cyclization is the method of choice. There is, however, a single requirement for the applicability of the phenol method for cyclization. At least one of the amino acid residues of the free dipeptide must prove to be sensitive to phenol for the cyclization procedure to occur.

The rate of a cyclization reaction, thus its yield in competition with side reactions, is related to the probability of juxtaposition of the ends of the open-chain precursor. This probability increases with the configurational and conformational stability of the ring to be formed and decreases with increases in the loss of internal freedom that results from ring formation.[168]

5.20.8.6.4 Crystallization

DKPs are soluble in hot absolute alcohol.[183] Young *et al.*[77] reported that DKPs are readily crystallized from ether, acetone, ether–acetone, and methanolic solutions. DKPs that separate out on the concentration of the alcoholic solutions or extracts may be recrystallized from alcohol or ethyl acetate. Crystals of polar DKPs can be obtained by slow evaporation of aqueous solutions.[202]

5.20.8.7 Factors Affecting the Formation of Diketopiperazines

5.20.8.7.1 The generation from peptides, peptide derivatives, and protein precursors

The rates of formation of various cyclic peptides and DKPs have been documented and shown to be affected by a wide range of physicochemical and structural parameters. Goolcharran and Borchardt[97] examined the effects of exogenous (i.e., pH, temperature, buffer species, and concentration) and endogenous (i.e., primary sequences) factors affecting the rate of cyclic dipeptide formation, using the dipeptide analogues of X-Pro-*p*-nitroaniline (X-Pro-*p*NA; where X represents the amino acid residue of the respective cyclic dipeptide).

The studies were performed under a variety of aqueous buffer solutions over the pH ranges of 1–10. Two possible pathways for the degradation of Phe-Pro-*p*NA were highlighted by Goolcharran and Borchardt.[97] The first degradation pathway involved the intramolecular aminolysis reaction yielding the cyclic dipeptide, cyclo(Phe-Pro). This reaction has been shown to occur readily for a number of peptides in aqueous solutions, thereby producing cyclic dipeptides. Cyclo(Phe-Pro) was shown to undergo pH-dependent hydrolysis at either of the two carbonyl groups, thereby yielding the two dipeptides Phe-Pro-OH and Pro-Phe-OH.

The intramolecular aminolysis reaction was also shown to be involved in the formation of (Gly-Gly) and the biosynthetic pathway of the biologically active cyclic dipeptide cyclo(His-Pro), found throughout the CNS, peripheral tissue, and body fluids.[36,203]

5.20.8.7.2 Exogenous factors affecting the rate of cyclic dipeptide formation

The rate of disappearance of Phe-Pro-*p*NA and the formation of cyclo(Phe-pro) was shown to follow pseudo-first-order kinetics, exhibiting significant dependence on pH. Additionally, the rate of formation of cyclo(Phe-Pro) at all pH values increased with increasing temperature, obeying the Arrhenius equation.[97] A pH–rate profile, generated from the results obtained by Goolcharran and Borchardt[97] suggests that the rate of cyclic dipeptide formation depends on the degree of ionization of the N-terminal amino group. At or below a pH of 3, where the reactant exists predominantly in the protonated form (pK_a of the agent was determined to be ~6.1), the rate of cyclic dipeptide formation was found to be independent of pH. In the pH range of 4–6 where the N-terminal amino groups become less protonated, the rate of cyclic dipeptide formation increased with pH. Interestingly, the rate of the intramolecular aminolysis reaction was shown to plateau between the pH of 6 and 8. Between these pH values, the free N-terminal amino groups are available for cyclization; however, the hydroxide ion concentration is not high enough to catalyze the reaction. Under these conditions, the formation of the cyclic dipeptide is suggested to occur by a neutral-catalyzed pathway, similar to that shown by Steinberg and Bada[175] for the formation of cyclic dipeptides from linear dipeptides. Under neutral conditions, the hydrolysis of glycylglycine proceeded exclusively by a mechanism involving the uncatalyzed attack by water on the peptide bond, thereby resulting in the formation of cyclo(Gly-Gly).[175]

Finally, in the pH range of 9–10, the slope of the pH profile was unit positive, indicating specific hydroxide ion catalysis. It is thus apparent that the unprotonated N-terminal group imparts more reactively to the

intramolecular aminolysis reaction than the protonated form. In addition to the pH of the solution, the nature of the buffer species was also shown to influence the rate of DPK formation. For Phe-Pro-*p*NA, Goolcharren and Borchardt[97] showed that above a pH of 5, the buffer type and concentration affects the catalysis significantly. No significant effect of buffer type and concentration on the rate of degradation was recorded below a pH of 5.

5.20.8.7.3 Endogenous factors affecting the rate of cyclic dipeptide formation

To form DKPs, the dipeptide precursors have to adopt a folded conformation, rather than a more stable, extended form in which the amide bond is in the favored *trans*-configuration.[204] Suitable evidence is provided by the high frequency of appearance of cyclic dipeptides derived from Pro-containing peptides. These peptides are forced into the folded form by the presence of the pyrrolidine ring and are therefore more prone to the formation of the cyclic dipeptide by intramolecular aminolysis reactions.[205] It is therefore not surprising to find a large number of Pro-containing endogenous cyclic dipeptides.[27,36] The steric accessibility in the cyclization step is known to be important and is highlighted by the relative instabilities of cephalosporins. DPKs have been shown to form spontaneously from cephalosporins. In contrast, they are rarely formed from penicillins. The added stability of penicillins is due to the presence of additional methyl groups (providing steric hindrance) that render the β-lactam function inaccessible to the side-chain amino group, thereby preventing the cyclization from occurring readily.[205]

Modification of the amino acid residues located on the N-terminal side of Pro was shown to have a major influence on the rate of cyclic dipeptide formation. For the series of dipeptide analogues of X-Pro-*p*NA, the half-lives of cyclic dipeptide formation in 0.5 mol l^{-1} phosphate buffer (pH 7) at 37 °C were reported as follows: X = Gly 5.1 days, X = Val 2.5 days, X = Ala 1.1 days, X = β-cyclohexylalanine 0.8 days, X = Arg 0.7 days, and X = Phe 0.5 days. Increased bulkiness of alkyl and aryl substituents have been previously shown to increase the rate of cyclization due to intramolecular reactions.[206] This however does not seem true for the series studied by Goolcharran and Borchardt[97] as the Ala analogue cyclized twice as fast as the bulkier analogue. From the study it is evident that simple steric bulk of substituents alone cannot be used to effectively explain the effects involved in the formation of cyclic dipeptides from various peptide precursors.

Sequence inversion and racemization have been associated with uncatalyzed formation of the cyclic dipeptides and has been shown to greatly complicate the kinetics of formation. Cyclic dipeptide formation, by uncatalyzed processes, is rapid enough to pose an apparent threat to the stability of proteins and a possible rationale for the posttranslational N-acetylation of proteins that have been observed in higher organisms.[203]

The rate of DKP formation will also depend on the carbonyl ester protecting groups or the structures of the peptide–resin linkage in the solid-phase mode.[207] Furthermore, cyclization is a concentration-independent reaction and demands the use of dilute solutions.[196]

5.20.8.8 Naming of Peptides

When naming peptides, the names of acyl groups ending in 'yl' are used. If Pro and Tyr were to condense so that Pro acylates Tyr, the dipeptide formed is named prolyltyrosine.[10]

5.20.9 Biological Relevance

5.20.9.1 Diketopiperazines as Drugs

Bacterial resistance to conventional antibiotics has become a serious problem in infection control, and has led to intensive research efforts to develop an effective novel antimicrobial agent.[208] Antimicrobial peptides have already played a crucial role in pharmaceutical research as biomedically useful agents or as lead compounds for drug development.[209] More specifically, cyclic peptides have shown some potential as a possible new class of antimicrobial agent.[208,210] In a study undertaken by Fernandez-Lopez *et al.*[20] various cyclic peptides were shown to be effective in killing both Gram-positive and Gram-negative bacteria. The mechanism by which cyclic peptides exert their antimicrobial action is believed to be a result of pore formation in the plasma membrane of bacteria, leading to cellular content leakage. This is achieved by the cyclic peptides forming tubes in the lipid bacterial membrane and subsequent depolarization of the cell membrane. Cyclic peptide antibiotics

have been shown to be well tolerated and appear to have no deleterious effects.[208] The rapid action of these peptides, together with the fact that they act on membrane integrity rather than on vital biosynthetic processes, proposes that resistance to this type of agent might be slower to develop than conventional antibiotics.[210]

Cyclic dipeptides have already been shown to exhibit antitumor activity.[211–213] In a study undertaken by Graz *et al.*[211] the effect of nine cyclic dipeptides on HT-29 cancer cells was investigated. The study was based on the theory that the use of maturation-inducing compounds may induce differentiation of neoplastic cells, thereby stimulating normal cells to faster recovery. The respective cyclic dipeptides were shown to have great specificity for the neoplastic cells, with little effect on human gastrointestinal mucosa, thereby limiting severe adverse effects. Rhee[213] has explored the various mechanisms by which cyclic dipeptides may have anticarcinogenic activity. Cyclic dipeptides having anticarcinogenic potential are thought to activate cellular systems, which would intercept and detoxify carcinogens, or may stimulate DNA damage repair, and/or may eradicate heavily damaged cells through apoptosis. Inhibition of DNA topoisomerase I activity has been proposed as one potential molecular target for cyclic dipeptides. Many cyclic dipeptides have this antitumor effect against numerous human cell lines thereby necessitating the screening of DPKs as potential cytotoxic agents.[213] Cancer neoplasms or malignancies are often associated with alterations in hemostasis, having a profound impact on the hemostatic system. Clinically, these alterations manifest in thrombotic disorders, including deep-vein thrombosis, pulmonary embolism, migratory thrombophlebitis, nonbacterial thrombotic endocarditis, and arterial thrombosis. Abnormalities in haemostatic parameters are present in about 95% of patients with some form of metastatic disease, and in 50% of patients with cancer, with cancer being the most common acquired cause of thromboembolism.[214,215]

In a study conducted by Szardenings *et al.*[216] various combinatorial libraries of DPKs scaffolds were created to design and evaluate the activity of DPKs as inhibitors of the matrix metalloproteinases, namely, collegenase-1 and gelatinase B. This study created structure–activity relationships (SAR) for side chains attached to a DPK core structure. These enzymes are therapeutic targets with indications in the treatment of cancer, arthritis, autoimmunity, and cardiovascular disease.

The use of structurally rigid DKPs as bioactive models for opioid receptor antagonists has been proposed. These compounds are used in the elucidation of the binding requirements and will lead to the design of highly selective molecules with potential clinical application for diseases of the opioid system. These include the treatment of autism, alcohol dependency, and modulation of immunity.[217] Further studies by Baures *et al.*[218] has proposed the use of L-prolyl-leucylglycinamide (PLG) DPK analogue five and six to modulate dopamine D_2 receptors by increasing the affinity of the receptor for agonists and by increasing the percentage of D_2 receptors that exist in the high-affinity state. Gly-containing DKPs have been shown to exhibit marked antioxidant effects, demonstrating neuroprotection properties. They are also capable of inhibiting calcium-induced necrotic cell death, as well as cell death induced by $FeSO_4$, a free radical generator.[219]

DKPs are ubiquitous in nature, with cyclo(Pro-Leu), cyclo(Pro-Phe), and cyclo(Ala-Leu) being a few examples of cyclic dipeptides of microbial origin. To date, only four naturally occurring cyclic dipeptides exhibit biological activity, that is, cyclo(Leu-Gly), cyclo(Tyr-Arg), cyclo(Asp-Pro), and cyclo(His-Pro). Only cyclo(His-Pro) is endogenous in mammals, and is found throughout the CNS and gastrointestinal tract as well as in a variety of body fluids including milk, blood, cerebrospinal fluid, semen, and urine.[36,220–222] The endogenous pathway of cyclo(His-Pro) formation has been considered to be the cleavage of thyroid releasing hormone (TRH) by pyroglutamyl aminopeptidase to produce His-Pro-NH_2, which is nonenzymatically converted into cyclo(His-Pro).[223]

5.20.9.2 Biological Activity of Individual Diketopiperazines

Cyclo(His-Pro) has been found to exhibit the following pharmacological effects:

1. Inhibition of prolactin secretion[224–232] and hormonal regulation.[233]
2. Antagonist of ethanol narcosis in mice.[234–237]
3. Production of hypothermia in rats.[238–241]
4. Appetite suppressant: Suppression of stress-induced eating, starvation-induced eating, and spontaneous eating over a 10-h period in mice.[242–246]

5. Decreases gastric mucosa blood flow.[247]
6. Increases the sedative effects of pentobarbital.[248]
7. Plays a key role in the perception of pain induced by thermal, chemical, physical, and mechanical stimuli.[249–252]
8. Antagonist of THC-induced analgesia (delta-9-tetrahydrocannabinol).[249]
9. Decreases motor activity.[249]
10. Reduces the sedative effects of ethanol and ketamine.[235]
11. Increases amphetamine-induced stereotypic behavior.[253]
12. Modulation of glucose metabolism.[254]

Although the mechanism for cyclo(His-Pro)'s biological action is not clear, it is assumed that multiple mechanisms are responsible for its biological action.[36]

Bicyclomycin (cyclo(Leu-Ile)), also known as bicozamycin, is a small cyclic dipeptide isolated from the culture broths of *Streptomyces sapporonensis* and *Streptomyces aizunensis*.[23,255] Owing to the fact that bicyclomycin displays low toxicity and is readily available from fermentation, it has been commercially introduced as an effective agent against nonspecific diarrhea in humans and bacterial diarrhea in calves and pigs. Its spectrum of activity covers mainly Gram-negative bacteria such as *Escherichia coli*, *Klebsiella*, *Shigella*, *Salmonella*, and *Neisseria*.[255] The mode of action is not well understood, but it appears to be different from other known mechanisms and therefore displays no cross-resistance to other antibiotics.[23,27]

Cyclo(Leu-Trp), a bitter compound isolated from the fermentation of milk casein by *Bacillus subtilis*, opened up the field to flavor and fragrance properties. It was further noted that dipeptides became more bitter when blockage of both the amino and carboxyl groups occurred or the dipeptide was converted into a DKP. This phenomenon opened the field of taste exhibition.[119]

Cyclo(His-D-Leu) acts as a hydrolytic catalyst.[28] Cyclo(Leu-Gly) blocks the development of (1) physical dependence on morphine, (2) tolerance to the pharmacological effects of β-endorphin, (3) tolerance to haloperidol-induced catalepsy and hypothermia, and (4) dopaminergic supersensitivity after chronic morphine administration.[36,256]

Cyclo(Tyr-Arg), a synthetic analogue of kyortorphin (an endogenous analgesic peptide), and its *N*-methyl tyrosine derivatives are more potent than kyotorphin in the mouse tail pressure test.[257]

Cyclo(Arg-Lys) and cyclo(Asp-Lys) were identified as immunomodulation agents.[258]

The combined synergistic effects of cyclo(Leu-Pro) and cyclo(Phe-Pro) were effective against five vancomycin-resistant enterococci (VRE) strains: *Enterococcus faecium* (K-99-38), *E. faecalis* (K-99-17), *E. faecalis* (K-99-258), *E. faecium* (K-01-312), and *E. faecalis* (K-01-511) with MIC values of 0.25–1 mg l^{-1}. It also showed activity against *E. coli*, *Staphylococcus aureus*, *Micrococcus luteus*, *Candida albicans*, and *Cryptococcus neoformans* with MIC values of 0.25–0.5 mg l^{-1}. This combination also showed mutagenic activity against *Salmonella typhimurium* TA98 and TA100 strains in a *Salmonella* mutation assay.[213]

Ström *et al.*[259] have isolated a *Lactobacillus plantarum* strain (MiLAB 393) with antifungal properties from grass silage. Three antifungal substances produced by the strain have been isolated, that is, cyclo(Phe-Pro), cyclo(Phe-*trans*-4-OH-Pro), and 3-phenyllactic acid.

The antitumor agent compound 593 A, was isolated from *Streptomyces griseoluteus*. This compound possesses the β-chloramine function of the nitrogen mustards, a group of compounds whose derivatives have been incorporated into many synthetic drugs used in cancer chemotherapy.[60]

Phenylahistin, a fungal DKP metabolite composed of Phe and isoprenylated dehydrohistidine exhibits potent *in vitro* and *in vivo* antitumor activity.[260]

Verticillen A, an antibiotic isolated from *Verticillium* spp., has been identified as an antitumor agent.[261]

Cyclo(Phe-Pro) was shown to inhibit the growth of HT-29 ($IC_{50} = 4.04 \pm 1.15$), HeLa ($IC_{50} = 2.92 \pm 1.55$), and MCF-7 cells ($IC_{50} = 6.53 \pm 1.26$) and induced apoptosis in HT-29 colon cancer cells. Cyclo(Tyr-Pro) exhibited a greater growth inhibitory effect in MCF-7 cells ($P < 0.01$) compared with HT-29 and HeLa cells.[262] It also induced caspase-3 activation and induction of PARP cleavage.[263]

Both cyclo(Gly-Phe) and cyclo(Ala-Phe) showed anticonvulsant activities.[264]

At elevated temperatures, extremes in pH, and/or moisture, the dipeptide sweetner aspartame is converted into cyclo(Asp-Phe).[36] The dehydro derivatives of cyclo(Phe-Phe) were found to inhibit cell division,[265] while cyclo(Asp-Pro) caused inhibition of dietary fat intake.[266]

Cyclo(Pro-Phe) isolated from *Rosellinia necatrix* is potent in blocking physical dependence to morphine in mice.[256]

Cyclo(Pro-Tyr), also known as maculosin, is a phytotoxin produced by the fungus *Alternaria alternata*, and has been found to cause black necrotic lesions on the leaves of spotted knapweed.[267] The mechanism underlying the phytotoxic action of cyclo(Pro-Tyr) may lie in its ability to inhibit ribulose-1,5-diphosphate carboxylase.[268]

Other cyclic dipeptides studied by Milne *et al.*[212] showed that cyclo(Phe-Pro), cyclo(Tyr-Pro), cyclo(Trp-Trp), and cyclo(Trp-Pro) are biological active in prokaryotes and eukaryotes. In addition, it was found that these compounds had the following effects:

1. Cyclo(Trp-Trp) and cyclo(Trp-Pro) blocked calcium channels.
2. Cyclo(Trp-Pro) and cyclo(Tyr-Pro) blocked delayed-rectifier potassium channels.
3. Cyclo(Phe-Pro), cyclo(Tyr-Pro), and cyclo(Trp-Trp) all significantly enhanced the expression of the biochemical differentiation marker ($P < 0.05$).
4. Antibacterial activity of the selected DKPs was concentration dependent.

Cyclo(Pro-Trp) and cyclo(Phe-Pro) showed broad-spectrum antibacterial properties, whereas cyclo(Trp-Pro) and cyclo(Trp-Trp) exhibited broad-spectrum antifungal properties. The maturation of the gastrointestinal cells was enhanced by cyclo(Trp-Trp), cyclo(Trp-Pro), cyclo(Phe-Pro), and cyclo(Tyr-Pro).[269]

Cyclo(Phe-Cys) and cyclo(Tyr-Cys) significantly reduced the growth of HeLa cells. Screening concentrations of 10 and 100 $\mu mol\,l^{-1}$ cyclo(Phe-Cys) reduced the viable Hela cells by $43.92 \pm 0.02\%$ ($P < 0.0001$) and $57.79 \pm 0.02\%$ ($P < 0.0001$), respectively, while the 10 and 100 $\mu mol\,l^{-1}$ screening concentrations of cyclo(Tyr-Cys) led to a reduction in viable cells of $54.61 \pm 0.02\%$ ($P < 0.0001$) and $61.36 \pm 0.02\%$ ($P < 0.0001$). A volume of 10 $\mu mol\,l^{-1}$ cyclo(Phe-Cys) resulted in a $30.56 \pm 0.03\%$ ($P < 0.0001$) reduction in HT-29 cell viability, increasing to $51.29 \pm 0.01\%$ ($P < 0.0001$) inhibition at a concentration of 100 $\mu mol\,l^{-1}$, while cyclo(Tyr-Cys) showed inhibition of $32.41 \pm 0.03\%$ ($P < 0.0001$) and $52.27 \pm 0.02\%$ ($P < 0.0001$) at 10 and 100 $\mu mol\,l^{-1}$, respectively. Cyclo(Phe-Cys) inhibited the growth of MCF-7 cells, producing a $36.49 \pm 0.02\%$ ($P < 0.0001$) inhibition at 10 $\mu mol\,l^{-1}$ and a $56.08 \pm 0.01\%$ ($P < 0.0001$) inhibition at 100 $\mu mol\,l^{-1}$, while cyclo (Tyr-Cys) inhibited cell viability by $35.15 \pm 0.01\%$ ($P < 0.0001$) at 10 $\mu mol\,l^{-1}$ and $52.42 \pm 0.02\%$ ($P < 0.0001$) at 100 $\mu mol\,l^{-1}$.[270]

The DKPs cyclo(His-Ala) and cyclo(His-Gly) proved to have promising anticancer activity comparable with that observed for cisplatin. Cyclo(His-Ala) in particular demonstrated an ability to inhibit tumor growth in HT-29, HeLa, and MCF-7 cancer cell lines. Cyclo(His-Gly) however, only had a marked effect on MCF-7 carcinoma cells at a concentration of 100 $\mu mol\,l^{-1}$. Both DKPs at a concentration of 0.5 $mmol\,l^{-1}$ were effective against two of the three Gram-positive bacteria, that is, *Bacillus* and *S. aureus*. Cyclo(His-Gly) was more effective against *Klebsiella pneumoniae* than cyclo(His-Ala). Cyclo(His-Ala) and cyclo(His-Gly) inhibited the growth of *C. albicans* by 66.3 and 47%, respectively.

Cardiac activity with respect to heart rate, coronary flow rate, and ventricular pressure indicated that both DKPs had significant effects on cardiac performance. Cyclo(His-Gly) had a greater effect on coronary flow rate during the heart rate experiments, decreasing coronary flow, while cyclo(His-Ala) had a greater effect decreasing coronary flow in the ventricular pressure experiments. Cyclo(His-Gly) proved to have a significant effect on thrombin-substrate binding throughout the concentration range, while cyclo(His-Ala) appeared to only have a significant effect at lower concentrations. Neither DPK had any significant effects on fibrinolysis. Thrombin-induced and adenosine diphosphate(ADP)-induced aggregation were inhibited to a greater degree by cyclo(His-Gly).[271]

Results indicated that cyclo(Phe-Tyr) and cyclo(Tyr-Tyr) are relatively inactive as antibacterial agents, while cyclo(Phe-Tyr) proved to be effective in the various tumor cell lines, for example, HT-29 (60.6% inhibition), HeLa (73.4% inhibition), and MCF-7 (75.6% inhibition). While having insignificant effects on delayed-rectifier inward potassium channels, calcium blocking activity was noted in both agents, with cyclo(Tyr-Tyr) being voltage dependent. Effects on the isolated rat heart showed cyclo(Phe-Tyr) to decrease heart rate, coronary flow rate, and ventricular pressure, while cyclo(Tyr-Tyr) showed the opposite activity, increasing heart rate and coronary flow rate while having no effect on ventricular pressure. Both compounds

bound to the μ-opioid receptor, with cyclo(Phe-Tyr) acting as an agonist, while the characterization of cyclo(Tyr-Tyr) was more complex and may be of mixed agonist/antagonist nature or a partial agonist.[272]

The results obtained from the histone acetylation and phosphorylation studies suggest that cyclo(Trp-Trp) and cyclo(Pro-Trp) induce differential gene expression through different signal transduction pathways. Cyclo(Trp-Trp) induced the highest level of acetylation of histones whereas cyclo(Phe-Pro) induced high levels of phosphorylation of histones.[211]

Cyclo(His-Tyr) showed significant activity in hematological studies, where it produced a significant prolongation of blood clotting time, slowing of clot lysis, and inhibition of ADP-induced platelet adhesion and aggregation. Cyclo(His-Phe) exhibited significant antitumor activity against HeLa, WHCO3, and MCF-7 cell lines, causing greatest reduction of cell viability in cervical carcinoma cells (56.8% inhibition). Cyclo(His-Tyr) did show some significant concentration-dependent reduction in cell viability in WHCO3 cells. Both cyclic dipeptides caused blocking of sodium and calcium ion channels, but opening of inward rectifying potassium ion channels. Cyclo(His-Phe) caused a significant slowing of heart rate and a decrease in coronary flow rate, while cyclo(His-Tyr) significantly increased heart rate. Both compounds caused an increase in ventricular pressure in isolated studies on the rat heart. Cyclo(His-Tyr) showed notable antibacterial activity against *Streptococcus pyogenes*, *E. coli*, and *Pseudomonas aeruginosa* and both DPKs showed excellent antifungal activity, especially cyclo(His-Phe).[273]

Compounds identified as cell cycle inhibitors are Trp-Pro DPK alkaloids isolated from the fungus *Aspergillus fumigatus*, namely, spirotryprostatin A and spirotryprostatin B.[274] These compounds show potential as M-phase inhibitors of the mammalian cell cycle.[275]

The fungal metabolites TAN-1496 A, C, and E (isolated from the culture broth of *Microsphaeropsis* sp. FL-16144) have been shown to have specific inhibitory activity against DNA topoisomerase 1. DNA topoisomerase 1 is an enzyme responsible for DNA metabolism and have been proposed as an intracellular target for cancer chemotherapy.[276]

A calpain inhibitor, the DPK of *N*-dimethyltyrosine, was isolated from *Streptomyces griseus* and this compound showed activity in the calpain assay as described by Alvarez *et al.*[277] Calpain is a cytosolic protease regulated by calcium and is distributed in mammalian and avian cells. Calpain catalyzes proteolysis of target protein in cells, causing changes in metabolic processes such as the activation of protein kinase C, neuropeptide metabolism, and the activation of platelets. It is proposed that these inhibitors can be used in the treatment of neurodegenerative diseases.

Cyclo(D-Trp-L-Pro) showed potential as Ca^{2+} channel antagonist, while cyclo(L-Trp-L-Pro), cyclo(L-Trp-D-Pro), and cyclo(D-Trp-D-Pro) acted as Ca^{2+} agonists. Only cyclo(D-Trp-L-Pro) showed independence of the membrane potential. No effect on the inward rectifier K^+ current was noted for any of the isomers.[278] This may explain why none of the isomers showed a prolonged QRS complex.[278,279] All isomers showed antagonistic effects on the Na^+ channel. Cyclo(L-Trp-L-Pro) and cyclo(D-Trp-D-Pro) showed no significant effect on the heart rate, while cyclo(L-Trp-D-Pro) caused a positive chronotropic effect, with cyclo(D-Trp-L-Pro) exhibiting negative chronotropic effects. An increase in coronary flow was observed for cyclo(L-Trp-D-Pro) and cyclo(D-Trp-L-Pro). All four isomers were capable of reducing the time spent in both ventricular arrhythmia (VA) and ventricular tachycardia (VT), as well a reducing the time taken for the heart to return to a normal sinus rhythm (SR).[278]

Only cyclo(D-Trp-L-Pro) is hepatocyte specific in its cytotoxicity, while the other three isomers are also cytotoxic for other cell types.[280] When using the MTT assay it was found that cyclo(L-Trp-L-Pro) is most effective against Gram-positive bacteria (*S. aureus* (47.4% inhibition) and *Streptococcus* (45.5% inhibition)) at $1\,mg\,ml^{-1}$. Cyclo(L-Trp-D-Pro) was highly effective against *E. coli* at all concentrations tested (0.125, 0.25, 0.5, $1\,mg\,ml^{-1}$) (64–74% inhibition), while cyclo(D-Trp-L-Pro) effectively inhibited the growth of *Streptococcus* at $1\,mg\,ml^{-1}$ (44.4%). A measure of 0.5 and $1\,mg\,ml^{-1}$ cyclo(D-Trp-D-Pro) reduced the viability of *C. albicans* to $69.963 \pm 1.24\%$ and $65.331 \pm 4.77\%$, respectively. An amount of $10^{-3}\,mol\,l^{-1}$ cyclo(L-Trp-D-Pro) and cyclo(D-Trp-L-Pro) showed the greatest potential as anticancer agents against HeLa cells over a 72-h period causing $\pm 50\%$ reduction in viability.[281]

Cyclo(Met-Tyr) and cyclo(Met-Trp) produced a marked inhibitory effect toward platelet adhesion induced by both ADP and thrombin. A moderate cytotoxic effect was displayed by cyclo(Met-Trp) against HeLa (55.85% inhibition), HT-29 (39.80% inhibition), and MCF-7 cancer cells (54.96% inhibition) whereas

cyclo(Met-Tyr) had very little effect in this regard (4.74–15.47% inhibition). Antibacterial studies indicated that cyclo(Met-Tyr) ($P = 0.0133$) has therapeutic potential as an antifungal (when utilizing XTT), while both agents showed antibacterial activity (when utilizing both MTT and TTC), with a greater effect toward Gram-negative organisms especially *E. coli* and *P. aeruginosa*. When cyclo(Met-Tyr) and cyclo(Met-Trp) were tested against *P. aeruginosa* utilizing MTT, *P*-values of 0.0031 and 0.0053 were, respectively, obtained. Cyclo(Met-Tyr) ($P = 0.0033$) and cyclo(Met-Trp) ($P = 0.0019$) produced inhibitory effects on *P. aeruginosa* when assayed with TTC. Cyclo(Met-Tyr) ($P = 0.0352$) and cyclo(Met-Trp) ($P = 0.0055$) showed inhibitory activity when the compounds were tested against *E. coli* utilizing MTT. Determining the effects of cyclo(Met-Tyr) and cyclo(Met-Trp) on *E. coli* utilizing TTC, resulted in *P*-values of 0.0003 and 0.0017, respectively. Cyclo(Met-Trp) displayed inhibitory activity toward the Gram-positive organism *S. pyogenes* (when utilizing MTT) ($P = 0.0159$). Both DPKs exhibited an antagonistic effect toward calcium and sodium channels.[282]

Cyclo(Trp-Tyr) and cyclo(D-Trp-Tyr) exhibited very high inhibitory activity against *S. pyogenes*. Ion channel studies indicated that the compounds were able to act as calcium channel agonists and to increase coronary flow rate without any significant effect on the heart rate. A volume of 20 mmol l^{-1} cyclo(Trp-Tyr) inhibited growth of HeLa (90.77%), MCF-7 (62.19%), and HT-29 cancer cells (26.22%). The hematological studies indicated the ability of the compounds to inhibit both ADP and thrombin-induced platelet aggregation.[283]

Both cyclo(Gly-Tyr) and cyclo(Gly-Phe)significantly inhibited ADP- and thrombin-induced aggregation of isolated platelets. Both DKPs displayed efficacy as antibacterial agents against several bacterial cultures (*B. subtilis*, *E. coli*, *Klebsiella* sp., *Streptococcus* sp., *S. aureus*, *P. aeruginosa*) with cyclo(Gly-Tyr) showing high growth inhibition against *Streptococcus* sp. (87%). Cyclo(Gly-Tyr) caused growth inhibition of 47.5, 35.3, and 38.3% against HeLa, MCF-7, and HT-29 cells, respectively. Cyclo(Gly-Phe) caused marginal growth inhibition that varied from 5.6 to 10.3%. Both compounds also showed activity as antagonists of calcium channels, but had minimal activity on delayed-rectifier potassium channels.[284]

Cyclo(Met-Gly) and cyclo(Met-Met) displayed efficacy as antifungal and antibacterial agents, with Gram-negative activity for both compounds against *E. coli and P. aeruginosa*, while cyclo(Met-Met) had an inhibitory effect on *S. pyogenes* ($P = 0.0002$). The effect of the DKPs on *E. coli* were determined by utilizing TTC and the inhibitory effects shown by cyclo(Met-Gly) and cyclo(Met-Met) yielded *P*-values of 0.0001 and <0.0001, respectively. Determining the effect of cyclo(Met-Gly) and cyclo(Met-Met) on *E. coli* as quantified by MTT, resulted in *P*-values of 0.0160 and 0.0003, respectively. Cyclo(Met-Gly) ($P = 0.0066$) and cyclo(Met-Met) ($P = 0.0040$) showed inhibitory effects against *P. aeruginosa* when assayed with TTC. The antitumor activity against colonic, breast, and cervical cancer was negligible for both DKPs. Calcium and sodium channel studies revealed marked activity in an antagonistic capacity for both channel types and by both compounds.[285]

Cyclo(Phe-Cys) and cyclo(Tyr-Cys) displayed significant inotropic and negative chronotropic effects, especially cyclo(Phe-Cys). Cyclo(Phe-Cys) and cyclo(Tyr-Cys) reduced heart rate by 39.85% ($P = 0.0033$) and 21.4% ($P = 0.0063$), respectively, displaying negative chronotropic activity. Cyclo(Phe-Cys) reduced coronary flow rate by 16.3% ($P = 0.0029$), while a 12.7% ($P = 0.0039$) reduction in coronary flow rate was observed for cyclo(Tyr-Cys). Cyclo(Phe-Cys) displayed the greatest increase in ventricular pressure ($P = 0.0053$) compared with cyclo(Tyr-Cys) ($P = 0.0140$). Both DKPs showed considerable activity in the hematological studies, where cyclo(Phe-Cys) and cyclo(Tyr-Cys) had significant effects on the calcium coagulation assay. Both DKPs slowed the rate of clot formation. The activity against *S. pyogenes* was limited with only cyclo(Tyr-Cys) (500 μmol l^{-1}) showing statistically significant activity (10.58% inhibition). Both DKPs were active against *B. subtilis*, but cyclo(Phe-Cys) (48.33% inhibition) exhibited more activity than cyclo(Tyr-Cys) (23.32% inhibition).[286]

Cyclo(Gly-Leu) and cyclo(Gly-Ile) inhibited ADP-induced aggregation of platelets significantly at a concentration of 100 μmol l^{-1}. At this concentration, a percentage inhibition of $36.66 \pm 1.89\%$ ($P = 0.003$) for cyclo(Gly-Leu) and $29.60 \pm 2.53\%$ ($P = 0.0129$) for cyclo(Gly-Ile) were noted. Cyclo(Gly-Leu) reduced heart rate by 15.29% ($P = 0.0263$), while cyclo(Gly-Ile) caused a decrease in heart rate of 9.05% ($P = 0.0389$). Cyclo(Gly-Leu) decreased the coronary flow rate by 49.95% ($P = 0.00912$) compared with a 30.24% ($P = 0.024$) decrease by cyclo(Gly-Ile). After 30 min of exposure, cyclo(Gly-Leu) showed a 15.78% reduction ($P = 0.0309$) in ventricular pressure compared with the 23.44% ($P = 0.01187$) reduction by cyclo(Gly-Ile). Cyclo(Gly-Leu) (0.5 mg ml^{-1}) caused a $42.50 \pm 3.87\%$ ($P = 0.0015$) growth inhibition of *K. pneumoniae* and

cyclo(Gly-Ile) (0.5 mg ml^{-1}) caused a 51.67 ± 3.82% ($P = 0.001$) growth inhibition. Cyclo(Gly-Ile) showed a higher growth inhibition of *C. albicans* (25.74 ± 0.58% at 1 mg ml^{-1} and 16.31 ± 0.53% at 0.5 mg ml^{-1}) when compared with cyclo(Gly-Leu) (17.76 ± 0.73% at 1 mg ml^{-1} and 10.29 ± 0.45% at 0.5 mg ml^{-1}).[287]

Both cyclo(Gly-Leu) and cyclo(Gly-Ile) at 10 and 100 µmol l^{-1} exhibited moderate effects in inhibiting HT-29, MCF-7, and HeLa cancer cell lines, except for cyclo(Gly-Ile) that showed very little activity against HT-29 cells at a concentration of 10 µmol l^{-1}. The greatest activity was noted for 100 µmol l^{-1} cyclo(Gly-Leu) against HeLa cells (28.64 ± 1.51% inhibition, $P = 0.0018$).[270]

Cyclo(Gly-Thr) at concentrations of 10 and 100 µmol l^{-1} exhibited growth inhibition of 15.55 ± 1.31% and 24.61 ± 2.37% for HeLa cells, 12.27 ± 1.06% and 30.84 ± 1.16% for HT-29 cells, 8.36 ± 0.24% and 8.73 ± 0.76% for MCF-7 cells, respectively. Cyclo(Gly-Ser) at concentrations of 10 and 100 µmol l^{-1} exhibited growth inhibition of 21.15 ± 1.16% and 16.75 ± 1.43% for HeLa cells, 20.09 ± 1.26% and 21.45 ± 1.66% for HT-29 cells, 15.03 ± 1.03% and 13.39 ± 1.23% for MCF-7 cells, respectively ($P < 0.05$).

Cyclo(Gly-Ser) at a screening concentration of 1 mg ml^{-1} showed significant inhibition against *S. aureus* ($P = 0.0386$). In addition, cyclo(Gly-Ser) at a screening concentration of 1 mg ml^{-1} did inhibit the growth of *B. subtilis* ($P = 0.0237$) and *C. albicans* ($P = 0.0001$).

Both cyclic dipeptides at all screening concentrations were found to significantly enhance ADP-induced platelet aggregation ($P < 0.0001$) with cyclo(Gly-Thr) demonstrating greater platelet aggregation activity over cyclo(Gly-Ser).

Cyclo(Gly-Thr) decreased heart rate by 14.20% ($P = 0.0198$) 30 min after perfusion with the drug, while cyclo(Gly-Ser) resulted in a slightly larger reduction in heart rate, decreasing heart rate by 19.12% ($P = 0.0188$). Cyclo(Gly-Thr) decreased coronary flow rate by 9.65% ($P = 0.0246$) 30 min after perfusion with the drug. Cyclo(Gly-Ser) resulted in a larger reduction in coronary flow, decreasing coronary flow rate by 45.79% ($P = 0.0021$). Both cyclic dipeptides decreased ventricular pressure when compared with the control, with cyclo(Gly-Thr) producing a greater reduction in ventricular pressure ($P = 0.0231$) than cyclo(Gly-Ser) ($P = 0.0236$).[288]

Both cyclo(Gly-L-Val) and cyclo(Gly-D-Val) proved to have moderate broad antibacterial activity and *P*-values of 0.0144 (48.68 ± 4.30% inhibition) and 0.0131 (49.74 ± 3.88% inhibition) were obtained for cyclo(Gly-L-Val) and cyclo(Gly-D-Val) at a final concentration of 0.667 mg ml^{-1}, against *S. aureus*. *P*-values of 0.0476 (43.42 ± 3.61% inhibition) and 0.0233 (38.56 ± 3.25% inhibition) were obtained for cyclo(Gly-L-Val) and cyclo(Gly-D-Val) at a concentration of 0.334 mg ml^{-1}, respectively.

P-values of 0.0280 (8.88 ± 1.13% inhibition) and 0.0005 (10.71 ± 1.02% inhibition) obtained for cyclo(Gly-L-Val) and cyclo(Gly-D-Val) at a concentration of 0.667 mg ml^{-1}, respectively, indicate that both cyclic dipeptides produced significant inhibition of *E. coli*. *P*-values of 0.0111 (42.82 ± 4.14% inhibition) and 0.0003 (59.10 ± 6.33% inhibition) were calculated for cyclo(Gly-L-Val) and cyclo(Gly-D-Val) at a concentration of 0.667 mg ml^{-1} against *K. pneumoniae*, while *P*-values of 0.0477 (41.22 ± 3.83% inhibition) and 0.0009 (53.33 ± 3.22% inhibition) were calculated for cyclo(Gly-L-Val) and cyclo(Gly-D-Val) at a concentration of 0.334 mg ml^{-1}, respectively. Cyclo(Gly-L-Val) ($P = 0.0208$; 25.59 ± 2.10% inhibition) and cyclo(Gly-D-Val) ($P = 0.0006$; 39.73 ± 3.83% inhibition) at 0.067 mg ml^{-1} displayed promising antifungal activity against *C. albicans*.[289]

Cyclo(Gly-L-Val) at a concentration of 10 µmol l^{-1} significantly inhibited the growth of HeLa cells (12.58 ± 1.23%) ($P = 0.0219$), whereas no significant inhibitory effect on HT-29 ($P = 0.0617$) and MCF-7 cell lines ($P = 0.0529$) was observed. At 100 µmol l^{-1} cyclo(Gly-L-Val) reduced the growth of all three cell lines with the greatest activity being noted for HT-29 at 23.90 ± 2.25% inhibition ($P = 0.0031$), while an inhibition of 23.46 ± 2.04% ($P = 0.0049$) and 21.19 ± 2.04% ($P = 0.0107$) was noted for HeLa and MCF-7 cells, respectively. Cyclo(Gly-D-Val) at 10 µmol l^{-1} produced significant inhibition of cell growth for all three cell lines, with the most pronounced activity being noted for MCF-7 (19.78 ± 1.86%) ($P = 0.0015$). An inhibition of 19.61 ± 0.82% ($P = 0.0231$) and 14.29 ± 1.40% ($P = 0.0392$) was observed for HeLa and HT-29 cells, respectively. The greatest activity noted for cells exposed to cyclo(Gly-D-Val) at 100 µmol l^{-1} was against MCF-7 with an inhibition of 21.94 ± 2.17% ($P = 0.0014$).[270]

Cyclo(Gly-L-Val) reduced the heart rate by 3.90% ($P = 0.0121$) after 5 min of perfusion when compared to the control, reaching a 24.01% reduction in heart rate after 30 min. After 5 min of perfusion, cyclo(Gly-D-Val) caused a 3.78% ($P = 0.0057$) reduction in heart rate, reaching a 14.61% reduction in heart rate after 30 min.

Cyclo(Gly-L-Val) caused a reduction in coronary flow rate of 43.72% ($P = 0.0002$) after 30 min of perfusion, while cyclo(Gly-D-Val) caused a reduction of 30.90% ($P = 0.0001$). The ventricular pressure was reduced by 39.63% ($P = 0.0037$) after 30 min of exposure to cyclo(Gly-L-Val). For cyclo(Gly-D-Val) a 16.74% ($P = 0.0190$) reduction in ventricular pressure was observed after 30 min.[289]

Biological studies revealed that cyclo(Ser-Ser) and cyclo(Ser-Tyr) inhibited the growth of HeLa, HT-29, and MCF-7 cancer cells as well as the growth of certain selected Gram-positive, Gram-negative, and fungal microorganisms. Cyclo(Ser-Ser) at concentrations of 100 and 10 $\mu mol\,l^{-1}$ exhibited growth inhibition of $23.17 \pm 2.04\%$ ($P < 0.0001$) and $5.37 \pm 0.42\%$ ($P = 0.0004$) for HeLa cells, $2.82 \pm 0.27\%$ ($P = 0.0005$) at 100 $\mu mol\,l^{-1}$ for HT-29 cells, and $8.06 \pm 0.71\%$ ($P < 0.0001$) and $13.26 \pm 0.94\%$ ($P < 0.0001$) at 100 and 10 $\mu mol\,l^{-1}$ for MCF-7 cells, respectively. Cyclo(Ser-Tyr) at concentrations of 100 and 10 $\mu mol\,l^{-1}$ exhibited growth inhibition of $20.37 \pm 0.76\%$ ($P < 0.0001$) and $3.29 \pm 0.21\%$ ($P = 0.0020$) for HeLa cells, $5.30 \pm 0.40\%$ ($P < 0.0001$) and $4.59 \pm 0.31\%$ ($P < 0.0001$) for HT-29 cells, $6.27 \pm 0.86\%$ ($P = 0.0003$) and $7.08 \pm 0.68\%$ ($P < 0.0001$) for MCF-7 cells, respectively.

Cyclo(Ser-Ser) at concentrations of 100, 50, 25, and 12.5 $\mu mol\,l^{-1}$ caused growth inhibition of $20.64 \pm 1.93\%$ ($P = 0.0002$), $12.97 \pm 1.19\%$ ($P = 0.0014$), $9.20 \pm 0.88\%$ ($P = 0.0069$), and $13.73 \pm 0.62\%$ ($P = 0.0009$) for *B. subtilis*; $27.25 \pm 2.57\%$ ($P < 0.0001$), $20.69 \pm 1.15\%$ ($P < 0.0001$), $25.97 \pm 2.58\%$ ($P < 0.0001$), and $18.37 \pm 1.15\%$ ($P < 0.0001$) for *S. aureus*; $11.64 \pm 1.01\%$ ($P < 0.0001$), $9.10 \pm 0.91\%$ ($P = 0.0003$), $7.19 \pm 0.69\%$ ($P = 0.0009$), and $7.94 \pm 0.59\%$ ($P = 0.0005$) for *E. coli*; $7.31 \pm 0.45\%$ ($P = 0.0295$) at 100 $\mu mol\,l^{-1}$ for *P. aeruginosa*; $11.62 \pm 1.09\%$ ($P = 0.0008$), $11.94 \pm 0.92\%$ ($P = 0.0006$) at 100 and 50 $\mu mol\,l^{-1}$ for *C. albicans*, respectively. Cyclo(Ser-Tyr) at concentrations of 100, 50, 25, and 12.5 $\mu mol\,l^{-1}$ exhibited growth inhibition of $17.95 \pm 1.54\%$ ($P = 0.0003$), $17.72 \pm 1.38\%$ ($P = 0.0003$), $9.32 \pm 0.92\%$ ($P = 0.0066$), and $8.92 \pm 0.79\%$ ($P = 0.0078$) for *B. subtilis*; $14.33 \pm 1.40\%$ ($P < 0.0001$), $10.62 \pm 0.89\%$ ($P < 0.0001$), $15.42 \pm 1.09\%$ ($P < 0.0001$), and $19.07 \pm 1.48\%$ ($P < 0.0001$) for *S. aureus*; $9.86 \pm 0.76\%$ ($P = 0.0002$), $8.01 \pm 0.72\%$ ($P = 0.0005$), $6.64 \pm 0.52\%$ ($P = 0.0013$), and $4.72 \pm 0.46\%$ ($P = 0.0065$) for *E. coli*; $11.51 \pm 0.30\%$ ($P = 0.0042$), $8.65 \pm 0.53\%$ ($P = 0.0153$), $9.77 \pm 0.80\%$ ($P = 0.0093$), and $11.69 \pm 0.76\%$ ($P = 0.0041$) for *P. aeruginosa*; $9.17 \pm 0.84\%$ ($P = 0.0024$), $9.78 \pm 0.54\%$ ($P = 0.0016$), $8.53 \pm 0.45\%$ ($P = 0.0031$), and $9.90 \pm 0.79\%$ ($P = 0.0016$) for *C. albicans*, respectively.

Cyclo(Ser-Ser) decreased the heart rate by 7.53% ($P < 0.0001$) after 30 min, while a decrease of 11.16% ($P < 0.0001$) was observed for cyclo(Ser-Tyr). Cyclo(Ser-Ser) reduced the coronary flow rate by 24.81% ($P < 0.0001$) after the 30th min of perfusion. A reduction of 54.74% ($P < 0.0001$) in coronary flow rate was noted for cyclo(Ser-Tyr). Cyclo(Ser-Ser) and cyclo(Ser-Tyr) decreased the ventricular pressure by 9.95% ($P = 0.0001$) and 20.97% ($P < 0.0001$) after the 30th min of perfusion, respectively.

Cyclo(Ser-Tyr) exhibited significant activity in the hematological studies, where it increased the rate of calcium induced-coagulation, and decreased the rate of streptokinase-induced fibrinolysis.[290]

Cyclo(Cys-Ile) and cyclo(Cys-Val) did not exhibit broad antimicrobial activity, demonstrating significant inhibition against only a few specific organisms. Cyclo(Cys-Ile) exhibited minimal inhibition ($1.93 \pm 0.173\%$ at 1 $mmol\,l^{-1}$) ($P = 0.2455$) against *S. aureus*, with only cyclo(Cys-Val) at a screening concentration of 1 $mmol\,l^{-1}$ ($33.34 \pm 1.86\%$) ($P = 0.0017$) and 0.5 $mmol\,l^{-1}$ ($19.26 \pm 1.56\%$) ($P = 0.0216$) showing significant inhibition against *S. aureus*.

Cyclo(Cys-Val) at a screening concentration of 1 and 0.5 $mmol\,l^{-1}$ did exhibit inhibition against *B. subtilis* ($43.10 \pm 3.88\%$ at 1 $mmol\,l^{-1}$) ($P = 0.00019$) and ($26.57 \pm 2.56\%$ at 0.5 $mmol\,l^{-1}$) ($P = 0.00234$). Both DKPs were found to inhibit *E. coli*: cyclo(Cys-Ile) and cyclo(Cys-Val) produced a maximum inhibition of $12.43 \pm 1.23\%$ ($P = 0.00038$) and $15.96 \pm 1.40\%$ ($P = 0.0028$), respectively, at a screening concentration of 1 $mmol\,l^{-1}$. At a screening concentration of 0.5 $mmol\,l^{-1}$, cyclo(Cys-Ile) and cyclo(Cys-Val) inhibited *E. coli* growth by $7.49 \pm 0.74\%$ ($P = 0.0021$) and $8.30 \pm 0.82\%$ ($P = 0.0038$), respectively. Cyclo(Cys-Ile) and cyclo(Cys-Val) at concentrations of 1 and 0.5 $mmol\,l^{-1}$ did not inhibit the growth of *C. albicans*.

A volume of 100 $\mu mol\,l^{-1}$ cyclo(Cys-Ile) reduced heart rate by 12.14% ($P = 0.0022$) 30 min after perfusion with the drug. Cyclo(Cys-Val) reduced the heart rate by 8.11% ($P = 0.0021$). Cyclo(Cys-Ile) reduced coronary flow rate by 92.34% ($P = 0.00012$) 30 min after perfusion with the drug, while cyclo(Cys-Val) reduced coronary flow rate by 84.05% ($P = 0.00014$).

During ventricular pressure experiments the heart rate was kept constant by pacing. Even with the heart rate constant, both DKPs displayed a decrease in coronary flow. Cyclo(Cys-Ile) was found to decrease coronary

flow rate by 83.94% ($P = 0.00048$) 30 min after perfusion with the drug. Cyclo(Cys-Val) resulted in a smaller reduction in coronary flow, decreasing coronary flow rate by 73.91% ($P = 0.00074$). Cyclo(Cys-Ile) caused a decline in ventricular pressure, reducing ventricular pressure by 13.68% ($P = 0.00011$) 30 min after drug perfusion. Cyclo(Cys-Val) reduced ventricular pressure by 16.79% ($P = 0.0036$). Both DKPs at all screening concentrations were found to significantly enhance ADP-induced platelet aggregation ($P < 0.05$ for all screening concentrations), with cyclo(Cys-Val) demonstrating greater platelet aggregation activity over cyclo(Cys-Ile).

When comparing the percentage inhibition of all three cancer cell lines, cyclo(Cys-Val) at a screening concentration of 0.1 mmol l^{-1} was found to cause the greatest percentage inhibition ($26.99 \pm 2.33\%$) ($P = 0.0039$) against MCF-7 cells. Cyclo(Cys-Val) also demonstrated significant inhibition against HeLa cell growth at a concentration of 0.1 mmol l^{-1} ($16.67 \pm 1.32\%$) ($P = 0.00093$). Cyclo(Cys-Ile) only demonstrated significant inhibition against HeLa cell growth at a screening concentration of 0.1 mmol l^{-1}, showing a percentage inhibition of $15.67 \pm 1.47\%$ ($P = 0.0064$).[291]

Glossary

affinity Tendency to bind to receptors. The greater the affinity between the drug and the receptor, the greater the efficacy of the drug.
agonist A drug which binds selectively to a particular type of receptor causing activation.
aliphatic An organic compound having an alkane carbon skeleton.
amino acids Bifunctional compounds and the building blocks of proteins. α-Amino acid consists of an amino group, a carboxyl group, an R-group, and a hydrogen atom which is bonded to the α-carbon.
amino group An NH_2 group.
aminolysis Nucleophilic substitution by an amine.
antagonist A drug which binds selectively to a particular type of receptor without activating it.
antiproliferative To inhibit the multiplication of cells.
apoptosis Programmed cell death.
carboxyl group The COOH functional group.
configuration The particular arrangement of atoms in space that is characteristic of a given stereoisomer.
conformation Means structural arrangement or one of the 'shapes' that a molecule can assume.
cyclic dipeptides Heterocyclic compounds comprising of two amino acid residues linked to a six-membered diketopiperazine ring structure.
cyclization Condensing of the NH_{2-} and COOH-end groups of a linear dipeptide.
esterification The reaction of a carboxylic acid with an alcohol to form an ester.
glucogenic Amino acids catabolized to pyruvate, α-ketoglutarate, succinyl CoA, fumarate, or oxaloacetate.
heterodetic The ring system is formed from amides and other heteroatom linkages.
heteromeric When one or more of the building units of the phenyl-containing homomeric ring are not amino acids.
homodetic Homomeric cyclic peptides in which the amino acids of the ring are joined through amide bonds.
homomeric The ring system/structure is built entirely of amino acids.
isomer Molecules that have the same number of atoms (molecular formula) but have different chemical structures and therefore different properties.
isomerization To change a compound into an isomer.
ketogenic Amino acids with carbon skeletons that serve as precursors of ketone bodies.
N-deprotection Removal of N-protecting groups.
nonpeptidomimetics Nonpeptide analogues, usually small molecules.
nonpolar Hydrophobic molecules that are poorly soluble in water.
oligopeptide A few amino acids joined by peptide bonds.
peptide bond When two amino acids combine in a condensation reaction that releases water, the dipeptide is held together by an amide linkage (also known as the peptide bond).

peptidomimetic (pseudopeptide) A compound that can imitate or block the biological effect of a peptide at the receptor level.
polar Water loving or hydrophilic. Describes molecules that are water soluble.
prodrugs Inactive drugs, but once metabolized they become active.
racemization Converting from an optically active compound or mixture to one that is racemic.
retro-inverso The reverse synthesis of a peptide using D-amino acids results in peptide analogues wherein the direction of the peptide bonds is reversed and the N- and the C-termini are interchanged.
specificity The ability of a receptor or an enzyme to discriminate among competing ligands or substrates.
stereochemistry Deals with structures in three dimensions/studies that take into account the spatial aspects of molecules.
tandem reaction Consecutive series of *intramolecular organic reactions* that often proceed through highly *reactive intermediates*. It is also known as a cascade – or domino reaction.
zwitterion A dipolar ion, with spatially separated negative and positive charges.

Abbreviations

ADP	adenosine diphosphate
5-HT	5-hydroxytryptamine
ATP	adenosine-5′-triphosphate
B. subtilis	*Bacillus subtilis*
BOP	1-benzotriazolyl-tri-dimethylaminophosphonium hexafluorophosphate
C. albicans	*Candida albicans*
C. neoformans	*Cryptococcus neoformans*
CD	circular dichroism
CNS	central nervous system
DCC	dicyclohexylcarbodiimide
DEPC	diethylphosphoryl cyanide
DKP	2,5-diketopiperazine
Dmab	4-{*n*-1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl]amino}benzyl ester
DMF	dimethylformamide
DNA	deoxyribonucleic acid
DOP	2,5-dioxopiperazine
E. coli	*Escherichia coli*
E. faecalis	*Enterococcus faecalis*
E. faecium	*Enterococcus faecium*
EEDQ	1-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline
Fmoc	9-fluorenylmethyloxycarbonyl
HBTU	*O*-benzotriazolyl-tetramethylisouronium hexafluorophosphate
HeLa	cervical carcinoma cell line
HPLC	high-performance liquid chromatography
HT-29	human colon carcinoma cell line
IC_{50}	inhibitory concentration (50%)
IIDQ	2-isobutyloxy-1-isobutyloxycarbonyl-1,2-dihydroquinoline
IR	infrared spectroscopy
K. pneumoniae	*Klebsiella pneumoniae*
M. luteus	*Micrococcus luteus*
m/e	mass/charge ratio

MCF-7	human breast cell line
MIC	minimal inhibitory concentration
MS	mass spectrometry
MSH	melanocyte stimulating hormone
MTT	3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide
NMM	*N*-methyl morpholine
NMR	nuclear magnetic resonance
ORD	optical rotatory dispersion
P. aeruginosa	*Pseudomonas aeruginosa*
PAF	platelet-activating factor
PEPT1	low-affinity/high-capacity peptide transporter
PEPT2	high-affinity/low-capacity peptide transporter
PKU	phenylketonuria
PLG	L-prolyl-leucylglycinamide
PPA	propane phosphonic acid anhydride
S. aureus	*Staphylococcus aureus*
S. griseus	*Streptomyces griseus*
S. pyogenes	*Streptococcus pyogenes*
S. typhimurium	*Salmonella typhimurium*
SAM	*S*-adenosyl methionine
SAR	structure–activity relationship
SPPS	solid-phase synthesis
SR	sinus rhythm
***t*-Boc**	*t*-butyloxycarbonyl
TFA	trifluoroacetic acid
THC	delta-9-tetrahydrocannabinol
THF	tetrahydrofuran
TRF	thyroid-releasing factor
Trityl	triphenylmethyl
TTC	2,3,5-triphenyltetrazolium chloride
VA	ventricular arrhythmia
VRE	vancomycin-resistant enterococci
VT	ventricular tachycardia
WHCO3	esophageal cancer cell line
XTT	2,3-bis[2-methoxy-4-nitro-5-sulfophenyl]-2*H*-tetrazolium-5-carboxanilide

Nomenclature

Arg or R	arginine
Asn or N	asparagine
Asp or D	aspartic acid
Ca^{2+}	calcium ions
CH_2Cl_2	dichloromethane
CH_3CN	acetonitrile
$CHCl_3$	chloroform
Cys or C	cysteine
Et_3N	triethylamine

$FeSO_4$	ferrous sulfate
Gln or Q	glutamine
Glu or E	glutamic acid
Gly or G	glycine
His or H	histidine
Ile or I	isoleucine
K^+	potassium ions
Leu or L	leucine
Lys or K	lysine
Met or M	methionine
mg l^{-1}	milligram per liter
mg ml^{-1}	milligram per milliliter
mmol l^{-1}	millimole per liter
Na^+	sodium ions
Phe or F	phenylalanine
Pro or P	proline
Ser or S	serine
Thr or T	threonine
Trp or W	tryptophan
Tyr or Y	tyrosine
Val or V	valine
μmol l^{-1}	micromole per liter

References

1. Medinfo. Lack of new medicines opens fields to illness. **2003**.
2. R. C. Hider, *Pharm. J.* **1998**, *261*, 393–399.
3. R. P. Rozek; N. Tully, *J. World Intellect. Prop.* **1999**, *2* (5), 293–319.
4. P. N. Kourounakis; E. Rekka, Introduction. In *Advanced Drug Design and Development: A Medicinal Chemistry Approach*; P. N. Kourounakis, E. Rekka, Eds.; Ellis Harwood: New York, 1994; pp 1–141.
5. P. L. Myers, *Curr. Opin. Biotech.* **1997**, *8*, 701–707.
6. H. Kessler, *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 512–523.
7. M. A. Gallop; R. W. Barrett; W. J. Dower; S. P. A. Fodor; E. M. Gordon, *J. Med. Chem.* **1994**, *37* (9), 1233–1256.
8. B. A. Berkowitz; B. G. Katzung, Basic & Clinical Evaluation of New Drugs. In *Basic and Clinical Pharmacology*, 8th ed.; B. G. Katzung, Ed.; Lange Medical Books/McGraw-Hill: New York, 2001; p 64.
9. W. Bernhagen, *Manuf. Chem. Aerosol News* **1995**, *66*, 27–30.
10. C. Mohan, Proteins, Peptides and Amino Acids. http://www.emdbiosciences.com (accessed 19 May 2008).
11. L. Stryer, *Biochemistry*, 3rd ed.; W. H. Freeman and Company: New York, 1988; Chapters 2, 21, 24, pp 15–42, 495–515, 575–600.
12. H. Kessler; K. Wagner; M. Mill, Analysis of Peptides Conformation by NMR Spectroscopy. *Proceedings of the 9th International Symposium on Medicinal Chemistry*, Berlin, 1986; pp 143–157.
13. V. V. Suresh Babu. One Hundred Years of Peptide Chemistry. http://www.ias.ac.in/resonance (accessed 3 March 2006).
14. A. Giannis; T. Kolter, *Angew. Chem. Int. Engl.* **1993**, *32*, 1244–1267.
15. V. J. Hruby, *Nat. Rev. Drug Discov.* **2002**, *1* (11), 847–858.
16. B. C. Katzung, *Basic and Clinical Pharmacology*, 8th ed.; McGraw-Hill Companies: North America, New York, 2001; pp 179–198, 287, 362, 753–776, 923.
17. J. Matsoukas; T. Mavromoustas, *Biomed. Health Res.* **1999**, *22*, 300–312.
18. J. A. Vida; M. Gordon, Conformationally Directed Drug Design: Peptides and Nucleic Acids as Templates or Targets. In ACS Symposium Series 251; American Chemical Society: Washington, DC, 1984; pp 9–21.
19. R. E. W. Hancock; D. S. Chappel, *Antimicrob. Agents Chemother.* **1999**, *43* (6), 1317–1323.
20. S. Fernandez-Lopez; H.-S. Kim; E. C. Choi; M. Delgado; J. R. Granja; A. Khasanov; K. Kraehenbuehl; G. Long; D. A. Weinberger; K. M. Wilcoxen; M. R. Ghadiri, *Nature* **2001**, *412*, 452–455.
21. H. P. Rang; M. M. Dale, *Pharmacology*, 2nd ed.; Churchill Livingstone: Tokyo, 1991; pp 1–955.
22. T. Mizuma; S. Masubuchi; S. Awazu, *J. Pharm. Pharmacol.* **1998**, *50*, 167–172.
23. M. E. Wolff, *Burger's Medicinal Chemistry and Drug Discovery, Volume 2: Therapeutic Agents*, 5th ed.; John Wiley and Sons, Inc.: New York, 1996; pp 1–670.

24. D. M. Jones; B. J. Leckie; L. Svensson; M. Szelke, A Novel Non-Hydrolyzable Isostere of the Peptide Transition State and a New Synthesis of the Hydroxyl-Ethylene Isostere. In *Peptides*; J. E. Rivier, R. G. Marshall, Eds.; ESCOM: Leiden, 1990; p 971.
25. L. Lakiewicz; D. B. Sherman; A. F. Spatola, Subtle Amide Bond Surrogates: The Effect of Backbone Thioamides on the Physical Properties, Conformation, and Biological Activities of Peptides. In *Peptides*; J. E. Rivier, R. G. Marshall, Eds.; ESCOM: Leiden, 1990; p 976.
26. P. V. Pallai; S. Richman; R. S. Struthers; M. Goodman, *Int. J. Peptide Protein Res.* **1983**, *21*, 84.
27. P. G. Sammes, *Chem. Org. Naturstoffe (Wien)* **1975**, *32*, 51–118.
28. M. J. O. Anteunis, *Bull. Soc. Chim. Belg.* **1978**, *87*, 626–650.
29. V. Crescenzi; A. Cesaro; E. Russo, *Int. J. Peptide Protein Res.* **1973**, *5*, 427–434.
30. R. Ramani; V. Sasisekharan; K. Venkatesan, *Int. J. Peptide Protein Res.* **1977**, *9*, 277–292.
31. K. P. Fu; H. C. Neu, *Antimicrob. Agents Chemother.* **1978**, *13*, 358.
32. H. C. Neu; K. P. Fu; N. Aswapokee; P. Asqapokee; K. Kung, *Antimicrob. Agents Chemother.* **1979**, *16*, 150.
33. T. Miyoshi; N. Miyairi; H. Aoki; M. Kohnsaka; H. Sakai; H. J. Imanaka, *J. Antibiot.* **1972**, *25*, 569.
34. Y. Wei; D. Pei, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1073.
35. T. Andoh, *Biochimie* **1998**, *80*, 235.
36. C. Prasad, *Peptides* **1995**, *16*, 151–164.
37. S. Johne; D. Groger, *Pharmazie* **1977**, *32*, 1–14.
38. U. Kertscher; M. Bienert; E. Krause; N. F. Sepetov; B. Mehlis, *Int. J. Peptide Protein Res.* **1993**, *41*, 207–211.
39. M. Nagayama; O. Takaoka; K. Iomata; Y. Yamagata, *Orig. Life Evol. Biosphys.* **1990**, *20*, 249–257.
40. C. Prasad, *Ann. N. Y. Acad. Sci.* **1989**, *553*, 232–251.
41. S. Tamura; A. Suzuki; A. Aoka; N. Otaki, *Agric. Biol. Chem. Jpn.* **1964**, *28*, 650–654.
42. S. V. Vitt; E. A. Paskonova; M. B. Saporovskaia; V. M. Belikov, *Prikl. Biokhim. Mikrobiol.* **1990**, *26*, 279–282.
43. S. Capasso; A. Vergara; L. Mazzarella, *J. Am. Chem. Soc.* **1998**, *120* (9), 1990–1995.
44. X. Fu; M. L. G. Ferreira; F. J. Schmitz; M. Kelly-Borges, *J. Nat. Prod.* **1998**, *61*, 1226–1231.
45. G. S. Jayatilake; M. P. Thornton; A. C. Leonard; J. E. Grimwade; B. J. Baker, *J. Nat. Prod.* **1996**, *59*, 293–296.
46. M. Adamsczeski; A. R. Reed; P. Crews, *J. Nat. Prod.* **1995**, *58*, 201–208.
47. M. Bodanszky; G. F. Singler; A. Bodanszky, *J. Am. Chem. Soc.* **1973**, *95*, 2352–2360.
48. C. L. Atkins; J. B. Neilands, *Biochemistry* **1968**, *7*, 3734–3737.
49. R. Brown; C. Kelly; S. E. Wibberley, *J. Org. Chem.* **1965**, *30*, 277–286.
50. J. L. Johnson; W. G. Jackson; T. E. Eble, *J. Am. Chem. Soc.* **1951**, *73*, 2947–2952.
51. M. O. Forster; W. B. Saviole, *J. Chem. Soc.* **1922**, *121*, 816–820.
52. P. S. Steyn, *Tetrahedron* **1973**, *29*, 107–111.
53. T. Takahara; F. D. Battaina; D. D. Ross; S. A. Akman; N. R. Bachur; K. R. Bailey; E. Lakatos; A. Peterkofsky, *Endocrinology* **1983**, *56*, 312–319.
54. Y.-S. Chen, *Bull. Agric. Chem. Soc. Jpn.* **1960**, *24*, 372–377.
55. Y. Kodaira, *Agric. Biol. Chem. (Tokyo)* **1961**, *25*, 261–270.
56. S. H. Park; G. A. Strobel, *Biochim. Biophys. Acta* **1994**, *1199*, 13–19.
57. H. Plieninger; H. Herzog, *Monatsh. Chem.* **1967**, *98*, 807–811.
58. J. H. Birkenshaw; Y. S. Mohammed, *Thom. Biochem. J.* **1962**, *85*, 523–528.
59. F. Caesar; K. Janssen; E. Mutschler, *Pharm. Acta Helv.* **1969**, *44*, 676–685.
60. N. P. Jensen; C. O. Gitterman; T. Y. Chen; B. H. Arison; J. L. Beck, Isolation of a new antitumor antibiotic from *Streptomyces griseoluteus. Chem. Eng. News* **1973**, 4 April, p 24.
61. M. Y. Karpeiskh; Y. N. Breusov; R. M. Khomatov; E. S. Severin; O. C. Polyanovskh, *Biokhimiya* **1963**, *28*, 342–347.
62. A. S. Khoklov; G. G. Loshkin, *Tetrahedron Lett.* **1963**, *27*, 1881–1885.
63. B. Heinemann; M. A. Kaplan; R. D. Muir; I. R. Hooper, *Antibiot. Chemother.* **1953**, *3*, 1239–1244.
64. A. Dansi; A. Dal Pozzo; V. Mariotti; B. Bonferoni; M. Piccioni, *Stärke* **1970**, *22*, 305–309.
65. G. P. Rizzi, Heat-Induced Flavour Formation from Peptides. In *Thermal Generation of Aromas*; T. H. Parliament, R. J. McGorrin, C. T. Ho, Eds.; American Chemical Society: Washington, DC, 1989; Vol. 409, Chapter 16, pp 172–181.
66. F. Roudot-Algaron; D. Le Bars; J. Einhorn; J. Adda; J. C. Gripon, *J. Food Sci.* **1993**, *58*, 1005–1009.
67. S. A. Eriksen; I. S. Fagerson, Non-Volatile Nitrogen Compounds in Hydrolysed Vegetable Protein. In *Cereals for Food and Beverages: Recent Progress in Cereal Chemistry and Technology*; G. E. Inglett, L. Munck, Eds.; Academic Press: New York, 1980; pp 395–408.
68. K. Takahashi; M. Tadenuma; K. Kitamoto; S. Sato, *Agric. Biol. Chem.* **1974**, *38*, 927–932.
69. T. Kawai; Y. Ishida; H. Kakiuchi; N. Ikeda; T. Higshida; S. Nakamura, *J. Agric. Food Chem.* **1991**, *39*, 770–777.
70. M. Gautschi; J. P. Schmid, *J. Agric. Food Chem.* **1997**, *45*, 3183–3189.
71. M. Ginz; U. H. Engelhardt, *J. Agric. Food Chem.* **2000**, *48*, 3528–3532.
72. M. Bodanszky, *Principles of Peptide Synthesis*, 2nd ed.; Springer-Verlag: Berlin, Heidelberg, New York, 1993; pp 1–318.
73. R. Schwyzer; B. Iselin; W. Rittel; P. Sieber, *Helv. Chim. Acta* **1956**, *39*, 872.
74. K. D. Kopple; C. A. Bush, *Tetrahedron Lett.* **1972**, *6*, 483–486.
75. I. L. Karie, X-ray Analysis: Conformation of Peptides in the Crystalline State. In *The Peptides. Analysis, Synthesis, Biology*; E. Gross, J. Meienhofer, Eds.; Academic Press: New York, 1981; Vol. 4, pp 4–8.
76. G. E. Krejcarek; B. H. Dominy; R. G. Lawton, *Chem. Commun. (Camb.)* **1968**, *22*, 1450–1452.
77. P. E. Young; V. Madison; E. R. Blout, *J. Am. Chem. Soc.* **1976**, *98* (17), 5365–5371.
78. P. M. Hardy; B. Ridge, Recent Advances in the Chemistry of Cyclic Peptides. In *Progress in Organic Chemistry*; W. Carruthers, J. K. Sutherland, Eds.; Butterworths: London, 1973; Vol. 8, p 129.
79. C. H. Hassal; A. Thomas, *Chem. Britain* **1971**, *7*, 145.
80. Y. A. Ovchinnikov; V. T. Ivanov, *Tetrahedron* **1975**, *31*, 2177–2209.
81. W. Pardridge, *Ann Rep. Med. Chem.* **1985**, *20*, 305.
82. H. Bhargava, *Eur. J. Pharmacol.* **1984**, *100*, 109.

83. R. Chipkin; M. Latranyl, *Eur. J. Pharmacol.* **1984**, *100*, 239.
84. C. Toniolo, *Int. J. Peptide Protein Res.* **1990**, *35*, 287–300.
85. R. Ramani; K. Venkatesan; R. E. Marsh; W.-J. Hu Kung, *Acta Crystallogr.* **1976**, *B32*, 1051–1056.
86. M. Henczi; D. F. Weaver, *Rapid Commun. Mass Spectrom.* **1995**, *9*, 800–803.
87. C. J. Dinsmore; D. C. Beshore, *Tetrahedron* **2002**, *58*, 3297–3312.
88. J. Haüsler; R. Jahn; U. Schmidt, *Chem. Ber.* **1978**, *111*, 361–366.
89. Y. Imanishi; T. Sugihara; M. Tanihara; T. Higashimura, *Chem. Lett.* **1975**, *4* (3), 261–264.
90. Y. Imanishi, For a Review on Cyclic Peptides Including DKPs. In *Advances in Polymer Sciences*; Springer Verlag: Heidelberg, 1976; Vol. 20, pp 1–77.
91. J. N. Lambert; J. P. Mitchell; K. D. Roberts, *J. Chem. Soc. Perkin Trans. 1* **2001**, 471–484.
92. N. Shimazaki; I. Shima; M. Okamoto; K. Yoshida; K. Hemmi; M. Hashimoto, *Lipids* **1991**, *26* (12), 1175–1178.
93. E. Lioy; J. Suarez; F. Guzman; S. Siegrist; G. Pluschke; M. E. Patarroyo, *Angew. Chem. Int. Ed.* **2001**, *40* (14), 2631–2635.
94. T. Johnson; M. Liley; T. J. Cheeseright; F. Begum, *J. Chem. Soc. Perkin Trans. 1* **2000**, 2811–2820.
95. D. X. Wang; M. T. Liang; G. J. Tian; H. Lin; H. Q. Lui, *Tetrahedron Lett.* **2002**, *34*, 865–867.
96. A. H. Sijpkes; G. J. van der Kleut; S. C. Gill, *Biophys Chem.* **1994**, *52*, 75–82.
97. C. Goolcharran; R. T. Borchardt, *J. Pharm. Sci.* **1998**, *87*, 283–288.
98. G. D. Grant, The Medicinal Chemistry of Cyclo(Trp-Trp), Cyclo(Gly-Trp), and Cyclo(Gly-Gly). Ph.D. Thesis, University of Port Elizabeth: Port Elizabeth, 2002.
99. S. D. Bull; S. G. Davies; R. M. Parkin; F. Sanchez-Sancho, *J. Chem. Soc. Perkin Trans. 1* **1998**, 2313–2320.
100. A. Siwicka; K. Wojtasiewicz; A. Zawadzka; J. K. Maurin; Z. Czarnocki; *Tetrahedron Asymmetry* **2005**, *16*, 2071–2073.
101. D. J. Triggle, The Transition from Agonist to Antagonist Activity: Symmetry and Other Considerations. In *Practice in Medicinal Chemistry*; C. G. Wermuth, Ed.; Academic Press: London, 1996; p 550.
102. O. Grahl-Nielson, *Tetrahedron Lett.* **1969**, *10* (33), 2827.
103. C. M. Deber; F. A. Bovey; J. P. Carver; E. R. Blout, *J. Am. Chem. Soc.* **1970**, *92* (21), 6191–6198.
104. T. Mizuma; S. Masubuchi; S. Awazu, *J. Pharm. Pharmacol.* **1997**, *49*, 1067–1071.
105. Y. J. Fei; Y. Kanai; S. Nussberger; V. Ganapathy; V. Leiback; F. H. Romero; M. F. Singh; S. K. Boron; M. A. Hediger, *Nature* **1994**, *368*, 563–566.
106. D. A. Matthews; D. Burston, *Clin. Sci.* **1984**, *67*, 541–549.
107. A. Rubino; M. Filed; H. Shwachman, *J. Biol. Chem.* **1971**, *246*, 3542–3548.
108. V. Ganapathy; J. F. Mendicino; H. L. Frederick, *J. Biol. Chem.* **1981**, *256*, 118–124.
109. Y. J. Fei; V. Ganapathy; F. H. Leibach, *Prog. Nucleic Acid Res. Mol. Biol.* **1998**, *58*, 239–261.
110. D. Meredith; C. A. Boyd, *Cell. Mol. Life Sci.* **2000**, *57* (5), 754–778.
111. I. Rubio-Aliaga; H. Daniel, *Trends Pharmacol. Sci.* **2002**, *23* (9), 434–440.
112. H. Shen; D. Smith; T. Yang; Y. Huang; J. Schnermann; F. Brosius, *Am. J. Physiol. Renal Physiol.* **1999**, *276*, F658.
113. D. E. Smith; A. Pavlova; U. V. Berger; M. A. Hediger; T. Yang; Y. G. Huang; J. B. Schnermann, *Pharm. Res.* **1998**, *15*, 1244–1249.
114. S. Nussberger; M. Hediger, *Exp. Nephrol.* **1995**, *3* (4), 211–218.
115. M. Palacin; R. Estevez; J. Bertran; A. Zorzano, *Physiol. Rev.* **1998**, *78*, 969–1054.
116. T. Mizuma; T. Narasaka; S. Awazu, *J. Pharm. Pharmacol.* **2002**, *54* (9), 1293–1296.
117. T. Mizuma; T. Narasaka; S. Awazu, *Biol. Pharm. Bull.* **2003**, *26* (11), 1625–1628.
118. M. Delaforge; G. Bouillé; A. Jaouen; C. K. Jankowski; C. Lamouroux; C. Bensoussan, *Peptides* **2001**, *22*, 557–565.
119. T. Shiba; K. Nunami, *Tetrahedron Lett.* **1974**, *6*, 509–512.
120. N. Minamiura; Y. Matsumura; T. Yamamoto, *J. Biochem.* **1972**, *72*, 841–848.
121. T. Matoba; T. Hata, *Agric. Biol. Chem.* **1972**, *36*, 1423–1431.
122. K. Bláha; J. Smolíková; A. Vítek, *Collect. Czech. Chem. Commun.* **1966**, *31*, 4296.
123. T. Miyazawa, *J. Mol. Spectrosc.* **1960**, *4*, 155–167.
124. L. J. Bellamy, *The Infrared Spectra of Complex Molecules*; Methuen and Co. Ltd.: London, 1957; Chapter 12, pp 1856–1876.
125. J. Szafranek; Z. Palacz; Z. Grzonka, *Org. Mass. Spectrum* **1976**, *11*, 920–930.
126. H. J. Svec; G. A. Junk, *J. Am. Chem. Soc.* **1964**, *86*, 2278–2282.
127. G. Peterson, In *Recent Advances in the Design and Synthesis of Peptidomimetics*. Evans Group Literature Seminar; Harvard University: Cambridge,13 April, 2001; pp 1–4.
128. L. E. Webb; C.-F. Lin, *J. Am. Chem. Soc.* **1971**, *93*, 3818.
129. R. Degeilh; R. E. Marsh, *Acta Crystallogr.* **1959**, *12*, 1007.
130. J. Vicar; J. Smolikova; K. Blaha, *Collect. Czech. Chem. Commun.* **1972**, *37*, 4060.
131. Z. Siemion, *Org. Magn. Reson.* **1971**, *3*, 545.
132. K. D. Kopple; M. Ohnishi, *J. Am. Chem. Soc.* **1969**, *91* (4), 962–970.
133. K. D. Kopple; D. H. Marr, *J. Am. Chem. Soc.* **1967**, *89*, 6192–6200.
134. E. H. Strickland; M. Wilcheck; J. Horwitz; C. Billups, *J. Biol. Chem.* **1970**, *245*, 4168.
135. H. Edelhoch; R. E. Lippoldt; M. Wilcheck, *J. Biol. Chem.* **1968**, *243*, 4799–4805.
136. J. Coillet; B. Pullman; B. Marigret, *Biopolymers* **1971**, *10*, 221.
137. N. Funasaki; S. Hada; S. Neya, *Anal. Chem.* **1993**, *65*, 1861–1867.
138. R. Jankowska; J. Ciarkowski, *Int. J. Peptide Protein Res.* **1987**, *30*, 61–78.
139. A. K. Bose; M. S. Manhas; R. F. Tavares; J. M. Van der Veen; H. Fujiwara, *Heterocycles* **1977**, *7*, 1227–1270.
140. G. Gawne; G. W. Kenner; N. H. Rogerts; R. C. Sheppard; K. Titlestad, Diamagnetic Shielding Effects and the Conformations of Small Peptides. In *Peptides*; E. Bricas, Ed.; North-Holland: Amsterdam, 1968; pp 28–39.
141. R. Deslauriers; Z. Grzonka; K. Schaumburg; T. Shiba; R. Walter, *J. Am. Chem. Soc.* **1975**, *97* (18), 5093–5099.
142. V. Madison; P. E. Young; E. R. Blout, *J. Am. Chem. Soc.* **1976**, *98*, 5358–5363.
143. T. Yamazaki; K.-I. Nunami; M. Goodman, *Biopolymers* **1991**, *31* (13), 1513–1528.
144. J. P. Carver; E. R. Blout, *Treatise on Collagen*; Academic Press: New York, 1967; Vol. 1, pp 441–526.
145. P. H. Von Hippel; K.-Y. Wong, *Biochemistry* **1963**, *2*, 1387–1398.

146. E. Adams, Proline. *McGraw-Hill Encyclopedia of Science & Technology,* 5 th ed.; McGraw-Hill Book Company: USA, 1982; Vol. 11, pp 1–789.
147. R. Balasubramanian; A. V. Lakshminarayanan; M. N. Sabesan; G. Tegoni; K. Venkatesan; G. N. Ramachandran, *Int. J. Protein Res.* **1971**, *3*, 25–33.
148. T. Ashida; M. Kakudo, *Bull. Chem. Soc. Jpn.* **1974**, *47* (5), 1129–1133.
149. K. S. Pitzer; W. E. Donath, *J. Am. Chem. Soc.* **1959**, *81*, 3213.
150. D. L. Nelson; M. M. Cox, *Lehninger Principles of Biochemistry*, 4th ed.; W. H. Freeman and Company: New York, 2005; Chapters 3, 17, 18, 22, 23, pp 75–115, 631–655, 656–689, 833–880, 881–922.
151. T. Devenyi; G. Gergely, *Amino Acids, Peptides and Proteins*; Elsevier Scientific Publishing Company: Amsterdam, 1974; p 632.
152. R. Meier, *Schweizerische Zeitschrift für GanzheitsMedizin* **2005**, *17* (5), 290–293.
153. A. E. Harper; N. N. Yoshimura, *Nutrition* **1993**, *9* (5), 460–469.
154. P. J. Reeds, Dispensable and Indispensable Amino Acids for humans. http://jn.nutrition.org;*J. Nutr.* **2000**, *130* (7), 1835S–1840S. PMID 10867060 http://www.ncbi.nlm..nih.gov (accessed 2 April 2008).
155. M. W. Harrold, *Am. J. Pharm. Ed.* **1996**, *60*, 192–197.
156. C. M. Deber; V. Madison; E. R. Blout, *Acc. Chem. Res.* **1976**, *9*, 106–113.
157. B. López-Corcuera; A. Geerlings; C. Aragón, *Mol. Membr. Biol.* **2004**, *18* (1), 13–20.
158. J. C. Sheenan; K. Hasspacher; Y. L. Yeh, *J. Am. Chem. Soc.* **1959**, *81*, 6086.
159. R. H. Mazur; J. M. Schlatter, *J. Org. Chem.* **1963**, *28*, 1025.
160. A. L. Lehninger; D. L. Nelson, M. M. Cox, *Principles of Biochemistry*, 2nd ed.; Worth Publishers: New York, 1993; pp 1–1013.
161. V. Madison, *Biopolymers* **1977**, *16*, 2671–2691.
162. L. Crawford; W. N. Lipscomb; C. G. Schellman, *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 538.
163. C. J. Brandl; C. M. Deber, *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 917–921.
164. J. F. Brandts; H. R. Halvorson; M. Brennan, *Biochemistry* **1975**, *14* (22), 4953–4963.
165. F. A. Bovey; A. I. Brewster; D. J. Patel; A. E. Tonelli; D. A. Torchia, *Acc. Chem. Res.* **1972**, *5*, 193.
166. E. Smith; M. Bergmann, *J. Biol. Chem.* **1944**, *153*, 627.
167. E. Abderhalden; E. Komm, *Z. Physiol. Chem.* **1924**, *139*, 147–152.
168. K. D. Kopple, *J. Pharm. Sci.* **1972**, *61* (9), 1345–1356.
169. D. M. Blow, *Acc. Chem. Res.* **1976**, *9*, 145–152.
170. S. Feurté; D. Tomé; D. W. Gietzen; P. C. Even; S. Nicolaïdis; G. Fromentin, *Nutr. Neurosci.* **2004**, *5* (4), 269–278.
171. B. G. Katzung, *Basic and Clinical Pharmacology*, 6th ed.; Prentice-Hall International: New York, 1995; pp 1–1046.
172. A. Meister, *Biochemistry of the Amino Acids*, 2nd ed.; Academic Press: New York, 1995; pp 1–199.
173. B. F. Gisin; R. B. Merrifield, *J. Am. Chem. Soc.* **1972**, *94*, 3102–3106.
174. V. A. Robinson-Fuentes; T. M. Jefferies; S. K. Branch, *J. Pharm. Pharmacol.* **1997**, *49*, 843–851.
175. M. Steinberg; J. L. Bada, *J. Org. Chem.* **1983**, *48*, 2295–2298.
176. J. E. Battersby; W. S. Hancock; E. Canova-Davies; J. Oeswein; J. O'Connor, *Int. J. Peptide Prot. Res.* **1994**, *44*, 215–222.
177. J. A. Straub; A. Akiyama; P. Parmar; G. F. Musso, *Pharm Res.* **1995**, *12*, 305–308.
178. A. R. Oyler; R. E. Naldi; J. R. Lloyd; D. A. Graden; C. J. Shaw, *J. Pharm. Sci.* **1991**, *80*, 271–275.
179. M. Tanihara; T. Hiza; Y. Imanishi; T. Higashimura, *Bull. Chem. Soc. Jpn.* **1983**, *56*, 1155–1160.
180. E. Fischer, *Ber. Deutsch. Chem. Ges.* **1906**, *39*, 2893.
181. K. Bláha, *Collect Czech. Chem. Commun.* **1969**, *34*, 4000.
182. D. E. Nitecki; B. Halpern; J. W. Westley, *J. Org. Chem.* **1968**, *33*, 864.
183. N. Lichtenstein, *J. Am. Chem. Soc.* **1938**, *60*, 560–563.
184. H. F. Schott; J. B. Larkin; L. B. Rockland; M. S. Dunn, *J. Org. Chem.* **1947**, *12*, 490.
185. K. D. Kopple; H. G. Ghazarian, *J. Org. Chem.* **1968**, *33* (2), 862–864.
186. H. Zahn; D. Brandenburg, *Annalen* **1966**, *692*, 220.
187. G. P. Slater, *Chem Ind. (London)* **1969**, *32*, 1092.
188. K. Suzuki; Y. Sasaki, N. Endo; Y. Mihara, *Chem. Pharm. Bull.* **1981**, *29* (1), 233–237.
189. P. J. Milne; D. W. Oliver; H. M. Roos, *J. Crystallogr. Spectrosc. Res.* **1992**, *22*, 224–230.
190. D. V. Patel; E. M. Gordon, *Drug Discov. Today* **1966**, *1* (4), 134–144.
191. A. K. Szardenings; T. S. Burkoth; H. H. Lu; D. W. Tien; D. A. Campbell, *Tetrahedron* **1997**, *53*, 6573.
192. A. Haywood, The Medicinal Chemistry of the Tryptophan-Containing Cyclic Dipeptide: Cyclo(Trp-Pro). Ph.D. Thesis, University of Port Elizabeth, 2000.
193. T. Kanmera; S. Lee; H. Aoyagi; N. Izumiya, *Tetrahedron Lett.* **1979**, *20* (46), 4483.
194. J. C. Estevez; J. W. Burton; R. J. Estevez; H. Ardron; M. R. Wormald; R. A. Dwek; D. Brown; G. W. J. Fleet, *Tetrahedron Assymmetry* **1998**, *9*, 2137.
195. J. McMurray, *Organic Chemistry*, 4th ed.; Brooks/Cole Publishing Company: United States of America, 1996; pp 437–446, 1078.
196. M. Bodanszky; Y. S. Klausner; M. A. Ondetti, *Peptide Synthesis*, 2nd ed.; John Wiley & Sons: London, Great Britain, 1976; pp 1–195.
197. Y. B. Alakhov; A. A. Kiryushkin; V. M. Lipkin; G. W. A. Milne, *J. Chem. Soc. D.* **1970**, *7*, 406b–407.
198. I. M. Mcdonald; D. J. Dunstone; S. B. Kalindjian; I. D. Linney; C. M. R. Low; M. J. Pether; K. I. M. Steel; M. J. Tozer; J. G. J. Vinter, *J. Med. Chem.* **2000**, *43*, 3518.
199. J. Häusler; U. Smidt, *Chem. Ber.* **1974**, *107*, 145.
200. M. Farloni; G. Giacomelli; A. Porcheddu; M. Taddei, *Eur. J. Org Chem.* **2000**, *8*, 1669.
201. L. Diafi; J. Couquelet; P. J. Tronche, *J. Heterocycl. Chem.* **1990**, *27*, 2181.
202. E. Benedetti; R. E. Marsh; M. Goodman, *J. Am. Chem. Soc.* **1976**, *98* (21), 6676–6684.
203. A. Radzicka; R. Wolfenden, *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.
204. M. Tsuboi; T. Shimanouchi; S. Miyushima, *J. Am. Chem. Soc.* **1959**, *81*, 1406.
205. J. M. Indelicato; T. T. Norvilas; W. J. Wheeler, *J. Chem. Soc. Chem. Commun.* **1972**, *8*, 1162.

206. R. T. Borchardt; L. A. Cohen, *J. Am. Chem. Soc.* **1972**, *94*, 9175–9182.
207. M. Fresno; J. Alsina; M. Royo; G. Barany; F. Alberico, *Tetrahedron Lett.* **1998**, *39*, 2639–2642.
208. J. Goldberg, *Trends Biotechnol.* **2001**, *19* (10), 379.
209. L. Yanga; R. Tan; Q. Wang; W. Huangc; Y. Yinc, *Tetrahedron Lett.* **2002**, *42* (37), 6545–6548.
210. J. Bradbury, *Lancet Infect. Dis.* **2001**, *1* (2), 71.
211. C. J. M. Graz; G. D. Grant; S. C. A. Brauns; A. Hunt; H. Jamie; P. J. Milne, *J. Pharm. Pharmacol.* **2000**, *52*, 75–82.
212. P. J. Milne; A. L. Hunt; K. Rostoll; J. J. van der Walt; C. J. M. Graz, *J. Pharm. Pharmacol.* **1998**, *50*, 1331–1337.
213. K.-H. Rhee, *Int. J. Antimicrob. Agents* **2004**, *24* (5), 423–427.
214. P. Di micco; M. Romanol; A. Niglio; P. Nozzolillo; P. Federicol; P. Petronella; L. Nunziata; R. Torella, *Dig. Liver Disord.* **2001**, *33*, 546–550.
215. S. Iqbal, *The Breast* **2000**, *9*, 264–266.
216. A. K. Szardenings; V. Antonenko; D. A. Campbell; N. DeFrancisco; S. Ida; L. Shi; N. Sharkov; D. Tien; Y. Wang; M. Navre, *J. Med. Chem.* **1999**, *42* (8), 1348–1357.
217. S. D. Bryant; G. Balboni; S. Guerrini; S. Salvadori; R. Tomatis; L. H. Lazarus, *Biol. Chem.* **1997**, *378*, 107–114.
218. P. W. Baures; W. H. Ojala; W. J. Costain; M. C. Ott; A. Pradhan; W. B. Gleason; R. K. Mishra; R. L. Johnson, *J. Med. Chem.* **1997**, *40* (22), 3594–3600.
219. K. R. C. Prakasha; Y. Tanga; A. P. Kozikowski; J. L. Flippen-Andersonb; S. M. Knoblachc; A. I. Fadenc, *Bioorg. Med. Chem.* **2002**, *10* (9), 3043–3048.
220. I. M. D. Jackson; L. D. Adelman; T. L. Munsat; S. Forte; R. M. Lechan, *Neurology* **1986**, *36*, 1218–1223.
221. M. Mori; T. Mallik; C. Prasad; J. F. Wilber, *Biochem. Biophys. Res. Commun.* **1982**, *109*, 541–547.
222. M. D. Dayhoff, *Atlas of Protein Sequence and Structure;* National Biomedical Research Foundation: Silver Spring, MD, 1969; Vol. 4.
223. K. Miyashita; M. Murakami; M. Yamada; T. Iriuchijima; M. Mori, *J. Biol. Chem.* **1993**, *268* (28), 20863–20865.
224. K. Bauer; K. J. Graaf; A. Faivre-Bauman; S. Beier; A. Tixier-Vidal; H. Kleinkauf, *Nature* **1978**, *274*, 174–175.
225. F. Battaini; P. S. Dannies; A. Peterkofsky, *Life Sci.* **1984**, *35*, 2519–2527.
226. C. H. Emerson; S. Alex; L. E. Braverman; M. S. Safran, *Endocrinology* **1981**, *109*, 1375–1379.
227. S. W. Lamberts; T. J. Visser, *Eur. J. Pharmacol.* **1981**, *71*, 337–341.
228. A. Enjalbert; S. Arancibia; M. Ruberg; M. Priam; C. Kordon, *Eur. J. Pharmacol.* **1979**, *58*, 97–98.
229. S. Melmed; H. E. Carlson; J. M. Hershman, *Clin. Endocrinol.* **1982**, *16*, 97–100.
230. C. Prasad; J. F. Wilber; V. Akerstrom; A. Banerji, *Life Sci.* **1980**, *27*, 1979–1983.
231. G. Brabant; E. J. Wicking; E. Nieschlag, *Acta Endocrinol.* **1981**, *98*, 189–194.
232. H. Jikihara; H. Ikegami; K. Koike; K. Wada; I. Morishige; H. Kurachi; K. Hirota; A. Miyake; O. Tanizawa, *Endocrinology* **1993**, *132* (3), 953–958.
233. P. Fragner; O. Presset; N. Bernad; J. Martinez; C. Roze; S. Aratan-Spire, *J. Am. Chem. Soc.* **1997**, *273*, E1127–E1132.
234. C. Prasad; T. Matsui; A. Peterkofsky, *Nature* **1977**, *68*, 142–144.
235. W. A. Banks; A. J. Kastin; J. B. Jasper, *Pharmacol. Biochem. Behav.* **1992**, *43*, 939–941.
236. W. A. Banks; A. J. Kastin; V. Akerstrom; J. B. Jaspan, *Am. J. Physiol.* **1993**, *264*, E723–E729.
237. R. E. Chipkin; J. Stewart; K. Channabasavaish, *Pharmacol. Biochem. Behav.* **1980**, *12*, 93–98.
238. J. J. Jacobs; C. Prasad; J. F. Wilber, *Brain Res.* **1982**, *250*, 205–209.
239. C. Prasad; T. Marsui; J. Williams; A. Peterkofsky, *Biochem. Biophys. Res. Commun.* **1978**, *85*, 1582–1587.
240. C. Prasad, *Brain Res.* **1987**, *437*, 345–348.
241. M. J. Kukla; H. J. Breslin; C. R. Bowden, *J. Med. Chem.* **1985**, *8* (11), 1745–1747.
242. J. E. Morley; A. S. Levine; C. Prasad, *Brain Res.* **1981**, *210*, 475–478.
243. J. F. Wilber; D. Rogers; T. Iriuchijima; C. Prasad, *Trans. Assoc. Am. Phys.* **1986**, *99*, 245–249.
244. L. M. Kow; D. W. Pfaff, *Pharmacol. Biochem. Behav.* **1991**, *38*, 365–369.
245. L. M. Kow; D. W. Pfaff, *Pharmacol. Biochem. Behav.* **1991**, *38*, 359–364.
246. C. Prasad; A. Hamdi; J. W. Brock; C. W. Hilton, Cyclo(His-Pro) and Food Intake. In *The Science of Food Regulation*; G. Bray, D. Ryan, Eds.; LSU Press: Baton Rouge, LA, 1992; pp 277–289.
247. L. D. Koskinen, *Br. J. Pharmacol.* **1986**, *87*, 509–519.
248. T. Suzuki; T. Sakurada; T. Tadano; K. Kisara; Y. Sasaki; K. Suzuki, *J. Pharmacobio-dyn.* **1981**, *4*, 377–379.
249. H. N. Bhargava; G. A. Matwyshyn, *Eur. J. Pharmacol.* **1980**, *68*, 147–154.
250. H. N. Bhargava, *Neuropharmacology* **1981**, *20*, 699–702.
251. P. D. Butler; J. R. Bodner, *Peptides* **1987**, *8*, 299–307.
252. S. Kawamura; S. Sakurada; T. Sakurada; K. Kisara; Y. Sasaki; K. Suzuki, *Eur. J. Pharmacol.* **1985**, *112*, 287–294.
253. C. Prasad, *Behav. Brain Res.* **1988**, *27*, 241–245.
254. I. K. Hwang; V. L. W. Go; D. M. Harris; I. Yip; K. W. Kang; M. K. Song, *Diabetes Obes. Metab.* **2003**, *5*, 317–324.
255. R. M. Williams; R. W. Armstrong; J. S. Dung, *J. Med. Chem.* **1985**, *28*, 733–740.
256. R. Walter; R. F. Ritzmann; H. N. Bhargava; L. B. Flexner, *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76* (1), 518–520.
257. S. Sakurada; T. Sakurada; H. Jin; T. Sato; K. Kisara; Y. Sakai; K. Suzuki, *J. Pharm. Pharmacol.* **1982**, *34*, 750–751.
258. S. Blažičková; J. Rovenský; E. Kasafirek; M. Buc, *Int. J. Immunother.* **1994**, *3*, 89–93.
259. K. Ström; J. Sjögren; A. Broberg; J. Schnürer, *Appl. Environ. Microbiol.* **2002**, *68* (9), 4322–4327.
260. K. Kanoh; S. Kohno; J. Katada; Y. Hayashi; M. Muramatsu; I. Uno, *Biosci. Biotechnol. Biochem.* **1999**, *63* (6), 1130–1133.
261. K. Katagiri; K. Sato; S. Hayakawa; T. Matsushima; H. Minato, *J. Antibiotics* **1970**, *23*, 420–422.
262. S. C. Brauns; P. Milne; R. Naude; M. van de Venter, *Anticancer Res.* **2004**, *24*, 1713–1720.
263. S. C. Brauns; G. Dealtry; P. Milne; R. Naude; M. van de Venter, *Anticancer Res.* **2005**, *25*, 4197–4202.
264. M. B. Szkaradzinska; M. Kubicki; P. W. Codding, *Acta Crystallogr.* **1994**, *50*, 565–569.
265. H. Kanzaki; D. Imura; T. Nitoda; K. Kawazu, *J. Biosci. Bioeng.* **2000**, *90*, 86–89.
266. L. Lin; S. Okada; D. A. York; G. A. Bray, *Peptides* **1994**, *15*, 849–854.
267. A. C. Stierle; J. H. Cardellina II; G. A. Stroebel, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8008–8011.

268. H. M. Miziorko; G. H. Lorimer, *Annu. Rev. Biochem.* **1983**, *52*, 507–535.
269. M. Graz; A. Hunt; H. Jamie; G. Grant; P. Milne, *Pharmazie* **1999**, *10*, 772–775.
270. E. van der Merwe; D. Huang; D. Peterson; G. Kilian; P. J. Milne; M. van de Venter; C. Frost, *Peptides* **2008**, *29*, 1305–1311.
271. F. R. Lucietto; P. J. Milne; G. Kilian; C. L. Frost; M. van de Venter, *Peptides* **2006**, *27*, 2706–2714.
272. G. Kilian; H. Jamie; S. C. A. Brauns; K. Dyason; P. J. Milne, *Pharmazie* **2005**, *60* (4), 305–309.
273. K. McCleland; P. J. Milne; F. R. Lucietto; C. Frost; S. C. Brauns; M. van de Venter; J. Du Plessis; K. Dyason, *J. Pharm. Pharmacol.* **2004**, *56*, 1143–1153.
274. H. Wang; T. Usui; H. Osada; A. Ganesan, *J. Med. Chem.* **2000**, *43* (8), 1577–1585.
275. C. Cui; H. Kakeya; H. Osada, *Tetrahedron* **1996**, *52* (39), 12651–12666.
276. Y. Funabashi; T. Horiguchi; S. Inuma; S. Tanida; S. Harada, *J. Antibiot.* **1994**, *47* (11), 1202–1218.
277. M. E. Alvarez; D. R. Houck; C. B. White; J. E. Brownell; M. A. Bobko; C. A. Rodger; M. B. Stawicki; H. H. Sun; A. M. Gillum; R. Cooper, *J. Antibiot.* **1994**, *47* (11), 1195–1201.
278. H. Jamie; G. Kilian; K. Dyason; P. J. Milne, *J. Pharm. Pharmacol.* **2002**, *54*, 1659–1665.
279. S. A. Rees; M. J. Curtis, *Br. J. Pharmacol.* **1993**, *108*, 139–145.
280. H. Jamie; G. Kilian; P. J. Milne, *Pharmazie* **2002**, *57*, 638–642.
281. H. Jamie, The Medicinal Chemistry of the Isomers of the Cyclic Dipeptide Cyclo(Trp-Pro). Thesis,Philosophiae Doctor, Faculty of Health Sciences, University of Port Elizabeth, 2002.
282. S. Jones, The Medicinal Chemistry of the Cyclic Dipeptides Cyclo(Met-Trp) and Cyclo(Met-Tyr). Dissertation, Magister Scientiae, Faculty of Health Sciences, University of Port Elizabeth, 2002.
283. C. W. Versluis, The Medicinal Chemistry of Cyclo(L-Trp-L-Tyr) and Cyclo(D-Trp-L-Tyr). Dissertation, Magister Scientiae, Faculty of Health Sciences, University of Port Elizabeth, 2002.
284. R. Olivier, The Medicinal Chemistry of the Cyclic Dipeptides Cyclo(Gly-Tyr) and C yclo(Gly-Phe). Dissertation, Magister Scientiae, Faculty of Health Sciences, University of Port Elizabeth, 2002.
285. R. Pitchen, The Medicinal Chemistry of the Cyclic Dipeptides Cyclo(Met-Met) and Cyclo(Met-Gly). Dissertation, Magister Scientiae, Faculty of Health Sciences, University of Port Elizabeth, 2002.
286. E. Van der Merwe, The Medicinal Chemistry of Cyclo(Phe-Cys) and Cyclo(Tyr-Cys). Dissertation, Magister Scientiae, Faculty of Health Sciences, University of Port Elizabeth, 2005.
287. D. Huang, The Medicinal Chemistry of the Selected Glycine-Containing Cyclic Dipeptides: Cyclo(Gly-Leu) and Cyclo(Gly-Ile). Dissertation, Magister Scientiae, Faculty of Health Sciences, Nelson Mandela Metropolitan University, 2006.
288. T. Cunningham, The Medicinal Chemistry of the Selected Threonine Containing Diketopiperazines: Cyclo(Gly-Thr) and Cyclo(Gly-Ser). Dissertation, Magister Scientiae, Faculty of Health Sciences, Nelson Mandela Metropolitan University, 2006.
289. D. C. Peterson, The Medicinal Chemistry of the Selected Glycine-Containing Cyclic Dipeptides: Cyclo(Gly-L-Val) and Cyclo(Gly-L-Val). Dissertation, Magister Scientiae, Faculty of Health Sciences, Nelson Mandela Metropolitan University, 2006.
290. A. L. Kritzinger, The Medicinal Chemistry of Cyclo(Ser-Ser) and Cyclo(Ser-Tyr). Dissertation, Magister Scientiae, Faculty of Health Sciences, Nelson Mandela Metropolitan University, 2007.
291. H. W. Janse van Rensburg, The Medicinal Chemistry of Cyclo(Cys-Ile) and Cyclo(Cys-Val). Dissertation, Magister Scientiae, Faculty of Health Sciences, Nelson Mandela Metropolitan University, 2007.

Biographical Sketches

Pieter J. Milne was born in Caledon, South Africa in 1955. He completed his undergraduate studies in 1977. After working for 5 years as a pharmacist, he pursued his Masters and doctoral studies under the direction of Dr. L. M. Gerritsma and Professor D. W. Oliver, respectively. In 1994, he was appointed as a lecturer in Pharmaceutical Chemistry at the University of Port Elizabeth, now known as the Nelson Mandela Metropolitan University (NMMU). He became head of the Department of Pharmacy in 2007 and is also the director of the Cyclic Peptide Research Unit (CPRU).

Gareth Kilian completed his undergraduate B.Pharm. degree at the University of Port Elizabeth in 1999. In 2002 he completed his M.Sc. degree in medicinal chemistry *Cum Laude*, also at the University of Port Elizabeth, were his work was directed on the medicinal chemistry of aromatic cyclic dipeptides. He has since worked as a community service pharmacist in 2002 at Butterworth Hospital in the Eastern Cape, after which he was employed as a lecturer in Pharmaceutics at the NMMU, where he currently works. He is a Ph.D. candidate at the NMMU under Professor Pieter Milne, also working on cyclic dipeptides. He has coauthored several peer-reviewed manuscripts in international journals as well as supervising and cosupervising a number of M.Sc. candidates over the past 6 years in the fields of medicinal chemistry and pharmaceutical formulation.

5.21 Ubiquitin-Dependent Protein Degradation

Ashok N. Hegde, Wake Forest University Health Sciences, Winston Salem, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

5.21.1 Discovery of the Ubiquitin–Proteasome Pathway

Eukaryotic cells possess two major systems for protein breakdown, the lysosomal system and the ubiquitin–proteasome pathway. Initially, it was assumed that the lysosome was responsible for most of the intracellular protein degradation. Lysosomal proteolysis, which degrades proteins inside a vesicle, could not account for vastly different half-life of proteins and change in rate of degradation of proteins in response to physiological stimuli. In addition, it was found that intracellular protein degradation required ATP, which did not fit with what was known about lysosomal protein degradation. By the 1980s, it became apparent that a proteolytic mechanism distinct from the lysosomal system must exist in eukaryotic cells, which eventually led to the discovery of the ubiquitin–proteasome pathway.

The quest for identification of the intracellular protein degradation mechanism began with attempts to reconstitute ATP-dependent proteolysis in a cell-free system.[1] Avram Hershko began studying proteolysis using reticulocyte lysate system, which had been established by Etlinger and Goldberg.[2] Avram Hershko, his student Aaron Ciechanover, and Hershko's collaborator Irwin Rose worked out the role of ATP using reticulocyte lysates and model substrates such as lysozyme and globin. It was found that a heat-stable factor called ATP-dependent proteolysis factor (APF-1) was conjugated to substrates.[3–5] APF-1 was later identified to be ubiquitin by Wilkinson *et al.*[6] The pioneering work by Hershko, Ceichanover, and Rose was recognized by a Nobel Prize in chemistry in 2004.[7–9]

5.21.2 The Ubiquitin–Proteasome Pathway

In the ubiquitin–proteasome pathway, the substrate proteins are marked by covalent linkage to ubiquitin for degradation by a proteolytic complex, the proteasome.

5.21.2.1 Ubiquitin

Ubiquitin is a small protein of 76 amino acids (**Figure 1**). Ubiquitin had been isolated from thymus prior to its identification as APF-1.[10] The evidence available at the time indicated that it was a universal constituent of living cells. Indeed, ubiquitin is present in all eukaryotic cells and is one of the highly evolutionarily conserved proteins. For example, the yeast and human ubiquitins differ only at three amino acid residues.[11] Amino acid sequences of the sea slug *Aplysia* and human ubiquitin are 100% identical.[12] Ubiquitin is found in all eukaryotes but not in prokaryotes. Ubiquitin is encoded by a polyubiquitin gene in which several ubiquitin coding sequences are tandemly linked without any introns in between. The number of ubiquitin coding repeats is usually 5 or 6.[13] In species such as *Trypanosoma cruzi*, however, polyubiquitin genes with 52 tandemly linked ubiquitin coding sequences exist.[14] Ubiquitin is also encoded by genes in which the ubiquitin coding sequence is fused to sequences encoding small ribosomal subunits.[15] In the cell, ubiquitin is found as a single protein (monoubiquitin).

5.21.2.2 Ubiquitin Conjugation to Substrate Proteins

Ubiquitin marks the cellular proteins for degradation. When a protein specifically degrades in the cell, ubiquitin attaches to an ε-amino group of lysine residues in the substrate protein. First, a single ubiquitin is attached. Then, to an internal lysine residue in the ubiquitin a second ubiquitin is attached and thus by sequential linkages of monoubiquitins, a polyubiquitin chain grows. The polyubiquitinated substrate is then recognized by the proteasome and is degraded to small peptides and amino acids. Ubiquitin itself is not degraded but disassembled by a class of enzymes called deubiquitinating enzymes (DUBs).[16,17]

The process of degradation of a substrate protein can be divided into two steps: (1) covalent attachment of ubiquitin to the substrate (often referred to an ubiquitin conjugation or ubiquitination) and (2) degradation of the polyubiquitinated substrate and disassembly of the polyubiquitin chain and recycling of free ubiquitin (**Figure 2**).

The ubiquitin-conjugation step is a highly regulated step catalyzed by the action of three classes of enzymes called E1, E2, and E3 (so named because of the order of their elution on anion-exchange chromatography). E1, the ubiquitin-activating enzyme activates the free ubiquitin in an ATP-dependent reaction. Activated ubiquitin is then transferred to an E2, which is generally referred to as an ubiquitin-carrier enzyme. An enzyme belonging to a class of enzymes called E3s then ligates the activated ubiquitin to the substrate (ubiquitin ligases). The disassembly of polyubiquitinated chains is carried by DUBs (**Figure 2**).

Ubiquitin is attached to substrates through its C-terminal glycine residue. Attachment of a single ubiquitin (monoubiquitination; see **Figure 3**) to proteins usually signals a conformational change in a protein. Attachment of a single ubiquitin to multiple lysine residues in a substrate (multiple monoubiquitination) usually marks the protein substrates in the plasma membrane for endocytosis. The tag that marks a protein substrate for degradation is composed of ubiquitin attached to each other. After the first ubiquitin is attached to the substrate, a second ubiquitin is attached to an internal lysine residue in the first ubiquitin (Lys48). The Lys48-linked polyubiquitin chain (polyubiquitination; **Figure 3**) marks proteins for degradation by the proteasome.[16] Other unusual linkages of ubiquitins with each other at different lysines in the ubiquitin sequence are known to occur and will be discussed later in this chapter.

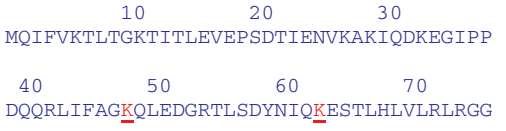


Figure 1 Amino acid sequence of ubiquitin. The sequence of human ubiquitin protein is shown using single-letter codes for amino acids. The key lysine (K) residues at the 48th and 63rd position are given in red letters.

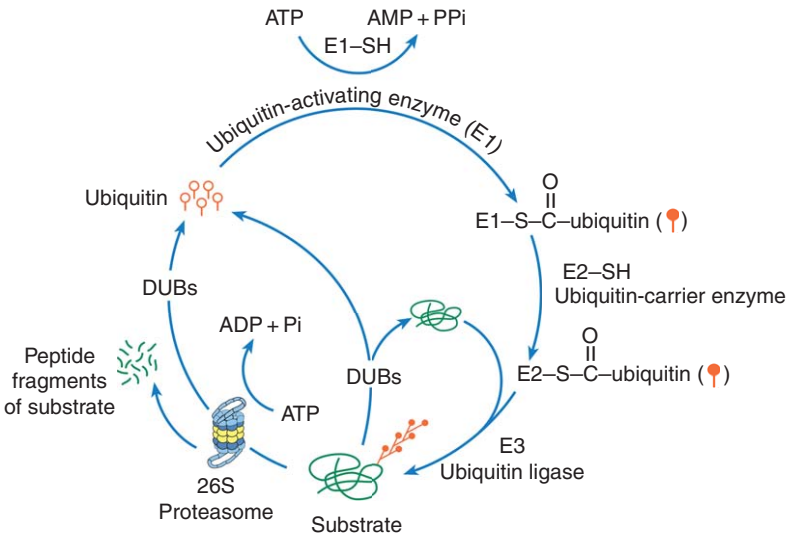


Figure 2 The ubiquitin-proteasome pathway. Steps in ubiquitin conjugation and degradation of ubiquitinated proteins. In this proteolytic pathway, ubiquitin (single ubiquitin molecule is represented by open circles with straight tails) is selectively and covalently linked to the substrate. The enzymatic process of attaching ubiquitin to substrates is called ubiquitination or ubiquitin conjugation and depends on the action of three different classes of enzymes E1, E2, and E3. First ubiquitin is activated by E1 to form an ubiquitin–AMP intermediate. Activated ubiquitin (closed circles with straight tails) is passed on to E2 (ubiquitin-carrier enzymes). E2s transfer ubiquitin to an E3 (ubiquitin ligase) that ligates the activated ubiquitin to the substrate. To the ubiquitin attached to the substrate another ubiquitin is attached and thus through successive linkages of ubiquitin a polyubiquitin chain forms. Polyubiquitinated substrates are degraded by a proteolytic complex called the 26S proteasome in an ATP-dependent reaction. Ubiquitin is not degraded but the polyubiquitin chain is disassembled and ubiquitin is recycled by deubiquitinating enzymes (DUBs). Before degradation by the proteasome, ubiquitination is reversible. DUBs can disassemble the polyubiquitin chain if a substrate is ubiquitinated erroneously and prevents the degradation of the substrate.

As to the ubiquitin terminology, 'polyubiquitin' is a well-accepted term to indicate ubiquitins linked to each other through internal lysines. As described above a 'polyubiquitin' tag marks protein substrate for degradation through the proteasome. There is chance for some confusion, however. The ubiquitin gene, which has ubiquitin coding sequences tandemly linked to each other, refers to the polyubiquitin gene. Therefore, I avoid saying polyubiquitin to refer to the gene and instead refer to the gene as 'tandemly linked ubiquitin gene'. In addition, previously the term 'multiubiquitin' was used to refer to what is now called 'polyubiquitin'. Finally, the term 'ubiquitylation' has been used in some publications. Because 'ubiquitination' makes intuitive sense, I prefer to use this term as do many others.

5.21.3 Ubiquitin-Conjugating Enzymes: E1, E2, and E3

Activation of ubiquitin by E1 requires ATP. E1 transfers the activated ubiquitin to an E2. Of the three classes of ubiquitin-conjugating enzymes (UBCs), E1 is the least physiologically regulated. In addition, because E1 activates ubiquitin, which is attached to all the substrates degraded by the ubiquitin–proteasome pathway, there is no substrate specificity in the action of this enzyme. E2s are more selective and believed to interact with specific E3s. Based on our knowledge of ubiquitination reaction, it can be asserted that E3 are the enzymes that possess substrate specificity.

5.21.3.1 E1

Based on studies in several species, it appears that there is only one El gene, although it can generate two El isoforms because of alternative translation initiation sites. E1 enzymes, which are about 105 K in molecular weight, recognize the C-terminal glycine residue in ubiquitin. The activation of ubiquitin is ATP dependent. ATP is hydrolyzed to generate AMP and a high-energy thiol-ester ubiquitin–AMP intermediate is formed.[1,18]

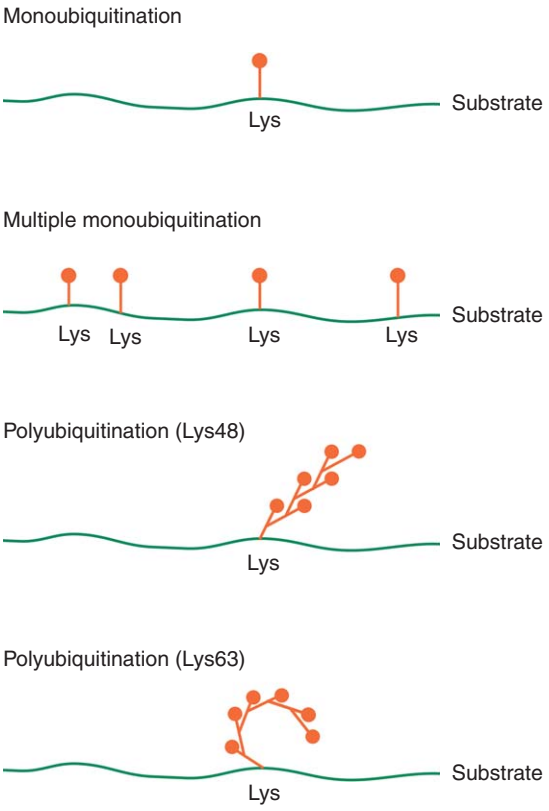


Figure 3 Multiple modes of ubiquitination. The substrate protein has been stretched out. A single ubiquitin can be attached to a single lysine residue (Lys) in the substrate (monoubiquitination), or single ubiquitins can be attached to multiple Lys residues in the substrate (multiple monoubiquitination). Other types of ubiquitination include attachment of a chain in which the second and subsequent ubiquitins are attached to the 48th lysine residue in the ubiquitin molecule (polyubiquitination Lys48) or 63rd lysine residue in the ubiquitin molecule (polyubiquitination Lys63).

The E1 protein is highly conserved through evolution. For example, human and yeast E1s are 53% identical. E1 contains a nuclear localization signal and is found in the nucleus in addition to other subcellular compartments.[19,20] This enzyme is also known to be phosphorylated and translocated to the nucleus in a cell cycle-dependent fashion.[21] Since only phosphorylated form of E1 is translocated to the nucleus despite the presence of the nuclear localization signal in the E1 sequence, it is conceivable that phosphorylation is an obligatory additional signal for nuclear localization. Evidence obtained by Stephen *et al.*[21] shows that phosphorylation facilitates either transport to the nucleus or its retention in the nucleus.

5.21.3.2 E2s

These enzymes are also called ubiquitin-carrier proteins or UBCs. Originally, these enzymes were believed only to carry the activated ubiquitin and pass it onto the E3s. Recent studies, however, suggest that at least some E2s can directly conjugate ubiquitin to substrates.[16]

E2 are structurally and functionally diverse. Early biochemical work by using reticulocyte system revealed five distinct proteins with properties of E2.[22] With the advent of genome sequencing, this observation has been corroborated. Even simple eukaryotes like yeast (*Saccharomyces cerevisiae*) have 13 genes potentially encoding E2s. The number of E2s in mammals is estimated to be in the range of 25–30.

Most E2s have a core domain of around 14 K, which is about 35% conserved between different E2s. The other part of the enzymes appears to be variable. Although most E2s are small (~36 K), some notable exceptions exist. For example, an E2, called BIR-repeat-containing ubiquitin-conjugating enzyme (BRUCE) is a gigantic 528-K protein.[23]

The diversity of E2 generates some degree of specificity in the ubiquitin-conjugating reaction. E2s bind to E3s selectively. Since the diversity of E3s is even greater than that of E2s, the combination of E2s and E3s potentially can generate a high degree of specificity. The heterogeneity of E2s is reflected in their subcellular localization as well. Although several E2s are present in the cytosol, some E2s are localized to other subcellular compartments. For example, BRUCE is localized to the Golgi apparatus and a yeast E2 called Ubc6 is anchored to the membrane of the endoplasmic reticulum.[24–27]

5.21.3.3 E3s

E3s are the enzymes that specifically recognize the substrates. E3s can be single proteins or complex of proteins. Single subunit E3s can accept ubiquitin in a thioester linkage from E2s and ligate ubiquitin to the substrate. When E3 contains multiple subunits, it is generally believed that the enzyme brings the E2 and the substrate together and facilitates the transfer of ubiquitin to the substrate. However, we recall that E3s are the most diverse among the UBCs and the least characterized. Therefore, it is possible that E3 catalytic mechanisms other than those described above might exist in nature.

There are three major classes of E3s: (1) HECT domain E3s, (2) really interesting new gene (RING) finger E3s, and (3) complex multisubunit E3s (**Figure 4**).

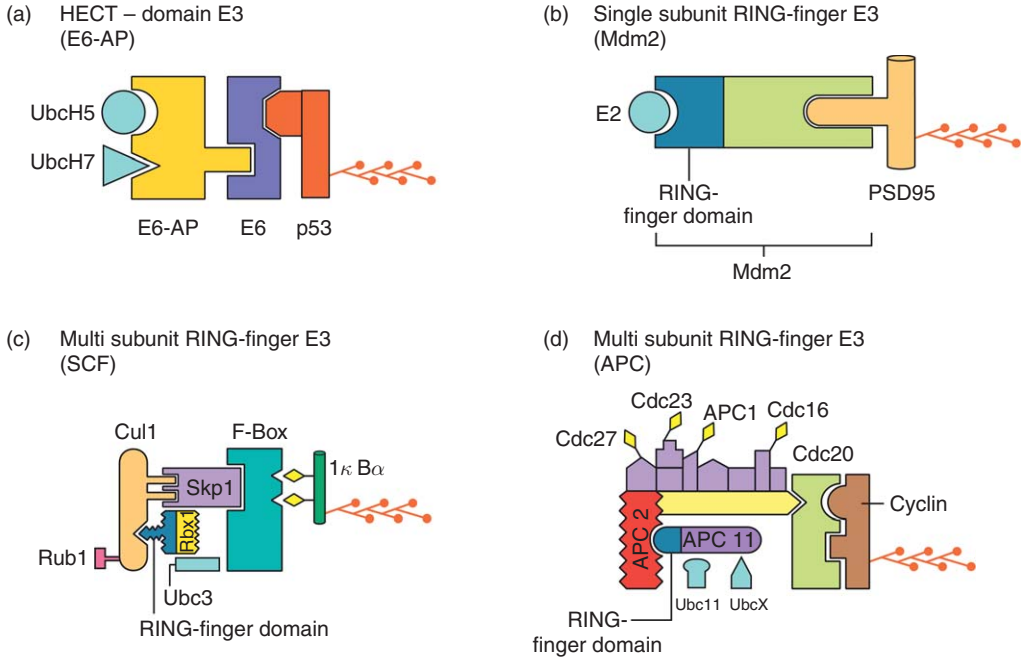


Figure 4 Classes of ubiquitin ligases (E3s). (a) HECT-domain E3. E6-AP ubiquitin ligase in combination with E6 protein and one of the two E2s (UbcH5 or UbcH7) ligates ubiquitin to the p53 tumor suppressor protein. (b) Single-subunit RING finger E3. Mdm2 ligates ubiquitin to PSD95 with the help of an E2 enzyme. (c) Multisubunit RING-finger E3. SCF ligases contain the substrate recognition site on an F box protein. Skp1 is an adaptor that joins the F box protein to Cul1. Ring-finger domain is on Rbx 1. The E2 is Ubc3. Cul 1 is modified by Rub1, another ubiquitin-like protein leading to an increase in the activity of the ligase complex. The substrate is phosphorylated (diamonds) IκBα. (d) Multisubunit RING-finger E3. APC is a more complex example of multisubunit RING-finger E3s and has a subunit composition distinct from that of SCF. Cdc20 protein in APC has the substrate (cyclin) recognition site. The RING-finger domain is on APC11. The E2s Ubc11 or UbcX can function with the APC ligase. In addition, several adaptor proteins, some labeled (Cdc27, Cdc23, APC 1, Cdc16) and some unlabeled, interact with Cdc20 and APC11. Diamonds on the adaptor subunits indicate phosphorylation. Polyubiquitin chain is shown on the substrates in each panel.

5.21.3.3.1 HECT domain E3s

The prototypical E3 in this class is the ubiquitin ligase called E6-AP that ligates ubiquitin to the tumor suppressor protein p53. E6 is a protein encoded by oncogenic strains of human papilloma virus. E6 associates with a cellular protein called E6-associated protein (E6-AP). The C-terminal region of E6-AP contains the catalytic domain of the ubiquitin ligase. E6-AP ligase can function with either of the E3s called UbcH5 and UbcH7 (**Figure 4(a)**). Later studies found that a family of proteins ubiquitin ligases with homology to the catalytic domain of E6-AP exists. These ubiquitin ligases came to be called *h*omologous to *E*6-AP carboxyl-terminus (HECT) domain E3s.[28,29]

In addition to the HECT domain, there is another domain in many E3s called the WW domain. The WW domain is thus named because of the characteristic tryptophan (W is the single letter code for the amino acid tryptophan) believed to be critical for protein–protein interaction. The WW domain-containing E3s also tend to have a C2 domain.[16,30] The presence of C2 domain is highly relevant to nervous system function because C2 domain responds to the elevation of intracellular Ca^{2+} and helps in translocation to the plasma membrane. Therefore, presence of this domain in neuronal HECT E3s might be critical in ligating ubiquitin to neurotransmitter receptors or proteins associated with them.

5.21.3.3.2 RING-finger E3s

These E3s are called RING-finger E3s because they contain a RING-finger domain, which consists of seven cysteine residues and one histidine residue forming a single-folded domain binding two zinc ions.[31–34] The arrangement of metal-binding residues in the RING-finger domain contrasts with the tandem arrangement in zinc-finger domain found in many proteins. Zinc-finger domain typically consists of two β-sheets each carrying a cysteine residue, and an α-helix carrying two histidine residues, which together bind a single zinc ion.[35]

The RING-finger motif was originally discovered by sequence database searches using the N-terminal sequence of a new gene called *r*eally *i*nteresting *n*ew *g*ene 1 (RING1).[36] Although numerous other proteins were found to have the RING-finger motif, the biological function of these proteins remained elusive. During the past few years, several ubiquitin ligases were found to contain the RING finger. It is now generally believed that RING-finger motif in ubiquitin ligases is critical for the transfer of ubiquitin to substrates or to RING-finger proteins themselves.

The RING-finger category of E3s can be subdivided into RING-finger E3s with a single subunit or RING-finger E3s with multiple subunits.

5.21.3.3.2(i) Single subunit RING-finger E3s Single subunit RING-finger E3s contain the RING-finger domain and the substrate recognition site in the same protein. One of the well-characterized single subunit RING-finger E3s is Mdm2 that ubiquitinates p53 in normal cells. As discussed above E6-AP, an HECT ubiquitin ligase, ubiquitinates p53 in human papilloma virus (HPV)-infected cells. A recent study showed that in HPV-infected cells E6-AP ubiquitinates p53. Although Mdm2 is present in HPV-infected cells, it does not mediate ubiquitination of p53.[37] Other studies using antisense oligonucleotides directed against E6-AP showed that E6-AP is essential for the degradation of HPV-positive cells but not in HPV-negative cells.[38,39] Conversely, decreasing Mdm2 expression or expression of Mdm2-inactivating peptides decreased p53 degradation in HPV-negative cells but not in HPV-positive cells.[39,40] Interestingly, the structural determinants on p53 that are recognized by E6-AP and Mdm2 are different from each other (illustrated as differently shaped protrusions on p53 in **Figures 4(a)** and **4(b)**). For example, p53 that is a substrate for E6-AP has an asparagine at position 268 whereas p53 that is a substrate for Mdm2 has an aspartate at position 268.[37]

Another well-studied single subunit RING-finger E3s is c-Cbl, an ubiquitin ligase that catalyzes attachment of ubiquitin to receptor tyrosine kinases. Based on the crystal structure of c-Cbl, it appears that the function of the RING finger is to serve as a scaffold to position of the E2 called UbcH7 and the substrate (receptor tyrosine kinase) properly for ubiquitin ligation of the substrate. A RING-finger E3 widely known to neuroscientists is parkin, the gene that mutated a recessive form of juvenile Parkinson's disease. Parkin has a RING-finger domain at its C-terminus that is critical for ubiquitin-ligating function. In addition, parkin has an ubiquitin-like (UbL) domain at its N-terminus. The UbL domain in parkin and other proteins is probably critical for interaction with the proteasome.[41,42]

5.21.3.3.2(ii) Multisubunit RING-finger E3s

1. SCF (Skp1-*c*ullin-*F*-box protein) complex: A well-characterized multisubunit RING-finger E3 is the SCF complex that ubiquitinates IκBα (**Figure 4(c)**). At the heart of the SCF complex is the RING-finger domain-containing protein Rbx1. The SCF-type ligases have another invariant protein called cullin. The theme appears to be that the cullins interact with linker proteins such as Skp1 to recruit substrate-interacting proteins such as the F-box proteins. There are at least five different cullins in mammals. There are several F-box proteins as well. Although it is not clear how many F-box proteins exist in mammals, the budding yeast genome comprises 17 F-box proteins. Therefore, just with the cullin F-box combination alone, it would be possible to generate close to a hundred E3s with differing specificities. Although the regulation of SCF ligase is not completely understood, two mechanisms of posttranslational regulation have been discovered so far. One is the covalent linkage of ubiquitin-related protein Rub1 to a cullin (Cul1 in **Figure 4(c)**). The second mode appears to be regulation of levels of F-box proteins through ubiquitin-mediated degradation through an autocatalytic mechanism.[43,44]
2. Anaphase-promoting complex: This class of E3s is typified by the anaphase-promoting complex (APC) (**Figure 4(d)**). Although APC could be included in the 'multisubunit RING-finger E3' category because of the presence of a subunit with a RING-finger domain (APC11), this ubiquitin ligase is distinct from the SCF ligase in overall subunit combination. For example, instead of one adaptor found in SCF ligases (such as Skp1), APC has multiple subunits that serve as adaptors. In addition, unlike SCF ligases, substrate phosphorylation is not an important determinant for specific-substrate recognition by the APC ligase. Rather, substrate specificity of APC ligases appears to be modulated by incorporation of 'specificity factors' into the ligase complex. For example, Cdc20 (see **Figure 4(d)**) enables APC to degrade substrates at the onset of anaphase such as the anaphase inhibitor Pds1p[45,46] whereas substitution of Cdc20 with another specificity factor called Hct1 enables APC to degrade a different set of substrates such as mitotic cyclins late in the anaphase.[47] APC acts together with an E2 Ubc11 or UbcX. One of the widely studied substrates of APC is mitotic cyclin. This substrate has a short stretch of nine amino acids called the 'destruction box', which is critical for recognition by the APC ubiquitin ligase.[48,49]

5.21.3.4 E4s

Another class of proteins called E4s that elongate polyubiquitin chain has been discovered. Protein product of a gene previously known as ubiquitin fusion degradation protein 2 (UFD2) in yeast was found to catalyze ubiquitin chain assembly along with E1, E2, and E3 and was named E4.[50] A characteristic feature of E4s is that these enzymes contain a conserved motif called the U-box (named because it is present in UFD2). A recent study showed that U-box proteins have ubiquitin ligase activity that is dependent on E1 and E2 but independent of E3.[51] Therefore, some E4s might be ubiquitin ligases. In support of this idea, U-boxes are considered modified RING-finger domains.[52] Moreover, comparative nuclear magnetic resonance studies of the U-box and the RING-finger domain revealed that the two domains are structurally similar.[53] If E4s are indeed ubiquitin ligases, they might represent a subfamily of E3s. Since there have not been many studies on E4s and the first E4 discovered function is being a cofactor for an E3,[50] it might be premature to conclude that E4s belong to a special class of E3s. U-box proteins/E4s are much fewer in number compared to E3s. For example, in the human genome the estimated number of U-box-containing proteins is 19.[54]

5.21.4 Regulation of the Ubiquitin–Proteasome Pathway

Proteolysis by the ubiquitin–proteasome pathway can be regulated at the ubiquitin-conjugation step or at the proteasome step. Since the specificity of ubiquitination lies at the conjugation step, clearly regulation of the conjugation process is important in determining whether or not a substrate is targeted for degradation. Regulation of proteasome has a global effect on degradation of cellular substrates.

5.21.4.1 Regulation of Ubiquitin Conjugation

Ubiquitin requires to be conjugated to the right substrate, at the right place in the cell, and at the right time in order to control physiological processes properly. Commitment of a substrate protein to ubiquitin–proteasome-mediated degradation is regulated by (1) modifying the substrate, (2) modulation of ubiquitin ligase activity, and (3) removal of ubiquitins.

5.21.4.1.1 Modification of the substrate

Protein substrates are degraded in the cell at specific times in response to physiological stimuli. In addition, degradation of substrates is probably spatially restricted within a cell. Based on accumulated evidence, it appears that the vulnerability or resistance to ubiquitin–proteasome-mediated degradation is regulated usually by a posttranslational modification. The protein substrates are modified in two main ways: (1) by phosphorylation or (2) by allosteric modifications.

5.21.4.1.1(i) Phosphorylation of the substrate Phosphorylation of a substrate can make it vulnerable for ubiquitination or resistant to ubiquitination. For example, yeast cyclins Cln2 and Cln and the cyclin-dependent kinase inhibitor p27Kip1, transcriptional regulators IκBα and β-catenin, are ubiquitinated after phosphorylation. In neurons, ubiquitination of p35, a neuronal-specific activator of cyclin-dependent kinase 5 (Cdk5), is stimulated when the protein is phosphorylated by Cdk5 within the active kinase complex.[55]

It is instructive to consider examples of some substrates as to how the regulation at the level of substrate works. The transcription factor NF-κB is inhibited by IκBα, which binds to NF-κB and keeps it in an inactive form in the cytosol. NF-κB is activated by numerous external stimuli such as cytokines, ionizing radiation, and neuronal injury. Activation of NF-κB is initiated by proteolysis of IκBα, which releases NF-κB to be translocated to the nucleus where it initiates transcription. Ubiquitination of IκBα requires phosphorylation on Ser32 and Ser36. Upon phosphorylation, IκBα is recognized by a specific multisubunit RING-finger ligase complex called $SCF^{\beta\text{-}TrCP}$. Activity of $SCF^{\beta\text{-}TrCP}$ ligase seems to be constitutive, while the kinases that phosphorylate IκBα are activated by the stimuli known to induce NF-κB-mediated transcription.[56–59]

A second example is that of Sic1, an inhibitor of cyclin B–CDK (cyclin-dependent kinase) complexes. Sic1 inhibits progression from G1 to S phase (DNA synthesis) in the cell cycle.[60] Rapid degradation of Sic1 at the end of the G1 phase allows the cell to initiate DNA replication. It was found that Sic1 is phosphorylated by G1 cyclin–CDK complex. (Note that the cyclin–CDK complex required for G1 is different from the one required for S phase, i.e., cyclin B–CDK.) Phosphorylated Sic1 is recognized by SCF^{Cdc4} ubiquitin ligase. Affinity of SCF^{Cdc4} to phosphorylated Sic1 increases with the number of phosphorylated residues. Since the activity of SCF^{Cdc4} is constitutive, it is thought that Sic1 phosphorylation is the key site of regulation in ubiquitination and degradation of Sic1. This is a remarkable instance of how regulation of degradation of a substrate by the ubiquitin pathway plays a role in temporal regulation of a physiological process. In the cell cycle, the protein kinase required for progression of one stage (G1) in the cell cycle G1 cyclin–CDK complex phosphorylates and causes the ubiquitin-mediated degradation of the inhibitor of the next stage (S phase) of the cell cycle.[61–63]

Phosphorylation of a substrate can make a stable protein vulnerable to degradation as seen above, or phosphorylation can have the opposite effect of stabilizing a short-lived protein. An example of physiological regulation by phosphorylation and ubiquitination is the stabilization of the *Aplysia* transcription factor known as CCAAT enhancer binding protein (Ap-C/EBP) in neurons.Yamamoto *et al.*[64] found that phosphorylation by MAP kinase is required for DNA binding by Ap-C/EBP. Interestingly, MAP kinase phosphorylation also makes the transcription factor resistant to ubiquitin-mediated degradation. Once Ap-C/EBP function is completed, it is degraded by the ubiquitin pathway, thereby restricting the expression of the transcription factor to a narrow time window. Based on the available data, it can be inferred that Ap-C/EBP is dephosphorylated upon completion of its role in transcription and dephosphorylation leads to the degradation of Ap-C/EBP. Another instance of MAP kinase phosphorylation that renders the substrate resistant to ubiquitination is that of the transcription factor jun.[65] The quantity of phosphorylated c-jun in vertebrate neurons has shown to be increased through the activation of NMDA-type glutamate receptors.[66] Similar to the example of transcription factor C/EBP, jun phosphorylation activates its DNA binding and its ability to activate transcription, and causes it to be resistant to ubiquitin-mediated degradation. Several other substrates of the ubiquitin

pathway are regulated by phosphorylation as well. For example, degradation of the protein product of the protooncogene c-*mos* is inhibited by phosphorylation on Ser3. Antiapoptotic protein Bcl2 is stable when it is phosphorylated. Stimuli that induce apoptosis cause Bcl2 to be dephosphorylated and degraded.[67,68]

How does phosphorylation of the substrate control ubiquitination? In the case of protein such as c-jun, phosphorylation perhaps masks a signal in the substrate recognized by the E3 ligase such that the substrate is unavailable for ubiquitination. In contrast, in those substrates that become susceptible to degradation after phosphorylation such as yeast cyclins, it is probable that phosphorylation unmasks a sequence or tertiary structure in the substrate, which the E3 ligase can recognize for ubiquitination.

5.21.4.1.1(ii) Allosteric modification of the substrate Although less well-studied, a mechanism for making a substrate susceptible for ubiquitination is allosteric modification by ligands. A physiological example is that of degradation of (R) subunits of PKA. R subunits are substrates for ubiquitination and degradation by the proteasome.[69] Degradation of R subunits leads to persistent activation of PKA without persistence in cAMP elevation and bridges the short-term action of neurotransmitter 5-HT to gene expression.[70,71] R subunit has two cAMP-binding sites. Without cAMP binding, R subunits are resistant to ubiquitination. Mutation studies have revealed that for ubiquitination of R subunits, binding of cAMP to both sites is essential. For example, R subunit mutants that bind cAMP to only one site are not efficiently degraded.[70]

5.21.4.1.2 Modulating the activity of ubiquitin ligases

Ubiquitin ligases largely control the substrate specificity of ubiquitin-conjugation reaction. The temporal specificity of ubiquitin conjugation to substrates by these enzymes is provided by regulation of the ligase activity. Activity of ubiquitin ligases can be modulated by posttranslational modification such as phosphorylation and by allosteric modification of the enzyme, or by attachment to UbL proteins.

5.21.4.1.2(i) Modulation of ubiquitin ligases by phosphorylation Regulation of ubiquitin ligase activity by phosphorylation has been shown in the studies on a multisubunit ligase, APC. As the name implies, this complex is critical for cell cycle progression into anaphase. A recent study shows that a form of APC is also expressed in postmitotic neurons.[72]

APC can be activated by Cdc2 kinase.[73] Cdc2 kinase appears to exert its effect by activating another protein kinase called polo-like kinase. In *Xenopus* and humans, phosphorylation of four different APC subunits, APC1, CDC16, CDC23, and CDC27, has been shown to be increased during mitosis. In neurons, APC might have a role in ubiquitinating different substrates from the ones ubiquitinated during cell cycle progression. An observation that lends credence to this notion is that levels of polo-like kinases Fnk and Snk dramatically increase with stimuli that produce LTP and other forms of synaptic plasticity.[74]

Activity of at least one single subunit RING-finger ubiquitin ligase, c-Cbl, is known to be regulated by phosphorylation. c-Cbl ubiquitinates the epidermal growth factor receptor (EGFR). Tyrosine phosphorylation of c-Cbl at a site next to the RING-finger domain stimulates the ligase to ubiquitinate the EGF receptor.[75,76]

Phosphorylation of ubiquitin ligases could have inhibitory effect as well. In the fission yeast *Schizosaccharomyces pombe*, PKA blocks APC activity. Moreover, the inhibitory effect of PKA seems to be dominant over the stimulatory effect of the polo-like kinase. Even if APC has been activated by polo-like kinase, addition of mammalian PKA to APC-containing fractions inhibits ubiquitination of the substrate cyclin B.[77,78]

Phosphorylation of ubiquitin ligases is also regulated by phosphatases. For example, type I protein phosphatases (PP1) are necessary for progression into anaphase. In addition, in *S. pombe*, mutations in dis2+ a gene that encodes a catalytic subunit of PP1 have deleterious effect. Although the exact mechanism of phosphatase action in promoting activity of APC is not clear, perhaps they act by counteracting the negatively regulating protein kinases such as PKA.[79,80]

5.21.4.1.2(ii) Allosteric modification of ubiquitin ligases In addition to phosphorylation, ubiquitin ligase activity can be stimulated by allosteric activation. An ubiquitin ligase called Ubr1 targets transcription factor Cup9, which is a negative regulator of di-/tripeptide transporter Ptr1 gene. Ubr1 has three sites at which it can

bind other molecules. It is believed that site III on Ubr1 binds to the substrate Cup9. Peptides that bind to site I or II can allosterically stimulate ligase activity of Ubr1 toward Cup9.[81]

5.21.4.1.2(iii) Modulation of ubiquitin ligase activity by attachment of UbL proteins Activity of the ligases is also modified by posttranslational modification by covalent linkage of UbL proteins. Linkage of an UbL protein to an E3 ubiquitin ligase appears to modulate the activity of the ligase. For example, an UbL protein called Rub1 (related to ubiquitin 1) is conjugated to proteins of the cullin family, which are part of a multisubunit ubiquitin ligase called Skp1-cullin-F-box protein complex (SCF). Conjugation to Rub1 is required for maximal activity of the SCF ligase14.[82]

5.21.4.1.3 Removal of ubiquitins

Ubiquitin–proteasome-mediated degradation can be regulated by removal of ubiquitin. When a protein is polyubiquitinated, it is targeted to the proteasome for degradation unless the ubiquitin chains are removed by the action of DUBs. Deubiquitination by DUBs serves two purposes: (1) reversing ubiquitination of a protein or (2) disassembling the polyubiquitin chains before the ubiquitinated proteins are channeled to the 26S proteasome. Disassembly of polyubiquitin chains at the proteasome step is probably a rate-limiting step for degradation. Since the pore of the proteasome catalytic chamber is small (13 Å),[18] the polyubiquitin tag needs to be removed before the substrate is fed into the catalytic core. Otherwise, the catalytic chamber tends to be 'clogged', thereby reducing the rate of degradation.[83]

5.21.5 Combinatorial Coding of Specificity in Ubiquitin Conjugation

Thus far, the evidence available indicates that the conjugation of ubiquitin to a substrate is a highly specific reaction. How is this specificity achieved? As explained above, ubiquitin conjugation requires three enzymes, E1, E2, and E3. E1 is common to all ubiquitination reactions because this enzyme activates ubiquitin. There is some degree of specificity at the E2 step. The E3s are the most specific to a given substrate, however. Initially it was thought that there is a specific E3 for each substrate. This situation would be untenable because of the coding burden it places on the genome. Rather, the specificity is derived from the combination of recognition modules as shown in **Figure 5**. A given single or multisubunit E3 has a domain specific to a given substrate. In some instances, an E3 ligates ubiquitin to only one substrate. In other cases, an E3 ligates ubiquitin to more than one substrate. When an E3 has activity toward more than one substrate, it is thought that the domain on an E3 that interacts with one substrate is distinct from the domain that an E3 uses to interact with another substrate. It is estimated that the human genome contains about 1000 genes encoding E3s.[84] The estimate for the number of genes coding for E2s is around 25–30. Considering that there are about 30 000 genes in the human genome, E2s and E3s together potentially could generate a unique combination for every gene! Besides the unique E2–E3 combinations, specificity can be generated by the state of the substrate (vulnerable or resistant to degradation) as well as regulation of E3s in many ways as described above. Thus, the ubiquitin-conjugation machinery can be highly specific to a given substrate.

5.21.5.1 Spatial and Temporal Specificity of Ubiquitin–Proteasome-Mediated Degradation

For a given substrate to be degraded, E1, E2, and E3s must be available in the vicinity of the substrate. If the availability of E2s or E3 is spatially restricted, degradation of a given substrate would occur in one subcellular compartment where all the components are available and not in others where one or more components might be missing. Moreover, spatial restriction of E2s and E3s can be conditional. For example, sequestering an E2 or E3 would render the enzyme temporarily unavailable in a subcellular location. Substrates or the factors such as protein kinases that regulate E3s might be sequestered as well.

Since the enzyme that is most specific in the ubiquitin-conjugation reaction, E3, is also highly regulated, temporal regulation of ubiquitination is easily achieved. For example, if a given E3 is activated by a protein kinase, activation of the E3 is tied to the stimulation of the kinase. For example, phosphorylation activates the APC ubiquitin ligase.[49,85] An interesting example is that of serum-inducible kinase (SNK). SNK stimulates

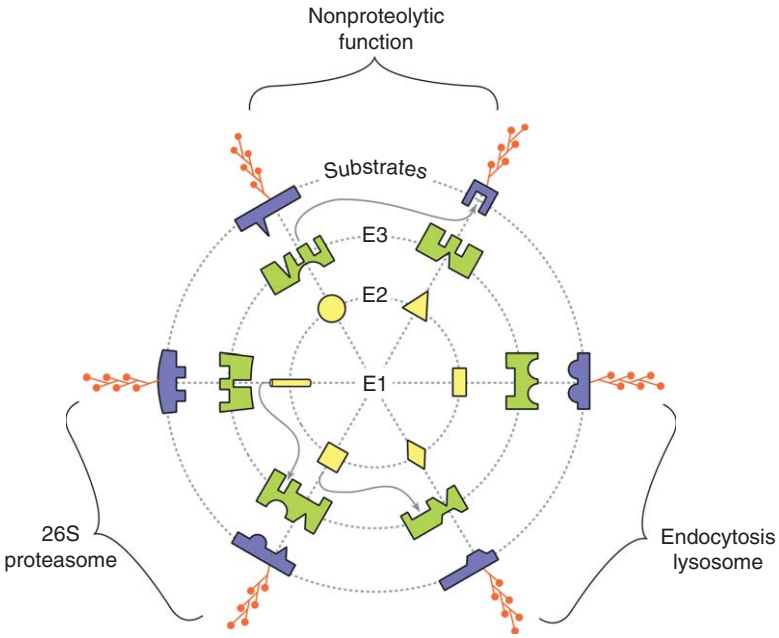


Figure 5 Combinatorial coding of specificity in ubiquitin conjugation. E1 (center) is common to all ubiquitination reactions and has no substrate specificity. E2s preferentially interact with some E3s but not others. An E3, which may be a single molecule or a complex of molecules is believed to be specific for each substrate. E2s, E3s, and substrates generate a large number of combinations to 'code' for the specificity of ubiquitination reaction. The shapes on each concentric circle (from inside to the outside) represent specific E2s, E3s, and substrates. The notches or projection in each shape represent specific domains in the enzymes or substrates. Radiating dotted lines indicate shapes that fit into each other indicating specific interactions. Occasionally, an E2 can interact with more than one E3 and a given E3 can ubiquitinate more than one substrate (wavy arrows between circles). These interactions are specific because they occur through different recognition domains in these molecules. The ubiquitinated proteins can undergo degradation by the proteasome or endocytosis or could have a nonproteolytic role.

ubiquitin–proteasome-mediated degradation of a protein called spine-associated Rap guanosine triphosphatase (GTPase) activating protein (SPAR), which is critical for controlling spine shape in neurons. Apparently, SNK itself is short-lived and is degraded by the proteasome. SNK protein is produced in dendritic spines by translation of mRNA delivered from the cell body. Therefore, instability of SNK provides a temporal control of SPAR degradation.[86] In the nervous system, protein kinases are often stimulated by generation of second messengers which in turn are linked to the action of neurotransmitters or arrival of an action potential. E3 might also directly sense second messengers such as Ca^{2+}. For example, the ubiquitin-ligase neuronal precursor cell-expressed developmentally downregulated 4 (Nedd4) that conjugates ubiquitin to epithelial and cardiac Na^{+} channels has a C2 domain, which might respond to an increase in Ca^{2+}.[87–89] Response to transient elevation in Ca^{2+} by molecules critical to synaptic plasticity has been well documented.[90,91] Temporal specificity of ubiquitination can also be achieved by controlling the vulnerability of the substrate to ubiquitination by posttranslational modification such as phosphorylation. Phosphorylation makes jun and C/EBP resistant to ubiquitin–proteasome-mediated degradation.[16,64,65]

5.21.6 The Proteasome

The term proteasome is used to describe two kinds of multisubunit proteolytic complexes, the 26S and 20S, based on their sedimentation coefficient. The 26S proteasome degrades ubiquitinated protein substrates. The 26S complex contains the 20S as a core and regulatory 'caps' on either end like a dumb bell. Each cap of the 26S proteasome is known as the 19S regulatory complex (19S RC). The 20S core is a cylindrical structure consisting of the catalytic part of the proteasome.[92,93]

5.21.6.1 The Catalytic 20S Core

Our knowledge of the proteasome organization comes from the studies on the crystal structure of proteasome from the archaebacterium *Thermoplasma acidophilum,* and the yeast *S. cerevisiae*. It appears that the proteasome is more ancient than ubiquitin because archaebacteria have the proteasome but not ubiquitin. *Thermoplasma acidophilum* has two genes encoding α- and β-subunits. The subunits are arranged in four stacked rings to form the catalytic cylinder with the two middle rings consisting of β-subunits, which are sandwiched between two rings of α subunits. In the *T. acidophilum* proteasome, both the α- and β-subunits are present in seven copies each and assembled in a symmetrical fashion $\alpha_7\beta_7\beta_7\alpha_7$.[92,94] This general structure is preserved in eukaryotes but the α- and β-subunits have diverged into seven different subunits each. In yeast the 20S core is made up of two outer rings with seven α-subunits ($\alpha1$–$\alpha7$) in each ring and two inner rings consisting of seven β-subunits each ($\beta1$–$\beta7$) (**Figure 6(a)**).

The catalytic core of the proteasome is a threonine protease. Based on the crystal structure of the proteasome, it was concluded that proteasome functions through a new kind of proteolytic mechanism. In

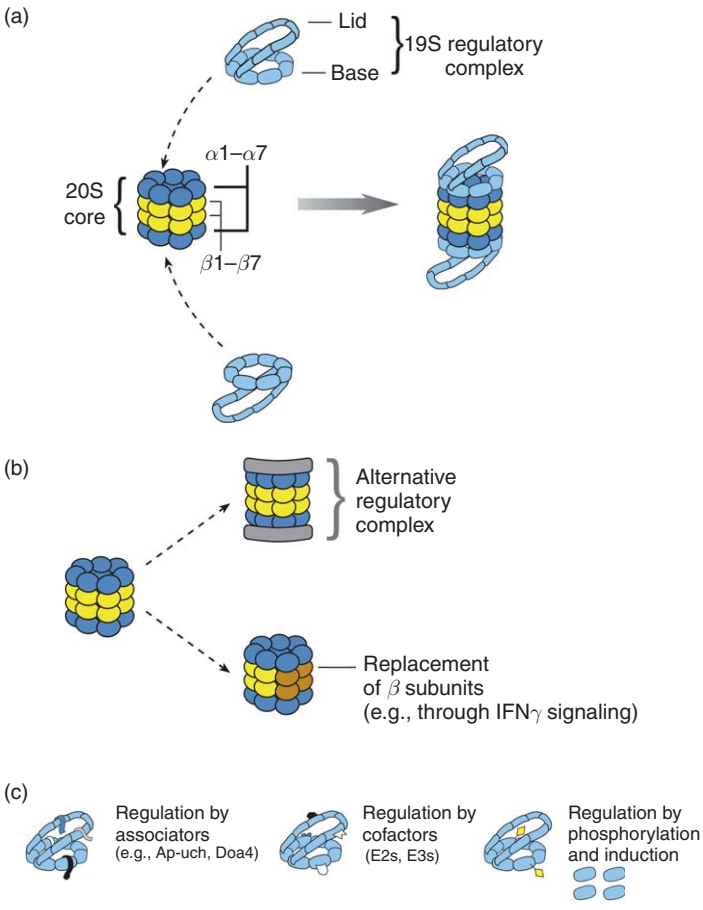


Figure 6 Proteasome assembly and regulation. (a) Assembly of the proteasome. The proteasome consists of the 20S core and the 19S cap. The 20S proteasome contains the catalytic sites and comprises four stacked rings. The inner two rings have seven β-subunits ($\beta1$–$\beta7$) and the outer rings have seven α-subunits ($\alpha1$–$\alpha7$). The 19S regulatory cap is made of base and a lid each with eight subunits. (b) Regulation by alternative regulatory complexes or substitution of core subunits. Proteasome can be regulated by attachment of alternative regulatory complexes (such as PA28, PA26, or 11S; shown as flat attachment to the core subunits) instead of the usual 19S regulatory complex. Activity of the 20S core is regulated by substitution of different subunits. For example, through IFNγ signaling three β-subunits ($\beta1$, $\beta2$, and $\beta5$) are replaced (two of which are shown in the figure). (c) Regulation of the 19S cap. 19S regulatory complex can be regulated by proteins that associate with it (e.g., Ap-uch, Doa4), proteins that act as cofactors (E2s and E3s) or induction or phosphorylation of the subunits of the complex themselves.

this mechanism, the active site nucleophile is the hydroxyl group on the threonine residue at the N-terminus of the β-subunit.[92] This catalytic mechanism seems to be conserved across evolution. The antibiotic lactacystin and the active form β-lactone, which are specific and irreversible inhibitors of the proteasome bind to the N-terminal threonine residue in the β-subunit of the mammalian proteasome.[95]

The catalytic core is the part that ultimately cleaves the ubiquitinated protein into small peptides anywhere from 3 amino acids to 32 amino acids long. The peptides generated are probably further hydrolyzed to generate free amino acids by other proteases and amino peptidases. The proteases that break down the products of the proteasome action include but not limited to Thimet oligo peptidase,[96] tricorn protease,[97] and tripeptidyl protease II.[98]

The 20S proteasome can exist not only as a core of 26S, but also as a separate population that cannot degrade ubiquitinated proteins.[99] However, the 20S proteasome by itself has chymotrypsin-like, trypsin-like, and postglutamyl peptidase activities which cleave after hydrophobic, basic, and acidic residues, respectively. The peptide-hydrolyzing activity of the 20S proteasome can be modulated by an 11S regulatory cap.[100]

5.21.6.2 19S Regulatory Complex

The 19S RC recognizes the polyubiquitinated substrate and channels the substrate into the catalytic 20S core of the proteasome. It also has the capacity to regulate the activity of the catalytic core and determine the nature of the degradation process. Usually one 19S RC is attached to either end of the catalytic core. The subunits of the 19S RC are highly conserved across evolution.[101–105] Two subcomplexes can be recognized within the 19S RC called the base and the lid (**Figure 6(a)**).

5.21.6.2.1 The base of the proteasome

The base consists of six ATPase subunits (Rpt1–Rpt6) and two non-ATPase subunits (Rpn1 and Rpn2). The Rpt subunits are highly conserved through evolution (up to 75% identical between yeast and humans). The ATPase subunits have a domain called the ATPases associated with different cellular activities (AAA) domain in the center. These subunits are homologous to each other with highest degree of homology in the AAA domain. Some of the Rpt subunits contact the α-ring in the catalytic core of the proteasome and are believed to channel the substrate into the catalytic chamber for degradation.[106–110]

5.21.6.2.2 The lid of the proteasome

The lid comprises eight subunits arranged in the manner of a ring (**Figure 6(a)**) that can attach itself to the base of the 19S RC or detach as a complex from the base. All the subunits in the lid are non-ATPase subunits. The exact function of the lid subunits is not known. Degradation of polyubiquitinated proteins requires lid attachment indicating that lid performs an essential function in ubiquitin–proteasome-mediated degradation.[111] Also, in archaebacteria that lack ubiquitin, the proteasome is devoid of the lid.[112] The lid subunits share a characteristic sequence of 200 amino acids called the proteasome, COP9, initiation factor 3 (PCI) domain.[113] COP9 signalosome is a huge complex of proteins originally discovered in the context of photomorphogenesis in plants. Translation initiation factor 3 is a protein complex critical for translation of mRNAs. In addition, the subunits of the lid have a 120 amino acid-long Mpr1p/Pad1p N-terminus (MPN) domain that is important for the structure of Rpn8 and Rpn11 subunits.[16] In addition, a subset of MPN domain-containing proteins has a motif of five polar residues called the MPN+ motif. The MPN+ motif has been shown to be critical for the function of Rpn11.[114]

5.21.7 Regulation of the Proteasome

Proteasome activity can be regulated in two main ways. One is regulation by cofactors or proteins that are loosely associated with it and by induction or phosphorylation of the subunits especially those of the 19S RC. The proteasome can also be regulated by a change in the composition of the intrinsic subunits (**Figure 6(c)**).

5.21.7.1 Regulation by Cofactors and Loosely Associated Factors

In addition to the intrinsic subunits, proteins that interact with the proteasome complex regulate its activity.[115] Often the cofactors are components of the ubiquitin pathway. Proteins such as chaperones and heat-shock proteins also assist in proteasome-mediated degradation of proteins.

Both E2s and E3s have been found to interact with the proteasome. For example, UBCs (E2s) Ubc1, Ubc2, and Ubc4 coimmunoprecipitate with the proteasome. E3s such as Ubr1 and Ufd4 have been shown to physically interact with subunits of the 19S regulatory complex. APC and SCF ubiquitin ligases copurify with the 19S regulatory complex. An ubiquitin ligase called Hul5 is associated with the proteasome as well.[116] The cofactors interact with the proteasome directly or through other proteins such as Cic1 and hPLICs that recruit the cofactors to the proteasome. For example, hPLIC-1 and hPLIC-2, the human counterparts of yeast Dsk2 interact with the proteasome as well as specific E3s E6-AP and β-TrCP. Two other proteins, Rad23 and BAG1 are also known to interact with the proteasome.

The DUBs interact with the proteasome as well. A DUB called USP14 is known to interact with the regulatory complex of the proteasome.[117] Genetic and biochemical experiments carried out using *S. cerevisiae* showed that the yeast orthologue of USP14, Ubp6 binds to the proteasome. Since binding to the proteasome increases the activity of Ubp6 by 300-fold and Ubp6 is thought to assist in the removal of polyubiquitin chain just before the substrate is degraded by recycling ubiquitin for further use.[116]

Other DUBs have also been reported to associate with the proteasome. A yeast DUB that is related to the human tre-2 oncogene associates with the proteasome and assists in the disassembly of the polyubiquitin chain.[118] The *Aplysia* homologue of UCH-L1 (Ap-uch) associates with the proteasome and improves proteolytic activity. In *in vitro* experiments, it was shown that Ap-uch cleaves the first ubiquitin attached to the substrate.[71]

Among the factors that interact with the proteasome are heat-shock proteins and chaperones. It has been known for a long time that the ubiquitin–proteasome pathway degraded misfolded or unfolded proteins. The general belief is that the heat-shock proteins refold the misfiled proteins into the right conformation. Misfolded proteins not rescued by the heat-shock proteins are thought to be substrates for the ubiquitin–proteasome pathway. It was not clear, however, which ubiquitin ligase (E3) specifically recognized misfolded proteins. Recently, a chaperone called BAG1 has been shown to interact with both the proteasome and the heat-shock protein Hsp70.[119] Also, a protein called CHIP is known to ubiquitinate unfolded proteins.[120] CHIP contains a protein sequence motif called the U-box[52] and a RING-finger domain. CHIP can interact with either Hsp70 or Hsp90, and together with either of the heat-shock proteins can ligate ubiquitin to misfolded proteins.[51,121] The cooperation between heat-shock proteins and the proteasome is also required to degrade aberrant membrane proteins. For example, the cystic fibrosis transmembrane regulator (CFTR) is a membrane protein that is often misfolded. CHIP and Hsc70 together recognize the CFTR protein and target it to proteasome-mediated degradation.[122]

Proteasome activity can also be regulated by substitution of subunits of the 20S core with subunits such as the ones induced by interferon-γ (IFNγ). In addition, alternative regulatory complexes with different composition than the 19S RC can be attached to the 20S core, which alters the substrate specificity and activity of the proteasome (**Figure 6(b)**).[16]

5.21.7.1.1 Ubiquitin-like domain of proteasome-interacting proteins contains a characteristic motif

A characteristic feature of several proteasome-interacting proteins is that they possess an N-terminal UbL domain. The UbL domain, as the name implies, has a high degree of homology to ubiquitin. Although studying parkin, which has an N-terminal UbL domain, Upadhya and Hegde[41] investigated whether the UbL domain of the proteins known to interact with the proteasome have any common features. They found that parkin and a subset of proteins with the UbL domain contain a 5-amino acid motif within the UbL domain. They termed this motif the 'proteasome-interacting motif (PIM)'. Several of these proteins such as Rad23, Dsk2, and Ubp6 have been shown to interact with the proteasomem,[16] although it remains to be experimentally determined how many of the PIM residues are absolutely essential for interaction with the proteasome. The PIM is probably predictive of a protein's ability to associate with the proteasome. The prediction made by the discovery of this motif has been experimentally confirmed for parkin, which is an ubiquitin ligase. Sakata *et al.*[42] found that parkin indeed associates with the Rpn10 subunit of the proteasome. Interestingly, Upadhya and Hegde[41] found

that the PIM is present in several transcription factors. Since the proteasome is known to be present in the nucleus and recent studies indicate that the proteasome might have a role in transcription, it is possible that PIM helps recruit cofactors to assist the proteasome in its nuclear functions.

5.21.7.2 Regulation of the Proteasome by Induction and Phosphorylation of Subunits and Subcellular Distribution

The subunits of the 26S proteasome are not fixed but change in response to the physiological condition of the cell. The capacity of the 26S proteasome to degrade ubiquitinated proteins can be regulated by changes in (1) the total amount of the proteasome, (2) its subunit composition and amount of each subunit, and (3) its subcellular distribution. The changes in the total amount of 26S can be brought about by the extent of 19S cap binding to the 20S core. This could occur by an increase in the amount of 19S as well as by increased association of the existing 19S with the 20S core.[100] During metamorphosis of *Manduca*, flight muscles develop and intersegmental muscles are destroyed. The destruction of intersegmental muscles is brought about by an increase in ubiquitin-dependent proteolysis as a result of extensive hormone-dependent reprogramming of the 19S regulatory complex. It has been shown that the multiubiquitin binding subunit (MBP/S5a) and the ATPases MSS1 (S7) and S4 are induced significantly during this period.[123,124] Increase in the amount of 26S by enhanced binding of 19S to the 20S without an increase in the total amount of 20S occurs during the metaphase–anaphase transition in the meiotic cell cycle.[125,126] In addition to the regulation of total proteasome content, the activity of the proteasome also appears to be regulated by alterations in its subcellular distribution. For example, during ascidian (marine animals commonly called sea squirts) embryonic development, the distribution of the proteasome changes in a cell cycle-dependent manner. The proteasome localized in the nucleus during interphase disappears from the nucleus during prophase, and in telophase the proteasome is again localized in the newly formed nucleus.[127]

Proteasome activity can also be regulated by phosphorylation. For example, Yang *et al.*[128] have shown that phosphorylation events are necessary for the assembly of the 26S proteasome. Several subunits including MSS1, S4, S6, and S12 of the 19S RC have been shown to be phosphorylated.[129] Recently it has been demonstrated that assembly of the proteasome requires phosphorylation of Rpt6, an ATPase subunit.[130]

5.21.8 COP9 Signalosome

The COP9 signalosome (CSN) is a multisubunit complex with close similarity to the lid of the 26S proteasome. The CSN was discovered in the context of plant development regulated by light. Mutation of the locus constitutive photomorphogenesis 9 (cop9) was found to confer phenotype light-grown phenotype on *Arabidopsis* plants grown in dark. Biochemical purification of the CSN revealed that it contained multiple proteins.[131,132] According to the standardized nomenclature, the CSN subunits are named CSN1 through CSN8[133] (**Table 1**). The CSN subunits are evolutionarily conserved from the fission yeast *S. pombe* to humans. In the budding yeast, *S. cerevisiae*, an orthologue for only CSN5 was

Table 1 CSN subunits and proteins interacting with them

CSN subunit	*Related proteasome subunit*	*Interacting proteins*	*Reference(s)*
CSN1	Rpn7	Inositol 1,3,4-trisphosphate 5/6 kinase, eIF3c, Rpn6 proteasome subunit	134–136
CSN2	Rpn6	Cullin 1, cullin 2, thyroid hormone receptor	137–139
CSN3	Rpn3	IKKγ, eIF3c	140–142
CNS4	Rpn5	COP10 (ubiquitin-conjugating enzyme)	142
CSN5	Rpn11	c-jun, Uch-L1 cyclin-dependent kinase inhibitor	143–145
CSN6	Rpn8	Rbx/Roc1/Hrt1, eiF3e	141, 146, 147
CNS7	Rpn9	eIF3e	141, 148
CSN8	Rpn12	eIF3c, COP10 (ubiquitin-conjugating enzyme)	135, 142

initially identified. With the advent of genomics, a CSN complex was identified in the budding yeast, and the subunits specific to this organism were designated CSN9, CSN10, and CSN11.

5.21.8.1 Similarities of CSN and the Proteasome Subunits

Several subunits of CSN share close similarities with the proteasome lid subunits. Bioinformatic analysis showed that CSN subunits share homology with proteasome lid subunits and subunits of the eukaryotic translation initiation complex eIF3.[111] CSN1, CSN2, CSN3, CSN4, CSN7a, CSN7b, and CSN8 contain a proteasome, COP7, eIF3 (PCI) domain. CSN5 and CSN6 contain an Mpr1p/Pad1p N-terminus (MPN) domain. The PCI and the MPN domains are also present in some of the proteasome lid subunits and in components of eIF3.[113,149] The CSN and the proteasome subunits share greater degree of similarity with each other than similarity to the eIF3 subunits.

5.21.8.2 CSN and Ubiquitin Ligase Function

The connection between CSN and ubiquitin ligase became evident with the discovery that a subunit of the SCF ligases cullin 1 interacts with CSN. This interaction appears to be evolutionarily conserved. Cullin 1 from yeast, plants, and humans was found to interact with CSN. In *Arabidopsis*, at least CSN–cullin 1 interaction seems to positively regulate ubiquitin-mediated degradation.[150] It was found that degradation of a substrate called PsIAA6 by a specific ligase SCF^{TIR1} required intact CSN function. These findings have made it clear that CSN is essential for proper functioning of many ligases containing cullin.[151] For example, the SCF ligase that degrades the cell cycle regulatory protein p27Kip1 requires CNS for its function. Other cullin-containing ligases such as VHL-elonginB/C-CUL2-Rbx1 ubiquitin ligase complex and BTB (Broad-complex/Tramtrack/Bric-a-brac)-POZ (poxvirus and zinc finger) domain-containing ubiquitin ligases also interact with CSN.

5.21.8.3 CSN and Removal of Nedd8 Moiety

Another function of CSN is cleavage of Nedd8 attached to cullin in E3 ligases. Evidence gathered thus far indicates that all cullins are modified by attachment to the ubiquitin-like protein Nedd8 (also called Rub1). In CSN mutants, neddylated cullins accumulate. Purified CSN can deneddylate cullins suggesting that CSN has an isopeptidase activity.[137] The subunit with the isopeptidase activity was found to be CSN5. It was known that CSN5 contains a motif for metalloprotease activity called JAMM (JAB1, MPN, MOV34) in its MN domain.[152] Mutations in the JAMM domain abolished the deneddylation activity and the ability to rescue the CSN5 loss of function mutants in fission and budding yeast and *Drosophila*. CSN5 by itself is unable to catalyze deneddylation of cullin and therefore it has been suggested that the isopeptidase activity of CSN5 requires association with other subunits in the CSN complex.

5.21.8.4 Association of Protein Kinases and DUBs with CSN

Around the same time as bioinformatics revealed similarities between CSN, proteasome, and eIF3, Dubiel and coworkers, while trying to identify new components of the 26S proteasome, identified a novel protein complex that possessed protein kinase activity.[153] They found that the new protein complex was capable of phosphorylating c-jun and $I\kappa KB\alpha$ and the precursor of $NF\kappa$-B was called p105. Other protein kinases are found to interact with CSN as well. For example, inositol 1,3,4,-trisphosphate 5/6 kinase, casein kinase 2, and protein kinase D associate with CSN.[154]

5.21.9 Deubiquitinating Enzymes

Ubiquitination reaction is reversible like phosphorylation until the ubiquitinated protein is committed to degradation by the proteasome. The reversibility is less clear with respect to endocytotic degradation, that is, internalization of plasma membrane proteins through endocytosis and their degradation through the lysosome.

Based on the knowledge of the endocytotic pathway, it is reasonable to assume that ubiquitination is reversible until the endocytosed membrane proteins such as neurotransmitter receptors are routed to the multivesicular body for lysosomal degradation.

Ubiquitin is removed from substrates by enzymes called DUBs. Based on protein sequence and the molecular size, DUBS can be classified into two general classes: (1) low-molecular-weight (20–30 K) ubiquitin C-terminal hydrolases (UCHs) and (2) high-molecular-weight (~100 K) ubiquitin-specific proteases (UBPs, also called USPs).[155,156] There are numerous DUBs in almost every eukaryotic organism studied. Among the DUBs, UBPs belong to a large family containing diverse genes, whereas the UCH family has fewer genes. For example, in yeast *S. cerevisiae* there are 27 UBPs and 1 UCH.[157] In the human genome, there are 63 genes encoding UBPs and 4 genes that code for UCHs.[158] UCHs and UBPs subserve different functions in the eukaryotic cell. Although the current name for these enzymes, that is, DUBs, emphasizes the removal of ubiquitin from substrates, some DUBs especially UCHs function to process linearly linked ubiquitin precursors and generate monoubiquitin.

UCHs are cysteine proteases in that the critical residue in the catalytic site is a cysteine. In addition, histidine and aspartate residues are critical for catalytic activity. All UCHs contain these residues even if they do not share a high degree of homology elsewhere in the sequence. For example, the *Aplysia* UCH (Ap-uch) critical for the induction of long-term facilitation has only 39% homology to its human counterpart UCH-L1.[71] Ap-uch and UCH-L1 both contain the catalytic cysteine, histidine, and aspartate residues at similar positions in the molecule. UCHs cleave small peptide chains linked to the C-terminus of ubiquitin. UBPs can cleave the isopeptide bond between ubiquitins in a polyubiquitin chain and the isopeptide bond between the ubiquitin and the substrate.[156]

DUBs are important for generating free ubiquitin at various steps of the ubiquitin–proteasome pathway. Ubiquitin is encoded by the tandemly linked polyubiquitin gene. In the cell, ubiquitin is always present as free monoubiquitin. Therefore, it is inferred that polyubiquitin is processed by DUBs to generate monoubiquitin. In addition to the polyubiquitin gene, ubiquitin is also encoded by fusion of two ribosomal subunits called L40 and S27. These gene products are believed to be processed by DUBs as well. Cleavage of isopeptide bond in the ubiquitin chains linked through Lys48 of ubiquitin serves two purposes. One is to recycle the ubiquitin after it has been used for marking a substrate for ubiquitination.[159] Another function is to 'edit' the errors made by the UBCs and reverse the ubiquitination reaction so that the substrate is no longer degraded (**Figure 7**).[156] In addition, editing function of DUBs probably serves to reverse the monoubiquitin attachment that marks membrane proteins for endocytosis.

5.21.9.1 Substrate Specificity of DUBs

Since numerous DUBs are present in eukaryotic organisms (**Table 2**), it is probable that they possess substrate specificity. UCHs have been studied in some detail with respect to their substrate specificity. Two major UCHs in mammals are UCH-L1 and UCH-L3. Larsen *et al.*[180] showed that UCH-L1 cleaves linear polyubiquitin molecules more efficiently than UCH-L3. In contrast, UCH-L3 appears to prefer ubiquitin fused to small ribosomal proteins (see **Figure 7**). The tissue distribution of the two UCHs is indicative of their functional specialization as well. UCH-L1 is a neuronal-specific enzyme whereas UCH-L3 is expressed primarily in hematopoietic tissues. Another UCH called UCH-L2 has wide tissue distribution.[181]

Substrate specificity of UBPs has been studied mainly with respect to the type of ubiquitin linkage cleaved. For example, most UBPs cleave isopeptide linkages such as those in a polyubiquitin chain attached to the substrate. An UBP called isopeptidase T cleaves ubiquitin that is closest to the substrate protein. Since ubiquitin is linked to the substrate through its C-terminus and isopeptidase T requires a free C-terminus of ubiquitin for its action, it has been inferred that isopeptidase T acts only after the degradation of the protein substrate by the proteasome and removal of the peptide remnant attached to the first ubiquitin.[182] Thus, it is probable that isopeptidase T functions to disassemble polyubiquitin chains after the degradation of the polyubiquitinated substrate by the proteasome. An UBP in the PA700 regulatory complex of the proteasome also cleaves isopeptide linkages of ubiquitin. In sharp contrast to isopeptidase T, the PA700-associated UBP cleaves the ubiquitin that is farthest from the substrate.[183]

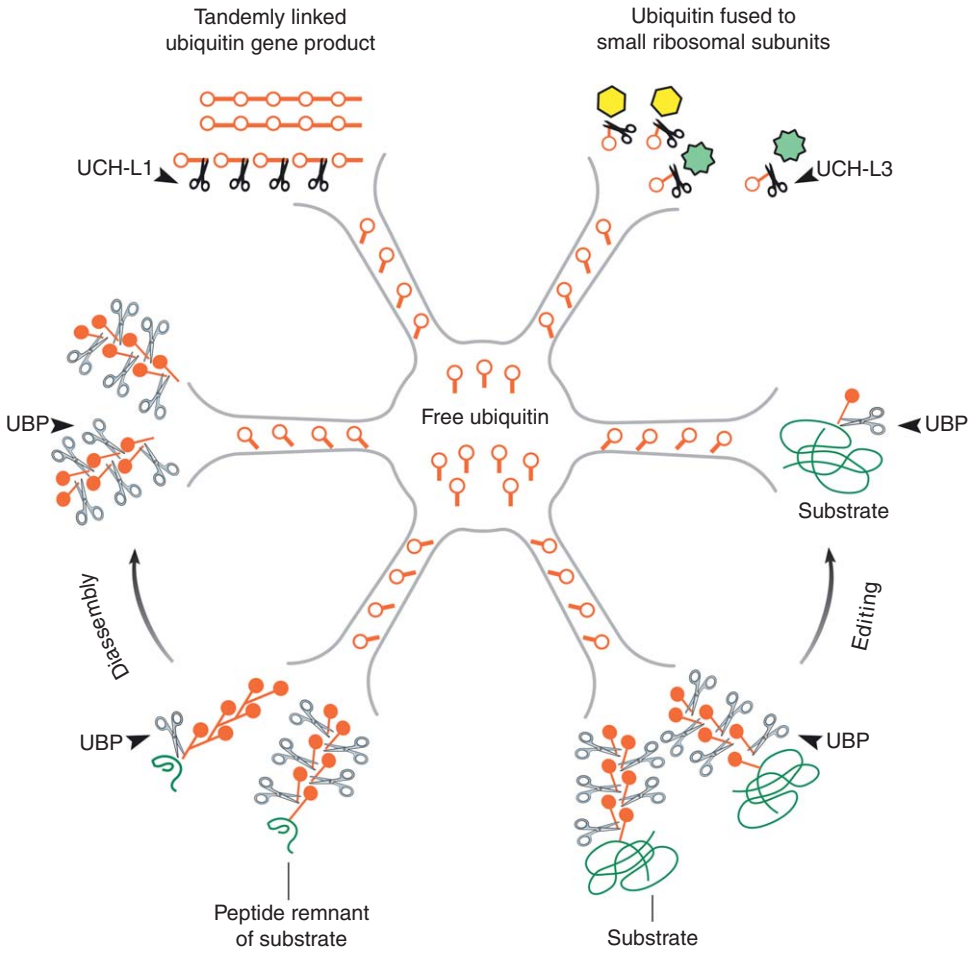


Figure 7 Multiple roles of the deubiquitinating enzymes. Deubiquitinating enzymes (DUBs) of the UCH type (dark scissors) process ubiquitin precursors. UCH-L1 generates monoubiquitins from tandemly linked ubiquitin gene product. UCH-L3 acts on ubiquitin synthesized as a protein fused to small ribosomal subunits. DUBs of the UBP type (shaded scissors) process ubiquitins linked in isopeptide linkage in polyubiquitin chains. DUBs also reverse the ubiquitination on erroneously targeted substrates (editing). Another important function of DUBs is disassembly of polyubiquitin chains as the ubiquitinated substrate is degraded. Ubiquitin attached to substrates after activation are indicated as lollipop-like structures with filled circles. Free ubiquitin or ubiquitin unit in precursor is shown with open circles.

Substrate specificity of UBPs has been studied mainly with respect to the type of ubiquitin linkage cleaved. For example, most UBPs cleave isopeptide linkages such as those in a polyubiquitin chain attached to the substrate. An UBP called isopeptidase T cleaves ubiquitin that is closest to the substrate protein. Since ubiquitin is linked to the substrate through its C-terminus and isopeptidase T requires a free C-terminus of ubiquitin for its action, it has been inferred that isopeptidase T acts only after the degradation of the protein substrate by the proteasome and removal of the peptide remnant attached to the first ubiquitin.[182] Thus, it is probable that isopeptidase T functions to disassemble polyubiquitin chains after the degradation of the polyubiquitinated substrate by the proteasome. An UBP in the PA700 regulatory complex of the proteasome also cleaves isopeptide linkages of ubiquitin. In sharp contrast to isopeptidase T, the PA700-associated UBP cleaves the ubiquitin that is farthest from the substrate.[183]

The studies on specificity of UBPs with respect to isopeptide linkages of ubiquitin do not address the physiological substrate specificity, however. Although diverse substrates are polyubiquitinated, the nature of the isopeptide linkages of ubiquitin is the same in all substrates. Therefore, physiological specificity must arise from noncatalytic parts of the UBP molecules that interact with the substrates. A recent study shows that two

Table 2 Classes of DUBs, their substrates, and roles

Enzyme	*Substrate*	*Description/Function*	*Reference(s)*
UCHs			
UCH-L1	Linear polyubiquitin	Neuron-specific	160
		Important for synaptic plasticity	71
UCH-L3	Ubiquitin fused to small ribosomal proteins	Regulates membrane recycling of epithelial sodium channel	161
UBPs			
CYLD	NEMO, TRAF2/6	Functions in NF-B signaling cylindromatosis tumor suppressor	162–165
USP2	Fatty acid synthase	Functions in androgen signaling	166, 167
USP7	Mdm2, p53, histone H2B	Regulates p53 levels, also called HAUSP	168–170
USP8	NRDP1	Regulates endocytosis	171
USP11	BRCA2	Controls DNA repair	
USP14	Polyubiquitin attached to substrates	Functions in the central nervous system; associates with the proteasome	172, 116
USP18	Not identified	Cleaves ISG-15, dysregulation causes brain cell injury	173, 174
USP21	Not identified	Cleaves Ub and Nedd8	175
Other DUBs			
Ataxin-3		Linked to Machado–Joseph disease	176, 177
A20	RIP	Also has Ub ligase fuction, role in NF-B signaling	178
CSN5	Cullins	Cleaves Nedd8	152, 179

isoforms of UBP in the testis called UBP-t1 and UBP-t2 contain identical catalytic core regions but divergent N-terminal regions. Although the cores alone can cleave peptide (linear) linkages of ubiquitin, the full UBP-t1 and UBP-t2 do not cleave peptide links of ubiquitin efficiently. Moreover, UBP-t1 has higher activity on isopeptide-linked triubiquitin compared to the core alone. On the other hand, UBP-t2 has lower activity on triubiquitin compared to the core.[184] Also, the N-terminal regions of the two isoforms target the UBPs to different subcellular locations.[185] Another example of physiological specificity is that of the cylindromatosis tumor suppressor gene (CYLD). The function of CYLD was unknown until recently. Three groups showed that the CYLD gene encodes a DUB that removes ubiquitin from tumor necrosis factor receptor-associated factor 2 (TRAF2).[162,164,165] TRAF2 can be ubiquitinated through Lys48-linked polyubiquitin chain, which leads to degradation[186] or through Lys63-linked ubiquitin chain which mediates signaling that activates NF-κB. CYLD appears to deubiquitinate TRAF2 with only Lys63-linked polyubiquitin chains.[165] There are DUBs that exhibit preference for Lys48-linked polyubiquitin chains. For instance, USP8 and USP14 cleave Lys48-linked but not Lys63-linked polyubiquitin chains. The newly discovered type of DUBs, such as OUT-domain-containing A20, functions equally well on Lys48-linked and Lys63-linked polyubiquitin chains *in vitro*. By contrast, it deubiquitinates only Lys63-linked substrates *in vivo*.

5.21.9.1.1 Specificity of DUBs with respect to substrates

Another level of specificity of DUBs is with respect to the substrates that are deubiquitinated. Although not much is known about how DUBs choose their targets, domains outside the catalytic core might play a role in determining the substrate to which a given DUB binds. For example, a CAP-Gly domain (a conserved glycine-rich domain found in some cytoskeleton-associated proteins (CAPs)) of CYLD functions in its interaction with NF-κB essential modulator (NEMO), a putative CYLD substrate.

DUBs appear to recognize specific substrates to subserve an editing function, that is, to remove ubiquitins from an erroneously ubiquitinated substrate. This is particularly true of ubiquitin ligases that are auto-ubiquitinated. For example, the enzyme USP8 antagonizes self-ubiquitination of Nrdp1, a RING-finger ligase that regulates the cellular levels of ErbB3 and ErbB4 receptor tyrosine kinases. An interesting example of a DUB regulating an ubiquitin ligase directly is that of Mdm2 a ligase that target p53 for ubiquitination. Mdm2 can self-ubiquitinate and the ubiquitinated Mdm2 is highly unstable. When the expression of a DUB that is known to deubiquitinate and stabilize p53, called herpes virus-associated ubiquitin-specific protease (HAUSP)

is nearly ablated p53 is stabilized as expected. Surprisingly, under these conditions p53 is activated leading to the conclusion that Mdm2 is stabilized in a p53-independent manner. The deubiquitinating action of HAUSP appears to be specific to Mdm2. In HeLa cells, in which p53 is degraded by the E6-AP ligase, p53 is not activated.

Overall specificity of substrate ubiquitination and deubiquitination is also regulated by different types of interaction between E3s and DUBs. For example, von Hippel–Landau protein (pVHL) is an ubiquitin ligase that targets hypoxia-inducible factor 1α (HIF-1α) for degradation. Two DUBs called VDU1 (pVHL-interacting deubiquitinating enzyme 1), and VDU2 have been shown to interact with pVHL. Of these two DUBs only VDU2 can deubiquitinate and stabilize HIF-1α. Another layer of control exists for the ubiquitination of HIF-1α as well. The DUBs, VDU1 and VDU2, are also substrates for the pVHL ubiquitin ligase. Thus by targeting VDU2 for ubiquitination and reducing its cellular amounts, pVHL can ensure HIF-1α degradation.

Another instance of close association of a DUB with E3 ligase is that of Rsp5, a HECT ubiquitin ligase and a DUB called Ubp2. The interaction between Rsp5 and Ubp2 is mediated by a UBA (ubiquitin-associated) domain-containing protein called Rup1. Rsp5 and Ubp2 appear to ubiquitinate and deubiquitinate the same substrates such as the transcription of Spt23.

5.21.10 Ubiquitination and Endocytosis

Recently, a role for ubiquitination in targeting membrane proteins for endocytosis has been discovered. Although ubiquitination of the platelet-derived growth factor receptor (PDGFR) was discovered more than 10 years ago,[187] it is only recently that detailed investigation into the role of ubiquitination in endocytosis has been carried out. One well-studied instance of endocytosis is that of the EGFR. Following ligand binding, EGFR is ubiquitinated. It is generally accepted that an ubiquitin ligase called Cbl ubiquitinates EGFR. In addition to ubiquitinating EGFR, Cbl also ubiquitinates a protein called CIN85 that complexes with another protein, endophilin. The CIN85–endophilin complex is believed to regulate the negative curvature of the membrane required during the early steps of endocytosis.[188]

Another protein called EGFR pathway substrate clone 15 (Eps15) has been shown to be critical for endocytosis as well.[189] Antibodies against Eps15 prevent EGFR internalization.[190] Recently, a family of proteins called Eps15-interacting proteins (epsins) has been identified.[191,192] Epsins are believed to be adaptor proteins that couple the receptor to the clathrin-coated pits. Both Eps15 and epsins contain ubiquitin-interacting motifs (UIMs);[193] which have been shown to be critical for monoubiquitination of these proteins.[194,195] The principles learned by studying EGFR might be generally applicable to other receptors such as the neurotransmitter receptors (**Figure 8**).

Unlike the polyubiquitin chain that marks the substrate for proteasome-mediated degradation, endocytosis appears to be mainly mediated by monoubiquitination. Attachment of single ubiquitin occurs at multiple sites on the receptor proteins in the plasma membrane (see **Figure 3**). Recently, Haglund *et al.*[196] and Mosesson *et al.*[197] obtained unequivocal evidence that the EGFR and PDGFR are monoubiquitinated at multiple sites and this signal is enough to target these receptor tyrosine kinases for endocytosis and degradation through the lysosome. The role of monoubiquitination as a tag for endocytosis of a *S. cerevisiae* G-protein-coupled receptor (GPCR) has been established as well.[198] There are exceptions to the 'monoubiquitination for endocytosis' rule, however. For example, for internalization of general amino acid permease in yeast and the V2-type vasopressin receptor in mammals, polyubiquitination is believed to be the signal.[199,200]

How does monoubiquitination target the membrane proteins for endocytosis and degradation through the lysosome? It is believed that monoubiquitination of receptors facilitates their binding to ubiquitinated epsins, which bind to adaptor proteins that are part of clathrin-coated pits. Ubiquitinated receptor then undergoes endocytosis and is incorporated into the endosomes, which in turn is sequestered into the multivesicular body. The membrane of the multivesicular bodies becomes continuous with lysosomes leading to degradation of the receptor. It has been shown that targeting the internalized receptor to the multivesicular body also requires ubiquitination. Moreover, a 350-kDa complex called endosomal sorting complex required for transport (ESCRT-I) that recognizes the ubiquitinated receptors has been identified. Function of ESCRT-I is essential for sorting the endocytosed receptor into the multivesicular body (**Figure 8**).[201] Two other complexes,

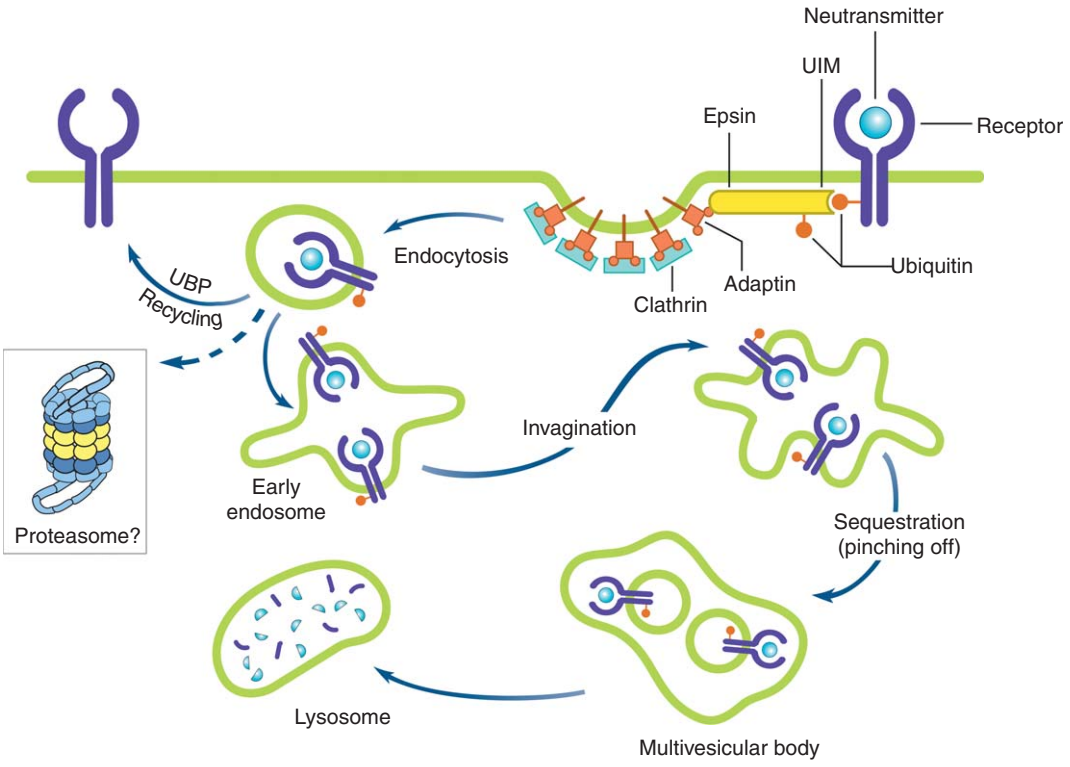


Figure 8 Ubiquitin and endocytosis. Receptors on the plasma membrane undergo monoubiquitination as a result of ligand (e.g., neurotransmitter). Ubiquitinated receptors bind to proteins called epsins, which in turn interact with adaptor proteins (adaptin) bound to clathrin-coated pits. Ubiquitination also functions to sort the internalized membrane protein into early endosomes, which directs them to degradation by lysosome through the multivesicular body. If ubiquitin from the endocytosed receptors is removed by an UBP, the receptor recycles back to the membrane. Proteasome inhibitors block endocytotic degradation of some proteins such as glutamate receptor subunits indicating a possible role for the proteasome.

ESCRT-II and ESCRT-III are thought to be necessary for the continued sorting into the multivesicular body.[202,203] Recent studies indicate that ubiquitin might play a role in delivering proteins from the *trans*-golgi network (TGN) to the early endosome. In yeast, the protein general acid permease (Gap1) is delivered directly from the TGN to the lysosome under nutrient conditions in which Gap1 is not required at the plasma membrane. Proteins called Golgi-localized γ-ear-containing, Arf-binding (GGA) proteins have been shown to bind ubiquitin. It has been suggested that GGAs bind the ubiquitin tag on Gap1 and direct ubiquitinated Gap1 from the TGN to the endosome.[204] Investigations carried out by another group showed that a specific GGA protein called GGA3 is critical for sorting EGFRs from early endosomes to the MVB.[205]

Many questions regarding the role of ubiquitination in endocytosis remain unanswered. It is not clear, for example, whether multiple monoubiquitinated receptors follow a different initial endocytotic route compared to polyubiquitinated proteins. In addition, it is not understood whether receptors with a single ubiquitin attachment and receptors with multiple monoubiquitin attachment bind to different adaptor proteins. The question of the relative role of ubiquitin-dependent endocytosis compared to ubiquitin-independent endocytosis requires answers as well. Some mammalian receptors such as the transferrin receptors (TfRs) undergo ubiquitin-independent internalization, which is a typical example of constitutive endocytosis. In contrast, several receptors such as EGFR and neurotransmitter receptors such as AMPA (α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid) receptors undergo ligand-dependent endocytosis, which is often referred to as regulated endocytosis. As of now, no rules regarding the role of ubiquitin in the two main types of endocytosis have emerged. For example, in yeast, constitutive endocytosis of Ste3p requires ubiquitination and leads to degradation, whereas ligand-induced endocytosis of Ste3p is ubiquitin-independent and leads to

recycling.[206] In the case of EGFR, both constitutive and ligand-induced endocytoses require ubiquitin and both result in degradation of the receptor. It is probable that some components other than ubiquitin determine the switch between constitutive and regulated endocytosis. For example, tyrosine phosphorylation of Eps15 is required for ligand-induced endocytosis but not for constitutive endocytosis.[207] It is presumed that ubiquitin is probably a key determinant in deciding whether the internalized receptors are routed to the MVB or recycled back to the plasma membrane.

5.21.10.1 Endocytosis and Synaptic Function

Internalization through ubiquitin-mediated signaling and sorting into the multivesicular body probably plays a critical role in controlling the neurotransmitter receptor number on the plasma membrane (hence synaptic function) in the nervous system. In support of this idea, Burbea *et al.*[208] found in *Caenorhabditis elegans* that GLR-1, a homologue of the mammalian GluR1, which is part of the AMPA-type glutamate receptor, is internalized through an ubiquitin-mediated mechanism. GLR-1 was found to be ubiquitinated *in vivo*. When ubiquitin was overexpressed, the quantity of GLR-1 on the neuronal surface was reduced. The effect of ubiquitin overexpression was blocked by mutations in *unc-11* gene, which encodes a clathrin adaptin protein (AP180). Mutation of a specific lysine residue to arginine in the cytoplasmic tail of GLR-1 was reduced. All these results together suggest that GLR-1 is endocytosed through an ubiquitin- and clathrin-dependent mechanism. Furthermore, mutation of the lysine residues that prevents the ubiquitination of GLR-1 affected locomotion in *C. elegans*. These observations indicate that prevention of ubiquitin-mediated GLR-1 endocytosis leads to an increase in synaptic strength resulting from a higher number of GLR-1 on the neuronal surface.

Since the cytoplasmic tails of four mammalian AMPA receptor subunits (GluR1–GluR4) and *C. elegans* GLR-1 have a stretch of 16 conserved amino acids homologous to the region in the yeast Ste2p and Ste6p proteins that have been shown to be signals for ubiquitination and endocytosis, it is highly probable that GluR subunits in other species are endocytosed through an ubiquitin-dependent mechanism as well. In support of this idea, ubiquitination of AMPA receptor subunits in the mouse brain has been found. (S. C. Upadhya and A. N. Hegde, unpublished observations). Moreover, it was found that internalization of GluR1 and GluR2 subunits of the AMPA receptor is inhibited by the introduction of ubiquitin chain elongation mutant (Lys48 mutated to Arg48).[209] Intriguingly, they observed that a proteasome inhibitor MG132 reduced internalization of AMPA receptors. One would expect the ubiquitinated GluRs to be either recycled back to the plasma membrane or routed to the lysosome for degradation. Involvement of the proteasome, if proven, would add a new twist to the process of endocytosis. Similar effect of MG132 on degradation of internalized vasopressin receptors[200] as well as the quantity of GABA receptors[210] have been reported (see below for description).

Other neurotransmitter receptors have been shown to be internalized through an ubiquitin-dependent process as well. Using *Xenopus* oocyte expression system, Buttner *et al.*[211] demonstrated that inhibitory glycine receptors are ubiquitinated at the plasma membrane and internalized. The internalization process generates fragments of 35 and 13 kDa. In their experiments, application of concanamycin, which blocks acidification of lysosomal and endosomal compartments by inhibiting vesicular H^+-ATPases, prevents cleavage whereas the proteasome inhibitor lactacystin has no effect on generation of the smaller fragments of glycine receptor. $GABA_A$ receptor number on the neuronal membrane appears to be determined by an ubiquitin-mediated process as well. Evidence for a possible role of ubiquitin in $GABA_A$ receptor internalization is indirect. The $GABA_A$ receptor interacts with the UbL protein Plic-1 that stabilizes the $GABA_A$ receptor. Application of the proteasome inhibitor lactacystin leads to a significant increase in the steady-state levels of α1- and β3-subunits of the $GABA_A$ receptor.[210] The inference drawn with respect to $GABA_A$ internalization is that association with Plic-1 prevents the receptor from getting ubiquitinated and being routed to either the lysosome or the proteasome for degradation. It is not clear what roles proteasome and lysosome play in $GABA_A$ receptor degradation.[212] The Plic-1 UbL domain contains a proteasome-interacting motif.[41] Since Plic-1 also binds the $GABA_A$ receptor α- and β-subunits through a different domain (ubiquitin-associated (UBA) domain that is different from UbL domain), it is possible that the UbL domain in Plic-1 is utilized for routing the internalized $GABA_A$ receptor subunits for degradation through the proteasome.

5.21.10.2 Endocytosis of G-Protein-Coupled Receptors

Ubiquitination of plasma membrane receptors probably has a widespread role in the brain. Recently, mammalian GPCRs have been shown to be endocytosed through an ubiquitin-mediated mechanism. Two receptors, the β2-adrenergic receptor (β2-AR) and the V2-type vasopressin receptor (V2-VR; V2-VR was used rather than V2R to avoid confusion with a family of vomeronasal type 2 receptors called V2Rs), have been studied in detail.[200,213,214] The endocytosis of GPCRs mediated by the ubiquitin signal differs from the endocytosis of other receptors in that an adaptor protein called β-arrestin plays a role. Moreover, differential ubiquitination appears to target the endocytosed receptor for recycling or route them to the lysosome for degradation.

The β2-AR undergoes ubiquitination upon stimulation with isoproterenol in Chinese hamster fibroblast cells. Ubiquitination of the receptor occurred within 15 min of exposure to the agonist and the receptors remained ubiquitinated for up to an hour. β-Arrestin2 was also ubiquitinated. Ubiquitination of β-arrestin was rapid and transient occurring within 1 min of the agonist exposure and declining after about 10 min. Yeast two-hybrid screen revealed that β-arrestin2 interacts with Mdm2 ubiquitin ligase. In *in vitro* reactions, Mdm2 catalyzes the ubiquitination of both β2-AR and β-arrestin2. Furthermore, Mdm2-catalyzed ubiquitination of β-arrestin2 is required for the internalization of β2-AR.[214] Ubiquitination of β2-AR and β-arrestin2 is thought to help sequester the receptor into clathrin-coated pits and endocytosis. About 15 min after endocytosis, however, only β2-AR is seen in endocytic vesicles. β-Arrestin2 is thought to dissociate from the receptor presumably because of deubiquitination of β-arrestin2. Although no direct proof for causative action of a DUB on β-arrestin2 dissociation has been obtained, the disappearance of ubiquitination of β-arrestin2 (hence inferred deubiquitination) correlates with dissociation of β-arrestin2 from β2-AR. Dissociation of β-arrestin2 from the receptor routes the endocytic vesicle containing the receptor for recycling back to the plasma membrane (**Figure 9(a)**).

In contrast to the example of β2-AR, endocytosis of V2-VR typifies the regulated ubiquitin-mediated endocytotic removal and destruction of the receptor, which effectively reduces the receptor number on the plasma membrane (**Figure 9(b)**). In response to stimulation by the agonist, arginine–vasopressin V2-VR is internalized. Agonist stimulation induces ubiquitination of V2-VR as well as of β-arrestin2. Ubiquitination of β-arrestin2 as well as of V2-VR is persistent. A mutant in which the Lys268 is mutated to arginine (K268R-V2-VR) is not ubiquitinated. Unstimulated wild-type V2-VR as well as the mutant receptor (K268R-V2-VR) is degraded at similar rates. Interestingly, upon agonist stimulation, degradation of the wild-type receptor is greatly enhanced, whereas the degradation of the mutant receptor is not significantly affected.[200] Based on a smear of high molecular weight, ubiquitin conjugates, Martin *et al.*[200] have suggested that V2-VR is polyubiquitinated. Since mutation of a single Lys residue (Lys268) eliminates ubiquitination of V2-VR, polyubiquitination of V2-VR is a possibility. It would be interesting to determine how the endocytosis of polyubiquitinated plasma membrane receptors differs from that of monoubiquitinated receptors.

How is endocytosed V2-VR degraded? Is it routed to the lysosome? Although earlier studies showed that V2-VRs were delivered to the lysosome after internalization, Martin *et al.*[200] found that ubiquitinated V2-VR could only be detected in the presence of a proteasome inhibitor MG132, thus suggesting that the ubiquitinated V2-VR was degraded by the proteasome. Is it possible, however, that both the proteasome and lysosome have a role in degradation of V2-VR. Requirement for both proteasome- and lysosome-mediated degradations has been reported for other plasma membrane receptors such as insulin-like growth factor receptor,[215] estrogen receptor,[216] and growth hormone receptor.[217]

5.21.10.3 Endocytotic Degradation: Where Do the Lysosome and the Proteasome Fit?

The traditional view of endocytotic degradation is that the degradation occurs in the lysosome. Besides, as discussed above, ubiquitin appears to target the endocytosed proteins to multivesicular body for eventual degradation in the lysosome. A role for the proteasome is also indicated based on several studies. For example, endocytotic degradation of interleukin-2 receptor complex,[218] growth hormone receptor,[217] δ opioid receptor,[219] vasopressin receptor,[200] and $GABA_A$ receptor subunits[210] is blocked by proteasome inhibitors. In addition, proteasome inhibitors have been shown to block endosomal sorting of membrane proteins to the

lysosome.[220] There is a caveat to these studies, however. Prolonged application of proteasome inhibitors leads to depletion of the ubiquitin pool.[209] Therefore, proteasome inhibitors might have an adverse effect on ubiquitination of receptors and hence indirectly block lysosomal degradation of receptors. If there is a role for the proteasome, it might be to partially degrade some proteins which otherwise block routing to the lysosome through the multivesicular body. Since proteasome is known to act on retrotranslocated ER proteins, it might act on the proteins in the early endosome in a similar fashion. In support of this idea, the investigations by Rocca *et al.*[221] suggest that proteasome inhibitors might have a global effect on intracellular trafficking. A least possibility is that endocytosed proteins are degraded through the lysosome or the proteasome the choice which is determined by differential ubiquitination.

5.21.11 The Ubiquitin–Proteasome Pathway and Endoplasmic Reticulum-Associated Degradation

Proteins destined for secretion or the plasma membrane are translocated into the endoplasmic reticulum (ER), the organelle that mediates the appropriate delivery of proteins. The proteins enter the ER in an unfolded state and the ER modifies and folds the proteins to their final biologically active conformation. In the ER, proteins are believed to undergo a quality-control check that distinguishes properly folded proteins from misfolded proteins or unassembled subunits of a given protein.[222] How are the proteins that do not pass the quality-control mechanism degraded? Although it was thought that proteases in the ER lumen might degrade the misfolded proteins, there has not been sufficient experimental evidence to support this idea. Several studies carried out over the past decade suggest that the ubiquitin–proteasome pathway is responsible for the degradation of improperly folded proteins in the ER.[223–228]

Since the components of the ubiquitin–proteasome pathway do not exist in the ER, how does the proteolysis of badly folded proteins occur? One idea is that proteins are retrotranslocated from the ER to the cytoplasm. There is some evidence to indicate that the complex that is responsible for translocating the proteins into the ER, Sec61 translocation complex (translocon), might also be responsible for retrotranslocation.[229,230] BiP, a chaperone protein, is believed to have a critical role in retrotranslocation.[230] Almost all the retrotranslocated proteins are polyubiquitinated and degraded by the proteasome[231] although a protein, mutated pro-α-factor, is not ubiquitinated but is degraded by the proteasome.[232]

In addition to playing a role in degradation of improperly folded proteins, polyubiquitination and the proteasome might play a role in the retrotranslocation. Polyubiquitination might act as a 'ratcheting' mechanism to pull the misfolded protein from the ER membrane into the cytoplasm. In support of this notion, retrotranslocation proceeds only if polyubiquitin chain has reached a certain length.[233] Moreover, impaired ubiquitination results in increased association of mutant ribophorin with Sec61 in the mammalian ER.[234] Besides, it has been suggested that proteasome, by cleaving cytoplasmic domains and loops between membrane-spanning regions of proteins inserted into the ER membrane, could aid in retrotranslocation of the ER membrane proteins into the cytosol.[235]

5.21.12 The Ubiquitin–Proteasome Pathway and Transcription

5.21.12.1 Ubiquitin and Transcription

A link between ubiquitin and transcription was suggested several years ago when ubiquitination of histones was discovered. Histone A24, which was considered a variant of histone Histone2A (H2A), was found to be ubiquitinated and its levels altered in response to external stimuli. Similarly, ubiquitination of Histone 2B (H2B) also increases under certain physiological conditions, such as the growth phase of the slime mold *Physarum polycephalum*. In the slime mold, 6–7% of H2A and H2B are ubiquitinated. In the yeast *S. cerevisiae*, approximately 10% of H2B is ubiquitinated. Mutation of H2B in the residues that are ubiquitinated induces defects in mitotic cell growth and meiosis. Rad6 (Ubc2) has been identified as the UBC (E2) that is critical for ubiquitination of Histone2B.[236] The ubiquitin ligase (E3) required for ubiquitination of H2B has been identified to be a RING-finger-containing enzyme called Bre1.[237,238]

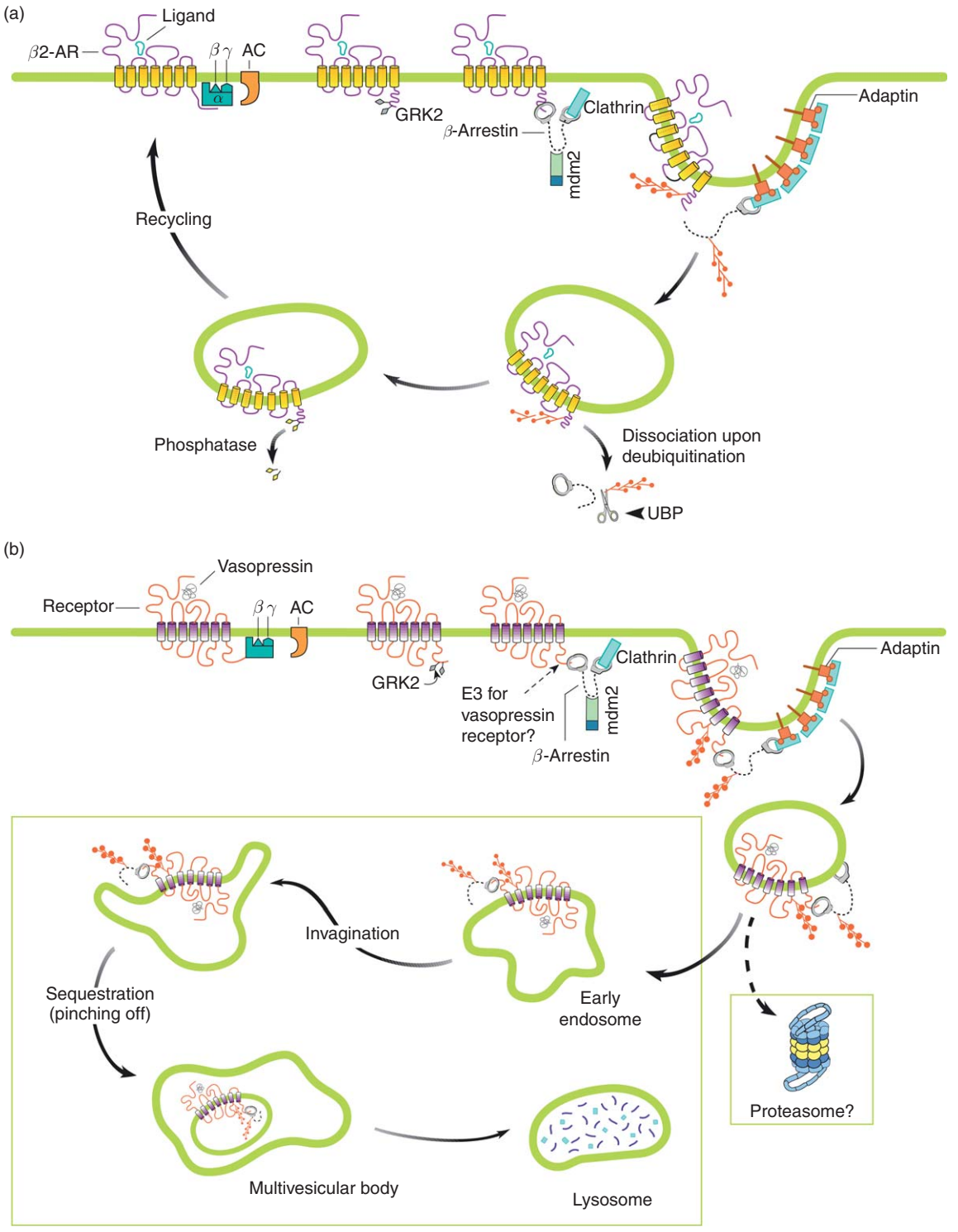

(a)
Ligand
β2-AR
βγ
AC
α
GRK2
β-Arrestin
Clathrin
mdm2
Adaptin
Recycling
Phosphatase
Dissociation upon
deubiquitination
UBP
(b)
Vasopressin
Receptor
βγ
AC
GRK2
E3 for
vasopressin
receptor?
β-Arrestin
mdm2
Clathrin
Adaptin
Invagination
Sequestration
(pinching off)
Early
endosome
Proteasome?
Multivesicular body
Lysosome

Figure 9

5.21.12.1.1 Histone ubiquitination and transcription

What is the role of histone ubiquitination in transcription? It is thought that histone modification such as ubiquitination, is a code deciphered by other histones or other regulatory proteins and the net outcome is determined to be either gene silencing or gene transcription.[239] In support of this idea, ubiquitination of Lys123 of H2B is a prerequisite for H3 methylation of Lys4 in yeast, a critical step in regulation of gene expression. Mutation of Lys123 in H2B impaired gene silencing at telomeres. It has been argued that H2B ubiquitination might serve as a 'master switch' that controls site-specific histone methylations that control telomeric gene silencing.[240]

Another speculation has been put forth. Because the turnover of methyl groups such as the modification on lysine residues is low, ubiquitination of histone N-termini might serve as a signal for proteolysis of methylated histones such that dynamic regulation of the chromatin is possible. Since no histone demethylases have been identified, ubiquitin-mediated proteolysis might be a way to reverse the effects of histone methylation.[239]

The enzymes that conjugate ubiquitin to H2B were discovered in the budding yeast. First, the E2 was identified as Rad6.[236] A RING-finger protein called Bre1 was subsequently identified to be the E3 for H2B ubiquitination.[237,238] Mutation of Bre1 had the same phenotypic effect that of Rad6 such as transcriptional silencing. Bre1 and Rad6 form a complex with another protein called Lge1 in *S. cerevisiae*. Lge1 is essential for H2B ubiquitination. In *S. pombe*, the homologues of Rad6, Bre1, and Lge1 are Rhp6, Brl2, and Shf1, respectively.[241] Shf1 is also required for H2B ubiquitination.[242] In mammals, the identity of the enzymes catalyzing H2B ubiquitination is less clear. HR6A and HR6B are human homologues of Rad6.[243,244] The Rad6 homologue in mice Hrb6a and Hrb6b, which are identical to their human counterparts, can rescue Lys4 H3 methylation in yeast carrying rad6 deletions. The data with Hrb6b null mutant mice suggest that Hrb6a and Hrb6b might have redundant functions because the mutants are viable and exhibit H2B ubiquitination comparable to wild-type mice.[245] Bre1 homologues have been found in *Drosophila* and humans. Two proteins RNF20 and RNF40 in humans are homologous to Bre1 but only RNF20 affects H2B ubiquitination. RNF20 and RNF40 have been shown to form a complex *in vivo*.[246,247]

Additional E2s and E3 could potentially ubiquitinate H2B in mammalian cells. An E2, UbcH6, can catalyze H2B ubiquitination *in vitro*. With respect to E3a, it has been found that Mdm2, which is known to be a ligase for the tumor suppressor protein p53, can ubiquitinate H2B *in vitro*.[248] Overexpression of Mdm2 leads to ubiquitination of both H2A and H2B although ubiquitination does not seem to be specific for the lysine residue (Lys120) that is usually modified by ubiquitin linkage. Other E3 ligases such as BRCA1 can ubiquitinate H2B *in vitro* but it remains to be seen whether the same holds true *in vivo*.[249–251]

The DUBs that remove ubiquitin from H2B in *S. cerevisiae* are called Ubp8 and Ubp10. Ubp8 is a component of the SAGA (Spt-Ada-Gcn5-acetyltransferase) complex and deubiquitinates H2B *in vitro* and *in vivo*.[252] The orthologues of Ubp8 and Ubp10 have been identified in *Drosophila* and humans and called Nonstop and Usp22, respectively. Ubp10 functions independently of the SAGA complex in yeast. Ubp8 and Ubp10 appear to have nonredundant function with respect to H2B ubiquitination. Deletion of both Ubp8 and Ubp10 leads to higher levels of ubiquitinated H2B than either deletion alone. Ubp8 and Ubp10 differ in their physiological function as well. Whereas Ubp8 has a role in transcriptional activation, Ubp10 functions in telomeric and rDNA silencing.

Figure 9 Role of ubiquitin in endocytosis of G-protein-coupled receptors. (a) Internalization and recycling of G-protein-coupled receptors (GPCRs). After ligand binding, the β-adrenergic receptor (β2-AR) is phosphorylated by β-adrenergic receptor kinase (GRK2). The β2-AR is then ubiquitinated along with the protein β-arrestin2. Mdm2 ubiquitin ligase ubiquitinates β-arrestin2 and possibly the same ligase ubiquitinates β2-AR as well. The binding of ubiquitinated β-arrestin2 to ubiquitinated β2-AR sequesters the two proteins into clathrin-coated pits because β-arrestins bind to clathrin. After internalization, β-arrestin2 is deubiquitinated perhaps by a UBP (shaded scissors), which causes it to dissociate from the receptor. β2-AR recycled back to the plasma membrane after a phosphatase removes the phosphate groups. (b) Internalization and degradation of GPCRs. The initial sequence of events for ubiquitin-mediated internalization for this class of GPCRs (such as the vasopressin receptor) is the same as in panel (a). After internalization, however, ubiquitinated β-arrestin2 remains associated with the ubiquitinated receptor, which targets the proteins to the lysosome for degradation. Proteasome inhibitors have also been shown to block degradation of endocytosed vasopressin receptor although the mechanism by which proteasome participates in the degradation of the vasopressin receptor is not clear.

Recently, another DUB called USP3 has been shown to deubiquitinate both H2A and H2B.[253] USP3 appears to play a critical role in progression of the S-phase of the cell cycle and in maintaining integrity of the genome.

H2B ubiquitination follows the early steps of transcription initiation and elongation. The E3 ligase Bre1 and the UBC Rad6 are recruited to the promoters that are ready to commence transcription. Initially, Bre1 is recruited by its interaction with transcription activators such as Gal4 in yeast or p53 in human cells. Bre1 in turn recruits Rad6 and its binding partner Lge1.[237,238,254] In addition to the recruitment of Rad6, H2B ubiquitination requires components of the PAF complex which consists of Paf1, Rtf1, Ctr9, Leo1, and Cdc73 (**Figure 10**). PAF is known to associate with initiating and elongating RNA polymerase II. H2B ubiquitination appears to require transcription as shown by *in vitro* studies. Additional evidence for the requirement of transcription for H2B ubiquitination comes from the fact that disruption of other factors plays a role in transcription elongation. These factors include the Bur1/Bur2 cyclin-dependent protein kinase complex (BUR complex) and hHR6A.[255] Also, the initial process of transcription elongation appears to be essential for H2B ubiquitination. Transcription elongation by RNA polymerase II requires sequential phosphorylation by CDK7 in human cells (Kin28 in yeast) followed by CDK9 (Ctk1 in yeast) on its C-terminal heptapeptide repeat sequences. Loss of Kin28 eliminates H2B ubiquitination whereas loss of Ctk1 has no effect suggesting that these early phosphorylation steps that drive transcription forward are required for H2B ubiquitination.[256]

5.21.12.1.2 Histone ubiquitination and histone methylation

H2B monoubiquitination is an obligatory step in Lys4 H3 and Lys79 H3 methylation in yeast as well as higher eukaryotes (**Figure 10**). The connection between monoubiquitination and H3 methylation functions unidirectionally: deletions and mutations that block H2B monoubiquitination reduce the level of H3 methylation but deletions and mutations that adversely affect H3 methylation have no effect on H2B monoubiquitination. It has also been established that H2B monoubiquitination specifically affects di- and trimethylation of Lys4 H3 and Lys79 H3 methylation but does not affect monomethylation. Besides, mutations in the BUR complex, which decrease H2B ubiquitination, only affect trimethylation and not dimethylation of Lys4 H3. How is H2B monoubiquitination linked to H3 methylation? The answer to this question came from studies on COMPASS, a protein complex with histone methyl transferase activity. COMPASS consists of an essential subunit called Set1 and seven other subunits.[257]

A role for yet-another protein complex called the Cccr4–Not has also been suggested in the process that links H2B ubiquitination Lys4 H3 methylation. Cccr4–Not functions in mRNA production and processing. It has nine core subunits in yeast. These include an mRNA deadebylases CCR4 and Caf1, Not 1-5, Caf40, and Caf130. The Cccr4–Not complex regulates mRNA degradation as well as transcription. Mutating the components of the Cccr4–Not complex causes a reduction in trimethylation of Lys4 H3 but not mono- or

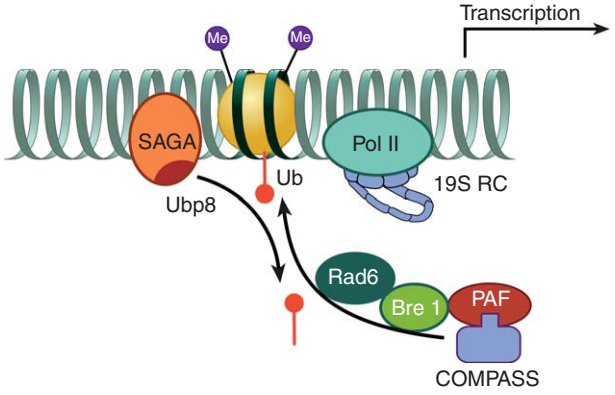


Figure 10 Role of histone monoubiquitination in transcription. The schematic figure shows attachment of single ubiquitin to histone H2B (yellow ball; the spiral indicates the DNA in chromatin) through the action of a complex containing the ubiquitin-conjugating enzyme (E2) called Rad6, the ubiquitin ligase Bre1 and the associated PAF complex, and the COMPASS protein. Deubiquitination of histone H2B occurs through a DUB associated with the SAGA complex called Ubp8. Histone H2B monoubiquitination is linked to its methylation (Me).

dimethylation. Moreover, the subunits of the Cccr4–Not complex interact with the 19S regulatory complex of the proteasome and the BUR complex. The evidence for a role of Cccr4–Not complex in regulating H2B ubiquitination is equivocal. For example, mutations in the components of the Cccr4–Not complex reduce the recruitment of the 19S proteasome to the promoters and reduce Lys4 H3 trimethylation, neither recruitment nor CTD phosphorylation of RNA polymerase II is affected. Besides, with respect to the effect of mutation in the Cccr4–Not complex subunits on H2B ubiquitination, one study found evidence in favor of a linkage,[258] whereas the other obtained evidence against the effect of Cccr4-Not on levels of ubiquitinated H2B.[259]

Based on numerous studies on yeast as well as higher eukaryotes, a picture of connection between H2B ubiquitination and Lys4 H3 methylation is beginning to emerge. The data are consistent with a model in which COMPASS associates with RNA polymerase II phosphorylated on Ser5 of its CTD through its interactions with the PAF complex.[257] In the absence of H2B ubiquitination, COMPASS lacks its key Cps35 subunit, which is capable of only monomethylating Lys4 H3. Following the initial monomethylation, Rad6 H2B-ubiquitin-conjugating activity is promoted by the BUR complex and through the association of Rad6 with the PAF complex and COMPASS. H2B ubiquitination facilitates the recruitment of Cps35, which enables COMPASS to catalyze di- and trimethylation of Lys4 H3. The di- and trimethylation reaction requires the 19S proteasome and possibly the Cccr4–Not complex. These protein complexes might also play a role in the recruitment or activation of Cps35 and perhaps that of the SAGA histone acetyl transferase complex. SAGA deubiquitinates H2B through the action of Ubp8 after Lys9/14 acetylation. It should be noted that although recruitment of Cps35 to the promoter requires H2B ubiquitination, it does not require the COMPASS complex itself.[257] Therefore, it is highly probable that control of Cps35 recruitment to the promoter region might be a critical step in linking H2B ubiquitination to Lys4 H3 di- and trimethylation.

Recent evidence suggests that H2B ubiquitination may have a role in transcription independent of Lys4 H3 methylation.[241] Data from experiments on transcription elongation indicate that H2B ubiquitination possibly assists a histone chaperone called facilitates chromatin transcription (FACT) and facilitates movement of RNA polymerase II through nucleosomal templates.[260] In *S. pombe* loss of H2B ubiquitination phenotypes was not observed in cells that lack Lys4 H3 methylation. H2B ubiquitination appears to have a greater role in gene regulation than Lys4 H3 methylation at least in the fission yeast. It remains to be seen whether this holds true in higher eukaryotes.

H2B ubiquitination seems to be a dynamic signal rather than a static one for transcription because both H2B ubiquitination and deubiquitination are required for transcription to go forward. It has been found that removal of ubiquitin from H2 by Ubp8, a DUB contained within SAGA, is essential for recruitment of the kinase Ctk1. In support of the idea, deletion of the E3 ligase Bre1 that ubiquitinates H2 rescued the Ctk1 recruitment defect that is seen with loss of Ubp8. It is thought that H2 ubiquitination acts as a barrier that prevents Ctk1 recruitment to RNA polymerase II during transcription elongation.[261] Affinity-purified Ctk1 binds to histone H2A and H2B but does not bind to H2B containing an ubiquitin tag indicating that Ctk1 interacts directly with histones and ubiquitination of H2B prevents this interaction. H2B ubiquitination might be an early checkpoint for transcription elongation. How does transcription proceed beyond this checkpoint? It is believed that removal of the ubiquitin moiety from H2B by Ubp8 would allow recruitment of Ctk1, Ser2 phosphorylation of the CTD of RNA polymerase II. Phosphorylation of CTD allows binding of Lys36 methyltransferase Set2 and methylation of Lys36 pushes transcription forward. Both Ubp8 and Rad6 have been shown to associate with RNA polymerase II which suggests that multiple round of ubiquitination and deubiquitination might be required for transcription elongation.

H2A monoubiquitination also plays an important role in transcription. Two E3 ligases in humans called Ring1B and 2A-HUB catalyze attachment of ubiquitin to H2A. Ring1B ligase appears to be responsible for much of H2A monoubiquitination because knockdown of this ligase greatly reduces ubiquitinated H2A levels.[262–264] The H2A ligases are largely associated with transcriptional silencing. Ring1b associates with three different repressive complexes: Polycomb repressive complex (PRC1), E2F-6.com-1, and the FBXL10-BcoR complex. PRC1 contains Ring1A, Ring1B, and Bim-1. Although all three subunits in the PRC1 complex are potential E3 ligases, only Ring1B has ubiquitin ligase activity specific to H2A. Bim-1 seems to stimulate the Ring1B ligase activity.[265] A second repressor complex with which Ring1B associates, E2F-6.com-1, contains E2F-6, Mga, Max, DP-1, HP1γ, MBLR, h-I(3) mbt-like protein, YAF2, Eu-HMTase1, and NG36/G9a. Of these, NG36/G9a is a Lys9 H3 methyltransferase.[266] It is not clear whether Ring1B in E2F-6.com-1 has

H2A-ubiquitinating activity. A third complex called the FBXL10-BcoR that contains Ring1B has also been identified. This complex contains several additional proteins such as CK2α, YAF2, Skp1, HP1γ, Cbx8, NPSC-1, Bmi-1, BcoR, and FBXL10.[267,268] NPSC1 has been shown to stimulate H2A-ubiquitinating activity of RingB1. There are other ligases such as 2A-HUB that ubiquitinate H2A as well. DUBs that remove ubiquitin from H2A have been identified. These include Ubp-M, 2A-DUB, and USP21.

5.21.12.2 Proteasome and Transcription

Proteasome has two types of roles in transcription. One is through degradation of ubiquitinated proteins and the second is by means of its nonproteolytic role. Many transcription factors are short-lived proteins and several have been shown to be substrates for the ubiquitin–proteasome pathway.[269] Regulation of transcription factor amount by ubiquitin–proteasome-mediated degradation has been shown to control many physiological processes. For example, Runx2, a transcription factor that plays a critical role in bone development is ubiquitinated and degraded by the proteasome.[270] Other transcription factors regulated by the ubiquitin–proteasome pathway include the repressor of cAMP-responsive element binding protein,[271] sterol regulatory element binding protein,[272] signal transducers and activator of transcription (STAT) factors,[273] muscle transcription factors MyoD and its inhibitor Id1,[274] transcription factors that respond to light signal in plants,[275] and several nuclear receptors such as estrogen receptor-α.[276]

Studies on the transcription factor myc revealed an interesting connection between ubiquitin–proteasome-mediated proteolysis and transcriptional activation. The myc transcriptional activation domain overlaps the sequence that targets myc for degradation. Furthermore, this overlap between the transcriptional activation domain and sequences targeting for ubiquitin–proteasome-mediated degradation was found to be true for several other transcription factors such as fos, jun, p53, β-catenin, and myb.[277] Furthermore, using a herpes simplex virus transcriptional regulatory protein VP16, it was shown that ubiquitination by an ubiquitin ligase called Met30 is required for transcriptional activation. Moreover, transcriptional activation and degradation are intimately linked. It has been proposed that ubiquitination activates transcription and, at the completion of transcription, the ubiquitinated transcription activator is degraded by the proteasome. Thus, ubiquitination acts as an in-built timer for the transcription of genes.[278] Additional evidence has been found for the role of the ubiquitin–proteasome pathway in transcription. A subunit of an ubiquitin ligase complex SCF, Skp2, has been shown to regulate proteasome-mediated degradation of c-myc and to act as a cofactor for c-myc-regulated transcription.[279,280] Interestingly, two subunits of the proteasome, Rpt3 and Rpt6, were shown to be recruited to the Cyclin D2 promoter in a myc-dependent manner.[280] Recent studies show that inducible transcription factors are subject to proteasome-mediated regulation as well. The studies on inducible yeast transcription factors Gcn4, Gal4, and Ino2/4 showed that ubiquitin–proteasome-mediated transcription factors is required for the transcription of target genes. Interestingly, proteasome-mediated degradation of these transcription activators was also required for their association with Pol II.[281] In addition, the proteasome appears to precisely regulate amounts of proteins in the SWI/SNF (mating-type switch/sucrose nonfermenting) chromatin remodeling complex.[282] A key function of the proteasome is probably the proteolytic removal of stalled Pol II (see **Figure 1**) to terminate transcription and to remove Pol II stalled by DNA damage.[283]

5.21.12.3 Nonproteolytic Role of 19S Subunits of the Proteasome in Transcription

The proteasome also appears to have nonproteolytic roles in transcription. The 19S subunits of the proteasome are recruited to activated yeast GAL1–10 promoter of yeast. Moreover, it was shown that the 19S complex of the proteasome can activate RNA polymerase II-mediated transcription elongation in a proteolysis-independent manner.[284–286] Recently, transcription of the activator of the APC ubiquitin ligase has been shown to require association of the proteasome with the promoter. This study showed that the proteasome subunit Rpt6 is critical for transcription.[287] In addition, two ATPases in the 19S RC, Rpt4 and Rpt6, have been shown to link histone H2B ubiquitination to histone H3 methylation, which distinguishes active from silent chromatin.[288] In this context, it is interesting to note that some proteasome subunits are homologous to transcription regulators. For example, the proteasome subunit Mov34 (Sug1 in yeast and S12 in human) shares homology with the transcriptional regulator JAB1/Pad1. The proteasome subunit S5a is homologous to the basal transcription

factor TFIIH.[149] 19S RC of proteasome helps recruit SAGA to target promoters. Studies on *S. cerevisiae* showed that the 19S RC can enhance targeting of a transcriptional coactivator named SAGA to promoters.[289] SAGA is a huge complex (1.8 MDa) with Gcn5 histone acetyltransferase as a catalytic subunit along with many other chromatin-modifying proteins. Experiments using the gel-shift assay revealed the role of 19S RC in targeting of SAGA to Gal4 promoter. Addition of Gal4 DNA-binding domain (Gal4DBD) alone did not bind SAGA to the promoter. With the addition of 19S RC to a Gal4DBD bound to an activation domain (Gal4–VP16), significant binding of SAGA was observed. Which part of the 19S proteasome, base or the lid, is critical for SAGA targeting? It was found that the lid complex did not stimulate SAGA targeting whereas the base complex alone was able to enhance SAGA targeting to the promoter albeit to a lesser extent than the 19S RC. Characterization of SAGA targeting by 19S RC showed that ATP hydrolysis is required for SAGA targeting by 19S RC.[289] This is especially important given that the base complex consists of ATPases. The next set of experiments determined that 19S RC can also facilitate targeting of related transcriptional coactivator SAGA-like complex (SLIK). The observation with SLIK suggested that 19S RC may have a broader role in transcriptional activation. Initial experiments on SAGA used naked DNA. Does 19S RC target SAGA to the promoter under more natural conditions when the promoter is embedded in a nucleosome array? Indeed, the results showed that SAGA recruitment to the nucleosome array was increased in the presence of 19S RC in a Gal4–VP16-dependent manner. How does 19S RC stimulate SAGA binding to active promoters? It is possible that 19S RC stimulates the DNA binding activity or histone acetyltransferase (HAT) activity of SAGA. The specificity of 19S RC action on SAGA was tested through experiments with SAGA in comparison with another HAT complex called NuA4. 19S RC stimulated SAGA binding to DNA but not NuA4 binding to DNA. Furthermore, 19S RC increased the HAT activity of SAGA but not that of NuA4. Because 19S RC increased DNA binding and HAT activity of SAGA even in the absence of transcriptional activators, the results suggested that 19S RC specifically acts on SAGA.[289] The ability of the 19S RC to modify properties of SAGA suggests that the two complexes might physically interact. To test possible physical interaction, Lee *et al.*[289] used yeast strains with TAP-tagged Spt7 (part of the SAGA complex) or Epl1 (part of the NuA4 complex). TAP (tandem affinity purification)-tagging technique uses two affinity tags such as immunoglobulin-binding domain of *Staphylococcus aureus* and a calmodulin-binding peptide. When Spt7–TAP was pulled down with calmodulin resin, Sug1 ATPase, which is part of the 19S RC, was detected in the precipitated samples. Sug1 was not detected in the samples precipitated with Epl1-TAP. When purified 19S RC- and HA-tagged SAGA complex were mixed *in vitro* and immunoprecipitated with anti-HA antibody, both SAGA and 19S RC were immunoprecipitated.

Does 19S RC have an effect on SAGA function *in vivo*. Previously it was shown that deletion of SAGA Gcn5 catalytic subunit resulted in global reduction in acetylation of histone H3. Given the role of 19S RC in SAGA recruitment, the next question was whether the 19S RC plays a role in global acetylation. The answer to this question came from experiments in which a specific subunit of the base complex of 19S RC called Sug1 was mutated. The experiments utilized a yeast strain with a mutation in the ATPase domain of Sug1 (sug1–25) and another one with a mutation outside the Sug1 ATPase domain (sug1–3). Histone acetylation was almost completely abolished in sug1–25 mutants but not in sug1–3 mutants.[289] Additional evidence for the role of 19S RC in SAGA targeting came from genetic experiments that tested the growth of yeast strains carrying a deletion mutation in gcn5 gene and an ATPase domain mutation in the Sug1 gene. Although the growth of the single mutants was comparable to that of the wild type, the growth of double mutants with defects in both gcn5 and sug1 genes was impaired. Besides, growth was retarded in a strain carrying a mutation in the sug1 gene and deletion of a gene called spt20, which is required for maintaining the integrity of the SAGA complex. The specificity of the genetic interactions was demonstrated by showing that neither a mutation in another HAT Sas3 nor mutation in ubp8, a gene that encodes an ubiquitin-specific protease, has any growth-retarding effect. Thus, 19S RC and SAGA interact functionally as shown by the genetic studies.[289]

5.21.12.4 The Ubiquitin–Proteasome Pathway and Transcription: Possible Roles in Synaptic Plasticity

The function of different components of the ubiquitin–proteasome pathway in synaptic plasticity is considered in detail elsewhere in this chapter (Section 5.21.16.2). The possible mechanistic roles of ubiquitin and the proteasome in transcription as relevant to synaptic plasticity are discussed here.

The dependence of long-term synaptic plasticity on gene transcription and new protein synthesis has been extensively studied over the past several years. However, the lacuna is the elucidation of the regulation of transcription that occurs in neurons that are modified by learning. For example, what molecular processes determine the threshold for transcription? What determines the narrow window of gene expression observed in many long-term memory paradigms? In many instances, the initial stimuli that the neuron receives such as binding of a neurotransmitter to its receptor are the same for short- and long-term synaptic plasticity. Nevertheless, long-term synaptic plasticity requires induction of gene expression. Long-term synaptic plasticity-producing protocols somehow must initiate a transcription cascade. Thus far, the molecular processes that control threshold for transcription have not been elucidated. It is possible that the ubiquitin–proteasome pathway operates to regulate the threshold for gene induction. Minimally, the ubiquitin–proteasome pathway could play a role in regulating transcription at two steps. One is the relief of transcription repression and the other is the control of transcriptional activation. For example, it is well established that CREB-mediated gene expression is required for long-term memory formation in *Aplysia*, *Drosophila*, and mice.[290–292] It is also known that through alternative splicing, CREB gene generates activators and repressors.[293] The repressors are normally at a higher stoichiometry in the cell than the activators.[294,295] For CREB-mediated gene expression to go forward, the repression has to be relieved. In *Aplysia* neurons, we have found that the repressor CREB1b is degraded by the ubiquitin–proteasome pathway whereas the activator CREB1a is stable compared to the repressor.[271] The regulation of CREB1b degradation probably determines the threshold for induction of CREB-mediated gene expression and hence a threshold for long-term facilitation. With respect to the removal of activator by proteolysis, previous work in *Aplysia* has shown that C/EBP, a transcription factor induced during long-term facilitation, is degraded by the ubiquitin–proteasome pathway.[64] The degradation of C/EBP determines a time window during which transcription must take place to induce long-term facilitation. The newly discovered role of the proteasome in transcription might be relevant to synaptic plasticity as well. We have found that proteasome inhibitors block induction of BDNF mRNA.[296]

5.21.13 Ubiquitin-Like Proteins

Since the discovery of ubiquitin as a polypeptide modifier of substrate proteins, several other proteins related to ubiquitin have been discovered (**Table 3**). Some of the proteins such as small ubiquitin-related modifier (SUMO) and Rub1 are conjugated to substrate proteins in a manner similar to conjugation of ubiquitin. It is not clear whether this process is generally applicable to all UbL proteins. SUMO and Rub1 have been investigated in some detail and the findings are described below.

Table 3 Ubiquitin-like proteins and their function

Protein	*E1*	*E2*	*E3*	*Description*	*References*
SUMO1–4	Uba2	Ubc9	Siz1, Siz2, Mms21	Targets numerous substrates; functions in nuclear transport and transcriptional regulation	297, 298
Rb1/Nedd8	Uba3, Ula1	Ubc12	Dcs1	Neddylation of cullins activates SCF ligases; functions in cell cycle, synapse function, and other physiological processes	299, 300
Atg8	Atg7	Atg3	No E3	Required for autophagy; vesicular transport required	301, 302
Atg12	Atg7	Atg3	No E3	Required for autophagy required	302, 303
ISG15	UbeL1	UbcH8	Herc5, Efp	Functions in transcription and pre-mRNA processing	304, 305

5.21.13.1 SUMO

Modification of substrate protein by SUMO is referred to as sumoylation. The first substrate for sumoylation to be identified was RanGAP1, a GTPase-activating protein. Sumoylation of RanGAP1 was found to be essential for nucleocytoplasmic trafficking. Sumoylation RanGAP1 enables it to bind to RanBP2, a subunit of the nuclear pore complex.[307,308] It is believed that binding of RanGAP1 to the nuclear pore complex facilitates nuclear protein import. In support of this idea in yeast cells deprived of SUMO conjugation, nuclear import is adversely affected.[308]

Transcription regulators are also known to be sumoylated. One of the early studies of this phenomenon showed that promyelocytic leukemia protein (PML) is a substrate for SUMO conjugation. Once SUMO is attached to the PML protein it is directed to a subdomain of the nucleus called the PML oncogenic domain (POD). It is thought that POD localization of the PML protein allows it to recruit other proteins such as transcription factors. Transcription factors in the POD can activate or inhibit transcription.[309] Another transcription factor known to be sumoylated is Sp3.[310,311] SUMO also has roles in chromatin condensation and interphase chromosome organization.[312]

SUMO conjugation can also directly regulate signal transduction pathways. A remarkable instance of control of signaling occurs in *Dictyostelium*. When *Dictyostelium* is starved of nutrients, amoeboid cells produce cAMP, which triggers chemotaxis and aggregation of surrounding cells. A key part of the signaling cascade MEK1 is the substrate for both sumoylation and ubiquitination. In response to cAMP, MEK1 is rapidly sumoylated, which prevents its degradation and allows it to transduce signals. Soon after signal transduction is accomplished, MEK1 is de-sumoylated and translocated to the nucleus where it is ubiquitinated and degraded.[313]

In the nervous system, so far there is evidence of sumoylation of at least one protein critical for synaptic plasticity. Long *et al.*[314] found that *Drosophila* calcium–calmodulin-dependent kinase II (CaMKII) is conjugated to SUMO *in vivo*. Disruption of SUMO conjugation leads to various overt phenotypes although the exact physiological role of sumoylation of CaMKII is not understood.

5.21.13.1.1 SUMO conjugation pathway

SUMO proteins are similar to ubiquitin in their folded structure but possess only about 20% homology to the amino acid sequence of ubiquitin. The number of SUMO genes varies among eukaryotes. Organisms such as yeast, *C. elegans*, and *Drosophila* contain a single SUMO gene in their genome whereas plants and vertebrates possess several SUMO genes. In humans, there are four different SUMO genes (SUMO-1–4).[315] SUMO proteins are conjugated to substrate proteins (**Figure 11**) in a manner similar to ubiquitin. SUMO proteins are expressed as immature precursors with C-terminal extensions of varying lengths (2–11 amino acids) following a Gly–Gly sequence motif that is the hallmark of the mature C-terminus of SUMO. The immature form is processed to the mature form by SUMO-specific isopeptidases (sentrin-specific proteases; SENPs). The C-terminus of the processed SUMO with its Gly–Gly sequence is exactly like the C-terminus of ubiquitin. The processed SUMO is activated by an E1-activating enzyme. The SUMO E1, unlike the E1 that activates ubiquitin, is a heterodimer, which consists of AOS1 and UBA2. The SUMO E1 forms an AMP-intermediate and passes SUMO on to a conjugating enzyme UBC9. The available evidence suggests that UBC9 is the only SUMO E2 from yeast to humans. UBC9 passes SUMO onto one of the several E3s. For several years after the initial discovery of sumoylation, it was not clear whether or not SUMO attachment required E3-like proteins. Work from several groups led to the identification of Siz1 and Siz2 proteins in yeast and PIAS1 in mammals as SUMO E3s.[316–318] A large number of SUMO E3 ligases have a characteristic SP-RING (Siz/PIAS RING) motif which is critical for the ligase function.[319] The SP-RING ligases contain the PIAS family of proteins. These proteins have a conserved 400 amino acid long N-terminal domain in addition to the SP-RING. Siz1 and Siz2 from yeast belong to this family. In mammals, five PIAS proteins exist. These are PIAS1, two splice variants of PIAS2 called PIASxα, PIASxβ, PIAS3, and PIASy. The other SP-RING is MMS21, which has been found in yeast as well as humans. MMS21 protein is part of an octomeric protein complex called SMC5–SMC6 that is required for vegetative growth and DNA repair in yeast and for maintaining telomere length in ALT cancer cells. Another SP-RING ligase called Zip3 has been found in yeast.

A second class of SUMO ligases is typified by RanBP2, which has no similarity to any known ubiquitin ligases. RanBP2 interacts with SUMO and it is thought that RanBP2 facilitates ligation of SUMO by positioning

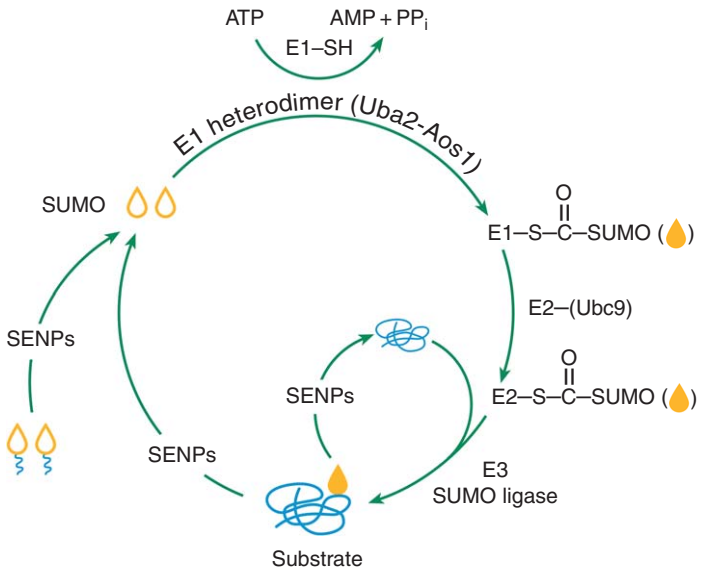


Figure 11 The SUMO conjugation pathway. SUMO (open drop-shaped objects) is synthesized with a C-terminal extension. SUMO-specific isopeptidases (SENPs) cleave the C-terminal extension from the SUMO precursor to produce a mature protein with glycine at the C-terminal. The mature SUMO is activated by an E1 heterodimer consisting of Uba2 and AOS1 in an ATP-dependent reaction. The action of E1 produces a SUMO–AMP intermediate (filled drop-shaped object), which is passed onto an E2 called Ubc9. E2 passes SUMO onto an E3 ligase, which attaches SUMO to the substrate. SUMO attachment is reversible and can be removed by SENPs.

the UbC9-SUMO thioester close to the Lys residue in the substrate.[307] Although no *in vivo* target for RanBP2 has been identified so far, it promotes sumoylation of proteins such as histone deacetylase HDAC4 but not of its binding partner RanGAP1.[320]

A third category of SUMO ligases includes a protein called Pc2 that belongs to the human Polycomb group (PcG) of proteins. PcGs form large multimeric complexes (PcG bodies) that function in gene silencing. Pc2 recruits the transcriptional corepressor CtBP to PcG bodies. Different lines of experiments have suggested that Pc2 is a SUMO E3 although whether it posseses a true catalytic activity remains to be established.[321,322]

Is there a specific consensus site for SUMO attachment in substrate proteins? With studies on just a handful of substrates such as RanGAP1, IκBα, p53, and c-jun the SUMO-acceptor site was found to be ΨKxE (where Ψ is an aliphatic branched amino acid and x is any amino acid).[323] The identification of a sumoylation consensus site is noteworthy because no such site for ubiquitination is found. Because ubiquitination is vastly more complex with hundreds of ligases, a common consensus sequence for ubiquitination may not exist. It must be pointed out, however, that the recent identification of the TEK box motif for APC suggests that consensus motifs for ubiquitin-acceptor sites might be found at least for some ligases.[324]

SUMO attachment can be removed by SUMO-specific isopeptidases just as UCHs and UBPs remove ubiquitin from substrate proteins. Thus, sumoylation is also a reversible modification. SUMO-specific isopeptidases possess a Cys residue in their catalytic core similar to UCHs and UBPs. Two SUMO-specific isopeptidases in yeast are called Ulp1 and Ulp2.[325] In humans, six Ulp homologues have been identified and are given the name sentrin-specific proteases (SENPs) because SUMO was previously known as 'sentrin'.

What is the physiological consequence of sumoylation? It appears that there is not a single outcome of sumoylation of a substrate. Sumoylation can affect the stability or subcellular localization of the substrate. Attachment of SUMO also affects protein–protein interactions either positively or negatively. For example, sumoylated RabGAP1 and p300 interact with RanBP2 and HDAC6, respectively.[326] An instance of negative influence of sumoylation is that of the transcription repressor ZNF76.[327] The sumoylation site of ZNF76 overlaps with its binding site for the TATA-binding protein. What are the mechanisms by which sumoylation affects protein–protein interaction? The positive effects might be due to the availability of additional molecular surface for

interaction. The negative effect might result from masking of binding sites by SUMO attachment or conformational alterations that do not favor binding to other proteins.

5.21.13.2 Rub1

Rub1 also known as Nedd8 shares about 53% sequence similarity with ubiquitin.[328] Rub1 conjugation to protein substrates is also highly similar to that of ubiquitin conjugation. Conjugation of Rub1 requires ULA1 and UBA3, two yeast proteins similar to E1, and UBC12, which is an E2. Interestingly, it was found that cullin, a subunit of multisubunit RING-finger E3s, was a major substrate for Rub1[329] (see also **Figure 2(c)**).

Conjugation of Rub1 to cullin appears to be essential for ubiquitin ligase activity. When Rub1 is conjugated to the cullin1 in the SCF ligase complex that ubiquitinates IκBα (see **Figure 2(b)**), the ubiquitin ligase activity of the SCF is enhanced. Rub1 conjugation can be reversed in a manner similar to conjugation of ubiquitin. The Rub1-cleaving activity is not similar to DUBs, however. Rub1 conjugated to substrates is cleaved by a huge complex of proteins called the COP9 signalosome (which has a subunit composition similar to that of a subcomplex of the 19S cap of the proteasome). A subunit called CSN5 is believed to have the isopeptidase activity against Rub1 conjugated to substrates. CSN5 has a metalloprotease sequence, which is also found in the related proteasome subunit Rpn11 that exhibits isopeptidase activity toward ubiquitin conjugates.[152]

5.21.14 Unusual Linkages of Ubiquitin

The ubiquitin signal that marks the substrates for proteolysis is the polyubiquitin chain in which ubiquitins are linked to each other through an isopeptide bond formed at Lys48. Other linkages such as polyubiquitin chain formed through isopeptide bond at Lys63 are known to occur[330] (**Table 4**). Polyubiquitin chains formed through Lys63 appear to have a role in signaling pathways. In mammalian cells, activation of the transcription factor NF-κB requires degradation of the inhibitor IκB. IκB becomes vulnerable to degradation when it is phosphorylated by a kinase called IκB kinase (IKK). IKK itself is activated by another protein kinase complex called TRIKA2. The activation of TRIKA2 has been shown to occur through attachment of Lys63-linked polyubiquitin chains in a proteasome-independent fashion. Formation of Lys63-linked polyubiquitin chains is catalyzed by the tumor necrosis factor receptor-associated factor 6 (TRAF6), an ubiquitin ligase.[333] Activation of the first kinase by Lys63-linked polyubiquitin chains in a protein kinase cascade could be a general mechanism to initiate complex signaling by kinases.[334,335] Lys29-linked polyubiquitin chains have also been found. This linkage, however, appears to have a role in assembling free polyubiquitin chains to the physiological significance of which is not understood.[332]

Although a role for unusual linkages of ubiquitin in signaling in the neurons has not been reported yet, given the complexity of the nervous system, we might expect several discoveries in the coming years.

Table 4 Types of ubiquitin linkages and their function

Protein	*Substrate(s)*	*Function*	*Reference*
Monoubiquitin	Neurotransmitter and G-protein-coupled receptors	Endocytosis Routing to the lysosome	208 201
Polyubiquitin (Lys48)	Numerous	Proteasome-mediated degradation	16
Polyubiquitin (Lys63)	TRAF6, plasma membrane proteins	DNA repair endocytosis	331
Polyubiquitin (Lys29)	Free Ub	Preassembly of polyUb chains	332
Polyubiquitin (Lys11)	Cyclin B1	Enables human APC to form polyubiquitin chains on specific substrates	324

5.21.15 Ubiquitin and Signaling

Ubiquitin attachment has also been found to function in signaling by modification of signaling molecules such as receptors and protein kinases. This function of ubiquitin is different from those in endocytosis and proteasome-mediated degradation. A key feature of ubiquitination that mediates signaling is that polyubiquitination is through K63 linkage rather than K48 linkage that marks substrates for proteasome-mediated degradation.

A well-studied example is that of IκB kinase (IKK) which mainly transduces signals that activate the transcription factor NF-κB. IKK is a multisubunit complex that contains four different subunits IKK-1 (also known as IKKα), IKK-2 (IKKβ), a regulatory subunit called NEMO (also known as IKKγ), and an additional subunit critical for substrate targeting called ELKS (because it is rich in glutamate [E], leucine [L], lysine [K], and serine [S]).[336,337]

Ubiquitination plays a role in recruitment of the IKK-signaling complex to the activated receptors such as tumor necrosis factor α (TNFα). When TNFα binds to its receptor TNF-R, three different proteins are recruited to TNF-R: receptor-interacting protein (RIP), TNF-R associated factor 2 (TRAF2), and cellular inhibitor of apoptosis 1 (cIAP1). Both RIP and TRAF2 are polyubiquitinated upon activation of TNF-R. TRAF2 is a RING-finger-containing ubiquitin ligase that ubiquitinates itself. Polyubiquitination of RIP and TRAF2 allows the IKK complex to be recruited to the activated TNF-R. To recruit the IKK complex to the activated TNF-R, NEMO needs to interact with the polyubiquitin chains of RIP. The interaction between NEMO and polyubiquitinated RIP is mediated by the NEMO ubiquitin-binding domain. After recruitment of RIP, TRAF2, and cIAP1, additional molecules are believed to be recruited to TNF-R.

Ubiquitination plays a nonproteolytic role in activation of IKK through a pool of free NEMO as well. Such a role for ubiquitination of NEMO was discovered in studies of cellular response to DNA-damaging agents. It was previously known that a protein kinase called ataxia telangiectasia mutated (ATM) kinase recognizes double-stranded breaks and stimulates NF-κB pathway through an RIP- and NEMO-dependent mechanism.[338] It was puzzling how a nuclear event could trigger activation of IKK, which is a cytoplasmic kinase. It was found that under conditions that induce DNA damage NEMO is sumoylated and the sumoylated NEMO accumulates in the nucleus.[339] Under these conditions, NEMO did not associate with the other subunits of the IKK complex suggesting that a free pool of NEMO independent of IKK exists in the cell. How does NEMO translocate back to the cytoplasm? The signal for export of NEMO from the nucleus to the cytoplasm is monoubiquitination of NEMO. The trigger for monoubiquitination of NEMO is phosphorylation of a specific site in NEMO (serine 85).[340]

In the nervous system, K63-linked ubiquitination regulates the function of TrkA, the receptor for nerve growth factor (NGF).[341] TrkA residing at the plasma membrane promotes neuronal survival, whereas internalized TrkA induces neuronal differentiation. It was observed that ubiquitination of TrkA was required for signaling because preventing ubiquitination blocked neuronal differentiation upon NGF stimulation. Different types of ubiquitination of TrkA appear to have different effects on TrkA. For example, polyubiquitination of TrkA by the E3 ligase Nedd4-2 is critical for downregulation of the receptor through degradation but not for internalization.[342] It is possible that K63-linked ubiquitination facilitates internalization of TrkA, whereas polyubiquitination (presumably through K-48 linkage) assists in degradation of the receptor.[343]

5.21.16 Physiological Functions of the Ubiquitin–Proteasome Pathway

Ubiquitin–proteasome-mediated proteolysis plays a critical role in many physiological processes including cell cycle and development. In the nervous system, proteolysis plays a role in the formation of synaptic connections during development as well as in alteration of synaptic strength (synaptic plasticity) that occurs during learning and memory. The role of proteolysis in cell cycle and synaptic plasticity will be considered in detail below.

5.21.16.1 Role of Ubiquitin–Proteasome-Mediated Proteolysis in the Cell Cycle

Cell cycle is one of the physiological processes in which the role of ubiquitin–proteasome-mediated proteolysis is well established. With the advent of yeast mutants that interfered with various phases of the cell cycle, cyclin-dependent kinases (Cdks) were found to have a critical role in regulating the cell cycle. Typically, Cdks activated by regulatory proteins are known as cyclins. Different Cdk–cyclin complexes are formed at specific stages of the cell cycle such as the S-phase (in which DNA synthesis occurs) and the metaphase. The transition from metaphase to anaphase depends on degradation of cyclins.[344,345] Systematic biochemical studies showed that cyclins were substrates for the ubiquitin–proteasome pathway (**Table 5**).

In higher eukaryotes, at the onset of S phase cyclin A accumulates which stimulates DNA synthesis. The amount of cyclin A continues to be high after the S phase because of its role in chromosome condensation. Cyclin A is degraded when cells enter prometaphase.[358] The level of another cyclin called cyclin B rises during G2 phase, which helps to complete the chromosome condensation and spindle assembly, which allow transition to metaphase. Cyclin B is degraded by APC during metaphase.[359]

APC contributes to the regulation of different phases of the cell cycle, which depends precisely on degrading the substrates in order. How does the precise ordering of degradation of substrates come about? *In vitro* ubiquitination experiments indicate that ordering of substrate degradation probably occurs because of the way in which the APC attaches ubiquitins to its substrates. APC links multiple ubiquitins (polyubiquitination) to some substrates in one step. In other words, such substrates need to bind to APC only once. For example, geminin, a negative regulator of DNA synthesis during the S phase is polyubiquitinated in one step. By contrast, certain other substrates of APC are ubiquitinated sequentially such that polyubiquitination requires several rounds of ubiquitin attachment and therefore requires APC to bind to the substrates multiple times. Cyclin A is one such substrate.[360]

A recent study showed that several APC substrates contain a motif comprising threonine (T), glutamate (E), and lyinse (K) residues that has been termed the TEK-box.[324] A similar TEK-box is also present around Lys11 of ubiquitin. The TEK-box is believed to be a new interaction motif critical for recognizing APC substrates for ubiquitination. It is important to note that TEK-box-containing substrates receive a K11 ubiquitin linkage as opposed to the more traditional K48 linkage.[324]

5.21.16.1.1 Regulation of APC-mediated degradation

Unlike ligases such as SCF, APC is generally known to degrade its substrates regardless of their posttranslational modification. Recent studies, however, indicate that at least in some instances substrate phosphorylation protects them from being marked for degradation by APC. For example, in human cells CD6, a protein critical for replication is phosphorylated by the kinase cyclin E-CDK2. This phosphorylation masks a stretch of amino acids that constitute the destruction box (D-box) that is recognized by APC.[344,361]

Table 5 Key APC substrates degraded during the cell cycle

Phase of the cell cycle	*Substrate*	*Function*	*Reference(s)*
Prophase	Cyclin A	Stimulation of DNA synthesis	346
	NEK2A	Protein phosphorylation	347
	HOXC10	Transcription	348
Metaphase	Cyclin B	Chromosome condensation and spindle assembly	349
	Securin	Inhibition of the protease separase	350
	CDKN1B (Kip 1)	Inhibition of Cdks	351
	Geminin	Inactivation of replication Factor CTD1	352
Anaphase	TPX2	Localization of Aurora A kinase to mitotic spindles	353
Exit from mitosis	Plk1	Spindle assembly	354
	Aurora A	Spindle formation and separation of centrosomes	355
	Aurora B	Coordination of chromosome segregation and cytokinesis	356, 357
	Annilin	An actin-binding protein required for cytokinesis	357

Another mechanism by which APC is regulated during cell cycle is through spatial restriction of the ligase to subcellular compartments. Although a large portion of APC is present in the cytoplasm, APC has been detected in kinetochores, the structures on chromosomes where spindle fibers attach for pulling the chromosomes apart during cell division. APC is also found in centrosomes and in the microtubules of spindle fibers. As in other systems such as the nervous system, local regulation of ligases might also occur during cell cycle. For example, in some mutants of *Drosophila* in which the centrosome can dissociate from spindle fibers, degradation of cyclin B occurs in the detached centrosomes but not on spindles.[361]

5.21.16.2 Ubiquitin–Proteasome Pathway and Synaptic Plasticity

Although ubiquitin was used as a marker for brain pathology, no physiological or pathological role for ubiquitin in the nervous system was found until about a decade ago. The first discovery of ubiquitin–proteasome-mediated degradation of a physiologically relevant substrate in the nervous system was that of R subunits of PKA.[69] Since then several substrates of the ubiquitin–proteasome pathway in the nervous system have been identified (**Table 6**).

5.21.16.2.1 Degradation of the R subunits of PKA

Evidence of a role for the ubiquitin–proteasome pathway in synaptic plasticity was derived from an investigation on persistent activation of PKA. Work on the biochemical mechanism of long-term facilitation[372] in *Aplysia* indicated that PKA was persistently activated in the absence of elevated cAMP. How is PKA activated in the absence of sustained increase in cAMP? It was found that the R subunits of PKA were decreased without any change in the catalytic (C) subunit during induction of long-term facilitation. Since there was no change in mRNA for either the R subunit or the C subunit, it was concluded that R subunits were diminished perhaps through proteolysis. What is the mechanism of R subunit degradation? Hegde *et al.*[69] found through a series of biochemical experiments that R subunits were substrates for ubiquitination and proteasome-mediated degradation.

Moreover, an UCH (Ap-uch) that interacts with the proteasome was found to be induced by 5-HT, the neurotransmitter that induced long-term facilitation. Ap-uch was found to be critical for the induction of long-term facilitation.[71] Subsequently, Chain *et al.*[373] showed that at sensory–motor neuron synapses, injection of lactacystin, a specific proteasome inhibitor blocked induction of long-term facilitation. Since R subunit inhibits the activity of C subunits of PKA, the results were interpreted to suggest that the ubiquitin–proteasome pathway operates to remove inhibitory constraints on the formation of long-term memory. This has been

Table 6 Molecules of the ubiquitin–proteasome pathway and their physiological function in the nervous system

Molecule	*Description*	*Reference(s)*
Substrates		
R subunit	Subunit of cAMP-dependent protein kinase crucial for synaptic plasticity in *Aplysia*	69
GluR1	Transmitter receptor critical for regulating synaptic strength	209
Synaptophysin	Presynaptic protein with a role in transmitter release	362
Ubiquitin-conjugating enzymes (E2s)		
Bendless	E2 that controls synapse formation in *Drosophila*	363, 364
UbcD1	E2 that controls dendritic pruning in *Drosophila*	365
Ubiquitin ligases		
E6-AP	Ubiquitin ligase (E3) important for synaptic plasticity, contextual memory	366
APC	A multisubunit RING-finger ligase critical for axon morphogenesis, synaptic size, and activity	367–369
LIN23	The substrate-binding component (F-box) of SCF ligase that regulated glutamate receptor abundance	370
Deubiquitinating enzymes (DUBs)		
Ap-uch	Deubiquitinating enzyme with a role in long-term synaptic plasticity in *Aplysia*	71
Fat facets	Deubiquitinating enzyme controlling synapse development in *Drosophila*	371

corroborated by work carried out on the rat hippocampus. Lopez-Salon *et al.*[374] demonstrated that bilateral infusion of lactacystin to the CA1 region of the rat hippocampus caused total retrograde amnesia for a one-trial avoidance learning. They also showed that total ubiquitination increases in the hippocampus 4 h after the training. These results are consistent with the idea that a decrease in some critical inhibitory proteins during long-term memory formation is mediated by the ubiquitin–proteasome pathway.

5.21.17 Diseases Associated with the Ubiquitin–Proteasome Pathway

Many diseases have been linked to components of the ubiquitin–proteasome pathway. These diseases affect several tissues and systems of the body. Major diseases connected to ubiquitin–proteasome-mediated proteolysis are different types of cancer and numerous diseases and disorders of the brain.

5.21.17.1 Cancer

Given that proteolysis by the ubiquitin–proteasome pathway plays a critical role in the control of cell proliferation, it is expected to have a role in cancer, which is basically uncontrolled cell proliferation. Many substrates, UBCs, ubiquitin ligases, DUBs, and the proteasome are all implicated in cancer pathogenesis (**Table 7**).

Among the substrates of proteolysis connected to cancer are many transcription factors that were discovered as tumor promoters (oncogenes) or tumor suppressors. For example, c-myc, c-fos, and c-jun are oncogenes linked to many types of cancer. c-Myc is mutated or translocated in B-cell lymphomas.[393] C-fos and c-jun together constitute AP-1 family of transcription factors that are thought to play a role in colorectal cancer.[394] Protein products of all these genes are known to be degraded in an ubiquitin–proteasome-dependent manner.[269] Protein kinases that are cell surface receptors such as EGF-R as well as cytoplasmic protein kinases such as Src are substrates for ubiquitin-mediated degradation.[188,395]

Among the tumor suppressor genes, p53 is one of the most studied because of its role in cell cycle progression. In normal cells, basal levels of nuclear p53 are regulated by the RING-finger ligase Mdm2.[396,397] In addition to Mdm2, other ligases such as CHIP, COP1, and PriH2 also ubiquitinate p53 and target it to degradation by the proteasome.[398] Modifications of p53 through attachment of ubiquitin-like molecules such as SUMO and NEDD appear to finely regulate ubiquitin-mediated degradation of p53,[399] although the precise fashion in which they regulate p53 degradation and functional consequence of these modifications to p53 function is not clear. For example, sumoylation has been shown to activate as well as repress p53-mediated transcription.

Table 7 The components of the ubiquitin–proteasome pathway implicated in cancer

Protein	*Description*	*Cancer*	*Reference(s)*
Substrates			
p53	Tumor suppressor/transcription factor	Non-small-cell lung cancer, colorectal cancer	375, 376
p27 (Kip)	Cyclin-dependent kinase inhibitor	Malignant melanoma, lymphoma	377, 378
Ligases			
β-TrCP	F-Box protein; substrate-recognizing part of SCF ligase	Breast cancer, colon cancer, pancreatic cancer, prostrate cancer	379–382
Skp2	Adaptor subunit of SCF ligase	Breast cancer, colon cancer, cervical cancer, lung cancer	383–390
DUBs			
HAUSP	Deubiquitinated p53	Non-small-cell lung cancer	391
Ubiquitin-like proteins			
PIAS3	SUMO (E3) ligase	Lung cancer, breast cancer, prostrate cancer, colorectal cancer, brain tumor	392

Among the ubiquitin ligases apart from Mdm2, APC and SCF ligases are linked to cancer. For example, APC is implicated in colorectal cancer. Components of SCF ligases, however, seem to be dysregulated in numerous types of cancer. For example, Skp2 is dysregulated in breast cancer, cervical cancer, lung cancer, multiple myeloma, ovarian cancer, and prostate cancer.[400–403] Regulation of β-TrCP is disrupted in breast cancer, colon cancer, gastric cancer, pancreatic cancer, and melanoma.[403–405] Since DUBs oppose the action of ubiquitin ligases it might be expected that DUBs play a role in some cancers as well. For example, down-regulation of HAUSP, a DUB which deubiquitinates p53, is implicated in non-small-cell lung cancer.[391] Several ubiquitin-like proteins such as SUMO, NEDD8, and ISG15 are implicated in different types of cancer as well.[406]

5.21.17.2 Nonnervous System Diseases Associated with the Ubiquitin–Proteasome Pathway

Anomalies in ubiquitin conjugation as well as destruction by the proteasome are linked to several diseases of the body (**Table 8**).

5.21.17.2.1 Cystic fibrosis

Cystic fibrosis is a disease that affects the lungs and the digestive system. It is caused by mutations in the *CFTR* gene.[416] The CFTR protein has 1480 amino acids and forms a chloride channel. Hundreds of mutations of the *CFTR* gene have been described. Mutations affect production of the CFTR proteins or any of the steps of folding, processing through the ER, and insertion into the plasma membrane. A common mutation that occurs in 70% of the cases is a deletion mutation (ΔF508).[417,418] Even though the chloride conductance of $CFTR^{\Delta F508}$ is normal, the disease is caused because the protein does not reach the plasma membrane.[419] Normally, the CFTR protein is cotranslationally inserted into the ER membrane, glycosylated in the golgi apparatus and then inserted into the plasma membrane. The ΔF508 mutation results in a misfolded CFTR protein, which is not processed through ER.[420] It is removed from the ER through retrotranslocation and degraded by the proteasome. Mutant proteins with cytoplasmic misfolding are bound to Hsp70. The ubiquitin ligase CHIP then recognizes the Hsp70-bound CFTR protein. The CHIP ligase ubiquitinates and targets the misfolded CFTR protein for degradation.[421,422] The E2 enzyme that assists the CHIP ligase has been identified to be UbcH5a.[122] Degradation of misfolded $CFTR^{\Delta F508}$ results in CFTR deficiency resulting in chronic airway obstruction in the lungs, mucus buildup, and infection. Another major symptom is defects in digestion owing to functional impairment of the pancreas.[417]

Table 8 Role of the ubiquitin–proteasome pathway in nonnervous system diseases

Disease/disorder	*Defective component*	*Reference(s)*
Substrates		
Liddle syndrome	Mutation in the β/γ-subunit of renal sodium channel	407
Cystic fibrosis	Misfolded CFTR protein	122
Fanconi anemia	Mutation in the FANCD2 protein	408
E3 ligase		
Von Hippel–Lindau (VHL) disease	Mutation in the VHL protein	409
DUBs		
Cylindramatosis	Mutation in CYLD	162, 164, 165
Proteasome		
Transient ischemia	Decreased activity of 26S proteasome	410, 411
Sjogren's syndrome	Autoimmune reactivity against proteasome subunits β1i and β5i	412
Hepatitis B	Inhibition of 20S and 26S proteasome	413
Cachexia	Increased activity of the 20S and 26S proteasome	414, 415

5.21.17.2.2 The Liddle syndrome

The Liddle syndrome is characterized by early-onset hypertension, which mimics the effect of excessive aldosterone produced by adrenal glands. The disorder is named after Grant Liddle who first described the symptoms of the disease in a young woman with severe hypertension without an increase in aldosterone.[423] The patient's relatives had this disorder as well indicating a genetic component, which was later determined to be autosomal dominant and the β-subunit of the renal epithelial sodium channel (β-ENaC) was found.[88,407] The connection to ubiquitin-mediated protein degradation and the disease was made with the discovery that stability of ENaC was determined by ubiquitination. ENaC has three subunits α, β, and γ. Only α and γ (but not the β-subunits) are ubiquitinated.[87] The ligase responsible for ENaC ubiquitination is Nedd4.[89] Mutations in the binding site of Nedd4 to the ENaC channel prevent Nedd4 from recognizing the channel. Therefore, the channel is stabilized and the increased number of ENaCs causes excessive reabsorption of Na^+ leading to severe hypertension.

5.21.17.3 Diseases of the Nervous System Linked to the Ubiquitin–Proteasome Pathway

Alterations in the ubiquitin–proteasome pathway have been connected to several neurodegenerative diseases (**Table 9**). In some instances, mutations in specific genes have been linked to the etiology of the disease. Although the perturbations in ubiquitin–proteasome-mediated proteolysis lead to pleiotropic effects on neurons including cell death or degeneration, one of the early effects is believed to be synaptic malfunction.

5.21.17.3.1 Alzheimer's disease

Alzheimer's disease (AD) is characterized by the appearance of neurofibrillary tangles and plaques at advanced stages. Although multiple genetic disturbances are believed to cause AD, however, a major cause of the disease is the buildup of the toxic Aβ peptide. One of the earliest symptoms of the disease is memory loss and cognitive dysfunction. Cognitive deficits are correlated with the loss of synapses.[437] Systematic quantitative studies have revealed that within 2–4 years of the onset of clinical manifestation of AD, there was 25–35% decrease in the density of synapses. Even at the advanced stages of the disease, loss of synapses is more robustly correlated with the disease than the plaques and tangles. Animal models of AD mimic cognitive impairment seen in AD. For example, in transgenic mice expressing mutant Alzheimer β-amyloid precursor protein (APP), learning and memory is impaired at 9–10 months of age although no tangles are seen in the brains of these mice.[438] In mice carrying various mutations of APP, both *in vitro* and *in vivo* LTP is impaired much before detectable Aβ deposits are observed. Besides, cerebral microinjection of oligomers of Aβ peptide inhibits *in vivo* LTP in rats.[439]

How does the ubiquitin–proteasome pathway contribute to pathogenesis of AD? It is thought that the causative factors in AD, namely, the Aβ peptide, or the paired helical filaments (PHF) of tau protein, impair proteasome function. In *in vitro* experiments, Aβ peptide has been shown to inhibit the proteasome. In the brains of AD patients, proteasome function has been shown to be reduced mostly in the areas critical for

Table 9 Ubiquitin–proteasome pathway and diseases/disorders of the nervous system

Disease	*Component linked to the disease*	*Reference(s)*
Alzhemer's disease	Ubiquitin (abnormal form)	424
	Inhibition of proteasome	425
Amyotrophic lateral sclerosis	Aggregated superoxide dismutase	426
Angelman syndrome	Mutation in E6-AP (UBE3A) ubiquitin ligase	427, 428
Ataxia	USP14 (DUB)	172
Gracile axonal dystrophy	Deletion in the UCH-L1 gene	429
Huntington's disease	Abnormal nuclear inclusions of ubiquitinated proteins	430
Parkinson's disease	Parkin (E3, ubiquitin ligase) UCH-L1	431–433
	α-Synuclein (substrate)	
Schizophrenia	Decreased expression of UCH-L1 and UBP14	434, 435
Wallerian degeneration	Possible microtubule fragmentation through the ubiquitin–proteasome pathway	436

long-term memory formation such as hippocampus, parahippocampal gyrus superior, and middle temporal gyri and inferior parietal lobule but not in other areas such as the occipital lobe.[440] A recent study showed that PHF of tau in brains of AD patients, as well as *in vitro* assembly of PHF using human recombinant tau protein, both inhibited proteasome activity.[425]

Taking the studies on early cognitive impairment in AD patients and animal models of AD together with the investigations on impairment of the ubiquitin–proteasome pathway in AD, one could argue that the early synaptic defects result from impairment of proteolysis. If proper ubiquitin–proteasome-mediated degradation at the synapse is required for plasticity as well as maintaining the integrity of the synapse, perturbations in these functions might first lead to synaptic dysfunction and eventually loss of synapses.

5.21.17.3.2 Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a disease of the neurons that control muscle movement (motor neurons). Degeneration of neurons causes muscle atrophy eventually impairing the movement of people afflicted with the disease.

Mutations of *superoxide dismutase1* (*SOD1*) have been associated with some cases of autosomal dominant familial ALS and some cases of sporadic ALS. The mutant SOD1 protein, unlike the wild type, is degraded by ubiquitin–proteasome pathway.[426,441] Overexpression of the putative SOD1 E3 ligase dorfin can inhibit cell death induced by the mutant SOD1 protein.[442] Gene-expression profiling of spinal cords from sporadic ALS patients indicated that genes associated with the ubiquitin–proteasome pathway (dorfin and ubiquitin-like protein 5), oxidative toxicity, transcription, neuronal differentiation, and inflammation might function in the pathogenesis of sporadic ALS.[443] It is possible that the expression of dorfin ligase is increased in an attempt to enhance clearance of the mutant SOD1. Recent studies also reported that heat-shock protein 70 (Hsp-70) or heat-shock cognate Hsc70, and CHIP play a role in proteasomal degradation of mutant SOD1.[444] In addition, Urushitani *et al.*[445] found that oxidative damage increases the degree of ubiquitination of mutant SOD1 protein and proteasome activity decreases following 1 week of expression of a mutant *SOD1* gene in cultured cells. Toxicity of mutant SOD1 protein aggregates is controversial, however. Lee *et al.*[446] found that aggregates of mutant SOD1 did not cause cell death. Besides, others found that motor neurons from wild-type animals and motor neurons from transgenic mice with mutant *SOD1* were equally viable.[447] Some studies, however, have demonstrated that proteasome inhibition results in increased cell death in human cells expressing mutant SOD1 protein.[448,449] The role of protein aggregates in familial ALS remains unclear. As with other neurodegenerative diseases, the interplay between SOD1 protein aggregates, ubiquitin–proteasome pathway, and disease progression remains to be elucidated.

5.21.17.3.3 Angelman syndrome

Angelman syndrome (AS) is a neurological disorder with symptoms such as mental retardation, unusually happy demeanor, susceptibility to epileptic seizures, and abnormal gait.[450] Occurrence of AS is estimated to be 1 in 15 000 births. In about 65–75% of AS patients, maternal deletions at chromosome 15q11–q13 are found. Other types of genetic abnormalities such as uniparental disomy and imprinting mutations are also observed in AS, each of which accounts for about 3–5% of the cases. It has been found that the defects occur in a gene called UBE3A.[427,428] Point mutations in UBE3A are found in about 4–6% of the AS cases.[366] UBE3A gene encodes an ubiquitin ligase that had been previously identified as E6-AP (E6-associated protein) ubiquitin ligase. E6-AP is the cellular protein that associates with a human papilloma virus protein called E6. E6-AP in association with E6 degrades the tumor suppressor protein p53.[451] Apart from p53, E6-AP is known to attach ubiquitin to at least three other substrates, RAD23, a human homologue of a yeast DNA repair protein;[452] multicopy maintenance protein 7 (MCM7) which is thought to function in chromosome replication;[453] and E6-AP itself.[454]

5.21.17.3.4 Huntington's disease

Huntington's disease (HD) is caused by mutations in a gene called Huntingtin. The disease is caused by abnormal expansion of CAG repeats, which encode long stretches of glutamine (polyglutamine). In addition to HD, there are several other known polyglutamine diseases such as spinocerebellar ataxia, and spinal and bulbar muscular atrophy.[84]

HD can also be viewed as a disease of the synapse. In HD patients, cognitive deficits appear much before the clinical symptoms of the disease.[455–459] This phenomenon is reproduced in mouse models of HD as well. For example, in transgenic mice carrying exon 1 of the human HD mutation, hippocampal plasticity is significantly altered, and deficits in spatial learning are observed. Deficits in synaptic plasticity occur before the overt neurological phenotype is observed.[460] Introduction of HD-like CAG repeats into murine huntingtin causes behavioral abnormalities as well.[461] Another study reported abnormalities in synaptic vesicle fusion machinery in HD mutant mice.[462] Even ectopic expression of N-terminal fragment with 150 glutamine residues in *Aplysia* neurons impaired long-term facilitation without affecting basal synaptic transmission or short-term facilitation.[463]

Is the ubiquitin–proteasome pathway impaired in HD? Altered proteasome function was observed when mutant huntingtin was expressed in Neuro2A cells under the control of an inducible promoter.[464] A convincing demonstration was obtained using an expression of huntingtin with a stretch of 103 glutamines in human embryonic kidney (HEK) 293 cells. Expression of mutant huntingtin with 103 glutamines in HEK 293 cells caused aggregate formation, accumulation of ubiquitinated proteins, and cell cycle arrest, whereas expression of shorter stretch of polyglutmaine (25 glutamines) had markedly less effect on all these parameters.[465]

How do polyglutamine-containing proteins inhibit the proteasome? One explanation is that polyglutamine-containing proteins act as direct inhibitors of the proteasome.[84] An alternative explanation is that the ubiquitin–proteasome pathway is overwhelmed by the load of aggregated or misfolded proteins. The nuclear inclusions in HD show ubiquitin immunoreactivity, suggesting that perhaps ubiquitin conjugation to aggregated protein is taking place and the proteasome is unable to efficiently degrade the ubiquitinated proteins. In support of this idea, protein inclusions from isolated neurons and brains of conditional HD transgenic mice can be recovered if the transgene is turned off. The disappearance of inclusion is proteasome dependent because the proteasome inhibitor lactacystin inhibits the reversal process.[466] In addition, when green fluorescent protein containing polyglutamine is expressed in SH-SY5Y cells, basal proteasome activity is not impaired but the ability of the proteasome to respond to stress such as heat shock is dramatically impaired.[467]

Based on the results from various studies described above, it is clear that the failure of the ubiquitin–proteasome pathway ultimately contributes to cell death or degeneration in polyglutamine diseases. This process is probably progressive. Initial formation of aggregates with proteins with long polyglutamine stretches perhaps impairs proteasome activity, which in turn leads to accumulation of more protein aggregates, and thus misfolded proteins could build up in the cell. Initially, however, impairment of the proteasome probably has an effect on synaptic properties of the neuron because as discussed elsewhere in this chapter, proteasome activity in various subcellular compartments of the neuron is essential for normal synaptic function and plasticity.

5.21.17.3.5 Parkinson's disease

Four different genes have been linked to Parkinson's disease (PD): *α-synuclein*, *UCH-L1*, *DJ-1*, and *parkin*.[431,432,468–470] Protein products of all these genes are linked to the ubiquitin–proteasome pathway. α-Synuclein is part of the Lewy bodies, the intracellular inclusions seen in the brains of Parkinson's disease patients. Lewy bodies contain high amounts of ubiquitinated proteins including ubiquitinated α-synuclein. UCH-L1 is an enzyme of the ubiquitin pathway. DJ-1 is a substrate for ligation to SUMO, a protein related to ubiquitin. The amino acid sequence of parkin protein contains an UbL domain at its N-terminus. In addition, parkin has been shown to be an ubiquitin ligase. Three of the genes linked to Parkinson's disease have direct connection to synapses. Parkin is present in postsynaptic densities. The *Aplysia* homologue of UCH-L1 is present in presynaptic terminals (A. N. Hegde *et al.*, unpublished observations). Synuclein is associated with synaptic vesicles.[471]

Recent evidence indicates that PD might affect synaptic function adversely. In parkin-deficient mice, higher stimulation of corticostriatal afferents were required to evoke EPSPs in striatal spiny neurons suggesting that excitability of spiny neurons was reduced.[472] Also, parkin, which is an ubiquitin ligase[473] has been shown to interact with and ubiquitinate synaptotagmin XI.[474]

5.21.17.3.6 Spinocerebellar ataxias

The key symptoms of spinocerebellar ataxias (SCAs) include loss of balance, motor coordination, and malfunction of the cerebellum. There are several types of SCAs of which types 1, 2, 3, 6, 7, and 17 are

CAG-repeat diseases in that the genes responsible have long tracts (40 or more) of CAGs. The proteins encoded by the genes causing SCAs therefore possess long polyglutamine tracts. As with HD, pathological symptoms of SCAs are believed to result from misfolding of proteins and accumulation of which causes neuronal dysfunction followed by degeneration. SCAs present diverse clinical symptoms and accordingly the genes implicated in SCAs differ from each other with respect to structure and function. The genes linked to SCAs 1, 2, 3, and 7 are called *ataxin-1*, *ataxin-2*, *ataxin-3*, and *ataxin-7*. Among these ataxins 1, 3, and 7 are substrates for ubiquitin–proteasome-mediated degradation. Recent studies implicate another kind of link between the ubiquitin–proteasome pathway and SCA 3 (also called Machado–Joseph disease). The ataxin-3 protein has been shown to be a DUB and therefore it is highly probable that ubiquitin–proteasome-mediated degradation is somehow linked to pathophysiology of SCAs. Ataxin-3 has a conserved N-terminal Josephin domain (named after the Machado–Joseph disease) and two ubiquitin-interacting motifs (UIMs) and interacts with the human homologue of Rad23 (HHR23), a protein that is known to translocate substrates to the proteasome. The solution structure of the Josephin domain indicates that the surface that interacts with HHR23 also is the part of ataxin-3 that binds to UBA domains and the proteasome. The expansion of polyglutamine in ataxin-3 might disrupt the activity of the proteasome and thus interfere with proteolysis.

5.21.18 Future Perspectives

Despite enormous progress made on various aspects of the ubiquitin–proteasome pathway since its discovery in the 1980s, many important questions remain unanswered. For example, scanty information is known regarding the spatial and temporal regulation of protein degradation in many organ systems such as the nervous system. In addition, the relationship between proteolysis and various diseases is not understood in mechanistic terms. Besides, the development of genetically modified animals that carry selective alterations in components of the ubiquitin–proteasome pathway has not progressed much. Many exciting discoveries might be expected in the years ahead with respect to the physiological and pathological roles of the ubiquitin–proteasome pathway.

Abbreviations

AAA	ATPases associated with different cellular activities
AD	Alzheimer's disease
ALS	amyotrophic lateral sclerosis
AMPA	α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid
APC	anaphase-promoting complex
Ap-C/EBP	*Aplysia* CCAAT enhancer binding protein
Ap-CAM	*Aplysia* homologue of neural cell adhesion molecule
APF-1	ATP-dependent proteolysis factor
APP	amyloid precursor protein
Ap-uch	*Aplysia* ubiquitin C-terminal hydrolase
AS	Angelman syndrome
BRUCE	BIR-repeat-containing ubiquitin-conjugating enzyme
BTB	broad-complex/Tramtrack/Bric-a-brac
CPEB	cytoplasmic polyadenylation element binding protein
CREB	cAMP-responsive element binding protein
CSN	COP9 signalosome
DUB	deubiquitinating enzyme
E6	a protein encoded by oncogenic strains of human papilloma virus
E6-AP	E6-associated protein
EGFR	epidermal growth factor receptor
ENaC	epithelial sodium channel
Eps15	EGFR pathway substrate clone 15
ESCRT-I	endosomal sorting complex required for transport

GPCR	G-protein-coupled receptor
H2A	histone 2A
H2B	histone 2B
HD	Huntington's disease
HECT	homologous to E6-AP C-terminus
LIN-23	abnormal cell lineage protein 23
LTP	long-term potentiation
MPN	Mpr1p/Pad1p N-terminus
NEDD	neuronal precursor cell-expressed developmentally downregulated 4
NEMO	NF-κB essential modulator
NGF	nerve growth factor
PCI	proteasome, COP9, initiation factor 3
PD	Parkinson's disease
PDGFR	platelet-derived growth factor receptor
PIM	proteasome-interacting motif
PKA	cAMP-dependent protein kinase
PML	promyelocytic leukemia protein
POD	PML oncogenic domain
PSD	postsynaptic density
R subunit	regulatory subunit
RING	really interesting new gene
Rub1	related to ubiquitin 1
19S RC	19S regulatory complex
SAGA	Spt-Ada-Gcn5-acetyltransferase
SCA	spinocerebellar ataxia
SCF complex	Skp1-cullin-F-Box protein complex
SENP	sentrin-specific proteases
SOD1	superoxide dismutase1
SUMO	small ubiquitin-related modifier
tPA	tissue plasminogen activator
UBC	ubiquitin-conjugating enzyme
UbL	ubiquitin-like protein
UBP	ubiquitin-specific protease
UCH	ubiquitin C-terminal hydrolase
UIM	ubiquitin-interacting motif
V2-VR	V2-type vasopressin receptor
β2-AR	β2-adrenergic receptor

References

1. A. Hershko; A. Ciechanover, *Annu. Rev. Biochem.* **1992**, *61*, 761–807.
2. J. D. Etlinger; A. L. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 54–58.
3. A. Ciehanover; Y. Hod; A. Hershko, *Biochem. Biophys. Res. Commun.* **1978**, *81*, 1100–1105.
4. A. Hershko; A. Ciechanover; I. A. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3107–3110.
5. A. Hershko; A. Ciechanover; H. Heller; A. L. Haas; I. A. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1783–1786.
6. K. D. Wilkinson; M. K. Urban; A. L. Haas, *J. Biol. Chem.* **1980**, *255*, 7529–7532.
7. A. Hershko, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 5932–5943.
8. A. Ciechanover, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 5944–5967.
9. I. Rose, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 5926–5931.
10. G. Goldstein; M. Scheid; U. Hammerling; D. H. Schlesinger; H. D. Niall; E. A. Boyse, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 11–15.

11. D. Finley; V. Chau, *Annu. Rev. Cell Biol.* **1991**, *7*, 25–69.
12. A. N. Hegde; B. M. Broome; M. Qiang; J. H. Schwartz, *Brain Res. Mol. Brain Res.* **2000**, *76*, 424–428.
13. R. T. Baker; P. G. Board, *Nucleic Acids Res.* **1987**, *15*, 443–463.
14. J. Swindle; J. Ajioka; H. Eisen; B. Sanwal; C. Jacquemot; Z. Browder; G. Buck, *EMBO J.* **1988**, *7*, 1121–1127.
15. R. T. Baker; P. G. Board, *Nucleic Acids Res.* **1991**, *19*, 1035–1040.
16. M. H. Glickman; A. Ciechanover, *Physiol Rev.* **2002**, *82*, 373–428.
17. A. N. Hegde; A. DiAntonio, *Nat. Rev. Neurosci.* **2002**, *3*, 854–861.
18. A. Hershko; A. Ciechanover, *Annu. Rev. Biochem.* **1998**, *67*, 425–479.
19. P. M. Handley; M. Mueckler; N. R. Siegel; A. Ciechanover; A. L. Schwartz, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 258–262.
20. J. P. McGrath; S. Jentsch; A. Varshavsky, *EMBO J.* **1991**, *10*, 227–236.
21. A. G. Stephen; J. S. Trausch-Azar; A. Ciechanover; A. L. Schwartz, *J. Biol. Chem.* **1996**, *271*, 15608–15614.
22. C. M. Pickart; I. A. Rose, *J. Biol. Chem.* **1985**, *260*, 1573–1581.
23. H. P. Hauser; M. Bardroff; G. Pyrowolakis; S. Jentsch, *J. Cell Biol.* **1998**, *141*, 1415–1422.
24. T. Biederer; C. Volkwein; T. Sommer, *EMBO J.* **1996**, *15*, 2069–2076.
25. R. K. Plemper; D. H. Wolf, *Trends Biochem. Sci.* **1999**, *24*, 266–270.
26. N. W. Bays; R. G. Gardner; L. P. Seelig; C. A. Joazeiro; R. Y. Hampton, *Nat. Cell Biol.* **2001**, *3*, 24–29.
27. T. Biederer; C. Volkwein; T. Sommer, *Science* **1997**, *278*, 1806–1809.
28. J. M. Huibregtse; M. Scheffner; P. M. Howley, *Mol. Cell Biol.* **1993**, *13*, 775–784.
29. J. M. Huibregtse; M. Scheffner; S. Beaudenon; P. M. Howley, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2563–2567.
30. D. Rotin, *Curr. Opin. Nephrol. Hypertens.* **2000**, *9*, 529–534.
31. K. L. Lorick; J. P. Jensen; S. Fang; A. M. Ong; S. Hatakeyama; A. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11364–11369.
32. M. Tyers; A. R. Willems, *Science* **1999**, *284*, 601, 603–604.
33. P. K. Jackson; A. G. Eldridge; E. Freed; L. Furstenthal; J. Y. Hsu; B. K. Kaiser; J. D. Reimann, *Trends Cell Biol.* **2000**, *10*, 429–439.
34. C. A. Joazeiro; A. M. Weissman, *Cell* **2000**, *102*, 549–552.
35. J. Miller; A. D. McLachlan; A. Klug, *EMBO J.* **1985**, *4*, 1609–1614.
36. A. J. Saurin; K. L. Borden; M. N. Boddy; P. S. Freemont, *Trends Biochem. Sci.* **1996**, *21*, 208–214.
37. A. Hengstermann; L. K. Linares; A. Ciechanover; N. J. Whitaker; M. Scheffner, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1218–1223.
38. P. Beer-Romero; S. Glass; M. Rolfe, *Oncogene* **1997**, *14*, 595–602.
39. M. Traidej; L. Chen; D. Yu; S. Agrawal; J. Chen, *Antisense Nucleic Acid Drug Dev.* **2000**, *10*, 17–27.
40. S. Hietanen; S. Lain; E. Krausz; C. Blattner; D. P. Lane, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8501–8506.
41. S. C. Upadhya; A. N. Hegde, *Trends Biochem. Sci.* **2003**, *28*, 280–283.
42. E. Sakata; Y. Yamaguchi; E. Kurimoto; J. Kikuchi; S. Yokoyama; S. Yamada; H. Kawahara; H. Yokosawa; N. Hattori; Y. Mizuno; K. Tanaka; K. Kato, *EMBO Rep.* **2003**, *4*, 301–306.
43. P. Zhou; P. M. Howley, *Mol. Cell* **1998**, *2*, 571–580.
44. J. M. Galan; M. Peter, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9124–9129.
45. R. Visintin; S. Prinz; A. Amon, *Science* **1997**, *278*, 460–463.
46. H. H. Lim; P. Y. Goh; U. Surana, *Curr. Biol.* **1998**, *8*, 231–234.
47. M. Shirayama; W. Zachariae; R. Ciosk; K. Nasmyth, *EMBO J.* **1998**, *17*, 1336–1349.
48. H. Yu; J. M. Peters; R. W. King; A. M. Page; P. Hieter; M. W. Kirschner, *Science* **1998**, *279*, 1219–1222.
49. A. M. Page; P. Hieter, *Annu. Rev. Biochem.* **1999**, *68*, 583–609.
50. M. Koegl; T. Hoppe; S. Schlenker; H. D. Ulrich; T. U. Mayer; S. Jentsch, *Cell* **1999**, *96*, 635–644.
51. S. Hatakeyama; M. Yada; M. Matsumoto; N. Ishida; K. I. Nakayama, *J. Biol. Chem.* **2001**, *276*, 33111–33120.
52. L. Aravind; E. V. Koonin, *Curr. Biol.* **2000**, *10*, R132–R134.
53. M. D. Ohi; C. W. Vander Kooi; J. A. Rosenberg; W. J. Chazin; K. L. Gould, *Nat. Struct. Biol.* **2003**, *10*, 250–255.
54. C. Patterson, *Sci. STKE* **2002**, *2002*, E4.
55. G. N. Patrick; P. Zhou; Y. T. Kwon; P. M. Howley; L. H. Tsai, *J. Biol. Chem.* **1998**, *273*, 24057–24064.
56. F. Margottin; S. P. Bour; H. Durand; L. Selig; S. Benichou; V. Richard; D. Thomas; K. Strebel; R. Benarous, *Mol. Cell* **1998**, *1*, 565–574.
57. A. Yaron; A. Hatzubai; M. Davis; I. Lavon; S. Amit; A. M. Manning; J. S. Andersen; M. Mann; F. Mercurio; Y. Ben-Neriah, *Nature* **1998**, *396*, 590–594.
58. M. Hart; J. P. Concordet; I. Lassot; I. Albert; S. R. del los; H. Durand; C. Perret; B. Rubinfeld; F. Margottin; R. Benarous; P. Polakis, *Curr. Biol.* **1999**, *9*, 207–210.
59. M. Karin; Y. Ben-Neriah, *Annu. Rev. Immunol.* **2000**, *18*, 621–663.
60. E. Schwob; T. Bohm; M. D. Mendenhall; K. Nasmyth, *Cell* **1994**, *79*, 233–244.
61. R. Verma; R. S. Annan; M. J. Huddleston; S. A. Carr; G. Reynard; R. J. Deshaies, *Science* **1997**, *278*, 455–460.
62. D. Kornitzer; A. Ciechanover, *J. Cell Physiol.* **2000**, *182*, 1–11.
63. P. Nash; X. Tang; S. Orlicky; Q. Chen; F. B. Gertler; M. D. Mendenhall; F. Sicheri; T. Pawson; M. Tyers, *Nature* **2001**, *414*, 514–521.
64. N. Yamamoto; A. N. Hegde; D. G. Chain; J. H. Schwartz, *J. Neurochem.* **1999**, *73*, 2415–2423.
65. A. M. Musti; M. Treier; D. Bohmann, *Science* **1997**, *275*, 400–402.
66. M. A. Schwarzschild; R. L. Cole; S. E. Hyman, *J. Neurosci.* **1997**, *17*, 3455–3466.
67. M. Nishizawa; K. Okazaki; N. Furuno; N. Watanabe; N. Sagata, *EMBO J.* **1992**, *11*, 2433–2446.
68. M. Nishizawa; N. Furuno; K. Okazaki; H. Tanaka; Y. Ogawa; N. Sagata, *EMBO J.* **1993**, *12*, 4021–4027.
69. A. N. Hegde; A. L. Goldberg; J. H. Schwartz, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7436–7440.
70. D. G. Chain; A. N. Hegde; N. Yamamoto; B. Liu-Marsh; J. H. Schwartz, *J. Neurosci.* **1995**, *15*, 7592–7603.
71. A. N. Hegde; K. Inokuchi; W. Pei; A. Casadio; M. Ghirardi; D. G. Chain; K. C. Martin; E. R. Kandel; J. H. Schwartz, *Cell* **1997**, *89*, 115–126.

72. C. Gieffers; B. H. Peters; E. R. Kramer; C. G. Dotti; J. M. Peters, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11317–11322.
73. S. Lahav-Baratz; V. Sudakin; J. V. Ruderman; A. Hershko, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9303–9307.
74. G. Kauselmann; M. Weiler; P. Wulff; S. Jessberger; U. Konietzko; J. Scafidi; U. Staubli; J. Bereiter-Hahn; K. Strebhardt; D. Kuhl, *EMBO J.* **1999**, *18*, 5528–5539.
75. G. Levkowitz; H. Waterman; E. Zamir; Z. Kam; S. Oved; W. Y. Langdon; L. Beguinot; B. Geiger; Y. Yarden, *Genes Dev.* **1998**, *12*, 3663–3674.
76. G. Levkowitz; H. Waterman; S. A. Ettenberg; M. Katz; A. Y. Tsygankov; I. Alroy; S. Lavi; K. Iwai; Y. Reiss; A. Ciechanover; S. Lipkowitz; Y. Yarden, *Mol. Cell* **1999**, *4*, 1029–1040.
77. N. Hisamoto; K. Sugimoto; K. Matsumoto, *Mol. Cell Biol.* **1994**, *14*, 3158–3165.
78. S. Kotani; S. Tugendreich; M. Fujii; P. M. Jorgensen; N. Watanabe; C. Hoog; P. Hieter; K. Todokoro, *Mol. Cell* **1998**, *1*, 371–380.
79. H. Yamano; J. Gannon; T. Hunt, *EMBO J.* **1996**, *15*, 5268–5279.
80. H. Yamada; K. Kumada; M. Yanagida, *J. Cell Sci.* **1997**, *110* (Pt. 15), 1793–1804.
81. G. C. Turner; F. Du; A. Varshavsky, *Nature* **2000**, *405*, 579–583.
82. M. Hochstrasser, *Nat. Cell Biol.* **2000**, *2*, E153–E157.
83. T. Hadari; J. V. Warms; I. A. Rose; A. Hershko, *J. Biol. Chem.* **1992**, *267*, 719–727.
84. A. Ciechanover; P. Brundin, *Neuron* **2003**, *40*, 427–446.
85. J. M. Peters; R. W. King; C. Hoog; M. W. Kirschner, *Science* **1996**, *274*, 1199–1201.
86. D. T. Pak; M. Sheng, *Science* **2003**, *302*, 1368–1373.
87. O. Staub; I. Gautschi; T. Ishikawa; K. Breitschopf; A. Ciechanover; L. Schild; D. Rotin, *EMBO J.* **1997**, *16*, 6325–6336.
88. H. Abriel; J. Loffing; J. F. Rebhun; J. H. Pratt; L. Schild; J. D. Horisberger; D. Rotin; O. Staub, *J. Clin. Invest.* **1999**, *103*, 667–673.
89. H. Abriel; E. Kamynina; J. D. Horisberger; O. Staub, *FEBS Lett.* **2000**, *466*, 377–380.
90. T. W. Abrams; Y. Yovell; C. U. Onyike; J. E. Cohen; H. E. Jarrard, *Learn. Mem.* **1998**, *4*, 496–509.
91. C. U. Onyike; A. H. Lin; T. W. Abrams, *J. Neurochem.* **1998**, *71*, 1298–1306.
92. J. Lowe; D. Stock; B. Jap; P. Zwickl; W. Baumeister; R. Huber, *Science* **1995**, *268*, 533–539.
93. M. Groll; L. Ditzel; J. Lowe; D. Stock; M. Bochtler; H. D. Bartunik; R. Huber, *Nature* **1997**, *386*, 463–471.
94. M. Bochtler; L. Ditzel; M. Groll; C. Hartmann; R. Huber, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 295–317.
95. G. Fenteany; R. F. Standaert; W. S. Lane; S. Choi; E. J. Corey; S. L. Schreiber, *Science* **1995**, *268*, 726–731.
96. C. L. Silva; F. C. Portaro; V. L. Bonato; A. C. de Camargo; E. S. Ferro, *Biochem. Biophys. Res. Commun.* **1999**, *255*, 591–595.
97. N. Tamura; F. Lottspeich; W. Baumeister; T. Tamura, *Cell* **1998**, *95*, 637–648.
98. T. Yao; R. E. Cohen, *Curr. Biol.* **1999**, *9*, R551–R553.
99. M. Rechsteiner; L. Hoffman; W. Dubiel, *J. Biol. Chem.* **1993**, *268*, 6065–6068.
100. O. Coux; K. Tanaka; A. L. Goldberg, *Annu. Rev. Biochem.* **1996**, *65*, 801–847.
101. G. N. DeMartino; C. R. Moomaw; O. P. Zagnitko; R. J. Proske; M. Chu-Ping; S. J. Afendis; J. C. Swaffield; C. A. Slaughter, *J. Biol. Chem.* **1994**, *269*, 20878–20884.
102. W. Dubiel; K. Ferrell; M. Rechsteiner, *Mol. Biol. Rep.* **1995**, *21*, 27–34.
103. D. Finley; K. Tanaka; C. Mann; H. Feldmann; M. Hochstrasser; R. Vierstra; S. Johnston; R. Hampton; J. Haber; J. Mccusker; P. Silver; L. Frontali; P. Thorsness; A. Varshavsky; B. Byers; K. Madura; S. I. Reed; D. Wolf; S. Jentsch; T. Sommer; W. Baumeister; A. Goldberg; V. Fried; D. M. Rubin; A. Toh-e, *Trends Biochem. Sci.* **1998**, *23*, 244–245.
104. M. H. Glickman; D. M. Rubin; V. A. Fried; D. Finley, *Mol. Cell Biol.* **1998**, *18*, 3149–3162.
105. H. Holzl; B. Kapelari; J. Kellermann; E. Seemuller; M. Sumegi; A. Udvardy; O. Medalia; J. Sperling; S. A. Muller; A. Engel; W. Baumeister, *J. Cell Biol.* **2000**, *150*, 119–130.
106. C. Richmond; C. Gorbea; M. Rechsteiner, *J. Biol. Chem.* **1997**, *272*, 13403–13411.
107. C. Gorbea; D. Taillandier; M. Rechsteiner, *J. Biol. Chem.* **2000**, *275*, 875–882.
108. A. Davy; P. Bello; N. Thierry-Mieg; P. Vaglio; J. Hitti; L. Doucette-Stamm; D. Thierry-Mieg; J. Reboul; S. Boulton; A. J. Walhout; O. Coux; M. Vidal, *EMBO Rep.* **2001**, *2*, 821–828.
109. H. Fu; N. Reis; Y. Lee; M. H. Glickman; R. D. Vierstra, *EMBO J.* **2001**, *20*, 7096–7107.
110. R. Hartmann-Petersen; K. Tanaka; K. B. Hendil, *Arch. Biochem. Biophys.* **2001**, *386*, 89–94.
111. M. H. Glickman; D. M. Rubin; O. Coux; I. Wefes; G. Pfeifer; Z. Cjeka; W. Baumeister; V. A. Fried; D. Finley, *Cell* **1998**, *94*, 615–623.
112. J. A. Maupin-Furlow; H. L. Wilson; S. J. Kaczowka; M. S. Ou, *Front Biosci.* **2000**, *5*, D837–D865.
113. K. Hofmann; P. Bucher, *Trends Biochem. Sci.* **1998**, *23*, 204–205.
114. V. Maytal-Kivity; N. Reis; K. Hofmann; M. H. Glickman, *BMC Biochem.* **2002**, *3*, 28.
115. R. Verma; S. Chen; R. Feldman; D. Schieltz; J. Yates; J. Dohmen; R. J. Deshaies, *Mol. Biol. Cell* **2000**, *11*, 3425–3439.
116. D. S. Leggett; J. Hanna; A. Borodovsky; B. Crosas; M. Schmidt; R. T. Baker; T. Walz; H. Ploegh; D. Finley, *Mol. Cell* **2002**, *10*, 495–507.
117. A. Borodovsky; B. M. Kessler; R. Casagrande; H. S. Overkleeft; K. D. Wilkinson; H. L. Ploegh, *EMBO J.* **2001**, *20*, 5187–5196.
118. F. R. Papa; M. Hochstrasser, *Nature* **1993**, *366*, 313–319.
119. A. J. McClellan; J. Frydman, *Nat. Cell Biol.* **2001**, *3*, E51–E53.
120. J. Demand; S. Alberti; C. Patterson; J. Hohfeld, *Curr. Biol.* **2001**, *11*, 1569–1577.
121. J. Jiang; C. A. Ballinger; Y. Wu; Q. Dai; D. M. Cyr; J. Hohfeld; C. Patterson, *J. Biol. Chem.* **2001**, *276*, 42938–42944.
122. G. C. Meacham; C. Patterson; W. Zhang; J. M. Younger; D. M. Cyr, *Nat. Cell Biol.* **2001**, *3*, 100–105.
123. S. P. Dawson; J. E. Arnold; N. J. Mayer; S. E. Reynolds; M. A. Billett; C. Gordon; L. Colleaux; P. M. Kloetzel; K. Tanaka; R. J. Mayer, *J. Biol. Chem.* **1995**, *270*, 1850–1858.
124. K. Takayanagi; S. Dawson; S. E. Reynolds; R. J. Mayer, *Biochem. Biophys. Res. Commun.* **1996**, *228*, 517–523.
125. H. Kawahara; H. Yokosawa, *Dev. Biol.* **1994**, *166*, 623–633.
126. T. Tokumoto, *Int. Rev. Cytol.* **1999**, *186*, 261–294.
127. H. Kawahara; H. Yokosawa, *Dev. Biol.* **1992**, *151*, 27–33.
128. Y. Yang; K. Fruh; K. Ahn; P. A. Peterson, *J. Biol. Chem.* **1995**, *270*, 27687–27694.
129. G. G. Mason; R. Z. Murray; D. Pappin; A. J. Rivett, *FEBS Lett.* **1998**, *430*, 269–274.

130. K. Satoh; H. Sasajima; K. I. Nyoumura; H. Yokosawa; H. Sawada, *Biochemistry* **2001**, *40*, 314–319.
131. N. Wei; X. W. Deng, *Annu. Rev. Cell Dev. Biol.* **2003**, *19*, 261–286.
132. C. Schwechheimer, *Biochim. Biophys. Acta* **2004**, *1695*, 45–54.
133. X. W. Deng; W. Dubiel; N. Wei; K. Hofmann; K. Mundt, *Trends Genet.* **2000**, *16*, 289.
134. Y. Sun; M. P. Wilson; P. W. Majerus, *J. Biol. Chem.* **2002**, *277*, 45759–45764.
135. B. Karniol; A. Yahalom; S. Kwok; T. Tsuge; M. Matsui; X. W. Deng; D. A. Chamovitz, *FEBS Lett.* **1998**, *439*, 173–179.
136. S. F. Kwok; J. M. Staub; X. W. Deng, *J. Mol. Biol.* **1999**, *285*, 85–95.
137. S. Lyapina; G. Cope; A. Shevchenko; G. Serino; T. Tsuge; C. Zhou; D. A. Wolf; N. Wei; A. Shevchenko; R. J. Deshaies, *Science* **2001**, *292*, 1382–1385.
138. H. Cohen; A. Azriel; T. Cohen; D. Meraro; S. Hashmueli; D. Bech-Otschir; R. Kraft; W. Dubiel; B. Z. Levi, *J. Biol. Chem.* **2000**, *275*, 39081–39089.
139. X. Yang; S. Menon; K. Lykke-Andersen; T. Tsuge; X. Di; X. Wang; R. J. Rodriguez-Suarez,; H. Zhang; N. Wei, *Curr. Biol.* **2002**, *12*, 667–672.
140. X. Hong; L. Xu; X. Li; Z. Zhai; H. Shu, *FEBS Lett.* **2001**, *499*, 133–136.
141. A. K. Hoareau; V. Bochard; S. Rety; P. Jalinot, *FEBS Lett.* **2002**, *527*, 15–21.
142. G. Suzuki; Y. Yanagawa; S. F. Kwok; M. Matsui; X. W. Deng, *Genes Dev.* **2002**, *16*, 554–559.
143. F. X. Claret; M. Hibi; S. Dhut; T. Toda; M. Karin, *Nature* **1996**, *383*, 453–457.
144. K. Tomoda; Y. Kubota; Y. Arata; S. Mori; M. Maeda; T. Tanaka; M. Yoshida; N. Yoneda-Kato; J. Y. Kato, *J. Biol. Chem.* **2002**, *277*, 2302–2310.
145. O. L. Caballero; V. Resto; M. Patturajan; D. Meerzaman; M. Z. Guo; J. Engles; R. Yochem; E. Ratovitski; D. Sidransky; J. Jen, *Oncogene* **2002**, *21*, 3003–3010.
146. S. Mahalingam; V. Ayyavoo; M. Patel; T. Kieber-Emmons; G. D. Kao; R. J. Muschel; D. B. Weiner, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3419–3424.
147. C. Schwechheimer; X. W. Deng, *Trends Cell Biol.* **2001**, *11*, 420–426.
148. A. Yahalom; T. H. Kim; E. Winter; B. Karniol; A. G. von Arnim; D. A. Chamovitz, *J. Biol. Chem.* **2001**, *276*, 334–340.
149. L. Aravind; C. P. Ponting, *Protein Sci.* **1998**, *7*, 1250–1254.
150. C. Schwechheimer; G. Serino; X. W. Deng, *Plant Cell* **2002**, *14*, 2553–2563.
151. C. Schwechheimer; G. Serino; J. Callis; W. L. Crosby; S. Lyapina; R. J. Deshaies; W. M. Gray; M. Estelle; X. W. Deng, *Science* **2001**, *292*, 1379–1382.
152. G. A. Cope; G. S. Suh; L. Aravind; S. E. Schwarz; S. L. Zipursky; E. V. Koonin; R. J. Deshaies, *Science* **2002**, *298*, 608–611.
153. M. Seeger; R. Kraft; K. Ferrell; D. Bech-Otschir; R. Dumdey; R. Schade; C. Gordon; M. Naumann; W. Dubiel, *FASEB J.* **1998**, *12*, 469–478.
154. S. Uhle; O. Medalia; R. Waldron; R. Dumdey; P. Henklein; D. Bech-Otschir; X. Huang; M. Berse; J. Sperling; R. Schade; W. Dubiel, *EMBO J.* **2003**, *22*, 1302–1312.
155. A. D'Andrea; D. Pellman, *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 337–352.
156. K. D. Wilkinson, *Semin. Cell Dev. Biol.* **2000**, *11*, 141–148.
157. A. Y. Amerik; S. J. Li; M. Hochstrasser, *Biol. Chem.* **2000**, *381*, 981–992.
158. S. S. Wing, *Int. J. Biochem. Cell Biol.* **2003**, *35*, 590–605.
159. S. Swaminathan; A. Y. Amerik; M. Hochstrasser, *Mol. Biol. Cell* **1999**, *10*, 2583–2594.
160. K. D. Wilkinson; K. M. Lee; S. Deshpande; P. Duerksen-Hughes; J. M. Boss; J. Pohl, *Science* **1989**, *246*, 670–673.
161. M. B. Butterworth; R. S. Edinger; H. Ovaa; D. Burg; J. P. Johnson; R. A. Frizzell, *J. Biol. Chem.* **2007**, *282*, 37885–37893.
162. E. Trompouki; E. Hatzivassiliou; T. Tsichritzis; H. Farmer; A. Ashworth; G. Mosialos, *Nature* **2003**, *424*, 793–796.
163. W. Reiley; M. Zhang; S. C. Sun, *J. Biol. Chem.* **2004**, *279*, 55161–55167.
164. T. R. Brummelkamp; S. M. Nijman; A. M. Dirac; R. Bernards, *Nature* **2003**, *424*, 797–801.
165. A. Kovalenko; C. Chable-Bessia; G. Cantarella; A. Israel; D. Wallach; G. Courtois, *Nature* **2003**, *424*, 801–805.
166. E. Graner; D. Tang; S. Rossi; A. Baron; T. Migita; L. J. Weinstein; M. Lechpammer; D. Huesken; J. Zimmermann; S. Signoretti; M. Loda, *Cancer Cell* **2004**, *5*, 253–261.
167. K. Oishi; K. Miyazaki; K. Kadota; R. Kikuno; T. Nagase; G. Atsumi; N. Ohkura; T. Azama; M. Mesaki; S. Yukimasa; H. Kobayashi; C. Iitaka; T. Umehara; M. Horikoshi; T. Kudo; Y. Shimizu; M. Yano; M. Monden; K. Machida; J. Matsuda; S. Horie; T. Todo; N. Ishida, *J. Biol. Chem.* **2003**, *278*, 41519–41527.
168. M. Li; C. L. Brooks; N. Kon; W. Gu, *Mol. Cell* **2004**, *13*, 879–886.
169. J. M. Cummins; C. Rago; M. Kohli; K. W. Kinzler; C. Lengauer; B. Vogelstein, *Nature* **2004**, *428*, 1.
170. J. A. van der Knaap; B. R. Kumar; Y. M. Moshkin; K. Langenberg; J. Krijgsveld; A. J. Heck; F. Karch; C. P. Verrijzer, *Mol. Cell* **2005**, *17*, 695–707.
171. X. Wu; L. Yen; L. Irwin; C. Sweeney; K. L. Carraway III, *Mol. Cell Biol.* **2004**, *24*, 7748–7757.
172. S. M. Wilson; B. Bhattacharyya; R. A. Rachel; V. Coppola; L. Tessarollo; D. B. Householder; C. F. Fletcher; R. J. Miller; N. G. Copeland; N. A. Jenkins, *Nat. Genet.* **2002**, *32*, 420–425.
173. K. J. Ritchie; M. P. Malakhov; C. J. Hetherington; L. Zhou; M. T. Little; O. A. Malakhova; J. C. Sipe; S. H. Orkin; D. E. Zhang, *Genes Dev.* **2002**, *16*, 2207–2212.
174. O. A. Malakhova; M. Yan; M. P. Malakhov; Y. Yuan; K. J. Ritchie; K. I. Kim; L. F. Peterson; K. Shuai; D. E. Zhang, *Genes Dev.* **2003**, *17*, 455–460.
175. L. Gong; T. Kamitani; S. Millas; E. T. Yeh, *J. Biol. Chem.* **2000**, *275*, 14212–14216.
176. B. Burnett; F. Li; R. N. Pittman, *Hum. Mol. Genet.* **2003**, *12*, 3195–3205.
177. H. Scheel; S. Tomiuk; K. Hofmann, *Hum. Mol. Genet.* **2003**, *12*, 2845–2852.
178. I. E. Wertz; K. M. O'Rourke; H. Zhou; M. Eby; L. Aravind; S. Seshagiri; P. Wu; C. Wiesmann; R. Baker; D. L. Boone; A. Ma; E. V. Koonin; V. M. Dixit, *Nature* **2004**, *430*, 694–699.
179. R. Groisman; J. Polanowska; I. Kuraoka; J. Sawada; M. Saijo; R. Drapkin; A. F. Kisselev; K. Tanaka; Y. Nakatani, *Cell* **2003**, *113*, 357–367.
180. C. N. Larsen; B. A. Krantz; K. D. Wilkinson, *Biochemistry* **1998**, *37*, 3358–3368.

181. K. D. Wilkinson; S. Deshpande; C. N. Larsen, *Biochem. Soc. Trans.* **1992**, *20*, 631–637.
182. K. D. Wilkinson; V. L. Tashayev; L. B. O'Connor; C. N. Larsen; E. Kasperek; C. M. Pickart, *Biochemistry* **1995**, *34*, 14535–14546.
183. Y. A. Lam; G. N. DeMartino; C. M. Pickart; R. E. Cohen, *J. Biol. Chem.* **1997**, *272*, 28438–28446.
184. H. Lin; L. Yin; J. Reid; K. D. Wilkinson; S. S. Wing, *J. Biol. Chem.* **2001**, *276*, 20357–20363.
185. H. Lin; A. Keriel; C. R. Morales; N. Bedard; Q. Zhao; P. Hingamp; S. Lefrancois; L. Combaret; S. S. Wing, *Mol. Cell Biol.* **2000**, *20*, 6568–6578.
186. X. Li; Y. Yang; J. D. Ashwell, *Nature* **2002**, *416*, 345–347.
187. S. Mori; C. H. Heldin; L. Claesson-Welsh, *J. Biol. Chem.* **1992**, *267*, 6429–6434.
188. P. Soubeyran; K. Kowanetz; I. Szymkiewicz; W. Y. Langdon; I. Dikic, *Nature* **2002**, *416*, 183–187.
189. A. E. Salcini; H. Chen; G. Iannolo; C. P. De; P. P. Di Fiore, *Int. J. Biochem. Cell Biol.* **1999**, *31*, 805–809.
190. R. Carbone; S. Fre; G. Iannolo; F. Belleudi; P. Mancini; P. G. Pelicci; M. R. Torrisi; P. P. Di Fiore, *Cancer Res.* **1997**, *57*, 5498–5504.
191. H. Chen; S. Fre; V. I. Slepnev; M. R. Capua; K. Takei; M. H. Butler; P. P. Di Fiore; C. P. De, *Nature* **1998**, *394*, 793–797.
192. J. A. Rosenthal; H. Chen; V. I. Slepnev; L. Pellegrini; A. E. Salcini; P. P. Di Fiore; C. P. De, *J. Biol. Chem.* **1999**, *274*, 33959–33965.
193. K. Hofmann; L. Falquet, *Trends Biochem. Sci.* **2001**, *26*, 347–350.
194. S. Polo; S. Sigismund; M. Faretta; M. Guidi; M. R. Capua; G. Bossi; H. Chen; C. P. De; P. P. Di Fiore, *Nature* **2002**, *416*, 451–455.
195. C. E. Oldham; R. P. Mohney; S. L. Miller; R. N. Hanes; J. P. O'Bryan, *Curr. Biol.* **2002**, *12*, 1112–1116.
196. K. Haglund; S. Sigismund; S. Polo; I. Szymkiewicz; P. P. Di Fiore; I. Dikic, *Nat. Cell Biol.* **2003**, *5*, 461–466.
197. Y. Mosesson; K. Shtiegman; M. Katz; Y. Zwang; G. Vereb; J. Szollosi; Y. Yarden, *J. Biol. Chem.* **2003**, *278*, 21323–21326.
198. J. Terrell; S. Shih; R. Dunn; L. Hicke, *Mol. Cell* **1998**, *1*, 193–202.
199. S. B. Helliwell; S. Losko; C. A. Kaiser, *J. Cell Biol.* **2001**, *153*, 649–662.
200. N. P. Martin; R. J. Lefkowitz; S. K. Shenoy, *J. Biol. Chem.* **2003**, *278*, 45954–45959.
201. D. J. Katzmann; M. Babst; S. D. Emr, *Cell* **2001**, *106*, 145–155.
202. M. Babst; D. J. Katzmann; W. B. Snyder; B. Wendland; S. D. Emr, *Dev. Cell* **2002**, *3*, 283–289.
203. M. Babst; D. J. Katzmann; E. J. Estepa-Sabal; T. Meerloo; S. D. Emr, *Dev. Cell* **2002**, *3*, 271–282.
204. P. M. Scott; P. S. Bilodeau; O. Zhdankina; S. C. Winistorfer; M. J. Hauglund; M. M. Allaman; W. R. Kearney; A. D. Robertson; A. L. Boman; R. C. Piper, *Nat. Cell Biol.* **2004**, *6*, 252–259.
205. R. Puertollano; J. S. Bonifacino, *Nat. Cell Biol.* **2004**, *6*, 244–251.
206. L. Chen; N. G. Davis, *Traffic* **2002**, *3*, 110–123.
207. S. Confalonieri; A. E. Salcini; C. Puri; C. Tacchetti; P. P. Di Fiore, *J. Cell Biol.* **2000**, *150*, 905–912.
208. M. Burbea; L. Dreier; J. S. Dittman; M. E. Grunwald; J. M. Kaplan, *Neuron* **2002**, *35*, 107–120.
209. G. N. Patrick; B. Bingol; H. A. Weld; E. M. Schuman, *Curr. Biol.* **2003**, *13*, 2073–2081.
210. F. K. Bedford; J. T. Kittler; E. Muller; P. Thomas; J. M. Uren; D. Merlo; W. Wisden; A. Triller; T. G. Smart; S. J. Moss, *Nat. Neurosci.* **2001**, *4*, 908–916.
211. C. Buttner; S. Sadtler; A. Leyendecker; B. Laube; N. Griffon; H. Betz; G. Schmalzing, *J. Biol. Chem.* **2001**, *276*, 42978–42985.
212. B. Luscher; C. A. Keller, *Nat. Cell Biol.* **2001**, *3*, E232–E233.
213. S. K. Shenoy; R. J. Lefkowitz, *J. Biol. Chem.* **2003**, *278*, 14498–14506.
214. S. K. Shenoy; P. H. McDonald; T. A. Kohout; R. J. Lefkowitz, *Science* **2001**, *294*, 1307–1313.
215. A. Vecchione; A. Marchese; P. Henry; D. Rotin; A. Morrione, *Mol. Cell Biol.* **2003**, *23*, 3363–3372.
216. Z. Nawaz; D. M. Lonard; A. P. Dennis; C. L. Smith; B. W. O'Malley, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1858–1862.
217. P. van Kerkhof; R. Govers; C. M. Alves dos Santos; G. J. Strous, *J. Biol. Chem.* **2000**, *275*, 1575–1580.
218. A. Yu; T. R. Malek, *J. Biol. Chem.* **2001**, *276*, 381–385.
219. K. Chaturvedi; P. Bandari; N. Chinen; R. D. Howells, *J. Biol. Chem.* **2001**, *276*, 12345–12355.
220. P. van Kerkhof; C. M. Alves dos Santos,; M. Sachse; J. Klumperman; G. Bu; G. J. Strous, *Mol. Biol. Cell* **2001**, *12*, 2556–2566.
221. A. Rocca; C. Lamaze; A. Subtil; A. Dautry-Varsat, *Mol. Biol. Cell* **2001**, *12*, 1293–1301.
222. L. Ellgaard; A. Helenius, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 181–191.
223. T. Sommer; S. Jentsch, *Nature* **1993**, *365*, 176–179.
224. T. J. Jensen; M. A. Loo; S. Pind; D. B. Williams; A. L. Goldberg; J. R. Riordan, *Cell* **1995**, *83*, 129–135.
225. C. L. Ward; S. Omura; R. R. Kopito, *Cell* **1995**, *83*, 121–127.
226. M. M. Hiller; A. Finger; M. Schweiger; D. H. Wolf, *Science* **1996**, *273*, 1725–1728.
227. E. J. Wiertz; D. Tortorella; M. Bogyo; J. Yu; W. Mothes; T. R. Jones; T. A. Rapoport; H. L. Ploegh, *Nature* **1996**, *384*, 432–438.
228. M. S. Gelman; E. S. Kannegaard; R. R. Kopito, *J. Biol. Chem.* **2002**, *277*, 11709–11714.
229. M. Pilon; R. Schekman; K. Romisch, *EMBO J.* **1997**, *16*, 4540–4548.
230. R. K. Plemper; S. Bohmler; J. Bordallo; T. Sommer; D. H. Wolf, *Nature* **1997**, *388*, 891–895.
231. J. S. Thrower; L. Hoffman; M. Rechsteiner; C. M. Pickart, *EMBO J.* **2000**, *19*, 94–102.
232. E. D. Werner; J. L. Brodsky; A. A. McCracken, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13797–13801.
233. E. Jarosch; C. Taxis; C. Volkwein; J. Bordallo; D. Finley; D. H. Wolf; T. Sommer, *Nat. Cell Biol.* **2002**, *4*, 134–139.
234. M. de Virgilio; H. Weninger; N. E. Ivessa, *J. Biol. Chem.* **1998**, *273*, 9734–9743.
235. A. A. McCracken; J. L. Brodsky, *Bioessays* **2003**, *25*, 868–877.
236. K. Robzyk; J. Recht; M. A. Osley, *Science* **2000**, *287*, 501–504.
237. W. W. Hwang; S. Venkatasubrahmanyam; A. G. Ianculescu; A. Tong; C. Boone; H. D. Madhani, *Mol. Cell* **2003**, *11*, 261–266.
238. A. Wood; N. J. Krogan; J. Dover; J. Schneider; J. Heidt; M. A. Boateng; K. Dean; A. Golshani; Y. Zhang; J. F. Greenblatt; M. Johnston; A. Shilatifard, *Mol. Cell* **2003**, *11*, 267–274.
239. T. Jenuwein; C. D. Allis, *Science* **2001**, *293*, 1074–1080.
240. I. Bach; H. P. Ostendorff, *Trends Biochem. Sci.* **2003**, *28*, 189–195.
241. J. C. Tanny; H. Erdjument-Bromage; P. Tempst; C. D. Allis, *Genes Dev.* **2007**, *21*, 835–847.
242. M. A. Osley, *Brief. Funct. Genomic Proteomic* **2006**, *5*, 179–189.
243. M. H. Koken; P. Reynolds; I. Jaspers-Dekker; L. Prakash; S. Prakash; D. Bootsma; J. H. Hoeijmakers, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8865–8869.

244. H. P. Roest; J. van Klaveren; J. de Wit; C. G. van Gurp; M. H. Koken; M. Vermey; J. H. van Roijen; J. W. Hoogerbrugge; J. T. Vreeburg; W. M. Baarends; D. Bootsma; J. A. Grootegoed; J. H. Hoeijmakers, *Cell* **1996**, *86*, 799–810.
245. W. M. Baarends; E. Wassenaar; J. W. Hoogerbrugge; S. Schoenmakers; Z. W. Sun; J. A. Grootegoed, *J. Cell Sci.* **2007**, *120*, 1841–1851.
246. J. Kim; S. B. Hake; R. G. Roeder, *Mol. Cell* **2005**, *20*, 759–770.
247. B. Zhu; Y. Zheng; A. D. Pham; S. S. Mandal; H. Erdjument-Bromage; P. Tempst; D. Reinberg, *Mol. Cell* **2005**, *20*, 601–611.
248. A. S. Coutts; N. B. La Thangue, *Cell Cycle* **2007**, *6*, 827–829.
249. A. Chen; F. E. Kleiman; J. L. Manley; T. Ouchi; Z. Q. Pan, *J. Biol. Chem.* **2002**, *277*, 22085–22092.
250. D. L. Mallery; C. J. Vandenberg; K. Hiom, *EMBO J.* **2002**, *21*, 6755–6762.
251. Y. Xia; G. M. Pao; H. W. Chen; I. M. Verma; T. Hunter, *J. Biol. Chem.* **2003**, *278*, 5255–5263.
252. J. A. Daniel; M. S. Torok; Z. W. Sun; D. Schieltz; C. D. Allis; J. R. Yates III; P. A. Grant, *J. Biol. Chem.* **2004**, *279*, 1867–1871.
253. V. M. Weake; J. L. Workman, *Mol. Cell* **2008**, *29*, 653–663.
254. C. F. Kao; C. Hillyer; T. Tsukuda; K. Henry; S. Berger; M. A. Osley, *Genes Dev.* **2004**, *18*, 184–195.
255. R. N. Laribee; N. J. Krogan; T. Xiao; Y. Shibata; T. R. Hughes; J. F. Greenblatt; B. D. Strahl, *Curr. Biol.* **2005**, *15*, 1487–1493.
256. G. A. Hartzog; J. W. Tamkun, *Genes Dev.* **2007**, *21*, 3209–3213.
257. J. S. Lee; A. Shukla; J. Schneider; S. K. Swanson; M. P. Washburn; L. Florens; S. R. Bhaumik; A. Shilatifard, *Cell* **2007**, *131*, 1084–1096.
258. K. W. Mulder; A. B. Brenkman; A. Inagaki; N. J. van den Broek; H. T. Timmers, *Nucleic Acids Res.* **2007**, *35*, 2428–2439.
259. R. N. Laribee; Y. Shibata; D. P. Mersman; S. R. Collins; P. Kemmeren; A. Roguev; J. S. Weissman; S. D. Briggs; N. J. Krogan; B. D. Strahl, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5836–5841.
260. R. N. Laribee; S. M. Fuchs; B. D. Strahl, *Genes Dev.* **2007**, *21*, 737–743.
261. A. Wyce; T. Xiao; K. A. Whelan; C. Kosman; W. Walter; D. Eick; T. R. Hughes; N. J. Krogan; B. D. Strahl; S. L. Berger, *Mol. Cell* **2007**, *27*, 275–288.
262. R. Cao; Y. Tsukada; Y. Zhang, *Mol. Cell* **2005**, *20*, 845–854.
263. H. Wang; L. Wang; H. Erdjument-Bromage; M. Vidal; P. Tempst; R. S. Jones; Y. Zhang, *Nature* **2004**, *431*, 873–878.
264. W. Zhou; P. Zhu; J. Wang; G. Pascual; K. A. Ohgi; J. Lozach; C. K. Glass; M. G. Rosenfeld, *Mol. Cell* **2008**, *29*, 69–80.
265. J. Wei; L. Zhai; J. Xu; H. Wang, *J. Biol. Chem.* **2006**, *281*, 22537–22544.
266. H. Ogawa; K. Ishiguro; S. Gaubatz; D. M. Livingston; Y. Nakatani, *Science* **2002**, *296*, 1132–1136.
267. D. Frescas; D. Guardavaccaro; F. Bassermann; R. Koyama-Nasu; M. Pagano, *Nature* **2007**, *450*, 309–313.
268. C. Sanchez; I. Sanchez; J. A. Demmers; P. Rodriguez; J. Strouboulis; M. Vidal, *Mol. Cell Proteomics* **2007**, *6*, 820–834.
269. A. Ciechanover; J. A. DiGiuseppe; B. Bercovich; A. Orian; J. D. Richter; A. L. Schwartz; G. M. Brodeur, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 139–143.
270. R. Shen; M. Chen; Y. J. Wang; H. Kaneki; L. Xing; R. J. O'Keefe; D. Chen, *J. Biol. Chem.* **2006**, *281*, 3569–3576.
271. S. C. Upadhya; T. K. Smith; A. N. Hegde, *J. Neurochem.* **2004**, *91*, 210–219.
272. A. Sundqvist; M. T. Bengoechea-Alonso; X. Ye; V. Lukiyanchuk; J. Jin; J. W. Harper; J. Ericsson, *Cell Metab.* **2005**, *1*, 379–391.
273. D. Ungureanu; O. Silvennoinen, *Sci. STKE* **2005**, *2005*, e49.
274. J. M. Lingbeck; J. S. Trausch-Azar; A. Ciechanover; A. L. Schwartz, *Oncogene* **2005**, *24*, 6376–6384.
275. U. Hoecker, *Curr. Opin. Plant Biol.* **2005**, *8*, 469–476.
276. D. M. Lonard; Z. Nawaz; C. L. Smith; B. W. O'Malley, *Mol. Cell* **2000**, *5*, 939–948.
277. S. E. Salghetti; M. Muratani; H. Wijnen; B. Futcher; W. P. Tansey, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3118–3123.
278. S. E. Salghetti; A. A. Caudy; J. G. Chenoweth; W. P. Tansey, *Science* **2001**, *293*, 1651–1653.
279. S. Y. Kim; A. Herbst; K. A. Tworkowski; S. E. Salghetti; W. P. Tansey, *Mol. Cell* **2003**, *11*, 1177–1188.
280. N. von der Lehr; S. Johansson; S. Wu; F. Bahram; A. Castell; C. Cetinkaya; P. Hydbring; I. Weidung; K. Nakayama; K. I. Nakayama; O. Söderberg; T. K. Kerppola; L. G. Larsson, *Mol. Cell* **2003**, *11*, 1189–1200.
281. J. R. Lipford; G. T. Smith; Y. Chi; R. J. Deshaies, *Nature* **2005**, *438*, 113–116.
282. J. Chen; T. K. Archer, *Mol. Cell Biol.* **2005**, *25*, 9016–9027.
283. T. G. Gillette; F. Gonzalez; A. Delahodde; S. A. Johnston; T. Kodadek, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5904–5909.
284. A. Ferdous; F. Gonzalez; L. Sun; T. Kodadek; S. A. Johnston, *Mol. Cell* **2001**, *7*, 981–991.
285. F. Gonzalez; A. Delahodde; T. Kodadek; S. A. Johnston, *Science* **2002**, *296*, 548–550.
286. A. Ferdous; T. Kodadek; S. A. Johnston, *Biochemistry* **2002**, *41*, 12798–12805.
287. M. C. Morris; P. Kaiser; S. Rudyak; C. Baskerville; M. H. Watson; S. I. Reed, *Nature* **2003**, *423*, 1009–1013.
288. E. Ezhkova; W. P. Tansey, *Mol. Cell* **2004**, *13*, 435–442.
289. D. Lee; E. Ezhkova; B. Li; S. G. Pattenden; W. P. Tansey; J. L. Workman, *Cell* **2005**, *123*, 423–436.
290. P. K. Dash; B. Hochner; E. R. Kandel, *Nature* **1990**, *345*, 718–721.
291. R. Bourtchuladze; B. Frenguelli; J. Blendy; D. Cioffi; G. Schutz; A. J. Silva, *Cell* **1994**, *79*, 59–68.
292. J. C. Yin; J. S. Wallach; M. Del Vecchio; E. L. Wilder; H. Zhou; W. G. Quinn; T. Tully, *Cell* **1994**, *79*, 49–58.
293. D. Bartsch; A. Casadio; K. A. Karl; P. Serodio; E. R. Kandel, *Cell* **1998**, *95*, 211–223.
294. J. F. Habener; C. P. Miller; M. Vallejo, *Vitam. Horm.* **1995**, *51*, 1–57.
295. P. Sassone-Corsi, *Annu. Rev. Cell Dev. Biol.* **1995**, *11*, 355–377.
296. C. Dong; S. C. Upadhya; L. Ding; T. K. Smith; A. N. Hegde, *Learn. Mem.* **2008**, *15*, 335–347.
297. E. S. Johnson, *Annu. Rev. Biochem.* **2004**, *73*, 355–382.
298. R. Geiss-Friedlander; F. Melchior, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 947–956.
299. O. Kerscher; R. Felberbaum; M. Hochstrasser, *Annu. Rev. Cell Dev. Biol.* **2006**, *22*, 159–180.
300. J. Herrmann; L. O. Lerman; A. Lerman, *Circ. Res.* **2007**, *100*, 1276–1291.
301. N. Mizushima; T. Noda; T. Yoshimori; Y. Tanaka; T. Ishii; M. D. George; D. J. Klionsky; M. Ohsumi; Y. Ohsumi, *Nature* **1998**, *395*, 395–398.
302. Y. Ohsumi, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 211–216.
303. A. R. Thompson; J. H. Doelling; A. Suttangkakul; R. D. Vierstra, *Plant Physiol.* **2005**, *138*, 2097–2110.

304. N. V. Giannakopoulos; J. K. Luo; V. Papov; W. Zou; D. J. Lenschow; B. S. Jacobs; E. C. Borden; J. Li; H. W. Virgin; D. E. Zhang, *Biochem. Biophys. Res. Commun.* **2005**, *336*, 496–506.
305. C. Zhao; C. Denison; J. M. Huibregtse; S. Gygi; R. M. Krug, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10200–10205.
306. V. Bernier-Villamor; D. A. Sampson; M. J. Matunis; C. D. Lima, *Cell* **2002**, *108*, 345–356.
307. A. Pichler; A. Gast; J. S. Seeler; A. Dejean; F. Melchior, *Cell* **2002**, *108*, 109–120.
308. K. Stade; F. Vogel; I. Schwienhorst; B. Meusser; C. Volkwein; B. Nentwig; R. J. Dohmen; T. Sommer, *J. Biol. Chem.* **2002**, *277*, 49554–49561.
309. D. C. Schwartz; M. Hochstrasser, *Trends Biochem. Sci.* **2003**, *28*, 321–328.
310. S. Ross; J. L. Best; L. I. Zon; G. Gill, *Mol. Cell* **2002**, *10*, 831–842.
311. A. Sapetschnig; G. Rischitor; H. Braun; A. Doll; M. Schergaut; F. Melchior; G. Suske, *EMBO J.* **2002**, *21*, 5206–5215.
312. K. L. Hari; K. R. Cook; G. H. Karpen, *Genes Dev.* **2001**, *15*, 1334–1348.
313. A. Sobko; H. Ma; R. A. Firtel, *Dev. Cell* **2002**, *2*, 745–756.
314. X. Long; L. C. Griffith, *J. Biol. Chem.* **2000**, *275*, 40765–40776.
315. F. Melchior, *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 591–626.
316. E. S. Johnson; A. A. Gupta, *Cell* **2001**, *106*, 735–744.
317. T. Kahyo; T. Nishida; H. Yasuda, *Mol. Cell* **2001**, *8*, 713–718.
318. Y. Takahashi; A. Toh-e; Y. Kikuchi, *Gene* **2001**, *275*, 223–231.
319. M. Hochstrasser, *Cell* **2001**, *107*, 5–8.
320. O. Kirsh; J. S. Seeler; A. Pichler; A. Gast; S. Muller; E. Miska; M. Mathieu; A. Harel-Bellan; T. Kouzarides; F. Melchior; A. Dejean, *EMBO J.* **2002**, *21*, 2682–2691.
321. M. H. Kagey; T. A. Melhuish; D. Wotton, *Cell* **2003**, *113*, 127–137.
322. M. H. Kagey; T. A. Melhuish; S. E. Powers; D. Wotton, *EMBO J.* **2005**, *24*, 108–119.
323. M. S. Rodriguez; C. Dargemont; R. T. Hay, *J. Biol. Chem.* **2001**, *276*, 12654–12659.
324. L. Jin; A. Williamson; S. Banerjee; I. Philipp; M. Rape, *Cell* **2008**, *133*, 653–665.
325. S. J. Li; M. Hochstrasser, *Nature* **1999**, *398*, 246–251.
326. D. Girdwood; D. Bumpass; O. A. Vaughan; A. Thain; L. A. Anderson; A. W. Snowden; E. Garcia-Wilson; N. D. Perkins; R. T. Hay, *Mol. Cell* **2003**, *11*, 1043–1054.
327. G. Zheng; Y. C. Yang, *J. Biol. Chem.* **2004**, *279*, 42410–42421.
328. M. Hochstrasser, *Genes Dev.* **1998**, *12*, 901–907.
329. D. Liakopoulos; G. Doenges; K. Matuschewski; S. Jentsch, *EMBO J.* **1998**, *17*, 2208–2214.
330. R. M. Hofmann; C. M. Pickart, *Cell* **1999**, *96*, 645–653.
331. C. M. Pickart; D. Fushman, *Curr. Opin. Chem. Biol.* **2004**, *8*, 610–616.
332. L. D. Mastrandrea; J. You; E. G. Niles; C. M. Pickart, *J. Biol. Chem.* **1999**, *274*, 27299–27306.
333. L. Deng; C. Wang; E. Spencer; L. Yang; A. Braun; J. You; C. Slaughter; C. Pickart; Z. J. Chen, *Cell* **2000**, *103*, 351–361.
334. C. Wang; L. Deng; M. Hong; G. R. Akkaraju; J. Inoue; Z. J. Chen, *Nature* **2001**, *412*, 346–351.
335. C. S. Shi; J. H. Kehrl, *J. Biol. Chem.* **2003**, *278*, 15429–15434.
336. H. Häcker; M. Karin, *Sci. STKE* **2006**, *2006*, re13.
337. H. Sebban; S. Yamaoka; G. Courtois, *Trends Cell Biol.* **2006**, *16*, 569–577.
338. N. Li; S. Banin; H. Ouyang; G. C. Li; G. Courtois; Y. Shiloh; M. Karin; G. Rotman, *J. Biol. Chem.* **2001**, *276*, 8898–8903.
339. T. T. Huang; S. M. Wuerzberger-Davis; Z. H. Wu; S. Miyamoto, *Cell* **2003**, *115*, 565–576.
340. Z. H. Wu; Y. Shi; R. S. Tibbetts; S. Miyamoto, *Science* **2006**, *311*, 1141–1146.
341. Y. Zhang; D. B. Moheban; B. R. Conway; A. Bhattacharyya; R. A. Segal, *J. Neurosci.* **2000**, *20*, 5671–5678.
342. J. C. Arevalo; J. Waite; R. Rajagopal; M. Beyna; Z. Y. Chen; F. S. Lee; M. V. Chao, *Neuron* **2006**, *50*, 549–559.
343. D. Mukhopadhyay; H. Riezman, *Science* **2007**, *315*, 201–205.
344. J. M. Peters, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 644–656.
345. B. R. Thornton; T. M. Ng; M. E. Matyskiela; C. W. Carroll; D. O. Morgan; D. P. Toczyski, *Genes Dev.* **2006**, *20*, 449–460.
346. N. den Elzen; J. Pines, *J. Cell Biol.* **2001**, *153*, 121–136.
347. M. J. Hayes; Y. Kimata; S. L. Wattam; C. Lindon; G. Mao; H. Yamano; A. M. Fry, *Nat. Cell Biol.* **2006**, *8*, 607–614.
348. D. Gabellini; I. N. Colaluca; H. C. Vodermaier; G. Biamonti; M. Giacca; A. Falaschi; S. Riva; F. A. Peverali, *EMBO J.* **2003**, *22*, 3715–3724.
349. R. W. King; J. M. Peters; S. Tugendreich; M. Rolfe; P. Hieter; M. W. Kirschner, *Cell* **1995**, *81*, 279–288.
350. A. Hagting; N. den Elzen; H. C. Vodermaier; I. C. Waizenegger; J. M. Peters; J. Pines, *J. Cell Biol.* **2002**, *157*, 1125–1137.
351. D. M. Gordon; D. M. Roof, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12515–12520.
352. T. J. McGarry; M. W. Kirschner, *Cell* **1998**, *93*, 1043–1053.
353. S. Stewart; G. Fang, *Mol. Cell Biol.* **2005**, *25*, 10516–10527.
354. C. Lindon; J. Pines, *J. Cell Biol.* **2004**, *164*, 233–241.
355. A. Castro; S. Vigneron; C. Bernis; J. C. Labbe; C. Prigent; T. Lorca, *EMBO Rep.* **2002**, *3*, 1209–1214.
356. S. Stewart; G. Fang, *Cancer Res.* **2005**, *65*, 8730–8735.
357. W. M. Zhao; G. Fang, *J. Biol. Chem.* **2005**, *280*, 33516–33524.
358. N. Furuno; N. den Elzen; J. Pines, *J. Cell Biol.* **1999**, *147*, 295–306.
359. S. Sigrist; H. Jacobs; R. Stratmann; C. F. Lehner, *EMBO J.* **1995**, *14*, 4827–4838.
360. M. Rape; S. K. Reddy; M. W. Kirschner, *Cell* **2006**, *124*, 89–103.
361. C. Acquaviva; F. Herzog; C. Kraft; J. Pines, *Nat. Cell Biol.* **2004**, *6*, 892–898.
362. T. C. Wheeler; L. S. Chin; Y. Li; F. L. Roudabush; L. Li, *J. Biol. Chem.* **2002**, *277*, 10273–10282.
363. M. G. Muralidhar; J. B. Thomas, *Neuron* **1993**, *11*, 253–266.
364. C. E. Oh; R. McMahon; S. Benzer; M. A. Tanouye, *J. Neurosci.* **1994**, *14*, 3166–3179.
365. C. T. Kuo; S. Zhu; S. Younger; L. Y. Jan; Y. N. Jan, *Neuron* **2006**, *51*, 283–290.
366. Y. H. Jiang; D. Armstrong; U. Albrecht; C. M. Atkins; J. L. Noebels; G. Eichele; J. D. Sweatt; A. L. Beaudet, *Neuron* **1998**, *21*, 799–811.

367. P. van Roessel; D. A. Elliott; I. M. Robinson; A. Prokop; A. H. Brand, *Cell* **2004**, *119*, 707–718.
368. P. Juo; J. M. Kaplan, *Curr. Biol.* **2004**, *14*, 2057–2062.
369. J. Stegmuller; Y. Konishi; M. A. Huynh; Z. Yuan; S. Dibacco; A. Bonni, *Neuron* **2006**, *50*, 389–400.
370. L. Dreier; M. Burbea; J. M. Kaplan, *Neuron* **2005**, *46*, 51–64.
371. A. DiAntonio; A. P. Haghighi; S. L. Portman; J. D. Lee; A. M. Amaranto; C. S. Goodman, *Nature* **2001**, *412*, 449–452.
372. S. M. Greenberg; V. F. Castellucci; H. Bayley; J. H. Schwartz, *Nature* **1987**, *329*, 62–65.
373. D. G. Chain; A. Casadio; S. Schacher; A. N. Hegde; M. Valbrun; N. Yamamoto; A. L. Goldberg; D. Bartsch; E. R. Kandel; J. H. Schwartz, *Neuron* **1999**, *22*, 147–156.
374. M. Lopez-Salon; M. Alonso; M. R. Vianna; H. Viola; T. Mello e Souza; I. Izquierdo; J. M. Pasquini; J. H. Medina, *Eur. J. Neurosci.* **2001**, *14*, 1820–1826.
375. H. Lind; S. Zienolddiny; P. O. Ekstrom; V. Skaug; A. Haugen, *Int. J. Cancer* **2006**, *119*, 718–721.
376. C. Menin; M. C. Scaini; G. L. De Salvo; M. Biscuola; M. Quaggio; G. Esposito; C. Belluco; M. Montagna; S. Agata; E. D'Andrea; D. Nitti; A. Amadori; R. Bertorelle, *J. Natl. Cancer Inst.* **2006**, *98*, 285–288.
377. Y. Katagiri; Y. Hozumi; S. Kondo, *J. Dermatol. Sci.* **2006**, *42*, 215–224.
378. M. S. Lim; A. Adamson; Z. Lin; B. Perez-Ordonez; R. C. Jordan; S. Tripp; S. L. Perkins; K. S. Elenitoba-Johnson, *Blood* **2002**, *100*, 2950–2956.
379. V. S. Spiegelman; W. Tang; A. M. Chan; M. Igarashi; S. A. Aaronson; D. A. Sassoon; M. Katoh; T. J. Slaga; S. Y. Fuchs, *J. Biol. Chem.* **2002**, *277*, 36624–36630.
380. A. Ougolkov; B. Zhang; K. Yamashita; V. Bilim; M. Mai; S. Y. Fuchs; T. Minamoto, *J. Natl. Cancer Inst.* **2004**, *96*, 1161–1170.
381. S. Muerkoster; A. Arlt; B. Sipos; M. Witt; M. Grossmann; G. Kloppel; H. Kalthoff; U. R. Folsch; H. Schafer, *Cancer Res.* **2005**, *65*, 1316–1324.
382. A. V. Gerstein; T. A. Almeida; G. Zhao; E. Chess; I. Shih; K. Buhler; K. Pienta; M. A. Rubin; R. Vessella; N. Papadopoulos, *Genes Chromosomes Cancer* **2002**, *34*, 9–16.
383. M. Slotky; M. Shapira; O. Ben-Izhak; S. Linn; B. Futerman; M. Tsalic; D. D. Hershko, *Breast Cancer Res.* **2005**, *7*, R737–R744.
384. F. Traub; M. Mengel; H. J. Lück; H. H. Kreipe; R. von Wasielewski, *Breast Cancer Res. Treat.* **2006**, *99*, 185–191.
385. H. Sonoda; H. Inoue; K. Ogawa; T. Utsunomiya; T. A. Masuda; M. Mori, *Clin. Cancer Res.* **2006**, *12*, 1215–1220.
386. M. Shapira; O. Ben-Izhak; S. Linn; B. Futerman; I. Minkov; D. D. Hershko, *Cancer* **2005**, *103*, 1336–1346.
387. M. Gstaiger; R. Jordan; M. Lim; C. Catzavelos; J. Mestan; J. Slingerland; W. Krek, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5043–5048.
388. N. Inui; K. Kitagawa; S. Miwa; T. Hattori; K. Chida; H. Nakamura; M. Kitagawa, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 978–984.
389. S. Yokoi; K. Yasui; M. Mori; T. Iizasa; T. Fujisawa; J. Inazawa, *Am. J. Pathol.* **2004**, *165*, 175–180.
390. C. Q. Zhu; F. H. Blackhall; M. Pintilie; P. Iyengar; N. Liu; J. Ho; T. Chomiak; D. Lau; T. Winton; F. A. Shepherd; M. S. Tsao, *Clin. Cancer Res.* **2004**, *10*, 1984–1991.
391. D. Masuya; C. Huang; D. Liu; T. Nakashima; H. Yokomise; M. Ueno; N. Nakashima; S. Sumitomo, *J. Pathol.* **2006**, *208*, 724–732.
392. L. Wang; S. Banerjee, *Oncol. Rep.* **2004**, *11*, 1319–1324.
393. P. Leder; J. Battey; G. Lenoir; C. Moulding; W. Murphy; H. Potter; T. Stewart; R. Taub, *Science* **1983**, *222*, 765–771.
394. K. Milde-Langosch, *Eur. J. Cancer* **2005**, *41*, 2449–2461.
395. A. Sanjay; W. C. Horne; R. Baron, *Sci. STKE* **2001**, *2001*, E40.
396. D. P. Xirodimas; C. W. Stephen; D. P. Lane, *Exp. Cell Res.* **2001**, *270*, 66–77.
397. K. H. Vousden; C. Prives, *Cell* **2005**, *120*, 7–10.
398. C. L. Brooks; W. Gu, *Mol. Cell* **2006**, *21*, 307–315.
399. A. M. Bode; Z. Dong, *Nat. Rev. Cancer* **2004**, *4*, 793–805.
400. S. Signoretti; L. Di Marcotullio; A. Richardson; S. Ramaswamy; B. Isaac; M. Rue; F. Monti; M. Loda; M. Pagano, *J. Clin. Invest.* **2002**, *110*, 633–641.
401. J. Shaughnessy, *Hematology* **2005**, *10* (Suppl. 1), 117–126.
402. F. Zhan; S. Colla; X. Wu; B. Chen; J. P. Stewart; W. M. Kuehl; B. Barlogie; J. D. Shaughnessy, Jr., *Blood* **2007**, *109*, 4995–5001.
403. D. Frescas; M. Pagano, *Nat. Rev. Cancer* **2008**, *8*, 438–449.
404. C. J. Kim; J. H. Song; Y. G. Cho; Y. S. Kim; S. Y. Kim; S. W. Nam; N. J. Yoo; J. Y. Lee; W. S. Park, *APMIS* **2007**, *115*, 127–133.
405. J. Liu; K. G. Suresh Kumar; D. Yu; S. A. Molton; M. McMahon; M. Herlyn; A. Thomas-Tikhonenko; S. Y. Fuchs, *Oncogene* **2007**, *26*, 1954–1958.
406. D. Hoeller; C. M. Hecker; I. Dikic, *Nat. Rev. Cancer* **2006**, *6*, 776–788.
407. R. A. Shimkets; D. G. Warnock; C. M. Bositis; C. Nelson-Williams; J. H. Hansson; M. Schambelan; J. R. Gill, Jr.,; S. Ulick; R. V. Milora; J. W. Findling, *Cell* **1994**, *79*, 407–414.
408. C. Jacquemont; T. Taniguchi, *BMC Biochem.* **2007**, *8* (Suppl. 1), S10.
409. M. A. Maynard; M. Ohh, *Curr. Cancer Drug Targets* **2005**, *5*, 345–356.
410. J. N. Keller; F. F. Huang; H. Zhu; J. Yu; Y. S. Ho; T. S. Kindy, *J. Cereb. Blood Flow Metab.* **2000**, *20*, 1467–1473.
411. A. L. Bulteau; K. C. Lundberg; K. M. Humphries; H. A. Sadek; P. A. Szweda; B. Friguet; L. I. Szweda, *J. Biol. Chem.* **2001**, *276*, 30057–30063.
412. S. Krause; U. Kuckelkorn; T. Dorner; G. R. Burmester; E. Feist; P. M. Kloetzel, *Ann. Rheum. Dis.* **2006**, *65*, 1021–1027.
413. Z. Hu; Z. Zhang; E. Doo; O. Coux; A. L. Goldberg; T. J. Liang, *J. Virol.* **1999**, *73*, 7231–7240.
414. W. E. Mitch; R. Medina; S. Grieber; R. C. May; B. K. England; S. R. Price; J. L. Bailey; A. L. Goldberg, *J. Clin. Invest.* **1994**, *93*, 2127–2133.
415. S. C. Hobler; A. Williams; D. Fischer; J. J. Wang; X. Sun; J. E. Fischer; J. J. Monaco; P. O. Hasselgren, *Am. J. Physiol.* **1999**, *277*, R434–R440.
416. J. M. Rommens; M. C. Iannuzzi; B. Kerem; M. L. Drumm; G. Melmer; M. Dean; R. Rozmahel; J. L. Cole; D. Kennedy; N. Hidaka, *Science* **1989**, *245*, 1059–1065.
417. S. M. Rowe; S. Miller; E. J. Sorscher, *N. Engl. J. Med.* **2005**, *352*, 1992–2001.
418. J. Zielenski; L. C. Tsui, *Annu. Rev. Genet.* **1995**, *29*, 777–807.

419. D. M. Cyr, *Nat. Struct. Mol. Biol.* **2005**, *12*, 2–3.
420. J. M. Younger; L. Chen; H. Y. Ren; M. F. Rosser; E. L. Turnbull; C. Y. Fan; C. Patterson; D. M. Cyr, *Cell* **2006**, *126*, 571–582.
421. H. McDonough; C. Patterson, *Cell Stress Chaperones* **2003**, *8*, 303–308.
422. S. Murata; T. Chiba; K. Tanaka, *Int. J. Biochem. Cell Biol.* **2003**, *35*, 572–578.
423. M. Botero-Velez; J. J. Curtis; D. G. Warnock, *N. Engl. J. Med.* **1994**, *330*, 178–181.
424. F. W. van Leeuwen; D. P. de Kleijn; H. H. van den Hurk; A. Neubauer; M. A. Sonnemans; J. A. Sluijs; S. Koycu; R. D. Ramdjielal; A. Salehi; G. J. Martens; F. G. Grosveld; J. Peter; H. Burbach; E. M. Hol, *Science* **1998**, *279*, 242–247.
425. S. Keck; R. Nitsch; T. Grune; O. Ullrich, *J. Neurochem.* **2003**, *85*, 115–122.
426. J. A. Johnston; M. J. Dalton; M. E. Gurney; R. R. Kopito, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12571–12576.
427. T. Kishino; M. Lalande; J. Wagstaff, *Nat. Genet.* **1997**, *15*, 70–73.
428. T. Matsuura; J. S. Sutcliffe; P. Fang; R. J. Galjaard; Y. H. Jiang; C. S. Benton; J. M. Rommens; A. L. Beaudet, *Nat. Genet.* **1997**, *15*, 74–77.
429. K. Saigoh; Y. L. Wang; J. G. Suh; T. Yamanishi; Y. Sakai; H. Kiyosawa; T. Harada; N. Ichihara; S. Wakana; T. Kikuchi; K. Wada, *Nat. Genet.* **1999**, *23*, 47–51.
430. S. W. Davies; M. Turmaine; B. A. Cozens; M. DiFiglia; A. H. Sharp; C. A. Ross; E. Scherzinger; E. E. Wanker; L. Mangiarini; G. P. Bates, *Cell* **1997**, *90*, 537–548.
431. T. Kitada; S. Asakawa; N. Hattori; H. Matsumine; Y. Yamamura; S. Minoshima; M. Yokochi; Y. Mizuno; N. Shimizu, *Nature* **1998**, *392*, 605–608.
432. E. Leroy; R. Boyer; G. Auburger; B. Leube; G. Ulm; E. Mezey; G. Harta; M. J. Brownstein; S. Jonnalagada; T. Chernova; A. Dehejia; C. Lavedan; T. Gasser; P. J. Steinbach; K. D. Wilkinson; M. H. Polymeropoulos, *Nature* **1998**, *395*, 451–452.
433. H. Shimura; M. G. Schlossmacher; N. Hattori; M. P. Frosch; A. Trockenbacher; R. Schneider; Y. Mizuno; K. S. Kosik; D. J. Selkoe, *Science* **2001**, *293*, 263–269.
434. F. A. Middleton; K. Mirnics; J. N. Pierri; D. A. Lewis; P. Levitt, *J. Neurosci.* **2002**, *22*, 2718–2729.
435. C. A. Altar; L. W. Jurata; V. Charles; A. Lemire; P. Liu; Y. Bukhman; T. A. Young; J. Bullard; H. Yokoe; M. J. Webster; M. B. Knable; J. A. Brockman, *Biol. Psychiatry* **2005**, *58*, 85–96.
436. Q. Zhai; J. Wang; A. Kim; Q. Liu; R. Watts; E. Hoopfer; T. Mitchison; L. Luo; Z. He, *Neuron* **2003**, *39*, 217–225.
437. D. J. Selkoe, *Science* **2002**, *298*, 789–791.
438. K. Hsiao; P. Chapman; S. Nilsen; C. Eckman; Y. Harigaya; S. Younkin; F. Yang; G. Cole, *Science* **1996**, *274*, 99–102.
439. D. M. Walsh; I. Klyubin; J. V. Fadeeva; W. K. Cullen; R. Anwyl; M. S. Wolfe; M. J. Rowan; D. J. Selkoe, *Nature* **2002**, *416*, 535–539.
440. J. N. Keller; K. B. Hanni; W. R. Markesbery, *J. Neurochem.* **2000**, *75*, 436–439.
441. E. K. Hoffman; H. M. Wilcox; R. W. Scott; R. Siman, *J. Neurol. Sci.* **1996**, *139*, 15–20.
442. J. Niwa; S. Ishigaki; N. Hishikawa; M. Yamamoto; M. Doyu; S. Murata; K. Tanaka; N. Taniguchi; G. Sobue, *J. Biol. Chem.* **2002**, *277*, 36793–36798.
443. S. Ishigaki; J. Niwa; Y. Ando; T. Yoshihara; K. Sawada; M. Doyu; M. Yamamoto; K. Kato; Y. Yotsumoto; G. Sobue, *FEBS Lett.* **2002**, *531*, 354–358.
444. M. Urushitani; J. Kurisu; M. Tateno; S. Hatakeyama; K. Nakayama; S. Kato; R. Takahashi, *J. Neurochem.* **2004**, *90*, 231–244.
445. M. Urushitani; J. Kurisu; K. Tsukita; R. Takahashi, *J. Neurochem.* **2002**, *83*, 1030–1042.
446. J. P. Lee; C. Gerin; V. P. Bindokas; R. Miller; G. Ghadge; R. P. Roos, *J. Neurochem.* **2002**, *82*, 1229–1238.
447. A. S. Vlug; D. Jaarsma, *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **2004**, *5*, 16–21.
448. K. Aquilano; G. Rotilio; M. R. Ciriolo, *J. Neurochem.* **2003**, *85*, 1324–1335.
449. D. H. Hyun; M. Lee; B. Halliwell; P. Jenner, *J. Neurochem.* **2003**, *86*, 363–373.
450. C. A. Williams, *Brain Dev.* **2005**, *27*, 88–94.
451. J. M. Huibregtse; M. Scheffner; P. M. Howley, *EMBO J.* **1991**, *10*, 4129–4135.
452. S. Kumar; A. L. Talis; P. M. Howley, *J. Biol. Chem.* **1999**, *274*, 18785–18792.
453. C. Kuhne; L. Banks, *J. Biol. Chem.* **1998**, *273*, 34302–34309.
454. U. Nuber; S. E. Schwarz; M. Scheffner, *Eur. J. Biochem.* **1998**, *254*, 643–649.
455. E. Mohr; P. Brouwers; J. J. Claus; U. M. Mann; P. Fedio; T. N. Chase, *Mov. Disord.* **1991**, *6*, 127–132.
456. T. Foroud; E. Siemers; D. Kleindorfer; D. J. Bill; M. E. Hodes; J. A. Norton; P. M. Conneally; J. C. Christian, *Ann. Neurol.* **1995**, *37*, 657–664.
457. K. W. Lange; B. J. Sahakian; N. P. Quinn; C. D. Marsden; T. W. Robbins, *J. Neurol. Neurosurg. Psychiatry* **1995**, *58*, 598–606.
458. A. D. Lawrence; B. J. Sahakian; J. R. Hodges; A. E. Rosser; K. W. Lange; T. W. Robbins, *Brain* **1996**, *119* (Pt. 5), 1633–1645.
459. A. D. Lawrence; J. R. Hodges; A. E. Rosser; A. Kershaw; C. ffrench-Constant; D. C. Rubinsztein; T. W. Robbins; B. J. Sahakian, *Brain* **1998**, *121* (Pt. 7), 1329–1341.
460. K. P. Murphy; R. J. Carter; L. A. Lione; L. Mangiarini; A. Mahal; G. P. Bates; S. B. Dunnett; A. J. Morton, *J. Neurosci.* **2000**, *20*, 5115–5123.
461. P. F. Shelbourne; N. Killeen; R. F. Hevner; H. M. Johnston; L. Tecott; M. Lewandoski; M. Ennis; L. Ramirez; Z. Li; C. Iannicola; D. R. Littman; R. M. Myers, *Hum. Mol. Genet.* **1999**, *8*, 763–774.
462. A. J. Morton; R. L. Faull; J. M. Edwardson, *Brain Res. Bull.* **2001**, *56*, 111–117.
463. J. A. Lee; C. S. Lim; S. H. Lee; H. Kim; N. Nukina; B. K. Kaang, *J. Neurochem.* **2003**, *85*, 160–169.
464. N. R. Jana; E. A. Zemskov; G. Wang; N. Nukina, *Hum. Mol. Genet.* **2001**, *10*, 1049–1059.
465. N. F. Bence; R. M. Sampat; R. R. Kopito, *Science* **2001**, *292*, 1552–1555.
466. E. Martín-Aparicio; A. Yamamoto; F. Hernández; R. Hen; J. Avila; J. J. Lucas, *J. Neurosci.* **2001**, *21*, 8772–8781.
467. Q. Ding; J. J. Lewis; K. M. Strum; E. Dimayuga; A. J. Bruce-Keller; J. C. Dunn; J. N. Keller, *J. Biol. Chem.* **2002**, *277*, 13935–13942.
468. M. H. Polymeropoulos, *Ann. N.Y. Acad. Sci.* **2000**, *920*, 28–32.
469. V. Bonifati; P. Rizzu; M. J. van Baren; O. Schaap; G. J. Breedveld; E. Krieger; M. C. Dekker; F. Squitieri; P. Ibanez; M. Joosse; J. W. van Dongen; N. Vanacore; J. C. van Swieten; A. Brice; G. Meco; C. M. van Duijn; B. A. Oostra; P. Heutink, *Science* **2003**, *299*, 256–259.

470. M. R. Cookson, *Neuron* **2003**, *37*, 7–10.
471. J. Lotharius; P. Brundin, *Nat. Rev. Neurosci.* **2002**, *3*, 932–942.
472. M. S. Goldberg; S. M. Fleming; J. J. Palacino; C. Cepeda; H. A. Lam; A. Bhatnagar; E. G. Meloni; N. Wu; L. C. Ackerson; G. J. Klapstein; M. Gajendiran; B. L. Roth; M. F. Chesselet; N. T. Maidment; M. S. Levine; J. Shen, *J. Biol. Chem.* **2003**, *278*, 43628–43635.
473. H. Shimura; N. Hattori; S. Kubo; Y. Mizuno; S. Asakawa; S. Minoshima; N. Shimizu; K. Iwai; T. Chiba; K. Tanaka; T. Suzuki, *Nat. Genet.* **2000**, *25*, 302–305.
474. D. P. Huynh; D. R. Scoles; D. Nguyen; S. M. Pulst, *Hum. Mol. Genet.* **2003**, *12*, 2587–2597.

Biographical Sketch

Ashok N. Hegde received his Ph.D. from the Centre for Cellular and Molecular Biology, Hyderabad, India, where he studied signal transduction by ras protein. He discovered a novel histidine phosphorylation stimulated by ras protein in liver cells. He changed his field to neuroscience at the postdoctoral level. Dr. Hegde obtained his postdoctoral training from the Center for Neurobiology and Behavior of Columbia University College of Physicians and Surgeons, New York, USA under the mentorship of leading neuroscientists Drs. James Schwartz and Eric Kandel. During his postdoctoral years, Dr. Hegde discovered a role for ubiquitin–proteasome-mediated protein degradation in synaptic plasticity that underlies learning and memory. Dr. Hegde's laboratory investigates the key unanswered questions pertaining to the role of the ubiquitin–proteasome pathway in synaptic plasticity, learning and memory, and how impaired proteolysis is linked to neurodegenerative diseases such as Alzheimer's disease.