

6.01 Overview and Introduction

Robert Woodward and Peng George Wang, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|--------|--------------|---|
| 6.01.1 | Introduction | 1 |
| 6.01.2 | Overview | 1 |

6.01.1 Introduction

Given the prevalence of sugars in nature, it is no surprise that a thorough understanding of carbohydrate chemistry coupled with glycobiology is often essential to elucidating the details of complex biological processes such as polysaccharide biosynthesis, the glycosylation of proteins in both prokaryotic and eukaryotic cells, and glycoconjugate-induced immunological activity. This has enabled the extension of previously existing synthetic strategies, as well as the development of novel methods, for the preparation of complicated carbohydrate-based biomolecules. For example, many excellent organometallic transformations have found applicability in glycoconjugate chemistry. Furthermore, the advance of physical organic chemistry has yielded a deeper understanding of the mechanism of glycosylation reactions, consequently allowing several new glycosylation approaches to be developed. These advances, coupled with those made in biology and biotechnology, have thus further solidified the necessity of effectively interweaving chemistry and biology. It is here, where chemistry and biology meet, that many more breakthroughs across various disciplines will undoubtedly be made.

6.01.2 Overview

The 20 additional chapters that comprise this volume provide a detailed look at carbohydrates and their uses both in a laboratory setting and in nature, while also examining a variety of other important biomolecules and associated processes. The first three chapters, contributed by Peng George Wang, Ashraf Brik, and Zhongwu Guo, consider how carbohydrates can be utilized for synthetic purposes. Both enzymatic and chemical modes of synthesis are discussed with regard to their use for complex carbohydrate, glycoprotein, and vaccine production. Applications of mass spectrometry to analysis of glycans are then explored in the next two chapters as authors Kay-Hooi Khoo and Jianjun Li discuss some of the more advanced technologies in the field.

Chapters 6.07–6.09 offer a look at the burgeoning field of chemical glycobiology. A broad look at this area is first presented by Jennifer Kohler in Chapter 6.07 with topics ranging from chemoenzymatic approaches to understanding glycan structure and function to the use of photocross-linkers for the covalent trapping of interactions between glycans. The final two chapters of this section, written by Alan Elbein and Howard Hang, take a more focused approach, considering how certain processes can be studied through the use of inhibitors and probes.

The next section of this volume introduces how carbohydrates are utilized in a variety of biosynthetic pathways. Specifically, author Miguel Valvano first considers O-antigen biosynthesis, a process that is essential in the production of lipopolysaccharide. A discussion of the generation of glycoproteins in both mammalian and bacterial systems follows in the next two chapters with contributions from Inka Brockhausen and Mario Feldman. Subsequently, the structure and biosynthesis of the mycobacterial cell wall and glycosaminoglycans are examined by Dean Crick and Jian Liu in Chapters 6.13 and 6.14, respectively. The reader is then offered a section by Jack Preiss, which details the processes of bacterial and mammalian glycogen biosynthesis, while also examining the production of starch in plants. Finally, Rajai Atalla's contribution, Chapter 6.16, focuses upon celluloses and their various roles in nature.

In the closing chapters of this volume, a variety of other important biomolecules and related processes are highlighted. Specifically, Norman Lewis and Daneel Ferreira describe lignins and proanthocyanidins in Chapters 6.17 and 6.18, specifically looking at the chemistry and biology of these highly aromatic molecules.

The final three chapters then direct attention to the areas of DNA and RNA with Darrell Davis first illustrating the importance and utility of nucleoside analogs. Chapter 6.20, written by George Garcia, subsequently examines how RNA is modified enzymatically and the corresponding consequences thereof before the structures and roles of riboswitches in regulatory processes are presented by Tina Henkin in Chapter 6.21.

As is quite apparent, the topics that appear within this volume vary quite extensively. However, whether it is a discussion of the transfer of an oligosaccharyl moiety to a serine residue in O-glycosylation or the regulation of transcription termination through a riboswitch, it is hoped that the reader will not only come away with an appreciation for all of the biomolecules discussed, but more importantly recognize the diversity of processes that these small molecules and associated polymers can perform and/or regulate.

Biographical Sketches



Robert Woodward obtained a B.A. in chemistry (2006) and a B.S. in biology (2006) from The Ohio State University. He is an NIH Chemistry–Biology Interface Program Predoctoral Fellow currently pursuing a Ph.D. in organic chemistry at The Ohio State University. His research focuses on the applications of synthesis for studying complex biological processes such as polysaccharide biosynthesis and oligosaccharide transfer.



Peng George Wang obtained a B.S. in chemistry (1984) from Nankai University, China, and a Ph.D. in organic chemistry (1990) from the University of California, Berkeley. He then conducted postdoctoral research at the Scripps Research Institute before becoming an assistant professor in 1994 at the University of Miami in Coral Gables. From 1997 to 2003, he was a faculty member at Wayne State University. Since then, he has served in the Departments of Biochemistry and Chemistry at The Ohio State University as Ohio Eminent Scholar in Macromolecular Structure and Function.

Research in the Wang laboratory is predominately focused on four areas of glycoscience. Glycochemistry: Work is centered on the generation of uncommon sugar libraries as well as the synthesis of key intermediates in carbohydrate-based biological processes, which are essential for their study. Glycobiology: Biochemical characterization of carbohydrate-active enzymes and investigations of the biological functions of carbohydrates in human diseases, immunity, and general microbiology are performed. Glycotechnology: Biosynthetic pathways are engineered for the synthesis of glycopharmaeuticals, polysaccharides for vaccine development, and biomedically important human glycoproteins. Glycoanalysis: Analysis of carbohydrate composition, sequence, structure, and their interaction with proteins through MS, NMR, QCM, and other analytical methods is conducted.

6.02 Enzymatic Synthesis of Complex Carbohydrates

Wei Zhao and Tiehai Li, Nankai University, Tianjin, China

Robert Woodward, Chengfeng Xia, and Peng George Wang, The Ohio State University, Columbus, OH, USA

Wanyi Guan, Shandong University, Shandong, China

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|---|----|
| 6.02.1 | Introduction | 5 |
| 6.02.2 | Glycosidases | 8 |
| 6.02.2.1 | Three Mechanisms of Glycosidases | 8 |
| 6.02.2.2 | Glycosidases in Carbohydrate Synthesis | 8 |
| 6.02.2.3 | Conclusion | 14 |
| 6.02.3 | Glycosynthases | 15 |
| 6.02.3.1 | Mechanisms of Glycosynthases | 15 |
| 6.02.3.2 | Glycosynthases in Carbohydrate Synthesis | 17 |
| 6.02.3.3 | Conclusion | 22 |
| 6.02.4 | Glycosyltransferases | 23 |
| 6.02.4.1 | Sugar Nucleotide Biosynthetic Pathway | 23 |
| 6.02.4.2 | Leloir Glycosyltransferases | 25 |
| 6.02.4.2.1 | Basic principle | 25 |
| 6.02.4.2.2 | Enzyme-based complex saccharide synthesis | 25 |
| 6.02.4.2.3 | Conclusion | 46 |
| 6.02.5 | Outlook | 47 |
| References | | 49 |

6.02.1 Introduction

Recently, oligosaccharides and glycoconjugates have emerged as a new and challenging research area at the interface of biology and chemistry. Carbohydrates, which constitute one of the most abundant types of biomolecules, play a very broad set of roles in biological science, especially in physiological and pathological processes, molecular recognition, signal transduction, cell communication, cell differentiation, and developmental events.^{1–6} In fact, carbohydrate complexes have been widely used as potential pharmaceuticals for the prevention of infection, the neutralization of toxins, and the immunotherapy of cancer. Therefore, further growth in research on the biological functions of the varied glycan structures will, undoubtedly, be closely tied to the availability of bioactive carbohydrates.^{7–9}

Complex saccharides are highly diverse in structure and biological functions. They are essential for many fields of research, for example, biochemical studies in glycobiology, as potential drugs directed to enzymes or receptors involved in their function and metabolism, and as advanced materials due to their biocompatibility, structure-forming capacity, and environmentally benign properties.^{10–15} The development of efficient synthetic methodologies for their preparation has therefore been in high demand. Methods for both chemical and enzymatic syntheses have experienced notable advances in the last few years with the aim of producing either polysaccharides resembling the natural products or novel polysaccharide mimics for biomedical applications and biomaterials development.

Although chemical tools have been proven indispensable for studies in glycobiology,¹⁶ most oligosaccharides and glycoconjugates and their derivatives are very difficult and costly to generate by chemical synthetic approaches, due to severe limitations in complex carbohydrates syntheses. Similar to proteins and DNA, glycans too consist of limited types of building blocks, monosaccharides. However, oligomers come in a far greater

diversity of structures because of varying glycosidic linkage patterns, stereochemistry, and branching. In addition, unlike DNA and peptide syntheses, which are commonly performed on solid phase using commercial instruments, comparable methodologies for oligosaccharides are still in their infancy.^{17,18} Given that all oligosaccharide synthetic protocols are plagued with the often tedious protection and deprotection steps, one-pot enzymatic systems with high regio- and stereoselectivities are clearly attractive alternatives.^{19–22} It is no wonder that Hurtley *et al.*²³ predicted as early as in 2001 in *Science* that what has rescued the chemistry and biology of carbohydrates, the Cinderella from the shadows, is no fairy godmother but a plethora of new synthetic and analytic methods. Over the years, enzymatic approaches have been gaining popularity for the synthesis of oligosaccharides and glycoconjugates, and it is becoming increasingly feasible to produce complex carbohydrates on a large scale, following enzymatic biosynthetic pathways. The biosynthesis of naturally occurring oligo- and polysaccharides is a complex process that involves the formation of glycosidic bonds between their constituent monosaccharide units and side chain modifications to produce specific functional group derivatives.

Among the numerous enzymes associated with carbohydrate processing in cells, the enzymes used in enzymatic synthesis belong to three categories: glycosidases, glycosynthases, and glycosyltransferases (**Table 1**).

Glycosidases are enzymes that hydrolyze oligosaccharides and polysaccharides *in vivo*. Under appropriate conditions, however, the activated intermediates can be intercepted by other sugars to form new glycosidic bonds. They can form glycosidic linkages under *in vitro* conditions in which a carbohydrate hydroxyl moiety acts as a more efficient nucleophile than water itself. They have been of tremendous benefit in the enzymatic synthesis of oligosaccharides due to their availability, stability, organic solvent compatibility, and low cost.^{24–26} Nevertheless, traditional glycosidase-catalyzed transglycosylations still suffer from low yields and poor unpredictable regioselectivities.

Glycosynthases, a class of glycosidase mutants, have been developed to enhance the enzymatic activity toward the synthesis of oligosaccharides, through mutation of a single catalytic carboxylate nucleophile to a neutral amino acid residue (Ala or Ser). Conversion of one of the free carboxylates to a methyl group provides an active site that retains the correct steric environment for the formation of a reactive glycosyl donor but lacks a key catalytic group for cleavage. The resulting enzymes have no hydrolytic activity, but increased activity toward the synthesis of oligosaccharides, using glycosylfluorides as activated donors.^{27–30} The efficiency of glycosidases is emphasized by the fact that some can accept unnatural substrates, displaying modifications of the sugar moiety and/or a variety of aglycone groups.

Glycosyltransferases are enzymes that can transfer sugar moiety to a defined acceptor, so as to construct a specific glycosidic linkage. This ‘one enzyme–one linkage’ concept makes glycosyltransferases useful and important in the construction of glycosidic linkages in complex saccharide synthesis.^{31–33} Glycosyltransferases can be further divided into two groups: the transferases of the Leloir pathway and those of non-Leloir pathways. The Leloir transferases are responsible for the synthesis of most glycoconjugates in cells, especially in mammalian systems,^{34–36} and are the focus of this chapter.

Table 1 Enzymatic formation of glycosidic bonds

| Donor + Acceptor $\xrightarrow{\text{Enzymes}}$ Products | | | |
|--|------------------------------|--|---|
| Enzymes | Glycosyl donor | Advantage | Disadvantage |
| Glycosidase | Nitrophenyl glycoside | Easy to perform Low cost | Low yield Low regioselectivity |
| Glycosynthase | Glycosylfluoride | High yield | Hard to obtain enzyme Difficult to predict results |
| Leloir glycosyl-transferase | Sugar nucleotide | High yield High regio- and stereoselectivities Essential for important sequences | High cost |
| Non-Leloir glycosyltransferase | Sugar phosphate or glycoside | High yield High regio- and stereoselectivities | Not useful for important sugar sequences |

6.02.2 Glycosidases

6.02.2.1 Three Mechanisms of Glycosidases

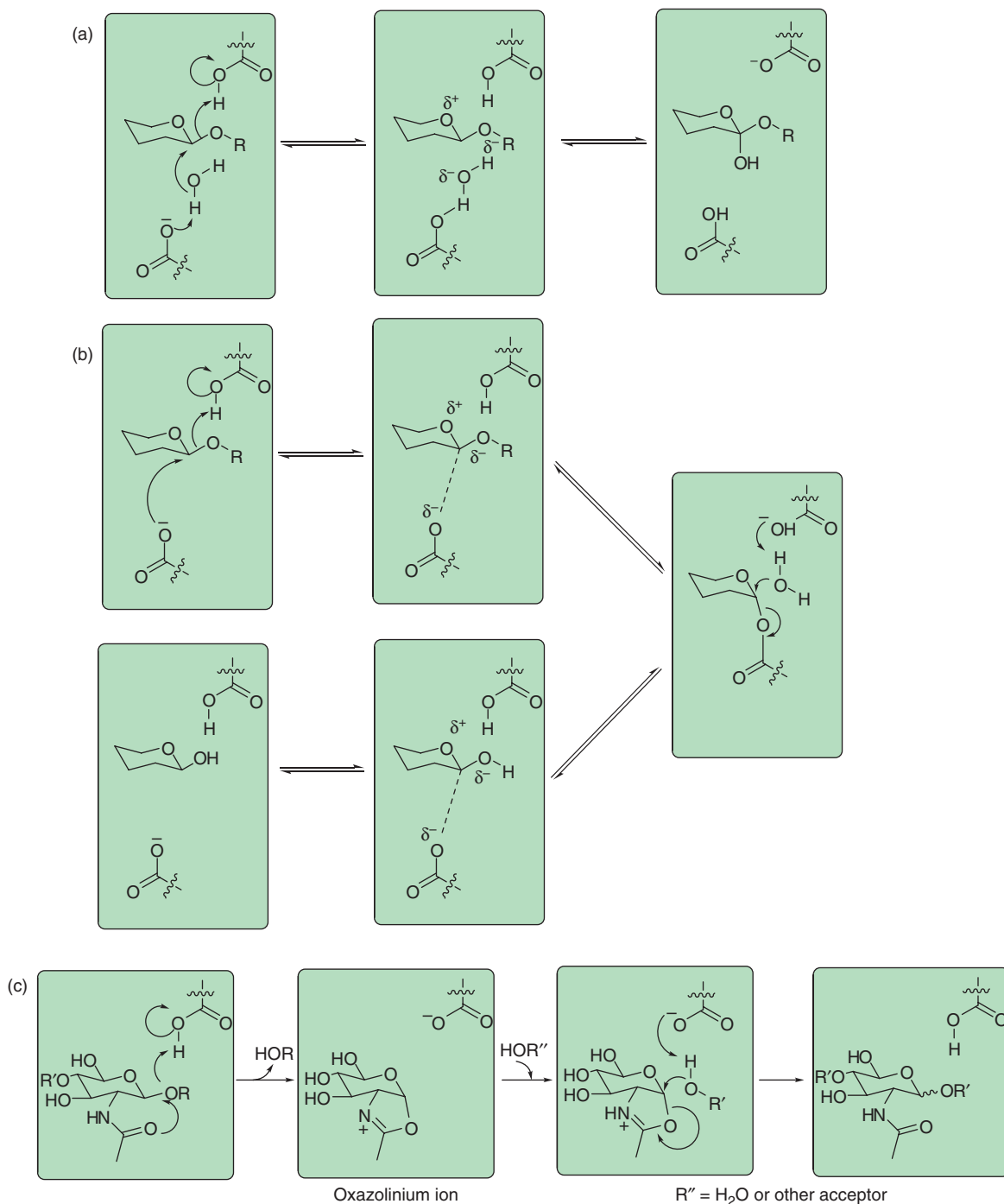
Glycosidases are degrading enzymes that catalyze the hydrolysis of glycosidic bonds *in vivo*,⁵⁶ but their normal hydrolytic reaction can be reversed under appropriate conditions. Therefore, glycosidases have been extensively studied as biocatalysts for oligo- and polysaccharide biosynthesis. They are stable enzymes and easy to produce, and a large number of enzymes from different organisms with different specificities are available. In addition, the glycosyl donors required are inexpensive compounds and easy to obtain in the gram scale. Through X-ray crystallographic analysis and site-directed mutagenesis, three reaction mechanisms have been found to exist for all glycosidases. The first is the inverting mechanism, which generally involves the nucleophilic attack of water at the anomeric center concomitant with the acid-catalyzed departure of the leaving group. This reaction occurs via a single-displacement mechanism wherein one carboxylic acid acts as a general base and the other as a general acid (**Scheme 3(a)**).⁵⁷ The second is the retaining mechanism, which occurs via a double-displacement mechanism wherein one carboxylic acid acts as a general acid–base and the other as a nucleophile. All reactions proceed in two ordered steps. In the first step, the carboxyl oxygen at the anomeric center, acting as nucleophile, attacks the substrate, resulting in the formation of a transient covalent glycosyl–enzyme intermediate. The other carboxyl group concomitantly facilitates departure of the aglycone leaving group by providing a general acid. In the second step, the residue acting as a general acid in the first step now acts as a general base, promoting the attack of water at the anomeric center, cleaving the intermediate to yield the hemiacetal product with retained stereochemistry (**Scheme 3(b)**). When used for synthetic applications, glycosidases often display only moderate regioselectivity and conversion yields. Nevertheless, they are generally considered to be useful because they are readily available and catalytically versatile.^{58–61} The efficiency of glycosidases is emphasized by the fact that some can accept unnatural substrates, displaying modifications of the sugar moiety and/or a variety of aglycone groups.^{62–64} The third mechanism is also a retaining in stereochemistry, which involves assistance from the neighboring 2-acetamido group of the substrate. So, this reaction is often named as a substrate-assisted mechanism (**Scheme 3(c)**).⁶⁵

All of the three mentioned mechanisms involve oxocarbenium ion-like transition states and a pair of carboxylic acids at the active site; however, they differ in several aspects. The inverting mechanism is a one-step reaction that results in the formation of a product with inverted stereochemistry at the anomeric center. The other two alternatives are retaining in stereochemistry at the anomeric center and differ from each other primarily in the nature of the intermediate; in the second mechanism, this species is a covalent enzyme adduct, whereas in the third case it is believed to be a bicyclic oxazoline or oxazolinium ion.⁶⁶

6.02.2.2 Glycosidases in Carbohydrate Synthesis

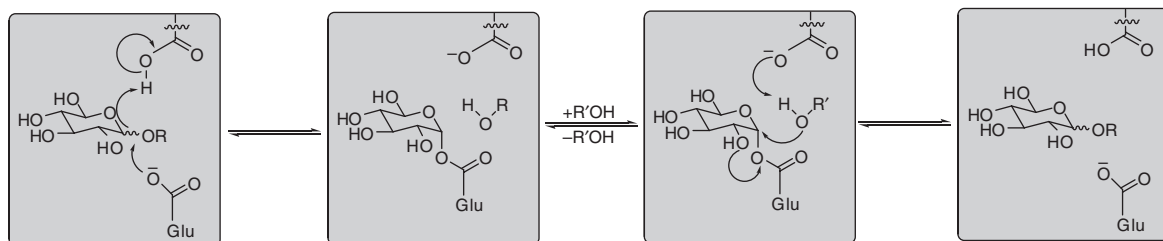
Carbohydrates play important structural and functional roles in numerous physiological processes, including various disease states.⁹ Most biologically important glycoconjugates (oligosaccharides, glycoproteins, and glycolipids) and their derivatives are difficult to obtain in large quantities, whether from natural or synthetic sources. The use of glycosidases for carbohydrate synthesis is currently being pursued by several research groups. The enzymatic action of retaining glycosidases is based on the formation of a glycosyl ester of enzyme, which quickly reacts with a nucleophile present in the reaction mixture to form the products. If the nucleophile is a water molecule, then the hydrolysis reaction takes place. Transglycosylation product formation is observed if other nucleophiles are present in the reaction mixture. The formation of self-transglycosylation products usually occurs when the glycon moiety of the enzyme glycosyl ester is transferred to another molecule of the substrate itself. Under appropriate conditions, however, the activated intermediates can be intercepted by other sugars to form new glycosidic bonds (**Scheme 4**).⁵⁶ Reverse hydrolysis (equilibrium-controlled synthesis) and transglycosylation (kinetically controlled process) are two mechanisms used in glycosidase-catalyzed synthesis of complex saccharides. Equilibrium-controlled synthesis offers only modest yields of oligosaccharide products, while kinetically controlled synthesis, which requires a retaining glycosidase, provides better yields (10–40%).

Compared with glycotransferases, glycosidases have many advantages in carbohydrate synthesis. Due to the high specificity, glycosyltransferases are incapable of synthesizing analogues of the naturally occurring complex



Scheme 3 General glycosidase mechanisms. (a) An inverting β -glycosidase. (b) A retaining β -glycosidase proceeding through an intermediate with a 4C1 conformation. (c) Substrate-assisted mechanism.

saccharides, which would be useful tools for the exploration of oligosaccharide functions. Alternatively, carbohydrate synthesis by glycosidases is especially useful where a glycosyltransferase is not available or difficult to obtain. Glycosidases also have the advantage of not requiring expensive sugar nucleotides as the sugar donor. They can generate glycosidic bonds using relatively simple glycosyl donors and readily available



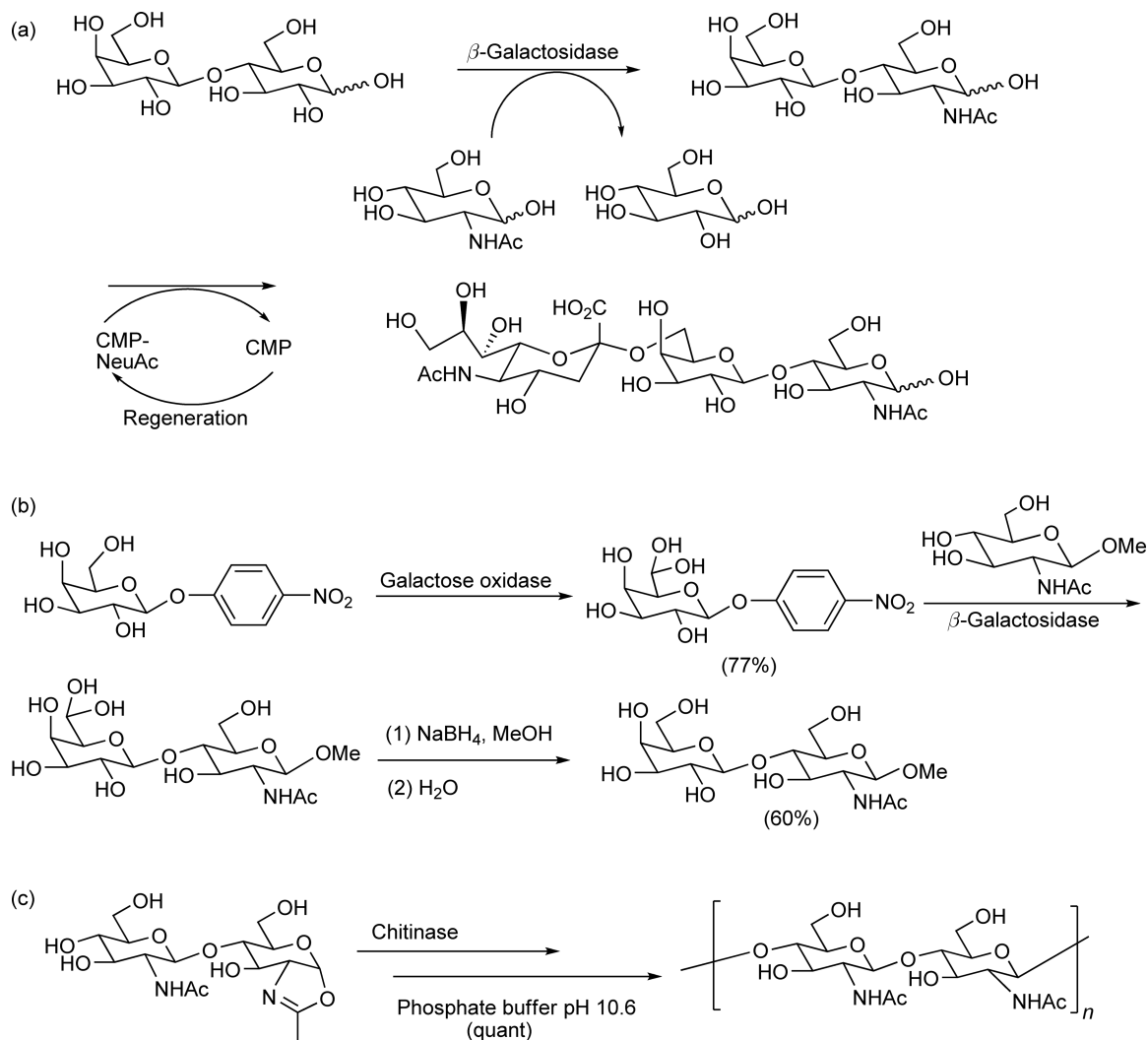
Scheme 4 Catalytic mechanisms of retaining glycosidases.

robust enzymes. Therefore, they offer the opportunity to synthesize oligosaccharides inexpensively. In addition, glycosidases are abundant and can often be used directly without purification. Moreover, they also have many advantages in the enzymatic synthesis of oligosaccharides due to their availability, stability, organic solvent compatibility, and low cost.^{25,26,67}

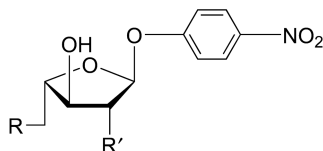
Traditional glycosidase-catalyzed transglycosylation reactions still suffer from low yields and poor regioselectivities. So, it is not generally economical for large-scale carbohydrate synthesis. A variety of strategies have been employed in the glycoside syntheses, including either thermodynamic-controlled (increasing substrate concentrations, decreasing product concentrations by absorption, elevating reaction temperatures, adding water-miscible organic cosolvents) or kinetic-controlled (using activated glycosyl donors and exogenous nucleophiles) protocols, and several large-scale glycosidic bond-forming reactions have been reported in the past years.^{68–72}

Some of the novel strategies are depicted in **Scheme 5**. Glycosidase-catalyzed synthesis of disaccharides, for example, can be coupled *in situ* with a glycosyltransferase reaction to improve the overall yield (**Scheme 5(a)**).⁷³ Another interesting way to improve the yield and facilitate product isolation of glycosidase-catalyzed glycosidation reactions has been demonstrated in the galactosidase-catalyzed synthesis of *N*-acetyllactosamine, one of the intermediates in the synthesis of SLe^x.²² The key was the use of 6-oxo *p*-nitrophenyl galactose, prepared by enzymatic oxidation of the corresponding galactose derivative with galactose oxidase, as the glycosyl donor. The 6-oxo derivatives are less prone to be hydrolyzed and result in improved yields of the 6'-oxo disaccharide. Reduction of the aldehyde with sodium borohydride afforded the desired product and also facilitated isolation due to the formation of a boron complex (**Scheme 5(b)**). Chitinase, which normally works to hydrolyze the polymer chitin, has been used to synthesize artificial chitin in quantitative yield.⁷⁴ The key to this polymerization reaction is the use of a transition state analogue substrate and performing the reaction at a high pH where the enzyme can activate the substrate but cannot hydrolyze the products (**Scheme 5(c)**).

α -L-Arabinofuranosidase of bacterial origin (AbfD3), a versatile glycosidase of carbohydrate-acting enzymes, has been successfully employed for the synthesis of novel alkyl-glycosides⁷⁵ and furanose-containing disaccharides.^{76,77} Evidences proved that AbfD3 can catalyze the self-condensation of both *p*-nitrophenyl α -L-arabinofuranoside and *p*-nitrophenyl β -D-galactofuranoside. Similarly, mixed disaccharides could be obtained when these *p*-nitrophenyl furanosides serve as donors and β -D-xylosides serve as acceptors. Depending on the donor/acceptor ratio, the reactions occurred with variable degrees of regioselectivity toward both (1 \rightarrow 2) and (1 \rightarrow 3) linkages, which reflects the glycosidic linkage specificity of AbfD3 operating in hydrolytic mode.⁷⁸ Nugier-Chauvin and coworkers⁷⁹ have investigated the specificity of an α -L-arabinofuranosidase using C2- and C5-modified α -L-arabinofuranosides. They attempted to manipulate the regioselectivity of AbfD3-catalyzed glycosylation reactions by synthesizing a series of unnatural donors that display structural modifications at their C2 or C5 position. Three donors, *p*-nitrophenyl 5-deoxy-5-fluoro- α -L-arabinofuranoside (**1**), *p*-nitrophenyl 5-deoxy- α -L-arabinofuranoside (**2**), and *p*-nitrophenyl 2-deoxy- α -L-arabinofuranoside (**3**), were synthesized (**Scheme 6**). Subsequently, these glycosides were tested to determine if these deoxygenated analogues could be recognized as substrates by AbfD3 in enzyme-catalyzed reactions. To test the suitability of modified arabinofuranosyl donors for AbfD3-catalyzed synthesis, compounds **1–3** were incubated with AbfD3. The results clearly demonstrated that, besides cleavage, the AbfD3 furanosidase is able to catalyze transglycosylation reactions with C5-modified *p*-nitrophenyl arabinofuranosides and confirmed that AbfD3 possesses the ability



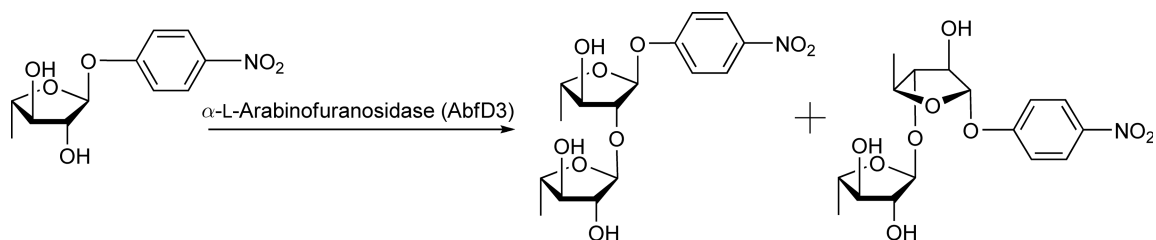
Scheme 5 Novel strategies for glycosidase-catalyzed transglycosylation reactions. (a) Glycosidase-catalyzed synthesis coupled *in situ* to improve overall yields. (b) Utilization of a 6-oxo *p*-nitrophenyl galactose donor as a means to improve the yield. (c) Synthesis of artificial chitin in quantitative yield at a high pH.



| | | | |
|-------------------------------|----|----|----------|
| pNP α -L-Araf | OH | OH | |
| pNP 5-F- α -L-Araf | F | OH | 1 |
| pNP 5-deoxy- α -L-Araf | H | OH | 2 |
| pNP 2-deoxy- α -L-Araf | OH | H | 3 |

Scheme 6 *p*-Nitrophenyl-activated C2- and C5-modified α -L-arabinofuranosides.

to synthesize both α -(1 \rightarrow 2)- and α -(1 \rightarrow 3)-linked regioisomeric disaccharides from the C5 deoxygenated substrate **2** (Scheme 7). However, when the 2-deoxy analogue **3** was incubated with AbfD3 under the same conditions, HPTLC analysis failed to reveal any reaction products. Moreover, the free sugar could not be

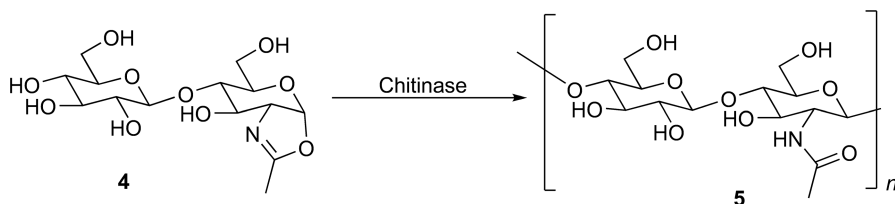


Scheme 7 Structures of the disaccharides that were obtained from the AbfD3-catalyzed condensation of **2**.

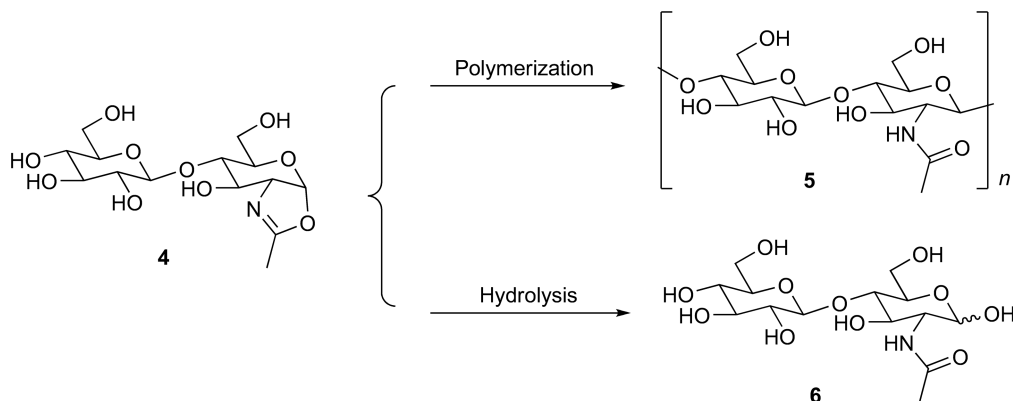
detected with a general acidic reagent, confirming that AbfD3 is not active toward this analogue. The preparation of three original arabinofuranosyl donors helps to demonstrate that the presence of OH-5 is not required for AbfD3 substrate recognition. On the other hand, these results also underline the importance of OH-2 for substrate recognition of α -L-arabinofuranosidase (AbfD3).

Cellulose and chitin are linear polysaccharides of D-glucose (Glc) or N-acetyl-D-glucosamine (GlcNAc) connected through a β -(4) glycosidic linkage. The structural difference between chitin and cellulose is the C2 substituent of a pyranoside unit: an acetamido group on the chitin molecule and a hydroxy group on the cellulose molecule. It was reported that the C2 acetamido groups provide various bioactivities for chitin. A synthetic blend polymer based on these two natural polysaccharides showed good mechanical properties as material for wound healing and for improving the quality of soil and water. Kobayashi *et al.*⁸⁰ have investigated the enzymatic polymerization to a novel cellulose–chitin hybrid polysaccharide, having a glucose–N-acetylglucosamine repeating unit. A sugar oxazoline monomer of Glc- β -(1 \rightarrow 4)GlcNAc (**4**) was designed as a transition-state analogue substrate (TSAS) monomer for chitinase catalysis. Monomer **4** was recognized by chitinase from *Bacillus* sp., giving rise to a cellulose–chitin hybrid polysaccharide (**5**) via ring-opening polyaddition with perfect regio- and stereochemistry (**Scheme 8**).

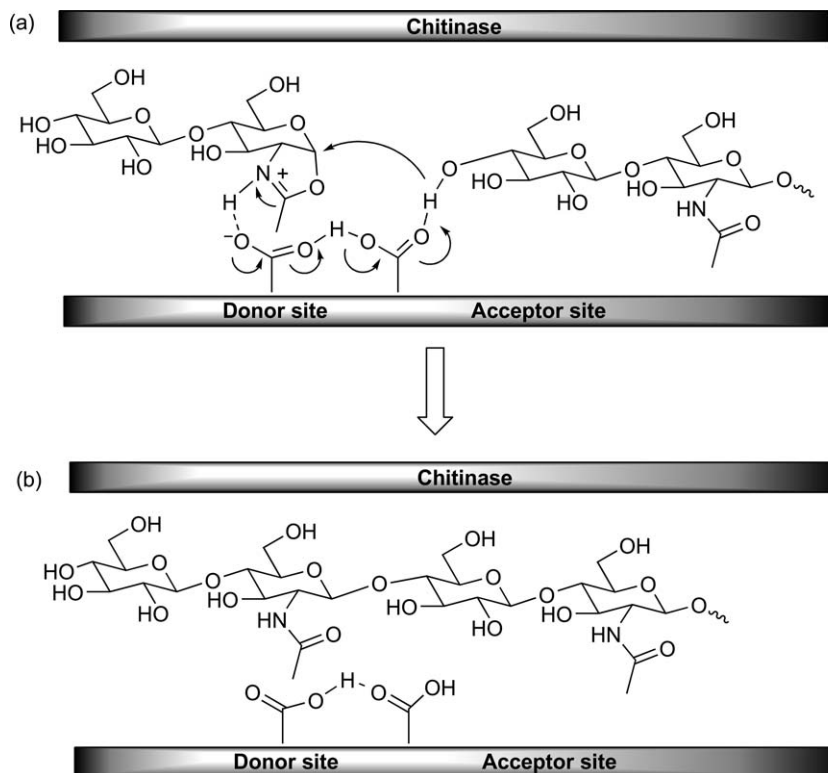
The molecular weight (Mn) of **5** corresponds to 22 saccharide units. With the enzyme catalysis, two kinds of reactions are possible, that is, enzymatic polymerization of **4** to provide the corresponding polymer **5** and hydrolysis of **4** enzymatically and nonenzymatically to afford hydrolysate **6** via the oxazoline ring opening (**Scheme 9**). Monomer **4** bearing an oxazoline structure is an activated, high-energy form of GlcNAc; therefore, it hydrolyzes gradually in aqueous media without enzyme. The mechanism of chitinase-catalyzed



Scheme 8 Enzymatic polymerizations to a cellulose–chitin hybrid polysaccharide.



Scheme 9 Two kinds of possible reactions in the enzymatic reaction of monomer **4**.

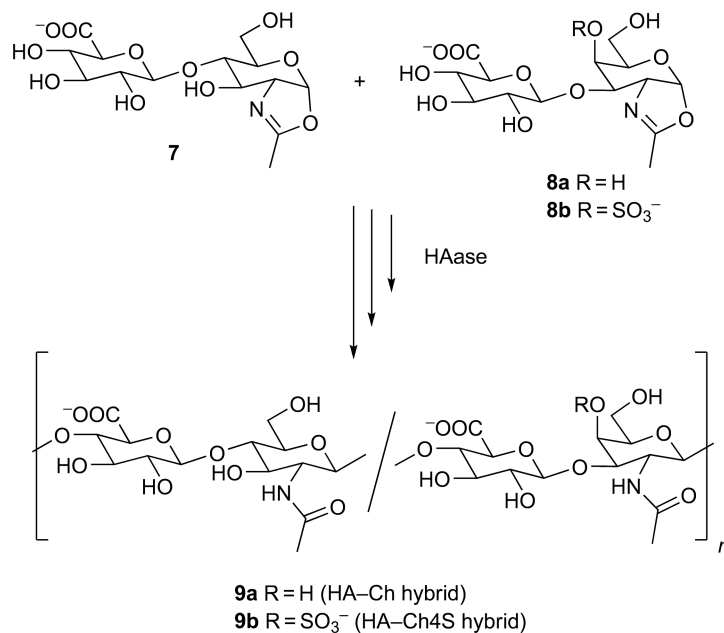


Scheme 10 Reaction mechanism of chitinase from *Bacillus* sp. (a) Formation of the corresponding oxazolinium ion. (b) Formation of the $\beta(1 \rightarrow 4)$ -glycosidic linkage.

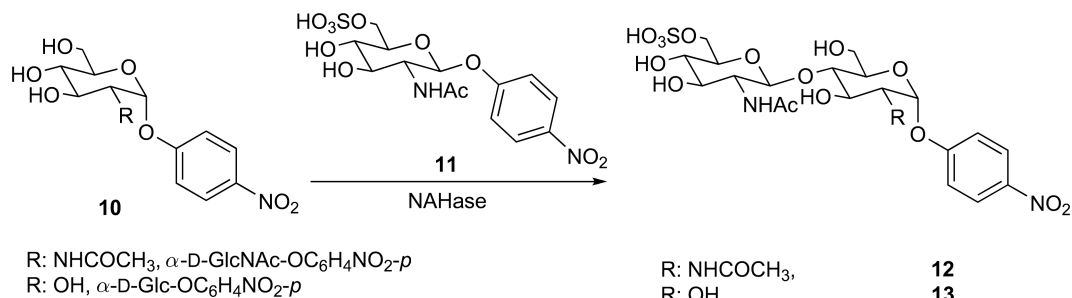
polymerization has been proposed (Scheme 10). In the reaction, monomer 4, which already has an oxazoline moiety, is recognized and protonated at the donor site of the enzyme to form the corresponding oxazolinium ion (stage a). The anomeric carbon of the oxazolinium is attacked from the β -side by the C4 hydroxy group of Glc in another monomer or in the growing chain end placed at the acceptor site, resulting in the formation of a $\beta(1 \rightarrow 4)$ -glycosidic linkage (stage b). Repetition of this glycosylation is a ring-opening polyaddition, providing a cellulose–chitin hybrid polysaccharide. Such an oxazoline-type monomer was named a TSAS monomer.⁷⁴

Recently, a novel hybrid glycosaminoglycan (GAG)⁸¹ consisting of chondroitin sulfate and dermatan sulfate (ChS/DS hybrid) was found *in vivo*. Because of its unique bioactivities, particularly in the development of the central nervous system, this kind of hybrid glycosaminoglycan has attracted much attention.⁸² Kobayashi *et al.*⁸³ investigated the enzymatic polymerization to hybrid glycosaminoglycans. They reported a facile synthesis of hybrid GAGs, particularly HA–Ch (9a) and HA–Ch-4-sulfate (Ch4S) hybrids (9b) by HAase-catalyzed copolymerizations of different sugar oxazoline monomers (7, 8a, and 8b) (Scheme 11).⁸⁴ As mentioned above, the monomers were designed according to the concept of a TSAS monomer. The reaction presented here is one example of copolymerization, in which the monomers producing the polymers with different main-chain structures have been enzymatically copolymerized. According to this method, the hybrid with the sulfate group was efficiently produced, which is difficult to generate via a biosynthetic synthase system. Their study is the first step leading to the tailor-made synthesis of hybrid GAGs with ordered sequences, which will be a potent tool in pharmaceutical uses, and open a new door to easy access to intramolecular polysaccharide hybridization.

β -*N*-Acetylhexosaminidase (NAHase) is a good tool-enzyme in the enzymatic synthesis of unusual complex saccharides. This enzymatic method is attractive because NAHases are generally inexpensive and readily available, and some of them can catalyze the transfer of GlcNAc or GalNAc to glycoside acceptors under controlled conditions.⁸⁵ Recently, Zeng *et al.*⁸⁶ developed a useful method for the synthesis of sulfated disaccharides by using NAHase from *Aspergillus oryzae*. In their report, they carried out the enzyme-catalyzed synthesis of *p*-nitrophenyl sulfated disaccharides with β -D-(6-sulfo)-GlcNAc units by using the transglycosylation of NAHase. When 10 and



Scheme 11 Hyaluronidase-catalyzed copolymerizations to HA–Ch (**9a**) and HA–Ch4S (**9b**) hybrids.



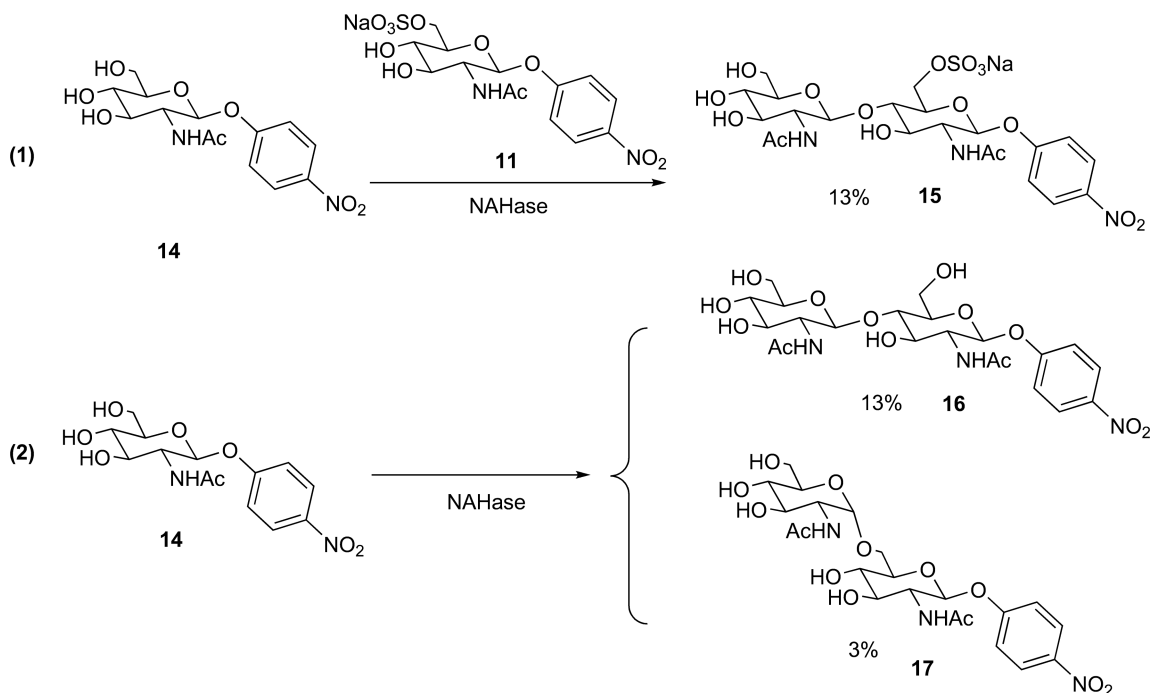
Scheme 12 Synthesis of sulfated disaccharides by β -*N*-acetylhexosaminidase.

α -D-GlcNAc-OC₆H₄NO₂-*p* were used as substrates, as shown in **Scheme 12**, the HPLC chromatogram showed that a new peak as a transfer product was formed during the incubation. The transfer product **11** was isolated from the reaction mixture in 94% yield (based on the donor glycoside). In a similar manner, **12** was synthesized from **10** and α -D-Glc-OC₆H₄NO₂-*p* as substrates.

When substrates **10** and **13** were incubated with NAHase, three transfer products were formed (**Scheme 13**). These results indicated that the efficiency of the transglycosylation was strongly dependent on the molar ratio of the acceptor to donor. With a molar ratio 1:6, the amount of **14** was about six times that with a molar ratio 1:1. It was also proposed that the enzyme preferentially mediated the formation of by-products (**15** and **16**) in the first stage of the enzyme-mediated transfer. Therefore, the source of the sulfate group becomes the donor or the acceptor depending on the easiness of β -D-(6-sulfo)-GlcNAc transfer, indicating that the direction of the transfer can be controlled at reducing or nonreducing terminal.

6.02.2.3 Conclusion

A rule of thumb in classical complex saccharide synthesis is that the addition of each monosaccharide unit requires on an average seven steps. Although there have been striking improvements in synthetic methods in recent years, the synthesis of a trisaccharide, for example, can still take 10 or more steps.⁸⁷ The application of glycosidases to the synthesis of complex hetero-oligosaccharides represents a considerable saving in time and materials, and hence in overall cost.



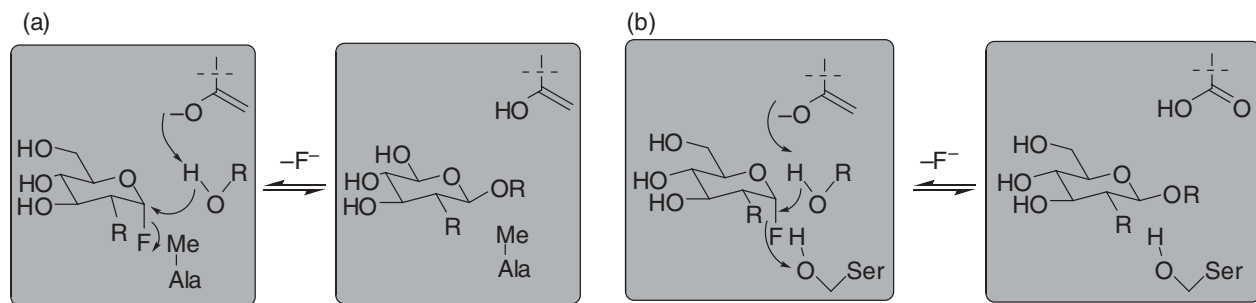
Scheme 13 Synthesis of disaccharides **14**, **15**, and **16** by β -*N*-acetylhexosaminidase.

The need for efficient synthesis of complex saccharides has stimulated major advances in glycosidase research. The more classical approach of substrate engineering aptly complements the construction of ‘superactive’ mutant glycosidases by means of rational design, or through directed evolution strategies using ingenious screening methods, such as a yeast three-hybrid system. The portfolio of genetically engineered glycosidases has expanded with novel activities, and the original glycosynthase concept was broadened to include mutants exercising other catalytic mechanisms. Particular carbohydrate structures that are in high demand can be produced by selective modification of sugar structures using specific glycosidases. The future multidisciplinary research on the interface of biochemistry, synthetic chemistry, and modeling will hopefully result in engineering new glycosidases for industrial uses and in overcoming more glycosidase-induced pathologies, which will be spurred by intensive structural and mechanistic studies on selected enzymes of interest.

6.02.3 Glycosynthases

6.02.3.1 Mechanisms of Glycosynthases

Based on the knowledge of structure–function relationships and the mechanisms of glycosidase-catalyzed reactions, modern mutagenesis technology has played an important role in enhancing the enzymatic activity toward the synthesis of saccharides.⁸⁸ Glycosynthases, a class of glycosidase mutants, have their catalytic nucleophile substituted by a nonnucleophilic residue and are able to catalyze the transglycosylation of glycosyl fluoride donors with opposite anomeric configuration compared to the normal substrate of the wild-type enzyme. As shown in **Scheme 14(a)**, mutation of the catalytic nucleophile disables the enzyme as a hydrolase because no glycosyl–enzyme intermediate can be formed. However, an activated glycosyl donor with an anomeric configuration opposed to that of the donor substrate in the wild-type reaction (i.e., an α -glycosyl fluoride for a β -glycosidase) mimics the glycosyl–enzyme intermediate and is therefore able to react with an acceptor. The cavity created in the active site by mutation of the carboxylate residue acting as nucleophile in the wild-type enzyme by a smaller residue allows binding of the glycosyl fluoride with the opposite anomeric configuration. As with the wild-type enzyme, transglycosylation is kinetically favored, but the



Scheme 14 Catalytic mechanisms of glycosynthases. (a) Mechanism of glycosynthases (E358A mutant). (b) Mechanism of glycosynthases (E358S mutant).

transglycosylation reaction is now irreversible because of the lack of the catalytic nucleophile; the product is no longer hydrolyzed and accumulates to give high transglycosylation yields. This approach represents an important improvement compared to kinetically controlled transglycosylation by wild-type enzymes.

Withers and coworkers⁸⁹ described the first glycosynthase by changing the glutamate residue at position 358 of a β -glucosidase from *Agrobacterium* sp. (Abg) to an alanine. With that mutant, β 1,4-glucosides up to pentasaccharides have been obtained in gram scale using α -glucosyl fluoride as donor substrates. The same concept has been adopted for other glycosidases. An E134A mutant of a 1,3-1,4- β -glucanase from *B. licheniformis*, obtained by Malet and Planas,⁹⁰ was able to catalyze the regio- and stereospecific glycosylation of diverse glucosides with α -laminaribiosyl fluoride affording 90% yields without any hydrolysis of the products.

Driguez and coworkers⁹¹ showed that an E197A mutant of a retaining cellulase endoglucanase I (Cel7B) from the fungus *Humicola insolens* was highly efficient for the synthesis of β 1,4-linkage for complex saccharides.

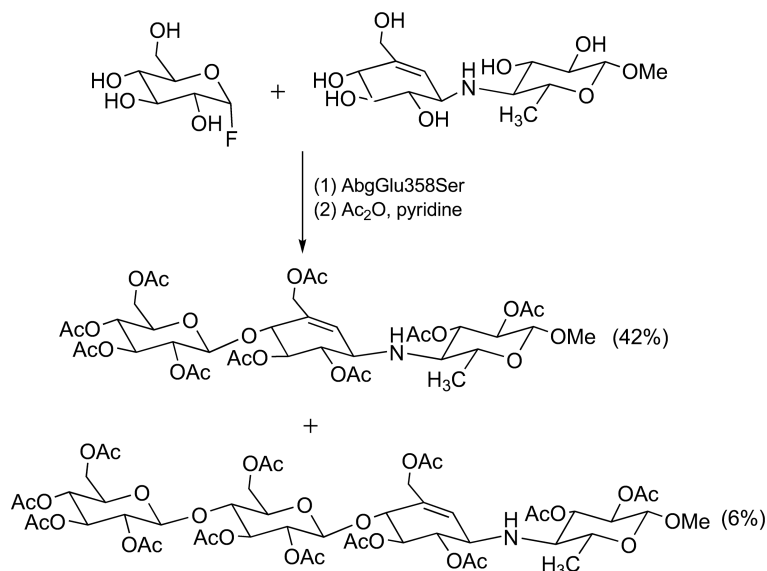
Nevertheless, the reactions catalyzed by the alanine mutants are relatively slow; thus, large quantities of glycosynthases and/or extended incubation times are required. To enhance the reaction rates, an E358S mutant was developed (**Scheme 14(b)**). The new glycosynthase was superior to the original E358A mutant with a dramatic 24-fold increase in the rate constant. Thus, it resulted in improved yields (>80%), reduced reaction times, and an enhanced synthetic repertoire in the synthesis of galactosides and glucosides. For example, *p*-nitrophenyl β -*N*-acetylglucosamine (PNP-LacNAc), a valuable precursor to various cell-surface antigens, was obtained from a weak acceptor *p*-nitrophenyl β -*N*-acetylglucosamine (PNP-GlcNAc) with this new mutant.⁹² The rationale for the improved catalytic activity is that the serine hydroxyl in the new mutant stabilizes the anomeric fluorine of the α -glycosyl fluoride donor through a hydrogen bond in the glycosylation transition state.

6.02.3.2 Glycosynthases in Carbohydrate Synthesis

Directed mutagenesis of the catalytic nucleophile (E338) of the *Thermus thermophilus*-retaining Tt- β -Gly glycosidase afforded the E338A, E338S, and E338G mutants, which were unable to catalyze the hydrolysis of their transglycosylation products.⁹³ These mutants were also inactive in the hydrolysis of the α -D-glycopyranosyl fluorides used as donors, enabling the regiospecific glycosylation of galactosyl- and glucosyl- β -(1 \rightarrow 3)-glycosides in high yields. Complex saccharides with (1 \rightarrow 3)- β -glycosidic linkages were obtained when substituted or unsubstituted α -laminaribiosyl fluoride donors were condensed with saccharide β -D-hexopyranoside acceptors using a (1 \rightarrow 3)- β -D-glycosynthase from barley.^{94,95}

A new class of glycosynthases,⁹⁶ Ta- β -GlyE386G from *Thermosphaera aggregans* and Ss- β -GlyE387G from *Sulfolobus solfataricus*, were applied to synthesize 4-methylumbelliferyl disaccharides and for the galactosylation of α - and β -xylosides of 4-penten-1-ol. The results showed that these two glycosynthases had the ability to construct glycosylated products in the presence of low excesses of acceptors at pH below neutrality. Recently, another new method for the reactivation of Ss- β -GlyE387G and two novel mutated β -glycosidases from *T. aggregans* (Ta- β -Gly) and *Pyrococcus furiosus* (CelB) was reported.⁹⁷ At pH 3.0 and low concentrations of sodium formate buffer, the three glycosynthases showed k_{cat} values that were similar to those of the wild-type enzymes and 17-fold higher than those observed under the usual reactivation conditions. Moreover, at acidic pH, the three reactivated mutants showed wide substrate specificity and improved efficiency in the synthetic glycosylation reactions. The data reported suggested that the reactivation conditions modified the ionization state of the catalytic acid/base carboxylic acid in the active site of these enzymes. The glycosynthases produced 2NP-oligosaccharides from 2-nitrophenyl-Glc. Kinetic analyses of the *H. insolens* Cel7B E197A and E197S glycosynthases showed that the E197S mutant is considerably more active than the corresponding alanine mutant due to a 40-fold increase in k_{cat} .⁹⁸

Currently, glycosynthases have become available for the construction of β -(1 \rightarrow 3)-, β -(1 \rightarrow 4)-, and β -(1 \rightarrow 6)-linked complex saccharides. The natural substrate structure of the original glycoside hydrolase decides the kind of donor sugar. The acceptor is often an aryl glycoside that exhibits good binding property in both the +1 and +2 subsites of the original glycoside hydrolases and glycosynthases. Fortunately, glycosynthases are capable of constructing unpredictable linkages for which they were not designed before. Watts and coworkers⁹⁹ explored the use of 6-*O*-aryl ethers or esters as acceptors, whose alternative acceptor substrate-binding orientation might present an alternative hydroxyl group for functionalization by the glycosyl donor, compared



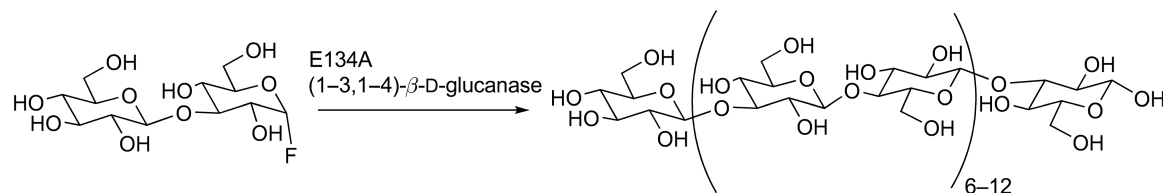
Scheme 15 Enzymatic synthesis of oligosaccharides.

with the respective aryl glycosides. Glycosynthase (Abg E358S)-mediated glycosylations between an α -D-galactopyranosyl fluoride donor and 6-O-benzyl-, 6-O-benzoyl-, or 6-O-(4-nitrobenzyl)-D-glucopyranose acceptors were carried out to build (1 \rightarrow 2)- β - and (1 \rightarrow 3)- β -D-glycosylated disaccharide products rather than the characteristic β -(1 \rightarrow 4)-linkage when aryl β -D-glucopyranoside was used as an acceptor. Afterward, the glycosynthase AbgGlu358Ser was used to construct the trisaccharide and tetrasaccharide derivatives of the cellulose inhibitor methyl- β -acarviosin. Interestingly, both the trisaccharide and the tetrasaccharide proved to be more effective inhibitors of various cellulases than the parent disaccharide (**Scheme 15**).

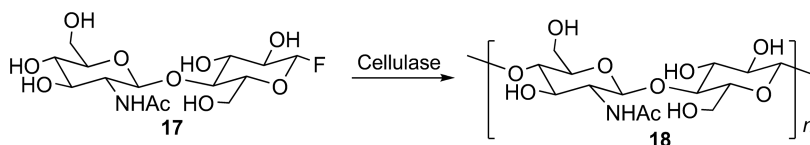
The catalytic activity of the glycosynthase mutated from *Bacillus licheniformis* (1 \rightarrow 3, 1 \rightarrow 4)- β -D-glucanase to catalyze self-condensation of sugar donors was tested for polymerization of α -laminaribiosyl fluoride to afford a novel linear glucan with a repeating unit of 4- β -Glc-3- β -Glc.¹⁰⁰ The product consisted of a mixture of oligosaccharides, the most abundant having a degree of polymerization (DP) of 12, as shown in **Scheme 16**.

Regiospecifically, the endoglycosynthase is capable of constructing β -(1 \rightarrow 4)-glycosidic bond with α -glycosyl fluoride donors and oligosaccharide acceptors containing glucose or xylose on the nonreducing end. Yields are 70–90% with aryl monosaccharide and cellobioside acceptors, but only 25–55% with laminaribiosides, due to competitive inhibition of the β -(1 \rightarrow 3)-linked saccharide acceptor for the donor subsites of the enzyme.¹⁰¹ Directed mutagenesis of a glycosynthase from *Agrobacterium* sp. has been developed to increase both its catalytic activity (up to 40-fold) and its donor substrate tolerance (e.g., α -Gal-F, α -Man-F, α -Xyl-F).¹⁰²

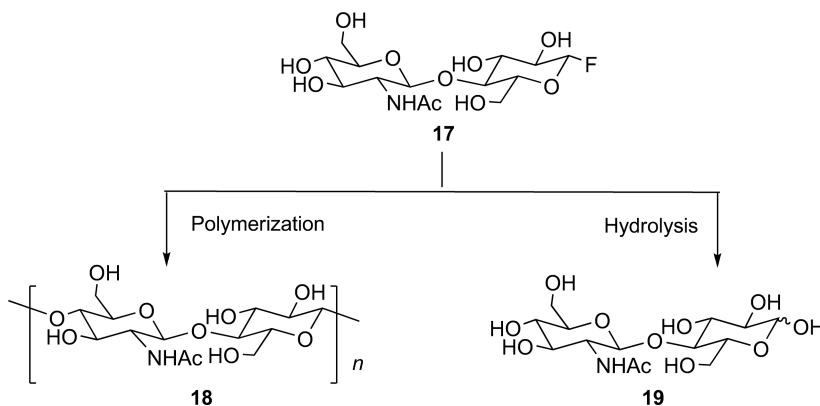
Kobayashi *et al.*⁸⁰ have also investigated other modes of enzymatic polymerization to produce a novel cellulose–chitin hybrid polysaccharide. A sugar fluoride monomer of GlcNAc β (1 \rightarrow 4)Glc (**17**) was designed as a TSAS monomer for polymerization catalyzed by cellulase from *Trichoderma viride*. In the polymerization, monomer **17** was recognized by cellulase from *T. viride*, leading to a cellulose–chitin hybrid polysaccharide **18** with perfect regioselectivity and stereochemistry (**Scheme 17**).



Scheme 16 Structure of a novel linear glucan.



Scheme 17 Enzymatic polymerizations to a cellulose–chitin hybrid polysaccharide.



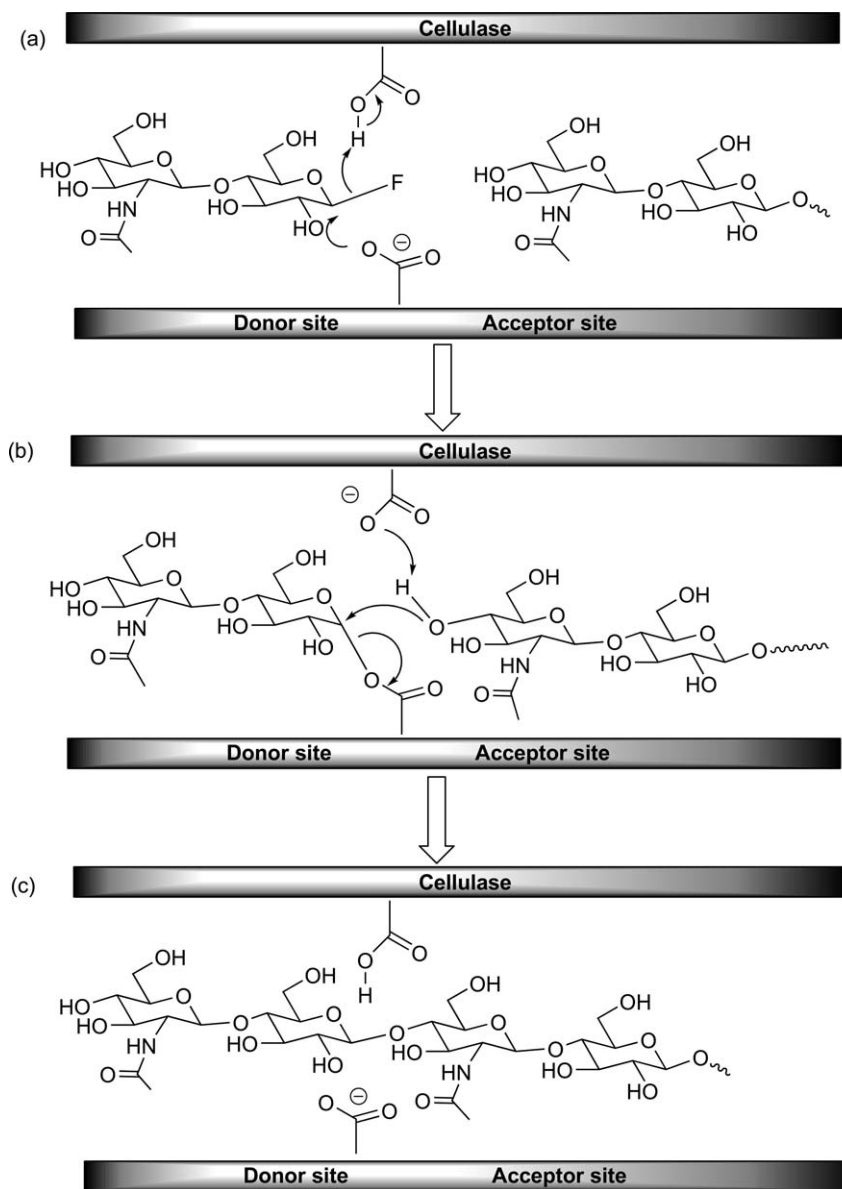
Scheme 18 Two kinds of reactions that are possible during enzymatic reaction of monomer **17**.

Monomer **17** is a high-energy substrate monomer activated at the anomeric carbon by a fluorine atom, which acts as a good leaving group, giving rise to a corresponding oxocarbenium intermediate.¹⁰³ Two kinds of reactions are possible during enzymatic reaction: enzymatic polycondensation of **17** to polymer **18** and hydrolysis of **17** to **19** (Scheme 18).

Cellulase from *T. viride* belonging to the glycoside hydrolase family 5 normally catalyzes cleavage of the (1→4)- β -glucoside linkage in cellulose. The reaction mechanism of cellulase has also been proposed (Scheme 19). In the polymerization, first, monomer **17** is placed at the donor site of the enzyme through recognition by cellulase. The high-energy substrate is then nucleophilically attacked by the carboxylate (catalytic nucleophile) from the α -side (stage a), resulting in the formation of a glycosyl enzyme intermediate (or transition state) (stage b). The anomeric carbon of the intermediate is attacked nucleophilically from the β -side by the 4-OH group of GlcNAc in the growing chain end or in another monomer molecule placed at the acceptor site (stage b), resulting in the formation of a $\beta(1 \rightarrow 4)$ -glycosidic linkage (stage c). Surprisingly, this type of enzymes could easily recognize the monomer despite their high substrate-specific character. Recent investigations revealed that the C2 substitutions of substrates located at -2 and $+1$ subsites seem to be not important for the cellulase catalysis.

The nonnucleophilic mutant E383A β -glucosidase from *Streptomyces* sp. has proven to be an efficient glycosynthase enzyme, catalyzing the condensation of α -glucosyl and α -galactosyl fluoride donors to a variety of acceptors. Faijes *et al.*¹⁰⁴ investigated the relationship between the regioselectivity of the new glycosidic linkage and the acceptor substrate structure. Glycosynthase catalytic activity of the E383A mutant β -glucosidase from *Streptomyces* sp. was assayed by a reaction between α -glucosyl fluoride and *p*-nitrophenyl glycoside, which produced a disaccharide. As shown in Table 2, their results underline that the regioselectivity of the new glycosidic linkage depends unexpectedly on the acceptor substrate. With aryl monosaccharide acceptors, $\beta(1 \rightarrow 3)$ -linked disaccharides are obtained in good yields, thus expanding the synthetic products available with the *exo*-glycosynthases. With xylopyranosyl acceptors, regioselectivity is poorer and results in the formation of a mixture of $\beta(1 \rightarrow 3)$ - and $\beta(1 \rightarrow 4)$ -linked products. The regioselectivity of glycosidic bond formation is switched from $\beta(1 \rightarrow 3)$ with monosaccharide acceptors to $\beta(1 \rightarrow 4)$ with disaccharides. This suggests that enzyme–ligand interactions in subsite $+2$ have a decisive effect on the orientation of the acceptor.

Withers¹⁰⁵ has reported a glycosynthase, derived from a retaining xylanase, that synthesizes a range of xylo-oligosaccharides. The catalytic domain of the retaining endo-1,4- β -xylanase from *Cellulomonas fimi* (CFXcd) was successfully converted to the corresponding glycosynthase using site-directed mutagenesis, bearing Gly or Ser at the catalytic nucleophile position (Glu235). The glycosynthase catalytic activity of the purified CFXcd

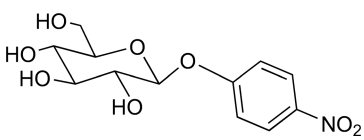
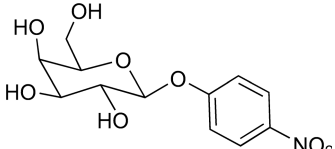
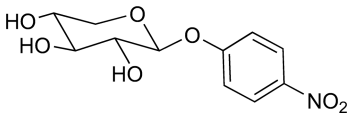
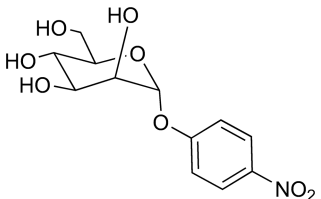
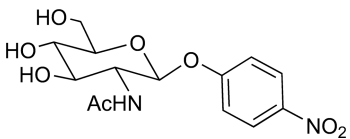
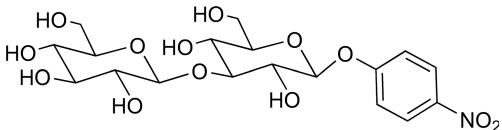


Scheme 19 Reaction mechanism of cellulase from *Trichoderma viride*. (a) Formation of a glycosyl enzyme. (b) Nucleophilic attack of the intermediate from the β -face. (c) Formation of a $\beta(1 \rightarrow 4)$ -glycosidic linkage.

mutants CFXcd-E235G was directly tested with α X2F as a donor and PNPX2 (xylobioside) as an acceptor. The mutant enzyme (CFXcd-E235G) was found to catalyze the transfer of a xylobiosyl moiety from α -xylobiosyl fluoride to either *p*-nitrophenyl β -xylobioside or benzylthio β -xylobioside to afford oligosaccharides ranging in length from tetra- to dodecasaccharides. TLC analysis of the reaction between α X2F and PNPX2 catalyzed by CFXcd-E235G revealed that reaction had terminated yielding a range of oligosaccharides from PNP β -D-xylotetraoside (PNPX4) up to PNP β -D-xylododecaoside (PNPX12), corresponding to five additions of X2 (Table 3). This expansion of the glycosynthase repertoire allows the regio- and stereoselective synthesis of $\beta(1 \rightarrow 4)$ -linked xylo-oligosaccharides in high yields. Their methodology provides a convenient process for the production of linear xylo-oligosaccharides of defined length.

Humicola insolens mutant Cel7B E197A was regarded as a powerful endoglycosynthase for the synthesis of artificial cellulose and its derivatives.^{91,106,107} Cottaz and coworkers showed that H209 and A211 are the main residues located around the position 2 of a β -D-glucosaminyl unit positioned in +1 subsite.¹⁰⁸ Encouraged by

Table 2 Glycosynthase reactions of E383A β -glucosidase using α -glucosyl fluoride donor and different acceptors

| Acceptor | Disaccharide product (% yield) | Reactivity (%) |
|---|--|----------------|
|  | β -(1 \rightarrow 3) (96) | 100 |
|  | β -(1 \rightarrow 3) (100) | 3.9 |
|  | β -(1 \rightarrow 3) (54) β -(1 \rightarrow 4) (28) | 0.6 |
|  | β -(1 \rightarrow 3) (53) | 1 |
|  | Nr | |
|  | β -(1 \rightarrow 4) (80) | 21 |

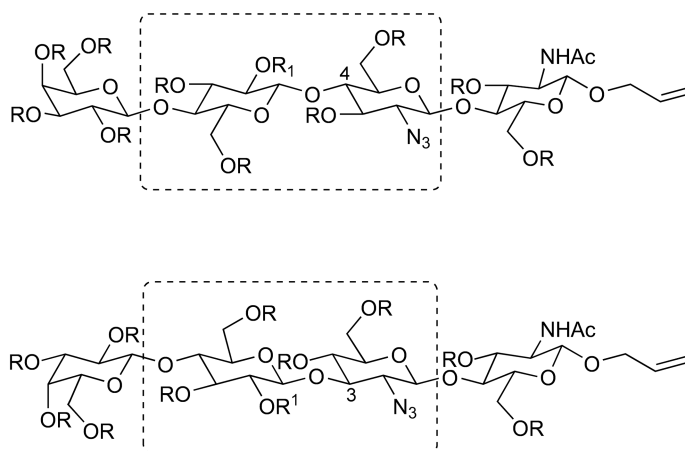
Nr, no reaction.

this result, they prepared and characterized three mutants of *H. insolens* Cel7B E197A glycosynthase: Cel7B E197A H209A, Cel7B E197A H209G, and Cel7B E197A H209A A211T.¹⁰⁹ These *H. insolens* Cel7B glycoside hydrolase mutants were evaluated for the coupling of lactosyl fluoride donor on *O*-allyl *N*^I-acetyl-2^{II}-azido- β -chitobioside acceptor. Their results showed that double mutants Cel7B E197A H209A and Cel7B E197A H209G preferentially had the abilities to catalyze the formation of a β -(1 \rightarrow 4) linkage between the two disaccharides, while single mutant Cel7B E197A and triple mutant Cel7B E197A H209A A211T produced predominantly the β -(1 \rightarrow 3)-linked tetrasaccharide. From these structure analyses shown in **Scheme 20**, it can be assumed that H209A or H209G mutation probably created a pocket in the +1 subsite such that the 2-azido-glucosyl residue can be appropriately positioned to form a β -(1 \rightarrow 4) bond with the lactosyl donor. In contrast, for the triple mutant Cel7B E197A H209A A211T, the +1 subsite has been remodeled in such a manner that the *O*-allyl *N*^I-acetyl-2^{II}-azido- β -chitobioside acceptor can bind in a single mode to catalyze exclusively the

Table 3 Yields of glycosylation products catalyzed by CFXcd-E235G

| [β -D-Xyl-(1-4)- β -D-Xyl-(1-4)-] $_n$ - β -D-Xyl-(1-4)-Xyl- β -R | R = PNP | | R = BT | |
|--|---------------|-----------|---------------|-----------|
| | Quantity (mg) | Yield (%) | Quantity (mg) | Yield (%) |
| n = 0 | 2.8 | 14 | 2.5 | 13 |
| n = 1 | 10.2 | 30 | 8.3 | 26 |
| n = 2 | 9.0 | 19 | 10.5 | 23 |
| n = 3 | 4.5 | 7 | 7.1 | 12 |
| n = 4 | 2.7 | 4 | 2.8 | 4 |
| n = 5 | 1.1 | 1 | 0.9 | 1 |
| Total purification yields of aryl saccharides | | 75 | | 79 |
| Total percentages of transfer products | | 81 | | 84 |

BT, benzylthio; PNP, 4-nitrophenyl.

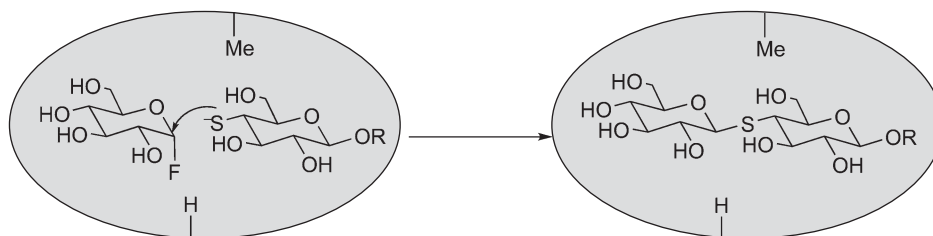
**Scheme 20** Two kinds of products catalyzed by *Humicola insolens* mutant.

formation of the β -(1 \rightarrow 3) products. Their result constitutes the first modulation of the regioselectivity through site-directed mutagenesis for an endoglycosynthase for enzyme-catalyzed complex saccharide synthesis.

Glycosynthases are mutants of glycosidases, which lack the nucleophilic active site carboxylic acid; however, thioglycosylases are variants of retaining glycosidases in which the acid–base catalyst has been mutated. These mutants of retaining glycosidases have the ability to couple sufficiently reactive donors with nucleophilic thiosugar acceptor to produce a thioglycosidic linkage with retention of configuration.¹¹⁰ The highly activated leaving group does not require protonation to facilitate formation of the covalent glycosyl enzyme intermediate. Absence of the catalytic base greatly decelerates the rate of deglycosylation by the hydroxyl moiety of an acceptor substrate. This is circumvented by the use of acceptors containing a thiol at the attacking position where the nucleophilic thiol or thiolate can facilitate deglycosylation (i.e., selective thioglycoside formation) in the absence of the acid–base catalyst. A logical extension of glycosynthases and thioglycosylases are thioglycosynthases, that is, double mutant, retaining glycosidases that lack both the catalytic nucleophile and the catalytic acid/base residues (**Scheme 21**). These can be used to efficiently catalyze thioglycoside formation from a glycosyl fluoride donor and thiosugar acceptors.¹¹¹

6.02.3.3 Conclusion

Glycosynthases are classified as mutated retaining glycosidases where the active site carboxylate nucleophile has been replaced by a nonnucleophilic amino acid side chain (such as alanine). Synthesis of oligosaccharides by the glycosynthase enzymes, which do not hydrolyze the products and use inexpensive



Scheme 21 General mechanism for thioglycosyltransferases.

sugars, is an alternative to scale-up of oligosaccharide synthesis. Glycosynthases derived by mutation of catalytic acid–base residue of retaining glycosidases are able to couple donors that are sufficiently reactive in the absence of the acid catalyst with nucleophilic thiosugar acceptor to produce a thioglycosidic linkage with retention of configuration.

Through the use of glycosynthase mutants, it is possible to access different mimetics of natural polysaccharides and artificial polysaccharides with well-defined structures and morphologies. Compared to polymerization by native glycosidases with activated donors as glycosyl fluorides or derivatives, the main advantage of the glycosynthase strategy is the lack of product hydrolysis, higher polymer yields, and higher degrees of polymerization. The impressive amount of glycosidases available clearly indicates that the potential biodiversity of the glycosynthases is still largely unexplored and that potential applications of these enzymes will increase in the near future. The glycosynthase approach is poised to be further investigated for its application to the large-scale production of oligosaccharides. Results obtained to date suggest a promising future for these new biocatalysts. Further directed evolution as well as discovery of new glycosidases with different specificities will expand the scope of glycosynthases.

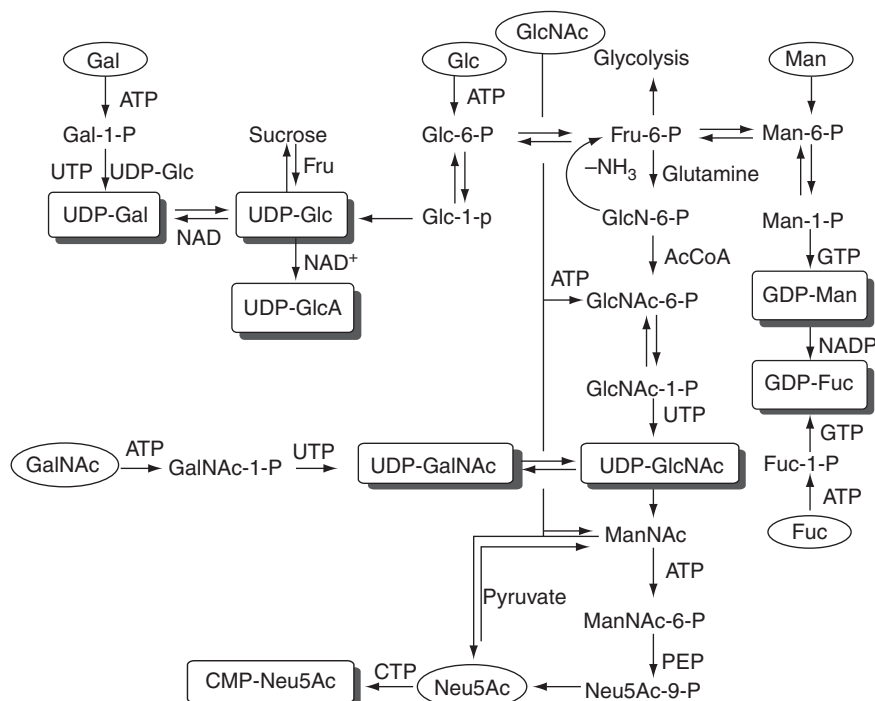
6.02.4 Glycosyltransferases

6.02.4.1 Sugar Nucleotide Biosynthetic Pathway

The structures of oligosaccharides and glycoconjugates can be quite complex; however, the number of important saccharide sequences is limited, and they are made of monosaccharide building blocks. The common monosaccharides are only nine: glucose (Glc), galactose (Gal), glucuronic acid (GlcA), *N*-acetylglucosamine (GlcNAc), xylose (Xyl), *N*-acetylgalactosamine (GalNAc), mannose (Man), fucose (Fuc), and sialic acid (Neu5Ac). To construct the sugar chains, the monosaccharides have to be activated by attachment to nucleoside phosphates. The role of sugar nucleotides in complex saccharide synthesis was first discussed in 1950 by the Nobel laureate L. F. Leloir. He showed that a nucleotide triphosphate such as UTP reacts with a glycosyl 1-phosphate to form the high-energy donor that could participate in complex saccharide synthesis. After half a century, the biosynthetic pathways of nine common sugar nucleotides (UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc, UDP-GlcA, UDP-Xyl, GDP-Man, CMP-Neu5Ac, and GDP-Fuc) are now well established. They are also known as primary sugar nucleotides, for they are generated *in vivo* from sugar-1-phosphate. Other sugar nucleotides are known as secondary sugar nucleotides because they are synthesized by modification of primary sugar nucleotides. All of the biosynthetic pathways are not equally active in different types of cells. However, these pathways have been conserved, to some extent, throughout evolution across both prokaryotic and eukaryotic species. The general biosynthesis pathway provides the basic guidelines on how to construct them *in vitro* (Scheme 22).^{112,113}

Regardless of the sugar and its origin, all monosaccharides must be activated by a kinase (reaction 1 below), or generated from a previously synthesized sugar nucleotide (reactions 2 and 3 below).

1. Sugar + ATP → sugar-P + NTP → sugar-NDP + PP_i
2. Sugar (A)-NDP + sugar (B)-NDP
3. Sugar (A)-NDP + sugar (B)-1-P → sugar (B) NDP + sugar (A)-1-P



Scheme 22 The whole biosynthetic pathway of sugar nucleotides. ATP, adenosine triphosphate; Gal-1-P, galactose-1-phosphate; UTP, uridine triphosphate; UDP, uridine diphosphate; NAD, nicotinamide adenine dinucleotide; Fru, fructose; AcCoA, acetyl coenzyme-A; PEP, phosphoenolpyruvate; CTP, cytidine triphosphate; NADP, nicotinamide adenine dinucleotide phosphate; GTP, guanosine triphosphate.

For Glc, GlcNAc, and Man, the routes to the corresponding sugar nucleotides are (1) sugar to sugar-6-P by a kinase, (2) sugar-6-P to sugar-1-P by a mutase, and (3) sugar-1-P to NDP-sugar by a pyrophosphorylase.

Till date, there are no C6 sugar kinases reported for Gal and GalNAc and Fuc. For fucose, the reason is obvious that there is no 6-OH group. For Gal and GalNAc, the 6-OH is quite sterically hindered for enzymatic phosphorylation due to the existence of the axial 4-OH group. Thus, these three monosaccharides are phosphorylated on the anomeric position, converting them directly to sugar-1-phosphates. There are several enzymes that interconvert different sugar nucleotides, such as UDP-Gal 4'-epimerase (GalE), UDP-GalNAc 4'-epimerase (GalNAcE), UDP-Glc 4'-epimerase, UDP-GlcNAc 4'-epimerase, and galactose-1-phosphate uridylyltransferase. It should be noted that UDP-GlcA universally comes from the oxidation of UDP-Glc. These enzymes are highly useful in sugar nucleotide regeneration systems, since they can be employed to convert one sugar nucleotide into another, or provide both sugar nucleotides at the same time. The biosynthesis of CMP-NeuAc is different from the rest of the common sugar nucleotides. CMP-NeuAc can be synthesized directly from Neu5Ac and CTP without a sugar-1-phosphate intermediate. Neu5Ac may come from two source pathways. The first pathway involves three steps from ManNAc to Neu5Ac, and consumes one molecule of ATP and one molecule of phosphoenolpyruvate (PEP). The second pathway produces Neu5Ac directly from ManNAc via an aldolase-catalyzed condensation with pyruvate. Since sugar nucleotides serve only as cofactors in the overall glycosylation reaction, the best way to minimize the donor expense would be to recycle them *in situ*. Such systems for nine common sugar donors were broadly applied to reduce the cost of sugar nucleotide donors and to prevent the inhibitory effects of nucleotides accumulated during the glycosyltransferase-catalyzed reactions.⁸⁸

6.02.4.2 Leloir Glycosyltransferases

6.02.4.2.1 Basic principle

In nature, glycosyltransferases are responsible for glycoside bond formation. They catalyze the transfer of a monosaccharide from a sugar nucleotide donor (in Leloir glycosyltransferases) to an acceptor, acting processively in homopolysaccharide biosynthesis or in combination with other glycosyltransferases to produce heteropolysaccharides. These enzymes can be further divided into two groups based on differences in the substrate structures: the transferases of the Leloir pathway and those of non-Leloir pathways. The Leloir pathway enzymes require sugar nucleotides as glycosylation donors, while non-Leloir glycosyltransferases, which utilize sugar phosphates as glycosyl donors, have been applied synthetically in a much more limited fashion and will not be covered further. The Leloir transferases are responsible for the synthesis of most glycoconjugates in cells, especially in mammalian systems.⁸⁸

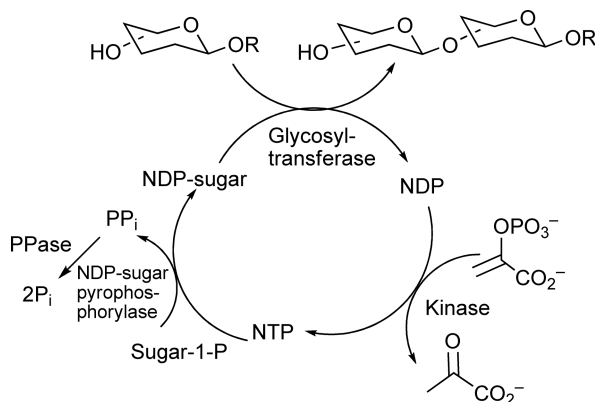
Glycosyltransferases of the Leloir pathway are responsible for the synthesis of most cell-surface glycoforms in mammalian systems.¹¹⁴ These enzymes transfer a given carbohydrate from the corresponding sugar nucleotide donor substrate to a specific hydroxyl group of the acceptor sugar. These enzymes exhibit very strict stereo- and regioselectivity. Moreover, they can transfer with either retention or inversion of configuration at the anomeric carbon of the sugar residue. A large number of eukaryotic glycosyltransferases have been cloned to date, and, in general, they exhibit exquisite linkage and substrate specificity. It is remarkable that such a considerable number of mammalian enzymes have converged on only nine general sugar nucleotides as glycosyl donor substrates. Glucosyl- and galactosyltransferases employ substrates activated with uridine diphosphate as the anomeric leaving group (α -UDP-Glc, α -UDP-GlcNAc, α -UDP-GlcUA, α -UDP-Gal, α -UDP-GalNAc), whereas fucosyl- and mannosyltransferases utilize guanosine diphosphate (β -GDPFuc, α -GDP-Man). Sialyltransferases are unique in that the glycosyl donor is activated by cytidine monophosphate (β -CMP-NeuAc). Sialyltransferases are unique in the sense that the glycosyl donor is activated by cytidine monophosphate (CMP-Neu5Ac). Preparative-scale syntheses of relevant sugar nucleotides have been developed previously, and most are now commercially available.¹¹⁵

As Leloir glycosyltransferases are highly regio- and stereospecific with respect to glycosidic linkage formation and provide products in high yield, they are often the catalyst of choice for glycoconjugate synthesis. However, glycosyltransferase-based syntheses suffer from two major drawbacks. First of all, the nucleoside diphosphates generated during the reaction are potent glycosyltransferase inhibitors. Furthermore, sugar nucleotide expense can become a burden if large-scale synthesis is required. The feedback inhibition problem can be solved by the addition of a phosphatase into the reaction, which results in breakdown of the NDP product.¹¹⁶ However, to circumvent sugar nucleotide expense and avoid product inhibition simultaneously, multienzyme recycling systems have been developed. In this case, NDP-sugars are required in only catalytic quantities, as they are generated *in situ* from inexpensive starting materials. Moreover, the NDPs are recycled to NDP-sugars, thereby avoiding product inhibition (Scheme 23).¹¹⁷

Glycosyltransferase availability is occasionally considered a third drawback. Although many glycosyltransferases can now be purchased from commercial sources (Table 4), an enzyme specific for every desired glycosidic linkage is not available at the present time. In some instances, those that are not available commercially can be isolated from tissue sources.^{118,119}

6.02.4.2.2 Enzyme-based complex saccharide synthesis

Recent advances in the area of enzymatic oligosaccharide synthesis are emerging from the identification and cloning of a large number of bacterial glycosyltransferases with many different donor, acceptor, and linkage specificities.^{120–122} Most eukaryotic glycosyltransferases are not active within prokaryotic expression system due to the absence of posttranslational modifications, including glycosylation. Bacterial glycosyltransferases, for example, are normally not glycosylated proteins. It has been demonstrated that these enzymes are more easily expressed as a soluble and active form in prokaryotic expression system such as *E. coli*. In addition, bacterial glycosyltransferases seem to have relatively broader acceptor substrate specificities, thereby offering



Scheme 23 Recycling of sugar nucleotides.

Table 4 Commercially available glycosyltransferases

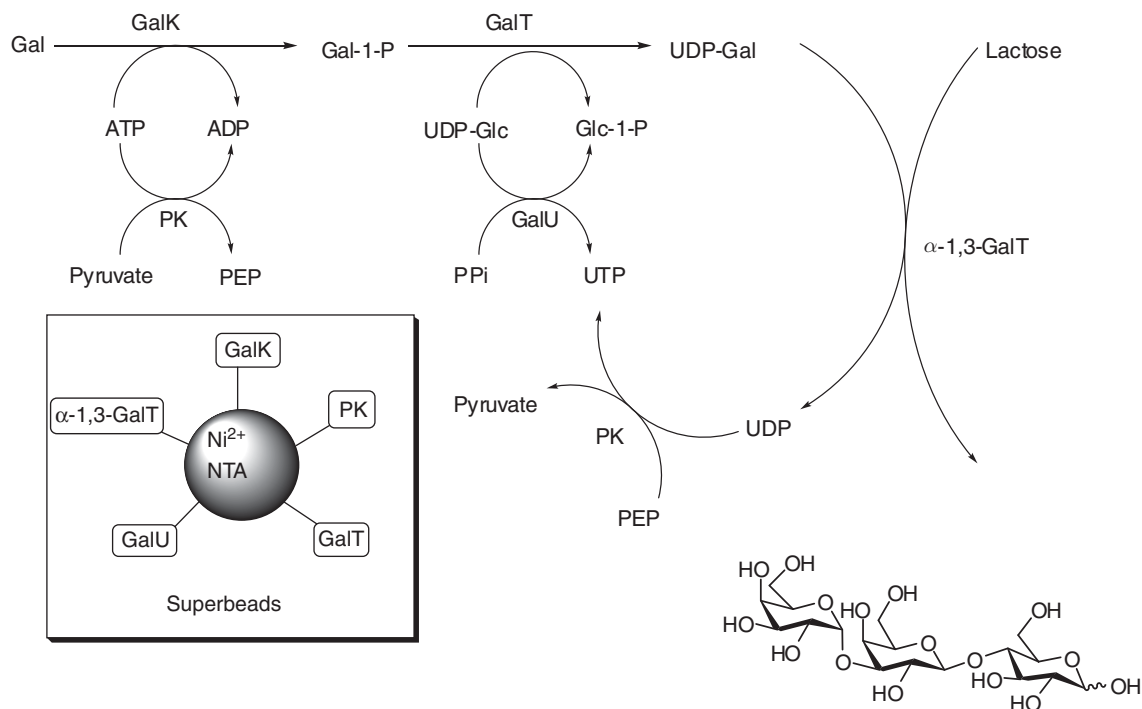
| Glycosyltransferase | Supplier |
|---|--|
| β 1,4-Galactosyltransferase | Calbiochem, Fluka, Sigma, Glyko, Worthington |
| α 1,3-Galactosyltransferase | Calbiochem |
| α 2,6-Sialyltransferase | Calbiochem, Sigma |
| α 2,3-Sialyltransferase | Calbiochem |
| α 1,3-Fucosyltransferase (III, IV, V, VI, VII) | Calbiochem |
| α 1,2-Mannosyltransferase | Calbiochem |

tremendous advantages to mammalian enzymes in the chemoenzymatic synthesis of complex saccharides and their analogues for the development of antiadhesion therapies for infectious diseases.^{123,124}

The availability of new glycosyltransferases has increased the demand for sugar nucleotides, which have been a problem in the production of oligosaccharides and glycoconjugates. One way to overcome this hurdle is to use a multiple enzymatic system with *in situ* UDP-GalNAc regeneration, from inexpensive starting materials. Since the pioneering work by Wong *et al.* on *in vitro* enzymatic synthesis of *N*-acetyllactosamine with the regeneration of UDP-galactose, several glycosylation cycles with regeneration of sugar nucleotides have been developed using either native or recombinant enzymes.^{40,125–128}

It has also been demonstrated that expensive substrates such as UDP-Gal can be readily prepared *in situ* by enzymatic conversion of the relatively inexpensive sugar nucleotide uridine 5'-diphospho- α -D-glucopyranose (UDP-Glc) using a UDP-Gal 4-epimerase enzyme. This system, coupled with an appropriate UDP-Gal transferase, provides more economic access to enzymatically galactosylated compounds. In these multienzyme systems, to increase enzyme efficiency and also avoid multiple fermentations for separate enzyme preparations, fusion proteins^{126–128} have been constructed that contain both the Gal-epimerase and Gal-transferase enzymes. The use of these fused enzyme systems has increased in the recent years as their catalysis of sequential reactions can have a kinetic advantage over the mixture of two separated enzymes since the product of the first enzyme travels a shorter distance before being captured by the next enzyme in the sequence.

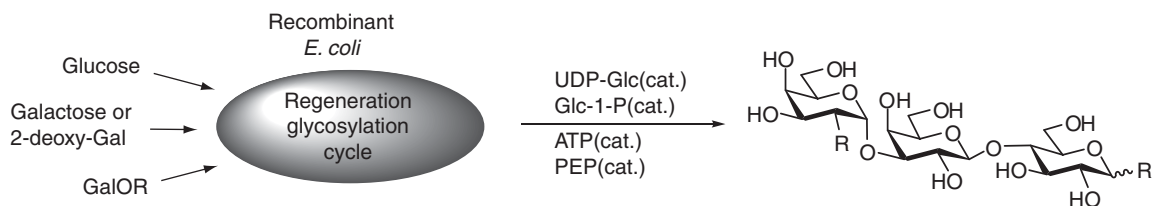
The isolation of recombinant enzymes is generally a rather laborious operation. A rapidly emerging method for the large-scale synthesis of complex carbohydrates is the use of metabolically engineered microorganisms. For the production of complex oligosaccharides, Wang's group has developed cell-free 'superbeads' technology.^{129–132} This approach involves immobilization of all the enzymes along the biosynthetic pathway onto beads. The beads are used as catalysts to produce larger oligosaccharides in a cell-free system. For example, in order to make α -Gal trisaccharide, Wang and colleagues have immobilized all the necessary enzymes, including GalK, GalT, GalUTP, PykF, and galactosyltransferase, for *in situ* regeneration of donors onto



Scheme 24 Biosynthetic pathway for UDP-Gal regeneration and α -Gal production using enzymes immobilized on superbeads.

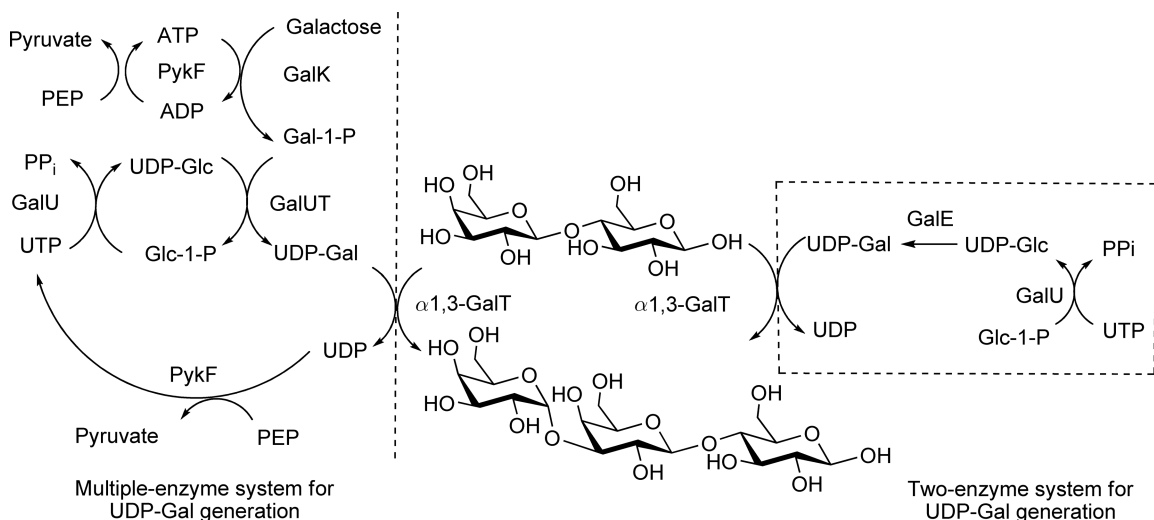
Ni-containing agarose resins (beads). These agarose beads function as stable and versatile synthetic reagents, which can be used and regenerated to synthesize a variety of oligosaccharides and glycoconjugates (Scheme 24). This methodology of enzyme immobilization has all the advantages of solid-phase organic synthesis, such as easy separation, increased stability, reusability, and improved kinetics. Larger oligosaccharide synthesis involves more corresponding glycosyltransferases together with necessary sugar nucleotide regenerating beads. Using this approach, it was possible to produce oligosaccharides containing more than eight sugar units.¹³²

Also, the Wang's 'superbug' approach can make use of engineered bacteria through fermentation to provide all the necessary enzymes along the biosynthetic pathway, starting from monosaccharide through oligosaccharides. The superbug production of oligosaccharides is a two-step procedure, distinct from the commonly employed fermentative processes. The first step involves the growth of the recombinant *E. coli* NM522 cells and the subsequent expression of the enzymes. In the second step, the cells are harvested from the culture media, permeabilized, and employed as biocatalysts in the reaction. This two-step process avoids the possible inhibition of cell growth by substrates and product and allows the use of high cell concentrations in the reaction (i.e., high catalyst concentrations) and facile manipulation of substrate concentrations. This approach relies on a single microbial strain transformed with a single artificial gene cluster of all the biosynthetic genes and uses the metabolism of the engineered bacteria to provide the necessary bioenergetics (ATP or PEP) to drive a glycosylation cycle. The advantage is that whole cells are used as biocatalysts in the reaction system without any laborious enzyme purification. Only catalytic amount of ATP is needed for this whole cell synthesis. Obviously, this makes the superbug production of oligosaccharide the most cost-effective method.^{133–135} In fact, this biotechnology for mass production of glycoconjugates has matured to such a level that it can provide a variety of products at a fraction of current commercial prices. Currently, the superbug technology can produce oligosaccharides of less than four sugar units efficiently (Scheme 25). Since they have incorporated the most common sugar nucleotide biosynthetic cycles into the superbug, the construction of new superbug for new oligosaccharide simply involves replacing/inserting the corresponding new glycosyltransferase gene(s) into the plasmid. Thus, the superbug can be constructed to produce a variety of oligosaccharides, depending on the availability of the glycosyltransferases.



Scheme 25 α -Gal superbug.

It was reported that isoglobotrihexosylceramide (iGb3) can stimulate both human $V\alpha 24$ NKT cells and mouse $V\alpha 14$ NKT cells, and it was suggested that iGb3 acted as an endogenous ligand for α NKT cells. However, current research is hampered by the rather limited access to iGb3. Purification of iGb3 from natural sources requires unique biochemical and analytical apparatuses and typically furnishes milligram quantities of the product in low yield.¹³⁶ Wang and colleagues explored efficient enzymatic syntheses of free trisaccharides. Generally, enzymatic glycosylation is one of the most practical methods for oligosaccharide synthesis.^{20,137} Currently, the most effective approach to isoglobotrihexose and globotrihexose synthesis is the direct enzymatic synthesis using UDP-Gal and lactose catalyzed by a single galactotransferase. However, the high cost of UDP-Gal limits application of the synthesis on a large scale. Multiple-enzyme sugar nucleotide regeneration systems have been extensively developed to avoid using costly stoichiometric amounts of sugar nucleotides.⁵⁵ The enzymatic syntheses of globotrihexose and isoglobotrihexose using superbead or superbug techniques were reported by Wang and coworkers.^{129,138} In these systems, UDP-Gal is generated through recycling of UTP by multiple enzymes as follows: UDP-Gal reacts with lactose to produce trisaccharide and the by-product UDP. Then, the UDP is recycled by the enzyme PykF to regenerate UTP. The UTP is further converted to UDP-Gal catalyzed by the enzymes GalU and GalPUT (**Scheme 26**). Nowadays, UTP is commercially available in large quantities at a low price. Therefore, the direct use of UTP in stoichiometric amounts can dramatically simplify the multiple-enzyme UDP-Gal generation system. A simpler two-enzyme UDP-Gal generation system to produce trisaccharides was developed (**Scheme 26**).¹³⁹ Briefly, glucosyl-1-phosphate reacts with UTP catalyzed by GalU to yield UDP-Glc. UDP-Glc is converted to UDP-Gal by epimerase, GalE.¹⁴⁰ A galactotransferase transfers the galactose from UDP-Gal to lactose and produces trisaccharide. For globotrihexose synthesis, a mixture of enzymes (GalE, GalU, and LgtC) was added to the solution of Glc-1-P, UTP, and lactose containing 0.01 mol l^{-1} Tris-HCl and 0.01 mol l^{-1} MnCl_2 . The reaction mixture was stirred at room temperature for 2 days. The removal of proteins followed by Sephadex G-15 gel column separation provided globotrihexose. They also successively applied this system on solid phase to produce isoglobotrihexose. The His6-tagged enzymes GalE,



Scheme 26 Efficient chemoenzymatic syntheses of iGb3 and Gb3.

GalU, and α -1,3-GalT were immobilized onto a Ni²⁺ resin column, and then a solution of Glu-1-P, UTP, and lactose containing 0.01 mol l⁻¹ Tris-HCl and 0.01 mol l⁻¹ MnCl₂ was circulated through the column driven by a pump at room temperature. After the reaction was completed, the solution was pumped out for purification, and the column was ready for the next batch of reaction.

Wang¹⁴¹ has also reported the enzymatic synthesis of an important tumor-associated carbohydrate antigen, Globo-H hexasaccharide. Starting with Lac-OBn as the initial acceptor, this approach employs three glycosyltransferases (LgtC, LgtD, and WbsJ) to synthesize the glycosidic linkages.

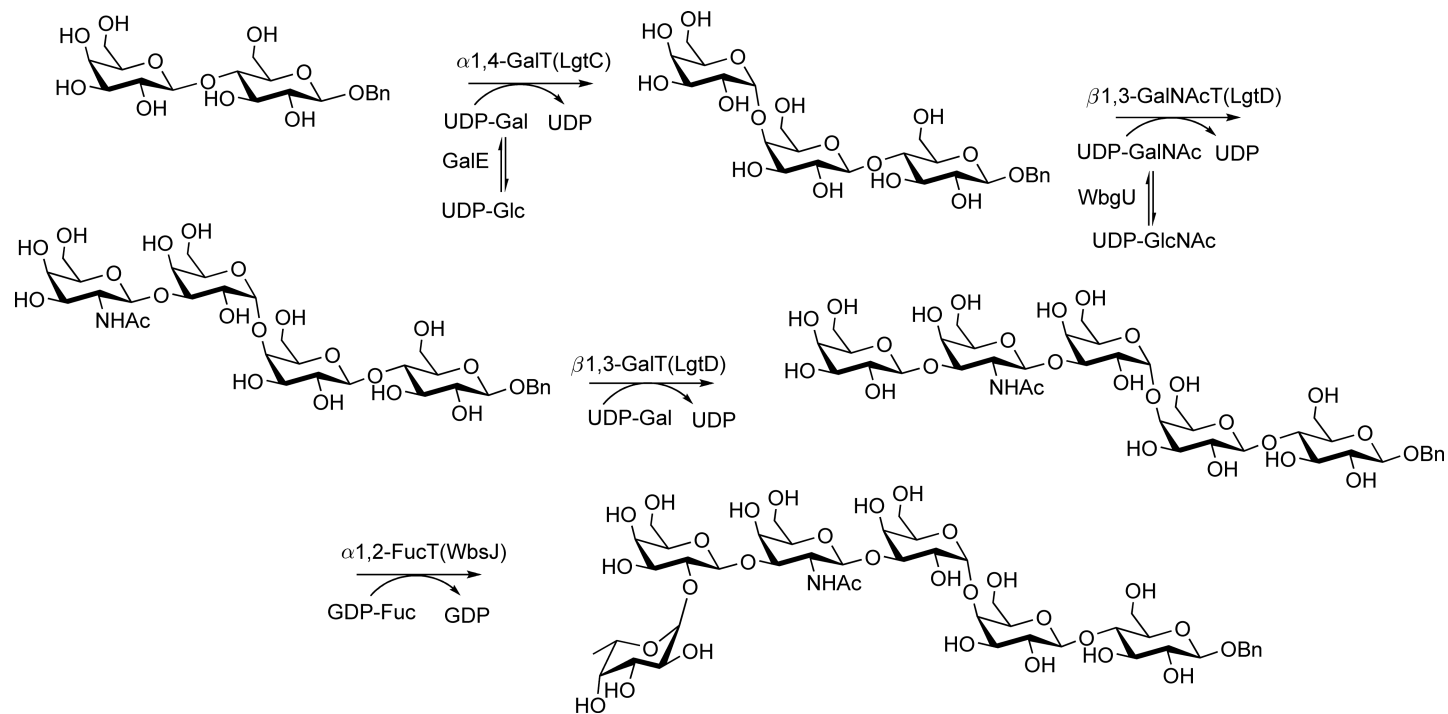
In this approach (Scheme 27), two other enzymes (GalE and WbgU) were employed in the glycosylation reactions to convert inexpensive donor substrates, uridine 5'-diphosphoglucose (UDP-Glc) and uridine 5'-diphospho-*N*-acetylglucosamine (UDP-GlcNAc), to expensive donors, UDP-Gal and uridine 5'-diphospho-*N*-acetyl-galactosamine (UDP-GalNAc), respectively. The whole synthetic route contains three glycosyltransferases and two epimerases with an overall yield of 57%. In view of high stereo- and regioselectivity under mild conditions without protecting group manipulation, this study paves a way for large-scale enzymatic synthesis of the Globo-H antigen. The yield of each coupled reaction is excellent (Table 5).

Advances in the area of glycosyltransferase-catalyzed synthesis are emerging from the identification and cloning of a large number of bacterial glycosyltransferases with various donor, acceptor, and linkage specificities. Glycosyltransferases from bacterial sources can be easily cloned and expressed in *E. coli* as soluble recombinant proteins. They also have a broader range of acceptors in comparison with mammalian glycosyltransferases. For example, several reactions have been successfully achieved by Neose Technologies, Inc. (Scheme 28).⁴⁴ Using an α 1,4-galactosyltransferase from *Neisseria gonorrhoeae*, 5 g of globotriose (Gb₃) was synthesized from lactose and UDP-Gal. Tetrasaccharide Gb₄ (1.5 g) was then synthesized with an overall yield of 60% from UDP-GalNAc and Gb₃ using a recombinant β 1,3-*N*-acetyl-D-galactosaminyl-transferase from *N. gonorrhoeae*. Also starting from lactose, trisaccharide LNT-2 (GlcNAc β 1,3Gal β 1,4Glc) was obtained in a yield of 250 g using a recombinant β 1,3-*N*-acetyl-D-glucosaminyltransferase, and tetrasaccharide LNT-3 (300 g) was then synthesized from LNT-2 catalyzed by β 1,4-galactosyltransferase. Finally, pentasaccharide LST_D (50 g) was generated from LNT-3 and NeuAc α 2,3Gal β 1,4Glc using the recombinant α 2,3-transialidase.

In general, it is difficult to express mammalian glycosyltransferases in *E. coli*, but some examples have been reported.¹⁴² Wang's group used a truncated bovine α 1,3-galactosyltransferase for the synthesis of a series of α -Gal epitope derivatives for xenotransplantation. The enzyme was expressed successfully in *E. coli* after the deletion of the transmembrane domain.⁴⁰

In addition, protein fusion technology has been employed to improve the efficiency of carbohydrate production. A hybrid bifunctional protein (fusion enzyme) has been constructed by Wakarchuk and colleagues by fusing a sugar nucleotide-generating enzyme (CMP-sialic acid synthetase) upstream of a glycosyltransferase (α 2,3-sialyltransferase) with a nine amino acid linker. The complete polypeptide chain was constitutively expressed in *E. coli* and subsequently used in the production of 3'-sialyllactose (77 g) (Scheme 29). Another example of fused bifunctional proteins was the linking of bovine α 1,3-galactosyltransferase with UDP-Gal 4-epimerase. The fused proteins, with the epimerase at either the N or the C terminus, were expressed in *E. coli* and applied in the synthesis of medically important α -Gal derivatives in 100 mg scales.¹⁴³⁻¹⁴⁵

Thorson and colleagues have investigated a promiscuous variant of galactokinase (GalK) (by directed mutagenesis) to generate novel sugar-1-phosphates.¹⁴⁶ A promiscuous variant of *rmlA*-encoded α -D-glucopyranosyl phosphate thymidyltransferase (*Ep*) from *Salmonella enterica* LT₂ has also been used to convert these sugar-1-phosphates into their corresponding sugar nucleotides. Finally, they have utilized the glycosyltransferase (GtfE), which was able to catalyze the first glycosylation step of the vancomycin biosynthetic pathway, to transfer over 30 of these uncommon sugar nucleotides onto the vancomycin aglycone scaffold to form the target products.¹⁴⁷ In a further extension of this approach, a three-step chemoenzymatic strategy for diversifying the glycone portion of vancomycin structures has been carried out using the *Salmonella* nucleotidyltransferase (*Ep*) to catalyze the conversion of 6-azido-6-deoxy glucose-1-phosphate into the sugar nucleotide.¹⁴⁸ GtfE-catalyzed glycosylation of the vancomycin aglycone then provides azido-labeled adduct, which could then be further diversified using the chemoselective Cu(I)-mediated 1,3-dipolar cycloaddition reaction with a parallel library of terminal alkynes to provide a range of triazole-linked derivatives, which were further evaluated for antibacterial activity (Scheme 30). A comparative study of chemical and chemoenzymatic methods for the glycosylation of vancomycin-like glycopeptides has been reported.¹⁴⁹ In this approach, daunosamine sulfoxide



Scheme 27 Enzymatic synthesis of Globo-H-OBn from Lac-OBn via four steps.

Table 5 Yields of each enzyme-catalyzed glycosylation step

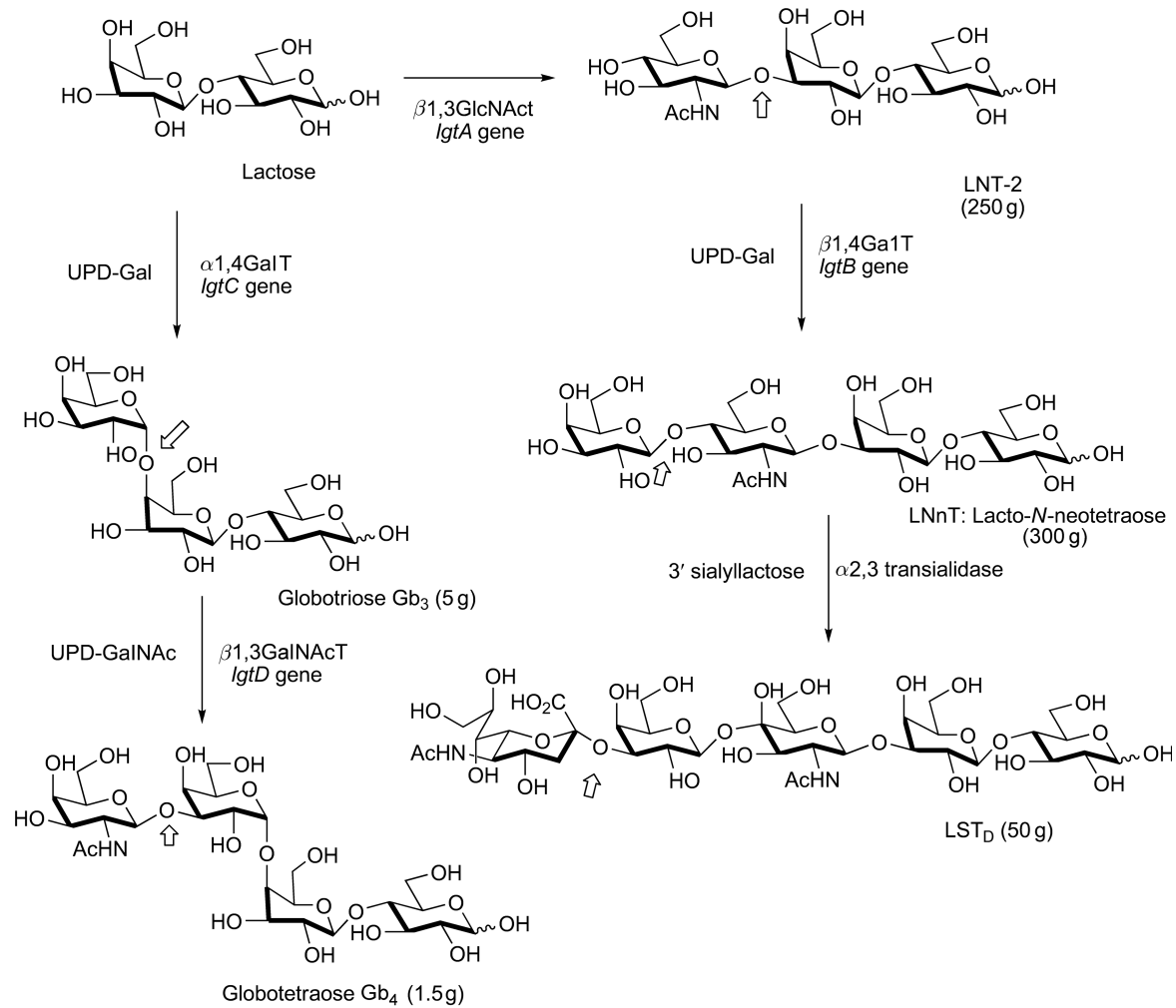
| Step | Enzyme(s) | Product | Yield (%) |
|------|-----------|-------------|-----------|
| 1 | GalE,LgtC | Gb3-OBn | 90 |
| 2 | LgtD-WbgU | Gb4-OBn | 78 |
| 3 | LgtD | Gb5-OBn | 85 |
| 4 | WbsJ | Globo-H-OBn | 95 |

was used as the activated donor sugar for the chemical route, whereas enzymatic transfer reaction was carried out using the vancomycin biosynthetic glycosyltransferase (GftD) in the presence of a fivefold excess of TDP- β -daunosamine in 70% yield. The enzymatic method was significantly shorter; however, the slow turnover rate of the unnatural substrate demands large quantities of biocatalyst, which currently limits this route to small-scale preparations (1–5 mg).

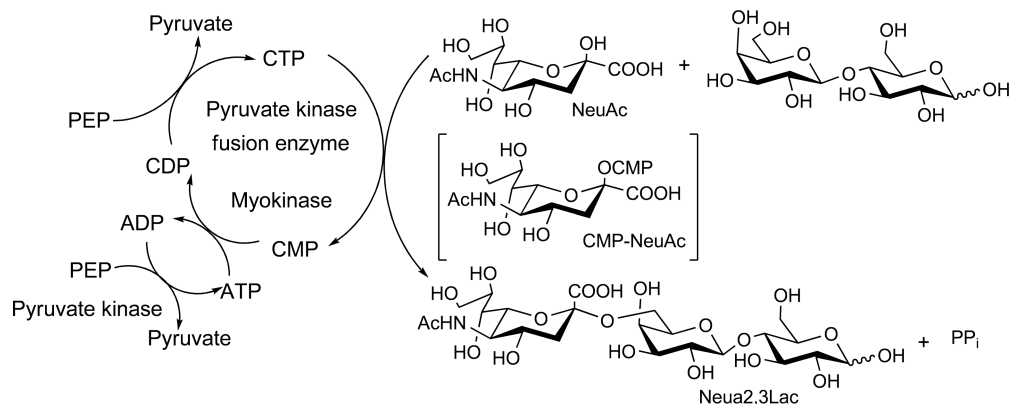
Thayer and Wong¹⁵⁰ reported a combined one-pot enzymatic glycosylation strategy for the synthesis of vancomycin analogues containing different monosaccharides such as glucose and glucosamine (**Scheme 31**). Starting from α -D-glucose-1-phosphate (Glc-1-P), thymidyltransferase (Ep) condenses Glc-1-P with uridine 5'-triphosphate (UTP) to form UDP-Glc and inorganic pyrophosphate (PP_i). Since the equilibrium of this reaction lies toward Glc-1-P and UTP, the desired UDP-Glc is sequestered by hydrolysis of PP_i with pyrophosphatase into two equivalents of inorganic phosphate (P_i). UDP-Glc then acts as substrate for GtfE for glycosylation of vancomycin aglycone to desvancosaminyl vancomycin, along with the release of uridine 5'-diphosphate (UDP). Pyruvate kinase then converts UDP into UTP by using one equivalent of PEP as the phosphoryl donor. The one-pot enzymatic glycosylation strategy should also be applicable to other glycopeptide antibiotic peptide scaffolds with the proper biosynthetic antibiotic glycosyltransferase. This chemoenzymatic method may be useful for creating analogues of other glycosylated natural products, for example, other glycopeptides, polyketides, and hybrid nonribosomal peptide/polyketide compounds.

Cyclodextrins (CDs) have the ability to include the hydrophobic area of a compound, and the formed complex becomes soluble in water due to hydroxyl groups orientated on the surface of CDs.^{151–153} Thus, it was proposed that CDs would have similar properties as surfactants in enzymatic glycosylations. Nishimura *et al.*¹⁵⁴ studied the utility of simple CDs for the synthesis of neoglycolipids using glycosyltransferase. In this approach, the authors attempted to use cell-media synthesis, which was named cellular-chemoenzymatic synthesis, as a triple-factor combination method, to construct oligosaccharide easily. Three types of CD, α -CD, β -CD, and γ -CD, which are composed of six, seven, and eight glucose units, respectively, were employed to support enzymatic glycosylation. A glucosamine derivative (**21**) was chosen as a model material as shown in **Scheme 32**. In this approach the results showed that the CDs showed potent efficiency for the blotting of hydrophobic substrates to water-soluble polymers and for the synthesis of neoglycolipids using glycosyltransferase. Interestingly, modified CDs, like methylated CDs with higher solubility, were also available. These merits were superior to surfactants commonly used in such a case. Based on their results, it was possible to make a bridge between oligosaccharide syntheses using cell function *in vivo* and enzymatic synthesis *in vitro* and well-established cellular-chemoenzymatic synthesis. The authors demonstrated the efficiency of cyclodextrins for the enzyme-catalyzed biosynthesis of neoglycolipids that are barely water soluble and suggested the CDs would act as a novel and efficient supporting material for enzymatic glycosylation.

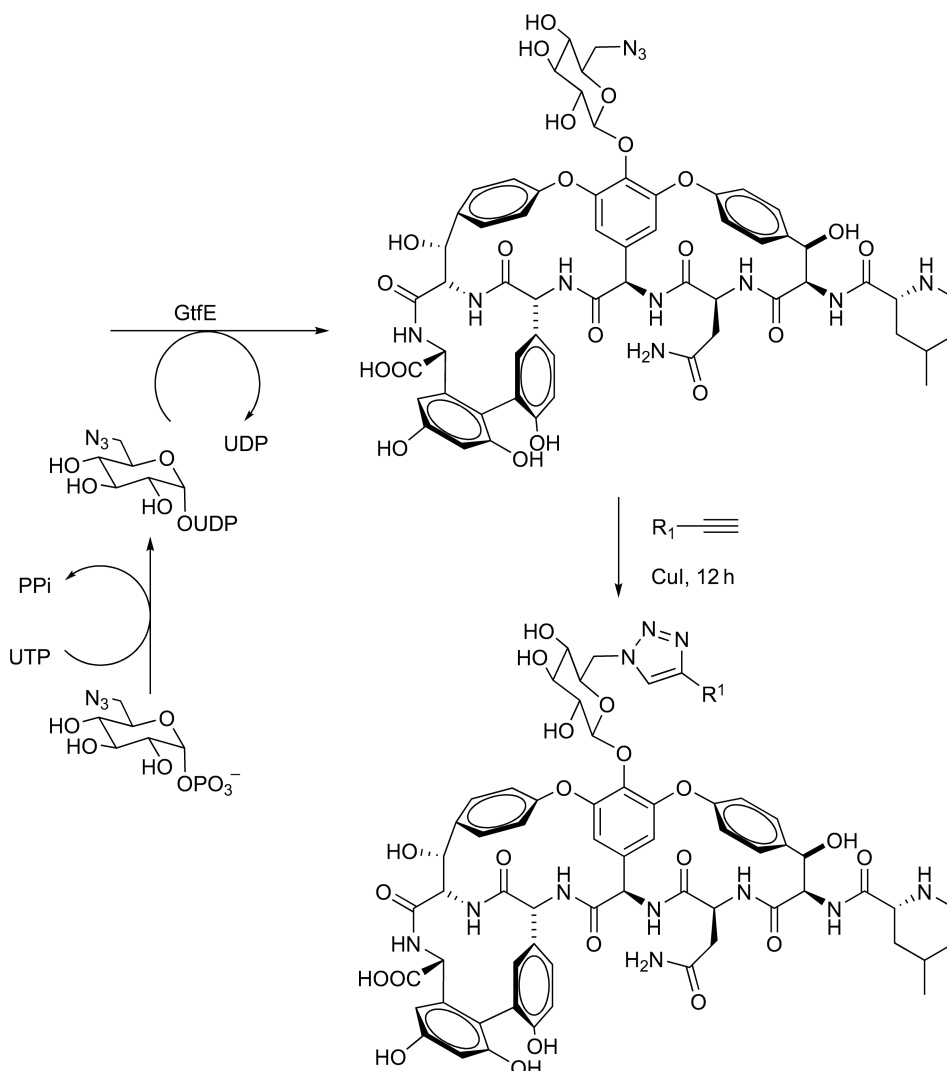
Zeng *et al.*¹⁵⁵ reported a convenient chemoenzymatic approach for the synthesis of *p*-aminophenyl glycosides of sialyl-*N*-acetyllactosaminide from *p*-nitrophenyl *N*-acetyl- β -D-glucosaminide as starting material with β -D-galactosidase, chemical reduction, and sialyltransferases. Sialyl-*N*-acetyllactosaminides were selected as model sugars because sialylated products are known to serve as the cell-surface receptor determinants for viruses, bacterial toxins, and lectins.¹⁵⁶ As shown in **Scheme 33**, intermediates **24**, **26**, and **28** were synthesized from **23** by using β -D-galactosidase from *Bacillus circulans*, recombinant rat α -(2 \rightarrow 3)-sialyltransferase, and rat liver α -(2 \rightarrow 6)-sialyltransferase. The desired trisaccharides of sialyl-*N*-acetyllactosaminide **27** and **29** could be obtained after reduction of the *p*-nitrophenyl group. Maybe due to the presence of Neu5Ac residues,



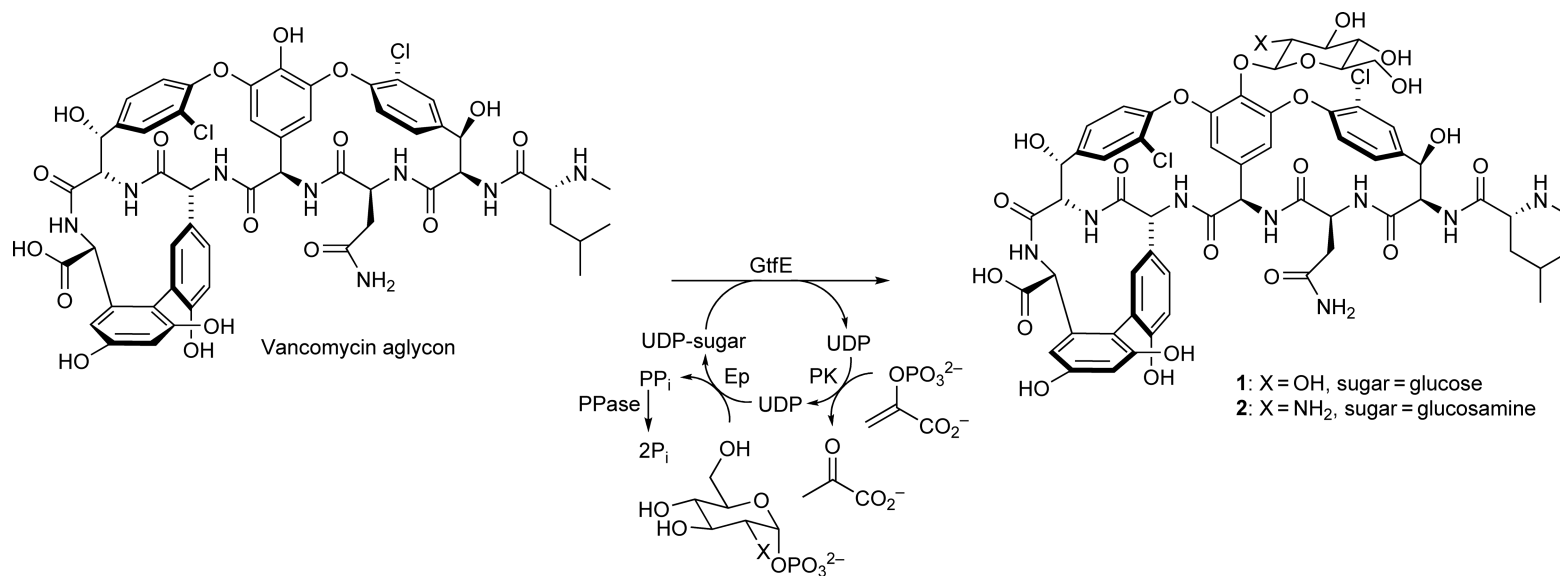
Scheme 28 The production of oligosaccharides Gb₃, Gb₄, LNT-2, LNnT, and LST_D through successive reactions using recombinant bacterial glycosyltransferases.



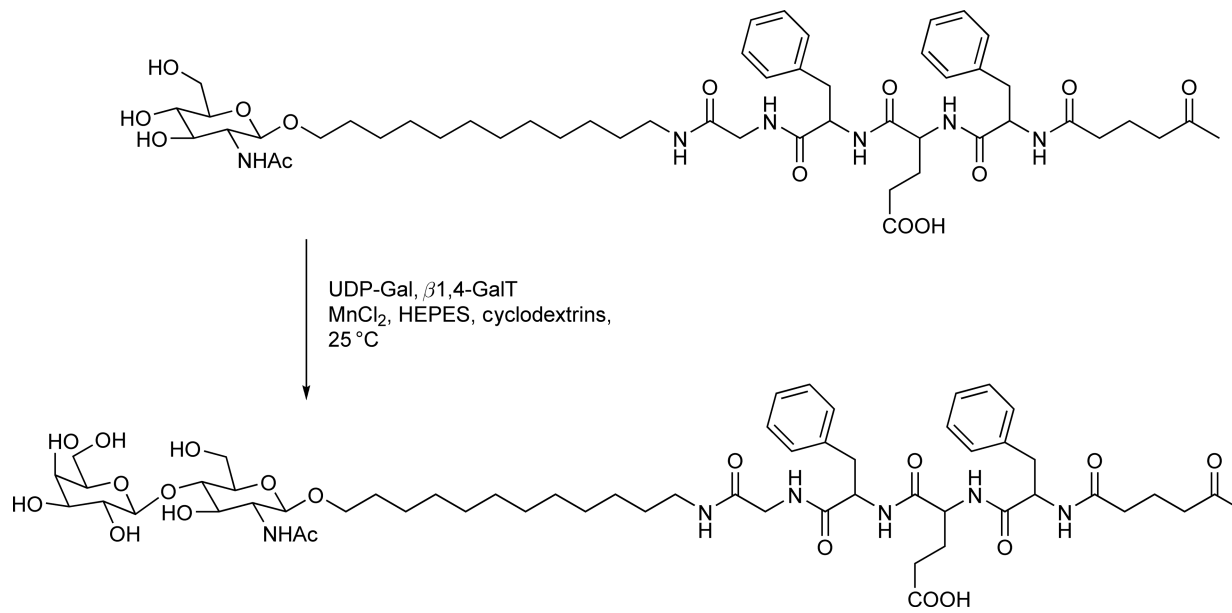
Scheme 29 Large-scale synthesis of sialylated lactose using a CMP-sialic acid synthetase/ α 2,3-sialyltransferase fusion enzyme with *in situ* regeneration of CMP-NeuAc.



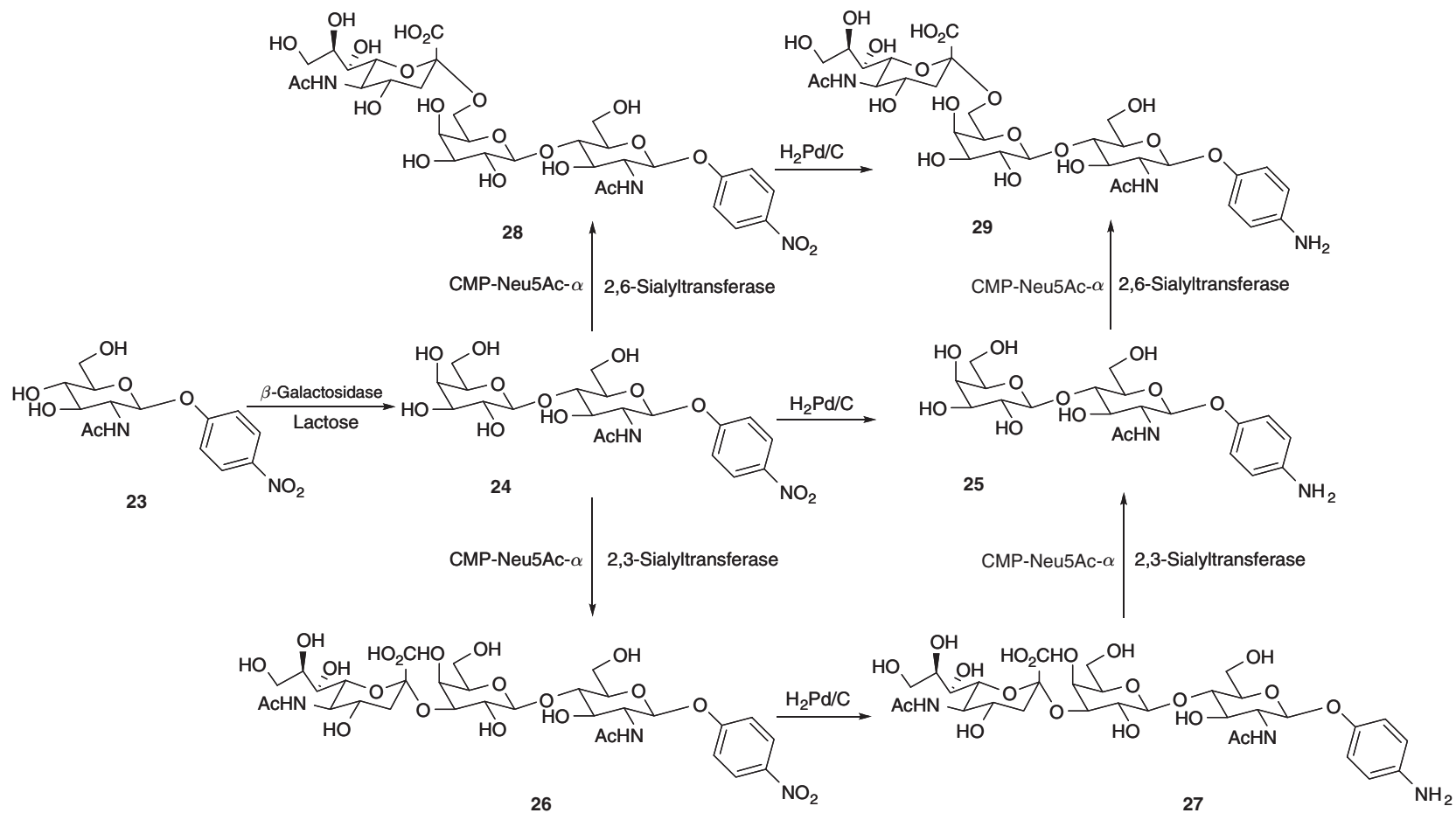
Scheme 30 Three-step chemoenzymatic strategy for vancomycin triazole-linked derivatives.



Scheme 31 One-pot enzymatic glycosylation of vancomycin aglycone to desvancosaminyl vancomycin and analogue with glycosyltransferase (GtfE), thymidyltransferase (Ep), pyruvate kinase (PK), and inorganic pyrophosphatase (PPase).



Scheme 32 Study of cyclodextrins' effect on the synthesis of *N*-acetylglucosamine derivative **21** by β 1,4-galactosyltransferase.

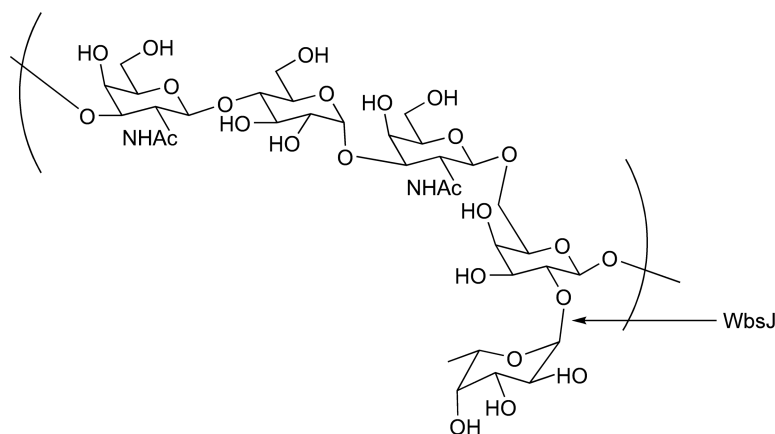


Scheme 33 Synthetic scheme for **29** and **27** from *p*-nitrophenyl *N*-acetyl- β -D-glucosaminide by using β -galactosidase, chemical reduction, and sialyltransferases.

this chemoenzymatic procedure led to quite low yields (about 15% by HPLC), compared with that of the reduction of intermediate **24**. Therefore, the authors tried another procedure for the synthesis of target trisaccharides. Initially, **24** was easily converted to its *p*-aminophenyl derivative **25** in 76% yield. Then, **25** was sialylated by recombinant rat α -(2 \rightarrow 3)-sialyltransferase and rat liver α -(2 \rightarrow 6)-sialyltransferase with CMP-Neu5Ac sodium salt as the donor to obtain the desired trisaccharides of sialyl-*N*-acetylglucosaminide **27** and **29** in 87% yield based on the donor used.¹⁵⁷

Furthermore, the synthesized glycosides were biotinylated to afford biotin-labeled sugars with an aminohexanosyl group and a phenyl group as the spacers between the biotin and glycan. The biotin-labeled sugars have been demonstrated to be useful for immobilization and assay of the carbohydrate–lectin interactions by surface plasmon resonance (SPR) technology. The results indicated that sialylated sugars were converted to their corresponding biotinyl sugars without any removal of sialic acid residues, which indicate the feasibility of this method for sialylated glycosides.

α 1,2-FucTs (fucosyltransferases) belong to glycosyltransferase family 11, catalyzing an inversion reaction by transfer of fucose from GDP- β -L-fucose to a galactose (Gal) residue to form an α 1,2-linkage. α 1,6-FucTs, however, are categorized into glycosyltransferase family 23, transferring fucose from GDP- β -L-fucose to an *N*-acetylglucosamine (GlcNAc) residue to form an α 1,6-linkage, as shown in Scheme 34.¹⁵⁸ Wang and coworkers reported the characterization of a novel α 1,2-fucosyltransferase of *E. coli* O128:B12 and functional investigation of its common motif.¹⁵⁹ The WbsJ gene from *E. coli* O128:B12 encodes an α 1,2-fucosyltransferase responsible for adding a fucose onto the galactose residue of the O-antigen repeating unit via an α 1,2 linkage. The WbsJ gene was overexpressed in *E. coli* BL21 (DE3) as a fusion protein with glutathione *S*-transferase (GST) at its N-terminus. GST-WbsJ fusion protein was purified to homogeneity via GST affinity chromatography followed by size exclusion chromatography. The enzyme showed broad acceptor specificity with Gal β 1,3GalNAc (T antigen). Gal β 1,4Man and Gal β 1,4Glc (lactose) are better acceptors than Gal β -*O*-Me and galactose, as shown in Table 6. Gal β 1,4Fru (lactulose), a natural sugar, was furthermore found to be the

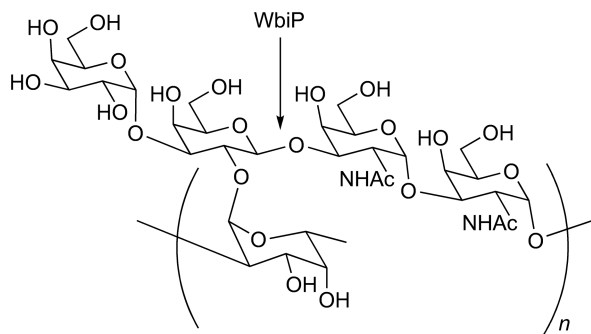


Scheme 34 Structure of *Escherichia coli* O128:B12 O-antigen repeating unit. WbsJ: α 1,2-fucosyltransferase.

Table 6 Acceptor substrate specificity of purified GST-WbsJ^a

| Acceptor (10 mmol l ⁻¹) | Rel act. (%) | Acceptor (10 mmol l ⁻¹) | Rel act. (%) |
|--|----------------|--------------------------------------|----------------|
| Gal β 1,4Glc | 100 \pm 1.4 | Gal | 35.7 \pm 1.1 |
| Gal β 1,4Glc β -O-N ₃ | 137 \pm 3.7 | Gal β -O-Me | 68.8 \pm 4.0 |
| Gal β 1,4Glc β -O-ph | 84.7 \pm 4.3 | Gal β 1,4GlcNAc | 12.4 \pm 0.2 |
| Gal β 1,4Glc β -S-ph | 207 \pm 9.7 | Gal β 1,4Fru | 380 \pm 14.9 |
| Gal β 1,4Glc β -1-NAc | 199 \pm 11.2 | Gal β 1,4Gal | ND |
| Gal α 1,4Gal | ND | Gal β 1,4Man | 162 \pm 4.7 |
| Gal α 1,4Gal β 1,4Glc | ND | Gal β 1,3GalNAc α -O-Bn | 202 \pm 6.7 |
| Gal α 1,3Gal β 1,4Glc | ND | Gal β 1,3GalNAc α -O-Me | 215 \pm 6.9 |

^a The α 1,2-fucosyltransferase activities of GST-WbsJ with different acceptors were determined at 37 °C for 2 h using 20 mmol l⁻¹ Tris-HCl, pH 7.0, 1 mmol l⁻¹ ATP, 0.3 mmol l⁻¹ GDP- β -L-fucose, GDP-L-[U-¹⁴C]-fucose (7000 cpm), 20 mmol l⁻¹ acceptor and 10 μ g of enzyme; Rel act, relative activity.

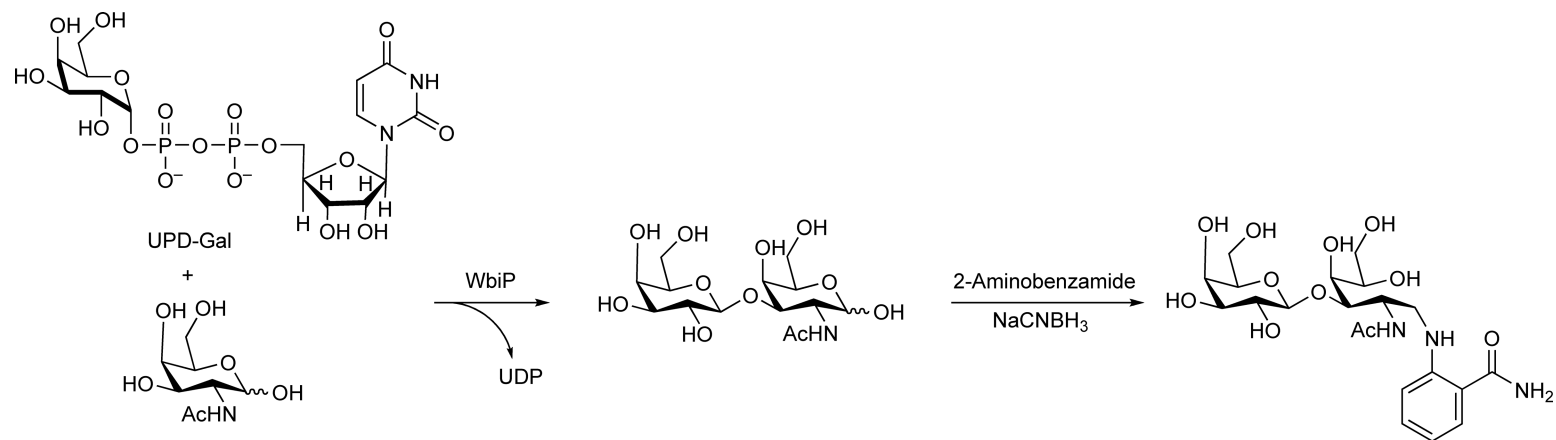


Scheme 35 O-Polysaccharide structure of *Escherichia coli* O127 and its biosynthetic gene cluster.

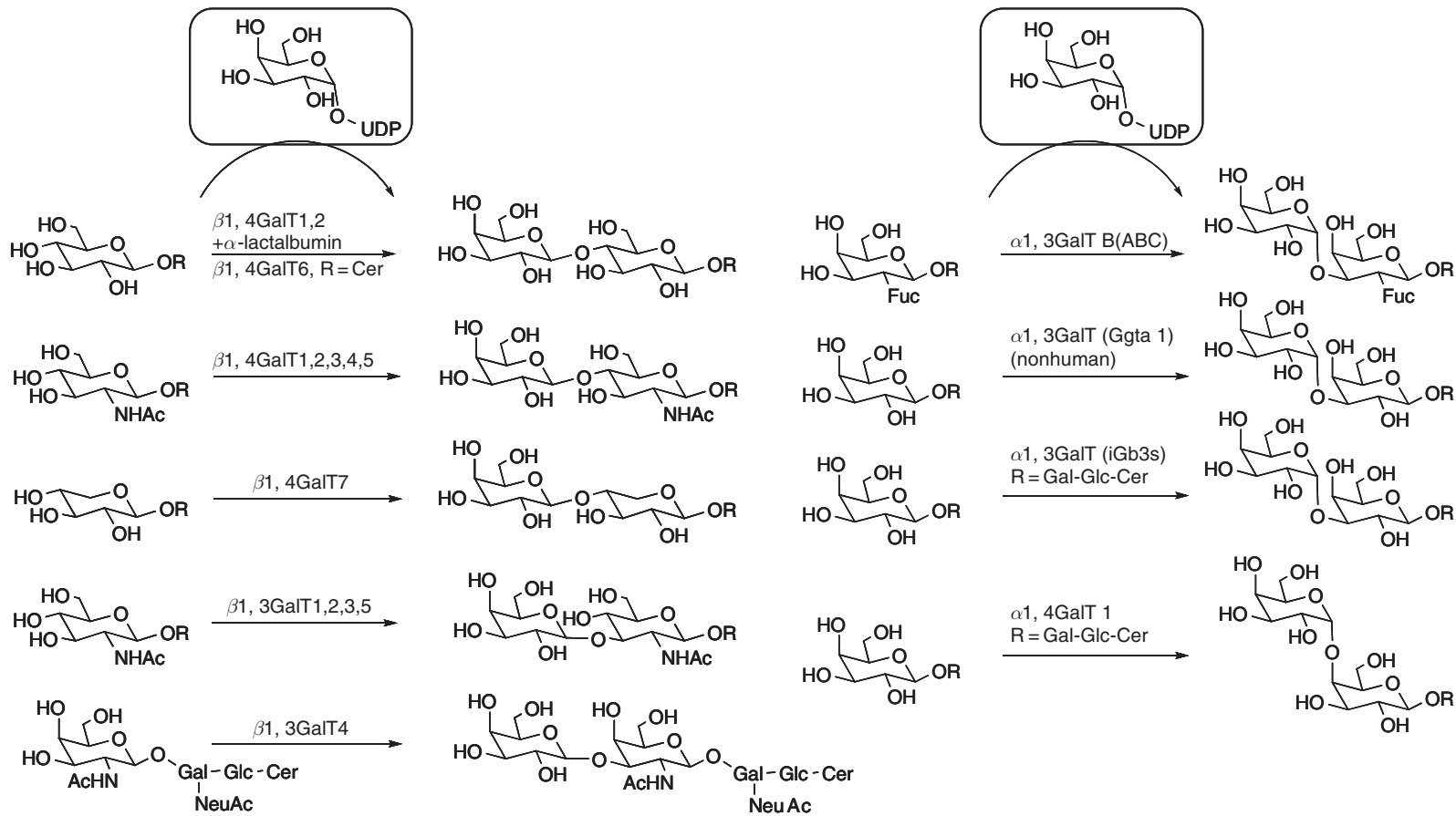
best acceptor for GST-WbsJ with a reaction rate four times faster than that of lactose. In addition, the α 1,2-fucosyltransferase activity of GST-WbsJ was found to be independent of divalent metal ions such as Mn^{2+} or Mg^{2+} . The results suggest that the common motif shared by both α 1,2-fucosyltransferases and α 1,6-fucosyltransferases has similar functions. Enzymatic synthesis of fucosylated sugars in milligram scale was successfully performed using Gal β -O-Me and Gal β 1,4Glc β -N₃ as acceptors.

Wang and coworkers¹⁶⁰ also characterized a bacterial β 1,3-galactosyltransferase (WbiP) from *E. coli* O127. *Escherichia coli* O127 belongs to the O serogroup of enteropathogenic *E. coli* (EPEC) strains, which are important causes of infantile diarrhea in developing countries.¹⁶¹ *Escherichia coli* O127 was reported to possess high human blood group H (O) activity. The elucidation of the chemical structure of the cell-surface O-antigen¹⁶² confirmed that *E. coli* O127 mimicked the expression of human blood group H antigen (Scheme 35). The O-antigen biosynthetic gene cluster contains multiple genes needed for the assembly of *E. coli* O127 polysaccharide structures. Among them, three genes (*orf3*, *orf12*, and *orf13*) encode putative glycosyltransferases involved in the synthesis of repeating oligosaccharide units. *Orf12* (WbiP) contains a conserved domain found in glycosyltransferase family 2. It shows 59% protein sequence identity and 82% similarity to WbnJ from *E. coli* O86, which was previously characterized as a β -1,3-galactosyltransferase.¹⁶³ To investigate the enzymatic function of WbiP, Wang and coworkers tested a panel of radiolabeled sugar donors (UDP-Gal, UDP-GalNAc, GDP-Fuc, and UDP-Glc) with GalNAc as an acceptor. The results showed that WbiP displayed high activity with UDP-Gal as the sugar donor, nearly 10-fold reduced efficiency with UDP-GalNAc, and no activity with GDP-Fuc and UDP-Glc. Thus, the donor specificity demonstrated that WbiP encodes a galactosyltransferase, consistent with the bioinformatics analysis. To confirm the galactosyltransferase activity of WbiP, they set up a 50 μ l scale reaction, which contained 0.05 mg of GalNAc, 5 μ l of 50 $mmol\ l^{-1}$ cold UDP-Gal, and 0.02 mg of WbiP (20 μ l). The reaction product was then labeled with 2-aminobenzamide (2-AB) and analyzed to via normal-phase analytical HPLC columns (Scheme 36).

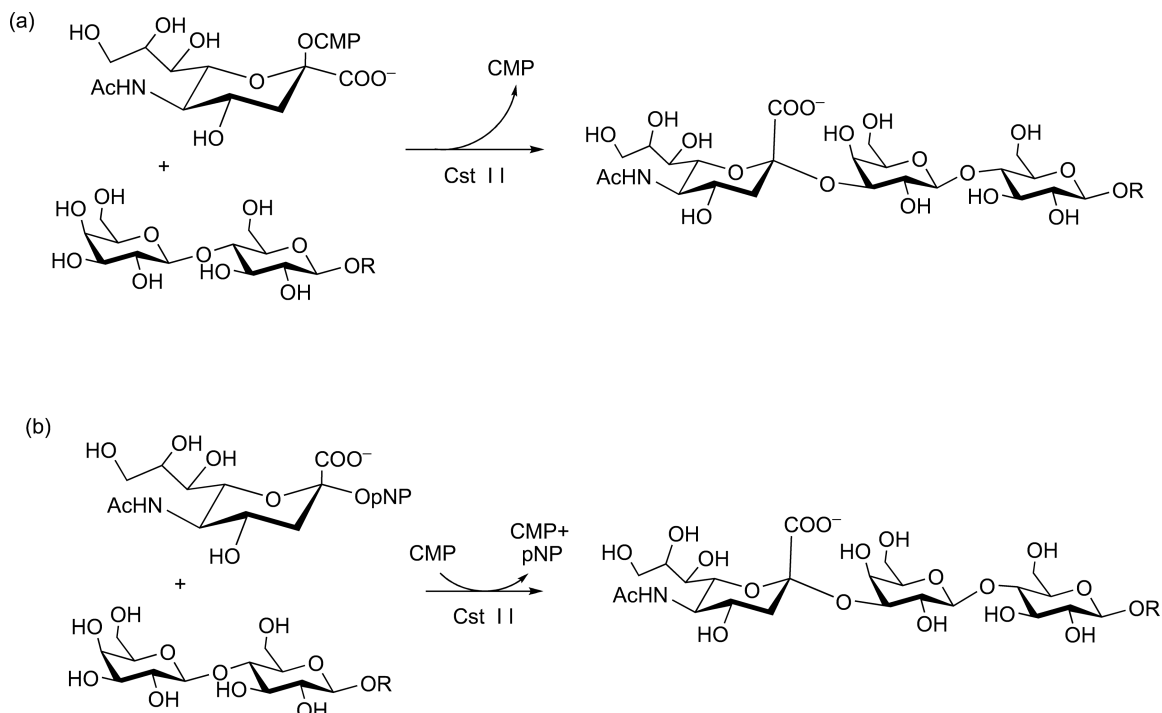
Depending on the stereochemical outcome at the anomeric center relative to that of the donor sugar, glycosyltransferases are classified as retaining and inverting classes, as with the glycosidases. All these enzymes use UDP-D-Gal as the donor, but generate β 1-4, β 1-3, α 1-3, and α 1-4 linkages to accepting templates in various types of glycoconjugates (Scheme 37).¹⁶⁴ While the mechanism of inverting glycosyltransferases is generally accepted, that of retaining glycosyltransferases remains a topic of considerable debate within the field.¹⁶⁵⁻¹⁶⁷ Withers and coworkers¹⁶⁸ described investigations into the donor substrate specificity of representatives of both retaining and inverting classes of glycosyltransferases. By the use of glycosyl donors containing aromatic leaving groups linked with opposite anomeric configurations compared to those of the natural donor substrates, both an inverting (Cst II) and a retaining (LgtC) glycosyltransferase were found to catalyze glycosylation reactions of natural acceptor substrates in the presence of the corresponding nucleotide. Cst II is an inverting bifunctional α -2,3/2,8 sialyltransferase from *Campylobacter jejuni*. It uses cytidine 5'-monophosphate (CMP) β -D-sialic acid as a donor substrate and transfers the sialic acid moiety with net inversion of anomeric configuration to the 3' hydroxyl of terminal lactose-containing acceptors (Scheme 38(a)). Subsequently, Cst II will use this newly formed 3'-sialyl lactose product as an acceptor and transfer another sialic acid moiety to the 8'' hydroxyl to form a tetrasaccharide. To determine whether this



Scheme 36 WbiP-catalyzed reaction and HPLC identification of products.



Scheme 37 Substrate specificity in the mammalian galactosyltransferase superfamily (GalT) for natural glycoconjugates as acceptor substrates and UDP-Gal as sugar nucleotide donor.

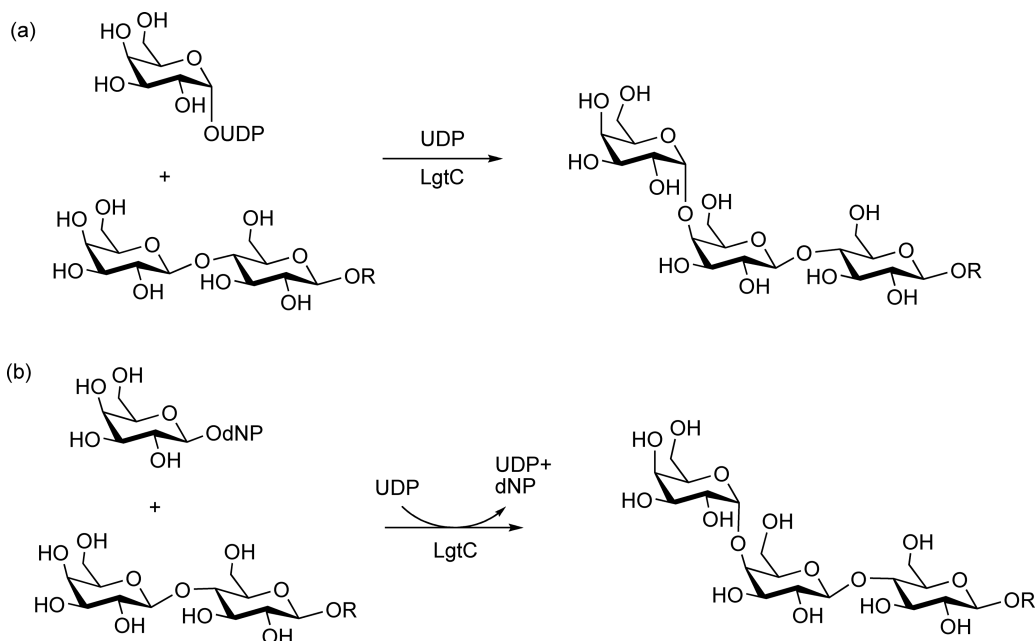


Scheme 38 Reaction catalyzed by Cst II using (a) the natural donor substrate CMP β -D-sialic acid or (b) the alternative donor substrate pNP α -D-sialic acid.

inverting enzyme could use an alternative source of activated sialic acid as a substrate, they investigated the glycosylation of the α -linked *c*phenyl (pNP) derivative (**Scheme 38(b)**). Because Cst II is an inverting enzyme, it was presumed that the aromatic ring of the nitrophenyl substituent could be accommodated within the active site, taking the place of the galactose ring of the acceptor, which must be located on the face of the sialic acid donor. With pNP α -D-sialoside bound in such a manner, CMP could also be accommodated within the active site on the β face of sialic acid, allowing for the direct displacement of *p*-nitrophenolate by CMP and the *in situ* formation of β -linked CMP sialic acid, which could be used as the donor substrate in a subsequent transfer reaction following the release of pNP. The results showed that Cst II was able to catalyze transfer to the fluorescein–lactose conjugate acceptor but only in the presence of CMP.

To further explore the generality of this substrate complexity, a similar strategy was applied to the retaining α -1,4 galactosyltransferase LgtC from *Neisseria meningitidis*.¹⁶⁹ LgtC is able to catalyze the transfer of galactose moiety to the 4' hydroxyl of terminal lactose-containing acceptor substrates using UDP α -D-galactose as the donor with overall net retention of anomeric configuration (**Scheme 39(a)**). To test the limits of donor substrate complexity of LgtC, pNP α -D-galactoside and dNP α -D-galactoside were used as the surrogate donor substrate. However, no result was observed under these conditions. Subsequently, a similar strategy analogous to that for Cst II was applied to LgtC, in which an activated leaving group of anomeric configuration opposite to that of the natural substrate was tested (**Scheme 39(b)**). As both the leaving group and the incoming nucleophile are present on the same face of the donor substrate in the active site of LgtC, the aromatic leaving group would not occupy the acceptor site. Using this more activated analogue, a trisaccharide was observed. The results of this work illustrated how the donor substrate promiscuity of both inverting and retaining glycosyltransferases can be used, providing the starting point of an alternative strategy for their application as synthetic tools. These enzymes appear to be able to transfer the glycosyl substrates with alternative activated leaving groups of opposite anomeric configuration, compared to the natural donor.

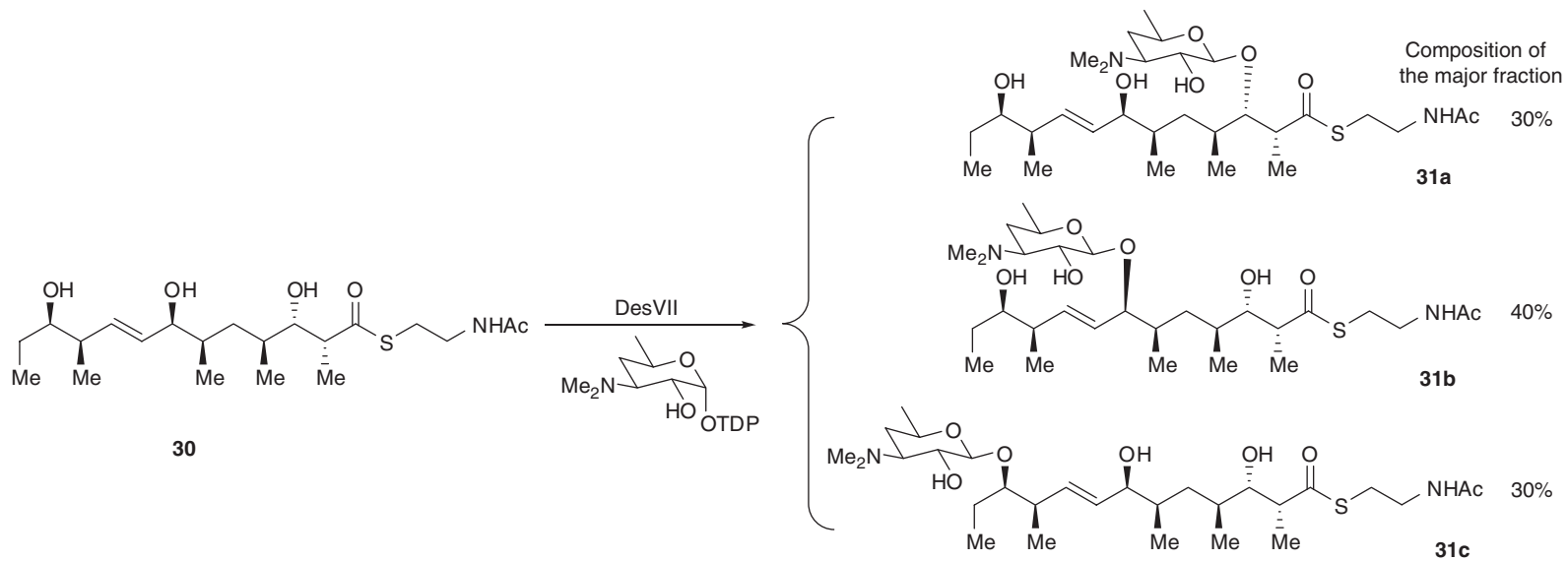
In the last few years, macrolide glycosyltransferases (GTs), toward both the sugar and aglycone substrates, gained interest as applicative enzymes for the convenient synthesis of biologically relevant new macrolide analogues that can potentially be used as drugs. Among the growing number of *in vitro* studies of related GTs, DesVII, which catalyzes the attachment of D-desosamine to 10-deoxymethynolide or narbonolide in the



Scheme 39 Reaction catalyzed by LgtC using (a) the natural substrate UDP α -D-galactose or (b) the alternative substrate 2,4-dinitrophenyl β -D-galactoside.

biosynthesis of methymycin, neomethymycin, narbomycin, and pikromycin in *Streptomyces venezuelae*, is one of the most extensively studied macrolide GTs.^{169,170} Liu and coworkers reported the biosynthetic potential of DesVII/DesVIII to explore the range of aglycone substrate flexibilities of this enzyme pair.¹⁷¹ The detailed *in vitro* analysis of sugar substrate specificity of DesVII/DesVIII had been carried out and the results showed that both L- and D-sugars are recognized as substrates and variant substitutions at C3 and C4 are tolerated, but deoxygenation at C6 is required.¹⁷² When a reaction mixture containing a linear precursor **30** and a 1.2-fold excess of TDP-D-desosamine was incubated with DesVII/DesVIII, TLC analysis of the reaction mixture revealed the presence of at least four polar compounds ($R_f = 0.04, 0.19, 0.32,$ and 0.37) in addition to the starting material ($R_f = 0.51$). The overall conversion of the linear precursor **30** to the more polar products was estimated to be nearly 50%. After separation by flash silica gel chromatography, compound **31** was isolated as the major component ($R_f = 0.19$), which was shown by extensive NMR analysis to be a mixture of three monodesosaminylated products (**31a**, **31b**, and **31c**, Scheme 40). So there is no regioselectivity of DesVII/DesVIII in the reaction with the linear precursor **30**. Their study clearly demonstrates that the macrolide glycosyltransferase, DesVII/DesVIII, can recognize and process not only cyclic substrates of different ring sizes, but also a variety of linear substrates with reduced, but measurable activities. When multiple hydroxyl groups are present, essentially no regioselectivity is observed for the linear substrates. Interestingly, only singly glycosylated products were generated in all cases. However, when cyclic substrates with multiple glycosylation sites were tested, the enzyme displayed excellent regioselectivity in most cases. Their results showed that the factors governing the glycosylation outcome for the cyclic substrates may be multifold, including the substitution pattern near the glycosylation site and the ring conformation. Also, the specificity/selectivity may be sensitive to the substituents distant from the glycosylation site.

Generally, members of the genus *Mycobacterium* are responsible for major health problems in humans, such as tuberculosis, leprosy, and opportunistic mycobacterioses in immunocompromised individuals.¹⁷³ Structure analysis showed that the backbone of the mycobacterial cell wall consists of a complex of covalently linked arabinogalactans (AGs) and peptidoglycans. The D-galactofuran, composed of ~ 30 alternating (1 \rightarrow 5)- and (1 \rightarrow 6)-linked repeating β -D-galactofuranose units (Gal_f),¹⁷⁴ is attached to the peptidoglycan on the cell surface by a dedicated linker unit, α -L-rhamnopyranosyl-(1 \rightarrow 3)-N-acetyl- α -D-glucosaminyl-phosphate (Rha-(1 \rightarrow 3)-GlcNAc-P).¹⁷⁵ Several arabinan chains, containing about 23 repeating D-arabinofuranose

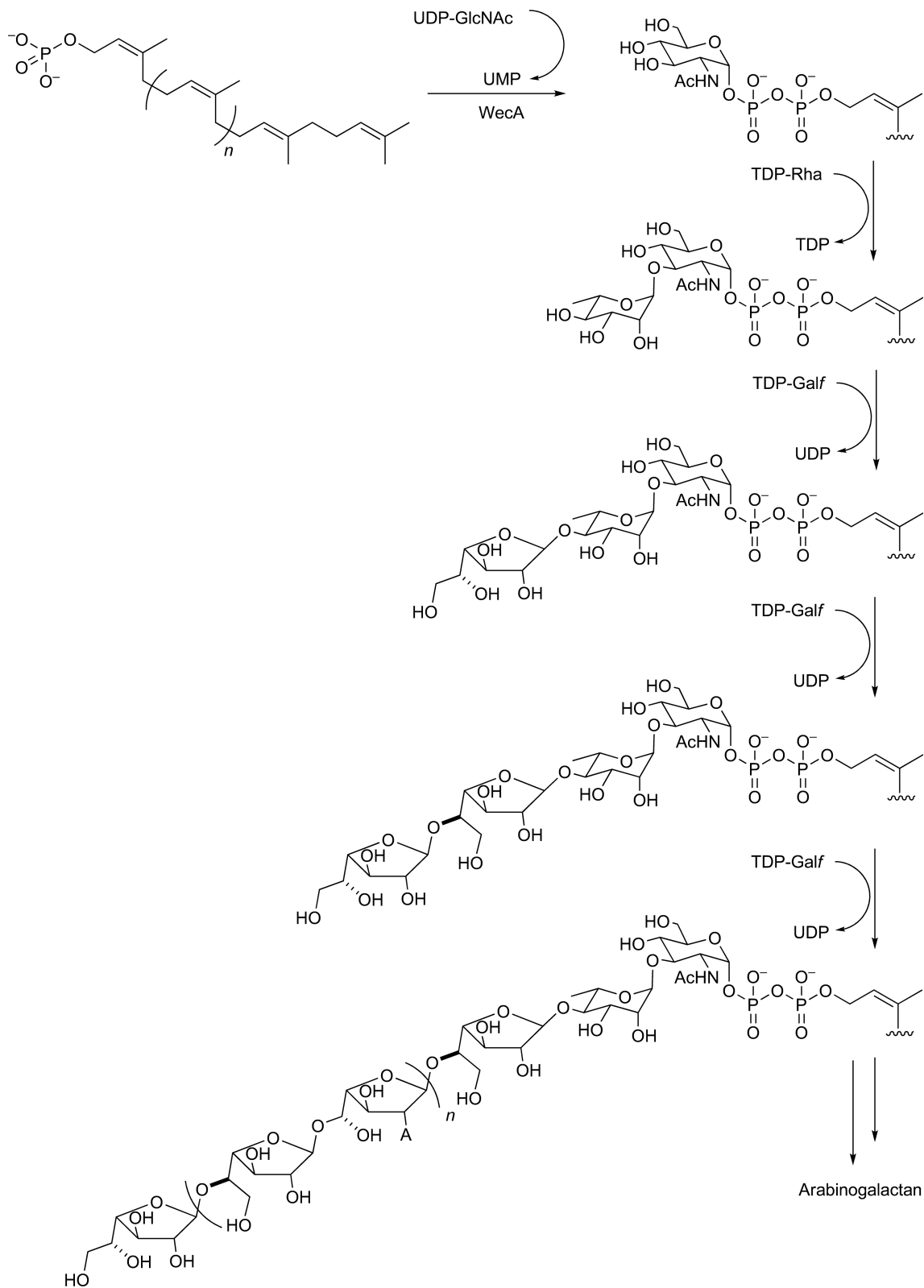


Scheme 40 Desosamylation of the linear polyketide substrate **30**.

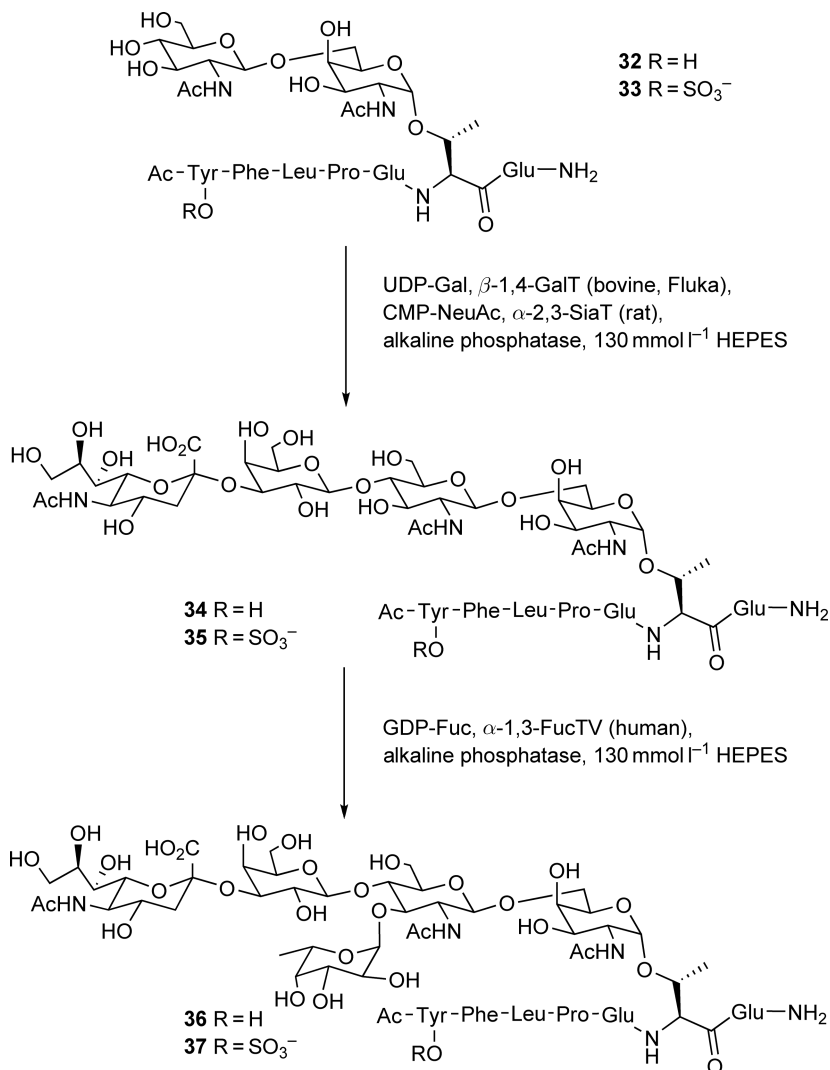
(Araf) units each, composed mostly of (1 → 5)-linked α -D-Araf units with branching introduced by 3,5- α -D-Araf, are bound to the galactan. Brennan and coworkers¹⁷⁶ reported the biosynthetic origins of the AG using cell extracts of *Mycobacterium smegmatis*. Their results showed that the synthesis of AG begins with the transfer of UDP-GlcNAc to the polyprenyl phosphate, decaprenyl-P (C_{50} -P), followed by the addition of Rha from dTDP-Rha, forming the linker region and initiating the biogenesis of AG fragments.¹⁷⁷ The polymerization of the AG then takes place on this C_{50} -P-P-GlcNAc-Rha unit by enzyme-catalyzed single sugar additions. Subsequently, the complete sequence of the *Mycobacterium tuberculosis* H37Rv genome¹⁷⁸ enabled the identification of several enzymes involved in this biosynthesis process. Rv1302 from *M. tuberculosis* H37Rv was regarded as the gene involved in the initial step. Subsequently, the enzyme-catalyzed addition of Rha moiety is activated by the rhamnosyltransferase WbbL (Rv3265c).¹⁷⁹ Based on the assortment of glycosyl linkages within the galactan, there are likely two or more galactosyltransferases (GalTr) involved in the following galactan polymerization pathway. However, to date, only a single candidate GalTr gene (Rv3808c) has been implicated. Evidence showed that this bifunctional enzyme catalyzes the synthesis of the alternating 5- and 6-linked, linear galactan in a processive manner (**Scheme 41**).

Brennan¹⁸⁰ presented the evidence that Rv3782 is an additional galactosyltransferase responsible for the initial transfer of Galf residues from their donor, UDP-Galf, onto the C_{50} -P-P-GlcNAc-Rha unit, initiating the subsequent polymerization events catalyzed by Rv3808c, a bifunctional galactosyltransferase. The possibility of Rv3782 being a putative galactosyltransferase implicated in galactan synthesis arose from a number of observations. First, there was amino acid sequence similarity to portions of the known GalTr Rv3808c. Second, according to the rule of carbohydrate-active enzymes classification, it was a glycosyltransferase (GT) of the nucleotide sugar-requiring, inverting GT-2 family of the GT-A superfamily. Third, Rv3782 was described as a possible AG biosynthetic gene cluster in *M. tuberculosis*. In this biosynthesis pathway, it appears that GL-2 is the primary acceptor of Galf donated by UDP-Galf catalyzed by Rv3782. To examine this hypothesis directly, GL-2 was generated *in vitro* and used as the glycolipid substrate in reactions containing nonradioactive UDP-Gal and membranes from control cells or the overproducer Rv3782 strain. The results showed determinately that the GL-2 was an effective acceptor for Gal transfer and that membranes from the overproducing strain were significantly more effective in the reaction. However, the primary product of the reaction was not GL-3, but GL-4, containing more than one additional Gal residue. Therefore, Rv3782 may be responsible for the initial conversion of GL-2 to GL-3, and in the reaction mixtures as constructed, there may be an immediate conversion of GL-3 to GL-4 promoted by a second endogenous galactosyltransferase (GalTr). Alternatively, Rv3782 may be a bifunctional enzyme catalyzing the synthesis of the 5-linked and the 6-linked Galf residues of GL-4. Data presented in their paper indicate that Rv3782 is the initial galactosyltransferase catalyzing addition of the first and/or second galactose moiety on the lipid-associated linker region (decaprenyl-P-P-GlcNAc-Rha) in the series of Gal polymerization steps leading ultimately to cell wall galactan structures. Apparently, it is a bifunctional enzyme capable of generating the initial 5- and 6-linked Galf residues but not capable of extending the chain beyond this level.

Enzymatic glycosylation for the synthesis of complex saccharides in glycopeptide has also gained attention. Lin and Wong have developed an enzymatic one-pot three-step glycosylation strategy for the synthesis of sLe^x moiety of truncated PSGL-1 glycopeptide with and without sulfation.¹⁸¹ Glycopeptides **32** and **33** were achieved from disaccharide-linked threonine as the building block via solid-phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl group (Fmoc) chemistry.¹⁸² Following the N-terminal acetylation, cleavage from the resin, and full deprotection, the basic hydrolysis of the acetate protecting groups gave glycopeptides **32** and **33**. With the glycopeptides as starting materials, the one-pot synthesis of sLe^x by glycosyltransferase-catalyzed glycosylation was investigated. Because LacNAc is a good substrate for both SiaT and FucT, and Le^x is not a good substrate for SiaT, the sLe^x moiety was constructed by adding the galactose, sialic acid, and fucose groups sequentially (**Scheme 42**). Glycopeptide **32** was treated with GalT and SiaT in the presence of donor substrates UDP-Gal and CMP-NeuAc in the N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) buffer to give **34** in good yield. Due to the fact that sulfated glycopeptide **33** is not a good substrate for SiaT, the enzymatic glycosylation reaction of **33** did not proceed as smoothly as for the unsulfated substrate **32**.¹⁸³ After the addition of more glycosyltransferases and extension of the reaction time, compound **35** was obtained in good yield. FucT and GDP-Fuc were incubated with **34** and **35**, respectively. The complete sLe^x glycopeptides of truncated PSGL-1 with or without sulfation, **36** and **37**, were afforded in 68 and 63% yield, respectively.



Scheme 41 Proposed pathway for the biosynthesis of mycobacterial arabinogalactan.

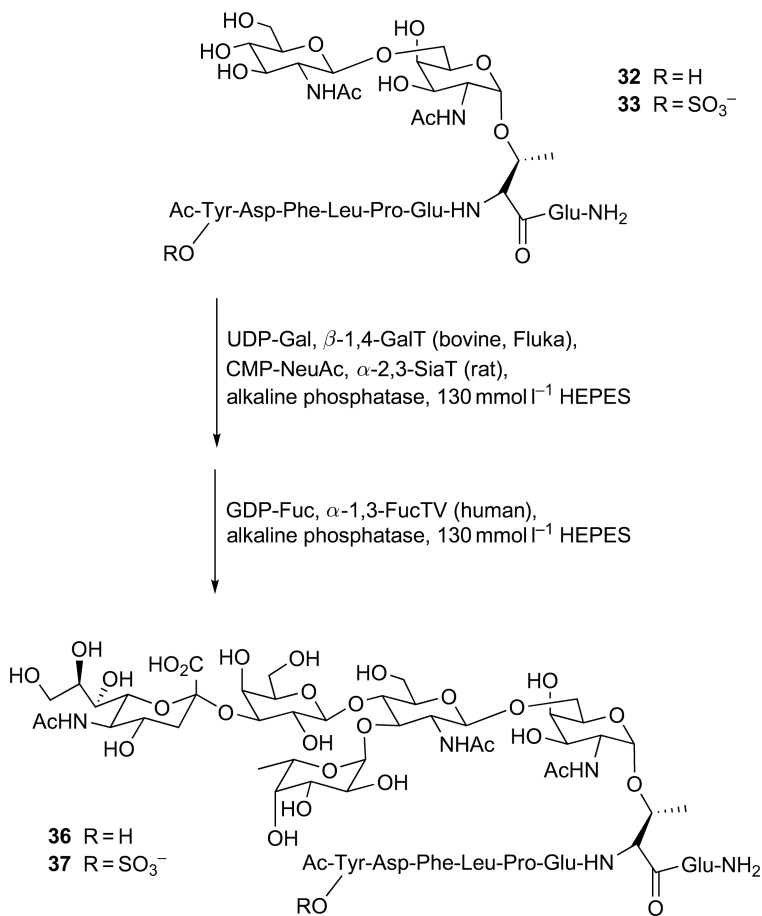


Scheme 42 Synthesis of unsulfated and sulfated glycopeptides **36** and **37**.

The aforementioned enzymatic reaction conditions were utilized to construct the sLe^x-containing glycopeptides using the multienzyme one-pot, three-step strategy (**Scheme 43**). Substrates **32** and **33** were incubated with GalT and SiaT in the presence of UDP-Gal and CMP-NeuAc. Subsequently, FucT and GDP-Fuc were added and incubated for another 48 h. After purification, the truncated PSGL-1 glycopeptides with or without sulfation, **36** and **37**, were obtained with 43 and 25% yield, respectively.

6.02.4.2.3 Conclusion

Recent advances in the area of enzymatic oligosaccharide synthesis are emerging from the identification and cloning of a large number of bacterial glycosyltransferases with many different donor, acceptor, and linkage specificities. Most eukaryotic glycosyltransferases are not active within prokaryotic expression system due to the absence of posttranslational modifications, including glycosylation. Bacterial glycosyltransferases, for example, are normally not glycosylated proteins. It has been demonstrated that these enzymes are more easily expressed as soluble and active form in prokaryotic expression system such as *E. coli*. In addition, bacterial glycosyltransferases seem to have relatively broader acceptor substrate specificities, thereby offering



Scheme 43 Multienzyme one-pot synthesis of unsulfated and sulfated glycopeptides.

tremendous advantages to mammalian enzymes in the chemoenzymatic synthesis of oligosaccharides and their analogues for the development of antiadhesion therapies for infectious diseases.

Rapid advances in the cloning and expression of glycosyltransferase genes, especially from bacteria, could open the way to overcoming difficulties in the mass production of oligosaccharides. The large-scale production of oligosaccharides using either glycosyltransferases isolated from engineered microorganisms or whole cells as an enzyme source could promote a new era in the field of carbohydrate synthesis. The increasing availability of new gene sequencing data is likely to stimulate the discovery of new biocatalysts with enhanced substrate tolerance and/or multiple catalytic functions. Such discoveries, coupled with refinements in methodologies for directed evolution and enzyme-catalyzed dynamic combinatorial library synthesis, will see further significant developments in the near future.

6.02.5 Outlook

As is evident from the work that is reviewed here, the application of biocatalysts in preparative carbohydrate synthesis has continued to expand and diversify in recent years. Of particular note is the reengineering of biocatalyst properties (e.g., substrate specificity, stability) by the process of directed evolution. Significant progress in the study of the enzymatic biosynthesis of complex carbohydrate has been made with the development in protein purification, molecular genetics, and new methods of enzymological analysis. Bioinformatics provides a large number of putative candidates for carbohydrate-active enzymes. The combined

enzymatic and genetic approach has overcome formidable obstacles ubiquitous to the study of oligosaccharides and polysaccharides, and has begun to yield new information, which definitively addresses basic enzymological issues relevant to these complex carbohydrates. Advances in this field will include the finding of novel enzymes and the modification of the known proteins to provide better catalysts. Further genetic engineering efforts will be focused on the recombinant expression of the enzymes involved in the biosynthetic pathways of glycoconjugates and the utilization of engineered microorganisms as 'living factories' to produce desired carbohydrates in large scales. Further development of enzymatic methods will allow synthetic biochemists to create important molecular tools for biochemical, biophysical, and medical applications.

These methodologies offer a range of possibilities for the synthesis of biomaterials. New perspectives such as the ability of using modified donors to introduce functional groups into polysaccharide structures, controlling the degree of polymerization, and the preparation of novel composite materials by enzymatic polymerization in the presence of other polymers are just emerging, and a number of applications in advanced biomaterials are awaiting to come in the near future.

Abbreviations

| | |
|-----------------------|--|
| AcCoA | acetyl coenzyme-A |
| ATP | adenosine triphosphate |
| CD | cyclodextrin |
| CFX | <i>Cellulomonas fimi</i> |
| CMP | cytidine monophosphate |
| CMP-NeuAc | cytidine 5'-monophospho- <i>N</i> -acetyl- α -D-neuraminic acid |
| CTP | cytidine triphosphate |
| GAG | glycosaminoglycan |
| GalK | galactokinase |
| GalNAc | <i>N</i> -acetylgalactosamine |
| GalNAcE | GalNAc 4'-epimerase |
| GalT | galactosyltransferase |
| GDP-Man | guanidine 5'-diphospho- α -D-mannose |
| GlcA | glucuronic acid |
| GlcNAc | <i>N</i> -acetyl-D-glucosamine |
| GT | glycosyltransferase |
| GTP | guanosine triphosphate |
| HAase | hyaluronidase |
| LacNAc | <i>N</i> -acetylactosamine |
| NAD | nicotinamide adenine dinucleotide |
| NADP | nicotinamide adenine dinucleotide phosphate |
| NAHase | β - <i>N</i> -acetylhexosaminidase |
| Neu5Ac | sialic acid |
| PEP | phosphoenolpyruvate |
| PK | pyruvate kinase |
| pNP | <i>p</i> -nitrophenyl |
| PNP-LacNAc | <i>p</i> -nitrophenyl β - <i>N</i> -acetylactosamine |
| PNPX2 | PNP β -D-xyloside |
| PP_i | pyrophosphate |
| PykF | pyruvate kinase from the <i>E. coli</i> K-12 genome |
| UDP | uridine diphosphate |
| UDP-Gal | uridine 5'-diphospho- α -D-galactose |
| UDP-GalNAc | uridine 5'-diphospho- <i>N</i> -acetyl- α -D-galactosamine |
| UDP-Glc | uridine 5'-diphospho- α -D-glucose |

| | |
|-------------------|--|
| UDP-GlcA | uridine 5'-diphospho- α -D-glucuronic acid |
| UDP-GlcNAc | uridine 5'-diphospho-N-acetyl- α -D-glucosamine |
| UTP | uridine triphosphate |

Nomenclature

| | |
|------------|-----------|
| Fuc | fucose |
| Gal | galactose |
| Glc | glucose |
| Man | mannose |

References

1. R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683–720.
2. J. W. Dennis; M. Granovsky; C. E. Warren, *BioEssays* **1999**, *21*, 412–421.
3. R. Apweiler; H. Hermjakob; N. Sharon, *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
4. P. M. Rudd; M. R. Wormald; R. L. Stanfield; M. Huang; N. Mattsson; J. A. Speir; J. A. DiGennaro; J. S. Fetrow; R. A. Dwek; I. A. Wilson, *J. Mol. Biol.* **1999**, *293*, 351–366.
5. S. I. Hakomori Si, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 225–232.
6. O. Seitz, *ChemBioChem* **2000**, *1*, 214–246.
7. C.-H. Wong; R. L. Halcomb; Y. Ichikawa; T. Kajimoto, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 412–432.
8. C.-H. Wong; R. L. Halcomb; Y. Ichikawa; T. Kajimoto, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 521–546.
9. K. M. Koeller; C.-H. Wong, *Glycobiology* **2000**, *10*, 1157–1169.
10. D. D. Allison; K. J. Grande-Allen, *Tissue Eng.* **2006**, *12*, 2131–2140.
11. N. Volpi, *Curr. Med. Chem.* **2006**, *13*, 1799–1810.
12. G. W. Yip; M. Smollich; M. Götte, *Mol. Cancer Ther.* **2006**, *5*, 2139–2148.
13. M. Prabaharan; J. F. Mano, *Macromol. Biosci.* **2006**, *6*, 991–1008.
14. V. L. Finkenstadt, *Appl. Microbiol. Biotechnol.* **2005**, *67*, 735–745.
15. S. Roseman, *J. Biol. Chem.* **2001**, *276*, 41527–41542.
16. C. R. Bertozzi; L. L. Kiessling, *Science* **2001**, *291*, 2357–2364.
17. P. Sears; C.-H. Wong, *Science* **2001**, *291*, 2344–2350.
18. O. J. Plante; E. R. Palmacci; P. H. Seeberger, *Science* **2001**, *291*, 1523–1527.
19. K. M. Koeller; C.-H. Wong, *Nature* **2001**, *409*, 232–240.
20. N. Wymer; E. J. Toone, *Curr. Opin. Chem. Biol.* **2000**, *4*, 110–119.
21. M. M. Palcic, *Curr. Opin. Biotechnol.* **1999**, *10*, 616–624.
22. D. H. Crout; G. Vic, *Curr. Opin. Chem. Biol.* **1998**, *2*, 98–111.
23. S. Hurlley; R. Service; P. Szuromi, *Science* **2001**, *291*, 2337.
24. D. H. G. Crout; P. Critchley; D. Muller; M. Scigelova; S. Singh; G. Vic, *Spec. Publ.—R. Soc. Chem.* **1999**, *246*, 15–23.
25. F. Van Rantwijk; O. M. Woudenberg-van; R. A. Sheldon, *J. Mol. Catal., B Enzym.* **1999**, *6*, 511–532.
26. D. J. Vocadlo; S. G. Withers, *Methods Mol. Biol.* **2000**, *146*, 203–222.
27. S. G. Withers, *Can. J. Chem.* **1999**, *77*, 1–11.
28. H. D. Ly; S. G. Withers, *Annu. Rev. Biochem.* **1999**, *68*, 487–522.
29. D. L. Jakeman; S. G. Withers, *Trends Glycosci. Glycotechnol.* **2002**, *14*, 13–25.
30. J. F. Tolborg; L. Petersen; K. J. Jensen; C. Mayer; D. L. Jakeman; R. A. Warren; S. G. Withers, *J. Org. Chem.* **2002**, *67*, 4143–4149.
31. S. Roseman, *Chem. Phys. Lipids* **1970**, *5*, 270–297.
32. W. M. Watkins, *Carbohydr. Res.* **1986**, *149*, 1–12.
33. E. J. Hehre, *Carbohydr. Res.* **2001**, *331*, 347–368.
34. U. M. Unligil; J. M. Rini, *Curr. Opin. Struct. Biol.* **2000**, *10*, 510–517.
35. G. J. Davies; B. Henrissat, *Biochem. Soc. Trans.* **2002**, *30*, 291–297.
36. M. Kaneko; S. Nishihara; H. Narimatsu; N. Saitou, *Trends Glycosci. Glycotechnol.* **2001**, *13*, 147–155.
37. M. M. Palcic; K. Sujino; X. Qian, *Carbohydr. Chem. Biol.* **2000**, *2*, 685–703.
38. K. Sujino; T. Uchiyama; O. Hindsgaul; N. O. L. Seto; W. W. Wakarchuk; M. M. Palcic, *J. Am. Chem. Soc.* **2000**, *122*, 1261–1269.
39. M. Fukuda; M. F. Bierhuizen; J. Nakayama, *J. Glycobiol.* **1996**, *6*, 683–689.
40. J. Fang; J. Li; X. Chen; Y. Zhang; J. Wang; Z. Guo; W. Zhang; L. Yu; K. Brew; P. G. Wang, *J. Am. Chem. Soc.* **1998**, *120*, 6635–6638.

41. N. Q. Palacpac; S. Yoshida; H. Sakai; Y. Kimura; K. Fujiyama; T. Yoshida; T. Seki, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4692–4697.
42. S. Shibatani; K. Fujiyama; S. Nishiguchi; T. Seki; Y. Maekawa, *J. Biosci. Bioeng.* **2001**, *91*, 85–87.
43. M. Malissard; S. Zeng; E. G. Berger, *Glycoconj. J.* **1999**, *16*, 125–139.
44. K. F. Johnson, *Glycoconj. J.* **1999**, *16*, 141–146.
45. O. Blixt; I. van Die; T. Norberg; D. H. van den Eijnden, *Glycobiology* **1999**, *9*, 1061–1071.
46. M. Izumi; G.-J. Shen; S. Wacowich-Sgarbi; T. Nakatani; O. Plettenburg; C.-H. Wong, *J. Am. Chem. Soc.* **2001**, *123*, 10909–10918.
47. P. L. DeAngelis, *Glycobiology* **2002**, *12*, 9R–16R.
48. C.-H. Wong; S. L. Haynie; G. M. Whitesides, *J. Org. Chem.* **1982**, *47*, 5416–5418.
49. Y. Ichikawa; R. Wang; C.-H. Wong, *Methods Enzymol.* **1994**, *247*, 107–127.
50. C.-H. Wong; R. Wang; S. Ichikawa, *J. Org. Chem.* **1992**, *57*, 4343–4344.
51. T. Endo; S. Koizumi; K. Tabata; S. Kakita; A. Ozaki, *Carbohydr. Res.* **1999**, *316*, 179–183.
52. T. Endo; S. Koizumi, *Curr. Opin. Struct. Biol.* **2000**, *10*, 536–541.
53. L. Elling; M. Grothus; M. R. Kula, *Glycobiology* **1993**, *3*, 349–355.
54. Y. Ichikawa; Y. C. Lin; D. P. Dumas; G. J. Shen; E. Garcia-Junceda; M. A. Williams; R. Bayer; C. Ketcham; L. E. Walker, *J. Am. Chem. Soc.* **1992**, *114*, 9283–9298.
55. P. Wang; G. J. Shen; Y. F. Wang; Y. Ichikawa; C.-H. Wong, *J. Org. Chem.* **1993**, *58*, 3985–3990.
56. S. Takayama; G. J. McGarvey; C.-H. Wong, *Annu. Rev. Microbiol.* **1997**, *51*, 285–310.
57. C. S. Rye; S. G. Withers, *Curr. Opin. Chem. Biol.* **2000**, *4*, 573–580.
58. P. Monsan; F. Paul, *FEBS Microbiol. Rev.* **1995**, *16*, 187–192.
59. M. Scigelova; S. Singh; D. H. G. Crout, *J. Mol. Catal., B Enzym.* **1999**, *6*, 483–494.
60. R. Lee; D. Mansey; A. Watson; K. Duncan; C. Rithner; M. McNeil, *Anal. Biochem.* **1996**, *242*, 1–7.
61. G. L. Cote; B. Y. Tao, *Glycoconj. J.* **1990**, *7*, 145–162.
62. S. J. Williams; S. G. Withers, *Carbohydr. Res.* **2000**, *327*, 27–46.
63. P. Fialova; L. Weignerova; J. Rauvolfova; V. Prikylova; A. Pisvejcova; R. Ettrich; M. Kuzma; P. Sedmera; V. Kren, *Tetrahedron* **2004**, *60*, 693–701.
64. A. Tauss; P. Greimel; K. Rupitz; A. J. Steiner; A. E. Stutz; S. G. Withers; T. M. Wrodnigg, *Tetrahedron, Asymmetry* **2005**, *16*, 159–165.
65. L. X. Wang, *Carbohydr. Res.* **2008**, *343*, 1509–1522.
66. M. S. Macauley; G. E. Whitworth; A. W. Debowski; D. Chin; D. J. Vocadlo, *J. Biol. Chem.* **2005**, *280*, 25313–25322.
67. D. H. G. Crout; T. A. MacManus; J. M. Ricca; S. Singh; P. Critchley; W. T. Gibson, *Pure Appl. Chem.* **1992**, *64* (8), 1079–1084.
68. S. Singh; M. Scigelova; G. Vic; D. H. G. Crout, *J. Chem. Soc. Perkin Trans. 1* **1996**, 1921–1926.
69. S. Singh; M. Scigelova; P. Critchley; D. H. G. Crout, *Carbohydr. Res.* **1997**, *305*, 363–370.
70. K. Ajisaka; H. Fujimoto; M. Miyasato, *Carbohydr. Res.* **1998**, *309*, 125–129.
71. K. Koizumi; T. Tanimoto; Y. Okada; S. Takeyama; K. Hamayasu; H. Hashimoto; S. Kitahata, *Carbohydr. Res.* **1998**, *314*, 115–325.
72. A. Ismail; S. Soultani; M. Ghouli, *Biotechnol. Prog.* **1998**, *14*, 874–878.
73. B. W. Murray; S. Takayama; J. Schultz; C. H. Wong, *Biochemistry* **1996**, *35* (34), 11183–11195.
74. S. Kobayashi; T. Kiyosada; S.-I. Shoda, *J. Am. Chem. Soc.* **1996**, *118*, 13113–13114.
75. C. Rémond; M. Ferchichi; N. Aubry; R. Plantier-Royon; C. Portella; M. J. O'Donohue, *Tetrahedron Lett.* **2002**, *43*, 9653–9655.
76. C. Rémond; R. Plantier-Royon; N. Aubry; E. Maes; C. Bliard; M. J. O'Donohue, *Carbohydr. Res.* **2004**, *339*, 2019–2025.
77. C. Rémond; R. Plantier-Royon; N. Aubry; M. J. O'Donohue, *Carbohydr. Res.* **2005**, *340*, 637–644.
78. T. Debeche; N. Cummings; I. Connerton; P. Debeire; M. J. O'Donohue, *Appl. Environ. Microbiol.* **2000**, *66*, 1734–1736.
79. G. Lopez; C. Nugier-Chauvin; C. Rémond; M. J. O'Donohue, *Carbohydr. Res.* **2007**, *342*, 2202–2211.
80. S. Kobayashi; A. Makino; H. Matsumoto; S. Kuni; M. Ohmae; T. Kiyosada; K. Makiguchi; A. Matsumoto; M. Horie; S.-I. Shoda, *Biomacromolecules* **2006**, *7*, 1644–1656.
81. R. Raman; V. Sasisekharan; R. Sasisekharan, *Chem. Biol.* **2005**, *12*, 267–277.
82. X. F. Bao; S. Nishimura; T. Mikami; S. Yamada; N. Itoh; K. Sugahara, *J. Biol. Chem.* **2004**, *279*, 9765–9776.
83. H. Ochiai; S.-I. Fujikawa; M. Ohmae; S. Kobayashi, *Biomacromolecules* **2007**, *8*, 1802–1806.
84. S. Kobayashi; M. Ohmae; H. Ochiai; S. Fujikawa, *Chem. Eur. J.* **2006**, *12*, 5962–5971.
85. L. Weignerova; P. Vavruskova; A. Pisvejcova; J. Thiem; V. Kren, *Carbohydr. Res.* **2003**, *338*, 1003–1008.
86. X. Zeng; Y. Sun; H. Ye; J. Liu; H. Uzawa, *Biotechnol. Lett.* **2007**, *29*, 1105–1110.
87. C. Bucke, *J. Chem. Technol. Biotechnol.* **1996**, *67*, 217–220.
88. J. Zhang; B. Wu; Z. Y. Liu; P. Kowal; X. Chen; J. Shao; P. G. Wang, *Curr. Org. Chem.* **2001**, *5*, 1169–1176.
89. L. F. Mackenzie; Q. Wang; R. A. J. Warren; S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.
90. C. Malet; A. Planas, *FEBS Lett.* **1998**, *440*, 208–212.
91. S. Fort; V. Boyer; L. Greffe; G. J. Davies; O. Moroz; L. Christiansen; M. Schulein; M. Schülein; S. Cottaz; H. Driguez, *J. Am. Chem. Soc.* **2000**, *122*, 5429–5437.
92. C. Mayer; D. L. Zechel; S. P. Reid; R. A. J. Warren; S. G. Withers, *FEBS Lett.* **2000**, *466*, 40–44.
93. A. S. Rowan; C. J. Hamilton, *Nat. Prod. Rep.* **2006**, *23*, 412–443.
94. J. Drone; H. Y. Feng; C. Tellier; L. Hoffmann; V. Tran; C. Rabiller; M. Dion, *Eur. J. Org. Chem.* **2005**, *10*, 1977–1983.
95. J. K. Fairweather; M. Hrmova; S. J. Rutten; G. B. Fincher; H. Driguez, *Chem. Eur. J.* **2003**, *9*, 2603–2610.
96. A. Trincone; A. Giordano; G. Perugino; M. Rossi; M. Moracci, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4039–4042.
97. G. Perugino; A. Trincone; A. Giordano; J. van der Oost; T. Kaper; M. Rossi; M. Moracci, *Biochemistry* **2003**, *42*, 8484–8493.
98. V. M. A. Ducros; C. A. Tarling; D. L. Zechel; M. Schulein; S. G. Withers; G. Davies, *J. Chem. Biol.* **2003**, *10*, 619–628.
99. R. V. Stick; K. A. Stubbs; A. G. Watts, *Aust. J. Chem.* **2004**, *57*, 779–786.
100. M. Fajjes; T. Ima; V. Bulone; A. Planas, *Biochem. J.* **2004**, *380*, 635–641.
101. M. Fajjes; X. Pérez; O. Pérez; A. Planas, *Biochemistry* **2003**, *42*, 13304–13318.

102. Y. W. Kim; S. S. Lee; R. A. J. Warren; S. G. Withers, *J. Biol. Chem.* **2004**, *279*, 42787–42793.
103. S. Kobayashi; M. Ohmae, *Adv. Polym. Sci.* **2006**, *194*, 159–210.
104. M. Fajjes; M. Saura-Valls; X. Pérez; M. Conti; A. Planas, *Carbohydr. Res.* **2006**, *341*, 2055–2065.
105. Y. W. Kim; D. T. Fox; O. Hekmat; T. Kantner; L. P. McIntosh; R. A. J. Warren; S. G. Withers, *Org. Biomol. Chem.* **2006**, *4*, 2025–2032.
106. S. Fort; S. Cottaz; H. Driguez; L. Christiansen; M. Schülein, *Isr. J. Chem.* **2000**, *40*, 218–221.
107. V. Boyer; S. Fort; T. P. Frandsen; M. Schülein; S. Cottaz; H. Driguez, *Chem. Eur. J.* **2002**, *8*, 1389–1394.
108. S. Blanchard; S. Cottaz; P. M. Coutinho; S. Patkar; J. Vind; H. Boer; A. Koivula; H. Driguez; S. Armand, *J. Mol. Catal., B Enzym.* **2007**, *44*, 106–116.
109. S. Blanchard; S. Armand; P. Coutinho; S. Patkar; J. Vind; E. Samain; H. Driguez; S. Cottaz, *Carbohydr. Res.* **2007**, *342*, 710–716.
110. J. Mullegger; M. Jahn; H. M. Chen; R. A. J. Warren; S. G. Withers, *Protein Eng. Des. Sel.* **2005**, *18*, 33–40.
111. L. A. Rowe; M. L. Geddie; O. B. Alexander; I. Matsumura, *J. Mol. Biol.* **2003**, *332*, 851–860.
112. J. Kruszewska; A. Janik; U. Lenart; G. Palamarczyk, *Acta Biochim. Pol.* **1999**, *46*, 315–324.
113. J. S. Rush; K. Panneerselvam; C. J. Waechter; H. H. Freeze, *Glycobiology* **2000**, *10*, 829–835.
114. L. F. Leloir, *Science* **1971**, *172*, 1299–1303.
115. J. E. Heidlas; K. W. Williams; G. M. Whitesides, *Acc. Chem. Res.* **1992**, *25*, 307–314.
116. C. Unverzagt; H. Kunz; J. C. Paulson, *J. Am. Chem. Soc.* **1990**, *112*, 9308–9309.
117. Y. Ichikawa; R. Wang; C.-H. Wong, *Methods Enzymol.* **1994**, *247*, 107–127.
118. M. M. Palcic, *Methods Enzymol.* **1994**, *230*, 300–316.
119. K. M. Koeller; C.-H. Wong, *Chem. Rev.* **2000**, *100*, 4465–4493.
120. J. Shao; J. Zhang; P. Kowal; Y. Lu; P. G. Wang, *Biochem. Biophys. Res. Commun.* **2002**, *295*, 1–8.
121. J. Shao; J. Zhang; P. Kowal; P. G. Wang, *Appl. Environ. Microbiol.* **2002**, *68*, 5634–5640.
122. J. Zhang; P. Kowal; J. Fang; P. Andreatana; P. G. Wang, *Carbohydr. Res.* **2002**, *337*, 969–976.
123. N. Sharon; I. Ofek, *Glycoconj. J.* **2000**, *17*, 659–664.
124. N. Sharon; I. Ofek, *Crit. Rev. Food Sci. Nutr.* **2002**, *42*, 267–272.
125. S. L. Haynie; G. M. Whitesides, *Appl. Biochem. Biotechnol.* **1990**, *23*, 155–170.
126. C. H. Hokke; A. Zervosen; L. Elling; D. H. Joziassse; D. H. van den Eijnden, *Glycoconj. J.* **1996**, *13*, 687–692.
127. Y. Ichikawa; L. C. J. Liu; G. J. Shen; C.-H. Wong, *J. Am. Chem. Soc.* **1991**, *113*, 6300–6302.
128. A. Zervosen; L. Elling, *J. Am. Chem. Soc.* **1996**, *118*, 1836–1840.
129. X. Chen; J. Fang; J. Zhang; Z. Liu; J. Shao; P. Kowal; P. Andreatana; P. G. Wang, *J. Am. Chem. Soc.* **2001**, *123*, 2081–2082.
130. J. Nahalka; Z. Liu; P. Gemeiner; P. G. Wang, *Biotechnol. Lett.* **2002**, *24*, 925–930.
131. J. Nahalka; Z. Liu; X. Chen; P. G. Wang, *Chem.—Eur. J.* **2003**, *9*, 372–377.
132. J. Zhang; B. Wu; Y. Zhang; P. Kowal; P. G. Wang, *Org. Lett.* **2003**, *5*, 2583–2586.
133. J. Shao; J. Zhang; P. Kowal; Y. Lu; P. G. Wang, *Chem. Commun.* **2003**, *12*, 1422–1423.
134. X. Chen; Z. Liu; J. Zhang; W. Zhang; P. Kowal; P. G. Wang, *ChemBioChem* **2002**, *3*, 47–53.
135. P. G. Wang; X. Chen; J. Zhang; P. Kowal; P. R. Andreatana, In Abstracts of Papers, Proceedings of the 222nd ACS National Meeting, Chicago, IL, USA, 26–30 August 2001: MEDI-126, 2001.
136. S. Teneberg; J. Angstroem; A. Ljungh, *Glycobiology* **2004**, *14*, 187–196.
137. M. M. Palcic, *Curr. Opin. Biotechnol.* **1999**, *10*, 616–624.
138. J. Zhang; P. Kowal; X. Chen; P. G. Wang, *Org. Biomol. Chem.* **2003**, *1*, 3048–3053.
139. Q. Yao; J. Song; C. Xia; W. Zhang; P. G. Wang, *Org. Lett.* **2006**, *8*, 911–914.
140. X. Chen; P. Kowal; S. Hamad; H. Fan; P. G. Wang, *Biotechnol. Lett.* **1999**, *21*, 1131–1135.
141. D. M. Su; H. Eguchi; W. Yi; L. Li; P. G. Wang; C. Xia, *Org. Lett.* **2008**, *10*, 1009–1012.
142. N. O. L. Seto; M. M. Palcic; O. Hindsgaul; D. R. Bundle; S. A. Narang, *Eur. J. Biochem.* **1995**, *234*, 323–328.
143. X. Chen; Z. Liu; J. Wang; J. Fang; H. Fan; P. G. Wang, *J. Biol. Chem.* **2000**, *275*, 31594–31600.
144. J. Fang; X. Chen; W. Zhang; J. Wang; P. Andreatana; P. G. Wang, *J. Org. Chem.* **1999**, *64*, 4089–4094.
145. J. Fang; X. Chen; W. Zhang; A. Janczuk; P. G. Wang, *Carbohydr. Res.* **2000**, *329*, 873–878.
146. D. Hoffmeister; J. Yang; L. Liu; J. S. Thorson, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13184–13189.
147. H. C. Losey; J. Q. Jiang; J. B. Biggins; M. Oberthur; X. Y. Ye; S. D. Dong; D. Kahne; J. S. Thorson; C. T. Walsh, *Chem. Biol.* **2002**, *9*, 1305–1314.
148. X. Fu; C. Albermann; C. S. Zhang; J. S. Thorson, *Org. Lett.* **2005**, *7*, 1513–1515.
149. C. Leimkuhler; Z. Chen; R. G. Kruger; M. Oberthur; W. Lu; C. T. Walsh; D. Kahne, *Tetrahedron: Asymmetry* **2005**, *16*, 599–603.
150. D. A. Thayer; C.-H. Wong, *Chem. Asian J.* **2006**, *1*, 445–452.
151. W. van Uden; H. Oeij; H. J. Woerdenbag; N. Pras, *Plant Cell Tissue Organ Cult.* **1993**, *34*, 169–175.
152. W. van Uden; H. Oeij; H. J. Woerdenbag; N. Pras, *Plant Cell Tissue Organ Cult.* **1994**, *38*, 103–113.
153. V. Nieder; M. Kutzer; V. Kren; R. G. Gallego; J. P. Kamerling; L. Elling, *Enzyme Microb. Technol.* **2004**, *34*, 407–414.
154. I. Nagashima; H. Shimizu; T. Matsushita; S.-I. Nishimura, *Tetrahedron Lett.* **2008**, *49*, 3413–3418.
155. X. Zeng; Y. Sun; H. Ye; J. Liu; X. L. Xiang; B. Zhou; H. Uzawa, *Carbohydr. Res.* **2007**, *342*, 1244–1248.
156. T. Angata; A. Varki, *Chem. Rev.* **2002**, *102*, 439–469.
157. X. Zeng; Y. Sun; H. Uzawa, *Biotechnol. Lett.* **2005**, *27*, 1461–1465.
158. V. Chazalet; K. Uehara; R.-A. Geremia; C. Breton, *J. Bacteriol.* **2001**, *183*, 7067–7075.
159. M. Li; X. Liu; J. Shao; J. Shen; Q. Jia; W. Yi; J. K. Song; R. Woodward; C. S. Chow; P. G. Wang, *Biochemistry* **2008**, *47*, 378–387.
160. W. Yi; R. S. Perali; H. Eguchi; E. Motari; R. Woodward; P. G. Wang, *Biochemistry* **2008**, *47*, 1241–1248.
161. M. Rivas; E. Miliwebsky; L. Balbi; B. Garcia; N. Leardini; M. Tous; G. Chillemi; A. Baschkier; L. Strugo, *Medicina* **2000**, *60*, 249–252.
162. G. Widmalm; K. Leontein, *Carbohydr. Res.* **1993**, *247*, 255–262.
163. W. Yi; J. Shao; L. Zhu; M. Li; M. Singh; Y. Lu; S. Lin; H. Li; K. Ryu; J. Shen; H. Guo; Q. Yao; C. A. Bush; P. G. Wang, *J. Am. Chem. Soc.* **2005**, *127*, 2040–2041.
164. C. A. Weijers; M. C. Franssen; G. M. Visser, *Biotechnol. Adv.* **2008**, *26*, 436–456.

165. K. Persson; H. D. Ly; M. Dieckelmann; W. W. Wakarchuk; S. G. Withers; N. C. Strynadka, *Nat. Struct. Biol.* **2001**, *8*, 166–175.
166. L. N. Gastinel; C. Bignon; A. K. Misra; O. Hindsgaul; J. H. Shaper; D. H. Joziase, *EMBO J.* **2001**, *20*, 638–649.
167. L. L. Lairson; C. P. Chiu; H. D. Ly; S. He; W. W. Wakarchuk; N. C. Strynadka; S. G. Withers, *J. Biol. Chem.* **2004**, *279*, 28339–28344.
168. L. L. Lairson; W. W. Wakarchuk; S. G. Withers, *Chem. Commun.* **2007**, *311*, 365–367.
169. S. A. Borisova; L. Zhao; C. E. Melancon, III; C.-L. Kao; H. W. Liu, *J. Am. Chem. Soc.* **2004**, *126*, 6534–6535.
170. C. L. Kao; S. A. Borisova; H. J. Kim; H. W. Liu, *J. Am. Chem. Soc.* **2006**, *128*, 5606–5607.
171. S. A. Borisova; H. J. Kim; X. Pu; H.-W. Liu, *ChemBioChem* **2008**, *9*, 1554–1558.
172. S. A. Borisova; C. Zhang; H. Takahashi; H. Zhang; A. W. Wong; J. S. Thorson; H.-W. Liu, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2748–2753.
173. V. Jarlier; H. Nikaido, *FEMS Microbiol. Lett.* **1994**, *123*, 11–18.
174. M. Daffe; P. J. Brennan; M. McNeil, *J. Biol. Chem.* **1990**, *265*, 6734–6743.
175. M. McNeil; M. Daffe; P. J. Brennan, *J. Biol. Chem.* **1990**, *265*, 18200–18206.
176. K. Mikusova; T. Yagi; R. Stern; M. R. McNeil; G. S. Besra; D. C. Crick; P. J. Brennan, *J. Biol. Chem.* **2000**, *275*, 33890–33897.
177. K. Mikusova; M. Mikus; G. S. Besra; I. Hancock; P. J. Brennan, *J. Biol. Chem.* **1996**, *271*, 7820–7828.
178. S. T. Cole; R. Brosch; J. Parkhill; T. Garnier; C. Churcher; D. Harris; S. V. Gordon; K. Eiglmeier; S. Gas; C. E. Barry, III; F. Tekaiia; K. Badcock; D. Basham; D. Brown; T. Chillingworth; R. Connor; R. Davies; K. Devlin; T. Feltwell; S. Gentles; N. Hamlin; S. Holroyd; T. Hornsby; 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.
179. J. A. Mills; K. Motichka; M. Jucker; H. P. Wu; B. C. Uhlik; R. J. Stern; M. S. Scherman; V. D. Vissa; F. Pan; M. Kundu; Y. F. Ma; M. McNeil, *J. Biol. Chem.* **2004**, *279*, 43540–43546.
180. P. J. Brennan, *J. Bacteriol.* **2006**, *188*, 6592–6598.
181. K. T. Huang; B. C. Wu; C. C. Lin; S. C. Luo; C. Chen; C. H. Wong; C. C. Lin, *Carbohydr. Res.* **2006**, *341*, 2151–2155.
182. K. M. Koeller; M. E. B. Smith; R. F. Huang; C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 4241–4242.
183. K. M. Koeller; M. E. B. Smith; C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 742–743.

Biographical Sketches



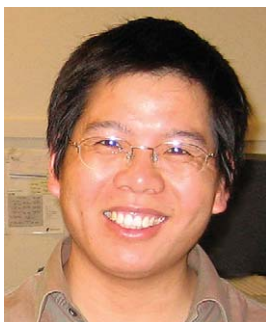
Wei Zhao obtained his BS in chemistry (1999) from Shandong University and his Ph.D. in environmental chemistry (2005) from the Research Center for Eco-Environmental Sciences (RCEES), Chinese Academy of Sciences. He is currently an associate professor in the College of Pharmacy at Nankai University, China. His research interests include glycoscience with an emphasis on carbohydrate chemistry, glycopharmaeaceutical science, carbohydrate-based vaccines, drugs and immunotherapy, as well as the structural and mechanism-based design and synthesis of inhibitors and ligands.



Tiehai Li obtained his BS in chemistry (2005) from Hubei Normal University, China. He is currently pursuing a Ph.D. at Nankai University, China. His research focuses on bioorganic chemistry, enzymatic synthesis of complex carbohydrates, and synthesis of new glycosidase inhibitors.



Robert Woodward obtained his BA in chemistry (2006) and his BS in biology (2006) from The Ohio State University. He is an NIH Chemistry-Biology Interface Program Predoctoral Fellow currently pursuing his Ph.D. in organic chemistry at The Ohio State University. His research focuses on the applications of synthesis for studying complex biological processes such as polysaccharide biosynthesis and oligosaccharide transfer.



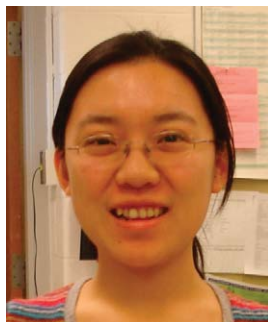
Chengfeng Xia obtained his BS in chemistry (1997) from Peking University and his Ph.D. (2002) from the Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences. He is currently a postdoctoral researcher at The Ohio State University. His research has focused on

the total synthesis of natural products, as well as on the structure–activity relationship study of glycolipids, the antigens of a newly discovered subpopulation of T cells, natural killer T cells.



Peng George Wang obtained his BS in chemistry (1984) from Nankai University, China, and his Ph.D. in organic chemistry (1990) from the University of California, Berkeley. He then conducted postdoctoral research at the Scripps Research Institute before becoming an assistant professor in 1994 at the University of Miami. From 1997 to 2003, he was a faculty member at Wayne State University. Since then, he has served in the Departments of Biochemistry and Chemistry at The Ohio State University as Ohio Eminent Scholar in Macromolecular Structure and Function.

Research in the Wang laboratory is predominately focused on four areas of glycoscience. Glycochemistry: Work is centered on the generation of uncommon sugar libraries as well as the synthesis of key intermediates in carbohydrate-based biological processes that are essential for their study. Glycobiology: Biochemical characterization of carbohydrate-active enzymes and investigations of the biological functions of carbohydrates in human diseases, immunity, and general microbiology are performed. Glycotechnology: Biosynthetic pathways are engineered for the synthesis of glycopharmaeuticals, polysaccharides for vaccine development, and biomedically important human glycoproteins. Glycoanalysis: Analysis of carbohydrate composition, sequence, structure, and their interaction with proteins through MS, NMR, QCM, and other analytical methods is conducted.



Wanyi Guan obtained her BS in biological sciences (2005) from Shandong University, China. She is currently pursuing her Ph.D. in glycobiology at the National Glycoengineering Research Center and State Key Laboratory of Microbial Technology at Shandong University, China. Her research focuses on the examination of substrate specificity in sugar kinases.

6.03 New Strategies for Glycopeptide, Neoglycopeptide, and Glycoprotein Synthesis

Ashraf Brik, Ben-Gurion University of the Negev, Beer Sheva, Israel

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|-------------------|---|----|
| 6.03.1 | Introduction | 55 |
| 6.03.1.1 | Nature of Glycopeptide and Glycoprotein | 56 |
| 6.03.1.2 | Neoglycopeptide | 57 |
| 6.03.2 | Synthesis of Glycopeptides | 58 |
| 6.03.2.1 | Solution-Phase Synthesis | 59 |
| 6.03.2.2 | Solid-Phase Synthesis | 61 |
| 6.03.2.3 | Enzymatic Synthesis | 66 |
| 6.03.2.4 | Chemical Ligation | 69 |
| 6.03.3 | Synthesis of Neoglycopeptide | 75 |
| 6.03.3.1 | C-Linked Glycopeptides | 75 |
| 6.03.3.2 | Oxime-Linked Glycopeptides | 77 |
| 6.03.3.3 | S-Linked Glycopeptides | 78 |
| 6.03.3.4 | Triazole-Linked Glycopeptides | 79 |
| 6.03.4 | Glycoprotein Synthesis | 79 |
| 6.03.4.1 | Chemical Approaches | 80 |
| 6.03.4.2 | Biochemical Approaches | 83 |
| 6.03.5 | Conclusions | 84 |
| References | | 85 |

6.03.1 Introduction

The attachment of a sugar residue to a protein side chain is believed to be the most complicated co- or posttranslational event.¹ As many as 13 various monosaccharides and 8 different amino acids (AAs) can be involved in this process.² The construction of a range of monosaccharides into glycans of varying lengths and sequences and their attachment to the amino acid side chains via anomeric bond lead to the formation of a large chemical diversity of glycoprotein structures. These products have an impressive number of glycopeptide bonds that would include N- and O-glycosidation, C-mannosylation, phosphoglycation and glypiation,² which are distributed among naturally occurring glycoproteins found in all living organisms. Not surprisingly, at least 16 enzymes are known to be involved in their formation at various intracellular sites, that is, endoplasmic reticulum, Golgi apparatus, cytosol, and nucleus.² On the other hand, several enzymes are known to cleave the N- and O-glycosidic bonds, that is, endoglycosidases and glycosidases, which appear to play many important physiological roles.¹

Glycoproteins and glycopeptides are fascinating natural products associated with several important biological processes, including protein folding, secretion, cell targeting and adhesion, stability in circulation, signal transduction, and many intercellular communication processes.^{1,3–5} Consequently, the majority of human proteins are glycosylated. Moreover, changes in the glycosylation pattern of a cell are often a signature of disease states, as in the case of cancer, rheumatoid arthritis, and some immunological diseases.⁶ In the drug discovery arena, glycoproteins are considered to be an important class of the biotechnology products.^{7,8} A well-studied example is that of erythropoietin (EPO), one of the most successful biotechnology drugs to date, which is used to treat anemia caused by bone marrow suppression. In EPO and other therapeutic glycoproteins, glycosylation is vital in several aspects, one of which is to ensure prolonged circulatory half-life in the blood.^{7,8}

6.03.1.1 Nature of Glycopeptide and Glycoprotein

A glycopeptide consists of a carbohydrate linked to a peptide composed of L- and/or D-amino acids. This would also include a biologically important class of glycopeptides that are nonribosomally synthesized, for example, vancomycin and bleomycin. However, this class of natural products will not be covered in this chapter as excellent reviews on this topic can be found elsewhere.⁹

Most of the naturally occurring glycosidic bonds can be divided into two main groups: the *N*- and *O*-glycoside. The β -glycosylamine linkage of the 2-acetamido-2-deoxy-D-glucose (GlcNAc) to Asn represents the most ubiquitous carbohydrate–peptide bond, found in a wide range of organisms ranging from Archaea to mammals.¹⁰ In this case, the glycosylation can only occur at Asn residue that is embedded in the consensus tripeptide sequence Asn-AA-Thr/Ser, AA being any amino acid except proline (Figure 1).¹¹ The *N*-glycosylation process starts with the oligosaccharyltransferase, which transfers a common oligosaccharide unit from dolichol phosphate to an Asn side chain. Subsequently, glycosylhydrolases trim the oligosaccharide to a pentasaccharide core, which will later be enzymatically glycosylated into various structures.

Glycosidic bond wherein the glycan unit is attached to the hydroxyl group, for example, *O*-glycoside, occurs in a large variety of proteins. Contrary to the *N*-glycosidic bonds, where it is limited to the β -configuration, the *O*-glycosidic bonds could adopt both α - and β -configurations.

In this class of glycopeptides, the glycan unit is attached mainly to the side chains of Ser, Thr, Tyr, and to a lesser extent to hydroxylysine and hydroxyproline.^{1,12} In mammals and other eukaryotes, the biosynthesis of *O*-glycosides often displays the 2-acetamido-2-deoxy-D-galactose (GalNAc) core unit attached through an α -*O*-linkage to the β -hydroxyl of Ser and Thr.^{13,14} Following this event, the GalNAc core is then diversified through a variety of glycosyltransferases furnishing the mucin-type *O*-glycosides (Figure 2).^{15,16} Mucins are glycoproteins that are

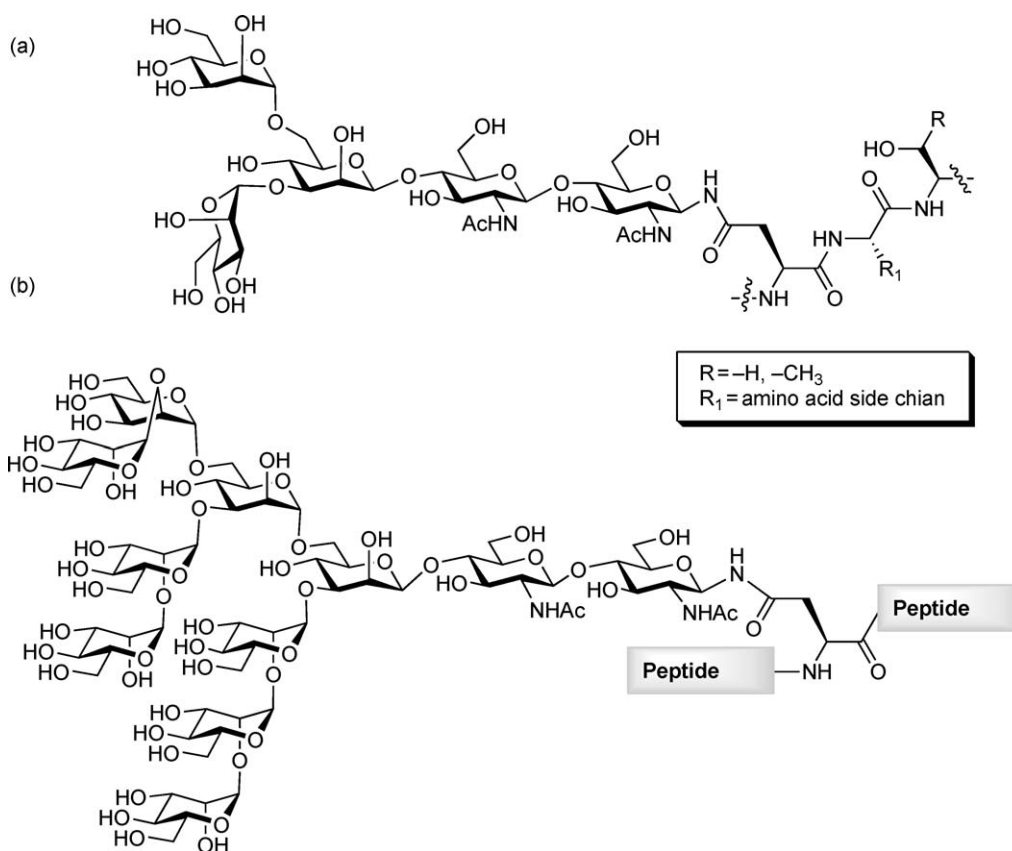


Figure 1 (a) Man₃GlcNAc₂ structure in the consensus sequences Asn-AA-Thr/Ser (AA – any amino acid but proline). (b) High mannose as an example of *N*-glycan in glycopeptides.

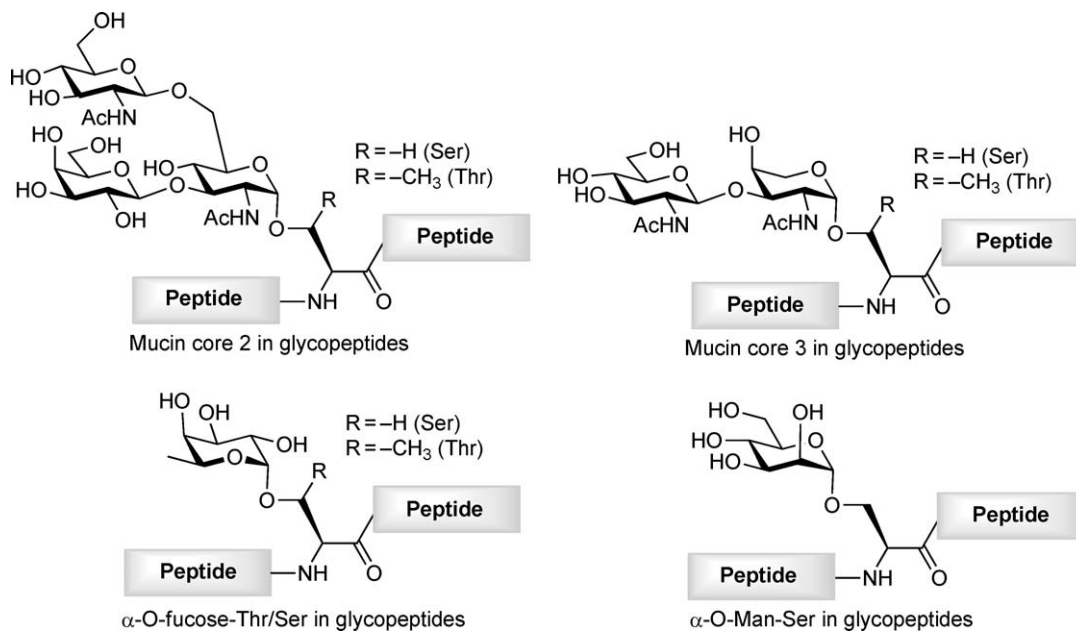


Figure 2 Examples of O-glycosidic bonds found in glycopeptides and glycoproteins.

excessively O-glycosylated, which can bear up to 20 monosaccharides per glycan. These glycoproteins can be found on the surface of epithelial cell types and are considered as an important class of tumor-associated antigens.¹⁷

While most of the peptide-linked monosaccharides can be further substituted by other sugars, the core GlcNAc attached via β -O-glycosidic bond does not become further elaborated. Thus, the sugar remains a simple monosaccharide attached to the protein. In return, this allows the use of radiolabeled UDP-Gal with galactosyltransferase as an effective probe for its presence, even in minute amounts.¹⁸ The GlcNAc core can be found in nuclear cytoskeletal proteins and the transcription factor and its addition seems to be involved with transcriptional regulations similar to phosphorylation.¹⁹ Several other important types of O-linked carbohydrates, for example, Gal- α -Ser/Thr, Fuc- α -Ser/Thr, and Man- α -Ser, have also been found and reported to have important biological functions.²

Protein glycosylation involves many enzymatic reactions in complex biosynthesis pathways that are not controlled by templates or transcription. Consequently, unpredictable glycoprotein products are formed, which possess many different glycoforms differing only in the glycan structure.⁵ While nature adapted well to this heterogeneity of products, scientists are struggling with this outcome in the context of the studies that are aiming to comprehend the biological functions of glycoproteins at the molecular level. Moreover, this complexity makes it difficult to generate protein-based medicines. Indeed, most therapeutic glycoprotein, for example, EPO, are administered as mixtures of glycoforms, the active form of which are often unknown. Therefore, there is a desperate need for methods that facilitate the preparations of homogeneous materials to assist researchers in their goals to understand the exact function/role of these sugars. With this in mind, scientists have tackled these problems from different directions; some of these strategies are merely based on chemical synthesis while others combine chemical and biochemical approaches.^{20–23} It is the aim of this chapter to review these strategies focusing on recent chemical approaches and pointing out the pros and cons of different methods, synthetic challenges, and possible solutions.

6.03.1.2 Neoglycopeptide

Natural product synthesis is the centerpiece of organic chemistry because it has always been the ultimate testing ground for new concepts and new synthetic methods.²⁴ Despite recent advances in carbohydrate chemistry^{24–27} and enzymatic glycosylation,^{28,29} glycopeptides and glycoproteins synthesis stands as one of

the most exigent tasks that synthetic chemists could encounter. The challenges could arise not only due to the large size of these macromolecules, but also due to the complexity of oligosaccharide chemistry. In this regard, the synthetic scheme would often require extensive protecting group manipulations and difficult glycosyl coupling conditions that are carefully designed to generate the correct regio- and stereochemistry of the glycosidic bond. Moreover, the convergent approach used to assemble the entire protein is challenging and is often faced with several obstacles. Yet, in order to meet the urgent need for glycoprotein material with a well-defined structure for biological studies and as novel medicines, a new field of glycoprotein mimetics has emerged.

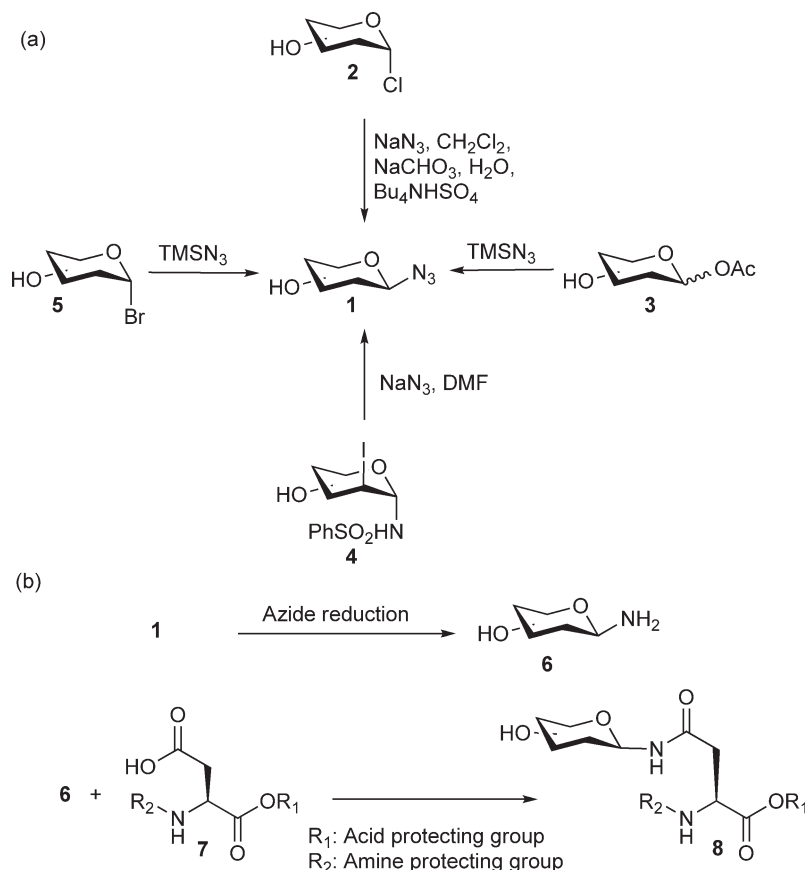
In the glycopeptide mimetic field, also known as neoglycopeptides, the native structure is sacrificed to mimicry, where the glycosidic bond is replaced with a more stable bond; for the benefit of easier synthesis.³⁰ The stability issue of the sugar peptide linkage is very important when dealing with glycoprotein therapeutic. One of the primary paths of glycoprotein degradation involves an enzymatic cleavage of the glycan from the peptide. As a result, replacement of the glycosidic bond with a more stable linkage without dramatically affecting the biological activity would be a reasonable way to follow. Indeed, researchers have found that such an effort wherein the glycosidic bond was replaced with C–S or C–C bond led to a more stable product against enzymatic degradation.^{31–33} Moreover, such substitutions were also found not to affect the biological and conformational properties by a measurable state compared to the native molecules. In this chapter, we will visit the synthetic efforts invested in this field pointing out surrogates based on the carbon–carbon, thioether, oxime, and triazole substitution as promising replacements of the glycosidic bond.

6.03.2 Synthesis of Glycopeptides

Chemical synthesis of glycopeptides, whether carried out on solid support or in solution, requires the construction of the glycosidic bond with the correct stereochemistry to one of the amino acid side chains. In the N-linked glycopeptides, the glycosidic bond is an amide; thus, most of the methods applied to assemble this bond use an amide-forming reaction between an amine from the sugar part and the carboxylate side chain of the aspartic acid. As a result, various methods were reported to access this precursor, one of which is the direct preparation from unprotected sugars using ammonia in methanol or ammonium hydrogen carbonate.³⁴ However, one may conclude from surveying the studies in this field that the synthesis of glycosyl azide **1** (**Scheme 1(a)**) is one of the most desirable building blocks in introducing the amine moiety. The stability of the azide functionality at the anomeric center, the facile introduction of the azide function, and the availability of effective methods for the conversion to the amine moiety stand behind favoring the use of glycosyl azide.

The synthesis of the glycosyl azide can be achieved via different approaches using an azide as a nucleophile and an activated anomeric center (**Scheme 1(a)**). For example, when treated with various azide salts in organic solvents and in presence of phase transfer catalyst, glycosyl chloride **2** leads to the formation of the glycosyl azide **1** in high yields (**Scheme 1(a)**).^{35–38} Alternatively, glycosyl bromide **5** or acetates **3** can be used with trimethylsilyl azide (TMSN₃) to give the desired product.^{39–41} A different route for constructing such a system starting from the glycal precursor has also been reported.⁴² Once the glycosyl azide is assembled, the reduction of the azide to the corresponding amine **7** is achieved using hydrogenation,⁴³ Staudinger reaction,⁴⁴ hydride reduction,⁴⁵ and propanedithiol-mediated reduction (**Scheme 1(b)**).⁴⁶ At this stage, the glycosylamine is attached to the side chain of the protected aspartic acid to form the preglycosylated asparagines **8**, ready for incorporation into the peptide sequence (**Scheme 1(b)**).

An example for the use of a glycosyl azide in the synthesis of an amino acid glycan building block for glycopeptide synthesis is shown in **Scheme 2**. Here, the Bertozzi group started the synthesis with the known monosaccharide **9**, which was converted to the disaccharide **10** in four steps.⁴⁷ Next, **10** was treated with the thioglycoside **11** followed by protecting group manipulations to afford trisaccharide **12**. Selective glycosylation with the donor **13** at the 3'-position, using SnCl₂/AgOTf as a catalyst, produced the tetrasaccharide in high yield and excellent selectivity. The 6'-position of oligosaccharide was selectively glycosylated with **13** to give pentasaccharide **14**. At this stage, the azide was reduced with propanedithiol to give, with complete retention of stereochemistry, the β -glycosylamine **15**. Peptide bond formation was then achieved using a protected aspartic



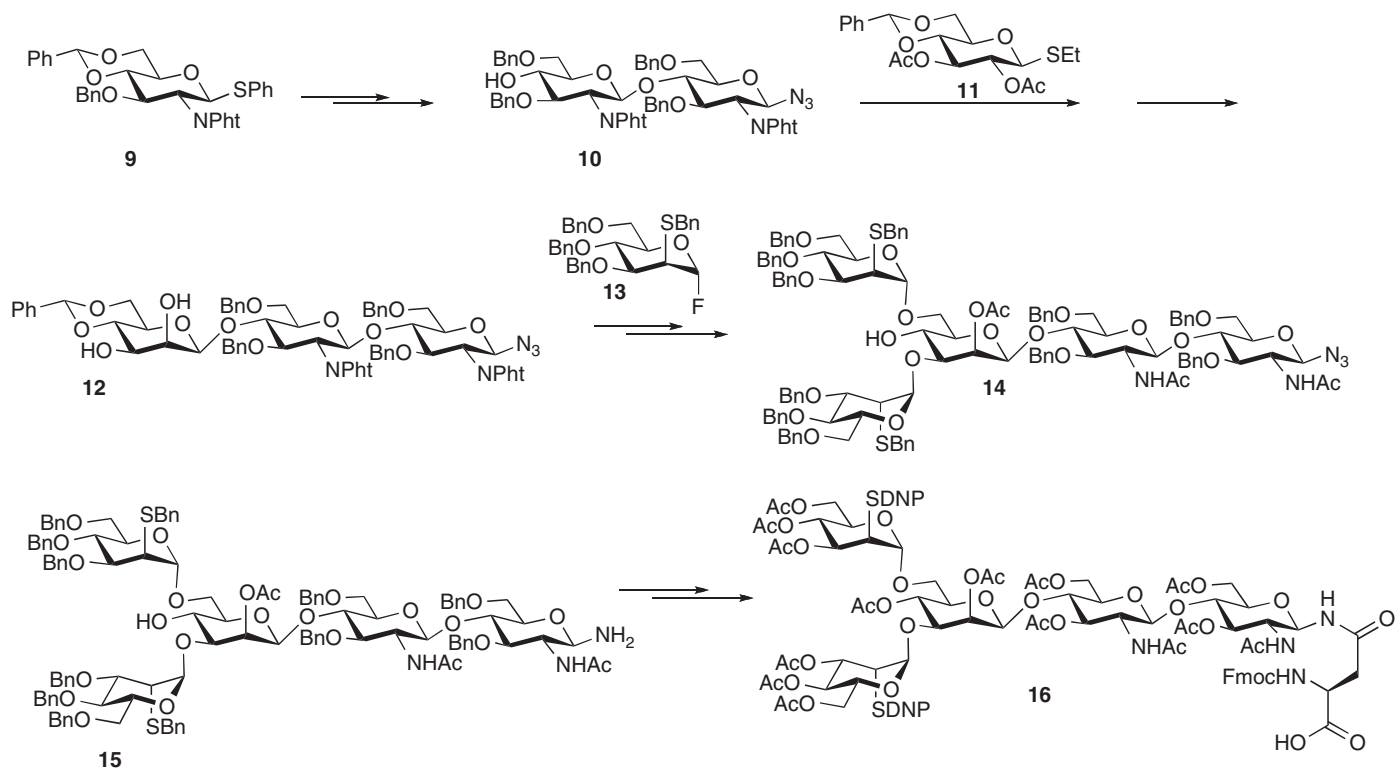
Scheme 1 (a) Synthesis of glycosyl azide. (b) Reduction of glycosyl azide to the corresponding glycosylamine followed by coupling with protected aspartic acid.

acid derivative, which after protecting group manipulations led to the glycosyl amino acid building block **16** ready for solid-phase peptide synthesis (SPPS). Previous to this work, the synthesis of CD52 containing the conserved N-linked pentasaccharide core was completed by Ogawa and coworkers.^{48,49}

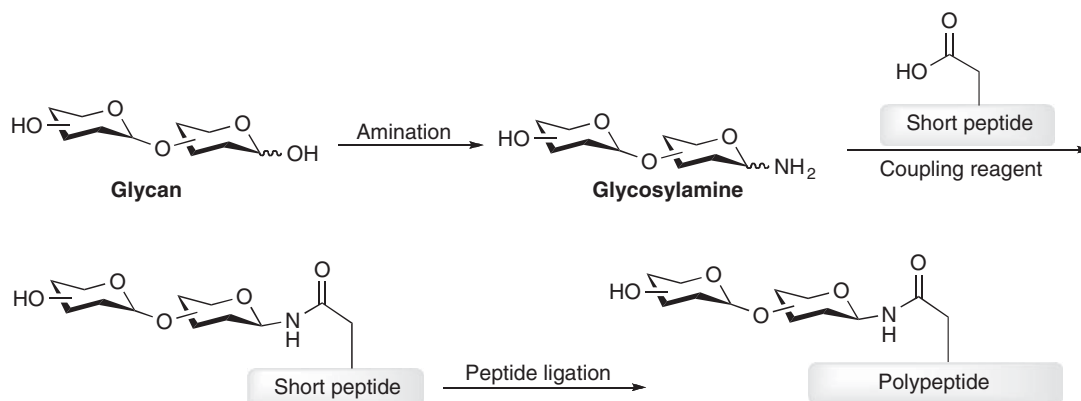
6.03.2.1 Solution-Phase Synthesis

Glycopeptide synthesis of short glycopeptide sequences could be performed in solution phase.^{50,51} Indeed, early developments in the field were focused on solution phase due to the lack of suitable linkers and difficult assessment of complete reactions on solid support. Importantly, the synthesis of glycopeptide in solution allows the use of a limited amount of the precious glycosyl amino acid. Another advantage of using the solution-phase synthesis is that the reaction can be monitored through classic methods such as thin layer chromatography (TLC) and nuclear magnetic resonance (NMR). Moreover, after each step product isolation can be accomplished by simple diethyl ether precipitation and filtration, thus taking advantage of the free glycan which acts as a phase tag to facilitate product isolation. Finally, the use of solution phase avoids the use of acid conditions that often are needed to release the glycopeptide from the solid support and unmask the amino acid side chains.

The direct coupling of glycosylamines to aspartic acid containing peptides in solution phase was studied and applied for the synthesis of a complex glycopeptide. This convergent approach for the chemical synthesis of asparagine-linked glycopeptides relies on the synthesis of protected peptides with orthogonally protected aspartic acid side chain, which upon removal permits activation and the coupling of glycosylamine (**Scheme 3**).⁵² The major limitation of this approach is the size of the used protected peptide wherein solubility problems in organic solvents and aspartamide formation may emerge during the activation step.



Scheme 2 Synthesis of glycosyl amino acid building block by applying glycosyl azide.



Scheme 3 The convergent approach of the N-linked glycopeptide.

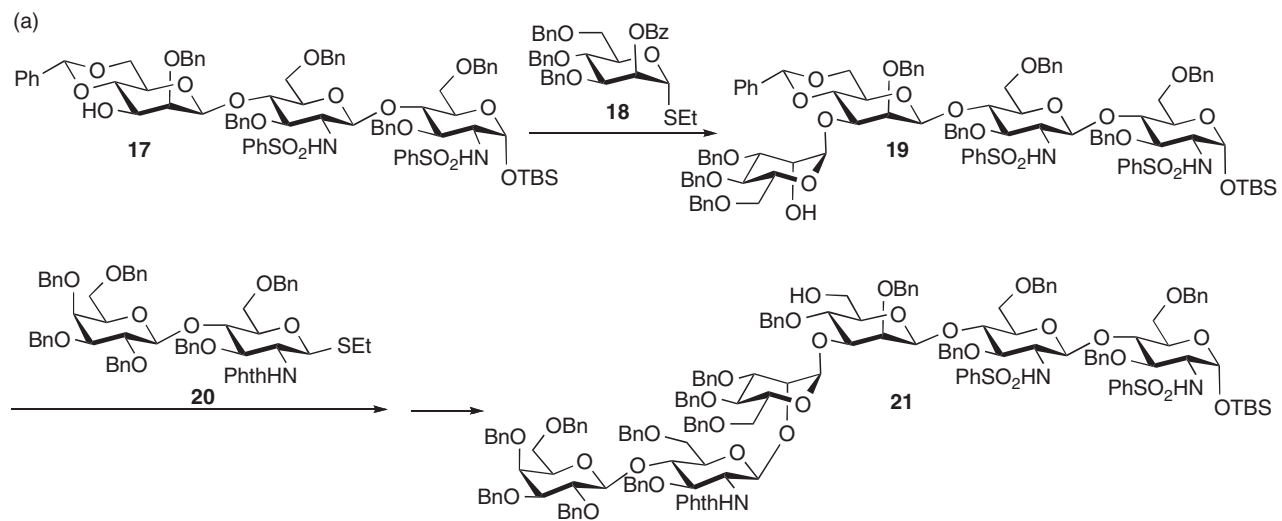
Using the convergent approach, the Danishefsky group accomplished the first synthesis of hybrid- and high-mannose-type gp120 glycopeptide fragments, which could lead to the design of a new human immunodeficiency virus (HIV) vaccine based on carbohydrate structures.^{53,54} Here, the authors synthesized mature hybrid-type HIV gp120 glycopeptide fragments by applying a synthetic strategy which included 3-mannosylations to construct the core trisaccharide and a fragment-coupling strategy to install the ‘upper’ high-mannose domain. Thus, the synthesis started with trisaccharide acceptor **17**, which was glycosylated with donor **18** to give **19** followed by another glycosylation with lactosamine thiol donor **20** to give the ‘lower’ part hexasaccharide **21** (**Scheme 4(a)**). Further protected group manipulations and glycosylations of the hexamer acceptor with trisaccharide donor **22** afforded the nonasaccharide **23**. After liberating the anomeric hydroxyl site and global deprotection steps followed by acetylation of the free amine, the glycan with the anomeric amine **24** was obtained by following the Kotchetkov amination procedure.³⁴ The glycan was then conjugated with the gp120 peptide segments through direct aspartylation to a key asparagine (332) residue to assemble the desired glycopeptides (**Scheme 4(b)**).

Glycopeptides bearing O-linkages were also synthesized in solution phase. In general the preparation of this family, in particular glycopeptides with complex glycans, is faced with the selectivity issue of constructing the α -isomer in the formation of the O-Ser/Thr mucin-type linkage. Solutions to this obstacle have been sought wherein Winterfled and Schmidt reported that nitroglycals of mono- and disaccharide react to form the α -Ser/Thr linkage in high selectivity, thus accessing Tn and Sialyl Tn antigens.⁵⁵ A general approach referred to as the cassette strategy has been developed to overcome the selectivity problem.^{56–59} In this approach, the desired α -Ser/Thr linkage is formed prior to the addition of branching residues onto the core GalNAc of an α -glycosyl amino acid. Using the cassette strategy, several tumor-related antigens were prepared.⁶⁰

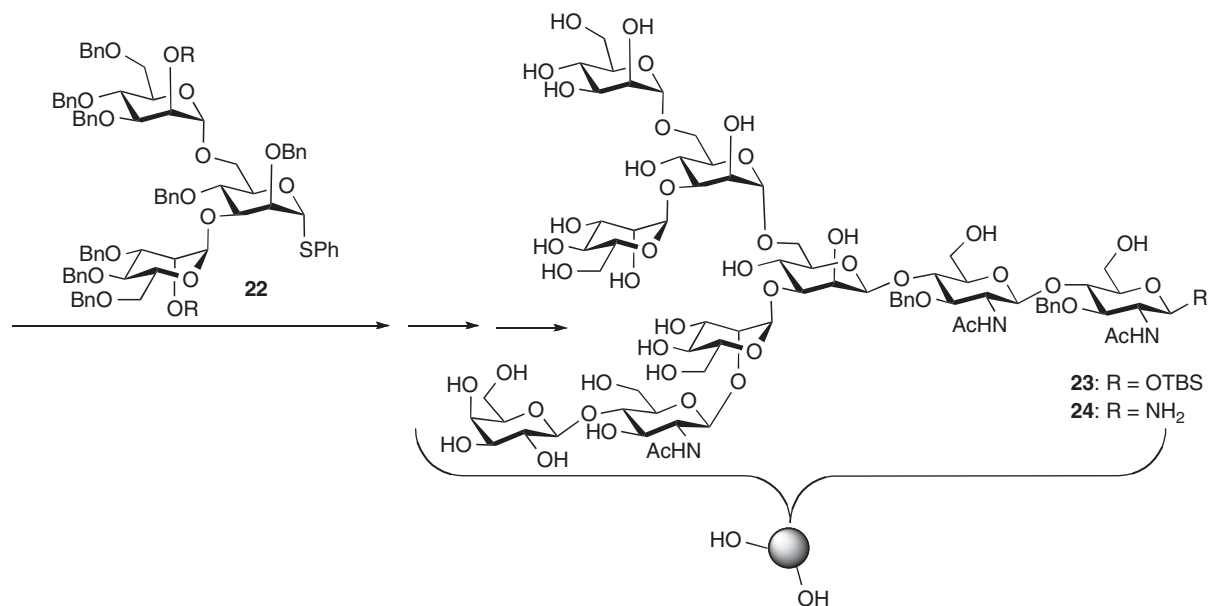
For example, cassette coupling and protecting group manipulations provided trifluoromethylsulfonyl (TF) disaccharide **28**, which was evolved in solution to the trimeric TF cluster **29** (**Scheme 5**). Here, epoxide **25** derived from glycal **25** proved to be a powerful donor in reaction with cassette **26** to afford the β -linked disaccharide **27** (**Scheme 5**). Similarly, 2,6-STF and STn clusters and the clustered LewisY were also prepared, conjugated to keyhole limpet hemocyanin (KLH) or bovine serum albumin (BSA), and their immunological properties were evaluated as anticancer vaccine.^{61,62}

6.03.2.2 Solid-Phase Synthesis

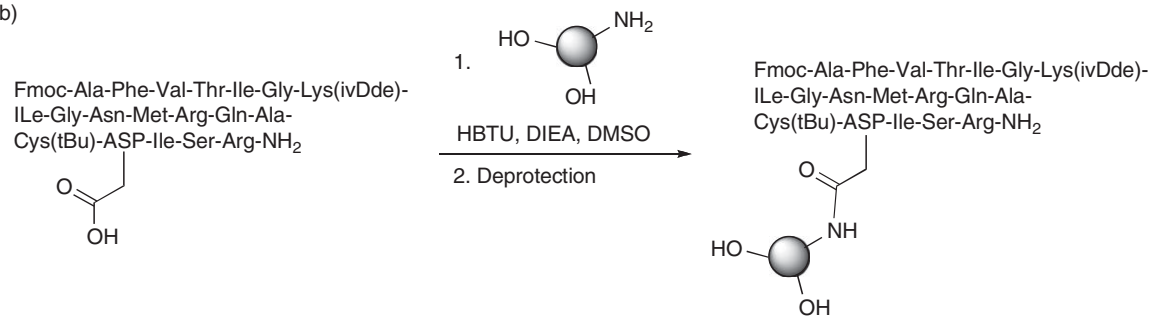
Solid-phase glycopeptide synthesis offers a rapid assembly, without the need of the cumbersome isolation of the intermediates, of the peptide component and the potential access of large glycopeptide fragments in combinatorial formats. However, it requires the synthesis of the preglycosylated amino acid and the use of a large excess of this precious building block to achieve a quantitative coupling. Various synthetic parameters in the SPPS should be taken into consideration when targeting glycopeptide. For example, the linker that anchors the glycopeptide with the solid support should be cleaved under mild conditions that will not harm the glycan



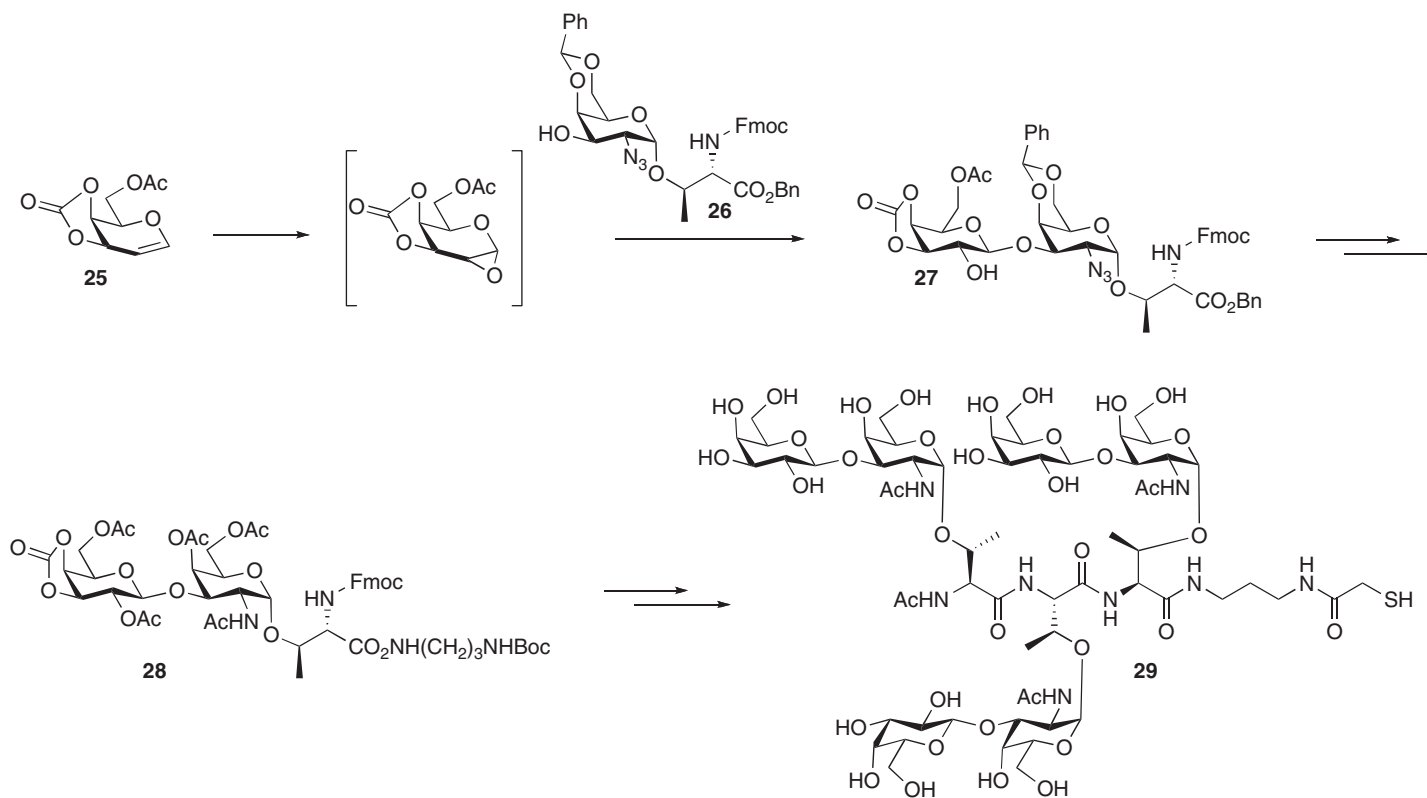
Scheme 4 (continued)



(b)



Scheme 4 Synthesis of hybrid- and high-mannose-type gp120 glycopeptide fragments via solution phase synthesis of the glycan component (a) followed by aspartylation (b).

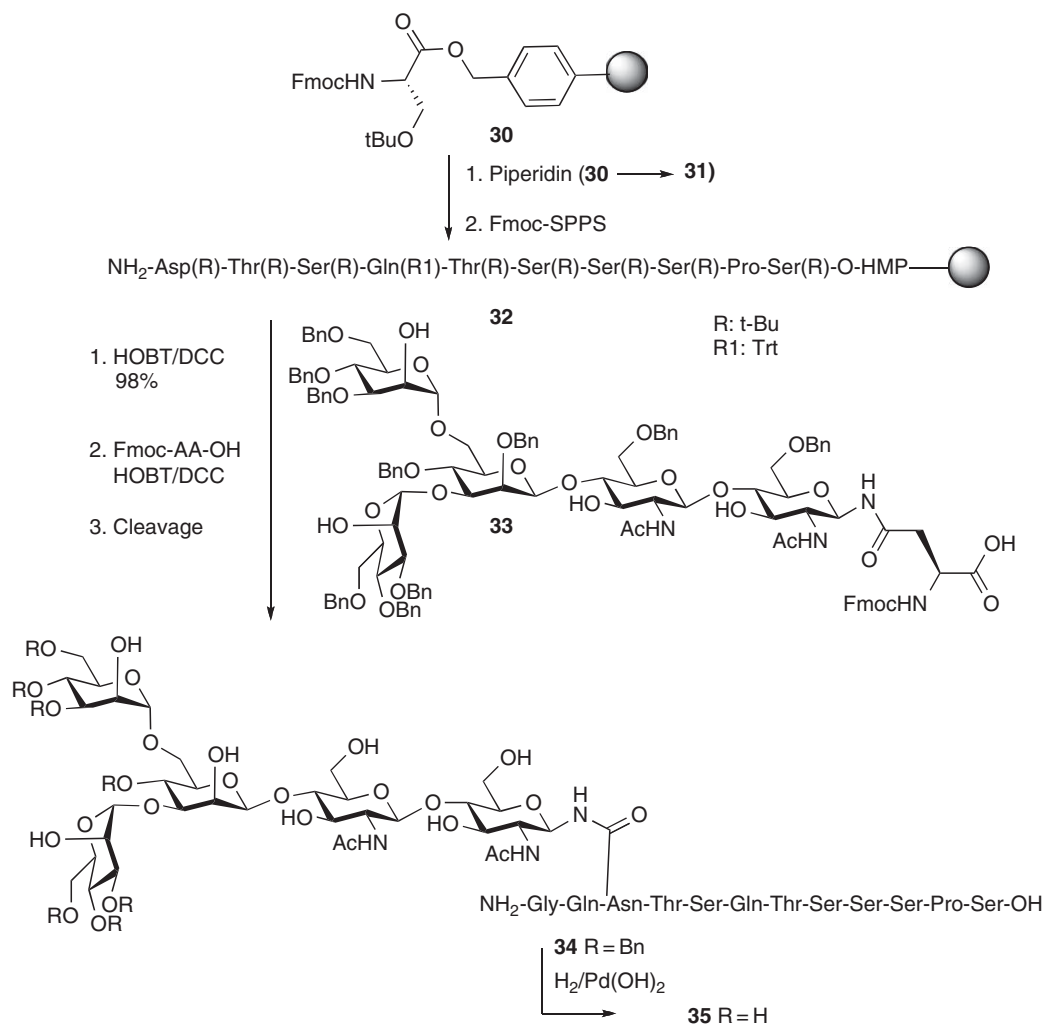


Scheme 5 Synthesis of trimeric TF cluster using the cassette approach.

structure. Thus, linkers that applied in *tert*-butyloxycarbonyl (Boc)-SPPS should not be used while others such as Rink and PAL, which are compatible with the fluoren-9-ylmethyloxycarbonyl (Fmoc)-strategy, are often suitable for this task. As for the solid support, there are no particular requirements besides having a high swelling capacity to allow for an efficient synthesis.

There are several examples in the literature on using SPPS to access a wide range of glycopeptides.^{63–71} The presented studies in this section and the examples that will be shown in the coming sections will testify to the power of solid-phase synthesis in this field. Ogawa and coworkers reported one of the earliest synthetic works in constructing a complex N-linked glycopeptide on solid support, the glycopeptide part **35** of the CD52 antigen (Scheme 6).⁴⁹ The CD52 antigen is a glycoprotein that is expressed on virtually all human lymphocytes. Monoclonal antibodies against this antigen are potent effectors of complement-mediated lysis and have been widely used *in vivo* and *in vitro* for the control of graft versus host disease and for the prevention of bone marrow transplant rejection. Structurally, its peptide chain is extremely short and consists of only 12 amino acid residues with only one N-glycosylation site and no O-glycosylation.

The solid-phase synthesis started on Fmoc-Ser(*t*Bu)-O-HMP-resin **30** following Fmoc-SPPS strategy (Scheme 6). The preglycosylated amino acid, Asn-linked core pentasaccharide with benzyl (Bn) protective groups **33**, was synthesized according to a slightly modified version of a reported procedure and was coupled to the free amine of the growing chain Asp⁴-Ser¹²-resin **32** using *N,N'*-dicyclohexylcarbodiimide/1-hydroxy-1*H*-benzotriazole



Scheme 6 Solid-phase synthesis of CD52 antigen.

(DCC/HOBT) coupling conditions with 2.4 equivalent of **32**. The use of the Bn group as a protecting group for oligosaccharides is due to the highly reactivity of the benzylated sugars in glycosylation reactions and the stability of Bn groups toward basic conditions. This would eliminate the possible side reactions during the elongation of the peptide chain. After completing the solid-phase synthesis the glycopeptide was released from the resin with the simultaneous removal of the amino acid side chain by treating the loaded resin with a 95% aqueous trifluoroacetic acid (TFA) solution containing 2.5% 1,2-ethanedithiol. Final debenzilation by hydrogenolysis using 20% Pd(OH)₂ on charcoal in 50% aqueous ethanol and purification gave the desired product **35** in high yield and purity.

O-linked glycopeptides were also synthesized on solid support, wherein numerous examples can be found in the literature.^{72–77} One such example is the synthesis of the epithelial mucin MUC4, which is the largest among the mucins known so far containing a variable number of tandem repeats of 16 amino acids and comprising 2334–6334 amino acids.⁷⁸ Because of the numerous O-glycans, in particular within the tandem repeat region, it adopts a stretched structure reaching 1.1–2.1 mm beyond the cell membrane. MUC4 is assumed to maintain important functions in cell–cell communication. In tumor cells of different tissues, an abnormal expression of MUC4 was observed, and MUC4 is the only tumor-associated mucin found on pancreatic adenocarcinoma cells.⁷⁹ Thus, MUC4 constitutes an important target molecule for the induction of tumor-selective immune responses.

Kunz and coworkers reported the solid-phase synthesis of glycopeptides derived from the tandem repeat sequence of MUC4 with tumor-associated carbohydrate antigens.⁸⁰

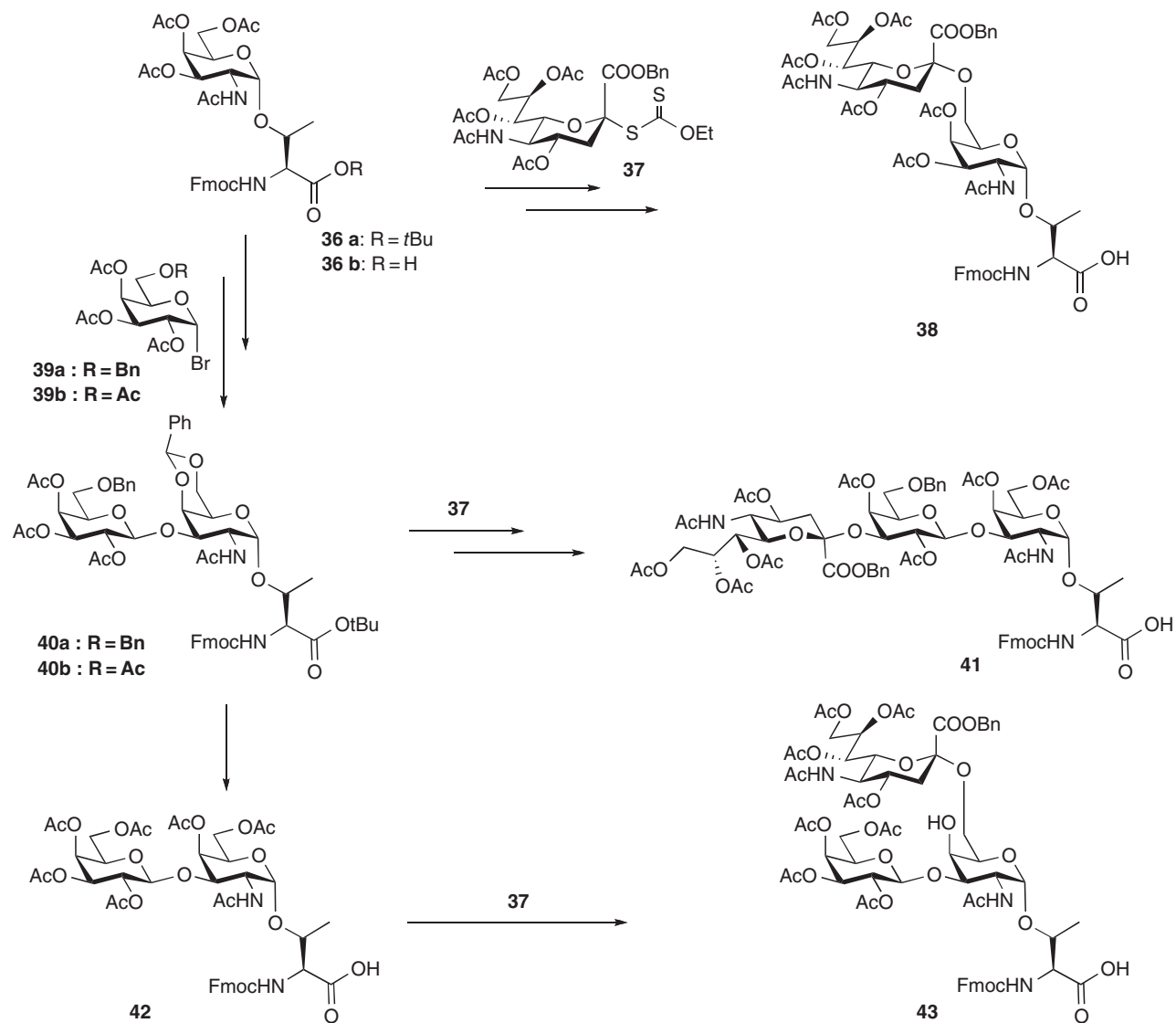
The required glycosyl amino acid building blocks containing the tumour-associated saccharide antigens **36b**, **38**, **41**, **42**, and **43** were synthesized according to a straightforward biomimetic strategy through stepwise extension of the saccharide side chain of Fmoc-protected galactosamine threonine *tert*-butyl ester **36a** (Scheme 7). Sialyl-Tn antigen threonine building block **38** was synthesized through careful transesterification of O-acetylated T_N threonine conjugate **36a** with NaOMe/MeOH at pH < 8.5, followed by a reaction with xanthate **37**, acetylation, and acidolysis of *tert*-butyl ester. Building block **40** was synthesized in high yield by introducing 4,6-benzylidene group to T_N threonine conjugate **36a**, followed by galactosylation with **39**. Removal of the benzylidene and *tert*-butyl protecting groups in **40** furnished the glycosylated amino acid **42**. The latter was transformed to (2, 6)-sialyl-T threonine building block **43** by regioselective sialylation, and acidolysis of the *tert*-butyl ester.

For preparation of compound **41**, precursor **40** was subjected to transesterification with NaOMe/MeOH, regioselective sialylation of the 3'-OH group with **37**, followed by the removal of the 4,6-benzylidene and the *tert*-butyl protecting groups. Over nine MUC4 glycopeptides, wherein one or two tumor-associated carbohydrate antigens, were assembled on solid support using the Fmoc-strategy (Table 1).⁸⁰ This study emphasizes the power of solid-phase synthesis over solution phase in constructing a library of glycopeptides for biological studies.

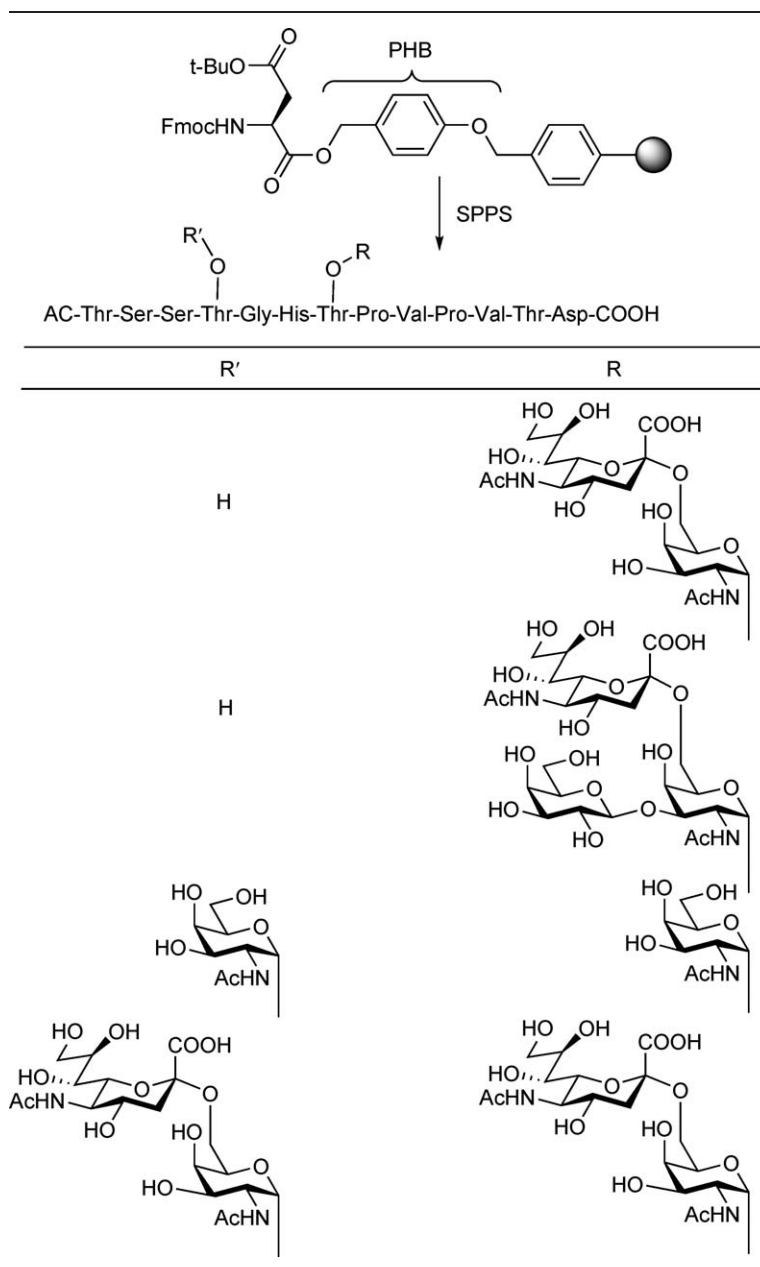
6.03.2.3 Enzymatic Synthesis

At least 16 enzymes are known to be involved in glycoprotein formation in various intracellular sites, that is, endoplasmic reticulum, golgi apparatus, cytosol, and nucleus.² Inspired by nature, chemists have used enzymes as catalysts in the synthesis of glycopeptides and glycoproteins.⁸¹ Enzymes have been proven to be very useful in several aspects in glycopeptide synthesis including, for example, involvement in glycosidic linkage formation, glycan elaboration, and glycopeptide ligation. In this section, we are surveying the role of specific enzymes such as glycosyltransferases and endoglycosidases in glycopeptide synthesis by illustrating a few examples with emphasis on recent developments. In Section 6.03.4, a few other examples related to this topic will also be given.

The enzymatic transfer of individual monosaccharides to a preformed simple O-linked glycan is advantageous over chemical glycosylation, which often requires protecting group manipulation and the formation of the glycosidic bond with the correct stereoselectivity. As a result, the use of glycosyltransferases to perform this task has gained considerable attention in the synthesis of O-linked glycopeptides containing an O-linked sialyl-Lewis-X (SLe^x) tetrasaccharide, a glycopeptide fragment of P-selectin glycoprotein ligand-1 (PSGL-1), and tumor-associated Tn (GalNAc α -Thr/Ser) and T (Gal α 1-3GalNAc α -Thr/Ser) antigens and their sialylated forms. In an effort to construct specific glycoforms of SLe^x of interest to the study of cell–cell adhesion, the O-linked glycopeptide containing β -linked GlcNAc **44** was elaborated sequentially with

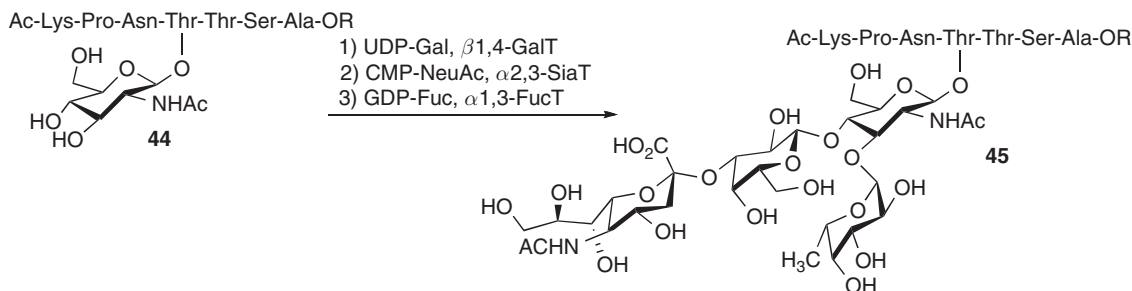


Scheme 7 Synthesis of MUC4 building blocks for SPPS.

Table 1 Examples of synthetic MUC4 glycopeptides

β -1,4-galactosyl-, α -2,3-sialyl-, and α -1,3-fucosyltransferases (FucTs) to yield an SLe^x tetrasaccharide containing glycopeptide **45** (Scheme 8).²⁸ Notably, the enzymatic glycosylation was carried out both in solution and on solid support, highlighting the flexibility of glycosyltransferases in chemoenzymatic synthesis for glycopeptide. Glycopeptides with a spacer between the SLe^x or SLe^x mimetic glycan and peptide have also been prepared using a similar strategy.^{82,83}

The chemoenzymatic synthesis of glycopeptide fragments of PSGL-1 applying glycosyltransferases was also reported. PSGL-1 is a membrane-bound counter-receptor for P-selectin and involved in leukocyte initiation in the inflammatory response. Cummings and colleagues reported the chemoenzymatic synthesis of several PSGL-1 glycopeptide fragments by sequential enzymatic glycosylation.^{84–86} Wong and coworkers synthesized



Scheme 8 Enzyme-assisted synthesis of the glycopeptide bearing the SLe^x tetrasaccharide.

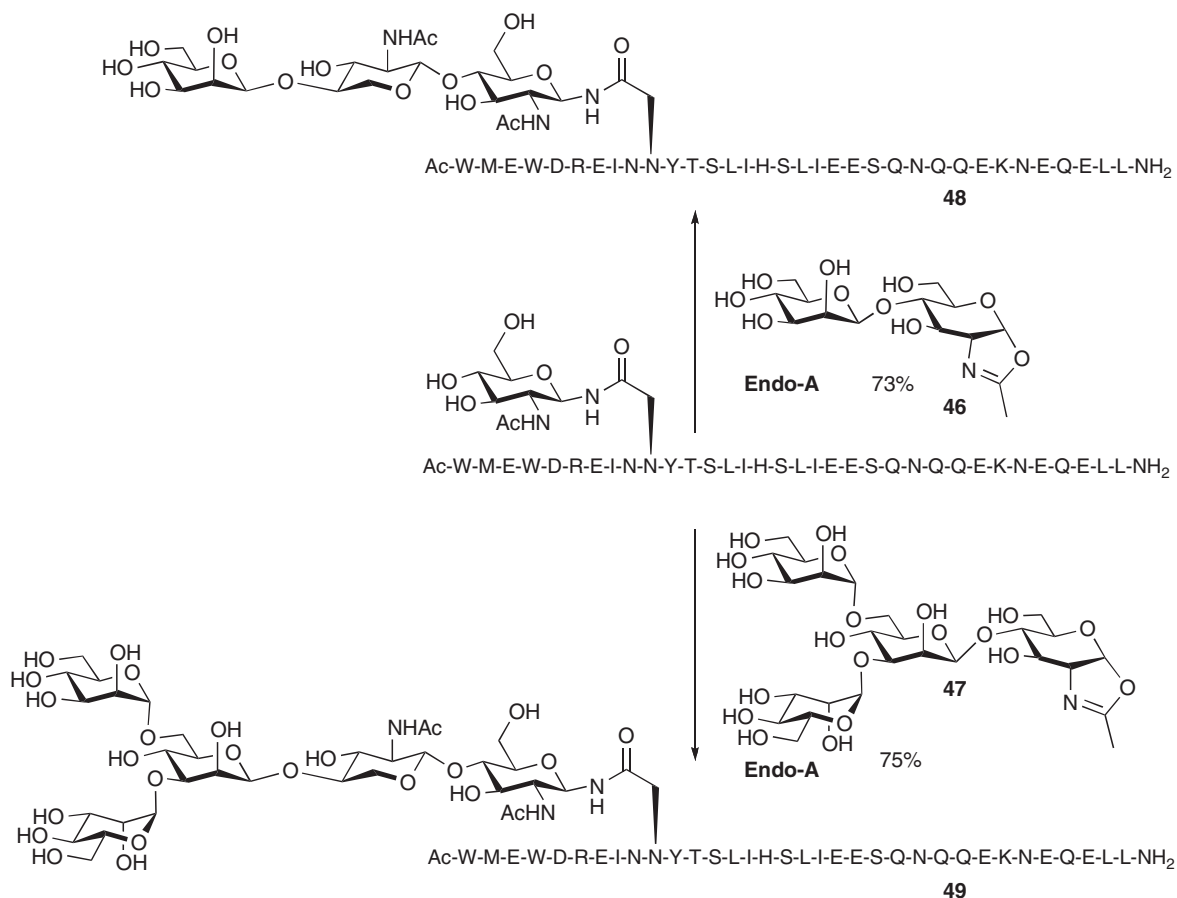
an associated fragment of PSGL-1, in which a synthetic glycopeptide containing a disaccharide was expanded to a pentasaccharide by glycosyltransferases.^{87,88} In glycopeptides containing mucins, glycosyltransferases have also found useful applications as, for example, with the chemoenzymatic syntheses of the α 2-3sialylated T-antigen, including the glycosyl amino acid Neu5Ac α 2-3Gal β 1-3GalNAc α -threonine and a neoglycopeptide, using the sialyltransferase (SiaT) ST3Gal I.^{89,90}

The use of glycosyltransferases in the chemoenzymatic synthesis of N-linked glycopeptides with complex glycans has also been examined. For example, Unverzagt and coworkers reported the chemoenzymatic synthesis of a diantennary complex-type N-linked glycosyl asparagines by means of sequential, one-pot glycan elaboration with β 1,4-galactosyl- and α 2,6-sialyltransferases.^{91,92} Another example can be found in the work of Imperiali and coworkers where the oligosaccharyltransferase, which uses dolichol-linked tetradecasaccharide donor to transfer the saccharide to an asparagine's side chain of a protein, was used for the glycosylation of a tetrapeptide.⁹³

Another group of enzymes that is useful in glycopeptide and glycoprotein synthesis is the endoglycosidases, which are known to cleave the glycosidic bond in an oligosaccharide or glycoconjugate. However, under kinetic conditions, the transglycosylase activity of endoglycosidases is favored over cleavage.⁹⁴ The endo- β -N-acetylglucosaminidases (ENGase) (including Endo-A, Endo-CE, Endo-D, Endo-F, Endo-H, and Endo-M) cleave the *N,N*-diacetylchitobiosyl 1,4-glycosidic linkage at the reducing terminus of the N-linked glycan, leaving one *N*-acetylglucosamine unit on the peptide or protein. For transglycosylase activity, these endoglycosidases require an asparagine-linked glycan donor. In contrast to common glycosyltransferases and exoglycosidases that transfer only monosaccharides, Endo-A and Endo-M can transfer a large intact oligosaccharide to a GlcNAc-peptide acceptor in a single step to form a new glycopeptide. This allows for a highly convergent glycopeptide synthesis without the need of protecting groups. Recently, Wang and coworkers used oligosaccharide oxazolines as donors for constructing *N*-glycopeptide fragments of HIV-1 gp41 and gp120 with Endo-A.⁹⁵ Di- and tetrasaccharide oxazolines, **46** and **47**, were transferred efficiently (yields of 73–82%) to GlcNAc-peptide acceptors with stereo- and regiospecificity to give glycopeptides **48** and **49**, respectively (Scheme 9). This method is based on the assumption that the ENGase-catalyzed reaction proceeds via a mechanism involving an oxazolinium ion intermediate.^{96,97}

6.03.2.4 Chemical Ligation

In general, peptide and glycopeptide synthesis, whether carried out on solid support or in solution, is limited to short peptide fragments. Despite being advantageous over solution phase in accessing longer peptides, SPPS is still limited to peptide chains of ~40–50 amino acids. This corresponds to a small number of proteins and glycoproteins. As a result, ligation methods have been sought to produce the desired sequence from smaller peptide segments.⁹⁸ The breakthrough in chemical ligation occurred in 1994 when Kent and coworkers introduced the native chemical ligation (NCL) method.⁹⁹ Since its development, NCL has been used to construct over 300 biologically important proteins which include pharmaceutical proteins, proteins made from D-amino acids, and proteins with posttranslational modification, for example, glycosylation and phosphorylation.⁹⁸ The ingenuity of this approach is the lack of an external scaffold to promote amide bond

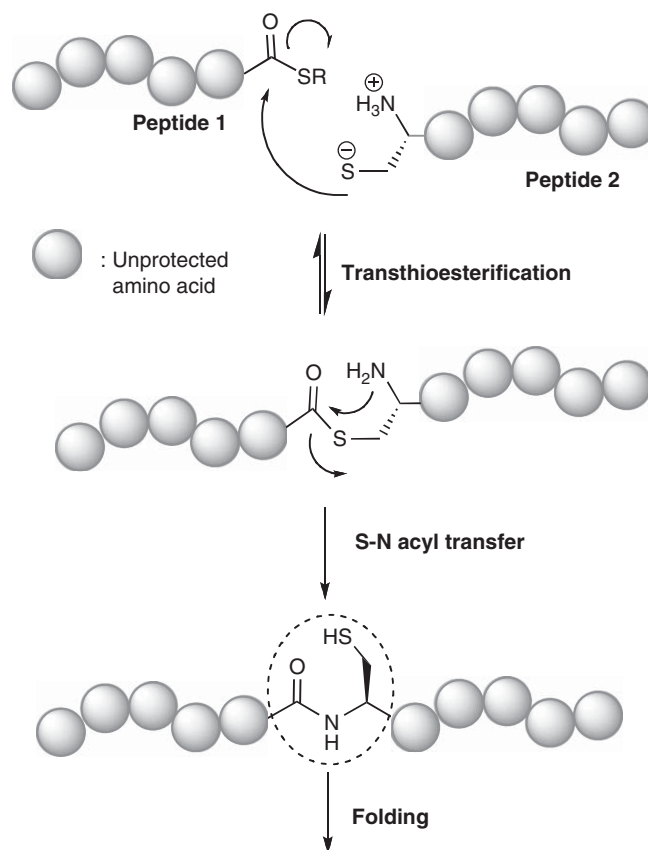


Scheme 9 Enzymatic transfer of di- and tetrasaccharide oxazolines to GlcNAc-peptide acceptors.

formation at the ligation site. The method relies on the thiol side chain of a cysteine residue to bring the N-terminal amine of the first peptide to close proximity of the thioester C-terminal peptide. Once the intermediate is formed, it spontaneously rearranges through a five-membered ring intermediate to form the native amide bond (**Scheme 10**).¹⁰⁰

The impressive successes that many laboratories have had with the use of NCL to construct the protein of interest have inspired several researchers to apply this technology to the synthesis of glycopeptides and glycoproteins. In this section and in the following one, we are focusing on applying chemical ligation, in reticular NCL, in glycopeptides and glycoproteins, presenting the successes and the challenges in the field. The challenges of applying NCL in the glycopeptide synthesis reside in two main issues: first the synthesis of the thioester peptides bearing glycan on one of the amino side chains and second the extension of NCL beyond the cysteine residue.

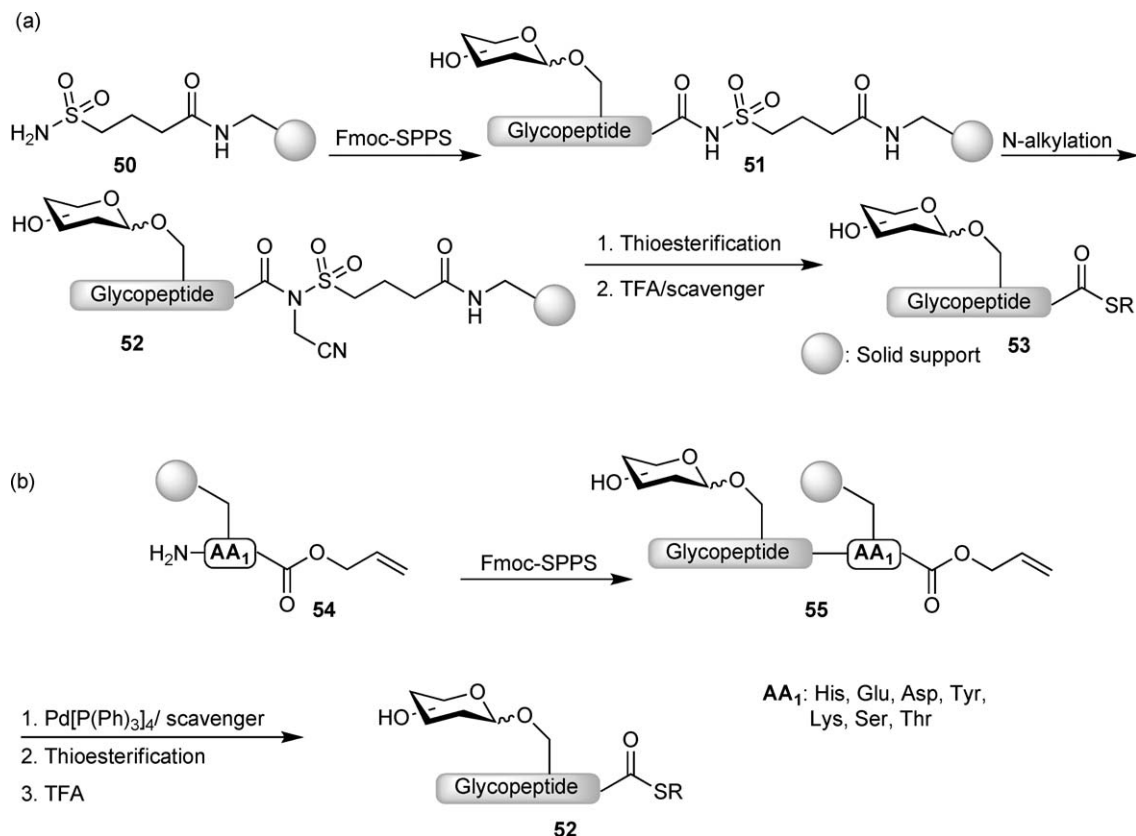
Efficient synthesis of peptide thioester requires the protocol of Boc-SPPS involving acidic procedures such as the repeated exposure of the resin-peptide to TFA and final cleavage/deprotection with hydrofluoric acid (HF).¹⁰¹ These conditions could severely damage the integrity of the glycan structure and therefore exclude the application of Boc-chemistry to glycopeptide synthesis. As a result, glycopeptides are synthesized primarily using Fmoc-based chemistry which requires only a single cleavage/deprotection step with TFA. Unfortunately, the recurring exposure of the resin-bound peptide to base during Fmoc deprotection at each step of the synthesis results in the cleavage of the thioester. With this in mind, Bertozzi and coworkers developed a new method, which relies on using the alkanesulfonamide 'safety-catch' **50** to prepare a thioester peptide **52** via Fmoc-SPPS protocol (**Scheme 11(a)**).¹⁰² Following this development, various methods have



Scheme 10 Proposed mechanism for NCL.

been introduced to construct thioester peptides.¹⁰³ For example, Barany *et al.* developed the backbone amide linker (BAL) strategy and orthogonal allyl protection of the C-terminal carboxylic group to generate thioester after chain assembly.¹⁰⁴ Hilvert and coworkers prepared peptide thioesters using Lewis acid (Me₂AlCl)-catalyzed thiolysis of ester resins.¹⁰⁵ Wong and coworkers have recently expanded the scope of the side-chain anchoring strategy to the synthesis of glycopeptide thioester.¹⁰⁶ The method relies on the side-chain immobilization of a variety of Fmoc-amino acids, protected at their C-termini, on solid supports.¹⁰⁷ Once the C-terminal amino acid is anchored to give **54**, the peptide **55** is constructed using SPPS according to the Fmoc protocol. After unmasking the C-terminal carboxylate, either thiols or amino acid thioesters were coupled to afford, after cleavage, glycopeptide thioester **52** in high yield (**Scheme 11(b)**).

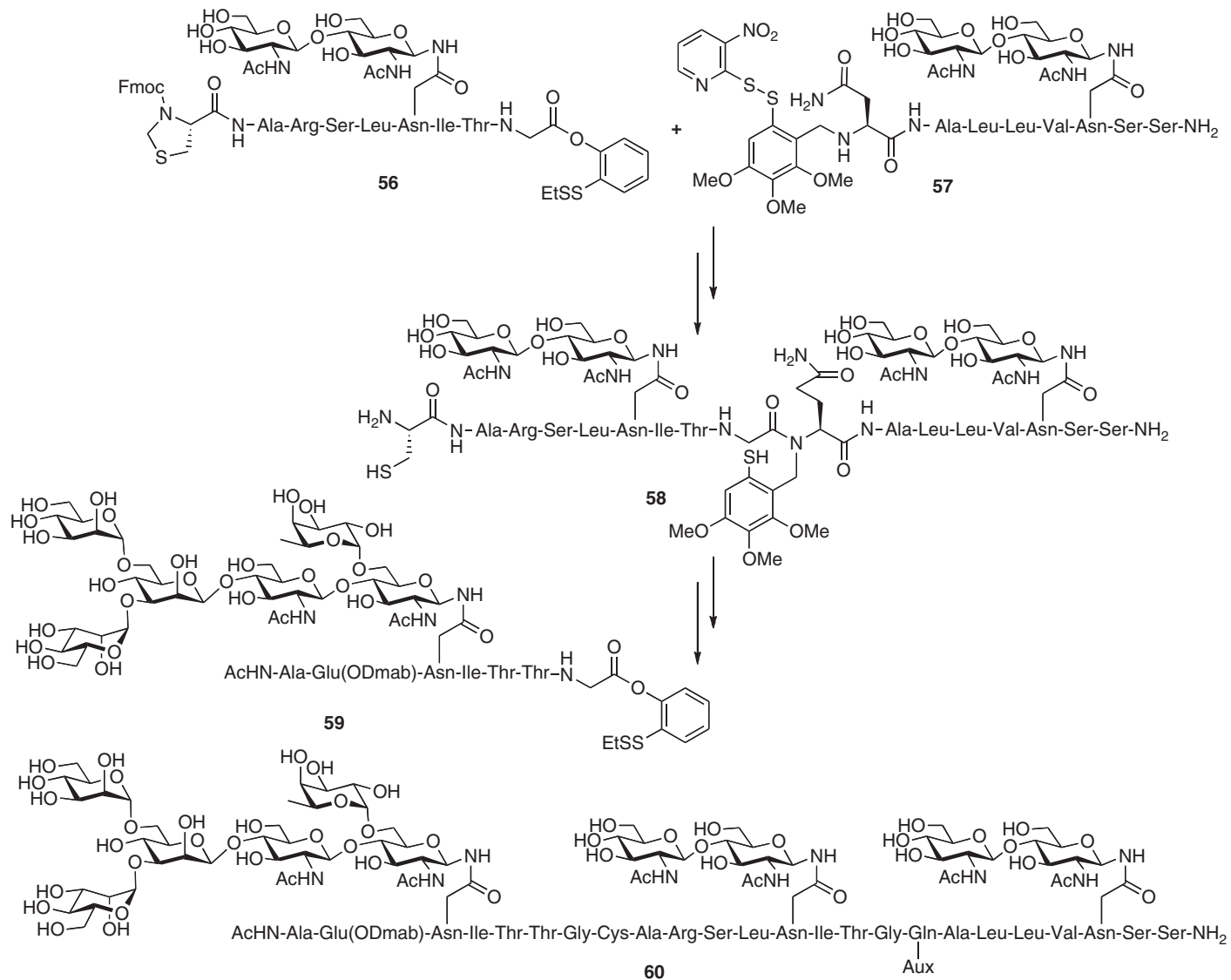
The second challenge that chemists face when using NCL is the limitation of the ligation junction to the cysteine residue. Although a number of proteins contain native cysteine residues in synthetically useful positions (ideally at every 30–40 amino acids of the protein sequence), the majority of proteins do not have the cysteine residue. Moreover, in many cases the cysteine location within the peptide sequence is not beneficial for the synthesis. To overcome this obstacle several strategies were developed.^{108–112} One such approach is the use of removable auxiliaries to mimic the cysteine function at the ligation site.^{113–115} Peptides with N-linked auxiliaries such as 1-phenyl-ethanethiol, 2-mercaptoethyl, and 2-mercapobenzyl have been investigated and successfully applied to the synthesis of complex glycopeptides.¹¹⁶ The Danishefsky group has extensively used the 2-mercaptobenzyle auxiliary to synthesize glycopeptide segments toward the total synthesis of the EPO glycoprotein.^{117,118} **Scheme 12** summarizes part of this work where both cysteine ligation and free cysteine ligation, that is, use of an auxiliary, were applied to prepare the desired glycopeptides. The two segments **56** and **57** were conjugated by first applying auxiliary-mediated ligation step at Gln–Gly junction to give **58** which was ligated with **59**, after removing the protecting group on the cysteine residue to allow for



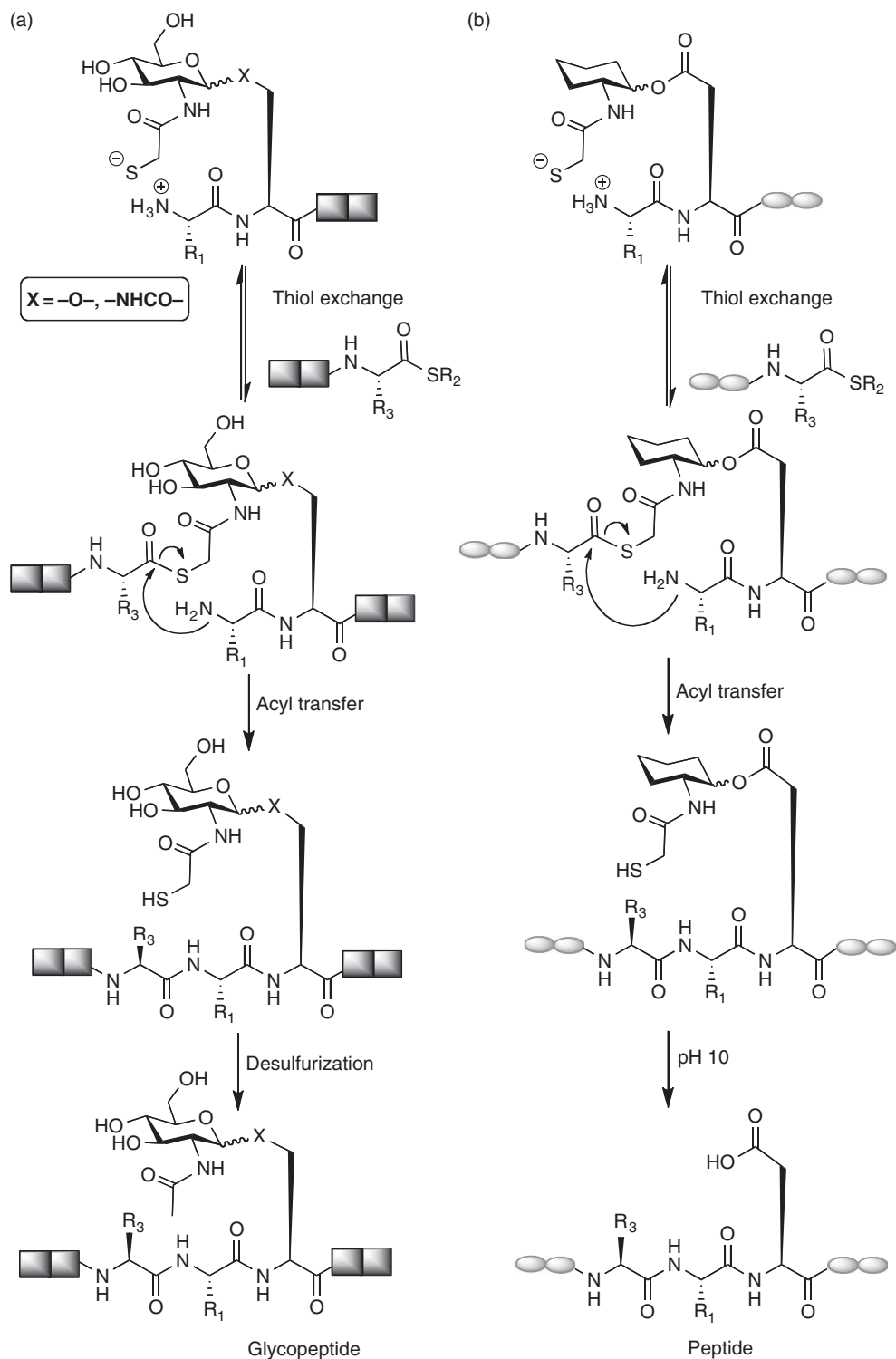
Scheme 11 (a) Synthesis of peptide thioester by applying the alkanesulfonamide 'safety-catch'. (b) Synthesis of peptide thioester by applying side-chain anchoring strategy.

an NCL, to form the glycosylated peptide **60**. The auxiliary can be removed at this stage using methyl *p*-nitrobenzene sulfonate to methylate the free thiol followed by exposure of the peptide-auxiliary to 9% TFA (**Scheme 12**).

In the above-described auxiliaries, the attachment of the auxiliary to the N-terminus peptide generates a secondary amine. Due to the increased steric hindrance on the amine that is involved in the acyl transfer, only ligation at Gly–Gly or Gly–AA (AA represents mainly unhindered amino acids) junction delivered the desired products, limiting their scope in various systems.¹¹⁶ In an effort to overcome the above limitation we have recently developed a new ligation approach named sugar-assisted ligation (SAL) for the synthesis of O- and N-linked glycopeptides (**Scheme 13(a)**).^{119–122} Our strategy uses a similar principle of NCL and an auxiliary-assisted ligation in terms of the entropic activation/proximity effect. However, SAL uses a thiol auxiliary placed on sugar rather than on the side chain of the amino acid, for example, cysteine. In SAL, instead of anchoring an auxiliary to the N-terminus of the peptide, we take advantage of the already existing sugar by slightly modifying its acetamido group on the C2 position to facilitate the ligation (**Scheme 13(a)**). Molecular dynamic studies revealed that the sugar moiety imposes its restricted conformation on the reacting groups of the thioester intermediate, thus acting as a rigid scaffold to facilitate the S→N rearrangement via 14-(O-linked glycopeptide) or 15-membered ring intermediate (N-linked glycopeptide).¹²³ Notably, the ring size of the intermediate in SAL is similar to the one in the dibenzofuran template reported by Kemp and coworkers.¹²⁴ Despite being much larger than the size of the intermediate produced by the 1-phenyl-2-mercaptoethyl, 2-mercapobenzyl, and ethanethiol auxiliaries (six-membered ring),^{113–115} the efficiency of SAL in terms of rate and the sequence tolerance at the ligation junction is superior. In the previous auxiliaries, the N-terminal nucleophile is secondary; however, in our strategy the nucleophile is a primary amine. This leads to a less hindered transition state, which may explain the tolerance of the ligation junction to a variety of amino acids. Another important observation in SAL is the effect of the



Scheme 12 The use of an auxiliary in glycopeptide ligation.



basicity of some of the side chain amino acids on the ligation rate. For example, the side chain of asparagine is as hindered as the aspartic acid; however, the rate of the ligation is three times slower. This can be explained by the ability of the carboxylate side chain, which serves as a general base in the reaction pathway.

Upon completion of the ligation reaction, the thiol handle can be removed by desulfurization to regenerate the unmodified sugar. Alternatively, the modified sugar can be further elaborated with glycosyltransferases to extend its glycan structure, or alkylate it by labeling reagents such as fluorescent dyes. Interestingly, in the three most prevalent occurring glycopeptides, N-linked, β -O-linked, and α -O-linked, the sugar that is directly attached to the peptide is equipped with the *N*-acetyl moiety at the C2 position. This would allow introducing the thiol handle, regardless of the glycoform type, to assist the ligation. Alternatively, the thiol handle can be attached to the sugar oxygen groups, allowing for the ligation of glycopeptides containing sugars without a nitrogen group and the basic removal of the auxiliary rather than the use of desulfurization.¹²⁵ More recently, inspired by the role of sugar in SAL, our group used a removable sugar mimic, which can be attached via ester bond to serine, threonine, aspartic acid, and glutamic acid to facilitate ligation followed by a saponification step for auxiliary removal (**Scheme 13(b)**).¹²⁶

In an effort to extend NCL beyond the cysteine residue, the traceless Staudinger ligation was developed.^{127–129} In this strategy, an azide at the N-terminus of peptide I reacts with a phosphinothioester located at the C-terminal of peptide II to form an iminophosphorane, which then undergoes an intramolecular S-N acyl transfer to form an amido phosphonium salt intermediate. Hydrolysis of the amidophosphonium salt produces an amide without any residual atoms from the phosphine prosthetic group. Recently, Wong and coworkers extended this method to the synthesis of short glycopeptides by using subtilisin to catalyze selective N-azidation of an unprotected polypeptide.¹³⁰ In principle, this strategy could be used to functionalize large expressed peptides with the azide function so it can be engaged in the traceless Staudinger ligation.

6.03.3 Synthesis of Neoglycopeptide

Driven by their ease of preparation and higher stability in comparison to glycopeptides, neoglycopeptides have been a synthetic target for many research groups.³⁰ Various glycosidic bond surrogates that have already been developed were recently extended to the synthesis of neoglycoproteins. This would include, for example, *C*- and *S*-glycosides as well as oxime and triazole bonds (**Figure 3**). In this section we will be reviewing the synthesis of such glycopeptide mimetics and the extension of these methods to access homogenous glycoproteins with comparable biological activities to their natural counterparts.

6.03.3.1 C-Linked Glycopeptides

Carbohydrate derivatives wherein the glycosidic oxygen is substituted with a carbon atom called *C*-glycosides. The chemistry of *C*-glycoside synthesis is well established and was covered by several reviews.¹³¹ A variety of *C*-glycosyl amino acids have been described; yet, the incorporation of such building blocks in the context of glycopeptides and

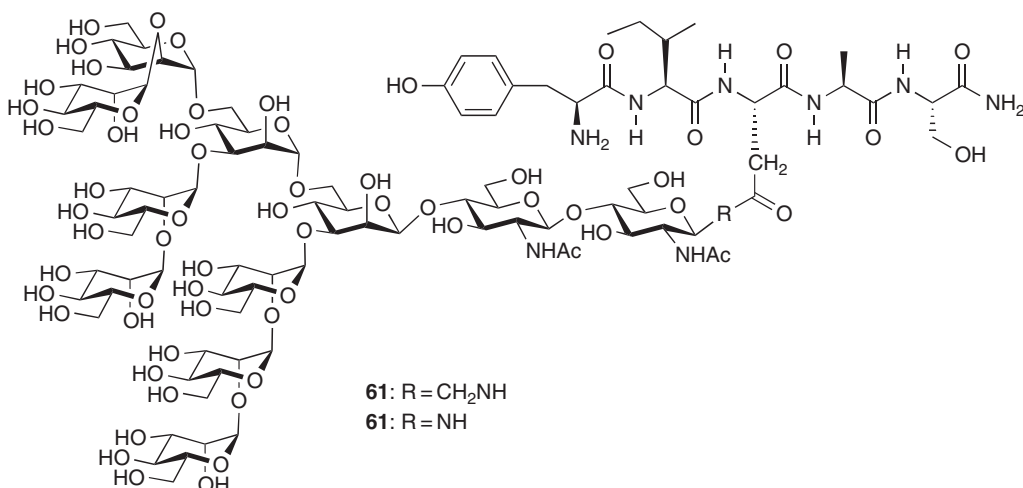
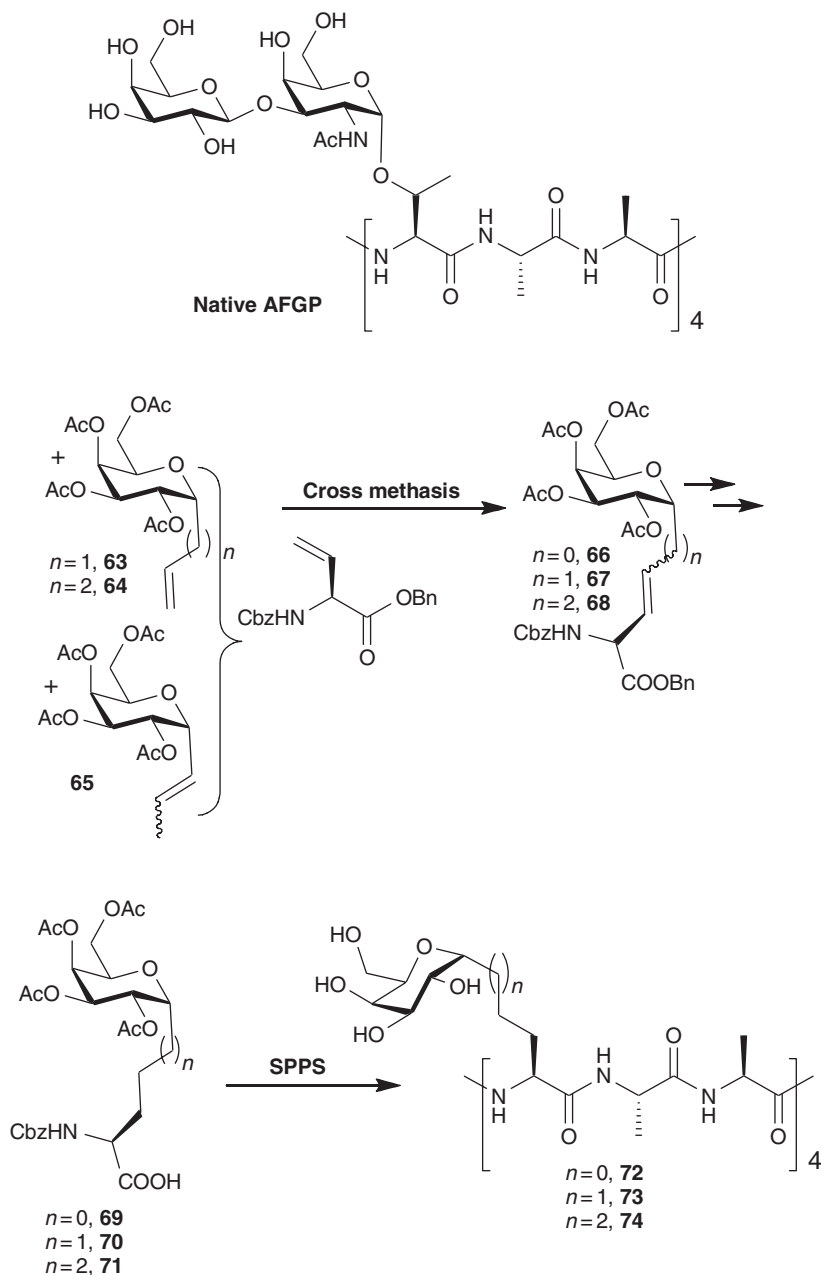


Figure 3 High-mannose C- and N-linked glycopeptides.

glycoproteins is still in early stages of development.^{132–137} Lee and coworkers applied a chemical and enzymatic approach to synthesize high-mannose C- and N-linked glycopeptides (**61**, **62**, **Figure 3**).¹³⁸ The synthesis included the chemical synthesis of GlcNAc containing peptides followed by an enzymatic transfer of the Man₉GlcNAc onto the core GlcNAc. Notably, compound **61** was found to resist enzymatic cleavage at the asparagine residue due to the insertion of the methylene group and inhibited various glycoamidases of plants, bacteria, and animals. The synthesis of a different class of C-linked glycopeptide containing C-glycosyl tyrosines was also reported.¹³⁹

Another example related to the synthesis of the C-linked glycopeptide can be found in the recent studies on antifreeze glycoproteins (AFGPs) done by Ben and coworkers. AFGPs, a subclass of biological antifreezes, are inhibiting the growth of ice and protecting living organisms in subzero environments (**Scheme 14**).¹⁴⁰ As a



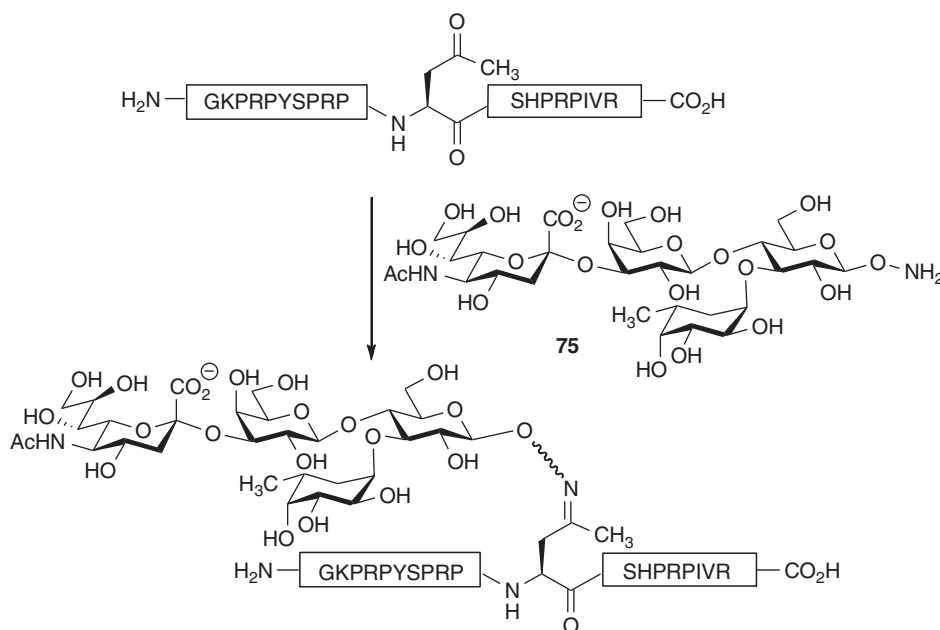
Scheme 14 Synthesis of C-linked AFGP.

result, these compounds have many potential uses in medical and industrial applications. Yet, the development of such a system, which takes advantage of the AFGPs properties, is hampered by their limited bioavailability and the inherent instability of the C–O glycosidic bond. Hence, the substitution of the C–O by the more stable C–C bond holds great potential in stabilizing these glycopeptides. In this regard, cross metathesis was used to prepare C-linked building blocks starting from C-allylated galactose 63–65 to give C-linked amino acids 69–71. SPPS based on Fmoc strategy was then used to incorporate these amino acids to generate C-linked AFGP analogues 72–74. Neoglycopeptide 72 was found to be the most potent recrystallization inhibitor resembling the native AFGP. Recent studies by the same group have shown that the configuration of the carbohydrate moiety in C-linked AFGP analogues is extremely important and modulates recrystallization inhibition activity.¹⁴¹

6.03.3.2 Oxime-Linked Glycopeptides

While C–C bond construction is considered to be a challenge, particularly in systems like glycopeptides, the oxime bond formation falls into the category of chemoselective ligation reaction, where no protecting groups are needed. The method originated from peptide ligation wherein various ligation chemistries were developed, one of which is the oxime ligation.⁹⁸ Here, the aminoxy group reacts in high selectivity with aldehyde and ketones in an aqueous environment, ideally at pH 4.5. Recently, Dawson and coworkers reported that aniline is capable of catalyzing this reaction at physiological conditions.¹⁴² In neoglycopeptide synthesis, a peptide bearing ketone or aldehyde on one of the amino acid side chains reacts, under aqueous conditions, with carbohydrate functionalized with aminoxy group to form an oxime-linked glycopeptide. For example, by applying chemical and established enzymatic methods the Bertozzi group prepared very complex oligosaccharides starting with simple aminoxy sugars. The group then constructed neoglycopeptides containing motifs found in naturally occurring N- and O-linked glycopeptides by coupling the nucleophilic sugars 75 to a synthetic peptide fashioned with a ketone group (Scheme 15).¹⁴³

Elaboration of the glycan-peptide to create a more complex glycopeptide using oxime chemistry was also reported.¹⁴⁴ This method is an alternative to replacing the glycosidic bond where in some cases the saccharide that is most proximal to the peptide backbone can profoundly alter the local peptide structure.^{145,146} Thus, creating unnatural linkage at this position may dramatically influence the glycopeptide structure. On the other



Scheme 15 Oxime ligation for the synthesis of complex neoglycopeptide.

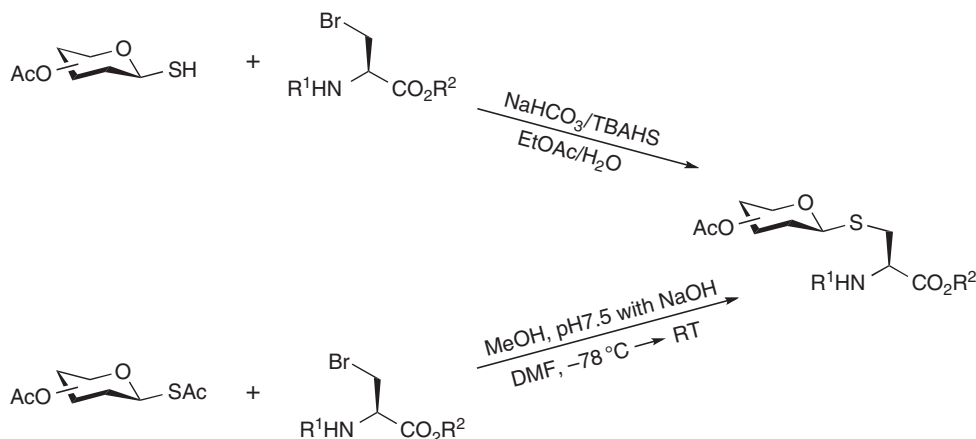
hand, previous studies have shown that the replacement of the glycosidic bond with an oxime linkage has no detrimental effect on function, even in a molecule whose activity is normally dependent on glycosylation. For example, the oxime-linked analogue of drosocin was found to be comparable to the native molecule in bacteriostatic activity.¹⁴⁷ It must be noted that the oxime-based chemoselective ligation strategy has an advantage over several other traditional methods for neoglycoprotein synthesis in that they are site-selective. Moreover, these ligation reactions are orthogonal to thiol alkylations, allowing the conjugation of two different oligosaccharides to ketone and thiol groups found within the same peptide sequence.¹⁴⁸

6.03.3.3 S-Linked Glycopeptides

Replacement of the glycosidic bond with sulfur results in the corresponding S-linked glycopeptide. This family is known to have better chemical stability and to be more resistant to glycosidases.¹⁴⁹ Moreover, members of the closely related S-linked oligosaccharides, which were used as enzyme inhibitors,¹⁵⁰ are suggested to be better immunogens than their native counterparts. As with the C-linked glycopeptides, C–S bond shows solution conformations and biological activities similar to the native structure.¹⁵¹

The synthesis of S-linked glycopeptides can be achieved mainly via two general approaches. First, by using anomeric thiolate nucleophile in reaction with an amino acid side chain decorated with a leaving group to synthesize the S-linked glycosyl amino acids which can be incorporated in SPPS. Second, by using glycosyl iodoacetamide to react with the cysteine side chain, thus generating N-linked glycopeptide mimics. In the first approach, the major side reaction is β -elimination to generate the dihydroalanine followed by Michael addition resulting in a diastereomeric mixture at the α -carbon. To overcome this issue, various procedures have emerged such as the use of the two-phase system for the alkylation reaction containing phase transfer catalyst or the use of one-pot reaction to generate the thiolate anion with NaOH in anhydrous methanol (Scheme 16).^{152,153} Using the latter approach several S-linked glycosyl amino acids were prepared in high yields, which were included in the SPPS scheme to generate the tyrocidine analogue.¹⁵⁴ Recently, Davis and coworkers reported a new method for the synthesis of S-linked glycosyl amino acid, starting from disulfide-linked glycosyl amino acids, by a desulfurization reaction using P(NMe)₂.¹⁵⁵ This enables the conversion of readily synthesized disulfide-linked glycosyl amino acids, glycopeptides, and glycoproteins into the corresponding thioether-linked glycoconjugates. Crich and coworkers utilized the allylic selenosulfide rearrangement, as a method for chemical ligation to thiols, for the synthesis of S-linked glycopeptides.¹⁵⁶

Glycosyl iodoacetamide found useful applications in the synthesis of N-linked glycoprotein mimics wherein the cysteine residue in peptide reacts selectively in solution with the modified glycan.¹⁵⁷ Despite its use in the synthesis of the glycosylated proteins,¹⁵⁸ this method is limited when more than one cysteine residue exists in the protein target.¹⁵⁹ However, through the use of chemical synthesis one may differentiate between two or



Scheme 16 Reported methods for two-step synthesis of S-linked and one-pot approach.

more cysteine residues. Working toward this goal, Flitsch and coworkers reported solid-phase synthesis of thioether-linked glycopeptide mimics for an application to glycoprotein semisynthesis.¹⁶⁰

6.03.3.4 Triazole-Linked Glycopeptides

Triazole-linked glycopeptides are relatively a new class of neoglycopeptides that incorporate the triazole unit via click chemistry by applying the copper (I)-catalyzed 1,2,3-triazole. Click chemistry has emerged as a new strategy for the rapid and efficient assembly of molecules with diverse functionalities.¹⁶¹ Among these reactions is the copper (I)-catalyzed 1,2,3-triazole synthesis.^{162,163} It guarantees a reliable synthesis of 1,4-disubstituted 1,2,3-triazole compounds in high yield, regioselectivity, and purity. Moreover, the reaction works in aqueous media and tolerates virtually all functional groups without protection. The Cu(I)-catalyzed 1,3-dipolar cycloaddition of azides and alkynes have found many applications in carbohydrate chemistry, including the preparation of simple glycoside and oligosaccharide mimetics, glycomacrocycles, glycopeptides, glycoclusters, and carbohydrate arrays.¹⁶⁴

In the neoglycopeptide field, the application of 1,2,3-triazole formation reaction to connect the glycan to the peptidic component is witnessing an increased interest. For example, the Danishefsky group prepared potential carbohydrate-based anticancer vaccines by using cycloaddition chemistry.¹⁶⁵ The key strategy involves decorating the oligosaccharide with an azide and the polypeptide with pendant alkynyl functionality. The two components were then joined through 1,3-dipolar cycloaddition. The method included oligosaccharide constructs with several azide linkages, which led to simultaneous cycloaddition to peptide-based acetylenes (Figure 4).¹⁶⁵ Macmillan and coworkers prepared a triazole-linked glycopeptide, using the 1,3-dipolar cycloaddition, which was extended via NCL with the glycopeptide thioester derived from EPO.¹⁶⁶

Recently, Davis and workers published one of the most impressive studies on neoglycopeptides/neoglycoproteins synthesis by applying click chemistry.¹⁶⁷ In an effort to expand the diversity of chemical protein modification and to allow posttranslational mimicry, the group took on the challenge of the attachment of multiple modifications to bacterially expressed protein scaffolds. Toward this goal, a strategy of a dual combination of selective synthetic methodology and site-directed mutagenesis was adopted to allow for modification at multiple, predetermined sites in peptide sequences. Thus, the triplet codons for cysteine and methionine were chosen to code for the different chemical tags. The methionine was replaced by azidohomocysteine as to permit Huisgen cycloaddition, while the cysteine residue was used to conjugate the glycomethanethiosulphonates for the second modification (Scheme 17). This approach was used to perform differential chemical posttranslational modifications to reconstitute function in a mimic of human protein known to possess differential posttranslational modifications important to its function, PSGL-1.¹⁶⁷

6.03.4 Glycoprotein Synthesis

Unlike protein synthesis, which is carried out through recombinant DNA-based expression to produce a homogenous product, glycoprotein biosynthesis is not under complete genetic control. Instead, glycoprotein is the product of a secondary metabolism leading to a large heterogeneity of glycoforms. As a result, chemical approaches have been sought to overcome these limitations. Glycoprotein synthesis stands as one of the most challenging targets which organic chemists could encounter. The difficulties could arise not only because of their large size (a typical protein molecule is ~30 kDa in size and consists of two ~15 kDa domains), but also because of the complexity of oligosaccharide chemistry, which often requires extensive protecting group manipulations and difficult glycosyl coupling conditions to generate the correct regio- and stereochemistry of the glycosidic bond. Moreover, the convergent approach that is usually adopted to assemble the entire fragments toward the final product is challenging and is often faced with several obstacles. In previous sections, we surveyed the chemical and enzymatic methods for glycopeptide synthesis; in the following section we are presenting the various studies that have adopted these technologies to assemble the large glycopeptides corresponding to folded glycoproteins.

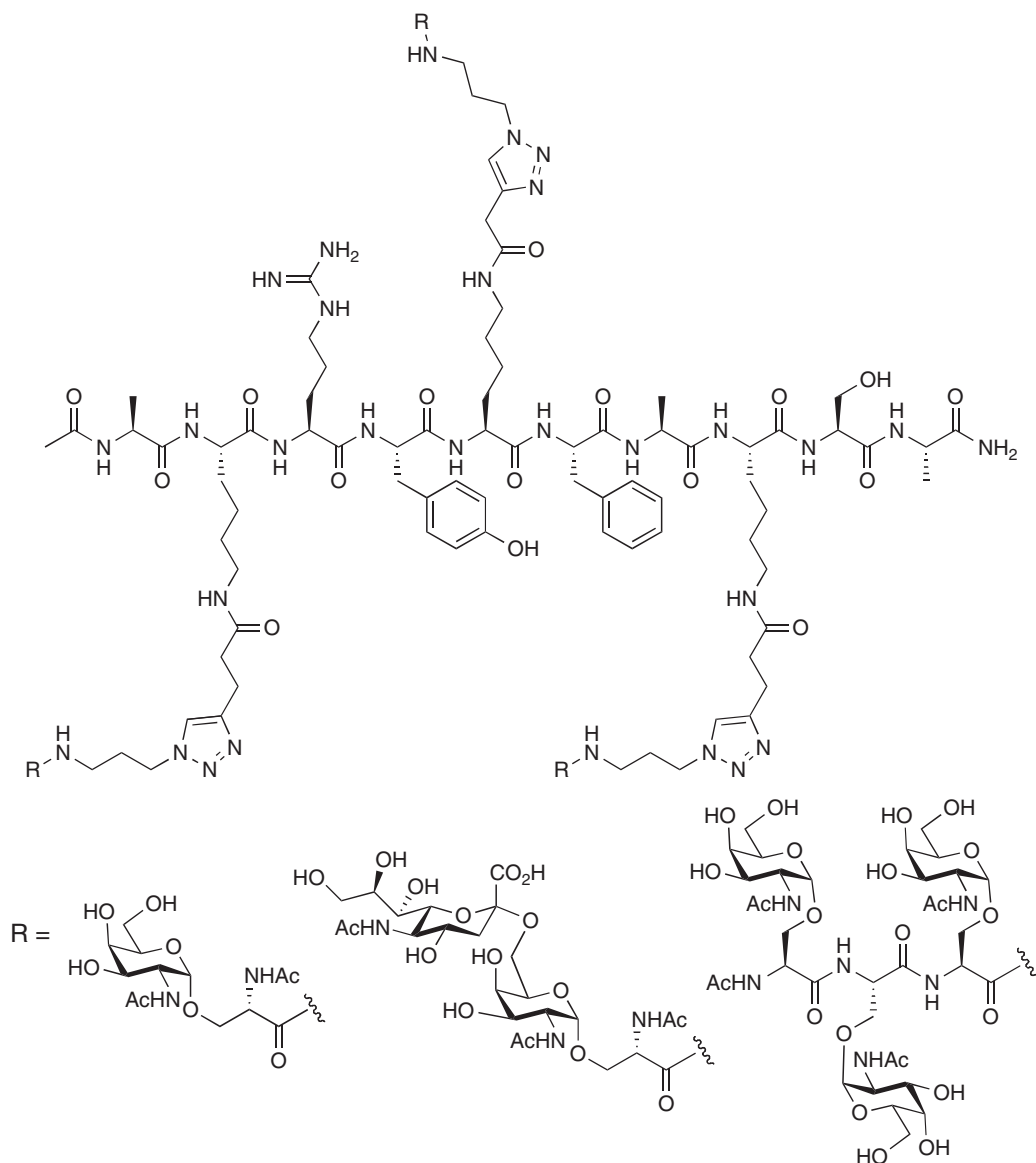
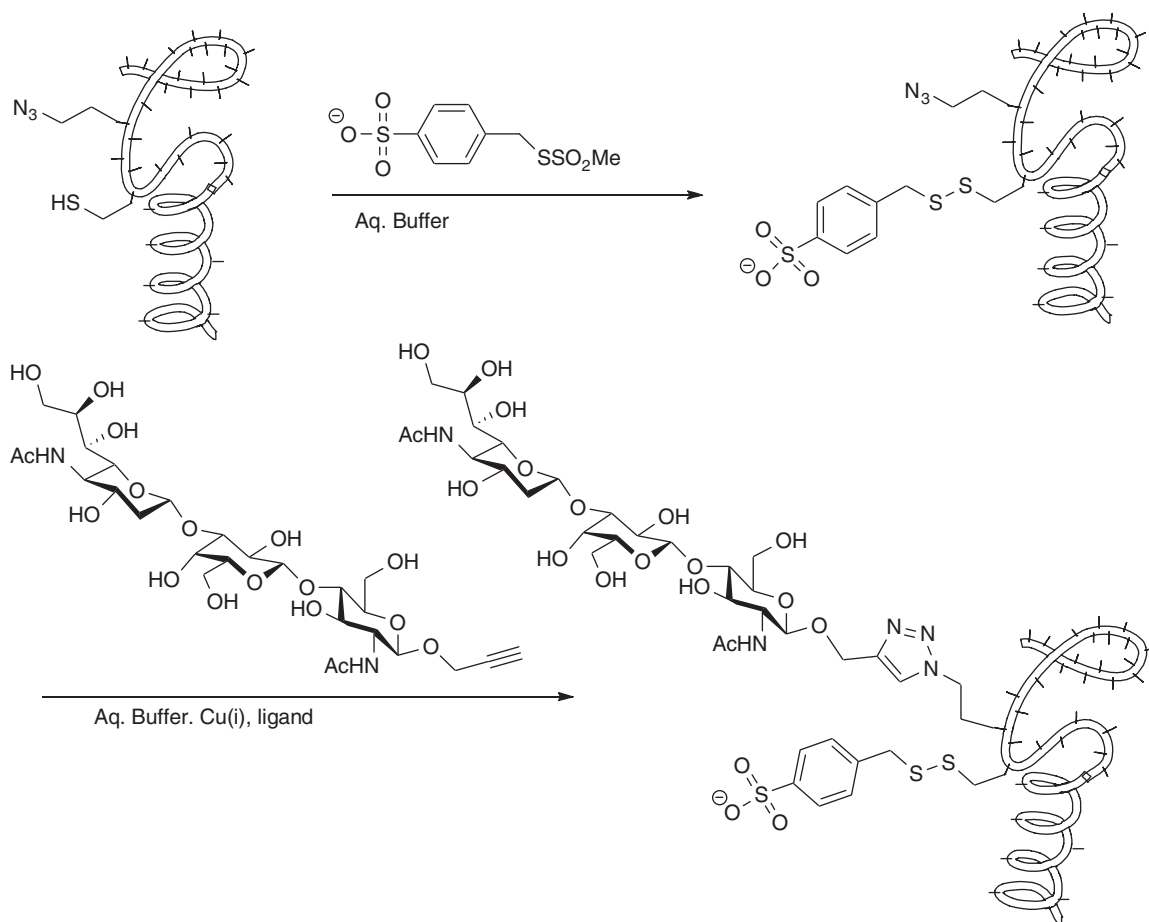


Figure 4 Synthesis of neoglycopeptides via click chemistry.

6.03.4.1 Chemical Approaches

In the field of chemical synthesis of glycoproteins, ligation methods, particularly NCL, stand as the preferred strategy to construct the peptidic part fashioned with the glycan structure. For example, Bertozzi and coworkers used NCL for the synthesis of diptericin, an 82-residue antibacterial glycoprotein, from two glycopeptide segments.¹⁰² Here, the 24-residue glycopeptide thioester was assembled using the alkanesulfonamide ‘safety-catch’ linker by the Fmoc-SPPS protocol. This segment was ligated with the cysteine bearing peptide to assemble the full-length diptericin. It must be noted that the native diptericin is devoid of cysteine residues, which are obligatory for NCL. Thus, Gly25 within the putative interdomain linker was replaced with Cys to afford a ligation site. Nevertheless, the glycoprotein was able to block bacterial growth with IC₅₀ similar to the native version of diptericin.

Our group took on the challenge of synthesizing the above-described glycoprotein, yet without the use of cysteine residue for NCL.¹⁶⁸ In this study, we took advantage of SAL to promote ligation at noncysteinyll residue. In our synthetic design, we dissected the protein sequence into three segments, which was assembled



Scheme 17 Differential multisite chemical protein modification.

sequentially by SAL and NCL (**Scheme 18**). The residues Gly⁵²-Val⁵³, which are next to glycosylation site Thr⁵⁴, represented an ideal ligation site for SAL as this junction is located near the middle of the sequence. Performing SAL generated glycopeptide Cys³⁷-Phe⁸², which was subjected to NCL to give the full-sized dipteracin. By including NCL to our synthetic scheme, the potential difficulty related to the synthesis of 52-residue glycopeptide thioester was avoided only if SAL was used. To take advantage of NCL, Ala³⁷ was temporarily mutated to cysteine. Finally a desulfurization step was performed to reduce both the cysteine and the thiol handle furnishing the full-sized dipteracin.

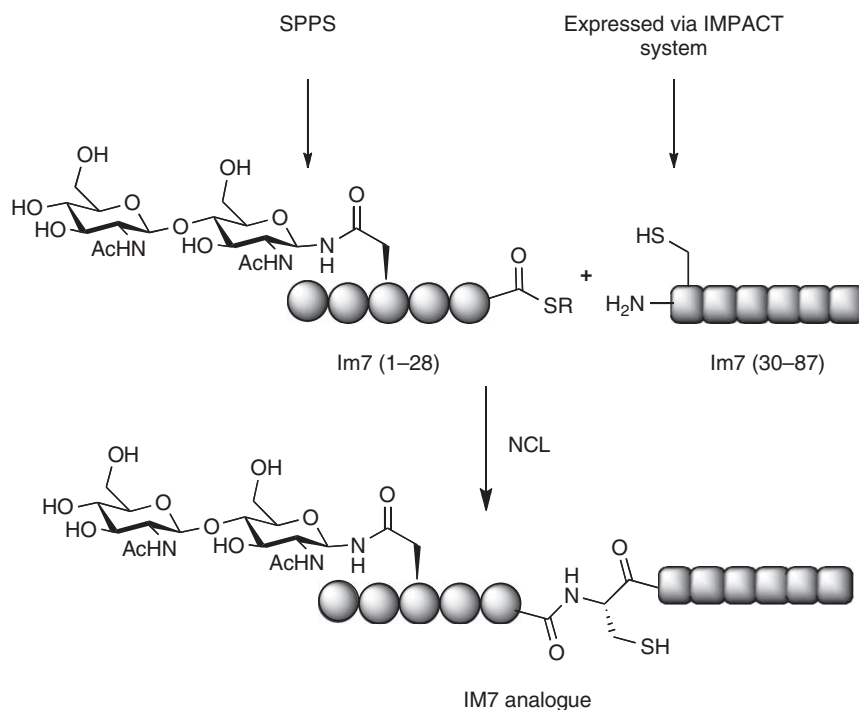
Recently, in a joint effort, Kajihara and Dawson reported the synthesis of a single glycoform of monocyte chemotactic protein-3 (MCP-3), a CC chemokine that consists of 76 amino acids and one N-glycosylation site, sialyloligosaccharide.¹⁶⁹ In their study, the team utilized a three-segment NCL approach, where Fmoc- and Boc-SPPS protocols were used to assemble the sialylglycopeptide-thioester segment. The synthesis of the glycopeptide bearing the thioester moiety at the C-terminal was carried out using the Boc-SPPS protocol; however, the use of HF was avoided by using minimal side-chain protection and direct thiolysis of the resin-bound peptide.

Using chemical synthesis, yet without relying on NCL, the Nakahara group has adopted a highly convergent strategy for the synthesis of MUC2 tandem repeat model composed of 141 amino acids with 42 GalNAc moieties.¹⁷⁰ In this approach, peptide thioester corresponding to a MUC2 tandem unit was prepared and consecutively joined by the activation of a thioester group using silver ions. This impressive work represents one of the largest glycoproteins made so far using chemical synthesis. However, in most cases this method would require protected peptides in organic solvents to achieve the desired chemoselectivity, which could limit the general utility of this strategy.

6.03.4.2 Biochemical Approaches

The high chemoselectivity and efficiency of the NCL to construct large peptides and glycopeptides inspired the development of a biochemical method named ‘Expressed Protein Ligation’ (EPL).¹⁷¹ In this approach, the C-terminal thioester segment can be obtained by thiolysis of a corresponding protein–intein fusion or polypeptide fragment containing N-terminal cysteine-expressed protein. For example, the Wong lab has used Tobacco Etch Virus (TEV) protease-cleavable fusion proteins to release cysteine proteins from suitable precursors.¹⁷² Notably, EPL technology allows access to large proteins with posttranslational modifications, for example, glycoproteins. One of the limitations of using the EPL approaches is that the sugar part should be located toward the C-terminus if the peptide thioester is expressed or toward the N-terminus if the cysteine fragment is expressed. In the latter approach, an additional limitation could arise due to the requirement of cysteine residue at the ligation junction where the use of auxiliaries to assist the ligation cannot be utilized. Several studies have already shown the utility of EPL in manipulating proteins for biochemical and biophysical studies, but so far only a few examples have been reported for glycoprotein synthesis. Recently, the Imperiali group prepared glycosylated Im7 analogue for protein folding studies using semisynthesis wherein the N-glycosylated thioester peptide (28 residues) was prepared by Fmoc-SPPS, while the larger fragment (59 residues) with N-terminal cysteine was expressed (Scheme 19).¹⁷³ Similarly, Macmillan and Bertozzi used EPL to access the unglycosylated fragment of GlyCAM-1 bearing the cysteine residue at the N-terminus for the synthesis of GlyCAM-1 glycoforms with as many as 13 *N*-acetylgalactosamine residues at predetermined positions.¹⁷⁴

The use of enzymatic approaches in glycoprotein synthesis is not limited to facilitate peptide ligation, but can also be used to remodel the carbohydrate moiety of a glycoprotein. As described in previous sections, various enzymes are known to catalyze the glycosidic linkage formation to generate more complex glycan structures. For example, the oligosaccharides of RNase B, occurring in different glycoforms, were removed leaving only GlcNAc as an acceptor for further glycosylation. Following this step, β -1,4-galactosyltransferase and α -2,3-sialyltransferase were used to introduce sLe^x epitope or the Hg-derivative for crystallographic studies.¹⁷⁵ Using a different set of enzymes (Endo-A, Endo-M) Takegawa and coworkers reported the efficient



Scheme 19 Semisynthetic strategy for glycosylated Im7 analogue.

transfer of (Man)₆GlcNAc to a partially deglycosylated RNase B.¹⁷⁶ More recently, the Wang group used RNase B as a model system to validate the usefulness of employing oligosaccharide oxazolines as donor substrates.¹⁷⁷ Thus, the oligosaccharides of natural RNase B were trimmed enzymatically leaving only the innermost GlcNAc followed by transferring tetrasaccharide or hexasaccharide oxazolines to furnish the (Man)₃GlcNAc₂-RNase B and the (Gal)₂Man₃GlcNAc₂-RNase B as a homogenous product in high yield.

Recombinant DNA-based expression allows protein substitutions to be limited among the common 20 amino acids. Consequently, considerable efforts have focused on methods to generate homogeneous glycoproteins using *in vitro* translation¹⁷⁸ and pathway engineering.¹⁷⁹ The pathway engineering method in yeast will be useful for making N-linked glycoproteins, and if combined with enzymatic glycosyltransfer reactions *in vitro*, complex human glycoproteins may be prepared. The sites of glycosylation in a human glycoprotein, especially the O-glycosylation site, however, may be different from that in yeast.

A promising new strategy for controlled *in vivo* synthesis of proteins with unnatural amino acids was established using suppressor tRNA technology to add new building blocks to the genetic code.¹⁸⁰ This newly developed technology has been used by our laboratory to express myoglobin containing GlcNAc serine at a defined position in good yield and high fidelity by evolving an orthogonal *Methanococcus jannaschii* Tyr-tRNA synthetase (MjTyrRS) and Tyr-tRNA pair that does not cross-react with endogenous *E. coli* tRNA and aminoacyl-tRNA synthetases.¹⁸¹ Similarly, the Wong group used this approach to incorporate mucin-type GalNAc- α -Thr into myoglobin in *E. coli* by evolving a specific MjTyrRS pair.¹⁸² The presented approach, if proven to be scalable, holds great potential for the synthesis of homogeneous glycoproteins for structure–function relationship studies as well as for the production of therapeutic glycoproteins.

6.03.5 Conclusions

The field of glycopeptide and homogenous glycoprotein synthesis is of current interest to many laboratories worldwide. With over 50% of human proteins being glycosylated there is increasing evidence that glycoproteins are playing a central role in many important biological processes and disease developments. The recent discovery that HIV-neutralizing antibodies (Abs), 2G12, recognize a conserved and unusually dense cluster of oligomannose residues on the gp120 of HIV-1 highlights the importance of glycoproteins in biological processes¹⁸³ and in the design of novel immunogens.¹⁸⁴ In addition, glycoproteins hold great interest in drug discovery as presented by over 60 recombinant glycoproteins that are in development stages. Yet, there are remaining challenges to overcome in several aspects of glycan, glycopeptide, and glycoprotein synthesis to obtain homogenous products in a more effective manner. Nevertheless, we have witnessed over the past decade the development of several promising methods based on chemical and biological approaches, thus allowing the preparation of neoglycopeptides, glycopeptides, and glycoproteins for biological studies. These newly developed tools have laid the ground for further development of more general and efficient technologies for the synthesis of homogenous glycoproteins and their mimetics.

Abbreviations

| | |
|-------------|---------------------------------------|
| AA | amino acid |
| Ac | acetyl |
| Acm | acetamidomethyl |
| AFGP | antifreeze glycoprotein |
| BAL | backbone amide linker |
| Bn | benzyl |
| Bz | benzoyl |
| Boc | <i>tert</i> -butyloxycarbonyl |
| BSA | bovine serum albumin |
| CMP | cytidine monophosphate |
| DCC | <i>N,N'</i> -dicyclohexylcarbodiimide |
| DMF | <i>N,N</i> -dimethylformamide |

| | |
|-------------------------|---|
| DNP | dinitrophenyl |
| DMSO | dimethyl sulfoxide |
| DIEA | <i>N</i> -ethyl- <i>N,N</i> -diisopropylamine |
| ENGase | endo- β - <i>N</i> -acetylglucosaminidase |
| EPO | erythropoietin |
| EtOAc | Ethyl acetate |
| Fuc | fucose |
| FucT | α -1,3-fucosyltransferase |
| GDP | guanosine 5'-diphosphate |
| GlcNAc | 2-acetamido-2-deoxyglucose |
| GalNAc | 2-acetamido-2-deoxy-D-galactose |
| Gal | galactose |
| GalT | β -1-4-galactosyltransferase |
| gp | glycoprotein |
| EPL | expressed protein ligation |
| Fmoc | fluoren-9-ylmethoxycarbonyl |
| HBTU | o-(benzotriazole-1-yl)- <i>N,N,N,N'</i> -tetramethyluronium hexafluorophosphate |
| HOBT | 1-hydroxy-1 <i>H</i> -benzotriazole |
| HIV | human immunodeficiency virus |
| HF | hydrofluoric acid |
| NMR | nuclear magnetic resonance |
| KLH | keyhole limpet hemocyanin |
| Man | mannose |
| MCP-3 | monocyte chemotactic protein-3 |
| MjTyrRS | <i>Methanococcus jannaschii</i> Tyr-tRNA synthetase |
| NCL | native chemical ligation |
| NeuAc | <i>N</i> -acetylneuraminic acid |
| Ph | phenyl |
| Phth | phthalooyl |
| PSGL-1 | P-selectin glycoprotein ligand-1 |
| RNA | ribonucleic acid |
| SAL | sugar-assisted ligation |
| SiaT | sialyltransferase |
| SLe^x | sialyl-Lewis-X |
| SPPS | solid-phase peptide synthesis |
| TBAHS | tetra- <i>n</i> -butylammonium hydrogen sulfate |
| TBS | tributyl silyl |
| TEV | tobacco etch virus |
| Tf | trifluoromethanesulfonyl |
| TFA | trifluoroacetic acid |
| TLC | thin layer chromatography |
| TMSN₃ | trimethylsilyl azide |
| UDP | uridine diphosphate |

References

1. R. A. Dwek, *Chem. Rev.* **1996**, 96, 683–720.
2. R. G. Spiro, *Glycobiology* **2002**, 12, 43–56.
3. A. Varki, *Glycobiology* **1993**, 3, 97–130.
4. H. Lis; N. Sharon, *Eur. J. Biochem.* **1993**, 218, 1–27.

5. A. Helenius; M. Aebi, *Science* **2001**, *291*, 2364–2369.
6. D. H. Dube; C. R. Bertozzi, *Nat. Rev. Drug Discovery* **2005**, *4*, 477–488.
7. K. J. Doores; D. P. Gambelin; B. G. Davis, *Chem. Eur. J.* **2006**, *12*, 656–665.
8. A. M. Sinclair; S. Elliott, *J. Pharmacol. Sci.* **2005**, *94*, 1626–1635.
9. F. Walter; S. Schoof; R. D. Sussmüth, *Top. Curr. Chem.* **2007**, *267*, 143–185.
10. J. Montreuil, *Adv. Carbohydr. Chem. Biochem.* **1980**, *37*, 157–223.
11. G. W. Hart; K. Brew; G. A. Grant; R. A. Bradshaw; W. J. Lennarz, *J. Biol. Chem.* **1979**, *254*, 9747–9753.
12. R. G. Spiro, *J. Biol. Chem.* **1967**, *242*, 4813–4823.
13. R. G. Spiro, *Adv. Protein Chem.* **1973**, *27*, 349–467.
14. J. E. Sadler, *Biology of Carbohydrates*, Wiley: New York, **1984**; Vol. 2, pp 199–288.
15. Y. Shimizu; S. Shaw, *Nature* **1993**, *366*, 630–631.
16. G. J. Strous; J. Dekker, *Biochem. Mol. Biol.* **1992**, *27*, 57–92.
17. J. Hilkens; S. V. Lightenberg; H. L. Vos; S. V. Litvinov, *Trends Biochem. Sci.* **1992**, *17*, 359–363.
18. G. D. Holt; R. S. Haltiwanger; C.-R. Torres; G. W. Hart, *J. Biol. Chem.* **1987**, *262*, 14847–14850.
19. G. W. Hart, *Annu. Rev. Biochem.* **1997**, *66*, 315–335.
20. B. G. Davis, *Chem. Rev.* **2002**, *102*, 579–602.
21. O. Seitz, *ChemBioChem* **2000**, *1*, 214–246.
22. M. R. Pratt; C. R. Bertozzi, *Chem. Soc. Rev.* **2005**, *34*, 58–68.
23. A. Brik; S. Ficht; C.-H. Wong, *Curr. Opin. Chem. Biol.* **2006**, *10*, 638–644.
24. K. C. Nicolaou; E. J. Sorensen, *Classics in Total Synthesis*, Wiley-VCH: Weinheim, 1996.
25. S. J. Danishefsky; J. R. Allen, *Angew. Chem. Int. Ed.* **2000**, *39*, 836–863.
26. P. H. Seeberger; W.-C. Haas, *Chem. Rev.* **2000**, *100*, 4349–4394.
27. P. Sears; C.-H. Wong, *Science* **2001**, *291*, 2344–2350.
28. O. Seitz; C.-H. Wong, *J. Am. Chem. Soc.* **1997**, *119*, 8766–8776.
29. C. Unverzagt, *Carbohydr. Res.* **1998**, *305*, 423–431.
30. L. A. Marcaurelle; C. R. Bertozzi, *Chem. Eur. J.* **1999**, *5*, 1383–1390.
31. T. Bar; R. R. Schmidt, *Liebigs Ann. Chem.* **1991**, *2*, 185–187.
32. D. Horton; J. D. Wander, *Carbohydrates: Chemistry and Biochemistry*, 2nd ed.; Academic Press: New York, 1990; Vol. 4B, pp 799–842.
33. M. A. Sparks; K. W. Williams; G. M. Whitesides, *J. Med. Chem.* **1993**, *36*, 778–783.
34. L. M. Likhoshero; O. S. Novikova; V. A. Derevitskaya; N. K. Kochetkov, *Carbohydr. Res.* **1986**, *146*, C1–C5.
35. C. Auge; C. Gautheron; H. Pora, *Carbohydr. Res.* **1989**, *193*, 288–293.
36. C. H. Bolton; R. W. Jeanloz, *J. Org. Chem.* **1963**, *28*, 3228–3230.
37. C. Unverzagt; H. Kunz, *J. Prakt. Chem.* **1992**, *334*, 570–578.
38. F. D. Tropper; F. O. Andersson; S. Braun; R. Roy, *Synthesis* **1992**, *7*, 618–620.
39. L. Szilagy; Z. Gyoergydeak, *Carbohydr. Res.* **1985**, *143*, 21–41.
40. K. Mastsubara; T. Mukaiama, *Chem. Lett.* **1994**, *15*, 247–250.
41. E. D. Soli; P. DeShong, *J. Org. Chem.* **1999**, *64*, 9724–9726.
42. J. Y. Roberge; X. Beebe; S. J. Danishefsky, *J. Am. Chem. Soc.* **1998**, *120*, 3915–3927.
43. J. Thiem; T. Wiemann, *Angew. Chem. Int. Ed.* **1990**, *29*, 80–82.
44. R. S. Clark; S. Banerjee; J. K. Coward, *J. Org. Chem.* **1990**, *55*, 6275–6285.
45. M. A. E.-M. Shaban; R. W. Jeanloz, *Bull. Chem. Soc. Jpn.* **1981**, *54*, 3570–3576.
46. H. Bayley; D. N. Standing; J. R. Knowles, *Tetrahedron Lett.* **1978**, *39*, 3633–3634.
47. M. R. Pratt; C. R. Bertozzi, *J. Am. Chem. Soc.* **2003**, *125*, 6149–6159.
48. Z. W. Guo; Y. Nakahara; T. Ogawa, *Bioorg. Med. Chem.* **1997**, *5*, 1917–1924.
49. Z. W. Guo; Y. Nakahara; T. Ogawa, *Angew. Chem. Int. Ed.* **1997**, *36*, 1464–1466.
50. S. Wen; Z. Guo, *Org. Lett.* **2001**, *3*, 3773–3776.
51. J. Xue; Z. Guo, *J. Org. Chem.* **2003**, *68*, 2713–2720.
52. S. T. Ansfield; P. T. Lansbury, *J. Org. Chem.* **1990**, *55*, 5560–5562.
53. M. Mandal; V. Y. Dudkin; X. Geng; S. J. Danishefsky, *Angew. Chem. Int. Ed.* **2004**, *43*, 2557–2561.
54. X. Geng; V. Y. Dudkin; M. Mandal; S. J. Danishefsky, *Angew. Chem. Int. Ed.* **2004**, *43*, 2562–2565.
55. G. A. Winterfeld; R. R. Schmidt, *Angew. Chem. Int. Ed.* **2001**, *40*, 2654–2657.
56. E. Meinjohanns; M. Meldal; H. Paulsen; A. Schleyer; K. Bock, *J. Chem. Soc., Perkin Trans. 1* **1996**, *10*, 985–993.
57. N. Mathieux; H. Paulsen; M. Meldal; K. Bock, *J. Chem. Soc., Perkin Trans. 1* **1997**, *16*, 2359–2368.
58. Y. Nakahara; H. Iijima; T. Ogawa, *Tetrahedron Lett.* **1994**, *35*, 3321–3324.
59. B. Liebe; H. Kunz, *Tetrahedron Lett.* **1994**, *35*, 8777–8778.
60. S. J. Danishefsky; J. R. Allen, *Angew. Chem. Int. Ed.* **2000**, *39*, 836–863.
61. D. Sames; X. T. Chen; S. J. Danishefsky, *Nature* **1997**, *389*, 587–591.
62. J. B. Schwarz; S. D. Kuduk; X. T. Chen; D. Sames; P. W. Glunz; S. J. Danishefsky, *J. Am. Chem. Soc.* **1999**, *121*, 2662–2673.
63. C. P. R. Hackenberger; M. K. O'Reilly; B. Imperiali, *J. Org. Chem.* **2005**, *70*, 3574–3578.
64. L. Jobron; G. Hummel, *Angew. Chem. Int. Ed.* **2000**, *39*, 1621–1624.
65. E. Meinjohanns; M. Meldal; H. Paulsen; R. A. Dwek; K. Bock, *J. Chem. Soc., Perkin Trans. 1* **1998**, *3*, 549–560.
66. B. Holm; S. Linse; J. Kihlberg, *Tetrahedron* **1998**, *54*, 11995–12006.
67. J. Broddefalk; K. Bergquist; J. Kihlberg, *Tetrahedron* **1998**, *54*, 12047–12070.
68. H. Kunz; B. Dombo, *Angew. Chem. Int. Ed.* **1998**, *27*, 711–713.
69. D. Vetter; D. Tumetly; S. K. Singh; M. A. Gallop, *Angew. Chem. Int. Ed.* **1995**, *34*, 60–63.
70. I. Christiansen-Brams; M. Meldal; K. Bock, *J. Chem. Soc., Perkin Trans. 1* **1993**, *13*, 1461–1471.
71. S. Lavielle; N. C. Ling; R. C. Guillemin, *Carbohydr. Res.* **1981**, *89*, 221–228.
72. S. Lavielle; N. C. Ling; R. Saltman; R. C. Guillemin, *Carbohydr. Res.* **1981**, *89*, 229–236.

73. H. Paulsen; G. Merz; U. Weichert, *Angew. Chem. Int. Ed.* **1988**, *27*, 1365–1367.
74. B. Luning; T. Norberg; C. Riversa-Baeza; J. Tejbrant, *Glytoconj. J.* **1991**, *8*, 450–455.
75. R. Polt; L. Szabo; T. Treiberg; Y. Li; V. J. Hruby, *J. Am. Chem. Soc.* **1992**, *114*, 10249–10258.
76. O. Seitz; H. Kunz, *Angew. Chem. Int. Ed.* **1995**, *34*, 803–805.
77. O. Seitz; H. Kunz, *J. Org. Chem.* **1997**, *62*, 813–826.
78. S. Nollet; N. Moniaux; J. Maury; D. Petitprez; P. Degand; A. Laine; N. Porchet; J.-P. Aubert, *Biochem. J.* **1998**, *332*, 739–748.
79. M. Andrianifahanana; N. Moniaux; B. M. Schmied; J. Ringel; H. Friess; M. A. Hollingsworth; M. W. Büchler; J.-P. Aubert; S. K. Batra, *Clin. Cancer Res.* **2001**, *7*, 4033–4040.
80. C. Brocke; H. Kunz, *Synlett* **2003**, *13*, 2052–2056.
81. D. A. Thayer; C.-H. Wong, *Top. Curr. Chem.* **2007**, *267*, 37–63.
82. F. Sallas; S. I. Nishimura, *J. Chem. Soc., Perkin Trans. 1* **2000**, *13*, 2091–2103.
83. M. Matsuda; S. I. Nishimura; F. Nakajima; T. Nishimura, *J. Med. Chem.* **2001**, *44*, 715–724.
84. A. Leppanen; P. Mehta; Y. B. Ouyang; T. Z. Ju; J. Helin; K. L. Moore; W. M. van Die; I. Canfield; R. P. McEver; R. D. Cummings, *J. Biol. Chem.* **1999**, *274*, 24838–24848.
85. A. Leppanen; S. P. White; J. Helin; R. P. McEver; R. D. Cummings, *J. Biol. Chem.* **2000**, *275*, 39569–39578.
86. A. Leppanen; S. P. White; J. Helin; R. P. McEver; R. D. Cummings, *Glycobiology* **2000**, *10*, 1106–1107.
87. K. M. Koeller; M. E. B. Smith; C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 742–743.
88. K. M. Koeller; M. E. B. Smith; C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 4241–4242.
89. G. Dudziak; N. Bezay; T. Schwientek; H. Clausen; H. Kunz; A. Liese, *Tetrahedron* **2000**, *56*, 5865–5869.
90. X. X. Zeng; Y. Nakaaki; T. Murata; T. Usui, *Arch. Biochem. Biophys.* **2000**, *383*, 28–37.
91. C. Unverzagt, *Angew. Chem. Int. Ed.* **1996**, *35*, 2350–2353.
92. C. Unverzagt, *Carbohydr. Res.* **1997**, *305*, 423–431.
93. V. W. F. Tai; B. Imperiali, *J. Org. Chem.* **2001**, *66*, 6217–6228.
94. R. B. Trimble; A. L. Tarentino, *J. Biol. Chem.* **1991**, *266*, 1646–1651.
95. B. Li; Y. Zeng; S. Hauser; H. Song; X. L. Wang, *J. Am. Chem. Soc.* **2005**, *127*, 9692–9693.
96. A. C. Terwisscha van Scheltinga; S. Armand; K. H. Kalk; A. Isogai; B. Henrissat; B. W. Dijkstra, *Biochemistry* **1995**, *34*, 15619–15623.
97. B. L. Mark; D. J. Vocadlo; S. Knapp; B. L. Triggs-Raine; S. G. Withers; M. N. James, *J. Biol. Chem.* **2001**, *276*, 10330–10337.
98. P. E. Dawson; S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
99. P. E. Dawson; T. W. Muir; I. Clarklewis; S. B. H. Kent, *Science* **1994**, *266*, 776–779.
100. E. C. B. Johnson; S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.
101. T. M. Hackeng; J. H. Griffin; P. E. Dawson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068–10073.
102. Y. Shin; K. A. Winans; B. J. Backes; S. B. H. Kent; J. A. Ellman; C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, *121*, 11684–11689.
103. J. A. Camarero; A. R. Mitchell, *Protein Pept. Lett.* **2005**, *12*, 723–728.
104. C. M. Gross; D. Lelievre; C. K. Woodward; G. Barany, *J. Pept. Res.* **2005**, *65*, 395–410.
105. D. Swinnen; D. Hilvert, *Org. Lett.* **2000**, *2*, 2439–2442.
106. S. Ficht; R. J. Payne; R. T. Guy; C.-H. Wong, *Chem. Eur. J.* **2008**, *14*, 3620–3629.
107. P. Wang; L. P. Miranda, *Int. J. Pept. Res. Ther.* **2005**, *11*, 117–123.
108. G. S. Belligere; P. E. Dawson, *J. Am. Chem. Soc.* **1999**, *121*, 6332–6333.
109. J. P. Tam; J. Xu; K. D. Eom, *Biopolymers* **2001**, *60*, 194–205.
110. L. Z. Yan; P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533.
111. R. Okamoto; Y. Kajihara, *Angew. Chem. Int. Ed.* **2008**, *47*, 1–6.
112. D. Crich; A. Banerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.
113. L. E. Cane; S. J. Bark; J. D. Warren; G. Chen; Z. Hua; S. J. Danishefsky, *J. Am. Chem. Soc.* **1996**, *118*, 5891–5896.
114. D. W. Low; M. G. Hill; M. R. Carrasco; S. B. H. Kent; P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554–6559.
115. J. Offer; C. Boddy; P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642–4646.
116. D. Macmillan; D. W. Anderson, *Org. Lett.* **2004**, *6*, 4659–4662.
117. B. Wu; J. Chen; J. D. Warren; G. Chen; Z. Hua; S. J. Danishefsky, *Angew. Chem. Int. Ed.* **2006**, *45*, 4116–4125.
118. J. Chen; G. Chen; B. Wu; Q. Wan; Z. Tan; Z. Hua; S. J. Danishefsky, *Tetrahedron Lett.* **2006**, *47*, 8013–8016.
119. A. Brik; Y.-Y. Yang; S. Ficht; C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627.
120. A. Brik; S. Ficht; Y.-Y. Yang; C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 15026–15033.
121. L. C. Hsieh-Wilson, *Nature* **2007**, *445*, 31–33.
122. A. Brik; C.-H. Wong, *Chem. Eur. J.* **2007**, *13*, 5670–5675.
123. R. J. Payne; S. Ficht; S. Tang; A. Brik; Y.-Y. Yang; D. A. Case; C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 13527–13536.
124. R. I. Carey; D. S. Kemp, *J. Org. Chem.* **1993**, *58*, 2216–2222.
125. S. Ficht; R. J. Payne; A. Brik; C.-H. Wong, *Angew. Chem. Int. Ed.* **2007**, *46*, 5975–5979.
126. M.-Y. Lutsky; N. Nepomniaschii; A. Brik, *Chem. Commun.* **2008**, *10*, 1229–1231.
127. E. Saxon; J. I. Armstrong; C. R. Bertozzi, *Org. Lett.* **2000**, *2*, 2141–2143.
128. B. L. Nilsson; L. L. Kiessling; R. T. Raines, *Org. Lett.* **2001**, *3*, 9–12.
129. R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91–118.
130. L. Liu; Z.-Y. Hong; C.-H. Wong, *ChemBioChem* **2006**, *7*, 429–432.
131. Y. Du; R. J. Linhardt, *Tetrahedron* **1998**, *54*, 9913–9959.
132. F. Burkhart; M. Hoffmann; H. Kessler, *Angew. Chem. Int. Ed.* **1997**, *36*, 1191–1192.
133. M. Hoffmann; F. Burkhart; G. Hessler; H. Kessler, *Helv. Chim. Acta* **1996**, *79*, 1519–1532.
134. L.-X. Wang; J.-Q. Fan; Y. C. Lee, *Tetrahedron Lett.* **1996**, *37*, 1975–1978.
135. A. Dondoni; A. Marra; A. Massi, *J. Org. Chem.* **1999**, *64*, 933–944.
136. T. Fuchss; R. R. Schmidt, *Synthesis* **1998**, *5*, 753–758.
137. L. Lay; M. Meldal; F. Nicotra; L. Panza; G. Russo, *J. Chem. Soc., Chem. Commun.* **1997**, *15*, 1469–1470.
137. A. Dondoni; P. P. Giovannini; D. Perrone, *J. Org. Chem.* **2005**, *70*, 5508–5518.

138. L.-X. Wang; M. Tang; T. Suzuki; K. Kitajima; Y. Inoue; S. Inoue; J.-Q. Fan; Y. C. Lee, *J. Am. Chem. Soc.* **1997**, *119*, 11137–11146.
139. A. J. Pearce; S. Ramaya; S. N. Thorn; G. B. Bloomberg; D. S. Walter; T. Gallagher, *J. Org. Chem.* **1999**, *64*, 5453–5462.
140. S. Liu; R. N. Ben, *Org. Lett.* **2005**, *7*, 2385–2388.
141. P. Czechura; R. Y. Tam; E. Dimitrijevic; A. V. Murphy; R. N. Ben, *J. Am. Chem. Soc.* **2008**, *130*, 2928–2929.
142. A. Dirksen; T. M. Hackeng; P. E. Dawson, *Angew. Chem. Int. Ed.* **2006**, *45*, 7581–7584.
143. E. C. Rodriguez; L. A. Marcaurrelle; C. R. Bertozzi, *J. Org. Chem.* **1998**, *63*, 7134–7135.
144. E. C. Rodriguez; K. A. Winans; D. S. King; C. R. Bertozzi, *J. Am. Chem. Soc.* **1997**, *119*, 9905–9906.
145. A. H. Andreotti; D. Kahne, *J. Am. Chem. Soc.* **1993**, *115*, 3352–3353.
146. R. Liang; A. H. Andreotti; D. Kahne, *J. Am. Chem. Soc.* **1995**, *117*, 10395–10396.
147. L. A. Marcaurrelle; E. C. Rodriguez; C. R. Bertozzi, *Tetrahedron Lett.* **1998**, *39*, 8417–8420.
148. F. Peri; F. Nicotra, *Chem. Commun.* **2004**, *6*, 624–627.
149. B. Capon, *Chem. Rev.* **1969**, *69*, 407–496.
150. H. Driguez, *Top. Curr. Chem.* **1997**, *187*, 85–116.
151. Y. Kishi, *Pure Appl. Chem.* **1993**, *32*, 771–778.
152. X. M. Zhu; K. Pachamuthu; R. R. Schmidt, *J. Org. Chem.* **2003**, *68*, 5641–5651.
153. X. M. Zhu; R. R. Schmidt, *Chem. Eur. J.* **2004**, *10*, 875–887.
154. D. A. Thayer; H. N. C. Galan; C.-H. Wong, *Angew. Chem. Int. Ed.* **2005**, *44*, 4596–4599.
155. G. J. L. Bernardes; E. J. Grayson; S. Thompson; J. M. Chalker; J. C. Errey; F. El Oualid; T. D. W. Claridge; B. G. Davis, *Angew. Chem. Int. Ed.* **2008**, *47*, 1–5.
156. D. Crich; V. Krishnamurthy; K. T. Hutton, *J. Am. Chem. Soc.* **2006**, *128*, 2544–2545.
157. N. J. Davis; S. L. Flitsch, *Tetrahedron Lett.* **1991**, *32*, 6793–6796.
158. D. Macmillan; R. M. Bill; K. A. Sage; D. Fern; S. L. Flitsch, *Chem. Biol.* **2001**, *8*, 133–145.
159. S. Y. C. Wong; G. R. Gulie; R. A. Dwek; G. Arsequell, *Biochem. J.* **1994**, *300*, 843–850.
160. D. Macmillan; A. M. Daines; M. Bayrhuber; S. L. Flitsch, *Org. Lett.* **1991**, *4*, 1467–1470.
161. H. C. Kolb; M. G. Finn; K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
162. V. V. Rostovtsev; L. G. Green; V. V. Fokin; K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
163. C. W. Tornøe; C. Christensen; M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064.
164. S. Dedola; S. A. Nepogodie; R. A. Field, *Org. Biomol. Chem.* **2007**, *5*, 1006–1017.
165. Q. Wan; J. Chen; G. Chen; S. J. Danishefsky, *J. Org. Chem.* **2006**, *71*, 8244–8249.
166. D. Macmillan; J. Blanc, *Org. Biomol. Chem.* **2006**, *4*, 2847–2850.
167. S. I. van Kasteren; H. B. Kramer; H. H. Jensen; S. J. Campbell; J. Kirkpatrick; N. J. Oldham; D. C. Anthony; B. G. Davis, *Nature* **2007**, *446*, 1105–1109.
168. Y.-Y. Yang; S. Ficht; A. Brik; C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 7690–7701.
169. N. Yamamoto; Y. Tanabe; R. Okamoto; P. E. Dawson; Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130*, 501–510.
170. H. Hojo; Y. Matsumoto; Y. Nakahara; E. Ito; Y. Suzuki; M. Suzuki; A. Suzuki; Y. Nakahara, *J. Am. Chem. Soc.* **2005**, *127*, 13720–13725.
171. T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
172. T. J. Tolbert; C.-H. Wong, *Angew. Chem. Int. Ed.* **2002**, *41*, 2171–2174.
173. C. P. R. Hackenberger; C. T. Friel; S. E. Radford; B. Imperiali, *J. Am. Chem. Soc.* **2005**, *127*, 12882–12889.
174. D. Macmillan; C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2004**, *43*, 1355–1359.
175. K. Witte; P. Sears; C.-H. Wong, *J. Am. Chem. Soc.* **1997**, *119*, 2114–2118.
176. K. Fujita; N. Tanaka; M. Sano; I. Kato; Y. Asada; K. Takegawa, *Biochem. Biophys. Res. Commun.* **2000**, *267*, 134–138.
177. B. Li; H. Song; S. Hauser; L. X. Wang, *Org. Lett.* **2006**, *8*, 3081–3084.
178. T. Arslan; S. V. Mamaev; N. V. Mamaeva; S. M. Hecht, *J. Am. Chem. Soc.* **1997**, *119*, 10877–10887.
179. S. R. Hamilton; P. Bobrowicz; B. Bobrowicz; R. C. Davidson; H. Li; T. Mitchell; J. H. Nett; S. Rausch; T. A. Stadheim; H. Wischniewski; S. Wildt; T. U. Gerngross, *Science* **2003**, *301*, 1244–1246.
180. L. Wang; P. G. Schultz, *Angew. Chem. Int. Ed.* **2005**, *44*, 34–66.
181. 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.
182. R. Xu; S. R. Hanson; Z. Zhang; Y. Y. Yang; P. G. Schultz; C.-H. Wong, *J. Am. Chem. Soc.* **2004**, *126*, 15654–15655.
183. C. N. Scanlan; R. Pantophlet; M. R. Wormald; E. O. Saphire; R. Stanfield; I. A. Wilson; H. Katinger; R. A. Dwek; P. M. Rudd; D. R. Burton, *J. Virol.* **2002**, *76*, 7306–7321.
184. L. X. Wang, *Curr. Opin. Drug Discov. Dev.* **2006**, *9*, 194–206.

Biographical Sketch



Dr. Ashraf Brik received his BS from Ben-Gurion University (BGU) in Israel, his MS from the Technion-Israel Institute of technology, and his PhD in chemistry from the Technion with Professor Ehud Keinan jointly with Professor Philip Dawson from The Scripps Research Institute (TSRI). He received a postdoctoral fellowship from the Israel Science Foundation and spent three years as a postdoctoral fellow in the laboratory of Professor Chi-Huey Wong at TSRI. Dr. Brik was promoted to a senior research associate and worked for an additional two years with Professor Wong. In 2007, Dr. Brik returned to his alma mater as a Sr. Lecturer (Assistant Professor) in the Chemistry Department at BGU. His current research involves the design and synthesis of proteins containing peptidomimetic motifs and developing novel chemistries to access proteins with posttranslational modifications for biological studies. His group is also working in the design and synthesis of pharmaceutically relevant peptides and peptidomimetics. Dr. Brik is the recipient of Ma'of Fellowship, the Marie Curie International Re-Integration award (EU 6th program).

6.04 Carbohydrate Vaccines

Qianli Wang and Zhongwu Guo, Wayne State University, Detroit, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.04.1 | Introduction | 91 |
| 6.04.2 | Tumor-Associated Carbohydrate Antigens | 92 |
| 6.04.2.1 | TACA Expression | 92 |
| 6.04.2.2 | Mechanism of TACA Expression | 93 |
| 6.04.2.3 | TACAs in Cancer Diagnosis and Immunotherapy | 94 |
| 6.04.3 | TACA-Based Cancer Vaccine Development | 94 |
| 6.04.3.1 | Immunogenicity of Carbohydrates and Immunotolerance to TACAs | 94 |
| 6.04.3.2 | Synthesis of TACAs and TACA Conjugates | 95 |
| 6.04.3.2.1 | Chemical synthesis of TACAs and TACA conjugates | 95 |
| 6.04.3.2.2 | <i>In vitro</i> enzymatic synthesis of TACAs and TACA conjugates | 95 |
| 6.04.3.2.3 | <i>In vivo</i> metabolic engineering to obtain TACAs and TACA conjugates | 96 |
| 6.04.3.3 | Immunological Adjuvant Application | 97 |
| 6.04.3.4 | Semisynthetic Glycoconjugate Cancer Vaccines | 98 |
| 6.04.3.4.1 | Semisynthetic monovalent glycoconjugate vaccines | 99 |
| 6.04.3.4.2 | Semisynthetic monovalent clustered glycoconjugate vaccines | 102 |
| 6.04.3.4.3 | Semisynthetic multivalent glycoconjugate vaccines | 105 |
| 6.04.3.5 | Fully Synthetic Glycoconjugate Vaccines | 107 |
| 6.04.3.5.1 | Two-component fully synthetic glycoconjugate vaccines | 107 |
| 6.04.3.5.2 | Three-component fully synthetic glycoconjugate vaccines | 110 |
| 6.04.3.5.3 | Four-component fully synthetic glycoconjugate vaccines | 112 |
| 6.04.3.6 | Cancer Vaccines Made of Unnatural TACA Analogs | 113 |
| 6.04.3.6.1 | Cancer vaccines made of chemically modified sialo-TACAs | 113 |
| 6.04.3.6.2 | Cancer immunotherapy based on modified sialo-TACAs vaccines and cancer cell glycoengineering | 114 |
| 6.04.4 | Conclusion | 115 |
| References | | 116 |

6.04.1 Introduction

Cancer is one of the most notorious health problems worldwide. Despite great progress in its diagnosis and treatment by traditional methods, such as chemotherapy, radiotherapy, and surgery, it is still regarded as an essentially incurable disease. With the aim of discovering more effective and safer treatments for cancer, many new therapeutic strategies and techniques have been investigated, among which cancer immunotherapy has attracted significant attention in the past few decades. The ultimate objective of cancer immunotherapy is to educate the patient's immune system to recognize and target the antigens uniquely expressed or significantly overexpressed on cancer cells, which are known as tumor-associated antigens (TAAs), for cancer cell elimination, while leaving normal cells unaffected. This goal can be achieved either by using an antibody specific for a TAA or by provoking the patient's immune system with a cancer vaccine.

TAAs, which are usually peptides/proteins and carbohydrates, including glycopeptides, glycoproteins, and glycolipids, are important markers on the tumor cell surface that have been widely used for cancer diagnosis.¹⁻³ TAAs are also important molecular targets and templates for the design and development of cancer immunotherapy. Among various TAAs identified so far, tumor-associated carbohydrate antigens (TACAs) have attracted much attention because they are usually richly expressed by tumors and they are also often the most exposed TAAs on the cancer cell surface.^{4,5} Moreover, it has been well recognized that TACAs present on the cancer cell surface are closely correlated to tumor progression and tumor metastasis.⁶⁻⁸ However, a major

problem associated with TACAs is that carbohydrates are usually poorly immunogenic and induce T cell-independent immune response,⁹ while T cell-mediated immunity is crucial for antitumor activity.¹⁰ To develop functional cancer vaccines based on TACAs, it is necessary to establish effective strategies to overcome this and other related issues.

This chapter will summarize the recent progress in the development of cancer vaccines or cancer immunotherapies based on TACAs. However, due to the large body of literature reported in this area, it is impossible to cover every aspect in great detail here. Moreover, there have been many reviews dealing with the topic from various angles.^{10–22} In particular, much of the earlier work has been covered thoroughly by several recent reviews.^{11–14} Therefore, while citing these reviews frequently, this chapter will only briefly introduce the TACAs commonly employed for cancer vaccine exploration and will mainly focus on the current strategies utilized to improve the immunogenicity of TACAs in the effort to develop effective cancer vaccines.

6.04.2 Tumor-Associated Carbohydrate Antigens

6.04.2.1 TACA Expression

All cells are covered by a thick layer of carbohydrates, known as the glycocalyx, which consists of many different carbohydrate epitopes. The unique or overexpressed carbohydrate structures on cancer cells are termed TACAs. The Thomsen–Friedenreich (TF) antigen was the first TACA identified with the help of a specific monoclonal antibody (mAb) and mass spectrometry. Since then, a number of TACAs have been characterized.^{23–27} **Figure 1** shows some of the representative TACAs.

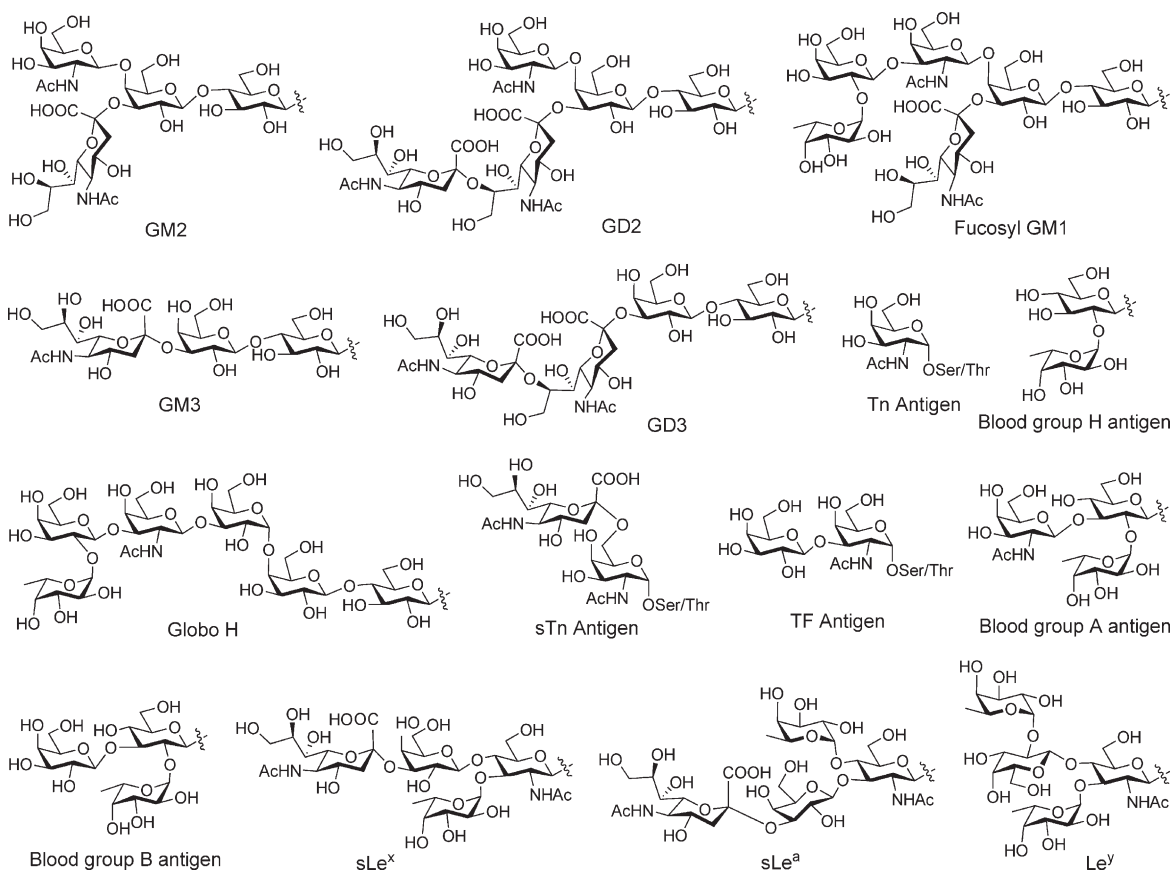


Figure 1 The structures of some representative TACAs.

Table 1 Common expression patterns of TACAs on malignant tissues

| <i>Tumors</i> | <i>Antigens</i> |
|------------------------|--|
| B cell lymphoma | GM2, GD2 |
| Breast | GM2, globo H, TF(c), Le ^y |
| Colon | GM2, TF(c), Le ^y , Tn, sTn(c), sLe ^a , blood group ABH |
| Lung | GM2, globo H, Le ^y , blood group ABH |
| Melanoma | GM2, GD2, GD3, GD3L |
| Neuroblastoma | GM2, GD2, GD3L, PSialA |
| Ovary | GM2, globo H, TF(c), Le ^y , sTn(c) |
| Prostate | GM2, globo H, TF(c), Le ^y , sTn(c), Tn(c) |
| Sarcoma | GM2, GD2, GD3, GD3L |
| Small cell lung cancer | GM2, globo H, sLe ^a , Fuc-GM1, PSialA |
| Stomach | GM2, sLe ^a , Le ^a , Le ^y |

PSialA, $\alpha(2,8)$ polysialic acid.

Reproduced from Table 1 in S. F. Slovin; S. J. Keding; G. Ragupathi, *Immunol. Cell Biol.* **2005**, *83*, 418–428.

TACAs are usually displayed as constituents of either glycolipids or glycoproteins. In the case of glycolipids, TACAs, such as GM2, GM3, GD2, GD3, fucosyl GM1, sLe^a, sLe^x, Le^y, and globo H, are attached to the lipid bilayer of the cell membrane by hydrophobic forces through a ceramide moiety. In glycoproteins, TACAs, such as TF, Tn, sTn, Le^y, and globo H, are generally attached to the hydroxyl group of serine and threonine residues through an O-linkage in mucins or to the asparagine residue through an N-linkage in other glycoproteins.

TACAs are widely detected in malignant tissues throughout the body including lung, breast, colorectal, ovarian, prostate, and pancreas. Each type of malignant tissue is characterized by a distinct set of changes in glycan expression (Table 1). For example, Zhang *et al.*^{28,29} screened a variety of malignant and normal tissues by immunohistochemistry using a panel of mAbs against carbohydrate antigens. They found that melanoma, sarcoma, and neuroblastoma expressed a broad range of carbohydrate antigens, such as GD2 and GD3. In the brain, GM2, GD2, and GD3 are abundantly expressed, while GD2 is also expressed on some peripheral nerves and in the spleen and lymph nodes. Globo H, Le^y, TF, Tn, and sTn are expressed at the secretory borders of a variety of epithelial tissues. Le^x and sLe^x are not only expressed at the secretory borders of many epithelial tissues but also on polymorphonuclear leukocytes. Aberrant expression of blood group antigens (ABH) has been observed in gastrointestinal, lung, cervical, oral epithelial, urothelial, and colon cancer.^{8,30,31}

6.04.2.2 Mechanism of TACA Expression

TACA expression on cancer cells is a complex event, and its mechanisms are still poorly understood. It is likely that TACAs are the result of deregulated changes of glycosyltransferases in the Golgi compartments of cancerous cells. For example, it has been observed that changes in glycosyltransferase levels can lead to modifications in the core structures of N-linked and O-linked glycans, as evidenced by the elevated expression of β 1,6-linked branching of complex oligosaccharides in the cell surface glycoprotein gp130 caused by the increased activity of *N*-acetylglucosaminyltransferase V (GlcNAc-TV, also known as MGAT5, the enzyme for β 1,6GlcNAc branching).³² Similarly, in lung cancer cells, a high expression of GD2 has been associated with an upregulation of the GD2 synthase gene (β 4GalNAc-T).³³ On the other hand, a reduced activity of A transferase (α 3GalNAc-T, processing the blood group antigens H (O) to A), results in a decreased expression of the A antigen, and consequently an increased expression of the H antigen.³⁴

In addition to changes in the core structure of glycans, increased branching creates additional sites for the attachment of terminal *N*-acetylneuramic acid (NeuNAc, more commonly known as sialic acid), which, together with the overexpression of sialyltransferases in tumor tissues, ultimately results in the global increase in the expression levels of NeuNAc on cancer cell surfaces^{35–40} and the observation that many TACAs are sialylated oligosaccharides, as shown in Figure 1. Similarly, the overexpression of fucosyltransferases in tumor tissues leads to an increase in the expression of fucosylated glycans on the cancer cell surface as well.^{41–43} For example, the presence of sTn antigen in LMCR rat colon cancer cells has been associated with a higher

ST6GalNAc activity,⁴⁴ and the overexpression of sLe^a is related to an increase of α 3Sial-T and α 4Fuc-T activities in colon and gastric cancer tissues, respectively.^{45,46} Moreover, melanoma cell lines expressing high levels of α 8Sial-T result in accumulation of GD3.⁴⁷

6.04.2.3 TACAs in Cancer Diagnosis and Immunotherapy

Investigations of the biological activities of aberrant TACAs expressed on cancer tissues indicated that these antigenic determinants are functionally related to the malignant behavior and the metastatic potential of cancer cells.^{6,48} For example, the increased expression level of NeuNAc on cancer cells is believed to correlate with tumor metastasis.^{35–40} High levels of Tn, sTn, TF, and Le^{a/x} are associated with poor prognosis and decreased survival of cancer patients,^{8,49–52} and are also correlated with the metastatic potential of cancer.⁵³ The serum levels of sLe^x and sLe^a in gastrointestinal, pancreatic,⁵⁴ prostate,⁵⁵ and colorectal⁵⁶ cancer patients were shown to correlate with tumor burden and bad prognosis.⁵⁷ Therefore, TACAs constitute powerful tools as tumor markers for the clinical diagnosis and prognosis of cancer.

TACAs expressed on the human cancer cell surface have also shown their value as suitable targets for immune attack against cancer for both active immunotherapy (using vaccines) and passive immunotherapy (using mAbs). Treatment of cancer using mAbs has shown great promise, and several mAbs are now available in the market.⁵⁸ Although none of these mAbs is directed to carbohydrate structures so far, some preclinical and clinical studies of mAbs against gangliosides have shown very encouraging results.^{59,60} On the other hand, cancer vaccines based on TACAs have also attracted great interest and achieved significant progress, which are discussed in detail here.

6.04.3 TACA-Based Cancer Vaccine Development

6.04.3.1 Immunogenicity of Carbohydrates and Immunotolerance to TACAs

TACAs are anticipated to be good vaccine candidates owing to their abundant expression and exposed location on the cancer cell surface. However, there are many difficulties associated with the development of TACA-based cancer vaccines.²² First, carbohydrates are notoriously poor immunogens.^{61,62} Moreover, the mechanisms by which carbohydrates interact with the immune system are largely unknown. To make carbohydrates immunogenic, they have to be linked covalently to proper immunologically active carriers. More importantly, the patient's immune system is typically tolerant to TACAs, meaning that TACAs are perceived as 'self' or 'normal' antigens and are not recognizable by the human immune system. Thus, although some TACAs can render cancer cells mildly antigenic (i.e., capable of eliciting antibodies), they are rarely immunogenic (i.e., not capable of recruiting immune effectors to kill cells).⁶¹ The exact mechanisms that cause immunotolerance to tumors during the natural history of tumor development have not yet been fully elucidated. The low levels of expression of some TACAs in normal tissues or at a specific stage of development^{63,64} and the structural mimicry of TACAs to normal antigens^{22,65} are at least partially responsible. In addition, even if some TACAs are not 'self' or 'normal' antigens, they are surrounded by 'normal' antigens providing a disguise that reduces the immune response. Another obstacle to the development of TACA-based cancer vaccines arises from the degree of heterogeneity of carbohydrates expressed on the tumor cell surfaces, even within a particular cancer type.^{28,29} Finally, successful tumor immunotherapy might require a T cell-mediated immune response, in addition to a B cell response. However, most TACAs raise a T cell-independent humoral response, so that upon vaccination with TACA-containing constructs, the immune system produces short-lived B cell immunoglobulins IgM with poor memory and without the support typically provided by T cells. All these factors represent obstacles that need to be addressed when designing TACA-based cancer vaccines.

Since immunotolerance is a common and central problem in cancer immunology, many attempts at generating TACA-based cancer vaccines have been focused on improving the immunogenicity of TACAs and breaking the immunotolerance to TACAs.^{12,15,65} For this purpose and for the construction of functional cancer vaccines, much attention has been paid to the selection of proper antigens and vaccine adjuvants, development of procedures for the synthesis of TACAs and their conjugates, exploration of novel vaccine concepts or designs, and chemical and/or enzymatic modification of TACAs to augment their immunogenicity, and so on. Each of these efforts is discussed in detail below.

6.04.3.2 Synthesis of TACAs and TACA Conjugates

It has been well established that covalently linking carbohydrates to a carrier protein may significantly improve their immunogenicity and even convert them from T cell-independent antigens to T cell-dependent antigens.⁹ Thus, for the development of useful carbohydrate-based vaccines, the most widely accepted strategy is to conjugate target carbohydrate antigens to a carrier protein to form glycoproteins. However, TACAs and TACA-containing molecules required for inclusion into a vaccine design are not readily available from natural sources. Even if, in principle, a TACA were available from natural sources, the measures necessary for its isolation, purification, and identification would be very tedious, low-yielding, and, in the end, impractical. Consequently, it is essential to develop efficient methods for generating sufficient quantities of TACAs and TACA conjugates for various investigations in the development of cancer vaccines. Currently, TACAs and TACA conjugates are obtained through three general methods, including chemical synthesis, enzymatic (or chemoenzymatic) synthesis, and bacterial metabolic pathway engineering.

6.04.3.2.1 Chemical synthesis of TACAs and TACA conjugates

Chemical synthesis could theoretically provide well-defined glycoconjugates for vaccine development, and the central and most challenging task in this process is the preparation of TACAs. Fortunately, many methods and strategies have been established in the past few decades for the efficient synthesis of oligosaccharides in carbohydrate chemistry. Since the progress in this field has been the topic of many reviews,^{66–71} it will not be discussed in detail in this chapter. However, it is worth pointing out that many novel synthetic strategies for carbohydrates have significantly facilitated access to TACAs. For example, the one-pot synthetic strategy^{68,72} has been successfully used to prepare globo H,^{73,74} Le^y,⁷⁵ sLe^x,⁷⁶ fucosylGM1,⁷⁷ and other TACAs^{72,78} in good overall yields. There are three general concepts that have been frequently used for one-pot carbohydrate synthesis, that is, (1) chemoselective glycosylation, which exploits the different reactivity of armed and disarmed glycosyl donors and acceptors,^{68,79} (2) orthogonal glycosylation, which is based on the selective activation of a glycosyl donor containing a specific leaving group in the presence of a glycosyl acceptor that bears a different leaving group and can be activated under different conditions for the next step of glycosylation,⁸⁰ and (3) preactivation-based glycosylation, in which the glycosyl donor is activated before the addition of the glycosyl acceptor that contains a leaving group suitable for the next step of glycosylation.^{81,82} Another promising strategy is solid-phase oligosaccharide and glycoconjugate synthesis,^{83–85} which has been used to generate an oligosaccharide library^{86,87} and the synthesis of globo H,⁸⁸ Gb3,⁸⁸ Le^x,^{89,90} Le^y,⁹⁰ Le^x–Le^y,⁹⁰ and GM3.⁹¹ Despite significant developments in solid-phase carbohydrate synthesis in the past two decades,^{83,86,92–94} solution-phase synthesis is still much more widely adopted for the preparation of TACAs and other complex oligosaccharides.

For the assembly of glycoconjugates, there are three main strategies as outlined in **Figure 2**. The first strategy, which is most commonly used for small glycopeptides, involves solid-phase linear assembly of the glycopeptide chain employing glycosylated amino acids as key building blocks (**Figure 2(a)**).⁹⁵ The second strategy is to couple unprotected oligosaccharides directly to an unprotected peptide or protein (**Figure 2(b)**), which can be utilized to obtain very complex glycoconjugates. In this case, to have chemically defined structures, site-selective glycosylation should be applied, which may be achievable via enzymatic methods.^{68,96} The third strategy is based on convergent condensation of unprotected or partially protected peptide or glycopeptide segments by either chemoselective ligation or enzymatic methods (**Figure 2(c)**). The ligation methods useful for this synthetic strategy include native chemical ligation,^{97–100} expressed protein ligation,¹⁰¹ protease-catalyzed ligation,^{102,103} and traceless Staudinger ligation.^{104,105} All these strategies have been extensively exploited, either alone or combined, for the production of glycopeptides and/or glycoproteins.^{106–110}

6.04.3.2.2 In vitro enzymatic synthesis of TACAs and TACA conjugates

Despite great advances in synthetic carbohydrate chemistry in the past few decades, the chemical synthesis of oligosaccharides, including most TACAs, still remains a time-consuming practice, because it involves iterative glycosylation reactions, numerous protection/deprotection steps, and tedious intermediate purification processes. Enzymatic synthesis, on the other hand, provides a way to construct oligosaccharides in a highly stereo- and regioselective manner under neutral aqueous conditions without functional group protection. As a result,

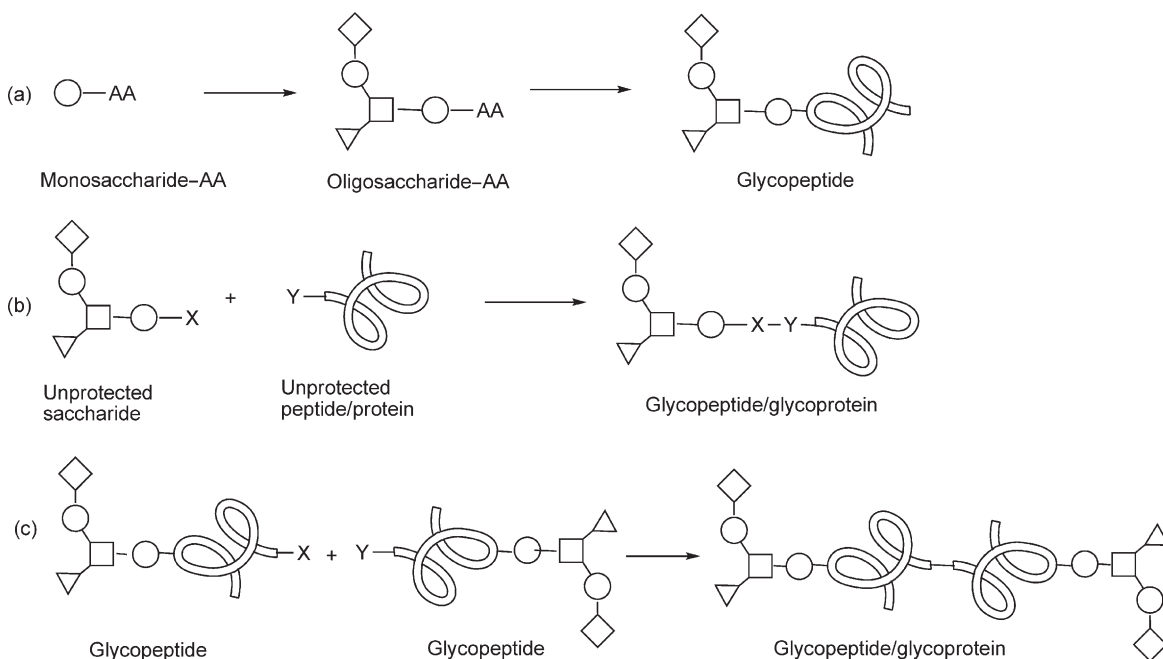


Figure 2 Strategies for the synthesis of glycopeptides and glycoproteins for carbohydrate-based cancer vaccine development. Glycopeptide synthesis by (a) solid-phase assembly using glycosylated amino acids as building blocks; (b) site-selective peptide-glycan coupling reaction; (c) convergent glycopeptides-glycopeptide ligation.

enzymatic carbohydrate synthesis has become a very active area,⁶⁸ especially with the improved access to various enzymes required.

Enzymes employed in carbohydrate and glycoconjugate synthesis are usually glycosidases and glycosyltransferases, and recent progress in molecular biology makes it rather easy to obtain these enzymes. Thus, recombinant glycosyltransferases have been successfully used to prepare a number of TACAs, such as GM3,^{111,112} GM2,¹¹³ GM1,¹¹³ GD1a,¹¹³ Gb3,¹¹⁴ Gb4,¹¹⁵ globo H,¹¹⁶ sTn, and sTF.^{117–119} A recombinant polypeptide, *N*-acetylgalactosaminyltransferase (ppGalNAc-T1), was utilized to synthesize several MUC6-Tn glycoproteins on a semipreparative scale.¹²⁰ These glycoproteins, which displayed a high level of Tn antigens, were recognized by two anti-Tn mAbs that are specific to human cancer cells. The combination of a series of enzymes was employed to construct highly complex TACAs and glycoconjugates. For instance, the enzymatic synthesis of the globo H hexasaccharide has recently been achieved with an overall yield of 57% starting from the initial acceptor Lac-OBn.¹²¹ The entire synthesis was accomplished with three glycosyltransferases: LgtC, an α -1,4-galactosyltransferase; LgtD, a β -1,3-galactosyl/ β -1,3-*N*-acetylgalactosaminyltransferase; and WbsJ, an α -1,2-fucosyltransferase. Similarly, mono- and dendrimeric sLe^x antigens were enzymatically synthesized by several groups.^{122–124} sTF glycopeptides¹²⁵ and GM3 glycosphingolipids¹¹¹ were constructed by using both glycosidases and glycosyltransferases. For sTF glycopeptide synthesis, galactose was first attached to GalNAc-containing peptides via a transglycosylation reaction using a recombinant β -galactosidase from *Bacillus circulans*, and then, NeuNAc was attached with a recombinant sialyltransferase derived from rat liver. Enzymatic strategies were also applied to solid-phase syntheses and one-pot reaction to prepare TACAs such as sLe^x,^{122,126,127} GM2,¹¹³ GD2,¹¹³ GM1,¹¹³ and sTF.^{128,129}

6.04.3.2.3 In vivo metabolic engineering to obtain TACAs and TACA conjugates

Over the past few years, the production of tumor-associated O-linked glycoproteins with more defined glycan structures by genetic engineering of host organisms has been extensively explored.¹³⁰ The use of human tumor cell lines transfected by mucin-encoding plasmids has been shown to be a valid approach, since the O-glycan profile of these recombinant mucins is comparable to the one found in endogenous mucins.¹³¹ Also, after

cotransfection of a wild-type glycosyltransferase, ppGalNAc-T4, CHO-K1 cells produce MUC1 with an increased number of O-glycosylated sites.¹³² Using a Cellferm-pro system and transfected MUC1-CHO K1 cell in protein-free medium, the production of an MUC1 glycoprotein was achieved in a high yield (100 mg day^{-1}).¹³³

To avoid recombinant enzyme production and purification as well as the use of expensive materials, such as CMP and activated sugars in the synthesis, bacterial metabolic pathway engineering has recently emerged as a powerful method for large-scale synthesis of oligosaccharides.⁶⁸ In this case, glycosylation reactions are performed by whole cells that overexpress the genes encoding the appropriate glycosyltransferases and enzymes for sugar-nucleotide biosynthesis. Using this technique, large-scale *in vivo* syntheses of GM3,¹³⁴ GM2,¹³⁵ GM1,¹³⁵ GD3,¹³⁶ Gb4,¹³⁷ and Gb5¹³⁸ in living bacteria have been achieved. For GM3 synthesis, a single metabolically engineered *Escherichia coli* strain devoid of β -galactosidase and NeuNAc aldolase activities but with overexpressed *Neisseria meningitidis* genes for α -2,3-sialyltransferase and CMP-NeuNAc synthase was used. When cultured with this *E. coli* strain, the exogenously added lactose and NeuNAc were actively internalized by the *E. coli* β -galactosidase and NeuNAc permeases and accumulated in the cytoplasm without being degraded. NeuNAc was then converted into CMP-NeuNAc and transferred onto lactose to form GM3 with a maximal yield of 2.6 g l^{-1} .¹³⁴ The *in vivo* syntheses of GM2 and GM1 were achieved by a similar approach but using different *E. coli* strains. The strain used for GM2 synthesis overexpressed additionally the genes for a β -1,4-GalNAc transferase and a UDP-GlcNAc C4 epimerase necessary to provide UDP-GalNAc in the cell. The strain used for the synthesis of GM1 overexpressed one more gene for β -1,3-galactosyltransferase. In high-cell-density cultures, the production yields for GM2 and GM1 were 1.25 g l^{-1} and 0.89 g l^{-1} , respectively.¹³⁵ Meanwhile, using an alternative *Campylobacter jejuni* CstII sialyltransferase, which exhibits both α -3 and α -8 sialyltransferase activities,¹³⁹ the *in vivo* synthesis of GD3 was achieved together with the production of GM3 and GT3. By varying the lactose and NeuNAc initial concentrations and the culture time, the maximal production yields of GD3 (0.83 g l^{-1}) and GT3 (0.91 g l^{-1}) were observed after 9 and 24 h of reaction, respectively.¹³⁶ GM3 was also synthesized by a whole-cell reaction through the combination of recombinant *E. coli* strains and *Corynebacterium ammoniagenes*.¹⁴⁰ First, *C. ammoniagenes* and two *E. coli* strains that overexpressed the genes of CMP-NeuNAc synthetase and CTP synthetase, respectively, were used for the production of CMP-NeuNAc. *C. ammoniagenes* contributed to the formation of UTP from orotic acid. After 27 h of reaction starting with orotic acid and NeuNAc, CMP-NeuNAc was accumulated at a concentration of 17 g l^{-1} . When *E. coli* cells that overexpressed the α -2,3-sialyltransferase gene of *Neisseria gonorrhoeae* were put into the CMP-NeuNAc production system, GM3 was accumulated at 33 g l^{-1} after 11 h of reaction starting with orotic acid, NeuNAc, and lactose. Other than GM3, almost no oligosaccharide by-products were observed from this reaction. The production of GM3 at a 5-l fermenter scale was almost the same as that at a beaker scale, indicating the high potential of GM3 production on an industrial scale. The preparative scale synthesis of sTn through the coupling of recombinant bacterial strains has also been reported.¹⁴¹ In this case, sTn was obtained at 45 g l^{-1} yield after 25 h of reaction starting from orotic acid, NeuNAc, and 2-acetamide-2-deoxy-D-galactose.

GM2 and GM3 derivatives bearing an allyl or a propargyl aglycon were also efficiently biosynthesized on the gram scale by metabolically engineered *E. coli* cells in the presence of the corresponding lactoside acceptors and NeuNAc.¹⁴² GM2 and GM3 bearing the allyl group can be transformed into an aldehyde for further conjugation with carrier proteins by the classical reductive amination method, while the alkyne-containing GM2 and GM3 can be utilized directly to couple with carrier proteins by click chemistry to form cancer vaccines.¹⁴³

6.04.3.3 Immunological Adjuvant Application

To improve the immunogenicity of a carbohydrate vaccine and enhance the duration of the immune response, adjuvants capable of eliciting T cell help and modulating antibody subclass distribution are necessary for cancer vaccine application. An immunological adjuvant is a substance that is itself nonimmunogenic but can significantly augment the immune response when administered together with the antigen. Currently, the only adjuvants acceptable for use in humans are the aluminum-based salts such as alum (aluminum hydroxide) and a squalene-oil-water emulsion (MF59).¹⁴⁴

During the last two decades a number of new adjuvants have been prepared and are in various phases of preclinical and clinical testing.¹⁴⁵ Based on their principal mechanisms of action, these adjuvants can be broadly

classified into two categories: immunostimulatory adjuvants and vaccine delivery systems. Immunostimulatory adjuvants are derived predominantly from pathogens and often represent pathogen-associated molecules, such as lipopolysaccharide (LPS), monophosphoryl lipid A (MPL), and CpG DNA, which activate cells of the innate immune system. In contrast, vaccine delivery systems are generally particles, for example, emulsions, micro-particles, iscoms, and liposomes, and function mainly to target associated antigens into antigen-presenting cells.

For example, MPL, which is derived from LPS of gram-negative bacteria, has been used extensively in clinical trials as a component in prophylactic and therapeutic vaccines targeting infectious disease, cancer, and allergies.¹⁴⁵ MPL has been formulated into emulsions to enhance its potency. The Ribi adjuvant system (MPL plus trehalose dimycolate) has emerged as a leading adjuvant with an enhanced efficacy and safety profile in both preclinical and clinical studies.¹⁴⁶ It has also been observed that helper T cell (Th)-dependent response was induced in mice when malaria parasites were emulsified in Ribi adjuvant.¹⁴⁷ QS-21, a plant-derived complex saponin from the South American tree *Quillaja saponaria* Molina cortex, is another immunological adjuvant used in antitumor vaccines.¹⁴⁸ Microgram quantities of this amphiphilic substance in combination with the antigen-carrier conjugates lead to enhancement in both antibody and cell-mediated immune response in a host of promising anticancer and antiviral vaccines.^{149,150} Similarly, tripalmitoyl-*S*-glyceryl-cysteinylserine (Pam3Cys), a lipopeptide derived from the immunologically active N-terminal sequence of an *E. coli* lipoprotein, has been widely used as a carrier and potent adjuvant for vaccines.^{151–155} Recent studies have shown that Pam3Cys exerts its activities through the interaction with Toll-like receptor 2 (TLR-2).¹⁵⁶ Pam3Cys is nonimmunogenic and has no toxic side effects nor does it cause tissue damage in animals. In addition, its adjuvanticity is comparable to the classical Freund's adjuvant. With this molecule, high-titer antibodies against HIV protein gp120 were obtained without the use of external adjuvants.¹⁵⁷

Liposomes are another type of adjuvant for cancer vaccine application. Liposome vesicles may be effectively used to deliver soluble TACAs to macrophages and other antigen-presenting cells through various routes of inoculation. The strong lipophilic property of liposomes enables them to be adsorbed in large numbers onto the surface of macrophages. This process can augment the immune response by mechanisms involving macrophage enlisting of Th.¹⁵⁸ **Table 2** summarizes some adjuvants capable of enhancing immune response to carbohydrate-based cancer vaccines.

6.04.3.4 Semisynthetic Glycoconjugate Cancer Vaccines

Due to the poor immunogenicity of TACAs, they must be coupled to appropriate carriers, usually proteins, to become potentially functional vaccines. The vaccines thus formed are called semisynthetic vaccines and have been intensively studied. The Livingston and Danishefsky groups found that keyhole limpet hemocyanin (KLH) is the most effective immunogenic carrier for TACAs, as compared to a variety of other proteins.^{159,160} Moreover, it was found that covalent coupling of TACAs to KLH is much more effective than simply mixing them together.^{159,161} Hence, after TACAs are obtained, they are usually conjugated to KLH for vaccine development. According to the number and type of TACAs attached to a carrier protein, conjugate vaccines can be classified into three main categories including monovalent conjugates, which contain a single type of TACA, monovalent cluster conjugates, which contain one type of TACA presented in clusters, and unimolecular multivalent conjugates, which contain several types of TACAs.

Table 2 Selected adjuvants capable of enhancing immune responses to TACAs

| Adjuvant classes | Adjuvants |
|-----------------------------|--|
| Mineral salts | Aluminum hydroxide, aluminum phosphate, calcium phosphate |
| Immunostimulatory adjuvants | Cytokines (e.g., IL-2, IL-12, GM-CSF), saponins (e.g., QS-21), MPL derivatives, bacterial DNA, MDP derivatives |
| Lipid particles | Emulsions (e.g., Freund's, SAF, MF59), liposomes, iscoms, virosomes, lipopeptide |
| Particulate adjuvants | Poloxamer particles, virus-like particles |
| Mucosal adjuvants | Cholera toxin (CT), mutant toxins (e.g., LTK63, LTR72) |

As already reviewed by Jennings and Sood⁹ and Pozsgay,¹⁶² many coupling methods have been developed for carbohydrate–protein conjugation, and the most generally accepted conjugation strategy is to modify the carbohydrate antigen to contain a reactive functionality that will then selectively react with either of the free amino, thio, and carboxylic groups of the carrier protein. In principle, all these methods can be utilized to couple TACAs and carrier proteins, and **Figure 3** depicts some common methods that have been adopted to prepare glycoconjugate vaccines discussed in this chapter. One of the most common conjugation methods makes use of the reductive amination reaction (**Figure 3(a)**). In this case, the carbohydrate antigen is usually modified to contain an aldehyde functionality that can react with the free amino groups (usually the side chains of lysine residues) of the carrier protein in the presence of NaBH₃CN to create a stable amino linkage between the carbohydrate antigen and carrier protein. Protein–carbohydrate conjugation via a bifunctional acyl group, which has one end linked to the carbohydrate antigen and the other end to the carrier protein via an amide linkage as shown in **Figure 3(b)**, is another commonly used method. The conjugation can also be effectively achieved by selective reactions between the thiol groups of the cysteine residues of the carrier protein and a bromoacetylated carbohydrate antigen (**Figure 3(c)**) or a carbohydrate antigen that is modified to contain an activated C=C double bond (**Figure 3(d)**). Coupling the carboxyl group of the carrier protein with functional groups such as hydrazide provides alternative methods for carbohydrate–protein conjugation (**Figure 3(e)**).

It is worth noting that the linker between the carbohydrate antigen and the carrier protein may have a significant influence on the immunological properties of the resulting conjugate vaccine. On one hand, the linker itself may induce a strong immune response. On the other hand, some linkers suppress the immune response to the carbohydrate antigen.^{163,164} Consequently, in the design of conjugate cancer vaccines, it should be helpful to use linkers that are similar to the structural motifs of natural proteins or carbohydrates, so that the linkers would not affect the immunogenicity of the carbohydrate antigen.¹⁶⁵

6.04.3.4.1 Semisynthetic monovalent glycoconjugate vaccines

Among various monovalent TACA–KLH conjugate vaccines explored so far, GM2–KLH (**Figure 4**) is one of the two vaccines that reached phase III clinical trials. GM2 was first identified as a potential target for cancer immunotherapy when IgM antibodies against GM2 were associated with a prolonged disease-free interval and an improved survival of melanoma patients.^{166,167} Moreover, tumor regression was observed in melanoma patients after treatment with GM2-specific mAbs.¹⁶⁸ Based on these clinical observations, Livingston and coworkers synthesized the covalently linked GM2–KLH conjugate and evaluated its antitumor efficiency in mice and melanoma patients either alone or with immunological adjuvants Bacillus Calmette–Guerin, Detox, or QS-21.^{13,149,169} It was found that GM2–KLH plus QS-21 not only induced the highest IgM titers, but also induced durable IgG antibodies in most patients.^{149,170} The IgG antibodies were of the IgG1 and IgG3 isotypes and were able to induce complement-mediated lysis of GM2-positive tumor cells and, in most cases, antibody-dependent cell-mediated cytotoxicity (ADCC).^{170–172} Based on these successful results, a randomized phase III clinical trial was initiated in patients with resected stage IIB–III melanoma. However, when coadministered with QS-21, GM2–KLH conjugate treatment was found to be no more effective than interferon- α 2b, the current standard therapy for high-risk melanoma.¹⁷³

Encouraged by the results of the GM2–KLH vaccine, some other ganglioside–KLH conjugates such as GM3–KLH,¹⁷⁴ GD2–KLH,¹⁷⁵ GD3–KLH,¹⁵⁰ and fucosyl GM1–KLH were prepared and applied to clinical trials as well. Fucosyl GM1 is the major ganglioside component contained in human small cell lung carcinoma (SCLC) tissue, expressed more frequently and more abundantly than either GM2 or GD3.¹⁷⁶ The immunohistochemistry studies using the F12 mAb indicated that the distribution of fucosyl GM1 in normal tissues is highly restricted and virtually nonexistent in any human cancer cell line other than SCLC,²⁸ making it an attractive vaccine target. Early phase I clinical trials have demonstrated that vaccination with fucosyl GM1–KLH induced an IgM antibody response against fucosyl GM1 and tumor cells expressing fucosyl GM1.¹⁷⁷

Globo H is a human breast cancer-associated hexasaccharide antigen identified and defined by murine mAb MBr1 generated by immunization of mice with human breast cancer cells MCF-7.^{178,179} Globo H has been synthesized using the glycal assembly method and then conjugated to KLH and investigated in a number of different tumors (**Figure 4**).^{180,181} Studies in mice showed that the globo H–KLH conjugate induced high titer IgM and IgG responses against globo H antigen.¹⁸² Furthermore, the antibodies reacted with globo H positive MCF-7 cells, but not globo H negative B78.2 cells, and were able to effectively induce complement-mediated

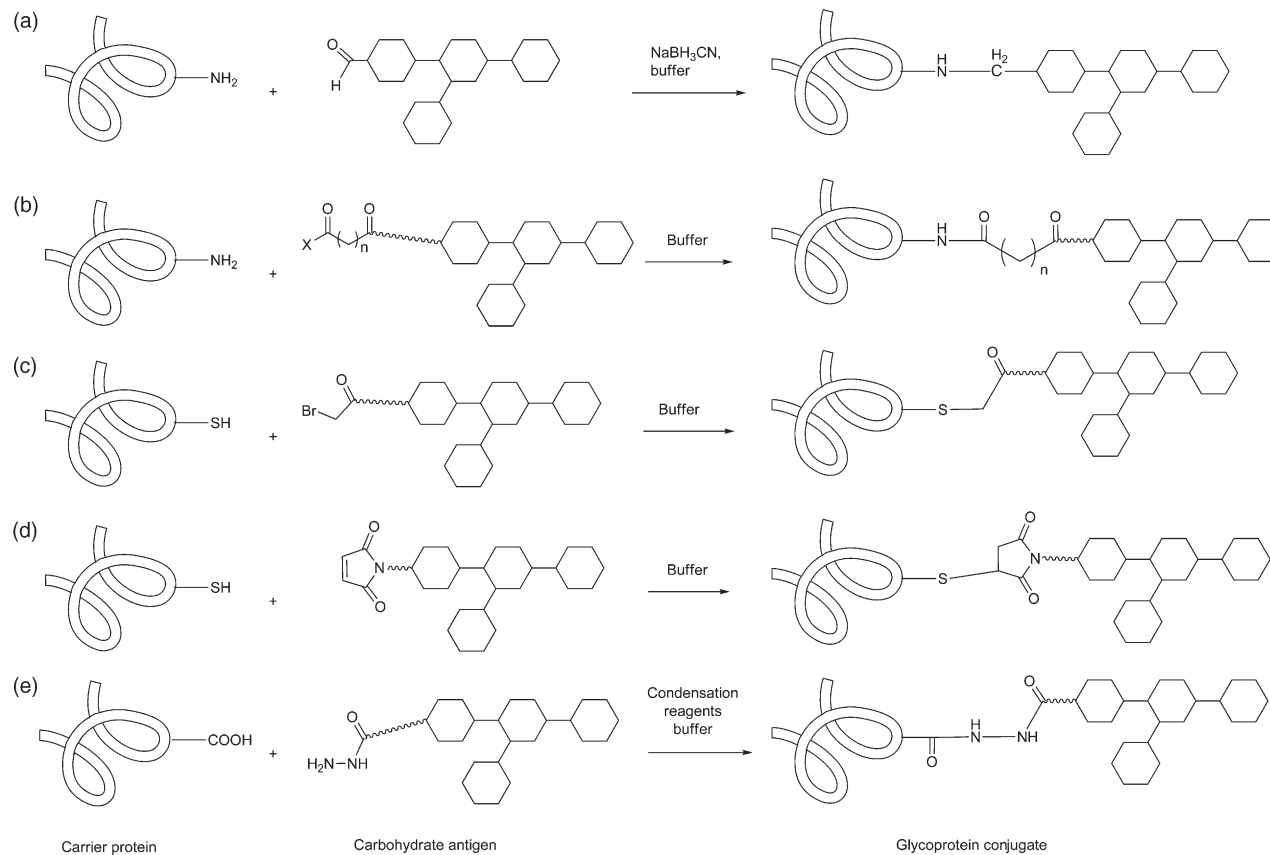


Figure 3 Representative conjugation methods for the preparation of semisynthetic conjugate vaccines. Protein-carbohydrate conjugation via: (a) reductive amination; (b) a bifunctional acyl linker; (c) reaction of protein thiol groups with bromoacetylated carbohydrates; (d) reaction of protein thiol groups with maleimide-modified carbohydrates; (e) reaction of protein carboxyl groups with hydrazide-modified carbohydrates.

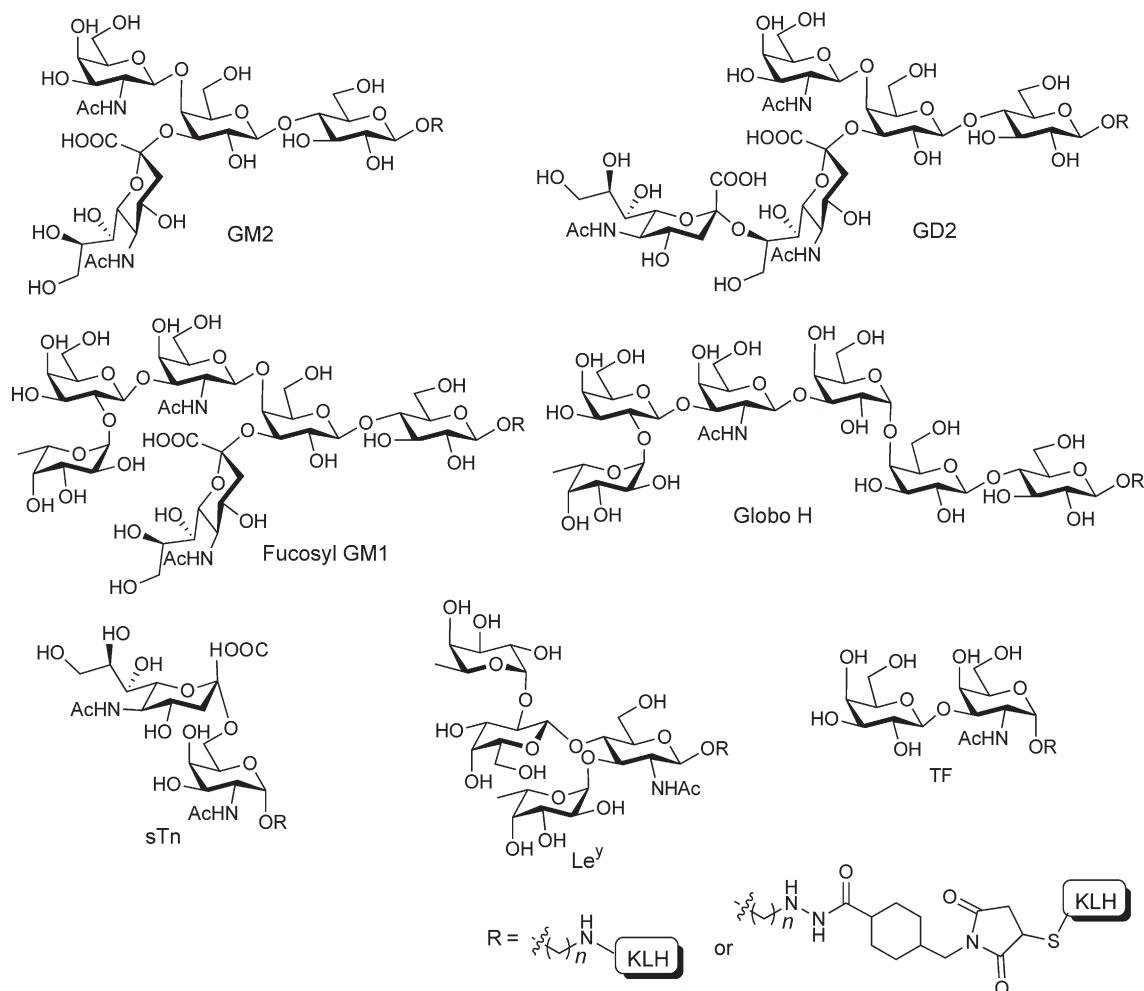


Figure 4 Structures of some semisynthetic monovalent TACA–KLH vaccines.

cytotoxicity (48% lysis). On the basis of these results, globo H–KLH vaccine was advanced to a full phase I trial in patients with progressive and recurring prostate cancers.^{181,183} All immunized patients exhibited good IgM responses against globo H, with the antibodies able to recognize globo H-expressing cell lines. In some cases, they were able to induce complement-mediated lysis. Another phase I trial was also initiated in patients with metastatic breast cancer without evidence of disease or with stable disease on hormone therapy.¹⁸⁰ Again, vaccination was able to stimulate the production of IgM antibodies in a majority of patients. Significant binding of the antibodies to MCF-7 was observed in 60% of the patients.

sTn–KLH (named Theratope by Biomira Inc., **Figure 4**) was another semisynthetic monovalent vaccine that entered phase III clinical trials. sTn is a mucin-associated carbohydrate antigen. It has attracted significant interest as a tumor antigen for several reasons. First, sTn is relatively tumor specific. Although sTn antigen is richly expressed on a number of tumors, it is rarely observed on normal tissues. Second, the expression of sTn antigen has been identified as an independent indicator for poor prognosis of cancer patients.^{50,184,185} Third, sTn appears to be associated with increased aggressiveness and metastatic potential of cancer,^{50,185} particularly in breast cancer.¹⁸⁶ A variety of clinical trials using sTn–KLH vaccines have been tested.^{187,188} Immunization of breast cancer patients with sTn–KLH plus Detox in combination with a low dose of cyclophosphamide induced high titers of IgM and IgG antibodies against synthetic sTn in most cases, but the titers were two- to four-fold lower in patients not pretreated with cyclophosphamide. IgM and IgG reactivity against ovine submaxillary mucin (OSM) was low. In another trial, patients with colon cancer were immunized with sTn–KLH plus Detox

or QS-21.¹⁸⁹ Similarly, sTn–KLH plus Detox and especially QS-21 induced high IgM and IgG antibody titers against sTn. Furthermore, when tested against tumor cells or natural mucins expressing sTn, IgM antibodies showed weak to moderate reactivity, while IgG antibodies were almost totally unreactive. In the phase III clinical trial, more than 1000 women with distant metastatic breast cancer were enrolled, but Theratope failed to meet the endpoints of time-to-disease progression and overall survival. However, patients treated with Theratope in conjunction with hormone therapy had improved survival rates with a time-to-disease progression of 8.3 months compared with 5.8 months for those on hormone therapy alone. This modest clinical efficacy could be due to the fact that Theratope elicits a B cell-mediated immune response but does not seem to trigger a T cell-mediated immune response.¹⁸⁶

Other semisynthetic monovalent vaccines such as Tn–KLH, TF–KLH, Le^y–KLH, and GD2-lactone–KLH were also prepared and used for the treatment of a variety of cancers. Their immune responses and therapeutic effects are briefly summarized in **Table 3**.

6.04.3.4.2 Semisynthetic monovalent clustered glycoconjugate vaccines

Although much progress has been made in the development of TACA-based monovalent vaccines for cancer, it was found that sometimes, especially for vaccines made of mucin-related TACAs, the provoked antibodies against the synthetic carbohydrate epitopes did not react as well with purified mucins or cancer cells.¹⁸⁹ In addition, findings in the field of glycohistology have demonstrated that mucins overexpressed on tumor cell surfaces often present clusters of two to five adjacent carbohydrate domains.²⁰¹ For a robust and efficient immune response to be generated, constructs with multiple repeats or clustered carbohydrate epitopes were required.²⁰² Therefore, scientists turned their attention to the development of cluster vaccines with the hope that these vaccines would better mimic the surface of targeted tumor cells.

Mucins comprise a family of large glycoproteins expressed on the surface of epithelial tissues. The amino acid sequence of mucins possesses a relatively high percentage of serine and threonine residues, often arranged in continuous arrays ranging from two to five. Despite the diversity found in mucin glycostructures, the appearance of an *N*-acetylgalactosamine moiety (GalNAc) α -linked to the serine/threonine residue appears to be highly conserved. Based on the properties described above, monovalent cluster vaccines are usually constructed by exposing carbohydrate epitopes on peptide backbone of serine and/or threonine.

Although successful methods have been developed for the formation of glycosylated amino acids, in many cases coupling reactions involving rather bulky glycosylated moieties can be low-yielding. In order to partially overcome this problem, a ‘cassette’ approach^{15,108} has been used. In the cassette strategy, a monosaccharide synthon is O-linked to a serine, a threonine, or a hydroxynorleucine residue with a differentiable acceptor site on the sugar. This construct serves as a general insert (cassette) that is joined to a target saccharide bearing a glycosyl donor function at its reducing end. In this way, the need for direct coupling of the serine side chain hydroxyl group to a fully elaborated, already complex saccharide donor is avoided. Using the cassette methodology, several trimeric clustered TACAs epitopes, including Tn(c),²⁰³ TF(c),²⁰³ sTn(c),²⁰⁴ 2,6-sTF(c),²⁰⁵ Le^y(c),^{109,206} and most recently Gb3(c)²⁰⁷ were accomplished. These clustered epitopes were then covalently linked to KLH using a heterobifunctional linker, *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS), providing the desired clustered constructs (**Figure 5**).

When Tn(c)–MBS–KLH was coadministered with QS-21 to mice, both IgM and IgG antibodies were generated after three immunizations.^{161,203} It was demonstrated that the sera obtained from the vaccinated mice showed clear IgM and IgG reactivities by flow cytometry and complement-dependent cytotoxicity (CDC) assays, evaluated using Tn(c) positive LS-C and Tn(c) negative LS-B colon cancer cells. A phase I clinical trial in prostate cancer patients showed that all patients immunized with Tn(c)–MBS–KLH/QS-21 developed high-titer antibody responses against Tn(c), confirming the immunogenicity of this Tn(c) vaccine. In addition, the levels of PSA observed in the treated patients either stabilized or declined, of which the clinical impact and relevance remains to be validated.¹⁹⁰

The TF(c)–MBS–KLH/QS-21 combination as cancer vaccine was also evaluated in phase I clinical trial in patients with biochemically relapsed prostate cancer.¹⁹² Four groups of five patients were treated with escalating doses of 1, 3, 10, and 30 μ g of TF(c)–MBS–KLH/QS-21. All doses induced high titers of IgM and IgG antibodies against TF(c). Interestingly, it was found that higher titers of antibodies developed at the lowest dose level (1 μ g), which was different from other previous dose-escalating phase I trials. An antitumor effect in the form of a change in posttreatment versus pretreatment log PSA slopes was also observed.

Table 3 TACA-based semisynthetic vaccines in clinical trials

| <i>Vaccine type</i> | <i>Vaccine</i> | <i>Cancer</i> | <i>Immune response / therapeutic effect</i> | <i>Reference(s)</i> | |
|-----------------------------|---------------------------|---|--|---|--|
| Semisynthetic monovalent | Tn-KLH | Prostate | IgM + IgG that (slightly recognize LSC cells); Decrease of PSA progression | 190 | |
| | TF-KLH | Ovary (metastatic stage) | IgM + IgG (recognize human tumor cells), CDC | 191 | |
| | | Prostate | mostly IgM (do not recognize human tumor cells); Decrease of PSA progression | 192 | |
| | sTn-KLH | Colorectum (metastatic stage) | IgM + IgG (do not recognize native TACA) | www.biomira.com 189 | |
| | | Breast, colorectum | Ab level correlate with improved survival | 193 | |
| | | Ovarian | IgM + IgG; Increased pre-ASI CA-125 serum levels were predictors of poor survival | 193 | |
| | Le ^y -KLH | GD3-KLH GD3-Lactone-KLH | Ovarian (stage I-IV) | Improved survival (chemotherapy + hormone therapy); TDP and OS endpoints not meet | www.biomira.com |
| | | | Melanoma (stage III, IV) | Mostly IgM (recognize human tumor cells), CDC | 194 |
| | | GD3-Lactone-KLH + BEC2 anti-idiotypic mAb | Melanoma (stage III, IV) | IgM + IgG | 150,195 |
| | | | Melanoma (stage III) | IgM + IgG; No correlation between Ab response and clinical outcome | 196 |
| | | GM2-KLH | Melanoma (stage III) | Not better than interferon- α ; DFS endpoints not meet | 173,197 |
| | | GD2-KLH GD2-Lactone-KLH | Melanoma (stage III, IV) | IgM + IgG (recognize native TACA), CDC | 175 |
| | Semisynthetic multivalent | Fuc-GM1-KLH | Small lung cell | IgM + IgG (recognize native TACA), CDC | 177,198 |
| | | | Prostate | mostly IgM (recognize native TACA); CDC; Decrease of PSA progression | 181 |
| Globo H-KLH | | Breast | mostly IgM (recognize native TACA); CDC | 180 | |
| | | Melanoma (stage III, IV) | IgM + IgG | 199 | |
| (MUC2-Tn)-KLH + Globo H-KLH | | Prostate | IgM + IgG; CDC | 200 | |
| | | Prostate | (MUC1-Tn)-KLH + Globo H-KLH + GM2-KLH + Tn(c)-KLH + TF(c)-KLH + Le ^y -KLH | Lower Ab titers compared with monovalent vaccine | 13 |

Ab, antibody; CDC, complement-dependent cytotoxicity; DFS, disease-free survival; OS, overall survival; PSA, prostate-specific antigen; TDP, time-to-disease progression.

Gb3–MUC5AC(c)–MBS–KLH (**Figure 5**), a vaccine consisting of clusters of both Gb3 antigen and MUC5AC peptide marker overexpressed by ovarian cancer cells,^{208–210} has recently been synthesized by Danishefsky and coworkers²⁰⁷ to target ovarian cancer. Structurally, MUC5AC consists of tandem repeats of an 8-amino acid sequence (TTSTTSAP), which is potentially responsible for the activation of T cells. On the basis of these observations, Gb3–MUC5AC(c)–MBS–KLH was designed to mimic the molecular architecture on ovarian cancer cell surfaces. Gb3–MUC5AC(c)–MBS–KLH thus designed was envisioned to have several potential advantages. First, a mucin-derived peptide fragment was incorporated as both a linker and a marker, which may behave not only as a B cell epitope for the production of antibodies against mucins, but also as a Th epitope to activate T cells. Furthermore, the tandem repeats of both Gb3 and MUC5Ac epitopes are anticipated to expose these B cell and Th epitopes to the maximum extent on the surface of KLH. Finally, vaccines composed of numerous carbohydrate antigens, such as Gb3, associated with a specific cancer type (ovarian cancer in this case) may provide a heightened and more varied response, thereby increasing the efficiency of binding to the target cells. The efficient synthesis of Gb3–MUC5AC(c)–MBS–KLH was enabled by the preparation of a Gb3–MUC5AC thioester cassette as a key building block for constructing three alternating repeats of Gb3 and MUC5AC, and then conjugating through MBS to KLH. The immunological evaluations of Gb3–MUC5AC(c)–MBS–KLH are currently executed.

6.04.3.4.3 Semisynthetic multivalent glycoconjugate vaccines

It has been well established that cancer cells can display several different TACAs on their surfaces and that during each stage of cellular development, a considerable amount of variation in the level and nature of antigenic expression may occur.^{28,29} Accordingly, a monovalent vaccine, whether clustered or not, may not be sufficient for targeting a population of transformed cells. On the other hand, a multivalent vaccine that targets several TACAs should, in principle, lead to a stronger and more specific immune response than a monovalent vaccine that targets a single TACA.^{107,211} In general, one could imagine two approaches for the construction of multivalent vaccines. The first one, termed polyvalent monomeric, involves the administration of a mixture of monovalent conjugate vaccines. The second one relies on the construction of a unimolecular multivalent vaccine consisting of multiple TACAs displayed on a single molecule.

Preclinical trials using polyvalent monomeric vaccines have demonstrated the viability of the first method (**Table 3**).^{212,213} For example, coadministration of GD3–KLH, Le^y–KLH, and the two peptidic-antigen conjugates MUC1–KLH and MUC2–KLH, along with QS-21 produced high titers of IgM and IgG antibodies that reacted specifically with the representative antigens in tumor cell lines expressing them regardless of the mixing methods and the injection sites.²¹³ Preclinical trials using a heptavalent–KLH conjugate vaccine containing the seven epithelial cancer antigens GM2, globo H, Le^y, TF(c), Tn(c), sTn(c), and glycosylated MUC1 gave similar results.²¹² On the basis of these results, a phase II clinical trial featuring three to seven individual TACA–KLH conjugates (all of which have been proven safe and effective in phase I trials) has been initiated in breast, ovarian, and prostate cancer patients. However, this approach suffers from several serious limitations. First, the polyvalent monomeric strategy requires the use of increased amounts of carrier proteins. In addition, a low-yielding final conjugation step for each monovalent–KLH vaccine is involved. Finally, it is necessary to validate each individual component of the vaccine mixture.

The problems associated with the polyvalent monomeric approach can be overcome by the construction of unimolecular multivalent vaccines. By combining three different TACAs linked to a peptide backbone through natural or nonnatural amino acids, two unimolecular multivalent vaccines were initially formed (**Figure 6**). One contained the Tn, TF, and Le^y antigens linked to a peptide backbone through serine hydroxy groups,²¹⁴ and the other consisted of the Tn, Le^y, and globo H antigens attached to the peptide backbone through the hydroxy group of the nonnatural amino acid hydroxynorleucine.²¹⁵ Preliminary studies indicated that both vaccines induced IgM and IgG antibodies against each TACA, with the unnatural globo H–Le^y–Tn vaccine more antigenic than the mucin-based TF–Le^y–Tn vaccine.¹⁰⁷

Prostate cancer has been revealed to express an abundance of blood group-related antigens such as Tn, sTn, TF, and Le^y, and ganglioside TACAs GM2 and globo H. To elicit an immune response directed specifically to prostate cancer, a unimolecular pentavalent vaccine that includes globo H, Le^y, sTn, Tn, and TF has been developed. The antigens were displayed on a peptide backbone, which was conjugated to KLH via a linker.²¹⁶ *In vitro* preclinical serological studies of this pentavalent vaccine, including enzyme-linked immunosorbent

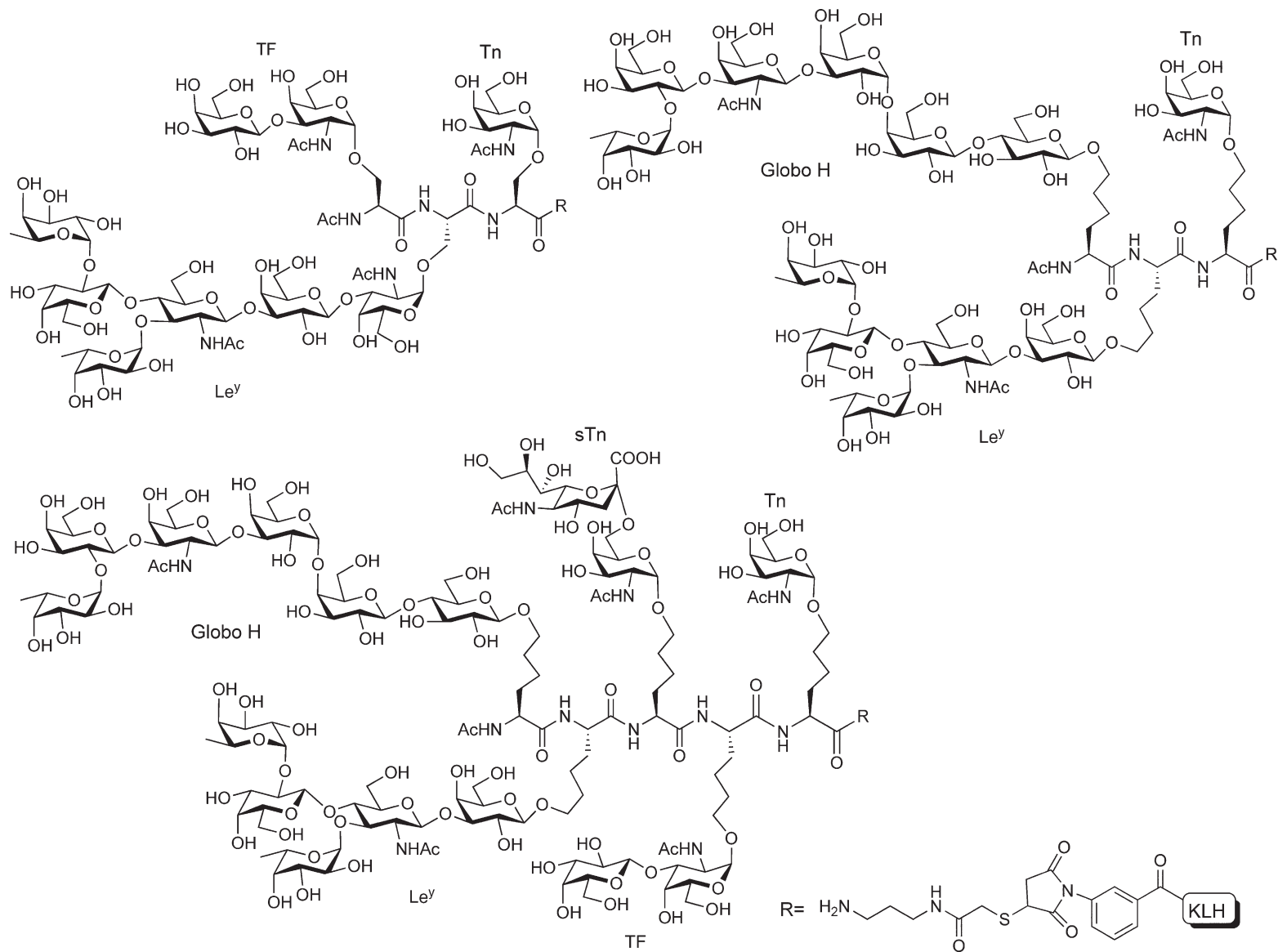


Figure 6 Structures of some semisynthetic unimolecular multivalent TACA-KLH vaccines.

assay (ELISA) and fluorescence activated cell sorting (FACS) analysis, were conducted. The cumulative data suggested that the immunological properties of the individual antigens are preserved in this highly elaborated vaccine. A hexavalent vaccine containing the Tn, TF, sTn, Le^y, globo H, and GM2 TACAs was also synthesized.²¹⁷ These vaccines constitute strong cases for evaluating the concept of the unimolecular multivalent semisynthetic vaccines.

Table 3 has also summarized some TACA-based semisynthetic multivalent conjugate vaccines in clinical stages, which are mainly polyvalent monomeric vaccines, as well as their immune responses and therapeutic effects in certain cancers. No semisynthetic unimolecular multivalent vaccine has reached clinical trials so far.

6.04.3.5 Fully Synthetic Glycoconjugate Vaccines

Although successful, semisynthetic vaccines have major limitations regarding their usage in humans.²¹⁸ These include (1) the ambiguity of the protein carrier and adjuvants in both composition and structure, (2) variable hapten density, and (3) irrelevant Abs production against the carrier protein, which often induces epitopic suppression.²¹⁹ On the other hand, fully synthetic vaccines have homogeneous and well-defined structures, a key quality for consistent studies and safe clinical evaluation. In addition, in fully synthetic vaccines, the adjuvant function can be incorporated directly in the molecule, which would avoid the use of external adjuvants. Hence, fully synthetic vaccines are of great interest in anticancer immunotherapy, and some have been tested in preclinical studies. Generally, fully synthetic vaccines are constructed by covalently linking TACAs to a CD4⁺ Th peptide epitope, and/or a CD8⁺ cytotoxic T lymphocytes (CTLs) peptide epitope, and/or an immunostimulant.

6.04.3.5.1 Two-component fully synthetic glycoconjugate vaccines

Following Tam's pioneering work on synthetic peptide vaccines with a built-in adjuvant against HIV,¹⁵⁷ Toyokuni *et al.*^{220,221} synthesized a fully synthetic Tn vaccine (**Figure 7**), in which dimeric Tn epitopes were conjugated to the Pam3Cys moiety through a spacer. It was demonstrated that this di-Tn-Pam3Cys conjugate could elicit not only high IgM but also significant anti-Tn IgG responses in mice, indicating Th cell enlistment. This vaccine was the first example demonstrating that a synthetic, small carbohydrate antigen can generate an immune response without the use of any protein carrier or additional adjuvant. A very similar tri-Tn-Pam3Cys construct was also synthesized and compared with the same trimeric (clustered) Tn epitope linked to KLH, namely Tn(c)-KLH.¹⁵ Tri-Tn-Pam3Cys alone, or in conjunction with QS-21, induced anti-Tn IgM and IgG responses in mice, although less efficiently than Tn(c)-KLH. In a phase I clinical trial in patients with biochemically relapsed prostate cancer, tri-Tn-Pam3Cys showed inferior antibody responses compared to Tn(c)-KLH.¹⁹⁰

Some fully synthetic Tn vaccines are linear and dendrimeric glycopeptides, which allows the collaboration of Tn antigen and a CD4⁺ T cell epitope-containing peptide for the stimulation of both B and T cells, required for efficient immune response. For example, the immunological studies of several linear glycopeptides containing mono-, tri-, or hexa-Tn antigen as a B cell epitope and a CD4⁺ T cell epitope of poliovirus (PV) demonstrated that glycopeptides Tn3-PV and Tn6-PV induced high titers of anti-Tn IgG antibodies in mice, with the former higher than the latter.²²² However, as revealed by both ELISA and FACS analysis, recognition of the native Tn antigen on Jurkat cells was better with Tn6-PV-induced antibodies than with Tn3-PV-induced antibodies (for both IgM and IgG). In addition, the position of the tri-Tn motif in the peptide sequence and the peptide backbone itself do not alter its antigenicity. These results indicate that short synthetic glycopeptides are able to induce anticancer antibody responses.

To further improve the antigenicity of the Tn-based glycopeptides, dendrimeric multiple antigenic O-linked glycopeptides (MAGs) carrying Tn antigen associated with PV T cell epitope (MAG:Tn-PV and MAG:Tn3-PV) were synthesized (**Figure 7**).^{223,224} The concept of MAG was inspired by Tam's multiple antigen peptide (MAP) in which the peptide backbone containing a T cell epitope was linked to each arm of the dendrimeric lysine core.^{225,226} It was shown that both MAG:Tn-PV and MAG:Tn3-PV were able to induce anti-Tn IgG antibodies that recognize human tumor cells expressing the Tn antigen. When used in active immunotherapy, the MAG:Tn3-PV was much more efficient than the MAG:Tn-PV in promoting the survival of tumor-bearing mice and resistance to tumor challenge. Furthermore, linear Tn6-PV was shown to be less efficient than MAG:Tn3-PV. MAG:Tn3-PV also induced a higher anti-Tn immune response than the

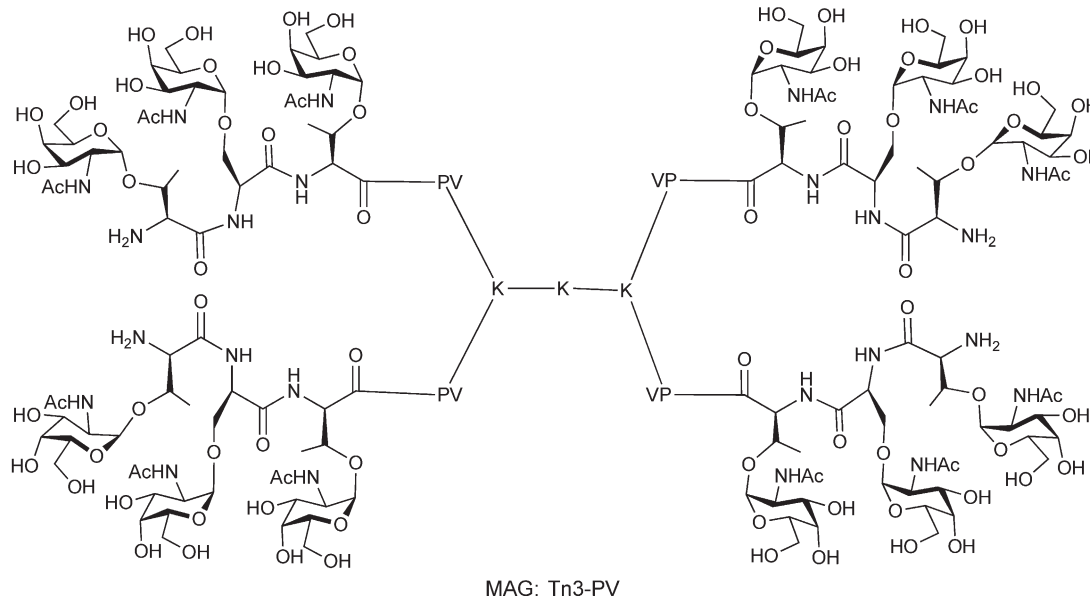
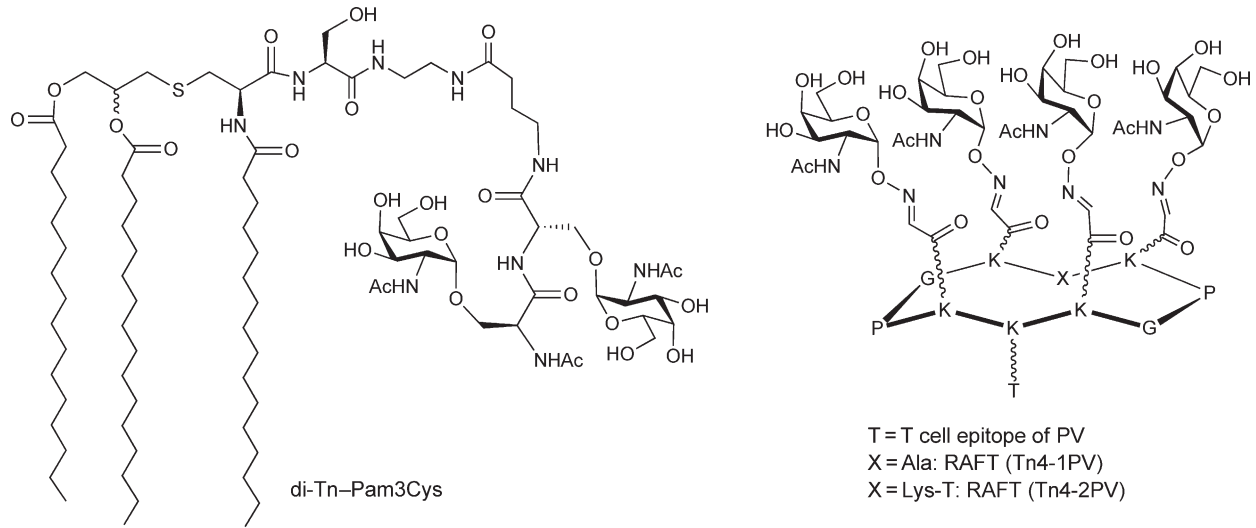


Figure 7 Representative structures of two-component fully synthetic vaccines.

Tn3–KLH conjugate.²²⁷ Therefore, both the clustering and the presentation form of Tn seem to be important parameters for stimulating efficient anti-Tn immune responses.

To apply the MAG system to human vaccination, two HLA-restricted T cell epitopes were chosen, one being a tetanus toxin (TT)-derived peptide²²⁸ and another a nonnatural engineered Th epitope, the Pan-HLA-DR-binding epitope (PADRE).²²⁹ The immunological evaluation in nonhuman primates (macaques and green monkeys) of MAG:Tn3-TT and MAG:Tn(S)3-PADRE showed that these two dendrimeric MAGs induced in all the animals strong anti-Tn IgG antibodies capable of specifically recognizing Tn-expressing human tumor cells. Moreover, these antibodies were able to mediate CDC or ADCC against Tn-positive human tumor cells.²²⁷

Dumy and coworkers also investigated the regioselectively addressable functionalized template (RAFT)-based multi-epitopic glycopeptide vaccines, which contained four copies of monomeric Tn antigen analog (Figure 7).²³⁰ RAFTs are topological templates composed of a cyclic decapeptide containing two Pro–Gly units as β -turn inducers that create conformational rigidity in solution.²³¹ The six lysine residues composing the β -strand provide six attachment points through their side chains in two independent functional faces, one face dedicated to the attachment of four Tn epitopes and the other to PV T cell peptide epitopes. Via a convergent chemoselective assembly, two RAFT vaccines, RAFT-Tn4-1PV and RAFT-Tn4-2PV, were formed. B- and T-antigenicity and immunogenicity of the two vaccines were investigated *in vitro* and *in vivo*. The studies clearly demonstrated that the carbohydrate part of the vaccines is recognized by Tn-specific mAbs 6E11 (IgG) and 83D4 (IgM). Moreover, like MAG systems, RAFT vaccines bearing Tn and PV epitopes are able to elicit an immune response in mice specifically directed against the native form of the Tn epitope expressed on human tumor cells. Altogether, the preliminary studies clearly demonstrated a promising potential of RAFT as a nonimmunogenic built-in vaccine carrier, which opens interesting perspectives to employing the RAFT scaffold for the development of TACA-based cancer vaccines.

Kunz²³² described a two-component vaccine composed of an MUC1 glycopeptide containing an sTn moiety and a TT-derived Th peptide epitope. This vaccine was found to induce proliferation of CD3⁺ cells.²³² Vaccines consisting of the mono-, di-, and tri-glycosylated MUC1 tandem repeat peptides and ovalbumin (OVA_{323–339}) T cell epitopes were also prepared.²³³ When administered to mice, both the monoglycosylated vaccine containing a single sTn antigen and the diglycosylated one having one sTn and one Tn antigen gave highly specific humoral immune responses. The poor immunogenicity of the triglycosylated vaccine with additional Tn antigen at the immunodominant PDTRP motif suggests that not only the saccharide but also the peptide backbone are important for the epitope recognition.

Schmidt and coworkers designed a fully synthetic GM2 conjugate in which GM2 was linked through a spacer to an immunostimulant (B cell stimulatory glycolipid BAY R100G) as shown in Figure 8.²³⁴ This neoglycoconjugate showed reactivities with several GM2 reactive antibodies as assessed by ELISA and immune thin-layer chromatography (ITLC). Vaccination of rabbits with this conjugate resulted in induction of antibodies against GM2, thus confirming the viability of this novel concept for the construction of a totally synthetic vaccine. Most recently, another fully synthetic GM2 vaccine (GM2-PV) was synthesized by efficient ligation of the alkyne-functionalized biosynthesized GM2¹⁴² with an azido CD4⁺ T cell epitope peptide using the Huisgen cycloaddition (Figure 8).¹⁴³ This glycopeptide can induce human tumor cell-specific antibodies after immunization in mice.

Using Le^y as a model, Kudryashov *et al.*²⁰⁶ prepared several fully synthetic vaccines based on Pam3Cys (Figure 9) and examined the role of epitope clustering, carrier structure, and adjuvant on the immunogenicity of these Le^y conjugates in mice. The vaccine containing a cluster of three contiguous Le^y epitopes and the immunostimulating Pam3Cys moiety was found to be superior to a similar construct containing only one Le^y epitope in eliciting antitumor cell antibodies. Because only IgM antibodies were produced by these vaccines, the effect on immunogenicity of coupling the glycopeptide to KLH, not Pam3Cys, was examined. The results showed that although both IgM and IgG antibodies were formed, the antibodies reacted only with the immunizing structure. Reexamination of the clustered Le^y–Pam3Cys conjugate with adjuvant QS-21 resulted in the identification of both IgG and IgM antibodies reacting with tumor cells, thus demonstrating the feasibility of a fully synthetic TACA-based anticancer vaccine in an animal model.

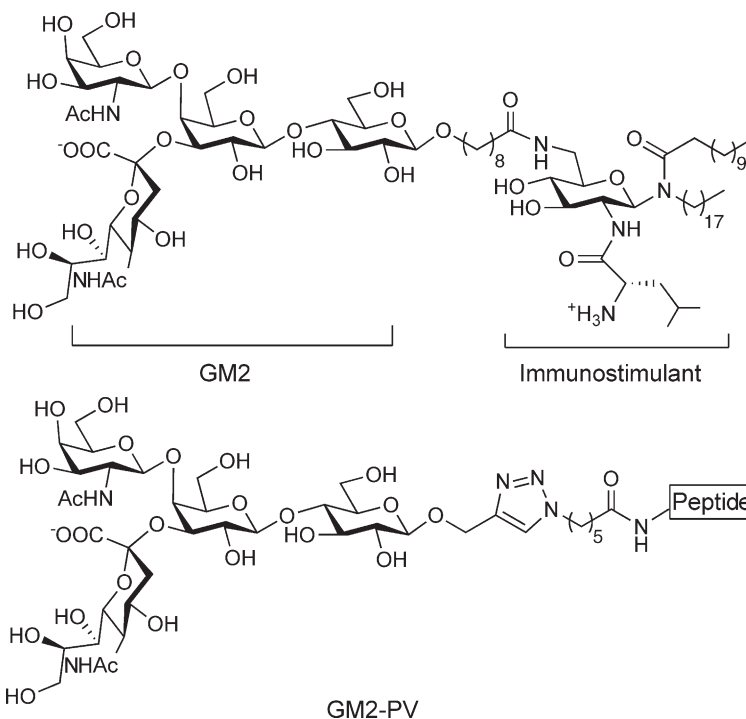


Figure 8 Structures of fully synthetic GM2 vaccines.

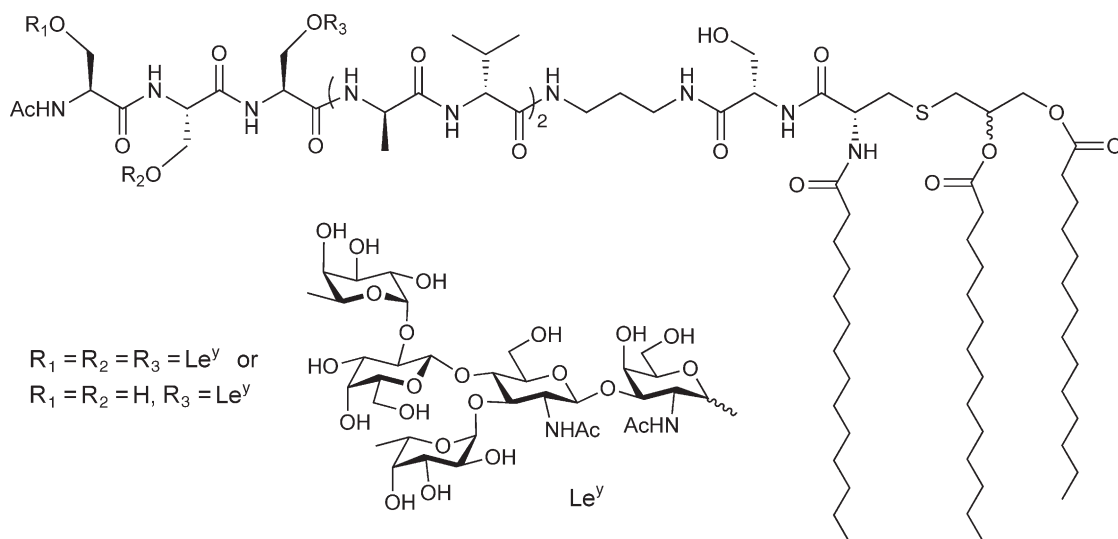


Figure 9 Structures of fully synthetic Le^Y vaccines.

6.04.3.5.2 Three-component fully synthetic glycoconjugate vaccines

As shown above, some Pam3Cys-containing fully synthetic vaccines such as Le^Y -Pam3Cys, di-Tn-Pam3Cys, and tri-Tn-Pam3Cys provoked no or only low levels of IgG immune response, probably due to the lack of T cell-stimulating epitopes in the structures. Although linear and dendrimeric fully synthetic glycopeptides contain T cell epitopes, an external adjuvant is still needed. To overcome this problem, Boons and coworkers proposed a three-component vaccine design consisting of a B cell epitope, namely TACA such as Tn antigen, an adjuvant Pam3Cys, and a Th epitope YAFKYARHANVGRNAFELFL (YAF) (**Figure 10**).²³⁵ YAF is a

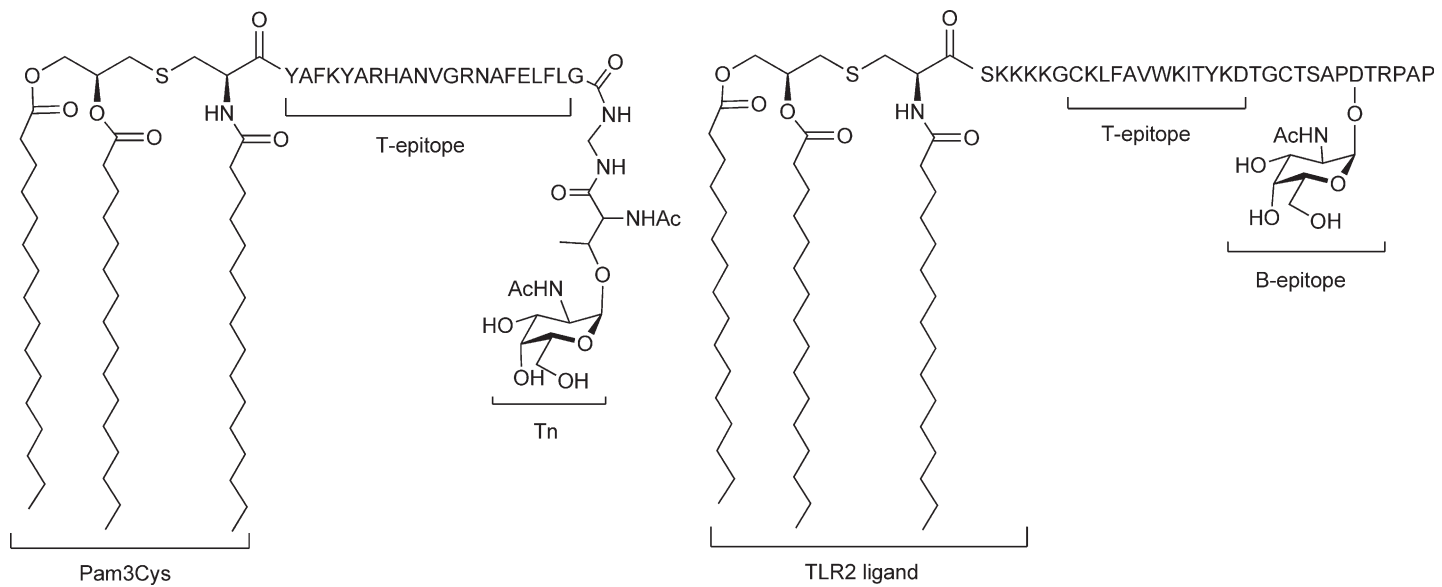


Figure 10 Boons' representative three-component fully synthetic vaccines.

20-amino acid peptide derived from an outer-membrane protein of *N. meningitidis* identified as an MHC class II restricted site for human T cells.²³⁶ It was envisaged that vaccines thus designed should induce the production of high levels of IgG antibodies against the Tn antigen without additional adjuvant. Prior to administration, the vaccine was localized in phospholipid-based liposomes. The mice that were immunized with the liposome preparation elicited low levels of IgM and IgG antibodies against the Tn antigen, while a little better result was found when the liposomes were coadministered with QS-21. The weak immune responses in mice of the three-component vaccine were thought to be the result of Th epitope mismatch, that is, the Th epitope employed in the study was known to be an MHC class II restricted epitope for humans, not for mice. Thus, a more efficient immune response and class switch to IgG antibodies may be expected when a murine Th epitope is employed.

To verify their point of view, Boons and coworkers synthesized two new three-component vaccines containing a tumor-associated glycopeptide of MUC1 and a well-documented mouse MHC class II restricted Th epitope, KLFAVWKITYKDT, derived from PV (Figure 10).²³⁷ Furthermore, of the two vaccines, one contained a built-in adjuvant Pam2CysSK4, which is a potent activator of TLR2 and TLR6, whereas the other contained Pam3CysSK4, which induces cellular activation through TLR1 and TLR2.²³⁸ After being incorporated into liposomes and given to mice, both vaccines elicited exceptionally high titers of IgG antibodies that recognized MUC1-expressing human cancer cells. Coadministration of these vaccines with QS-21 did not lead to a significant increase of IgG antibodies. It was also shown that incorporation of a TLR agonist is important for robust antigenic responses against tumor-associated glycopeptide antigens. In this respect, cytokines induced by the TLR2 ligand are important for maturation of immune cells, leading to robust antibody responses.²³⁹

6.04.3.5.3 Four-component fully synthetic glycoconjugate vaccines

Recently, Renaudet *et al.*²⁴⁰ developed a four-component vaccine displayed on the RAFT platform (Figure 11), consisting of a clustered Tn antigen, a CD4⁺ Th peptide epitope, a CD8⁺ CTL peptide epitope, and a built-in immunoadjuvant Pam3Cys. In this vaccine, the OVA_{257–264} peptide (SIINFEKL)²⁴¹ was used as a target CD8⁺ T cell epitope and PADRE was used as the CD4⁺ Th peptide epitope. The PADRE CD4⁺ Th epitope helps to prime and sustain both B and CTL responses.²⁴² When immunization was carried out in mice, this vaccine was well tolerated without local or general adverse reactions. Mice immunized by the vaccine elicited robust

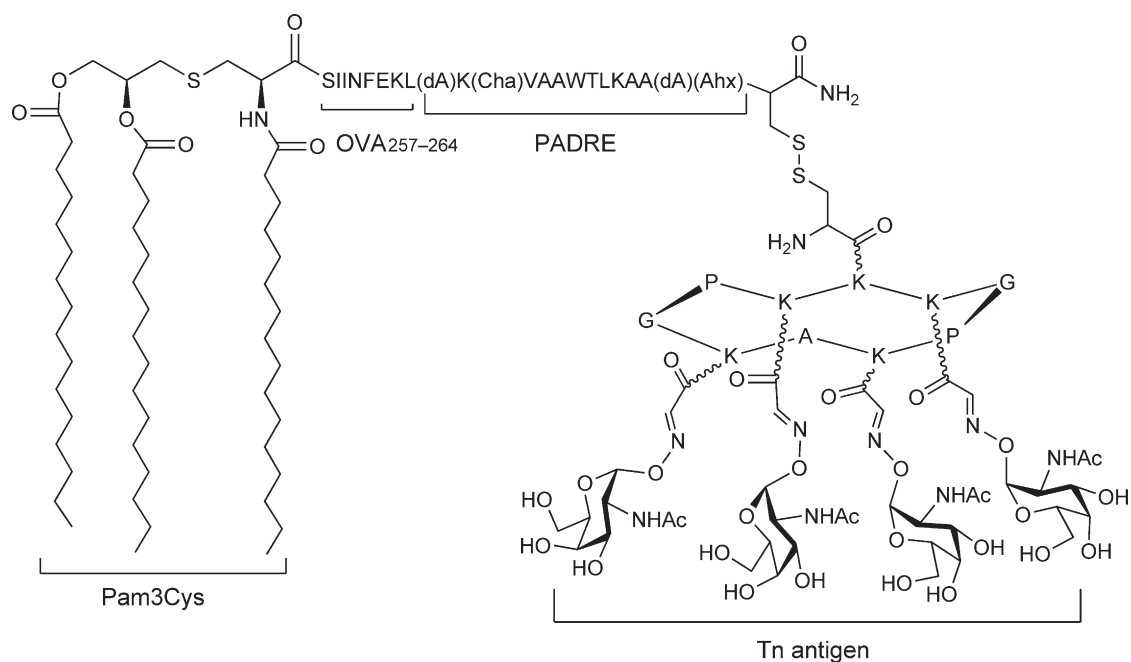


Figure 11 Four-component fully synthetic vaccine.

Tn-specific IgG/IgM antibodies capable of recognizing human tumor cells expressing Tn antigen. In addition, this vaccine induced strong PADRE-specific CD4⁺ T cell and OVA_{257–264}-specific IFN- γ -producing CD8⁺ T cell responses, highlighting correct APCs processing and T cell presentation of both Th and CTL epitopes displayed by the prototype glycolipopeptide structure. Moreover, immunization with this vaccine also resulted in reduction of tumor size in mice inoculated with syngeneic murine MO5 carcinoma cells and protection of animals from lethal carcinoma cell challenge. Finally, vaccination significantly inhibited the growth of preestablished MO5 tumors.²⁴³ These results suggested the potential of self-adjuncting glycolipopeptides as a platform for B cell, CD4⁺, and CD8⁺ T cell epitope-based therapeutic cancer vaccines.

6.04.3.6 Cancer Vaccines Made of Unnatural TACA Analogs

Despite the aforementioned effort and progress, for many TACAs, immunotolerance is still a major issue. As a result, vaccines made of natural TACAs often fail to induce a robust immune response, especially a T cell-dependent immune response, in cancer patients. Even though the mechanisms for immunotolerance are largely unclear, this problem is at least partially due to the structural similarity between TACAs and 'self' glycans. In fact, the majority of TACAs are only overexpressed 'self' antigens rather than being cancer-specific. It is thus natural to think of the possibility of improving the immunological properties of TACAs via the modification of their structures. It is anticipated that these more foreign-looking TACA derivatives would be more immunogenic than the natural TACAs and induce not only antigen-specific antibodies but also CTLs required for successful tumor immunotherapy.

There are two potential ways to realize the specific interaction between the target cancer cell and the provoked immune system after immunization with the synthetic vaccine made of an unnatural TACA analog. One way is to allow the provoked antibodies and immune cells to somehow recognize and cross-react with the natural TACAs on cancer cells, in which case the unnatural TACA analog used must be structurally similar or closely related to the target natural TACA. The other way is to force tumor cells to express the unnatural TACA analog in place of the natural TACA on the cancer cell surface, so that the prestimulated immune system would recognize and specifically react with cancer cells. Both concepts have been explored.

As indicated in **Figure 1**, many of the TACAs identified so far are sialo-TACAs having NeuNAc residues located at the nonreducing end. Therefore, NeuNAc is typically exposed on the cell surface and involved in various biological and pathological processes.^{39,244} It has also been established that oncogenesis is generally accompanied by NeuNAc overexpression.⁴⁰ As a result, chemical modification of the NeuNAc residue could be a rather broadly useful method for augmenting the immunogenicity of sialo-TACA-based vaccines.

6.04.3.6.1 Cancer vaccines made of chemically modified sialo-TACAs

Since GD2–KLH and GD3–KLH plus QS-21 failed to induce consistent, relevant antibody responses in cancer patients,⁶⁵ in order to construct functional vaccines out of GD3 and GD2, Livingston and coworkers converted them to GD3-lactone¹⁹⁵ and GD2-lactone,¹⁷⁵ respectively, and then conjugated these unnatural TACA derivatives with KLH. It was observed that the resulting conjugates had indeed much improved immunogenicity. In the case of GD3, after six melanoma patients were immunized with GD3-lactone–KLH conjugate and QS-21, IgM and IgG antibodies were detected against both GD3 and GD3-lactone. ELISA reactivity was confirmed by ITLC using GD3 and melanoma extracts. Sera obtained from four of the six patients showed cell surface reactivity by FACS, and two showed strong cell surface reactivity by immune adherence (IA) and complement lysis against the GD3 positive cell line SKMEL-28.

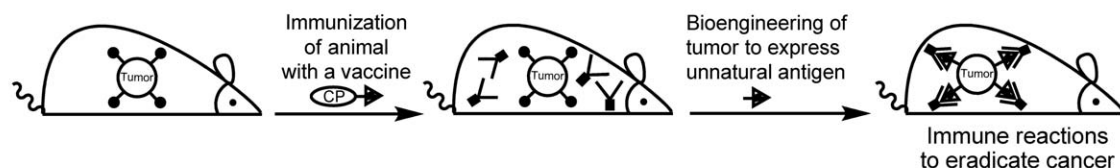
As a bacterial and cancer antigen, PSialA is essentially nonimmunogenic in human beings, because PSialA is abundantly expressed in the human embryo and infants.²⁴⁵ However, Jennings and coworkers have demonstrated that substituting the *N*-acetyl groups of PSialA by other acyl groups, for example, propanoyl and butanoyl groups, resulted in excellent immunogens.²⁴⁶ Livingston and coworkers conducted a preliminary clinical trial in which SCLC patients were vaccinated with PSialA–KLH or *N*-propanoyl PSialA–KLH (PSialANPr–KLH) conjugates. It was observed that the PSialANPr–KLH conjugate elicited specific antibodies that cross-reacted with PSialA, whereas the natural PSialA conjugate did not elicit any antibody.²⁴⁷

6.04.3.6.2 Cancer immunotherapy based on modified sialo-TACAs vaccines and cancer cell glycoengineering

Although GD3-lactone-KLH¹⁹⁵ and PSialANPr-KLH²⁴⁶ conjugates could induce antibodies that cross-reacted with GD3 and PSialA expressed on target cancer cells, it can be imagined that at least a portion of the antibodies, as well as the specific immune reactions, elicited by the unnatural antigens would not recognize and interact with their natural counterparts. As a result, the therapeutic efficiency of the vaccine treatment would be compromised. To overcome this problem, Guo and coworkers have proposed and explored a novel strategy for cancer immunotherapy in which glycoconjugate vaccines made of unnatural TACA analogs are combined with glycoengineering of TACAs on cancer cells.^{248,249} The basic principle of this strategy is depicted in **Figure 12(a)**. First, a glycoconjugate vaccine made of an unnatural analog of a TACA is prepared and employed to immunize animals or cancer patients. Once an immune response specific for the unnatural TACA analog is established, the animals or patients will be treated with an identically modified biosynthetic precursor of the target TACA to induce the expression of the unnatural TACA analog in place of the natural TACA on the cancer cell surface. Subsequently, the preprovoked immune system will recognize and eradicate the glycoengineered tumor cells. The new cancer immunotherapy can also be achieved by a passive immunization protocol (**Figure 12(b)**). First, animals or patients are treated with an unnatural TACA precursor to engineer the expression of the unnatural TACA analog on cancer cells. Meanwhile, a healthy individual is immunized with a synthetic vaccine made of the unnatural TACA analog for the preparation of specific antibodies, which will be used to treat animals or patients for cancer immunotherapy.

For this strategy to work, it has to meet two conditions. First, there must be a functional vaccine made of an unnatural TACA analog that can induce a specific and effective immune response. Second, there must be a suitable method for glycoengineering of cancer cells for the expression of the unnatural TACA analog. For this purpose, sialo-TACAs are ideal targets. As mentioned above, previous studies suggested that chemically modified sialic acids and sialo-oligosaccharides are significantly more immunogenic than their natural counterparts.^{245,247,248,250–252} Most importantly, several groups, including Reutter,^{253,254} Bertozzi,^{105,255,256} Varki,²⁵⁷ Guo,^{249,258,259} and Yarema,^{260–262} have demonstrated that cell surface NeuNAc and sialoglycoconjugates can be efficiently engineered with unnatural *N*-acetyl-D-mannosamine (ManNAc) derivatives as the biosynthetic precursors of artificial sialic acids and sialo-TACA analogs.

(a) Glycoengineered active immunotherapy of cancer



(b) Glycoengineered passive immunotherapy of cancer

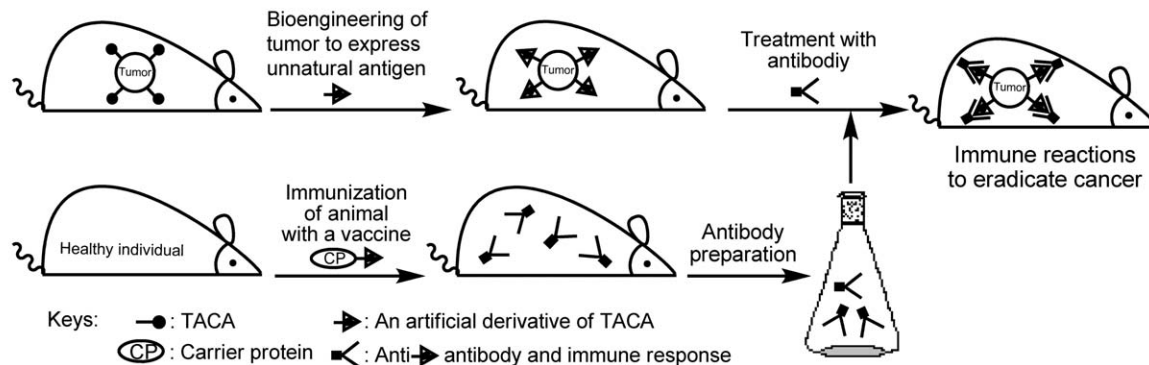


Figure 12 Immunotherapy based on cancer cell glycoengineering.

To identify the ideal precursor and vaccine combination for this strategy, a series of *N*-acyl derivatives of ManNAc and KLH conjugates of *N*-acyl sialic acids or TACA analogs containing unnatural *N*-acyl sialic acids were synthesized and examined as cell glycoengineering precursors and vaccines, respectively.^{165,174,248,251,252,258,259,263} It was revealed that *N*-phenylacetyl-neuraminic acid (NeuNPhAc)²⁵¹ and NeuNPhAc-containing TACA analogs, such as *N*-phenylacetyl-GM3 (GM3NPhAc)²⁴⁸ and *N*-phenylacetyl-sTn (sTnNPhAc),^{165,252} were highly immunogenic and induced high titers of antigen-specific IgG antibodies in mice. It was also disclosed that low micromolar concentrations of *N*-phenylacetyl-D-mannosamine (ManNPhAc) could effectively glycoengineer an array of cancer cell lines.^{258,259} The latter discovery is important in that a micromolar concentration of the precursor is manageable *in vivo*. Furthermore, in the presence of complements, GM3NPhAc–KLH-induced antisera and GM3NPhAc-specific mAb 2H3 exhibited strong and specific cytotoxicity to several melanoma cell lines after the cells were treated with ManNPhAc,^{258,259} while normal cells were not affected under the same conditions, suggesting the high selectivity of the new cancer immunotherapy. It was concluded that combined treatment using ManNPhAc and GM3NPhAc-specific antibodies or using ManNPhAc and vaccines made of GM3NPhAc is worth further investigation as a potentially useful immunotherapy for melanoma and other tumors that express GM3, which is actively pursued.²¹

This strategy was also verified by Jennings and coworkers.^{249,264} For example, when RMA cells, which express PSialA as a unique TACA, were treated with ManNPr, which served as an unnatural biosynthetic precursor of PSialA, the expression of *N*-propionyl-polysialic acid (PSialANPr) on cancer cell surfaces was observed. Moreover, the PSialANPr-specific mAb 13D9 showed strong and specific cytotoxicity to glycoengineered RMA cells, and the cytotoxicity was dependent on the concentrations of ManNPr used. The results of *in vivo* experiments have also demonstrated that administration of mAb 13D9 and ManNPr could inhibit RMA tumor growth and prevent tumor metastasis.²⁴⁹ Similarly, SKMEL-28 cells were glycoengineered with *N*-butanoyl-D-mannosamine (ManNBu) as a glycoengineering precursor to express unnatural *N*-butanoyl GD3 (GD3NBu).²⁶⁴ In the presence of complements, glycoengineered cancer cells were selectively lysed by GD3NBu-specific mAb 2A or antiserum obtained by immunizing mice with GD3NBu–KLH. Although less effective in the control of existing large size tumors in mice, mAb 2A in combination with ManNBu could effectively protect mice from SKMEL-28 tumor grafting.

6.04.4 Conclusion

The natural abundance and forefront locations of TACAs on cancer cell surfaces together with the important roles they play in the process of oncogenesis make TACAs ideal molecular targets for the design and development of cancer vaccines or cancer immunotherapies. However, the relatively poor immunogenicity of carbohydrates and the problem of immunotolerance to TACAs pose a significant hurdle to overcome for any TACA-based cancer vaccine to become realistic. Having said that, great effort and progress have been made in the past two to three decades to address these and related issues for the development of functional TACA-based cancer vaccines. The potential of using such vaccines for the treatment of cancer has been assessed and demonstrated, especially with semisynthetic glycoconjugate vaccines, which contain TACAs coupled to an immunologically active carrier protein. For example, a number of clinical trials have already been conducted with several semisynthetic glycoconjugate vaccines, and two of them even reached phase III clinical trials. While this strategy is still actively pursued, no vaccine has met the clinical endpoints so far. The failures are mainly due to the absence of T cell-mediated immune response in cancer patients treated with these vaccines.

To further augment the immunogenicity of TACA-based vaccines, especially for the generation of T cell-mediated immunity, novel strategies have been developed and explored. Presently, an active field in this regard is fully synthetic glycoconjugate vaccines, which can incorporate not only the TACA epitope but also some CD4⁺ Th and/or CD8⁺ CTL peptide epitopes and immunostimulant epitopes. It is expected that this design will make the resultant conjugates more efficient at inducing T cell-mediated immune response. Moreover, when coupled with solid-phase peptide synthesis, it is feasible to achieve an array of systematically designed small-molecule vaccine candidates, which are relatively easy to characterize and handle, for detailed structure–activity relationship (SAR) studies in order to discover more effective and functional vaccines.

Indeed, some fully synthetic TACA conjugates, such as the MAG-, RAFT-, and Pam3Cys-based TACA conjugates, have already shown potential as vaccine candidates in preclinical trials. However, it is worth noting that there are still very few examples of fully synthetic cancer vaccines reported as compared to semisynthetic glycoconjugates, and only one vaccine has been applied to clinical trials.¹⁹⁰ This is because fully synthetic cancer vaccines are a relatively new area and their preparation still represents a significant challenge. Nevertheless, there has been sufficient evidence to be optimistic that this area will witness major growth in the future, as more promising research results emerge and as fully synthetic glycoconjugates become more easily available.

Another promising direction in cancer vaccine development or cancer immunotherapy is the combination of synthetic glycoconjugates made of more immunogenic unnatural analogs of TACAs and cancer cell glycoengineering. Cell glycoengineering has been proved to be a particularly powerful tool in cell glycobiology, and it is reasonable to believe that more applications of this technique to cancer research and treatment will appear. Moreover, the marriage of this novel immunotherapeutic strategy and other strategies for cancer immunotherapy, such as fully synthetic glycoconjugate vaccine designs, may provide new growth points for the discovery of new and efficient cancer therapies.

Acknowledgment

Our research on cancer vaccines and cancer immunotherapies has been supported by NIH/NCI (CA95142).

References

1. R. Peracaula; S. Barrabes; A. Sarrats; P. M. Rudd; R. de Llorens, *Dis. Markers* **2008**, *25*, 207–218.
2. R. Saldova; M. R. Wormald; R. A. Dwek; P. M. Rudd, *Dis. Markers* **2008**, *25*, 219–232.
3. E. Dabelsteen, *J. Pathol.* **1996**, *179*, 358–369.
4. G. Ragupathi, *Cancer Immunol. Immunother.* **1996**, *43*, 152–157.
5. D. L. Morton; A. Barth, *CA Cancer J. Clin.* **1996**, *46*, 225–244.
6. S. Hakomori, *Curr. Opin. Immunol.* **1991**, *3*, 646–653.
7. S. Hakomori; Y. Zhang, *Chem. Biol.* **1997**, *4*, 97–104.
8. S. Hakomori, *Adv. Exp. Med. Biol.* **2001**, *491*, 369–402.
9. H. J. Jennings; R. K. Sood, Synthetic Glycoconjugates as Human Vaccines. In *Neoglycoconjugates: Preparation and Applications*; Y. C. Lee, R. T. Lee, Eds.; Academic Press: San Diego, 1994; pp 325–371.
10. C. H. Chen; T. C. Wu, *J. Biomed. Sci.* **1998**, *5*, 231–252.
11. L. Cipolla; F. Peri; C. Airoidi, *Anticancer Agents Med. Chem.* **2008**, *8*, 92–121.
12. T. Freire; S. Bay; S. Vichier-Guerre; R. Lo-Man; C. Leclerc, *Mini-Rev. Med. Chem.* **2006**, *6*, 1357–1373.
13. S. F. Slovin; S. J. Keding; G. Ragupathi, *Immunol. Cell Biol.* **2005**, *83*, 418–428.
14. J. D. Warren; X. D. Geng; S. J. Danishefsky, Synthetic Glycopeptide-Based Vaccines. In *Glycopeptides and Glycoproteins: Synthesis, Structure, and Application*; V. Wittmann, Ed.; Springer: Berlin, 2007; Vol. 267, pp 109–141.
15. S. J. Danishefsky; J. R. Allen, *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 836–863.
16. D. H. Dube; C. R. Bertozzi, *Nat. Rev. Drug Discovery* **2005**, *4*, 477–488.
17. D. M. Padoll, *TIPS* **1993**, *14*, 2021–2028.
18. L. J. Old, *Sci. Am.* **1996**, *275*, 136–143.
19. K. E. Hellstrom; P. Gladstone; I. Hellstrom, *Mol. Med. Today* **1997**, *3*, 286–290.
20. S. Ben-Efraim, *Tumour Biol.* **1999**, *20*, 1–24.
21. R. J. Bitton; M. D. Guthmann; M. R. Gabri; A. J. Carnero; D. F. Alonso; L. Fainboim; D. E. Gomez, *Oncol. Rep.* **2002**, *9*, 267–276.
22. P. O. Livingston, *Curr. Opin. Immunol.* **1992**, *4*, 624–629.
23. Z. Shriver; S. Raguram; R. Sasisekharan, *Nat. Rev. Drug Discovery* **2004**, *3*, 863–873.
24. A. Singhal; S. Hakomori, *Bioessays* **1990**, *12*, 223–230.
25. S. Hakomori, *Acta Anat.* **1998**, *161*, 79–90.
26. A. Kobata, Cancer Cells and Metastasis. Concluding Remarks. In *Glycoproteins and Disease*; J. Montreuil, J. F. G. Vliegthart, H. Schachter, Eds.; Elsevier: Amsterdam, 1996; pp 183–241.
27. S. Hakomori, Tumor-Associated Carbohydrate Antigens and Modified Blood Group Antigens. In *Glycoproteins and Disease*; J. Montreuil, J. F. G. Vliegthart, H. Schachter, Eds.; Elsevier: Amsterdam, 1996; pp 243–276.
28. S. Zhang; C. Cordon-Cardo; H. S. Zhang; V. E. Reuter; S. Adluri; W. B. Hamilton; K. O. Lloyd; P. O. Livingston, *Int. J. Cancer* **1997**, *73*, 42–49.
29. S. Zhang; H. S. Zhang; C. Cordon-Cardo; V. E. Reuter; A. K. Singhal; K. O. Lloyd; P. O. Livingston, *Int. J. Cancer* **1997**, *73*, 50–56.
30. I. Brockhausen, *Adv. Exp. Med. Biol.* **2003**, *535*, 163–188.
31. N. Fujitani; Y. Liu; S. Toda; K. Shirouzu; T. Okamura; H. Kimura, *Glycoconjugate J.* **2000**, *17*, 331–338.
32. J. W. Dennis; S. Laferte; C. Waghorne; M. L. Breitman; R. S. Kerbel, *Science* **1987**, *236*, 582–585.

33. S. Yoshida; S. Fukumoto; H. Kawaguchi; S. Sato; R. Ueda; K. Furukawa, *Cancer Res.* **2001**, *61*, 4244–4252.
34. B. Davidson; A. Berner; J. M. Nesland; B. Risberg; G. B. Kristensen; C. G. Trope; M. Bryne, *Hum. Pathol.* **2000**, *31*, 1081–1087.
35. R. Michalides; B. Kwa; D. Springall; N. van Zandwijk; J. Koopman; J. Hilkens; W. Mooi, *Int. J. Cancer: Suppl.* **1994**, *8*, 34–37.
36. Y. J. Kim; A. Varki, *Glycoconjugate J.* **1997**, *14*, 569–576.
37. J. Z. Kiss; G. Rougon, *Curr. Opin. Neurobiol.* **1997**, *7*, 640–646.
38. M. Suzuki; J. Nakayama; A. Suzuki; K. Angata; S. Chen; K. Sakai; K. Hagihara; Y. Yamaguchi; M. Fukuda, *Glycobiology* **2005**, *15*, 887–894.
39. F. A. Troy, II, *Glycobiology* **1992**, *2*, 5–23.
40. R. Takano; E. Muchmore; J. W. Dennis, *Glycobiology* **1994**, *4*, 665–674.
41. E. Staudacher; F. Altmann; I. B. Wilson; L. Marz, *Biochim. Biophys. Acta* **1999**, *1473*, 216–236.
42. R. Renkonen; P. Mattila; M. L. Majuri; J. Rabina; S. Toppila; J. Renkonen; L. Hirvas; J. Niittymaki; J. P. Turunen; O. Renkonen; T. Paavonen, *Glycoconjugate J.* **1997**, *14*, 593–600.
43. E. Miyoshi; K. Moriwaki; T. Nakagawa, *J. Biochem.* **2008**, *143*, 725–729.
44. I. Brockhausen; J. Yang; M. Lehotay; S. Ogata; S. Itzkowitz, *Biol. Chem.* **2001**, *382*, 219–232.
45. H. Ito; N. Hiraiwa; M. Sawada-Kasugai; S. Akamatsu; T. Tachikawa; Y. Kasai; S. Akiyama; K. Ito; H. Takagi; R. Kannagi, *Int. J. Cancer* **1997**, *71*, 556–564.
46. Y. Ikehara; S. Nishihara; T. Kudo; T. Hiraga; K. Morozumi; T. Hattori; H. Narimatsu, *Glycoconjugate J.* **1998**, *15*, 799–807.
47. S. Yamashiro; S. Ruan; K. Furukawa; T. Tai; K. O. Lloyd; H. Shiku, *Cancer Res.* **1993**, *53*, 5395–5400.
48. T. Muramatsu, *Glycobiology* **1993**, *3*, 291–296.
49. S. H. Itzkowitz; M. Yuan; C. K. Montgomery; T. Kjeldsen; H. K. Takahashi; W. L. Bigbee; Y. S. Kim, *Cancer Res.* **1989**, *49*, 197–204.
50. S. H. Itzkowitz; E. J. Bloom; W. A. Kokal; G. Modin; S. Hakomori; Y. S. Kim, *Cancer* **1990**, *66*, 1960–1966.
51. R. Kannagi; M. Izawa; T. Koike; K. Miyazaki; N. Kimura, *Cancer Sci.* **2004**, *95*, 377–384.
52. I. Takahashi; Y. Maehara; T. Kusumoto; M. Yoshida; Y. Kakeji; H. Kusumoto; M. Furusawa; K. Sugimachi, *Cancer* **1993**, *72*, 1836–1840.
53. R. Kannagi, *Glycoconjugate J.* **1997**, *14*, 577–584.
54. J. L. Magnani; Z. Steplewski; H. Koprowski; V. Ginsburg, *Cancer Res.* **1983**, *43*, 5489–5492.
55. T. Jorgensen; A. Berner; O. Kaalhus; K. J. Tvetter; H. E. Danielsen; M. Bryne, *Cancer Res.* **1995**, *55*, 1817–1819.
56. S. Akamine; T. Nakagoe; T. Sawai; T. Tsuji; K. Tanaka; S. Hidaka; S. Shibasaki; A. Nanashima; H. Yamaguchi; T. Nagayasu; T. Yasutake, *Anticancer Res.* **2004**, *24*, 2541–2546.
57. T. Nakagoe; T. Sawai; T. Tsuji; M. A. Jibiki; A. Nanashima; H. Yamaguchi; T. Yasutake; H. Ayabe; K. Arisawa, *Hepatogastroenterology* **2003**, *50*, 696–699.
58. L. G. Durrant; I. Spendlove, *Expert Opin. Emerg. Drugs* **2003**, *8*, 489–500.
59. M. W. Retter; J. C. Johnson; D. W. Peckham; J. E. Bannink; C. S. Bangur; K. Dresser; F. Cai; T. M. Foy; N. A. Fanger; G. R. Fanger; B. Woda; K. L. Rock, *Cancer Res.* **2005**, *65*, 6425–6434.
60. A. L. Yu; M. M. Uttenreuther-Fischer; C. S. Huang; C. C. Tsui; S. D. Gillies; R. A. Reisfeld; F. H. Kung, *J. Clin. Oncol.* **1998**, *16*, 2169–2180.
61. M. Fukuda, *Cancer Res.* **1996**, *56*, 2237–2244.
62. P. O. Livingston; G. Ragupathi, *Cancer Immunol. Immunother.* **1997**, *45*, 10–19.
63. A. F. Ochsenbein; P. Klenerman; U. Karrer; B. Ludewig; M. Pericin; H. Hengartner; R. M. Zinkernagel, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2233–2238.
64. R. M. Zinkernagel, *Semin. Immunol.* **2000**, *12*, 163–171.
65. P. O. Livingston, *Immunol. Rev.* **1995**, *145*, 147–166.
66. X. M. Zhu; R. R. Schmidt, *Angew. Chem. Int. Ed.* **2009**, *48*, 1900–1934.
67. B. G. Davis, *J. Chem. Soc. Perkin. Trans. 1* **2000**, 2137–2160.
68. K. M. Koeller; C. H. Wong, *Chem. Rev.* **2000**, *100*, 4465–4494.
69. G. J. Boons, *Drug Discovery Today* **1996**, *1*, 331–342.
70. G. J. Boons, *Tetrahedron* **1996**, *52*, 1095–1121.
71. H. Paulsen; I. Brockhausen, *Glycoconjugate J.* **2001**, *18*, 867–870.
72. Y. Wang; X. S. Ye; L. H. Zhang, *Org. Biomol. Chem.* **2007**, *5*, 2189–2200.
73. F. Burkhardt; Z. Zhang; S. Wacowich-Sgarbi; C. H. Wong, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1274–1277.
74. Z. Wang; L. Zhou; K. El-Boubbou; X. S. Ye; X. Huang, *J. Org. Chem.* **2007**, *72*, 6409–6420.
75. K. K. T. Mong; C. H. Wong, *Angew. Chem. Int. Ed.* **2002**, *41*, 4087–4090.
76. Z. Y. Zhang; K. Niikura; X. F. Huang; C. H. Wong, *Can. J. Chem.* **2002**, *80*, 1051–1054.
77. T. K. Mong; H. K. Lee; S. G. Duron; C. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 797–802.
78. H. Tanaka; M. Adachi; T. Takahashi, *Chemistry* **2005**, *11*, 849–862.
79. B. Fraser-Reid; Z. F. Wu; U. E. Udodong; H. Ottosson, *J. Org. Chem.* **1990**, *55*, 6068–6070.
80. P. Pornsuriyasak; A. V. Demchenko, *Tetrahedron-Asymmetry* **2005**, *16*, 433–439.
81. L. Huang; X. Huang, *Chemistry* **2007**, *13*, 529–540.
82. X. Huang; L. Huang; H. Wang; X. S. Ye, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 5221–5224.
83. P. H. Seeberger; W. C. Haase, *Chem. Rev.* **2000**, *100*, 4349–4394.
84. P. H. Seeberger; D. B. Werz, *Nat. Rev. Drug Discovery* **2005**, *4*, 751–763.
85. A. Holemann; P. H. Seeberger, *Curr. Opin. Biotechnol.* **2004**, *15*, 615–622.
86. R. Liang; L. Yan; J. Loebach; M. Ge; Y. Uozumi; K. Sekanina; N. Horan; J. Gildersleeve; C. Thompson; A. Smith; K. Biswas; W. C. Still; D. Kahne, *Science* **1996**, *274*, 1520–1522.
87. M. J. Sofia; N. Allanson; N. T. Hatzenbuehler; R. Jain; R. Kakarla; N. Kogan; R. Liang; D. Liu; D. J. Silva; H. Wang; D. Gange; J. Anderson; A. Chen; F. Chi; R. Dulina; B. Huang; M. Kamau; C. Wang; E. Baizman; A. Branstrom; N. Bristol; R. Goldman; K. Han; C. Longley; S. Midha; H. R. Axelrod, *J. Med. Chem.* **1999**, *42*, 3193–3198.
88. D. B. Werz; B. Castagner; P. H. Seeberger, *J. Am. Chem. Soc.* **2007**, *129*, 2770–2771.

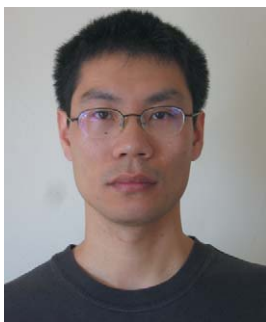
89. O. J. Plante; E. R. Palmacci; P. H. Seeberger, *Science* **2001**, *291*, 1523–1527.
90. K. R. Love; P. H. Seeberger, *Angew. Chem. Int. Ed.* **2004**, *43*, 602–605.
91. E. R. Swanson; P. H. Seeberger, *Abstr. Pap. Am. Chem. Soc.* **2002**, *224*, U268–U268.
92. K. C. Nicolaou; N. Watanabe; J. Li; J. Pastor; N. Winssinger, *Angew. Chem. Int. Ed.* **1998**, *37*, 1559–1561.
93. P. Sears; C. H. Wong, *Science* **2001**, *291*, 2344–2350.
94. O. J. Plante; E. R. Palmacci; P. H. Seeberger, *Adv. Carbohydr. Chem. Biochem.* **2003**, *58*, 35–54.
95. H. Herzner; T. Reipen; M. Schultz; H. Kunz, *Chem. Rev.* **2000**, *100*, 4495–4538.
96. H. J. Gijzen; L. Qiao; W. Fitz; C. H. Wong, *Chem. Rev.* **1996**, *96*, 443–474.
97. R. S. Goody; K. Alexandrov; M. Engelhard, *ChemBioChem* **2002**, *3*, 399–403.
98. P. E. Dawson; S. B. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
99. P. E. Dawson; T. W. Muir; I. Clark-Lewis; S. B. Kent, *Science* **1994**, *266*, 776–779.
100. J. D. Warren; J. S. Miller; S. J. Keding; S. J. Danishefsky, *J. Am. Chem. Soc.* **2004**, *126*, 6576–6578.
101. T. W. Muir; D. Sondhi; P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710.
102. K. Witte; O. Seitz; C. H. Wong, *J. Am. Chem. Soc.* **1998**, *120*, 1979–1989.
103. C. H. Wong; M. Schuster; P. Wang; P. Sears, *J. Am. Chem. Soc.* **1993**, *115*, 5893–5901.
104. E. Saxon; J. I. Armstrong; C. R. Bertozzi, *Org. Lett.* **2000**, *2*, 2141–2143.
105. E. Saxon; C. R. Bertozzi, *Science* **2000**, *287*, 2007–2010.
106. G. A. Winterfeld; R. R. Schmidt, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 2654–2657.
107. G. Ragupathi; D. M. Coltart; L. J. Williams; F. Koide; E. Kagan; J. Allen; C. Harris; P. W. Glunz; P. O. Livingston; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13699–13704.
108. N. Mathieux; H. Paulsen; M. Meldal; K. Bock, *J. Chem. Soc. Perkin. Trans. 1* **1997**, 2359–2368.
109. P. W. Glunz; S. Hintermann; L. J. Williams; J. B. Schwarz; S. D. Kuduk; V. Kudryashov; K. O. Lloyd; S. J. Danishefsky, *J. Am. Chem. Soc.* **2000**, *122*, 7273–7279.
110. S. J. Keding; A. Endo; S. J. Danishefsky, *Tetrahedron* **2003**, *59*, 7023–7031.
111. S.-I. Nishimura; K. Yamada, *J. Am. Chem. Soc.* **1997**, *119*, 10555–10556.
112. K. J. Lee; S. Mao; C. Sun; C. Gao; O. Blixt; S. Arrues; L. G. Hom; G. F. Kaufmann; T. Z. Hoffman; A. R. Coyle; J. Paulson; B. Felding-Habermann; K. D. Janda, *J. Am. Chem. Soc.* **2002**, *124*, 12439–12446.
113. O. Blixt; D. Vasilii; K. Allin; N. Jacobsen; D. Warnock; N. Razi; J. C. Paulson; S. Bernatchez; M. Gilbert; W. Wakarchuk, *Carbohydr. Res.* **2005**, *340*, 1963–1972.
114. K. F. Johnson, *Glycoconjugate J.* **1999**, *16*, 141–146.
115. J. Shao; J. Zhang; P. Kowal; P. G. Wang, *Appl. Environ. Microbiol.* **2002**, *68*, 5634–5640.
116. Z. Wang; M. Gilbert; H. Eguchi; H. Yu; J. S. Cheng; S. Muthana; L. Y. Zhou; P. G. Wang; X. Chen; X. F. Huang, *Adv. Synth. Catal.* **2008**, *350*, 1717–1728.
117. S. K. George; T. Schwientek; B. Holm; C. A. Reis; H. Clausen; J. Kihlberg, *J. Am. Chem. Soc.* **2001**, *123*, 11117–11125.
118. O. Blixt; K. Allin; L. Pereira; A. Datta; J. C. Paulson, *J. Am. Chem. Soc.* **2002**, *124*, 5739–5746.
119. A. L. Sorensen; C. A. Reis; M. A. Tarp; U. Mandel; K. Ramachandran; V. Sankaranarayanan; T. Schwientek; R. Graham; J. Taylor-Papadimitriou; M. A. Hollingsworth; J. Burchell; H. Clausen, *Glycobiology* **2006**, *16*, 96–107.
120. T. Freire; R. Lo-Man; F. Piller; V. Piller; C. Leclerc; S. Bay, *Glycobiology* **2006**, *16*, 390–401.
121. D. M. Su; H. Eguchi; W. Yi; L. Li; P. G. Wang; C. Xia, *Org. Lett.* **2008**, *10*, 1009–1012.
122. O. Seitz; C. H. Wong, *J. Am. Chem. Soc.* **1997**, *119*, 8766–8776.
123. M. M. Palcic; H. Li; D. Zanini; R. S. Bhella; R. Roy, *Carbohydr. Res.* **1997**, *305*, 433–442.
124. F. Sallas; S. I. Nishimura, *J. Chem. Soc. Perkin. Trans. 1* **2000**, 2091–2103.
125. K. Aijisaka; M. Miyasato; C. Ito; Y. Fujita; Y. Yamazaki; S. Oka, *Glycoconjugate J.* **2001**, *18*, 301–308.
126. M. Schuster; P. Wang; J. C. Paulson; C. H. Wong, *J. Am. Chem. Soc.* **1994**, *116*, 1135–1136.
127. O. Blixt; T. Norberg, *J. Org. Chem.* **1998**, *63*, 2705–2710.
128. G. Dudziak; N. Bezay; T. Schwientek; H. Clausen; H. Kunz; A. Liese, *Tetrahedron* **2000**, *56*, 5865–5869.
129. G. F. Herrmann; Y. Ichikawa; C. Wandrey; F. C. A. Gaeta; J. C. Paulson; C. H. Wong, *Tetrahedron Lett.* **1993**, *34*, 3091–3094.
130. E. Grabenhorst; P. Schlenke; S. Pohl; M. Nimtz; H. S. Conradt, *Glycoconjugate J.* **1999**, *16*, 81–97.
131. S. Muller; F. G. Hanisch, *J. Biol. Chem.* **2002**, *277*, 26103–26112.
132. F. J. Olson; M. Backstrom; H. Karlsson; J. Burchell; G. C. Hansson, *Glycobiology* **2005**, *15*, 177–191.
133. T. Link; M. Backstrom; R. Graham; R. Essers; K. Zorner; J. Gatgens; J. Burchell; J. Taylor-Papadimitriou; G. C. Hansson; T. Noll, *J. Biotechnol.* **2004**, *110*, 51–62.
134. B. Priem; M. Gilbert; W. W. Wakarchuk; A. Heyraud; E. Samain, *Glycobiology* **2002**, *12*, 235–240.
135. T. Antoine; B. Priem; A. Heyraud; L. Greffe; M. Gilbert; W. W. Wakarchuk; J. Lam; E. Samain, *ChemBioChem* **2003**, *4*, 406–412.
136. T. Antoine; A. Heyraud; C. Bosso; E. Samain, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 1350–1352.
137. T. Antoine; C. Bosso; A. Heyraud; E. Samain, *Biochimie* **2005**, *87*, 197–203.
138. M. Randriantsoa; S. Drouillard; C. Breton; E. Samain, *FEBS Lett.* **2007**, *581*, 2652–2656.
139. M. Gilbert; M. F. Karwaski; S. Bernatchez; N. M. Young; E. Taboada; J. Michniewicz; A. M. Cunningham; W. W. Wakarchuk, *J. Biol. Chem.* **2002**, *277*, 327–337.
140. T. Endo; S. Koizumi; K. Tabata; A. Ozaki, *Appl. Microbiol. Biotechnol.* **2000**, *53*, 257–261.
141. T. Endo; S. Koizumi; K. Tabata; S. Kakita; A. Ozaki, *Carbohydr. Res.* **2001**, *330*, 439–443.
142. S. Fort; L. Birikaki; M. P. Dubois; T. Antoine; E. Samain; H. Driguez, *Chem. Commun.* **2005**, 2558–2560.
143. S. Bay; S. Fort; L. Birikaki; C. Ganneau; E. Samain; Y. M. Coic; F. Bonhomme; E. Deriau; C. Leclerc; R. Lo-Man, *ChemMedChem* **2009**, *4*, 582–587.
144. O. J. Finn, *Nat. Rev. Immunol.* **2003**, *3*, 630–641.
145. M. Singh; D. T. O'Hagan, *Pharm. Res.* **2002**, *19*, 715–728.
146. J. T. Evans; C. W. Cluff; D. A. Johnson; M. J. Lacy; D. H. Persing; J. R. Baldrige, *Expert Rev. Vaccines* **2003**, *2*, 219–229.
147. D. J. Rawlings; D. C. Kaslow, *J. Exp. Med.* **1992**, *176*, 1483–1487.
148. C. R. Kensil; U. Patel; M. Lennick; D. Marciani, *J. Immunol.* **1991**, *146*, 431–437.

149. F. Helling; S. Zhang; A. Shang; S. Adluri; M. Calves; R. Koganty; B. M. Longenecker; T. J. Yao; H. F. Oettgen; P. O. Livingston, *Cancer Res.* **1995**, *55*, 2783–2788.
150. F. Helling; A. Shang; M. Calves; S. Zhang; S. Ren; R. K. Yu; H. F. Oettgen; P. O. Livingston, *Cancer Res.* **1994**, *54*, 197–203.
151. K. H. Wiesmuller; G. Jung; D. Gillissen; C. Loffl; W. G. Bessler; T. Boltz, *Immunology* **1991**, *72*, 109–113.
152. W. G. Bessler; G. Jung, *Res. Immunol.* **1992**, *143*, 548–553.
153. K. H. Wiesmuller; W. G. Bessler; G. Jung, *Int. J. Pept. Protein Res.* **1992**, *40*, 255–260.
154. K. Deres; H. Schild; K. H. Wiesmuller; G. Jung; H. G. Rammensee, *Nature* **1989**, *342*, 561–564.
155. K. H. Wiesmuller; G. Jung; G. Hess, *Vaccine* **1989**, *7*, 29–33.
156. A. O. Aliprantis; R. B. Yang; M. R. Mark; S. Suggett; B. Devaux; J. D. Radolf; G. R. Klimpel; P. Godowski; A. Zychlinsky, *Science* **1999**, *285*, 736–739.
157. J. P. Defoort; B. Nardelli; W. Huang; J. P. Tam, *Int. J. Pept. Protein Res.* **1992**, *40*, 214–221.
158. A. M. Buiting; N. van Rooijen; E. Claassen, *Res. Immunol.* **1992**, *143*, 541–548.
159. F. Helling; P. O. Livingston, *Mol. Chem. Neuropathol.* **1994**, *21*, 299–309.
160. V. Kudryashov; H. M. Kim; G. Ragupathi; S. J. Danishefsky; P. O. Livingston; K. O. Lloyd, *Cancer Immunol. Immunother.* **1998**, *45*, 281–286.
161. E. Kagan; G. Ragupathi; S. S. Yi; C. A. Reis; J. Gildersleeve; D. Kahne; H. Clausen; S. J. Danishefsky; P. O. Livingston, *Cancer Immunol. Immunother.* **2005**, *54*, 424–430.
162. V. Pozsgay, *Adv. Carbohydr. Chem. Biochem.* **2001**, *56*, 153–199.
163. T. Buskas; Y. Li; G. J. Boons, *Chemistry* **2004**, *10*, 3517–3524.
164. F. Mawas; J. Niggemann; C. Jones; M. J. Corbel; J. P. Kamerling; J. F. Vliegenthart, *Infect. Immun.* **2002**, *70*, 5107–5114.
165. Q. Wang; S. A. Ekanayaka; J. Wu; J. Zhang; Z. Guo, *Bioconjugate Chem.* **2008**, *19*, 2060–2067.
166. P. C. Jones; L. L. Sze; P. Y. Liu; D. L. Morton; R. F. Irie, *J. Natl. Cancer Inst.* **1981**, *66*, 249–254.
167. P. O. Livingston; G. Y. Wong; S. Adluri; Y. Tao; M. Padavan; R. Parente; C. Hanlon; M. J. Calves; F. Helling; G. Ritter, *J. Clin. Oncol.* **1994**, *12*, 1036–1044.
168. R. F. Irie; T. Matsuki; D. L. Morton, *Lancet* **1989**, *1*, 786–787.
169. P. B. Chapman; D. M. Morrissey; K. S. Panageas; W. B. Hamilton; C. Zhan; A. N. Destro; L. Williams; R. J. Israel; P. O. Livingston, *Clin. Cancer Res.* **2000**, *6*, 874–879.
170. P. Livingston; S. Zhang; S. Adluri; T. J. Yao; L. Graeber; G. Ragupathi; F. Helling; M. Fleisher, *Cancer Immunol. Immunother.* **1997**, *43*, 324–330.
171. K. Kitamura; P. O. Livingston; S. R. Fortunato; E. Stockert; F. Helling; G. Ritter; H. F. Oettgen; L. J. Old, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2805–2809.
172. G. Ragupathi; N. X. Liu; C. Musselli; S. Powell; K. Lloyd; P. O. Livingston, *J. Immunol.* **2005**, *174*, 5706–5712.
173. J. M. Kirkwood; J. G. Ibrahim; J. A. Sosman; V. K. Sondak; S. S. Agarwala; M. S. Ernstoff; U. Rao, *J. Clin. Oncol.* **2001**, *19*, 2370–2380.
174. J. Xue; Y. Pan; Z. Guo, *Tetrahedron Lett.* **2002**, *43*, 1599–1602.
175. G. Ragupathi; P. O. Livingston; C. Hood; J. Gathuru; S. E. Krown; P. B. Chapman; J. D. Wolchok; L. J. Williams; R. C. Oldfield; W. J. Hwu, *Clin. Cancer Res.* **2003**, *9*, 5214–5220.
176. T. Brezicka; B. Bergman; S. Olling; P. Fredman, *Lung Cancer* **2000**, *28*, 29–36.
177. L. M. Krug; G. Ragupathi; C. Hood; M. G. Kris; V. A. Miller; J. R. Allen; S. J. Keding; S. J. Danishefsky; J. Gomez; L. Tyson; B. Pizzo; V. Baez; P. O. Livingston, *Clin. Cancer Res.* **2004**, *10*, 6094–6100.
178. S. Menard; E. Tagliabue; S. Canevari; G. Fossati; M. I. Colnaghi, *Cancer Res.* **1983**, *43*, 1295–1300.
179. S. Canevari; G. Fossati; A. Balsari; S. Sonnino; M. I. Colnaghi, *Cancer Res.* **1983**, *43*, 1301–1305.
180. T. Gilewski; G. Ragupathi; S. Bhuta; L. J. Williams; C. Musselli; X. F. Zhang; K. P. Bencsath; K. S. Panageas; J. Chin; C. A. Hudis; L. Norton; A. N. Houghton; P. O. Livingston; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3270–3275.
181. S. F. Slovin; G. Ragupathi; S. Adluri; G. Ungers; K. Terry; S. Kim; M. Spassova; W. G. Bornmann; M. Fazzari; L. Dantis; K. Olkiewicz; K. O. Lloyd; P. O. Livingston; S. J. Danishefsky; H. I. Scher, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5710–5715.
182. G. Ragupathi; T. K. Park; S. L. Zhang; I. J. Kim; L. Graber; S. Adluri; K. O. Lloyd; S. J. Danishefsky; P. O. Livingston, *Angew. Chem. Int. Ed.* **1997**, *36*, 125–128.
183. Z. G. Wang; L. J. Williams; X. F. Zhang; A. Zatorski; V. Kudryashov; G. Ragupathi; M. Spassova; W. Bornmann; S. F. Slovin; H. I. Scher; P. O. Livingston; K. O. Lloyd; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2719–2724.
184. M. Leivonen; S. Nordling; J. Lundin; K. von Boguslawski; C. Haglund, *Oncology* **2001**, *61*, 299–305.
185. H. Kobayashi; T. Terao; Y. Kawashima, *J. Clin. Oncol.* **1992**, *10*, 95–101.
186. L. A. Holmberg; B. M. Sandmaier, *Expert Rev. Vaccines* **2004**, *3*, 655–663.
187. B. M. Longenecker; M. Reddish; R. Koganty; G. D. MacLean, *Ann. N. Y. Acad. Sci.* **1993**, *690*, 276–291.
188. G. D. MacLean; M. Reddish; R. R. Koganty; T. Wong; S. Gandhi; M. Smolenski; J. Samuel; J. M. Nabholz; B. M. Longenecker, *Cancer Immunol. Immunother.* **1993**, *36*, 215–222.
189. S. Adluri; F. Helling; S. Ogata; S. Zhang; S. H. Itzkowitz; K. O. Lloyd; P. O. Livingston, *Cancer Immunol. Immunother.* **1995**, *41*, 185–192.
190. S. F. Slovin; G. Ragupathi; C. Musselli; K. Olkiewicz; D. Verbel; S. D. Kuduk; J. B. Schwarz; D. Sames; S. Danishefsky; P. O. Livingston; H. I. Scher, *J. Clin. Oncol.* **2003**, *21*, 4292–4298.
191. G. D. Maclean; M. B. Bowenyacyshyn; J. Samuel; A. Meikle; G. Stuart; J. Nation; S. Poppema; M. Jerry; R. Koganty; T. Wong; B. M. Longenecker, *J. Immunother.* **1992**, *11*, 292–305.
192. S. F. Slovin; G. Ragupathi; C. Musselli; C. Fernandez; M. Diani; D. Verbel; S. Danishefsky; P. Livingston; H. I. Scher, *Cancer Immunol. Immunother.* **2005**, *54*, 694–702.
193. G. D. MacLean; M. A. Reddish; R. R. Koganty; B. M. Longenecker, *J. Immunother. Emphasis Tumor Immunol.* **1996**, *19*, 59–68.
194. P. J. Sabbatini; V. Kudryashov; G. Ragupathi; S. J. Danishefsky; P. O. Livingston; W. Bornmann; M. Spassova; A. Zatorski; D. Spriggs; C. Aghajanian; S. Soignet; M. Peyton; C. O’Flaherty; J. Curtin; K. O. Lloyd, *Int. J. Cancer* **2000**, *87*, 79–85.
195. G. Ragupathi; M. Meyers; S. Adluri; L. Howard; C. Musselli; P. O. Livingston, *Int. J. Cancer* **2000**, *85*, 659–666.

196. P. B. Chapman; D. Wu; G. Ragupathi; S. Lu; L. Williams; W. J. Hwu; D. Johnson; P. O. Livingston, *Clin. Cancer Res.* **2004**, *10*, 4717–4723.
197. R. J. Bitton, *Curr. Opin. Mol. Ther.* **2004**, *6*, 17–26.
198. M. N. Dickler; G. Ragupathi; N. X. Liu; C. Musselli; D. J. Martino; V. A. Miller; M. G. Kris; F. T. Brezicka; P. O. Livingston; S. C. Grant, *Clin. Cancer Res.* **1999**, *5*, 2773–2779.
199. P. B. Chapman; D. Morrisey; K. S. Panageas; L. Williams; J. J. Lewis; R. J. Israel; W. B. Hamilton; P. O. Livingston, *Clin. Cancer Res.* **2000**, *6*, 4658–4662.
200. S. F. Slovin; G. Ragupathi; C. Fernandez; M. P. Jefferson; M. Diani; A. S. Wilton; S. Powell; M. Spassova; C. Reis; H. Clausen; S. Danishefsky; P. Livingston; H. I. Scher, *Vaccine* **2005**, *23*, 3114–3122.
201. I. Carlstedt; J. R. Davies, *Biochem. Soc. Trans.* **1997**, *25*, 214–219.
202. S. Zhang; L. A. Walberg; S. Ogata; S. H. Itzkowitz; R. R. Koganty; M. Reddish; S. S. Gandhi; B. M. Longenecker; K. O. Lloyd; P. O. Livingston, *Cancer Res.* **1995**, *55*, 3364–3368.
203. S. D. Kuduk; J. B. Schwarz; X. T. Chen; P. W. Glunz; D. Sames; G. Ragupathi; P. O. Livingston; S. J. Danishefsky, *J. Am. Chem. Soc.* **1998**, *120*, 12474–12485.
204. J. B. Schwarz; S. D. Kuduk; X. T. Chen; D. Sames; P. W. Glunz; S. J. Danishefsky, *J. Am. Chem. Soc.* **1999**, *121*, 2662–2673.
205. D. Sames; X. T. Chen; S. J. Danishefsky, *Nature* **1997**, *389*, 587–591.
206. V. Kudryashov; P. W. Glunz; L. J. Williams; S. Hintermann; S. J. Danishefsky; K. O. Lloyd, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3264–3269.
207. J. Zhu; Q. Wan; G. Ragupathi; C. M. George; P. O. Livingston; S. J. Danishefsky, *J. Am. Chem. Soc.* **2009**, *131*, 4151–4158.
208. R. L. Giuntoli, 2nd; G. C. Rodriguez; R. S. Whitaker; R. Dodge; J. A. Voynow, *Cancer Res.* **1998**, *58*, 5546–5550.
209. K. Kiguchi; Y. Iwamori; N. Suzuki; Y. Kobayashi; B. Ishizuka; I. Ishiwata; T. Kita; Y. Kikuchi; M. Iwamori, *Cancer Sci.* **2006**, *97*, 1321–1326.
210. C. A. Lingwood; A. A. Khine; S. Arab, *Acta Biochim. Pol.* **1998**, *45*, 351–359.
211. P. Livingston, *Clin. Cancer Res.* **2001**, *7*, 1837–1838.
212. G. Ragupathi; F. Koide; N. Sathyan; E. Kagan; M. Spassova; W. Bornmann; P. Gregor; C. A. Reis; H. Clausen; S. J. Danishefsky; P. O. Livingston, *Cancer Immunol. Immunother.* **2003**, *52*, 608–616.
213. G. Ragupathi; S. Cappello; S. S. Yi; D. Canter; M. Spassova; W. G. Bornmann; S. J. Danishefsky; P. O. Livingston, *Vaccine* **2002**, *20*, 1030–1038.
214. L. J. Williams; C. R. Harris; P. W. Glunz; S. J. Danishefsky, *Tetrahedron Lett.* **2000**, *41*, 9505–9508.
215. J. R. Allen; C. R. Harris; S. J. Danishefsky, *J. Am. Chem. Soc.* **2001**, *123*, 1890–1897.
216. S. J. Keding; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11937–11942.
217. G. Ragupathi; F. Koide; P. O. Livingston; Y. S. Cho; A. Endo; Q. Wan; M. K. Spassova; S. J. Keding; J. Allen; O. Ouerfelli; R. M. Wilson; S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 2715–2725.
218. R. Roy, New Trends in Carbohydrate-Based Vaccines. In *Drug Discovery Today: Technologies*; J. Owens, Ed.; Elsevier Science Ltd.: Oxford, 2004; Vol. 1, pp 327–336.
219. M. P. Schutze; C. Leclerc; M. Jolivet; F. Audibert; L. Chedid, *J. Immunol.* **1985**, *135*, 2319–2322.
220. T. Toyokuni; S. Hakomori; A. K. Singhal, *Bioorg. Med. Chem.* **1994**, *2*, 1119–1132.
221. T. Toyokuni; B. Dean; S. P. Cai; D. Boivin; S. Hakomori; A. K. Singhal, *J. Am. Chem. Soc.* **1994**, *116*, 395–396.
222. S. Vichier-Guerre; R. Lo-Man; S. Bay; E. Deriaud; H. Nakada; C. Leclerc; D. Cantacuzene, *J. Pept. Res.* **2000**, *55*, 173–180.
223. R. Lo-Man; S. Bay; S. Vichier-Guerre; E. Deriaud; D. Cantacuzene; C. Leclerc, *Cancer Res.* **1999**, *59*, 1520–1524.
224. R. Lo-Man; S. Vichier-Guerre; S. Bay; E. Deriaud; D. Cantacuzene; C. Leclerc, *J. Immunol.* **2001**, *166*, 2849–2854.
225. D. N. Posnett; H. McGrath; J. P. Tam, *J. Biol. Chem.* **1988**, *263*, 1719–1725.
226. J. P. Tam, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5409–5413.
227. R. Lo-Man; S. Vichier-Guerre; R. Perraut; E. Deriaud; V. Huteau; L. BenMohamed; O. M. Diop; P. O. Livingston; S. Bay; C. Leclerc, *Cancer Res.* **2004**, *64*, 4987–4994.
228. P. Panina-Bordignon; A. Tan; A. Termijtelen; S. Demotz; G. Corradin; A. Lanzavecchia, *Eur. J. Immunol.* **1989**, *19*, 2237–2242.
229. J. Alexander; J. Sidney; S. Southwood; J. Ruppert; C. Oseroff; A. Maewal; K. Snoke; H. M. Serra; R. T. Kubo; A. Sette; H. M. Grey, *Immunity* **1994**, *1*, 751–761.
230. S. Grigalevicius; S. Chierici; O. Renaudet; R. Lo-Man; E. Deriaud; C. Leclerc; P. Dumy, *Bioconjugate Chem.* **2005**, *16*, 1149–1159.
231. S. Peluso; T. Ruckle; C. Lehmann; M. Mutter; C. Peggion; M. Crisma, *ChemBioChem* **2001**, *2*, 432–437.
232. H. Kunz, *J. Pept. Sci.* **2003**, *9*, 563–573.
233. U. Westerlind; A. Hobel; N. Gaidzik; E. Schmitt; H. Kunz, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 7551–7556.
234. W. Dullenkopf; G. Ritter; S. R. Fortunato; L. J. Old; R. R. Schmidt, *Chem. Eur. J.* **1999**, *5*, 2432–2438.
235. T. Buskas; S. Ingale; G. J. Boons, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 5985–5988.
236. E. J. Wiertz; J. A. van Gaans-van den Brink; H. Gausepohl; A. Prochnicka-Chalufour; P. Hoogerhout; J. T. Poolman, *J. Exp. Med.* **1992**, *176*, 79–88.
237. C. Leclerc; E. Deriaud; V. Mimic; S. van der Werf, *J. Virol.* **1991**, *65*, 711–718.
238. R. Spohn; U. Buwitt-Beckmann; R. Brock; G. Jung; A. J. Ulmer; K. H. Wiesmuller, *Vaccine* **2004**, *22*, 2494–2499.
239. S. Ingale; M. A. Wolfert; T. Buskas; G. J. Boons, *ChemBioChem* **2009**, *10*, 455–463.
240. O. Renaudet; L. BenMohamed; G. Dasgupta; I. Bettahi; P. Dumy, *ChemMedChem* **2008**, *3*, 737–741.
241. O. Rotzschke; K. Falk; S. Stevanovic; G. Jung; P. Walden; H. G. Rammensee, *Eur. J. Immunol.* **1991**, *21*, 2891–2894.
242. L. BenMohamed; S. L. Wechsler; A. B. Nesburn, *Lancet Infect. Dis.* **2002**, *2*, 425–431.
243. I. Bettahi; G. Dasgupta; O. Renaudet; A. A. Chentoufi; X. Zhang; D. Carpenter; S. Yoon; P. Dumy; L. BenMohamed, *Cancer Immunol. Immunother.* **2009**, *58*, 187–200.
244. R. Schauer, *Adv. Exp. Med. Biol.* **1988**, *228*, 47–72.
245. H. J. Jennings, *Int. J. Infect. Dis.* **1997**, *1*, 158–164.
246. R. A. Pon; M. Lussier; Q. L. Yang; H. J. Jennings, *J. Exp. Med.* **1997**, *185*, 1929–1938.

247. L. M. Krug; G. Ragupathi; K. K. Ng; C. Hood; H. J. Jennings; Z. Guo; M. G. Kris; V. Miller; B. Pizzo; L. Tyson; V. Baez; P. O. Livingston, *Clin. Cancer Res.* **2004**, *10*, 916–923.
248. Y. Pan; P. Chefalo; N. Nagy; C. Harding; Z. Guo, *J. Med. Chem.* **2005**, *48*, 875–883.
249. T. Liu; Z. Guo; Q. Yang; S. Sad; H. J. Jennings, *J. Biol. Chem.* **2000**, *275*, 32832–32836.
250. G. Ritter; E. Boosfeld; M. J. Calves; H. F. Oettgen; L. J. Old; P. O. Livingston, *Immunobiology* **1990**, *182*, 32–43.
251. P. Chefalo; Y. Pan; N. Nagy; C. Harding; Z. Guo, *Glycoconjugate J.* **2004**, *20*, 407–414.
252. J. Wu; Z. Guo, *Bioconjugate Chem.* **2006**, *17*, 1537–1544.
253. H. Kayser; C. C. Geilen; C. Paul; R. Zeitler; W. Reutter, *FEBS Lett.* **1992**, *301*, 137–140.
254. C. Schmidt; P. Stehling; J. Schnitzer; W. Reutter; R. Horstkorte, *J. Biol. Chem.* **1998**, *273*, 19146–19152.
255. C. R. Bertozzi; L. L. Kiessling, *Science* **2001**, *291*, 2357–2364.
256. L. K. Mahal; K. J. Yarema; C. R. Bertozzi, *Science* **1997**, *276*, 1125–1128.
257. P. Tangvoranuntakul; P. Gagneux; S. Diaz; M. Bardor; N. Varki; A. Varki; E. Muchmore, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12045–12050.
258. P. Chefalo; Y. Pan; N. Nagy; Z. Guo; C. V. Harding, *Biochemistry* **2006**, *45*, 3733–3739.
259. Q. Wang; J. Zhang; Z. Guo, *Bioorg. Med. Chem.* **2007**, *15*, 7561–7567.
260. C. T. Campbell; S. G. Sampathkumar; K. J. Yarema, *Mol. Biosyst.* **2007**, *3*, 187–194.
261. S. G. Sampathkumar; A. V. Li; M. B. Jones; Z. Sun; K. J. Yarema, *Nat. Chem. Biol.* **2006**, *2*, 149–152.
262. S. G. Sampathkumar; M. B. Jones; K. J. Yarema, *Nat. Protocol.* **2006**, *1*, 1840–1851.
263. Y. Pan; T. Ayani; J. Nadas; S. Wen; Z. Guo, *Carbohydr. Res.* **2004**, *339*, 2091–2100.
264. W. Zou; S. Borrelli; M. Gilbert; T. Liu; R. A. Pon; H. J. Jennings *J. Biol. Chem.* **2004**, *279*, 25390–25399.

Biographical Sketches



Qianli Wang received his B.S. degree in pharmaceuticals (1996) and his Ph.D. degree (2003) in medicinal chemistry from the Shenyang Pharmaceutical University in China. He joined Professor Katja Michael's group at University of Hawaii at Manoa as a postdoctoral fellow in 2003 and then moved to Case Western Reserve University in 2005, where he spent about 10 months working with Professor M. Edward Medof in the Department of Pathology. Since 2005, he is working with Professor Zhongwu Guo at Wayne State University as a research associate. His current interests are mainly in carbohydrate chemistry and glycobiology, especially the development of carbohydrate-based cancer vaccines and novel cancer immunotherapies.



Zhongwu Guo received his degree of pharmacy (1984) and then M.S. degree in medicinal chemistry (1987) from the Second Military Medical University, and his Ph.D. degree (1991) in organic chemistry with Professor Aleksander Zamojski from the Institute of Organic Chemistry, Polish Academy of Sciences. Dr. Guo was appointed to the faculty at the Shanghai Institute of Organic Chemistry in 1994. From 1996 to 1997, he worked with Professor Tomoya Ogawa at RIKEN, Japan as a RIKEN Fellow, and then served as an assistant research officer at the National Research Council of Canada from 1997 to 1999. Dr. Guo was appointed assistant and associate professor of chemistry and oncology at Case Western Reserve University from 1999 to 2005, and is presently professor of chemistry and a member of Karmanos Cancer Institute at Wayne State University. His research interests are mainly in synthetic organic chemistry, carbohydrate chemistry, medicinal chemistry, and glycobiology.

6.05 MS-Based Glycoanalysis

Kay-Hooi Khoo, Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.05.1 | Introduction | 124 |
| 6.05.2 | Diversity in Eukaryotic Protein Glycosylation | 125 |
| 6.05.2.1 | Basic Tenets of Protein Glycosylation in the Context of MS Analysis | 125 |
| 6.05.2.1.1 | Factors determining the glycosylation structures – how much can be assumed <i>a priori</i> ? | 125 |
| 6.05.2.1.2 | Heterogeneity and glycoforms – how much is enough? | 126 |
| 6.05.2.1.3 | Site- and protein-specific glycosylation – why glycopeptides and glycoproteomics? | 126 |
| 6.05.2.2 | Structural Architecture and Diversity | 126 |
| 6.05.2.2.1 | Types of glycosylation | 127 |
| 6.05.2.2.2 | The common core structures | 127 |
| 6.05.2.2.3 | Terminal glycosylation from the perspective of MS analysis | 128 |
| 6.05.3 | Workflows in MS-Based Protein Glycosylation Analysis | 128 |
| 6.05.3.1 | Defining the Conceptual Framework | 129 |
| 6.05.3.2 | Pre-MS Sample Preparation Requirements | 130 |
| 6.05.3.2.1 | From biological source to solubilized glycoproteins | 130 |
| 6.05.3.2.2 | Glycan release and reducing end tagging | 131 |
| 6.05.3.2.3 | Permethylation and other related chemical derivatization | 132 |
| 6.05.3.2.4 | Glycan fractionation and targeted enrichment | 133 |
| 6.05.3.2.5 | Glycopeptide enrichment | 134 |
| 6.05.3.3 | Practical Considerations in MS-Based Glycan and Glycomic Analysis | 135 |
| 6.05.3.3.1 | Limitations in glycan mass profiling | 135 |
| 6.05.3.3.2 | MALDI versus ESI-MS and MS/MS | 135 |
| 6.05.3.3.3 | Other MS- and non-MS-based complementary techniques | 137 |
| 6.05.3.3.4 | Qualitative versus quantitative MS mapping | 137 |
| 6.05.4 | MS/MS Glycan Sequencing | 139 |
| 6.05.4.1 | Glycosidic Cleavages and Oxonium Ions | 139 |
| 6.05.4.1.1 | Neutral loss and identification of core structures | 140 |
| 6.05.4.1.2 | Nonreducing terminal oxonium and B ions | 141 |
| 6.05.4.1.3 | Elimination and linkage-specific Z ions | 142 |
| 6.05.4.2 | Cross-Ring and Other Linkage-Specific Cleavages | 142 |
| 6.05.4.2.1 | Negative ion mode sequencing of native glycans | 143 |
| 6.05.4.2.2 | MS ⁿ of permethylated glycans | 145 |
| 6.05.4.2.3 | High-energy CID-MS/MS on MALDI-TOF/TOF of permethylated glycans | 146 |
| 6.05.4.3 | Resolving the Glycomic Complexity by MS/MS | 147 |
| 6.05.4.3.1 | Identifying structural isomers by key diagnostic ions | 148 |
| 6.05.4.3.2 | Glycosylation and glycomic mapping at MS ² level | 148 |
| 6.05.5 | Targeted Glycomics, Glycoproteomics, and the Way Forward | 149 |
| 6.05.5.1 | Prospects for Targeted Glycomics | 149 |
| 6.05.5.2 | Prospects for Targeted Glycoproteomics | 150 |
| 6.05.5.2.1 | Current limitations in glycopeptide analysis | 150 |
| 6.05.5.2.2 | Prospects for emerging MS techniques and solutions | 151 |
| 6.05.5.3 | Concluding Remarks | 151 |
| References | | 152 |

6.05.1 Introduction

Mass spectrometry (MS) and various modes of separation techniques, coupled with chemical modifications and glycosidase digestions, are the cornerstones of the high-sensitivity approach in structural determination of complex glycans derived from biological sources. While conventional analyses performed on the glycans derived from isolated glycoproteins often entail additional use of techniques such as nuclear magnetic resonance (NMR) to complete the detailed structural characterization, particularly with respect to defining its stereochemistry, more recent advances in eukaryotic protein glycosylation analysis have been driven mostly by revolutionary technical developments in MS and proteomics. Notably, many of these glycosylation studies have been caught on the hypes and tides of 'omics' science and evolved into what is now commonly referred to as glycomics. To distinguish from its misnomer when applied only to analysis of glycans from isolated glycoprotein(s) of interest, glycomics in its true sense involves mapping the complexity and characteristics of the glycome of a cell, tissue, or organism, at a particular physiological or genetically altered state. A logical extension is then to further localize the identified glycans or glycoepitopes, particularly those novel ones or those with altered expression level, to its protein carriers, in what may be referred to as targeted glycoproteomics. These operational definitions are further elaborated in this chapter. Suffice here to point out the obvious, namely at such an -omic level, MS is unrivaled in combining precision with sensitivity, although not without its own shortcomings.

Currently, a handful of laboratories and experimental workflows dictate and set the trends in glycosylation analysis. Two seminal global initiatives have been completed recently under the auspices of the human disease glycomics/proteome initiative (HGPI),¹ which invited all experts in the field to use their favorite methods and workflows to complete the MS analysis of *N*- and *O*-glycans from the same centrally provided glycoprotein sample. The data obtained were then submitted to the coordinator and written into manuscripts for publication.² While the nature of this exercise prevented a true test of techniques against the most daunting task, the emerging picture is nonetheless most informative and educational to experts and newcomers alike. It somewhat objectively and accurately reflects the current strengths and limitations of various approaches, from which several guiding principles can be formulated. This chapter intends to build on that and extend beyond, with further considerations and thoughts given to the future of MS-based glycomics and glycoproteomics.

There is a long list of excellent recent articles covering various aspects of MS-based protein glycosylation analysis, ranging from giving a concise global view^{3–12} to focusing on more specialized aspects such as the various separation modes.^{13,14} Instead of providing yet another comprehensive account of all available methods, readers are referred to these articles and references cited therein. This chapter aims to deliver a more personal view of current practice and prospects in MS-based eukaryotic protein glycosylation analysis, with an emphasis on proven practicality in tackling real glycobiology problems, rather than dwelling specifically on the technical aspects and advances in MS instrumentation per se, although appropriate discussion and references will be given. It is written in a way that would hopefully give uninitiated readers a necessary background understanding and a clear path to tread when coming to terms with the need to analyze the glycan structures or glycoproteins at hand.

The need to confine the scope of this chapter to MS-based analysis of eukaryotic protein glycosylation is simply to avoid overlapping with other chapters in this volume. Partly, it is also reflecting a significant difference in the choice of analytical methods when tackling specifically this class of glycoconjugates versus the others. MS analyses of other eukaryotic glycosylation not specifically covered include those of polysialic acids, sulfated glycosaminoglycans and proteoglycans, the glycolipids including the glycosyl phosphatidylinositol (GPI) anchor, and the single *O*-GlcNAc modifications on a wide range of cytoplasmic and nuclear proteins. Analysis of bacterial and protozoan lipoglycans (see Chapter 6.06, this volume), and polysaccharides from bacterial, fungal, and plant cell walls, are also not specifically discussed. Naturally, some general principles described in this chapter apply equally well to those classes of glycoconjugates but each also involves its own specialized treatment and considerations. A convergent point for the MS analysis of all these disparate glycoconjugates, including eukaryotic glycoproteins, are the released glycans depolymerized to a size of below 5000 Da or so for them to be amenable to effective MS/MS sequencing. The route to this juncture is nevertheless not so straightforward for some of the aforementioned classes of glycoconjugates, and therein lie the specialized efforts in sample preparation that are beyond the current scope.

A consistent theme developed here is the emphasis on designing the right MS analysis strategies and workflows based on a sound, if not thorough, understanding of the biology of protein glycosylation. The glycobiology governing the norm of eukaryotic protein glycosylation will thus be first described from the perspectives of MS analysis in Section 6.05.2, before dwelling on the commonly adopted workflows in glycosylation analyses in Section 6.05.3. Section 6.05.4 will then be devoted to more advanced aspects of MS/MS analysis for glycan sequencing, before closing with a personal view on the future prospects of targeted glycomics and glycoproteomics in Section 6.05.5.

6.05.2 Diversity in Eukaryotic Protein Glycosylation

A fundamental difference between prokaryotic and eukaryotic glycosylation is the reliance of the latter on the endoplasmic reticulum (ER) and Golgi-resident glycosyltransferases to tread out well-conserved glycosylation pathways. From the most primitive protozoa to the highest order human, the eukaryotic protein glycosylation event follows essentially a few well-characterized pathways,^{15,16} resulting in readily recognized common core structures attached to either Asn (N-linked) or Ser/Thr (O-linked). Diversification resides mostly in the chemical nature of chain elongation, branching, and termination (see Section 6.05.2.2). In contrast, bacteria, by virtue of lacking the ER–Golgi assembly, do not follow such a protein glycosylation event and the increasingly being identified cases of bacterial protein glycosylation¹⁷ display a much more diverse range of esoteric structures with many monosaccharide residues not found in mammalian glycans.

Understanding these fundamental differences in eukaryotic and prokaryotic protein glycosylation is highly relevant when applying MS for structural analysis. Essentially, as will be elaborated in the later sections, current glycoanalysis and data interpretation are very much knowledge based, facilitated by a few well-accepted assumptions. Indeed, all the chemical details related to the conserved structural features have rarely been thoroughly reestablished nowadays, when analyzing mammalian structures of very limited amount using MS. This largely alleviates, although not eliminates, the otherwise intractable problems of not being able to effectively distinguish stereo- and structural isomers by MS analyses alone. NMR, on the other hand, may represent a better tool when sample amount is not limiting. A sound understanding of the glycobiology, of what are to be expected from the sample being analyzed, is therefore an important prerequisite, as described below.

6.05.2.1 Basic Tenets of Protein Glycosylation in the Context of MS Analysis

6.05.2.1.1 Factors determining the glycosylation structures – how much can be assumed *a priori*?

A simple concept regarding glycosylation is that its occurrence and eventual structure can neither be predicted *a priori* nor is it template driven akin to protein translation and mRNA transcription. The direct consequence is that this information can only be determined empirically, although prior knowledge of sample sources can limit the expected repertoire. A related concept is that glycosylation is determined by both protein- and host-specific factors. Encoded within the protein sequence is the eventual secondary and tertiary conformation to be adopted upon translation, which would constrain the microenvironment surrounding the exposed glycosylation site for the glycosylation machinery to act on. Thus, neither all sites nor all proteins synthesized by the same host cell carry the same glycan structures. The additional cues that result in instances of protein-specific glycosylation are mostly not well understood. In a more general sense, these protein-specific factors determine the potentials of site-specific glycosylation within the same protein. It is akin to the blueprint of the desirable glycan structures to be made. However, whether this blueprint is accurately and faithfully executed will be dependent on the host-specific factors.

Namely, if a particular nucleotide sugar substrate, transporter, and/or glycosyltransferase is not available, either naturally so or effected by an oncogene developmental event, aberrant glycosylation will occur, as in cancer or cell activation/differentiation. This host dependence also prescribes cell- or tissue-specific glycosylation structures on the same protein sequence simply because the availability of host materials and glycosyltransferase ensemble is different. Another well-known case example is that when a recombinant glycoprotein is expressed in a heterologous system, the eventual glycosylation structures will take on those of the host cells

instead of the native forms. A useful inference from the host-specific factors is that if one knows where the sample came from, one can rule out a certain structure or epitope while expecting others to be present. This not only simplifies MS analysis, but also affects the choice of sample preparation. A most obvious case is the use of peptide-*N*-glycosidase (PNGase) A and not PNGase F to release the *N*-glycans if the sample comes from a plant or insect or another lower animal for they may carry core α 3-fucosylation that renders them resistant to PNGase F digestion.¹⁸ It should also be pointed out that in nature, not all glycosidic linkages exist since the energetically unfavorable process of forming a particular glycosidic bond must be catalyzed by one or more corresponding glycosyltransferases. Thus, the knowledge-based interpretation will consider the assignment of certain linkages unlikely while other linkages including its anomeric configurations are to be expected under normal circumstances. In addition to what will be described in Section 6.05.2.2, readers are referred to the recently updated excellent textbook, *Essentials of Glycobiology*,¹⁶ which is also freely available online on the National Center for Biotechnology Information (NCBI) bookshelf, for an account of commonly occurring structures to accrue the requisite knowledge base. Additional references are not further cited here.

6.05.2.1.2 Heterogeneity and glycoforms – how much is enough?

Extending from the nontemplate-driven nature of glycosylation is the phenomenon of heterogeneity. Simply put, it is more common to find tens, if not hundreds, of different glycan structures attached to a particular site or protein. Each protein or peptide of the same sequence but differing in the glycan structures it carries is referred to as a glycoform. Resolving the full complexity of glycoforms is a daunting challenge in glycoanalysis, be it of a single site or the entire glycome. It is besieged by the same problems facing proteomics, namely, the sensitivity and dynamic range of MS instruments, and similarly calls for fractionation and enrichment as the only viable strategies to enable detecting those very minor but potentially most biologically relevant glycoepitopes. It also affects formulating the right answer to the perpetual question of how much starting material is enough for MS-based glycoanalysis. Perhaps a more appropriate counterquestion is how much depth one wants to fathom the glycome under investigation, with a sound understanding of the critical concept of ‘beyond detection limit.’ Equally important is to avoid the fallacy of discounting the presence based on failure to detect, namely, the false negatives. On the other hand, false positives are mostly due to isobaric and/or isomeric structures, particularly those not supported by MS/MS analysis and thus wrongly assigned by simply fitting the composition to the observed m/z values.

6.05.2.1.3 Site- and protein-specific glycosylation – why glycopeptides and glycoproteomics?

In most cases, the *raison d'être* for analyzing glycopeptides instead of just the released glycans and the deglycosylated peptides from a single isolated glycoprotein is to delineate site-specific glycosylation pattern, which may or may have glycobiology consequences. At the glycome level, this additionally translates into answering the specific questions of how widespread is a glycoepitope of interest distributed within the glycoproteome. It can be expected that, more likely than not, any alteration in glycosylation event would affect not just a single glycoprotein but several, if not the entire glycoproteome. One of the yet to be realized promises of glycoproteomics is thus to furnish the answers on which protein-specific factors may determine the distribution of a particular glycoepitope, and what would be the collective effects, synergistic or antagonistic, if all these proteins bearing the common epitopes were recognized and engaged simultaneously. Would the subtle site-specific distribution among the glycoproteins set the activation threshold or attenuate the critical outcome? Understanding the glycobiology issue in hand will help the decision making on whether to embark on the more challenging tasks of defining site-specific glycosylation and glycoproteomics, which at present can only be solved by MS analysis and no other techniques.

6.05.2.2 Structural Architecture and Diversity

As discussed above, MS analysis of protein glycosylation is very much driven by our knowledge of the common structures to expect. Due to the biosynthetic constraints imposed mainly but not solely by the regulated expression of various glycosyltransferases that need to act in concert to bring about a certain structure, the glycomic repertoire of a cell is essentially defined by its complement of expressed glycosyltransferases. Viewed

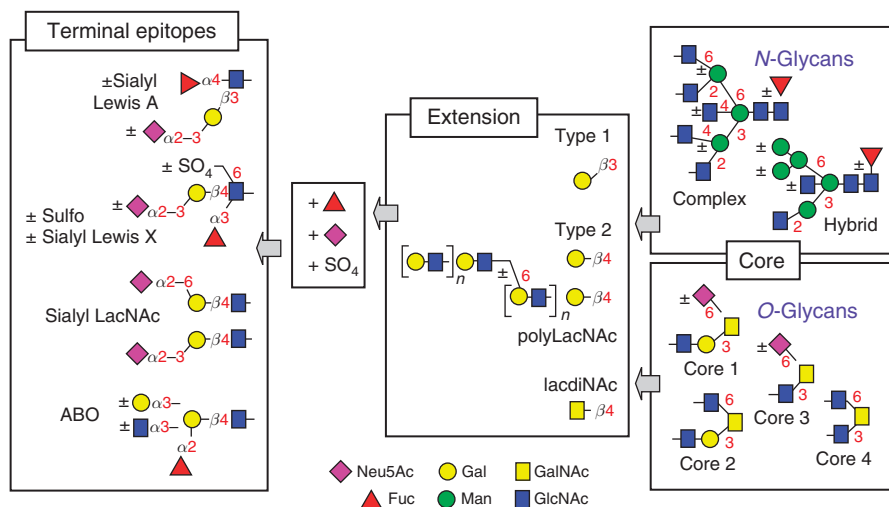


Figure 1 The common structural architecture and diversity of mammalian protein glycosylation. The core structures of the hybrid- and complex-type *N*-glycans and *O*-glycans can be similarly extended by type 1, type 2 (lacNAc), or the lacdiNAc chain, and then sialylated, fucosylated, and/or sulfated at different positions to give the unique terminal epitopes. Only a few representatives are illustrated here.

from the structural perspectives, especially in the context of MS analysis, the common structural architecture and diversity of mammalian protein glycosylation can best be summarized as depicted in **Figure 1** and further elaborated below.

6.05.2.2.1 Types of glycosylation

Eukaryotic protein glycosylation is mainly categorized into *N*- and *O*-glycosylation. *N*-Glycosylation at the consensus sequon Asn-Xxx-Ser/Thr is relatively well conserved throughout all phyla and corresponds essentially to only one kind, namely that initiated by a trimannosyl core structure. In contrast, *O*-glycosylation on Ser/Thr does not follow any strict consensus sequence. In addition to the most common mucin-type *O*-glycosylation initiated by GalNAc addition, other eukaryotic *O*-glycosylation includes *O*-fucosylation, *O*-glucosylation, and *O*-mannosylation, all of which can be further extended in a specific manner. *O*-Xylosylation occurs on proteoglycans to give rise first to a linkage region of tetrasaccharides and is then further polymerized into the sulfated glycosaminoglycans. The nuclear and cytoplasmic *O*-GlcNAcylation, on the other hand, is not further extended. There are a few other less commonly found *O*-glycosylations, which one should be aware of when the samples to be analyzed are derived from lower animals, plants, fungi, and protozoa.

6.05.2.2.2 The common core structures

The core structure of *N*-glycans is invariably the trimannosyl chitobiose unit, Man₃GlcNAc₂, which can be α 6-fucosylated at the reducing end GlcNAc. In plants, this core fucosylation is α 3 where it occurs, with and without additional β 2-xylosylation at the β -mannose. In insects, either or both types of core fucosylation can occur but not β 2-xylosylation, whereas in some other lower animals that have been looked at, including mollusks and nematodes, instances of all these core modifications could be found.

The simplest eukaryotic *N*-glycans are the so-called truncated or pauci-mannose structures, which correspond to this trimannosyl core, or its further trimmed-down versions to having only two or one mannose, with and without the aforementioned core modifications. These *N*-glycans represent the most commonly found structures in plants and insects but are not normally found in mammals and vertebrates. In higher animals, the high mannose-type structures retain 5–9 of the original mannose transferred *en bloc* as Glc₃Man₉GlcNAc₂ during biosynthesis but are not trimmed beyond that without concomitant addition of a GlcNAc to the Man on the 3-arm of the trimannosyl core. Further removal of the two additional Man on the 6-arm would allow the

addition of one or more GlcNAc to the Man on the 6-arm to give the complex-type *N*-glycans. If no additional GlcNAc is added to the 6-arm and only the 3-arm is extended, the resulting *N*-glycans are referred to as hybrid type. The addition of the GlcNAcs onto the Man follows a strict specificity with defined linkages and initiates the multiantennary nature of *N*-glycans.

The mucin-type *O*-glycans are attached to Ser/Thr through one of the many core-type structures. Among these, only cores 1 and 2, and to some extent, cores 3 and 4, are commonly found. Structurally, the former pair differs from the latter by having an extra Gal β 1-3 linked to the reducing end GalNAc, before being extended further or capped by sialylation. The additional β 6-GlcNAc branching on the 6-arm would make the *O*-glycans biantennary-like with important glycobiochemistry consequences; this can be contrasted with the complex-type *N*-glycans, which can have up to penta-antenna, with and without the additional bisecting GlcNAc on the β -Man.

The GlcNAcs on each of the antennae, *N*- and *O*-glycans alike, are normally β 4- or β 3-galactosylated to give rise to the common structural unit referred to as type 2 or type 1 chain, respectively. The type 2 chain, that is, Gal β 1-4GlcNAc or *N*-acetylglucosamine (LacNAc) can then be further extended to either linear or branched polyLacNAcs, whereas the corresponding type 1 chain, Gal β 1-3GlcNAc, is rarely extended. Less commonly, one or more of the GlcNAcs on the Man are incompletely galactosylated and thus remain as terminal GlcNAc. Another variation is its extension by β 4-GalNAc to give the so-called *N,N'*-diacetylglucosaminylamine (lacdiNAc) unit. Collectively, these disaccharide units form the peripheral chains elongated from the well-defined core structures and represent the least variable parts of the common *N*- and *O*-glycan structures. The *O*-linked Man and Fuc can likewise be extended by the LacNAc unit and then capped by sialylation.

6.05.2.2.3 Terminal glycosylation from the perspective of MS analysis

By far, structural variation and diversity reside in the kind of terminal glycosylation that is added onto the peripheral glycan chains extending out from the aforementioned common core structures. The resulting glycoepitopes, or glycotopes, are too numerous to be summarized here but are essentially derived from a combination of α 2/3/4-fucosylation, α 2-3/6-sialylation, and/or sulfation, among the more common ones for the mammals. It should be noted that depending on the sample preparation, chemical derivatization, and analytical methods employed, some of the nonsaccharide substituents such as O-acetylation and O-methylation may be rendered cryptically or lost completely, and thus their wider occurrence is yet to be fully appreciated. It is also likely that some novel substituents with physicochemical properties that defy MS analysis have yet to be detected and defined through our routine screening analyses. Still other modifications are conveniently ignored or denigrated to artifacts as their mass increments cannot be fitted to any known molecular entities. Notwithstanding, the extensive data accumulated in the past on the common mammalian structures, together with current advances in identifying all the relevant glycosyltransferase genes from the genome, do furnish us a good command of the mammalian glycan repertoire, especially of humans and mice.

Much less is known about other eukaryotic glycosylation but what is known does support the notion that the core structures are relatively well conserved across the different phyla, whereas much divergence is observed for the terminal glycoepitopes. Unfortunately, there is as yet not a good, comprehensive, and searchable database that catalogues all those 'unusual' structures that have ever been characterized, although coordinated efforts are underway.¹⁹ On the other hand, it is unlikely that the full repertoire of glycan structures existing in nature would ever be completely characterized to allow us to claim such knowledge. Thus, moving away from the safety zone of mammalian glycans, one is increasingly being challenged with novel glycoepitopes and thereby no longer enjoys the luxury of knowledge-based assumptions. More efforts are actually required to nail down the structural details including linkages and stereochemistry. Even so, MS is still the most powerful analytical tool as it provides a distinct mass value as a starting point to facilitate the interpretation of other complementary data.

6.05.3 Workflows in MS-Based Protein Glycosylation Analysis

The general workflow in MS-based protein glycosylation analysis may be viewed as comprising three stages, from sample preparation to data acquisition and then data analysis (Figure 2), preceded first by a strategic planning. As described in Section 6.05.2, the strategic planning involves a thoughtful consideration of the sample sources and therefore what kind of glycans to be expected, since these will affect the sample preparation

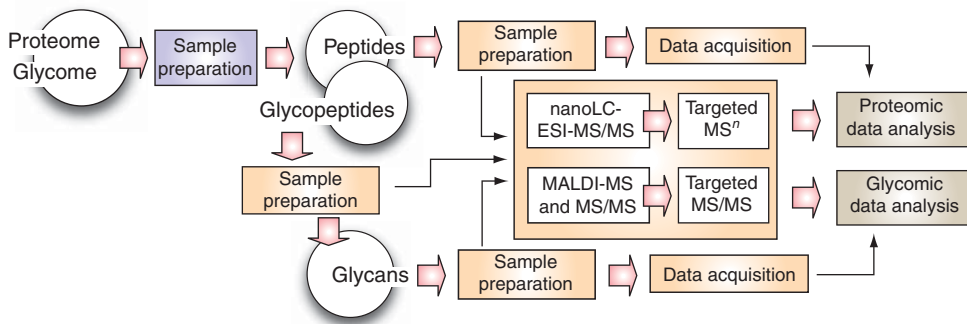


Figure 2 The main stages in the general workflow of an MS-based global-scale protein glycosylation analysis.

steps. It also involves the decision whether to analyze just the released glycans or to address the site-specific glycosylation. For glycomics, the total proteomic or subfractionated proteomic extracts would be subjected to glycan release, usually after proteolytic digestion to assist solubilization and facilitate efficient enzymatic de-N-glycosylation. The de-N-glycosylated peptides, along with the original glycopeptides/peptides, can both be fed into the glycoproteomic/proteomic workflows, in parallel with or subsequent to glycomic analysis.

6.05.3.1 Defining the Conceptual Framework

A complete protein glycosylation profile includes both the structures of the attached glycans and their respective attachment sites. While mapping the glycan diversity on either a single isolated glycoprotein or an entire cell or organism may be completed on its own by analyzing only the released glycans, site-specific glycosylation relies on analyzing the glycopeptides and normally cannot be completely defined without also analyzing separately the released glycans. This is simply because current MS analysis cannot yet afford critical structural details of glycans when performed at the glycopeptide or intact glycoprotein level. Thus, the full range of glycan diversity particularly in terms of structural and stereoisomers can only be efficiently resolved and defined at the glycan level.

In other words, glycan analysis or glycomics forms an integral part of glycosylation site analysis or glycoproteomics. In most cases, it is performed first and the information obtained is used to design the requisite workflow for glycopeptide analysis or glycoproteomics (Figure 3). Furthermore, interpretation of the

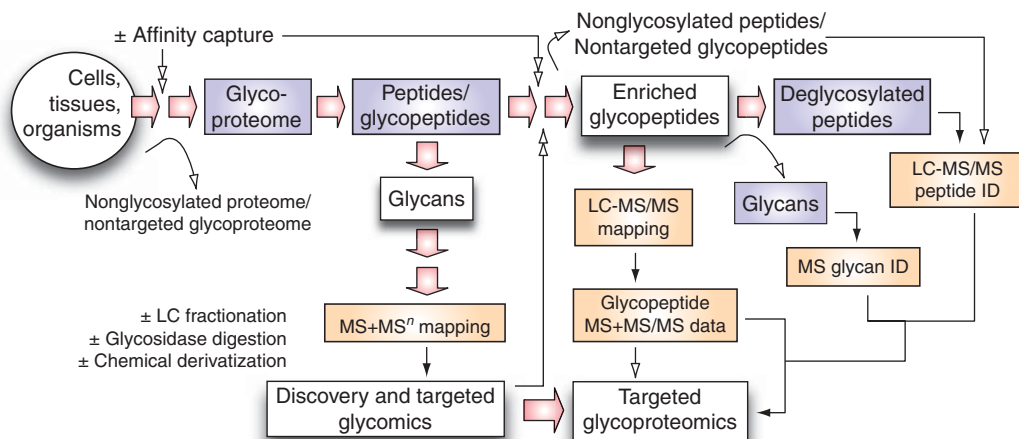


Figure 3 An integrated workflow of glycomics and glycoproteomics. Successful glycoproteomics is shown to require inputs from glycan, peptide, and glycopeptide MS and MS/MS data. It is also likely to require affinity capture at both the glycoprotein and the glycopeptide levels.

glycopeptide MS and MS/MS data, either manually or executed through any of the bioinformatics program under development, will always benefit from a knowledge of the range of glycans expected to be present on the particular subset of glycoproteome/glycopeptides under investigation. This can be used to constrain the search and match algorithm in the first instance, or as additional input parameters after the initial match to score and rank the hits.

By definition, glycoproteomics aims to identify the glycosylation heterogeneity associated with each of the glycosylation sites on each of the glycoproteins made by a cell, tissue, or organism. This is necessarily a very time- and labor-consuming daunting task, as well as technically challenging, if not insurmountable. The overriding considerations in whether to launch such an undertaking at all normally reside in not only the availability of sufficient sample amount, but also a sound justification for true biological importance. While such systematic glycoproteomics is the ultimate goal of omics studies in its true grandeur, a more realistic undertaking takes the form of a targeted approach in either of the two perspectives.

A comprehensive site-specific glycosylation analysis is only performed on one or several isolated glycoproteins of interest, with proven biological relevance, or simply selected randomly based on abundance or ease of isolation. Thus, in essence, the venture is downgraded to single glycoprotein analysis. Alternatively, information from comparative glycomic mapping, which normally identifies one or many novel glycoepitopes or biologically relevant glycoepitopes that are up- or downregulated in expression during a pathophysiological event, can be used to devise a targeted glycoproteomic approach. The question formulated then is which subset of the glycoproteome bears the particular glycoepitope of interest at which of their glycosylation sites, on which are specific glycan carriers. A key factor in enabling such an approach is obviously the ability to design a target enrichment or capture scheme against the glycoepitope of interest, which can ideally be applied at both glycoprotein and glycopeptide levels, along with meeting all other pre-MS analysis sample requirements.

6.05.3.2 Pre-MS Sample Preparation Requirements

As in proteomics, the practical usefulness of MS analysis, be it applied to a single glycoprotein or entire glycoproteome, is ultimately limited by the ability to detect those biologically important structures or glycoepitopes of low abundance, some of which are refractory to efficient ionization and detection. It is currently unknown how many orders of dynamic range are needed to cover both the most and the least abundant components. It is also generally acknowledged that different extraction, solubilization, glycan release, fractionation, and derivatization methods may bias against a particular subpopulation. Any combination of the sample preparation methods may be used and is generally not a problem for an initial glycomic screening, which – provided the starting material is of sufficient quantity – will yield data one way or the other. The more difficult issue is whether the resulting profile is at least quantitatively representative, if not comprehensive. It is common that further fine tuning of the sample preparation methods often leads to ‘discovery’ of additional glycan species and is in fact needed to ‘see’ some of the more difficult classes of glycans such as the sulfated glycans.

A general workflow will therefore prescribe a rough and quick first-screen strategy followed by more detailed or targeted analyses, which may require additional chromatographic fractionation and/or chemoenzymatic manipulations. Most laboratories specialized in glycosylation analysis have their own preferred methods of sample preparation and only general considerations will be given here to provide some starting guidelines.

6.05.3.2.1 From biological source to solubilized glycoproteins

Depending on the studies, the starting biological source can be biological fluids or cultured or isolated cells and tissues.^{7,20} Different developmental stages of small animals such as *Drosophila*,^{21,22} zebrafish,²³ and worms^{24,25} have also been directly subjected to direct homogenization and glycoproteomic extraction. For glycomic analysis, extensive subcellular fractionation is normally not necessary unless a distinction is to be made on glycans derived only from the plasma membrane glycoproteins and not those in transit or turnover and thus residing in ER/Golgi or cytosolic compartments. Otherwise, total lysates can be directly used for extraction. On the other hand, if the analysis of the membrane glycoprotein carriers is intended, a crude ultracentrifugation to obtain at least the microsomal fraction, separated away from the abundant cytosolic proteins, is desirable.

It should be noted that apart from secreted glycoproteins as found in biological fluids or cultured media, all other cellular glycoproteins are membrane glycoproteins. A popular workflow in extraction is to first delipidate the cell lysates by organic solvent, usually by a combination of chloroform, methanol, and water although other solvent composition with and without salts is possible. This fraction may be used for the analysis of glycolipids if so desired. However, optimizing the recovery of glycolipids normally entails repeated rounds of extractions with organic solvents of increasing polarity and may compromise the yield of glycoproteins from the remaining cell pellets. In contrast, only a simple step of delipidation with fairly nonpolar organic solvent is needed if only the glycoproteins are to be profiled. Following delipidation, the cell pellet is then commonly solubilized by extraction with urea or guanidine chloride, followed by dialysis. Solubilization by pronase or proteinase K digestion is not recommended since, in general, one aims to preserve the peptide backbone for glycoproteomic analysis, although it can be useful if one is concerned only with glycomics.

Alternatively, glycoproteins can be solubilized and extracted directly from the total lysates or microsomal fraction by detergents such as sodium dodecyl sulfate (SDS), 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), and Triton X-100 among the most commonly used, without the first delipidation step. This tends to produce cleaner extracts and sometimes is necessary if the conformation of glycoproteins needs to be preserved for subsequent affinity capture or activity assays in purification. However, a common problem for MS analysis is the subsequent removal of detergents, which is not trivial. The selective loss introduced by methods such as trichloroacetic acid, acetone, or ethanol precipitation, or passing through specialized detergent-removing columns including ion exchange, can be an important factor that needs consideration. Others have introduced gel-based fractionation and digestion to overcome the problem of detergent removal. Full resolving 1D sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) can be employed with an added advantage of reducing the complexity of total extracts into as many gel slices as needed, as commonly adopted in shotgun proteomic analysis. The drawback is that in-gel digestion is required followed by efficient extraction from the gel. Direct release of *N*-glycans from the gel without prior proteolytic digestion has also been commonly employed but the respective yield from all these different methods is difficult to assess and has not been systematically addressed.²⁶ On other occasions when an immuno-pulled-down glycoprotein is to be analyzed, this gel-based approach may represent the only viable means of effective purification despite its noted shortcomings.

In general, efficient solubilization and recovery of membrane glycoproteins in a form that is relatively free from salt and detergent for MS analysis is a daunting task. This raises the concern that the so-called glycomic analysis is not more than just profiling the most readily solubilized and extracted glycoprotein subsets. Many problematic classes such as the heavily glycosylated mucins or those integral membrane proteins with many transmembrane domains may be underrepresented at best. On a more positive note, it is fair to say that most glycoepitopes characteristic of a cell are likely to be carried on more than just a few proteins and therefore would still be represented in the acquired glycomic profile, although this presumption remains to be validated. Targeted analysis of single glycoproteins, on the other hand, would necessitate an efficient affinity capture in the overall isolation scheme. Here, the prime difficulty is first and foremost the very limited amount of the eventually purified endogenous glycoproteins and, in the process, certain glycoforms may be biased against and likewise underrepresented in the successfully recovered population.

6.05.3.2.2 Glycan release and reducing end tagging

As described in Section 6.05.3.1, for concerted glycomics and glycoproteomics analysis, or for simple site-specific glycosylation profiling of a single glycoprotein, the preferred next step in the workflow after obtaining the glycoproteins is to digest them into peptides/glycopeptides; this supplies the sample not only for proteomic and glycoproteomic identification but also for efficient glycan release in glycomic applications. It is generally understood that intact glycoproteins often cannot be enzymatically de-*N*-glycosylated efficiently without denaturing first in the presence of detergent. This approach nonetheless introduces detergent and can only be sensibly coupled with gel-based analysis and glycan release.

The *N*-glycans can be efficiently released intact from glycopeptides by PNGase F, or PNGase A if core α 3-fucosylation is expected. This converts the Asn in the consensus *N*-glycosylation sequon into Asp, with a one mass unit increment. This change serves as a useful mass tag to identify originally *N*-glycosylated peptides. Better still, a range of endo-*F*-glycosidases can be used along with endo-*H* to separately release complex and high mannose-type glycans, while leaving behind a GlcNAc attached to Asn. This has the advantage of a larger

mass increment (+203 u) for the deglycosylated peptide compared to the Asn to Asp conversion (+1 u) and avoids the trapping of spontaneous deamidation, which may mistag a non-N-glycosylated site. However, for all its promises, endo-F appears to be less efficient in N-glycan release and thus results in low recovery of N-glycans and incomplete peptide de-N-glycosylation for subsequent analyses. Herein lies the general pros and cons of enzymatic treatment. The main advantage offered is the nondegradative nature toward the released glycan chains and the deglycosylated peptides, compared with chemical methods such as hydrazinolysis. The drawback is its sometimes overspecificity and sensitivity to steric hindrance, thus resulting in incomplete digestion. In the case of O-glycans, the only O-glycanase available targets specifically the bare core 1 structure, Gal β 1-3GalNAc-Ser/Thr, which renders its use as a general reagent to release intact O-glycans infeasible. Thus, O-glycans are commonly released instead by the well-tested alkaline reductive elimination method, which unfortunately would destroy the de-O-glycosylated proteins/peptides. Other O-glycan chemical release methods including less degradative and nonreductive treatments have been proposed²⁷ and used to different extents by different groups, to variable success. The main shortcomings are normally low efficiency/recovery and/or unwarranted degradation.

For a number of laboratories, the nonreductive means of O-glycan release is essential to preserve the reducing end for subsequent reductive amination tagging with a fluorescent probe for liquid chromatography (LC) applications,^{28,29} in a manner similar to that employed for N-glycans. 2D or even 3D high-performance liquid chromatography (HPLC) mapping of such fluorescent-labeled glycans is a powerful approach to structural analysis complementary to MS analysis.^{30–32} Its main advantages include better orthogonal fractionation scheme afforded by wider choice of stationary phases, usually in the form of anion exchange, C18 reverse phase, and amide-based normal phase separations. It also allows relatively easy purification and quantification of each separated glycan species based on the fluorescent tag. On the other hand, its main limitation is the reliance on standards or previously established database for the 2D LC coordinates for peak identification, which is not possible when novel structures are encountered. Some of the applications have in fact been facilitated by MS analysis of the isolated glycan peaks. It is, however, an overall laborious approach and unlikely to resolve the full complexity of a glycome on its own. Recent developments have spotted a trend to miniaturize the fluorescent label approach for microscale handling and separation on capillary LC columns that can be interfaced directly to MS.^{33,34} This would enjoy the advantage of automated MS analysis while providing an added dimension of efficient LC fractionation, with ease of quantification. Considered in this context, this approach is ultimately limited by the chemistry of tagging and subsequent cleanup, which never approaches complete yield. It also necessitates an efficient online MS/MS sequencing capability that is compatible with the chromatographic timescale if structural identification or confirmation is required.

6.05.3.2.3 Permethylation and other related chemical derivatization

Reducing end tagging apart, another of the most commonly used chemical derivatizations is permethylation. Along with peracetylation, and their stable isotope counterparts, that is, perdeuteromethylation and perdeuterioacetylation, these chemical derivatizations essentially convert all free hydroxyl groups on the glycan chains into O-methyl or O-acetyl groups. An essential difference is that the O-methyls are stable to most subsequent chemical treatment including acid hydrolysis, whereas the O-acetyls, being esters, are readily reversible. Permethylation is the necessary first step in linkage analysis,^{35, 36} followed by hydrolysis, reduction, and peracetylation for gas chromatography-electron impact (GC-EI)-MS analysis. By converting a hydrophilic glycan into a hydrophobic molecule, permethylation allows simple cleanup from hydrophilic salts and contaminants and thus is well suited for robust glycomic applications in which a low amount of glycans derived from the cell extracts often come with a high level of many of the biological matrices. In comparison, MS analysis of native glycans often demands more tedious pre-MS cleanup to ensure good ionization amidst hydrophilic contaminants.

In general, both permethylation and peracetylation also increase the sensitivity of MS detection *per se* by up to 10 fold or more on a side-by-side comparison when equal amounts of native and derivatized analytes are applied to MS. The practical gain in sensitivity is, however, much more difficult to evaluate as many factors in biological sample handling come into play. Most notably when dealing with low or subpicomolar amount of sample, direct applications of the native glycan sample to MS analysis may turn out to be more successful than taking the same limiting amount through permethylation, concomitant with the loss introduced during the chemical derivatization and cleanup. The situation is somewhat different when dealing with nanomolar to

micromolar range of sample materials for which small loss during permethylation is negligible and more than compensated by the gain in other aspects, notably in deriving useful MS/MS data (see Section 6.05.4). This added advantage in affording more effective MS/MS sequencing along with its robustness and proven applications to real samples are in fact the prime considerations in continuing to favor permethylation today even with vastly improved MS and MS/MS sensitivity. Methods were developed to introduce online solid-phase methylation,³⁷ and analysis of permethylated glycans remains the preferred method adopted by several leading laboratories² that continue to solve novel glycan structures by MS analysis.

An important consideration to bear in mind when one chooses to work with a permethylated sample is that it will methyl esterify carboxylic groups, but will also remove any naturally occurring functional group attached to the glycan through ester bond, most notably the *O*-acetyls on sialic acids. Consequently, it abolishes the negative charges born by sialic acids and hexuronic acids, thereby stabilizes them and renders glycans with these residues more amenable to analysis in positive ion modes. At the same time, it will remove *O*-acetyl and other *O*-acyl substituents while rendering naturally occurring *O*-methyl groups cryptic. These should be 'uncovered' by parallel analysis of nonderivatized samples or be dealt with when finally zooming in on the analysis of glycopeptides. On the other hand, either the Hakomori methylation conditions³⁸ using the methylsulfinyl carbanion or the NaOH/DMSO slurry³⁹ methods have been shown to retain the phosphates⁴⁰ and sulfates.^{41–44} A phosphodiester will itself be methylated and the phosphate monoester may acquire one or two methyl groups, whereas the sulfates will not be methylated and thus retain its negative charge. These have important bearings in designing sample cleanup, fractionation, and enrichment schemes for the sulfated, phosphorylated, and/or sialylated glycans. A different strategy involves the analysis of native glycans but with their carboxylic groups on sialic acids being neutralized by methyl esterification⁴⁵ and other novel chemistry.^{46–48} These have the same desirable effects of stabilizing the sialic acids, abolishing their negative charges, and enabling them to be analyzed along with other neutral glycans in positive ion mode without being biased against or induced loss of sialic acids.

6.05.3.2.4 Glycan fractionation and targeted enrichment

As a starting point, additional fractionation of the released and permethylated glycans are normally not necessary prior to attempting a rapid survey MS mapping. It is more useful to have the feedback from such initial screening to inform the most sensible fractionation steps, if needed. As mentioned, permethylation followed by a simple organic solvent extraction is normally sufficient to get rid of most interfering hydrophilic impurities. It also neutralizes charges on sialic acids to allow simultaneous detection across the full mass range of most common classes of *N*- and *O*-glycans in positive ion mode. Depending on the complexity and the kind of potentially interesting or important glycoepitopes detected, judicious choice of fractionation scheme can then be made. This can be effected at the bulk preparative level or at the microscale level to facilitate MS detection. Further size separation of the permethylated glycans can be performed on reverse-phase LC but in general gives poor chromatographic resolution. Instead, it is practically more useful to have the C18 materials packed in solid-phase extraction cartridge or microtip for pre-MS microscale cleanup and simple class separation. Permethylated sulfated glycans, for example, can be partially separated away from neutral permethylated glycans and thus enriched. Similarly, a microtip packed with porous or nonporous graphitized carbon material⁴⁹ can be used to clean up nonderivatized native glycans prior to MS applications.

Conventional detailed glycan analysis carried out at a relatively larger scale often necessitates isolation of glycans through multidimensional chromatography to allow efficient NMR and compositional analysis. Anion-exchange chromatography takes advantage of the different sialic acid contents and is usually employed as the first dimension to separate the glycan pool into neutral, singly, doubly, triply, and multiply negatively charged fractions, corresponding to the number of sialic acids contained. Amine- or amide-based normal phase is then employed for size fractionation with the charge effects counteracted by the pairing ions included in the solvent buffer. Reverse-phase separation is only applicable to glycans tagged with hydrophobic fluorescent label such as aminopyridine³¹ or 2-aminobenzamide.³⁰ While these methods are still very much in use, current glycomic mapping prefers to have simple one-dimensional fractionation and, if possible, direct coupling to online LC-MS/MS, to increase throughput and sensitivity while relying on the high performance of current MS instruments to provide the resolution and precise sequencing needed. In this context, porous graphitized carbon-based microscale LC-MS/MS is increasingly employed to effect separation of structural isomers.^{50–54}

For MS-based analysis, this is arguably more useful than various modes of hydrophilic interaction liquid chromatography (HILIC), which provide mostly size fractionation.^{14,33,34,55} MS itself gives the highest resolution in terms of molecular mass and thus would be better complemented by LC fractionation that resolves not the size but the structural and stereoisomers of same masses. High pH anion exchange as performed on Dionex LC system,⁵⁶ which can likewise resolve isomeric structures, is also popular but the extensive online and/or offline desalting needed makes it less friendly for direct coupling to LC.⁵⁷ Finally, it is also desirable to have affinity-based fractionation to allow enrichment of a subglycomic population, usually effected through specific lectins and antibodies but only with limited success, with respect to the specificity achieved.

6.05.3.2.5 Glycopeptide enrichment

For effective MS analysis, glycopeptides as produced through proteolytic digest of a glycoprotein pool need to be enriched out from the dominance of peptides. This has in fact been one of the key areas in glycoproteomics that merits intensive research and development. Unfortunately, no single method emerges as the clear winner and each appears to demonstrate certain degree of success when applied to standard glycoproteins and selected glycoproteomic samples. It is fair to say that at a glycoproteomic level, any of the enrichment methods would give the desirable effect of allowing more glycopeptides to be detected on a global count. However, practical experience shows that there is no *a priori* rule to predict if a particular glycopeptide can be enriched as desired. A common technical problem is that certain glycopeptides will not be retained while others may not be eluted or recovered. Conversely, nonspecific binding of nonglycosylated peptides is inevitable. Among the stationary phases that rely on the added hydrophilicity conferred on glycopeptides by the attached glycan to effect the enrichment capture are various HILIC materials including cellulose, sepharose,⁵⁸ silica, aminopropyl, and zwitterionic types.^{59,60} Notably, none of these exhibit exclusive nonreactivity against the peptide itself and therefore, depending on the particular amino acid composition of the peptides, their weak to moderate or even strong binding to the HILIC phase is naturally unavoidable and is critically affected by the pH and ionic strengths employed for binding, wash, and subsequent elution.

In contrast to these materials that exploit the physicochemical properties of the glycans for global enrichment of glycopeptides, specific affinity capture can also be performed using one or more of the commercially available lectins.^{60–63} Some of these such as concanavalin A (ConA) and wheat germ agglutinin (WGA) have long been employed for conventional purifications of glycoproteins and are ranked among those lectins that have broad specificities against a range of glycans. Other lectins that target specifically the Fuc or sialic acids are very useful if the glycoproteomic experiment aims to define the changes related to fucosylation and sialylation, respectively. Still other animal lectins such as the galectins and mannan-binding proteins or monoclonal antibodies against specific glycotopes have been coupled to beads or gels for affinity capture. A common problem with these is the general low affinity and a requirement of multivalency for high avidity, which may not be reproduced on isolated glycopeptides. Often, an effective affinity probe for Western blot or flow cytometry applications may not translate into an effective tool for affinity capture, not to mention their prohibitively expensive cost.

On the other hand, the use of these protein-based affinity capture materials potentially allows a multilevel capture process to increase overall specificity. The same lectin can be employed at the glycoprotein, glycopeptide, and glycan levels. One can thus start with an already enriched targeted glycoproteome subset and resubmit its digested peptide/glycopeptide pool to another round of affinity capture by the same lectin. Nonretained peptides from this pool should in principle originate only from the targeted subglycoproteome and can be used for proteomic identification. Even allowing for the expected nonspecific binding, this would still narrow down the proteomic subset to be considered. In addition, glycomic mapping may initially be conducted on the subglycomic population enriched by the same lectin to obtain information on the expected ligands. It is also possible that several lectins be used in tandem or in mixture to further narrow down or increase, respectively, the targeted range of glycopeptides. The lectin affinity capture can also be used in combination with any of the HILIC-based enrichment, which is nondiscriminative against the glycan moieties.⁶⁰ This has the added advantage of possible desalting and removal of the high amount of monosaccharides often included in the eluting buffer. It should be noted that desalting and removal of other small-molecule contaminants from glycopeptides by C18 reverse-phase column are not always feasible as some of the glycopeptides, especially those with large glycan moiety relative to small peptides with charged amino acids, may not be retained. The HILIC column may be a good alternative but it should likewise be cautioned that binding and subsequent elution are not guaranteed.

6.05.3.3 Practical Considerations in MS-Based Glycan and Glycomic Analysis

As described in the previous sections, a preferred workflow following the release of glycans involves permethylation for a rapid mapping before attempting additional fractionations to reduce the glycomic complexity or to target certain subpopulation of interest. For reasons to be described in this section, the first survey mapping is preferably done by MALDI-MS supplemented by MALDI-MS/MS on selected peaks to confirm dubious peak assignment based on composition alone. Following this first step, many alternative routes are possible but generally boil down to treading one of the three pipelines (**Figure 3**). The first is aimed at confirming the glycan structures assigned by MS/MS (Section 6.05.4), usually with respect to linkage and anomeric configurations, while the second is to subject the glycans to further rounds of ever more sophisticated MS or MS/MS probing, with and without LC, to answer specifically if particular glycoepitopes or core structures not apparent from the initial mapping are present. It is anticipated that this second-tier target validation and identification will eventually be accompanied also by target quantification in the not too distant future although not currently feasible. Finally, a third route is to venture into targeted glycoproteomic studies based on the glycoepitopes detected to delineate their specific protein carriers.

6.05.3.3.1 Limitations in glycan mass profiling

A basic understanding of MS indicates that it measures the mass-to-charge ratio (m/z) of an ionized molecule. Knowing the charge can translate the m/z value into true molecular weight information. Current MS instruments boast of exceptionally high resolving power with respect to molecular masses but is essentially nondiscriminating with respect to isomers, unless the different isomers can be experimentally derivatized to derivatives of distinctive mass differences. In glycan terms, the structural diversity of the common mammalian glycans therefore degenerates simply into molecular entities comprising a combination of Hex, HexNAc, NeuAc/Gc, dHex, and less frequently, Pent and HexA, as their building block units. One cannot distinguish by MS, for example, α -Gal, β -Gal, α -Man, and β -Man. Instead, all these residues are simply designated as Hex. This somewhat simplifies the assignment as there are only a few residual masses to be considered and fit. With the knowledge of expected range of common structures, one can readily assign a measured m/z value to a particular composition, from which the most likely structures may be inferred.

It is, however, obvious that definitive structural assignment is not possible by such straightforward one-dimensional MS mapping. A common problem in assignment based on MS data alone, either manually or by using software programs, is that there are simply too many possible structural isomers that would fit the molecular mass data, even discounting the stereochemistry. The common practice is thus to apply matching constraints based on our knowledge of glycobiology along the lines described in Section 6.05.2. One would limit the combinatorial permutations of different monosaccharide residues to certain numbers and kinds and further manually narrow down the possibilities by eliminating those unlikely structural composition. As one can imagine, the more complex the glycome or glycosylation profile under investigation, and the larger the molecular weight, the likely isobaric and isomeric combination increase exponentially. These considerations place a premium on the need to obtain high-accuracy MS data and, above all, in acquiring MS/MS for as many individual peaks as possible within the life span of a limiting amount of analytes.

6.05.3.3.2 MALDI versus ESI-MS and MS/MS

The history of glycan MS analysis commenced in the 1980s and the early 1990s when fast atom bombardment (FAB) or liquid secondary ionization MS (LSIMS) on the sector type of instrument^{64,65} was the only viable MS ionization technique for efficient analysis of oligosaccharides. In addition to affording molecular ions for intact glycan molecules up to several kilodaltons, direct fragment ions, particularly those referred to as oxonium ions resulting from glycosidic cleavages (see Section 6.05.4), were often observed. These fragment ions are not true MS/MS ions in the sense that parent ions are not selected or isolated by the first mass analyzer and subjected to additional collision-induced dissociation (CID) events in a dedicated collision cell or ion trap, before being analyzed in a second mass analyzer tandem in space, or the same ion trap tandem in time. The beauty of a true MS/MS event is that all observed fragment ions can be attributed to a single parent ion and hence most informative for detailed sequencing of a glycan. However, FAB-MS/MS analysis as implemented then on the four-sector tandem mass spectrometers was mostly of insufficient sensitivity for it to be practically useful for

analyzing biological samples. In contrast, the abundant direct fragment ions produced in-source upon FAB ionization were used extensively to identify the range of terminal epitopes and polyLacNAc-based peripheral sequences present, much as one would with true MS/MS. The only drawback is that these fragment ions cannot be traced back to a particular parent ion unless the glycan of interest is first purified and analyzed in isolation.

Recent advances in MS instruments have largely supplanted both FAB ionization and sector instruments. For newcomers to this field, one is mostly acquainted with and has access only to matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) sources fitted on a range of mass analyzers with much improved performance, such as quadrupole/time-of-flight (Q/TOF), TOF/TOF, and ion traps that allow efficient implementation of CID-MS/MS. Both MALDI and ESI are much softer ionization techniques compared with FAB and deemed not to impart sufficient energy to induce much in-source fragmentation. However, it is well appreciated that direct fragmentation in the ionization source can still be observed albeit it is less prominent. In MALDI, this in-source prompt fragmentation can be further promoted or 'laser-induced' by elevating the laser energy above the normal ionization level needed.⁶⁶ In ESI, similar in-source fragmentation can be induced if the cone voltage is somewhat elevated above normal settings. From the application aspect, these direct fragment ions that come 'free' with MS mapping are very useful to derive diagnostic markers indicative of the presence of particular glycoepitopes within the population for the initial discovery screening. An important feature is that such fragment ions, as with all other molecular ions, can actually be selected for further MS/MS in what is commonly referred to as a pseudo-MS³ event. In practical terms, this would allow the oxonium ions representing the biologically relevant terminal epitopes to be further fragmented to derive additional linkage information.⁶⁷ The only limitations are that, in comparison with FAB-MS, they may not always be produced at a significant level, especially if the sample amount is low, and are often masked by a high-matrix background within the low-mass region where they occur.

All factors considered, MALDI-MS is the method of choice for an initial glycan or glycomic mapping. First, it is more rapid, robust, and easier to operate than ESI-MS for nonspecialists. Typically, 0.5–1 μl concentrated glycan sample is spotted onto the target plate in the presence of appropriate matrix solution, either premixed or mixed on-plate. Data acquisition for each sample is completed within several seconds to minutes, depending on the instrument. The use of a large format target plate allows applications of hundreds of spots and MS data to be automatically acquired for high-throughput screening. Any sample of bad quality due to high salt and other contaminant content will not affect the data acquisition of other samples. At a comparable sensitivity, offline nanospray (nanoESI) MS using a metal-coated borosilicate emitter to introduce a stable spray of analytes at 10–40 nl min^{-1} allows a longer analysis time of sometimes up to an hour or so for a few microliters of glycan sample, which is very useful for extensive multistages tandem MS (MSⁿ) acquisition experiments but too time-consuming and requiring much higher operation skill for a simple first screen. It is also less tolerant to salt, and 'dirty' samples often do not produce stable, long-lasting spray to allow data accumulation to good signal-to-noise ratio.

Second, for most practical applications, the glycans normally afford only singly charged molecular ions by MALDI, which facilitates direct assignment of molecular ion signals distributed across a full mass range, up to approximately m/z 6000 for reflectron mode before sensitivity tails off rapidly beyond that. Switching to linear mode can extend the mass range of detection at a cost of inferior resolution and accuracy. In contrast, ESI typically affords singly to multiply charged species with effective m/z values falling within a more restricted range, usually below m/z 2000 for most common applications. In principle, the formation of multiply charged species should afford a better detection of large-sized glycans at low m/z values. However, in complex mixtures, these larger glycans of lower abundance are in fact often masked or suppressed by the lower mass components. Added to that, a different combination of proton, ammonium, sodium, and/or potassium may act as the counter cations in multiply charged species, if samples are not desalted and handled properly, which spread the signal intensities across several distinctive peaks and further complicate the resulting spectra. High-resolution mass analyzer is therefore desirable to allow accurate mass and charge state determination of glycan signals in an often overcrowded ESI-mass spectrum. In this context, nanoESI-MS would benefit from coupling to online nanoLC to improve resolution by the time dimension. MS survey spectra from segments or the entire elution time span can then be summed together to present one or more overall MS profiles. Strong signals contributed by contaminant peaks that elute at a specific time would be averaged out more effectively than a single MS acquisition of continuously spraying analytes.

Third, in-source prompt fragmentation as discussed above is more readily formed in MALDI-MS without further adjustment of parameters and is very informative of the nonreducing terminal epitopes. Finally, the static nature of MALDI allows one with ample time to first interrogate and assign the spectra before returning to the sample spot to manually select particular peaks of interest for MS/MS experiments, provided the MS instrument is equipped to do so efficiently, such as that afforded by MALDI-Q/TOF or MALDI-TOF/TOF instruments. In contrast, ESI-MS usually requires all analysis be completed at once, and nonconsumed sample in the emitter may be difficult to recover for subsequent second-round applications. In summary, although both ESI and MALDI can be used for effective glycan mapping, MALDI-MS is more user-friendly for novice and general users, with more straightforward operation modes and data interpretation. The molecular mass can be readily assigned without considering the charged state and less overlapping isobaric peaks are encountered, which are thus more conducive for direct manual MS/MS acquisition on the tentatively assigned peaks.

6.05.3.3.3 Other MS- and non-MS-based complementary techniques

Apart from fragment ions produced either in-source or through MS/MS, sequencing and structural information can also be obtained through specific enzymatic and/or chemical cleavages coupled with MS detection of the products. Larger glycans such as the polyLacNAc can be cleaved into smaller fragments by endo- β -galactosidases while terminal residues can be sequentially removed by exo-glycosidases. The specificity introduced through the choice of enzyme can provide some linkage and stereochemistry information otherwise not afforded by MS alone. The partially degraded products can then be mapped by MS by virtue of the changes in mass value. Such an experimental approach has been in use since the early days of FAB-MS and still ranks among the most sensitive techniques in addressing, for example, if a terminal Hex is α -Gal or β -Gal. Specific β -galactosidase also exists that would cleave only a β 1,4-linked Gal residue under specific reaction conditions. Likewise, there are commercially available α 2,3-neuraminidase and α 1,2-fucosidase, just to name a few. Unfortunately, there is not a collection of exo-glycosidase available corresponding to each of the existing linkages and thus not all linkages can be defined this way. A major shortcoming of applying glycosidase digestions is that nonreactivity may not often necessarily be taken to establish the absence of that particular targeted residues as it may be caused instead by steric hindrance, nonactive enzyme upon storage, incorrect experimental conditions, and so on. Thus, a positive result is always more reliable than inference from a negative result. Another common problem is incomplete digestion despite exhaustive treatment, which cannot be distinguished from the presence of similar structures but with terminal residues different from the targeted ones.

While chemical cleavage may be less affected by steric hindrance, it is likewise subjected to incomplete and/or undesirable side reactions. Mild acid hydrolysis is commonly used for removal of labile substituents or glycosyl residues such as the fucose and sialic acids. Periodate oxidation, usually as the first step of the Smith degradation procedure, is another commonly employed partial degradation method. It would cleave the carbon-carbon bond between any two adjacent hydroxyl groups and render the cleaved glycosyl residues labile to mild acid hydrolysis. Normally, the aldehyde groups created by oxidative cleavage are reduced back to primary alcohol before subjecting the products with ring-open glycosyl residues to mild acid hydrolysis. Another variation is to exploit the higher susceptibility of the side chain of sialic acids, or the reducing end GalNAcitol on reduced *O*-glycans to periodate oxidation such that in the so-called mild periodate oxidation, only these are cleaved, leaving all other residues intact. This is a very useful chemical cleavage as applied to reduced *O*-glycans since it will cleave the GalNAcitol into two halves, retaining the substituents on the 3- and 6-arm, respectively. Importantly, the C2 and C4 remnants of the oxidatively cleaved GalNAcitol are of distinct masses and thus one can identify the respective series of products by MS mapping.⁶⁸ This is contrasted with complete periodate oxidation in Smith degradation in which all terminal glycosyl residues including all the fucose appendages on LacNAc-based glycan chains will be removed whereas the 3Gal 1-3/4GlcNAc unit remains intact and thus the *O*-glycan backbone can be easily defined.

6.05.3.3.4 Qualitative versus quantitative MS mapping

All the above-described methods in further chemoenzymatic manipulations and the fragment ions produced either directly or through true MS/MS are aimed toward establishing the exact glycan structures represented by the peaks in the MS profile. With the knowledge of identity, a further need is often to provide a quantitative

measure so as to detect up- or downregulated expression of a particular glycan or glycoform. Despite several attempts in this direction, MS mapping of glycans (and glycopeptides) remains largely qualitative. At a crude level and to a good approximation, the peak height or area afforded by MALDI-MS of the permethyl derivatives can be a good indicative index for relative amount and their alterations from sample to sample.² For nonmethylated glycans, it is necessary to at least methyl esterify or neutralize the carboxylic groups of sialic acids by chemical means in order to provide a quantitative measure in positive ion mode. Otherwise, the sialic acids would tend to be lost nonquantitatively and thus cannot be distinguished from those similar structures with a lower degree of sialylation. Even without the MS-induced loss, analysis in positive ion mode would still discriminate against negatively charged glycans, such as those with sulfates or phosphates. Conversely, the response of neutral glycans will be biased against in negative ion mode analysis. There is at present no simple way of efficiently detecting both the neutral and the negatively charged glycans in one measure and accurately determining their relative amounts by simply registering their peak intensities.

Considered carefully, the problems of quantification by MS are not trivial. In general, glycans of similar physicochemical properties but with a size distribution within a not-too-wide mass range may be assumed to give similar response factors under MS ionization and detection. Thus, for example, the relative amount of high mannose structures, Man₅₋₉GlcNAc₂, may be readily determined from their MS signal intensities as such. The sample applies to hybrid and complex type glycans within roughly a narrow mass range, provided the negatively charged sialic acids have been neutralized and stabilized as described above. It is, however, generally true that much larger glycans would give poorer MS response at higher mass range, and how accurate is a direct measurement of their signal intensities reflecting their actual abundance relative to lower molecular weight components has not been carefully established. Added to that, in MALDI, still larger glycans may only be detectable by switching to linear mode. In ESI, several differently charged species and thus signal series of different m/z values for the same glycan may be produced and need to be all detected efficiently for them to be summed together. At best, quantification by such only adds together all isobaric and isomeric components represented by a single nonresolved peak, which often also produces a skewed isotopic distribution due to signal overlapping and therefore imposes problems in delineating the respective isotopic signal clusters. The choice has to be made in whether to take just the monoisotopic peak, if it can be accurately resolved and determined, for quantification or whether deisotoping by any of the software would be needed to arrive at a single deconvoluted peak.

Once all these issues have been adequately settled, additional normalization considerations need to be given for comparing different samples. As in label-free quantitative proteomics, both biological and technical replicates are needed. Spiking with internal standards will also be desirable to correct for sample-to-sample variations even if one resorts solely to constant 'housekeeping' glycans for normalization. It is much simpler to express all abundance relative to a single common glycan for rough comparative measure. A different approach is to arbitrarily set the peak detection threshold to first filter out the low-level noise peak and then sum all intensities of all remaining peaks as a measure of total abundance and the basis of normalized comparison. All detected glycan components of interest can then be expressed as a percentage amount of this total for comparison across different datasets. To date, different approaches have been adopted in a few application examples but it is fair to say that no standard procedure has been systematically evaluated, let alone agreed upon. It nevertheless has caught the attention of the glycoanalysis community and the initiatives are underway. Alternatively, few groups have proposed stable isotope approach, introduced within the reductively aminated tag at the reducing end,^{69,70} or through permethylation with isotope-labeled methyl iodide.⁷¹⁻⁷³ Again, none has been widely adopted by the community and none is without its own pitfalls. Given the complexity of the glycomic profile, mixing together populations of differentially labeled samples only adds extra complexities through isobaric overlapping. The only conclusion that can be made at present is that the need for better quantification has been widely recognized but the solution is far from simple and will need more development work. In that respect, quantification of fluorescent-labeled glycans by HPLC is both simple and elegant by comparison. The shortcomings however are already described, residing mostly in inadequate resolution by one-dimensional separation and being unable to always identify all peaks. Multidimensional separation will be time- and labor-consuming and it is best to attempt absolute quantification using suitable standards for subsequent comparison across different samples.

6.05.4 MS/MS Glycan Sequencing

Although sequential exo-glycosidase digestions coupled with MS detection is effective in defining not only the glycan sequence in terms of Hex, HexNAc, and so on, but also its stereochemistry, its practical utility is ultimately limited by the availability of a wide array of glycosidases that are robust and specific enough. In addition, the entire process is lengthy and may require desalting considerations in between steps to facilitate MS detection. While single-step confirmation of assigned nonreducing terminal residue can be readily accomplished at high sensitivity, a full cycle of sequential trimming from the nonreducing end may not be. Further ambiguity in data interpretation arises due to the branched nature of most *N*- and *O*-glycans, especially since the sequential digestion often needs to be performed on a glycomic mixture rather than isolated individual glycans. Thus, by far, the way forward for efficient glycan or glycomic sequencing depends on high-sensitivity MS/MS analysis. Its success in turn relies on a sound knowledge of both the expected repertoire of glycan structures (Section 6.05.2.2) and the characteristic fragment ions produced under different modes of implementation (this section).

Early systematic studies of the fragmentation pattern afforded by glycans were mostly carried out on the permethyl derivatives out of necessity for EI- and FAB-ionization sources. A body of historical work performed then^{64,65,74–76} has collectively provided a mechanistic basis for much of our current understanding. The fundamental differences in the specific cleavages afforded by protonated versus other cationized parent ions, or low (<100 eV) versus high (up to several kiloelectronvolts) collision energy, or different collision gases, were noted based on FAB-CID-MS/MS carried out on either a triple quadrupole or a four-sector MS instrument.⁷⁷ The advent of ESI in the 1990s as fitted on the Q/TOFs and ion traps has since led to many MS/MS studies of the permethyl derivatives along the line of low-energy CID^{78–82} and, increasingly, is being extended to analyze directly underivatized glycan samples,^{83,84} especially in negative ion mode (for references see Section 6.05.4.2.1), or one that has been tagged at the reducing end^{85–89}. In parallel, MALDI has evolved from initial dependence on post-source decay (PSD) fragmentation^{90,91} to true MS/MS capability enabled on the Q/TOF⁹² and TOF/TOF,^{93–96} for low- and high-energy CID, respectively. These more recent developments in nanoESI and MALDI along with a range of higher performance mass analyzers have practically overcome the sensitivity problem of analyzing native glycans, which is now a viable option in both positive and negative ion modes, particularly in high-throughput LC-ESI-MS/MS applications for rapid screening and sequencing.^{33,51,52,97,98}

On the other hand, detailed structural analysis, especially one that needs to critically define not only the sequence but also the linkage positions, continues to rely much on the distinct advantages offered by MS/MS analysis of the permethyl derivatives. Following a rapid profiling at the MS level, direct selection of a few of the most prominent molecular ion signals of interest for low-energy CID-MS/MS would most efficiently establish the overall core structures, peripheral extension, and terminal epitopes based on a complementary set of sequence informative glycosidic cleavage ions (Section 6.05.4.1). This can then be followed up by further isolating one or more of the primary fragment ions for additional stages of fragmentation until linkage-specific ions can be identified (Section 6.05.4.2.2). Alternatively, such linkage-informative cleavage ions can also be produced at the MS² level by resorting to high-energy CID analysis on a MALDI-TOF/TOF (Section 6.05.4.2.3).

6.05.4.1 Glycosidic Cleavages and Oxonium Ions

MALDI- and ESI-MS analyses of glycans in positive ion mode afford mostly sodiated molecular ions, from which the most readily produced fragment ions are those derived from glycosidic cleavages, referred to as the B, C, Y, and Z ions according to the nomenclature proposed by Domon and Costello⁹⁹ (Figure 4(a)). For a neutral glycan without any amino sugar, charged residue, or labile substituent, glycosidic cleavage is expected to occur randomly at each residue along the chain. However, in the case of *N*- and *O*-glycans, which carry HexNAc at specific locations in the core moieties, along the peripheral chains, and at each branch point (Section 6.05.2.2; Figure 1), as well as terminal sialic acids, a very distinctive fragmentation pattern is observed instead. Most importantly, glycosidic cleavages are directed preferentially at HexNAc and NeuAc/NeuGc residues, giving rise to dominant pairs of B and Y ions. This applies to both native and chemically derivatized glycans but is

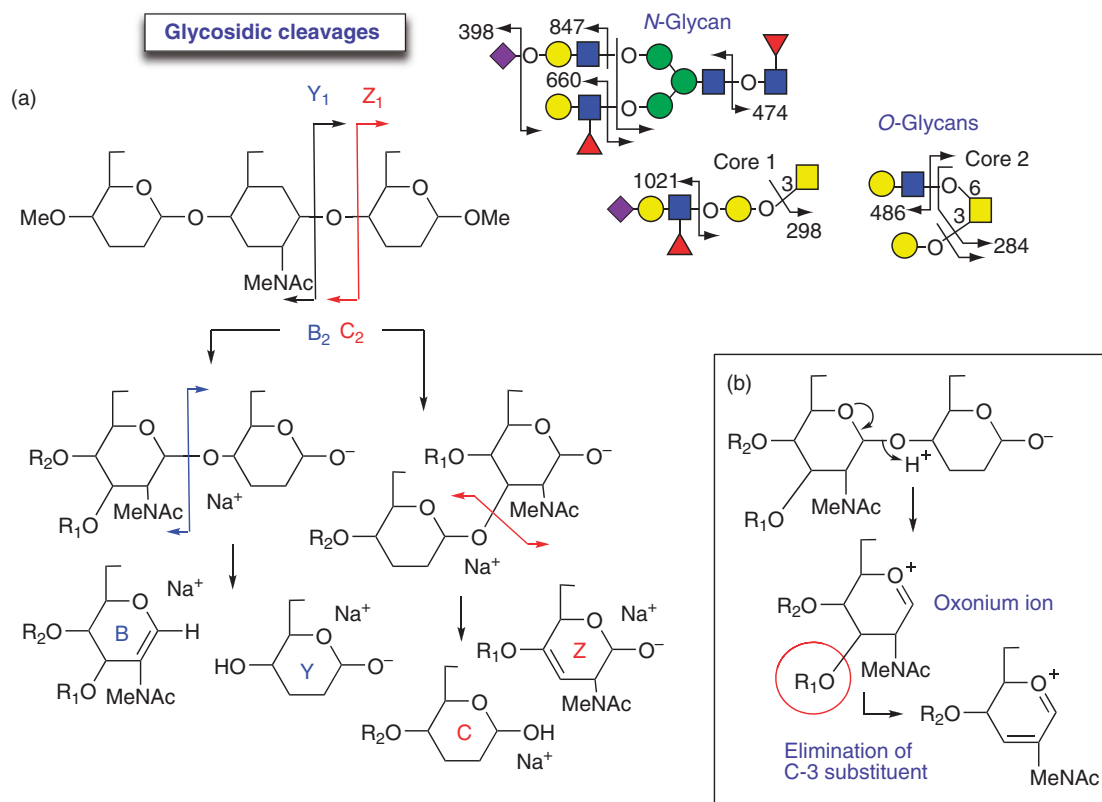


Figure 4 Major glycosidic cleavage pathways. The sodiated B and Y ions are the major fragment ions afforded by the sodiated molecular ions (a) although C and Z ions can also be formed at certain sites. The specification of a HexNAc at the cleavage site in the schematic drawings is meant to emphasize that such glycosidic cleavages are preferentially directed to HexNAc for the permethylated derivatives. The N- and O-glycan cartoons illustrate the resulting key ions that are informative of core and terminal structures. For protonated molecular ions (b), the oxonium ions formed at the HexNAcs are the most prominent key ions, which are often accompanied by elimination at position 3.

most pronounced for permethylated glycans, to the extent that cleavages at other sites are largely disfavored under low-energy CID-MS/MS commonly performed on Q/TOF or ion trap.

6.05.4.1.1 Neutral loss and identification of core structures

Translated into practical terms, the sodiated molecular ions of N- and O-glycans are thus expected to demonstrate the facile neutral losses of sialic acids and the entire nonreducing terminal epitopes delimited by HexNAc to give abundant Y ions (Figure 5). Multiple losses of such antennary chains or substituents from the parent ions are commonly observed in low-energy CID, particularly under MALDI-Q/TOF-MS/MS, leading eventually to prominent fragment ions corresponding to the bare core structures (Figure 5(b)). For native and reducing end-tagged glycans, such neutral losses are very useful in defining the presence of certain terminal glycosyl substituents such as sialylation and fucosylation but critical information on their localization to a particular terminal epitope may be missing. To illustrate this point, consider a biantennary N-glycan with a terminal disialyl LacNAc, NeuAc-Hex-(NeuAc)HexNAc, on one antennae and a fucosylated LacNAc on the other. Neutral losses of the sialic acids and Fuc are likely to be predominant, followed by subsequent losses of LacNAc, in preference to direct loss of a disialyl LacNAc or Lewis X (LeX) moiety, the Y ions of which may or may not be detected to unambiguously establish their presence.

In contrast, for permethylated glycans, direct losses of the entire sialylated and/or fucosylated epitopes are always to be expected since the propensity for cleavage at HexNAc is much elevated to the level on par with that at sialic acids and higher than that at Fuc. Importantly, since all OH groups not engaging in linkages are methylated,

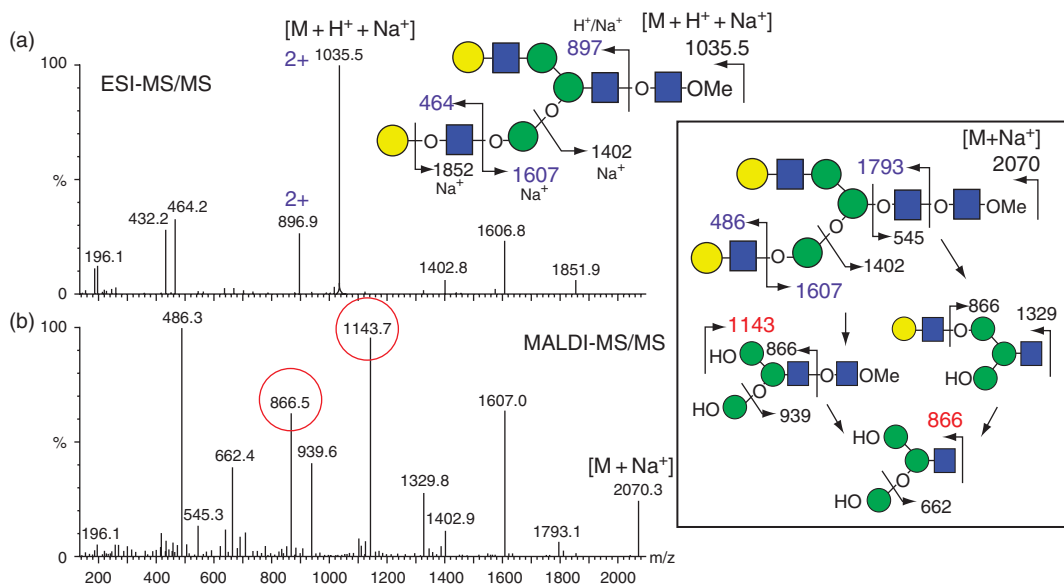


Figure 5 A typical low-energy CID-MS/MS spectrum for a permethylated biantennary *N*-glycan acquired under ESI as a doubly charged protonated/sodiated molecular ion (a) or MALDI as a singly charged sodiated molecular ion (b). Note the common as well as the different ions produced. The schematic drawings illustrate how successive neutral losses would give rise to the ions corresponding to the trimannosyl core with a number of free OH groups. All cleavage ions drawn here are B and Y ions.

subsequent loss of LacNAc through additional cleavages in this case example would be defined by the presence of extra free OH groups, namely, (OH)Hex-(OH)HexNAc or Hex-(OH)HexNAc, which can be distinguished from a bona fide Hex-HexNAc terminal epitope. This free OH tagging of each cleavage site enables one to readily define the presence of branch point and constitutes one of the several distinct advantages in MS/MS analysis of permethyl derivatives. It is very informative as well with respect to defining the presence of polyLacNAc versus distribution of the extra LacNAc units to additional antennae. Furthermore, the ion corresponding to the core structure after losing all the antennary extension will be 'tagged' by multiple deficits of 14 mass units in its m/z values, each of which denotes a cleavage site and therefore the number of antennary branchings.

6.05.4.1.2 Nonreducing terminal oxonium and B ions

As described above, identification of the facile neutral losses from the sodiated molecular ions are often sufficient to define the kind of antennary structures extending from the core, which will be further corroborated by detecting the corresponding B ions at the low-mass region. Since the m/z values of these B ions depend only on the glycosyl constituents of the nonreducing terminal epitope, they truly qualify as one of the most useful sets of diagnostic fragment ions that are reliably produced at high abundance. It should be noted that similar but nonsodiated B ions are in fact the predominant fragment ions afforded by protonated parents, be it glycans or glycopeptides.¹⁰⁰ Dated back to the days of FAB-MS,^{64,65} these B ions are the single most important fragment ions, and sometimes the only ones, that are produced directly in-source during the ionization process. The proposed mechanism invoked the formation of an oxonium ion (Figure 4(b)) and was named as the A-type ion in the nomenclature popularized by Dell.^{64,101} This designation should however be now discouraged as it is easily confused with the widely adopted Domon and Costello nomenclature.⁹⁹ Nonetheless, the mechanism underlying the formation of the oxonium ions are likely to be different from that of the sodiated B ions produced from sodiated parents and thus there is merit in retaining the distinction. Interestingly, despite the propensity in forming sodiated molecular ions, prompt in-source fragmentation under MALDI likewise produces mostly the low-mass nonsodiated oxonium ions. In addition, MALDI-MS/MS of sodiated parents performed on TOF/TOF and ESI-MS/MS of multiply charged parent ions, which are both sodiated and protonated, frequently afford both sodiated B ions and the corresponding oxonium ions, at variable relative intensities from run to run.

The B ions and/or the oxonium ions reliably inform the ensemble of nonreducing terminal epitopes that are presented by the particular glycan serving as the parent ion in MS/MS analysis. When induced as direct in-source fragment ions, be it implemented under FAB-, MALDI-, or ESI-MS, they serve equally well to map the nonreducing termini of a glycomic sample. In the case of permethylated *N*-glycans, the glycosidic bond between the HexNAcs in the chitobiose core represents another site that is prone to cleavage. The resulting Y_1 ion is informative with respect to whether there is a core fucosylation, whereas the corresponding B ion often forms another series of ions parallel to those Y ions produced through successive neutral losses from the parent ions (**Figure 5(b)**).¹⁰² Thus, the B ion itself can exhibit further neutral loss and likewise will benefit from being permethylated to allow discrimination between primary fragment ions and those arising from additional cleavages. Taken all these into consideration, mapping of nonreducing terminal epitopes through a complementary set of abundant B and Y ions is thus relatively straightforward. The remaining task and the more difficult aspects are further definitions of the specific linkages among the constituent residues of the epitopes defined by the diagnostic B ions. For example, a terminal epitope identified as $\text{Fuc}_1\text{Hex}_1\text{HexNAc}_1$ may correspond to H type 1 or 2 ($\text{Fuc-2Gal-3/4GlcNAc}$) or Lewis X/A (Gal-(Fuc)GlcNAc); a $\text{NeuAc}_1\text{Hex}_1\text{HexNAc}_1$ can be $\text{NeuAc2-3/6Gal-3/4GlcNAc}$. These are important issues of glycobiological significance, but less readily resolved by identifying the abundant B and Y ions alone.

6.05.4.1.3 Elimination and linkage-specific Z ions

For the sodiated permethylated glycans, the sodiated C ions at 18 mass units higher than the corresponding B ions can, in principle, be less restricted to HexNAc sites but are normally less abundant and often not detected at all, with two notable exceptions.¹⁰² First, the C ion derived from cleavage at the Hex of the nonreducing terminal Hex-HexNAc unit in *N*- and *O*-glycans is readily formed. It can be nonsubstituted or further glycosylated, giving rise to another set of prominent diagnostic ions corresponding to, for example, Hex-Hex-OH (from $\text{Gal}\alpha 1\text{-3Gal}$ epitope), Fuc-Hex-OH (from H epitope or Lewis B/Y), and NeuAc-Hex-OH (from sialylated LacNAc), which inform the presence of these terminal epitopes. Second, in cases where an internal Hex is 3-linked to HexNAc, the sodiated C ions originating from glycosidic cleavage at the respective Hex residues become more prominent. Important instances where this fragmentation applies and proves to be extremely useful are those R-Gal-OH ions from internal type 1 chain ($\text{R-Gal}\beta 1\text{-3GlcNAc}\beta 1\text{-}$), the presence of which is indicative of the respective 3-linkages, but care should be taken as similar C ions can also be produced from a terminal type 2 unit, as described above, albeit less abundant.

Of better diagnostic values are the Z ions corresponding to elimination of substituents at position 3 in place of the more commonly seen Y ions.^{102,103} For example, further elimination of either a 3-linked Fuc or a 3-linked Gal from the sodiated B ion, Gal-(Fuc)GlcNAc , affords an easy way to differentiate between a terminal Lewis X and A epitopes. This is in analogy with the further elimination of substituents at C3 of HexNAc in the oxonium ion, extensively used in FAB-MS to define type 1 and 2 linkages (**Figure 4(b)**).⁶⁴ In addition, only glycan structures containing a 3-linked Fuc will readily give Z ion corresponding to its elimination from the respective parents whereas neutral loss of 2-linked Fuc as in H antigen ($\text{Fuc}\alpha 1\text{-2Gal}$), or 4-linked Fuc as in Lewis A, would give Y ions. For cores 1 and 2 of *O*-glycans, Z_1 ions are characteristically produced from the elimination of the glycosyl substituent from the 3-arm of GalNAc, be it a single Gal residue or an entire chain.¹⁰⁴ This would give a Z_1 ion corresponding to a $\Delta\text{GalNAcitol}$ residue for a core 1 structure, or one at 14 mass units lower carrying an extra free OH group for a core 2 structure (**Figure 4(a)**).

6.05.4.2 Cross-Ring and Other Linkage-Specific Cleavages

Overall, the advantages and usefulness of HexNAc-driven glycosidic cleavages afforded by permethyl derivatives (Section 6.05.4.1) are their reliability and sequence informative character, matching nicely the basic architecture of mammalian *N*- and *O*-glycans that are founded on Hex-HexNAc extension from limited sets of core structures (Section 6.05.2.1). The B and Y ions are the most consistently observed fragment ions afforded by different MS instruments from one laboratory to another. By restricting the fragmentation to a few key sites, a low-energy CID-MS/MS spectrum is elegantly simple and all major fragment ion signals can be readily assigned. However, the consequential drawback is the lack of cleavages at nonamino sugars and thus sequence information is lacking in between HexNAc residues. This is normally not a problem unless one has to deal with

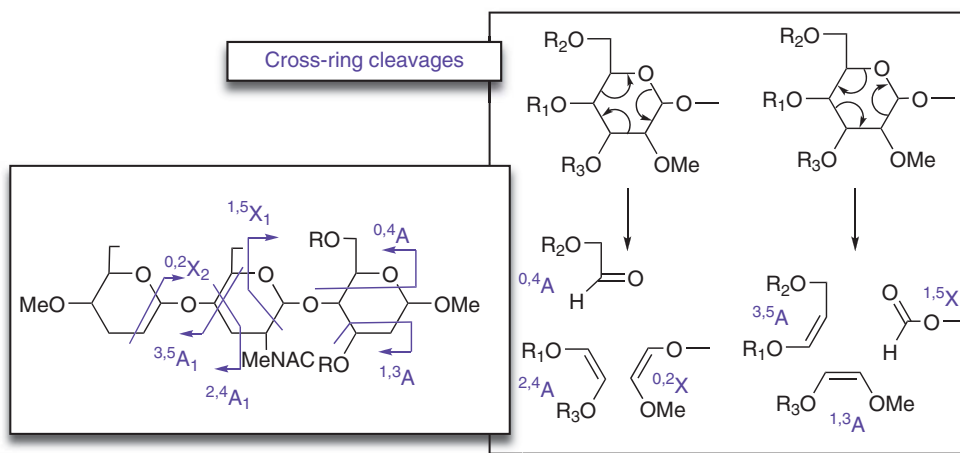


Figure 6 The most commonly observed cross-ring cleavage ions with the proposed mechanism for their formation. The subscript denotes that the cleavages occur at the reducing end (A ions) or nonreducing end (X ions) of the n th residue from either end. Other cleavages are possible and can be named accordingly but their mechanistic basis is less clear.

unusual terminal epitopes that carry extra Hex residue(s), for example, NeuAc₁Fuc₁Hex₂HexNAC₁ from zebrafish.²³ A more critical shortcoming, as noted before, is the lack of linkage information. In general, glycosidic cleavages, except for the Z ions described above (Section 6.05.4.1.3), are only sequence informative and not linkage specific.

The key to securing linkage information is to induce the so-called cross-ring cleavage ions (Figure 6) and other concerted cleavages around the ring that would eliminate only the glycosyl substituents at specific positions (Figure 7) under specific modes of CID implementation. Often the mechanistic basis for these latter cleavages is not well understood and not readily proven, and the same ion may even be assigned differently by different groups. However, as long as they are well verified against standards under similar MS/MS conditions and found to be consistently produced where expected, they can be elevated to the status of key fragment ions supportive of inferred linkages. With low-energy CID commonly acquired on the Q/TOFs and ion traps in positive ion mode, these fragment ions are normally present only in relatively very low abundance for them to be reliably useful, but can be effectively promoted by switching to MS/MS analysis of native glycans in negative ion mode (Section 6.05.4.2.1) or by subjecting a few primary fragment ions of interest through successive stages of fragmentation (Section 6.05.4.2.2). Alternatively, MALDI-TOF/TOF, which alone among the current generation of MS instruments affords true high-energy CID-MS/MS, has also been capitalized to provide the much-needed linkage information (Section 6.05.4.2.3).

6.05.4.2.1 Negative ion mode sequencing of native glycans

As noted before, MS/MS analysis of nonderivatized glycans in positive ion mode is generally less effective and therefore not recommended. First, the sialic acids need to be stabilized by one of the several chemical methods that neutralize the negative charge,^{45–48} to render those sialylated glycans more amenable to positive ion mode MS detection in the first place before they can be isolated for MS/MS. Even so, the neutral loss of the sialic acid is normally the most preferred cleavage at the expense of other useful ions. Its removal by chemical means or sialidase is therefore commonly practiced, which has the added advantage of using an α 2–3-specific sialidase to provide an easy way of determining linkage. Second, without an effective *O*-Me tag, single cleavages are not distinguishable from multiple cleavages and the assignment is problematic with respect to branching pattern and localizing the terminal sialylation and fucosylation. With additional facile loss of a H₂O moiety, internal fragment ions can often be confused with those primary ions retaining either terminus of the glycans. It is thus useful only for confirming certain aspects of the initial compositional assignment derived from MS mapping and the structures thereby inferred. Similar range of diagnostic B ions can be detected along with some

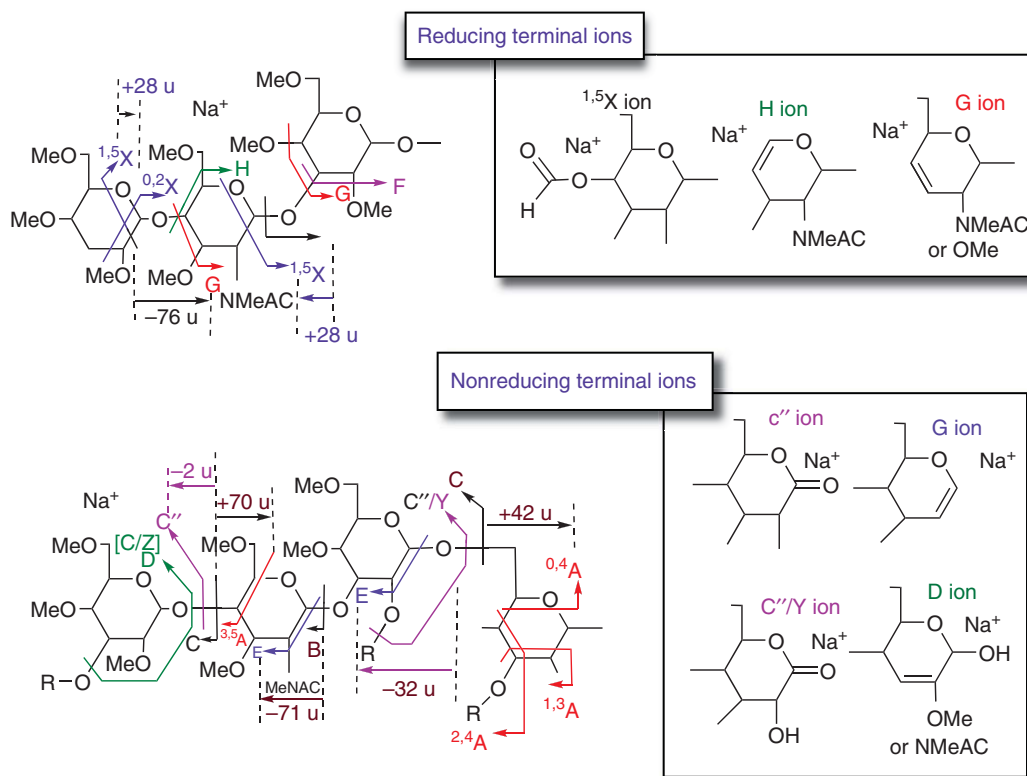


Figure 7 The most consistently observed cleavage ions under high-energy CID-MS/MS as performed on a MALDI-TOF/TOF, which are grouped into either the reducing terminal or the nonreducing terminal ions. Their interrelationship in terms of mass increment for a permethyl derivative is also illustrated.

characteristic neutral losses. In most cases, definitive sequencing is not attainable and linkage-specific fragment ions are not reliably obtained or assigned.

Switching to negative ion mode analysis, however, provides a better solution, which has been extensively studied and applied to a range of native glycans.^{105–112} Collectively, these works showed that the B and Y ions that are so dominant in positive ion mode can be significantly suppressed, except for those resulting from the loss of terminal sialic acids in sialylated glycans. Interestingly, the α 2–3 versus α 2–6 linkage of the terminal sialic acids can be distinguished by virtue of an additional ion occurring at 16 mass units higher than the B₁ ion for NeuAc at m/z 290, which are only present at m/z 306 if the NeuAc is α 2–6-linked.^{110,113} In general, however, this linkage-specific ion is weak in intensity and a productive negative ion mode MS/MS analysis of underivatized multisialylated glycans with unstabilized sialic acid residues remains difficult.^{107,112}

The problem of sialic acids apart, MS/MS of native glycans in negative ion mode does produce more nonreducing terminal fragment ions informative of sequence and/or linkage. Among the more consistently noted are the glycosidic cleavage C ion, double-cleavage D ion, and various cross-ring cleavage A ions, particularly the ^{0,2}A and ^{2,4}A ions. For the N-glycans, the ^{0,2}A and ^{2,4}A ions are most readily formed at the reducing end GlcNAcs of the chitobiose core, which can inform the presence of core fucosylation. Other A ions can also be detected but unfortunately for native glycans, the mass increment of an ^{0,4}A ion from the corresponding C ion is the same as those of ^{2,4}A and ^{1,3}A ions, which renders them not specific enough to distinguish a 6- from 3-linkage to a Hex. Thus, along with the ^{0,2}A ion, these A ions are mostly only sequence informative, whereas ^{3,5}A ion is not normally observed.^{110,112}

The D ions were originally designated for the specific cleavages observed under high-energy CID or PSD for native N-glycans in positive ion mode,^{91,114} which correspond to a glycosidic cleavage at the reducing end of the core branching β -mannose in concert with the elimination of the entire 3-arm substituent from its 3 position.

Its m/z value thus informs the 6-arm extension and is inclusive of the bisecting GlcNAc at position 4, where present. A companion ion at 221 mass units lower will additionally be observed and thought to correspond to further elimination of the bisecting GlcNAc ($203 + 18$). In its absence, the D ion will be accompanied instead by an ion at -18 mass units, corresponding to the elimination of a H_2O moiety. Remarkably, this concerted cleavage is reproduced in negative ion mode by mere low-energy CID.^{107,112} In addition, it was found not to be restricted to cleavages at the core mannose but will extend to other glycosyl residues that carry glycosyl substituent at C3 position, such as the 3-linked GlcNAc in type 1 chain, as first noted by Chai *et al.*¹⁰⁵ in their negative ion mode low-energy CID-MS/MS analysis of milk oligosaccharides. It should however be cautioned that since multiple cleavages at two or more distinct sites abound with instruments such as Q/TOF, a genuine D ion that results from concomitant elimination of the 3-linked substituent of a particular glycosyl residue undergoing glycosidic cleavage may not be distinguishable from a B/Y or C/Z double-cleavage ion resulting from two unrelated cleavages at two distinct sites. More recent applications of high-energy CID-MS/MS by MALDI-TOF/TOF^{93,94,96,102,103} have since demonstrated that the D ions can indeed be formed at the core mannose and other residues of the native and permethylated glycans and rank among the most prominent and useful linkage-specific concerted cleavage ions afforded (see Section 6.05.4.2.3).

6.05.4.2.2 MS^n of permethylated glycans

While analysis in negative ion mode additionally provides some useful linkage information, it does not fully alleviate the common problems associated with MS and MS/MS analysis of native glycans, as discussed above. By far more novel and/or biologically important structures are solved by analyzing the permethyl derivatives in positive ion mode. The superiority of mass resolution and the accuracy afforded by Q/TOFs place these mass analyzers as instrument of choice for implementing low-energy CID, with either MALDI or ESI, compared to the highly popular and ever improving ion traps, due to the development in proteomics. While coupling to fourier transformed-ion cyclotron resonance (FT-ICR) or Orbitraps has provided high mass resolution and accuracy at the MS level, the quality of an MS^2 spectrum is still inherently inferior to one obtained on a Q/TOF. Not only is it often dominated by fewer Y ions, the abundance of the low-mass B ions additionally suffers from the low-mass cutoff rule associated with ion traps.

However, a very attractive feature of an ion trap is the ability to perform additional stages of MS/MS in an MS^n experiment. This is very compatible with the time scale of an offline nanoESI-MS, which often allows analysis time approaching an hour in length, with only a few microliters of the analyte in solution. The practical advantage of MS^n resides in the ability to isolate any fragment ion corresponding to a structural motif of interest and to induce on it cross-ring and other cleavages that would otherwise not occur when more favorable fragmentation pathways are available. In principle, such an approach can be equally applied to native and derivatized glycans alike but for reasons already described, it is more productively implemented on permethylated glycans^{80,82,115,116} to take advantage of the *O*-Me versus free OH tags for branch sites and linkage identification. Often this involves stripping down to a disaccharide backbone with only a single linkage site to determine. Among the most useful ones are the various A ions including $^{3,5}\text{A}$, $^{2,4}\text{A}$, and $^{0,4}\text{A}$ ions that can be used to define the site of attachment. This approach has been used to discriminate between Gal-3Gal-GlcNAc versus Gal-4Gal-GlcNAc by further MS/MS on the isolated Gal-3/4Gal fragments¹¹⁷ or to define the NeuAc2-3/6Gal linkages. At this level, other less commonly observed ions may be additionally detected and some proved to be linkage-specific enough after careful evaluation against authentic standards. For example, an innovative and alternative approach to the above case examples is to isolate instead the (HO)Gal-GlcNAc or (OH)Gal ion for additional MS/MS, which produces linkage-specific ions depending on the location of the free OH group and hence the original attachment site.^{117,118} Another useful case is to induce cleavages between the nonamino sugars that would otherwise be difficult to obtain when cleavage at the delimiting HexNAc is dominating. It is also useful to produce the bare core structures through successive stages of isolation and activation, and finally to enumerate the number of antennary branchings by the number of free OH groups carried.

Ironically, some of the information eventually obtained can often be obtained through a direct MS/MS analysis on a Q/TOF, which promotes extensive multiple cleavages corresponding to clipping off the antennary substituents (Section 6.05.4.1.1). Other linkage-specific cleavages are equally well produced by high-energy CID MS/MS on a TOF/TOF (see Section 6.05.4.2.3). In contrast, the dominant neutral losses

of sialic acids and the HexNAc-driven cleavages often render an ion trap MS² profile not informative enough and further stages of MSⁿ become a necessity. Since it involves sequential stages of intelligence-based selection of the precursors among the many signals, which are often not the base peaks, productive MSⁿ automation is not apparent for high-throughput analysis. This problem may eventually be overcome either by experience and knowledge as provided by an expert operator or by software development that would furnish the decision tree operation and then collate all the acquired MSⁿ data, possibly also for spectral matching against an extensive library of standard MSⁿ spectral collection.^{119–122} The extremely high sensitivity offered by the current generation of high-performance ion traps and their hybrids has already made MSⁿ beyond MS³ and MS⁴ practically feasible,¹¹⁶ and many other innovative applications are constantly evolving, especially in coupling to data-dependent LC-MS/MS analysis. It is also probably the only way to resolve an unassigned peak commonly encountered in any MS/MS analysis by virtue of the capability to further isolate it for fragmentation. This would help identifying novel or unexpected structures, as well as mining the depth of the glycome (Section 6.05.4.3).

6.05.4.2.3 High-energy CID-MS/MS on MALDI-TOF/TOF of permethylated glycans

Short of MSⁿ implementation, high-energy CID MS² as acquired on MALDI-TOF/TOF represents the only other practical solution to promote linkage-specific cleavages, which is much faster and efficient once one is familiar with the anticipated range of key fragment ions (Figure 7). It should be noted that not all TOF/TOF instruments work equally well to give the desirable high-energy CID characteristics. At lower or suboptimal collision energy, the TOF/TOF-MS/MS spectra can be very similar to one that is acquired on a Q/TOF with one notable exception; namely, most ions are primary fragment ions resulting from a single collision event.^{123,124} Thus, it does not produce successive losses nor the ion corresponding to trimannosyl core structures be obtained. However, this empirical rule is not absolute as sialylated and/or smaller glycans may additionally parade internal fragment ions that apparently can only be derived from double cleavages. This apart, under optimum conditions, a high-energy CID spectrum can be distinctively different from a low-energy CID spectrum, particularly for *N*-glycans.^{96,102} For the *O*-glycans, milk oligosaccharides or glycans derived from lactosylceramides, most of the single glycosidic cleavage fragment ions afforded by Q/TOF including the linkage-specific Z ions are likewise observed but with additional ones that can only be produced through high-energy CID.^{66,93–95,103,104,125}

MALDI-TOF/TOF has been applied to both derivatized and nonderivatized glycans but for reasons already discussed above, fragment ions afforded by the sodiated permethyl derivatives are more distinctive and only these will be further described here. About the most drastic change as one switches to high-energy CID is the propensity to produce ring cleavage ^{1,5}X ions at almost every glycosyl residue, which provides a complete series of sequence-defining ions not delimited by HexNAcs. At HexNAcs, the corresponding Y ion at 28 units lower will also be formed and its reduced intensity in favor of the ^{1,5}X ion is indicative of the extent that a true high-energy CID has occurred. For the *N*-glycans, the ^{1,5}X_{1–3} ions are important in identifying core fucosylation, and the respective substituents on either arms of the trimannosyl core. The ^{1,5}X ion itself is, however, not linkage informative. The 6-arm versus 3-arm extension is better informed by the D ion formed at the β-mannose, in a fashion similar to that already described for the negative mode MS/MS analysis of a native *N*-glycan (Section 6.05.4.2.1). This D ion will normally be accompanied by the ^{3,5}A and ^{0,4}A ion pairs, albeit at much lower intensity, to critically establish the 6-arm extension. Interestingly, while a bisecting GlcNAc is identified by an ion at 221 units lower than the D ion for the native glycans, the permethyl derivatives would instead afford a diagnostic ion that is 321 units lower¹²⁶ only if it is bisected. The bisecting GlcNAc will be carried on the ^{3,5}A but not the ^{0,4}A ion. Further along the peripheral chain, the full sets of ^{1,5}X ions can be used to define in detail the terminal epitopes, aided by the Y ions corresponding to its neutral loss in its entirety. Conversely from the nonreducing end, sequence information will be given by the C ions or the C–2H ion at 2 mass units lower (C'' ion), which can be coupled with a loss of the glycosyl substituent at C2 to give the diagnostic C''/Y ion.¹⁰²

Spina *et al.*⁹⁵ were the first to note that high-energy CID MALDI-MS/MS additionally affords the concerted elimination of the substituents from two adjacent carbons of the ring and coined the term E, F, and G ions to extend the Domon and Costello nomenclature, taking into consideration also the D ions described by others. In addition, an ω ion arises through elimination of the C6, and its substituent concomitant with elimination of another MeOH moiety, as noted by Lemoine *et al.*⁷⁷ in their analysis with high-energy CID-FAB-MS/MS, can

be further incorporated to this series of 'new' fragment ions and properly referred to as the H ion¹⁰² (Figure 7). The reducing terminal G and H ions proved to be as useful as the complementary nonreducing terminal ^{3,5}A, ^{0,4}A, and ^{1,3}A ions in providing linkage information. Where an adjacent glycosyl substituent is absent, elimination of the attached chain would be accompanied by elimination of the OMe group instead in forming the G and H ions, which would differ by 14 units if only C4 is substituted but occurring at very different masses if either C6 or C3 is also substituted. The F ion, which relates to elimination of substituents at C2 and C3, is less useful as not many C2 glycosyl substituents are normally present in *N*- and *O*-glycans. Likewise, the E ions, which can be found to accompany the B ions at either 71 or 30 units lower depending on whether it is a HexNAc or Hex residue, are prominent where the otherwise characteristic B ions may be suppressed, but are not linkage informative either.

At first glance, the high-energy CID-MS/MS spectra are intimidatingly complex. However, by noting the simple rules formulated through systematic studies, one can navigate through the spectra with ease. The best approach is first to establish the overall sequence and core structures through a complementary low-energy CID-MS/MS and then look specifically for the expected reducing terminal ^{1,5}X ions and the G/H ion pairs, and the nonreducing terminal E/B ion pairs and A ions. In principle, an expert can simply assign *de novo* without resorting to another low-energy CID-MS/MS spectra but the two tend to be complementary and mutually corroborative. From experience, it is noted that the D ions at each residue, particularly those expected to carry glycosyl extension at C3, are among the most prominent key fragment ions that are readily assigned. For example, the presence of a Lewis A type structure, irrespective of whether it is internal or at the nonreducing terminal, would give a prominent ion at *m/z* 442, corresponding to sodiated Fuc-4GlcNAc after eliminating the R-Gal substituent from C3.¹⁰³ An internal LeX would give instead a D ion corresponding to R-Gal-4GlcNAc after eliminating the 3-linked Fuc. Both would additionally induce very prominent reducing terminal G ions through elimination of both the 3 and 4 substituents.

In comparison with the MSⁿ approach, high-energy CID-MS/MS, which is available only on TOF/TOF, is relatively straightforward for data acquisition and assignment as it involves no further decision making in precursor ion selection for subsequent stages of fragmentation. Its most notable two shortcomings are insufficient sensitivity in today's term and inadequate resolution in both the precursor ion selection and the high mass region of the resulting MS² spectrum. Thus, powerful as it is, there is currently not much scope for further improvement in performance, nor is it suitable for coupling to online LC-ESI-MS/MS. It is, however, very useful for advanced structural analysis and any sample amount that is adequate for conventional linkage analysis after permethylation is deemed sufficient for MALDI-TOF/TOF high-energy CID-MS/MS sequencing, with linkage information to be derived for each individual residue.

6.05.4.3 Resolving the Glycomic Complexity by MS/MS

As discussed in the preceding sections, the fragmentation afforded by the glycans under different modes of MS/MS can be summarized into a few distinctive pathways that would produce a number of sequence- or linkage-informative key ions. Any MS/MS spectral signal assignment would start by identifying these few key ions and is often sufficient to thus arrive at a deduced structure. Any other signal that does not 'fit' the known fragmentation pattern and/or assigned structures should ideally be subjected to further stages of MS/MS analysis although this is not always feasible or productive. Instead, it is best to capitalize on those that can be assigned based on the known fragmentation rules. The web-based Glyco-peakfinder program,¹²⁷ for example, can be used to calculate and assign all types of fragment ions including monosaccharide cross-ring cleavage products and multiply charged ions. Any other minor peaks that are not identified then may be better left unassigned instead of evoking some irrational multiple cleavages that would simply fit the mass. A more common problem is to deal with two or more mutually exclusive fragment ions that cannot be derived from a single unique structure and thus a need to invoke the presence of isobaric or isomeric structures. In fact, with the precursor ion isolation window afforded by the ion traps or the timed-ion selector of a TOF/TOF, it is not uncommon to let in precursors differing by 1 Da or more, particularly at higher mass, which would further complicate the analysis. Thus, for a truly complex glycomic sample, one needs to bear in mind that MS/MS analyses are often not performed on a single parent defined by a unique molecular mass and composition.

6.05.4.3.1 Identifying structural isomers by key diagnostic ions

The extreme heterogeneity of protein glycosylation would result in many isomeric structures differing in branching pattern, location of terminal substituents, and their linkages. The more capable the MS/MS technique to resolve these subtle structural differences, the more likely it is to reveal the extent of isomeric variations. In fact, MSⁿ technique has been employed to resolve *N*-glycan structural differences between metastatic and nonmetastatic tumor cells that are apparent only after the third or fourth stage of fragmentation.¹¹⁷ In this context, fragment ions that would be produced at different relative intensities due to different linkages or stereochemistry are normally not useful when both isomeric structures are present and isolated as a single precursor ion. More useful are those linkage-specific fragment ions that occur at different *m/z* values. Detection of as many such ions would inform the presence of as many structural isomers and hence evolved the concept of mapping the biologically important structural motifs by identifying each of the diagnostic ions. For a single isolated precursor, it is not unusual that several tens of isomeric structures may thus be enumerated, especially for mucin-derived larger *O*-glycans that vary not only in terminal glycosylation but also in core structures. For the multiantennary complex-type *N*-glycans, it is usually more difficult to precisely resolve all isomers differing only in positions of antennary substituents.

Suffice here to reiterate that the isomeric depth of each individual glycan can only be fully resolved if one resorts to those advanced MS/MS techniques that can afford linkage-specific ions, rather than relying on just the dominant primary B and Y ions. Collectively, the individual MS/MS data inform the true complexity of a glycome beyond the simplistic MS-level mapping. It is also possible in some MS instruments to effect 'total' fragmentation of all precursors without the selection. The beauty of this approach is the ability to derive a 'full' complement of diagnostic marker ions indicative of all possible terminal glycotopes and core structures that may be present. Conceptually, this may be all that is important for a glycomic mapping, although it is obviously less useful for true sequencing in structural characterization since the causal relationship cannot be established to attribute the observed fragment ions to a particular parent ion.

6.05.4.3.2 Glycosylation and glycomic mapping at MS² level

Conventional MS/MS analysis would require that a parent ion be detected in the first place for it to be isolated for fragmentation. With increasing demand on sensitivity and resolution for glycomics, minor components are often not detected above the background level sufficiently for them to be considered as signal at all. Even if detected, one is still faced with the problem of having the need to manually select too many peaks for MS/MS analysis within a limited life span of the analytes. The consequence is that only the major peaks or the targeted peaks of interest will be analyzed further, whereas those rendered cryptic due to overlapping with the isotopic clusters of major peaks or their usual array of satellite ions, or 'well-recognized' contaminant peaks, will be conveniently neglected. Conceptually, for a more comprehensive and unbiased glycomic coverage, it would be desirable to perform an automated global MS² analysis across the entire useful mass range where the glycans of interest are expected to fall, irrespective of the occurrence of MS signals. In fact, the presence of some components may be better detected at the MS² level based on the characteristic neutral loss and/or diagnostic B ions. The extra selectivity introduced can increase the relative sensitivity of detection by discriminating against noise. This has been elegantly demonstrated with real examples of applications using the built-in total ion mapping (TIM) functionality provided by a linear ion trap MS system.^{21,22,128}

In principle, any MS instrument that is capable of true MS/MS can be used to perform such a global glycomic mapping at the MS² level. The only caveat is that it requires sufficiently high scan speed to cover the desirable mass range and would be practically limited to offline nanoESI-based analysis, which would bring down the effective *m/z* values of high mass components by virtue of multiple charges. At its current iteration, TIM was acquired for every 2 Da with the overlapping precursor isolation window. The built-in software would then allow a reconstruction of an MS²-based profile by registering only those signals that contain the desirable neutral loss or product ions – in a way that a true parent ion or neutral loss scan would – except that a full MS/MS data is available for each of the identified signals, which should be manually examined to further filter out any false positives. TIM can be employed as a prescreen analysis to identify candidate signals including those that are not obvious by MS mapping for subsequent in-depth MSⁿ interrogation. Alternatively, it could be used as an additional level of screening to ensure that components with a particular structural motif of interest are not overlooked. Its completeness in glycomic coverage will, however, still be

limited by suppression of ionization in the first place of those difficult classes of glycans unless they are prefractionated. Furthermore, since it is only implemented at low-energy CID MS² level, it has the same limitation in not being able to filter specifically for epitopes that would require additional stages of MS/MS to distinguish their linkages.

6.05.5 Targeted Glycomics, Glycoproteomics, and the Way Forward

It is fair to say that with the current range of advanced mass analyzers fitted with either ESI or MALDI for high-precision and high-sensitivity MS and MS/MS analyses, protein glycosylation analysis at the level of a single isolated glycoprotein is ultimately limited only by sensitivity and hence the available quantity. A different level of complexity is introduced with systematic glycomics and glycoproteomics. One of the promises held by such discovery sciences is identification of new and exciting targets or markers, but is this noble aim practically feasible? Would such analysis lead to identification of more abundant components that have already been described instead? Can, for example, an oncodevelopmental change in the relative intensity of a particular structural or stereoisomer be sensibly identified by a systematic global-scale mapping either at the MS or MS² level? In the case of glycoproteomics, the mere prospect of resolving all site-specific glycoforms is daunting, if not unattainable. Arguably, any such systematic screening may, at best, sketch out the broad landscape of the glycoproteome so as to furnish the bearings for further navigation by a targeted approach.

6.05.5.1 Prospects for Targeted Glycomics

Targeted glycomics aims specifically to identify specific subsets of the glycome otherwise not detectable or apparent by the initial systematic mapping. The total ion mapping approach described in Section 6.05.4.3.2 is an attempt to comprehensively map the glycome at the MS² level, which allows subsequent interrogation of the presence or otherwise of a particular epitope. It facilitates the identification of components that are prone to neutral loss and produce abundant diagnostic fragment ions such that extra sensitivity is gained by the detection at the MS² level. The major conceptual limitation is that all the targeted glycans need to be competently ionized in the ESI source along with all other more abundant glycomic constituents, as well as any potential contaminants, in the first place. No selectivity in target detection is actually imposed at data acquisition level and it is unlikely to improve the odds of detecting those 'problematic' classes of glycans including the sulfated ones and those larger ones that carry polysialic acid chains or polyLacNAc chains. For these, target enrichment by judicious fractionation scheme is often the simplest and probably the best and only option available. In fact, TIM is well suited to be placed at further downstream of the pipeline, following an offline fractionation and enrichment scheme, as a final alternative step of comprehensive mapping.

For those less problematic glycans that are simply very low in abundance, the selected or multiple reaction monitoring (SRM or MRM) MS technique being recently popularized as an alternative targeted approach to global proteomic analysis¹²⁹ may be likewise adopted for targeted glycomics. In essence, this relies on detecting a predetermined, carefully selected, specific pair of precursor/fragment ion that would inform the presence of a particular biomolecule. Thus, one can monitor specifically over elution time at much higher sensitivity if a particular glycan carrying a particular epitope as defined by one or more diagnostic key ions is present among the glycomic mixtures that are subjected to an LC-MS analysis. With the current generation of triple quadrupole instrument needed for effective implementation of a scheduled SRM, hundreds of such pairs can be realistically monitored in one run without sacrificing sensitivity, provided the targeted glycans can be evenly spaced out along the LC elution time. Thus, one can program an LC-MS SRM run to seek out and quantify the presence of a large number of desirable target glycans. This is much more efficient than the conventional data-dependent LC-MS/MS acquisition, which is usually a random hit and miss for those low-abundance, underrepresented components. However, if a specific epitope can only be resolved from its isomeric variants by virtue of MS³ or even MS⁴, one would need to first somehow induce a total fragmentation at the MS level for a pseudo-MS² ion to be selected for MS³. Alternatively, the capillary LC resolution for glycans needs to be significantly improved by additional dimensions of offline and online separation such that almost all components would be detected, triggered for MS²,

and, if a targeted diagnostic ion representing a glycoepitope of interest be produced, further subjected to data-dependent MS³ for confirmation of true positives, as well as to further obtain linkage-specific fragment ions.

Successful micro- and nanoLC-MS/MS applications to native glycans are currently in place based on negative ion mode,^{51,52,54,130,131} or positive ion mode,⁵⁰ or alternative positive and negative ion modes⁵³ analysis, on a porous graphitized carbon (PGC) column, or positive ion mode analysis on normal phase amide column.³³ Impressive results have been obtained on analysis of both *N*- and *O*-glycans from single glycoproteins, as well as the highly heterogeneous *O*-glycans from mucin, but true tests against cell- and tissue-derived glycomic samples have yet to be demonstrated. The likely shortcomings would be insufficient one-dimensional chromatographic resolution, insufficient dynamic range in detecting and selecting the least abundant or the more 'problematic' classes of glycans for MS/MS, and insufficient MS/MS quality for unambiguous sequencing and linkage determination, as common to most MS/MS analysis of native glycans (Section 6.05.4.2.1). Successful glycomic applications of LC-MS/MS of permethylated glycans have yet to be demonstrated mostly due to their poor chromatographic resolution on a C18 column. In contrast, LC fractionation on PGC offers the advantage of resolving structural isomers by the elution time dimension, which would alleviate the pressure on MS/MS alone to complete the task. In addition, it is conceivable that with highly reproducible LC, collated analysis of the common repertoire of mammalian *N*- and *O*-glycans may eventually contribute to a retention time and MS/MS spectral library. Thus, common and abundant glycans may first be identified based on simple library matching search program, leaving only those unmatched for additional stages of manual analysis. With parallel rapid developments in MS-based shotgun proteomics, such LC-MS/MS-based glycomics along with implementation of label-free quantification would likely represent the way forward for high-throughput and comprehensive analysis.

6.05.5.2 Prospects for Targeted Glycoproteomics

While targeted glycomics is to extract the extra-targeted information from the glycome otherwise not readily obtained by a preceding systematic analysis, targeted glycoproteomics is essentially to first narrow down the scale to better focus on a glycoproteomic subset that is most relevant to the biological questions in hand. A separate glycomic analysis of the released glycans will always be needed (see Section 6.05.3.1) unless precise information on the glycan structural details, complete with linkage position and full isomeric variations, is already available or of no biological consequence to the intended research work. A quick targeted glycomic mapping executed in between each successive steps of glycoprotein/glycopeptide enrichment would also be most useful to determine the extent of enrichment efficiency and the remaining heterogeneity.

6.05.5.2.1 Current limitations in glycopeptide analysis

Ultimately, either systematic or targeted glycoproteomics, or even for a single glycoprotein of interest, analysis of glycopeptides is limited by its poor MS and MS/MS characteristics in giving the single most important information, namely, the peptide sequence ions that are of sufficient quality to allow online database search for direct identification. The relatively large size of most tryptic glycopeptides and their poor ionization efficiency, particularly those multisialylated ones, as well as their low abundance compounded further by the usual spreading of signal intensity across a range of glycoforms for each unique peptide, render them problematic for MS and MS/MS analysis (reviews of glycopeptide analysis are available^{10,132}). MALDI is usually applicable only to isolated glycopeptides from single glycoproteins and better detection is often registered with linear mode, which compromises mass accuracy and resolution particularly at high mass range, and the ability to perform MS/MS. Thus, while it can be employed for rapid screening of the enriched fractions and for occasional analysis of isolated, smaller glycopeptides,^{58,133–135} most efforts for global-scale analysis of glycoproteomic mixtures are focused on positive ion mode ESI analysis in the form of LC-MS or MS/MS.

Current ESI-CID-MS/MS of glycopeptides by any means are dominated by fragmentation of the glycan moiety in a fashion similar to that observed for positive ion mode MS/MS analysis of native glycans. Abundant oxonium ions at the low mass end are diagnostic of the presence of glycopeptides and have been capitalized in many ways to program precursor ion scans, data-dependent MS/MS acquisition of the glycopeptides, or at least to inform their presence.^{100,136–139} These are also relied on to filter out MS/MS data of glycopeptides from those of peptides in a typical LC-MS/MS run¹⁴⁰ and constitute an important input factor in most glycopeptide

identification and scoring algorithm.^{141–143} Apart from the ubiquitous m/z 204 and 366 ions corresponding to HexNAc⁺ and Hex-HexNAc⁺, respectively, other oxonium ions can inform the presence of specific terminal epitopes, for example, NeuAc/NeuGc sialic acids and the fucosylated Lewis epitopes. On the other hand, successive neutral loss from the parent ions may be of limited use in deducing sequence *de novo*. Overall, these glycan-related B and Y ions are sufficient to confirm the probable composition but not definitive enough in the presence of potentially many isobaric structures and certainly not informative for determining linkage. An important Y ion is one that arises from losing all glycosyl residues but the GlcNAc attached to the Asn. The m/z of this (peptide core + GlcNAc) ion, if detected, can inform the molecular mass of the peptide core and thus allows a rapid calculation of the glycan moiety as well as matches the peptide to potential tryptic peptides in the subglycoproteome pool that carries a consensus N-glycosylation sequon. Better still, if this ion can somehow be selected for further MS/MS, it will generate the peptide sequence ions. Unfortunately, although its presence among the MS² ion signals may be readily identified through manual or computer-assisted examination, there is currently no way to automate an intelligent-based data-dependent MS³ to target specifically this ion, which can rank from being a strong signal to almost absent.

6.05.5.2.2 Prospects for emerging MS techniques and solutions

Much excitement has been generated by the advent of electron capture dissociation (ECD)^{144,145} and more recently the electron transfer dissociation (ETD)^{146,147} modes of MS/MS, as applied to glycopeptides. Both are known to produce a series of c ion and z radical ion through cleavage of the N–C α bond of the peptide backbone while leaving the entire glycan moiety largely intact. In principle, this should be sufficient to allow peptide identification but these ions are often of low abundance due to low fragmentation efficiency and much ambiguity is associated with determining their accurate masses. To date, better success has been made with applications to O-glycopeptides from mucin^{148–150} and the IgA hinge region,^{151,152} with simple core 1 and 2 O-glycans, or the O-GlcNAc-modified glycopeptides.¹⁵³ For applications of ETD on N-glycans, proof of concept has been demonstrated against those glycopeptides with relatively simple truncated mannose type^{147,154} and recently extended to sialylated complex types.^{154–156} However, incorporation of ETD-MS/MS into the actual workflow of glycoproteomics, alternating with the normal CID-MS/MS scan to obtain both glycan and peptide sequence for *de novo* identification glycopeptides in one single or two parallel LC-MS/MS runs, has yet to be reported.

It is widely accepted that current ETD generally favors multiply charged precursor ions and is not effective for ions exceeding m/z 1400 or so.^{154,157} Multiply charged glycopeptides at m/z above this range when acted on tend to give abundant charge-reduced species without the effective peptide backbone fragmentation much coveted. A potential solution is to apply additional CID on the charged reduced species with reduced normalized collision energy while lowering the q value in an ion trap.¹⁵⁵ This has been demonstrated to increase the efficiency in obtaining more complete c and z ion series. Recent availability of the option to measure the ETD ions with the high-resolution and high-accuracy Orbitrap¹⁵⁸ may help to improve the mass measurement but its true application has yet to be tested. It remains to be seen if all these additional manipulations may yet compromise the attainable sensitivity to render it not practically useful. It is nevertheless anticipated that coupling to additional mode of ETD analysis in an LC-MS/MS run of enriched glycopeptides may represent the future norm. Bioinformatic software solutions better than that currently available will also be needed to take advantage of both the ETD and CID inputs in the MS/MS data of glycopeptides, along with the glycan and deglycosylated peptide data acquired separately, to score the best match hits in the automated data analysis.

6.05.5.3 Concluding Remarks

Notwithstanding the many technical hurdles briefly outlined above and waiting to be solved, this section on the prospects of MS-based protein glycosylation analysis would end with a highly optimistic tone in anticipation of all those possible future developments. The enabling technologies have taken shape, waiting to be further explored and put into real applications. Mammalian protein glycosylation analysis has rapidly evolved into both targeted glycomics and glycoproteomics out of necessity, which is more on how to intelligently reduce sample complexity while maintaining comprehensiveness of a global analysis to identify targets that matter

most in glycobiology. The future appears to be better invested with LC-MS/MS-based techniques just as the field of proteomics has over the past few years. Nonetheless, MALDI-MS- and MS/MS-based glycomics should still hold their respective niches for their unmatched simplicity and throughput. For nonmammalian protein glycosylation, where novel structures may continue to be uncovered, a complementary low- and high-energy CID-MS/MS as implemented on MALDI systems is often sufficient and most effective to extract as many structural details as possible by homing in on a few major components to get a representative overview of the respective glycomic characteristics. There are probably no *a priori* reasons to dig deeper, unlike the case with mammalian protein glycosylation, particularly of mice and humans, to which we have applied a host of genetic manipulations and need to identify even the subtlest glycoproteomic changes that are of pathophysiological consequence. Targeted analysis is most suited for such applications because we often anticipate what needs to be identified, yet are frustrated enough in not finding it through global analysis.

Abbreviations

| | |
|-----------------------|---|
| CID | collision-induced dissociation |
| ECD | electron capture dissociation |
| EI | electron impact |
| ER | endoplasmic reticulum |
| ESI | electrospray ionization |
| ETD | electron transfer dissociation |
| FAB | fast-atom bombardment |
| FT-ICR | fourier transformed-ion cyclotron resonance |
| GC | gas chromatography |
| GPI | glycosyl phosphatidylinositol |
| HILIC | hydrophilic interaction liquid chromatography |
| LacNAc | <i>N</i> -acetyllactosamine |
| LC | liquid chromatography |
| LeX | Lewis X |
| LSIMS | liquid secondary ionization MS |
| MALDI | matrix-assisted laser desorption ionization |
| MRM | multiple reaction monitoring |
| MS | mass spectrometry |
| MS/MS | tandem mass spectrometry, MS ² |
| MSⁿ | multistages tandem MS |
| nanoESI | nanospray |
| PGC | porous graphitized carbon |
| PNGase | peptide- <i>N</i> -glycosidase |
| PSD | post-source decay |
| Q/TOF | quadrupole/time-of-flight |
| SRM | selected reaction monitoring |
| TIM | total ion mapping |
| TOF | time-of-flight |

References

1. N. Taniguchi, *Mol. Cell Proteomics* **2008**, 7 (3), 626–627.
2. Y. Wada; P. Azadi; C. E. Costello; A. Dell; R. A. Dwek; H. Geyer; R. Geyer; K. Takechi; N. G. Karlsson; K. Kato; N. Kawasaki; K. H. Khoo; S. Kim; A. Kondo; E. Lattova; Y. Mechref; E. Miyoshi; K. Nakamura; H. Narimatsu; M. V. Novotny; N. H. Packer; H. Perreault; J. Peter-Katalinic; G. Pohlentz; V. N. Reinhold; P. M. Rudd; A. Suzuki; N. Taniguchi, *Glycobiology* **2007**, 17 (4), 411–422.
3. A. Dell; H. R. Morris, *Science* **2001**, 291 (5512), 2351–2356.

4. J. Zaia, *Mass Spectrom. Rev.* **2004**, 23 (3), 161–227.
5. D. J. Harvey, *Proteomics* **2005**, 5 (7), 1774–1786.
6. H. Geyer; R. Geyer, *Biochim. Biophys. Acta* **2006**, 1764 (12), 1853–1869.
7. J. Jang-Lee; S. J. North; M. Sutton-Smith; D. Goldberg; M. Panico; H. Morris; S. Haslam; A. Dell, *Methods Enzymol.* **2006**, 415, 59–86.
8. W. Morelle; K. Canis; F. Chirat; V. Faid; J. C. Michalski, *Proteomics* **2006**, 6 (14), 3993–4015.
9. W. Morelle; J. C. Michalski, *Nat. Protoc.* **2007**, 2 (7), 1585–1602.
10. M. Wuhrer; M. I. Catalina; A. M. Deelder; C. H. Hokke, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2007**, 849 (1–2), 115–128.
11. S. M. Haslam; S. Julien; J. M. Burchell; C. R. Monk; A. Ceroni; O. A. Garden; A. Dell, *Immunol. Cell Biol.* **2008**, 86 (7), 564–573.
12. J. Zaia, *Chem. Biol.* **2008**, 15 (9), 881–892.
13. Y. Mechref; M. V. Novotny, *Mass Spectrom. Rev.* **2009**, 28 (2), 207–222.
14. M. Wuhrer; A. R. de Boer; A. M. Deelder, *Mass Spectrom. Rev.* **2009**, 28 (2), 192–206.
15. J. D. Marth; P. K. Grewal, *Nat. Rev. Immunol.* **2008**, 8 (11), 874–887.
16. A. Varki; R. D. Cummings; J. D. Esko; H. H. Freeze; P. Stanley; C. R. Bertozzi; G. W. Hart; M. E. Etzler, *Essentials of Glycobiology*, 2nd edn; CSHL Press: Cold Spring Harbor, NY, 2009; p 784.
17. M. Abu-Qarn; J. Eichler; N. Sharon, *Curr. Opin. Struct. Biol.* **2008**, 18 (5), 544–550.
18. V. Tretter; F. Altmann; L. Marz, *Eur. J. Biochem.* **1991**, 199 (3), 647–652.
19. R. Ranzinger; S. Herget; T. Wetter; C. W. von der Lieth, *BMC Bioinf.* **2008**, 9 (1), 384.
20. S. M. Haslam; S. J. North; A. Dell, *Curr. Opin. Struct. Biol.* **2006**, 16 (5), 584–591.
21. K. Aoki; M. Perlman; J. M. Lim; R. Cantu; L. Wells; M. Tiemeyer, *J. Biol. Chem.* **2007**, 282 (12), 9127–9142.
22. K. Aoki; M. Porterfield; S. S. Lee; B. Dong; K. Nguyen; K. H. McGlamry; M. Tiemeyer, *J. Biol. Chem.* **2008**, 283 (44), 30385–30400.
23. Y. Guerardel; L. Y. Chang; E. Maes; C. J. Huang; K. H. Khoo, *Glycobiology* **2006**, 16 (3), 244–257.
24. C. H. Hokke; A. M. Deelder; K. F. Hoffmann; M. Wuhrer, *Exp. Parasitol.* **2007**, 117 (3), 275–283.
25. K. Paschinger; M. Gutternigg; D. Rendic; I. B. Wilson, *Carbohydr. Res.* **2008**, 343 (12), 2041–2049.
26. B. Kuster; T. N. Krogh; E. Mortz; D. J. Harvey, *Proteomics* **2001**, 1 (2), 350–361.
27. J. Peter-Katalinic, *Methods Enzymol.* **2005**, 405, 139–171.
28. L. Royle; T. S. Mattu; E. Hart; J. I. Langridge; A. H. Merry; N. Murphy; D. J. Harvey; R. A. Dwek; P. M. Rudd, *Anal. Biochem.* **2002**, 304 (1), 70–90.
29. K. Yamada; S. Hyodo; Y. K. Matsuno; M. Kinoshita; S. Z. Maruyama; Y. S. Osaka; E. Casal; Y. C. Lee; K. Takeuchi, *Anal. Biochem.* **2007**, 371 (1), 52–61.
30. J. C. Bigge; T. P. Patel; J. A. Bruce; P. N. Goulding; S. M. Charles; R. B. Parekh, *Anal. Biochem.* **1995**, 230 (2), 229–238.
31. N. Takahashi; H. Nakagawa; K. Fujikawa; Y. Kawamura; N. Tomiya, *Anal. Biochem.* **1995**, 226 (1), 139–146.
32. K. R. Anumula, *Anal. Biochem.* **2006**, 350 (1), 1–23.
33. M. Wuhrer; C. A. Koelman; A. M. Deelder; C. H. Hokke, *Anal. Chem.* **2004**, 76 (3), 833–838.
34. K. Deguchi; T. Keira; K. Yamada; H. Ito; Y. Takegawa; H. Nakagawa; S. Nishimura, *J. Chromatogr. A* **2008**, 1189 (1–2), 169–174.
35. B. Lindberg; J. Lonngren, *Methods Enzymol.* **1978**, 50, 3–33.
36. R. Geyer; H. Geyer, *Methods Enzymol.* **1994**, 230, 86–108.
37. P. Kang; Y. Mechref; M. V. Novotny, *Rapid Commun. Mass Spectrom.* **2008**, 22 (5), 721–734.
38. S. Hakomori, *J. Biochem.* **1964**, 55, 205–208.
39. I. Ciucanu; F. Kerek, *Carbohydr. Res.* **1984**, 131 (2), 209–217.
40. M. J. McConville; J. E. Thomas-Oates; M. A. Ferguson; S. W. Homans, *J. Biol. Chem.* **1990**, 265 (32), 19611–19623.
41. K. H. Khoo; H. R. Morris; R. A. McDowell; A. Dell; M. Maccarana; U. Lindahl, *Carbohydr. Res.* **1993**, 244 (2), 205–223.
42. T. Taguchi; M. Iwasaki; Y. Muto; K. Kitajima; S. Inoue; K. H. Khoo; H. R. Morris; A. Dell; Y. Inoue, *Eur. J. Biochem.* **1996**, 238 (2), 357–367.
43. J. Mitoma; X. Bao; B. Petryanik; P. Schaerli; J. M. Gauguier; S. Y. Yu; H. Kawashima; H. Saito; K. Ohtsubo; J. D. Marth; K. H. Khoo; U. H. von Andrian; J. B. Lowe; M. Fukuda, *Nat. Immunol.* **2007**, 8 (4), 409–418.
44. E. Garenaux; S. Y. Yu; D. Florea; G. Strecker; K. H. Khoo; Y. Guerardel, *Glycoconj. J.* **2008**, 25 (9), 903–915.
45. A. K. Powell; D. J. Harvey, *Rapid Commun. Mass Spectrom.* **1996**, 10 (9), 1027–1032.
46. S. Sekiya; Y. Wada; K. Tanaka, *Anal. Chem.* **2005**, 77 (15), 4962–4968.
47. M. Toyoda; H. Ito; Y. K. Matsuno; H. Narimatsu; A. Kameyama, *Anal. Chem.* **2008**, 80 (13), 5211–5218.
48. S. F. Wheeler; P. Domann; D. J. Harvey, *Rapid Commun. Mass Spectrom.* **2009**, 23 (2), 303–312.
49. N. H. Packer; M. A. Lawson; D. R. Jardine; J. W. Redmond, *Glycoconj. J.* **1998**, 15 (8), 737–747.
50. N. Kawasaki; S. Itoh; M. Ohta; T. Hayakawa, *Anal. Biochem.* **2003**, 316 (1), 15–22.
51. N. G. Karlsson; B. L. Schulz; N. H. Packer, *J. Am. Soc. Mass Spectrom.* **2004**, 15 (5), 659–672.
52. N. G. Karlsson; N. L. Wilson; H. J. Wirth; P. Dawes; H. Joshi; N. H. Packer, *Rapid Commun. Mass Spectrom.* **2004**, 18 (19), 2282–2292.
53. S. Itoh; N. Kawasaki; N. Hashii; A. Harazono; Y. Matsuishi; T. Hayakawa; T. Kawanishi, *J. Chromatogr. A* **2006**, 1103 (2), 296–306.
54. M. Backstrom; K. A. Thomsson; H. Karlsson; G. C. Hansson, *J. Proteome Res.* **2008**, 7, 538–545.
55. L. R. Ruhaak; C. Huhn; W. J. Waterreus; A. R. de Boer; C. Neuss; C. H. Hokke; A. M. Deelder; M. Wuhrer, *Anal. Chem.* **2008**, 80 (15), 6119–6126.
56. R. R. Townsend; M. R. Hardy; Y. C. Lee, *Methods Enzymol.* **1989**, 179, 65–76.
57. C. Bruggink; M. Wuhrer; C. A. Koelman; V. Barreto; Y. Liu; C. Pohl; A. Ingendoh; C. H. Hokke; A. M. Deelder, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2005**, 829 (1–2), 136–143.
58. Y. Wada; M. Tajiri; S. Yoshida, *Anal. Chem.* **2004**, 76 (22), 6560–6565.
59. P. Hagglund; J. Bunkenborg; F. Elortza; O. N. Jensen; P. Roepstorff, *J. Proteome Res.* **2004**, 3 (3), 556–566.
60. C. D. Calvano; C. G. Zamboni; O. N. Jensen, *J. Proteomics* **2008**, 71 (3), 304–317.

61. M. Kullolli; W. S. Hancock; M. Hincapie, *J. Sep. Sci.* **2008**, *31* (14), 2733–2739.
62. M. Madera; B. Mann; Y. Mechref; M. V. Novotny, *J. Sep. Sci.* **2008**, *31* (14), 2722–2732.
63. Y. Mechref; M. Madera; M. V. Novotny, *Methods Mol. Biol.* **2008**, *424*, 373–396.
64. A. Dell, *Adv. Carbohydr. Chem. Biochem.* **1987**, *45*, 19–72.
65. H. Egge; J. Peter-Katalinic, *Mass Spectrom. Rev.* **1987**, *6*, 331–393.
66. M. Wührer; A. M. Deelder, *Rapid Commun. Mass Spectrom.* **2006**, *20* (6), 943–951.
67. M. Terada; K. H. Khoo; R. Inoue; C. I. Chen; K. Yamada; H. Sakaguchi; N. Kadowaki; B. Y. Ma; S. Oka; T. Kawasaki; N. Kawasaki, *J. Biol. Chem.* **2005**, *280* (12), 10897–10913.
68. H. H. Huang; P. L. Tsai; K. H. Khoo, *Glycobiology* **2001**, *11* (5), 395–406.
69. J. Yuan; N. Hashii; N. Kawasaki; S. Itoh; T. Kawanishi; T. Hayakawa, *J. Chromatogr. A* **2005**, *1067* (1–2), 145–152.
70. N. Hashii; N. Kawasaki; S. Itoh; Y. Nakajima; T. Kawanishi; T. Yamaguchi, *Immunology* **2009**, *126*, 336–345.
71. P. Kang; Y. Mechref; Z. Kyselova; J. A. Goetz; M. V. Novotny, *Anal. Chem.* **2007**, *79* (16), 6064–6073.
72. G. Alvarez-Manilla; N. L. Warren; T. Abney; J. Atwood, 3rd; P. Azadi; W. S. York; M. Pierce; R. Orlando, *Glycobiology* **2007**, *17* (7), 677–687.
73. J. A. Atwood, III; L. Cheng; G. Alvarez-Manilla; N. L. Warren; W. S. York; R. Orlando, *J. Proteome Res.* **2008**, *7* (1), 367–374.
74. N. K. Kochetkov; O. S. Chizhov, *Adv. Carbohydr. Chem. Biochem.* **1966**, *21*, 39–93.
75. B. L. Gillece-Castro; A. L. Burlingame, *Methods Enzymol.* **1990**, *193*, 689–712.
76. W. J. Richter; D. R. Muller; B. Domon, *Methods Enzymol.* **1990**, *193*, 607–623.
77. J. Lemoine; B. Fournet; D. Despeyroux; K. R. Jennings; R. Rosenberg; E. de Hoffmann, *J. Am. Soc. Mass Spectrom.* **1993**, *4* (3), 197–302.
78. V. N. Reinhold; B. B. Reinhold; S. Chan, *Methods Enzymol.* **1996**, *271*, 377–402.
79. N. Vieux; E. de Hoffmann; B. Domon, *Anal. Chem.* **1997**, *69* (16), 3193–3198.
80. N. Vieux; E. de Hoffmann; B. Domon, *Anal. Chem.* **1998**, *70* (23), 4951–4959.
81. A. S. Weiskopf; P. Vouros; D. J. Harvey, *Rapid Commun. Mass Spectrom.* **1997**, *11* (14), 1493–1504.
82. A. S. Weiskopf; P. Vouros; D. J. Harvey, *Anal. Chem.* **1998**, *70* (20), 4441–4447.
83. D. J. Harvey, *J. Mass Spectrom.* **2000**, *35* (10), 1178–1190.
84. C. Robbe; C. Capon; B. Coddeville; J. C. Michalski, *Rapid Commun. Mass Spectrom.* **2004**, *18* (4), 412–420.
85. D. J. Harvey, *J. Am. Soc. Mass Spectrom.* **2000**, *11* (10), 900–915.
86. D. J. Harvey; T. S. Mattu; M. R. Wormald; L. Royle; R. A. Dwek; P. M. Rudd, *Anal. Chem.* **2002**, *74* (4), 734–740.
87. D. J. Harvey, *J. Mass Spectrom.* **2005**, *40* (5), 642–653.
88. W. Morelle; A. Page; J. C. Michalski, *Rapid Commun. Mass Spectrom.* **2005**, *19* (9), 1145–1158.
89. K. Deguchi; Y. Takegawa; H. Ito; N. Miura; S. Yoshioka; S. Nagai; H. Nakagawa; S. I. Nishimura, *Rapid Commun. Mass Spectrom.* **2005**, *20* (3), 412–418.
90. D. J. Harvey, *Mass Spectrom. Rev.* **1999**, *18* (6), 349–450.
91. D. J. Harvey, *J. Am. Soc. Mass Spectrom.* **2000**, *11* (6), 572–577.
92. D. J. Harvey; R. H. Bateman; R. S. Bordoli; R. Tyldesley, *Rapid Commun. Mass Spectrom.* **2000**, *14* (22), 2135–2142.
93. Y. Mechref; M. V. Novotny; C. Krishnan, *Anal. Chem.* **2003**, *75* (18), 4895–4903.
94. W. Morelle; M. C. Slomianny; H. Diemer; C. Schaeffer; A. V. Dorsselaer; J. C. Michalski, *Rapid Commun. Mass Spectrom.* **2004**, *18* (22), 2637–2649.
95. E. Spina; L. Sturiale; D. Romeo; G. Impallomeni; D. Garozzo; D. Waidelich; M. Glueckmann, *Rapid Commun. Mass Spectrom.* **2004**, *18* (4), 392–398.
96. E. Stephens; S. L. Maslen; L. G. Green; D. H. Williams, *Anal. Chem.* **2004**, *76* (8), 2343–2354.
97. K. A. Thomsson; N. G. Karlsson; G. C. Hansson, *J. Chromatogr. A* **1999**, *854* (1–2), 131–139.
98. K. A. Thomsson; H. Karlsson; G. C. Hansson, *Anal. Chem.* **2000**, *72* (19), 4543–4549.
99. B. Domon; C. E. Costello, *Glycoconj. J.* **1988**, *5*, 397–409.
100. M. J. Huddleston; M. F. Bean; S. A. Carr, *Anal. Chem.* **1993**, *65* (7), 877–884.
101. A. Dell; A. J. Reason; K. H. Khoo; M. Panico; R. A. McDowell; H. R. Morris, *Methods Enzymol.* **1994**, *230*, 108–132.
102. S. Y. Yu; S. W. Wu; K. H. Khoo, *Glycoconj. J.* **2006**, *23* (5–6), 355–369.
103. Y. Y. Fan; S. Y. Yu; H. Ito; A. Kameyama; T. Sato; C. H. Lin; L. C. Yu; H. Narimatsu; K. H. Khoo, *J. Biol. Chem.* **2008**, *283* (24), 16455–16468.
104. S. Y. Yu; K. H. Khoo; Z. Yang; A. Herp; A. M. Wu, *Glycoconj. J.* **2008**, *25* (3), 199–212.
105. W. Chai; V. Piskarev; A. M. Lawson, *Anal. Chem.* **2001**, *73* (3), 651–657.
106. A. Pfenninger; M. Karas; B. Finke; B. Stahl, *J. Am. Soc. Mass Spectrom.* **2002**, *13* (11), 1341–1348.
107. D. Sagi; J. Peter-Katalinic; H. S. Conradt; M. Nimitz, *J. Am. Soc. Mass Spectrom.* **2002**, *13* (9), 1138–1148.
108. D. J. Harvey, *J. Am. Soc. Mass Spectrom.* **2005**, *16* (5), 631–646.
109. D. J. Harvey, *J. Am. Soc. Mass Spectrom.* **2005**, *16* (5), 647–659.
110. W. Chai; V. E. Piskarev; B. Mulloy; Y. Liu; P. G. Evans; H. M. Osborn; A. M. Lawson, *Anal. Chem.* **2006**, *78* (5), 1581–1592.
111. D. J. Harvey; K. Baruah; C. N. Scanlan, *J. Mass Spectrom.* **2009**, *44*, 50–60.
112. D. J. Harvey; L. Royle; C. M. Radcliffe; P. M. Rudd; R. A. Dwek, *Anal. Biochem.* **2008**, *376* (1), 44–60.
113. S. F. Wheeler; D. J. Harvey, *Anal. Chem.* **2000**, *72* (20), 5027–5039.
114. D. J. Harvey; R. H. Bateman; M. R. Green, *J. Mass Spectrom.* **1997**, *32* (2), 167–187.
115. D. Ashline; S. Singh; A. Hanneman; V. Reinhold, *Anal. Chem.* **2005**, *77* (19), 6250–6262.
116. D. J. Ashline; A. J. Lapadula; Y. H. Liu; M. Lin; M. Grace; B. Pramanik; V. N. Reinhold, *Anal. Chem.* **2007**, *79* (10), 3830–3842.
117. J. M. Prien; L. C. Huysentruyt; D. J. Ashline; A. J. Lapadula; T. N. Seyfried; V. N. Reinhold, *Glycobiology* **2008**, *18* (5), 353–366.
118. R. M. Anthony; F. Nimmerjahn; D. J. Ashline; V. N. Reinhold; J. C. Paulson; J. V. Ravetch, *Science* **2008**, *320* (5874), 373–376.
119. Y. Takegawa; S. Ito; S. Yoshioka; K. Deguchi; H. Nakagawa; K. Monde; S. I. Nishimura, *Rapid Commun. Mass Spectrom.* **2004**, *18* (4), 385–391.
120. A. Kameyama; N. Kikuchi; S. Nakaya; H. Ito; T. Sato; T. Shikanai; Y. Takahashi; K. Takahashi; H. Narimatsu, *Anal. Chem.* **2005**, *77* (15), 4719–4725.

121. H. Zhang; S. Singh; V. N. Reinhold, *Anal. Chem.* **2005**, *77* (19), 6263–6270.
122. A. J. Lapadula; P. J. Hatcher; A. J. Hanneman; D. J. Ashline; H. Zhang; V. N. Reinhold, *Anal. Chem.* **2005**, *77* (19), 6271–6279.
123. M. Sutton-Smith; N. K. Wong; K. H. Khoo; S. W. Wu; S. Y. Yu; M. S. Patankar; R. Easton; F. A. Lattanzio; H. R. Morris; A. Dell; G. F. Clark, *Glycobiology* **2007**, *17* (6), 553–567.
124. P. Babu; S. J. North; J. Jang-Lee; S. Chalabi; K. Mackerness; S. R. Stowell; R. D. Cummings; S. Rankin; A. Dell; S. M. Haslam, *Glycoconj. J.* **2008**.
125. Y. Mechref; P. Kang; M. V. Novotny, *Rapid Commun. Mass Spectrom.* **2006**, *20* (8), 1381–1389.
126. H. S. Chen; J. M. Chen; C. W. Lin; K. H. Khoo; I. H. Tsai, *FEBS J.* **2008**, *275* (15), 3944–3958.
127. K. Maass; R. Ranzinger; H. Geyer; C. W. von der Lieth; R. Geyer, *Proteomics* **2007**, *7* (24), 4435–4444.
128. K. L. Abbott; K. Aoki; J. M. Lim; M. Porterfield; R. Johnson; R. M. O'Regan; L. Wells; M. Tiemeyer; M. Pierce, *J. Proteome Res.* **2008**, *7* (4), 1470–1480.
129. V. Lange; P. Picotti; B. Domon; R. Aebersold, *Mol. Syst. Biol.* **2008**, *4*, 222.
130. Y. Andersch-Bjorkman; K. A. Thomsson; J. M. Holmen Larsson; E. Ekerhovd; G. C. Hansson, *Mol. Cell Proteomics* **2007**, *6* (4), 708–716.
131. N. L. Wilson; L. J. Robinson; A. Donnet; L. Bovoetto; N. H. Packer; N. G. Karlsson, *J. Proteome Res.* **2008**, *7* (9), 3687–3696.
132. D. S. Dalpathado; H. Desaire, *Analyst* **2008**, *133* (6), 731–738.
133. M. Wuhler; C. H. Hokke; A. M. Deelder, *Rapid Commun. Mass Spectrom.* **2004**, *18* (15), 1741–1748.
134. J. Irungu; E. P. Go; Y. Zhang; D. S. Dalpathado; H. X. Liao; B. F. Haynes; H. Desaire, *J. Am. Soc. Mass Spectrom.* **2008**, *19* (8), 1209–1220.
135. K. Kubota; Y. Sato; Y. Suzuki; N. Goto-Inoue; T. Toda; M. Suzuki; S. Hisanaga; A. Suzuki; T. Endo, *Anal. Chem.* **2008**, *80* (10), 3693–3698.
136. M. A. Ritchie; A. C. Gill; M. J. Deery; K. Lilley, *J. Am. Soc. Mass Spectrom.* **2002**, *13* (9), 1065–1077.
137. J. Jebanathirajah; H. Steen; P. Roepstorff, *J. Am. Soc. Mass Spectrom.* **2003**, *14* (7), 777–784.
138. K. Sandra; B. Devreese; J. Van Beeumen; I. Stals; M. Claeysens, *J. Am. Soc. Mass Spectrom.* **2004**, *15* (3), 413–423.
139. B. Sullivan; T. A. Addona; S. A. Carr, *Anal. Chem.* **2004**, *76* (11), 3112–3118.
140. S. Itoh; A. Hachisuka; N. Kawasaki; N. Hashii; R. Teshima; T. Hayakawa; T. Kawanishi; T. Yamaguchi, *Biochemistry* **2008**, *47* (38), 10132–10154.
141. D. Goldberg; M. Bern; S. Parry; M. Sutton-Smith; M. Panico; H. R. Morris; A. Dell, *J. Proteome Res.* **2007**, *6* (10), 3995–4005.
142. S. Joenvaara; I. Ritamo; H. Peltoniemi; R. Renkonen, *Glycobiology* **2008**, *18* (4), 339–349.
143. O. Ozohanics; J. Krenyacz; K. Ludanyi; F. Pollreis; K. Vekey; L. Drahos, *Rapid Commun. Mass Spectrom.* **2008**, *22* (20), 3245–3254.
144. E. Mirgorodskaya; P. Roepstorff; R. A. Zubarev, *Anal. Chem.* **1999**, *71* (20), 4431–4436.
145. K. Hakansson; H. J. Cooper; M. R. Emmett; C. E. Costello; A. G. Marshall; C. L. Nilsson, *Anal. Chem.* **2001**, *73* (18), 4530–4536.
146. J. M. Hogan; S. J. Pitteri; P. A. Chrisman; S. A. McLuckey, *J. Proteome Res.* **2005**, *4* (2), 628–632.
147. M. I. Catalina; C. A. Koeleman; A. M. Deelder; M. Wuhler, *Rapid Commun. Mass Spectrom.* **2007**, *21* (6), 1053–1061.
148. M. Mormann; H. Paulsen; J. Peter-Katalinic, *Eur. J. Mass Spectrom. (Chichester, Eng.)* **2005**, *11* (5), 497–511.
149. C. Sihlbom; I. van Dijk Hard; M. E. Lidell; T. Noll; G. C. Hansson; M. Backstrom, *Glycobiology* **2009**, *19* (4), 375–381.
150. I. Perdivara; R. Petrovich; B. Alliquant; L. J. Deterding; K. B. Tomer; M. Przybylski, *J. Proteome Res.* **2009**, *8* (2), 631–642.
151. M. B. Renfrow; H. J. Cooper; M. Tomana; R. Kulhavy; Y. Hiki; K. Toma; M. R. Emmett; J. Mestecky; A. G. Marshall; J. Novak, *J. Biol. Chem.* **2005**, *280* (19), 19136–19145.
152. M. B. Renfrow; C. L. Mackay; M. J. Chalmers; B. A. Julian; J. Mestecky; M. Kilian; K. Poulsen; M. R. Emmett; A. G. Marshall; J. Novak, *Anal. Bioanal. Chem.* **2007**, *389* (5), 1397–1407.
153. N. Khidekel; S. B. Ficarro; P. M. Clark; M. C. Bryan; D. L. Swaney; J. E. Rexach; Y. E. Sun; J. J. Coon; E. C. Peters; L. C. Hsieh-Wilson, *Nat. Chem. Biol.* **2007**, *3* (6), 339–348.
154. W. R. Alley, Jr.; Y. Mechref; M. V. Novotny, *Rapid Commun. Mass Spectrom.* **2009**, *23* (1), 161–170.
155. S. L. Wu; A. F. Huhmer; Z. Hao; B. L. Karger, *J. Proteome Res.* **2007**, *6* (11), 4230–4244.
156. M. Wuhler; J. C. Stam; F. E. van de Geijn; C. A. Koeleman; C. T. Verrips; R. J. Dolhain; C. H. Hokke; A. M. Deelder, *Proteomics* **2007**, *7* (22), 4070–4081.
157. D. L. Swaney; G. C. McAlister; M. Wirtala; J. C. Schwartz; J. E. Syka; J. J. Coon, *Anal. Chem.* **2007**, *79* (2), 477–485.
158. G. C. McAlister; W. T. Berggren; J. Griep-Raming; S. Horning; A. Makarov; D. Phanstiel; G. Stafford; D. L. Swaney; J. E. Syka; V. Zabrouskov; J. J. Coon, *J. Proteome Res.* **2008**, *7* (8), 3127–3136.

Biographical Sketch



Kay-Hooi Khoo graduated from the Department of Biochemistry, Imperial College, London, in 1989 and went on to do his Ph.D. studies with Professor Anne Dell at the same department. Specializing in mass spectrometry applications to solving glycan structures, he obtained his doctorate degree in 1993 and stayed for another year and a half at Imperial College as a Wellcome Prize Fellow, before moving to Colorado State University in 1994 for another postdoctoral fellowship. At Colorado State University, he focused on analyzing the mycobacterial cell wall-associated glycoconjugates, a work that was initiated during his Ph.D. studies in collaboration with Dr. Delphi Chatterjee and Professor Patrick Brennan of the Mycobacterial Research Laboratories at Colorado State University. In 1996, he was recruited as an associate research fellow to the Institute of Biological Chemistry, Academia Sinica, Taiwan, where he is now a full research fellow. Dr. Khoo's research interests remain focused on applications of mass spectrometry to glycan analysis, but have expanded to include protein glycosylation, and more recently to glycomics, proteomics, and glycoproteomics. He is the director of the Core Facilities for Proteomics and Glycomics under the Taiwan National Research Program in Genomic Medicine and is engaged as participating investigator in the NIH Consortium for Functional Glycomics, and the HGPI of human proteome organisation (HUPO).

6.06 Glycoanalysis of Bacterial Glycome

Chuan Wang, Camille L. A. Hamula, and Xing-Fang Li, University of Alberta, Edmonton, AB, Canada

Xin Liu and Jianjun Li, National Research Council of Canada, Ottawa, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|------------------------------|-----|
| 6.06.1 | Introduction | 157 |
| 6.06.2 | Protein Glycosylation | 158 |
| 6.06.2.1 | O-Glycosylation | 158 |
| 6.06.2.2 | N-Glycosylation | 158 |
| 6.06.3 | Lipopolysaccharide | 159 |
| 6.06.3.1 | R-Type LPS | 160 |
| 6.06.3.1.1 | Analysis of O-deacylated LPS | 160 |
| 6.06.3.1.2 | Core analysis | 163 |
| 6.06.3.1.3 | Lipid A analysis | 164 |
| 6.06.3.2 | S-Type LPS | 165 |
| 6.06.4 | Capsular Polysaccharides | 168 |
| 6.06.5 | Conclusions | 169 |
| References | | 170 |

6.06.1 Introduction

The bacterial glycome encompasses all the glycan structures present in a cell. Glycomics is the systematic study of glycomes, including their genetic, physiologic, pathologic, and other aspects. Lipopolysaccharides (LPS) are the major components of the outer membrane of Gram-negative bacteria. LPS contains O-linked polysaccharides (O-PS) comprised of repeating oligosaccharide units. The O-PS repeating units can be referred to as the O-antigens due to their immunogenicity. Some bacteria produce capsular polysaccharides (CPS), which are referred to as the K-antigens to distinguish them from the O-antigens. Bacterial polysaccharides are widespread and most often are found to be associated with diseases. Examples include the lipooligosaccharide (LOS) produced by *Campylobacter jejuni* in Guillain-Barré syndrome (GBS) and the pneumococcal polysaccharides, which are thought to be responsible for the virulence of encapsulated strains of *Streptococcus pneumoniae*. The glycome of a cell also includes glycoproteins in addition to the cell surface polysaccharides. Proteins can undergo co-translational and posttranslational modifications, divided into two categories, O- and N-linked glycosylation. Protein glycosylation is an important determinant of many different biological processes, including protein-protein interactions, protein trafficking and folding, immune recognition, cell adhesion and migration, and intercellular signaling. In the context of bacterial glycomics, the identity of the entirety of carbohydrates is thus collectively referred to as the glycome, including N- and O-linked glycans, smooth (S-type LPS) and rough lipopolysaccharides (R-type LPS), and capsular polysaccharides. Hence, research in bacterial glycomics has to deal with an inherent level of complexity compounded by the relatively small amounts of material obtainable from biological sources. To this end, mass spectrometry (MS) has played an important role as a sensitive structural tool in the elucidation of trace-level peptide, oligosaccharide, glycolipid, and protein structures.

MS, especially in combination with advanced separation techniques, is one of the most powerful and versatile techniques for the structural analysis of bacterial glycomes. Modern mass spectral ionization techniques such as electrospray (ESI) and matrix-assisted laser desorption/ionization (MALDI) provide detection limits in the high atto- to low femto-mole range for the identification of peptides and complex carbohydrates. Structural characterization of these trace level components can be achieved using tandem MS. This provides a number of specific scanning functions such as product, precursor ion, and constant neutral loss scanning to

probe specific functionalities on target compounds. The power of ultrahigh sensitivity MS strategies for defining the primary structures of highly complex mixtures of glycoprotein glycoforms is set to revolutionize structural glycobiology in the coming postgenomic era.¹

6.06.2 Protein Glycosylation

Posttranslational glycosylation is a universal modification of proteins in eukaryotes, archaea, and bacteria. Protein glycosylation was previously considered to be restricted to eukaryotes until recent reports of a bacterial N-linked glycosylation pathway in *C. jejuni*.^{2,3} Through advances in analytical methods and genome sequencing, there have been increasing reports of both O- and N-linked protein glycosylation pathways in bacteria, particularly among mucosal-associated pathogens.⁴

6.06.2.1 O-Glycosylation

Compared to the N-glycosylation system (see Section 6.06.2.2), the molecular basis of O-glycosylation processes are still not well understood. The attachment site of glycans does not appear to be a consensus peptide sequence.⁵ Pilin from *Neisseria meningitidis*, *N. gonorrhoeae*, or some *Pseudomonas aeruginosa* strains, and flagella from *C. jejuni* are O-linked glycosylated proteins that have been extensively studied.

The bacterial flagellum has a remarkable structure, and can be found in both Gram-positive and Gram-negative bacteria. It is a long, thin filament that protrudes from the cell body and is comprised of an engine and a propeller that are joined by a flexible hook.⁶ Flagellar glycosylation commonly occurs in numerous bacterial pathogens.^{7–12} Failure to glycosylate the flagellin protein leads to loss of bacterial motility, and the organisms can no longer colonize their respective hosts.¹³

A ‘bottom-up’ approach is generally used to identify sites of flagellar glycosylation. Typically, flagellins are digested with trypsin and analyzed using liquid chromatography–mass spectrometry (LC–MS).^{6,8–10} The analysis of suspected glycopeptides can be achieved via the detection of characteristic oxonium ions from the cleavage of glycosidic bonds generated by in-source fragmentation or from selected precursor ions in tandem MS experiments (MS/MS).^{6,10,11} On the other hand, a ‘top-down’ strategy provides information on the extent of glycosylation, the molecular masses, and the structures of oligosaccharide residues on bacterial flagella.¹¹ In this approach, the intact proteins are analyzed first and their molecular masses are obtained, which can indicate the extent of glycosylation when amino acid sequences are available. The nature of the glycosylation can be obtained from MS/MS spectra of multiply charged protein ions. Confirmation of the potential structure of the carbohydrate residues can be achieved by acquiring second-generation MS/MS spectra of oxonium ions formed upon collisional activation in the orifice region of the electrospray interface. The top-down strategy is a rapid and relatively straightforward approach for routine screening of protein samples, and it can identify unusual carbohydrate residues and glycosylation changes.

6.06.2.2 N-Glycosylation

Until recent years, N-glycosylation was detected only in eukaryotes. Two studies have confirmed a bacterial N-linked glycosylation pathway in the human gastrointestinal pathogen *C. jejuni*.^{2,3} The N-linked glycosylation pathway in bacteria is less complex than in eukaryotes. In *C. jejuni*, a single gene cluster named *pgl* (protein glycosylation) is necessary and sufficient for the N-linked glycosylation pathway.³ The oligosaccharyltransferase PglB is an integral membrane protein whose sequence is highly similar to its counterpart in eukaryotes, STT3.⁵ PglB transfers oligosaccharides (GalNAc5GlcBac) from a lipid-pyrophosphate donor to asparagine side chains.³ The precise role of glycosylation systems in mucosal bacterial pathogens has yet to be resolved, but since many of these organisms dedicate a significant proportion of their relatively small genomes to glycosylating proteins, these modifications are likely to have crucial roles, similar to their eukaryotic counterparts.⁴

For eukaryotic systems, the specific functions and the relationship between their structure and functions are commonly studied using the approach of exoglycosidase cleavage and permethylation, that is, releasing

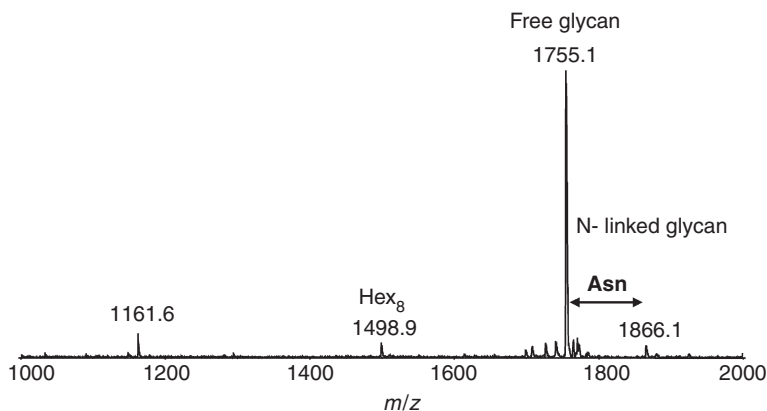


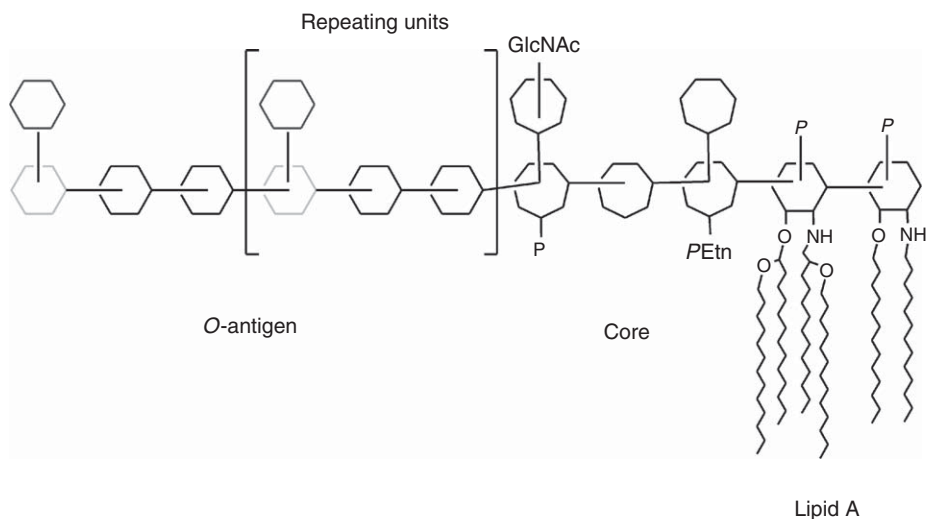
Figure 1 MALDI-TOF/TOF analysis of permethylated N-linked glycans from total protein extract of *C. jejuni* 11168H (1.0 mg). The protein extract was digested with pronase E for 24 h, and the digest was then purified using porous graphitic carbon cartridges followed by permethylation.

N-glycans using peptide-*N*-glycosidase F (PNGase F).¹ This technique is not applicable to prokaryotic cells because the carbohydrate-peptide linkages are different and the enzymes that release carbohydrate moieties are rarely available. This difficulty was solved using a universal method of isolating N-linked glycans from glycoproteins in complex bacterial and eukaryotic protein mixtures.¹⁴ This strategy relies on the MS of permethylated glycopeptides from proteins digested with pronase E and results in the production of Asn-linked glycans. This method enables the identification of the glycan structures of any glycoprotein. The strategy was validated using well-characterized eukaryotic standard glycoproteins such as ribonuclease B, ovalbumin, and avidin. The results indicate that glycans with a single amino acid were the major components when a higher ratio of pronase E to protein (2:1) and incubation time of 48 h were used. The Asn-glycans were further purified using porous graphitic carbon cartridges followed by permethylation. The MS data demonstrated that hydroxyl groups were methylated, while the amino group in asparagine underwent β -elimination. Both modifications were confirmed by the evidence of a mass increase of 111 Da in the molecular masses of permethylated Asn-glycans compared to the corresponding free oligosaccharides. This method was also applied to N-glycan analysis from total protein extracts of *C. jejuni* 11168H, as shown in [Figure 1](#).

In [Figure 1](#), ions at m/z 1755.0 and 1866.1 correspond to Glc1GalNAc5 Bac and Glc1GalNAc5 Bac-Asn, respectively. The ion at m/z 1498.9 is permethylated starch-like fragments. The MS/MS experiment for the ion at m/z 1866.1 confirmed the Glc1GalNAc5 Bac-Asn structure.¹⁴ In addition to the N-linked heptasaccharide in the whole cell extract of *C. jejuni*, an unexpected free heptasaccharide (m/z 1755.0) was detected, which was also found in the total protein extract without pronase E digestion. This confirmed that the free glycans were not artifacts due to the pronase E treatment or the permethylation process. Further studies revealed that the quantity of free glycans was associated with growth conditions. Hence, the new method enables not only N-linked glycan detection, but also the discovery of free glycans.

6.06.3 Lipopolysaccharide

LPS is the major component of the outer membrane of Gram-negative bacteria.^{15–22} These outer surface components are amphipathic molecules that consist of a hydrophilic oligosaccharide core linked to a hydrophobic glycolipid, referred to as lipid A. LPS from bacteria that colonize plates in a smooth form is called S-type. S-type LPS has a polysaccharide chain referred to as the *O*-antigen, which is composed of repeating units that are linked to the core region ([Scheme 1](#)). LPS from mucosal pathogens always lacks the *O*-polysaccharide (*O*-antigen). This kind of LPS is called R (rough)-type, and is often referred to as short-chain LPS or LOS. The polysaccharides have important structural and functional roles in the life of a bacterial cell. Probing the subtle changes in LPS structure as a result of phase variation or as a consequence of site-directed mutation requires



Scheme 1 Schematic structure of a lipopolysaccharide. Hexose, \square ; heptose, \heptagon ; GlcNAc, \heptagon with vertical line; PEtn, phosphoethanolamine; P, phosphate.

sensitive analytical tools.²³ Structural characterization of this diverse population of glycolipids is not only important for further understanding bacterial pathogenesis processes and biosynthetic mechanisms leading to their expression, but also for the development of antibodies specifically targeted toward immuno-determinant structures. Determining the composition and structures of complex carbohydrates is challenging because the natural heterogeneity of glycolipid distribution and the presence of functional groups appended to different sites result in a large number of related LPS isoforms.^{24–26}

Electrospray MS has played a key role in the characterization of these complex carbohydrates. However, the lack of on-line separation techniques often precludes the determination of sample heterogeneity. For example, isomeric glycoforms cannot be separated or identified by MS if they are not separated prior to MS identification. Capillary electrophoresis–mass spectrometry (CE–MS) has demonstrated superb capability for the determination of glycoform populations and the characterization of isomeric glycoforms.^{23–26} For example, the occurrence of isomeric LPS glycoforms differing by the location or presence of functional groups such as phosphate (P), phosphoethanolamine (PEtn), or pyrophosphoethanolamine (PPEtn) are reflected by corresponding changes in their electrophoretic mobilities.²⁵ The characterization of oligosaccharide sequence and branching information can be obtained using tandem MS.

6.06.3.1 R-Type LPS

Some bacteria, such as *Haemophilus influenzae* and *N. meningitidis*, express R-type LPS comprising a heterogeneous mixture of core oligosaccharides.^{27–29} LPS can be subjected either to O-deacylation to remove O-linked fatty acids from lipid A or to mild acid treatment to release the entire lipid A. These treatments increase the solubility for MS analysis and simplify the MS data for the elucidation of the structure of molecules.

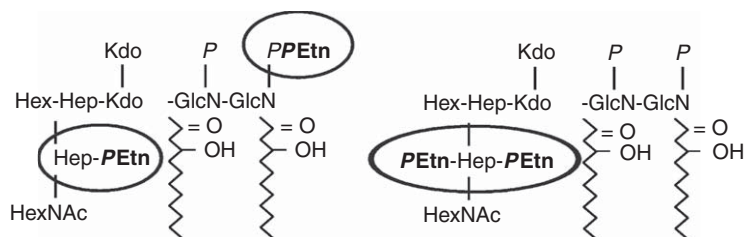
6.06.3.1.1 Analysis of O-deacylated LPS

CE–MS has been successfully used to characterize trace levels of O-deacylated LPS samples of *H. influenzae* and *N. meningitidis*.^{15,24–26} In those applications, anhydrous hydrazine treatment under mild conditions was employed to remove the O-linked fatty acids leaving only an N-linked fatty acid remaining attached to each glucosamine. Hydrazinolysis along with the enzyme treatment provides a method to determine sample impurity, and most importantly, improves the solubility and reduces the aggregation of LPS in aqueous

solutions. CE–MS provides highly sensitive analysis of O-deacylated LPS.²³ This hyphenated methodology facilitates the determination of closely related LPS glycoforms and isoforms by exploiting differences in their unique molecular conformations and ionic charge distributions by electrophoretic separation. Direct infusion of LPS samples into the mass spectrometer without any front-end separation generally gives very complex mass spectra, since LPS glycolipids often consist of a complex mixture of closely related glycoforms varying by the length and the sites of attachment of the oligosaccharide chains. When a CE system is coupled to the mass spectrometer, a contour profile of m/z vs time is obtained providing an additional dimension for the illumination of structural complexity.²³ For example, sialylated glycoforms were detected by direct analysis of *ex vivo* organisms from a middle ear infection, although the sialylated glycoforms were undetectable in the inoculum.¹⁵ CE–MS profiling of LPS-OH has been utilized for comparing the LPS glycoform populations of *H. influenzae* and *N. meningitidis* in different strains. In addition, CE–MS is also a powerful technique for separation and characterization of isomeric glycoforms.^{25,26} Using the sheath-solution interface, we could separate LPS in anionic forms and detect them with either negative or positive ion mode MS. The structural assignments obtained from tandem MS and CE–MS analyses have enabled the identification of isomeric glycoforms. For example, we have found that the variation in phosphoethanolamines resulted in different glycoforms with the same composition in both *H. influenzae* and *N. meningitidis*.^{25,26} The application of this technique to the analysis of LPS from the *galE* mutants of *N. meningitidis* strain BZ157 B5+ revealed the presence of isomeric glycoforms, in which the location of the functional group phosphoethanolamine was found to be either in the inner core or lipid A regions (**Scheme 2**).²⁵

The total ion electropherogram (TIE) and reconstructed ion electropherogram (RIE) at m/z 795.0 are presented in **Figure 2(a)**, together with the extracted mass spectrum (XIE) at 10.3 and 11.0 min (**Figures 2(b) and 2(c)**). From the XIE, two isoforms were well separated, one at 10.3 min and another at 11.0 min. The molecular masses extracted were derived as 2388.0 Da, having the composition of HexNAc₁ Hex₁ PEtn₂ Hep₂ Kdo₂ lipid A (2P). To characterize the structures of these two isomeric glycoforms, CE–MS/MS experiments were first conducted in the negative ion detection mode (data not shown). The tandem mass spectrum at 11.0 min gave the fragment ion at m/z 951.5, together with a doubly charged ion at m/z 475.0, thus indicating that the lipid A–OH had the expected molecular mass. In contrast, the extracted mass spectrum at 10.3 min produced fragment ions at m/z 1074.5 and 536.5 (data not shown). The difference in the molecular mass between the two lipid A–OH moieties implies an extra PEtn in the core region for the glycoform with a migration time of 10.3 min.

The electropherogram shown in **Figure 3(a)** corresponds to the CE–MS/MS analysis of doubly charged precursor ions at m/z 1195.0. The MS/MS spectrum (**Figure 3(a)**) showed a neutral loss of a lipid A at m/z 1436 (neutral moiety 952 Da) accompanied by consecutive losses of the Kdo residues at m/z 1216 and 996. Thus, the additional PEtn group could be assigned to the core oligosaccharide region of the molecule although its exact location was still unknown. Further structural details were obtained in an MS/MS experiment for the fragment ion at m/z 996, generated with a high orifice voltage (**Figure 3(b)**). The tandem mass spectrum was dominated by simple cleavage of each glycosidic bond enabling direct sequence assignment. The sequence consisting of HexNAc–(PEtn)Hep(PEtn)–Hep–Hex was confirmed by the observation of fragment ions at m/z 439, 631, and 793. Another series of fragment ions, m/z 396, 587, and 750, corresponding to a loss of the ethanolamine moiety (43 Da) also supported this assignment. Therefore, it was concluded that two PEtn residues were attached to the



Scheme 2 Structure of the core region from *Neisseria meningitidis* BZ157 B5+ showing the presence of isomeric glycoforms. Hex, hexose; Hep, heptose; Kdo, 3-deoxy-D-manno-octulosonic acid; GlcN, glucosamine; HexNAc, N-acetylhexosamine; PEtn, phosphoethanolamine; PPEtn, pyrophosphoethanolamine.

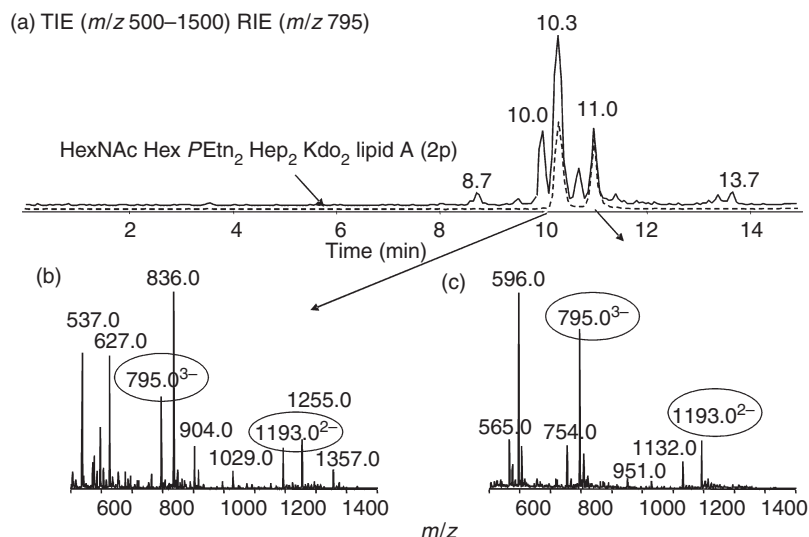


Figure 2 CE-MS analysis of O-deacylated LPS from *Neisseria meningitidis* strain BZ157 B5+. (a) TIE (m/z 500–1500) and RIE at m/z 795; (b) extracted mass spectrum at 10.3 min; (c) extracted mass spectrum at 11.0 min. Separation conditions: 10 nl injection of $100\ \mu\text{g ml}^{-1}$ of O-deacylated LPS, 5% methanol in $30\ \text{mmol l}^{-1}$ morpholine at pH 9.0 and +30 kV as separation voltage. From J. Li; A. D. Cox; D. Hood; E. R. Moxon; J. C. Richards, *Electrophoresis* **2004**, *25*, 2017. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

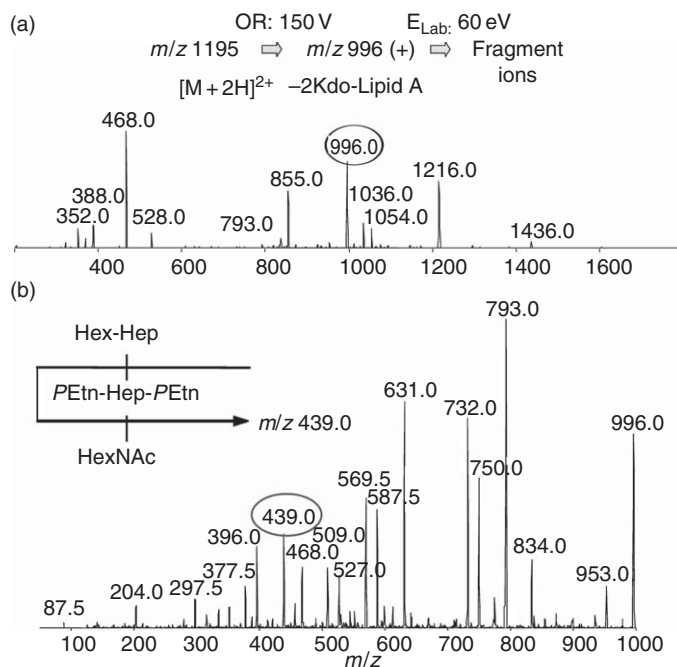


Figure 3 CE-MS/MS analysis of O-deacylated LPS from *Neisseria meningitidis* strain BZ157 B5+. (a) Extracted MS/MS spectrum of ion at m/z 1195. MS conditions: orifice voltage was set at 30 V, N_2 was used as collision gas and +100 V was set as collision energy; (b) extracted MS/MS/MS spectrum of ion at m/z 996. MS conditions: orifice was set at +180 V, N_2 was used as collision gas and +60 V was set as collision energy. Separation conditions: 10 nl injection of $100\ \mu\text{g ml}^{-1}$ of O-deacylated LPS, 5% methanol in $15\ \text{mmol l}^{-1}$ ammonium acetate at pH 9.0 and +20 kV. From J. Li; A. D. Cox; D. Hood; E. R. Moxon; J. C. Richards, *Electrophoresis* **2004**, *25*, 2017. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

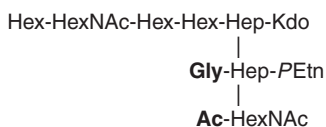
HepII residue of the LPS molecules.²⁵ This observation is in agreement with a previous nuclear magnetic resonance (NMR) study, which indicated that the HepII residue was substituted with *PEtn* moieties at both the 3- and 6-positions.³⁰

6.06.3.1.2 Core analysis

Compared to the highly variable *O*-antigens, the core region has limited variations. For example, there are five types of core structures in *E. coli*, designated K1, R1, R2, R3, and R4. Core OS is conceptually divided into the inner core and the outer core. Generally, core OS is not considered to be a virulent factor, but it has some indirect roles in the adhesion of certain bacteria.³¹ The inner core plays an essential role in the stability and integrity of the outer membrane, which may account for the conservation of the structure. The LPS of *N. meningitidis* has a conserved inner-core structure that contains two Kdos, two Heps, and one GlcNAc.²⁸ Additional heterogeneity can be found by the incorporation of noncarbohydrate substituents such as acetate (Ac) and glycine (Gly). Conventional analysis of bacterial LPS by CE-MS has generally employed the prior removal of *O*-acyl lipid chains, which is necessary for the effective solubilization and separation of the heterogeneous ensemble of LPS species. However, *O*-deacylation also causes the undesired removal of important glycan-associated *O*-linked modifications, such as *O*-acetate and *O*-linked amino acids. To overcome this problem, mild acid hydrolysis is normally employed as a complementary sample preparation procedure.^{32–34} Typically, the core oligosaccharides can be prepared with mild acid hydrolysis of the LPS (1.5% acetic acid, 100 °C, 2 h), followed by cooling and removal of the insoluble lipid A moiety by centrifugation. The carbohydrate-containing supernatant can later be lyophilized and analyzed by CE-MS.

Figure 4 shows the ESI-MS spectrum (positive ion detection mode) of oligosaccharides derived from *N. meningitidis* strain NGH15.³² The highlighted ions, m/z 820, 841, and 870, correspond to the compositions of Hex₃ HexNAc₂ Hep₂ *PEtn* anhydroKdo, Ac₁ Hex₃ HexNAc₂ Hep₂ *PEtn* anhydroKdo, and Ac₁ Gly₁ Hex₃ HexNAc₂ Hep₂ *PEtn* anhydroKdo, respectively. The mass difference of 17 Da for each glycoform corresponds to an ammonium adduct in positive detection mode. In addition, sodium adducts were also detected, that is, m/z 832. The extracted tandem mass spectra for the precursor ions at m/z 820, 841, and 870 are shown in **Figures 4(b)–4(d)**, respectively. As indicated in the figures, the structural information was limited to the consecutive losses of three Hex and one HexNAc residues. The positions of the acetates could not be obtained by MS/MS experiments. The location assignments of acetate and glycine were only obtained in a separate set of MS³ experiments in which the singly charged ions at m/z 949, 991, and 1049 were formed at a high orifice voltage and selected as precursors (**Figure 5**). The tandem MS spectra for the oligosaccharide gave the fragment ions at m/z 204, 316, 508, 711, and 949, corresponding to HexNAc, Hep-*PEtn*, Hep(*PEtn*)-HexNAc, and Hep(*PEtn*)-HexNAc-Kdo, respectively. For the spectrum of the core oligosaccharide containing one acetate residue, the fragment ion at m/z 204 disappeared and a new fragment ion was observed (m/z 246). A mass increment of 42 Da suggested the existence of an *O*-acetylated HexNAc. A similar result was obtained in the spectrum of ion at m/z 1049, in which a high abundant ion at m/z 246 was detected. Furthermore, the location of glycine could be obtained based on the detection of a fragment ion at m/z 373 and the disappearance of the ion at m/z 316 (their corresponding difference was 57 Da). From these fragment patterns, the structure was assigned as follows:

In general, information on the location of Gly and Ac can be obtained by tandem MS experiments in the positive ion mode. In the MS³ spectra, the occurrence of marker ions corresponding to specific residues, that is, m/z 246 vs m/z 204, and m/z 373 vs m/z 316, are of particular interest for the identification of the location of Ac and Gly (**Scheme 3**).



Scheme 3 Derived structure of core region from *Neisseria meningitidis* strain NGH15. Hex, hexose; HexNAc, *N*-acetylhexosamine; Hep, heptose; Kdo, 3-deoxy-*D*-manno-octulosonic acid; Gly, glycine; *PEtn*, phosphoethanolamine; Ac, acetic acid.

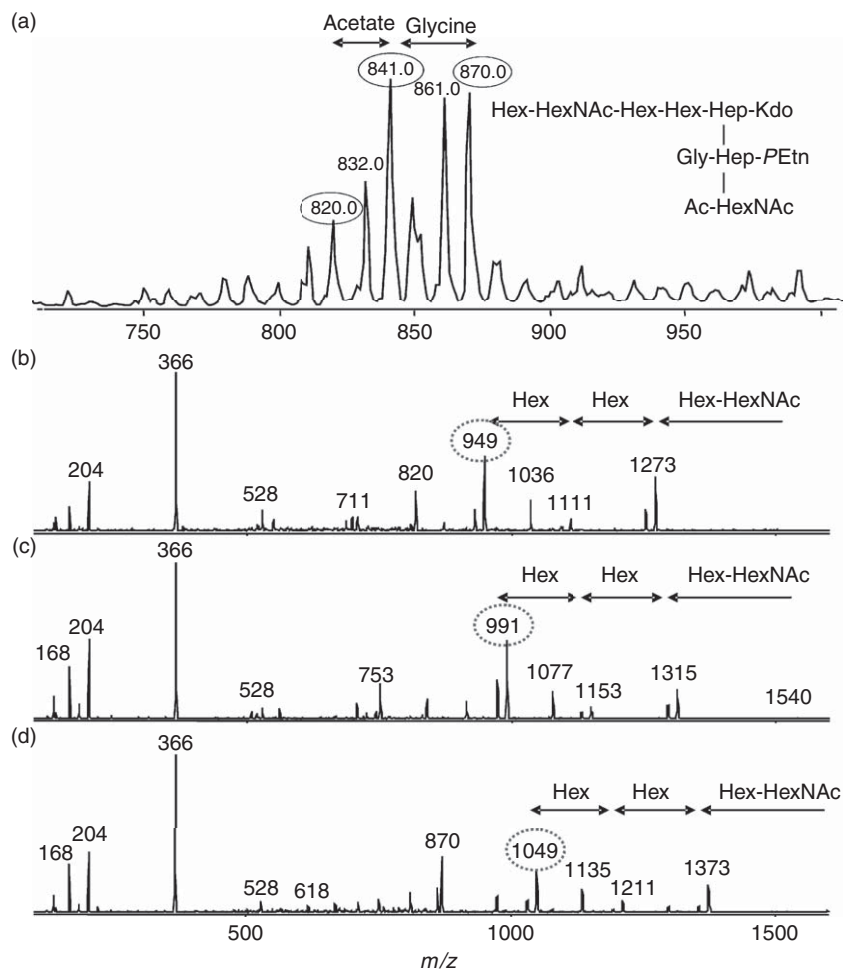


Figure 4 CE-MS analysis of core oligosaccharides from *Neisseria meningitidis* strain NGH15. (a) Extracted MS spectrum; (b) extracted MS/MS spectrum of ion at m/z 820 ($[M + H + NH_4]^{2+}$); (c) extracted MS/MS spectrum of ion at m/z 841 ($[M + H + NH_4]^{2+}$); (d) extracted MS/MS spectrum of ion at m/z 870 ($[M + H + NH_4]^{2+}$). The experiments were performed using an API 3000 mass spectrometer (Applied Biosystems/Sciex, Concord, Ontario, Canada) via a microspray interface coupling to a CE system. A sheath solution (isopropanol-methanol, 2:1) was delivered at a flow rate of $1 \mu\text{L min}^{-1}$. The orifice voltage was set at 30 V. Separations were obtained on approximately 90 cm length of bare fused-silica capillary using 15 mmol l^{-1} ammonium acetate in deionized water at pH 7.0.

6.06.3.1.3 Lipid A analysis

Lipid A is the primary immunostimulator of LPS and is responsible for most of its biological activities, including, the toxicity of LPS, and the activation of innate immune responses in a structure-dependent manner. It is a highly diverse molecule.²² Its endotoxic activity strongly depends on the type of hexosamine residues, the presence of phosphates and other negatively charged groups, and especially on the type, chain length, and location of fatty acids. We have previously characterized lipid A species using a combination of MS and mild acid hydrolysis.²⁰ This technique provides information about the heterogeneity of different species in the lipid A families and distribution of the fatty acids on each glucosamine unit. Recently, we developed a method for the analysis of intact LPS, from which the structural information can be obtained for both core oligosaccharides and lipid A moieties.^{35,36} Using in-source collision-induced dissociation in combination with open tubular liquid chromatography-mass spectrometry (OTLC-MS), intact LPS ions could be fragmented within the orifice-skimmer region of an electrospray mass spectrometer, and the generated lipid A fragments could be analyzed. **Figure 6(a)** shows the MS spectra of intact LPS from *H. influenzae* strain RM118. The ion at m/z 1824.9

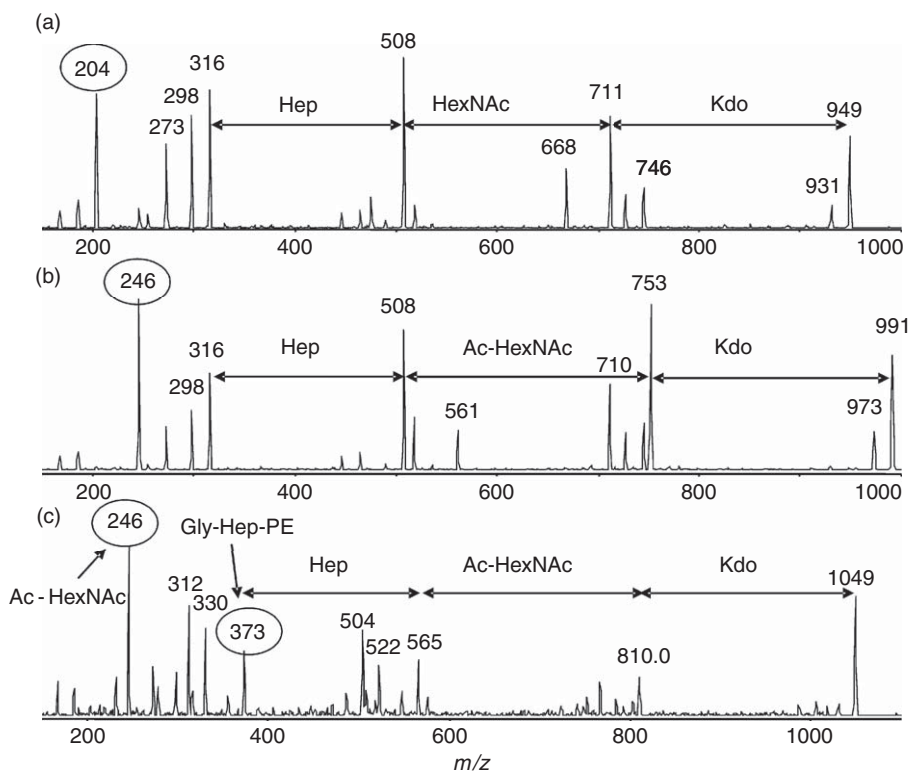


Figure 5 CE-MS analysis of core oligosaccharides from *Neisseria meningitidis* strain NGH15. In order to generate fragment ions at m/z 949, 991, and 1049, the orifice potential was increased from +30V to +180V. (a) Extracted MS/MS spectrum of ion at m/z 949; (b) extracted MS/MS spectrum of ion at m/z 991; (c) extracted MS/MS spectrum of ion at m/z 1049. Other conditions are the same as in **Figure 4**.

corresponded to a hexa-acyl lipid A structure, which comprised a β -GlcN-(1 \rightarrow 6)- α -GlcN with phosphate groups at C-1 and C-4' positions (inset). The N-2/N-2' and O-3/O-3' positions were substituted by amide-linked and ester-linked 3-hydroxytetradecanoic acid chains (14:0(3-OH)), respectively, while the hydroxyl group of 14:0(3-OH) fatty acid chains at N-2' and O-3' positions were further esterified by tetradecanoic acid chains (14:0). The ions at m/z 1596.6 and 1388.4 correspond to penta- and tetra-acylated diphosphorylated lipid A moieties, respectively.³⁶ The structural information of lipid A could be obtained from MS³ experiments on the ions at m/z 1824.9 (**Figure 6(b)**). Owing to the loss of an O-linked nonhydroxylated 14:0 or O-linked 14:0(3-OH) fatty acid, the parent ion produced fragment ions at m/z 1596.7 or 1580.9, respectively. The ion at m/z 1726.9 was formed by the loss of a phosphate group from the weak anomeric C-1 position. The fragment ions at m/z 1271.6 and 1253.6 were formed from the consecutive losses of a phosphate group from O-4' and two O-linked 14:0 fatty acids, substituting the 3-hydroxy group of 14:0(3-OH) at the 3' position and O-linked 3-hydroxylated 14:0(3-OH) at the O-3-position (**Scheme 4**, left panel). The ion at m/z 1045.6 corresponded to the loss of a phosphate group at C-1, O-linked 14:0[3-O(14:0)] from 3'-position and not from N-2' position due to the greater stability of acyl amide in the N-2' position, and unsubstituted 14:0(3-OH) from the O-3 position (**Scheme 4**, right panel). The origins of other fragment ions are outlined in **Scheme 4**.

6.06.3.2 S-Type LPS

LPS and CPS represent the first line of defense against complement and bacteriophages.^{19,22,37–41} In contrast to R-type LPS, S-type LPS has an O-antigen, which consists of up to 50 repeated units, attached to the core region. O-antigens contain the major antigenic determinants that distinguish various serotypes of bacteria, which may

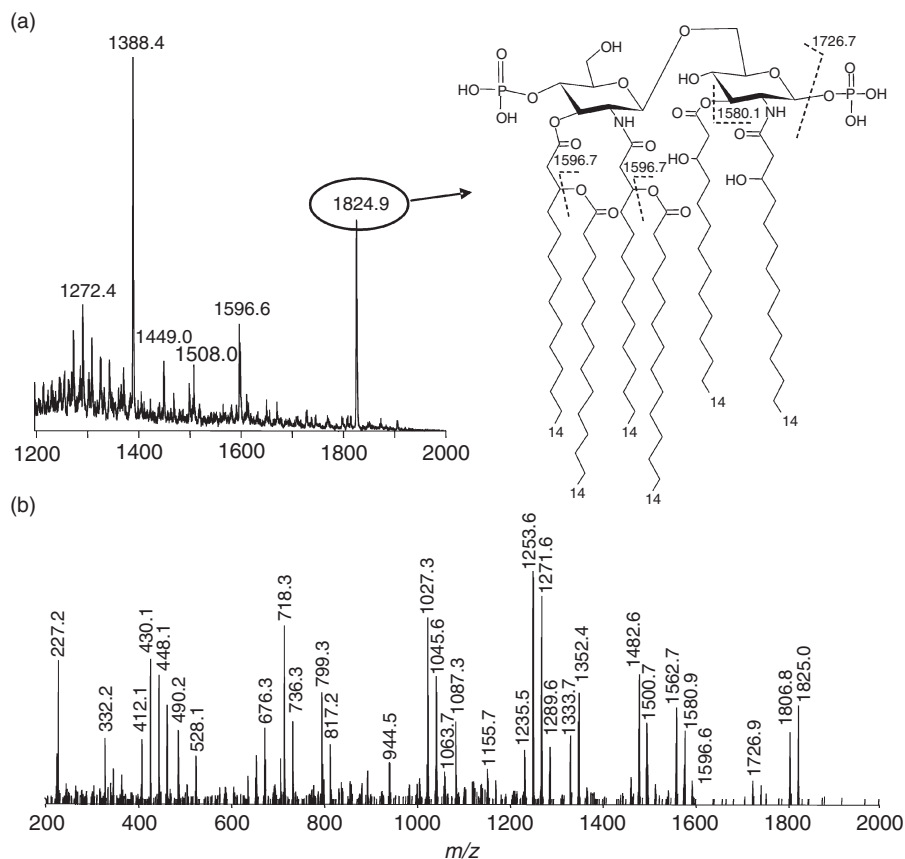
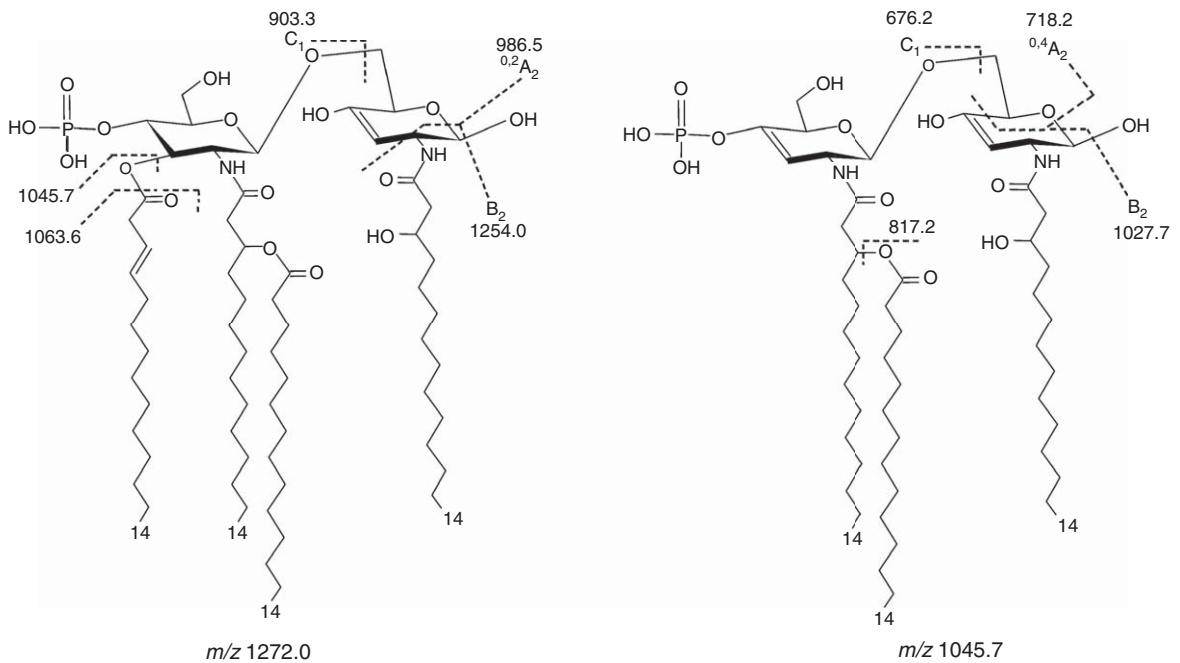


Figure 6 EA-OTLC-MS analysis of LPS from *Haemophilus influenzae* strain RM118 (negative ion mode), using a 4000 Q-Trap mass spectrometer (Applied Biosystems/Sciex, Concord, Ontario, Canada). (a) Extracted MS spectrum with high orifice potential of -300 V; (b) extracted MS/MS spectrum of ion at m/z 1824.9 ($[M-H]^-$). A separation voltage of 20 kV, together with a pressure of 500 mbar, was employed. A measure of $1.0 \mu\text{l}$ of samples were injected for all the experiments. For MS/MS analysis, N_2 was used as collision gas and -110 V was set as collision energy. From M. Dzieciatkowska; E. K. Schweda; E. R. Moxon; J. C. Richards; J. Li, *Electrophoresis* **2008**, 29, 2171. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

be related to human diseases.⁴² This highlights the need for analytical techniques for the microscale isolation, purification, and characterization of LPS to support research efforts in pathogenesis and vaccine development. ESI-MS is not able to generate multiply charged ions for detecting intact O-PS or CPS.⁴³ A recent study demonstrated that in-source collision-induced dissociation mass spectrometry (IS-CID-MS) could promote the formation of structurally relevant repeating units of heterogeneous O-PS and CPS; otherwise, these units would remain undetectable using conventional ESI conditions.⁴³ This approach has proved particularly useful for probing the subtle structural differences in monosaccharide composition and functionalities arising across bacterial serotypes.⁴¹

In this section, we mainly discuss the characterization methods for the O-antigen region in S-type LPS. Mild acid hydrolysis of the LPS expressed by *Salmonella riogrande* O:40 was examined using IS-CID-MS (Figure 7). The structure of the *S. riogrande* O:40 antigen was first determined from an analysis of the antigenic O-polysaccharide component of the LPS using ^1H - and ^{13}C -NMR spectroscopy, methylation analysis, and periodate degradation methods.⁴⁰ The O-polysaccharide was found to be a high-molecular weight branched polymer of repeating pentasaccharide units as shown in Scheme 5.

Under conventional MS experimental conditions, no ions associated with the O-chain were detected in the spectra. When the orifice voltage of the mass spectrometer was increased to $+400$ V, the characteristic fragment ions corresponding to the composition and sequence of the biopolymer were clearly detected. In the spectrum,



Scheme 4 Derived structures of lipid A from *Haemophilus influenza* strain RM118.

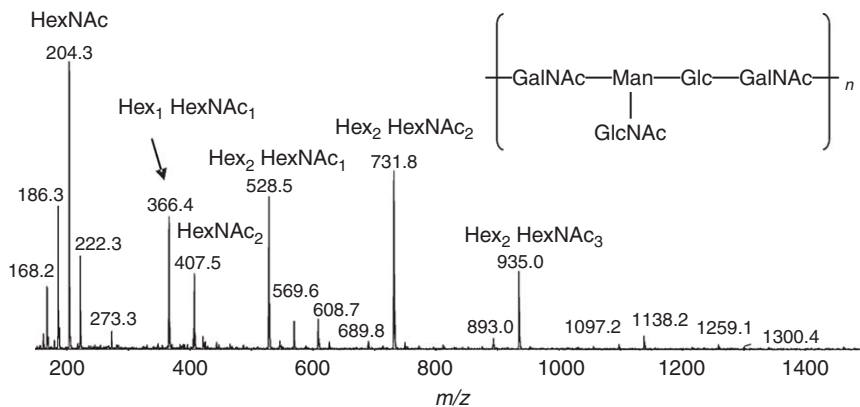
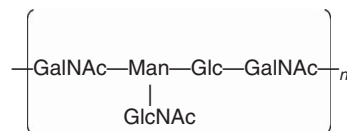


Figure 7 IS-CID-MS analysis of O-PS ($1.0 \mu\text{g} \mu\text{l}^{-1}$) from *Salmonella riogrande* O:40. The experiments were performed using a 4000 Q-Trap mass spectrometer (Applied Biosystems/Sciex, Concord, Ontario, Canada) via a CE-MS interface. A sheath solution (isopropanol-methanol, 2:1) was delivered at a flow rate of $1 \mu\text{l} \text{min}^{-1}$. The orifice voltage was set at +400 V. Separations were obtained on approximately 90 cm length of bare fused-silica capillary using 15 mmol l^{-1} ammonium acetate in deionized water at pH 7.0.



Scheme 5 Derived structure of O-polysaccharide from *Salmonella riogrande* O:40. GalNAc, *N*-acetylgalactosamine; Man, mannose; Glc, glucose; GlcNAc, *N*-acetylglucosamine.

the major peaks were identified and the corresponding sugar sequences were assigned. The peaks m/z 204.3, 366.4, 407.5, 528.5, 731.8, and 935.0 represented the sugar compositions of HexNAc, Hex₁ HexNAc₁, HexNAc₂, Hex₂ HexNAc₂, and Hex₂ HexNAc₃, respectively. These fragment ions could be used as fingerprints for bacterial serological classification.^{41,43}

6.06.4 Capsular Polysaccharides

Capsular polysaccharides are linked to the cell surface of bacteria through covalent attachment to phospholipids. Similar to the *O*-polysaccharides, the capsular polysaccharides also are composed of repeating units and can be either homo- or heteropolymers linked by glycosidic bonds. CPS molecules are known to play an important role in bacterial survival in the environment and often contribute to pathogenesis, for instance, in prevention of desiccation, adherence to each other or to the host cells, and resistance to host immunity. There is interest in CPS structures because of the possible applications in vaccine development. *Actinobacillus pleuropneumoniae* is a Gram-negative bacterium causing contagious pleuropneumonia in pigs. Strains of *A. pleuropneumoniae* have been assigned to 15 serotypes based on the unique structure of their corresponding antigenic CPS and LPS components. The IS-CID-MS technique was successfully used to probe the subtle structural differences in monosaccharide composition and functionalities arising from different serotypes.⁴³

As shown in Figure 8, the CPS from *A. pleuropneumoniae* serotypes 1 and 4 was composed of similar repeating units. For *A. pleuropneumoniae* serotype 1, major signals corresponding to one repeating unit (m/z 444.3, 486.3, and 504.3) or two repeating units (m/z 931.2, 949.3, 973.3, and 991.2) were observed. The

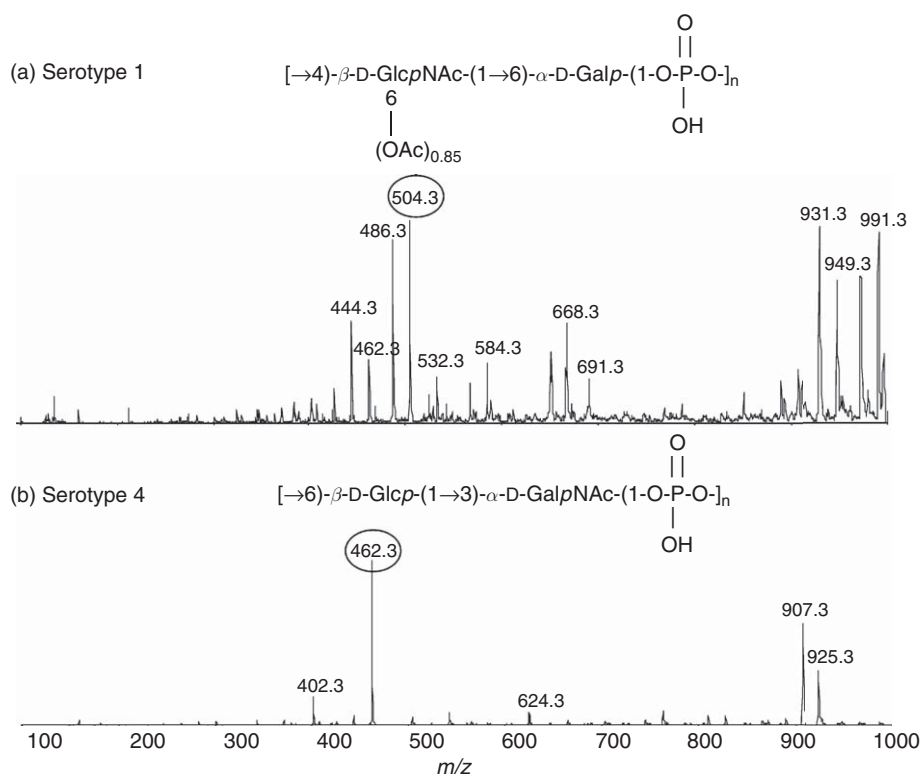


Figure 8 IS-CID-MS analysis of CPS ($1.0 \mu\text{g } \mu\text{l}^{-1}$) from *Actinobacillus pleuropneumoniae*. (a) Serotype 1; (b) serotype 4. The experiments were performed using an API 3000 mass spectrometer (Applied Biosystems/Sciex, Concord, Ontario, Canada) via a microspray interface coupling to a CE system. A sheath solution (isopropanol–methanol, 2:1) was delivered at a flow rate of $1 \mu\text{l min}^{-1}$. The orifice voltage was set at 200 V. Separations were obtained on approximately 90 cm length of bare fused-silica capillary using 15 mmol l^{-1} ammonium acetate in deionized water at pH 7.0.

existence of ions with m/z 486.3 and m/z 504.3 (difference of 18 Da) suggested that the in-source fragmentation produced intact repeating units, with an Ac group attached (**Figure 8(a)**). The relative intensity ratios of ions at m/z 444.3 and m/z 462.3 over ions at m/z 486.3 and m/z 504.3 suggested that at least more than half of the GlcNAc residues were O-acetylated. This assumption was further supported by the observation of minor intensity signals for ions at m/z 889.3 or m/z 907.3 corresponding to two O-deacetylated repeating units. The IS-CID-MS results clearly indicated the difference between CPS from serotype 1 and serotype 4. The ions at m/z 486.3 and m/z 504.3, corresponding to the O-acetylated repeating unit of CPS of serotype 1, were totally absent in the mass spectrum corresponding to the repeating unit of CPS of serotype 4 (**Figure 8(b)**). The ions at m/z 462.3 and m/z 907.3 corresponded to one repeating unit and two repeating units, respectively.

6.06.5 Conclusions

Bacterial surface glycoconjugates control a variety of biological events, including cell differentiation, cell adhesion, cell recognition, and immunological recognition. The heterogeneity and diversity of bacterial glycans present a challenge in the elucidation of their functions and the association of their structures with their functions. Characterizing glycans requires highly sensitive analytical techniques that are able to handle small amounts of samples and that can provide sufficient structural information for identification of the molecules. MS techniques have demonstrated their capability for sensitive bacterial glycome analysis with tandem mass spectra providing the compositional information about glycoconjugates, such as LPS, CPS, and glycoproteins. Recent advances have enabled studies of both the structures and the functions of bacterial glycomes in a systematic manner. However, the applications of MS techniques are limited to composition and sequence analysis. They cannot provide complete structural information about carbohydrates, such as linkage and conformation. Thus, integration of MS techniques with other methods and technologies, such as enzymatic methods and NMR, will become essential to provide further structural information.

Acknowledgments

We acknowledge the contributions of Drs. J. C. Richards, E. R. Moxon, E. K. H. Schweda, D. W. Hood, and E. Altman, which have been instrumental and have led to the applications and development of the methods described herein. We thank Andee Gravel for obtaining the MS spectrum (**Figure 8**). We thank M. Perry for providing O-PS of *S. riogrande* O:40. The authors also acknowledge funding support from the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Water Network, and Alberta Health and Wellness.

Abbreviations

| | |
|--------------|---|
| CE-MS | capillary electrophoresis–mass spectrometry |
| CID | collision-induced dissociation |
| CPS | capsular polysaccharide |
| EOF | electroosmotic flow |
| LOS | lipooligosaccharide |
| LPS | lipopolysaccharide |
| MS | mass spectrometry |
| MS/MS | tandem mass spectrometry |
| O-PS | O-linked polysaccharide |
| OS | oligosaccharide |
| P | phosphate |

| | |
|------------|------------------------------------|
| RIE | reconstructed ion electropherogram |
| TIE | total ion electropherogram |
| XIE | extracted mass spectrum |

Nomenclature

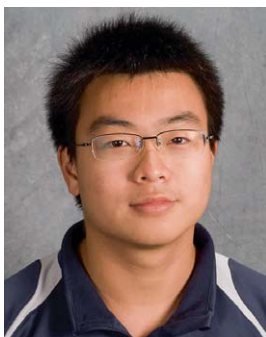
| | |
|---------------|-------------------------------------|
| Gal | galactose |
| GalNAc | <i>N</i> -acetylgalactosamine |
| Glc | glucose |
| GlcN | 2-amino-2-deoxy-D-glucose |
| Gly | glycine |
| Hep | L-glycero-D-manno-heptose |
| Hex | hexose |
| Kdo | 3-deoxy-D-manno-oct-2-ulosonic acid |
| Neu5Ac | <i>N</i> -acetylneuraminic acid |
| OAc | <i>O</i> -acetyl |
| PEtn | phosphoethanolamine |
| PPEtn | pyrophosphoethanolamine |

References

1. A. Dell; H. R. Morris, *Science* **2001**, *291*, 2351.
2. M. Wacker; D. Linton; P. G. Hitchen; M. Nita-Lazar; S. M. Haslam; S. J. North; M. Panico; H. R. Morris; A. Dell; B. W. Wren; M. Aebi, *Science* **2002**, *298*, 1790.
3. N. M. Young; J.-R. Brisson; J. Kelly; D. C. Watson; L. Tessier; P. H. Lanthier; H. C. Jarrell; N. Cadotte; F. St. Michael; E. Aberg; C. M. Szymanski, *J. Biol. Chem.* **2002**, *277*, 42530.
4. C. M. Szymanski; B. W. Wren, *Nat. Rev. Microbiol.* **2005**, *3*, 225.
5. C. M. Szymanski; S. M. Logan; D. Linton; B. W. Wren, *Trends Microbiol.* **2003**, *11*, 233.
6. F. F. Chevance; K. T. Hughes, *Nat. Rev. Microbiol.* **2008**, *6*, 455.
7. S. M. Logan, *Microbiology* **2006**, *152*, 1249.
8. M. Schirm; E. C. Soo; A. J. Aubry; J. Austin; P. Thibault; S. M. Logan, *Mol. Microbiol.* **2003**, *48*, 1579.
9. M. Schirm; M. Kalmokoff; A. Aubry; P. Thibault; M. Sandoz; S. M. Logan, *J. Bacteriol.* **2004**, *186*, 6721.
10. M. Schirm; S. K. Arora; A. Verma; E. Vinogradov; P. Thibault; R. Ramphal; S. M. Logan, *J. Bacteriol.* **2004**, *186*, 2523.
11. M. Schirm; I. C. Schoenhofen; S. M. Logan; K. C. Waldron; P. Thibault, *Anal. Chem.* **2005**, *77*, 7774.
12. P. Thibault; S. M. Logan; J. F. Kelly; J.-R. Brisson; C. P. Ewing; T. J. Trust; P. Guerry, *J. Biol. Chem.* **2001**, *276*, 34862.
13. P. Guerry; C. P. Ewing; M. Schirm; M. Lorenzo; J. Kelly; D. Pattarini; G. Majam; P. Thibault; S. Logan, *Mol. Microbiol.* **2006**, *60*, 299.
14. X. Liu; D. J. McNally; H. Nothhaft; C. M. Szymanski; J.-R. Brisson; J. Li, *Anal. Chem.* **2006**, *78*, 6081.
15. V. Bouchet; D. W. Hood; J. Li; J.-R. Brisson; G. A. Randle; A. Martin; Z. Li; R. Goldstein; E. K. Schweda; S. I. Pelton; J. C. Richards; E. R. Moxon, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8898.
16. N. J. Phillips; B. Schilling; M. K. McLendon; M. A. Apicella; B. W. Gibson, *Infect. Immun.* **2004**, *72*, 5340.
17. J. S. Plested; K. Makepeace; M. P. Jennings; M. A. Gidney; S. Lacelle; J.-R. Brisson; A. D. Cox; A. Martin; A. G. Bird; C. M. Tang; F. M. Mackinnon; J. C. Richards; E. R. Moxon, *Infect. Immun.* **1999**, *67*, 5417.
18. E. Pupo; N. J. Phillips; B. W. Gibson; M. A. Apicella; E. Hardy, *Electrophoresis* **2004**, *25*, 2156.
19. Z. Wang; E. Vinogradov; S. Larocque; B. A. Harrison; J. Li; E. Altman, *Carbohydr. Res.* **2005**, *340*, 693.
20. Z. Wang; J. Li; E. Altman, *Carbohydr. Res.* **2006**, *341*, 2816.
21. W. Yi; L. Zhu; H. Guo; M. Li; J. Li; P. G. Wang, *Carbohydr. Res.* **2006**, *341*, 2254.
22. S. I. Miller; R. K. Ernst; M. W. Bader, *Nat. Rev. Microbiol.* **2005**, *3*, 36.
23. J. Li; J. C. Richards, *Mass Spectrom. Rev.* **2007**, *26*, 35.
24. J. Li; R. W. Purves; J. C. Richards, *Anal. Chem.* **2004**, *76*, 4676.
25. J. Li; A. D. Cox; D. Hood; E. R. Moxon; J. C. Richards, *Electrophoresis* **2004**, *25*, 2017.
26. J. Li; A. D. Cox; D. W. Hood; E. K. Schweda; E. R. Moxon; J. C. Richards, *Mol. Biosyst.* **2005**, *1*, 46.
27. D. W. Hood; J. C. Richards; E. R. Moxon, *Biochem. Soc. Trans.* **1999**, *27*, 493.

28. J. S. Plested; K. Makepeace; M. P. Jennings; M. A. Gidney; S. Lacelle; J.-R. Brisson; A. D. Cox; A. Martin; A. G. Bird; C. M. Tang; F. M. Mackinnon; J. C. Richards; E. R. Moxon, *Infect. Immun.* **1999**, *67*, 5417.
29. E. R. Moxon; B. E. Gewurz; J. C. Richards; T. Inzana; M. P. Jennings; D. W. Hood, *Mol. Microbiol.* **1996**, *19*, 1149.
30. A. D. Cox; J. Li; J.-R. Brisson; E. R. Moxon; J. C. Richards, *Carbohydr. Res.* **2002**, *337*, 1435.
31. M. Jacques, *Trends Microbiol.* **1996**, *4*, 408.
32. A. D. Cox; J. Li; J. C. Richards, *Eur. J. Biochem.* **2002**, *269*, 4169.
33. J. Li; S. H. Bauer; M. Mansson; E. R. Moxon; J. C. Richards; E. K. Schweda, *Glycobiology* **2001**, *11*, 1009.
34. H. H. Yildirim; J. Li; J. C. Richards; D. W. Hood; E. R. Moxon; E. K. Schweda, *Carbohydr. Res.* **2005**, *340*, 2598.
35. M. Dzieciatkowska; D. Brochu; A. Van Belkum; A. P. Heikema; N. Yuki; R. S. Houlston; J. C. Richards; M. Gilbert; J. Li, *Biochemistry* **2007**, *46*, 14704.
36. M. Dzieciatkowska; E. K. Schweda; E. R. Moxon; J. C. Richards; J. Li, *Electrophoresis* **2008**, *29*, 2171.
37. D. J. McNally; H. C. Jarrell; J. Li; N. H. Khieu; E. Vinogradov; C. M. Szymanski; J.-R. Brisson, *FEBS J.* **2005**, *272*, 4407.
38. D. J. McNally; H. C. Jarrell; N. H. Khieu; J. Li; E. Vinogradov; D. M. Whitfield; C. M. Szymanski; J.-R. Brisson, *FEBS J.* **2006**, *273*, 3975.
39. D. J. McNally; M. P. Lamoureux; A. V. Karlyshev; L. M. Fiori; J. Li; G. Thacker; R. A. Coleman; N. H. Khieu; B. W. Wren; J.-R. Brisson; H. C. Jarrell; C. M. Szymanski, *J. Biol. Chem.* **2007**, *282*, 28566.
40. M. B. Perry; L. L. Maclean, *Carbohydr. Res.* **1992**, *232*, 143.
41. Z. Wang; X. Liu; A. Dacanay; B. A. Harrison; M. Fast; D. J. Colquhoun; V. Lund; L. L. Brown; J. Li; E. Altman, *Fish. Shellfish Immunol.* **2007**, *23*, 1095.
42. A. Kadioglu; J. N. Weiser; J. C. Paton; P. W. Andrew, *Nat. Rev. Microbiol.* **2008**, *6*, 288.
43. J. Li; Z. Wang; E. Altman, *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1305.

Biographical Sketches



Chuan Wang obtained his B.Sc. degree in biological sciences in 2007 from Wuhan University, China. He is currently a Ph.D. student at the University of Alberta, under the supervision of Dr. Xing-Fang Li. His research focuses on the development of new bioanalytical techniques for detection and characterization of viable but nonculturable (VBNC) pathogenic bacteria, particularly with regard to changes in gene expression and cell surface molecules associated with the VBNC state.



Camille L. A. Hamula obtained her B.Sc. degree in microbiology in 2003 from the University of Alberta and is currently pursuing her Ph.D. in medical sciences. Co-supervised by Dr. Chris Le and Dr. Xing-Fang Li and supported by scholarships from the Natural Sciences and Engineering Council of Canada and the Alberta Heritage Foundation for Medical Research, she focuses her research on the selection and applications of novel DNA aptamers for bacterial cells.



Xing-Fang Li is an Associate Professor in the Division of Analytical and Environmental Toxicology, Department of Laboratory Medicine and Pathology, University of Alberta. She obtained her B.Sc. in organic chemistry from Hangzhou University (1983), an M.Sc. in environmental chemistry from the Chinese Academy of Sciences (1986), a second M.Sc. in analytical chemistry from Brock University (1989), and a Ph.D. in environmental/analytical chemistry from the University of British Columbia (1994). She pursued NSERC postdoctoral research in bioanalytical chemistry at the University of Alberta and worked as a research scientist for MDS Sciex before returning to academia in 2001. Her current research interests include: (1) development of separation and MS techniques and applications to clinical and environmental studies; (2) microarray and affinity techniques for detection of viable microbial pathogens; and (3) exposure assessment and toxicology of new disinfection by-products in drinking water.



Dr. Xin Liu received his Ph.D. in analytical chemistry from Wuhan University (2002), China, and thereafter worked for Professor Lu at the College of Life Science as a postdoctoral fellow, innovating measurement approaches for plant hormone regulation. In 2004, he continued his postdoctoral research at York University, Canada, focusing on biomolecular interactions. He worked as an NSERC visiting fellow at the Institute for Biological Sciences, National Research Council of Canada in 2005–2008. His research interests focus on MS-based proteomics and glycomics.



Jianjun Li is a Senior Research Officer and the Head of the Glycoanalysis-MS Facility in the Glycobiology program at the Institute for Biological Sciences, National Research Council of Canada (NRC-IBS). He received a B.Sc. in chemistry and a Ph.D. in analytical chemistry from Wuhan University, China. Dr. Li's research has involved the development of high-resolution separation techniques coupled with MS and their application to the structural characterization of various biological molecules. From 1997 to 2001, he focused on the development of chip-based capillary electrophoresis–mass spectrometry (Chip-CE/MS) systems. He has successfully integrated different functional materials, such as C₁₈, MAb, and IMAC beads onto a chip channel to perform preconcentration chip-CE/MS with applications to functional proteomics. Since 2001, Dr. Li's research projects extended to glycomics. He has authored and co-authored over 130 original peer-reviewed scientific papers and three book chapters.

6.07 Chemical Glycobiology

Chad M. Whitman and Michelle R. Bond, Stanford University, Stanford, CA, USA

Jennifer J. Kohler, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|-------------------|---|-----|
| 6.07.1 | Introduction | 175 |
| 6.07.2 | Synthetic Glycans with Biological Activity | 176 |
| 6.07.2.1 | Role of Glycan Epitopes in Mediating Cell–Cell Interactions | 176 |
| 6.07.2.2 | Glycosaminoglycans | 177 |
| 6.07.2.3 | Glycolipids/Gangliosides | 181 |
| 6.07.2.4 | Carbohydrate Roles in Bacteria and Mycobacteria | 183 |
| 6.07.2.5 | Toxins | 187 |
| 6.07.3 | Chemoenzymatic Synthesis to Study Glycan Structure and Function | 187 |
| 6.07.3.1 | Alternative Nucleotide Sugar Donors | 189 |
| 6.07.3.2 | Chemoenzymatic Synthesis of Sialic Acids | 193 |
| 6.07.4 | Applications of Metabolic Oligosaccharide Engineering | 197 |
| 6.07.4.1 | Metabolic Flux of Unnatural Sugar Analogs and Effects on Cell Viability | 197 |
| 6.07.4.2 | Polysialic Acid Function | 200 |
| 6.07.4.3 | Metabolically Engineered Gangliosides for Cancer Immunotherapy | 202 |
| 6.07.5 | Covalently Trapping Glycan Interactions with Photocross-linkers | 204 |
| 6.07.5.1 | Sugar Probes for Lectin Characterization and Discovery | 205 |
| 6.07.5.2 | Cross-linking Glycolipids | 210 |
| 6.07.5.3 | Metabolic Incorporation of Cross-linking Sugars | 214 |
| References | | 218 |

6.07.1 Introduction

Carbohydrates are fundamental molecules found in all forms of life. Glucose (Glc) and its stored forms, glycogen and starch, are important energy sources. Pentoses are an essential component of nucleic acids. Intracellular sugars, either as free monosaccharides or attached to proteins or other metabolites, are key signaling molecules. Extracellular glycoconjugates are arbiters of cell–cell communication in multicellular organisms, where glycan-mediated interactions are essential to normal processes, such as development, and to disease states, such as metastatic cancers. Glycoconjugates are also an important component of prokaryotes, where they function in cell wall assembly and in extracellular dialogs with environmental cues.

Chemists' ability to synthesize naturally occurring molecules and their analogs is important to elucidating and manipulating the biological functions of these molecules. This is especially true for carbohydrates, where the heterogeneity of naturally occurring molecules makes it extremely challenging to assign functions to particular structures. The study of carbohydrates has a long history of being at the intersection of organic chemistry and biology. Early on, Lavoisier's experiments in chemical analysis revealed the elemental composition of sugar, and Emil Fisher's studies established the stereochemistry of these molecules. Recent years have seen great improvements in methodologies for synthesis of monosaccharide building blocks and their incorporation into glycoconjugates, including glycoproteins, glycolipids, and natural products. Access to these molecules enables chemical glycobiology research, in which the products of organic synthesis can be used to interrogate and control living systems.

This chapter focuses on the use of synthetically prepared molecules to study the function of glycans in the biological world. Since the field of chemical glycobiology is immense, we have chosen to highlight topics of recent interest rather than providing a comprehensive review. First, we present examples where syntheses of

homogeneous forms of naturally occurring glycans have been used to understand their biological roles. Next, we describe recent advances in chemoenzymatic synthesis that have expanded the library of monosaccharide building blocks. From there, the discussion turns to applications of glycan metabolic engineering that have expanded our knowledge of glycan function and metabolism. Finally, we narrate the utility of photoreactive functional groups in the capture and characterization of glycan-mediated interactions. The examples described here demonstrate the power and potential of chemical methods as they are brought to bear on the biology of glycosylation.

6.07.2 Synthetic Glycans with Biological Activity

Complex carbohydrates remain a daunting class of targets for synthetic chemistry.¹ Part of this challenge arises from the fact that each sugar molecule contains several hydroxyl groups of similar reactivity. The construction of oligosaccharides, particularly branched structures, necessitates the use of complex protecting group schemes that facilitate hydroxyl group differentiation. Along with hydroxyl group differentiation, another major hurdle is the formation of glycosidic bonds in a stereospecific, high-yielding fashion. Notwithstanding these obstacles, syntheses of quite complex, naturally occurring glycans and their analogs have been reported in recent years. Access to homogeneous, chemically defined carbohydrates enables experiments aimed at deconvoluting the biological roles of this class of molecules. In this section, we describe the biological utility of synthetically prepared glycans. The molecules discussed here include cell adhesion molecules, glycosaminoglycans (GAGs), glycolipids, and bacterial cell wall components.

6.07.2.1 Role of Glycan Epitopes in Mediating Cell–Cell Interactions

The ability of a cell to exchange information with its surroundings is dependent on the complex and heterogeneous array of carbohydrates that decorates the cell surface. Interactions between glycans and their protein ligands are essential in both normal and pathogenic states. For example, carbohydrates bearing 6-sulfated sialyl Lewis X (sLe^X) epitopes participate in a high-affinity interaction with the adhesion molecule L-selectin, a protein displayed on the surface of leukocytes.² The interaction between sLe^X carbohydrates and L-selectin is a key element of the inflammatory response, mediating attachment of bloodstream leukocytes to endothelial cells at the site of inflammation. Furthermore, interactions of sLe^X oligosaccharides with L-selectin and other cell surface receptors direct metastasis of lymphomas and are involved in the regulation of tumor growth. Due to their significant, yet incompletely understood, roles in cancer and the immune response, sLe^X and other glycan epitopes that regulate cell–cell adhesion are important targets for biological study and eventual therapeutic use. This section will highlight examples in which synthetically prepared glycans have been used to investigate cellular adhesion.

The Lewis blood group antigens are glycan structures commonly found on mammalian cell surfaces. They include A, B, H, Lewis X (Le^X), and Lewis Y (Le^Y) (**Figure 1**) antigens. Tumor cells often overexpress or inappropriately express one or more of these glycans. For this reason, Lewis blood group antigens are important synthetic targets that could be useful for elucidating mechanisms of biological processes and in cancer immunotherapy. Danishefsky *et al.*³ have synthesized immunogenic forms of cancer antigens, such as a keyhole limpet hemocyanin (KLH)-conjugated Le^Y. KLH is a complex, high-molecular-weight carrier protein commonly used in the production of antibodies due to its ability to elicit highly specific immunogenic responses to attached antigens. In a phase 1 clinical trial, ovarian cancer patients who had been immunized with KLH-Le^Y developed measurable titers of anti-Le^Y antibodies. However, these antibodies were not effective at recognizing clustered Le^Y, as it is normally displayed on mucins.⁴ To improve the likelihood of eliciting a productive immune response, the researchers synthesized polyvalent glycans, which array several Le^Y motifs on a single molecule.^{5–7} Immunotherapies have also been pursued with other glycan epitopes, such as the globo-H (globohexaosylceramide), Tn (α -GalNAc O-linked to Ser/Thr), and TF (Thomsen–Friedenreich) antigen. Tumor cells often display a variety of glycan antigens, for example, prostate cancer cells typically express globo-H, sTn, Tn, TF, and Le^Y antigens. To elicit more specific immune responses, Danishefsky and coworkers prepared vaccines consisting of multivalent displays of antigens on a single peptide backbone conjugated to KLH- or a Pam₃Cys (tripalmitoyl-S-glyceryl cysteine)-scaffold (**Figure 2**). Reported molecules include

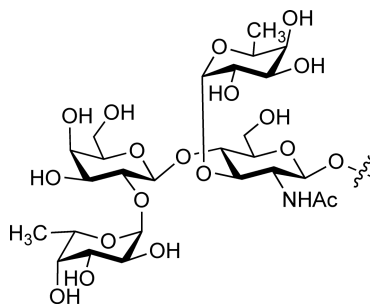


Figure 1 The Le^Y antigen.

trivalent, pentavalent, and hexavalent structures that display globo-H, Le^Y, GM2, sTn, TF, and Tn antigens.^{8–11} These polyvalent vaccines are immunogenic in mice, leading to increased IgM and IgG titers, and the results of these studies support carrying these types of molecules forward into human clinical trials. In addition, facile, large-scale syntheses of these glycans are becoming more accessible through the advent of automated methods.¹² Both solution-phase and solid-phase synthetic approaches have been used to synthesize many of the blood group antigens.^{13–16}

Another potential strategy to interfere with carbohydrate-mediated cell adhesion events is the use of glycan mimetics.^{17,18} Adopting this approach, Wong and coworkers synthesized sLe^X mimics that bind selectins. These glycan–peptide hybrids are designed to display the critical functional groups from sLe^X that are responsible for selectin recognition (**Figure 3**).^{19–21} The most potent inhibitor was found to bind selectins 10⁴-fold better than the naturally occurring glycan.^{22,23} Other inhibitors take advantage of multivalent display of glycan epitopes to create high-affinity ligands. Applying this strategy, Kiessling and coworkers synthesized multivalent glycan mimics that inhibited L-selectin-mediated leukocyte rolling, an essential step in the inflammatory response. They used ring-opening metathesis polymerization (ROMP) to produce a multivalent display of sulfated galactose (Gal) and Le^X analogs (**Figure 4**).^{24–30} The most effective of these inhibitors utilized a disulfated trisaccharide mimic of Le^X^{25,28,30} that inhibited leukocyte rolling more efficiently than sLe^X or 6-sulfo-sLe^X.³⁰ Other groups have made use of scaffolds such as polymerized liposomes³¹ and dendrimer-like glycopolymers³² to produce multivalent molecules with potential therapeutic properties.

Glycan-mediated interactions can also be inhibited by preventing production of the relevant glycan. Toward this goal, Esko and coworkers made use of disaccharide-aglycon primers that interfere with the synthesis of cells surface sLe^X.^{33–35} These molecules act as metabolic decoys because glycosyltransferases are occupied with glycosylating the primer and fail to modify their normal substrates. A variety of disaccharide primers were found to be effective inhibitors of sLe^X production.^{36,37} A peracylated version of GlcNAc(*N*-acetylglucosamine) β 1,3Gal *O*-linked to a naphthalenemethanol aglycone acts as a prodrug: once delivered to LS180 human colon carcinoma cells, the deacylated metabolite inhibited sLe^X expression and reduced the carcinoma cells' binding to thrombin-activated platelets and TNF- α activated endothelial cells (**Figure 5**).³⁸ This primer increased tumor cell susceptibility to leukocyte-mediated lysis and also reduced metastasis to the lungs from a subcutaneous tumor in a mouse model (see Chapter 6.04).^{38,39}

6.07.2.2 Glycosaminoglycans

GAGs are linear polysaccharide chains composed of repeating hexosamine disaccharide units. Commonly occurring GAGs include heparin, heparan sulfate (HS), chondroitin sulfate (CS), and dermatan sulfate (DS). GAGs play essential roles in neuronal growth, cancer, and angiogenesis. The biological activity of this diverse group of molecules is likely to derive from the array of functional groups that they present, the three-dimensional structure they adopt, and their mechanical properties. This section will highlight examples where synthetic GAG analogs are used to probe the biological function of these molecules. Numerous reviews cover this topic more comprehensively.^{40,41}

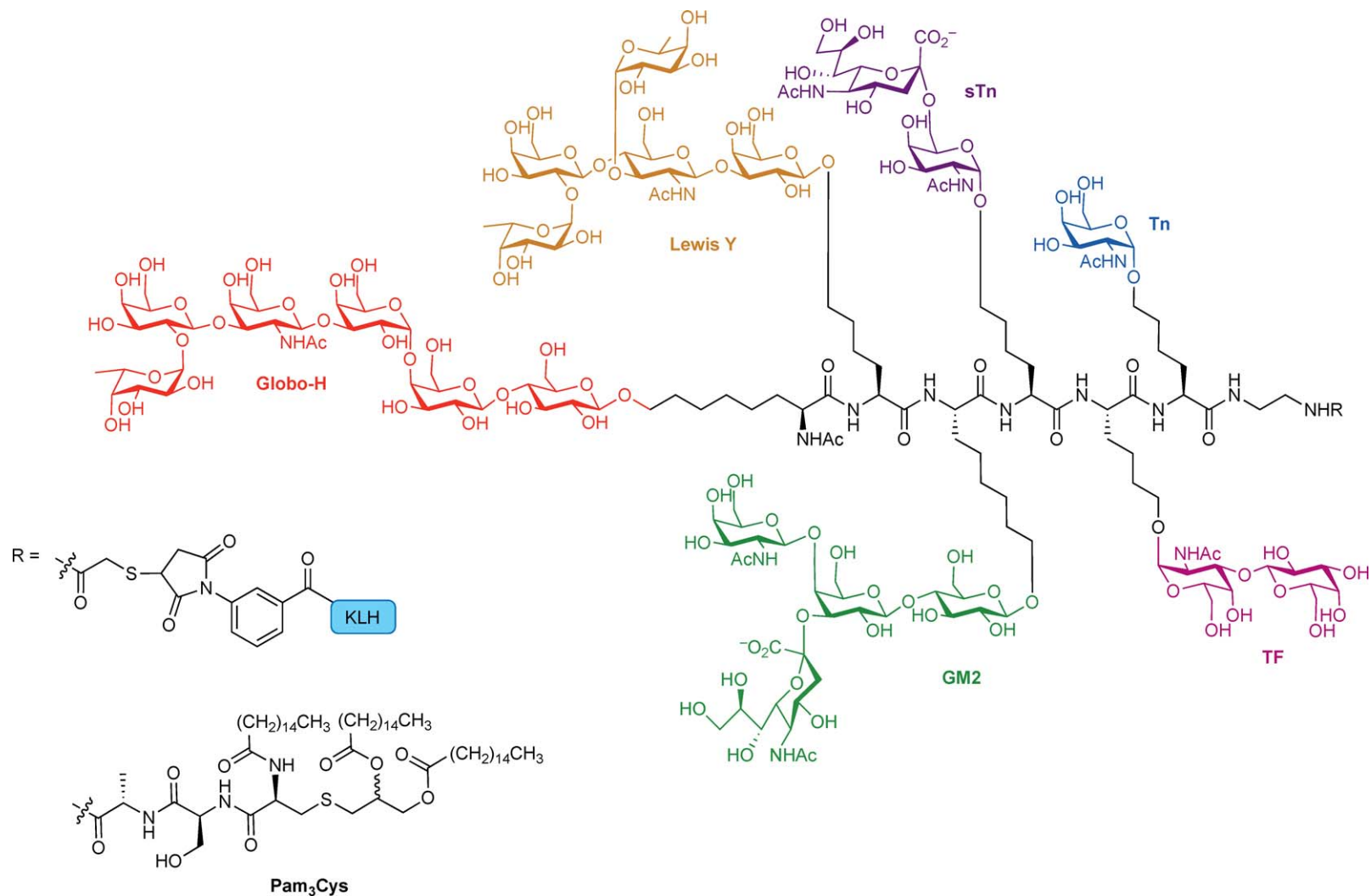


Figure 2 Hexavalent vaccine displaying several common cell surface antigens. The antigens are displayed on a single-peptide backbone and conjugated to two different scaffolds, KLH and Pam₃Cys (noted as R groups). Reprinted with permission from G. Ragupathi; F. Koide; P. O. Livingston; Y. S. Cho; A. Endo; Q. Wan; M. K. Spassova; S. J. Keding; J. Allen; O. Ouerfelli; R. M. Wilson; S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 2715–2725. Copyright (2006) American Chemical Society.

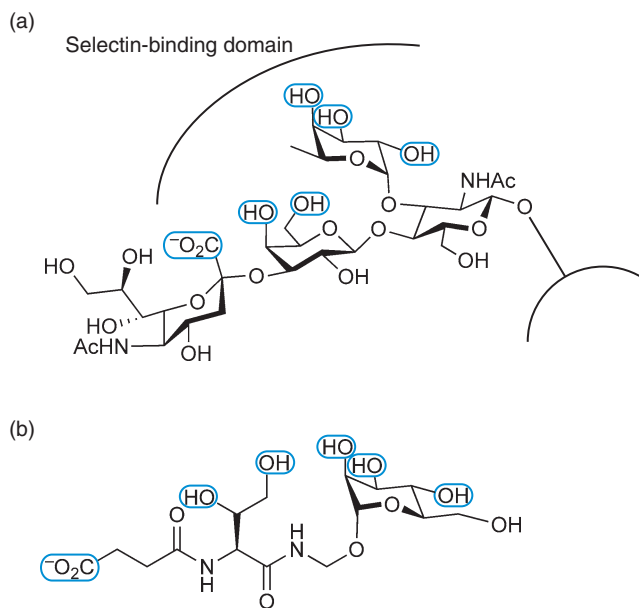


Figure 3 Design of sLe^X mimics. (a) Selectins recognize specific functional groups (highlighted in blue) on sLe^X. (b) Glycan–peptide mimic of sLe^X presents these functional groups in a similar geometry.

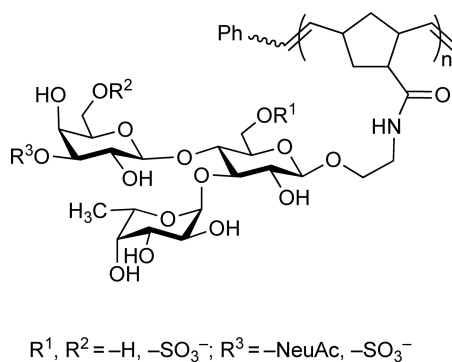


Figure 4 Polyvalent sLe^X is a potent inhibitor of L-selectin-mediated leukocyte rolling.

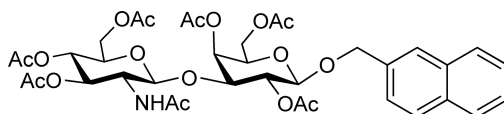


Figure 5 Structure of a peracetylated disaccharide primer that interferes with sLe^X synthesis.

GAGs share a common core tetrasaccharide that links them to a protein. GAG biosynthesis begins with the addition of xylose (Xyl) to specific serine (Ser) residues of core proteoglycans. This event is followed by sequential addition of three more sugars to complete the core tetrasaccharide, GlcA β 1–3Gal β 1–3Gal β 1–4Xyl–Ser (where GlcA is glucuronic acid). The repeating polymers appended to this tetrasaccharide determine the identity of the GAG: heparin and HS contain polymers of β 1–4GlcA α 1–4GlcNAc and β 1–4IdoA α 1–4GlcNAc, respectively; CS is composed of β 1–3GlcA β 1–4GalNAc (*N*-acetylgalactosamine); and DS is made up of β 1–3IdoA β 1–4GalNAc (where IdoA is iduronic acid). All of these polymers are heterogeneous and are subject to a variety of sulfation events.

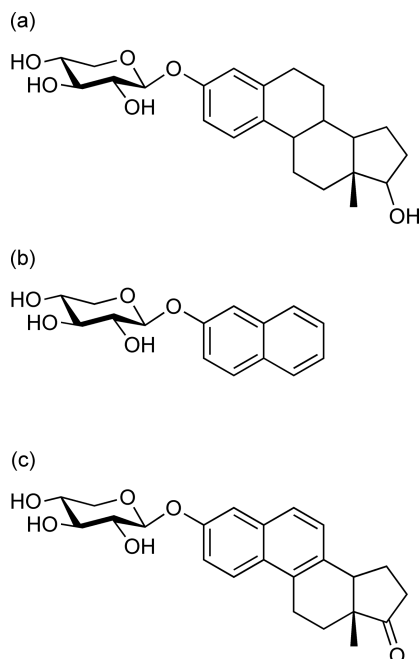


Figure 6 Three lipophilic β -D-xylosides that inhibit GAG synthesis: (a) 3-estradiol- β -D-xyloside, (b) 2-naphthol- β -D-xyloside, and (c) 3-D-equilenin- β -D-xyloside.

HS has been implicated in the mechanism for cellular uptake of polyamines.⁴² Because cells depend on polyamines for proper growth, depletion of cell surface GAGs is a potential therapeutic approach to combat uncontrolled tumor growth.⁴³ As described in Section 6.07.2.1, oligosaccharide primers can be used to divert cellular glycosyltransferases, thereby interfering with the production of normal cellular glycoconjugates. Similarly, β -D-xylosides function as primers for GAG synthesis. Simple β -D-xylosides are preferentially elongated as CS or DS chains and only weakly interfere with HS or heparin production. Esko and coworkers improved the ability of β -D-xylosides to function as HS primers by using lipophilic substituents such as estradiol,⁴⁴ 2-naphthol,⁴⁵ and 3-D-equilenin⁴⁵ in the aglycone moiety (Figure 6). Disaccharide primers provide even more selectivity toward GAG synthesis, but cellular delivery of these molecules requires the use of a larger lipophilic aglycone³⁴ or protection with acyl groups.⁴⁶ Using a naphthalenemethanol β -D-xyloside in combination with a polyamine biosynthesis inhibitor (DFMO – difluoromethylornithine), Esko and coworkers showed a decrease in cell proliferation both *in vitro* and *in vivo*, preventing tumor growth in a mouse model of induced metastasis.⁴³ GAG synthesis has also been implicated in the transformation of cellular prion protein (PrP^C) into its pathogenic form (PrP^{Sc}) in prion diseases. After treatment with an estradiol β -D-xyloside, the Esko group observed a strong reduction in the formation of PrP^{Sc}, demonstrating a link between HS biosynthesis and the production of PrP^{Sc}.⁴⁷ These examples, among others, demonstrate the utility of soluble primers in the study of GAG function.

CS plays important roles in neural development, functions as a recognition target for viral invasion, and assists in mechanisms of recovery from spinal cord injuries.^{48–50} Due to the complexity and heterogeneity of naturally occurring CS, deciphering its mechanistic roles remains a challenge. Recently, improved analytical techniques have made it possible to identify the structures that mediate specific biological processes: the CS-E tetrasaccharide (GlcA β 1–3GalNAc(4S,6S))₂ was shown to directly interact with L- and P-selectins, while related structures were implicated in neurite outgrowth.^{51,52} To better understand the role of specific GAG structures in neuronal growth, Hsieh-Wilson and coworkers synthesized CS disaccharides and tetrasaccharides with defined sulfation patterns. They showed that a CS-E tetrasaccharide (Figure 7) is the minimal motif required to promote neurite outgrowth.⁵³ Using a microarray printed with immobilized CS oligosaccharides, they demonstrated CS-E tetrasaccharide binding to several growth factors involved in neuronal development, including midkine,

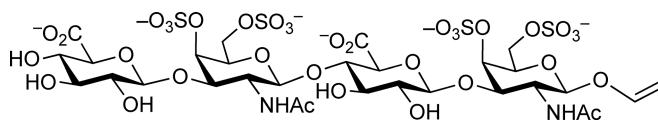


Figure 7 CS-E tetrasaccharide.

brain-derived neurotrophic factor (BDNF), and phospholipase C (PLC),^{54,55} and discovered that CS-E binds tumor necrosis factor α (TNF- α), a proinflammatory cytokine.⁵⁶ To investigate the effects of multivalency, Hsieh-Wilson and colleagues used ring-opening metathesis to polymerize CS disaccharides and tetrasaccharides. As observed for many other glycan motifs, multivalent display of CS-E enhanced its potency: neuroactivity of CS-E disaccharide polymers is enhanced by increases in polymer length (see Chapter 6.14).⁵⁷

6.07.2.3 Glycolipids/Gangliosides

Glycolipids are essential membrane components found in most organisms, including humans and the pathogens that infect us. Genetic defects in glycolipid metabolism are the cause of lysosomal storage diseases, a group of progressive, and often fatal disorders and have been implicated in neurological conditions, including epilepsy. Inappropriate glycolipid expression is characteristic of a number of malignancies and appears to play a role in the metastatic behavior of transformed cells. In addition, pathogens often synthesize unusual glycolipids whose antigenic determinants are essential for the development of host immunity. Here we review classes of glycolipids and glycolipid analogs that have been prepared synthetically and describe the use of these molecules in biological assays and therapeutic regimens.

The most commonly occurring lysosomal storage disorder is Gaucher's disease. Gaucher's disease is caused by mutations in β -glucosidase that impair its ability to catabolize glucosylceramide.⁵⁸ Inactive glucosylceramide accumulates in the lysosome, leading to variety of debilitating symptoms. Even small increases in β -glucosidase activity can improve symptoms, allowing this disease to be better managed. A common treatment for Gaucher's disease consists of β -glucosidase enzyme replacement. However, this treatment is extremely expensive and not all patients respond. An alternative strategy is to use chemical chaperones to facilitate folding of β -glucosidase into an active form. Toward this end, Kelly and coworkers have synthesized mimics of glucosylceramide (**Figure 8**). Early work using the *N*-(*n*-nonyl)deoxynojirimycin (NN-DNJ) chaperone led to a twofold improvement in the activity of a common disease-causing mutant, β -glucosidase N370S.⁵⁹ Adding a long adamantyl-modified alkyl chain or changing the iminosugar core to isofagomine increased the potency of the chaperone,^{60,61} and combining these features further enhanced chaperone activity.⁶² Intriguingly, isofagomine itself restored correct folding and

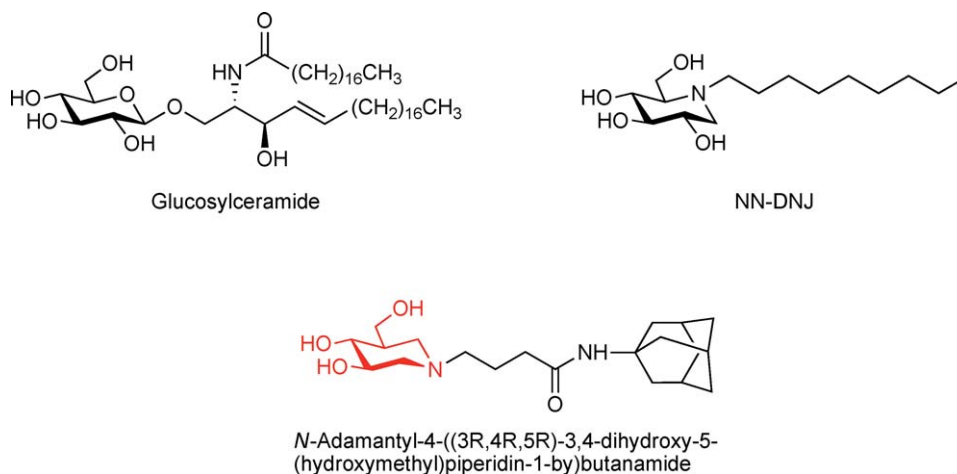


Figure 8 Glucosylceramide and two glucosylceramide mimics that function as chemical chaperones of β -glucosidase. The isofagomine core is shown in red.

lysosomal activity to the N370S mutant.⁶³ A crystal structure revealed that this molecule binds in the active site, locking the enzyme in a substrate-bound conformation.⁶⁴

While most known glycolipid disorders are associated with ineffective catabolism of these molecules, emerging evidence suggests that defects in glycolipid synthesis are the cause of some heritable diseases. Recently, a familial form of epilepsy was shown to be associated with an inability to synthesize GM3 ganglioside.⁶⁵ Administration of synthetic forms of the missing glycolipids is being explored as a treatment for this disease. More work will be required to determine whether there are other disorders of glycolipid synthesis that remain uncharacterized.

Glycophosphatidylinositols (GPIs) are complex glycolipids that are added posttranslationally to many eukaryotic proteins, anchoring these molecules in the membrane.⁶⁶ GPIs comprise a phospholipid tail and a pentasaccharide glycan core, which is modified with additional sugars and phosphoethanolamines and linked to the C-terminus of the attached protein (Figure 9). The core glycan of GPIs is conserved and consists of three mannose (Man) residues, a glucosamine, and a *myo*-inositol. Many functions have been ascribed to the GPI anchor,⁶⁶ but other than a clear significance in membrane targeting, its biological roles are not well understood. The first chemical synthesis of a GPI was reported in 1991 by Murakata and Ogawa.⁶⁷ Since then a number of other synthetic routes to GPI anchors and GPI-modified proteins have been explored.⁶⁸ To understand the functional significance of the conserved glycan core, Paulick *et al.*⁶⁹ recently prepared a panel of molecules that lacked three, four, or five of the sugars of the conserved pentasaccharide. These analogs were conjugated to green fluorescent protein (GFP) and incorporated into lipid bilayers or the plasma membrane of living cells.⁷⁰ Previous work had suggested that GPI-anchored molecules preferentially associate with lipid rafts, which are membrane microdomains of decreased fluidity.⁷¹ With this in mind, Paulick *et al.* measured the diffusion rates of these GPI analogs to learn more about their membrane environment. These researchers observed that deleting sugars from the glycan core led to corresponding decreases in the lateral mobility of these molecules, indicating that sugars contribute to the way in which GPI-anchored molecules interact with other membrane components.

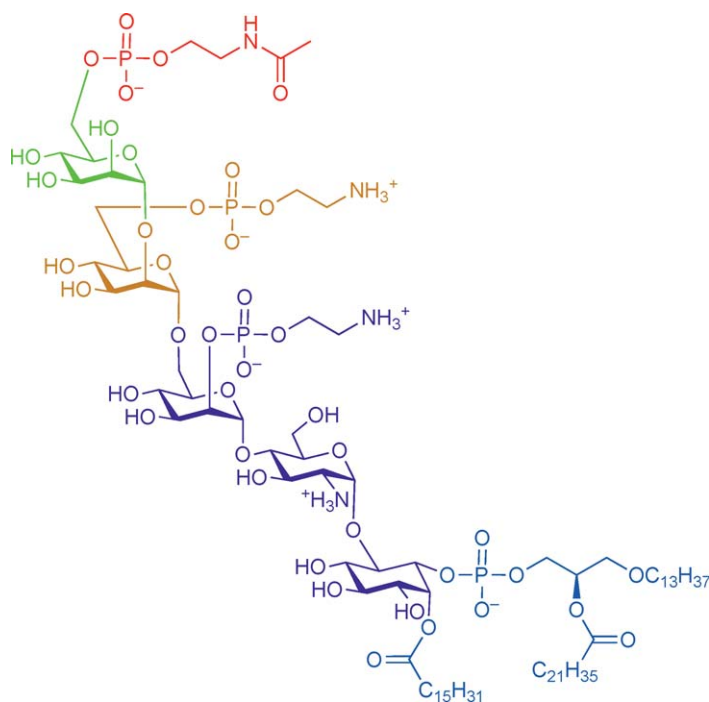


Figure 9 Structure of the GPI anchor found in eukaryotes (phosphoethanolamine linker shown in red, phospholipid tail shown in black). Paulick and coworkers examined several truncated forms of the GPI anchor to examine the functional significance of the core glycan structure. These include structures replacing three (cyan), four (orange and cyan), or five (green, orange, and cyan) of the sugars in the conserved pentasaccharide structure with appropriate ethylene glycol spacers.

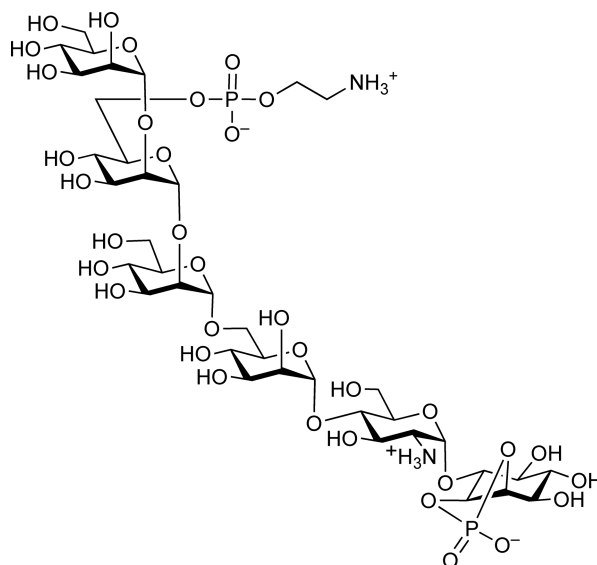


Figure 10 An antigenic *Plasmodium falciparum* GPI analog.

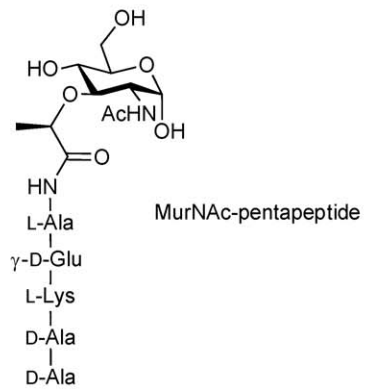
The malaria parasite *Plasmodium falciparum* also produces a cell surface GPI. This molecule functions as a proinflammatory toxin in malarial infection. Unlike mammalian GPIs, which are heavily modified, *P. falciparum* GPIs display an exposed core glycan that is antigenic in animals. To investigate the immune response to this molecule, Seeberger and coworkers prepared the *P. falciparum* GPI glycan synthetically and conjugated it to KLH (**Figure 10**).⁷² Mice injected with this conjugate produced anti-GPI antibodies and demonstrated substantially reduced mortality rates in a malarial infection model. Mice also exhibited a reduction in common malarial symptoms, including pulmonary edema and acidosis, demonstrating that this vaccine may also have desirable antitoxin properties. The convergent synthetic route developed by the Seeberger group provides access to fully lipidated GPI^{73,74} and a variety of other analogs.^{75,76} Using microarrays printed with these analogs, these investigators discovered that high levels of IgGs against Man₃-GPIs were present only in malaria-exposed individuals.⁷⁷ The Man₃-GPI epitope may therefore be an important component of future malarial vaccines.

6.07.2.4 Carbohydrate Roles in Bacteria and Mycobacteria

Similar to eukaryotes, bacteria also rely on carbohydrates for many facets of extracellular communication. Sugars serve as important signaling molecules for bacteria, indicating nutrient sources and inducing chemotactic behavior.⁷⁸ Bacteria also utilize carbohydrates to enclose and protect themselves. The cell wall of many Gram-positive bacteria is composed of a dynamic peptidoglycan structure, whose biosynthesis is critical to bacterial propagation. Many common antibiotics target the enzymes responsible for peptidoglycan synthesis. In some bacteria, including *Mycobacterium tuberculosis*, peptidoglycan is decorated with an additional arabinogalactan–mycolic acid complex that decreases the bacterium’s susceptibility to antibiotics. In this section, we discuss some of the diverse applications of chemical glycobiology in examining prokaryotic biology.

The peptidoglycan cell wall that surrounds Gram-positive bacteria is a rigid polymer composed of β 1–4-linked glycans cross-linked through peptide chains. Biosynthesis of peptidoglycan occurs in stages (**Figure 11**). Stage 1 culminates in the synthesis of UDP(uridine diphosphate)-MurNAc(*N*-acetylmuramic acid)-pentapeptide. Stage 2, which occurs on the cytosolic face of the membrane, consists of coupling to a membrane-anchored undecaprenyl (C₅₅), generating Lipid I, and the addition of another GlcNAc, generating Lipid II. In Stage 3, Lipid II translocates to the extracellular face of the cell membrane, where it is subject to further polymerization and cross-linking by various transglycosylases and transpeptidases. This results in the production of Lipid IV and cross-linked peptidoglycan chains. Commonly used antibiotics, including β -lactams and glycopeptides,

(a)



(b)

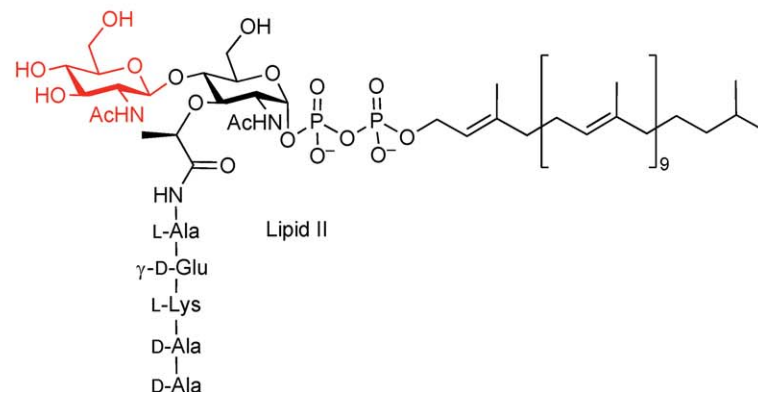
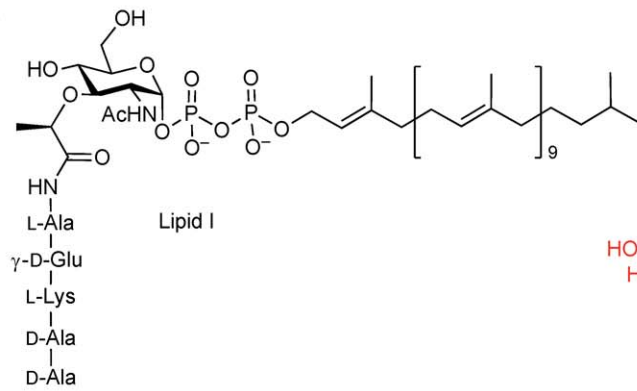


Figure 11 (Continued)

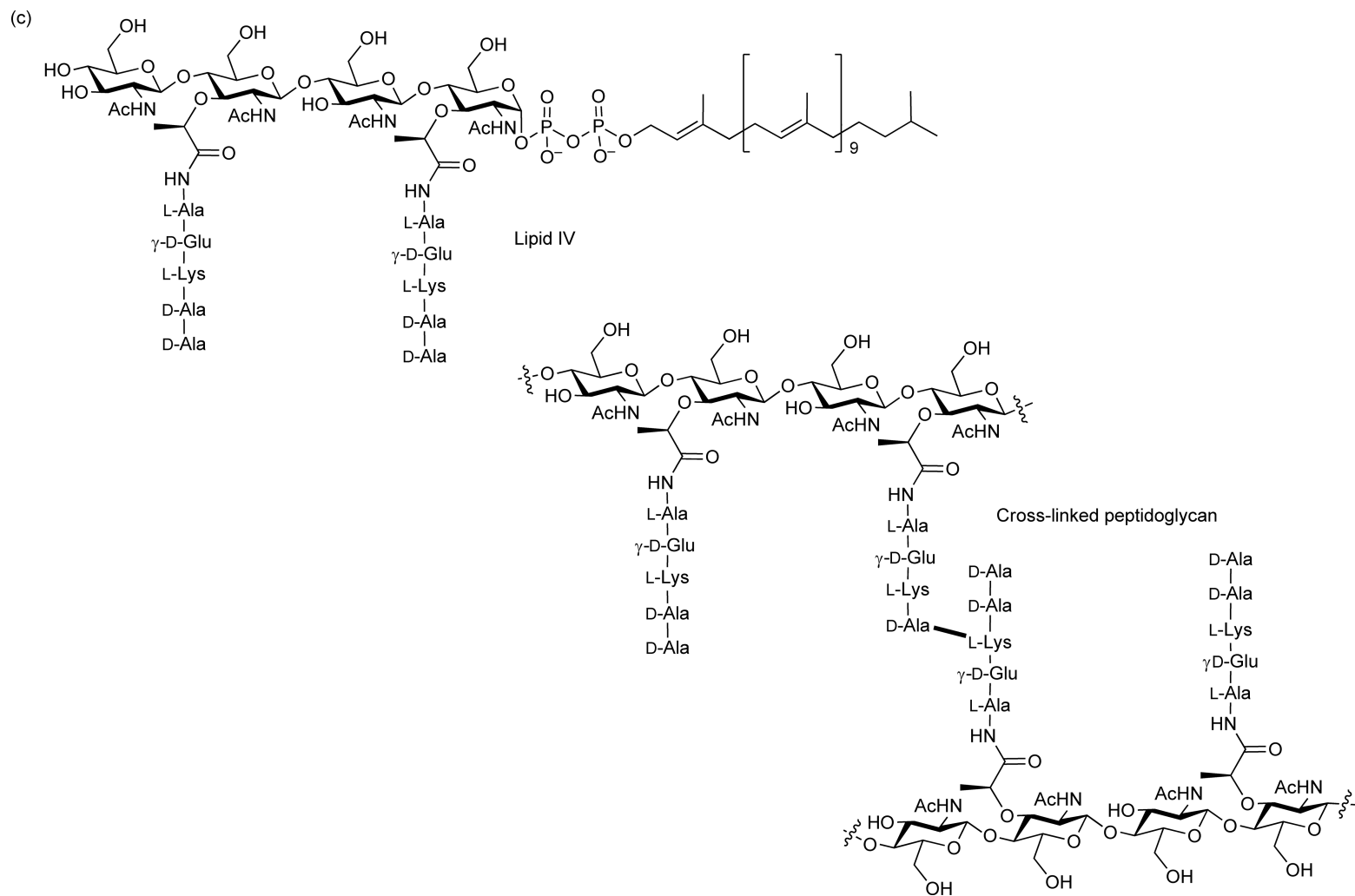


Figure 11 Structures of key molecules in peptidoglycan biosynthesis: (a) MurNAc-pentapeptide; (b) Lipid I and Lipid II (GlcNAc residue added by MurG is shown in red); (c) Lipid IV and the cross-linked peptidoglycan structure.

target enzymes involved in peptidoglycan biosynthesis. The rise of bacterial resistance to these antibiotics has motivated the search for new strategies to interfere with this pathway.

MurG catalyzes the transfer of GlcNAc to the Lipid I substrate, assembling the disaccharide–pentapeptide building block Lipid II. Due to the extreme lipophilic nature of the undecaprenyl chain, mechanistic characterization of MurG has remained elusive. Several groups have reported syntheses of Lipid I,^{79–81} but small yields have limited the use of these molecules in mechanistic studies. To aid this endeavor, Walker and coworkers prepared several shorter-chain lipid analogs of Lipid I. An early analog replaced the 55-carbon chain with a much shorter 10-carbon citronello chain,⁸² and facilitated studies of MurG's substrate selectivity.⁸³ Lipid I analogs with other lipid substitutions have enabled further exploration of MurG's activity.^{84–86}

The product of MurG, Lipid II, has also been prepared synthetically⁸⁷ and used in transpeptidation reactions.⁸⁸ Analogs of Lipid II have been used to study the activity of transglycosylases in peptidoglycan biosynthesis⁸⁹ and to characterize the mechanism of action of nisin, a bacterial peptidoglycan antibiotic.⁹⁰ Zhang *et al.*⁹¹ reported the synthesis of a heptaprenyl-Lipid IV analog that could be incorporated into larger peptidoglycan structures by *Escherichia coli* peptidoglycan glycosyltransferases. Using a version of this molecule with a Gal-block on the nonreducing end, they demonstrated that the peptidoglycan chain grows from the reducing end.⁹² These results are consistent with the mechanistic interpretations that have been proposed based on crystal structures of the peptidoglycan glycosyltransferases.^{93,94}

The peptidoglycan-based cell wall of *M. tuberculosis* is further modified by the addition of an arabinogalactan–mycolic acid complex (Figure 12). This additional barrier poses an added challenge for antibiotic design. The arabinogalactan chain is synthesized by two galactosyltransferases.⁹⁵ Belanova *et al.*⁹⁶ prepared lipid-shortened analogs of the rhamnose–GlcNAc linker and used them to demonstrate that the addition of the first two galactofuranose units is controlled by galactofuranosyltransferase 1, while further growth of the galactofuranose chain is regulated by galactofuranosyltransferase 2. Additional experiments demonstrated that two arabinofuranosyl transferases, EmbA and EmbB, catalyze the addition of the arabinan directly onto galactan before incorporation into the peptidoglycan.⁹⁷ In this way, synthetic access to glycan intermediates has provided insight into mechanisms of *M. tuberculosis* galactan biosynthesis and may aid in the design of antimycobacterial agents.

Environmental carbohydrates also affect the behavior of prokaryotes. Bacteria move in response to gradients of molecules in their surroundings. Chemotaxis occurs through a dynamic sensing mechanism that is regulated by extracellular chemoreceptors. Stimuli can act as either attractants or repellants. Simple sugar molecules such as Gal and Glc are known molecular attractants for bacteria. Kiessling and coworkers investigated the nature of chemotactic responses in *E. coli* using multivalent ligands of individual sugar molecules. Using ROMP, they synthesized multivalent ligands displaying up to 50 individual sugar molecules (Figure 13). These multivalent ligands elicited a more significant chemotactic response than Gal alone.^{98–101} The investigators hypothesized

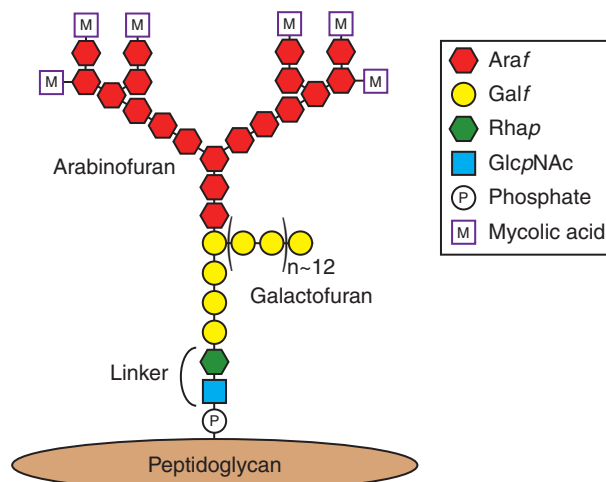


Figure 12 Cartoon of the arabinogalactan–mycolic acid-modified *Mycobacterium tuberculosis* peptidoglycan structure. Pyranose forms of sugars are indicated by *p*, furanose forms of sugars are indicated by *f*.

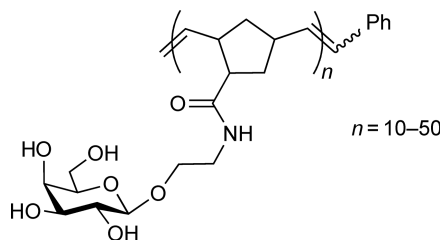


Figure 13 Multivalent galactose polymer used to investigate chemotactic responses in bacteria.

that the increase in chemotactic response could be attributed to the ability of a single multivalent molecule to simultaneously engage multiple cell surface chemoreceptors. Multivalent polymers were also shown to induce chemoreceptor clustering, as demonstrated by fluorescence colocalization experiments.^{98–100} Both targeted and nontargeted receptors exhibited clustering.⁹⁸ Moreover, increases in receptor clustering caused the nontargeted receptors to become more sensitive to their own ligands.⁹⁸ These responses indicate that inter-receptor communication occurs during the chemotactic response. Analogous responses were also observed in *Bacillus subtilis*, *Vibrio furnissii*, and *Spirochaete aurantia* (see Chapter 6.13).¹⁰¹

6.07.2.5 Toxins

Many bacterial toxins and viruses bind and infect their hosts by recognizing host cell surface glycans. Since protein–glycan interactions are typically weak and transient, toxins rely on multivalency to increase their affinity for glycan targets. To attempt to interfere with these binding events, several groups have prepared multivalent glycan structures optimized for toxin binding. This section describes multivalent glycan mimetics that are designed to bind bacterial toxins and interfere with their toxic effects.

Shiga toxin is an AB₅ bacterial toxin. The five B subunits of the assembled toxin cooperatively recognize glycolipids on the host cell surface, primarily the glycolipid GB₃ (globotriaosylceramide). Each of the individual B subunits displays three potential binding sites for GB₃. Following recognition, the catalytically active subunit A is endocytosed, ultimately disrupting protein synthesis in the invaded cell. Bundle and coworkers developed a decameric glycan display by attaching bivalent GB₃ ligands to each of the five hydroxyl groups of Glc. They dubbed the resulting molecule STARFISH (**Figure 14**).¹⁰² STARFISH inhibited Shiga toxin binding six orders of magnitude better than the GB₃ trisaccharide alone. STARFISH and a closely related inhibitor, DAISY, which differ in linker composition (**Figure 14**),¹⁰³ were tested as therapeutics in mice treated with two forms of Shiga-like toxins (Shiga-like toxin 1 and Shiga-like toxin 2). STARFISH significantly increased survival rates in mice treated with Shiga-like toxin 1, while DAISY was effective against both toxins.

In another strategy to develop reagents to treat Shiga intoxication,¹⁰⁴ Nishikawa *et al.*^{105,106} prepared a carbosilane dendrimer scaffold bearing the GB₃ epitope in various valencies (for an example molecule, see **Figure 15**). These ‘SUPER TWIG’ molecules were found to bind and inhibit both forms of Shiga-like toxin and completely suppressed the lethal effects of Shiga-like toxin 2 in mice.^{105,106} While these results were promising, the molecules could only be administered intravenously. To develop an oral agent, the investigators changed their scaffold to an acrylamide polymer backbone and generated copolymers with regularly spaced pendant GB₃ glycans.^{107,108} The new glycopolymer was demonstrated to bind Shiga toxin 2 as well as SUPER TWIG did and can be administered orally to mice, yielding similar suppression of the toxin’s effects.¹⁰⁸

6.07.3 Chemoenzymatic Synthesis to Study Glycan Structure and Function

Methods to synthesize novel glycan-containing molecules are desirable for a variety of applications. For example, new macrolide sugars may display attractive antibiotic properties, new polysaccharides may have unanticipated pharmaceutical or industrial uses, and modifications to naturally occurring glycoconjugates could allow for exploration of structure–function relationships. While synthetic chemistry represents one route to these molecules, carbohydrate synthesis is notoriously difficult and not always generalizable. An alternative to synthetic chemistry is

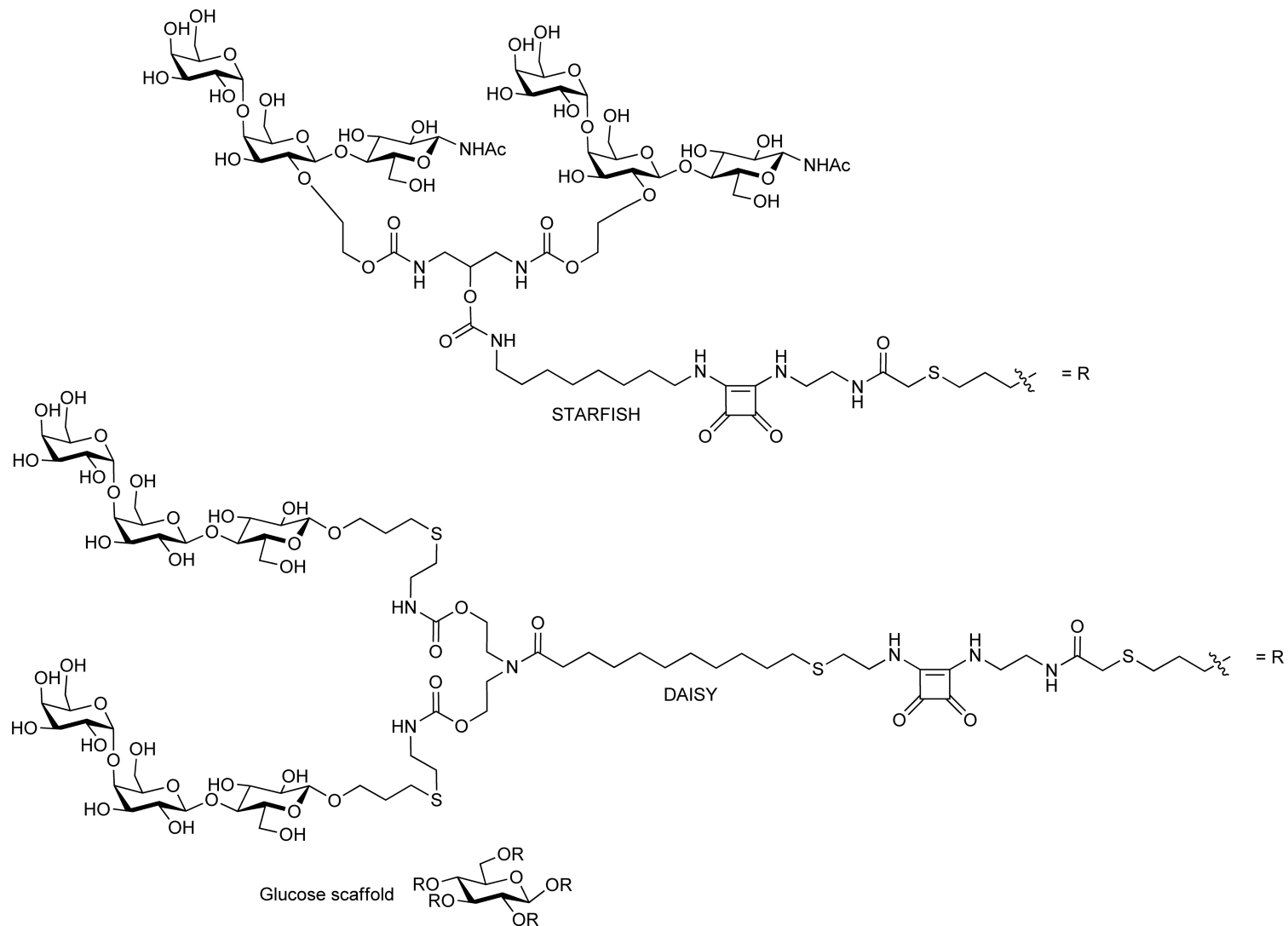


Figure 14 Structures of STARFISH and DAISY Shiga toxin inhibitors. The decavalent glycan display involves the attachment of a bivalent ligand (designated by R groups) onto a glucose scaffold.

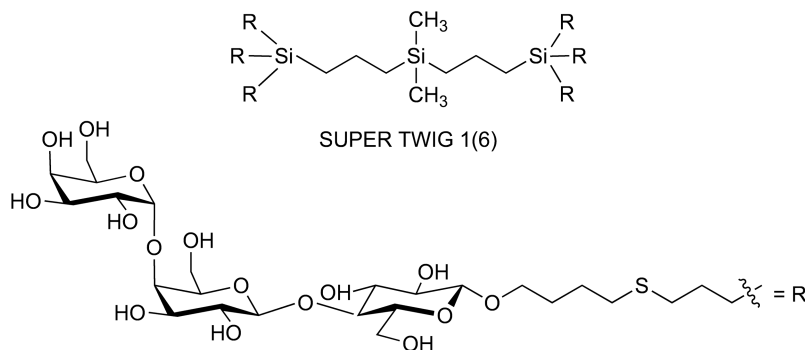


Figure 15 Structure of the SUPER TWIG 1(6) Shiga toxin inhibitor. The nomenclature is derived from the generation of the dendrimer and the number of glycan displays. Here, the hexavalent glycan display (designated by R groups) is presented on a first-generation carbosilane dendrimer scaffold.

to make use of glycan biosynthetic enzymes for chemoenzymatic synthesis of unnatural glycoconjugates. *In vitro*, many of these enzymes will accept unnatural substrates; in other cases, enzymes have been engineered to expand their substrate tolerance. Chemoenzymatic synthesis of glycoconjugates is a broad field and receives attention in several other chapters in this volume. Here we focus on two important classes of targets for chemoenzymatic synthesis: alternative nucleotide sugar donors and variant sialic acids (Sias).

6.07.3.1 Alternative Nucleotide Sugar Donors

High-energy nucleotide sugar donors are a key intermediate in glycan synthesis (Figure 16). Glycosyltransferases transfer sugars from nucleotide sugar donors to glycan, protein, lipid, or metabolite substrates. Synthesis of alternative nucleotide sugar donors that are accepted and utilized by glycosyltransferases is an essential step in unnatural glycan production. Biosynthetically, nucleotide sugars are the products of nucleotidyltransferases. Nucleotidyltransferases, also called sugar pyrophosphorylases, catalyze phosphodiester bond formation in the reaction of sugar-1-phosphates with nucleotidetriphosphates (Figure 16). The *in vitro* use of nucleotidyltransferases and glycosyltransferases for chemoenzymatic synthesis of glycoconjugates was pioneered in the 1980s and 1990s and is reviewed elsewhere.^{109–112} Preparative scale syntheses have been further enabled by methods to immobilize and recycle the necessary enzymes.¹¹³ In this section, we discuss how unnatural nucleotide sugars can be produced using new chemoenzymatic strategies or adaptations of existing chemoenzymatic techniques. Approaches include exploiting the permissivity of nucleotidyltransferases from bacterial and archaeal sources, mutagenesis of naturally occurring nucleotidyltransferases, taking advantage of the reversibility of glycosyltransferase-catalyzed reactions, and the introduction of azide functional groups for glycorandomization through click chemistry.

Although nucleotidyltransferases usually catalyze production of only one or a few nucleotide sugar products *in vivo*, their substrate scope is often limited only by substrate availability. When presented with unnatural sugar-1-phosphates or nucleotidetriphosphates, nucleotidyltransferases frequently accept these substrates, catalyzing production of unnatural nucleotide sugars (Figure 17). Studies of nucleotidyltransferase permissivity have examined

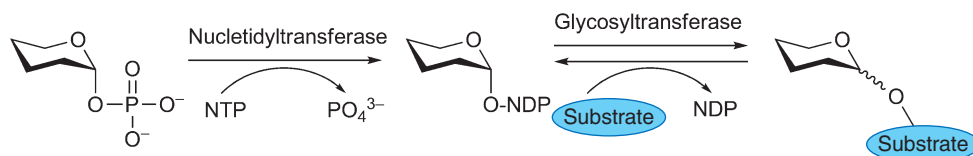


Figure 16 Enzymatic steps in glycoconjugate biosynthesis. Nucleotidyltransferases catalyze the reaction of sugar-1-phosphates with nucleotidetriphosphates to produce nucleotide sugar donors. These activated sugars are transferred to substrates by glycosyltransferases. The general structure of sugar-1-phosphate is meant to be applicable to all types of nucleotide sugars. Commonly occurring nucleotide sugars include UDP-glucose, UDP-GlcNAc, UDP-GlcA, UDP-Gal, UDP-GalNAc, UDP-xylose, GDP-fucose, and GDP-mannose. Sialic acid donors are monophosphates, e.g., CMP-NeuAc.

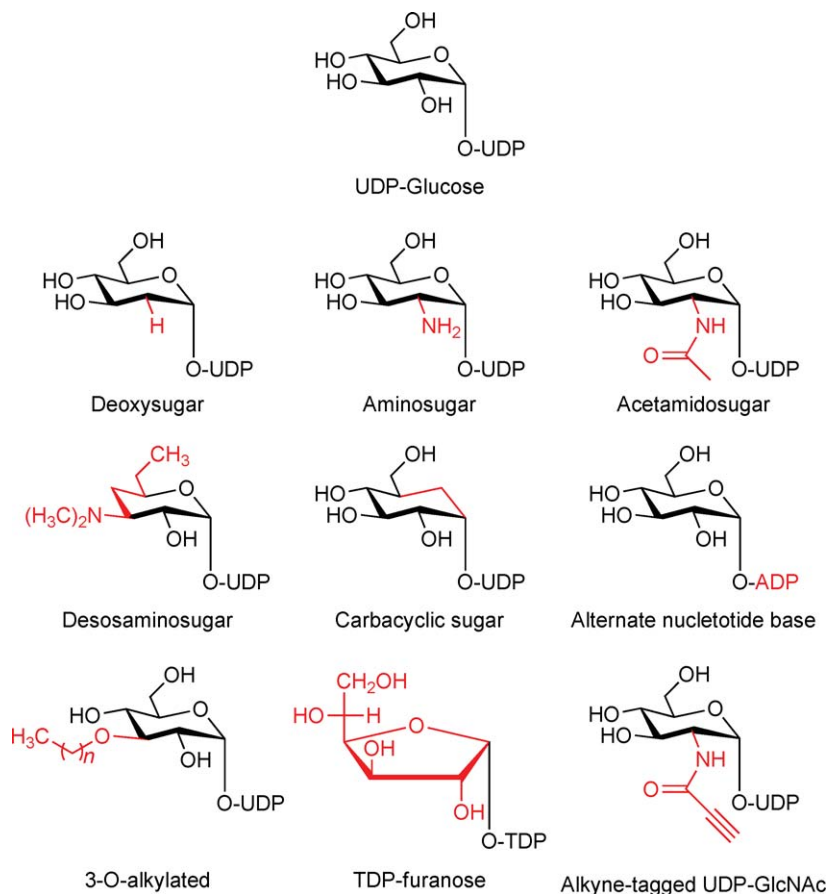


Figure 17 Some examples of alternative nucleotide sugar donors that have been produced chemoenzymatically. Differences from a natural nucleotide sugar donor structure, UDP-glucose, are highlighted in red.

enzymes from a variety of organisms. An early and extensive analysis of substrate scope was conducted with the α -D-glucopyranosyl phosphate thymidyltransferase from *Salmonella enterica* LT2 (E_p). This permissive enzyme produces appreciable quantities of unnatural TDP-sugars and UDP-sugars by accepting a number of deoxy sugars,¹¹⁴ amino- and acetamidoglycosyl donors,¹¹⁵ and desosamine analogs.¹¹⁶ *Escherichia coli* enzymes also demonstrate tolerance toward unnatural substrates: a glucose-1-phosphate uridylyltransferase is capable of producing carbacyclic UDP-glucose¹¹⁷ and several UDP- and dTDP(deoxythymidine diphosphate)-deoxyglucoses.¹¹⁸ Bacterial thymidyltransferases from *Streptococcus pneumoniae* R6 and *Aneurinibacillus thermoaerophilus* DSM 10155 are permissive for alternate nucleotides and can produce ADP(adenosine diphosphate)-, CDP(cytidine diphosphate)-, and GDP(guanosine diphosphate)-sugars.¹¹⁹ The same thymidyltransferases also accept 3-O-alkylated glucose-1-phosphates, leading to the production of lipophilic nucleotide sugar donors.¹²⁰ Furthermore, the *S. pneumoniae* enzyme is able to synthesize dTDP-furanoses.¹²¹ Broad substrate scope has also been demonstrated for plant enzymes, including UDP-sugar pyrophosphorylases from *Pisum sativum* L. (pea)¹²² and *Arabidopsis thaliana*.¹²³

Nucleotidyltransferases from extremophiles often exhibit unusually broad substrate tolerance. Pohl and coworkers have examined several enzymes from the thermophile *Pyrococcus furiosus*. One such nucleotidyltransferase accepts a variety of glycan-1-phosphates including Man, Gal, and, to a lesser extent, fucose, glucosamine, galactosamine, and GlcNAc.¹²⁴ Intriguingly, this enzyme also demonstrates acyltransferase activity, catalyzing the *N*-acetylation of glucosamine-1-phosphate. When provided with alternative acyl donors, the bifunctional enzyme can catalyze the two-step conversion of glucosamine into chloroacetyl- and alkyne-tagged UDP-GlcNAc analogs.¹²⁵ Other extremophile enzymes with broad substrate tolerance include a glucose-1-phosphate uridylyltransferase from the acidophile *Helicobacter pylori*,¹²⁶ a CMP(cytidine

monophosphate)-Neu5Ac (*N*-acetylneuraminic acid) synthase from the thermophile *Clostridium thermocellum*,¹²⁷ a nucleotidyltransferase from the thermophile *Thermus caldophilus* GK24,¹²⁸ and a mannosylglycerate synthase from the hypothermophile *Rhodothermus marinus*.¹²⁹

Notwithstanding the examples discussed above, there are limits to the natural permissivity of nucleotidyltransferases. To further expand the diversity of nucleotide sugar analogs, the Thorson group and others have created mutant enzymes that exhibit broadened substrate scope. Relying on the crystal structure of the *Salmonella* thymidyltransferase,¹³⁰ the Thorson and Nikolov groups prepared mutant thymidyltransferases that accept acetamido derivatives and epimers of the normal glucose-1-phosphate substrate.^{130,131} Similarly, mutations to the nucleotide-binding pocket yielded mutant enzymes that tolerated all eight naturally occurring NTPs (nucleotide triphosphates).¹³² An analogous strategy has been used to expand the tolerance of the sugar kinases that produce the sugar phosphate substrates of nucleotidyltransferases.^{133–135}

While much effort has focused on the nucleotidyltransferases, glycosyltransferases can also be utilized to produce unusual nucleotide sugar donors. Glycosyltransferases are generally described as catalyzing the transfer of a sugar from a nucleotide sugar donor to an acceptor, resulting in the formation of a new glycosidic bond between the sugar and the acceptor (Figure 16). However, glycosyltransferases also catalyze the reverse reaction: transfer of a sugar from a glycosylated molecule to a nucleotide, forming a nucleotide sugar. Recently, the Eguchi group exploited the reversibility of the reaction catalyzed by a glycosyltransferase. These investigators used the VinC glycosyltransferase to transfer sugars from a glycosylated substrate to a nucleotide diphosphate, resulting in nucleotide sugar formation.¹³⁶ Similarly, the Thorson group used a reverse glycosyltransfer reaction to prepare novel diphosphate nucleotide sugar donors from glycosylated natural products (Figure 18).^{137,138} Glycosyltransferase permissivity can be enhanced by mutagenesis, further expanding the scope of sugars that can be transferred.¹³⁹ By using natural products as sugar sources, this technology has the potential to produce an immense variety of nucleotide sugars.

The utility of alternative nucleotide sugar donors is dependent on their availability in large enough quantities to be used in subsequent reactions. Wang and coworkers have reported the ‘superbead’ method for large-scale syntheses of naturally occurring nucleotide sugar donors.^{113,140,141} Their approach relies on immobilizing and recycling the biosynthetic enzymes that produce these molecules. More recently, Errey *et al.*¹⁴² took advantage of the permissivity of a Gal-1-phosphate uridytransferase and showed that a similar immobilization strategy could be used to produce a variety of UDP-Gal analogs in good yield. These data suggest that a ‘superbead’ method could be an effective and scalable route to unnatural nucleotide sugar donors.

As a larger variety of unusual nucleotide sugars becomes available, chemoenzymatic methods are being developed to use these molecules in the synthesis of diverse glycoconjugates. Using a technique termed *in vitro* glycorandomization, Thorson and coworkers made use of permissive glycosyltransferases to transfer sugars from nucleotide sugar donors to aglycon substrates, creating libraries of natural products bearing diverse glycans. For example, the GtfE glycosyltransferase was used to rapidly produce 21 vancomycin analogs;¹⁴³ other natural products have been similarly diversified.^{144,145} By using nucleotide azidosugar donors, further diversification can be achieved rapidly: an azidosugar can be transferred to the vancomycin aglycon and then reacted with various alkynes, producing an even larger array of vancomycin analogs (Figure 19).^{143,146}

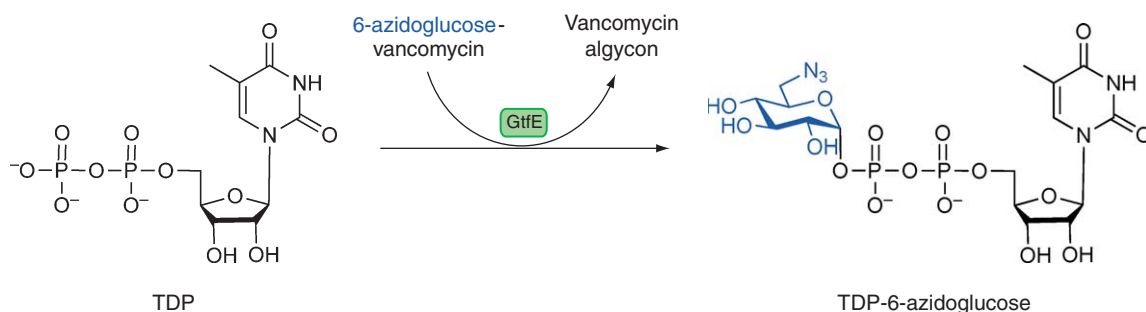


Figure 18 An aglycon exchange reaction can be used to produce unnatural nucleotide sugar donors. Shown here is an example using vancomycin.

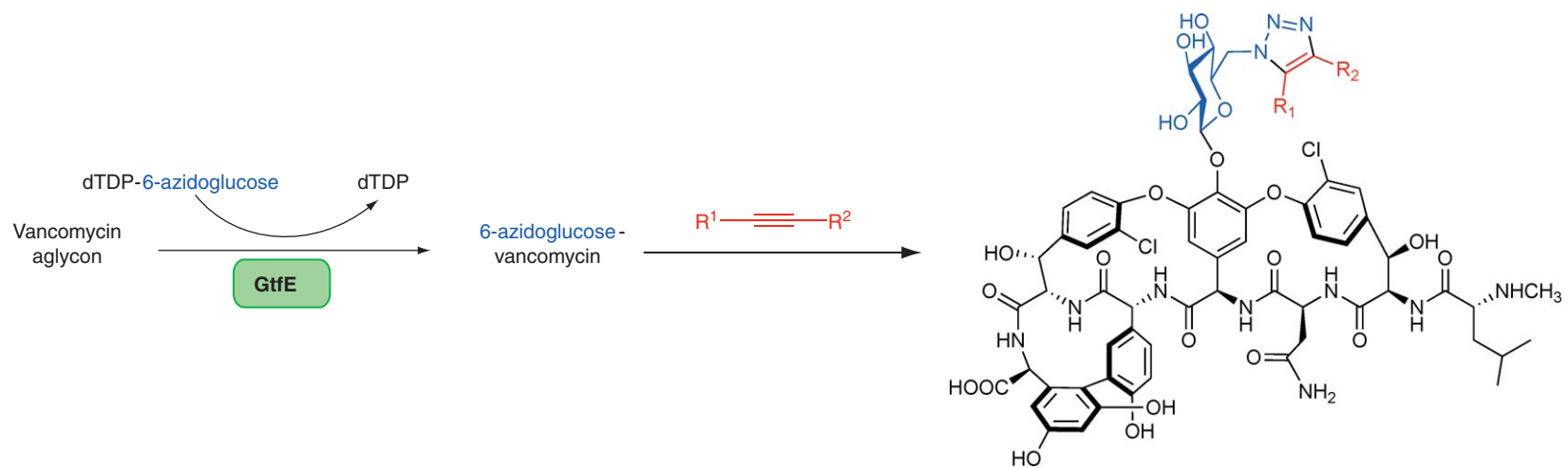


Figure 19 A glycosylation technique makes use of 'click chemistry' to create a variety of glycosidic products. Sugar is shown in blue, alkyne reagent is shown in red. Reprinted with permission from X. Fu; C. Albermann; C. S. Zhang; J. S. Thorson, *Org. Lett.* **2005**, 7, 1513–1515. Copyright (2005) American Chemical Society.

Alternative nucleotide sugar donors have been used in the synthesis of a wide variety of novel glycosylated products, too numerous to be treated comprehensively here. Some notable natural products that have been subjected to glycan diversification include novobiocin,^{139,144} calicheamicin,¹³⁸ avermectins,¹³⁷ oleandomycin,^{145,147} vicienistatin,¹³⁶ erythromycin,¹⁴⁷ tylosin,¹⁴⁷ methomycin/pikromycin,¹⁴⁸ and spinosyns.¹⁴⁹ Several oleandomycin and erythromycin analogs were shown to have improved antibacterial activity over their parent compound.¹⁴⁷ As the availability of novel nucleotide sugar donors grows, we expect this list to expand rapidly (see Chapter 6.19).

6.07.3.2 Chemoenzymatic Synthesis of Sialic Acids

The term sialic acid (Sia) refers to a group of nine-carbon α -keto acids derived from Neu5Ac, 5-glycolyneuraminic acid (Neu5Gc), and deaminated neuraminic acid (KDN) (Figure 20). A closely related keto-deoxy acid, 3-deoxy-D-manno-octulosonic acid (KDO) is commonly found in plants and in the extracellular polysaccharides that enclose capsulated bacteria. In vertebrates and some higher invertebrates, Sias are added to the nonreducing terminus of cell surface glycoconjugates, where they mediate recognition events. Sias are also found in the bacterial extracellular polysaccharides. Sia-containing glycoconjugates are commonly modified by postglycosylational processing, leading to over 50 naturally occurring variants of Sia (Figure 21).^{150,151} In this section we discuss chemoenzymatic methods in which Sia biosynthetic enzymes are used individually, or in combination with one another. These enzymes can catalyze the synthesis of a variety of natural and unnatural Sias and sialosides.¹⁵² We also note some of the biological uses of these compounds.

The Sia synthase, or aldolase, is responsible for the reversible aldol reaction that occurs between six-carbon mannosamines and pyruvate (Figure 22). In the forward direction, this condensation reaction creates a new carbon-carbon bond to produce nine-carbon Sias.¹⁵³ Aldolases capable of Sia synthesis are found in a variety of bacteria. The *E. coli* enzyme has proven to be extremely adaptable in chemoenzymatic reactions. Its substrate scope has been studied extensively and is described in a large body of literature published in the late 1980s and early 1990s. Taken together, these reports indicate that the enzyme exhibits a strong preference for pyruvate as

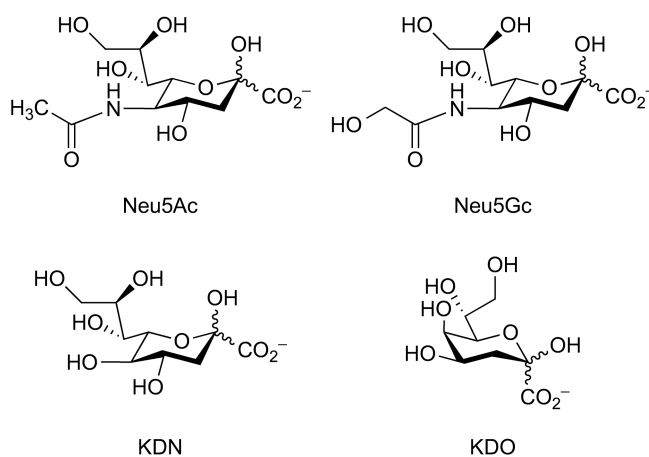


Figure 20 Sialic acids include Neu5Ac, Neu5Gc, and KDN. KDO is a closely related eight-carbon keto-deoxy acid that is found in several types of prokaryotes and plants.

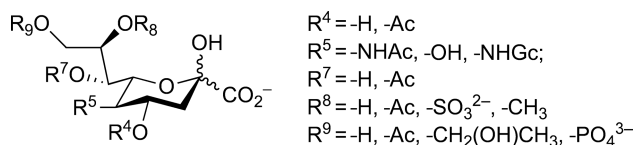


Figure 21 Modifications to sialic acids result in over 50 different naturally occurring molecules.

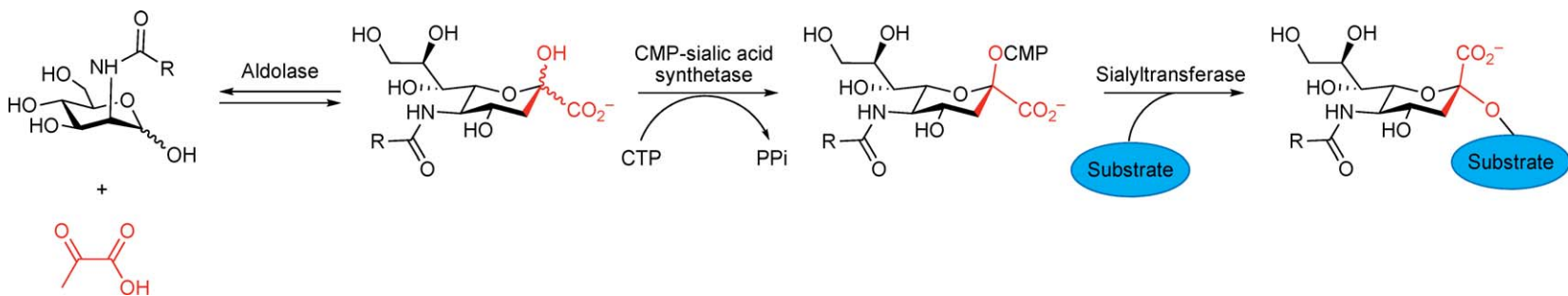


Figure 22 Biosynthesis of sialosides. Sialic acid aldolases catalyze the aldol condensation of *N*-substituted mannosamines with pyruvate to produce sialic acids. CMP-sialic acid synthetases produce CMP-sialic acid from CTP and sialic acid. Sialic acid is transferred to acceptor molecules by sialyltransferases.

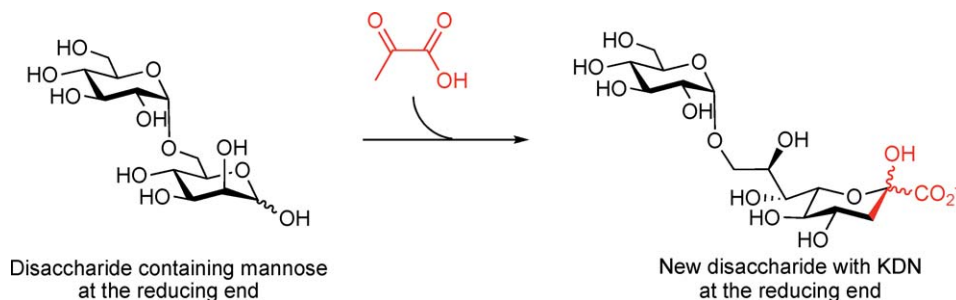


Figure 23 Aldolase-catalyzed synthesis of oligosaccharides bearing sialic acid at the reducing end. In this example, pyruvate (in red) was condensed with Glc α 1–6 Man to produce Glc α 1–9 KDN.

the donor, but a relaxed specificity toward acceptor molecules.^{154,155} The *E. coli* aldolase accepts a variety of naturally occurring hexoses and pentoses, as well as many unnatural ones. For example, this enzyme has been used to prepare 5-azido and 7-azido Sia analogs from the corresponding mannosamine precursors.^{156,157} *N*-substituted mannosamines are also well-tolerated by the aldolase.¹⁵⁸ Nonetheless, the enzyme does not accept all sugars: aldoses lacking a C-3 hydroxyl group are particularly poor substrates.¹⁵⁴ In addition, the reversibility of the aldolase-catalyzed reaction has been exploited to convert Sias into their corresponding Man derivatives and to interconvert Sias.^{159,160} More recently, the aldolase's permissivity to substitutions on the sugar substrate has been exploited by the Chen group to synthesize oligosaccharides that have Sia at the reducing end (**Figure 23**): the enzyme accepts a variety of disaccharides that have a Man or a Man derivative at the reducing terminus, and converts that terminal sugar into the corresponding KDN or Sia analog.^{161,162} These disaccharides mimic structures found in some pathogenic bacteria and can be further elaborated by other glycosyltransferases.

Mutagenesis of the aldolase has been employed to expand or redefine its substrate scope.^{163,164} The crystal structure of *E. coli* Sia aldolase was reported in 1994 and has served as an important resource in efforts to evolve this enzyme.¹⁶⁵ The Wong group reported the use of *in vitro* directed evolution to alter the enantioselectivity of Sia aldolase. Wild-type aldolase exhibits a strong bias for D-Sia and other D-sugars such as D-KDO (**Figure 20**). However, Wong and coworkers were able to use directed evolution to select for a mutant enzyme with the opposite enantioselectivity. The mutant enzyme, which contains eight amino acid substitutions, efficiently synthesizes L-KDO and has lost much of its activity toward D-sugars.^{166,167} Berry and colleagues used structure-guided saturation mutagenesis to evolve an enzyme that accepts four-carbon aldehydes to produce seven-carbon Sia analogs.^{168,169} This group also evolved a pair of enzymes that exhibit opposing preferences for the stereochemistry of the C-4 position in the substrate.¹⁷⁰

To biosynthetically or chemoenzymatically prepare unnatural sialosides, the unnatural Sias must first be converted to nucleotide sugar donors. CMP-Sia synthetases catalyze the reaction of Sia with CTP to yield CMP-Sia and pyrophosphate (**Figure 22**). The *E. coli* synthetase is commonly used in chemoenzymatic reactions,^{171,172} but its narrow substrate scope¹⁷³ limits its use in the synthesis of variant CMP-Sias. The *Neisseria meningitidis* enzyme exhibits a broader substrate tolerance,¹⁷⁴ while the synthetase from *C. thermocellum* offers the benefit of thermostability.¹²⁷ Recently, the Chen group has combined use of the *N. meningitidis* CMP-Sia synthetase and the *E. coli* Sia aldolase in a one-pot reaction to effect conversion of Man and ManNAc (*N*-acetylmannosamine) analogs into their Sia counterparts. Using this enzyme pair, they were able to prepare a wide variety of Sia derivatives with substitutions at C-3, C-5, C-8, and C-9.^{175,176}

Transfer of Sias from CMP-Sia to an acceptor substrate requires the action of a sialyltransferase (**Figure 22**). Formation of the glycosidic bond of Sia is a particularly challenging reaction, so even many otherwise nonenzymatic syntheses rely on an enzymatic transformation to add Sia (e.g., see Blixt *et al.*¹⁷⁷). While a number of mammalian sialyltransferases are commonly used chemoenzymatically, these enzymes generally exhibit a narrowly defined substrate scope, limiting their utility in chemoenzymatic reactions with unnatural Sias. Sialyltransferases from prokaryotes can be more permissive. Yamamoto *et al.*¹⁷⁸ have developed the use of an α 2–6 sialyltransferase from *Photobacterium damsela* for chemoenzymatic synthesis of a variety of sialosides. This enzyme can be purified in good yield from its natural host and adds Sia in an α 2–6 linkage to β -linked Gal

found at the nonreducing end of glycans.^{179,180} The *P. damsela* enzyme tolerates acceptor galactosides that are modified at C-2 or C-3, including galactosamines and glycopeptide substrates.^{160,181} It also has remarkable permissivity for CMP-Sia donors, accepting molecules with a large variety of substituents at C-5, C-7, C-8, and C-9.^{182,183} An α 2-3 sialyltransferase from *N. meningitidis* demonstrates unusual acceptor tolerance, having the ability to sialylate both α - and β -linked galactosides and to accept several *N*-substituted CMP-Sia donors.¹⁸⁴ A related proteobacterial pathogen, *Neisseria gonorrhoeae*, produces an α 2-3 sialyltransferase that transfers Sias with several different C-5 substituents to a broad range of α -linked galactoside and galactosaminoside acceptors.¹⁸⁵ Prokaryotic polysialyltransferases, responsible for synthesizing polymers of α 2-8 or α 2-9 polysialic acid (PSA), may also prove to be useful for the chemoenzymatic syntheses.^{186,187} Sialyltransferases have even been identified in viral genomes: Palcic and coworkers described the *in vitro* use of an α 2-3 sialyltransferase from myxoma virus to transfer Sia residues onto several fucosylated acceptors, but the donor permissivity of this enzyme has not yet been reported.¹⁸⁸

Aldolases, CMP-Sia synthetases, and/or sialyltransferases have been combined in a number of one-pot, chemoenzymatic syntheses of naturally occurring sialosides (e.g., see Ichikawa *et al.*¹⁸⁹). The use of permissive enzymes broadens the scope of these one-pot methods and has facilitated the synthesis of a number of unnatural sialosides. The Chen group reported a *Pasteurella multocida* sialyltransferase that can function as an α 2-3 sialyltransferase, α 2-6 sialyltransferase, α 2-3 sialidase, or α 2-3 *trans*-sialidase, depending on the pH of the reaction mixture.¹⁹⁰ Using this enzyme in a one-pot reaction in combination with *E. coli* aldolase and *N. meningitidis* CMP-Sia synthetase, they produced sialosides with a variety of substituents at C-5, including alkynes and azides (Figure 24(a)). C-9 azido-sialosides were prepared in one-pot syntheses that made use of *N. meningitidis* and mammalian sialyltransferases (Figure 24(b)).¹⁹¹ The highly permissive *P. damsela* α 2-6 sialyltransferase has been used in combination with the *E. coli* aldolase and the *N. meningitidis* CMP-Sia synthetase to transform Man and ManNAc analogs into sialosides modified at C-5, C-7, C-8, or C-9.^{183,192} While most one-pot reactions make use of mixtures of purified enzymes, other strategies include the use of gene fusions, such as one between CMP-Sia synthetase and α 2-3 sialyltransferase,¹⁹³ or a mixture of crude *E. coli* extracts from overexpressing strains.¹⁹⁴

Once chemoenzymatically produced, these unnatural sialosides have demonstrated utility in a variety of biological applications. A sialoside consisting of 9-*O*-acetylated sialyl Tn antigen (sTn) was produced by one-pot synthesis and conjugated to human serum albumin to be used to evaluate the role of acetylation in carbohydrate-protein interactions.¹⁹⁵ *Para*-nitrophenol-tagged sialosides can be used in a colorimetric assay to evaluate sialidase specificity.¹⁹⁶ Biotin-tagged sialosides have been immobilized on a streptavidin-coated surface to study protein-carbohydrate interactions with surface plasmon resonance techniques.¹⁹⁷ Size-defined

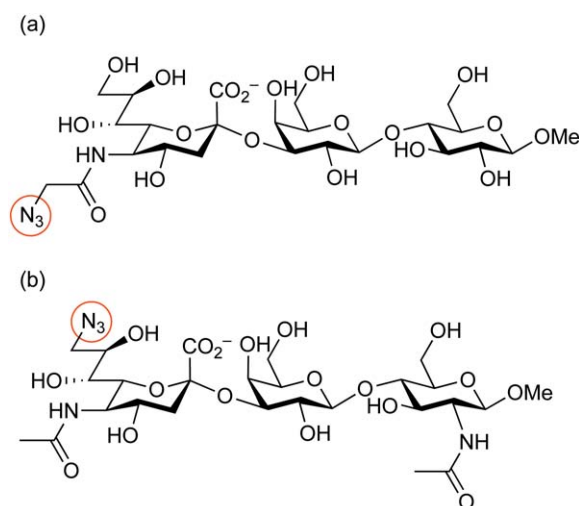


Figure 24 Examples of azide-modified sialosides prepared in one-pot syntheses. (a) A 9-azido-sialoside. (b) A 5-azido-sialoside.

polymers produced by a combination of chemoenzymatic methods and click chemistry have the potential to be used to study the role of Sia-containing polymers in bacterial pathogenesis and biofilm formation.¹⁹⁸ In addition to these representative examples, unnatural sialosides have the potential to be used to examine many other biological problems (see Chapter 6.02).

6.07.4 Applications of Metabolic Oligosaccharide Engineering

Metabolic oligosaccharide engineering refers to the ability to introduce small changes into cellular glycans through the use of unnatural monosaccharide analogs. To introduce these glycan modifications, synthetic analogs of monosaccharides are added to the media of cultured cells^{199,200} or injected into animals.^{200–202} Sugar analogs enter cells where they are metabolized to activated nucleotide sugar donors, then transferred to glycoconjugate substrates. Using this technique, a range of chemical modifications—extended alkyl chains, azides, ketones, thiols, and alkynes—have been introduced into Sia, fucose, GlcNAc, and GalNAc and displayed on cellular glycans.¹⁹⁹ The uses of this technology are numerous and are reviewed at further length in Chapter 6.09 entitled Molecular Probes for Protein Glycosylation, found in this volume. Here we have selected three emerging applications for metabolic oligosaccharide engineering for further discussion: the study of metabolic flux through carbohydrate assimilation pathways, elucidation of the neurobiological functions of PSA, and the development of immunological reagents for cancer therapy.

6.07.4.1 Metabolic Flux of Unnatural Sugar Analogs and Effects on Cell Viability

In experiments dating back to the 1970s, unnatural sugars have been used to inhibit or modulate the cellular synthesis of glycoconjugates (see Goon and Bertozzi²⁰³ for a review of early work in this field). More recently, the advent of sophisticated analytical techniques has made it possible to pinpoint the cellular fate of exogenously added monosaccharide analogs. As a result, the area of metabolic oligosaccharide engineering has enjoyed a renaissance that began with a 1992 report by Reutter and coworkers demonstrating that *N*-acyl-substituted mannosamines can be metabolized to the cell surface sialosides in a living rat.²⁰⁰ Further work established that cellular Sia engineering can be accomplished either by the use of Sia analogs, which the cell converts into CMP-Sia before transfer to glycoconjugates, or by intercepting the Sia biosynthetic pathway through the use of ManNAc analogs, which are first metabolized to Sia in three enzymatic steps (**Figure 25**).^{204,205} The growing use of this technique has motivated recent studies to better understand the consequences of treating cells or animals with these molecules. In this section we highlight the use of ManNAc analogs to study metabolic flux in the Sia pathway and describe some unanticipated effects that these molecules have on cellular metabolism.

Based on the observation that the Sia biosynthetic pathway is permissive for *N*-acyl substitutions, a variety of ManNAc analogs have been prepared (**Figure 26**). These molecules have been used to delineate the pathway's tolerance and to investigate analog-associated toxicity. Some ManNAc analogs bearing extended *N*-acyl chains can be effectively converted into their sialoside analogs,²⁰⁶ but side chains longer than five carbons are poorly metabolized.²⁰⁷ Furthermore, increasing *N*-acyl side chain length leads to decreased cell viability.²⁰⁸ The observed toxicity bears the hallmarks of an apoptotic process. The addition of substituents, such as ketones, to the *N*-acyl chain can also dramatically lower incorporation levels and increase toxicity (**Figure 27**). In contrast to ManNAc analogs, Sia analogs with large *N*-acyl substituents can be effectively metabolized into cell surface glycoconjugates.^{209,210} This suggests that ManNAc analog permissivity is limited by one of the three enzymatic steps responsible for converting ManNAc into Sia. In fact, phosphorylation of ManNAc by ManNAc-6-kinase has been identified as the bottleneck in metabolism of *N*-acyl-substituted ManNAc analogs.²⁰⁷

Many metabolic engineering experiments have utilized monosaccharide analogs that are acetylated on the free hydroxyl groups. These peracetylated compounds enter cells easily and are believed to be converted into free sugars by the action of nonspecific cellular esterases. As a result of their efficient uptake, peracetylated sugars can be used at concentrations nearly three orders of magnitude lower than their unprotected analogs.²¹¹ However, recent reports indicate that hydroxyl group modifications can have unanticipated effects on cell toxicity and metabolic flux.^{208,211–213} To further investigate this concept, Yarema and coworkers varied the size of the protecting group and measured the resulting changes in metabolic conversion and cell toxicity

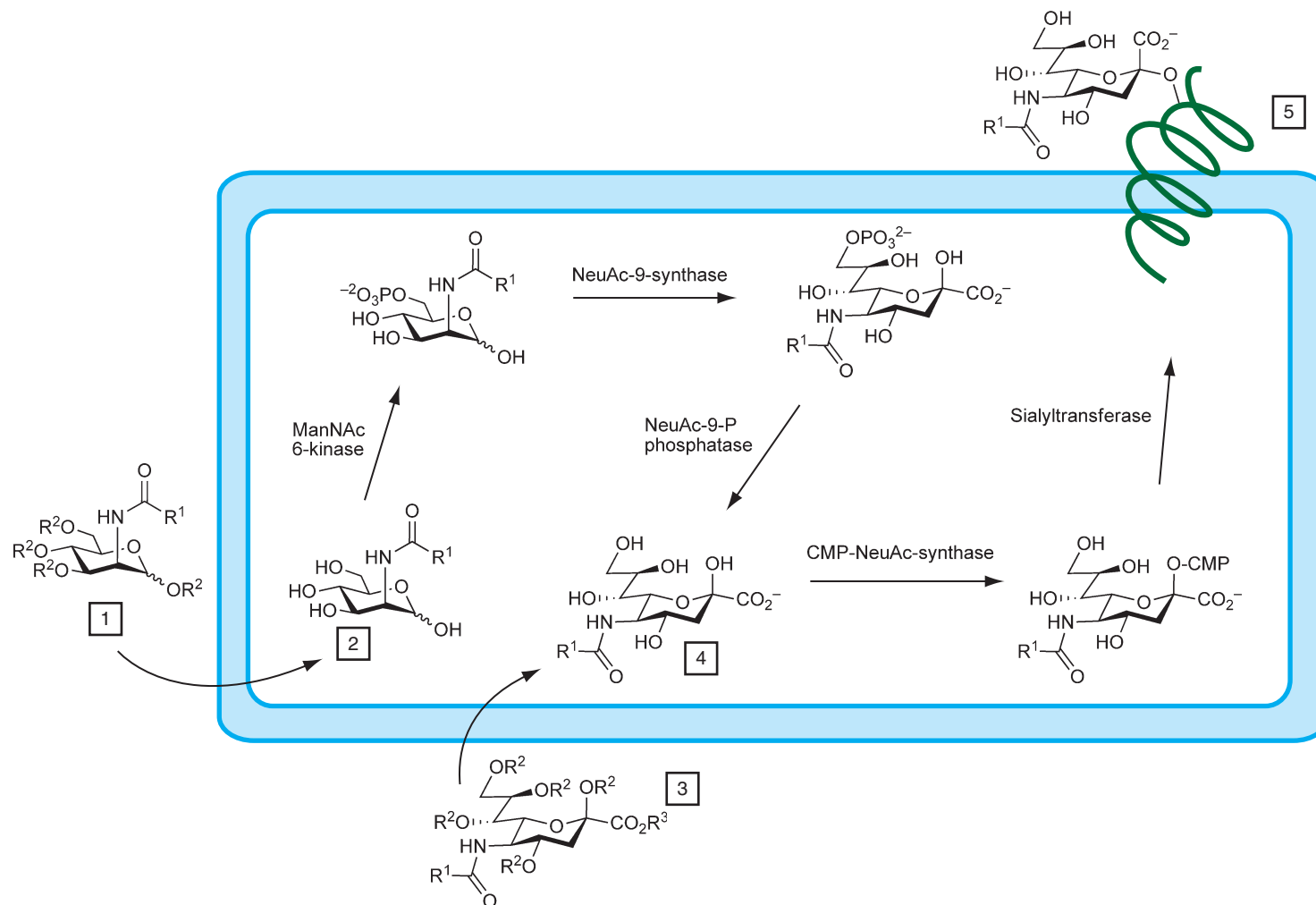


Figure 25 The sialic acid biosynthetic pathway. R^1 is $-\text{CH}_3$ in naturally occurring ManNAc and sialic acid, but a variety of analogs (see [Figure 26](#)) can also be incorporated through this pathway. R^2 and R^3 are both $-\text{H}$ in naturally occurring ManNAc and sialic acid, but compounds with protecting groups at these positions enter the cell more readily. Protected compounds commonly have an acetyl at R^2 and a $-\text{CH}_3$ at R^3 . ManNAc (1) and sialic acid (3) analogs enter the cell. If protecting groups are present, they are removed by nonspecific esterases to yield de-acylated sugars (2) and (4), respectively. Compound (2) is phosphorylated and converted into sialic acid (4) by a series of enzymatic transformations. The activated nucleotide sugar CMP-sialic acid is then produced by CMP-NeuAc-synthase. Finally, a sialyltransferase appends sialic acid (or an analog) to a growing glycan chain on a lipid or protein, yielding a sialoside (5). Compound (5) is then either inserted into the cell membrane or secreted.

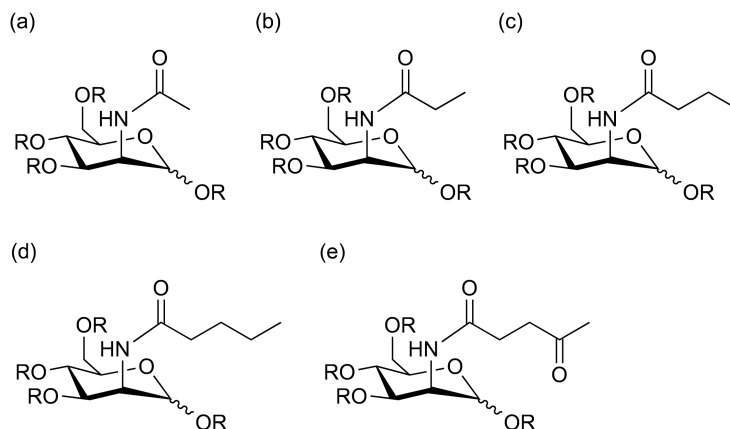


Figure 26 Structure of ManNAc and several common analogs used in sialic acid biosynthetic engineering (R = H or Ac): (a) ManNAc, (b) ManProp, (c) ManBut, (d) ManPent, (e) ManLev.

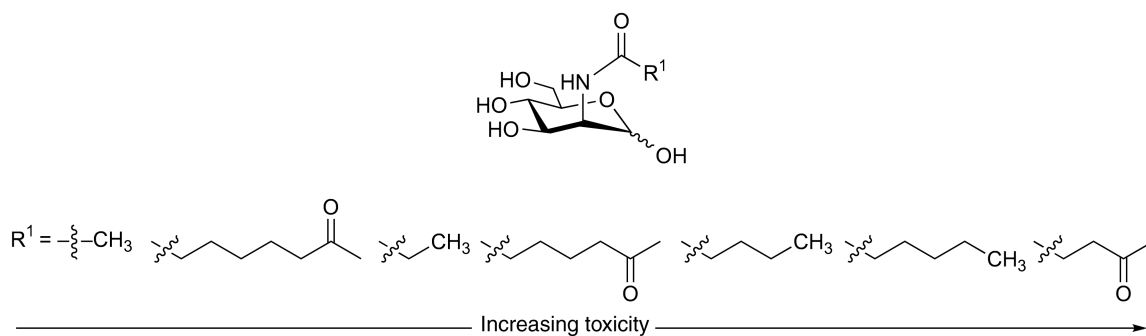


Figure 27 *N*-acyl modifications to ManNAc affect its cellular toxicity. Changes in chain length and the addition of ketones affect metabolic incorporation of the compounds as well as cell viability. The lethal dose for the analogs decreases from left to right.

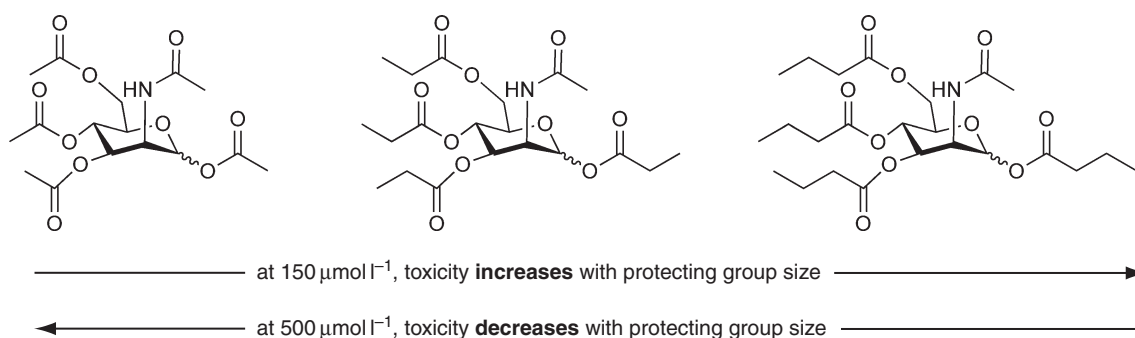


Figure 28 Increasing the size of ManNAc's protecting groups affects cellular viability. When the compounds are used at $150 \mu\text{mol l}^{-1}$, toxicity *increases* with protecting group size. When the compounds are used at $500 \mu\text{mol l}^{-1}$, toxicity *decreases* with protecting group size.

(**Figure 28**).²⁰⁸ When protected compounds were added to the cellular media at low concentrations ($150 \mu\text{mol l}^{-1}$), the addition of methylene units onto the protecting group improved the incorporation of Sia, with perbutyrylated ManNAc being most efficiently incorporated. This trend was reversed with a higher concentration of sugar analog ($500 \mu\text{mol l}^{-1}$), suggesting that multiple processes were at work. Cellular viability

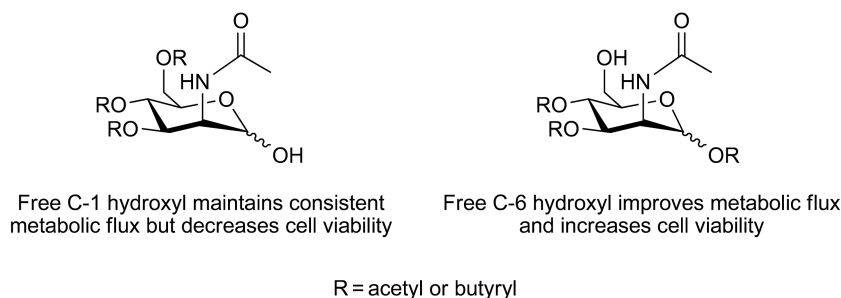


Figure 29 Protected ManNAc analogs with C-1 or C-6 free hydroxyls (R = acetyl or butyryl groups). A free hydroxyl at the C-1 position has little effect on metabolic flux but negatively impacts cell viability. A free hydroxyl at the C-6 position improves metabolic flux and increases cell viability.

also varied with larger protecting groups leading to increased toxicity. More toxic analogs not only failed to be incorporated into sialosides, but also inhibited overall flux through the Sia pathway.

Yarema and coworkers sought to deconvolute the complicated effects of ManNAc peracylation by investigating the roles of acyl groups placed at specific positions.²¹³ They prepared acylated ManNAc analogs with free hydroxyl groups at either C-1 or C-6 and measured their effects on metabolic flux through the Sia pathway and on cell growth (**Figure 29**). Molecules with a free hydroxyl group at C-1 demonstrated increased toxicity effects, while metabolic flux remained relatively unaffected. When the free hydroxyl was at the C-6 position, toxicity effects were relieved and metabolic flux improved. Similar effects were observed with protected GlcNAc molecules. These experiments provide a strategic guide for maximizing metabolic incorporation of analogs and also offer a cautionary message about the off-target effects that sugar analogs can cause.

Toxicity effects have been associated with metabolic oligosaccharide engineering since the earliest experiments were performed.²⁰³ In some instances, these effects may be mediated by the sugar analogs themselves, but removable protecting groups also have the potential to exhibit bioactivity. Recently, Yarema and coworkers investigated the effects that perbutyrylated ManNAc has on metabolic flux and cell viability.²¹² Short-chain fatty acids, such as the *n*-butyrate esters found on perbutyrylated ManNAc, can act as histone deacetylase (HDAC) inhibitors, inducing apoptosis in transformed cells. However, *n*-butyrate suffers from poor pharmacokinetic properties and high concentrations (up to 50 mmol l^{-1}) are usually required for activity. Sampathkumar *et al.* found that $80 \mu\text{mol l}^{-1}$ of perbutyrylated ManNAc was sufficient to cause cell death. Moreover, the cells displayed many characteristics commonly seen during apoptosis. Notably, neither separate administration of *n*-butyrate and ManNAc, nor treatment with other butyryl protected sugars was able to reproduce the toxicity effects observed with perbutyrylated ManNAc. These results point to a potential use of protected sugars as prodrugs for the delivery of short-chain fatty acids in cells.

Metabolic oligosaccharide engineering has also been exploited to obtain information about metabolic pathways that would be difficult to discern using other techniques. For example, Luchansky *et al.*²¹⁴ made use of ManNAc analogs to show GlcNAc 2-epimerase, which interconverts ManNAc and GlcNAc, functions both anabolically and catabolically, and its primary role may be to divert sugars from the Sia biosynthesis pathway. In another example, Yarema *et al.*²¹⁵ used a ManNAc analog, ManLev (*N*-levulinolmannosamine), to select for T-cell lines that displayed phenotypes of low or high cell surface expression of Sia analogs. Cells that expressed low levels of the Sia analog were found to have accumulated mutations in the UDP-GlcNAc 2-epimerase genes that were remarkably similar to those found in patients with sialuria, a severe congenital glycosylation disorder. Cells expressing high levels of Sia analogs had activated neo-expression of $\alpha 2$ -8-linked PSA, a change often observed in highly invasive tumors. In this way, metabolic oligosaccharide engineering might be used to develop model systems that phenocopy naturally occurring disorders and diseases.

6.07.4.2 Polysialic Acid Function

Metabolic oligosaccharide engineering has also been exploited for the study of PSA. PSA is a linear homopolymer of $\alpha 2$ -8-linked Sia residues that can extend to over 50 residues (**Figure 30**).^{216,217} PSA is synthesized

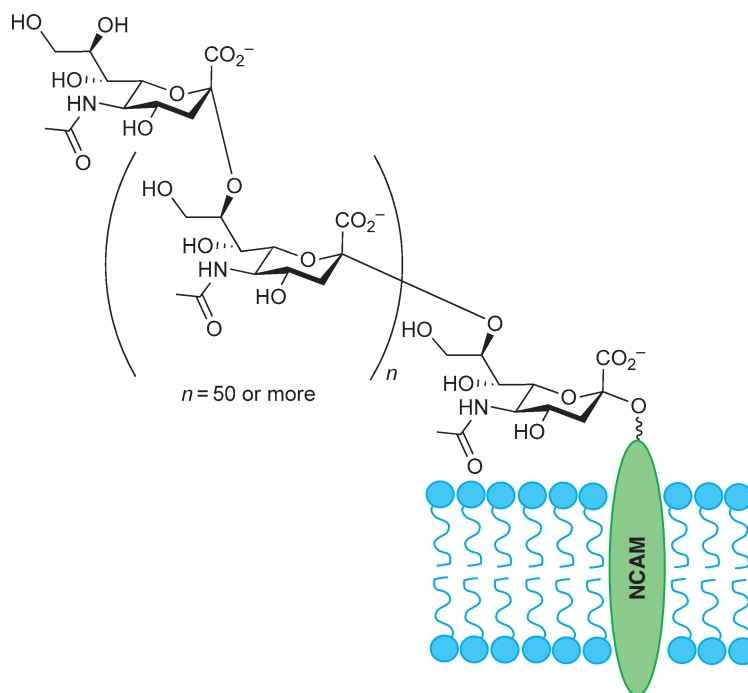


Figure 30 PSA chains are linear homopolymers of α 2–8-linked sialic acid residues, extending to over 50 residues. They are attached to the NCAM polypeptide.

by two α 2–8 sialyltransferases, PST (ST8Sia4 polysialyltransferase) and STX (ST8Sia2 polysialyltransferase), and is displayed exclusively on the neural cell adhesion molecule (NCAM). Normal expression of polysialylated NCAM (PSA-NCAM) occurs primarily on neuronal cells, where it plays essential roles in early development and in maintaining synaptic plasticity in adults. Abnormal expression of PSA-NCAM in non-neuronal cells is a characteristic of several human cancers. Metabolic oligosaccharide engineering provides an avenue for studying the function of PSA and the substrate specificity of the two sialyltransferases involved in its synthesis. In addition, metabolically engineered PSA can be recognized *in vivo* using antibodies that target the modified polymer.

Metabolic oligosaccharide engineering of PSA has been attempted with four different ManNAc analogs: ManLev, ManProp (*N*-propanoylmannosamine), ManBut (*N*-butanoylmannosamine), and ManPent (*N*-pentanoylmannosamine) (Figure 30). In 2000, Bertozzi and coworkers reported that NT2 neuronal cells cultured with ManLev produced PSA chains containing SiaLev (5-*N*-levulinoylneuraminic acid).²¹⁸ In these experiments, overall PSA expression levels were unaffected, but the modified *N*-acyl chains rendered unnatural PSA molecules resistant to sialidase cleavage. In analogous experiments, using the ManProp and ManBut precursors, Mahal *et al.*²¹⁹ found that *N*-propanoyl Sia was incorporated efficiently into PSA polymers, while ManBut treatment resulted in premature PSA chain termination. Furthermore, ManBut-treated neurons exhibited decreased neurite outgrowth, consistent with inhibition of PSA synthesis.²²⁰ More recently, two additional groups have examined the effects of unnatural *N*-acyl mannosamines on PSA synthesis, with divergent results. Horstkorte *et al.*²²¹ reported that ManProp, ManBut, and ManPent all inhibited PSA synthesis in NT2 cells. Confoundingly, Jennings and colleagues found that ManProp and ManBut were both effectively metabolized into Sia analogs that were incorporated into PSA displayed on NT2 neurons.²²² Furthermore, incorporation results varied when other cell lines were examined.

These three groups used different analytical techniques and reagents to evaluate unnatural sugar incorporation in PSA. This makes a direct comparison of results difficult. Nonetheless, conflicting observations about PSA incorporation may result from a variety of sources. Variable cell culture conditions may contribute to the discrepancies observed: Jennings and coworkers noted that ManNAc analogs were effectively incorporated in immature cells, but more mature neurons displayed limited *de novo* PSA synthesis.²²² In addition, the two polysialyltransferases, PST and STX, may be unequally capable of utilizing unnatural precursors. *In vitro* experiments conducted by Mahal *et al.* suggest that PST readily accepts unnatural sugars, while STX is less promiscuous. Similarly, Horstkorte *et al.*²²¹ found that cells expressing PST readily incorporated unnatural *N*-acyl mannosamines into PSA, while treatment with these analogs inhibits PSA production in cells expressing STX. Finally, unnatural *N*-acyl mannosamines may exhibit as-yet uncharacterized off-target effects (e.g., on growth or maturation rates) that affect NCAM or PSA production through other mechanisms.

Regardless of the disparities noted above, *N*-acyl modifications to PSA chains do not appear to alter the adhesive properties of cells. Incorporation of *N*-propanoyl sugars did not affect neurite outgrowth in neurons derived from chick dorsal root ganglia,²²⁰ nor did modifications affect induced neuronal clustering in rat tissue culture.²²³ These results suggest that ManNAc analogs could be used to specifically target PSA-expressing cells, while having little effect on normal neuronal cell behavior, including outgrowth and clustering.

In fact, several groups have used ManProp in the development of prototype therapeutic reagents. Gagiannis *et al.*²²⁴ reported the production of *N*-propionylated PSA in the brains of mice treated with peracetylated ManProp by intraperitoneal injection, indicating that peracetylated ManProp has some ability to cross the blood–brain barrier. These mice also exhibited a dramatic reduction in PSA levels, but the mechanism behind this change remains to be elucidated. In another example, Jennings and coworkers reported a method to immunotarget tumor cells using a therapy that combines sugar analog and antibody administration. Using an antibody, mAb 13D9, that specifically recognizes *N*-propionylated PSA,²²⁵ these investigators demonstrated selective targeting and killing of cells expressing the modified PSA.²²⁶ Furthermore, mice harboring PSA-expressing tumor cells showed decreased tumor growth following coadministration of ManProp and mAb 13D9.²²⁶ Although this report is preliminary, a small molecule-dependent immunotherapy could provide a powerful strategy for controlling metastases (see Chapter 6.09).

6.07.4.3 Metabolically Engineered Gangliosides for Cancer Immunotherapy

Changes in cell surface glycosylation are a hallmark of cancer. Specifically, all tumor cells exhibit altered ganglioside expression, commonly including overexpression of GM1, GM2, GM3, GD2, or GD3. Only low levels of these molecules can be detected in normal cells. These distinct changes in ganglioside expression offer a potential target for cancer therapeutics.^{227,228} As described in Section 6.07.2.1, ganglioside-based reagents have been explored as potential cancer vaccines, but can suffer from poor immunogenicity and have the potential to induce autoimmune effects. Alternatively, humanized antibodies that specifically recognize gangliosides have been used in passive immunotherapy, but challenges remain with this therapeutic approach. Another strategy takes advantage of oligosaccharide metabolic engineering to introduce slight structural changes into gangliosides. Even small changes can be sufficient to make the modified ganglioside appear foreign and immunogenic. In addition to improving a ganglioside's immunogenicity, a metabolic engineering approach ensures that the antiganglioside immune response will be active only when a modified sugar precursor is administered. In this way, the risk of an autoimmune response is minimized. In this section, we describe the use of metabolic engineering to convert GD3 and GM3 into effective immunotargets for cancer therapy.

GD3 expression is a marker for several types of human cancer. Anti-GD3 immunotherapies have shown promise in patients with metastatic melanoma and small cell lung cancer. However, not all treated individuals mount an anti-GD3 immune response.^{229,230} To attempt to improve GD3 immunogenicity, Jennings and coworkers have used metabolic engineering to introduce changes into the two Sia residues found in this molecule.²³¹ These researchers synthesized GD3 analogs that contained Sia residues bearing *N*-propionyl, *N*-butyryl, or *N*-benzoyl groups in place of the normally occurring *N*-acetyl. The GD3 analogs were conjugated to

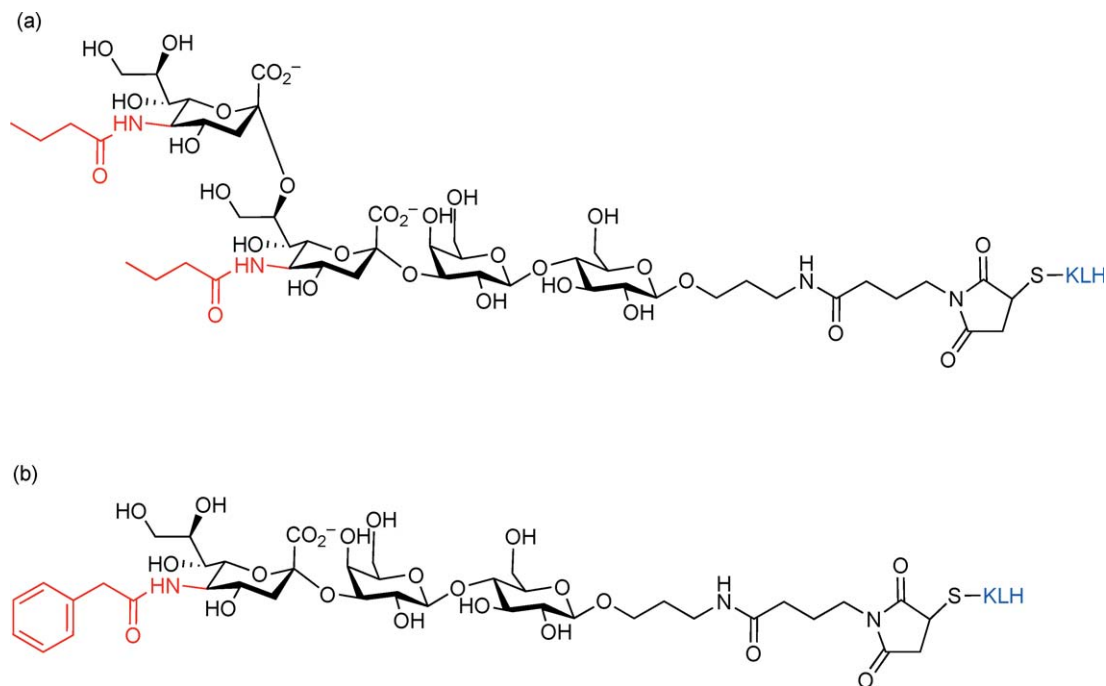


Figure 31 Structures of immunogenic ganglioside analogs. (a) Structure of GD3Bu-KLH (butyryl groups in red, KLH in blue). (b) Structure of GM3Pac-KLH (Pac in red, KLH in blue).

KLH to enhance their immunogenicity. Mice were injected with the conjugates and the resulting antisera were harvested for analysis. The *N*-butyryl analog (*N*-butanoyl GD3 – GD3Bu, **Figure 31(a)**) generated the strongest response and, as desired, the anti-GD3Bu serum did not cross-react with unmodified GD3. Jennings and coworkers also isolated a monoclonal antibody (mAb), 2A, that specifically recognizes GD3Bu. These investigators demonstrated that this antibody can specifically target melanoma cells that had been cultured with the *N*-butyrylated ManNAc precursor, ManBut (**Figure 26(c)**). More remarkably, mice harboring SK-MEL-28-derived tumors demonstrated suppressed tumor growth when treated with combination therapy comprised of the anti-GD3Bu antibody and ManBut. Unfortunately, the treatment did not eliminate established tumors. Nonetheless, these experiments provide a strong precedent for future work utilizing metabolic engineering in cancer immunotherapy.

Guo and coworkers employed a similar approach for development of an immunotherapy directed against another ganglioside, GM3.^{232–234} GM3 is a monosialylated ganglioside that is overexpressed in a variety of cancers, including breast cancer and melanoma. These researchers synthesized GM3-KLH analogs bearing *N*-propionyl, *N*-butyryl, *N*-isobutyryl, or *N*-phenylacetyl-modified Sia, injected these compounds into mice, and harvested the resulting antisera. Using an ELISA (enzyme-linked immunosorbent assay), they discovered that the antiserum against *N*-phenylacetyl GM3 contained the highest level of antigen-reactive antibodies and the lowest level of cross-reactivity toward natural GM3 (**Figure 31(b)**).²³³ These investigators then showed that this antiserum could specifically recognize melanoma cells that had been cultured with *N*-phenylacetyl-mannosamine), the metabolic precursor of *N*-phenylacetyl Sia. These results indicate that the cells were producing *N*-phenylacetyl GM3.²³² Strikingly, ManNPhAc was efficiently incorporated, even at micromolar concentrations, which bodes well for the use of this analog in metabolic labeling conducted *in vivo*. More recently, Guo and coworkers reported a mAb that specifically recognizes *N*-phenylacetyl GM3.²³⁴ This antibody has been shown to mediate ManNPhAc-dependent cytotoxicity. Most notably, a melanoma cell line expressing high levels of GM3 was found to be highly sensitive to the antibody, while an embryonic cell line expressing low levels of GM3 required a 200-fold higher antibody concentration to display the same cytotoxic effects. These results suggest that immunotherapy directed toward metabolically engineered gangliosides could provide an effective treatment with minimal toxic side effects.

6.07.5 Covalently Trapping Glycan Interactions with Photocross-linkers

Photoaffinity reagents are useful tools for probing molecular interactions that occur in complex mixtures. Latently reactive functional groups can be incorporated into biological reaction mixtures in a test tube, a cell, or an organism. Upon irradiation with light of an appropriate wavelength, the photoactive groups become highly reactive, forming new covalent bonds with nearby molecules. By isolating and characterizing these covalent adducts, one can obtain information about the environment surrounding the photoactive probe. Photoactive probes offer particular promise in the study of glycan-mediated interactions because of the transient nature of these binding events. Carbohydrate-dependent binding events are typically low affinity, with high micromolar or even millimolar equilibrium binding constants. As a result, the macromolecular complexes generally dissociate rapidly and do not survive the purification steps necessary for analysis. By covalently capturing these low-affinity complexes, one can obtain important information about ephemeral, yet essential, binding events.

Commonly used photoreactive cross-linkers with demonstrated utility in biological settings include benzophenones, aryl azides (AAzs), and diazirines (**Figure 32**).^{235,236} Benzophenones offer the advantage of cross-linking through a reactive diradical. If cross-linking does not occur, the diradical can relax back to the starting material and be reactivated, leading to highly efficient cross-linking. Photoactivation of AAzs results in the formation of a reactive nitrene, with concomitant loss of N₂. The nitrene rapidly inserts in neighboring C–H, N–H, or O–H bonds, forming covalent cross-links. Along with good photophysical properties, AAzs are advantageous because they can participate in chemoselective reactions, such as the Staudinger ligation²³⁷ or azide–alkyne cycloaddition,²³⁸ which can be exploited to track the incorporation of these photocross-linking sugars. A third class of cross-linkers are the diazirines. These functional groups react by losing N₂ to form a highly reactive carbene. Carbenes are nonselective cross-linkers and can insert in nearby C–H, N–H, or O–H bonds. Two types of diazirines are frequently used: alkyl diazirines, which have the advantage of small size, and trifluoromethyl phenyl diazirines, which are less susceptible to nonproductive rearrangement to diazo species. Recently, the use of a diazocyclopentadien-3-yl-carbonyl functional group to generate reactive carbenes has also been reported.²³⁹ Each cross-linker has distinct advantages and disadvantages, so the optimal choice is likely to be application-specific.

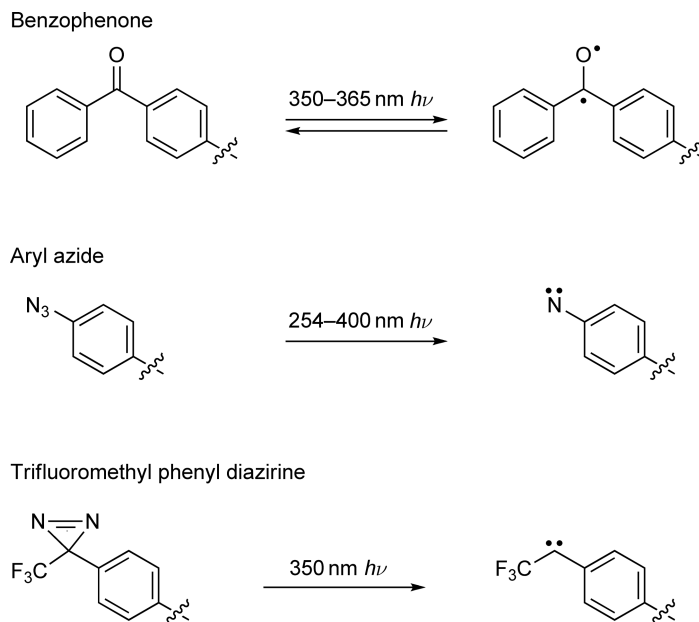


Figure 32 Structures of three common photocross-linking moieties. Benzophenone is photoactivated to a reactive diradical that can relax back to the starting material. Aryl azides and diazirines both irreversibly lose N₂ to react through nitrene and carbene species, respectively.

6.07.5.1 Sugar Probes for Lectin Characterization and Discovery

In important early work, Lee and Lee used photoaffinity probes to map sugar-binding sites on the asialoglycoprotein receptor, a Gal/GalNAc-specific lectin found in mammalian liver. The asialoglycoprotein receptor comprises multiple subunits, which reside in hepatocyte membranes and bind desialylated serum glycoproteins. Lee and coworkers prepared photoaffinity probes semisynthetically, beginning with a desialylated triantennary glycopeptide. This glycopeptide was isolated from bovine fetuin and was known to directly interact with the receptor through three terminal Gal residues. The C-6 positions of the Gal residues were selectively modified with AAzs and the peptide was ^{125}I -labeled through tyrosylation followed by radioiodination (**Figure 33**).²⁴⁰ Using this probe, these investigators discovered multiple Gal-binding sites on different subunits.^{240,241} To investigate the contributions of each Gal residue in the triantennary structure, the researchers used glycopeptides in which individual Gal residues were selectively modified with AAzs. In this way, they were able to demonstrate that two of the terminal residues were responsible for binding to the major subunit, while the third bound to the minor subunit.^{242,243} These results established that a precise geometry is required for binding between the triantennary glycopeptide and the asialoglycoprotein receptor, and provided some of the only information available regarding the supramolecular structure of this complex. Among the other early uses of photocross-linking glycan probes, another notable success was the use of an AAz-modified UDP-Gal analog to identify the donor-binding site in a β -1,4-galactosyltransferase.²⁴⁴

While experiments to define the Gal-binding sites on the asialoglycoprotein receptor exploited the natural glycan ligand, other groups have prepared multivalent ligands synthetically and modified these probes with photocross-linkers. For example, Hashimoto *et al.*²⁴⁵ reported the synthesis of bis-glucose derivatives tagged with diazirine and their utility in labeling Glc transporters. Similarly, the Lindhorst group synthesized di- and trivalent mannoside probes bearing the diazirine cross-linker and anticipates using them to define the Man-binding sites on FimH, which is found in the type I fimbriae of pathogenic *E. coli*.²⁴⁶ The multivalency of these probes is expected to improve their affinity for their protein-binding partners. One can also imagine using multivalent molecules with differing linker lengths to glean information about the spacing of carbohydrate-binding sites.

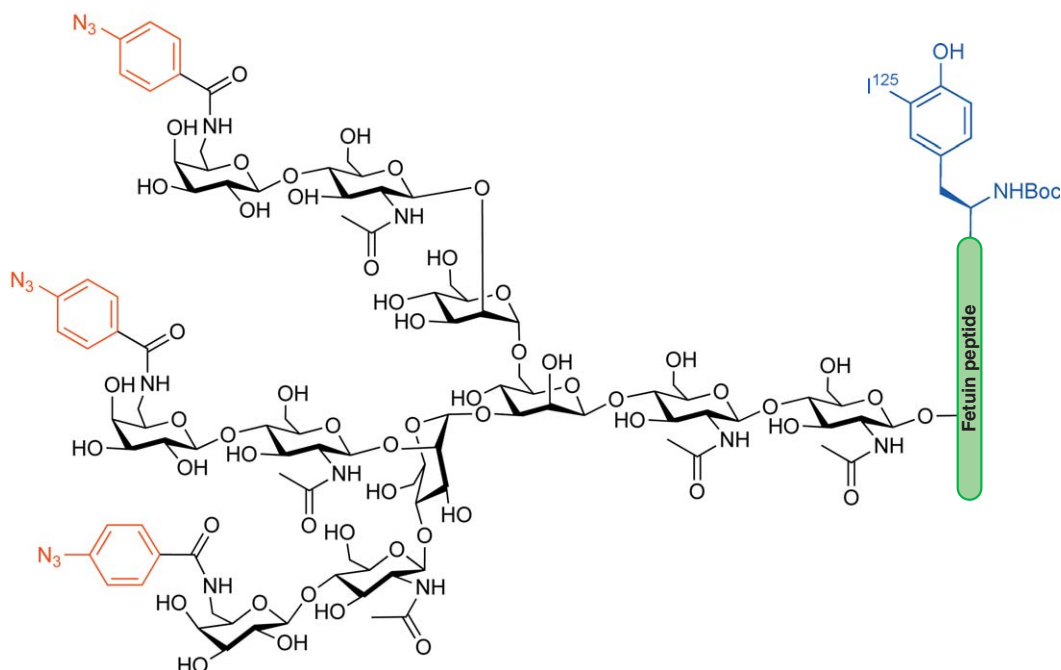


Figure 33 An early glycan photoaffinity probe. This molecule was used to map galactose-binding sites on the asialoglycoprotein receptor. The probe was prepared from a desialylated fetuin glycopeptide: galactose residues were modified with aryl azides (shown in red) and the peptide was labeled by radioiodination (shown in blue).

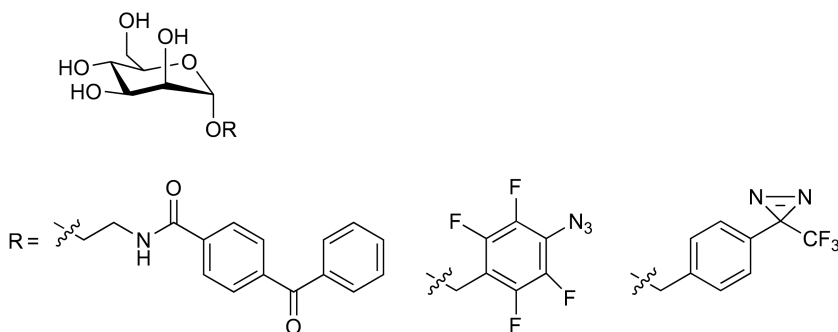


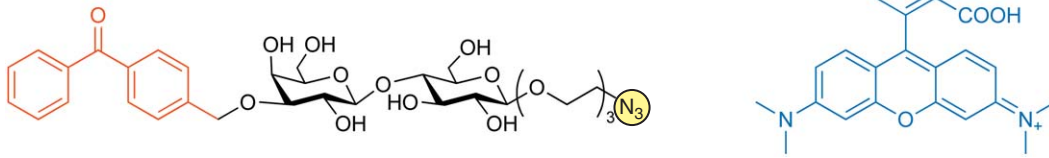
Figure 34 Panel of mannose-based photocross-linkers. The cross-linking activity of photoactive α -mannosides was compared in a simple model system. The trifluoromethyl phenyl diazine yielded the most effective cross-linking to a variety of amino acids.

A key component of all photoreactive carbohydrates is the photocross-linking functional group. As mentioned above, the ideal cross-linker will vary for different applications and cannot be predicted *a priori*. For this reason, it may be advantageous to synthesize probes in a manner that allows for facile introduction of a variety of photoreactive groups. Lindhorst and coworkers employed such an approach when they prepared mannosyl probes designed to target lectin domains. These investigators functionalized the anomeric position of Man with either a benzophenone, a tetrafluorinated AAz, or with a trifluoromethyl phenyl diazine group (Figure 34).²⁴⁷ All three reagents were capable of cross-linking to small peptides *in vitro*, but the most efficient photolabeling was observed with the diazine probe, suggesting that this molecule would be best suited for future experiments aimed at mapping Man-binding sites. In this case, the superior performance of the diazine was attributed to its relatively long wavelength for photoexcitation, good solubility, and preference for insertion into amino and hydroxide groups. Others have noted that diazirines produce more stable adducts than other cross-linkers.²⁴⁸

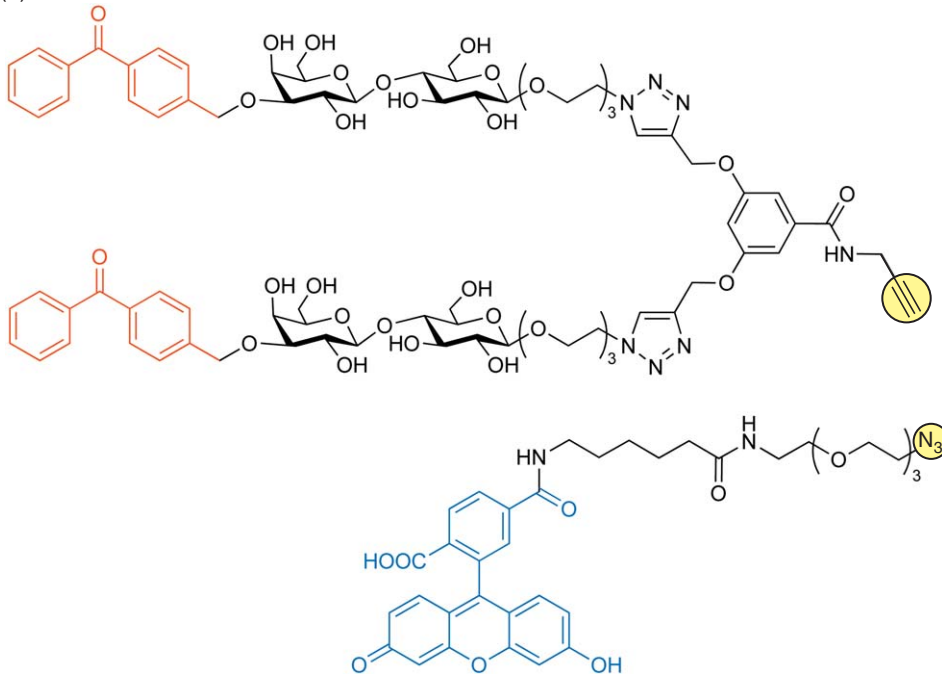
Once glycoconjugate complexes are photocross-linked, it is often desirable to purify these complexes for further analysis. A reagent that specifically recognizes one of the binding partners can be used to purify the cross-linked complex. Alternatively, one might wish to introduce a purification tag into the cross-linked complex. Although not widely used, the 5-azidonaphthalene-1-sulfonyl cross-linker offers the advantage of functioning both as a photocross-linking agent and as a fluorescent label. Pathak *et al.*^{249,250} prepared a series of disaccharides and tagged each with the 5-azidonaphthalene-1-sulfonyl group. The molecules function as substrates for mycobacterial glycosyltransferases and could potentially be used to fluorescently tag these enzymes. Another approach involves incorporating separate photocross-linkers and fluorophores into glycan probes. For example, Pieters and coworkers prepared a lactose reagent that includes a benzophenone at the C-6 position of Gal and an azide attached to the reducing end of the disaccharide through an ethylene glycol linker (Figure 35(a)).²⁵¹ Irradiation leads to covalent attachment of the probe and a protein-binding partner. Following cross-linking, the azide can engage in a chemoselective reaction with an alkyne-functionalized rhodamine tag, thereby fluorescently labeling the protein. In a proof-of-principle experiment conducted in a small protein mixture, rhodamine labeling of two lactose-binding galectins was demonstrated. A multivalent probe of similar design demonstrated improved specificity and improved detection levels (Figure 35(b)).²⁵² This new probe selectively labeled galectin-3 with fluorescein in the presence of other carbohydrate-binding proteins (CBPs) and in bacterial and mammalian cell lysates. Similarly, Shin and coworkers prepared monovalent and trivalent probes comprising sugars (either Man or fucose), a benzophenone cross-linker, and a fluorophore (Figure 35(c)).²⁵³ These reagents were used to selectively label known lectins even in the presence of competing *E. coli* proteins. The same molecules may also find utility in protein microarray analysis. Shin and coworkers covalently labeled a microarray surface with 20 proteins, a number of which had known glycan-binding properties. The microarray surface was interrogated with trivalent glycan probes functionalized with benzophenone and Cy3.²⁵⁴ Visualization of known protein-glycan interactions was greatly enhanced by irradiation, suggesting that cross-linking by the benzophenone aided in the capture of transient, yet specific binding events.

Other multifunctional probes incorporate both a photocross-linker and a purification tag. These reagents have the potential to be used to identify the portion of the lectin that functions in glycan recognition. For example,

(a)



(b)



(c)

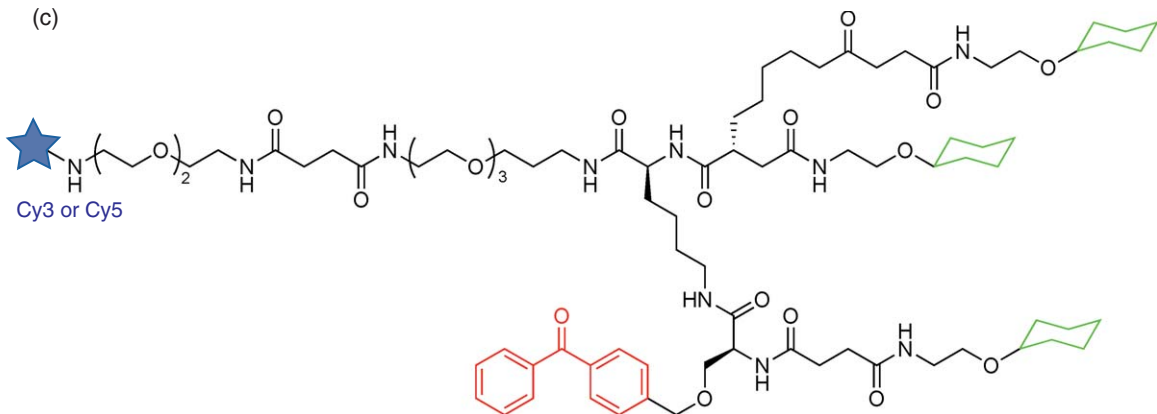


Figure 35 Glycan photocross-linking probes for fluorescent labeling. Photocross-linkers are shown in red and fluorophores are in blue. (a and b) Lactose molecules modified with a benzophenone cross-linker and a functional group that can react through ‘click’ chemistry. After photocross-linking, the adducts are reacted with an azide or alkyne-functionalized fluorophore to fluorescently label the protein-binding partner. (c) Multivalent mannose or fucose (represented in green) probes bear a benzophenone cross-linker and a Cy3 or Cy5 fluorescent label.

information about substrate-binding sites in bovine β 1,4-galactosyltransferase has been obtained through the use of a GlcNAc probe modified with a diazirine cross-linker and a biotin affinity tag.²⁵⁵ Additional examples of affinity-tagged photocross-linking probes have been reported: the Lee group's latest designs uses an AAz for cross-linking and a digoxigenin tag for detection or purification with commercially available antibodies (**Figure 36(a)**),²⁵⁶ Lindhorst and coworkers prepared mannosides that bear both a benzophenone cross-linker and a biotin affinity tag (**Figure 36(b)**)²⁴⁷ and Lee *et al.*²⁵³ have reported trivalent sugar probes that incorporate benzophenone and biotin. Probes containing more complex glycan structures have been prepared by Hatanaka *et al.*²⁵⁷ in 2000, this group reported the use of a biotinyl diazirine reagent (AffiLight-CHO) to label the reducing end of unprotected di-, tri-, and tetrasaccharides (**Figure 36(c)**). The same reagent has been used to modify large naturally occurring oligosaccharides.²⁵⁸ In one application of this technology, the resulting probes were used to covalently tag and identify the carbohydrate-binding subunit of sophoragin, a heterotetrameric protein found in the bark of the Japanese pagoda tree.²⁵⁹

Affinity-tagged, photocross-linking glycan probes also have the potential to be used for *de novo* identification of carbohydrate-binding probes. In an early use of this technology, Lauc *et al.*²⁶⁰ identified a novel Glc-binding lectin, CBP33, which is a stress-induced protein expressed in rat liver nuclei. To conduct these experiments,

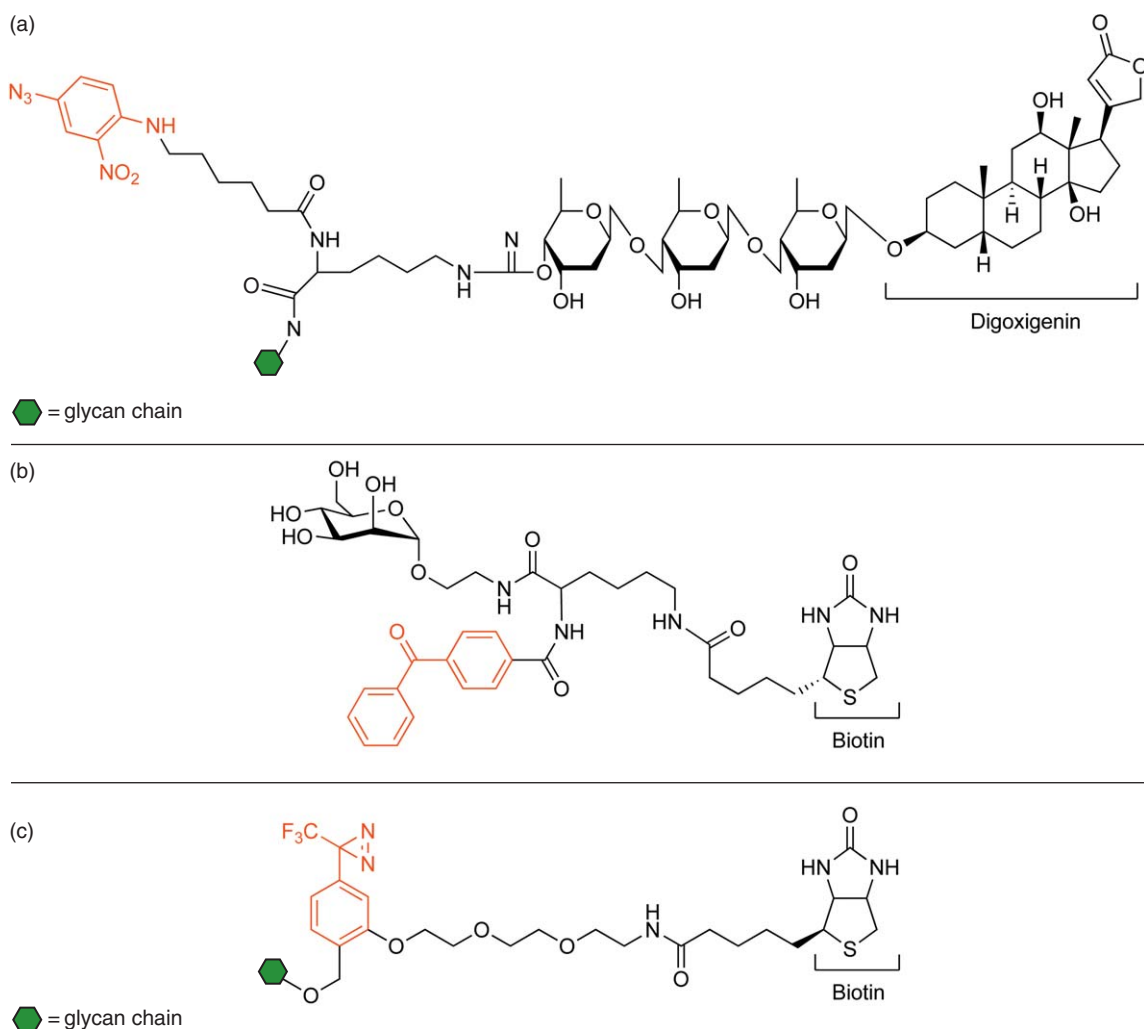


Figure 36 Glycan photocross-linking probes with purification tags. Photocross-linkers are shown in red. (a) Glycan probe utilizing nitrophenyl azide and a digoxigenin tag. (b) Mannoside-based probe bearing a benzophenone and a biotin tag. (c) AffiLight-CHO labeling reagent.

the investigators modified α -D-glucose at the anomeric position with a lysyl-lysine backbone bearing both an AAZ and digoxigenin tag. Using this probe and an antidigoxigenin antibody, the researchers were able to collect enough intact protein to generate antibodies and purify larger samples for characterization.

A significant advance for this class of reagents was the introduction of a cleavable linker between the photocross-linker and the affinity tag. This feature allows for facile release of the cross-linked peptide or protein and is especially important when the essentially irreversible biotin-streptavidin interaction is used for purification. For example, the Hatanaka group recently reported a reagent that incorporates diazirine-modified LacNAc (*N*-acetylglucosamine) and biotin on either side of an acylsulfonamide (Figure 37(a)).²⁶¹ After photocross-linking to a LacNAc-binding lectin, the cross-linked adduct was purified using immobilized streptavidin. The cross-linked protein was then released by cleavage of the acylsulfonamide with a mild base in the presence of ammonia. Other types of cleavable linkers can also be used in this manner, including a simple disulfide. Borén and coworkers demonstrated the use of a glycan probe to discover a key adhesive lectin of *H. pylori*, the causative agent of peptic ulcer disease.²⁶² *H. pylori* binds the Lewis b blood group antigen (Le^b). This interaction was known to be essential for colonization of human gastric mucosa but prior to the work by Ilver *et al.*²⁶² in 1998, the *H. pylori* lectin that recognized Le^b remained unidentified. To identify the Le^b -binding protein, this group prepared a probe in which albumin was directly conjugated to Le^b glycan and also linked to a dual aryl-azide/biotin reagent through a disulfide linkage (Figure 37(b)). After incubating the probe with *H. pylori* culture and performing UV irradiation, the researchers were able to isolate the Le^b -binding lectin using streptavidin-coated beads. Using this purification method, they obtained enough of this low abundance protein to sequence its N-terminus and identify the encoding gene. This technique, termed 'ReTagging' by its practitioners, has the potential for general use in the identification of glycan-recognizing proteins.

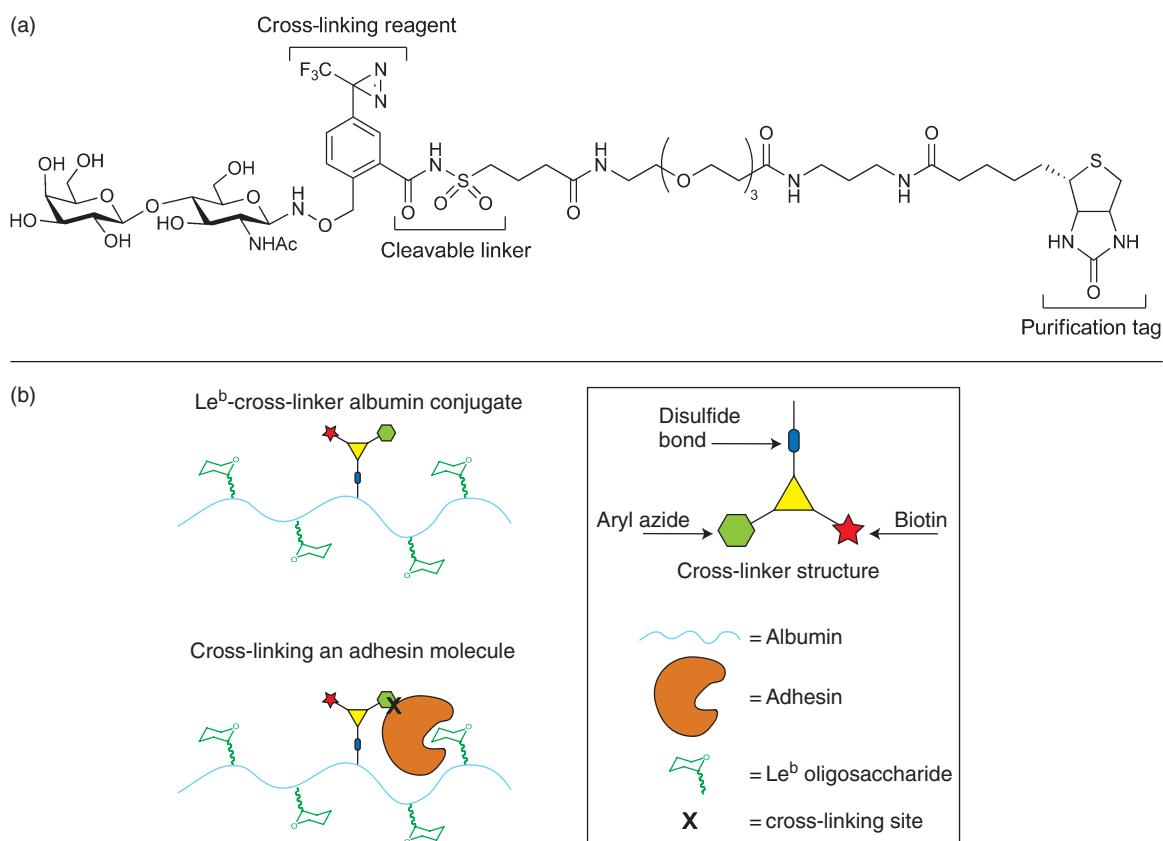


Figure 37 Bifunctional photocross-linking affinity purification reagents with cleavable linkers. (a) Diazirine-modified LacNAc probe with a biotin tag and a cleavable acylsulfonamide linker. (b) Aryl azide Le^b probe with a biotin tag that contains a cleavable disulfide linker.

6.07.5.2 Cross-linking Glycolipids

Glycolipids play integral roles in many cellular processes, including cell growth and differentiation, cell adhesion and recognition, and signal transduction. Despite the essential functions of these molecules, relatively little structural detail is known about the interactions in which glycolipids engage. Photocross-linkers offer a potentially powerful approach for capturing glycolipid–protein complexes and mapping the binding surfaces.²⁶³ Photoreactive groups have been incorporated in either the carbohydrate moiety, to detect glycan-based interactions, or in the fatty acid chains, to detect lipid-based interactions. Here we present a variety of glycolipid-based photocross-linking probes and describe how they have been used to characterize glycolipid–protein interactions.

Many pathogenic proteins recognize or invade host cells through interaction with cell surface glycolipids. For example, cholera toxin specifically binds cells displaying the ganglioside GM1. Following binding, the toxin is internalized and causes changes in ADP ribosylation that, in turn, lead to the dehydration that characterizes cholera infection. To study GM1 binding by cholera toxin, Pacuszka and Fishman developed a photoaffinity probe.²⁶⁴ They liberated the GM1 oligosaccharide from the glycolipid and performed reductive amination on the terminal Glc molecule to generate an amine at the former anomeric position. This position was functionalized with an aryl-azide photocross-linker and an ¹²⁵I radiolabel (**Figure 38**). Using this probe, the investigators demonstrated specific cross-linking to the B subunit of cholera toxin, which is known to recognize GM1, and not to the A subunit. This type of probe design has also been successful in detecting ganglioside interaction partners on the surface of S20Y murine neuroblastoma cells.²⁶⁵

A similar approach revealed even greater detail about the tetanus toxin–ganglioside interaction. Tetanus toxin binds specifically to the GD1b ganglioside, this recognition event is believed to be essential for the toxin's entry into host motor neurons. To identify the ganglioside recognition domains of tetanus toxin, Schnaar and coworkers used a GD1b-based photoaffinity probe.²⁶⁶ Beginning with naturally occurring GD1b, they attached an aryl-azide group radiolabeled with ¹²⁵I to the C-7 position of Sia through a cleavable disulfide linker (**Figure 39**). Photocross-linking was induced by irradiating the probe–toxin complex, followed by enzymatic digestion of the protein. MALDI–MS (matrix-assisted laser desorption–mass spectrometry) analysis revealed that the probe had cross-linked to a 34 amino acid C-terminal peptide, which was found to be covalently modified on a single residue, H1293. The location of the Sia-binding site was later confirmed by cocrystal structures of the toxin bound to lactose and to a GT1b analog.^{267,268}

Many ganglioside–protein-binding events are mediated by hydrophobic interactions between the fatty acid chains of gangliosides and hydrophobic residues of transmembrane proteins. Gangliosides that have been modified with photoreactive groups in the fatty acid chains have proven useful for the study of these interactions. Toward this end, Sonnino *et al.*^{269–271} have synthesized several photoaffinity probes based on GM1. Their general probe design incorporated a photoreactive nitrophenyl azide group at the terminus of the fatty acid and a radioactive isotope in the glycan, such as ³H in C-5 acetyl of Sia (**Figure 40(a)**). In initial experiments, the researchers demonstrated that these probes could be incorporated into cellular membranes by exogenous administration. At short incubation times, that is, less than 2 h, the observed photocross-linking was limited to extracellular membrane proteins. However, after a 24-h incubation, the GM1 reagent was internalized, as demonstrated by its metabolic

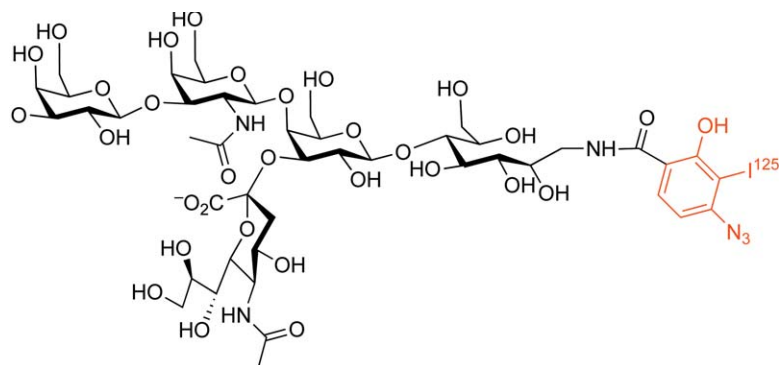


Figure 38 A GM1 oligosaccharide functionalized with a radiolabeled aryl-azide photocross-linker.

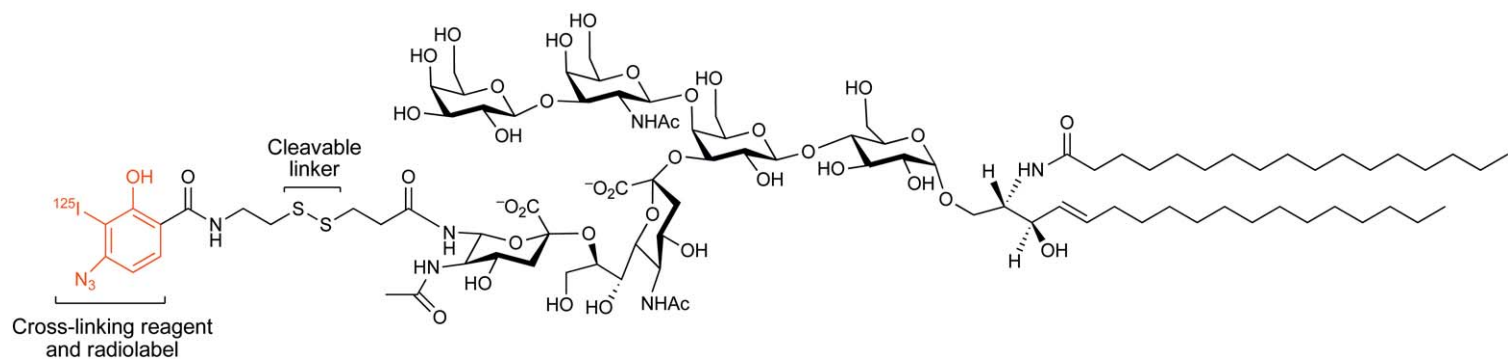
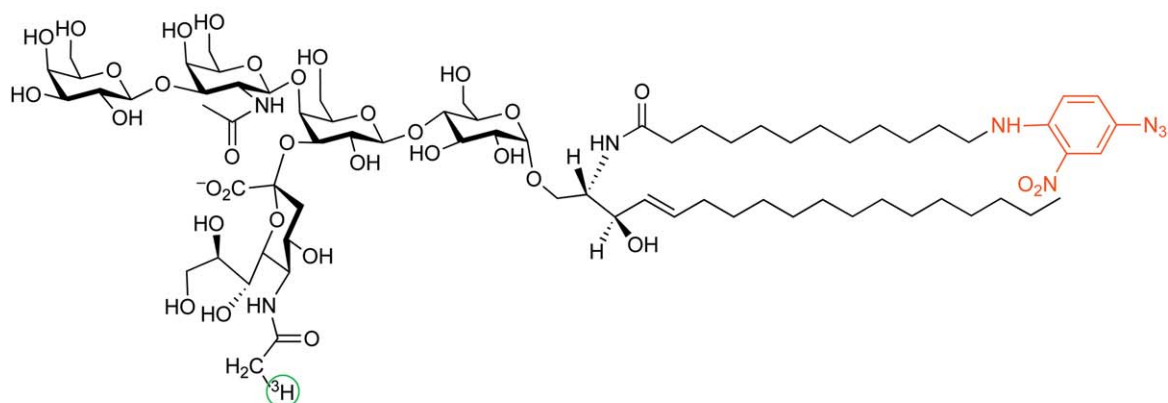
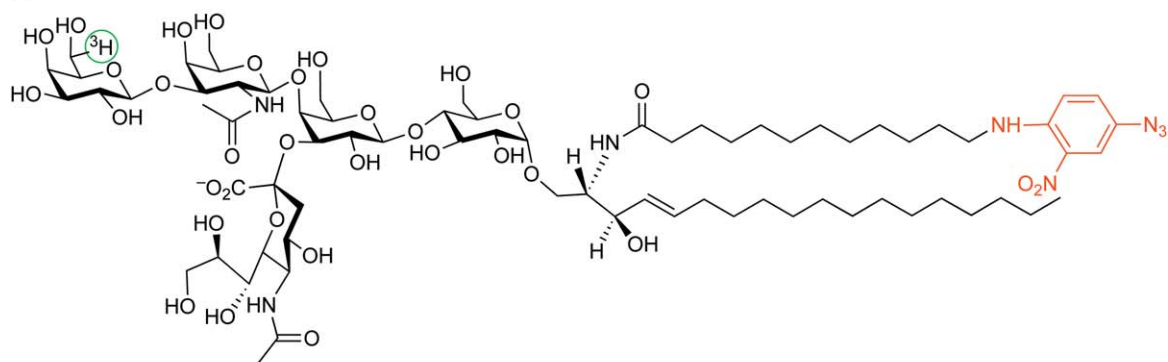


Figure 39 A cleavable GD1 ganglioside probe functionalized with a radiolabeled aryl azide photocross-linker.

(a)



(b)



(c)

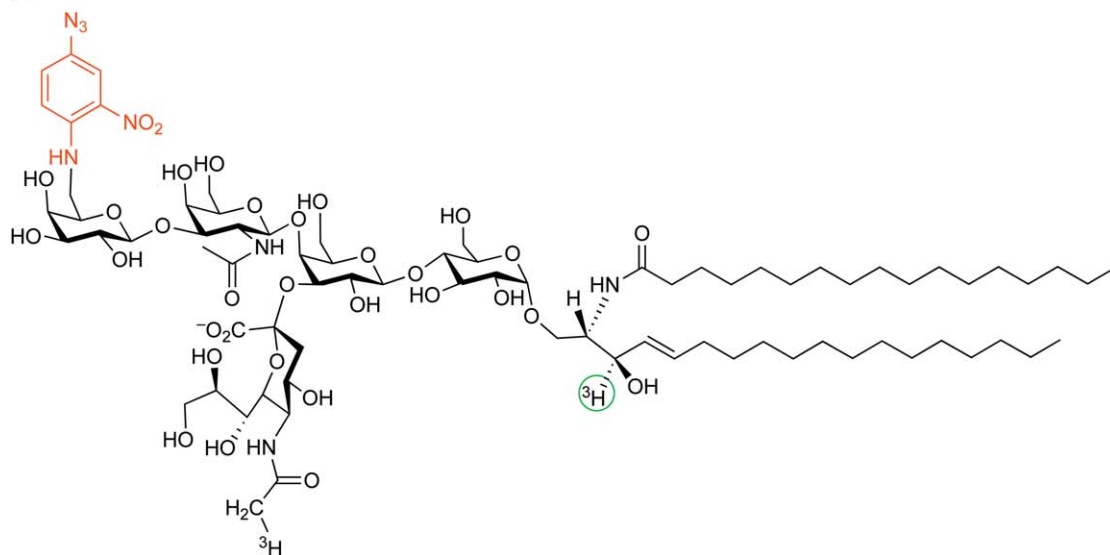


Figure 40 Radiolabeled GM1 ganglioside probes with photoreactive nitrophenyl azides. Photocross-linkers are shown in red, radiolabeled atoms are highlighted with green circles. (a and b) GM1 probes with the photoreactive probe incorporated at the terminus of the fatty acid. (c) GM1 probe with the photoreactive group incorporated on the terminal galactose at C-6.

conversion into GM2, GM3, and GD1a analogs and by detectable photocross-linking to intracellular proteins. These experiments provide a foundation for future use of these probes to study cellular ganglioside metabolism.

In subsequent experiments, similar radiolabeled, photocross-linking gangliosides were used to study the composition of caveolae. Caveolae are plasma membrane invaginations that have been implicated in endocytosis and signal transduction pathways. Using a GM1 probe²⁷⁰ ³H-labeled at the C-6 position of the terminal Gal (**Figure 40(b)**), Fra *et al.*²⁷² demonstrated selective cross-linking toward VIP21-caveolin, an essential structural component of caveolae. These experiments confirmed the presence of GM1 in caveolae, consistent with microscopy studies. Using a similar approach, Pitto *et al.*²⁷³ investigated the lipid environment surrounding caveolar GM1. The investigators prepared a GM1-based probe that incorporated an ¹²⁵I-labeled trifluoromethyl aryl diazirine photoreactive group at the terminus of the fatty acid. This probe was incorporated into A431 cells, which were subsequently irradiated to introduce cross-linking between GM1 and neighboring molecules. The detergent-resistant caveolae fraction was isolated and radioactive photocross-linked complexes were identified. Significant amounts of GM1–sphingomyelin cross-linking were observed, suggesting that these two lipids cosegregate in the caveolae membranes. In contrast, a later study using a radiolabeled, photoactive GM3 derivative demonstrated little presence of GM3 in the caveolae.²⁷⁴ After a 24-h incubation period, no cross-linking between GM3 and caveolae proteins was detected. These results suggest that, unlike GM1, GM3 segregates into a glycosphingolipid-rich region that is distinct from caveolae.

Another important role for glycolipids is in the regulation of signal transduction events. In investigations of the relevant glycosphingolipid–protein interactions, photocross-linking probes have been important tools. Using a radiolabeled, photoactivable GD1b probe, Prinetti *et al.*²⁷⁵ studied interactions between GD1b and Src-family protein tyrosine kinases specifically those that have been previously shown to be found in glycosphingolipid-enriched membrane domains. This group demonstrated a direct interaction between the glycosphingolipid probe and both c-Src (cellular-Src) and Lyn. This interaction likely occurs between the acyl chain of GD1b and the *N*-myristoyl chains that anchor c-Src and Lyn kinases in the lipid membrane. In contrast, the Csk kinase failed to cross-link to GD1b, suggesting that its localization to glycosphingolipid-enriched domains might be facilitated through protein–protein interactions rather than protein–glycosphingolipid interactions. These investigators also discovered that, in cultured neuronal cells, GM3, GM1, and GD1b interact with a GPI-anchored protein that they identified as TAG-1.^{276,277} Cross-linking between GM1 and TAG-1 was observed regardless of whether the GM1 photocross-linking group was placed in the fatty acid (**Figure 40(a)**) or the oligosaccharide (**Figure 40(c)**), underscoring the close association between these molecules. Another recent example by Kabayama *et al.* utilized photoaffinity probes to investigate ganglioside binding to the insulin receptor. These interactions may influence signal transduction through the plasma membrane. Using a radiolabeled, photoactivable GM3 probe, the researchers demonstrated that GM3 ganglioside specifically interacts with the β -subunit of the insulin receptor. Moreover, they found that a lysine residue located proximal to the transmembrane region is required for this interaction.²⁷⁸ A more surprising discovery was made by Palestini *et al.*²⁷⁹ using a radioactive and diazirine-modified GM1 molecule. This group detected cross-links between the GM1 probe and α - and β -tubulin. The complex was found to dissociate after base treatment, suggesting that GM1 interacted with a fatty acid-modified form of tubulin. The presence of lipid-anchored tubulin in glycosphingolipid-enriched membrane domains could have important implications for signal transduction and structural remodeling of the cell, aspects that remain to be investigated.

As described above, photocross-linking glycolipids have been used to detect a number of glycolipid–protein interactions. However, few of these studies provide molecular detail about the nature of these interactions. One notable exception was the use of GM2-based photoaffinity probes to provide information about the GM2-binding region of the GM2-activator protein (GM2AP), an essential cofactor needed for the degradation of GM2 by lysosomal glycosidases. The GM2 probe used in these experiments contained a trifluoromethyl phenyl diazirine in the fatty acid and ¹⁴C-labeled acetyl group at C-5 of Sia (**Figure 41**).²⁸⁰ After efficient photolabeling of GM2AP, the protein was trypsinized and radio-labeled fragments were isolated and identified by mass spectrometry (MS). This analysis revealed a direct interaction between the ceramide of GM2 and a surface loop on GM2AP. Previously reported crystal structures of GM2AP had shown that this loop was the most flexible surface of the protein and was likely to act as a ligand-binding domain.^{281,282}

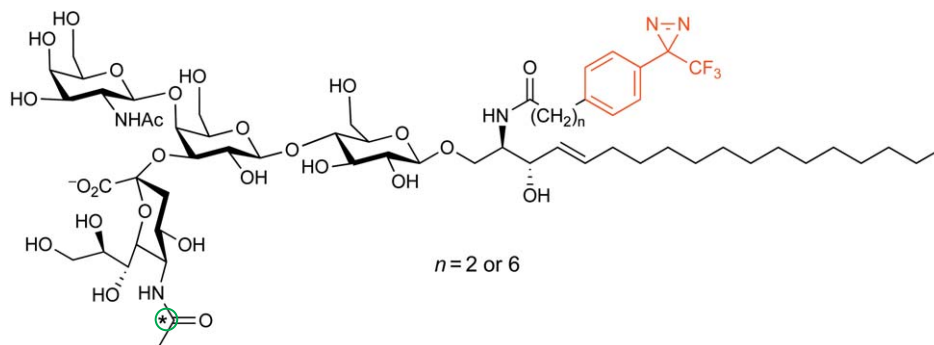


Figure 41 Radiolabeled GM2 ganglioside functionalized with an aryl-azide photocross-linker. Photocross-linker is highlighted in red, radiolabeled atom is highlighted by a green circle.

6.07.5.3 Metabolic Incorporation of Cross-linking Sugars

Metabolic engineering is a powerful method that has given researchers the ability to introduce unnatural functional groups at specific sites within cellular glycoconjugates. While the photocross-linking probes described in the previous two sections can be used for *in vitro* experiments or added exogenously to cells, metabolic incorporation of photocross-linking sugars provides the possibility of performing cross-linking experiments in relatively unperturbed cells or organisms. Metabolically incorporated photocross-linking sugars can be used to capture and characterize the transient macromolecular interactions that underlie essential cell–cell and cell–ligand-binding events.²³⁶ Here we describe three recent examples of metabolic incorporation of sugars bearing photocross-linking groups, and speculate upon the future utility of these molecules.

Two different Sia analogs bearing the aryl-azide cross-linker have been reported. The analog reported by the Bertozzi group places the AAz on the C-5 *N*-acyl chain (SiaNAAz–5-aryl azide-*N*-acetylneuraminic acid, **Figure 42(a)**),²⁰⁹ while the Paulson group installed the same cross-linker at C-9 (9-AAz-NeuAc, **Figure 42(b)**).²⁸³ Both of these molecules have been effectively incorporated into the cell surface glycoconjugates of cultured lymphocyte cell lines. In addition to these two photocross-linking Sias, the Bertozzi group reported a ManNAc analog (ManNAAz–*N*-aryl-azidemannosamine, **Figure 42(c)**) bearing the AAz on the *N*-acyl chain.²⁰⁹ This Sia precursor could also be metabolized to SiaNAAz and incorporated into cell surface glycoproteins, although incorporation was found to be less efficient than when SiaNAAz was used directly.

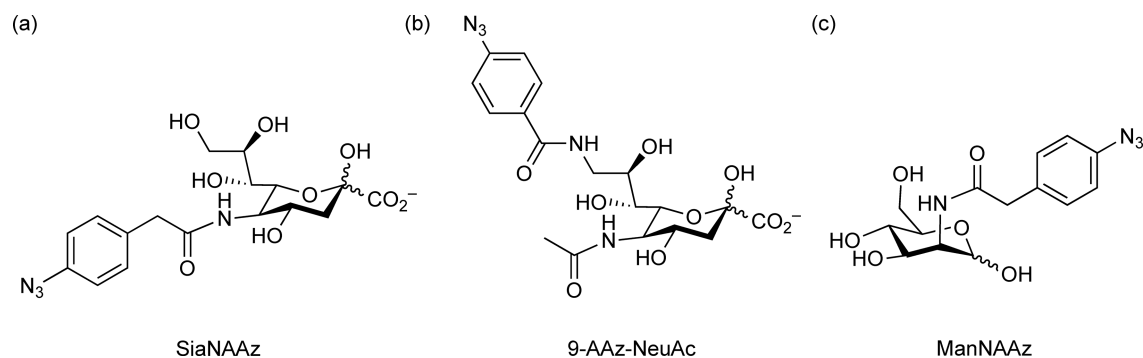


Figure 42 Structures of aryl-azide mannosamine and sialic acid analogs: (a) SiaNAAz, (b) 9-AAz-NeuAc, (c) ManNAAz. (a and c) SiaNAAz and ManNAAz bear an aryl azide on their *N*-acyl chain, are metabolized by cells, and incorporated into cell surface glycoconjugates in places of naturally occurring sialic acid. SiaNAAz is incorporated much more efficiently than ManNAAz. (b) When K20 BJAB cells are cultured with this sialic acid analog, it is efficiently incorporated into the cell surface glycoprotein CD22, where it can be photocross-linked to CD22's binding partners.

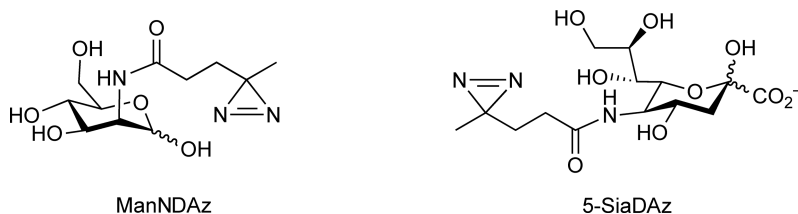


Figure 43 Structures of ManNDAz and SiaDAz. Both ManNDAz and SiaDAz are metabolized by BJAB cells and incorporated into the CD22 glycoprotein in the form of sialic acid, where they can be used as photocross-link CD22 multimers.

The modest incorporation efficiency of ManNAAz may be attributable to the size of the photocross-linker, which has the potential to sterically interfere with one of the three enzymatic steps that convert ManNAc into Sia (**Figure 25**). Indeed, we recently reported a ManNAc analog bearing the smaller diazirine photocross-linker on the *N*-acyl chain (ManNDAz – *N*-(4,4'-azi)-butanoylmannosamine, **Figure 43**) and showed that this molecule is efficiently incorporated into cell surface glycoproteins where it can be used to photocross-link glycoprotein multimers.²⁸⁴

Metabolically incorporated photocross-linking sugars offer a way to covalently capture the transient complexes formed by cellular glycoconjugates. The Paulson group utilized this technology to define the *cis* ligands of the CD22 (cluster of differentiation-22) glycoprotein found on the surface of B cells. *In vitro* assays had demonstrated that CD22 binds to a number of other glycoproteins, including CD45 and surface IgM, in a glycan-dependent manner. However, CD22's cellular binding partners remained ill-defined. The Paulson group cultured a B-cell line with 9-AAz-NeuAc and UV irradiated the cells to cross-link CD22 to its glycoprotein-binding partners. By analyzing the cross-linked complexes, the researchers were able to analyze CD22 in its normal cellular context and show that this glycoprotein preferentially self-associates, binding to the glycans of other CD22 molecules, but not to glycans attached to CD45 or surface IgM. Similar cross-linking has been observed with Sia bearing a diazirine cross-linking moiety (5-SiaDAz, **Figure 43**).²⁸⁴ These results suggest the presence of a membrane microdomain structure where CD22 is effectively sequestered, information that would have been difficult to obtain using more traditional immunoprecipitation techniques.

While 9-AAz-NeuAc was an effective reagent to study the binding partners of CD22, a larger panel of photocross-linking sugars will be required to study other glycoconjugate-binding events. Naturally occurring Sias are subject to a wide range of posttranslational modifications including acetylation, sulfation, phosphorylation, methylation, lactylation, and sialylation (**Figure 21**). Since these modifications can have dramatic effects on the binding properties of sialosides, photocross-linking sugars should be chosen appropriately to allow physiologically relevant modifications to take place. At the same time, nonphysiological modifications, such as the addition of aromatics to C-9 of Sia, have been demonstrated to enhance binding by several orders of magnitude.^{285,286} In some instances, it may be useful to conduct experiments with both C-9 and C-5 modified Sias and compare the results. In addition, not all glycoconjugates contain Sia, or the Sia may not be positioned in such a way that it can cross-link to a binding partner. Previous reports have demonstrated that it is possible to use metabolic engineering to incorporate GlcNAc, GalNAc, and fucose analogs bearing azides and alkynes; one can envision using a similar strategy to add photocross-linkers to these monosaccharides, providing tools to study a wider range of glycan-mediated-binding events.

To date, metabolic incorporation of photocross-linking sugars has been used for *in situ* studies of how CD22 interacts, or fails to interact, with its known binding partners. One can also envision other uses for these molecules. For example, it may be possible to discover the binding partners of a known (glyco)protein using MS-based protein identification methods. In addition, information about the sites of cross-linking could potentially be used to map the architecture of glycan-containing structures (**Figure 44**). The future development of this technology will be facilitated by improvements in MS that make high sensitivity and high accuracy measurements more routine (see Chapters 6.05 and 6.09).²⁸⁷

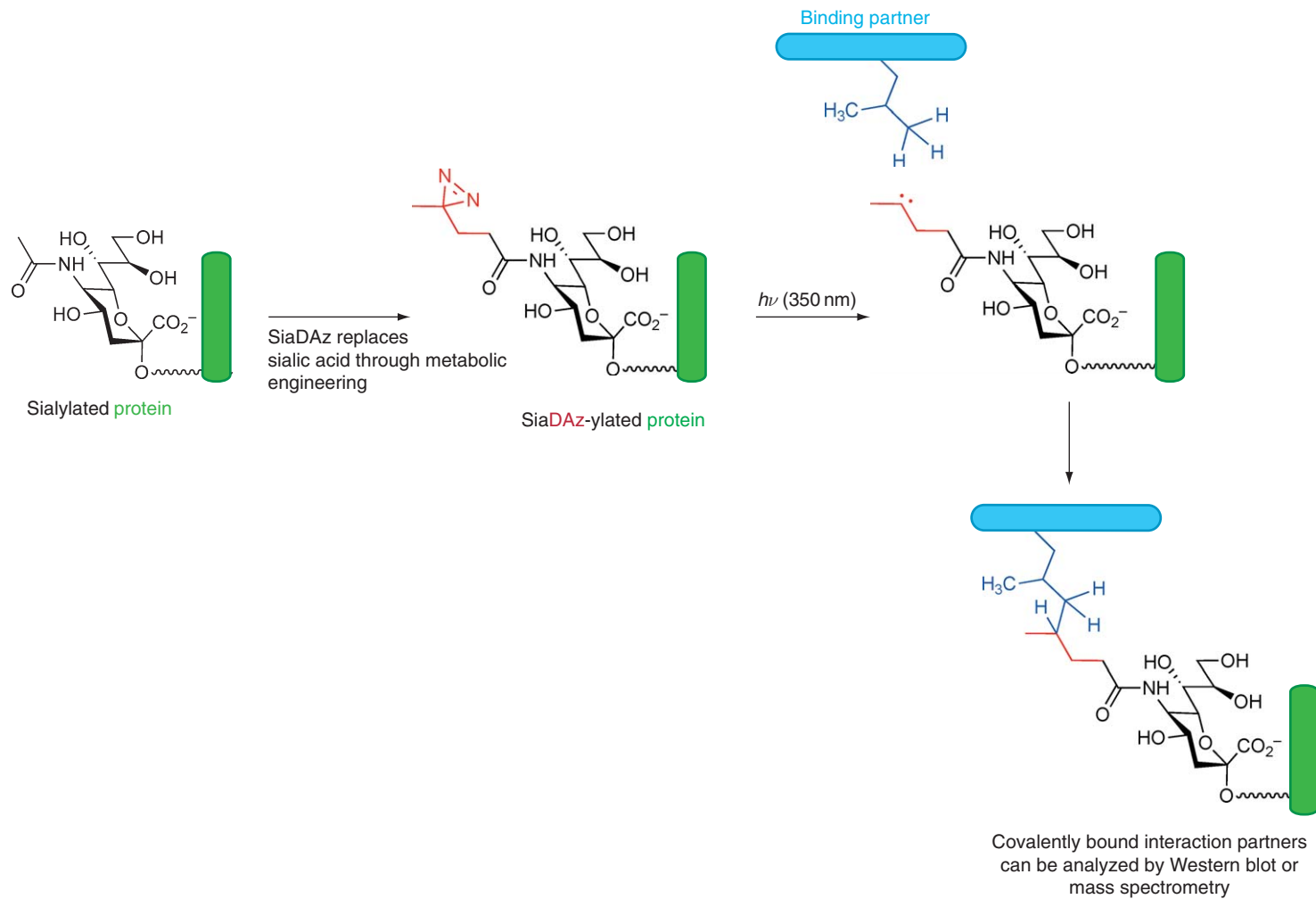


Figure 44 Strategy for identifying the binding partners of sialosides. Metabolic engineering can be used to replace sialic acids with their photocross-linking counterparts. Modified glycoconjugates can be subjected to *in situ* irradiation to form cross-links with their binding partners. The covalently bound complexes can be isolated and analyzed by immunochemical or mass spectrometry methods.

Abbreviations

| | |
|-----------------------|---|
| AAz | aryl azide |
| ADP | adenosine diphosphate |
| BDNF | brain-derived neurotrophic factor |
| CBP33 | carbohydrate-binding protein (MW = 33 kDa) |
| CD22 | cluster of differentiation-22 |
| CDP | cytidine diphosphate |
| CMP | cytidine monophosphate |
| CS | chondroitin sulfate |
| c-Src | cellular Src |
| DFMO | difluoromethylornithine |
| DS | dermatan sulfate |
| dTDP | deoxythymidine diphosphate |
| ELISA | enzyme-linked immunosorbent assay |
| Gal | galactose |
| GalNAc | <i>N</i> -acetylgalactosamine |
| Glc | glucose |
| GlcA | glucuronic acid |
| GlcNAc | <i>N</i> -acetylglucosamine |
| Globo-H | globohexaosylceramide |
| GAG | glycosaminoglycan |
| GB₃ | globotriaosylceramide |
| GDP | guanosine diphosphate |
| GFP | green fluorescent protein |
| GD3Bu | <i>N</i> -butanoyl GD3 |
| GM2AP | GM2-activator protein |
| GPI | glycophosphatidylinositol |
| HDAC | histone deacetylase |
| HS | heparan sulfate |
| IdoA | iduronic acid |
| KDN | deaminated neuraminic acid |
| KDO | 3-deoxy- <i>D</i> -manno-octulosonic acid |
| KLH | keyhole limpet hemocyanin |
| LacNAc | <i>N</i> -acetyllactosamine |
| Le^b | Lewis b antigen |
| Le^Y | Lewis Y antigen |
| mAb | monoclonal antibody |
| MALDI-MS | matrix-assisted laser desorption ionization-mass spectrometry |
| Man | mannose |
| ManBut | <i>N</i> -butanoylmannosamine |
| ManLev | <i>N</i> -levulinolmannosamine |
| ManPent | <i>N</i> -pentanoylmannosamine |
| ManNAAz | <i>N</i> -aryl-azidemannosamine |
| ManNDAz | <i>N</i> -(4,4'-azi)-butanoylmannosamine |
| ManNPac | <i>N</i> -phenylacetylmannosamine |
| ManProp | <i>N</i> -propanoylmannosamine |
| ManNAc | <i>N</i> -acetylmannosamine |
| MurNAc | <i>N</i> -acetylmuramic acid |
| NCAM | neural cell adhesion molecule |
| Neu5Ac | <i>N</i> -acetylneuraminic acid |
| Neu5Gc | <i>N</i> -glycolylneuraminic acid |

| | |
|--------------------------------|---|
| NN-DNJ | <i>N</i> -(<i>n</i> -nonyl)deoxynojirimycin |
| NTP | nucleotide triphosphate |
| Pam₃Cys | tripalmitoyl-S-glyceryl cysteine |
| PLC | phospholipase C |
| PrP^C | cellular prion protein, normal form |
| PrP^{Sc} | cellular prion protein, pathogenic form |
| PSA | polysialic acid |
| PST | ST8Sia4 polysialyltransferase |
| ROMP | ring-opening metathesis polymerization |
| Ser | serine |
| Sia | sialic acid |
| SiaLev | 5- <i>N</i> -levulinoylneuraminic acid |
| SiaNAAz | 5-aryl azide- <i>N</i> -acetylneuraminic acid |
| SiaPhAc | 5- <i>N</i> -phenylacetylneuraminic acid |
| sLe^x | sialyl Lewis X antigen |
| sTn | sialyl Tn antigen |
| STX | ST8Sia2 polysialyltransferase |
| TF | Thomsen–Friedenreich antigen |
| Tn | α -GalNAc O-linked to Ser/Thr |
| TNF-α | tumor necrosis factor α |
| UDP | uridine diphosphate |
| Xyl | xylose |

References

1. B. Ernst; G. W. Hart; P. Sinay, *Carbohydrates in Chemistry and Biology*; Wiley-VCH: Weinheim, New York, 2000.
2. S. D. Rosen, *Am. J. Pathol.* **1999**, *155*, 1013–1020.
3. S. J. Danishefsky; V. Behar; J. T. Randolph; K. O. Lloyd, *J. Am. Chem. Soc.* **1995**, *117*, 5701–5711.
4. P. J. Sabbatini; V. Kudryashov; G. Ragupathi; S. J. Danishefsky; P. O. Livingston; W. Bornmann; M. Spassova; A. Zatorski; D. Spriggs; C. Aghajanian; S. Soignet; M. Peyton; C. O'Flaherty; J. Curtin; K. O. Lloyd, *Int. J. Cancer* **2000**, *87*, 79–85.
5. P. W. Glunz; S. Hintermann; J. B. Schwarz; S. D. Kuduk; X. T. Chen; L. J. Williams; D. Sames; S. J. Danishefsky; V. Kudryashov; K. O. Lloyd, *J. Am. Chem. Soc.* **1999**, *121*, 10636–10637.
6. P. W. Glunz; S. Hintermann; L. J. Williams; J. B. Schwarz; S. D. Kuduk; V. Kudryashov; K. O. Lloyd; S. J. Danishefsky, *J. Am. Chem. Soc.* **2000**, *122*, 7273–7279.
7. V. Kudryashov; P. W. Glunz; L. J. Williams; S. Hintermann; S. J. Danishefsky; K. O. Lloyd, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3264–3269.
8. G. Ragupathi; F. Koide; P. O. Livingston; Y. S. Cho; A. Endo; Q. Wan; M. K. Spassova; S. J. Keding; J. Allen; O. Ouerfelli; R. M. Wilson; S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 2715–2725.
9. J. R. Allen; C. R. Harris; S. J. Danishefsky, *J. Am. Chem. Soc.* **2001**, *123*, 1890–1897.
10. G. Ragupathi; D. M. Coltart; L. J. Williams; F. Koide; E. Kagan; J. Allen; C. Harris; P. W. Glunz; P. O. Livingston; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13699–13704.
11. S. J. Keding; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11937–11942.
12. O. J. Plante; E. R. Palmacci; P. H. Seeberger, *Science* **2001**, *291*, 1523–1527.
13. K. R. Love; R. B. Andrade; P. H. Seeberger, *J. Org. Chem.* **2001**, *66*, 8165–8176.
14. K. R. Love; P. H. Seeberger, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 602–605.
15. K. R. Love; P. H. Seeberger, *J. Org. Chem.* **2005**, *70*, 3168–3177.
16. S. Hanashima; B. Castagner; D. Esposito; T. Nokami; P. H. Seeberger, *Org. Lett.* **2007**, *9*, 1777–1779.
17. J. Barbaro, *Curr. Org. Chem.* **2004**, *8*, 883–902.
18. P. Sears; C. H. Wong, *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 2300–2324.
19. T. G. Marron; T. J. Woltering; G. Weitz-Schmidt; C. H. Wong, *Tetrahedron Lett.* **1996**, *37*, 9037–9040.
20. M. W. Cappi; W. J. Moree; L. Qiao; T. G. Marron; G. Weitz-Schmidt; C. H. Wong, *Bioorg. Med. Chem.* **1997**, *5*, 1453.
21. C. Y. Tsai; W. K. C. Park; G. Weitz-Schmidt; B. Ernst; C. H. Wong, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2333–2338.
22. C. C. Lin; F. Moris-Varas; G. Weitz-Schmidt; C. H. Wong, *Bioorg. Med. Chem.* **1999**, *7*, 425–433.
23. C. H. Wong; F. Moris-Varas; S. C. Hung; T. G. Marron; C. C. Lin; K. W. Gong; G. Weitz-Schmidt, *J. Am. Chem. Soc.* **1997**, *119*, 8152–8158.
24. E. J. Gordon; J. E. Gestwicki; L. E. Strong; L. L. Kiessling, *Chem. Biol.* **2000**, *7*, 9–16.

25. E. J. Gordon; W. J. Sanders; L. L. Kiessling, *Nature* **1998**, *392*, 30–31.
26. E. J. Gordon; L. E. Strong; L. L. Kiessling, *Bioorg. Med. Chem.* **1998**, *6*, 1293–1299.
27. D. D. Manning; X. Hu; P. Beck; L. L. Kiessling, *J. Am. Chem. Soc.* **1997**, *119*, 3161–3162.
28. P. Mowery; Z. Q. Yang; E. J. Gordon; O. Dwir; A. G. Spencer; R. Alon; L. L. Kiessling, *Chem. Biol.* **2004**, *11*, 725–732.
29. R. M. Owen; J. E. Gestwicki; T. Young; L. L. Kiessling, *Org. Lett.* **2002**, *4*, 2293–2296.
30. W. J. Sanders; E. J. Gordon; O. Dwir; P. J. Beck; R. Alon; L. L. Kiessling, *J. Biol. Chem.* **1999**, *274*, 5271–5278.
31. R. E. Bruehl; F. Dasgupta; T. R. Katsumoto; J. H. Tan; C. R. Bertozzi; W. Spevak; D. J. Ahn; S. D. Rosen; J. O. Nagy, *Biochemistry* **2001**, *40*, 5964–5974.
32. S. M. Rele; W. X. Cui; L. C. Wang; S. J. Hou; G. Barr-Zarse; D. Tatton; Y. Gnanou; J. D. Esko; E. L. Chaikof, *J. Am. Chem. Soc.* **2005**, *127*, 10132–10133.
33. A. K. Sarkar; J. R. Brown; J. D. Esko, *Carbohydr. Res.* **2000**, *329*, 287–300.
34. A. K. Sarkar; T. A. Fritz; W. H. Taylor; J. D. Esko, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3323–3327.
35. A. K. Sarkar; K. S. Rostand; R. K. Jain; K. L. Matta; J. D. Esko, *J. Biol. Chem.* **1997**, *272*, 25608–25616.
36. J. R. Brown; M. M. Fuster; T. Whisenant; J. D. Esko, *J. Biol. Chem.* **2003**, *278*, 23352–23359.
37. T. K. K. Mong; L. V. Lee; J. R. Brown; J. D. Esko; C. H. Wong, *ChemBioChem* **2003**, *4*, 835–840.
38. M. M. Foster; J. R. Brown; L. C. Wang; J. D. Esko, *Cancer Res.* **2003**, *63*, 2775–2781.
39. J. R. Brown; M. M. Fuster; R. X. Li; N. Varki; C. A. Glass; J. D. Esko, *Clin. Cancer Res.* **2006**, *12*, 2894–2901.
40. C. I. Gama; L. C. Hsieh-Wilson, *Curr. Opin. Chem. Biol.* **2005**, *9*, 609–619.
41. M. S. M. Timmer; B. L. Stocker; P. H. Seeberger, *Curr. Opin. Chem. Biol.* **2007**, *11*, 59–65.
42. M. Belting; S. Persson; L. A. Fransson, *Biochem. J.* **1999**, *338*, 317–323.
43. M. Belting; L. Borsig; M. M. Fuster; J. R. Brown; L. Persson; L. A. Fransson; J. D. Esko, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 371–376.
44. F. N. Lugenwa; J. D. Esko, *J. Biol. Chem.* **1991**, *266*, 6674–6677.
45. T. A. Fritz; F. N. Lugenwa; A. K. Sarkar; J. D. Esko, *J. Biol. Chem.* **1994**, *269*, 300–307.
46. A. K. Sarkar; J. D. Esko, *Carbohydr. Res.* **1995**, *279*, 161–171.
47. O. Ben-Zaken; S. Tzaban; Y. Tal; L. Horonchik; J. D. Esko; I. Vlodyavsky; A. Taraboulos, *J. Biol. Chem.* **2003**, *278*, 40041–40049.
48. E. J. Bradbury; L. D. F. Moon; R. J. Popat; V. R. King; G. S. Bennett; P. N. Patel; J. W. Fawcett; S. B. McMahon, *Nature* **2002**, *416*, 636–640.
49. J. Iida; D. Pei; T. Kang; M. A. Simpson; M. Herlyn; L. T. Furcht; J. B. McCarthy, *J. Biol. Chem.* **2001**, *276*, 18786–18794.
50. K. Sugahara; T. Mikami; T. Uyama; S. Mizuguchi; K. Nomura; H. Kitagawa, *Curr. Opin. Struct. Biol.* **2003**, *13*, 612–620.
51. H. Kawashima; K. Atarashi; M. Hirose; J. Hirose; S. Yamada; K. Sugahara; M. Miyasaka, *J. Biol. Chem.* **2002**, *277*, 12921–12930.
52. S. Nadanaka; A. Clement; K. Masayama; A. Faissner; K. Sugahara, *J. Biol. Chem.* **1998**, *273*, 3296–3307.
53. S. E. Tully; R. Mabon; C. I. Gama; S. M. Tsai; X. W. Liu; L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2004**, *126*, 7736–7737.
54. C. I. Gama; S. E. Tully; N. Sotogaku; P. M. Clark; M. Rawat; N. Vaidehi; W. A. Goddard; A. Nishi; L. C. Hsieh-Wilson, *Nat. Chem. Biol.* **2006**, *2*, 467–473.
55. N. Sotogaku; S. E. Tully; C. I. Gama; H. Higashi; M. Tanaka; L. C. Hsieh-Wilson; A. Nishi, *J. Neurochem.* **2007**, *103*, 749–760.
56. S. E. Tully; M. Rawat; L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2006**, *128*, 7740–7741.
57. M. Rawat; C. I. Gama; J. B. Matson; L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2008**, *130*, 2959–2961.
58. M. Chen; J. Wang, *Arch. Pathol. Lab. Med.* **2008**, *132*, 851–853.
59. A. R. Sawkar; W. C. Cheng; E. Beutler; C. H. Wong; W. E. Balch; J. W. Kelly, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15428–15433.
60. A. R. Sawkar; M. Schmitz; K. P. Zimmer; D. Reczek; T. Edmunds; W. E. Balch; J. W. Kelly, *ACS Chem. Biol.* **2006**, *1*, 235–251.
61. A. R. Sawkar; S. L. Adamski-Werner; W. C. Cheng; C. H. Wong; E. Beutler; K. P. Zimmer; J. W. Kelly, *Chem. Biol.* **2005**, *12*, 1235–1244.
62. Z. Q. Yu; A. R. Sawkar; L. J. Whalen; C. H. Wong; J. W. Kelly, *J. Med. Chem.* **2007**, *50*, 94–100.
63. R. A. Steet; S. Chung; B. Wustman; A. Powe; H. Do; S. A. Kornfeld, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13813–13818.
64. R. L. Lieberman; B. A. Wustman; P. Huertas; A. C. Powe; C. W. Pine; R. Khanna; M. G. Schlossmacher; D. Ringe; G. A. Petsko, *Nat. Chem. Biol.* **2007**, *3*, 101–107.
65. M. A. Simpson; H. Cross; C. Proukakis; D. A. Priestman; D. C. A. Neville; G. Reinkensmeier; H. Wang; M. Wiznitzer; K. Gurtz; A. Verganelaki; A. Pryde; M. A. Patton; R. A. Dwek; T. D. Butters; F. M. Platt; A. H. Crosby, *Nat. Genet.* **2004**, *36*, 1225–1229.
66. M. G. Paulick; C. R. Bertozzi, *Biochemistry* **2008**, *47*, 6991–7000.
67. C. Murakata; T. Ogawa, *Tetrahedron Lett.* **1991**, *32*, 671–674.
68. Z. W. Guo; L. Bishop, *Eur. J. Org. Chem.* **2004**, 3585–3596.
69. M. G. Paulick; A. R. Wise; M. B. Forstner; J. T. Groves; C. R. Bertozzi, *J. Am. Chem. Soc.* **2007**, *129*, 11543–11550.
70. M. G. Paulick; M. B. Forstner; J. T. Groves; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20332–20337.
71. P. Sharma; R. Varma; R. C. Sarasij; Ira; K. Gousset; G. Krishnamoorthy; M. Rao; S. Mayor, *Cell* **2004**, *116*, 577–589.
72. L. Schofield; M. C. Hewitt; K. Evans; M. A. Siomos; P. H. Seeberger, *Nature* **2002**, *418*, 785–789.
73. X. Y. Liu; Y. U. Kwon; P. H. Seeberger, *J. Am. Chem. Soc.* **2005**, *127*, 5004–5005.
74. X. Y. Liu; P. H. Seeberger, *Chem. Commun. (Camb.)* **2004**, 1708–1709.
75. Y. U. Kwon; R. L. Soucy; D. A. Snyder; P. H. Seeberger, *Chem. Eur. J.* **2005**, *11*, 2493–2504.
76. P. H. Seeberger; R. L. Soucy; Y. U. Kwon; D. A. Snyder; T. Kanemitsu, *Chem. Commun. (Camb.)* **2004**, 1706–1707.
77. F. Kamena; M. Tamborini; X. Y. Liu; Y. U. Kwon; F. Thompson; G. Pluschke; P. H. Seeberger, *Nat. Chem. Biol.* **2008**, *4*, 238–240.
78. J. Adler; G. L. Hazelbauer; M. M. Dahl, *J. Bacteriol.* **1973**, *115*, 824–847.
79. G. Auger; M. Crouvoisier; M. Caroff; J. van Heijenoort; D. Blanot, *Lett. Pept. Sci.* **1997**, *4*, 371–376.
80. S. A. Hitchcock; C. N. Eid; J. A. Aikins; M. Zia-Ebrahimi; L. C. Blaszcak, *J. Am. Chem. Soc.* **1998**, *120*, 1916–1917.
81. M. S. VanNieuwenhze; S. C. Mauldin; M. Zia-Ebrahimi; J. A. Aikins; L. C. Blaszcak, *J. Am. Chem. Soc.* **2001**, *123*, 6983–6988.
82. H. B. Men; P. Park; M. Ge; S. Walker, *J. Am. Chem. Soc.* **1998**, *120*, 2484–2485.

83. S. Ha; E. Chang; M. C. Lo; H. Men; P. Park; M. Ge; S. Walker, *J. Am. Chem. Soc.* **1999**, *121*, 8415–8426.
84. L. Chen; H. Men; S. Ha; X. Y. Ye; L. Brunner; Y. Hu; S. Walker, *Biochemistry* **2002**, *41*, 6824–6833.
85. P. Cudic; D. C. Behenna; M. K. Yu; R. G. Kruger; L. M. Szewczuk; D. G. McCafferty, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 3107–3110.
86. H. Liu; T. K. Ritter; R. Sadamoto; P. S. Sears; M. Wu; C. H. Wong, *ChemBioChem* **2003**, *4*, 603–609.
87. M. S. VanNieuwenhze; S. C. Mauldin; M. Zia-Ebrahimi; B. E. Winger; W. J. Hornback; S. L. Saha; J. A. Aikins; L. C. Blaszcak, *J. Am. Chem. Soc.* **2002**, *124*, 3656–3660.
88. B. Schwartz; J. A. Markwalder; Y. Wang, *J. Am. Chem. Soc.* **2001**, *123*, 11638–11643.
89. X. Y. Ye; M. C. Lo; L. Brunner; D. Walker; D. Kahne; S. Walker, *J. Am. Chem. Soc.* **2001**, *123*, 3155–3156.
90. E. Breukink; H. E. van Heusden; P. J. Vollmerhaus; E. Swiezewska; L. Brunner; S. Walker; A. J. R. Heck; B. de Kruijff, *J. Biol. Chem.* **2003**, *278*, 19898–19903.
91. Y. Zhang; E. J. Fechter; T. S. A. Wang; D. Barrett; S. Walker; D. E. Kahne, *J. Am. Chem. Soc.* **2007**, *129*, 3080–3081.
92. D. L. Perlstein; Y. Zhang; T. S. Wang; D. E. Kahne; S. Walker, *J. Am. Chem. Soc.* **2007**, *129*, 12674–12675.
93. A. L. Lovering; L. H. de Castro; D. Lim; N. C. J. Strynadka, *Science* **2007**, *315*, 1402–1405.
94. Y. Q. Yuan; D. Barrett; Y. Zhang; D. Kahne; P. Sliz; S. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5348–5353.
95. K. Mikusova; M. Belanova; J. Kordulakova; K. Honda; M. R. McNeil; S. Mahapatra; D. C. Crick; P. J. Brennan, *J. Bacteriol.* **2006**, *188*, 6592–6598.
96. M. Belanova; P. Dianiskova; P. J. Brennan; G. C. Completo; N. L. Rose; T. L. Lowary; K. Mikusova, *J. Bacteriol.* **2008**, *190*, 1141–1145.
97. T. Yagi; S. Mahapatra; K. Mikusova; D. C. Crick; P. J. Brennan, *J. Biol. Chem.* **2003**, *278*, 26497–26504.
98. J. E. Gestwicki; L. L. Kiessling, *Nature* **2002**, *415*, 81–84.
99. J. E. Gestwicki; L. E. Strong; S. L. Borchardt; C. W. Cairo; A. M. Schnoes; L. L. Kiessling, *Bioorg. Med. Chem.* **2001**, *9*, 2387–2393.
100. J. E. Gestwicki; L. E. Strong; L. L. Kiessling, *Chem. Biol.* **2000**, *7*, 583–591.
101. A. C. Lamanna; J. E. Gestwicki; L. E. Strong; S. L. Borchardt; R. M. Owen; L. L. Kiessling, *J. Bacteriol.* **2002**, *184*, 4981–4987.
102. P. I. Kitov; J. M. Sadowska; G. Mulvey; G. D. Armstrong; H. Ling; N. S. Pannu; R. J. Read; D. R. Bundle, *Nature* **2000**, *403*, 669–672.
103. G. L. Mulvey; P. Marcatò; P. I. Kitov; J. Sadowska; D. R. Bundle; G. D. Armstrong, *J. Infect. Dis.* **2003**, *187*, 640–649.
104. K. Matsuoka; M. Terabatake; Y. Esumi; D. Terunuma; H. Kuzuhara, *Tetrahedron Lett.* **1999**, *40*, 7839–7842.
105. K. Nishikawa; K. Matsuoka; E. Kita; N. Okabe; M. Mizuguchi; K. Hino; S. Miyazawa; C. Yamasaki; J. Aoki; S. Takashima; Y. Yamakawa; M. Nishijima; D. Terunuma; H. Kuzuhara; Y. Natori, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7669–7674.
106. K. Nishikawa; K. Matsuoka; M. Watanabe; K. Igai; K. Hino; K. Hatano; A. Yamada; N. Abe; D. Terunuma; H. Kuzuhara; Y. Natori, *J. Infect. Dis.* **2005**, *191*, 2097–2105.
107. M. Watanabe; K. Igai; K. Matsuoka; A. Miyagawa; T. Watanabe; R. Yanoshita; Y. Samejima; D. Terunuma; Y. Natori; K. Nishikawa, *Infect. Immun.* **2006**, *74*, 1984–1988.
108. M. Watanabe; K. Matsuoka; E. Kita; K. Igai; N. Higashi; A. Miyagawa; T. Watanabe; R. Yanoshita; Y. Samejima; D. Terunuma; Y. Natori; K. Nishikawa, *J. Infect. Dis.* **2004**, *189*, 360–368.
109. H. J. M. Gijzen; L. Qiao; W. Fitz; C. H. Wong, *Chem. Rev.* **1996**, *96*, 443–473.
110. S. Hanson; M. Best; M. C. Bryan; C. H. Wong, *Trends Biochem. Sci.* **2004**, *29*, 656–663.
111. K. M. Koeller; C. H. Wong, *Chem. Rev.* **2000**, *100*, 4465–4493.
112. C. H. Wong; S. L. Haynie; G. M. Whitesides, *J. Org. Chem.* **1982**, *47*, 5416–5418.
113. J. Nahalka; Z. Y. Liu; X. Chen; P. G. Wang, *Chem. Eur. J.* **2003**, *9*, 373–377.
114. J. Q. Jiang; J. B. Biggins; J. S. Thorson, *J. Am. Chem. Soc.* **2000**, *122*, 6803–6804.
115. J. Q. Jiang; J. B. Biggins; J. S. Thorson, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1502–1505.
116. J. Q. Jiang; C. Albermann; J. S. Thorson, *ChemBioChem* **2003**, *4*, 443–446.
117. K. S. Ko; C. J. Zea; N. L. Pohl, *J. Am. Chem. Soc.* **2004**, *126*, 13188–13189.
118. K. S. Ko; C. J. Zea; N. L. Pohl, *J. Org. Chem.* **2005**, *70*, 1919–1921.
119. S. C. Timmons; R. H. Mosher; S. A. Knowles; D. Jakeman, *Org. Lett.* **2007**, *9*, 857–860.
120. M. P. Huestis; G. A. Aish; J. P. M. Hui; E. C. Soo; D. L. Jakeman, *Org. Biomol. Chem.* **2008**, *6*, 477–484.
121. S. C. Timmons; J. Hui; J. L. Pearson; P. Peltier; R. Daniellou; C. Nugier-Chauvin; E. Soo; R. T. Syvitski; V. Ferrieres; D. Jakeman, *Org. Lett.* **2008**, *10*, 161–163.
122. T. Kotake; D. Yamaguchi; H. Ohzono; S. Hojo; S. Kaneko; H. K. Ishida; Y. Tsumuraya, *J. Biol. Chem.* **2004**, *279*, 45728–45736.
123. T. Kotake; S. Hojo; D. Yamaguchi; T. Aohara; T. Konishi; Y. Tsumuraya, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 761–771.
124. R. M. Mizanur; C. J. Zea; N. L. Pohl, *J. Am. Chem. Soc.* **2004**, *126*, 15993–15998.
125. R. M. Mizanur; F. A. Jaipuri; N. L. Pohl, *J. Am. Chem. Soc.* **2005**, *127*, 836–837.
126. R. M. Mizanur; N. L. Pohl, *J. Mol. Catal., B Enzym.* **2008**, *50*, 13–19.
127. R. M. Mizanur; N. L. Pohl, *Appl. Microbiol. Biotechnol.* **2007**, *76*, 827–834.
128. J. Bae; K. H. Kim; D. Kim; Y. Choi; J. S. Kim; S. Koh; S. I. Hong; D. S. Lee, *ChemBioChem* **2005**, *6*, 1963–1966.
129. J. Flint; E. Taylor; M. Yang; D. N. Bolam; L. E. Tailford; C. Martinez-Fleites; E. J. Dodson; B. G. Davis; H. J. Gilbert; G. J. Davies, *Nat. Struct. Mol. Biol.* **2005**, *12*, 608–614.
130. W. A. Barton; J. Lesniak; J. B. Biggins; P. D. Jeffrey; J. Q. Jiang; K. R. Rajashankar; J. S. Thorson; D. B. Nikolov, *Nat. Struct. Biol.* **2001**, *8*, 545–551.
131. W. A. Barton; J. B. Biggins; J. Q. Jiang; J. S. Thorson; D. B. Nikolov, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13397–13402.
132. R. Moretti; J. S. Thorson, *J. Biol. Chem.* **2007**, *282*, 16942–16947.
133. D. Hoffmeister; J. Yang; L. Liu; J. S. Thorson, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13184–13189.
134. J. Yang; L. Liu; J. S. Thorson, *ChemBioChem* **2004**, *5*, 992–996.
135. J. Yang; X. Fu; J. C. Liao; L. Liu; J. S. Thorson, *Chem. Biol.* **2005**, *12*, 657–664.
136. A. Minami; K. Kakinuma; T. Eguchi, *Tetrahedron Lett.* **2005**, *46*, 6187–6190.
137. C. Zhang; C. Albermann; X. Fu; J. Thorson, *J. Am. Chem. Soc.* **2006**, *128*, 16420–16421.

138. C. Zhang; B. R. Griffith; Q. Fu; C. Albermann; X. Fu; I. Lee; L. Li; J. Thorson, *Science* **2006**, *313*, 1291–1294.
139. G. J. Williams; R. D. Goff; C. S. Zhang; J. S. Thorson, *Chem. Biol.* **2008**, *15*, 393–401.
140. Z. Y. Liu; J. B. Zhang; X. Chen; P. G. Wang, *ChemBioChem* **2002**, *3*, 348–355.
141. J. Shao; J. B. Zhang; J. Nahalka; P. G. Wang, *Chem. Commun. (Camb.)* **2002**, 2586–2587.
142. J. C. Errey; B. Mukhopadhyay; K. P. R. Kartha; R. A. Field, *Chem. Commun. (Camb.)* **2004**, 2706–2707.
143. X. Fu; C. Albermann; J. Q. Jiang; J. C. Liao; C. S. Zhang; J. S. Thorson, *Nat. Biotechnol.* **2003**, *21*, 1467–1469.
144. C. Albermann; A. Soriano; J. Q. Jiang; H. Vollmer; J. B. Biggins; W. A. Barton; J. Lesniak; D. B. Nikolov; J. S. Thorson, *Org. Lett.* **2003**, *5*, 933–936.
145. G. J. Williams; C. Zhang; J. S. Thorson, *Nat. Chem. Biol.* **2007**, *3*, 657–662.
146. X. Fu; C. Albermann; C. S. Zhang; J. S. Thorson, *Org. Lett.* **2005**, *7*, 1513–1515.
147. M. Yang; M. R. Proctor; D. N. Bolam; J. C. Errey; R. A. Field; H. J. Gilbert; B. G. Davis, *J. Am. Chem. Soc.* **2005**, *127*, 9336–9337.
148. L. S. Zhao; J. Ahlert; Y. Q. Xue; J. S. Thorson; D. H. Sherman; H. W. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 9881–9882.
149. L. Hong; Z. Zhao; C. E. Melancon; H. Zhang; H. Liu, *J. Am. Chem. Soc.* **2008**, *130*, 4954–4967.
150. H. Yu; X. Chen, *Org. Biomol. Chem.* **2007**, *5*, 865–872.
151. T. Angata; A. Varki, *Chem. Rev.* **2002**, *102*, 439–469.
152. I. Hemeon; A. J. Bennet, *Synthesis (Stuttg.)* **2007**, 1899–1926.
153. D. G. Comb; S. Roseman, *J. Am. Chem. Soc.* **1958**, *80*, 497–499.
154. W. Fitz; J. R. Schwark; C. H. Wong, *J. Org. Chem.* **1995**, *60*, 3663–3670.
155. M. J. Kim; W. J. Hennen; H. M. Sweers; C. H. Wong, *J. Am. Chem. Soc.* **1988**, *110*, 6481–6486.
156. C. Auge; S. David; A. Malleron, *Carbohydr. Res.* **1989**, *188*, 201–205.
157. D. C. M. Kong; M. Vonitzstein, *Tetrahedron Lett.* **1995**, *36*, 957–960.
158. C. C. Lin; C. H. Lin; C. H. Wong, *Tetrahedron Lett.* **1997**, *38*, 2649–2652.
159. M. J. Kiefel; J. C. Wilson; S. Bennett; M. Gredley; M. von Itzstein, *Bioorg. Med. Chem.* **2000**, *8*, 657–664.
160. T. Miyazaki; T. Sakakibara; H. Sato; Y. Kajihara, *J. Am. Chem. Soc.* **1999**, *121*, 1411–1412.
161. S. S. Huang; H. Yu; X. Chen, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 2249–2253.
162. H. Yu; X. Chen, *Org. Lett.* **2006**, *8*, 2393–2396.
163. A. Bolt; A. Berry; A. Nelson, *Arch. Biochem. Biophys.* **2008**, *474*, 318–330.
164. S. M. Dean; W. A. Greenberg; C. H. Wong, *Adv. Synth. Catal.* **2007**, *349*, 1308–1320.
165. T. Izard; M. C. Lawrence; R. L. Malby; G. G. Lilley; P. M. Colman, *Structure* **1994**, *2*, 361–369.
166. C. C. Hsu; Z. Y. Hong; M. Wada; D. Franke; C. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9122–9126.
167. M. Wada; C. C. Hsu; D. Franke; M. Mitchell; A. Heine; I. Wilson; C. H. Wong, *Bioorg. Med. Chem.* **2003**, *11*, 2091–2098.
168. G. J. Williams; T. Woodhall; A. Nelson; A. Berry, *Protein Eng. Des. Sel.* **2005**, *18*, 239–246.
169. T. Woodhall; G. Williams; A. Berry; A. Nelson, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 2109–2112.
170. G. J. Williams; T. Woodhall; L. M. Farnsworth; A. Nelson; A. Berry, *J. Am. Chem. Soc.* **2006**, *128*, 16238–16247.
171. J. L. C. Liu; G. J. Shen; Y. Ichikawa; J. F. Rutan; G. Zapata; W. F. Vann; C. H. Wong, *J. Am. Chem. Soc.* **1992**, *114*, 3901–3910.
172. S. L. Shames; E. S. Simon; C. W. Christopher; W. Schmid; G. M. Whitesides; L. L. Yang, *Glycobiology* **1991**, *1*, 187–191.
173. A. J. Humphrey; C. Fremann; P. Critchley; Y. Malykh; R. Schauer; T. D. H. Bugg, *Bioorg. Med. Chem.* **2002**, *10*, 3175–3185.
174. M. Knorst; W. D. Fessner, *Adv. Synth. Catal.* **2001**, *343*, 698–710.
175. H. Yu; H. Yu; R. Karpel; X. Chen, *Bioorg. Med. Chem.* **2004**, *12*, 6427–6435.
176. H. Chokhawala; H. Cao; H. Yu; X. Chen, *J. Am. Chem. Soc.* **2007**, *129*, 10630–10631.
177. O. Blixt; K. Allin; L. Pereira; A. Datta; J. C. Paulson, *J. Am. Chem. Soc.* **2002**, *124*, 5739–5746.
178. T. Yamamoto; H. Nagae; Y. Kajihara; I. Terada, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 210–214.
179. T. Yamamoto; M. Nakashizuka; H. Kodama; Y. Kajihara; I. Terada, *J. Biochem.* **1996**, *120*, 104–110.
180. T. Yamamoto; M. Nakashizuka; I. Terada, *J. Biochem.* **1998**, *123*, 94–100.
181. C. F. Teo; T. S. Hwang; P. H. Chen; C. H. Hung; H. S. Gao; L. S. Chang; C. H. Lin, *Adv. Synth. Catal.* **2005**, *347*, 967–972.
182. Y. Kajihara; T. Kamitani; R. Sato; N. Kamei; T. Miyazaki; R. Okamoto; T. Sakakibara; T. Tsuji; T. Yamamoto, *Carbohydr. Res.* **2007**, *342*, 1680–1688.
183. H. Yu; S. S. Huang; H. Chokhawala; M. C. Sun; H. J. Zheng; X. Chen, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 3938–3944.
184. M. Gilbert; A. M. Cunningham; D. C. Watson; A. Martin; J. C. Richards; W. W. Wakarchuk, *Eur. J. Biochem.* **1997**, *249*, 187–194.
185. M. Izumi; G. J. Shen; S. Wacowich-Sgarbi; T. Nakatani; O. Plettenburg; C. H. Wong, *J. Am. Chem. Soc.* **2001**, *123*, 10909–10918.
186. G. J. Shen; A. K. Datta; M. Izumi; K. M. Koeller; C. H. Wong, *J. Biol. Chem.* **1999**, *274*, 35139–35146.
187. L. M. Willis; M. Gilbert; M. F. Karwaski; M. C. Blanchard; W. W. Wakarchuk, *Glycobiology* **2008**, *18*, 177–186.
188. K. Sujino; R. J. Jackson; N. W. C. Chan; S. Tsuji; M. M. Palcic, *Glycobiology* **2000**, *10*, 313–320.
189. Y. Ichikawa; G. J. Shen; C. H. Wong, *J. Am. Chem. Soc.* **1991**, *113*, 4698–4700.
190. H. Yu; H. Chokhawala; R. Karpel; H. Yu; B. Y. Wu; J. B. Zhang; Y. X. Zhang; Q. Jia; X. Chen, *J. Am. Chem. Soc.* **2005**, *127*, 17618–17619.
191. O. Blixt; J. C. Paulson, *Adv. Synth. Catal.* **2003**, *345*, 687–690.
192. H. Yu; H. A. Chokhawala; S. S. Huang; X. Chen, *Nat. Protoc.* **2006**, *1*, 2485–2492.
193. M. Gilbert; R. Bayer; A. M. Cunningham; S. Defrees; Y. H. Gao; D. C. Watson; N. M. Young; W. W. Wakarchuk, *Nat. Biotechnol.* **1998**, *16*, 769–772.
194. T. Endo; S. Koizumi; K. Tabata; S. Kakita; A. Ozaki, *Carbohydr. Res.* **2001**, *330*, 439–443.
195. H. Yu; H. A. Chokhawala; A. Varki; X. Chen, *Org. Biomol. Chem.* **2007**, *5*, 2458–2463.
196. H. A. Chokhawala; H. Yu; X. Chen, *ChemBioChem* **2007**, *8*, 194–201.
197. M. J. Linman; J. D. Taylor; H. Yu; X. Chen; Q. Cheng, *Anal. Chem.* **2008**, *80*, 4007–4013.
198. S. Muthana; H. Yu; S. Huang; X. Chen, *J. Am. Chem. Soc.* **2007**, *129*, 11918–11919.
199. C. T. Campbell; S. G. Sampathkumar; K. J. Yarema, *Mol. Biosyst.* **2007**, *3*, 187–194.
200. H. Kayser; R. Zeidler; C. Kannicht; D. Grunow; R. Nuck; W. Reutter, *J. Biol. Chem.* **1992**, *267*, 16934–16938.
201. D. H. Dube; J. A. Prescher; C. N. Quang; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4819–4824.

202. J. A. Prescher; D. H. Dube; C. R. Bertozzi, *Nature* **2004**, *430*, 873–877.
203. S. Goon; C. R. Bertozzi, *J. Carbohydr. Chem.* **2002**, *21*, 943–977.
204. D. H. Dube; C. R. Bertozzi, *Curr. Opin. Chem. Biol.* **2003**, *7*, 616–625.
205. O. T. Keppler; R. Horstkorte; M. Pawlita; C. Schmidts; W. Reutter, *Glycobiology* **2001**, *11*, 11R–8R.
206. O. T. Keppler; P. Stehling; M. Herrmann; H. Kayser; D. Grunow; W. Reutter; M. Pawlita, *J. Biol. Chem.* **1995**, *270*, 1308–1314.
207. C. L. Jacobs; S. Goon; K. J. Yarema; S. Hinderlich; H. C. Hang; D. H. Chai; C. R. Bertozzi, *Biochemistry* **2001**, *40*, 12864–12874.
208. E. J. Kim; S. G. Sampathkumar; M. B. Jones; J. K. Rhee; G. Baskaran; S. Goon; K. J. Yarema, *J. Biol. Chem.* **2004**, *279*, 18342–18352.
209. S. J. Luchansky; S. Goon; C. R. Bertozzi, *ChemBioChem* **2004**, *5*, 371–374.
210. C. Oetke; R. Brossmer; L. R. Mantey; S. Hinderlich; R. Isecke; W. Reutter; O. T. Keppler; M. Pawlita, *J. Biol. Chem.* **2002**, *277*, 6688–6695.
211. M. B. Jones; H. Teng; J. K. Rhee; N. Lahar; G. Baskaran; K. J. Yarema, *Biotechnol. Bioeng.* **2004**, *85*, 394–405.
212. S. G. Sampathkumar; M. B. Jones; M. A. Meledeo; C. T. Campbell; S. S. Choi; K. Hida; P. Gomutputra; A. Sheh; T. Gilmartin; S. R. Head; K. J. Yarema, *Chem. Biol.* **2006**, *13*, 1265–1275.
213. U. Aich; C. T. Campbell; N. Elmouelhi; C. A. Weier; S. G. Sampathkumar; S. S. Choi; K. J. Yarema, *ACS Chem. Biol.* **2008**, *3*, 230–240.
214. S. J. Luchansky; K. J. Yarema; S. Takahashi; C. R. Bertozzi, *J. Biol. Chem.* **2003**, *278*, 8035–8042.
215. K. J. Yarema; S. Goon; C. R. Bertozzi, *Nat. Biotechnol.* **2001**, *19*, 553–558.
216. H. Hildebrandt; M. Muhlenhoff; B. Weinhold; R. Gerardy-Schahn, *J. Neurochem.* **2007**, *103*, 56–64.
217. U. Rutishauser, *Nat. Rev. Neurosci.* **2008**, *9*, 26–35.
218. N. W. Charter; L. K. Mahal; D. E. Koshland, Jr.; C. R. Bertozzi, *Glycobiology* **2000**, *10*, 1049–1056.
219. L. K. Mahal; N. W. Charter; K. Angata; M. Fukuda; D. E. Koshland, Jr.; C. R. Bertozzi, *Science* **2001**, *294*, 380–381.
220. N. W. Charter; L. K. Mahal; D. E. Koshland, Jr.; C. R. Bertozzi, *J. Biol. Chem.* **2002**, *277*, 9255–9261.
221. R. Horstkorte; M. Muhlenhoff; W. Reutter; S. Nohring; M. Zimmermann-Kordmann; R. Gerardy-Schahn, *Exp. Cell Res.* **2004**, *298*, 268–274.
222. R. A. Pon; N. J. Biggs; H. J. Jennings, *Glycobiology* **2007**, *17*, 249–260.
223. C. Faure; A. Chalazonitis; C. Rheaume; G. Bouchard; S. G. Sampathkumar; K. J. Yarema; M. D. Gershon, *Dev. Dyn.* **2007**, *236*, 44–59.
224. D. Gagiannis; R. Gossrau; W. Reutter; M. Zimmermann-Kordmann; R. Horstkorte, *Biochim. Biophys. Acta* **2007**, *1770*, 297–306.
225. R. A. Pon; M. Lussier; Q. L. Yang; H. J. Jennings, *J. Exp. Med.* **1997**, *185*, 1929–1938.
226. T. Liu; Z. Guo; Q. Yang; S. Sad; H. J. Jennings, *J. Biol. Chem.* **2000**, *275*, 32832–32836.
227. P. Fredman; K. Hedberg; T. Brezicka, *BioDrugs* **2003**, *17*, 155–167.
228. P. B. Chapman; D. Morrissey; K. S. Panageas; L. Williams; J. J. Lewis; R. J. Israel; W. B. Hamilton; P. O. Livingston, *Clin. Cancer Res.* **2000**, *6*, 4658–4662.
229. P. B. Chapman; D. Wu; G. Ragupathi; S. Lu; L. Williams; W. J. Hwu; D. Johnson; P. O. Livingston, *Clin. Cancer Res.* **2004**, *10*, 4717–4723.
230. G. Giaccone; C. Debruyne; E. Felip; P. B. Chapman; S. C. Grant; M. Millward; L. Thiberville; G. D'addario; C. Coens; L. S. Rome; P. Zatloukal; O. Masso; C. Legrand, *J. Clin. Oncol.* **2005**, *23*, 6854–6864.
231. W. Zou; S. Borrelli; M. Gilbert; T. Liu; R. A. Pon; H. J. Jennings, *J. Biol. Chem.* **2004**, *279*, 25390–25399.
232. P. Chefalo; Y. Pan; N. Nagy; Z. Guo; C. V. Harding, *Biochemistry* **2006**, *45*, 3733–3739.
233. Y. Pan; P. Chefalo; N. Nagy; C. Harding; Z. Guo, *J. Med. Chem.* **2005**, *48*, 875–883.
234. Q. Wang; J. Zhang; Z. Guo, *Bioorg. Med. Chem.* **2007**, *15*, 7561–7567.
235. J. Brunner, *Annu. Rev. Biochem.* **1993**, *62*, 483–514.
236. Y. Tanaka; M. R. Bond; J. J. Kohler, *Mol. Biosyst.* **2008**, *4*, 473–480.
237. E. Saxon; C. R. Bertozzi, *Science* **2000**, *287*, 2007–2010.
238. J. F. Lutz, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 1018–1025.
239. E. V. Tsibizova; E. L. Vodovozova; I. I. Mikhalyov; Y. G. Molotkovsky, *Russ. J. Bioorg. Chem.* **2002**, *28*, 152–157.
240. R. T. Lee; Y. C. Lee, *Biochemistry* **1986**, *25*, 6835–6841.
241. R. T. Lee; Y. C. Lee, *Biochemistry* **1987**, *26*, 6320–6329.
242. K. G. Rice; Y. C. Lee, *J. Biol. Chem.* **1990**, *265*, 18423–18428.
243. K. G. Rice; O. A. Weisz; T. Barthel; R. T. Lee; Y. C. Lee, *J. Biol. Chem.* **1990**, *265*, 18429–18434.
244. D. Aoki; H. E. Appert; D. Johnson; S. S. Wong; M. N. Fukuda, *EMBO J.* **1990**, *9*, 3171–3178.
245. M. Hashimoto; Y. Hatanaka; J. Yang; J. Dhesi; G. D. Holman, *Carbohydr. Res.* **2001**, *331*, 119–127.
246. M. Walter; M. Wiegand; T. K. Lindhorst, *Eur. J. Org. Chem.* **2006**, 719–728.
247. M. Wiegand; T. K. Lindhorst, *Eur. J. Org. Chem.* **2006**, 4841–4851.
248. E. Yoshida; H. Nakayama; Y. Hatanaka; Y. Kanaoka, *Chem. Pharm. Bull. (Tokyo)* **1990**, *38*, 982–987.
249. A. K. Pathak; V. Pathak; J. M. Riordan; S. S. Gurcha; G. S. Besra; R. C. Reynolds, *Carbohydr. Res.* **2004**, *339*, 683–691.
250. A. K. Pathak; V. Pathak; L. Seitz; S. S. Gurcha; G. S. Besra; J. M. Riordan; R. C. Reynolds, *Bioorg. Med. Chem.* **2007**, *15*, 5629–5650.
251. L. Ballell; K. J. Alink; M. Slijper; C. Versluis; R. M. Liskamp; R. J. Pieters, *ChemBioChem* **2005**, *6*, 291–295.
252. L. Ballell; M. van Scherpenzeel; K. Buchalova; R. M. Liskamp; R. J. Pieters, *Org. Biomol. Chem.* **2006**, *4*, 4387–4394.
253. M. R. Lee; D. W. Jung; D. Williams; I. Shin, *Org. Lett.* **2005**, *7*, 5477–5480.
254. M. R. Lee; S. Park; I. Shin, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5132–5135.
255. Y. Hatanaka; M. Hashimoto; Y. Kanaoka, *J. Am. Chem. Soc.* **1998**, *120*, 453–454.
256. G. Lauc; R. T. Lee; J. Dumiaie; Y. C. Lee, *Glycobiology* **2000**, *10*, 357–364.
257. Y. Hatanaka; U. Kempin; P. Jong-Jip, *J. Org. Chem.* **2000**, *65*, 5639–5643.
258. Y. Sadakane; Y. Hatanaka, *Anal. Sci.* **2006**, *22*, 209–218.
259. H. Ueda; H. Fukushima; Y. Hatanaka; H. Ogawa, *Biochem. J.* **2004**, *382*, 821–829.
260. G. Lauc; M. Fogel; B. Diehl-Seifert; H. C. Schroder; W. E. Muller, *Glycoconj. J.* **1994**, *11*, 541–549.

261. J. J. Park; Y. Sadakane; K. Masuda; T. Tomohiro; T. Nakano; Y. Hatanaka, *ChemBioChem* **2005**, *6*, 814–818.
262. D. Ilver; A. Arnqvist; J. Ogren; I. M. Frick; D. Kersulyte; E. T. Incecik; D. E. Berg; A. Covacci; L. Engstrand; T. Borén, *Science* **1998**, *279*, 373–377.
263. E. L. Vodovozova, *Biochemistry (Mosc.)* **2007**, *72*, 1–20.
264. T. Pacuszka; P. H. Fishman, *Glycobiology* **1992**, *2*, 251–255.
265. S. M. Fueshko; C. L. Schengrund, *J. Neurochem.* **1992**, *59*, 527–535.
266. R. E. Shapiro; C. D. Specht; B. E. Collins; A. S. Woods; R. J. Cotter; R. L. Schnaar, *J. Biol. Chem.* **1997**, *272*, 30380–30386.
267. P. Emsley; C. Fotinou; I. Black; N. F. Fairweather; I. G. Charles; C. Watts; E. Hewitt; N. W. Isaacs, *J. Biol. Chem.* **2000**, *275*, 8889–8894.
268. C. Fotinou; P. Emsley; I. Black; H. Ando; H. Ishida; M. Kiso; K. A. Sinha; N. F. Fairweather; N. W. Isaacs, *J. Biol. Chem.* **2001**, *276*, 32274–32281.
269. V. Chigorno; M. Valsecchi; D. Acquotti; S. Sonnino; G. Tettamanti, *FEBS Lett.* **1990**, *263*, 329–331.
270. S. Sonnino; V. Chigorno; D. Acquotti; M. Pitto; G. Kirschner; G. Tettamanti, *Biochemistry* **1989**, *28*, 77–84.
271. S. Sonnino; V. Chigorno; M. Valsecchi; M. Pitto; G. Tettamanti, *Neurochem. Int.* **1992**, *20*, 315–321.
272. A. M. Fra; M. Masserini; P. Palestini; S. Sonnino; K. Simons, *FEBS Lett.* **1995**, *375*, 11–14.
273. M. Pitto; J. Brunner; A. Ferraretto; D. Ravasi; P. Palestini; M. Masserini, *Glycoconj. J.* **2000**, *17*, 215–222.
274. V. Chigorno; P. Palestini; M. Sciannamblo; V. Dolo; A. Pavan; G. Tettamanti; S. Sonnino, *Eur. J. Biochem.* **2000**, *267*, 4187–4197.
275. A. Prinetti; N. Marano; S. Prioni; V. Chigorno; L. Mauri; R. Casellato; G. Tettamanti; S. Sonnino, *Glycoconj. J.* **2000**, *17*, 223–232.
276. N. Loberto; S. Prioni; A. Prinetti; E. Ottico; V. Chigorno; D. Karagogeos; S. Sonnino, *J. Neurochem.* **2003**, *85*, 224–233.
277. S. Prioni; L. Mauri; N. Loberto; R. Casellato; V. Chigorno; D. Karagogeos; A. Prinetti; S. Sonnino, *Glycoconj. J.* **2004**, *21*, 461–470.
278. K. Kabayama; T. Sato; K. Saito; N. Loberto; A. Prinetti; S. Sonnino; M. Kinjo; Y. Igarashi; J. Inokuchi, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 13678–13683.
279. P. Palestini; M. Pitto; G. Tedeschi; A. Ferraretto; M. Parenti; J. Brunner; M. Masserini, *J. Biol. Chem.* **2000**, *275*, 9978–9985.
280. M. Wendeler; J. Hoernschemeyer; D. Hoffmann; T. Kolter; G. Schwarzmann; K. Sandhoff, *Eur. J. Biochem.* **2004**, *271*, 614–627.
281. C. S. Wright; S. C. Li; F. Rastinejad, *J. Mol. Biol.* **2000**, *304*, 411–422.
282. C. S. Wright; Q. Zhao; F. Rastinejad, *J. Mol. Biol.* **2003**, *331*, 951–964.
283. S. Han; B. E. Collins; P. Bengtson; J. C. Paulson, *Nat. Chem. Biol.* **2005**, *1*, 93–97.
284. Y. Tanaka; J. J. Kohler, *J. Am. Chem. Soc.* **2008**, *130*, 3278–3279.
285. S. Kelm; R. Brossmer; R. Isecke; H. J. Gross; K. Strenge; R. Schauer, *Eur. J. Biochem.* **1998**, *255*, 663–672.
286. S. Kelm; J. Gerlach; R. Brossmer; C. P. Danzer; L. Nitschke, *J. Exp. Med.* **2002**, *195*, 1207–1213.
287. S. M. Patrie; D. E. Robinson; F. Y. Meng; Y. Du; N. L. Kelleher, *Int. J. Mass Spectrom.* **2004**, *234*, 175–184.

Biographical Sketches



Chad M. Whitman was born in 1982 in North Kansas City, Missouri. He received a B.A. degree in Chemistry from The Iowa State University of Science and Technology where he carried out research in Victor Shang-Yi Lin's lab developing novel mesoporous silica nanoparticle materials as antibacterial delivery agents. Currently, he is a fifth year chemistry graduate student at Stanford University. His research is focused on the application of metabolic oligosaccharide engineering with photocross-linking sugars to investigate cell-surface ganglioside interactions.



Michelle R. Bond received her B.S. in Chemistry from Gettysburg College in Gettysburg, PA. There, she synthesized and tested the competency of several novel dicationic phase-transfer catalysts. Currently, Michelle is a fourth year chemistry graduate student at Stanford University. She works in Jennifer Kohler's lab at UT Southwestern Medical Center in Dallas, TX. Her research focuses on utilizing biologically relevant carbohydrate analogs to perturb and understand biological functions within cells.



Jennifer J. Kohler was born in York, PA and attended Bryn Mawr College, where she carried out research in Professor Susan White's lab using chemical reagents to study RNA structure. As a graduate student in the chemistry department at Yale University, she worked in Professor Alanna Schepartz's lab, studying the kinetics and mechanism of protein–DNA interactions. After obtaining her Ph.D., she was an American Cancer Society postdoctoral fellow in the lab of Professor Carolyn Bertozzi at UC Berkeley, where she developed a method to regulate glycosylation through small molecule control. She is currently an assistant professor in the Division of Translational Research, Department of Internal Medicine at UT Southwestern Medical Center in Dallas, TX. Dr. Kohler is the recipient of the Camille and Henry Dreyfus New Faculty Award, an NSF CAREER Award, and the Alfred P. Sloan Fellowship. The Kohler group's research is motivated by two over-riding goals: (1) to obtain a more molecular understanding of the machinery of eukaryotic glycosylation and (2) to exploit that understanding in the development of biochemical tools to study the roles of glycosylation.

6.08 Alkaloid Glycosidase Inhibitors

Alan D. Elbein, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Russell J. Molyneux, Western Regional Research Center, ARS-USDA, Albany, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|-------------------|--|-----|
| 6.08.1 | Introduction | 225 |
| 6.08.2 | Chemistry of Alkaloid Glycosidase Inhibitors | 226 |
| 6.08.2.1 | Structural Classes | 226 |
| 6.08.2.1.1 | Monocyclic alkaloids | 226 |
| 6.08.2.1.2 | Bicyclic alkaloids | 231 |
| 6.08.2.2 | Occurrence and Isolation from Natural Sources | 235 |
| 6.08.2.2.1 | Occurrence | 235 |
| 6.08.2.2.2 | Isolation and structural determination | 236 |
| 6.08.3 | Glycosidase Inhibition | 237 |
| 6.08.3.1 | Glycosidase Inhibitory Activity | 237 |
| 6.08.4 | Biological Activity of Glycosidase Inhibitors | 239 |
| 6.08.4.1 | Mammalian Toxicity | 239 |
| 6.08.5 | Structure and Biosynthesis of N-Linked Glycoproteins | 240 |
| 6.08.5.1 | Introduction | 240 |
| 6.08.5.2 | Biosynthesis of N-Linked Oligosaccharides | 242 |
| 6.08.5.3 | Processing of N-Linked Glycoproteins and Effects of Glycosidase Inhibitors | 243 |
| 6.08.6 | Structure and Synthesis of Glycosphingolipids | 246 |
| 6.08.6.1 | Introduction | 246 |
| 6.08.6.2 | Degradation of Glycosphingolipids in the Lysosomes and Resulting Diseases | 247 |
| 6.08.6.3 | Use of Glycosidase Inhibitors to Treat Lysosomal Lipid Storage Diseases | 248 |
| 6.08.7 | Therapeutic Activities of Glycosidase Inhibitors | 251 |
| 6.08.7.1 | Introduction | 251 |
| 6.08.7.2 | Use of Inhibitors in Substrate Reduction Therapy | 251 |
| 6.08.7.3 | Glycosidase Inhibitors in Diabetes Therapy | 252 |
| 6.08.7.4 | Potentials for Therapy in Viral Diseases | 253 |
| References | | 255 |

6.08.1 Introduction

At the time of the previous review¹ polyhydroxy alkaloids with glycosidase inhibitory properties were a relatively new class of compounds, comprising about 50 members. New alkaloids were being discovered with some regularity and the group was rapidly expanding. Since 1999, the pace of discovery has slowed down and the research area has matured. Expansion has resulted primarily from new additions within the previously known structural classes, through variations on the theme of the number, distribution, and stereochemistry of hydroxyl groups or other moieties. Significant advances have come especially from discovery of new sources, better understanding of structure–activity relationships to glycosidase inhibitory properties, biological activity, and therapeutic potential. As with many classes of natural products, the polyhydroxy alkaloids have provided structural templates for a large number of synthetic analogues, now far exceeding in number those isolated from biological sources. Details of synthetic approaches and critical assessments of specific routes can be found in the comprehensive volume on glycosidase inhibitors edited by Stütz² and the valuable series of reviews on indolizidine and quinolizidine alkaloids by Michael.³ This chapter is confined to chemistry and bioactivity of the naturally occurring alkaloids, their occurrence and distribution, glycosidase inhibitory properties and effects on glycoprotein processing, and consequent biological properties. In order to avoid

reiteration of information incorporated in the previous review,¹ the alkaloids known at that time will not be discussed in detail unless significant new discoveries have been made with respect to their sources or biological activities.

Given the disparate structural skeletons incorporated into these alkaloids, it is not surprising that several different names have been used to describe them, including polyhydroxy alkaloids, iminosugars or azasugars, and nitrogenous glycomimetics. Imino- (or aza-) sugars should only be properly applied to the monocyclic members of the group, and glycomimetics does not imply any natural occurrence, so polyhydroxy alkaloids appear to be the most suitable and inclusive term for this class of alkaloids.

6.08.2 Chemistry of Alkaloid Glycosidase Inhibitors

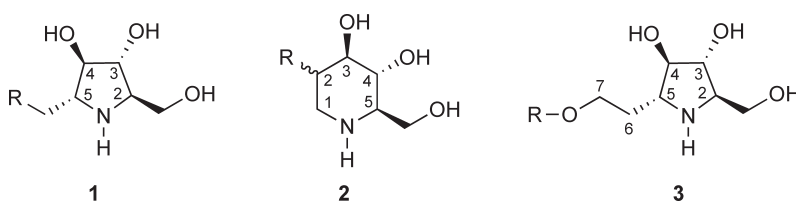
6.08.2.1 Structural Classes

Alkaloid glycosidase inhibitors can be broadly defined as monocyclic or bicyclic natural products possessing a heterocyclic nitrogen atom and at least two (but generally more) hydroxyl groups, falling into five different subclasses. The ability to inhibit one or more glycosidases, and consequently glycoprotein processing, is a common feature of their bioactivity which unites them within the broader class.

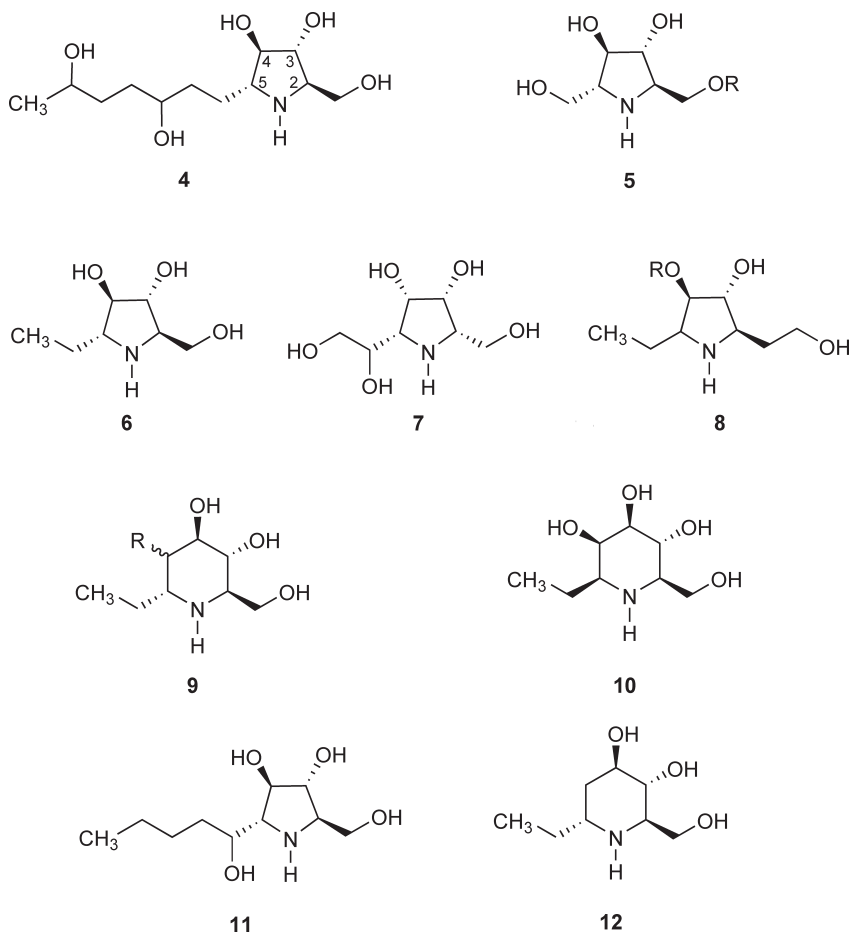
6.08.2.1.1 Monocyclic alkaloids

The monocyclic members of the polyhydroxy alkaloid group possess either five- or six-membered ring systems with a heterocyclic secondary nitrogen atom and fall into the pyrrolidine and piperidine classes, respectively. They have structural affinities to aminosugars and therefore are often referred to as iminosugars.

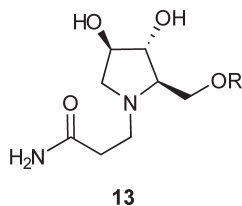
6.08.2.1.1(i) Pyrrolidines and piperidines Representative pyrrolidines are 2,5-dihydroxymethyl-3,4-dihydropyrrolidine (DMDP) (**1**, R = α -OH)⁴ and 6-deoxy-DMDP (**1**, R = H),⁵ whereas typical piperidines are 1-deoxynojirimycin (DNJ) (**2**, R = OH)⁶ and its *N*-methyl derivative,⁷ and 1-deoxymannojirimycin (DMJ) (**2**, R = β -OH). Further complexity is introduced by glycosylation of the hydroxyl groups, as in homoDMDP 7-*apioside* (**3**, R = apiose)⁸ and the gluco- and galacto-pyranosides of DNJ and DMJ.⁷



The number of members in the pyrrolidine class has been expanded by the isolation of 6-deoxy-6-*C*-(2,5-dihydroxyhexyl)-DMDP (**4**) and homoDMDP-7-*O*- β -D-xylopyranoside (**3**, R = xylose) from members of the plant family Hyacinthaceae,⁹ and DMDP-1-*O*- β -D-fructofuranoside (**5**, R = fructose) from *Baphia nitida* (Leguminosae),¹⁰ all of which co-occur with known polyhydroxy pyrrolidines and piperidines. *Scilla sibirica* (Hyacinthaceae) has proven to be a rich source of new pyrrolidine and piperidine alkaloids, including 7-deoxy-homoDMDP (**6**), 2,5-dideoxy-2,5-imino-glycero-D-galacto-heptitol (**7**), the 4-*O*- β -D-mannoside (**8**, R = mannose) and the 4-*O*- β -D-mannobioside (**8**, R = mannobiose) of 6-deoxy-homoDMDP (**8**, R = H), 7-deoxyhomonojirimycin (**9**, R = α -OH), and 7-deoxyhomomannojirimycin (**9**, R = β -OH),¹¹ whereas *S. socialis* contains β -1-*C*-ethyldeoxymannojirimycin (**10**).¹² A member of the closely related family Campanulaceae, *Adenophora triphylla* var. *japonica*, contains 6-*C*-butyl-DMDP (**11**) and α -1-*C*-ethylfagomine (**12**), DAB, DNJ, and DMJ.¹³ However, the pyrrolidine and piperidine alkaloids in *Scilla* species co-occur with a number of polyhydroxypyrrolidines; none of these bicyclic alkaloids were isolated from *A. triphylla*.

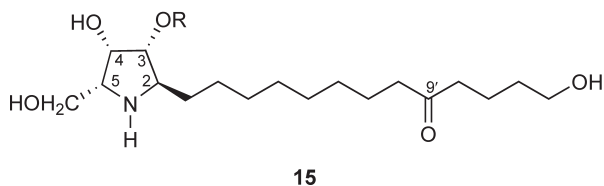
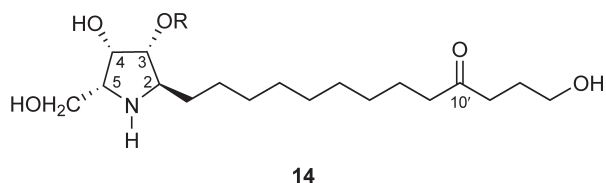


The fruits, leaves, and root bark of white mulberry trees (*Morus alba*; Moraceae) have yielded many polyhydroxy pyrrolidines and piperidines,^{7,14,15} especially DNJ and its glycosides. One unusual alkaloid is the *N*-propionamide derivative of (2*R*,3*R*,4*R*)-2-hydroxymethyl-3,4-dihydroxypyrrolidine (**13**) or its enantiomer.¹⁵

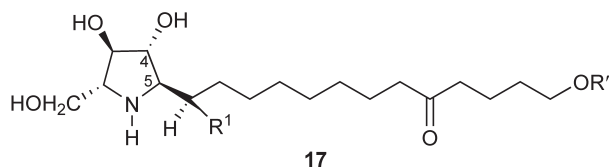
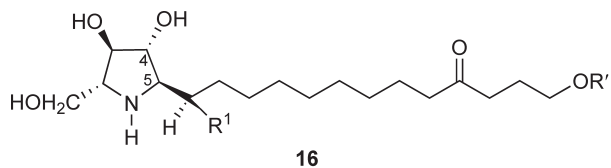


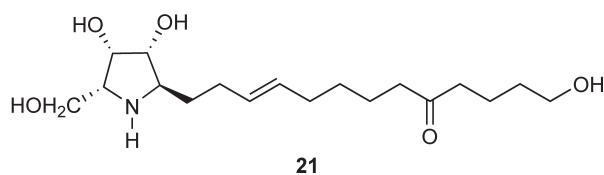
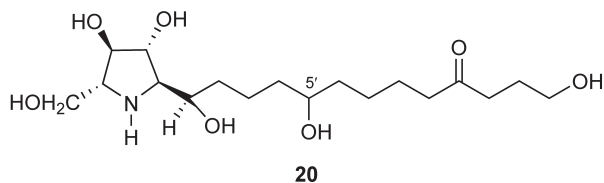
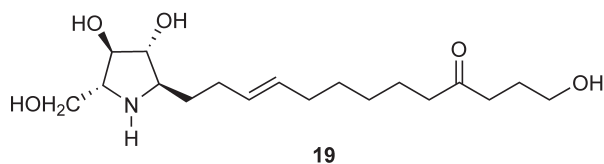
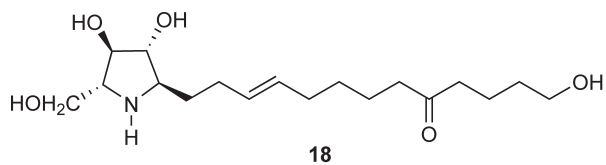
Investigation of three Thai medicinal plants, 'Non tai yak' (*Stemona tuberosa*; Stemonaceae), 'Thopthaep' (*Conarus ferrugineus*; Connaraceae), and 'Cha em thai' (*Albizia myriophylla*; Leguminosae) showed that all contained polyhydroxypiperidine alkaloids, several of which were new compounds.¹⁶ *Stemona tuberosa* roots contained 0.1% dry weight of α -homonojirimycin, accompanied by the lesser amounts of α -1-*C*-hydroxymethylfagomine and the pyrrolidine alkaloids 3-*O*- β -D-glucopyranosyl-DMDP and 2,5-dideoxy-2,5-imino-D-glucitol. High yields of DMJ and its derivatives were characteristic of the other two drugs, with *C. ferrugineus* yielding the new alkaloids 2-*O*- α -D-galactopyranosyl-DMJ, 3-*O*- β -D-glucopyranosyl-DMJ, 1,4-dideoxymannonojirimycin, 1,4-dideoxyallonojirimycin, and 1,4-dideoxyaltronojirimycin, whereas *A. myriophylla* produced the previously unknown compounds 2-*O*- β -D-glucopyranosyl-DMJ and 4-*O*- β -D-glucopyranosyl-DMJ.

Comprehensive investigation of constituents of the Japanese tree *Broussonetia kazinoki* (Moraceae), commonly known as 'kozo,' has led to the isolation and identification of broussonetines A–M and O–Z, M₁, U₁, J₁, and J₂, an extraordinary series of 31 dihydroxy-hydroxymethyl-pyrrolidine (14–30) alkaloids bearing a variously functionalized 13-carbon side chain at the 5-position of the pyrrolidine ring.^{17–25} Confusingly, broussonetines A (14, R = Glc) and B (15, R = Glc) are the 4-*O*-β-D-glucopyranosides of the aglycones, designated broussonetinines A (14, R = H) and B (15, R = H), whereas all other aglycones and glycosides in the series are named broussonetines.¹⁷ Furthermore, broussonetine N²³ is a pyrrolizidine alkaloid (*vide infra*), whereas broussonetines U²⁴ and U₁²⁵ are the only members of the group having a double bond within the pyrrolidine ring (i.e., a pyrroline). An additional problem is that the name broussonetine usurps that first used for a bisquinoyl butyrolactone alkaloid isolated from *Babylonia zeylanica*.²⁶ This situation demands a rationalization of the trivial names for these compounds before discovery of new compounds in the series introduces additional confusion. At the very least, broussonetinines A and B should be renamed as broussonetines A and B, respectively, and the latter renamed as broussonetine glucosides. Also, the polyhydroxypyrrolizidine alkaloid, broussonetine N, should be given a unique name that distinguishes it from all of the pyrrolidine alkaloids isolated from *B. kazinoki*.

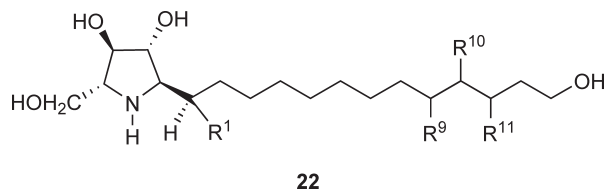


The 29 broussonetines with a pyrrolidine ring system can be categorized on the basis of their side-chain substitution patterns. The most common feature is a keto group in the side chain, at either C-9' or C-10', and a terminal –OH at C-13', which are present in broussonetines A, B, C (16, R¹ = R' = H), D (17, R = R' = H), E (16, R¹ = OH, R' = H), F (17, R¹ = OH, R' = H), K (16, R¹ = OH, R' = glucose), L (17, R¹ = OH, R' = glucose), O (18), P (19), T (20), and V (21); of these, broussonetines E, F, and T are all hydroxylated at C-1', with broussonetine T corresponding to broussonetine E but having a second hydroxyl group at C-5'. Broussonetines K and L are glucosylated at the terminal –OH position. Broussonetine Q is the only diglycoside, being the 4-*O*-glucoside of broussonetine K. Broussonetines O and P have unsaturation between the 3'- and 4'-positions of the side chain.

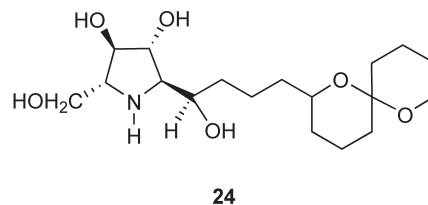
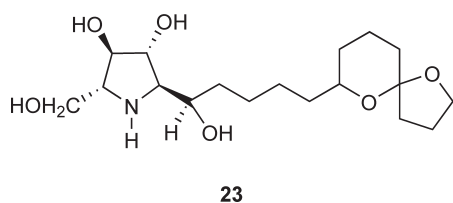


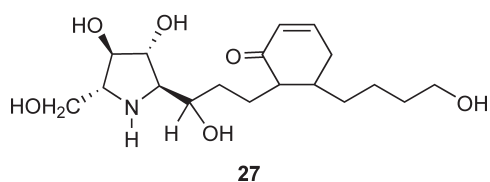
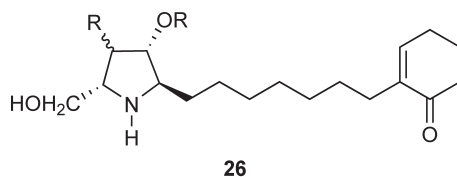
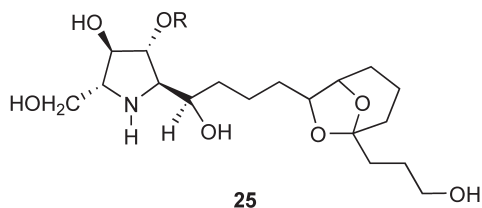


Four of the alkaloids, broussonetines M (**22**, $R^1 = R^9 = R^{10} = H$, $R^{11} = OH$), S (**22**, $R^1 = R^{10} = OD$, $R^9 = R^{11} = H$), M_1 (**22**, $R^1 = R^{10} = R^{11} = H$, $R^9 = OH$), and Y (**22**, $R^1 = R^{11} = H$, $R^9 = R^{10} = OH$), lack a keto group in the side chain, bearing either one or two hydroxyl groups at position C-1', C-9', C-10', or C-11'.

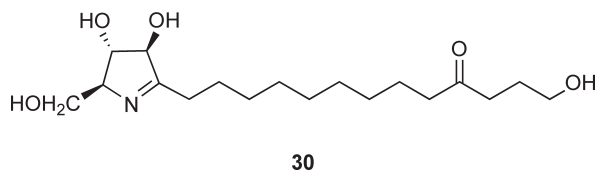
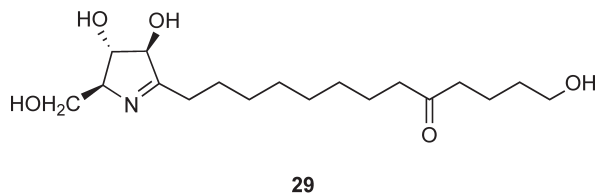
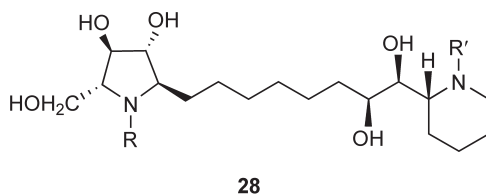


Broussonetines with heterocyclic oxygen functionalization are G (**23**) and H (**24**) that have terminal spiroketal functionalities, whereas Z (**25**) has a dioxabicyclo[3.2.1]octane ring system encompassing C-5' to C-10'. Broussonetines W (**26**, $R = OH$, $R' = H$) and X (**26**, $R = OH$, $R' = \text{glucose}$) have terminal cyclohexenone moieties, whereas in broussonetinine R (**27**), the same functionality is internal to the side chain.



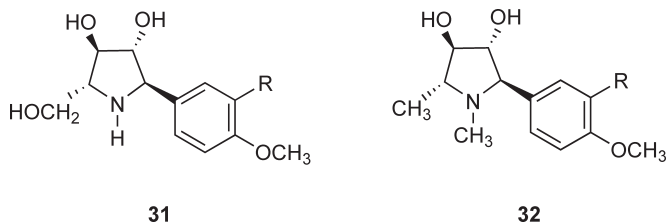


Broussonetines I (**28**, R = H, R' = Ac), J (**28**, R = R' = Ac), J₁ (**28**, R = Ac, R' = H), and J₂ (**28**, R = R' = H) have a terminal piperidine ring but this has no hydroxy groups appended, thus averting potential confusion as to whether these four alkaloids should be classified as polyhydroxypiperidines. Broussonetines U (**29**) and U₁ (**30**) are pyrrolines, carrying unsaturation within the pyrrolidine ring, and excluding these two alkaloids, 21 of the remaining 30 alkaloids have the same configuration (2*R*,3*R*,4*R*,5*R*) of the substituents around the pyrrolidine ring.

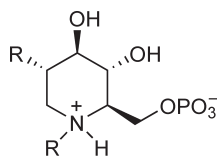


Two new trihydroxypyrrolidine alkaloids with an aromatic substituent, rather than an aliphatic side chain as in the broussonetinines, have been isolated from *Lobelia chinensis* (Campanulaceae). These alkaloids, named radica-mines A and B, were identified as 5-(3-hydroxy-4-methoxyphenyl)- and 5-(4-hydroxyphenyl)- derivatives of

(2*S*,3*S*,4*S*,5*S*)-2-hydroxymethyl-3,4-dihydroxy-pyrrolidine, respectively, by spectroscopic analysis.²⁷ The benzoate chirality method was used to support the assigned absolute configurations but a subsequent enantioselective synthesis established the absolute configurations of the pyrrolidine ring as (2*R*,3*R*,4*R*,5*R*), leading to the structures for radicamine A as (**31**, R = OH) and radicamine B as (**31**, R = H).²⁸ The radicamines have structural analogies to the *N*-methylated dihydroxypyrrolidine alkaloids (–)-codonopsinine (**32**, R = H) and (–)-codonopsine (**32**, R = OCH₃), which were isolated over 40 years ago from the Asian bellflower, *Codonopsis clematidea*, also a member of the Campanulaceae.^{29,30} The absolute stereochemistry of these two alkaloids was subsequently established as (2*R*,3*R*,4*R*,5*R*) by synthesis^{31,32} and X-ray crystallography.³³



Screening of a number of marine sponge species (primarily *Haliclona* and *Raispalia* spp.) exhibiting glycosidase inhibitory activity of their extracts, from diverse locations in Western Australia, Florida, Bahamas, and Micronesia, showed that they variously contained 1,4-dideoxy-1,4-imino-arabinitol (D-AB1) and its *xylo* isomer.³⁴ Two additional isomers, tentatively identified as the *ribo* and *lyxo* diastereomers, were also isolated and the configuration of the isolated D-AB1 was established by Marfey's method.³⁵ Additional polyhydroxy alkaloids occurring in sponges are the two new piperidine alkaloid derivatives 1-deoxynojirimycin-6-phosphate (**33**, R = H) and *N*-methyl-1-deoxynojirimycin-6-phosphate (**33**, R = CH₃) recently isolated from *Lendenfeldia chondrodes* collected from Yap, Micronesia.³⁶ However, the phosphate derivative showed greatly diminished α -glucosidase activity ($\sim 8000\times$ less) relative to the parent compound.³⁷

**33**

6.08.2.1.2 Bicyclic alkaloids

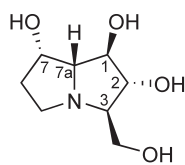
The bicyclic alkaloids fall into three structural classes: pyrrolizidines, indolizidines, and nortropanes. The pyrrolizidines have two fused five-membered rings with a bridgehead (tertiary) nitrogen atom; these may be formally regarded as an amalgamation of two hydroxylated pyrrolidine ring systems with a common nitrogen atom. In an analogous manner to the pyrrolizidine alkaloids, the indolizidine group may be formally regarded as a pyrrolidine ring fused with a piperidine ring, yielding a bicyclic 5/6 ring system. In contrast, the nortropane ring system can be visualized as fusion of a five-membered pyrrolidine ring with a six-membered piperidine ring at the positions α to the nitrogen atom of each monocyclic system; the nitrogen atom is therefore secondary, rather than tertiary, as in the pyrrolizidines and indolizidines.

Naturally occurring polyhydroxyquinolizidine alkaloids, analogous to the pyrrolizidines and indolizidines but possessing a 6/6 fused ring system, have yet to be discovered. However, a number of members of this class have been synthesized³⁸ and shown to be glycosidase inhibitors,² and their ultimate discovery should not be unexpected.

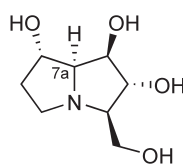
6.08.2.1.2(i) Pyrrolizidines In considering the polyhydroxypyrrolizidines, it is important to note that a very large class of widely distributed plant-derived pyrrolizidine alkaloids exist but these are much more complex in structure, consisting of a pyrrolizidine core (necine base) usually only dihydroxylated and possessing a double bond, which occur naturally mono- or diesterified with a variety of aliphatic necic

acids.³⁹ There is no evidence that these necine bases, which can be obtained from the parent alkaloids by hydrolysis, have any appreciable glycosidase inhibitory properties.

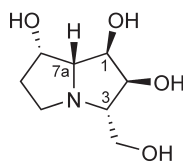
Polyhydroxypyrrrolizidine alkaloid glycosidase inhibitors have either four or five hydroxyl groups. Australine (**34**) was the first tetrahydroxylated pyrrolizidine isolated, co-occurring with the indolizidine alkaloid castanospermine in the monotypic leguminous tree *Castanospermum australe*,⁴⁰ whereas alexine (**35**) was isolated from the closely related *Alexa* species.⁴¹ Subsequently, the 1- and 3-epimers of australine were reported,^{42,43} and 7-*epi*-australine tentatively identified,⁴² but the latter was ultimately shown to be identical to australine by rigorous analysis of nuclear magnetic resonance (NMR) data⁴⁴ for both synthetic⁴⁵ and natural australines. This stimulated a careful search for other polyhydroxylated pyrrolizidines in *C. australe* seeds, leading to the identification of 2,3-*diepi*-australine (**36**), 2,3,7-*triepi*-australine (**37**), and 1-*epi*-australine 2-*O*- β -D-glucopyranoside (**38**).⁴⁶



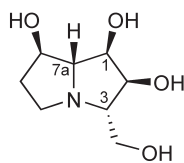
34



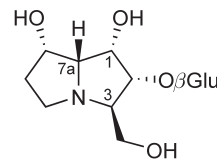
35



36

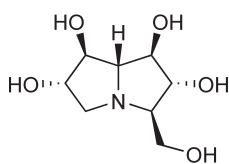


37

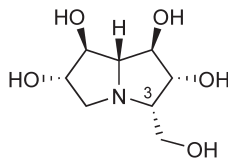


38

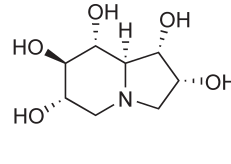
Pentahydroxy pyrrolizidine alkaloids are typified by casuarine (**39**)⁴⁷ and 3-*epi*-casuarine (**40**), the latter recently isolated from myrtle (*Myrtus communis*; Myrtaceae).⁴⁸ Uniflorine A from *Eugenia uniflora* (Myrtaceae), commonly known as the Surinam cherry or pitanga, was originally formulated as the indolizidine alkaloid (**41**)⁴⁹ but has recently been established as 6-*epi*-casuarine (**42**) by total synthesis.⁵⁰ Its congener uniflorine B, originally defined as a structurally improbable 5-hydroxyindolizidine (**43**), has correspondingly been shown to be spectroscopically identical to casuarine.⁵¹



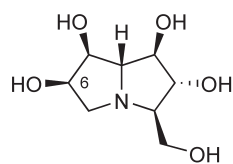
39



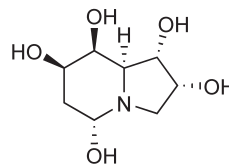
40



41



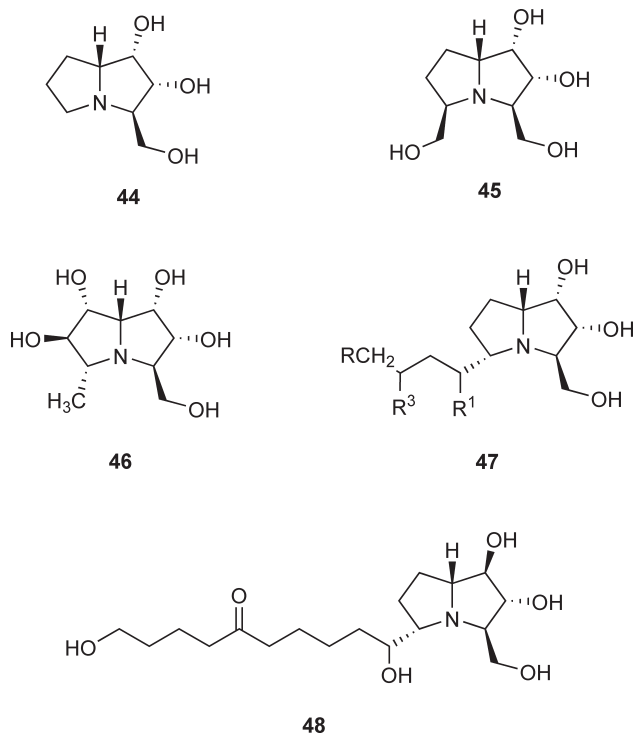
42



43

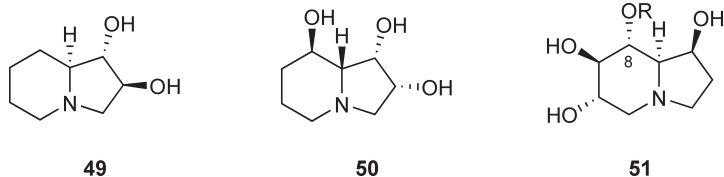
Other polyhydroxylated pyrrolizidines are the hyacinthacines, named after their occurrence in members of the plant family Hyacinthaceae (*Hyacinthoides non-scripta* and *Scilla campanulata*).⁹ The stereochemical disposition of substituents around the pyrrolizidine core was determined by nOe effects in the NMR spectra, but the

specific enantiomers were not established. Subsequently, four additional structurally analogous alkaloids were isolated from *Muscari armeniacum*,⁵² a close relative of *H. non-scripta*, and various epimeric hyacinthacines have also been discovered in *S. sibirica* and *Scilla socialis*.^{11,13} The letter designations after the trivial name conform to the convention used for the calystegines (*vide infra*), with 'A' corresponding to trihydroxylated, 'B' to tetrahydroxylated, and 'C' to pentahydroxylated alkaloids. Representative structures are hyacinthacines A₁ (**44**), B₁ (**45**), and C₁ (**46**). The total number of alkaloids designated as hyacinthacines now numbers 18 (A₁₋₇, B₁₋₆, and C₁₋₅), together with further four alkaloids bearing longer side chains at C-5, such as the 3-hydroxybutyl- (**47**, R = R¹ = H, R³ = OH), 1,3-dihydroxybutyl- (**47**, R = H, R¹ = R³ = OH), and 1,3,4-trihydroxybutyl- (**47**, R = R¹ = R³ = OH) derivatives of hyacinthacine A₁, co-occurring with α -5-C-(3-hydroxybutyl)-7-*epi*-australine in *Scilla peruviana*⁵³ and α -5-C-3-hydroxybutylhyacinthacine A₂ found in *S. socialis*.



These hyacinthacines and australines with C-5 side chains have structural features in common with broussonetine N (**48**),²³ a pyrrolizidine alkaloid bearing a 10-carbon side chain at the same position isolated from *B. kazinoki*. These 10 carbon atoms, in combination with the three carbon atoms of the nonhydroxylated moiety of the pyrrolizidine core, would correspond to the 13 carbon atoms found in the pyrrolizidine-based broussonetines (**14–30**). This implies biosynthetic formation from a presently unknown precursor through cyclization onto the nitrogen atom of the pyrrolizidine ring. A possible candidate might be an analogue of broussonetine V (**21**),²⁴ possessing an appropriate stereochemistry of substituents on the pyrrolizidine ring, because this alkaloid has a double bond functionalizing the appropriate 3'-position of the 13-carbon side chain.

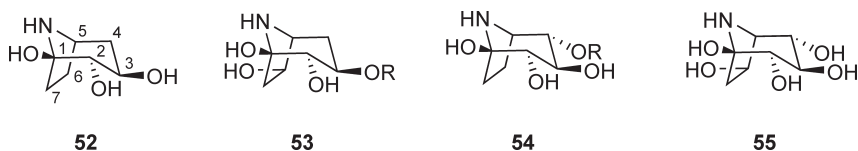
6.08.2.1.2(ii) Indolizidines Polyhydroxy alkaloids belonging to the indolizidine subclass are epitomized by the dihydroxylated lentiginosine (**49**), trihydroxylated swainsonine (**50**), and tetrahydroxylated castanospermine (**51**, R = -H); several epimers of the latter are known. Swainsonine and castanospermine have generated considerable interest due to the specificity of their glycosidase inhibition and consequent therapeutic potential, which is discussed later. Swainsonine is noteworthy within the indolizidine class as the only member with an 8a-*R* bridgehead configuration; although the lentiginosines co-occur with swainsonine, they have the opposite configuration at this position. Castanospermine 8-*O*- β -D-glucopyranoside (**51**, R = - β -D-glucopyranosyl) was first prepared by synthesis⁵⁴ but subsequently isolated from *C. australe* seeds.⁴⁶



There are no known naturally occurring pentahydroxyindolizidines, although several have been synthesized. Uniflorines A and B from *Eugenia uniflora* were originally formulated as 1,2,6,7,8-pentahydroxyindolizidine (**41**) and the structurally improbable 5-hydroxylated indolizidine (**43**), respectively, but these have now been shown to be the polyhydroxypyrrrolizidines, 6-*epi*-casuarine and casuarine (**39**).

6.08.2.1.2(iii) Nortropanes The most recently discovered group of bicyclic polyhydroxy alkaloids are the calystegines, based on the nortropane ring system, with a secondary nitrogen atom, rather than tertiary as in the pyrrolizidine and indolizidine classes. The large number of known tropane alkaloids with a tertiary N-methylated nitrogen atom adds a complication as to whether the calystegines should simply be regarded as demethylated tropanes but it is likely that they are biosynthesized through a modification of the normal tropane biosynthetic pathway and therefore constitute a unique class.⁵⁵ Furthermore, typical tropane alkaloids such as scopolamine and hyoscyne are less highly hydroxylated and occur as esters. In this respect, the situation is somewhat reminiscent of the distinction between the widely distributed esterified pyrrolizidine alkaloids and the polyhydroxypyrrrolizidines. Another unique feature is the presence in the calystegines of a 1-hydroxy group, leading to an aminoketal functionality. This functionality raises the question as to whether partial equilibration with the aminocycloheptanone structural form could occur.

The calystegines may be tri-, tetra-, and pentahydroxylated and have been conveniently divided into the calystegine A, B, and C subclasses, respectively, on the basis of the degree of hydroxylation. The most commonly found within each of these groups are calystegines A₃ (**52**), B₁ (**53**, R = H) and B₂ (**54**, R = H), and C₁ (**55**). Inspection of these structures reveals that a considerable diversity of compounds can theoretically result from different positional and stereochemical distribution of the hydroxyl groups, and at least 13 calystegines have so far been isolated from a variety of plant sources, together with calystegine N₁, which has a 1-amino substituent. However, there is a possibility that the latter could be an artifact of calystegine B₂, resulting from the use of ammonia in the isolation process, and its actual existence as a natural product needs to be established by demonstration of its existence within the plant. Two calystegine glycosides have been isolated, namely calystegine B₁ 3- β -glucoside (**53**, R = glucose) and calystegine B₂ 4- α -galactoside (**54**, R = galactose).¹⁶ Various aspects of the chemistry, biosynthesis, and biological activity of calystegines have been reviewed by Dräger.⁵⁵



The question as to whether other polyhydroxy alkaloids with similar structures as the calystegines should be properly regarded as nortropanes or merely as demethylated tropanes is difficult. Factors to be considered for their inclusion might be occurrence in plant families known to produce calystegines or other glycosidase inhibitors, lack of esterification, and glycosidase inhibitory activity. Evidence for their production by a biosynthetic route different from that of the tropanes would be conclusive proof for their inclusion as nortropanes. If the criterion is that they should also possess a 1-OH group, as in the calystegines, then all other known di- and trihydroxy nortropanes would be excluded; therefore, it seems most appropriate that the name calystegine be reserved for those alkaloids with the aminoketal functionality.

6.08.2.2 Occurrence and Isolation from Natural Sources

6.08.2.2.1 Occurrence

Polyhydroxy alkaloid glycosidase inhibitors have been isolated primarily from plant sources, and to a lesser extent microorganisms, but have occasionally been found in insects. The pyrrolidine alkaloid DMDP (**1**, R = OH) is particularly widely distributed and has been discovered in many disparate genera of the plant families Araceae, Campanulaceae, Euphorbiaceae, Hyacinthaceae, and Leguminosae, as well as in a *Streptomyces* species.⁵⁶ It has also been found in the body, wings, pupae, and eggs of the neotropical day-flying moth (*Urania fulgens*) together with HNJ, both being constituents of the insect's food plant, *Omphalea diandra*.⁵⁷ The insect appears to selectively concentrate these particular alkaloids, with levels 10-fold those in the plant, whereas other polyhydroxy glycosidase inhibitors occurred at approximately the same concentrations in both the moth and the plant. Similarly, silkworms (*Bombyx mori*) concentrate DNJ almost threefold relative to the levels in mulberry leaves on which they feed.¹⁵ Although DNJ is a potent inhibitor of the rat intestinal maltase, the midgut maltase of the silkworm is 400 times less sensitive to inhibition by the alkaloid, suggesting adaptation as a specialized herbivore, little affected by antifeedant or growth inhibitory properties that the compound might have. There is no evidence to date that insects produce such glycosidase-inhibitory alkaloids *a priori* and acquisition through the food plants appears to be the most probable route. The aposematic coloring of *U. fulgens* and its ability to fly during the day indicates that the acquired alkaloids serve a protective function against predation by birds.

DNJ has also been isolated from the lichen *Umbilicaria esculenta*⁵⁸ and its phosphate derivative from the marine sponge *L. chondrodes*.³⁶ Lichens are symbiotic associations of a fungus with a photosynthetic phycobiont and sponges commonly have endophytic fungi associated with them, which raises a question as to whether or not the alkaloid produced in these cases is a fungal metabolite. This possibility was proposed for the occurrence of D-AB1 and its diastereomers in other marine sponges.³⁴

At the time of the initial discovery of the first members of the bicyclic pyrrolizidine and indolizidine classes, there appeared to be a taxonomic relationship with plant genera such as *Swainsona*, *Astragalus*, *Oxytropis*, and *Castanospermum* in the family Leguminosae. This relationship can no longer be considered to exist, as the alkaloids have been discovered in an increasingly diverse number of plant families. For example, among the polyhydroxy pyrrolizidines, although the australines (**34–38**) are restricted to the legume *C. australe*, the structurally closely related casuarines (**39, 40, 42**) occur in the Casuarinaceae and Myrtaceae and the hyacinthacines (**44–47**) in the Hyacinthaceae. Similarly, the nortropanoid calystegines were initially found in *Calystegia* species of the Convolvulaceae^{59,60} but have now been identified in 14 additional genera of this plant family,⁶¹ especially *Ipomoea* species.^{62–65} Furthermore, they are now known to occur to some extent in Moraceae, and in both wild and cultivated *Erythroxylum* spp. of the Erythroxylaceae,⁶⁶ and to be widespread in the Solanaceae⁶⁷ and Brassicaceae.⁶⁸ The latter two plant families encompass important vegetable and fruit species such as potatoes, eggplants, and cabbages, which when considered together with their presence in mulberries and sweet potatoes (*Ipomoea batatas*) raises questions as to the overall intake and relative safety of calystegines in the human diet.⁶⁹ Calystegines have also been isolated from the Death's head hawkmoth (*Acherontia atropus*), presumably by larval acquisition from its solanaceous food plants, especially nightshades and potatoes, but preliminary evidence⁷⁰ suggests that the alkaloids consumed by larvae of the Convolvulus hawkmoth (*Agrius convolvuli*) are excreted and that none are transferred to the adult. This is consistent with the coloring of the two moth species, with the latter being much more cryptically colored, presumably because the inability to accumulate the calystegines results in a lack of protection against predation.

The situation with regard to the indolizidine alkaloid swainsonine (**50**) is particularly intriguing. Following its first discovery in the Australian legume genus *Swainsona*,⁷¹ it was subsequently found in North American *Astragalus* and *Oxytropis* species, known as locoweeds, and then in other members of the same genera in South America and Asia.^{72,73} It therefore seemed reasonable to assume that swainsonine was a characteristic alkaloid of certain genera of the Leguminosae. However, it was identified shortly thereafter in the unrelated microorganisms, *Rhizoctonia leguminicola* and *Metarhizium anisopliae*. More recently, swainsonine has been shown to co-occur with calystegines in *Ipomoea* species^{62–64,74,75} and in the Brazilian plant species, *Turbina cordata* (syn. *Ipomoea martii*) (Convolvulaceae)⁷⁶ and *Sida carpinifolia* (Malvaceae),⁷⁷ and even in the primitive plant *Lepidozamia peroffskyana*, a cycad.⁷⁸

Such widespread occurrence both geographically and across species from several different plant families, together with its isolation from both plants and microorganisms, raises questions as to its taxonomic significance as a phytochemical. Furthermore, its biosynthesis was shown to proceed by an identical pathway in both Diablo locoweed (*Astragalus oxyphythus*) and *R. leguminicola*.⁷⁹ Subsequent studies of its occurrence in white locoweed (*Oxytropis sericea*) have shown that it occurs in some populations whereas others are completely devoid of the alkaloid.⁸⁰ This led to the isolation of an endophyte from the locoweeds *O. sericea*, *Oxytropis lambertii*, and *Astragalus mollissimus* containing the alkaloid but not from those in which swainsonine was absent.⁸¹ When cultured *in vitro* the fungus, initially tentatively identified as an *Alternaria* but subsequently as an *Embellisia* species, produced significant quantities of swainsonine. In a comparative study, rats fed on the fungus and a locoweed (*O. lambertii*) showed indistinguishable symptoms, indicating that the plant is not an essential requirement for toxicity.⁸² A survey of major locoweed species has established the presence of the endophyte in *Astragalus lentiginosus*, *A. mollissimus*, *Astragalus pubentissimus*, *Astragalus wootoni*, and *O. sericea*.⁸³ It seems unlikely that the same endophyte is responsible for production of the alkaloid in swainsonine-containing plant species from diverse genera and in many other parts of the world, such as Australia, South America, and Asia, suggesting that there are probably numerous species of endophytes with the capability to produce it. Defining the endophytes and their plant associations may prove to be a fruitful area of future research.

The possibility that swainsonine is not a phytochemical metabolite raises questions as to the true source of structurally related glycosidase inhibitors such as the castanospermines and australines. One major difference is that the latter have so far been found to be highly restricted in their occurrence, having been isolated only from *C. australe* and *Alexa* spp., both closely related genera in the Leguminosae. This chemotaxonomic evidence suggests that castanospermines and australines are true phytochemicals and not endophytic metabolites. The situation with regard to swainsonine and the nortropane calystegines is much more problematic, as both classes of alkaloid can co-occur in *Ipomoea* species and are found at somewhat similar levels. In roots of *Calystegia sepium* producing the alkaloids a calystegine-catabolizing strain of *Rhizobium meliloti* is found but this is not present within plants that do not produce calystegines.⁸⁴ This suggests a strong symbiotic relationship but whether this is plant–fungal or inter-fungal is not known. On the contrary, the diversity of calystegine structures and occurrence, and their structural similarities to the tropane alkaloids, leads to the first assumption that they are true phytochemicals. This is supported by a recent comparative study of their content and composition in potatoes (*Solanum tuberosum*) from the groups Phureja and Tuberosum.⁸⁵ In whole tubers their concentration was of a similar order of magnitude in both groups, and the most abundant were calystegines A₃ and B₂. Calystegine concentrations in the peel and sprouts were up to 13 and 100 times higher, respectively, than in the flesh. Such considerations lead to the conclusion that calystegines are unlikely to be endophytic natural products. The reason for a plant–endophyte association to produce structurally different polyhydroxy alkaloids, each with specific glycosidase inhibitory properties remains subject to conjecture at the present time.

It now appears that alkaloidal glycosidase inhibitors may be widely distributed, and their discovery has been limited by their high water solubility and consequent failure to isolate them by conventional alkaloid extraction techniques. Although certain plant families such as Leguminosae, Solanaceae, and Convolvulaceae are obvious sources, general taxonomic predictions as to their distribution are probably of little value at the moment, and the possibility of endophyte involvement should always be considered. An alternative explanation for their widespread occurrence, consistent with their high water solubility, could be that they are metabolites of microorganisms in the rhizosphere that migrate to and are taken up by the roots and transported to other plant organs.

6.08.2.2.2 Isolation and structural determination

The polarity and hydrophilicity of polyhydroxy alkaloids require that different techniques be utilized for their isolation and purification than those usually adopted for conventional alkaloids that are soluble in organic solvents. These techniques have been extensively reviewed.^{78,86–88} In general, the alkaloids are extracted from the plant source by polar solvents such as water, methanol, or ethanol, then purified and at least partially separated by ion-exchange chromatography. Typically, strongly acidic cationic exchange resins are used to retain the basic alkaloids, whereas neutral and acidic compounds such as sugars and phenolics are eluted with water. Subsequent separation can be achieved by further chromatography on either weakly acidic Amberlite cationic CG50 (NH₄⁺ form) or strongly basic anionic CG400 (OH⁻ form) resins, eluted with water/aqueous ammonia or acetic acid, respectively. Specific adaptations of these resins and eluants may be required for each

plant matrix and the publications of Asano and coworkers provide useful models for separation of large numbers of co-occurring alkaloids from particular sources. Thin-layer chromatography (TLC), especially the centrifugal radial modification, is an alternative to ion-exchange chromatography for obtaining usable quantities of the alkaloids for biological testing.

One problem associated with the polyhydroxy alkaloids is that they are often not amenable to detection by conventional alkaloid colorimetric reagents such as Dragendorff's. In some cases, such as the indolizidines, their elution can be monitored by specific spray reagents.⁸⁹ Monitoring of eluants by glycosidase inhibition is problematic since it is time-consuming and difficult to set up a battery of assays to cover all of the enzymes that might be inhibited and other constituents, such as phenolic compounds, may have nonspecific inhibitory properties. To avoid misleading results the best practice would seem to be to determine glycosidase inhibition only with completely purified alkaloids.

Gas chromatography with mass spectrometric (GC-MS) detection has been a useful technique for establishing purity and gross structural features of the polyhydroxy alkaloids. Because the alkaloids are so polar, they must be derivatized to ensure that they are sufficiently volatile and for this purpose trimethylsilylation has been most successfully applied. Numerous reagents exist that are capable of forming trimethylsilyl (TMSi) derivatives of hydroxyl groups and/or primary and secondary amino groups. The most commonly used are hexamethyldisilazane (HMDS)/trimethylchlorosilane (TMCS) or *N*-methyl-*N*-trimethylchlorosilyl-fluoroacetamide (MSTFA), the latter having the advantage that all of the reaction by-products are volatile. In general, the GC retention times increase in proportion to the overall number of -OTMSi groups, and some fragmentation patterns are quite diagnostic. For example, a peak at m/z 217 indicates the presence of three adjacent -OH groups.

Although high-performance liquid chromatography (HPLC) would appear to be well suited for analysis of polyhydroxy alkaloids, the technique is limited by the difficulty of detecting the compounds because of their lack of UV absorption. However, the interface with liquid chromatography-mass spectrometry (LC-MS) offers much promise in surmounting this problem, in addition to providing fundamental information as to the molecular size, ring configuration, and presence of glycoside derivatives which may have a high molecular weight to survive GC-MS, even when derivatized. Tandem mass spectrometry (LC-MS/MS) has been investigated for analysis of a mixture of 12 polyhydroxy alkaloids and atmospheric pressure ionization (APCI) compared to electrospray ionization (ESI).⁹⁰ Negative mode APCI was found to be the most useful as it generated deprotonated molecular ions, the collision-induced dissociation (CID) spectra of which were particularly diagnostic of isomeric alkaloids. The technique has been successfully applied to a crude extract of bluebells (*H. non-scripta*), indicating the presence of two glycosides that were not detected by GC-MS.⁹¹ LC-MS with selected-ion monitoring has also been used to establish the configuration of isolated polyhydroxypyrrolidines by discriminating between derivatives prepared with Marfey's reagent.³⁴

Although MS provides valuable information regarding the gross structure of the alkaloids, the density of chiral centers in such relatively small molecules requires the power of NMR spectroscopy to elucidate the relative stereochemistry of the substituent hydroxyl groups. Because chemical shift values are highly dependent on pH of the NMR solution, this parameter has to be carefully controlled in measuring spectra and direct comparison with literature data for authentic samples can often be misleading. However, extraction of ^1H - ^1H three-bond coupling constants ($^3J_{\text{HH}}$) has proved useful, particularly for five-membered ring systems for which one-bond coupling constants ($^1J_{\text{HH}}$) are unreliable. The value of this technique has been demonstrated with australine diastereomers.⁴⁴ Determination of the absolute stereochemistry is a more difficult problem and usually depends upon X-ray crystallography, circular dichroism techniques, or stereospecific synthesis. It should be recognized that published structures often only represent the most likely enantiomers, based on biosynthetic considerations and congruence with previously isolated compounds in the particular class.

6.08.3 Glycosidase Inhibition

6.08.3.1 Glycosidase Inhibitory Activity

The glycosidase inhibitory specificity and potency of polyhydroxy alkaloids is highly dependent on the origin of any given enzyme, its purity, and conditions such as pH under which the assay is performed. For example, alkaloids that strongly inhibit yeast α -glucosidase (amyloglucosidase) may show little or no activity against

α -glucosidases from other sources and vice versa. Because of this it is often difficult to compare data for a newly isolated alkaloid with that in the literature or with another laboratory. Ideally, comparison of alkaloids having new or known structures with standards should be made side-by-side. It is also essential to ensure that assays are only performed on carefully purified alkaloids, since the presence of a minor component with very potent activity may greatly distort the inhibition profile. In spite of such inherent quantitative variation, in simple qualitative screening assays the specific enzyme(s) against which an individual alkaloid shows activity will usually be consistent from one laboratory to another. Nevertheless, it is important to examine all relevant publications on a particular alkaloid before firm conclusions are drawn as to its glycosidase inhibitory properties.

It is now recognized that relationships between alkaloid structure and enzyme inhibitory activity are impossible to predict *a priori*. This did not appear to be the case in the earliest phases of research on the group. For example, among the indolizidine alkaloids, swainsonine (**50**) has the same stereochemical arrangement of hydroxyl groups as in D-mannopyranose but with substitution of a nitrogen atom for the heterocyclic oxygen of the sugar. Its ability to inhibit α -mannosidase⁹² was therefore not unexpected. Similarly, castanospermine (**51**) appeared to be an aza-analogue of glucose and inhibited glucosidases as predicted.⁷³ However, this simplistic model soon failed when epimers of castanospermine and australines, the latter with a pyrrolizidine ring system, were isolated from *C. australe*, and all were found to inhibit amyloglucosidase regardless of the stereochemistry of their hydroxyl groups.^{40,42,43,93–95} It is interesting to speculate why the plant would biosynthesize so many isomers having the same inhibitory activity, with the major constituents of the indolizidine and pyrrolizidine classes, castanospermine and australine, ultimately predominating and showing the highest activity (K_i 8 and 6 $\mu\text{mol l}^{-1}$, respectively). A similar situation exists with regard to the calystegines, in which in spite of differing numbers, disposition, and stereochemistry of hydroxyl groups, almost all members of the group inhibit either β -glucosidase and/or α - and β -galactosidases.⁹⁶ On the contrary, calystegine C₂ is an aza-mannose (Man) analogue with respect to the six-membered ring moiety and is an inhibitor of α -mannosidase.⁹⁷

In the absence of obvious structure–activity relationships, molecular modeling methods need to be applied either to a series of structurally related alkaloids, such as the calystegines, or else to a group of inhibitors of a particular enzyme that cover a number of structural classes. In general, either of these approaches will require the synthesis of non-natural analogues to obtain a large enough library of compounds to draw substantive conclusions. An early attempt in this direction was the glycosidase inhibition comparison of 20 natural and synthetic analogues of DNJ, including some that were zwitterionic in character, and others in which the nitrogen atom was substituted by sulfur at various oxidation states.⁹⁸ This study indicated that for good activity the inhibitor should have a high positive charge on the anomeric carbon and the heteroatom and a half-chair conformation, or a chair conformation with the same orientation of hydroxyl groups as in the corresponding monosaccharide. The significant influence of ring conformation and charge may explain why certain structure–activity relationships are obvious from inspection of the alkaloid structure, whereas such simple correlations do not exist for similar gross structures. For pyrrolidine alkaloids, it has been shown by comparison of a series of enantiomers, epimers, and sulfur analogues of D-AB1 that N-alkylation and the presence of sulfate in the side chain are important factors in the specificity of amyloglucosidase inhibition.⁹⁹ Furthermore, L-AB1 was 24 times less effective than the D-enantiomer. The trend for D-pyrrolidine alkaloids to be competitive inhibitors of D-glycosidases in contrast to their L-enantiomers, which were noncompetitive inhibitors, whereas for iminofuranoses the L-enantiomers were more potent than the D-enantiomers had been recognized earlier^{100,101} and has been confirmed by synthesis of two-acetamido derivatives.¹⁰²

Generally accepted models for glycosidase inhibition fit best for the calystegines, which is perhaps not surprising since they provide a reasonable complete series of structurally related polyhydroxy alkaloids. This model requires the presence of two carboxylic acid groups at the active site of the enzyme for β -glucosidase inhibition, each with a specific function. The first is required for generation of the glycosyl cation intermediate, and the other for its stabilization.³⁵ It appears that calystegines B₁ (**53**, R = –H) and C₁ (**55**) may be protonated by the acidic group responsible for catalytic activity within the active site at the *exo* six-hydroxyl whereas calystegine B₂ (**54**, R = –H), which also shows similar inhibitory activity toward β -glucosidase, is bound to the glucosyl cation binding site through the four-hydroxyl group. Equatorial hydroxyl groups at the 2- and 3-positions are essential requirements. However, the mechanism of galactosidase inhibitory activity cannot be

explained in a similar manner, particularly since calystegine B₃ is the closest configurational analogue to D-galactose (Gal) but has no inhibitory activity against these enzymes.⁹⁶ Synthesis of non-natural analogues has provided some insight to the (–)-enantiomer of calystegine B₂ exhibiting no inhibition of any of the tested glycosidases and N-methylation of natural (+)-calystegine B₂ suppressing inhibition of β-glucosidase but with retention of activity toward α-galactosidase.¹⁰³ Incorporation of information of this type with detailed knowledge of the catalytic site for each of the enzymes will ultimately permit design and synthesis of the most effective inhibitors.

6.08.4 Biological Activity of Glycosidase Inhibitors

6.08.4.1 Mammalian Toxicity

Most of the early research on glycosidase inhibitors was driven by a need to identify the plant toxins responsible for poisoning of livestock.¹⁰⁴ A major stimulus was the isolation of swainsonine (**50**) from the poison peas (*Swainsona* spp.) of Australia⁷¹ and its discovery, together with lentiginosine and 2-*epi*-lentiginosine, in the locoweeds (*Astragalus* and *Oxytropis* spp.) of North America.^{72,105} Subsequently, the alkaloid was identified in *Oxytropis* spp. from South America¹⁰⁶ and parts of Asia,¹⁰⁷ thereby establishing ‘locoism’ as a worldwide phenomenon. The primary symptom of locoweed poisoning is neurological damage and this is fully accounted for biochemically and histopathologically by the potent α-mannosidase inhibitory activity of swainsonine, which disrupts glycoprotein processing by mannosidase II in the Golgi and causes neuronal vacuolation due to accumulation of mannose-rich oligosaccharides.¹⁰⁸ The clinical effects associated with locoweed consumption extend beyond neuronal damage and can vary markedly with animal species, sex, age, nutritional status, and environment, although the involvement of swainsonine in such effects is much less clearly defined.¹⁰⁹ Commonly observed signs of poisoning are emaciation, reproductive failure in both males and females, and congestive right-heart failure in grazing at high altitude.¹¹⁰ Maternal ingestion of swainsonine can have marked effects on the behavior of the offspring,¹¹¹ and the alkaloid appears to be transferred to nursing animals in the milk.¹¹²

The discovery of swainsonine in *Ipomoea* species of Australia,⁶² Mozambique,⁶³ and South America,^{74–76,113} co-occurring with calystegines, presents a more complex scenario with respect to poisoning of livestock. The question arises as to the particular biochemical lesions and signs of poisoning specifically caused by swainsonine and calystegines, respectively. Typical neurological damage and cytoplasmic vacuolation of neurons were observed whenever swainsonine was detected in the plant but syndromes of *Ipomoea* poisoning also present a ‘star-gazing’ attitude, and muscle twitching, tremors, and epileptiform seizures, which are not characteristic of locoweed poisoning. The latter signs are typical of phenocopies of the human genetic lysosomal storage defects Gaucher’s disease and Fabry’s disease, which might be predicted to be produced by inhibition of β-glucosidase and α-galactosidase by the calystegines. Histological examination of animal tissues showed vacuolation of Purkinje cells, also not seen in locoweed poisoning. However, feeding experiments in goats with *I. asarifolia* produced a tremorgenic syndrome, even though calystegines were absent from the plant and swainsonine occurred only in trace amounts (<0.01%).¹¹⁴ This suggests that there may be additional unidentified neurotoxins in *Ipomoea* species other than the glycosidase inhibitory alkaloids. A recent study in mice of the comparative pathology of swainsonine, castanospermine, and calystegines A₃, B₂, and C₁ supports this finding with swainsonine and castanospermine producing clinical and histological changes typical of their anticipated toxicity, whereas calystegine A₃ at the highest dose (100 mg kg⁻¹) caused only minor hepatic changes.¹¹⁵ However, rodents are resistant to induced lysosomal storage diseases and may therefore be a poor model for such comparisons. It is possible that calystegines do not play a primary role in livestock poisonings but rather have a secondary effect, perhaps by influencing the uptake and distribution of swainsonine or other glycosidase inhibitors through alteration of the activity of enzymes in the digestive system. Nevertheless, it is important that their precise influence be defined because they undoubtedly affect mammalian liver enzymes⁶⁹ *in vitro* and their presence in commonly consumed fruits and vegetables may have cryptic deleterious effects on human consumers.

Gastrointestinal problems are the most consistent feature of field cases of livestock (and occasional human) poisoning by *C. australe* in Australia. This is a predictable outcome of the ability of castanospermine (**51**), australine (**34**), and their isomers to inhibit α- and β-glucosidases. Although neurological damage has not been

reported, rodent feeding experiments with castanospermine produced hepatic vacuolation and glycogen accumulation, consistent with a phenotype of Pompe's disease or type II glycogenesis.¹¹⁶ Problems with digestion and ensuing lethargy have long been known to occur in livestock grazing English bluebells (*H. non-scripta*),¹¹⁷ and the presence of the glucosidase inhibitors, DMDP and homoDMDP, in the plant could be responsible for the syndrome.⁸ The hyacinthacines (44–47), more recently discovered in this plant, are only moderate inhibitors of the same enzymes and may be minor contributors to the problem.

The polyhydroxy alkaloids usually occur in the plants consumed by livestock at relatively low levels. This factor, in combination with their high water solubility, means that the ingested and absorbed dose is quite low, so that symptoms of poisoning develop relatively slowly. In this regard toxicity is much less dramatic than with many other alkaloidal toxins, where the animals may succumb within a few days or even hours of ingestion. Clinical signs may therefore only develop after several weeks and association of toxicity with a particular plant may be difficult or impossible. In other cases, consumption of the plant may be intermittent or short term with overt changes lacking. Toxicity may then only manifest as failure to thrive, minor digestive disturbances, or increased susceptibility to infection.

6.08.5 Structure and Biosynthesis of N-Linked Glycoproteins

6.08.5.1 Introduction

Glycoproteins represent a substantially large class of compounds since there are many different glycoproteins in a given cell, and these glycoproteins occur in all eukaryotic cells as well as in a number of prokaryotes.¹¹⁸ In addition, some of the individual glycoproteins may be present in significant amounts in a given cell. In fact, many of the proteins that are components of cell membranes are glycoproteins, and they may have essential functions as receptor molecules that capture various ligands for the cell, as transport proteins that are involved in intake of various compounds, or as structures that mediate molecular recognition and interactions between cells.¹¹⁹ Glycoproteins, of course, are proteins that contain covalently bound sugars, and these sugars may be present as a single monosaccharide attached to the protein, but more commonly they occur as oligosaccharides linked to the protein. The two most common types of linkages between protein and carbohydrate involve either O-linked oligosaccharides or N-linked oligosaccharides.

In the first case, the oligosaccharide is attached to the protein in an O-glycosidic linkage, whereby the sugar at the reducing end of the oligosaccharide is usually *N*-acetylgalactosamine (GalNAc), and it is attached to the hydroxyl group of a serine or threonine in the protein.¹²⁰ In the second case, called the N-glycosidic linkage, the sugar at the reducing end of the oligosaccharide is *N*-acetylglucosamine (GlcNAc), and this sugar is bonded to the amide nitrogen of an asparagine in the protein.¹²¹ The N-linked glycoproteins are the topic of this section since there are a number of naturally occurring, as well as synthetic, nitrogen-containing (i.e., alkaloidal-type) compounds that are potent and biologically useful inhibitors of enzymes that participate in the biosynthesis of these oligosaccharides.¹²² Our earlier version of this topic reviewed the activities of a number of these inhibitors on the synthesis of the N-linked-oligosaccharide chains.¹ Thus in this section, we will briefly review the structures and mechanisms of biosynthesis of these oligosaccharides, considering some of the newer studies that demonstrate the role of sugars in directing these glycoproteins to various sites in the cell and including the role of glucose and mannose in quality control in protein folding. Some of the alkaloid inhibitors cause alterations in the targeting of proteins and also affect the quality control system. In subsequent sections of this review, we point out how some of these inhibitors are tested as possible therapeutic agents against other types of complex carbohydrates, such as animal glycosphingolipids (GSLs), and their associated diseases, such as Gaucher and Tay–Sachs diseases.

The N-linked glycoproteins are involved, or play key roles, in many critical cellular functions including protein folding, protein targeting, receptor function, cell–cell recognition, protein stability, signaling, and other cellular functions. Although the carbohydrate portion of the glycoprotein does not play a key role in all of these interactions, specific carbohydrate structures are necessary in a number of these cases, such as protein folding, cell–cell recognition, protein degradation in the endoplasmic reticulum (ER), pathogenesis, and so on.¹²³ As demonstrated below, a number of the alkaloid glycosidase inhibitors have profound effects on a number of these physiological functions.

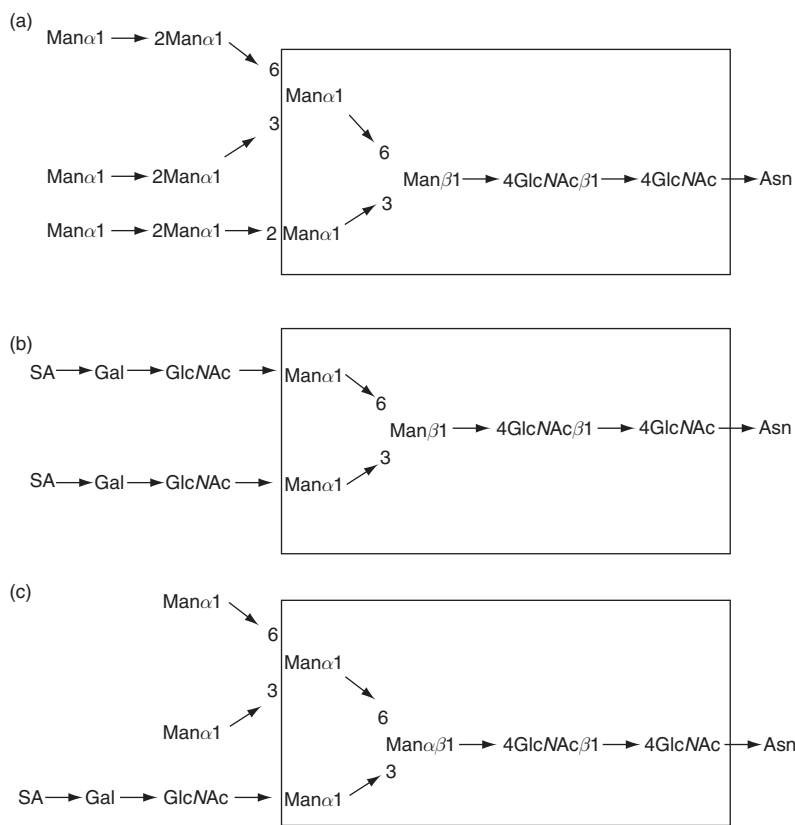


Figure 1 Generalized structures of N-linked oligosaccharides of glycoproteins. (a) High mannose type of oligosaccharide common in lower eukaryotes but also found in some mammalian glycoproteins. This oligosaccharide is the precursor to all other N-linked oligosaccharides. (b) Complex type of oligosaccharide. Complex oligosaccharides can have two branches (biantennary structure), three branches (triantennary), or four branches (tetraantennary). These types of structures are prominent on mammalian glycoproteins. (c) Hybrid type of oligosaccharide. These structures arise when the Golgi α -mannosidase II is inhibited (such as by swainsonine) or is absent (as in HEMPAS disease).

Generalized structures of the N-linked oligosaccharides are presented in **Figure 1**. The upper structure (**Figure 1(a)**) is referred to as a high mannose oligosaccharide since it contains nine mannose residues and two GlcNAcs. Eight of these mannoses are α -linked and are joined to each other in 1 \rightarrow 6, 1 \rightarrow 3, and 1 \rightarrow 2 glycosidic linkages.^{121,124} These mannoses are in a highly branched oligosaccharide that is then attached to the trisaccharide, Man β -1,4-GlcNAc β 1,4-GlcNAc, which connects this high mannose structure to specific asparagine residues in the N-linked glycoproteins.¹²⁴ These high mannose oligosaccharides are commonly found on glycoproteins of lower eukaryotes (yeast and fungi) and also in some viral glycoproteins, but are not so common in glycoproteins of mammalian cells.^{121,123} However, the high mannose oligosaccharide is also the precursor to all of the other N-linked oligosaccharides, and it is the structure that is initially synthesized in the ER and is attached to protein.¹²⁵ As outlined below, this high mannose structure is processed in the ER and Golgi apparatus of the cell by the sequential action of a number of glycosidases and glycosyltransferases that result in the production of the complex types of oligosaccharide. A typical example of a biantennary (two branches) complex oligosaccharide is shown in **Figure 1(b)**. In this case, six of the mannoses have been removed by glycosidases and replaced with other sugars, that is, GlcNAc, Gal, and sialic acid (NANA). Some complex chains can have three (i.e., triantennary complex structure) or four (tetraantennary) of these branches, and they can also contain other sugars, such as L-fucose (in place of sialic acid), or GalNAc, or other kinds of substitutions. A large number of different complex oligosaccharides have been isolated from various glycoproteins, and their structures have been determined.^{141,144,146}

Other N-linked oligosaccharide structures may also occur on specific glycoproteins or be produced under certain conditions. One such example is the hybrid-type oligosaccharide, as shown in **Figure 1(c)**. This structure is found in certain disease conditions, such as HEMPAS disease, where individuals lack the Golgi enzyme, mannosidase II (see discussion below), and therefore cannot remove the mannoses on the α 1,6 branch.¹²⁷ In that case, that branch cannot be processed further. Hybrid types of oligosaccharides are also produced in cultured animal cells when they are grown in the presence of the plant alkaloid, called swainsonine.¹²⁸ On the contrary, hybrid chains are not found on animal cell glycoproteins under normal conditions.

6.08.5.2 Biosynthesis of N-Linked Oligosaccharides

The biosynthesis of the N-linked oligosaccharide chain is initiated by the assembly of the high mannose type of oligosaccharide on the ER membrane (**Figure 2**). This assembly process involves the participation of a carrier molecule called dolichyl-phosphate (Dol-P), which is a component of the ER membrane, and is a lipid structure

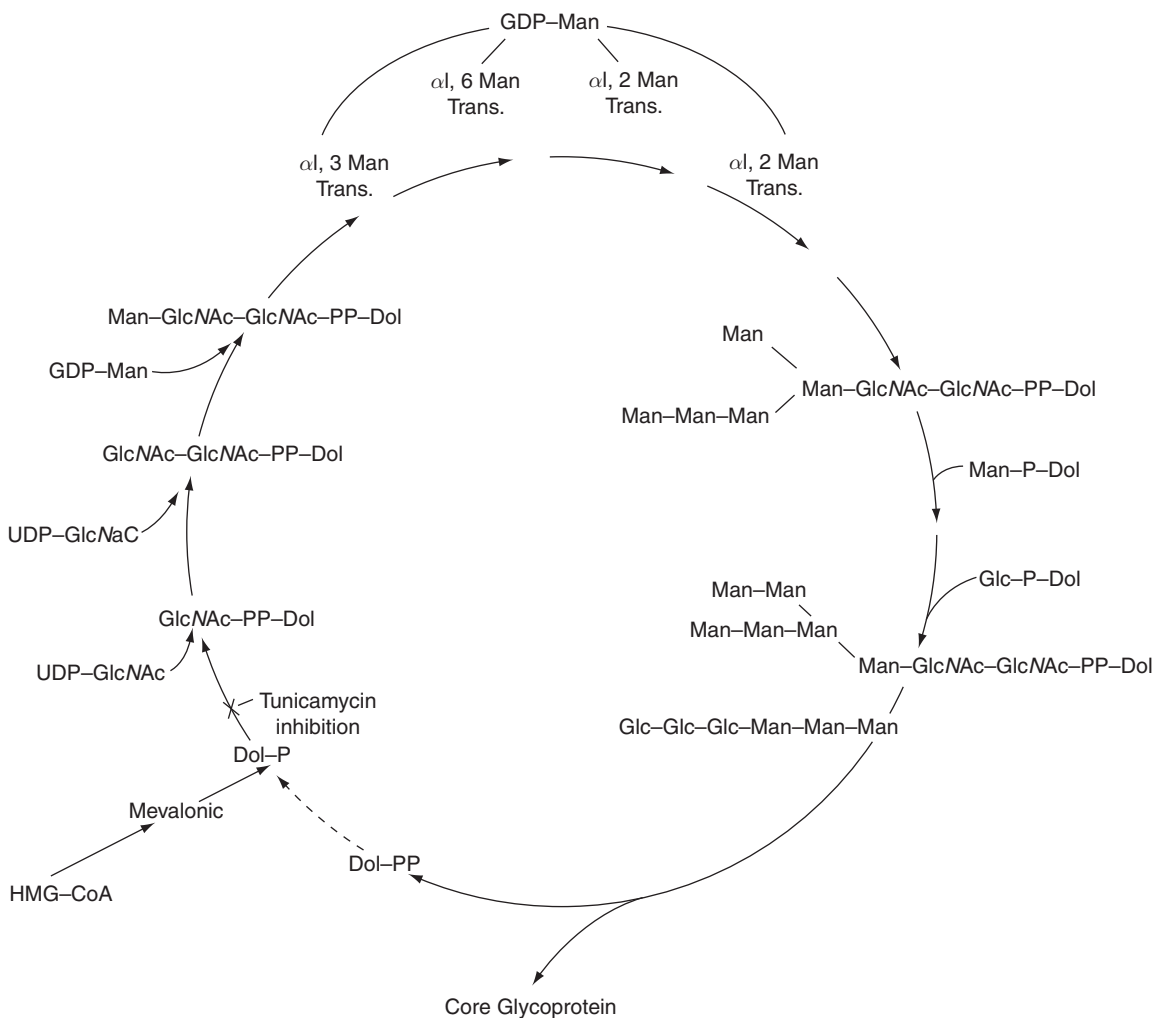


Figure 2 Pathway of biosynthesis of precursor lipid-linked oligosaccharide, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-dolichol}$. A series of glycosyltransferases, located on the membranes of the ER, sequentially transfer two **GlcNAcs**, nine **mannoses**, and three **glucoses** from their activated forms (**UDP-GlcNAc**, **GDP-Man**, **dolichyl-P-Man**, and **dolichyl-P-glucose**) to the membrane lipid, **Dol-P**, to form the precursor molecule, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-dolichol}$. The completed oligosaccharide is then transferred to specific asparagine residues of proteins as they are being synthesized in the ER. This transfer of oligosaccharides is cotranslational and is catalyzed by oligosaccharide transferase.

composed of some 20 polyprenol units linked together in a long linear chain.¹²⁹ Initially, this lipid is oriented with its phosphate group directed toward the cytoplasm, and the first seven sugars (i.e., two GlcNAcs and five mannoses) are added to the phosphate group of Dol-P through membrane-bound enzymes (glycosyltransferases), located on the cytoplasmic face of the ER. These enzymes utilize the sugar nucleotides, UDP-GlcNAc (uridine diphosphate *N*-acetylglucosamine) and GDP-Man, as substrates and they transfer these sugars in a sequential fashion to the Dol-P.^{129,130} Thus, a GlcNAc-1-P is first transferred from UDP-GlcNAc to form a Dol-P-P-GlcNAc, followed by a second GlcNAc to give Dol-P-P-GlcNAc-GlcNAc.^{130,131} This is followed by the sequential transfer of five mannose residues that are added from GDP-Man by five different mannosyltransferases to give a Man₅GlcNAc₂-P-P-Dol.¹³¹ The transfer of GlcNAc-1-P to Dol-P is strongly inhibited by the highly useful N-linked glycosylation inhibitor tunicamycin, an antibiotic produced by several streptomycetes.¹³² Tunicamycin completely prevents the synthesis of N-linked oligosaccharides, thereby preventing proteins that normally have N-linked oligosaccharides from having any sugars attached to them.¹³³

At the Man₅GlcNAc₂-PP-Dol stage of synthesis, the lipid carrier undergoes a 'flip-flop,' such that the oligosaccharide portion now becomes oriented in the lumen of the ER,¹³⁴ and four additional mannoses, as well as three glucoses, are added to form the completed oligosaccharide precursor, referred to as Glc₃Man₉GlcNAc₂-P-P-Dol.¹³⁵ These seven sugars are transferred from the lipid carriers, Dol-P-glucose and Dol-P-Man that are synthesized and utilized in the lumen of the ER.^{136,137} The purpose of the mannose is to complete the formation of the high mannose oligosaccharide (see **Figure 1**), whereas the addition of the glucoses serve two other important functions. The first function is to insure that only completely synthesized oligosaccharides will be transferred to proteins, since the oligosaccharide transferase (OST), which catalyzes the transfer of completed oligosaccharides to asparagine residues on newly synthesized proteins, has a much faster rate of transfer with oligosaccharides that contain the three terminal glucose residues.¹³⁸⁻¹⁴⁰ The addition of glucose to the precursor oligosaccharide also has a very important second function in the synthesis of many N-linked glycoproteins. Thus, many of these proteins, but not all, need help in folding into the proper and biologically active conformation, and a single glucose on the N-linked oligosaccharide of these proteins is the primary site to which the helper proteins (chaperones) bind (see below).^{141,142} The glucose also plays a role in quality control since misfolded proteins undergo ER-associated degradation (ERAD), which is largely regulated by their N-linked polymannose oligosaccharides, but the chaperone interaction with the Glc₁Man₉GlcNAc₂-protein sorts out persistently unfolded proteins for proteolysis.^{142,143}

6.08.5.3 Processing of N-Linked Glycoproteins and Effects of Glycosidase Inhibitors

Once the oligosaccharide is transferred to the asparagine residue on the protein and even before the protein has folded into its native conformation, the oligosaccharide begins to undergo processing or trimming reactions. Initially these reactions involve the removal of the glucose moieties by two ER glucosidases, called glucosidase I and glucosidase II. Glucosidase I removes the outermost α 1,2-linked glucose,¹⁴⁴ whereas glucosidase II removes the next two glucoses, both of which are linked in α 1,3-glycosidic bonds.¹⁴⁵ The second glucose is removed by glucosidase II fairly rapidly, but the third glucose, that is, the glucose that is linked to the mannose, is released much more slowly.^{137,143,145} Perhaps the reason for this slow release is to allow the oligosaccharide time to interact with calnexin if that glycoprotein needs a chaperone to help it fold. Several of the so-called processing inhibitors (castanospermine, deoxynojirimycin, australine, etc.) inhibit both glucosidase I and glucosidase II and therefore either completely or partially prevent the removal of the glucoses.^{143,146,147} As a result and as indicated above, these inhibitors prevent some proteins from leaving the ER at the normal rate, since they only fold very slowly into the proper conformation without aid (e.g., the low-density lipoprotein receptor¹⁴⁸ and various viral glycoproteins^{142,149}). These proteins then 'pile up' in the ER and are usually transferred to the cytoplasm and degraded by the ERAD system in combination with ubiquitin lyase and other degrading enzymes in the cytosol.¹⁵⁰

The role of glucose in this system is to help these proteins fold, and it does this by acting as a 'beacon' or recognition site to allow helper proteins to recognize and bind to N-linked glycoproteins that have not yet folded into the right conformation. These helper proteins or chaperones, called calnexin and calreticulum, are located in the lumen or on the membranes of the ER, and their function is to help other proteins fold into their proper and active conformation. Both of these chaperones are lectins that recognize and bind to

oligosaccharides on unfolded proteins that still have a single glucose attached to the high-mannose chain. Once bound, the chaperones utilize the energy of adenosine triphosphate to catalyze and expedite the folding of these ER proteins.¹⁵¹ Castanospermine has been a valuable tool to demonstrate that the glucose residues, on these N-linked glycoproteins, play an important role in a quality control system in the ER that is involved in degrading improperly folded glycoproteins.¹⁴² It is essential for the cell to be able to remove and destroy misfolded or unfolded proteins so that they do not interfere with the proper functioning of the ER. Thus, when castanospermine was given to cells producing various N-linked glycoproteins, it prevented (see **Figure 3**) the excision of glucose molecules from the N-linked oligosaccharide chains, and these proteins underwent accelerated degradation, that is, chaperones were not able to recognize and help these proteins fold.^{147,149} *In vitro* studies on the binding of calnexin and calreticulin to N-linked oligosaccharides demonstrated that this binding does not occur with di- and tri-glucosylated polymannose chains.^{152,153} These workers also showed that even after extensive trimming of α 1,2-mannoses from the α 1,6-branch of the oligosaccharides, calnexin still bound as long as the other branch still retained the single glucose residue.

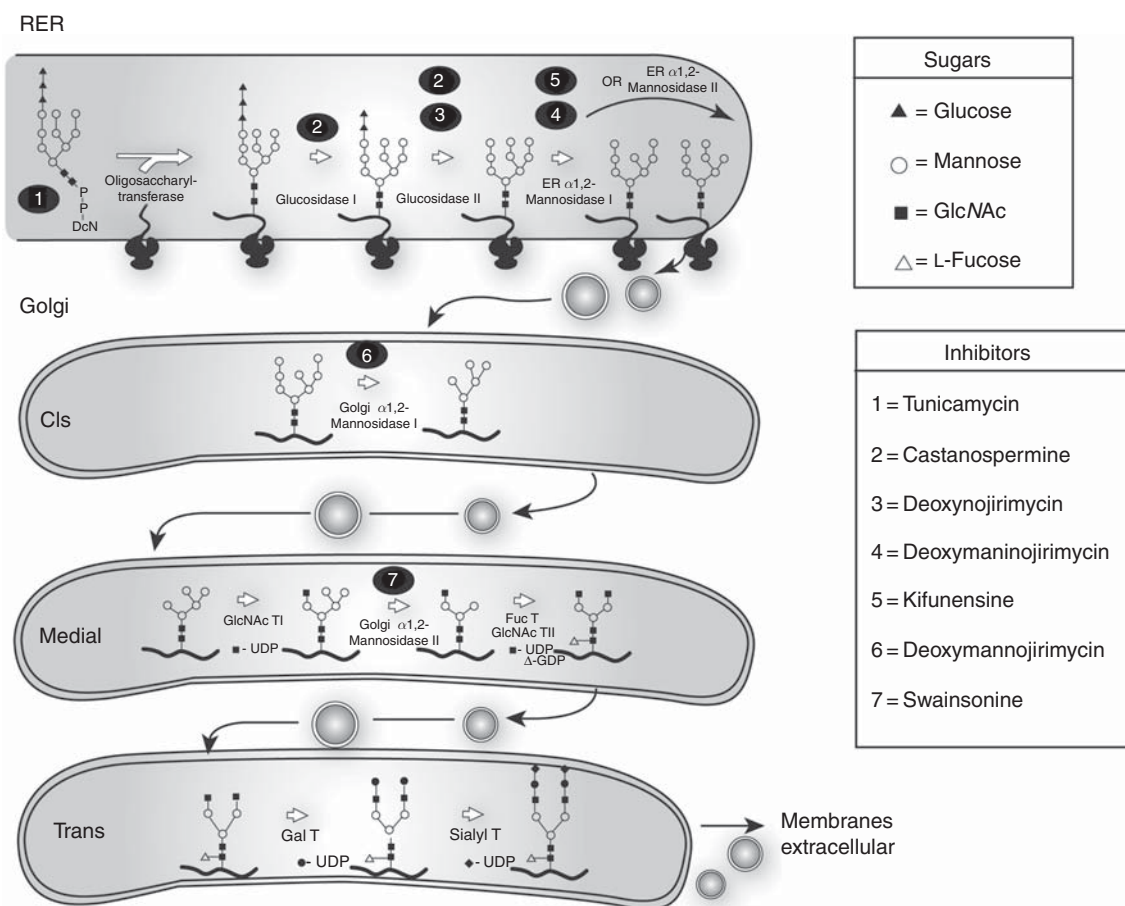


Figure 3 Reactions involved in processing of the precursor oligosaccharide to produce the high-mannose complex and hybrid types of oligosaccharides. Trimming of the protein-bound precursor oligosaccharide begins in the ER with the removal of all three glucoses and one mannose residue (top frame). Glucosidases I and II can be inhibited by the alkaloids castanospermine and deoxynojirimycin. The ER- α -mannosidase I can be inhibited by kifunensine and deoxymannojirimycin. If the protein has folded into its proper conformation, it is transported to the Golgi apparatus where additional mannoses are removed to give a $\text{Man}_5\text{GlcNAc}_2$ -protein. This mannosidase is susceptible to inhibition by deoxymannojirimycin as well as by kifunensine. A GlcNAc transferase then adds a GlcNAc to the mannose on the α 1,3 branch and this signal activates the Golgi α -mannosidase II to remove the two mannoses attached to the α 1,6-linked mannose. Other sugars can then be added in the trans Golgi by various glycosyltransferases to give the complex types of oligosaccharides.

As indicated above, the glucose on the polymannose chain plays an important role in folding and quality control in the ER. However, another part of this ER quality control involves the mannose residues, especially the terminal α 1,2-mannose on the middle branch¹⁴³ (referred to as M_{8B} as shown in **Figure 3**). Mannose trimming is initiated in the ER of cells by the α -mannosidase called ER α 1,2-mannosidase I, which specifically removes a single mannose from the middle branch to produce a $Man_8GlcNAc_2$ -protein.¹⁵⁴ Although this enzyme can remove other mannose residues when incubated with the high mannose oligosaccharide *in vitro*, in cells it appears to remove only a single mannose. However, this reaction is apparently essential to have properly folded proteins transported to the Golgi apparatus. It is also an important step in the quality control of proteins because once this mannose has been removed, calnexin loses its ability to interact with the oligosaccharide even if it still contains a single glucose.^{155,156} Thus, unfolded proteins with the Man_8 -B isomer can no longer be helped by this chaperone, but now become susceptible to the interaction by a mannose lectin (or several in this family) called EDEM (ER degradation-enhancing α -mannosidase-like protein), which binds to the oligosaccharide chains of unfolded proteins and transports them from the ER to the cytosol for degradation by the proteasome.¹⁵⁶

Interaction of this putative ER lectin called EDEM promotes the release of misfolded glycoproteins from calnexin.¹⁵⁶ Thus, when the ER-luminal concentration of EDEM was increased in mammalian cells by overexpression of the protein, there was a corresponding increase in the ER-associated degradation of misfolded glycoproteins and the release of mannose residues from these misfolded proteins.¹⁵⁷ This release of mannoses from these misfolded proteins was prevented by altering one amino acid, that is, E220Q, in the EDEM protein structure. This glutamic acid is indicated as a catalytic residue that is preserved in the GH47 family of α -1,2-mannosidases.¹⁵⁸ Interestingly, this substitution of the catalytic amino acid glutamic acid for a glutamine did not affect the lectin activity of the EDEM 1, but it did inhibit the release of mannoses from these proteins. In addition, the α 1,2-mannosidase inhibitor, kifunensine, also blocked the increase in mannose release in these overexpressed EDEM cells. These data suggest that EDEM(s) may have α -mannosidase activity.¹⁵⁷ The authors suggest that EDEM upregulation leads to accelerated removal of the mannose to which the single glucose is attached and this removal leads to *N*-glycans that lack mannose A, the glucose acceptor. This results in prevention of calnexin binding and accelerated degradation.

There is another ER mannosidase that is also involved in removing mannose from the $Man_9GlcNAc_2$ -protein and may also be involved in quality control.¹⁵⁹ This ER α 1,2-mannosidase II also removes a single mannose residue to produce a $Man_8GlcNAc_2$ -protein, but in this case the mannose is removed from the α 1,6-branch to give isomer M_{8C} .^{154,159} These two enzymes have a number of differences, but one that has been very well used to differentiate the two mannosidases and to determine their action *in vivo* is their sensitivity to kifuninsine.^{140,159} Kifuninsine is an alkaloid produced by the actinomycete *Kitasatosporia kifunense* that was initially shown to inhibit the Golgi mannosidase I but not the ER mannosidase.¹⁶⁰ More recent studies, showing that there are several different mannosidases in the ER, have shown that kifunensine inhibits ER mannosidase I but is inactive against ER mannosidase II.^{153,159} On the contrary, both of these mannosidases are sensitive to inhibition by deoxymannojirimycin, another alkaloidal, synthetic mannosidase inhibitor.¹⁶¹ The primary action of the ER mannosidase I is to remove a single α 1,2-mannose from the middle branch of the polymannose oligosaccharide in order to generate the $Man_8GlcNAc_2$ B isomer, whereas the mannosidase II enzyme also produces a $Man_8GlcNAc_2$ structure, but it removes the terminal α 1,2-mannose from the α 1,6 branch to generate the C isomer. Thus, although both enzymes can theoretically release other mannose residues, their major action, at least *in vivo*, appears to be the removal of a single, and very specific, mannose residue.¹⁵³ The kifunensine-sensitive ER mannosidase I appears to be the enzyme that is involved in quality control.

The mannosidase inhibitors, kifunensine and deoxymannojirimycin, have been elegantly used to help solve the structure of the catalytic domain of the human ER class I α 1,2-mannosidase.¹⁶² The crystal structure of the catalytic domain of this enzyme was studied in the presence and in the absence of these two inhibitors. Both inhibitors bind to the protein at the bottom of the active-site cavity with the required calcium ion coordinating the O-2' and O-3' hydroxyl groups and stabilizing the six-membered ring structure of both inhibitors in a 1C_4 conformation. The authors point out that this is the first direct evidence for the role of Ca^{2+} for this mannosidase. The absence of any change in conformation of the protein upon binding the inhibitors and the comparisons with other mannosidase-product complexes suggested to these workers that this mannosidase has a novel catalytic mechanism.¹⁶²

There is yet another α 1,2-mannosidase that is also in the ER as well as the Golgi apparatus, and this enzyme may also be involved in early processing.¹⁶³ This enzyme is also inhibited by both kifunensine and deoxymannojirimycin, and it can remove three or four mannoses to produce $\text{Man}_5\text{GlcNAc}_2$. It is undoubtedly involved in processing but its role, if any, in quality control and ERAD is not clear. This enzyme may recycle between these two compartments and may simply be involved in mannose processing. Finally there are at least two other α -mannosidases that have been localized to the Golgi, and these enzymes can cleave all four of the α 1,2-mannoses to produce the $\text{Man}_5\text{GlcNAc}_2$ -oligosaccharide.¹⁶⁴

Once the glycoprotein has folded correctly and been transported to the Golgi apparatus, an α 1,2-mannosidase can convert the high mannose oligosaccharide from the $\text{Man}_8\text{GlcNAc}_2$ (A-, B-, or C-isomer) into a $\text{Man}_5\text{GlcNAc}_2$ -protein. A GlcNAc transferase can then attach a single GlcNAc from UDP-GlcNAc to the single mannose that is attached in an α 1,3-linkage to the β -linked mannose.¹⁶⁵ This GlcNAc acts as a signal for another α -mannosidase called Golgi mannosidase II to remove the final two mannoses from the α 1,6 branch. These two mannoses are linked in α 1,3- and α 1,6-linkages, indicating that this enzyme is quite distinct from the α -mannosidases.¹⁶⁵ Interestingly, the first of the glycoprotein-processing inhibitors, a plant alkaloid called swainsonine, was found to inhibit this mannosidase but none of the above mannosidases.^{166,167} Swainsonine was initially identified as the cause of locoism in animals¹⁶⁸ and shown to be an inhibitor of the lysosomal α -mannosidase that is involved in turnover of glycoproteins.⁹²

Other processing reactions of these oligosaccharides occur in the trans-Golgi and include addition of other sugars, such as GlcNAc, Gal, sialic acid, L-fucose, and so on. Of course, if the oligosaccharide processing is blocked with the use of some of these inhibitors, then the protein may still fold into its native conformation and be transported to the Golgi and to its proper location. If so, the question of whether that protein can function normally will depend on whether a specific carbohydrate structure is involved in its function or is necessary for its targeting to a specific site in the cell. Some of the effects of these processing inhibitors on glycoprotein function were discussed in a previous review.¹ Section 6.08.7 also cites some of their effects on glycoprotein function.

6.08.6 Structure and Synthesis of Glycosphingolipids

6.08.6.1 Introduction

Glycosidase inhibitors are proving to be useful tools for research dealing with GSL lysosomal storage diseases, and such inhibitors may prove to be valuable in the therapy of these diseases as well.¹⁶⁹ GSL are ubiquitous components of all the cellular membranes of eukaryotic cells, and some 300 different glycolipid structures have now been identified.¹⁷⁰ Although these GSLs make only a small contribution to the total mass of the plasma membrane, they do make major contributions to various physiological functions of cells, including cell–cell adhesion, cell growth and its regulation, and differentiation, to name a few.^{171,172} All of the GSLs contain a ceramide (Cer) as the basic structure to which is attached a fatty acid and one or more sugars. As shown in **Figure 4**, Cer is a lipid-like structure formed initially by the condensation of palmitoyl-CoA and serine. The resulting product is called ketosphinganine which is then reduced to form sphinganine. This sphinganine is then acylated with a fatty acid to form dihydroceramide, and the dihydroceramide reduced again to give Cer.

Thus, Cer is the common and basic component of all of the GSLs, as well as of sphingomyelin.¹⁷³ **Figure 5** demonstrates how Cer is glycosylated with various sugars to produce a great variety of different glycolipids.¹⁷⁴ For example, the route for formation of gangliosides, that is, the sialic acid containing GSLs, involves the addition of glucose from UDP-glucose to produce glucosyl-ceramide (GluCer), followed by transfer of a Gal from UDP-Gal to the glucose to produce lactosyl-ceramide (LacCer). An *N*-acetylneuraminic acid (NeuNAc) is next added to form the first ganglioside, referred to as $\text{G}_{\text{M}3}$. As indicated in **Figure 5**, $\text{G}_{\text{M}3}$ can be lengthened by addition of an GalNAc to produce $\text{G}_{\text{M}2}$, and then another Gal can be added to give $\text{G}_{\text{M}1}$, and so on. Each of these glycolipids may have a specific physiological function in the cell, and each glycolipid turns over or is degraded in the cell at an individual rate, that is, at its own rate.

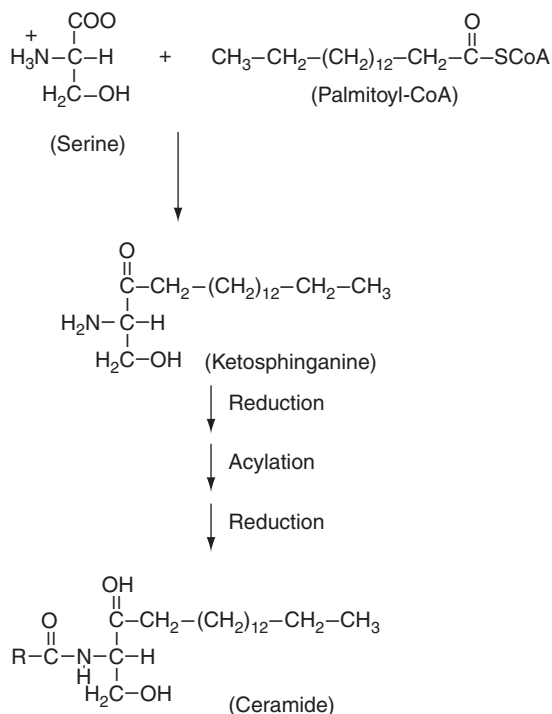


Figure 4 Pathway for the production of Cer, the lipid component of mammalian GSL. In the ER, palmitoyl-CoA is condensed with L-serine to produce ketosphinganine, which is then reduced, and acylated to form Cer.

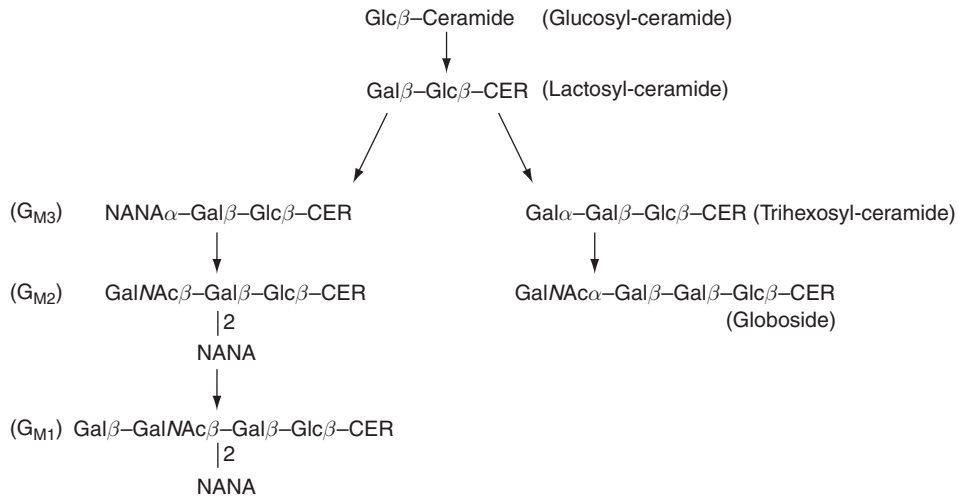


Figure 5 Pathways of biosynthesis of the GSL. Glycosyltransferases in the ER sequentially add sugars to produce the various glycolipids. Gangliosides (i.e., glycosphingolipids containing sialic acid) are produced by the addition of NANA to lactosyl-ceramide to form $\text{G}_{\text{M}1}$, which can then be further glycosylated to $\text{G}_{\text{M}2}$ and then $\text{G}_{\text{M}3}$. Globosides also come from lactosyl-ceramide by addition of another Gal and then a GalNAc.

6.08.6.2 Degradation of Glycosphingolipids in the Lysosomes and Resulting Diseases

Sphingolipids are turned over or degraded in the lysosomes of cells, and this degradation mechanism involves the stepwise removal of the sugars from the nonreducing end of the oligosaccharide chain, releasing one monosaccharide (or other substituent such as sulfate) at each enzymatic step.¹⁷⁵ The degradation pathway of $\text{G}_{\text{M}1}$ ganglioside is shown in **Figure 6**. In some respects, this pathway might appear to be a reversal of the

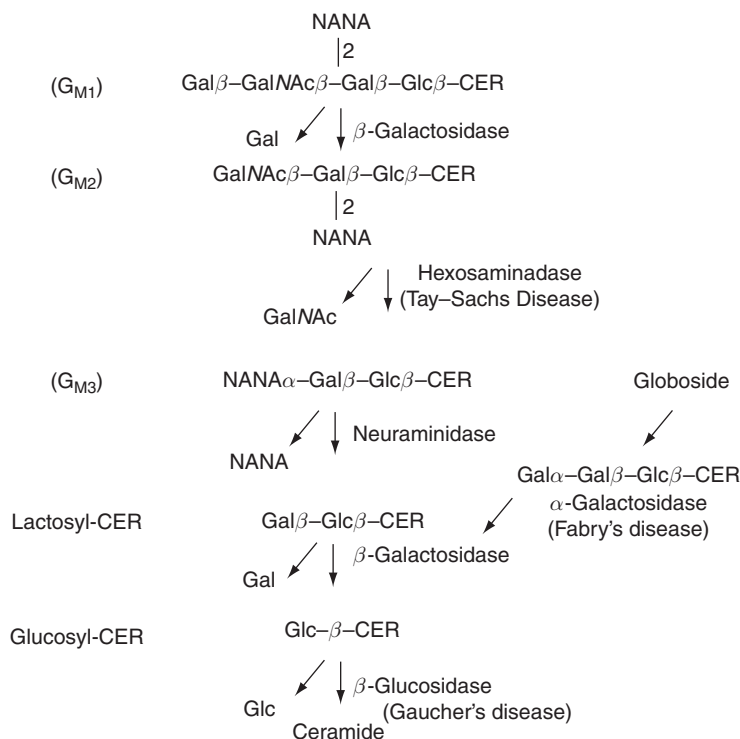


Figure 6 Pathways of degradation of the glycosphingolipids in the lysosomes. Ganglioside G_{M1} is degraded in the lysosomes by sequential removal of sugars by various glycosidases. Thus, β -galactosidase removes the terminal Gal to produce G_{M2} and this lipid is then the substrate for a hexosaminidase which removes Gal/NAc to produce G_{M1} . Absence of this hexosaminidase gives rise to the lipid storage disease called Tay-Sachs disease. G_{M3} is then converted into lactosyl-ceramide by action of the lysosomal neuraminidase that removes sialic acid. Another β -galactosidase removes the Gal to produce glucosyl-ceramide, the substrate that accumulates in Gaucher's disease.

synthetic pathway (Figure 5), except, of course, for the fact that biosynthesis involves a series of glycosyl-transferases that are located in the ER and that transfer activated sugars from their nucleoside diphosphate sugar derivative to another sugar, or another acceptor, to produce an oligosaccharide or a GSL.

On the contrary, degradation occurs in the lysosome of the cell, where specific glycosidases remove one sugar at a time from the nonreducing end of an oligosaccharide (or GSL) and release that sugar as a free monosaccharide. Since degradation occurs at the nonreducing end of an oligosaccharide, if that terminal sugar, for example, NeuNAc, is not released from G_{M1} , then the next sugar, that is, Gal, also cannot be cleaved, the therefore degradation stops at the G_{M1} stage. In these GSL storage diseases, the absence of activity of one of the enzymes involved in release of a specific sugar results in the accumulation of the substrate of the missing enzyme and causes lysosomal lipid storage diseases such as Tay-Sachs, Fabry's, or Gaucher's diseases.¹⁷⁶ Figure 6 presents a brief scheme showing the degradative pathway followed by most of these GSL derivatives as they are turned over in the lysosome. This figure also points out the site of the lesion in the above-mentioned lipid storage diseases.

6.08.6.3 Use of Glycosidase Inhibitors to Treat Lysosomal Lipid Storage Diseases

Thus far, there are relatively few options that are available for treating patients with GSL storage diseases. Enzyme replacement is one option and has been used with good success in patients with Gaucher's disease,¹⁷⁷ but may be much more problematic with some of the other storage diseases such as Fabry or Tay-Sachs. In the case of Gaucher's disease, this is primarily a disease of the macrophages. Of course, all of the cells of the Gaucher patient degrade their GSLs to produce the substrate of the defective enzyme, that is, GlcCer.

However, most of the defective cells in a Gaucher individual still produce enough residual enzyme, that is, glucocerebrosidase (also called glucosylceramidase or β -glucosidase), to degrade the small amounts of glucosyl-ceramide (GlcCer) being produced in their turnover pathways.¹⁷⁸

On the contrary, macrophages accumulate much more GlcCer than the other cells of this patient because macrophages phagocytose senescent erythrocytes, leukocytes, and apoptotic cells, as well as their own GSLs. This adds to their GSL burden and exceeds the capacity of their residual enzyme. Thus, macrophages of Gaucher patients accumulate substantial GlcCer in their lysosomes. The therapy for these patients has successfully involved intravenous injection of a placenta-derived β -glucosidase (CEREDASE), which is engineered to have its N-linked oligosaccharides terminate in mannose residues. Since macrophages have mannose receptors on their plasma membranes, CEREDASE is specifically taken up by these cells in a receptor-mediated fashion, which does not occur in other types of cells.¹⁷⁹ The production of this recombinant and engineered β -glucosidase is difficult and time-consuming, and therefore the therapy for treating Gaucher patients is very expensive. This has highlighted the potential problem of dealing with the other glycolipid storage diseases by enzyme replacement technology. Gaucher's disease is the most common of the GSL diseases and the cost for this treatment is almost prohibitive, so the cost and difficulty in developing similar therapies for Fabry and Tay-Sachs diseases may make such therapy highly unlikely. Added to this is the fact that most of these glycolipid storage diseases involve accumulation of GSL degradation products in cells of the central nervous system. Unfortunately, glycoprotein enzymes do not cross the blood-brain barrier, and therefore they cannot get into cells in the brain, indicating additional problems with enzyme replacement therapy. Therefore, the Gaucher model may only apply in a few cases, adding additional excessive costs to the development of such treatments.

The above problems have led to other strategies for therapy of these diseases. One such strategy is called substrate deprivation.^{180,181} The basis of this strategy is to reduce the biosynthesis of GSLs in these patients, so that the amount of glycolipid being produced is nearly the same as the rate of degradation of these molecules in the individual. In order to make this strategy work, one must identify compounds that block the biosynthesis of the GSLs, preferably at an early step in the biosynthetic pathway. Several inhibitors that act on the glucosyltransferase that catalyzes the transfer of glucose to Cer (i.e., glucosyl-ceramide synthase) have been described.¹⁸² The prototype molecule is PDMP (D,L-threo-1-phenyl-2-decanoylamino-3-morpholino-1-propanol).

A second derivative of PPMP with a C₁₆ fatty acid rather than the C₁₀ of PDMP showed better activity with an inhibitory constant on the isolated GlcCer synthase of 0.7 $\mu\text{mol l}^{-1}$. This inhibition of glucose transfer to Cer showed mixed-type inhibition for Cer, but was noncompetitive for UDP-glucose. Radin suggested the use of PDMP for the treatment of Gaucher's disease.¹⁸³

First, the problem with these compounds is that they are very hydrophobic and need to be in an organic solvent or detergent to be solubilized for administration to cells or organisms. Second, PDMP is rapidly metabolized in cells by the cytochrome P-450 detoxification system and therefore has a very short half-life in cells. Therefore there is a need for other inhibitors that do not have these drawbacks. On the basis of the similarities between these compounds and derivatives of deoxynojirimycin (DNJ), N-alkylated derivatives of DNJ were prepared and tested on the isolated GlcCer synthase.¹⁸⁴ These compounds have previously been shown to be potent inhibitors of the N-linked oligosaccharide processing enzymes, α -glucosidases I and II, with a K_i of 0.22 $\mu\text{mol l}^{-1}$.¹⁸⁵ Both the N-alkyl-DNJ and the N-alkyl-DGJ inhibited the Cer-specific glucosyltransferase with a K_i of about 7.4 $\mu\text{mol l}^{-1}$. The inhibition by N-alkyl-DNJ was critically dependent on the N-alkyl chain length with the minimum chain length being three, but activity was better with longer alkyl groups such as the butyl derivative.¹⁸⁶ *In vitro* studies indicated that both N-butyl-DNJ (NB-DNJ) and N-butyl-DGJ (deoxygalactonojirimycin) were competitive inhibitors for Cer as the substrate, but noncompetitive for UDP-glucose. Molecular modeling studies indicated a structural homology between three or more of the chiral centers in NB-DNJ (and NB-DGJ) with the *trans*-alkenyl chain of Cer. The authors suggest that in addition to mimicking the hydrophobic regions of Cer, N-alkylation of DNJ results in this compound being able to enter into the lipid phase and thus allows it to interact more readily with the membrane-bound glucosyltransferase.¹⁸⁷ Interestingly, the mannose, fucose, and GlcNAc analogues of NB-DNJ did not inhibit this enzyme indicating that the stereochemistry of the ring moiety is essential.¹⁸⁷

These inhibitors were tested in several *in vivo* systems to determine whether they did inhibit GSL synthesis in a living system. In a Gaucher model system, mouse macrophages were induced to accumulate GlcCer, by

incubating them with an inhibitor of the glucocerebrosidase (i.e., the lysosomal β -glucosidase). Interestingly when either NB-DNJ or NB-DGJ was added to the macrophages in the presence of the glucocerebrosidase inhibitor, they prevented the accumulation of GlcCer, suggesting that they blocked the accumulation of GSLs by inhibiting the synthetic route. Furthermore, when the lysosomes of these cells were examined in the electron microscope, there was no evidence of the electron-dense material (accumulated GSLs) seen in the absence of NB-DNJ, also indicating that there was no accumulation of GSL.¹⁸⁸ Mice treated with NB-DGJ or NB-DNJ showed an equivalent depletion of GSLs in their livers. However, NB-DNJ also inhibited glycogen catabolism in the liver, as well as intestinal sucrase and maltase. On the contrary, NB-DGJ did not inhibit glycogen metabolism or the activity of sucrase or maltase, although it inhibited GSL accumulation as well as NB-DNJ. NB-DGJ did, however, inhibit lactase whereas NB-DNJ did not. These workers concluded that NB-DGJ is a more selective inhibitor and should be evaluated for human therapy.¹⁸⁹

Fabry's disease is another lysosomal storage disease in which individuals have a deficiency in the α -galactosidase and therefore cannot remove the terminal α -galactosyl residue from the globoside, trehexosylceramide (Gal α 4Gal β 4GlcCer). Therefore this lipid accumulates in their lysosomes. This lipid accumulates in vascular endothelial cells and causes kidney failure with premature myocardial infarction and strokes. Some individuals with Fabry's disease still have some residual α -galactosidase activity, but studies on the properties of the Fabry enzyme with α -galactosidase from normal individuals indicated that although the kinetic properties of both were similar, the Fabry α -galactosidase was much less stable.^{187,190} This instability has been found to exist in most Fabry patients that fall into the cardiac variant class where clinical manifestations are limited to the heart.¹⁹¹ An α -galactosidase purified from Fabry patients with a Q279E mutation had the same K_m and V_{max} as the normal enzyme. However, the mutant enzyme lost most of its catalytic activity after incubation at 37 °C for 30 min whereas the enzyme from normal individuals was stable to these conditions. In cells from these patients, the mutant enzyme formed an aggregate in the ER and was degraded.^{187,192}

Since the mutant enzymes from Fabry patients have normal catalytic activity but are present at only low levels in the lysosomes and are also much less stable than the normal α -galactosidase, it appeared likely that the genetic defect affected the folding and translocation of the enzyme from the ER to the lysosome.^{193,194} It is well documented that improperly folded proteins are not transported, or are transported very poorly, to the Golgi apparatus, and the majority of these proteins accumulate in the ER and are degraded by the ERAD and ubiquitin systems^{195,196} (see Section 6.08.5.2). On the basis of information on protein folding and targeting and studies presented below on the effects of imino sugars, Asano *et al.* proposed a chemical chaperone theory to explain the effects of DGJ and other imino sugars in enhancing the α -galactosidase activity of lymphoblasts from Fabry patients with the Q279E mutation. They proposed that competitive inhibitors such as DGJ, when used at subinhibitory concentrations, act as specific chemical chaperones that help the unstable protein fold into the proper conformation and/or help the unstable protein to leave the ER and move to the Golgi apparatus.¹⁹⁷ These workers also tested a series of DGJ derivatives for activity with both the isolated α -galactosidase and Fabry lymphoblasts. Their results indicated that DGJ was the most potent inhibitor of the enzyme with an IC_{50} of $0.04 \mu\text{mol l}^{-1}$. Other inhibitors such as α -galacto-homonojirimycin, α -allo-homonojirimycin, and β -1-C-butyl-deoxygalactonojirimycin were also effective inhibitors, but so less than DGJ, with IC_{50} values of 0.21, 4.3, and $16 \mu\text{mol l}^{-1}$, respectively. N-alkylation, deoxygenation at C-2, and epimerization at C-3 of DGJ markedly lowered or abolished the inhibitory activity against the α -galactosidase. When DGJ, α -galacto-homonojirimycin, α -allo-homonojirimycin, and β -1-C-butyl-DGJ were added to the culture media of Fabry lymphocytes at $100 \mu\text{mol l}^{-1}$, they increased the intracellular activity of α -galactosidase by the following amounts: 14-, 5.2-, 2.4-, and 2.3-fold, respectively. There was a direct correlation between inhibition of α -galactosidase and increasing the activity of this enzyme in lymphocytes. The more potent the inhibitor was on the enzyme, the greater the effect on enhancing enzyme activity in the cells.¹⁸⁷ The authors indicate that these results show that the more potent inhibitors act as more effective and specific chemical chaperones for the mutant enzyme and also that the potent competitiveness of α -galactosidase are effective chaperones for Fabry's disease. These data suggest another role for these imino sugars as helper molecules to bind to specific glycosidases (and perhaps other proteins such as lectins) and help these proteins assume a stable and effective conformation for targeting and activity.¹⁹⁸

6.08.7 Therapeutic Activities of Glycosidase Inhibitors

6.08.7.1 Introduction

The ability of polyhydroxy alkaloids to inhibit glycosidases has inevitably resulted in interest in the therapeutic use of these compounds for the treatment of a number of diseases that result from disruption of glycoprotein processing in animal cells, and their potential has been comprehensively reviewed.^{199–202} An obvious application is to induce a particular disease state in animal models and then to study various treatments to alleviate or overcome the disease. Examples would be the genetic lysosomal storage disorders, mannosidosis, Pompe's, Fabry's, and Gaucher's diseases in humans. These are caused by complete absence, reduced activity, or incomplete trafficking or folding of the enzymes α -mannosidase, α -glucosidase, α -galactosidase, and β -glucocerebrosidase (a β -glucosidase), respectively. A first choice for treatment of such diseases has been enzyme replacement therapy, which has proven effective in many cases but is extremely expensive and must be frequently and repeatedly infused throughout the patient's lifetime.

6.08.7.2 Use of Inhibitors in Substrate Reduction Therapy

A less expensive and more convenient alternative in cases where there is some residual enzyme activity or misfolding would be the use of an orally available drug that could restore activity to a level sufficient to alleviate at least the more severe symptoms; this process has been named substrate reduction therapy (SRT). As with most carbohydrate-based drugs, there has been some skepticism as to whether the polyhydroxy alkaloids would have appropriate pharmacokinetics, given their high water solubility, which could lead to rapid excretion. However as discussed in the previous section on GSL metabolism, it has been shown that the preparation of relatively simple alkyl derivatives of moderate chain length can lead to prodrugs with sufficient retention to be effective but which are ultimately hydrolyzed to the active form. One example of a prodrug is *N*-nonyl-1-deoxynojirimycin, which produces a twofold increase in activity of β -glucocerebrosidase in cells derived from Gaucher patients.²⁰³ The alkaloid acts as a pharmacological chaperone, inducing correct folding of the protein and recovery of lost function. Appropriate targets must have a mutation which renders the protein unstable but not inactive so that the chaperone intercepts the mutated protein and stabilizes its fold, circumventing the retrotranslocation machinery of the secretory pathway whereby it would be degraded by the proteasome in the cytoplasm. Inhibitors of the targeted enzyme are able to act as chaperones either due to differences in the destination of the target whereby the inhibitor is ejected or because a relatively high concentration of substrate out-competes the inhibitor. Even in situations where the inhibitor is not ejected, for most lysosomal storage diseases it is preferable for a small amount of enzyme to be inhibited than for it to be quantitatively degraded. Miglustat (*N*-butyl-DNJ; Zavesca[®]) was approved for treatment of mild-to-moderate type 1 Gaucher's disease in 2002.²⁰⁴ The pharmacokinetics of the drug in Gaucher patients are favorable, with time to maximum observed plasma concentration ranging from 2 to 2.5 h and an effective half-life of approximately 6–7 h, demonstrating that converting the natural alkaloid DNJ into a prodrug reduces hydrophilicity.

Another alternative for pharmacological chaperones in SRT is to permanently incorporate the aliphatic side chain into the structure through synthesis so that it cannot be hydrolyzed; such compounds would be drugs *per se* and not merely prodrugs. Examples that have shown promise *in vitro* are α -1-*C*-octyl-deoxynojirimycin²⁰⁵ and α -1-*C*-octyl-1,5-dideoxy-1,5-imino-D-xylitol.²⁰⁶ Both of these compounds approximately double the activity of β -glucocerebrosidase without the side effect of inhibiting α -glucosidase. Similar intracellular enhancement of lysosomal α -galactosidase A activity by 1-deoxygalactonojirimycin (DGJ) and derivatives has been shown *in vitro* in Fabry lymphoblasts. DGJ has also been used to characterize and increase the activity at subinhibitory concentrations of 11 mis-sense mutants of the α -galactosidase A gene from Fabry's disease patients. These compounds and other α -galactosidase inhibitors, such as the calystegines, could therefore be useful for treating patients with Fabry's disease after structural modification to improve pharmacokinetic properties. It is reasonable to consider that similar approaches could be successful, with individual alkaloid glycosidase inhibitors targeted against other lysosomal storage diseases in an analogous manner.

6.08.7.3 Glycosidase Inhibitors in Diabetes Therapy

One of the goals of therapy in diabetes is to bring blood sugar levels back to the normal range and to remove or reduce insulin resistance in noninsulin-dependent diabetes mellitus (NIDDM). The symptoms and complications of diabetes are closely related to the degree of metabolic control and the blood sugar levels.²⁰⁷ One way to control blood sugar levels is by diet, and this is an important part of the therapy in diabetes. A major part of the diet is made up of starch and sucrose, and these carbohydrate components have to be broken down in the intestine to monosaccharides, glucose, and fructose, to be absorbed by the intestinal epithelial cells and enter into the circulatory system. The enzymes involved in degrading the major dietary carbohydrate components are glycosidases that reside on the surfaces of the intestinal cells.²⁰⁸ Thus, the idea of mimicking dietary control by inhibiting these α -glucosidases appeared to be a reasonable way to affect blood sugar levels.²⁰⁹ One inhibitor that proved to be quite effective was the pseudo-tetrasaccharide, acarbose (**Figure 7**). This compound is produced by the microorganism, *Actinoplanes*, and it is a potent inhibitor of the starch and glycogen-degrading brush-border enzymes, glucoamylase, dextrinase, maltase, and sucrase, both *in vivo* and *in vitro*.²¹⁰ When tested

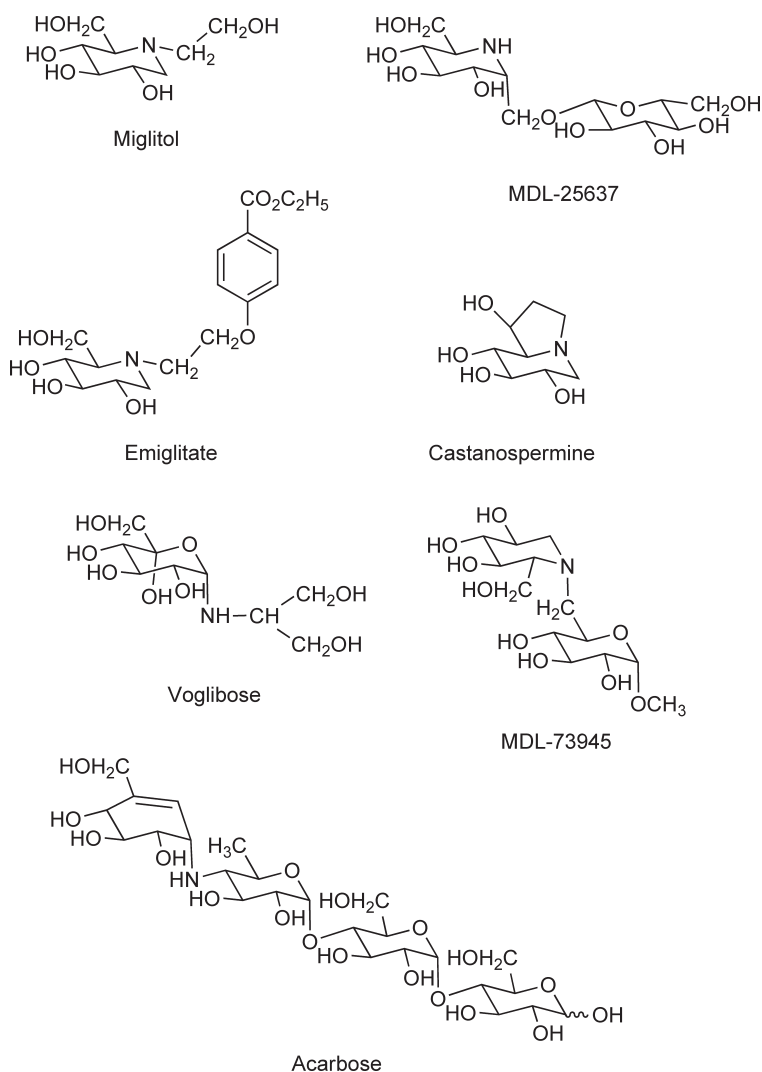


Figure 7 Glycosidase inhibitors that are being tested against diabetes. Acarbose is an inhibitor of amylases and inhibits utilization of starch and glycogen thereby lowering blood glucose levels. Other inhibitors may also have potential use by inhibiting key glycosidases, or other enzymes, that are involved in raising blood glucose levels.

in animal models and in human subjects, acarbose inhibited the carbohydrate digestive enzymes *in vitro* and also caused a significant decrease in the postprandial rise in blood sugar after giving the subjects a mixed carbohydrate feeding.²¹¹ A number of other inhibitors of intestinal brush-border α -glycosidases, which show a broad spectrum of activity, have been described. These include miglitol, emiglitate, voglibose, AI-5662, castanospermine, MDL-25637, and MDL-73945.²¹² Some of these are currently being tested in clinical trials. However, thus far, acarbose is still the most thoroughly tested. The difficulty with some of these glucosidase inhibitors is that inhibition of disaccharide breakdown in the intestine may result in diarrhea and flatulence in the patient because disaccharides such as sucrose, lactose, and maltose are good substrates for intestinal bacteria, and these organisms grow rapidly on these carbon sources and produce large amounts of acid and gas. This was found to be the case when castanospermine was fed to rats to inhibit the lysosomal α -glucosidase as a model system for the glycogen storage disease, Pompe's disease.²¹³ Nevertheless, these inhibitors may be valuable antidiabetics if given at the right dose and times. They may also serve as prototypes to provide synthetic chemists with a guide as to the necessary structure for inhibition and allow them to design more specific and more potent site-directed inhibitors.

6.08.7.4 Potentials for Therapy in Viral Diseases

Other potential applications for polyhydroxy alkaloid glycosidase inhibitors exist but are not yet in clinical practice. The replication of a number of viruses has been shown to be suppressed by α -glucosidase I inhibitors, by preventing the correct folding of their envelope glycoproteins. N-linked glycosylation inhibitors have been tested against most of the major viral diseases affecting humans, including human immunodeficiency virus (HIV), hepatitis B (HBV),²¹⁴ and hepatitis C, the latter using bovine viral diarrhea virus (BVDV) as a model,²¹⁵ in the absence of an effective HCV cell culture system. Castanospermine and a number of 6-O-esterified prodrug derivatives inhibit HIV *in vitro*, with 6-O-butanoylcastanospermine being the most effective having a cellular 'therapeutic index' similar to the reverse transcriptase inhibitors 3'-azido-3'-deoxythymidine (AZT) and 2',3'-dideoxycytidine (ddc).²¹⁶ The same compound has been tested orally in a mouse model against herpes simplex virus type 1 (HSV-1) and found to produce a 100-fold reduction of the virus load and significant delay in lesion development.²¹⁷ 6-O-Butanoylcastanospermine (Celgosivir) has also undergone trials and shown promise as an oral drug for patients with resistant hepatitis C.²¹⁸

Castanospermine has been shown to be effective against rarer but serious virus infections such as dengue fever, with all four serotypes being affected *in vitro* and mortality of mice infected with the virus being prevented at doses as low as 10 mg kg⁻¹ bodyweight/day.^{219,220} On the contrary, it was ineffective against either yellow fever virus or West Nile virus.

In addition to its antiviral activity, castanospermine demonstrates antiparasitic activity, preventing adhesion of *Plasmodium falciparum* to infected erythrocytes and providing protection against cerebral malaria.²²¹ The alteration of cellular recognition processes through modification of glycoprotein structures in the plasma membrane also accounts for its immunosuppressive properties, extending heart and renal allograft survival in rats.²²²

Both swainsonine and castanospermine have shown potential as anticancer agents and the rationale for using glycoprocessing inhibitors as antitumor and antimetastatic agents has been reviewed.²²³ Metastasis of B16-F10 murine melanoma cells²²⁴ and tumor growth induced in nude mice²²⁵ were inhibited by castanospermine but the alkaloid had no effect on the metastatic properties of prostate cancer cells in a rat model.²²⁶ Swainsonine has been more thoroughly investigated and found to enhance natural killer cell activity. When administered to mice in the drinking water at 3 mg ml⁻¹ for 24 h prior to injection of murine melanoma cells into the tail vein it reduced pulmonary colonization of the lungs by 80%,²²⁷ and human melanoma xenografts were reduced by 50% at 10 mg ml⁻¹.¹⁴³ In humans with advanced malignancies, a phase I study of swainsonine administered over 5 days by continuous infusion showed remission of head and neck tumors in one patient and symptomatic improvement in two other patients.²²⁸ In a subsequent phase IB clinical trial, the maximum tolerated dose for oral administration was shown to be approximately 300 mg kg⁻¹ day⁻¹ with 150 mg kg⁻¹ day⁻¹ as the oral chronic intermittent dose.²²⁹ Swainsonine administered to mice in drinking water was retained primarily in lymphoid tissue for at least 3 days with levels in the brain being low, indicating that for a short period of administration at the levels studied, exposure would be insufficient to produce neurological damage.²³⁰ If the same were to hold true

for humans, intravenous administration prior to and following surgery could prevent postoperative metastasis of tumor cells. Swainsonine has also significant immunomodulatory effects, increasing colony-forming unit (CFU) capacity of bone marrow cells in mice, with potential to overcome the bone marrow suppressive effects of cancer chemotherapy and radiotherapy.²³¹ It has been shown to protect mice bearing melanoma-derived tumors from cytotoxicity of the antitumor drug cyclophosphamide, without interfering with its chemotherapeutic activity, and to restore colony-forming cells in cultures of human hematopoietic cells treated with the myelosuppressive AIDS drug AZT, protecting them from toxicity.²³²

Unfortunately, the supply of swainsonine is quite restricted and consequently expensive, limiting its potential as a therapeutic agent. However, castanospermine can be obtained from *C. australe* seeds in yields of 0.3% fresh weight or higher, which makes it a cost-effective and renewable source material. Synthetic modification to produce a prodrug with better pharmacokinetics, by analogy with the use of 6-butanoylcastanospermine for treatment of hepatitis C could provide a sufficient quantity of the drug for treatment of large numbers of cancer patients at reasonable cost.

Abbreviations

| | |
|-----------------------|--|
| APCI | atmospheric pressure chemical ionization |
| ATP | adenosine triphosphate |
| Cer | ceramide |
| CFU | colony forming units |
| CID | collision-induced dissociation |
| DMDP | 2,5-dihydroxymethyl-3,4-dihydropyrrolidine |
| D-AB1 | 1,4-dideoxy-1,4-imino-arabinitol |
| DMJ | 1-deoxymannojirimycin |
| DNJ | 1-deoxynojirimycin |
| Dol-P | dolichyl-phosphate |
| Dol-P-Man | dolichyl-phosphoryl-mannose |
| EDEM | endoplasmic reticulum degradation-enhancing α -mannosidase-like protein |
| ERAD | endoplasmic reticulum-associated degradation |
| ER | endoplasmic reticulum |
| ESI | electrospray ionization |
| Gal | galactose |
| GalNAc | <i>N</i> -acetylgalactosamine |
| GC-MS | gas chromatography-mass spectrometry |
| GDP-mannose | guanosine diphosphate mannose |
| GlcNAc | <i>N</i> -acetylglucosamine |
| GM₁ | ganglioside M ₁ |
| GM₃ | ganglioside M ₃ |
| GluCer | glucosyl-ceramide |
| GSL | glycosphingolipids |
| HMDS | hexamethylsilazane |
| HIV | human immunodeficiency virus |
| HBV | hepatitis B virus |
| LacCer | lactosyl-ceramide |
| LC-MS/MS | liquid chromatography-tandem mass spectrometry |
| Man | mannose |
| MSTFA | <i>N</i> -methyl- <i>N</i> -trimethylchlorosilyl-fluoroacetamide |
| NANA | <i>N</i> -acetylneuraminic acid (sialic acid) |
| NIDDM | non-insulin-dependent diabetes mellitus |
| nOe | nuclear Overhauser effect |
| NMR | nuclear magnetic resonance |
| OST | oligosaccharyl transferase |

| | |
|-------------------|--|
| PDMP | D,L-threo-1-phenyl-2-decanoylamino-3-morpholino-1-propanol |
| SRT | substrate reduction therapy |
| TLC | thin-layer chromatography |
| TMCS | trimethylchlorosilane |
| TMSi | trimethylsilyl |
| UDP-GlcNAc | uridine diphosphate <i>N</i> -acetylglucosamine |

References

1. A. D. Elbein; R. J. Molyneux, *Comprehensive Natural Products Chemistry*; B. M. Pinto, Ed.; Elsevier: Amsterdam, 1999; Vol. 3, Chapter 3.07, p 129.
2. A. E. Stütz, Ed., *Iminosugars as Glycosidase Inhibitors*; Wiley-VCH: Weinheim, 1999.
3. J. P. Michael, *Nat. Prod. Rep.* **2008**, *25*, 139; previous reviews in the series.
4. A. Welter; J. Jadot; G. Dardenne; M. Marlier; J. Casimir, *Phytochemistry* **1976**, *15*, 747.
5. R. J. Molyneux; Y. T. Pan; J. E. Tropea; A. D. Elbein; C. H. Lawyer; D. J. Hughes; G. W. J. Fleet, *J. Nat. Prod.* **1993**, *56*, 1356.
6. S. Muraio; S. Miyata, *Agric. Biol. Chem.* **1980**, *44*, 219.
7. N. Asano; K. Oseki; E. Tomioka; H. Kizu; K. Matsui, *Carbohydr. Res.* **1994**, *259*, 243.
8. A. A. Watson; R. J. Nash; M. R. Wormald; D. J. Harvey; S. Dealler; E. Lees; N. Asano; H. Kizu; A. Kato; R. C. Griffiths; A. J. Cairns; G. W. J. Fleet, *Phytochemistry* **1997**, *46*, 255.
9. A. Kato; I. Adachi; M. Miyauchi; K. Ikeda; T. Komae; H. Kizu; Y. Kameda; A. A. Watson; R. J. Nash; M. R. Wormald; G. W. J. Fleet; N. Asano, *Carbohydr. Res.* **1999**, *316*, 95.
10. A. Kato; N. Kato; S. Miyauchi; Y. Minoshima; I. Adachi; K. Ikeda; N. Asano; A. A. Watson; R. J. Nash, *Phytochemistry* **2008**, *69*, 1261.
11. T. Yamashita; K. Yasuda; H. Kizu; Y. Kameda; A. A. Watson; R. J. Nash; G. W. J. Fleet; N. Asano, *J. Nat. Prod.* **2002**, *65*, 1875.
12. N. Asano; M. Nishida; S. Miyauchi; K. Ikeda; M. Yamamoto; H. Kizu; Y. Kameda; A. A. Watson; R. J. Nash; G. W. J. Fleet, *Phytochemistry* **2000**, *53*, 379.
13. A. Kato; N. Kato; I. Adachi; J. Hollinshead; G. W. J. Fleet; C. Kuriyama; K. Ikeda; N. Asano; R. J. Nash, *J. Nat. Prod.* **2007**, *70*, 993.
14. N. Asano; E. Tomioka; H. Kizu; K. Matsui, *Carbohydr. Res.* **1994**, *253*, 235.
15. N. Asano; T. Yamashita; K. Yasuda; K. Ikeda; H. Kizu; Y. Kameda; A. Kato; R. J. Nash; H. S. Lee; K. S. Ryu, *J. Agric. Food Chem.* **2001**, *49*, 4208.
16. N. Asano; T. Yamauchi; K. Kagamifuchi; N. Shimizu; S. Takahashi; H. Takatsuka; K. Ikeda; H. Kizu; W. Chuakul; A. Kettawan; T. Okamoto, *J. Nat. Prod.* **2005**, *68*, 1238.
17. M. Shibano; S. Kitagawa; G. Kusano, *Chem. Pharm. Bull.* **1997**, *45*, 505.
18. M. Shibano; S. Kitagawa; S. Nakamura; N. Akazawa; G. Kusano, *Chem. Pharm. Bull.* **1997**, *45*, 700.
19. M. Shibano; S. Nakamura; N. Akazawa; G. Kusano, *Chem. Pharm. Bull.* **1998**, *46*, 1048.
20. M. Shibano; S. Nakamura; M. Kubori; K. Minoura; G. Kusano, *Chem. Pharm. Bull.* **1998**, *46*, 1416.
21. M. Shibano; S. Nakamura; N. Motoya; G. Kusano, *Chem. Pharm. Bull.* **1999**, *47*, 472.
22. M. Shibano; D. Tsukamoto; R. Fujimoto; Y. Masui; H. Sugimoto; G. Kusano, *Chem. Pharm. Bull.* **2000**, *48*, 1281.
23. M. Shibano; D. Tsukamoto; G. Kusano, *Chem. Pharm. Bull.* **1999**, *47*, 907.
24. D. Tsukamoto; M. Shibano; R. Okamoto; G. Kusano, *Chem. Pharm. Bull.* **2001**, *49*, 492.
25. D. Tsukamoto; M. Shibano; G. Kusano, *Chem. Pharm. Bull.* **2001**, *49*, 1487.
26. A. A. L. Gunatilaka; S. Surendrakumar; R. H. Thomson, *Phytochemistry* **1984**, *23*, 929–931.
27. M. Shibano; D. Tsukamoto; A. Masuda; Y. Tanaka; G. Kusano, *Chem. Pharm. Bull.* **2001**, *49*, 1362.
28. C.-Y. Yu; M. H. Huang, *Org. Lett.* **2006**, *8*, 3021.
29. S. F. Matkhalikova; V. M. Malikov; S. Yunusov, *Yu. Khim. Prir. Soedin.* **1969**, *5*, 606.
30. S. F. Matkhalikova; V. M. Malikov; M. R. Yagudaev; S. Yunusov, *Yu. Khim. Prir. Soedin.* **1971**, *7*, 210.
31. H. Iida; N. Yamazaki; C. J. Kibayashi, *Org. Chem.* **1987**, *52*, 1956.
32. H. Yoda; T. Nakajima; K. Takabe, *Tetrahedron Lett.* **1996**, *37*, 5531.
33. C.-L. J. Wang; J. C. Calabrese, *J. Org. Chem.* **1991**, *56*, 4341.
34. J. P. Saludes; S. C. Lievens; T. F. Molinski, *J. Nat. Prod.* **2007**, *70*, 436.
35. P. Marfey, *Carlsberg Res. Commun.* **1984**, *49*, 591.
36. R. Sakai; H. Kamiya, *J. Antibiot.* **2006**, *59*, 507.
37. P. R. Scudder; R. A. Dwek; T. W. Rademacher; G. S. Jacob, *Eur. Pat. Appl. EP 413674*, **1991**.
38. W. Zou; M. Sandbhor; M. Bhasin, *J. Org. Chem.* **2007**, *72*, 1226; references cited therein.
39. T. Hartmann; L. Witte, *Alkaloids: Chemical and Biological Perspectives*; S. W. Pelletier, Ed.; Pergamon: Oxford, UK, 1995; Vol. 9, Chapter 4, p 155.
40. R. J. Molyneux, M. Benson, R. Y. Wong, J. E. Tropea, and A. D. Elbein, *J. Nat. Prod.*, **1988**, *51*, 1198.
41. R. J. Nash; L. E. Fellows; J. V. Dring; G. W. J. Fleet; A. E. Derome; T. A. Hamor; A. M. Scofield; D. J. Watkin, *Tetrahedron Lett.* **1988**, *29*, 2487.
42. R. J. Nash; L. E. Fellows; J. V. Dring; G. W. J. Fleet; A. Girdhar; N. G. Ramsden; J. M. Peach; M. P. Hegarty; A. M. Scofield, *Phytochemistry* **1990**, *29*, 111.
43. C. M. Harris; T. M. Harris; R. J. Molyneux; J. E. Tropea; A. D. Elbein, *Tetrahedron Lett.* **1989**, *30*, 5685.
44. M. R. Wormald; R. J. Nash; P. Hrcniar; J. D. White; R. J. Molyneux; G. W. J. Fleet, *Tetrahedron: Asymmetry* **1998**, *9*, 2549.

45. S. E. Denmark; B. Herbert, *J. Am. Chem. Soc.* **1998**, *120*, 7357.
46. A. Kato; E. Kano; I. Adachi; R. J. Molyneux; A. A. Watson; R. J. Nash; G. W. J. Fleet; M. R. Wormald; S. E. Denmark; H. Kizu; K. Ikeda; N. Asano, *Tetrahedron Asymmetry* **2003**, *14*, 325.
47. R. J. Nash; P. I. Thomas; R. D. Waigh; G. W. J. Fleet; M. R. Wormald; P. M. de Q. Lilley; D. J. Watkin, *Tetrahedron Lett.* **1994**, *35*, 7849.
48. J. Van Ameijde; G. Horne; M. R. Wormald; R. A. Dwek; R. J. Nash; P. W. Jones; E. L. Evinson; G. W. J. Fleet, *Tetrahedron Asymmetry* **2006**, *17*, 2702.
49. T. Matsumura; M. Kasai; T. Hayashi; M. Arisawa; Y. Momose; I. Arai; S. Amagaya; Y. Komatsu, *Pharm. Biol.* **2000**, *38*, 302.
50. T. Ritthiwigrom; S. G. Pyne, *Org. Lett.* **2008**, *10*, 2769.
51. A. S. Davis; S. G. Pyne; B. W. Skelton; A. H. White, *J. Org. Chem.* **2004**, *69*, 3139.
52. N. Asano; H. Kuroi; K. Ikeda; H. Kizu; Y. Kameda; A. Kato; I. Adachi; A. A. Watson; R. J. Nash; G. W. J. Fleet, *Tetrahedron Asymmetry* **2000**, *11*, 1.
53. N. Asano; K. Ikeda; M. Kasahara; Y. Arai; H. Kizu, *J. Nat. Prod.* **2004**, *67*, 846.
54. B. L. Rhinehart; K. M. Robinson; C.-H. R. King; P. Liu, *Biochem. Pharmacol.* **1990**, *39*, 1537.
55. B. Dräger, *Nat. Prod. Rep.* **2004**, *21*, 211.
56. S. Watanabe; H. Kato; K. Nagayama; H. Abe, *Biosci. Biotech. Biochem.* **1995**, *59*, 936.
57. G. C. Kite; J. M. Horn; J. T. Romeo; L. E. Fellows; D. C. Lees; A. M. Scofield; N. G. Smith, *Phytochemistry* **1990**, *29*, 103.
58. K.-A. Lee; M. S. Kim, *Can J. Microbiol.* **2000**, *46*, 1077.
59. D. Tepfer; A. Goldmann; N. Pamboukdjian; M. Maille; A. Lepingle; D. Chevalier; J. Dénarié; C. Rosenberg, *J. Bacteriol.* **1988**, *170*, 1153.
60. A. Goldmann; M.-L. Milat; P.-H. Ducrot; J.-Y. Lallemand; M. Maille; A. Lepingle; I. Charpin; D. Tepfer, *Phytochemistry* **1990**, *29*, 2125.
61. T. Schimming; B. Tofern; P. Mann; A. Richter; K. Jenett-Siems; B. Dräger; N. Asano; M. P. Gupta; M. D. Correa; E. Eich, *Phytochemistry* **1998**, *49*, 1989.
62. R. J. Molyneux; R. A. McKenzie; B. M. O'Sullivan; A. D. Elbein, *J. Nat. Prod.* **1995**, *58*, 878.
63. K. K. I. M. de Balogh; A. P. Dimande; J. J. van der Lugt; R. J. Molyneux; T. W. Naudé; W. G. Welms, *J. Vet. Diagn. Invest.* **1999**, *11*, 266.
64. M. Haraguchi; S. L. Gorniak; K. Ikeda; Y. Minami; A. Kato; A. A. Watson; R. J. Nash; R. J. Molyneux; N. Asano, *J. Agric. Food Chem.* **2003**, *51*, 4995.
65. T. Schimming; K. Jenett-Siems; P. Mann; B. Tofern-Reblin; J. Milson; R. W. Johnson; T. Deroin; D. F. Austin; E. Eich, *Phytochemistry* **2005**, *66*, 469.
66. A. Brock; S. Bieri; P. Christen; B. Dräger, *Phytochemistry* **2005**, *66*, 1231.
67. K. Bekkouche; Y. Daali; S. Cherkouki; J.-L. Veuthey; P. Christen, *Phytochemistry* **2001**, *58*, 455.
68. A. Brock; T. Herzfeld; R. Paschke; M. Koch; B. Dräger, *Phytochemistry* **2006**, *67*, 2050.
69. N. Asano; A. Kato; K. Matsui; A. A. Watson; R. J. Nash; R. J. Molyneux; L. Hackett; J. Topping; B. Winchester, *Glycobiology* **1997**, *7*, 1085.
70. R. J. Nash; A. A. Watson; N. Asano. In *Alkaloids: Chemical and Biological Perspectives*; S. W. Pelletier, Ed.; Pergamon: Oxford, UK, 1996; Vol. 11, p 345.
71. S. M. Colegate; P. R. Dorling; C. R. Huxtable, *Aust. J. Chem.* **1979**, *32*, 2257.
72. R. J. Molyneux; L. F. James, *Science* **1982**, *216*, 190.
73. R. J. Molyneux; L. F. James, Swainsonine, the Locoweed Toxin: Analysis and Distribution. In *Handbook of Natural Toxins*; R. F. Keeler, A. T. Tu, Eds.; Marcel Dekker: New York, 1991; Vol. 6, p 191.
74. R. C. Barbosa; F. Riet-Correa; E. F. Lima; R. M. T. Medeiros; K. M. R. Guedes; D. R. Gardner; R. J. Molyneux; L. E. H. de Melo, *Braz. J. Vet. Res.* **2007**, *27*, 409.
75. R. C. Barbosa; F. Riet-Correa; R. M. T. Medeiros; E. F. Lima; S. S. Barros; E. J. Gimeno; R. J. Molyneux; D. R. Gardner, *Toxicon* **2006**, *47*, 371.
76. A. F. M. Dantas; F. Riet-Correa; D. R. Gardner; R. M. T. Medeiros; S. S. Barros; B. L. Anjos; R. B. Lucena, *Toxicon* **2007**, *49*, 111.
77. E. M. Colodel; D. R. Gardner; P. Zlotowski; D. Dreimeier, *Vet. Hum. Toxicol.* **2002**, *44*, 177.
78. R. J. Nash, Naturally Occurring Glycosidase Inhibitors. In *Bioactive Natural Products*, 2nd ed.; S. M. Colegate, R. J. Molyneux, Eds.; CRC Press: Boca Raton, FL, 2008; p 407.
79. C. M. Harris; B. C. Campbell; R. J. Molyneux; T. M. Harris, *Tetrahedron Lett.* **1988**, *29*, 4815.
80. D. R. Gardner; R. J. Molyneux; M. H. Ralphs, *J. Agric. Food Chem.* **2001**, *49*, 4573.
81. K. Braun; J. Romero; C. Liddell; R. Creamer, *Mycol. Res.* **2003**, *107*, 980.
82. J. McLain-Romero; R. Creamer; H. Zepeda; J. R. Strickland; G. Bell, *J. Anim. Sci.* **2004**, *82*, 2167.
83. M. Ralphs; R. Creamer; D. Baucom; D. Gardner; S. Welsh; J. Graham; C. Hart; D. Cook; B. Stegelmeier, *J. Chem. Ecol.* **2008**, *34*, 32.
84. A. Goldmann; B. Message; D. A. Tepfer; R. J. Molyneux; O. Duclos; F.-D. Boyer; A. D. Elbein, *J. Nat. Prod.* **1996**, *59*, 1137.
85. D. W. Griffiths; T. Shepherd; D. Stewart, *J. Agric. Food Chem.* **2008**, *56*, 5197.
86. R. J. Molyneux, Polyhydroxy Indolizidines and Related Alkaloids. In *Methods in Plant Biochemistry: Volume 8 – Alkaloids and Sulphur Compounds*; P. G. Waterman, Ed.; Academic Press: London, 1993; p 511.
87. R. J. Molyneux; D. R. Gardner; L. F. James; S. M. Colegate, *J. Chromatogr. A* **2002**, *967*, 57.
88. R. J. Molyneux, *Phytochem. Anal.* **1993**, *4*, 193.
89. R. J. Molyneux; L. F. James; K. E. Panter; M. H. Ralphs, *Phytochem. Anal.* **1991**, *2*, 125.
90. M. J. Egan; G. C. Kite; E. A. Porter; M. S. J. Simmonds; S. Howells, *Analyst* **2000**, *125*, 1409.
91. M. J. Egan; E. A. Porter; G. C. Kite; M. S. J. Simmonds; S. Howells, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 195.
92. P. R. Dorling; C. R. Huxtable; S. M. Colegate, *Biochem. J.* **1980**, *191*, 649.
93. R. J. Molyneux; J. E. Tropea; A. D. Elbein, *J. Nat. Prod.* **1990**, *53*, 609.
94. R. J. Molyneux; J. N. Roitman; G. Dunnheim; T. Szumilo; A. D. Elbein, *Arch. Biochem. Biophys.* **1986**, *251*, 450.
95. R. J. Molyneux; M. Benson; J. E. Tropea; Y. T. Pan; G. P. Kaushal; A. D. Elbein, *Biochemistry* **1991**, *30*, 9981.
96. N. Asano; A. Kato; M. Miyauchi; H. Kizu; T. Tomimori; K. Matsui; R. J. Nash; R. J. Molyneux, *Eur. J. Biochem.* **1997**, *248*, 296.
97. A. Kato; N. Asano; H. Kizu; K. Matsui; S. Suzuki; M. Arisawa, *Phytochemistry* **1997**, *45*, 425.
98. T. Kajimoto; K. K.-C. Liu; R. L. Pederson; Z. Zhong; Y. Ichikawa; J. A. Porco, Jr.; C.-H. Wong, *J. Am. Chem. Soc.* **1991**, *113*, 6187.
99. Y. Minami; C. Kuriyama; K. Ikeda; A. Kato; K. Takebayashi; I. Adachi; G. W. J. Fleet; A. Kettawan; T. Okamoto; N. Asano, *Bioorg. Med. Chem.* **2008**, *16*, 2734.

100. A. Kato; N. Kato; E. Kano; I. Adachi; K. Ikeda; L. Yu; T. Okamoto; Y. Banba; H. Ouchi; H. Takahata; N. Asano, *J. Med. Chem.* **2005**, *48*, 2036.
101. N. Asano; K. Ikeda; L. Yu; A. Kato; K. Takebayashi; I. Adachi; I. Kato; H. Ouchi; H. Takahata; G. W. J. Fleet, *Tetrahedron Asymmetry* **2005**, *16*, 223.
102. J. S. S. Rountree; T. D. Butters; M. R. Wormald; R. A. Dwek; N. Asano; K. Ikeda; E. L. Evinson; R. J. Nash; G. W. J. Fleet, *Tetrahedron Lett.* **2007**, *48*, 4287.
103. A. Goldmann; B. Message; D. A. Tepfer; R. J. Molyneux; O. Duclos; F.-D. Boyer; A. D. Elbein, *J. Nat. Prod.* **1996**, *59*, 1137.
104. R. J. Molyneux; S. T. Lee; D. G. Gardner; K. E. Panter; L. F. James, *Phytochemistry* **2007**, *68*, 2973.
105. I. Pastuszak; R. J. Molyneux; L. F. James; A. D. Elbein, *Biochemistry* **1990**, *29*, 1886.
106. R. J. Molyneux; E. Gomez-Sosa, *Bol. Soc. Argent. Bot.* **1991**, *27*, 59.
107. R. J. Molyneux; L. F. James; M. H. Ralphs; J. A. Pfister; K. E. Panter; R. J. Nash. In *Poisonous Plants of the World: Agricultural, Phytochemical and Ecological Aspects*; S. M. Colegate, P. R. Dorling, Eds.; CAB International: Wallingford, UK, 1994; p 107.
108. L. F. James; A. D. Elbein; R. J. Molyneux; C. D. Warren, Eds., *Swainsonine and Related Glycosidase Inhibitors*; Iowa State University Press: Ames, IA, 1989.
109. B. L. Stegelmeier; L. F. James; K. E. Panter; M. H. Ralphs; D. R. Gardner; R. J. Molyneux; J. A. Pfister, *J. Nat. Toxins* **1999**, *8*, 35.
110. L. F. James; R. J. Molyneux; A. F. Alexander, Congestive Right Heart Failure in Cattle: High Mountain Disease and Factors Influencing Incidence. In *Handbook of Natural Toxins*; R. F. Keeler, A. T. Tu, Eds.; Dekker: New York, 1991; Vol. 6, p 635.
111. J. A. Pfister; T. Davidson; K. E. Panter; C. D. Cheney; R. J. Molyneux, *Small Ruminant Res.* **2005**, *65*, 70.
112. L. F. James; W. J. Hartley, *Am. J. Vet. Res.* **1977**, *38*, 1263.
113. M. Haraguchi; S. L. Gorniak; K. Ikeda; Y. Minami; A. Kato; A. A. Watson; R. J. Nash; R. J. Molyneux; N. Asano, *J. Agric. Food Chem.* **2003**, *51*, 4995.
114. R. M. T. Medeiros; R. C. Barbosa; F. Riet-Correa; E. F. Lima; I. M. Tabosa; S. S. de Barros; D. R. Gardner; R. J. Molyneux, *Toxicon* **2003**, *41*, 933.
115. B. L. Stegelmeier; R. J. Molyneux; N. Asano; A. A. Watson; R. J. Nash, *Toxicol. Pathol.* **2008**, *36*, 651.
116. R. Saul; J. J. Ghidoni; R. J. Molyneux; A. D. Elbein, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 93.
117. R. H. Thursby-Pelham, *Vet. Rec.* **1967**, *80*, 709.
118. R. K. Upreti; M. Kumar; V. Shankar, *Proteomics* **2003**, *3*, 363–379.
119. A. Varki, *Glycobiology* **1992**, *3*, 97–130.
120. K. L. Carraway; S. R. Hull, *Glycobiology* **1991**, *1*, 131–136.
121. R. Kornfeld; S. Kornfeld, *Annu. Rev. Biochem.* **1985**, *54*, 631–664.
122. A. D. Elbein; R. J. Molyneux. In *Iminosugars as Glycosidase Inhibitors*; A. E. Stutz, Ed.; Wiley-VCH: Weinheim, 1999; pp 216–240.
123. R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683–720.
124. D. F. Wyss; G. Wagner, *Curr. Opin. Biotechnol.* **1996**, *7*, 409–416.
125. E. Weerapana; B. Imperiala, *Glycobiology* **2006**.
126. P. W. Robbins; S. C. Hubbard; S. J. Turco; D. F. Wirth, *Cell* **1997**, *12*, 893–900.
127. M. N. Fukuda, *Glycobiology* **1991**, *1*, 9–16.
128. M. S. Kang; A. D. Elbein, *J. Virol.* **1983**, *46*, 60–69.
129. C. J. Waechter; W. J. Lennarz, *Annu. Rev. Biochem.* **1976**, *45*, 95–112.
130. N. Gao; M. A. Lehrman, *J. Biol. Chem.* **2002**, *277*, 39425–39435.
131. C. G. Frank; M. Aebi, *Glycobiology* **2005**, *15*, 1156–1163.
132. G. Tamura; I. Teiichiro, In *Tunicamycin*; G. Tamura, Ed.; Japan Scientific Societies Press: Tokyo, 1981.
133. A. Heifetz; R. W. Keenan; A. D. Elbein, *Biochemistry* **1979**, *18*, 2186–2192.
134. M. D. Snider; O. C. Rogers, *Cell* **1984**, *36*, 753–761.
135. J. F. Cipollo; R. B. Trimble; J. H. Chi; Q. Yan; N. Dean, *J. Biol. Chem.* **2001**, *276*, 21828–21840.
136. A. Helenius; M. Aebi, *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049.
137. S. J. Turco; B. Stetson; P. W. Robbins, *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4411–4414.
138. D. Karaoglu; D. J. Kelleher; R. Gilmore, *Biochemistry* **2001**, *40*, 12193–12206.
139. G. Yan; W. J. Lennarz, *J. Biol. Chem.* **2002**, *277*, 47692–47700.
140. C. A. Jakob; P. Burda; S. te Heeson; M. Aebi; J. Roth, *Glycobiology* **1998**, *8*, 155–164.
141. A. Helenius; E. S. Trombetta; D. N. Herbert; J. F. Simons, *Trends Cell Biol.* **1997**, *7*, 193.
142. R. G. Spiro, *Cell. Mol. Life Sci.* **2004**, *61*, 1025.
143. B. E. Kalz-Fuller; E. Bieberich; E. Bause, *Eur. J. Biochem.* **1995**, *231*, 344.
144. S. E. Trombetta; J. F. Simons; A. Helenius, *J. Biol. Chem.* **1966**, *271*, 27509.
145. Y. T. Pan; H. Hori; R. Saul; B. A. Sanford; R. J. Molyneux; A. D. Elbein, *Biochemistry* **1983**, *22*, 3975.
146. S. E. H. Moore; R. G. Spiro, *J. Biol. Chem.* **1993**, *268*, 3809.
147. E. Edwards; E. A. Sprague; J. L. Kelley; J. J. Kerbacher; C. J. Schwartz; A. D. Elbein, *Biochemistry* **1989**, *28*, 7679.
148. K. P. Kearse; D. B. Williams; A. Singer, *EMBO J.* **1994**, *13*, 3678.
149. E. D. Werner; J. L. Brodsky; A. A. McCracken, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13797.
150. S. H. Keller; J. Lindstrom; P. Taylor, *J. Biol. Chem.* **1998**, *273*, 17064.
151. A. Helenius; E. S. Trombetta; D. N. Herbert; J. F. Simons, *Trends Cell Biol.* **1997**, *7*, 193.
152. R. G. Spiro; Q. Zhu; V. D. Bhoyroo; H. D. Soling, *J. Biol. Chem.* **1996**, *271*, 11588.
153. S. Weng; R. G. Spiro, *J. Biol. Chem.* **1993**, *268*, 25656.
154. A. Herscovics, *Biochim. Biophys. Acta* **1999**, *1426*, 275.
155. R. Mancini; M. Aebi; A. Helenius, *J. Biol. Chem.* **2003**, *278*, 46895.
156. Y. Oda; N. Hosokawa; I. Wada; K. Nagata, *Science* **2003**, *299*, 1394.
157. S. Olivari; T. Cali; R. H. Salo; P. Paganetti; L. W. Ruddock; M. Molinari, *Biochem. Biophys. Res. Commun* **2006**, *349*, 1278.
158. W. Tempel; K. Karaveg; Z. T. Liu; J. Rose; B. C. Wang; K. Moreman, *J. Biol. Chem.* **2004**, *279*, 29774.
159. S. Weng; R. G. Spiro, *Arch. Biochem. Biophys.* **1996**, *325*, 113.
160. A. D. Elbein; J. E. Tropea; M. Mitchell; G. P. Kaushal, *J. Biol. Chem.* **1990**, *265*, 15599.

161. U. Fuhrmann; E. Bause; G. Legler; H. Ploegh, *Nature* **1984**, 307, 775–1758.
162. F. Vallee; K. Karaveg; A. Herscovics; K. W. Moreman; P. L. Howell, *J. Biol. Chem.* **2000**, 275, 41287.
163. E. Bause; W. Breuer; J. Schweden; R. Roeser; R. Geyer, *Eur. J. Biochem.* **1992**, 208, 451.
164. K. W. Moreman; R. B. Trimble; A. Herscovics, *Glycobiology* **1994**, 4, 113.
165. N. Harpaz; H. Schachter, *J. Biol. Chem.* **1980**, 255, 4894.
166. A. D. Elbein; R. Solf; P. R. Dorling; K. Vosbeck, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, 78, 7393.
167. D. P. R. Tulsiani; T. M. Harris; O. Touster, *J. Biol. Chem.* **1982**, 257, 7936.
168. P. R. Dorling; C. R. Huxtable; S. M. Colegate, *Neuropathol. Appl. Neurobiol.* **1978**, 4, 285.
169. T. D. Butters; R. A. Dwek; F. M. Platt, *Chem. Rev.* **2000**, 100, 4683.
170. K. Sandhoff; G. vonEchten, *Adv. Lipid Res.* **1993**, 26, 119.
171. S. Hakomori, *Annu. Rev. Biochem.* **1980**, 50, 733.
172. A. Varki, In *Essentials of Glycobiology*; A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, J. Marth, Eds.; Cold Spring Harbor Laboratory Press: 1999, p 115.
173. F. M. Platt; T. D. Butters, *Biochem. Pharmacol.* **1998**, 56, 421.
174. T. Kolter; K. Sandhoff, *Angew. Chem. Int. Ed.* **1999**, 38, 1532.
175. D. Hockstra; J. W. Kok, *Biochim. Biophys. Acta Rev. Biomembr.* **1992**, 1113, 277.
176. E. Neufeld, *Annu. Rev. Biochem.* **1991**, 60, 257.
177. E. Beutler, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 5384.
178. E. Beutler; G. Grabowski, In *The Metabolic and Molecular Basis of Inherited Diseases*; C. R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, Eds.; McGraw Hill: NY, 1995, p 2641.
179. N. W. Barton; F. S. Furbish; G. J. Murray; M. Garfield; R. O. Brady, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, 87, 1913.
180. J.-Q. Fan; S. Ishii; N. Asano; Y. Suzuki, *Nat. Med.* **1999**, 5, 112.
181. F. M. Platt; G. Reinkensmeier; R. A. Dwek; T. D. Butters, *J. Biol. Chem.* **1997**, 272, 19365.
182. J.-I. Inokuchi; N. S. Radin, *J. Lipid Res.* **1987**, 28, 565.
183. N. S. Radin, *Glycoconj. J.* **1996**, 13, 153.
184. F. M. Platt; G. R. Neises; R. A. Dwek; T. D. Butters, *J. Biol. Chem.* **1994**, 269, 8362.
185. E. Truscheit; I. Hillebrand; B. Junge; L. Miller; W. Puls; D. D. Schmidt, *Prog. Clin. Biochem. Med.* **1988**, 7, 17.
186. F. M. Platt; G. R. Neises; G. B. Karlsson; R. A. Dwek; T. D. Butters, *J. Biol. Chem.* **1994**, 269, 27108.
187. N. Asano; S. Ishii; H. Kizu; K. Ikeda; K. Yasuda; A. Kato; O. R. Martin; J.-Q. Fan, *Eur. J. Biochem.* **2000**, 267, 4179.
188. F. M. Platt; G. R. Neises; G. Reinkensmeier; M. J. Townsend; V. H. Perry; R. L. Proia; B. Winchester; R. A. Dwek; T. D. Butters, *Science* **1997**, 276, 428.
189. U. Anderson; T. D. Butters; R. A. Dwek; F. M. Platt, *Biochem. Pharmacol.* **2000**, 59, 821.
190. F. M. Platt; T. D. Butters, *Trends Glycosci. Glycotechnol.* **1995**, 7, 495.
191. U. Anderson; D. Smith; M. Jeyakumar; T. D. Butters; M. C. Borja; R. A. Dwek; F. M. Platt, *Neurobiol. Dis.* **2004**, 16, 506.
191. G. Remeo; G. M. Urso; A. Piszczane; E. Blum; A. de Falco; A. Ruffilli, *Biochem. Genet.* **1975**, 13, 615.
192. S. Ishii; R. Kase; H. Sakuraba; Y. Suzuki, *Biochem. Biophys. Res. Commun.* **1993**, 197, 1585.
193. S. Ishii; R. Kase; T. Okumiyama; H. Sakuraba; Y. Suzuki, *Biochem. Biophys. Res. Commun.* **1996**, 220, 812.
194. A. Helenius, *Mol. Biol. Cell* **1994**, 5, 253.
195. V. E. Bychkova; O. B. Ptitsyn, *FEBS Lett.* **1995**, 359, 6.
196. J.-Q. Fan; S. Ishii; N. Asano; Y. Suzuki, *Nat. Med.* **1999**, 5, 112.
197. T. D. Butters; R. A. Dwek; F. M. Platt, *Glycobiology* **2005**, 15, 43R.
198. S. Nakao; T. Takenaka; M. Maeda; C. Kodama; A. Tanaka; M. Tahara; A. Yoshida; M. Kuriyama; H. Hayashibe; H. Sakuraba; H. Tanaka, *N. Engl. J. Med.* **1995**, 333, 288.
199. N. Asano; R. J. Nash; R. J. Molyneux; G. W. J. Fleet, *Tetrahedron Asymmetry* **2000**, 11, 1645.
200. N. Asano; A. Kato; A. A. Watson, *Mini Rev. Med. Chem.* **2001**, 1, 145.
201. A. A. Watson; G. W. J. Fleet; N. Asano; R. J. Molyneux; R. J. Nash, *Phytochemistry* **2001**, 56, 265.
202. R. A. Dwek; T. D. Butters; F. M. Platt; N. Zitzmann, *Nat. Rev. Drug Discov.* **2002**, 1, 65.
203. A. R. Sawkar; W.-C. Cheng; E. Beutler; C.-H. Wong; W. E. Balch; J. W. Kelly, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 15428.
204. T. M. Cox; J. M. Aerts; G. Andria; M. Beck; N. Belmatoug; B. Bembi; R. Chertkoff; S. Vom Dahl; D. Elstein; A. Erikson; M. Giralt; R. Heitner; C. Hollak; M. Hrebicek; S. Lewis; A. Mehta; G. M. Pastores; A. Rolfs; M. C. Miranda; A. Zimran, *J. Inherit. Metab. Dis.* **2003**, 26, 513.
205. L. Yu; K. Ikeda; A. Kato; I. Adachi; G. Godin; P. Compain; O. Martin; N. Asano, *Bioorg. Med. Chem.* **2006**, 14, 7736.
206. P. Compain; O. R. Martin; C. Boucheron; G. Godin; L. Yu; K. Ikeda; N. Asano, *Chem. Bio. Chem.* **2006**, 7, 1356.
207. A. D. Newcomer; D. B. McGill, *Gastroenterology* **1966**, 51, 481.
208. W. Puls; W. U. Keup; H. P. Krause; G. Thomas; F. Hoffmeister, *Naturwissenschaften* **1977**, 64, 536.
209. D. D. Schmidt; W. Frommer; B. Junge, *Naturwissenschaften* **1977**, 64, 535.
210. W. Puls; H. Bischoff; H. Schutt, In *Delaying Absorption as a Therapeutic Principle in Metabolic Diseases*; W. Creutzfeldt, D. R. Folsch, Eds.; Stuttgart: Thieme Verlag, 1983, p 70.
211. S. Lee; S. Bustamante; O. Koldovsky, *Metabolism* **1983**, 32, 793.
212. H. Bischoff, *Eur. J. Clin. Invest.* **1994**, 24, 3.
213. R. Saul; J. I. Ghidoni; R. I. Molyneux; A. D. Elbein, *Proc. Natl. Acad. Sci. USA* **1985**, 82, 93.
214. A. Mehta; N. Zitzmann; P. M. Rudd; T. M. Block; R. A. Dwek, *FEBS Lett.* **1998**, 430, 17.
215. K. Wliitby; D. Taylor; D. Patel; P. Ahmed; A. S. Tyms, *Antivir Chem Chemother.* **2004**, 15, 141.
216. P. S. Sunkara; M. S. Kang; T. L. Bowlin; P. S. Liu; A. S. Tyms; A. Sjoerdsma, *Ann. NY. Acad. Sci.* **2005**, 616, 90.
217. C. G. Bridges; S. P. Ahmed; M. S. Kang; R. J. Nash; E. A. Porter; A. S. Tyros, *Glycobiology* **1995**, 5, 249.
218. L. A. Sorbera; J. Castaner; L. Garcia-Capdevila, *Drugs Future* **2005**, 30, 545.
219. K. Whitby; T. C. Pierson; B. Geiss; K. Lane; M. Engle; Y. Zhou; R. W. Doms; M. S. Diamond, *J. Virol.* **2005**, 79, 8698.
220. W. Schul; W. Liu; H.-Y. Xu; M. Flamand; S. G. Vasudevan, *J. Infect. Dis.* **2007**, 195, 665.

221. P. S. Wright; D. E. Cross-Doersen; K. K. Schroeder; T. L. Bowlin; P. P. McCarm; A. J. Bitonti, *Biochem. Pharmacol.* **1991**, *41*, 1855.
222. P. M. Grochowicz; A. D. Hibberd; Y. C. Smart; K. M. Bowen; D. A. Clark; W. B. Cowden; D. O. Willenborg, *Transpl. Immunol.* **1996**, *4*, 275.
223. P. E. Goss; M. A. Baker; J. P. Carver; J. W. Dennis, *Clin. Cancer Res.* **1995**, *1*, 935.
224. M. J. Humphries; K. Matsumoto; S. L. White; R. L. Molyneux; K. Olden, *Cancer Res.* **1986**, *46*, 5215.
225. G. K. Ostrander; N. K. Scribner; L. R. Rohrschneider, *Cancer Res.* **1988**, *48*, 1091.
226. C. S. Yee; E. D. Schwab; J. E. Lehr; M. Quigley; K. J. Pienta, *Anticancer Res.* **1997**, *17*, 3659.
227. M. J. Humphries; K. Matsumoto; S. L. White; R. J. Molyneux; K. Olden, *Clin. Exp Metastasis*, **1990**, *8*, 89.
228. J. W. Dennis; K. Koch; S. Yousefi; I. VanderElst, *Cancer Res.* **1990**, *50*, 1867.
229. P. E. Goss; J. Baptiste; B. Fernandes; M. Baker; J. W. Dennis, *Cancer Res.* **1994**, *54*, 1450.
230. P. E. Goss; C. L. Reid; D. Bailey; I. W. Dennis, *Clin. Cancer Res.* **1997**, *3*, 1077.
231. D. Bowen; W. M. Southerland; C. D. Bowen; D. E. Hughes, *Anticancer Res.* **1997**, *17*, 4345.
232. O. A. Oredipe; P. M. Furbert-Harris; I. Laniyan; W. R. Green; S. L. White; K. Olden; D. Parish-Gause; T. Vaughn; W. M. Griffin; R. Sridhar, *Int. Immunopharmacol.* **2003**, *3*, 445.

Biographical Sketches



Alan Elbein received a B.A. degree in Biology from Clark University, Worcester, MA, an M.S. degree in Microbiology from the University of Arizona, Tucson, and a Ph.D. in Microbial Biochemistry from Purdue University, W. Lafayette, IN. He then did postdoctoral research in Biochemistry at the University of Michigan, Ann Arbor, and the University of California, Berkeley, before joining the faculty of the Department of Biology at Rice University in Houston, in 1964, as an assistant professor and then associate professor. In 1970, he moved to the University of Texas Health Science Center in San Antonio as professor of Biochemistry, and in 1990 he became chairman of the Department of Biochemistry and Molecular Biology at the University of Arkansas for Medical Sciences in Little Rock. His research interests are on the biosynthesis and function of complex carbohydrates, especially N-linked glycoproteins, and include the identification and use of inhibitors of glycosylation and glycosidases such as plant alkaloids. He has also an active research project on the biosynthesis and metabolism of the disaccharide, trehalose, in mycobacteria as a target site for chemotherapy against tuberculosis. He has served as a member of the Metabolic Biology review board of the National Science Foundation and of the Physiological Chemistry Study Section and the Pathobiochemistry Study Section of the National Institutes of Health. He has served on the Editorial Boards of the following journals: *Journal of Bacteriology*, *Plant Physiology*, *Planta*, *Archives of Biochemistry and Biophysics*, *Journal of Biological Chemistry*, *European Journal of Biochemistry*, *FEBS Journal*, and *Glycobiology*. His laboratory has published over 200 peer-reviewed research papers and over 50 book chapters and/or reviews.



Russell Molyneux received his BSc (Chemistry) and his Ph.D. (Organic Chemistry) degrees from the University of Nottingham. Following postdoctoral and faculty appointments at the University of Oregon and Oregon State University, respectively, he joined the Western Regional Research Center (Agricultural Research Service/USDA) in Albany, California. In 1993, he was awarded a Sir Frederick McMaster Fellowship (CSIRO, Australia), was the 1998 recipient of the Award for Advancement of Agricultural and Food Chemistry (American Chemical Society), and is the 2006 recipient of the Spencer Award for Excellence in Food and Agricultural Chemistry (American Chemical Society). He is a Fellow of the Royal Society of Chemistry (FRSC) and the American Association for the Advancement of Science (AAAS). He was appointed an Associate Editor of the *Journal of Agricultural and Food Chemistry* in 1999. Dr. Molyneux's research interests have focused on the isolation and identification of natural products belonging to many structural classes, with particular emphasis on natural toxins, especially alkaloid constituents of native plants responsible for poisoning of livestock in the western United States. His research on the locoweed, swainsonine, led to an expansion of interest in structurally related polyhydroxy indolizidine, pyrrolizidine, and tropane alkaloids. He has published over 250 scientific papers, patents, and book chapters.

6.09 Molecular Probes for Protein Glycosylation

Howard C. Hang, The Rockefeller University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.09.1 | Introduction | 261 |
| 6.09.2 | Protein Glycosylation in Eukaryotes | 263 |
| 6.09.2.1 | N-Linked Glycosylation | 263 |
| 6.09.2.2 | O-Linked Glycosylation | 264 |
| 6.09.2.3 | Cytoplasmic O-GlcNAc | 265 |
| 6.09.2.4 | C-Mannosylation | 266 |
| 6.09.2.5 | Proteoglycans | 267 |
| 6.09.3 | Protein Glycosylation in Prokaryotes | 268 |
| 6.09.4 | Chemical Approaches toward Understanding Protein Glycosylation | 268 |
| 6.09.5 | Production of Homogeneous Glycoproteins | 269 |
| 6.09.5.1 | Methods for Glycoconjugate Synthesis | 270 |
| 6.09.5.2 | Semisynthesis of Glycoproteins | 273 |
| 6.09.5.3 | Expression of Homogeneous Glycoproteins | 273 |
| 6.09.6 | Chemical Probes for Perturbing Protein Glycosylation | 277 |
| 6.09.6.1 | Synthetic Modulators of Glycan-Protein Interactions | 277 |
| 6.09.6.2 | Perturbation of Glycan-Processing Enzymes with Small Molecules | 280 |
| 6.09.7 | Global Analysis of Protein Glycosylation | 282 |
| 6.09.7.1 | Specific Enrichment of Glycoproteins for Proteomic Analysis | 283 |
| 6.09.7.2 | Chemical Reporters of Protein Glycosylation | 283 |
| 6.09.7.3 | Mechanism-Based Probes for Profiling Glycan-Processing Enzymes | 288 |
| 6.09.7.4 | FRET-Sensors of Protein Glycosylation | 288 |
| 6.09.7.5 | Affinity Probes for Glycan-Binding Proteins | 289 |
| 6.09.8 | Microarray Platforms for Glycomics | 290 |
| 6.09.8.1 | Glycan Microarrays | 290 |
| 6.09.8.2 | Protein Microarrays: Lectins/Antibodies | 291 |
| 6.09.9 | Conclusions and Future Perspectives | 291 |
| References | | 291 |

6.09.1 Introduction

The modification of proteins with glycans regulates a wide range of functions in biology (**Figure 1**).¹ The study of glycoproteins was launched at the turn of the twentieth century by seminal studies from Karl Landsteiner and coworkers on human blood group antigens.² These pioneering studies revealed that specific ‘antigens’ on the surface of red blood cells governed the serological differences between individuals with various blood groups. The molecular identity of the antigens was therefore of tremendous interest to many physicians and scientists, which inspired the biochemical characterization of these antigens. Subsequent work from several laboratories revealed that the glycans attached to proteins and lipids were the molecular determinants of the blood group specificities.³ The discovery and medical significance of glycosylation on proteins and lipids brought a new perspective on carbohydrates beyond a source of nutrients and raised many questions regarding the role of glycans in biology. These questions are as follows. What are the structures and functions of glycans? How are glycoconjugates (glycoproteins, glycolipids, and proteoglycans) assembled? What types of proteins and lipids are glycosylated? How do glycans influence the function of proteins and lipids? What factors regulate and interact with glycoconjugates?

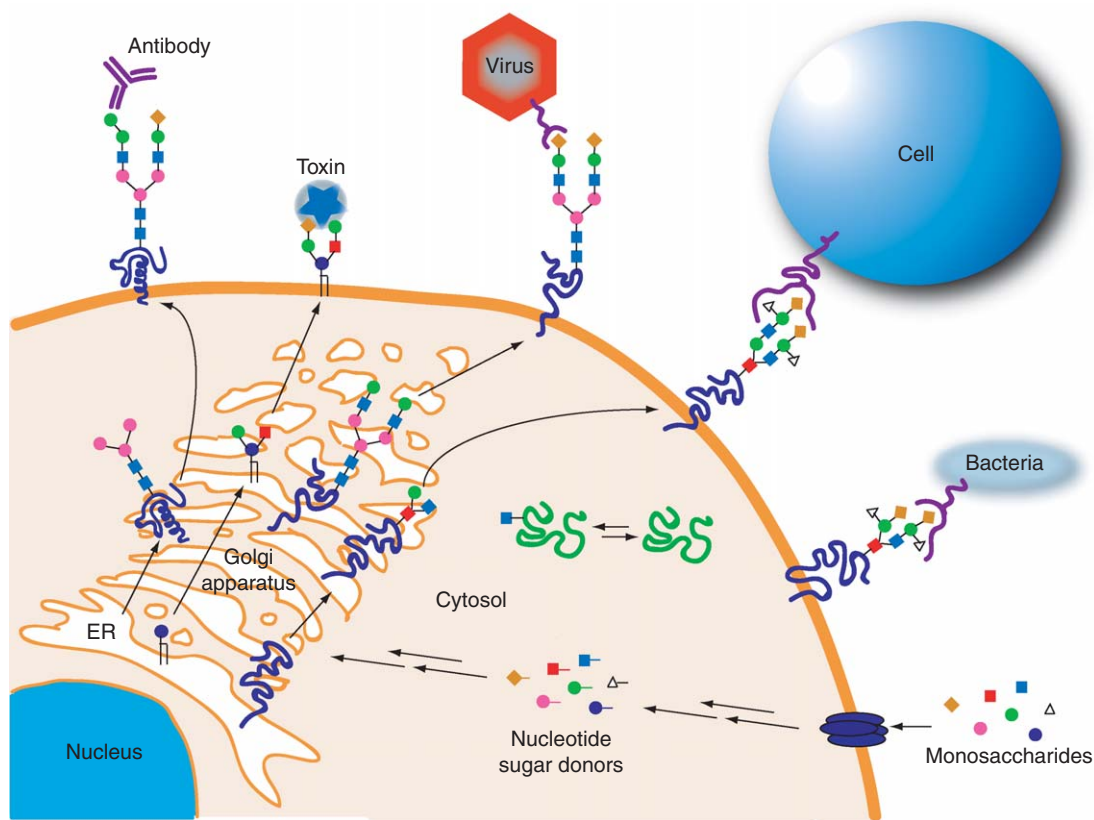


Figure 1 Biosynthesis of glycoconjugates in mammalian cells and their cellular interactions with antibodies, toxins, viruses, other cells, and bacteria.

Elucidating the structure and function of glycoconjugates in physiology and disease has been challenging. The biosynthesis of glycoconjugates is orchestrated by an assembly line of enzymes, which modify proteins or lipids traversing the secretory pathway (**Figure 1**). These glycan-processing enzymes are composed of glycosyltransferases and glycosidases that add and remove monosaccharides, respectively, onto proteins, lipids, and glycans (**Figure 2**). Glycoproteins and lipids can be further modified by other enzyme families such as sulfotransferases/sulfatases, acetyltransferases, and kinases. Glycoconjugates from natural sources are therefore heterogeneous, rendering the large-scale isolation of homogenous materials very difficult for detailed biochemical and structural studies. The production of glycoconjugates by recombinant methods is not

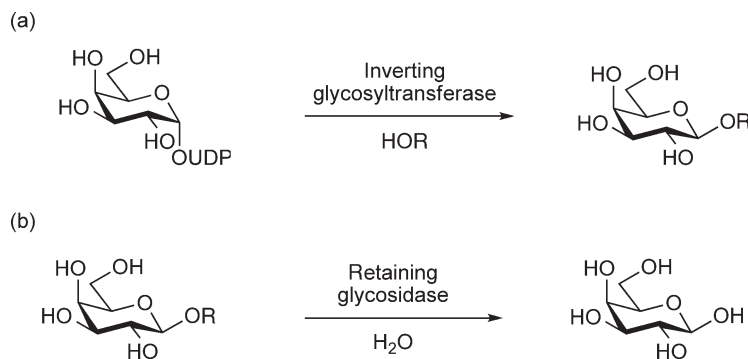


Figure 2 Examples of glycan-processing enzymes. (a) Inverting glycosyltransferase; (b) retaining glycosidase. Inverting and retaining description refers to the stereochemistry of the glycosidic linkage following enzymatic conversion.

straightforward. The glycosylation machinery differs dramatically between cell types and organisms, which precludes the overexpression of mammalian glycoconjugates in commonly used strains of laboratory yeast or bacteria. Furthermore, the template-independent nature of glycoconjugate assembly that mediates several enzyme families with overlapping substrate specificities makes the genetic manipulation of glycosylation pathways not trivial. Glycoconjugates have also been difficult to obtain by chemical synthesis. The size and complexity of glycoproteins and glycolipids has demanded the development of innovative synthetic and enzymatic methodologies for their construction.

Despite the challenges for the functional analysis of glycoconjugates, remarkable progress has been made that has revealed unique roles for glycans in many areas of biology¹ such as protein folding/trafficking,⁴ cell-cell adhesion,⁵ and host–pathogen interactions⁶ (Figure 1). The significance of glycosylation on physiology is underscored by a variety of human diseases associated with mutations in genes that regulate glycan function.^{7,8} These discoveries have highlighted the tremendous potential of glycans to control important biological processes, the extent of which we are only beginning to appreciate. The sequencing of genomes from diverse organisms combined with technical innovations in genetics, chemistry, and analytical methods have provided an unprecedented opportunity to understand biology in the twenty-first century. In this chapter, I will focus on the various types of protein glycosylation found in nature, the challenges for understanding their functions and development of molecular probes that should provide a more precise and global understanding of protein glycosylation.

6.09.2 Protein Glycosylation in Eukaryotes

Protein glycosylation in higher eukaryotes can be divided into three major classes: N-linked, O-linked, and C-linked, characterized by their glycosidic linkages to amino acid side chains.⁹ Bioinformatic analyses suggest that more than 50% of mammalian proteins are glycosylated,¹⁰ nevertheless the functions of these glycoproteins at the molecular and cellular level are only starting to be unraveled. The glycosylation of proteins generally improves their stability and facilitates their trafficking *in vivo*, however, the structure of the appended glycans can confer specific function to proteins and dramatically alter their activity. The glycan structures on proteins typically reflect the activities of glycan-processing enzymes within the cell at a given time, but can also be protein specific on the basis of their trafficking properties and interactions with other cellular factors. Below, I highlight some of the functions that have been revealed for protein glycosylation in eukaryotes and bacteria.

6.09.2.1 N-Linked Glycosylation

N-Linked glycosylation is defined by the attachment of β -*N*-acetylglucosamine (GlcNAc) via an amide linkage to the side chain of Asn in the primary amino acid sequon, AsnXaaSer/Thr (Figure 3).¹¹ The biosynthesis of N-linked glycoproteins is relatively well characterized.¹¹ The core structure of N-linked glycans is assembled on the cytoplasmic leaflet of the endoplasmic reticulum (ER) as a dolicol pyrophosphate substrate, flipped across the ER membrane, modified by a series of glycan-processing enzymes and transferred en bloc by the multisubunit oligosaccharyltransferase complex onto nascent polypeptides that are inserted into the ER. In the ER, N-linked glycosylation serves to ensure proper folding of proteins through the calnexin/calreticulin cycle.⁴ Terminally misfolded N-linked glycoproteins are sent for destruction through the ER-associated degradation pathway.¹² In contrast, properly folded N-linked glycoproteins continue through the secretory pathway where their N-linked glycans are enzymatically elaborated into complex N-linked glycans in the Golgi apparatus and sent to the plasma membrane or secreted. Alternatively, glycoproteins such as lysosomal hydrolases are modified with mannose-6-phosphate on their N-linked glycans and sorted into endocytic compartments.¹³

Outside of the cells, N-linked glycans can serve as ligands for glycan-binding proteins (lectins and antibodies) in cell-cell communication, regulate the aggregation of glycoproteins in the plasma membrane, and modulate the serum half-life of antibodies, cytokines, and hormones.¹ The specific properties of N-linked glycoproteins often govern specific glycan structures. For example, the generation of mice deficient in *N*-acetylglucosaminyltransferase (GlcNAcT) V that is responsible for complex N-linked glycan biosynthesis exhibited enhanced T-cell receptor clustering and signaling.¹⁴ Alternatively, deletion of the GlcNAcT IVa gene resulted in enhanced glucose transporter 2 (Glut-2) internalization into endocytic compartments due to

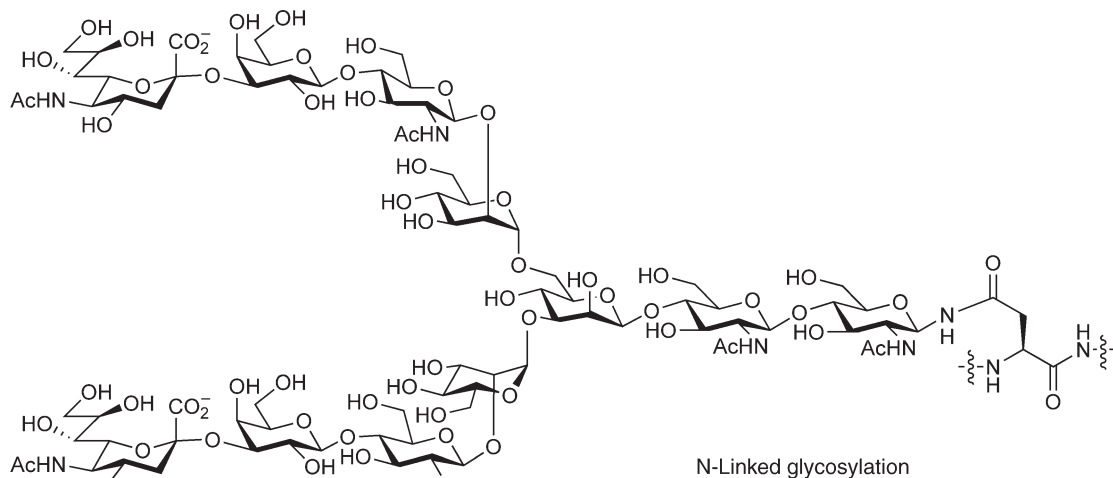


Figure 3 Example of N-linked glycosylation.

truncation of Glut-2 N-linked glycans.¹⁵ Decrease in Glut-2 cell surface expression on pancreatic beta cells impaired glucose-stimulated insulin secretion that led to metabolic dysfunction characteristic of type 2 diabetes.¹⁵ Interestingly, truncated N-linked glycans on the surface of viral particles stimulate specific antibody responses that neutralize human immunodeficiency virus (HIV).^{16–18} Therefore, N-linked glycosylation plays key roles in cellular homeostasis and when altered can give rise to disease such as autoimmunity or protection from pathogens.

6.09.2.2 O-Linked Glycosylation

In contrast to N-linked glycosylation, there are several different types of O-linked glycosylation in higher eukaryotes. Mucin-type glycosylation is the most abundant form of O-linked glycosylation, characterized by α -N-acetylgalactosamine (GalNAc) attached to the hydroxyl group of Ser/Thr side chains (**Figure 4**).^{19–21} A family of polypeptide N-acetyl- α -galactosaminyltransferases (ppGalNAcTs) initiates the biosynthesis of mucin-type O-linked glycans through the enzymatic transfer of GalNAc from UDP-GalNAc onto protein substrates.²² There are ~ 24 ppGalNAcT isoforms in both humans and mice, 15 in the fruit fly (*Drosophila melanogaster*), nine in the worm (*Caenorhabditis elegans*), and five in the unicellular parasite (*Toxoplasma gondii*). Genetic, cellular, and biochemical analyses suggest that the ppGalNAcTs are selectively expressed in different tissues and stages of development, function in the Golgi apparatus, and have differential as well as overlapping protein substrate specificities.²² For example, genetic ablation of murine ppGalNAcT-1 yields defects in the vascular response and humoral immunity,²³ whereas loss of ppGalNAcT-13 results in aberrant O-linked

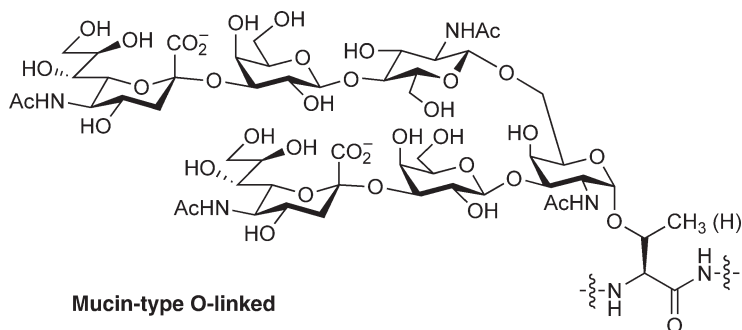


Figure 4 Example of mucin-type O-linked glycosylation.

glycosylation on syndecan-3 in neurons.²⁴ In the fruit fly, deletion of a ppGalNAcT isoform (pgant35A) is required for proper epithelial tube formation.²⁵ For further details on the mucin-type O-linked glycan biosynthesis, see Chapter 6.11.

After the initial attachment of GalNAc onto protein substrates this core monosaccharide is elaborated into a diverse array of complex O-linked glycans with discrete functions.^{1,19,20} In general, large mucin-type O-linked glycoproteins are categorized as 'mucins' that serve as lubricants and protective barriers in epithelial tissues.²⁶ On the other hand, O-linked glycans on secreted proteins (hormones and cytokines) often modulate their serum half-lives. O-Linked glycans on membrane proteins influence protein-protein interactions and have been shown to regulate their rates of internalization. Mucin-type O-linked glycans can also serve as specific ligands for receptors. For example, O-linked glycans on sialomucins are critical for leukocyte homing to sites of inflammation.²⁷ In cancerous tissues aberrant expression of mucin-type O-linked glycans are often observed.¹⁹ Although the precise mechanisms of defective O-linked glycosylation in cancer are not well understood, these altered O-linked glycans represent promising leads for cancer vaccine antigens.²⁸

The glycosylation of proteins with α -O-mannose is the only form of O-linked glycosylation in yeast and has been found predominantly in the brain of higher eukaryotes (**Figure 5(a)**).^{29,30} The several protein O-mannosyltransferases (POMT) are responsible for initiating O-mannosylation and are followed by downstream glycosyltransferases to afford complex glycans (**Figure 5(a)**). O-Mannosylation is essential in yeast as deletion of the POMTs interferes with many cellular processes.²⁹ In mammals, O-mannosylation appears to be a key component of extracellular matrix integrity.³⁰ Analysis of genetic defects associated with congenital muscular dystrophies revealed that O-mannosylation of α -dystroglycan is essential for interactions with lamina in the extracellular matrix.^{31,32} Whether O-mannosylation occurs on other substrates is still unknown.

Several other types of O-linked glycosylation have also been discovered in higher eukaryotes. The analysis of epidermal growth factor (EGF) repeats in a variety of proteins revealed an α -O-fucose modification of Ser/Thr residues within the consensus sequon CysXaa₍₃₋₅₎Ser/ThrCys (**Figure 5(b)**).³³ The requirement of O-fucosylation for Notch signaling during development implicated a regulatory role for O-fucosylation and the downstream glycosyltransferases such as Fringe (β 1-3-N-acetylglucosaminyltransferase) in these critical processes.^{34,35} Further analysis of EGF domains in proteins also revealed β -O-Glc modifications (**Figure 5(c)**),³³ confirming earlier reports described by Hase *et al.*^{36,37} on bovine blood coagulation factors VII and IX. These results have led to the proposed consensus sequence of CysXaaSerXaaProCys for O-glucosylation³³ and the identification of protein O-glucosyltransferase activity in Chinese hamster ovary cells.³⁸ Recently, the identification of Fringe as a β 1-3-GlcNAc transferase that modifies O-fucose residues on Notch has highlighted a new role for glycans in signal transduction and embryonic development.³⁴ The β -O-galactosylation of hydroxylysine (GlyXaaLys) (**Figure 5(d)**) and hydroxyproline (GlyXaaPro) was first discovered on collagen³⁹⁻⁴¹ and has been implicated as an antigenic epitope responsible for rheumatoid arthritis.⁴²⁻⁴⁴ β -O-Galactosylation was also identified on the core-specific lectin and mannan-binding proteins and was proposed to modulate the aggregation of these lectins.^{45,46}

6.09.2.3 Cytoplasmic O-GlcNAc

The identification of β -O-GlcNAc attached to Ser/Thr by Hart and coworkers revealed that nuclear/cytoplasmic and mitochondrial proteins were glycosylated (**Figure 6(a)**).⁴⁷ This unexpected discovery has led to the identification of many O-GlcNAc-modified proteins and the corresponding O-GlcNAc transferase (OGT) and O-GlcNAcases that regulate this modification.⁴⁸ Loss of O-GlcNAc transferase activities is lethal in plants as well as animals and impaired development in the worm. Interestingly, O-GlcNAcase deletion in the worm afforded similar phenotypes to O-GlcNAc transferase ablation, which suggests that both the addition and removal of O-GlcNAc are essential cellular processes.^{49,50} Furthermore, the discovery of β -O-GlcNAc at sites that can also be phosphorylated, suggests that a reciprocal relationship exists between these two dynamic modifications.⁵¹ Although the modification of nuclear/cytosolic proteins with β -O-GlcNAc has been known for two decades, its roles in metabolism, signal transduction, transcription, and proteolysis are only beginning to be unraveled.⁴⁸

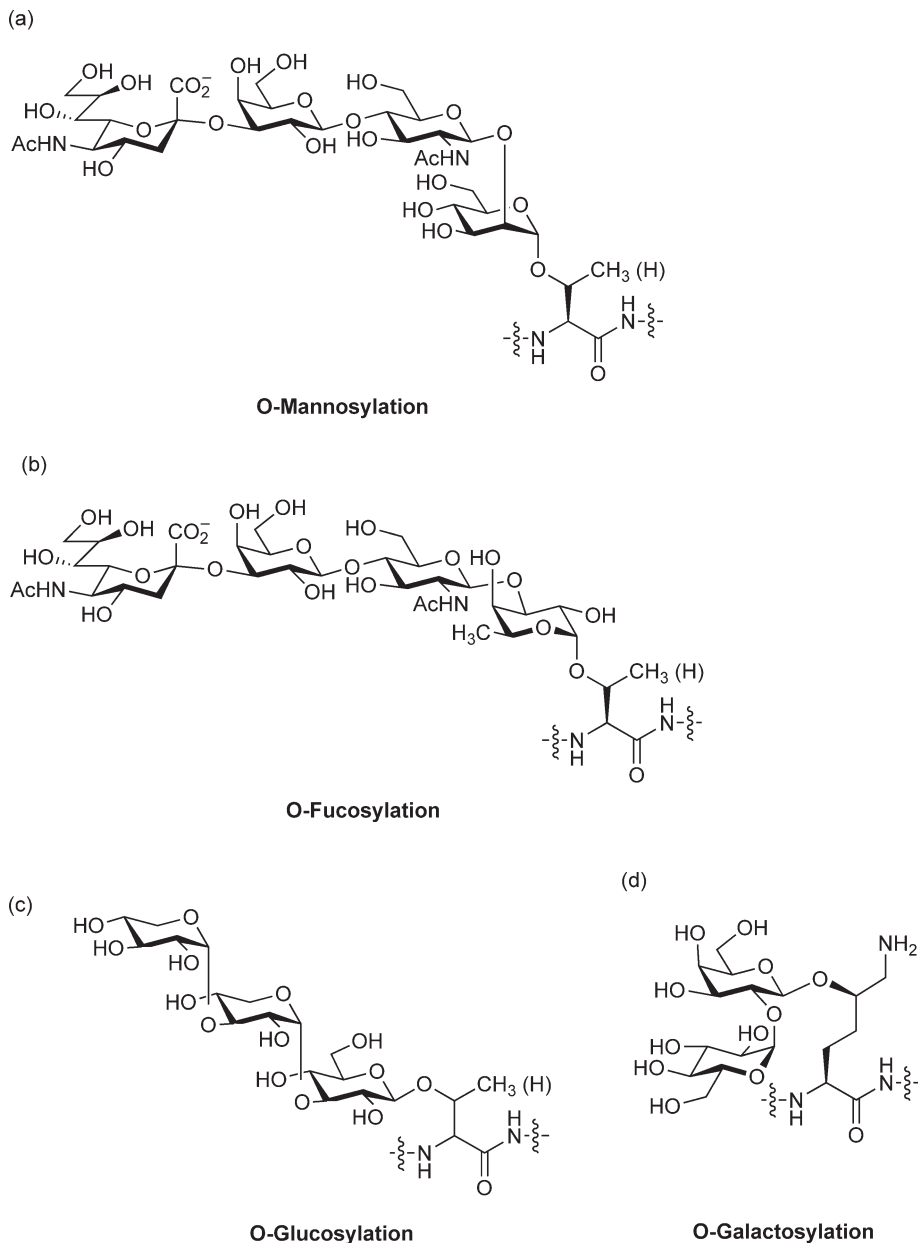


Figure 5 Other types of O-linked glycosylation in eukaryotes. (a) O-Mannosylation; (b) O-fucosylation; (c) O-glucosylation; (d) O-galactosylation.

6.09.2.4 C-Mannosylation

C-Linked glycosylation was first identified on ribonuclease 2 and is characterized by α -mannose (Man) attached to the 2-position of the indole side chain of tryptophan residues (**Figure 6(b)**).^{52,53} This modification has subsequently been identified on several proteins such as MUC5AC and MUC5B,⁵⁴ thrombospondin,⁵⁵ and even the Ebola virus soluble glycoprotein.⁵⁶ Analysis of protein databases for the putative C-linked mannosylation consensus sequence, TrpXaaXaaTrp, suggests that a variety of proteins may bear this modification.^{57–59} The main role of C-linked mannosylation on protein remains to be determined.

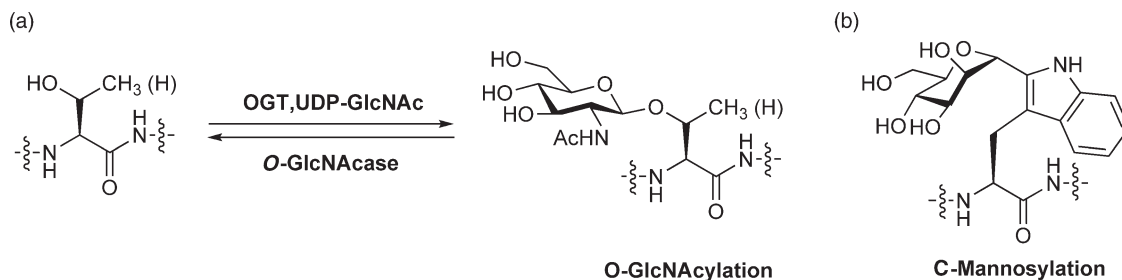


Figure 6 (a) Dynamic O-GlcNAcylation; (b) C-mannosylation.

6.09.2.5 Proteoglycans

Proteoglycans comprise another major class of glycoconjugates in higher eukaryotes that are composed of glycosylaminoglycans (GAGs) attached to protein through sites of O-linked glycosylation.⁶⁰ The core tetrasaccharide (GlcA β 1-3Gal β 1-3Gal β 1-4Xyl) of proteoglycans is attached to Ser residues and then elongated with GAG chains that are composed of repeating disaccharide units of GlcA β 1-4GlcNAc for heparan sulfate (HS) or GlcA β 1-4GalNAc for chondroitin sulfate (CS) (**Figure 7**). These structures are further diversified by epimerization (glucuronic acid (GlcA) to iduronic acid (IdA)), N-deacetylation (HS only) and sulfation to afford complex patterns of GAGs. For further details on the GAG biosynthesis, see Chapter 6.14.

The complex and diverse structures of GAGs attached to proteins provide a platform for sequestering signaling molecules such as chemokines and growth factors in the extracellular milieu of multicellular organisms. For example, HS proteoglycans are involved in morphogen signaling, recruitment of extracellular proteases,⁶¹ and are essential for brain development.⁶² CS proteoglycans are critical components of cartilage, where defects in CS biosynthesis result in diseases associated with loss of skeletal integrity.⁶³ CS also modulates axonal growth, as chondroitinase treatment facilitates recovery after injury to the central nervous system.^{64,65} Furthermore, disruption of CS biosynthesis in *C. elegans* interferes with cell division⁶⁶ and vulval morphogenesis.⁶⁷

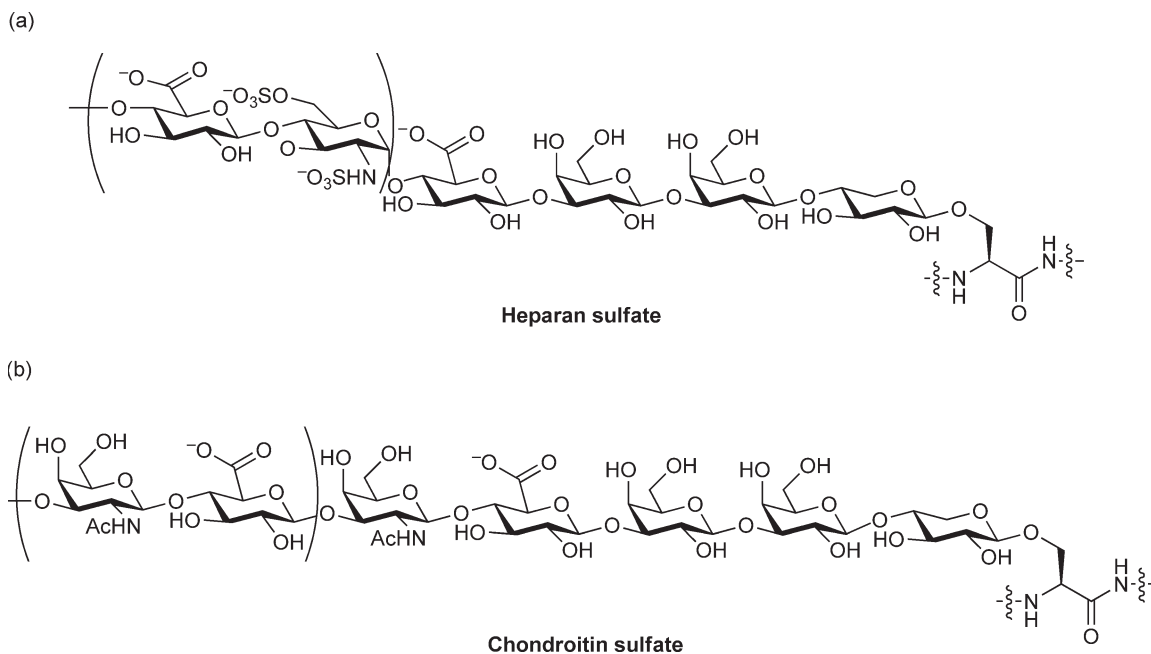


Figure 7 Example of proteoglycans. (a) Heparan sulfate; (b) chondroitin sulfate.

6.09.3 Protein Glycosylation in Prokaryotes

Protein glycosylation in bacteria was first identified in S-layer proteins from archaea and was thought to be absent in Gram-negative and Gram-positive bacteria.⁶⁸ Recently, the identification of glycoproteins from various pathogenic strains of bacteria and the specific genes responsible for protein glycosylation has demonstrated that most bacteria actually attach glycans to proteins.⁶⁹ Perhaps, the most striking protein glycosylation in bacteria occurs in the Gram-negative bacterial pathogen, *Campylobacter jejuni*, a major causative agent of human gastroenteritis and the neurodegenerative disorder known as Guillain–Barre syndrome.⁶⁹ Biochemical and genetic analyses revealed that *C. jejuni* contains a gene loci dedicated to N-linked glycosylation.⁷⁰ In fact, introduction of this gene locus into a laboratory strain of *Escherichia coli* endowed these normally protein glycosylation-incompetent bacteria with the capacity to produce N-linked glycoproteins.⁷¹ It should be noted that the N-linked glycan structure generated by bacteria is significantly different from eukaryotes (Figure 8(a)).⁶⁹ Whether N-linked glycosylation of proteins serves a similar function in bacteria as in eukaryotes remains to be determined, nonetheless, mutations in the N-linked glycosylation pathway attenuate the virulence of *C. jejuni* strains.⁶⁹

Bacteria also have several forms of O-linked glycosylation.⁶⁹ For example, *C. jejuni*, *Pseudomonas aeruginosa*, *Neisseria meningitidis*, and *Helicobacter pylori* all have O-linked glycans attached to their pili, extracellular protein polymers involved in cell adhesion and motility.⁶⁹ Protein glycosylation also influences the motor activity of the flagella in bacteria. In *P. aeruginosa*, the O-linked glycosylation of flagella appears to be critical for export of flagella subunits and virulence of the bacteria,⁷² whereas intracellular O-GlcNAcylation in *Listeria monocytogenes* regulates transcription of genes that control flagella function.⁷³ Protein glycosylation has also been reported in mycobacteria, although the precise molecular composition has not been fully characterized.⁷⁴ It has been suggested that dynamic glycosylation of extracellular proteins masks antigenic determinants, which would otherwise be detected by the immune system, thereby providing a mechanism for pathogenic bacteria to evade host immune responses.⁶⁹ The discovery of protein glycosylation in bacteria has revealed unique monosaccharides such as pseudaminic acid and fucosamine as well as their biosynthetic pathways.⁶⁹ The identification of protein glycosylation in bacteria has raised many questions about their functions in microbiology, their contribution to virulence of pathogens, and whether glycan-processing enzymes and bacterial glycoproteins may provide new targets for antibiotic and vaccine development, respectively. For further details on the bacterial protein glycosylation, see Chapter 6.12.

6.09.4 Chemical Approaches toward Understanding Protein Glycosylation

The survey of glycoprotein biosynthesis and function in the previous section highlights the tremendous progress that has been made in glycobiology.¹ These studies, however, have only begun to elucidate the mechanisms for many glycan-dependent processes. The genomics revolution has now provided the blueprint for many organisms and revealed the diversity of genes that may control glycan-dependent processes. With a nearly complete inventory of glycan-associated genes (i.e., glycosyltransferases, glycosidases, and lectins), the scientific community is poised to systematically dissect the function of these genes in normal physiology and disease. Despite the technical advances in genetics and chemistry to date, the realization of ‘functional glycomics’ will require the development of new methods to produce, perturb, and profile glycoconjugates in biology.

Chemistry has played a crucial role in understanding glycobiology.^{28,75,76} Chemical synthesis has provided homogeneous substrates to evaluate the specificity of glycan-processing enzymes and receptors. Moreover, the development of enzyme inhibitors and antagonists of glycan-binding proteins has afforded tools for evaluating glycan-dependent processes in a time- and dose-dependent manner. Collectively, these discoveries have paved the way for therapeutic strategies targeted at glycan-associated diseases. In the following sections, I will highlight chemical approaches to generate homogeneous glycoproteins, develop small molecules that perturb glycan-associated receptors and enzymes as well as profile glycoconjugates from complex mixtures. For further details on the chemical approaches to glycobiology, see Chapter 6.07.

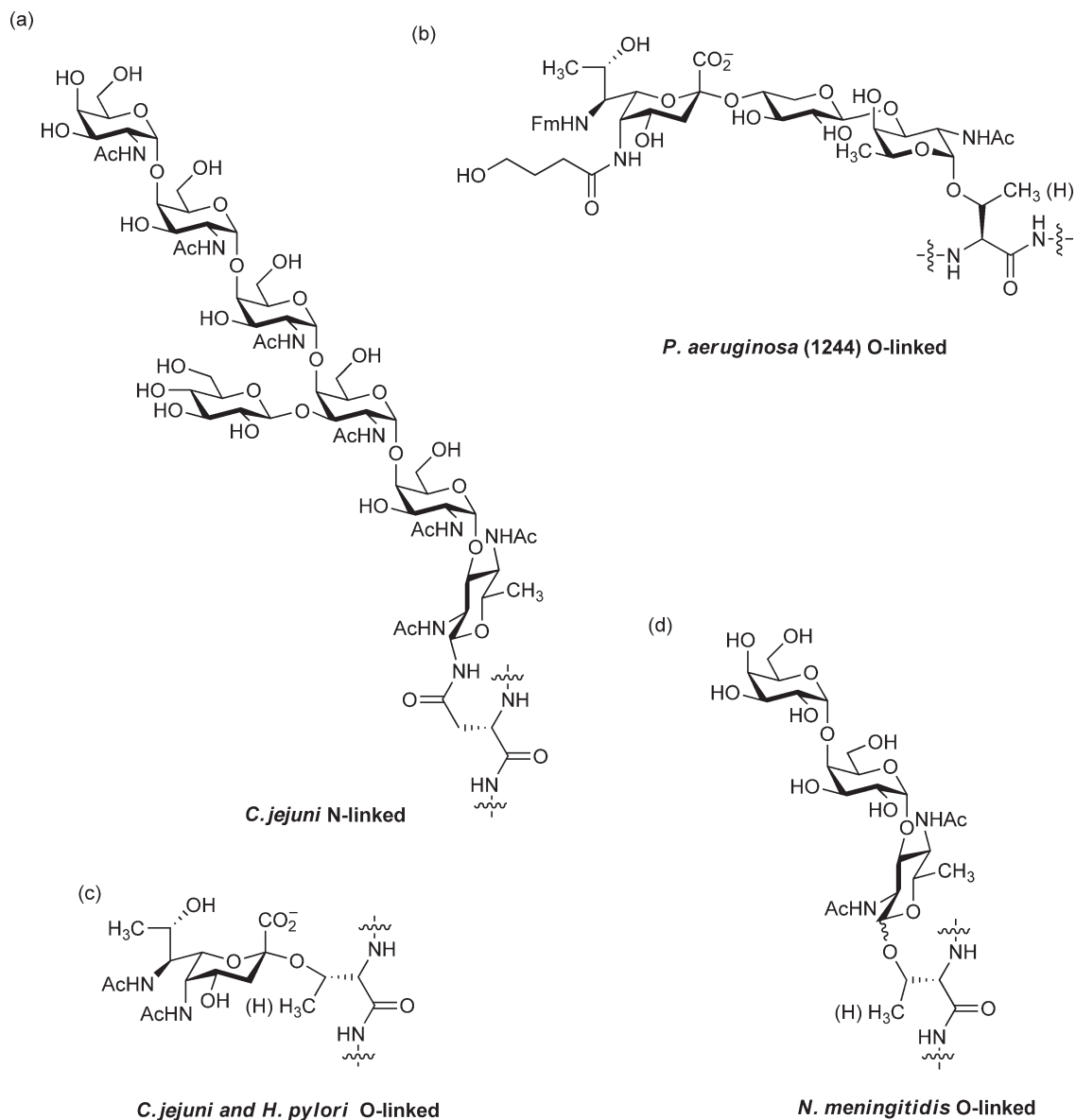


Figure 8 Examples of protein glycosylation in prokaryotes. (a) *Campylobacter jejuni* N-linked glycosylation; (b) *Pseudomonas aeruginosa* pilin O-linked glycosylation; (c) *C. jejuni* and *Helicobacter pylori* O-linked glycosylation; (d) *Neisseria meningitidis* O-linked.

6.09.5 Production of Homogeneous Glycoproteins

In the absence of efficient methods to isolate homogeneous glycoconjugates from natural sources, chemical synthesis provides the only means of obtaining homogeneous materials for biochemical and structural investigations.²⁸ The versatility of chemical synthesis also provides access to glycoconjugate derivatives for structure-function studies. Glycoproteins can be retrosynthetically deconstructed in three general ways (Figure 9); semisynthesis from glycopeptides (a), enzymatic elaboration of truncated glycoproteins (b), and directing coupling of complex glycans with proteins (c).^{77–79} The execution of these approaches has

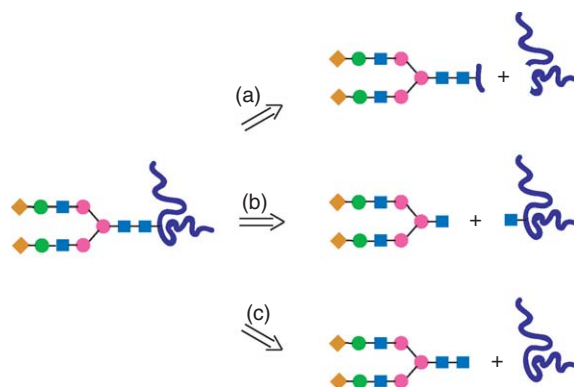


Figure 9 Strategies for the construction of glycoproteins. (a) Semisynthesis of glycopeptides; (b) elaboration of homogeneous glycoproteins; (c) site-specific condensation of complex glycans with proteins.

necessitated the development of robust methods for the synthesis of glycoconjugates and glycopeptides, semisynthesis, and expression of glycoproteins as well as innovative approaches to link complex glycans to proteins.

6.09.5.1 Methods for Glycoconjugate Synthesis

The size and complexity of glycoconjugates require sophisticated methods for regio- and stereo-selective glycosidic bond formation, a task that nature has evolved a large number of enzymes (~500 in higher eukaryotes) to perform. The invention of elegant activation methods for glycosidic bond formation and clever protecting group strategies have enabled the total synthesis of glycoconjugates with impressive size and complexity.^{28,80} In particular, Danishefsky and coworkers have synthesized a high mannose N-linked glycopeptide bearing the H-type 2 human blood group antigen (**Figure 10(a)**).⁸¹ These achievements in glycoconjugate synthesis have allowed the characterization of substrates for glycan-processing enzymes and ligands for glycan-binding proteins as well as the pursuit of glycan-based vaccines for immunotherapy.⁸²

For difficult chemical transformations, enzymatic methods can be used to complement organic synthesis for the construction of complex glycans.⁸³ Of note, Leppanen *et al.*⁸⁴ generated a series of glycosulfopeptides by chemoenzymatic synthesis to determine the precise binding epitope of the P-selectin (**Figure 10(b)**). The generality of enzymatic methods have been limited by the availability of enzymes and cost of nucleotide sugar donors in the past. However, improvements in protein expression methods and the construction of nucleotide sugar regenerating systems should reduce these barriers.⁸⁵ One major limitation of enzymatic synthesis has been the substrate tolerance of enzymes, but rational design⁸⁶ and directed evolution approaches⁸⁷ generate glycan-processing enzymes with altered or broadened substrate specificity to construct a wide variety of glycoconjugates.⁸⁸

The automated syntheses of oligonucleotides⁸⁹ and peptides⁹⁰ have been instrumental for providing access to homogeneous materials, which enabled a revolution in the field of molecular biology and recent advances in genomics and proteomics. Likewise, the development of automated methods for glycoconjugate syntheses is beginning to provide homogeneous materials for glycomics. The emergence of milder glycosylation conditions and new strategies for polymer-supported synthesis has enabled the construction of relatively large glycoconjugates and glycan-libraries on solid phase.⁹¹ From these methods, Seeberger and coworkers have developed an automated solid-phase oligosaccharide synthesizer that has enabled the rapid assembly of various glycoconjugates.⁹² For example, the automated solid-phase synthesis of a glycosylphosphatidylinositol (GPI) anchor structure facilitated the generation of a malaria vaccine candidate (**Figure 11**).^{93,94}

The introduction of the ‘arm’ and ‘disarm’ strategies for glycosidic bond formation has provided the foundations for one-pot multicomponent glycoconjugate synthesis in solution.⁹⁵ The extension of these reactivity principles to a large number of monosaccharide building blocks enabled Wong and coworkers to develop a one-pot programmable method for the synthesis of complex glycoconjugates.⁹⁶ The utility of the

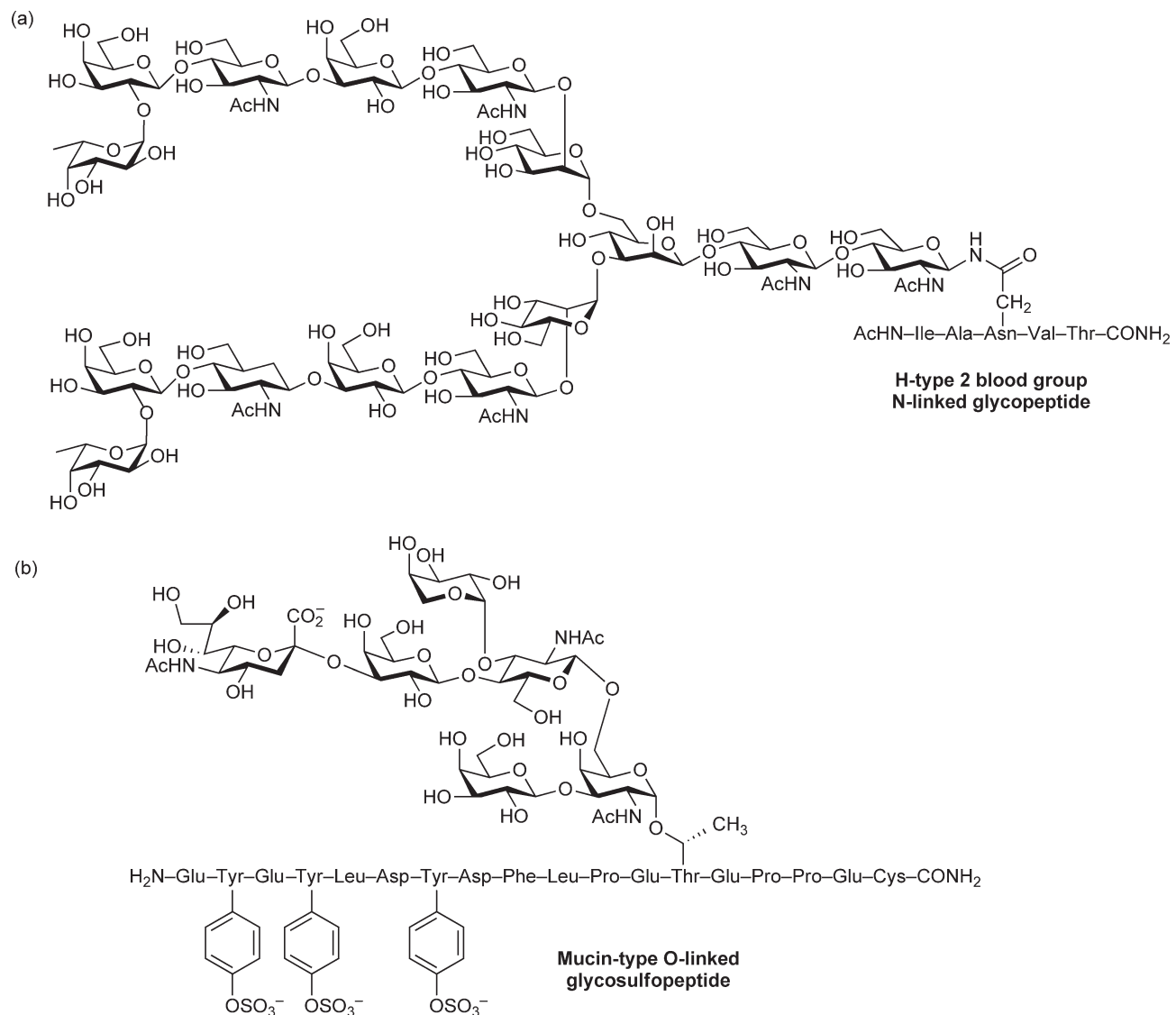


Figure 10 Highlights of glycopeptide synthesis. (a) H-type 2 blood group N-linked glycopeptide generated by chemical synthesis;⁸¹ (b) sulfated mucin-type O-linked glycopeptide generated by chemoenzymatic synthesis.⁸⁴

Automated solid-phase oligosaccharide synthesis

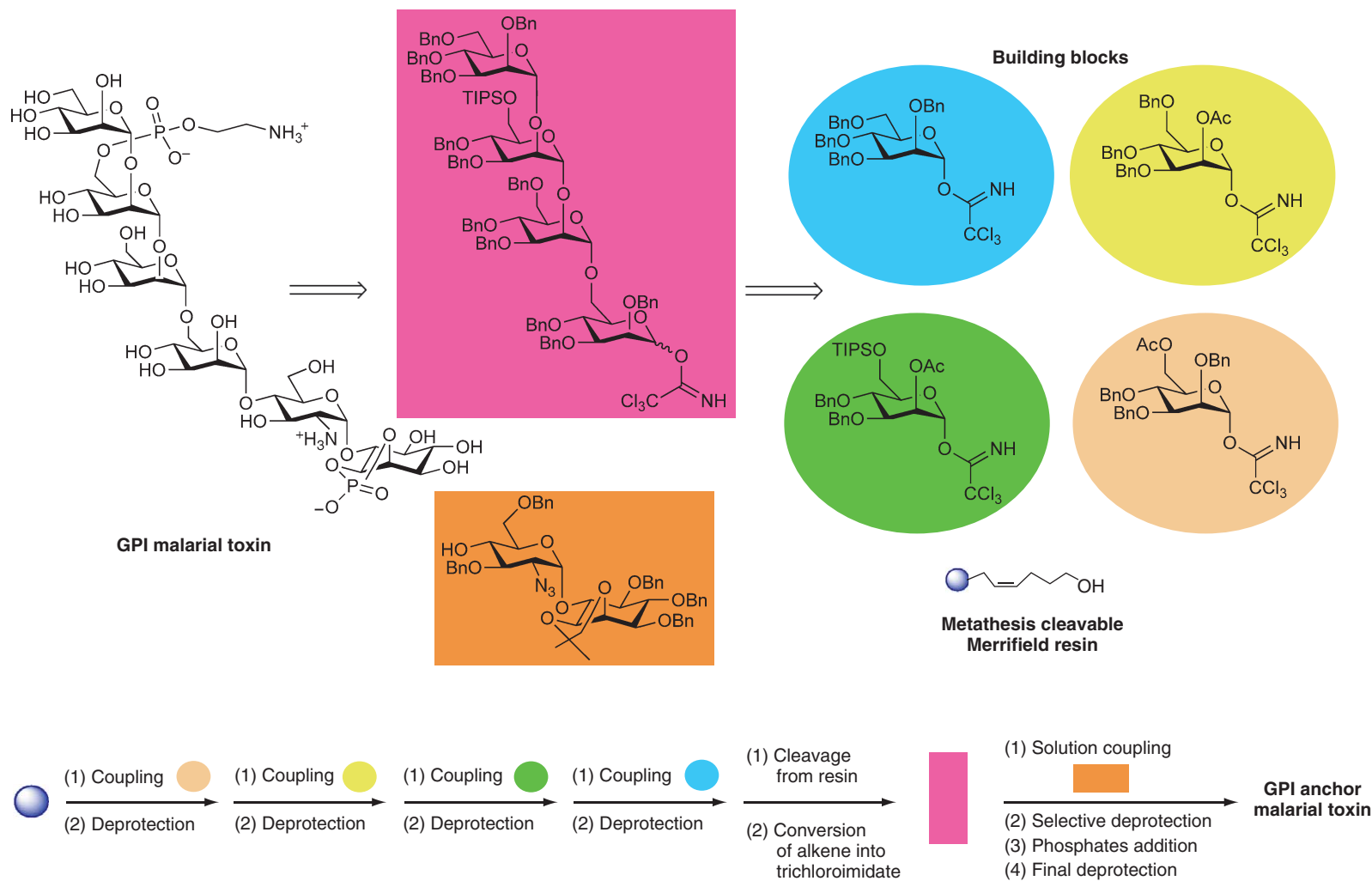


Figure 11 Automated oligosaccharide synthesis of GPI structure on solid phase.⁹³

automated one-pot programmable approach was demonstrated with the concise synthesis of the Globo H hexasaccharide (Figure 12).⁹⁷ Automated solid-phase synthesis and one-pot programmable approach hold tremendous promise for streamlining glycoconjugate synthesis and will be crucial for the system-wide analysis of glycan function. For further details on the glycoconjugate and glycopeptide synthesis, see Chapters 6.02 and 6.03.

6.09.5.2 Semisynthesis of Glycoproteins

The chemical production of glycoproteins largely is restricted by the size of polypeptides (~1 kDa) accessible by solid-phase peptide synthesis (SPPS). To overcome the limitations of SPPS, native and expressed protein ligation strategies have been developed to couple polypeptides bearing C-terminal thioesters with polypeptides containing N-terminal cysteine residues to afford larger proteins.⁹⁸ The application of these strategies with synthetic glycopeptides has enabled the construction of homogeneous glycopeptides/proteins with relatively large N-linked glycans^{99–101} and multiple O-linked glycans (Figure 13).¹⁰² Recently, Wong and coworkers have expanded the repertoire of glycoprotein semisynthesis methods by developing a sugar-assisted ligation approach where the reactive thiol group of cysteine residues has been moved onto the glycan itself.^{103–105} The reactive thiol group can be removed after the coupling of glycopeptide fragments, leaving no trace of the ligation site.¹⁰⁶ This method along with auxiliary-based ligation^{107–110} and auxiliary/cysteine-free¹¹¹ methods may broaden the scope of glycoprotein semisynthesis beyond the ligation site at cysteine residues.

6.09.5.3 Expression of Homogeneous Glycoproteins

Recombinant expression would be an ideal means of generating significant quantities of homogeneous glycoproteins for biochemical studies and therapeutic applications.¹¹² Unfortunately, glycoproteins from various overexpression systems have heterogeneous glycan structures that also vary between cell types and organisms. This heterogeneity and species difference has been recently addressed by combinatorial engineering of the budding yeast glycan-processing machinery to afford recombinant glycoproteins with uniform sialylated biantennary N-linked glycans.¹¹³ The ability to generate systematically recombinant glycoproteins with mammalian glycosylation patterns is a significant advance and presents exciting possibilities for production of homogeneous glycoproteins with defined structure.

Site-specific incorporation of unnatural amino acids *in vivo* presents a promising method for producing recombinant glycoproteins.¹¹⁴ The engineering of mutant aminoacyl-tRNA synthetases that utilize glycosylamino acids has made this possibility a reality in bacteria.¹¹⁴ The Wong and Schultz laboratories have successfully generated strains of bacteria capable of installing β -O-GlcNAc and α -O-GalNAc on Ser/Thr into recombinant proteins (Figure 14).^{115,116} The extension of this approach to other types of glycosylation, combined with enzymatic strategies⁷⁹ for the elaboration of core-glycan structures should enable production of homogeneous glycoproteins and glycoprotein mimetics.

The direct site-specific attachment of complex glycans onto full-length proteins with naturally occurring glycosidic linkages has not been possible. Nonetheless, chemoselective reactions have enabled the assembly of glycoprotein mimetics with unnatural linkages.^{77–79} For example, the unique reactivity of thiols has also been exploited for attaching synthetic glycans onto cysteine residues in peptides and proteins. Alternatively, glycans with anomeric thiols can be synthesized and coupled with aziridine-containing peptides.¹¹⁷ Other types of chemoselective reactions have also been exploited for generating glycopeptide/protein mimetics, which include hydrazone/oxime formation reactions^{77,118} and Huisgen [3 + 2] cycloaddition.¹¹⁹ Collectively, these approaches suggest that various types of homogeneous glycoproteins might be accessible in sufficient quantities in the near future for detailed biochemical studies that were previously impossible. For further details on the production of homogeneous glycoproteins, see Chapter 6.03.

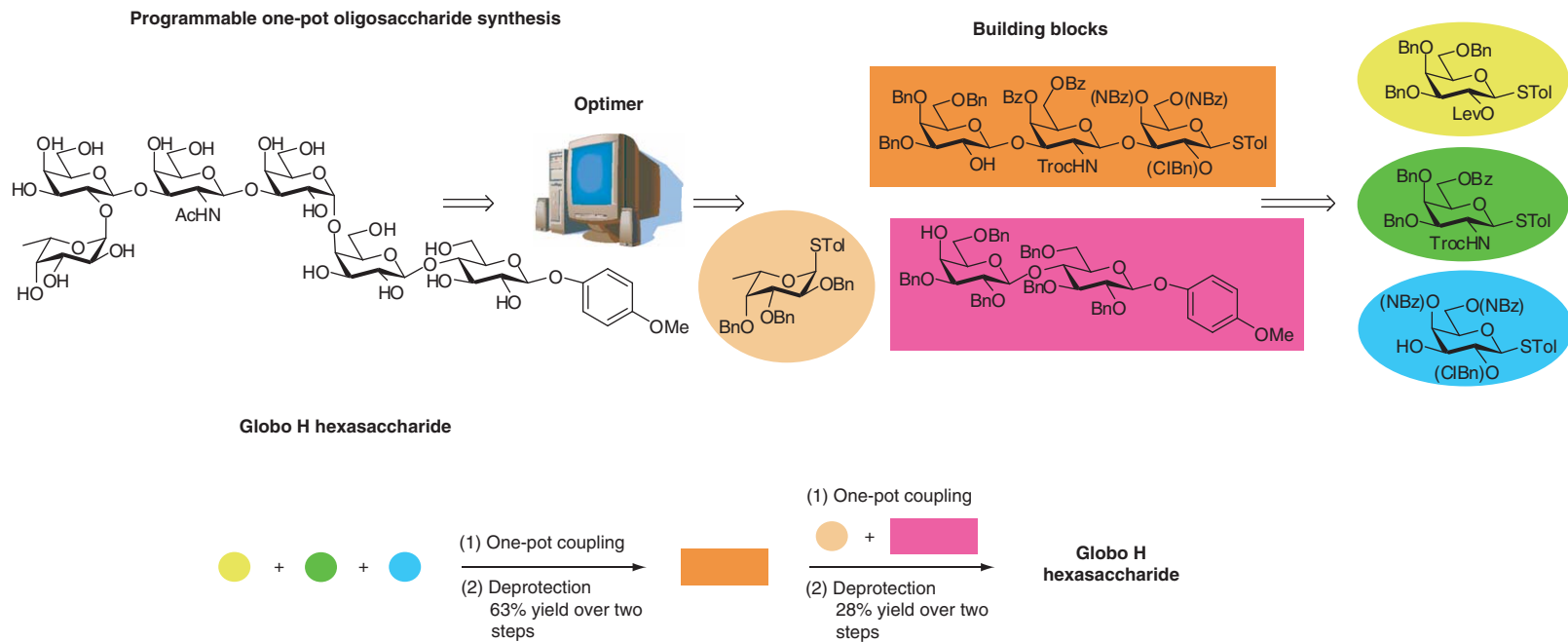


Figure 12 One-pot programmable synthesis of Globo H.⁹⁷

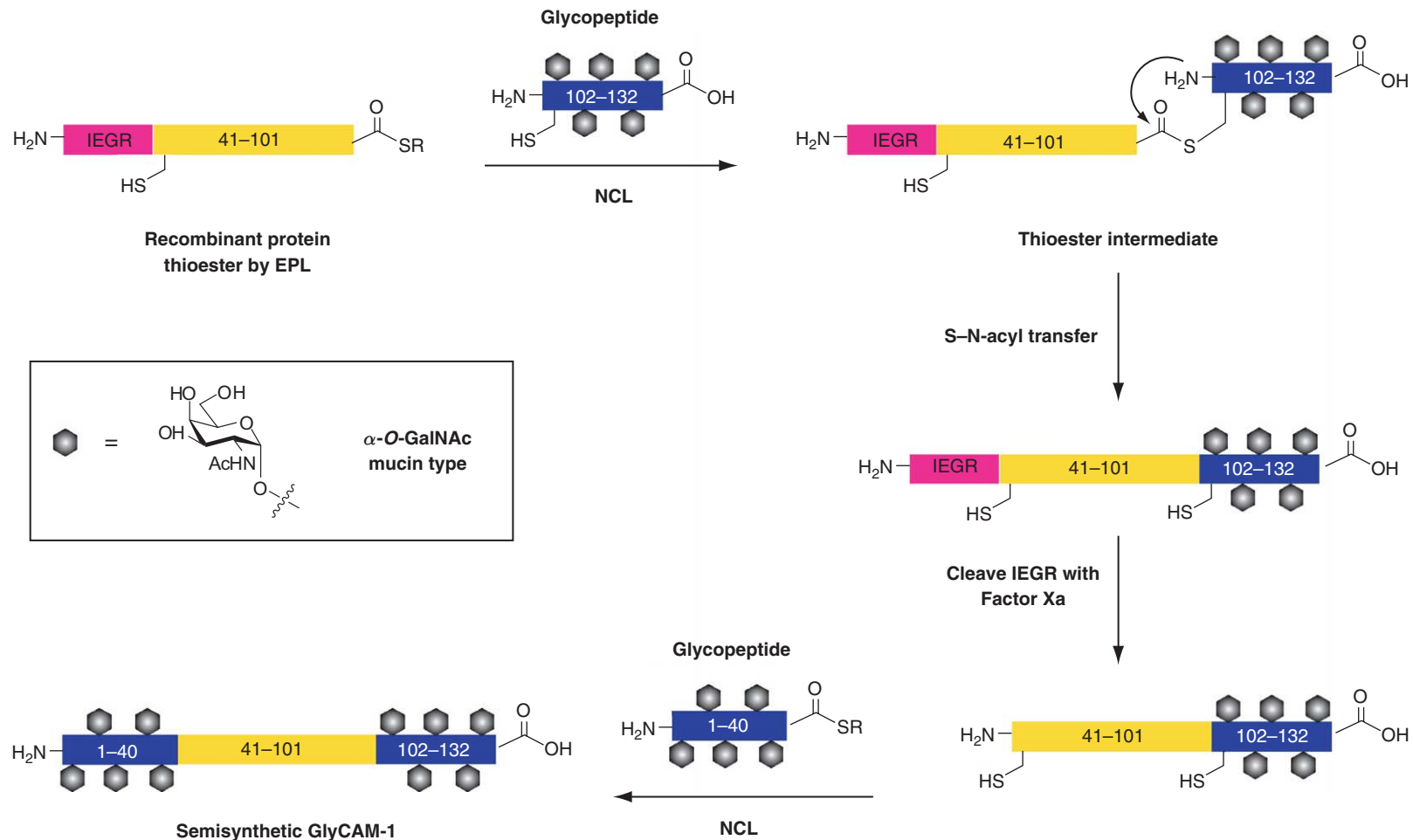


Figure 13 Semisynthesis of homogeneous glycosylated GlyCAM-1 by expression protein ligation (EPL) combined with native chemical ligation (NCL) of recombinant protein and synthetic glycopeptide fragments.¹⁰²

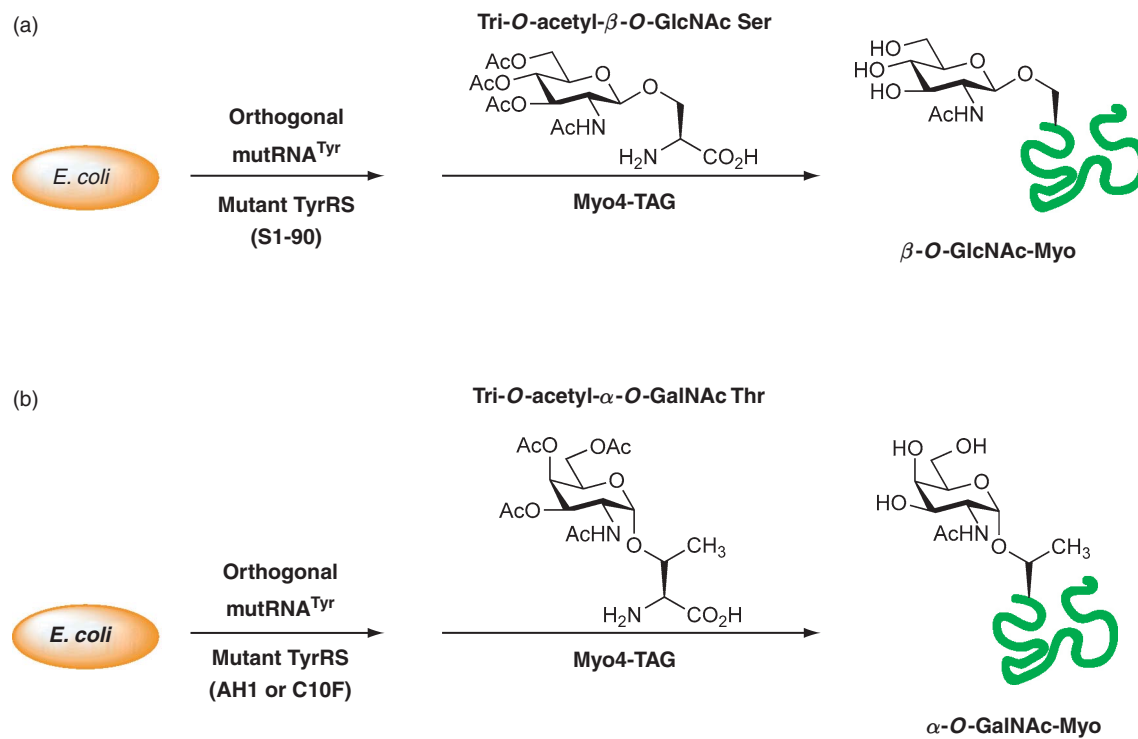


Figure 14 Expression of recombinant glycoproteins by unnatural amino acid mutagenesis. (a) β-O-GlcNAc-modified myoglobin;¹¹⁰ (b) α-O-GalNAc-modified myoglobin.¹¹¹

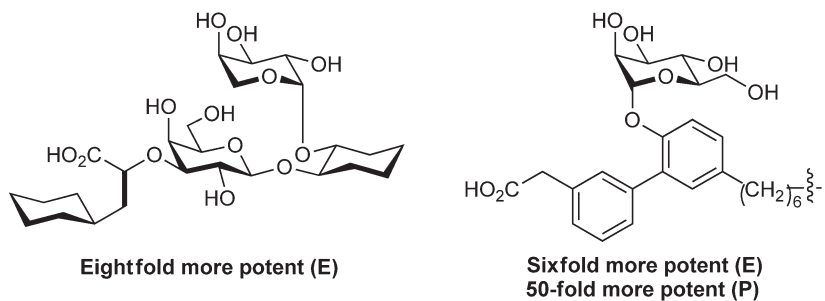
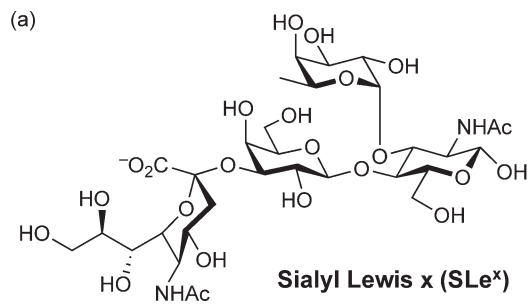
6.09.6 Chemical Probes for Perturbing Protein Glycosylation

Pharmacological and genetic perturbation of biological systems provide complementary approaches for elucidating biological function.¹²⁰ For example, the specific ablation of glycan-associated genes in various organisms has revealed new unique roles for glycans in many biological processes.¹²¹ Although the functional analysis of some glycan-associated genes has been hindered by embryonic lethality,¹²¹ the development of conditional and tissue-specific gene inactivation methods such as Cre-loxP site-specific recombination and tetracycline-inducible systems should enable functional analysis of these essential genes.¹²² In addition, the discovery of RNA interference and its application to mammalian systems presents exciting opportunities for genome-wide 'knock-down' of transcript levels *in vivo*.¹²³ These genetic methods target transcript levels that take days or weeks to observe phenotypes and may therefore be susceptible to downstream mechanisms of compensation that are not ideal for monitoring rapid physiological processes.¹²⁰ In contrast, small molecules provide a powerful means to modulate biological events on a rapid time scale that may also be tunable and reversible.¹²⁰ The identification of specific small molecule inhibitors of biological pathways is at the heart of drug discovery and has been very challenging.¹²⁴ Despite the hurdles in developing specific small molecule modulators of glycan-associated processes, progress has been made to pharmacologically perturb glycan–protein interactions and glycan-processing enzymes.^{75,76} Given that pharmacological and genetic perturbations do not necessarily afford similar phenotypes, the development of specific small molecules targeted at glycan-associated processes is not only critical for the development of therapeutics, but is also important for dissecting the glycan function in normal physiology.

6.09.6.1 Synthetic Modulators of Glycan–Protein Interactions

Glycan–protein interactions are ubiquitous in biology and represent important targets for therapeutic intervention. In general, glycan-binding proteins such as lectins recognize their oligosaccharide ligands with relatively shallow binding clefts, which result in relatively low binding affinities ($K_d \sim 10^{-3} - 10^{-4} \text{ mol l}^{-1}$) to their soluble oligosaccharide ligands.¹²⁵ These weak interactions reinforces the multivalent modes of binding that increase the overall avidity and specificity of glycan–protein interactions. The multivalent requirements for glycan–protein recognition are exemplified by the recent crystal structure of a HIV neutralizing antibody that binds truncated N-linked glycans on the surface of HIV particles.¹⁶ This structure determined by Calarese *et al.*¹⁶ revealed a dramatic alteration in the immunoglobulin-fold of the HIV neutralizing antibody compared to classical antibody–protein antigen complexes that enabled the multivalent-binding of the mannose-residues on N-linked glycans. The large surfaces of glycan–protein recognition are reminiscent of protein–protein interactions, which have made the development of low-molecular weight compounds that interfere with glycan–protein interactions very difficult. Nonetheless, remarkable advances toward small molecule glycan–protein antagonists have been made.

The discovery of the selectins (E, L, and P) as key players in the inflammatory cascade and the characterization of their ligands as sialyl Lewis x oligosaccharides (SLe^x, **Figure 15(a)**) generated tremendous enthusiasm for developing glycan–protein antagonists as anti-inflammatory therapeutics.¹²⁶ With limited structural data, the chemical synthesis of SLe^x derivatives mapped the key recognition elements of the selectins and afforded new synthetic antagonists.¹²⁷ For example, replacement of the sialic acid and GlcNAc residues with noncarbohydrate scaffolds afforded an antagonist with eightfold more potent E-selectin binding than SLe^x (**Figure 15(a)**).¹²⁸ Further simplification and dimerization of SLe^x derivatives afforded TBC1269, a P-selectin antagonist that was 50-fold more potent than SLe^x (**Figure 15(a)**).¹²⁹ The multivalent recognition of glycans by lectins has motivated the design of various polymer scaffolds for the multivalent display of carbohydrate ligands with improved potencies.¹³⁰ Remarkably, the presentation of a monomeric SLe^x mimetic on a polylysine scaffold improved the potency by 700-fold against E-selectin, resulting in an overall enhancement of $\sim 50\,000$ -fold over SLe^x (**Figure 15(b)**).¹³¹ Using ring-opening metathesis methods, Kiessling and coworkers generated a multivalent L-selectin antagonist with ~ 500 -fold improvement over SLe^x (**Figure 15(b)**).¹³² In addition, this multivalent L-selectin antagonist induces the cell surface proteolysis and shedding of L-selectin that provides a unique mechanism for receptor downregulation.¹³³



Small molecule SLe^x antagonists

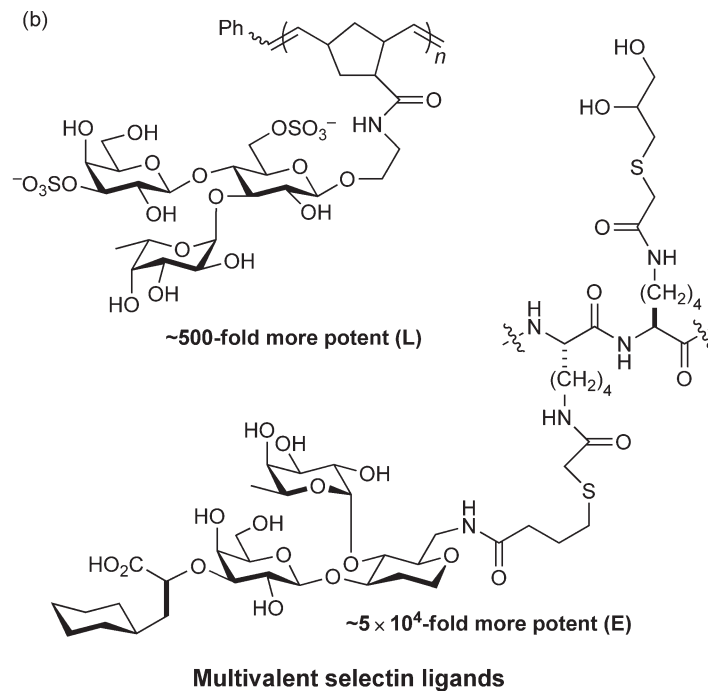
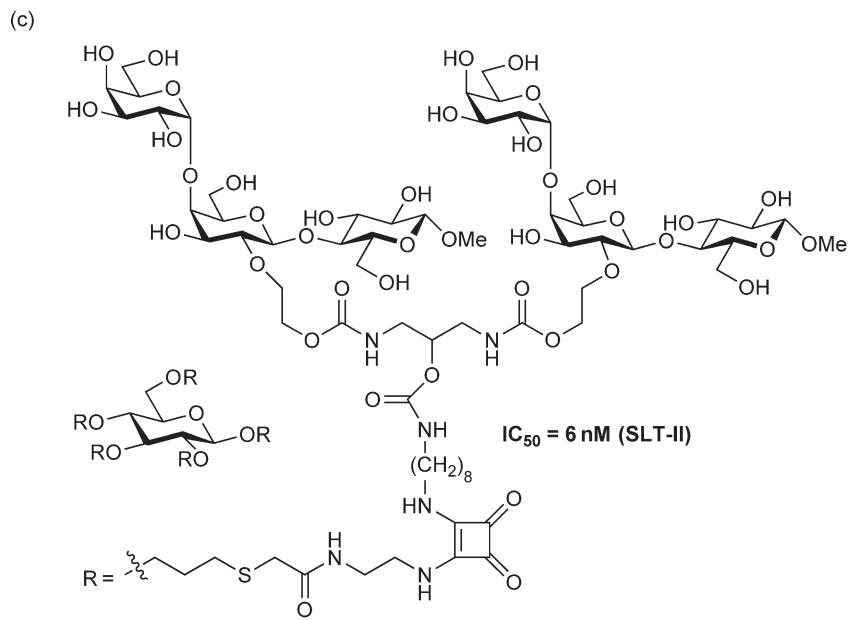
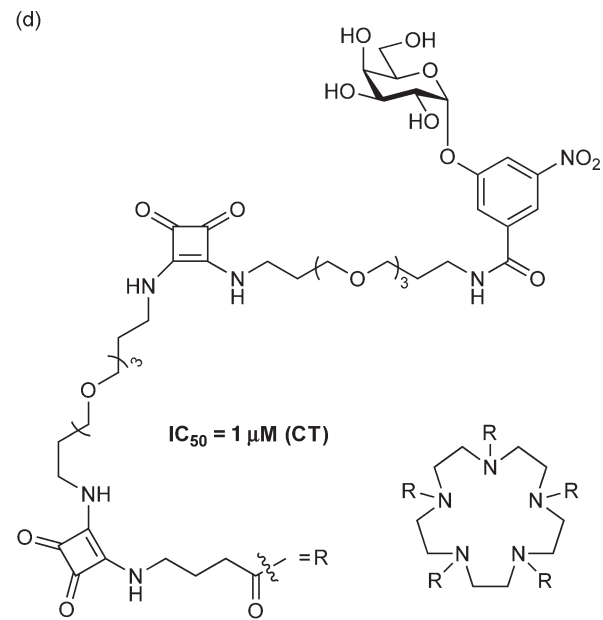


Figure 15 (Continued)



shiga-like toxin antagonist



Cholera toxin and Enterotoxin antagonist

Figure 15 Synthetic compounds that modulate glycan–protein interactions. (a) Sialyl Lewis x and small molecule antagonists;^{123,124} (b) multivalent selectin antagonists;^{126,127} (c) Shiga toxin-like antagonist;¹²⁹ (d) cholera toxin and enterotoxin antagonists.^{130,131}

Multivalent glycan-based antagonists have also been effective against glycolipid-binding toxins. In particular, the pentavalent display of the Shiga-like toxin¹³⁴ and cholera/enterotoxin^{135,136} ligands yielded multivalent antagonists with very potent inhibitory capacities compared to their monomeric units (Figures 15(c) and 15(d)). Alternatively, multivalent display of glycans can also cluster receptors on the surface of cells and induce specific signaling pathways such as bacterial chemotaxis.^{130,137}

Although these multivalent ligands are effective at modulating glycan–protein interactions, noncarbohydrate-based ligands would be more desirable for pharmacological perturbation *in vivo* and for therapeutic development. Toward this goal, Borrok and Kiessling¹³⁸ have developed a high-throughput fluorescence binding assay for glycan–protein binding and identified noncarbohydrate inhibitors of DC-SIGN, an important lectin expressed on dendritic cells involved in HIV and *Mycobacterium tuberculosis* recognition. The development of these synthetic modulators has broadened our understanding of glycan–protein interactions and revealed unique mechanisms of glycan-mediated signaling that should ultimately lead to novel therapeutics.

6.09.6.2 Perturbation of Glycan-Processing Enzymes with Small Molecules

Small molecules provide transient and dose-dependent modulation of glycan-processing enzymes, which can circumvent lethal phenotypes or compensatory mechanisms. For example, the discovery of natural product glycosidase inhibitors has provided powerful tools for modulating N-linked glycosylation in cells (Figure 16(a)).¹³⁹ Modifications of these compounds have subsequently afforded inhibitors of glycosphingolipid (GSL) biosynthesis, which prevent the accumulation of GSLs in lysosomal storage diseases (Figure 16(b)).¹⁴⁰ The genetic basis of lysosomal storage disorders are largely due to mutations in lysosomal glycosidases that give rise to misfolded proteins targeted for ER-associated degradation pathways.¹⁴¹ Remarkably, Fan *et al.*¹⁴² observed low doses of deoxygalactonorijimycin, a α -galactosidase inhibitor, administered to fibroblasts from Fabry patients accelerated the maturation of a mutant α -galactosidase A and restored its activity in lysosomal compartments. These observations have led to development of ‘chemical chaperones’ for misfolded glycosidases as treatments for neurodegenerative disorders¹⁴³ and Gaucher disease¹⁴⁴ (Figure 16(c)). Much progress has been made in glycosidase inhibitor design and is exemplified by the neuraminidase inhibitors that are currently on the market as anti-influenza drugs (Figure 16(d)).¹⁴⁵ For further details on glycosidase inhibitors, see Chapter 6.08.

Glycosyltransferase inhibitor design has historically been very challenging due to limited mechanistic and structural studies and have primarily focused on substrate-based mimetics and transition-state analogues.¹⁴⁶ Although nucleotide donor analogues have been useful for inhibition studies *in vitro*, they are typically charged and membrane-impermeable, precluding their use in living cells. Competitive substrates like benzyl- α -GalNAc (Figure 16(e))¹⁴⁷ and pNP- β -Xyl (Figure 16(f))¹⁴⁸ have effectively blocked the elaboration of proteoglycans and mucin-type O-linked glycans in cells, respectively. However, these compounds are often toxic to cells and nonspecifically target other downstream glycosyltransferases. The development of acetylated disaccharide decoys has partially circumvented these issues, reducing the effective concentration of competitive substrates needed and increasing specificity (Figure 16(g)).¹⁴⁹ Indeed, these compounds block the synthesis of terminal glycans like SLe^x in cells and present promising candidates for cancer therapy *in vivo*.¹⁵⁰

Unnatural monosaccharides exploited to interfere with glycosylation-dependent events. Pioneering work by Reutter and coworkers demonstrated analogues of *N*-acetylmannosamine (ManNAc), such as *N*-propanoylmannosamine (ManProp) are metabolized into their corresponding cell surface sialosides and blocked sialic acid-dependent viral infection (Figure 16(h)).¹⁵¹ Subsequent studies by Mahal *et al.*¹⁵² demonstrated that *N*-butanoylmannosamine (ManBut) was metabolized into sialosides and interfered with the biosynthesis of α 2-8-linked polysialic acid in cells (Figure 16(h)). In the absence of specific and cell-permeable sialyltransferase inhibitors, ManNAc analogues provide useful chemical tools for modulating sialic acid-dependent processes.

The rational design of glycosyltransferase inhibitors has been on substrate mimicry. For example, Lowary and coworkers synthesized 3-amino-3-deoxy(Fuc α 1-2)Gal β -O(CH₂)₇CH₃ as a competitive inhibitor of the blood group A *N*-acetylgalactosaminyltransferase (Figure 16(i)).¹⁵³ The administration of this compound to cells afforded only a modest reduction of the blood group A determinant expression, which was attributed to

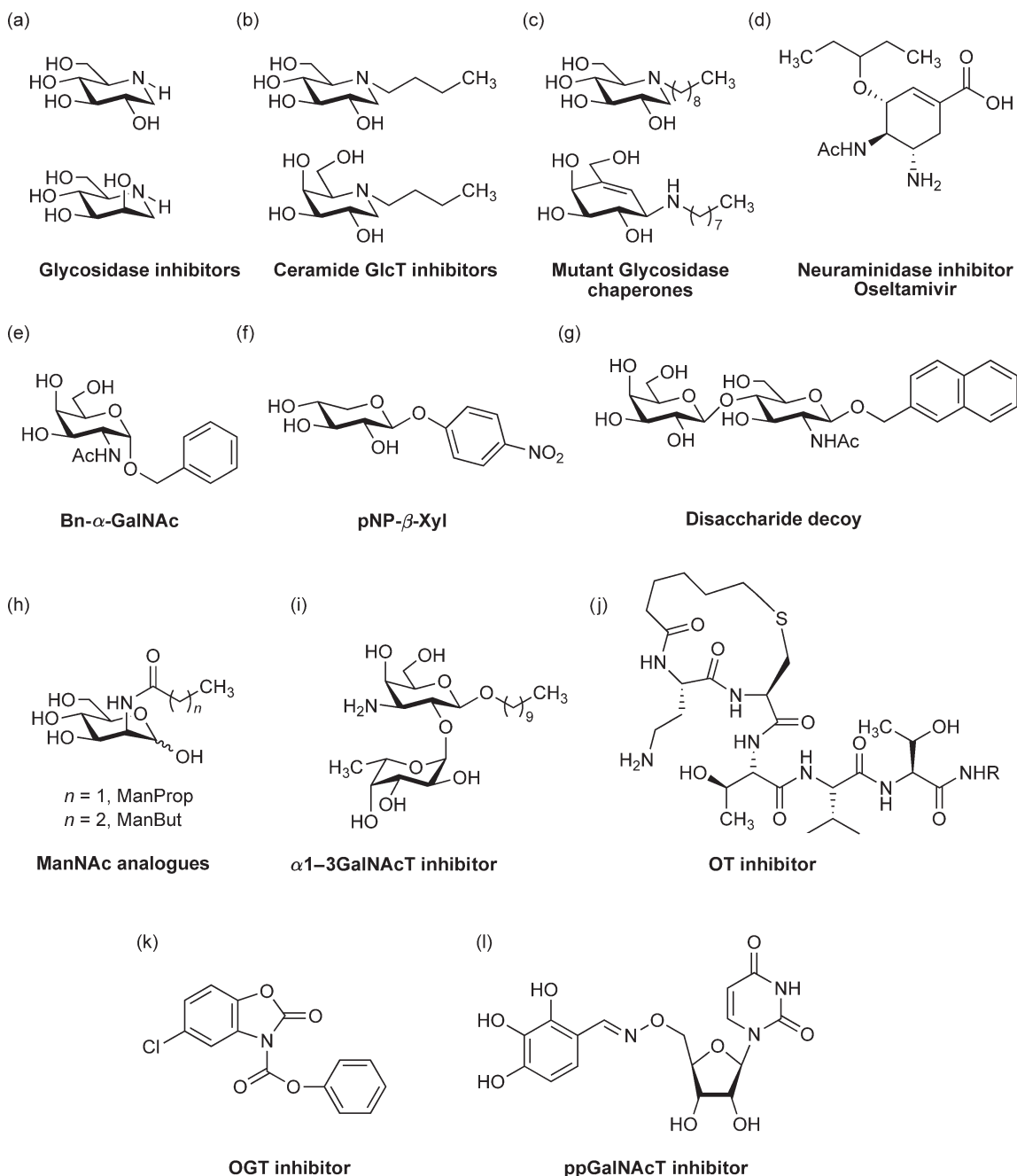


Figure 16 Compounds designed to interfere with glycoconjugate biosynthesis and processing. (a) Glycosidase inhibitors: deonorijimycin (top) and deomannonorijimycin (bottom);¹³⁴ (b) ceramide glucosyltransferase inhibitors: *N*-butyl deonorijimycin (top) and *N*-butyl deogalactorijimycin (bottom);¹³⁵ (c) chemical chaperones of mutant glycosidases;^{137–139} (d) anti-influenza neuraminidase inhibitor;¹⁴⁰ (e) Bn-α-GalNAc substrate-based inhibitor of mucin-type O-linked glycosylation elaboration;¹⁴² (f) pNP-β-Xyl substrate-based inhibitor of proteoglycan elaboration;¹⁴³ (g) disaccharide decoy inhibitor of SLe^x biosynthesis;^{144,145} (h) *N*-acetylmannosamine analogues: ManProp ($n = 1$) inhibitor of influenza infection,¹⁴⁶ and ManBut ($n = 2$) inhibitor of PSA biosynthesis;¹⁴⁷ (i) blood group A antigen α1-3GalNAcT inhibitor;¹⁴⁸ (j) OT inhibitor;¹⁴⁹ (k) OGT inhibitor;¹⁵¹ (l) broad-spectrum ppGalNAcT inhibitor.¹⁵⁴

poor bio-availability under physiological conditions. In an effort to inhibit N-linked glycosylation in cells, Imperiali and coworkers appended peptide import sequences onto substrate-based oligosaccharyltransferase (OT) inhibitors (**Figure 16(j)**).¹⁵⁴ Unfortunately, the need to penetrate both the plasma and ER membranes, coupled with proteolytic activities in the cytosol, undermined the efficacy of these peptide-based OT inhibitors in cells.

Given the challenges associated with the rational design of cell-permeable glycosyltransferase inhibitors, high-throughput glycosyltransferase assays have been developed to screen chemical libraries for new pharmacological leads. Using the bacterial glycosyltransferase MurG, Walker and coworkers developed a high-throughput fluorescence polarization assay for glycosyltransferases and identified a novel MurG inhibitor from approximately 50 000 compounds.¹⁵⁵ The Walker lab has also employed screening approaches to identify *O*-GlcNAc transferase inhibitors,^{156,157} one of which appears to interfere with cell cycle transition in cell extracts (**Figure 16(k)**).^{156,157} To target glycosyltransferases and other nucleotide sugars utilizing enzymes, a uridine-based library was synthesized by Winans and Bertozzi.¹⁵⁸ Evaluation of this focused library afforded a potent inhibitor of the UDP-GlcNAc-C₄-epimerase, a key metabolic enzyme that controls the synthesis of GalNAc-containing glycoconjugates in cells.¹⁵⁸ Subsequent screening of the uridine-based library with a microtiter plate assay for the ppGalNAcTs afforded two structurally related inhibitors of this glycosyltransferase family (**Figure 16(l)**).¹⁵⁹ Collectively, the growing number of glycosyltransferase structures,¹⁶⁰ the synthesis of focused small molecule libraries, and development of high-throughput assays for glycosyltransferases suggest cell-permeable glycosyltransferase inhibitors might be available soon. For further details on the enzymology of glycosyltransferases, see Chapter 8.11.

Chemical inducers of dimerization (CIDs) provide a powerful method to control protein-protein interactions with small molecules and have been applied to a variety of biological systems.¹⁶¹ This approach has been employed to control the Golgi localization of a fucosyltransferase (FucT) VII catalytic domain and consequently its activity in cells (**Figure 17**).¹⁶² The system designed was such that in the absence of rapamycin, the catalytic domain of a FucTVII-FRB chimera is secreted and minimal SLe^x is synthesized. Upon the addition of rapamycin, the FucTVII-FRB chimera is retained in the Golgi apparatus through association with FucTVII localization domain FKBP chimera, restoring FucTVII activity and SLe^x synthesis in cells. Until specific cell-permeable glycosyltransferase inhibitors are developed, this approach provides a rapid and conditional means to control glycosyltransferase activities in living cells.

6.09.7 Global Analysis of Protein Glycosylation

The advances in mass spectrometry (MS) for subfemtomolar detection of biomolecules have enabled the global analysis of protein glycosylation to be addressed.^{163–165} Glycoproteomics encompasses the large-scale identification of glycoproteins, the site(s) of glycosylation, and structure of glycans attached to each site from complex mixtures. Gel-based approaches for separation of proteins from cell lysates have begun to reveal the identity of many glycoproteins as well as the glycans attached to them from a variety of tissues.¹⁶⁶ However,

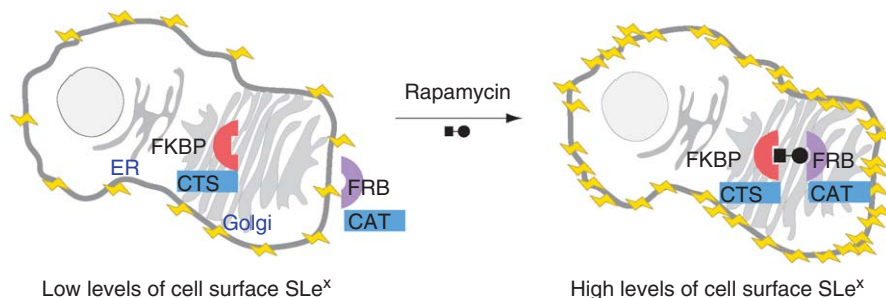


Figure 17 Modulation of cell surface glycosylation by small control of glycosyltransferase localization.¹⁵⁷ In the absence of rapamycin the catalytic domain of FucT VII is secreted. Upon addition of rapamycin the catalytic domain of FucT VII is retained via heterotrimeric complex in the Golgi, restoring the biosynthesis of SLe^x.

the inherent microheterogeneity of glycoproteins complicates their separation by gel-based methods and limits the repertoire of glycoproteins identified from crude lysates. Therefore, several approaches have been developed to selectively enrich different classes of glycoproteins from cell and tissue lysates. For further details on the glycoanalyses, see Chapters 6.05 and 6.06.

6.09.7.1 Specific Enrichment of Glycoproteins for Proteomic Analysis

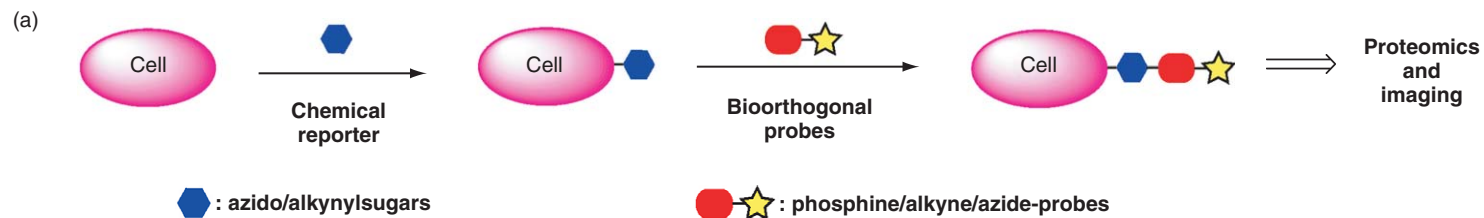
Lectins and antibodies provide selective means to enrich subsets of glycoproteins from complex mixtures (**Figure 18(a)**). For example, lectin-affinity chromatography enabled the large-scale identification of N-linked glycoproteins from complex mixtures by MS.¹⁶⁷ Alternatively, the availability of specific antibodies toward β -O-GlcNAc residues has revealed several nuclear/cytoplasmic proteins that bear this modification.¹⁶⁸ In theory, lectin- or antibody-affinity chromatography applies to many different types of glycoproteins. In practice, the generality of the approach remains to be established given the limited specificity of many lectins and antibodies as well as weak affinities for their glycan targets.

Chemical methods also contribute to glycoproteomics. Kaji *et al.*¹⁶⁹ has utilized the periodate oxidation of glycans followed by chemoselective reaction with hydrazide beads to capture glycopeptides from complex mixtures. Treatment of the beads with peptide N-glycosidase F (pNGase-F) specifically liberates peptides for MS analysis and allows the identification of N-linked glycoproteins. This approach also captures O-linked glycopeptides on solid phase. The development of mild conditions to cleave the O-linked glycopeptides from the resin should therefore allow the identification of O-linked glycoproteins as well.

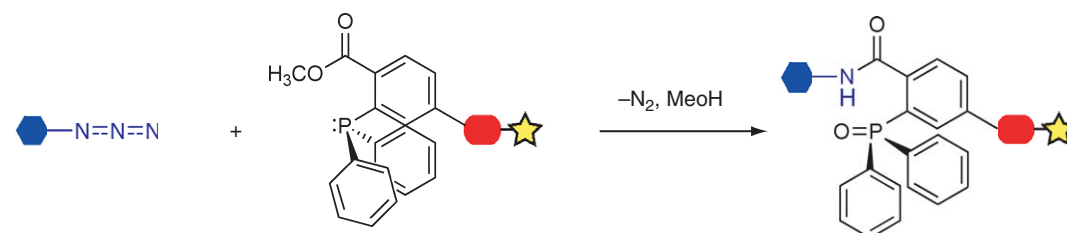
Unnatural substrates with selectively reactive chemical handles provide unique opportunities for the analysis of glycoconjugates that is not possible with radioactive reagents. For example, UDP-Gal modified with biotin at the 6-position was utilized by bovine β 1-4galactosyltransferase (β 1-4GalT), which enabled the nonradioactive detection of β 1-4GalT acceptor substrates from glycoprotein samples.¹⁷⁰ Unfortunately, the addition of biotin is not always tolerated by enzymes, as the 6-biotinylated derivative of UDP-GalNAc was not tolerated by the ppGalNAcTs.¹⁷⁰ Ketones or azides are relatively small chemical groups that introduce only subtle steric perturbations into substrates and often tolerated by enzymes. Using a mutant β 1-4GalT, Khidekel *et al.*¹⁷¹ demonstrated β -O-GlcNAc residues on proteins could be enzymatically modified with a ketone derivative of UDP-Gal and detected following aminoxy-biotin conjugation (**Figure 18(b)**). This approach has proven to be relatively robust and enabled the identification of several β -O-GlcNAc-modified proteins, their sites of modification and even allowed quantitative comparison of β -O-GlcNAc-modified proteins from brain tissues.¹⁷²

6.09.7.2 Chemical Reporters of Protein Glycosylation

The promiscuity of the glycoconjugate biosynthetic pathways have also been exploited for glycoproteomics.¹⁷³ The metabolic incorporation of unnatural monosaccharides adorned with uniquely reactive chemical handles enables selective labeling and enrichment of glycoconjugates for proteomic analysis (**Figure 19(a)**). Initial experiments demonstrated ketone-bearing monosaccharides could be metabolized into sialic acid- and GalNAc-containing glycoconjugates and selectively reacted with hydrazide reagents on surface of cells.^{174,175} Unfortunately, the modest incorporation of the ketone derivatives was not ideal for proteomic applications. The development of the Staudinger ligation, a chemoselective reaction between phosphines with alkyl or aryl azides, which form stable covalent adducts under physiological conditions has enabled the use of azide-modified substrates for metabolic labeling (**Figures 19(b) and 19(c)**).¹⁷⁶ The azide group is sterically less obtrusive and more readily tolerated by enzymes, which has enabled the use of azide-modified sugars for metabolic labeling of sialylated glycoproteins,^{176,177} β -O-GlcNAc-modified proteins,^{178–180} and mucin-type O-linked glycoproteins¹⁸¹ (**Figure 19(c)**). Importantly, the azide provides a chemical handle for reaction with a variety of phosphine reagents, which enables detection and enrichment of azide-modified biomolecules (**Figure 19(b)**).¹⁸² Several β -O-GlcNAc-modified proteins have been identified from mammalian cells using this approach.^{179,180} Proteomic studies on sialylated and mucin-type O-linked glycoproteins using these chemical reporters are in progress.¹⁸³ Remarkably, these chemical reporters also function in living animals and afforded selective labeling of glycoproteins in mice and zebrafish.^{184–186}



(b) **Staudinger ligation**



Click chemistry – Cu^I-mediated



Click chemistry – strain promoted

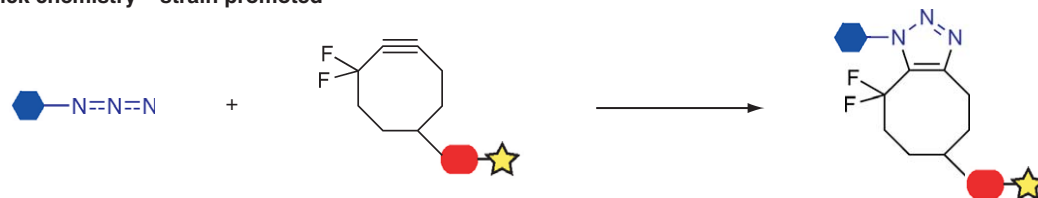
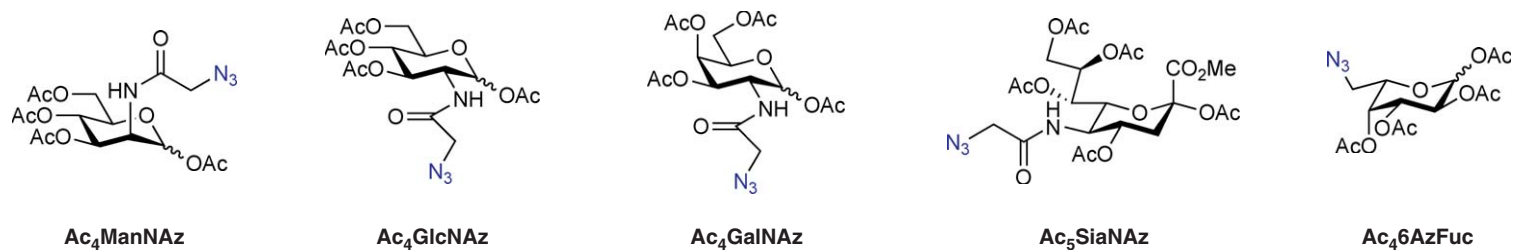


Figure 19 (Continued)

(c)

Azidosugars



Alkynylsugars

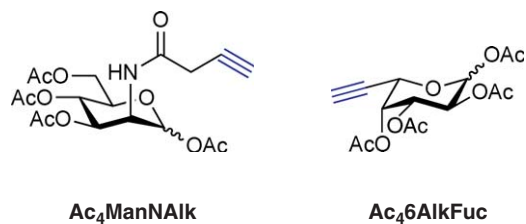


Figure 19 Analysis of glycoproteins in complex mixtures with chemical reporters. (a) Metabolic incorporation of chemical reporters, followed by bioorthogonal labeling with specific probes for proteomics and imaging applications; (b) bioorthogonal labeling methods: Staudinger ligation,¹⁷¹ Cu^I-mediated click chemistry^{181,182} and strain-promoted click chemistry;^{183,184} (c) chemical reporters for sialylated glycans: ManNAz, SiaNAz, ManNAIk, β -O-GlcNAc-modified proteins: GlcNAz, mucin-type O-linked glycoproteins: GalNAz, fucosylated glycans: 6AzFuc, 6AlkFuc. The chemical reporters are acetylated to facilitate cellular uptake.^{171–177,185–188}

The scope and utility of these chemical reporters is expanding. The application of the Huisgen [3 + 2] cycloaddition or ‘click chemistry’ reaction provides a complementary approach to the Staudinger ligation for selectively labeling azide-modified substrates in biological settings (Figure 19(b)).^{187,188} A comparative analysis of the bioorthogonal labeling reactions suggests that click chemistry may function better for proteomics applications.¹⁸⁹ Click chemistry typically requires the Cu^I-mediated catalysis, but the design of strained cycloalkyne probes has enabled rapid copper-free reactions, selective labeling of azide-modified glycoconjugates,^{190–194} and even allowed the dynamic imaging of protein glycosylation *in vivo*.¹⁸⁶ The chemical reporters for protein glycosylation have also been expanded to include fucose analogues^{195–197} as well as the use of alkynylsugars^{196–198} and thiosugars¹⁹⁹ (Figure 19(c)). Moreover, chemical reporters have also been used in bacteria,^{200–202} which suggests that these approaches could be extended for proteomic analysis of bacterial glycoproteins. The development of a robust bioorthogonal labeling and affinity enrichment protocol should facilitate the proteomic analysis of various glycoproteins targeted by these chemical reporters.

Chemical reporters can also facilitate *in vitro* studies. Cellular and biochemical studies demonstrated the azide analogues of UDP-GlcNAc and UDP-GalNAc were utilized by the β -O-GlcNAc transferase¹⁷⁸ and the ppGalNAcTs,²⁰³ respectively. To facilitate the biochemical analysis of glycosyltransferases, a high-throughput assay termed the ‘azido-ELISA’ (enzyme-linked immunosorbent assay) was developed to rapidly quantify the transfer of azide analogues onto biotinylated acceptor substrates (Figure 20(a)).²⁰³ The azido-ELISA and UDP-GalNAz enabled Pratt *et al.*²⁰⁴ to profile the glycopeptide substrate specificities of several ppGalNAcT isoforms, which revealed that most of the ppGalNAcTs prefer glycosylated peptide substrates with unique and overlapping substrates (Figure 20(b)). This systematic study demonstrated that the ppGalNAcT isoforms may be classified into subsets of glycosyltransferases based on their biochemical preference for nascent polypeptides, partially glycosylated or densely glycosylated substrates (Figure 20(b)).²⁰⁴ This approach has also been utilized for the analysis of β -O-GlcNAc transferase peptide substrates.²⁰⁵ In addition, the application of UDP-GalNAz and the Cu^I-catalyzed Huisgen [3 + 2] cycloaddition has facilitated the fluorescent detection of O-GlcNAc-modified proteins.²⁰⁶ Therefore, these *in vitro* chemical reporters provide useful alternatives to radioactive substrates for investigating the biochemical activities of glycosyltransferases in a high-throughput manner.

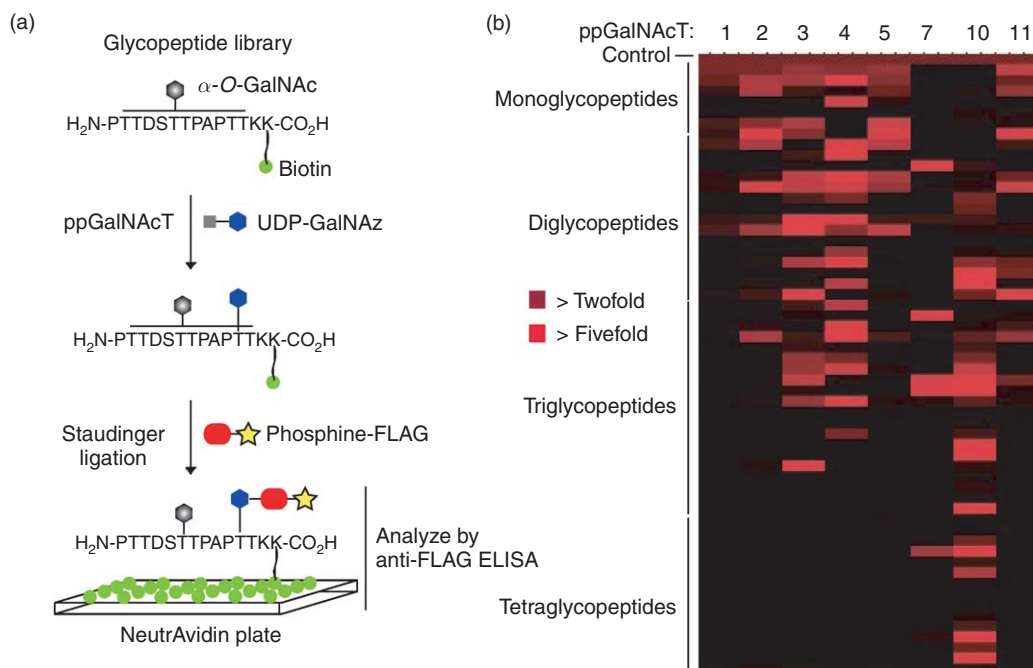


Figure 20 Rapid profiling of glycosyltransferase substrate specificities *in vitro* using chemical reporters. (a) Azido-ELISA schematic for profiling ppGalNAcT substrate specificities;¹⁹³ (b) comparative analysis of ppGalNAcT activities with a library of glycopeptide substrates.¹⁹⁴

6.09.7.3 Mechanism-Based Probes for Profiling Glycan-Processing Enzymes

The development of mechanism-based probes has yielded a powerful set of chemical tools to profile enzymes in complex mixtures and characterize nonannotated polypeptide sequences.²⁰⁷ Mechanism-based probes are typically generated by the functionalization of covalent enzyme inhibitors with visualization and/or affinity tags (fluorophores, biotin, or azides) that allow selective labeling and detection of enzymes in cell lysates or even in cells.²⁰⁷ The design of mechanism-based probes for glycosidases is based upon the pioneering studies from the Withers laboratory, which demonstrated that fluorosugars could covalently modify key glutamate residues in the active site of glycosidases.^{208,209} In particular, the synthesis of azide-modified fluorosugar probes has enabled the selective labeling and characterization of bacterial glycosidases in cell lysates (Figure 21(a)).^{210–214} Further development of these mechanism-based probes should enable the identification and comparative analysis of glycan-processing enzymes in normal physiology and disease.

6.09.7.4 FRET-Sensors of Protein Glycosylation

The dynamic analysis of specific enzyme activities in living cells has been difficult to achieve, but has recently become possible with the advent of Förster resonance energy transfer (FRET)-sensors.²¹⁵ These FRET-sensors are fluorescent protein chimeras that are designed to undergo large conformation shifts based upon the enzymatic modification of a specific substrate sequence, which in turn alters the fluorescent properties of the protein reporter.²¹⁵ The induced conformation change typically brings the donor chromophore in close proximity to an acceptor chromophore such that excitation of the donor leads to FRET and subsequent fluorescence emission from the acceptor.²¹⁵ FRET-sensors of this type have been developed for kinases, methyltransferases as well as calcium sensors.^{215–217} A FRET-sensor for the β -O-GlcNAc transferase has recently been reported by construction of a protein chimera composed of the GafD lectin domain fusion to a β -O-GlcNAc acceptor substrate flanked by the cyan and yellow fluorescent proteins (CFP and YFP,

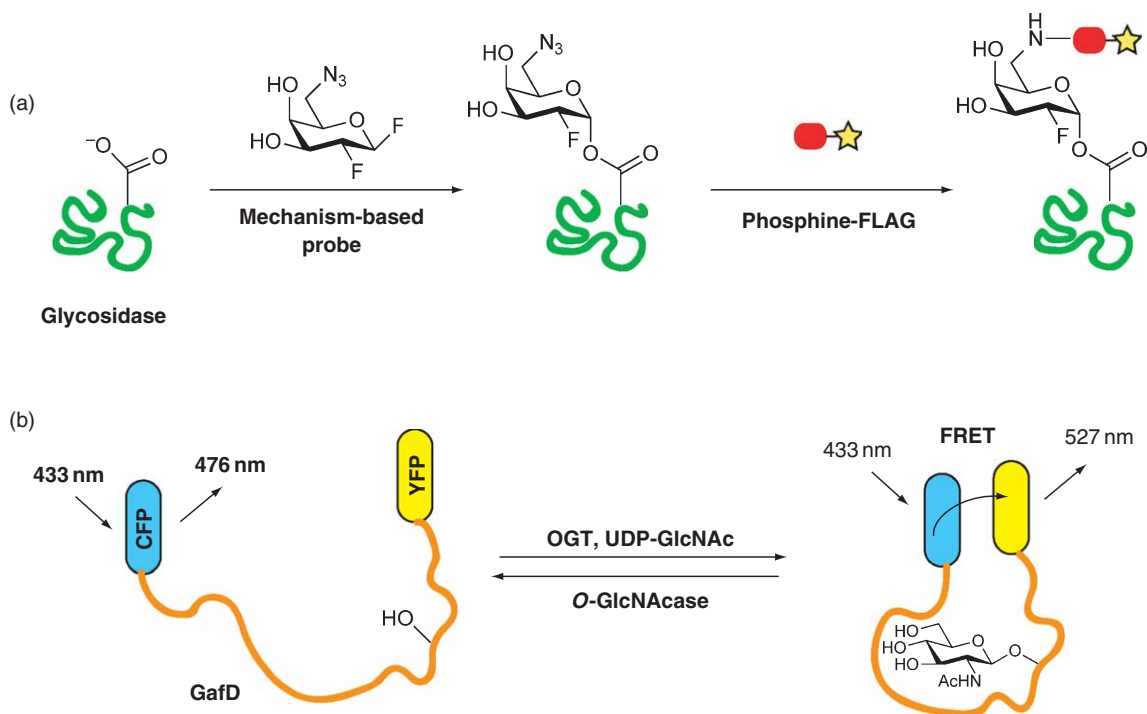


Figure 21 Chemical probes for monitoring glycan-processing enzymes. (a) Mechanism-based probe for labeling active exo-glycosidases in cell lysates;¹⁹⁹ (b) FRET-based reporter for OGT activity in cells.²⁰⁶

respectively) (Figure 21(b)).²¹⁸ When introduced into cells, this β -*O*-GlcNAc transferase FRET-sensor affords a means to monitor changes in β -*O*-GlcNAc transferase activity that may allow dynamic evaluation of this post-translational modification in living cells.²¹⁸

6.09.7.5 Affinity Probes for Glycan-Binding Proteins

Although genome sequencing and bioinformatics analysis has revealed genes encoding glycan-binding proteins, the functional characterization of glycan-binding proteins from complex mixtures is still a challenging task. Affinity chromatography with glycan-modified polymeric resins is one of the most robust methods for identifying glycan-binding proteins and was one of the first methods used to purify lectins.²¹⁹ As discussed previously, glycan-protein binding affinities are often weak and therefore may not necessarily be accessible by affinity chromatography. Several glycan-based photoaffinity probes have been synthesized to covalently trap lectins such as ricin,²²⁰ siglecs,²²¹ and galectins^{222,223} (Figure 22). Monosaccharides equipped with photoaffinity probes such as aryl azides²²¹ or diazirines²²⁴ can also be metabolically incorporated into glycoconjugates, which presents exciting opportunities for analyzing glycan-protein interactions in living cells. These preliminary studies are very promising and may indeed provide new reagents to characterize glycan-binding proteins.

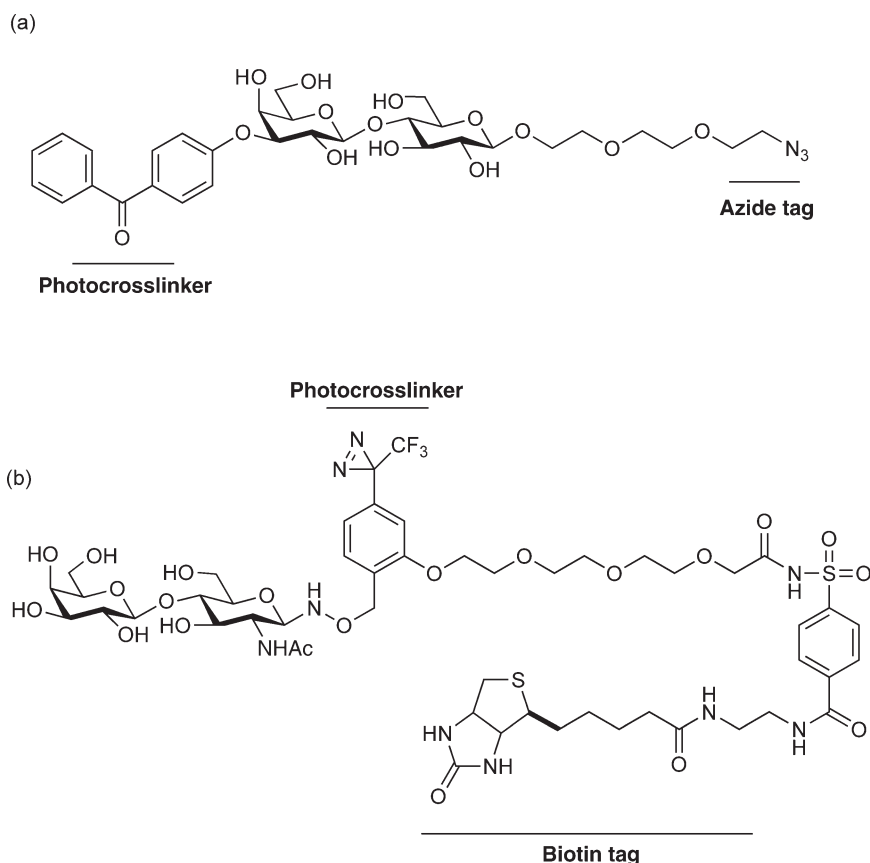


Figure 22 Photoaffinity probes for labeling glycan-binding proteins. (a) Lactose-based photoaffinity probe,²⁰⁹ (b) *N*-acetyllactosamine-based photoaffinity probe.²⁰⁸

6.09.8 Microarray Platforms for Glycomics

Deciphering the diverse array of specific glycan–protein interactions is a major challenge in biology. Understanding the molecular basis for this ‘glycan–protein code’ is critical for dissecting extracellular signaling events in multicellular organisms as well as interactions between pathogens and host cells. While some glycan–protein interactions have been elucidated, specific ligands for many lectins and antibodies have not been identified.²²⁵ The functional characterization of specific glycan–protein interactions will therefore require high-throughput technologies for surveying all the possible ligands for a particular lectin or antibody and vice versa.

6.09.8.1 Glycan Microarrays

Microarrays technologies have played an integral role in the high-throughput analysis of nucleic acids²²⁶ and proteins²²⁷ and are beginning to provide insight into glycan–protein interactions as well as the specificity of glycan-processing enzymes through the development of glycan microarrays^{228–230} (Figure 23(a)). Glycans are typically derivatized at their reducing end for immobilization on surfaces either by absorption or covalent chemistries.^{228–230} The methods that have been utilized to absorb glycans on surfaces include direct spotting on glass slides,²³¹ neoglycoconjugates on membranes,²³² biotinylated-glycans on avidin-coated plates,²³³ and fluoros-phase interactions.²³⁴ Alternatively, chemical strategies allow covalent immobilization of glycans through thiol,^{235,236} amine,^{237,238} and photolysis²³⁹ coupling methods. In addition, the 2,6-diaminopyridine derivatization of glycans is allowing multiplex analysis of glycan libraries by microarray, MS, and fluorescence imaging methods.^{240,241} Both naturally occurring and synthetic glycan libraries have been immobilized and used to interrogate the specificity of glycan-binding proteins, glycan-processing enzymes, and even whole cells.^{228–230} The analysis of these glycan arrays is usually performed with fluorescently labeled lectins or antibodies (Figure 23(a)). Notably, Stevens *et al.*^{242,243} were able to map the glycan specificity of hemagglutinin (HA) from different strains of influenza virus with glycan microarrays that revealed alterations in HA glycan-specificity that are associated with the highly pathogenic H5N1 influenza strain.

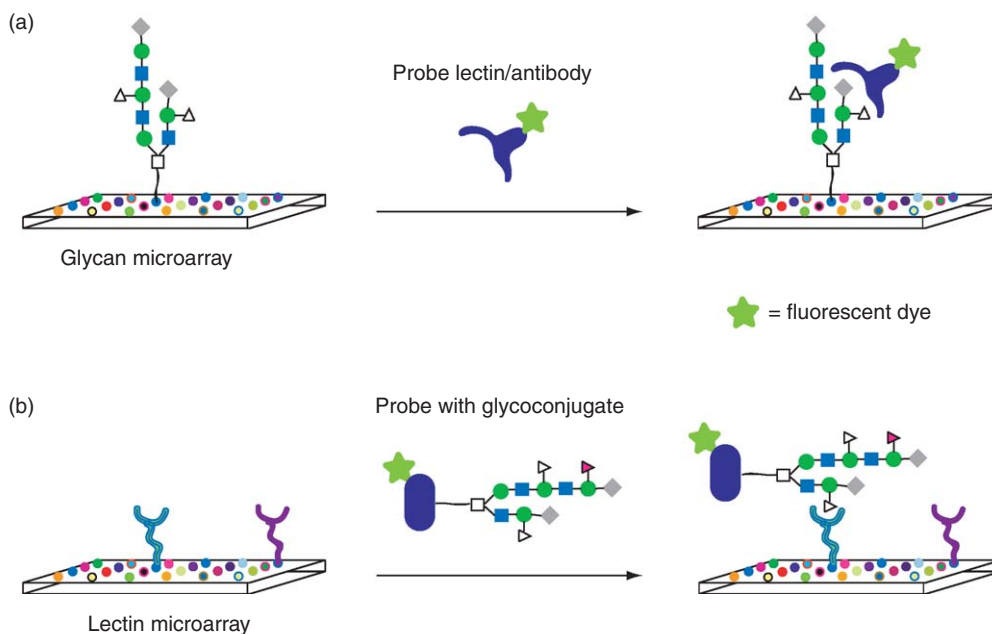


Figure 23 Microarray platforms for evaluating glycan–protein interactions. (a) Glycoconjugate microarrays;^{214–229} (b) lectin microarrays.^{230–234}

6.09.8.2 Protein Microarrays: Lectins/Antibodies

The high-throughput evaluation of glycan–protein interactions performs the reverse orientation with protein microarrays (Figure 23(b)). In this format, lectins or glycan-specific antibodies can be printed on glass slides by using protein microarray technologies and interrogated with fluorescently labeled glycoconjugates or cells (Figure 23(b)). Indeed, several lectin or glycan-antibody microarrays have been developed to explore the glycosylation pattern of glycoproteins and cells.^{244–247} These studies have demonstrated that lectin microarrays can yield specific recognition of glycoconjugates, which upon further development should afford a useful means to rapidly survey glycosylation patterns of purified glycoconjugates and whole cells.

The generation of glycan and lectin microarrays has begun to provide powerful tools for the high-throughput evaluation of glycan–protein interactions and enzyme specificities. These studies highlight the incredible advances in glycan microarray technologies that have now become integral components of functional glycomics core facilities ((<http://www.functionalglycomics.org>). Improvements in the isolation of glycoconjugates from natural sources and synthesis of glycoconjugate libraries will be critical for the construction of more comprehensive glycan microarrays. Likewise, more diverse lectin microarrays will require better methods for the functional expression of these proteins. These microarray platforms afford an unprecedented means for surveying all the possible interactions between glycans and proteins that will enable focused analysis of the glycan–protein interactions *in vivo*.

6.09.9 Conclusions and Future Perspectives

In summary, the rapid advances in chemistry provide more precise molecular probes to dissect protein glycosylation and glycobiology. The automation of glycoconjugate synthesis combined with improved methods for production of glycoproteins suggests that homogeneous materials will be more readily available for detailed biochemical and structural studies as well as large-scale analyses using microarray technologies. Chemical approaches also provide new methods to profile glycoconjugates and their associated proteins (lectins, glycosyltransferases, glycosidases, etc.) from complex mixtures that should facilitate characterization of glycan-associated processes in biology. Small molecules that perturb glycan-associated processes in living cells have historically been difficult to come by, but the emergence of chemical genetic methods and development of high-throughput screening methods yield new chemical tools for modulating glycan-associated processes *in vivo*. Collectively, these chemical approaches yield powerful tools to complement genetic methods for understanding glycobiology and identifying new pathways for therapeutic intervention.

Acknowledgments

H. C. H. acknowledges financial support from The Rockefeller University and The Irma T. Hirschl/Monique Weill-Caulier Trust.

Nomenclature

| | |
|-----------------------|---|
| glycan | polysaccharide or oligosaccharide |
| glycoconjugate | protein or lipid covalently modified with glycan(s) |
| semisynthesis | combination of chemical and enzyme-assisted synthesis |

References

1. K. Ohtsubo; J. D. Marth, *Cell* **2006**, *126* (5), 855–867.
2. K. Landsteiner, *Science* **1931**, *73* (1894), 403–409.
3. W. M. Watkins, *Science* **1966**, *152* (3719), 172–181.
4. A. Helenius; M. Aebi, *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049.

5. A. Varki, *Nature* **2007**, *446* (7139), 1023–1029.
6. L. E. Comstock; D. L. Kasper, *Cell* **2006**, *126* (5), 847–850.
7. H. H. Freeze, *Nat. Rev. Genet.* **2006**, *7* (7), 537–551.
8. G. Vogt; B. Vogt; N. Chuzhanova; K. Julenius; D. N. Cooper; J. L. Casanova, *Curr. Opin. Genet. Dev.* **2007**, *17* (3), 245–251.
9. R. G. Spiro, *Glycobiology* **2002**, *12* (4), 43R–56R.
10. E. Jung; A. L. Veuthey; E. Gasteiger; A. Bairoch, *Proteomics* **2001**, *1* (2), 262–268.
11. R. Kornfeld; S. Kornfeld, *Annu. Rev. Biochem.* **1985**, *54*, 631–664.
12. K. Romisch, *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 435–456.
13. P. Ghosh; N. M. Dahms; S. Kornfeld, *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (3), 202–212.
14. M. Demetriou; M. Granovsky; S. Quaggin; J. W. Dennis, *Nature* **2001**, *409* (6821), 733–739.
15. K. Ohtsubo; S. Takamatsu; M. T. Minowa; A. Yoshida; M. Takeuchi; J. D. Marth, *Cell* **2005**, *123* (7), 1307–1321.
16. D. A. Calarese; C. N. Scanlan; M. B. Zwick; S. Deechongkit; Y. Mimura; R. Kunert; P. Zhu; M. R. Wormald; R. L. Stanfield; K. H. Roux; J. W. Kelly; P. M. Rudd; R. A. Dwek; H. Katinger; D. R. Burton; I. A. Wilson, *Science* **2003**, *300* (5628), 2065–2071.
17. C. N. Scanlan; R. Pantophlet; M. R. Wormald; E. Ollmann Saphire; R. Stanfield; I. A. Wilson; H. Katinger; R. A. Dwek; P. M. Rudd; D. R. Burton, *J. Virol.* **2002**, *76* (14), 7306–7321.
18. E. O. Saphire; P. W. Parren; R. Pantophlet; M. B. Zwick; G. M. Morris; P. M. Rudd; R. A. Dwek; R. L. Stanfield; D. R. Burton; I. A. Wilson, *Science* **2001**, *293* (5532), 1155–1159.
19. I. Brockhausen, *EMBO Rep.* **2006**, *7* (6), 599–604.
20. H. C. Hang; C. R. Bertozzi, *Bioorg. Med. Chem.* **2005**, *13* (17), 5021–5034.
21. M. A. Tarp; H. Clausen, *Biochim. Biophys. Acta* **2008**, *1780* (3), 546–563.
22. K. G. Ten Hagen; T. A. Fritz; L. A. Tabak, *Glycobiology* **2003**, *13* (1), 1R–16R.
23. M. Tenno; K. Ohtsubo; F. K. Hagen; D. Ditto; A. Zarbock; P. Schaerli; U. H. von Andrian; K. Ley; D. Le; L. A. Tabak; J. D. Marth, *Mol. Cell Biol.* **2007**, *27* (24), 8783–8796.
24. Y. Zhang; H. Iwasaki; H. Wang; T. Kudo; T. B. Kalka; T. Hennes; T. Kubota; L. Cheng; N. Inaba; M. Gotoh; A. Togayachi; J. Guo; H. Hisatomi; K. Nakajima; S. Nishihara; M. Nakamura; J. D. Marth; H. Narimatsu, *J. Biol. Chem.* **2003**, *278* (1), 573–584.
25. E. Tian; K. G. Ten Hagen, *J. Biol. Chem.* **2007**, *282* (1), 606–614.
26. C. L. Hattstrup; S. J. Gendler, *Annu. Rev. Physiol.* **2008**, *70*, 431–457.
27. S. D. Rosen, *Annu. Rev. Immunol.* **2004**, *22*, 129–156.
28. D. P. Galonic; D. Y. Gin, *Nature* **2007**, *446* (7139), 1000–1007.
29. L. Lehle; S. Strahl; W. Tanner, *Angew. Chem. Int. Ed. Engl.* **2006**, *45* (41), 6802–6818.
30. R. Barresi; K. P. Campbell, *J. Cell Sci.* **2006**, *119* (Pt. 2), 199–207.
31. D. E. Michele; R. Barresi; M. Kanagawa; F. Saito; R. D. Cohn; J. S. Satz; J. Dollar; I. Nishino; R. I. Kelley; H. Somer; V. Straub; K. D. Mathews; S. A. Moore; K. P. Campbell, *Nature* **2002**, *418* (6896), 417–422.
32. D. E. Michele; K. P. Campbell, *J. Biol. Chem.* **2003**, *278* (18), 15457–15460.
33. R. J. Harris; M. W. Spellman, *Glycobiology* **1993**, *3* (3), 219–224.
34. R. S. Haltiwanger; P. Stanley, *Biochim. Biophys. Acta* **2002**, *1573* (3), 328–335.
35. D. J. Moloney; L. H. Shair; F. M. Lu; J. Xia; R. Locke; K. L. Matta; R. S. Haltiwanger, *J. Biol. Chem.* **2000**, *275* (13), 9604–9611.
36. S. Hase; S. Kawabata; H. Nishimura; H. Takeya; T. Sueyoshi; T. Miyata; S. Iwanaga; T. Takao; Y. Shimonishi; T. Ikenaka, *J. Biochem.* **1988**, *104* (6), 867–8.
37. S. Hase; H. Nishimura; S. Kawabata; S. Iwanaga; T. Ikenaka, *J. Biol. Chem.* **1990**, *265* (4), 1858–1861.
38. L. Shao; Y. Luo; D. J. Moloney; R. Haltiwanger, *Glycobiology* **2002**, *12* (11), 763–770.
39. W. T. Butler; L. W. Cunningham, *J. Biol. Chem.* **1965**, *240*, PC3449–PC3450.
40. W. T. Butler; L. W. Cunningham, *J. Biol. Chem.* **1966**, *241* (17), 3882–1388.
41. R. G. Spiro, *J. Biol. Chem.* **1967**, *242* (20), 4813–4823.
42. E. Michaelsson; V. Malmstrom; S. Reis; A. Engstrom; H. Burkhardt; R. Holmdahl, *J. Exp. Med.* **1994**, *180* (2), 745–749.
43. J. Backlund; S. Carlsen; T. Hoger; B. Holm; L. Fugger; J. Kihlberg; H. Burkhardt; R. Holmdahl, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (15), 9960–9965.
44. J. Backlund; A. Treschow; R. Bockermann; B. Holm; L. Holm; S. Issazadeh-Navikas; J. Kihlberg; R. Holmdahl, *Eur. J. Immunol.* **2002**, *32* (12), 3776–3784.
45. K. J. Colley; J. U. Baenziger, *J. Biol. Chem.* **1987**, *262* (21), 10290–10295.
46. Y. Ma; H. Shida; T. Kawasaki, *J. Biochem.* **1997**, *122* (4), 810–818.
47. C. R. Torres; G. W. Hart, *J. Biol. Chem.* **1984**, *259* (5), 3308–3317.
48. G. W. Hart; M. P. Housley; C. Slawson, *Nature* **2007**, *446* (7139), 1017–1022.
49. M. E. Forsythe; D. C. Love; B. D. Lazarus; E. J. Kim; W. A. Prinz; G. Ashwell; M. W. Krause; J. A. Hanover, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (32), 11952–11957.
50. J. A. Hanover; M. E. Forsythe; P. T. Hennessey; T. M. Brodigan; D. C. Love; G. Ashwell; M. W. Krause, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (32), 11266–11271.
51. C. Slawson; G. W. Hart, *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 631–636.
52. J. Krieg; W. Glasner; A. Vicentini; M. A. Doucey; A. Loffler; D. Hess; J. Hofsteenge, *J. Biol. Chem.* **1997**, *272* (42), 26687–26692.
53. A. Loffler; M. A. Doucey; A. M. Jansson; D. R. Muller; T. de Beer; D. Hess; M. Meldal; W. J. Richter; J. F. Vliegenthart; J. Hofsteenge, *Biochemistry* **1996**, *35* (37), 12005–12014.
54. J. Perez-Vilar; S. H. Randell; R. C. Boucher, *Glycobiology* **2004**, *14* (4), 325–337.
55. E. Muroi; S. Manabe; M. Ikezaki; Y. Urata; S. Sato; T. Kondo; Y. Ito; Y. Ihara, *Glycobiology* **2007**, *17* (9), 1015–1028.
56. D. Falzarano; O. Krokhin; G. Van Domselaar; K. Wolf; J. Seebach; H. J. Schnittler; H. Feldmann, *Virology* **2007**, *368* (1): 83–90.
57. J. Krieg; S. Hartmann; A. Vicentini; W. Glasner; D. Hess; J. Hofsteenge, *Mol. Biol. Cell* **1998**, *9* (2), 301–309.
58. A. Furmanek; J. Hofsteenge, *Acta Biochim. Pol.* **2000**, *47* (3), 781–789.
59. K. Julenius, *Glycobiology* **2007**, *17* (8), 868–876.
60. K. Sugahara; H. Kitagawa, *Curr. Opin. Struct. Biol.* **2000**, *10* (5), 518–527.
61. J. R. Bishop; M. Schuksz; J. D. Esko, *Nature* **2007**, *446* (7139), 1030–1037.

62. M. Inatani; F. Irie; A. S. Plump; M. Tessier-Lavigne; Y. Yamaguchi, *Science* **2003**, *302* (5647), 1044–1046.
63. N. B. Schwartz; M. Domowicz, *Glycobiology* **2002**, *12* (4), 57R–68R.
64. E. J. Bradbury; L. D. Moon; R. J. Papat; V. R. King; G. S. Bennett; P. N. Patel; J. W. Fawcett; S. B. McMahon, *Nature* **2002**, *416* (6881), 636–640.
65. L. D. Moon; R. A. Asher; K. E. Rhodes; J. W. Fawcett, *Nat. Neurosci.* **2001**, *4* (5), 465–466.
66. S. Mizuguchi; T. Uyama; H. Kitagawa; K. H. Nomura; K. Dejima; K. Gengyo-Ando; S. Mitani; K. Sugahara; K. Nomura, *Nature* **2003**, *423* (6938), 443–448.
67. H. Y. Hwang; S. K. Olson; J. D. Esko; H. R. Horvitz, *Nature* **2003**, *423* (6938), 439–443.
68. R. Novotny; A. Pfoesti; P. Messner; C. Schaffer, *Glycoconj. J.* **2004**, *20* (7–8), 435–447.
69. C. M. Szymanski; B. W. Wren, *Nat. Rev. Microbiol.* **2005**, *3* (3), 225–237.
70. M. Wacker; D. Linton; P. G. Hitchen; M. Nita-Lazar; S. M. Haslam; S. J. North; M. Panico; H. R. Morris; A. Dell; B. W. Wren; M. Aebi, *Science* **2002**, *298* (5599), 1790–1793.
71. M. F. Feldman; M. Wacker; M. Hernandez; P. G. Hitchen; C. L. Marolda; M. Kowarik; H. R. Morris; A. Dell; M. A. Valvano; M. Aebi, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (8), 3016–3021.
72. S. K. Arora; A. N. Neely; B. Blair; S. Lory; R. Ramphal, *Infect. Immun.* **2005**, *73* (7), 4395–4398.
73. A. Shen; H. D. Kamp; A. Grundling; D. E. Higgins, *Genes Dev.* **2006**, *20* (23), 3283–3295.
74. B. C. VanderVen; J. D. Harder; D. C. Crick; J. T. Belisle, *Science* **2005**, *309* (5736), 941–943.
75. J. A. Prescher; C. R. Bertozzi, *Cell* **2006**, *126* (5), 851–854.
76. C. R. Bertozzi; L. L. Kiessling, *Science* **2001**, *291* (5512), 2357–2364.
77. M. J. Grogan; M. R. Pratt; L. A. Marcaurette; C. R. Bertozzi, *Annu. Rev. Biochem.* **2002**, *71*, 593–634.
78. B. G. Davis, *Chem. Rev.* **2002**, *102* (2), 579–602.
79. A. Brik; S. Ficht; C. H. Wong, *Curr. Opin. Chem. Biol.* **2006**, *10* (6), 638–644.
80. K. C. Nicolaou; H. J. Mitchell, *Angew. Chem. Int. Ed. Engl.* **2001**, *40* (9), 1576–1624.
81. Z. G. Wang; X. Zhang; M. Visser; D. Live; A. Zatorski; U. Iserloh; K. O. Lloyd; S. J. Danishefsky, *Angew. Chem. Int. Ed. Engl.* **2001**, *40* (9), 1728–1732.
82. S. J. Danishefsky; J. R. Allen, *Angew. Chem. Int. Ed. Engl.* **2000**, *39* (5), 836–863.
83. C. S. Bennett; C. H. Wong, *Chem. Soc. Rev.* **2007**, *36* (8), 1227–1238.
84. A. Leppanen; L. Penttila; O. Renkonen; R. P. McEver; R. D. Cummings, *J. Biol. Chem.* **2002**, *277* (42), 39749–39759.
85. J. Zhang; X. Chen; J. Shao; Z. Liu; P. Kowal; Y. Lu; P. G. Wang, *Methods Enzymol.* **2003**, *362*, 106–124.
86. W. A. Barton; J. Lesniak; J. B. Biggins; P. D. Jeffrey; J. Jiang; K. R. Rajashankar; J. S. Thorson; D. B. Nikolov, *Nat. Struct. Biol.* **2001**, *8* (6), 545–551.
87. D. Hoffmeister; G. Drager; K. Ichinose; J. Rohr; A. Bechthold, *J. Am. Chem. Soc.* **2003**, *125* (16), 4678–4679.
88. J. S. Thorson; W. A. Barton; D. Hoffmeister; C. Albermann; D. B. Nikolov, *ChemBiochem* **2004**, *5* (1), 16–25.
89. M. H. Caruthers, *Science* **1985**, *230* (4723), 281–285.
90. R. B. Merrifield, *Science* **1965**, *150* (693), 178–185.
91. P. H. Seeberger; D. B. Werz, *Nat. Rev. Drug Discov.* **2005**, *4* (9), 751–763.
92. O. J. Plante; P. H. Seeberger, *Curr. Opin. Drug Discov. Devel.* **2003**, *6* (4), 521–525.
93. M. C. Hewitt; D. A. Snyder; P. H. Seeberger, *J Am Chem Soc* **2002**, *124* (45), 13434–13436.
94. L. Schofield; M. C. Hewitt; K. Evans; M. A. Siomos; P. H. Seeberger, *Nature* **2002**, *418* (6899), 785–789.
95. D. R. Mootoo; P. Konradsson; J. E. Udodong; B. Fraser-Reid, *J. Am. Chem. Soc.* **1998**, *110*, 5583–5585.
96. P. Sears; C. H. Wong, *Science* **2001**, *291* (5512), 2344–2350.
97. F. Burkhardt; Z. Zhang; S. Wacowich-Sgarbi; C. H. Wong, *Angew. Chem. Int. Ed. Engl.* **2001**, *40* (7), 1274–1277.
98. T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
99. V. Y. Dudkin; M. Orlova; X. Geng; M. Mandal; W. C. Olson; S. J. Danishefsky, *J. Am. Chem. Soc.* **2004**, *126* (31), 9560–9562.
100. C. P. Hackenberger; C. T. Friel; S. E. Radford; B. Imperiali, *J. Am. Chem. Soc.* **2005**, *127* (37), 12882–12889.
101. N. Yamamoto; Y. Tanabe; R. Okamoto; P. E. Dawson; Y. Kajihara, *J. Am. Chem. Soc.* **2008**, *130* (2), 501–510.
102. D. Macmillan; C. R. Bertozzi, *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (11), 1355–1359.
103. Y. Y. Yang; S. Ficht; A. Brik; C. H. Wong, *J. Am. Chem. Soc.* **2007**, *129* (24), 7690–7701.
104. A. Brik; Y. Y. Yang; S. Ficht; C. H. Wong, *J. Am. Chem. Soc.* **2006**, *128* (17), 5626–5627.
105. R. J. Payne; S. Ficht; S. Tang; A. Brik; Y. Y. Yang; D. A. Case; C. H. Wong, *J. Am. Chem. Soc.* **2007**, *129* (44), 13527–13536.
106. S. Ficht; R. J. Payne; A. Brik; C. H. Wong, *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (31), 5975–5979.
107. C. Chatterjee; R. K. McGinty; J. P. Pellois; T. W. Muir, *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (16), 2814–2818.
108. D. W. Low; M. G. Hill; M. R. Carrasco; S. B. Kent; P. Botti, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (12), 6554–6559.
109. J. Offer; C. N. Boddy; P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124* (17), 4642–4646.
110. Q. Wan; S. J. Danishefsky, *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (48), 9248–9252.
111. G. Chen; Q. Wan; Z. Tan; C. Kan; Z. Hua; K. Ranganathan; S. J. Danishefsky, *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (39), 7383–7387.
112. Y. Chiba; Y. Jigami, *Curr. Opin. Chem. Biol.* **2007**, *11* (6), 670–676.
113. S. R. Hamilton; R. C. Davidson; N. Sethuraman; J. H. Nett; Y. Jiang; S. Rios; P. Bobrowicz; T. A. Stadheim; H. Li; B. K. Choi; D. Hopkins; H. Wischniewski; J. Roser; T. Mitchell; R. R. Strawbridge; J. Hoopes; S. Wildt; T. U. Gerngross, *Science* **2006**, *313* (5792), 1441–1443.
114. L. Wang; J. Xie; P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
115. Z. Zhang; J. Gildersleeve; Y. Y. Yang; R. Xu; J. A. Loo; S. Uryu; C. H. Wong; P. G. Schultz, *Science* **2004**, *303* (5656), 371–373.
116. R. Xu; S. R. Hanson; Z. Zhang; Y. Y. Yang; P. G. Schultz; C. H. Wong, *J. Am. Chem. Soc.* **2004**, *126* (48), 15654–15655.
117. D. P. Galonic; N. D. Ide; W. A. van der Donk; D. Y. Gin, *J. Am. Chem. Soc.* **2005**, *127* (20), 7359–7369.
118. H. Liu; L. Wang; A. Brock; C. H. Wong; P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125* (7), 1702–1703.
119. S. I. van Kasteren; H. B. Kramer; H. H. Jensen; S. J. Campbell; J. Kirkpatrick; N. J. Oldham; D. C. Anthony; B. G. Davis, *Nature* **2007**, *446* (7139), 1105–1109.
120. M. A. Shogren-Knaak; P. J. Alaimo; K. M. Shokat, *Annu. Rev. Cell Dev. Biol.* **2001**, *17*, 405–433.

121. J. B. Lowe; J. D. Marth, *Annu. Rev. Biochem.* **2003**, *72*, 643–691.
122. M. Lewandoski, *Nat. Rev. Genet.* **2001**, *2* (10), 743–755.
123. G. J. Hannon, *Nature* **2002**, *418* (6894), 244–251.
124. B. R. Stockwell, *Nature* **2004**, *432* (7019), 846–854.
125. W. I. Weis; K. Drickamer, *Annu. Rev. Biochem.* **1996**, *65*, 441–473.
126. N. Kaila; B. E. Thomas, IV, *Med. Res. Rev.* **2002**, *22* (6), 566–601.
127. E. E. Simanek; G. J. McGarvey; J. A. Jablonowski; C. H. Wong, *Chem. Rev.* **1998**, *98* (2), 833–862.
128. R. Banteli; B. Ernst, *Bioorg. Med. Chem. Lett.* **2001**, *11* (4), 459–462.
129. T. P. Kogan; B. Dupre; H. Bui; K. L. McAbee; J. M. Kassir; I. L. Scott; X. Hu; P. Vanderslice; P. J. Beck; R. A. Dixon, *J. Med. Chem.* **1998**, *41* (7), 1099–1111.
130. J. E. Gestwicki; L. E. Strong; L. L. Kiessling, *Chem. Biol.* **2000**, *7* (8), 583–591.
131. G. Thoma; R. O. Duthaler; J. L. Magnani; J. T. Patton, *J. Am. Chem. Soc.* **2001**, *123* (41), 10113–10114.
132. W. J. Sanders; T. R. Katsumoto; C. R. Bertozzi; S. D. Rosen; L. L. Kiessling, *Biochemistry* **1996**, *35* (47), 14862–14867.
133. E. J. Gordon; W. J. Sanders; L. L. Kiessling, *Nature* **1998**, *392* (6671), 30–31.
134. P. I. Kitov; J. M. Sadowska; G. Mulvey; G. D. Armstrong; H. Ling; N. S. Pannu; R. J. Read; D. R. Bundle, *Nature* **2000**, *403* (6770), 669–672.
135. Z. Zhang; E. A. Merritt; M. Ahn; C. Roach; Z. Hou; C. L. Verlinde; W. G. Hol; E. Fan, *J. Am. Chem. Soc.* **2002**, *124* (44), 12991–12998.
136. E. A. Merritt; Z. Zhang; J. C. Pickens; M. Ahn; W. G. Hol; E. Fan, *J. Am. Chem. Soc.* **2002**, *124* (30), 8818–8824.
137. J. E. Gestwicki; L. L. Kiessling, *Nature* **2002**, *415* (6867), 81–84.
138. M. J. Borrok; L. L. Kiessling, *J. Am. Chem. Soc.* **2007**, *129* (42), 12780–12785.
139. A. D. Elbein, *Annu. Rev. Biochem.* **1987**, *56*, 497–534.
140. R. A. Dwek; T. D. Butters; F. M. Platt; N. Zitzmann, *Nat. Rev. Drug Discov.* **2002**, *1* (1), 65–75.
141. R. J. Desnick; M. M. Kaback, *Adv. Genet.* **2001**, *44*, 349–356.
142. J. Q. Fan; S. Ishii; N. Asano; Y. Suzuki, *Nat. Med.* **1999**, *5* (1), 112–115.
143. J. Matsuda; O. Suzuki; A. Oshima; Y. Yamamoto; A. Noguchi; K. Takimoto; M. Itoh; Y. Matsuzaki; Y. Yasuda; S. Ogawa; Y. Sakata; E. Nanba; K. Higaki; Y. Ogawa; L. Tominaga; K. Ohno; H. Iwasaki; H. Watanabe; R. O. Brady; Y. Suzuki, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (26), 15912–15917.
144. A. R. Sawkar; W. C. Cheng; E. Beutler; C. H. Wong; W. E. Balch; J. W. Kelly, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (24), 15428–15433.
145. K. McClellan; C. M. Perry, *Drugs* **2001**, *61* (2), 263–283.
146. P. Sears; C. H. Wong, *Cell Mol. Life Sci.* **1998**, *54* (3), 223–252.
147. S. F. Kuan; J. C. Byrd; C. Basbaum; Y. S. Kim, *J. Biol. Chem.* **1989**, *264* (32), 19271–19277.
148. N. B. Schwartz, *J. Biol. Chem.* **1977**, *252* (18), 6316–6321.
149. M. M. Fuster; J. R. Brown; L. Wang; J. D. Esko, *Cancer Res.* **2003**, *63* (11), 2775–2781.
150. J. R. Brown; M. M. Fuster; T. Whisenant; J. D. Esko, *J. Biol. Chem.* **2003**, *278* (26), 23352–23359.
151. O. T. Keppler; R. Horstkorte; M. Pawlita; C. Schmidt; W. Reutter, *Glycobiology* **2001**, *11* (2), 11R–18R.
152. L. K. Mahal; N. W. Charter; K. Angata; M. Fukuda; D. E. Koshland, Jr.; C. R. Bertozzi, *Science* **2001**, *294* (5541), 380–381.
153. S. Laferte; N. W. Chan; K. Sujino; T. L. Lowary; M. M. Palcic, *Eur. J. Biochem.* **2000**, *267* (15), 4840–4849.
154. P. D. Eason; B. Imperiali, *Biochemistry* **1999**, *38* (17), 5430–5437.
155. J. S. Helm; Y. Hu; L. Chen; B. Gross; S. Walker, *J. Am. Chem. Soc.* **2003**, *125* (37), 11168–11169.
156. B. J. Gross; B. C. Kraybill; S. Walker, *J. Am. Chem. Soc.* **2005**, *127* (42), 14588–14589.
157. B. J. Gross; J. G. Swoboda; S. Walker, *J. Am. Chem. Soc.* **2008**, *130* (2), 440–441.
158. K. A. Winans; C. R. Bertozzi, *Chem. Biol.* **2002**, *9* (1), 113–129.
159. H. C. Hang; C. Yu; K. G. Ten Hagen; E. Tian; K. A. Winans; L. A. Tabak; C. R. Bertozzi, *Chem. Biol.* **2004**, *11* (3), 337–345.
160. U. M. Unligil; J. M. Rini, *Curr. Opin. Struct. Biol.* **2000**, *10* (5), 510–517.
161. P. A. Clemons, *Curr. Opin. Chem. Biol.* **1999**, *3* (1), 112–115.
162. J. J. Kohler; C. R. Bertozzi, *Chem. Biol.* **2003**, *10* (12), 1303–1311.
163. A. Dell; H. R. Morris, *Science* **2001**, *291* (5512), 2351–2356.
164. F. Kjeldsen; K. F. Haselmann; B. A. Budnik; E. S. Sorensen; R. A. Zubarev, *Anal. Chem.* **2003**, *75* (10), 2355–2361.
165. J. Cox; M. Mann, *Cell* **2007**, *130* (3), 395–398.
166. B. Kuster; T. N. Krogh; E. Mortz; D. J. Harvey, *Proteomics* **2001**, *1* (2), 350–361.
167. H. Zhang; X. J. Li; D. B. Martin; R. Aebersold, *Nat. Biotechnol.* **2003**, *21* (6), 660–666.
168. L. Wells; K. Vosseller; R. N. Cole; J. M. Cronshaw; M. J. Matunis; G. W. Hart, *Mol. Cell Proteomics* **2002**, *1* (10), 791–804.
169. H. Kaji; H. Saito; Y. Yamauchi; T. Shinkawa; M. Taoka; J. Hirabayashi; K. Kasai; N. Takahashi; T. Isobe, *Nat. Biotechnol.* **2003**, *21* (6), 667–672.
170. T. Bulter; T. Schumacher; D. J. Namdjou; R. Gutierrez Gallego; H. Clausen; L. Elling, *ChemBiochem* **2001**, *2* (12), 884–894.
171. 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* (52), 16162–16163.
172. N. Khidekel; S. B. Ficarro; P. M. Clark; M. C. Bryn; D. L. Swaney; J. E. Rexach; Y. E. Sun; J. J. Coon; E. C. Peters; L. C. Hsieh-Wilson, *Nat. Chem. Biol.* **2007**, *3* (6), 339–348.
173. D. H. Dube; C. R. Bertozzi, *Curr. Opin. Chem. Biol.* **2003**, *7* (5), 616–625.
174. L. K. Mahal; K. J. Yarema; C. R. Bertozzi, *Science* **1997**, *276* (5315), 1125–1128.
175. H. C. Hang; C. R. Bertozzi, *J. Am. Chem. Soc.* **2001**, *123* (6), 1242–1243.
176. E. Saxon; C. R. Bertozzi, *Science* **2000**, *287* (5460), 2007–2010.
177. S. J. Luchansky; S. Argade; B. K. Hayes; C. R. Bertozzi, *Biochemistry* **2004**, *43* (38), 12358–12366.
178. D. J. Vocadlo; H. C. Hang; E. J. Kim; J. A. Hanover; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (16), 9116–9121.
179. R. Sprung; A. Nandi; Y. Chen; S. C. Kim; D. Barma; J. R. Falck; Y. Zhao, *J. Proteome Res.* **2005**, *4* (3), 950–957.
180. A. Nandi; R. Sprung; D. K. Barma; Y. Zhao; S. C. Kim; J. R. Falck; Y. Zhao, *Anal. Chem.* **2006**, *78* (2), 452–458.

181. H. C. Hang; C. Yu; D. L. Kato; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (25), 14846–14851.
182. J. A. Prescher; C. R. Bertozzi, *Nat. Chem. Biol.* **2005**, *1* (1), 13–21.
183. S. T. Laughlin; C. R. Bertozzi, *Nat. Protoc.* **2007**, *2* (11), 2930–2944.
184. J. A. Prescher; D. H. Dube; C. R. Bertozzi, *Nature* **2004**, *430* (7002), 873–877.
185. D. H. Dube; J. A. Prescher; C. N. Quang; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (13), 4819–4824.
186. S. T. Laughlin; J. M. Baskin; S. L. Amacher; C. R. Bertozzi, *Science* **2008**, *320* (5876), 664–667.
187. Q. Wang; T. R. Chan; R. Hilgraf; V. V. Fokin; K. B. Sharpless; M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125* (11), 3192–3193.
188. A. E. Speers; G. C. Adam; B. F. Cravatt, *J. Am. Chem. Soc.* **2003**, *125* (16), 4686–4687.
189. N. J. Agard; J. M. Baskin; J. A. Prescher; A. Lo; C. R. Bertozzi, *ACS Chem. Biol.* **2006**, *1* (10), 644–648.
190. N. J. Agard; J. A. Prescher; C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126* (46), 15046–15047.
191. J. M. Baskin; J. A. Prescher; S. T. Laughlin; N. J. Agard; P. V. Chang; I. A. Miller; A. Lo; J. A. Codelli; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (43), 16793–16797.
192. J. A. Codelli; J. M. Baskin; N. J. Agard; C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130* (34), 11486–11493.
193. E. M. Sletten; C. R. Bertozzi, *Org. Lett.* **2008**, *10* (14), 3097–3099.
194. X. Ning; J. Guo; M. A. Wolfert; G. J. Boons, *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (12), 2253–2255.
195. D. Rabuka; S. C. Hubbard; S. T. Laughlin; S. P. Argade; C. R. Bertozzi, *J. Am. Chem. Soc.* **2006**, *128* (37), 12078–12079.
196. T. L. Hsu; S. R. Hanson; K. Kishikawa; S. K. Wang; M. Sawa; C. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (8), 2614–2619.
197. M. Sawa; T. L. Hsu; T. Itoh; M. Sugiyama; S. R. Hanson; P. K. Vogt; C. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (33), 12371–12376.
198. S. R. Hanson; T. L. Hsu; E. Weerapana; K. Kishikawa; G. M. Simon; B. F. Cravatt; C. H. Wong, *J. Am. Chem. Soc.* **2007**, *129* (23), 7266–7267.
199. S. G. Sampathkumar; A. V. Li; M. B. Jones; Z. Sun; K. J. Yarema, *Nat. Chem. Biol.* **2006**, *2* (3), 149–152.
200. B. Schilling; S. Goon; N. M. Samuels; S. P. Gaucher; J. A. Leary; C. R. Bertozzi; B. W. Gibson, *Biochemistry* **2001**, *40* (42), 12666–12677.
201. S. Goon; B. Schilling; M. V. Tullius; B. W. Gibson; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (6), 3089–3094.
202. R. Sadamoto; K. Niikura; P. S. Sears; H. Liu; C. H. Wong; A. Suksomcheep; F. Tomita; K. Monde; S. Nishimura, *J. Am. Chem. Soc.* **2002**, *124* (31), 9018–9019.
203. H. C. Hang; C. Yu; M. R. Pratt; C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126* (1), 6–7.
204. M. R. Pratt; H. C. Hang; K. G. Ten Hagen; J. Rarick; T. A. Gerken; L. A. Tabak; C. R. Bertozzi, *Chem. Biol.* **2004**, *11* (7), 1009–1016.
205. T. M. Leavy; C. R. Bertozzi, *Bioorg. Med. Chem. Lett.* **2007**, *17* (14), 3851–3854.
206. P. M. Clark; J. F. Dweck; D. E. Mason; C. R. Hart; S. B. Buck; E. C. Peters; B. J. Agnew; L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2008**, *130* (35), 11576–11577.
207. M. J. Evans; B. F. Cravatt, *Chem. Rev.* **2006**, *106* (8), 3279–3301.
208. C. S. Rye; S. G. Withers, *Curr. Opin. Chem. Biol.* **2000**, *4* (5), 573–580.
209. D. J. Vocadlo; G. Yu; J. Davies; R. Laine; S. G. Withers, *Nature* **2001**, *412* (6849), 835–838.
210. K. A. Stubbs; A. Scaffidi; A. W. Debowski; B. L. Mark; R. V. Stick; D. J. Vocadlo, *J. Am. Chem. Soc.* **2008**, *130* (1), 327–335.
211. D. J. Vocadlo; C. R. Bertozzi, *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (40), 5338–5342.
212. S. J. Williams; O. Hekmat; S. G. Withers, *Chembiochem* **2006**, *7* (1), 116–124.
213. O. Hekmat; C. Florizone; Y. W. Kim; L. D. Eltis; R. A. Warren; S. G. Withers, *Chembiochem* **2007**, *8* (17), 2125–2132.
214. J. Zhang; R. E. Campbell; A. Y. Ting; R. Y. Tsien, *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (12), 906–918.
215. C. W. Lin; A. Y. Ting, *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (22), 2940–2943.
216. C. W. Lin; C. Y. Jao; A. Y. Ting, *J. Am. Chem. Soc.* **2004**, *126* (19), 5982–5983.
217. L. D. Carrillo; L. Krishnamoorthy; L. K. Mahal, *J. Am. Chem. Soc.* **2006**, *128* (46), 14768–14769.
218. H. Lis; N. Sharon, *Annu. Rev. Biochem.* **1973**, *42*, 541–574.
219. J. J. Park; Y. Sadakane; K. Masuda; T. Tomohiro; T. Nakano; Y. Hatanaka, *Chembiochem* **2005**, *6* (5), 814–818.
220. S. Han; B. E. Collins; P. Bengtson; J. C. Paulson, *Nat. Chem. Biol.* **2005**, *1* (2), 93–97.
221. L. Ballell; M. van Scherpenzeel; K. Buchalova; R. M. Liskamp; R. J. Pieters, *Org. Biomol. Chem.* **2006**, *4* (23), 4387–4394.
222. L. Ballell; K. J. Alink; M. Slijper; C. Versluis; R. M. Liskamp; R. J. Pieters, *Chembiochem* **2005**, *6* (2), 291–295.
223. Y. Tanaka; J. J. Kohler, *J. Am. Chem. Soc.* **2008**, *130* (11), 3278–3279.
224. P. R. Crocker; J. C. Paulson; A. Varki, *Nat. Rev. Immunol.* **2007**, *7* (4), 255–266.
225. D. J. Lockhart; E. A. Winzler, *Nature* **2000**, *405* (6788), 827–836.
226. H. Zhu; M. Bilgin; R. Bangham; D. Hall; A. Casamayor; P. Bertone; N. Lan; R. Jansen; S. Bidlingmaier; T. Houfek; T. Mitchell; P. Miller; R. A. Dean; M. Gerstein; M. Snyder, *Science* **2001**, *293* (5537), 2101–2105.
227. J. C. Paulson; O. Blixt; B. E. Collins, *Nat. Chem. Biol.* **2006**, *2* (5), 238–248.
228. T. Feizi; W. Chai, *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (7), 582–588.
229. M. D. Disney; P. H. Seeberger, *Chem. Biol.* **2004**, *11* (12), 1701–1707.
230. D. Wang; S. Liu; B. J. Trummer; C. Deng; A. Wang, *Nat. Biotechnol.* **2002**, *20* (3), 275–281.
231. S. Fukui; T. Feizi; C. Galustian; A. M. Lawson; W. Chai, *Nat. Biotechnol.* **2002**, *20* (10), 1011–1017.
232. O. E. Galanina; M. Mecklenburg; N. E. Nifantiev; G. V. Pazyrnina; N. V. Bovin, *Lab Chip* **2003**, *3* (4), 260–265.
233. K. S. Ko; F. A. Jaipuri; N. L. Pohl, *J. Am. Chem. Soc.* **2005**, *127* (38), 13162–13163.
234. S. Park; M. R. Lee; S. J. Pyo; I. Shin, *J. Am. Chem. Soc.* **2004**, *126* (15), 4812–4819.
235. D. M. Ratner; E. W. Adams; M. D. Disney; P. H. Seeberger, *Chembiochem* **2004**, *5* (10), 1375–1383.
236. M. Schwarz; L. Spector; A. Gargir; A. Shtevi; M. Gortler; R. T. Altstock; A. A. Dukler; N. Dotan, *Glycobiology* **2003**, *13* (11), 749–754.
237. O. Blixt; S. Head; T. Mondala; C. Scanlan; M. E. Huflejt; R. Alvarez; M. C. Bryan; F. Fazio; D. Calarese; J. Stevens; N. Razi; D. J. Stevens; J. J. Skehel; I. van Die; D. R. Burton; I. A. Wilson; R. Cummings; N. Bovin; C. H. Wong; J. C. Paulson, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (49), 17033–17038.

238. S. Angeloni; J. L. Ridet; N. Kusy; H. Gao; F. Crevoisier; S. Guinchard; S. Kochhar; H. Sigrist; N. Sprenger, *Glycobiology* **2005**, *15* (1), 31–41.
239. A. R. de Boer; C. H. Hokke; A. M. Deelder; M. Wuhrer, *Anal. Chem.* **2007**, *79* (21), 8107–8113.
240. B. Xia; Z. S. Kowar; T. Ju; R. A. Alvarez; G. P. Sachdev; R. D. Cummings, *Nat. Methods* **2005**, *2* (11), 845–850.
241. J. Stevens; O. Blixt; T. M. Tumpey; J. K. Taubenberger; J. C. Paulson; I. A. Wilson, *Science* **2006**, *312* (5772), 404–410.
242. J. Stevens; O. Blixt; J. C. Paulson; I. A. Wilson, *Nat. Rev. Microbiol.* **2006**, *4* (11), 857–864.
243. K. L. Hsu; K. T. Pilobello; L. K. Mahal, *Nat. Chem. Biol.* **2006**, *2* (3), 153–157.
244. K. T. Pilobello; L. Krishnamoorthy; D. Slawek; L. K. Mahal, *ChemBiochem* **2005**, *6* (6), 985–989.
245. K. T. Pilobello; D. E. Slawek; L. K. Mahal, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (28), 11534–11539.
246. S. Chen; T. Zheng; M. R. Shortreed; C. Alexander; L. M. Smith, *Anal. Chem.* **2007**, *79* (15), 5698–5702.
247. T. Zheng; D. Peelen; L. M. Smith, *J. Am. Chem. Soc.* **2005**, *127* (28), 9982–9983.

Biographical Sketch



Howard Hang received a Bachelor of Science in Chemistry with honors at the University of California, Santa Cruz in 1998. At Santa Cruz, he worked on the total synthesis of natural products under the guidance of Prof. Joseph Konopelski. He then joined the laboratory of Prof. Carolyn Bertozzi in the Department of Chemistry at the University of California, Berkeley. His Ph.D. dissertation described the development of chemical tools to study protein glycosylation in mammalian cells. After receiving his Doctorate of Philosophy in Chemistry from the University of California, Berkeley in 2003, he performed postdoctoral studies in chemical biology and immunology with Prof. Hidde Ploegh at Harvard Medical School and the Whitehead Institute of Biomedical Research. In 2007, he joined the faculty at The Rockefeller University as an Assistant Professor and Head of Laboratory. The Hang laboratory focuses on the development of chemical tools to study microbial pathogenesis.

6.10 O Antigen Biosynthesis

Miguel A. Valvano, The University of Western Ontario, London, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|---|-----|
| 6.10.1 | Introduction | 297 |
| 6.10.2 | Biogenesis of O Antigen | 298 |
| 6.10.2.1 | Initiation Reaction | 298 |
| 6.10.2.1.1 | Polyisoprenyl-phosphate <i>N</i> -acetylhexosamine-1-phosphate transferases | 299 |
| 6.10.2.1.2 | Polyisoprenyl-phosphate hexose-1-phosphate transferases | 301 |
| 6.10.2.2 | Elongation/Translocation/Polymerization | 302 |
| 6.10.2.2.1 | Wzy/Wzx-dependent pathway | 304 |
| 6.10.2.2.2 | ABC transporter-dependent pathway | 308 |
| 6.10.2.2.3 | Synthase-dependent pathway | 308 |
| 6.10.2.3 | Ligation Reaction | 309 |
| 6.10.3 | Future Prospects | 310 |
| References | | 310 |

6.10.1 Introduction

Lipopolysaccharide (LPS) is a surface molecule unique to Gram-negative bacteria that plays a key role as an elicitor of innate immune responses, ranging from localized inflammation to disseminated sepsis.¹ LPS is a major constituent of the outer leaflet of the Gram-negative bacterial outer membrane,^{2,3} and consists of lipid A-core oligosaccharide (OS), and O-specific polysaccharide or O antigen.^{3,4} Lipid A, the membrane-embedded portion of LPS, forms the majority of the outer lipid leaflet of the outer membrane. Lipid A is made of a β -1,6-linked glucosamine disaccharide, which becomes phosphorylated and acylated with a variable number of fatty and hydroxyfatty acid chains.⁵ All the steps in lipid A synthesis have been well characterized (for a recent review see Raetz⁶).

The core OS, made of hexoses, *glycero-manno*-heptose, and *keto*-deoxy-octulosonic acid,⁷ is assembled on preformed lipid A by the sequential transfer of sugar components. The core OS can be subdivided into inner and outer core domains. The outer core usually consists of hexoses and hexosamines while the inner core is, depending on the particular species, composed of one to three residues of 3-deoxy-D-*manno*-octulosonic acid, and two or three residues of L-*glycero*-D-*manno*-heptose.^{7,8}

LPS plays an important role in maintaining the structural integrity of the bacterial outer membrane by interacting with outer membrane proteins as well as divalent cations (for a review see Nikaido⁹). Also, LPS provides an effective diffusion barrier. This is due in part to the low fluidity state of the hydrocarbon regions of the LPS molecules and to strong lateral interactions between LPS molecules.⁹ In addition, phosphate groups covalently attached to heptose residues in the inner core participate in ionic interactions with divalent cations, especially Mg²⁺, which contribute to create a hydrophilic surface 'lattice' that when combined to the highly ordered state of the hydrophobic interior, provide a barrier preventing the passage of hydrophobic substances such as detergents, dyes, and antibiotics across the outer membrane.⁹⁻¹¹

O antigens are polymers of OS repeating units. The chemical composition, structure, and antigenicity of O antigens vary widely among Gram-negative bacteria, giving rise to a large number of O-serotypes.^{12,13} The O antigen, which is the most surface-exposed LPS moiety, also contributes to pathogenicity by protecting infecting bacteria from bactericidal host responses like killing by serum complement and phagocytosis.^{4,14-16}

The biogenesis of LPS is a complex process involving various steps that occur at the plasma membrane followed by the translocation of LPS molecules to the bacterial cell surface. LPS biosynthesis employs a large number of enzymes.^{6,7,17,18} The core OS is assembled on preformed lipid A by sequential glycosyl transfer of monosaccharides, while the O antigen is independently assembled on undecaprenyl-phosphate (Und-P), a

polyisoprenoid lipid to which O antigen is linked via a phosphodiester bond.⁴ These pathways eventually converge by the ligation of the O antigen onto outer core domain of the lipid A-core OS acceptor, with the concomitant release of Und-PP.^{4,5,7,18–21} Und-PP is recycled into Und-P by a poorly characterized pathway, also conserved in all cells,^{22–28} which involves the hydrolysis of the terminal phosphate at the periplasmic side of the membrane. The C₅₅ Und-P is also essential for the biosynthesis of peptidoglycan and enterobacterial common antigen (ECA), a surface glycolipid similar to the O antigen but not commonly attached to lipid A-core OS.²⁹ In eukaryotes, the polyisoprenoid lipid carrier is a C₉₅ dolichyl-P.²²

The genes governing the synthesis of lipid A and substrates for the assembly of the inner core components (*glycero-manno*-heptose and *keto-deoxy*-octulosonic acid) are scattered throughout the chromosome.^{4,5,20,30–33} For example, genes encoding enzymes involved in the ADP-*glycero-manno*-heptose synthesis pathway such as *gmbA*³⁰ and *bltE*³² are located far apart from each other on the chromosome. In contrast, the genes encoding functions for the assembly of the outer core OS and the O-specific polysaccharide are clustered.^{4,20,34} Mutations in most core biosynthesis genes (*waa*, formerly *rfa*³⁵) lead to rough mutants with an incomplete core, which lacks the site for the attachment of *O polysaccharides*. *Escherichia coli* mutants lacking heptose in the LPS display a more dramatic phenotype known as ‘deep rough’. This phenotype is characterized by hypersensitivity to novobiocin, detergents, and bile salts,³⁶ as well as defects in F plasmid conjugation and generalized transduction by the bacteriophage P1.^{37–39} All of these defects result from reduced amounts of outer membrane proteins, some of which serve as surface receptors for conjugation and bacteriophage attachment^{39–44} or as channel components of efflux systems.^{45,46} The impaired stability of the outer membrane in deep rough mutants is associated at least in part with the absence of phosphate groups, since mutations in genes encoding LPS core kinases also show pleiotropic phenotypes similar to those found in heptose-deficient mutants.^{47,48}

Mutations in any of the *wb** (formerly *rfb*³⁵) genes, which are involved in the synthesis of the *O polysaccharide*, result in rough mutants that have a complete core structure.⁴ In some cases, *wb** genes may be entirely plasmid-encoded,^{49,50} while in other cases plasmid-mediated functions, in addition to chromosomal genes, are required for the biosynthesis of the *O polysaccharide*.^{51,52} Minimally, *wb** gene clusters encode nucleotide sugar synthases (for biosynthesis of the nucleotide sugar precursors specific to O antigens), and glycosyltransferases (for the sequential and specific addition of sugars that make the O repeating unit). Additional genes encoding functions involved in the assembly of the *O polysaccharide* are also present in these clusters, such as *wzy* (O antigen polymerase) and *wzx* (putative O antigen flippase; see below) in some systems, and *wzm* (membrane component of ATP-binding cassette (ABC) transporter) and *wzt* (ATP-binding component of ABC transporter) in others.³⁴ In systems containing *wzx* and *wzy*, the average size distribution of the *O polysaccharide* chain is modulated by the product of *wzz*,^{53–55} a gene usually located outside but in the proximity of *wb** clusters. An O-antigen ligase activity encoded by a gene located within the core OS cluster, *waaL*,²⁰ is required for the transfer of the *O polysaccharide* onto lipid A-core OS.

This chapter will review the current mechanisms strictly operating in the biogenesis of the O-specific LPS. More comprehensive information on the biosynthesis and assembly of other LPS components and capsular polysaccharides can be found in recent reviews.^{6,17,56,57}

6.10.2 Biogenesis of O Antigen

The biogenesis of O antigens can be mechanistically conceived into four interconnected stages (**Figure 1**): (1) initiation reactions, (2) processing of the O antigen, which involves the elongation, polymerization, and membrane translocation of O repeating subunits, (3) the ligation reaction to the lipid A-core OS, and (4) the recycling of the Und-PP polyisoprenoid carrier into Und-P to reinitiate biosynthesis.

6.10.2.1 Initiation Reaction

The initiation reaction for O antigen subunit biosynthesis occurs at the interface of the plasma membrane and the cytosol where the nucleotide sugar precursors are available. The reaction involves the formation of a phosphodiester bond between a membrane-associated polyprenyl phosphate and a cytosolic UDP-sugar with the release of UMP. Depending on the specific microorganism, this reaction is catalyzed by two different

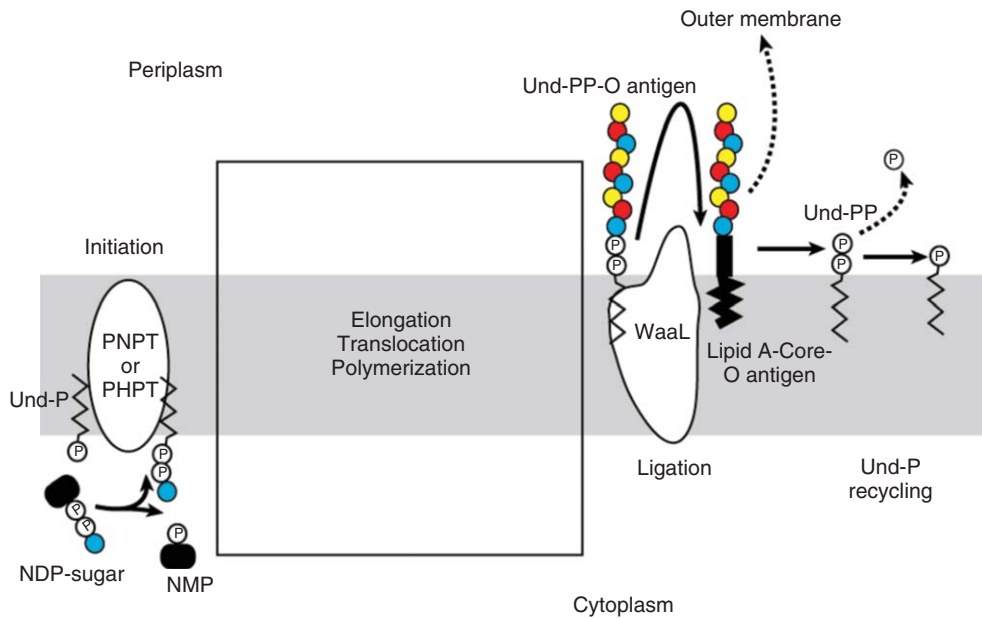


Figure 1 General steps in the biogenesis of O antigens. The gray rectangle denotes the inner membrane of the bacterial cell. The initiation reaction requires a nucleotide-diphosphate sugar and Und-P and is catalyzed by PHPT or PNPT enzymes. This reaction results in the formation of a Und-PP-linked sugar and the release of nucleotide-monophosphate. The large box denotes a number of different reactions that involve the elongation/translocation/polymerization of O repeating subunits (see **Figure 2** for details), which result in the production of a polymeric O antigen linked to Und-PP (Und-PP-O antigen) that is localized to the periplasmic side of the inner membrane. The ligation reaction, catalyzed by WaaL, results in the formation of a complete LPS molecule (lipid A-core oligosaccharide-O antigen), which is further translocated to the outer membrane and becomes surface exposed. The remaining Und-PP is recycled to Und-P by dephosphorylation reactions and reused in the synthesis of Und-PP-linked polymers.

families of proteins. One of these families corresponds to the polyisoprenyl-phosphate *N*-acetylhexosamine-1-phosphate transferases (PNPT family^{58–60}) comprising proteins that are present both in prokaryotes and in eukaryotes. The other family corresponds to the polyisoprenyl-phosphate hexose-1-phosphate transferases (PHPT family), and its prototype member is WbaP from *Salmonella enterica*. This family of proteins has no known homologues in eukaryotic cells.

6.10.2.1.1 Polyisoprenyl-phosphate *N*-acetylhexosamine-1-phosphate transferases

Eukaryotic PNPTs are localized in the membrane of the endoplasmic reticulum (ER) where they catalyze the first step in N-linked glycoprotein biosynthesis resulting in a Dol-PP-GlcNAc intermediate.⁶⁰ In contrast, bacterial PNPTs such as WecA, MraY, WbpL, and WbcO utilize different *N*-acetylhexosamine substrates and they also differ in their susceptibility to selective inhibitors.⁵⁸ Several regions of conserved amino acid sequence can be found in bacterial and eukaryotic members of the PNPT family. It is plausible that all the members of this family utilize a common enzymatic mechanism for the formation of the phosphodiester bond. However, bacterial and eukaryotic PNPTs differ in their substrate specificity for various *N*-acetylhexosamine substrates and also they can discriminate the type of polyisoprenoid phosphate.⁶¹ Und-P contains 11 isoprene units all of which are fully unsaturated, while Dol-P can be made of 15–19 isoprene units that have a saturated α -isoprene.⁶² The α -isoprene is the phosphorylated end of the molecule, which participates in the phosphodiester bond formation with the *N*-acetylhexosamine-1-P. Therefore, the ability of eukaryotic and bacterial enzymes to exquisitely discriminate their lipid substrate is likely a reflection of evolutionary divergence.

WecA is a tunicamycin-sensitive UDP-GlcNAc:Und-P GlcNAc-1-phosphate transferase.^{63,64} Tunicamycin, a nucleoside antibiotic that is thought to resemble the UDP-GlcNAc-polyisoprenoid lipid reaction intermediate,^{65,66} has the ability to inhibit the function of WecA, MraY, and the eukaryotic PNPTs.^{66,67} The *wecA* gene is located within the ECA gene cluster and serves to initiate both O antigen and ECA synthesis.^{29,63,68} ECA is a

surface glycolipid found on the outer membrane of most enteric bacteria.²⁹ WecA is usually required for the initiation of the biosynthesis of O-specific polysaccharide that have GlcNAc or *N*-acetylgalactosamine (GalNAc) in the O subunit^{64,69–73} and that are translocated across the plasma membrane via the Wzx/Wzy-dependent pathway (see Section 6.10.2.2.1). WecA is also necessary for the initiation of the biosynthesis of *O* polysaccharides transported by Wzy/Wzx-independent pathways (see Sections 6.10.2.2.2 and 6.10.2.2.3), such as *E. coli* O8 and O9,⁶³ *Klebsiella pneumoniae* O1,⁷⁴ *Serratia marcescens* O16,⁷⁵ and *S. enterica* serovar Borreze O:54.⁷⁶ *Pseudomonas aeruginosa* produces two forms of LPS, designated A-band and B-band LPS. WbpL is a PNPT member with a seemingly dual substrate recognition since it appears to be required for the initiation of B-band LPS synthesis with a FucNAc residue and A-band synthesis with either a GlcNAc or GalNAc residue.⁷⁷ The deduced amino acid sequence of WbpL shows high identity to WbcO from *Yersinia enterocolitica* that initiates O antigen synthesis by adding FucNAc to the lipid carrier Und-P.⁷⁸ The *MraY* protein catalyzes the formation of Und-P-P-*N*-acetylmuramyl-pentapeptide, the first step in the lipid cycle reactions in biosynthesis of bacterial cell wall peptidoglycans.⁷⁹ The topology of the *E. coli* as well as the *Staphylococcus aureus* *MraY* transferase was established experimentally⁸⁰ and has served as a template to model the topology of WecA.⁸¹ The predicted topology of *MraY* is also similar to eukaryotic PNPTs.⁶⁰

PNPT proteins are all polytopic membrane proteins.^{58,60,80,82} Limited information is currently available on the structural motifs or critical amino acid residues in these enzymes that may be important for substrate recognition and/or catalysis. It is plausible that regions of the protein extending into the aqueous environment of the cytosolic face of the plasma membrane (or the ER membrane) are involved in recognition and interaction with the nucleotide sugar substrates. Indeed, bacterial and eukaryotic WecA homologues share discrete regions of conserved amino acid sequence located in segments of the protein that are exposed to the cytosolic face of the plasma membrane or the membrane of the ER.^{58,59,83} Bacterial homologues also carry a large cytosolic loop containing some conserved residues that may be important for the recognition of the nucleotide sugar substrates.^{58,83}

In WecA, 11 transmembrane (TM) domains and 5 external and 5 internal loops have been predicted using several robust algorithms.^{58,82} While the first three N-terminal TM helices could be deleted without affecting membrane localization of WecA, deletion of the last predicted TM helix of WecA affected protein stability.⁸³ It is possible that TM-XI specifically interacts with another internal TM region to stabilize the protein in the membrane, as the replacement of this region with a similar TM helix from the MalB protein did not correct the stability phenotype.⁸³ The topological model of WecA was recently confirmed by the substituted cysteine accessibility method,⁸⁴ which also permitted to define better the boundaries of cytosolic loops that appear to contain critical residues for enzyme activity. Also, the location of the C-terminus in the cytosol was confirmed by a C-terminal fusion to the green fluorescent protein, which resulted in a chimeric protein that localized to the bacterial membrane and exhibited fluorescence.⁸⁴

Several studies on WecA characterized conserved aspartic acid residues in the predicted cytoplasmic loops II (D90 and D91) and III (D156 and D159). WecA derivatives with amino acid replacement at these sites were assayed by *in vivo* complementation of O antigen biosynthesis, as well as by *in vitro* transfer and UDP-GlcNAc-binding abilities.⁸¹ From these analyses, it was proposed that D90 and D91 are important in forwarding the Und-PP-linked O subunit to the next step in the assembly of the polysaccharide, namely the translocation reaction, while D156 and D159 form part of the enzyme's catalytic site. Although initial studies predicted that D156 and D159 are required for ionic interactions with metal divalent cations (especially Mg^{2+} or Mn^{2+}) that are essential for phosphoryl transfer reactions,⁸⁵ a more detailed examination suggests that they are probably interacting directly with the nucleotide UDP or the sugar GlcNAc. The experimentally determined kinetic parameters for the UDP-GlcNAc substrate supported these conclusions.⁸⁴ Moreover, WecA has a higher affinity for Mn^{2+} than for Mg^{2+} , with K_M values of 0.6 $mmol\ l^{-1}$ and 3 $mmol\ l^{-1}$, respectively. These values are in the same range of physiological concentrations of Mn^{2+} and Mg^{2+} in the bacterial cells; thus, it is likely that Mn^{2+} is the preferred ion for enzyme activity.

A conserved short sequence motif, His-Ile-His-His (HIHH), and a conserved arginine were identified in WecA at positions 279–282 and 265, respectively.⁸³ This region is located within the cytosolic loop V that is present in all bacterial homologues of WecA. Both HIHH279–282 and the Arg265 are reminiscent of the His-Ile-Gly-His (HIGH) motif and a nearby upstream lysine, which contribute to the three-dimensional architecture of the nucleotide-binding site among various enzymes displaying nucleotidyltransferase activity.⁸⁶

Alternatively, the H279 and Arg265 could be required for coordination with Mn^{2+} ions, and could play a role in maintaining the structure of the catalytic site of the enzyme. Thus, it was hypothesized that these residues may play a role in the interaction of WecA with UDP-GlcNAc.⁸³ The high level of conservation of H279 and Arg265 among bacterial WecA homologues that utilize several different UDP-*N*-acetylhexosamine substrates underscores the functional importance of these residues.

Virtually no information is available on specific residues involved in the recognition of Und-P by WecA and other proteins involved in O antigen synthesis. Albright *et al.*⁸⁷ proposed a dolichol recognition sequence (DRS) based on the alignment of yeast glycosyltransferases from the same family as WecA. The DRS was a 13 amino acid peptide with the consensus sequence LFVXFXXIPXFFY predicted to be in TM spanning regions. The DRS was also identified in mammalian GPTs and demonstrated to be required for enzyme activity.^{88–91} Upon identification of the DRS in the *E. coli* K1 proteins NeuE and KpsM, which use Und-P as a substrate, the motif was renamed “polyisoprenol recognition sequence” (PIRS).^{92,93} PIRS peptides were constructed in an attempt to demonstrate interaction between PIRS-containing enzymes and polyisoprenoids and the NMR 3D-structures of the PIRS peptides with both Und-P and Dol-P were obtained.^{93–95} Molecular modeling techniques revealed specific interactions between hydrophobic residues in the PIRS peptides and the isoprenoids. TMs 4 and 5 in WecA are the only regions of the protein with an amino acid composition rich in large aromatic amino acids, similar to the PIRS motif. Amino acid replacements in these residues, specifically Phe-143 and Trp-146, resulted in WecA mutants that lead to reduced O antigen surface production and also have reduced enzyme activity *in vitro* in the presence of excess Und-P (E. Haggerty, E. Ciepchal, and M.A. Valvano, unpublished). Molecular modeling of these TMs suggest a putative hydrophobic cleft formed by the amino acid side chains that could serve as an ideal docking site for Und-P providing multiple points for hydrophobic interactions to occur between the isoprenoid and the enzyme. Interestingly, the model also predicts that the cytoplasmic loop 2, containing the conserved D90 and D91 residues that have a yet undefined but essential role in catalysis, would be located at the end of the putative cleft between TMs 4 and 5. Although not conclusive, these observations pinpoint a role for specific TMs in interactions with Und-P, which may lead to the appropriate orientation of the phosphate groups of this lipid that participate in the phosphodiester reaction at the interface of the membrane and the cytosol. More detailed characterization of WecA awaits the determination of its tri-dimensional structure.

6.10.2.1.2 Polyisoprenyl-phosphate hexose-1-phosphate transferases

WbaP, a member of the PHPT family, is responsible for the transfer of galactosyl-1-phosphate from UDP-galactose to Und-P in a reversible reaction, which is the first step of O antigen synthesis in *S. enterica*.⁹⁶ No homologues to WbaP are found in eukaryotic systems. In contrast, WbaP shows high-level sequence similarity throughout their entire lengths with proteins involved in the synthesis of exopolysaccharides, such as GumD (*Xanthomonas campestris*), AmsG (*Erwinia amylovora*), Orf14 (*K. pneumoniae*), and WcaJ (*E. coli* K-12), which are all demonstrated or predicted UDP-glucose:Und-P glucose-1-phosphate or UDP-galactose:Und-P galactose-1-phosphate transferases.^{97–100}

The WbaP protein is a large hydrophobic protein, which has a number of potential membrane-spanning domains.⁹⁶ However, the predicted WbaP topology suggests a different model than WecA.¹⁰¹ WbaP comprises at least three predicted structural domains: an N-terminal region containing four TM helices, a large central periplasmic loop, and a C-terminal domain containing the last TM helix and a large cytoplasmic tail. The contribution of each region to WbaP function was recently examined using a series of mutant WbaP proteins and assessing their ability to complement the O antigen synthesis in $\Delta wbaP$ mutants of *S. enterica* serovars Typhi and Typhimurium.¹⁰¹ Truncated forms of WbaP lacking the periplasmic loop exhibited altered chain length distribution in the O antigen polymerization, suggesting that this central domain is involved in modulating chain length distribution of the *O polysaccharide*. The N-terminal and periplasmic domains were dispensable for complementation of O antigen synthesis *in vivo*, suggesting that the C-terminal domain carries the sugar-phosphate transferase activity. Indeed, the C-terminal half of WbaP shows high-level sequence similarity to ExoY of *Rhizobium meliloti*¹⁰² as well as with several transferases involved in the formation of group II capsules in Gram-positive bacteria.¹⁰³ However, despite that a construct expressing the C-terminus of WbaP complemented *in vivo* the synthesis of O antigen in the $\Delta wbaP$ mutant, membrane extracts containing WbaP derivatives without the N-terminal domain failed to transfer radioactive galactose from UDP-Gal into

a lipid-rich fraction.¹⁰¹ These results suggest that the N-terminal region of WbaP, containing four TM domains, is essential for the insertion or stability of the protein in the bacterial membrane. Thus, it may be possible that the domain structure of WbaP enables this protein not only to function in the transfer of Gal-1-P to Und-P but also to establish critical interactions with additional proteins required for the correct assembly of O antigen in *S. enterica*.

Wang and Reeves have proposed that the *S. enterica* serovar Typhimurium WbaP has two functions.¹⁰³ One function is the actual UDP-galactose:Und-P galactose-1-phosphate transferase reaction that is involved in the first step of O antigen synthesis. This function is primarily attributable to the C-terminal portion of the protein. The other function is manifested at a later step (but prior to the flippase reaction) and involves the release of Und-PP-galactose from WbaP, which is mediated by the N-terminal region rich in TM domains. These conclusions are based on the analysis of *wbaP(T)* mutations, which occur within the first half of the *wbaP* gene.¹⁰³ *wbaP(T)* mutants retain UDP-galactose:Und-P galactose-1-phosphate transferase activity but they cannot complete the formation of O-specific polysaccharide. This scenario is very similar to that observed with the WecA mutants in residues D90 and D91, which also retain transferase activity *in vitro* but cannot mediate O antigen synthesis *in vivo*.⁸¹ Together, these observations suggest that membrane-bound PNPT and PHPT proteins, albeit different in terms of sequence similarity and predicted topology, may provide a membrane scaffold for the additional reactions mediated by loosely membrane-associated glycosyltransferases that complete the formation of the O-specific subunits and the ensuing translocation/elongation process. In this proposed model, the growing Und-PP-linked subunit would remain associated to PNPT (or PHPT) proteins, perhaps in complex with other membrane proteins involving in subunit translocation across the plasma membrane and polymerization.

An interesting feature in PHPTs is the large periplasmic loop, which in homologues from Gram-positive bacteria should be exposed to the bacterial surface.^{104–106} Examination of the predicted secondary structure of the WbaP periplasmic loop reveals a periodicity of α -helices and β -strands.¹⁰¹ Most of the residues that are conserved in the alignments with WbaP homologues from other bacterial species are also present within the predicted α -helices and β -strands (K. Patel and M.A. Valvano, unpublished), supporting the idea that the periplasmic loop of WbaP could interact with other proteins for the O antigen synthesis, with the O antigen itself, or both. Results from genetic and biochemical experiments with *S. pneumoniae* capsule assembly genes suggest a possible function for this region. Under conditions where assembly of the complete capsule subunit is defective, suppressor mutations that reduce the activity of the initiating glycosyltransferase CpsE (WbaP homologue) can be isolated, most of which map in the surface-exposed ('periplasmic') domain.¹⁰⁵ Therefore, this region could 'sense' the formation and proper assembly of the Und-PP-linked unit and exert a feedback control over the initiation reaction. Consistent with this view, WbaP mutants with deletions in this region caused reduced production of O antigen, a change in the chain length of the polymers and they also display reduced enzyme activity *in vitro*¹⁰¹ (Patel and Valvano, unpublished).

6.10.2.2 Elongation/Translocation/Polymerization

After the formation of the first Und-PP-linked sugar, additional sugars must be added to complete the O subunit. These reactions form glycosidic bonds catalyzed by specific glycosyltransferases, which are either soluble or associated proteins to the plasma membrane by ionic interactions. Bacterial glycosyltransferases involved in the synthesis of exopolysaccharides have an abundance of positively charged amino acids, which result in predicted high *pI*s and account, at least in part, for the ionic interactions of these proteins with the plasma membrane. Glycosyltransferases are classified in different families (<http://afmb.cnrs-mrs.fr/CAZY>; for reviews see Campbell *et al.*,¹⁰⁷ Coutinho *et al.*,¹⁰⁸ and Lairson *et al.*¹⁰⁹). The cytoplasmic location of these enzymes is consistent with the notion that the assembly of Und-PP-linked O subunits occurs at the cytosolic face of the plasma membrane. However, the ligation of Und-PP-linked *O polysaccharides* to lipid A-core OS occurs at the periplasmic face of the plasma membrane.^{110,111} As a result, the biogenesis of *O polysaccharides* requires a process of assembly that includes a mechanism whereby Und-PP-linked saccharides are translocated across the plasma membrane.

At least three different mechanisms for the assembly and translocation of O-specific polysaccharides have been described (Figure 2). One of them involves the synthesis of O repeating subunits by the sequential addition of monosaccharides at the nonreducing end of the molecule, a process that takes place on the cytosolic

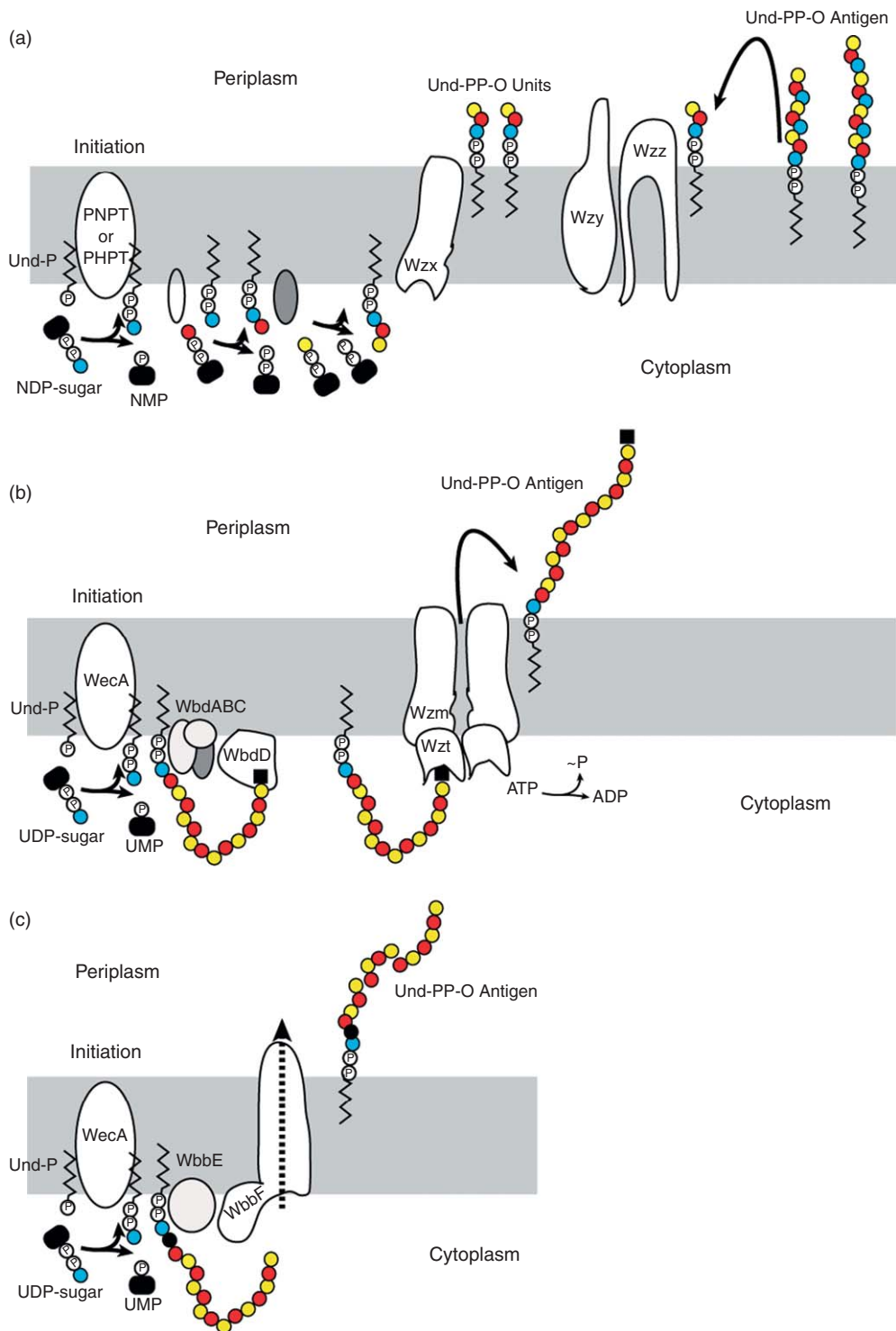


Figure 2 Pathways for O antigen assembly. (a) Wzy-dependent pathway. After the initiation reaction, glycosyltransferases extend the O antigen until the completion of the repeating unit, which is translocated across the membrane by Wzz. In the periplasmic side of the membrane nascent units are polymerized by Wzy and the control of the chain length distribution is carried out by Wzz. (b) ABC transporter-dependent pathway. In this pathway the polymer is formed intracellularly and terminated by the addition of a termination signal (black square), which also couples the polymer to the Wzt protein. ATP hydrolysis is required for export across the membrane. (c) Synthase-dependent pathway. After initiation, one more 'adaptor' sugar is added (black circle) and the WbbF bifunctional polymerase that is also responsible for its vectorial translocation across the inner membrane extends the rest of the polymer.

side of the plasma membrane.²¹ These subunits are translocated across the plasma membrane, and they subsequently become polymerized by a mechanism involving the successive addition of the reducing end of the growing polymer to the nonreducing end of Und-P-P-linked subunits that is mediated by the Wzy protein (**Figure 2(a)**). The Und-P-linked polymer is then ligated 'en bloc' to the lipid A-core OS by reactions occurring on the periplasmic face of the membrane.^{110–112} This pathway, also referred to as the *wzy* (polymerase)-dependent pathway, occurs in the synthesis of the majority of O antigens, especially in those made of repeating units of different sugars (heteropolymeric O antigens).³⁴ The pathway also involves another protein, Wzx, which is a putative flippase, and it is always present in the gene clusters containing the *wzy* gene. Therefore, it would be more appropriate to refer to this pathway as the Wzy/Wzx-dependent pathway.

A second mechanism for O antigen biosynthesis involves the formation of a polymeric O antigen by reactions taking place on the cytosolic face of the plasma membrane (**Figure 2(b)**), which are mediated by the sequential action of glycosyltransferases elongating the polysaccharide at the nonreducing end.¹¹³ The nascent polysaccharide is transported across the plasma membrane by a two-component ATP-binding cassette transporter,^{34,114} and subsequently ligated to lipid A-core OS. This pathway has been observed especially in O antigens made of repeating units of the same sugar (homopolymeric O antigens) such as those from *E. coli* O8 and O9²¹ and *K. pneumoniae* O1,¹¹⁵ as well as in *E. coli* group 2 and 3 exopolysaccharide capsules.¹¹³ Finally, a third mechanism involves a synthase (**Figure 2(c)**) and has only been described to date in *S. enterica* serovar Borreze.^{52,76}

A unique aspect in the biogenesis of complex carbohydrate structures in all types of cells is the participation of polyisoprenol lipids. Polyisoprenol-linked sugars are involved in the biosynthesis of eukaryotic glycoproteins, bacterial cell walls, and surface polysaccharides.^{22,116} Once assembled, polyisoprenol-linked sugars must cross the lipid bilayer for further processing. In the case of protein glycosylation, a topological model has been proposed involving the transmembrane movement of dolichol-linked sugars across the ER membrane.¹¹⁷ A similar model involving transmembrane 'flipping' of polyisoprenol-linked sugars has been proposed for the synthesis of bacterial cell wall peptidoglycan, teichoic acids, and ECA in bacteria as well as *N*-protein glycosylation in eukaryotes (**Figure 2**). However, how lipid-linked carbohydrates are translocated from one leaflet of the lipid bilayer to the other remains unclear. It has been shown that the unassisted transbilayer movement of polyisoprenol-linked sugars in liposomes is extremely slow,^{118,119} suggesting the need for protein-assisted translocation.^{120,121}

6.10.2.2.1 Wzy/Wzx-dependent pathway

In contrast to the Wzy/Wzx-independent O antigens (see Section 6.10.2.2.2), no obvious ABC transporters have been identified in *wzy*-dependent systems. At least three proteins (Wzx, Wzy, and Wzz) are involved in this export pathway but currently, there is no information concerning the manner in which these proteins interact with one another to facilitate the formation of predicted functional complexes. Once the individual Und-PP-linked O subunits are formed, they must be exported to the site of polymerization at the periplasmic face of the plasma membrane (**Figure 2(a)**).

6.10.2.2.1(i) O antigen translocase All *wzx/wzy*-dependent O antigen clusters studied to date contain a gene that encodes a plasma membrane protein designated Wzx that has been postulated as a candidate for the O unit flippase or translocase.¹²² Based on comparison of predicted hydrophobicity, Wzx proteins were classified within a family of integral membrane proteins with 12 predicted TM helices.¹²³ The membrane topology was experimentally confirmed for the Wzx protein of *S. enterica* serovar Typhimurium.¹²⁴ Wzx proteins share very little primary amino acid sequence similarity, and their genes have also very poor nucleotide sequence homology, to the point that they can be used as genetic markers for distinguishing among specific O antigens.^{73,125} The involvement of Wzx proteins in the translocation of Und-PP-linked O subunits was based on experiments using a heterologous O antigen expressing system in *S. enterica*. *In vivo* radiolabeling of O antigen precursors in the presence of a *wzx* mutation suggested the accumulation of Und-PP-linked O subunits with an apparent location on the cytoplasmic face of the plasma membrane.¹²² However, the small amount of material involved makes definitive localization difficult, since the *wzx* block only resulted in approximately 50% of the radiolabeled material being accumulated on the cytoplasmic face of the plasma membrane. It is possible that Wzx facilitates the transit of Und-PP-linked O subunits to the periplasmic face of the plasma

membrane in a similar manner as a permease, and with the help of proton motive force as a source of energy. This could explain the absence of typical amino acid motifs in the primary sequence of Wzx that are indicative of ATP- or GTP-binding sites. However, fractionation experiments using a *wzx* mutant in *E. coli* O7 and immunoblot analysis of the cell fractions with O7-specific antibodies, revealed an overall reduction of O7-specific precursors at the plasma membrane instead of an accumulation of precursors as would be predicted for a transporter (C. L. Marolda and M. A. Valvano, unpublished observations).

O antigen translocase activities are difficult to measure directly. Addressing the topological orientation of lipids in a membrane and the purification of the translocases are both difficult as they depend on the use of detergents that perturb native membrane structures. Although a translocase assay may reveal protein-dependent lipid flipping, the major challenge relies on maintaining the specificity of the assay for a particular protein. For example, only in one case an assay using a soluble analogue of Und-PP-GlcNAc was linked to a genetically identified translocase, WzxE.¹²⁶ However, the same strain contains another translocase, Wzx_{O16}, and despite that both WzxE and Wzx_{O16} are interchangeable,¹²⁷ that assay did not detect an activity for Wzx_{O16}.¹²⁶ The biochemical assays to date require detergents to prepare membrane extracts, which generates denatured membrane proteins that in turn can provide continuity between the leaflets of the membranes and thereby serve as artifactual conduits for polar moieties.^{128–130} The lack of a definitive biochemical assay to determine ‘flipping’ of the Und-PP-linked O subunits complicates the conclusive determination of the function of Wzx. Supporting a role for O antigen translocation, recent work has demonstrated that PglK and Wzx are interchangeable.¹³¹ However, PglK is an ABC transporter implicated in the flipping of Und-PP-linked saccharides from *Campylobacter jejuni*,¹³¹ which are the glycan component of an N-linked bacterial glycosylation machinery.^{132,133}

Since the membrane translocation of Und-PP-linked O subunits must be a conserved process, the absence of any obvious conserved motifs in the primary amino acid sequences of Wzx proteins is intriguing. One possible explanation for the abundant differences among Wzx proteins could be the requirement for the recognition of specific O subunits, which are highly variable in terms of structure and sugar composition. However, a complete Und-PP-linked O subunit is not required for ‘flipping’ since a single sugar, GlcNAc, can be incorporated into the lipid A-core OS of *E. coli* K-12 by a process requiring *wecA* and *wzx*.¹³⁴ This demonstrates that Und-PP-linked with the first sugar of the O unit is the minimal substrate for Wzx-dependent translocation. Using a genetic system based on reconstructing O16 antigen synthesis in *E. coli* K-12, we showed that Wzx homologues from different bacteria complement an *E. coli* K-12 Δwzx mutant.¹³⁵ More specifically, Wzx proteins from O antigen systems that use Und-PP-GlcNAc or Und-PP-GalNAc for the initiation of the biosynthesis of the O repeat can fully complement the formation of O16 LPS. Partial complementation was seen with Wzx from *P. aeruginosa*, a system that uses Und-PP-FucNAc in the initiation reaction, while the complementation with the Wzx protein from *S. enterica* (that uses Und-PP-Gal) was only possible under high levels of protein expression. Therefore, it would appear that Wzx proteins, like the initiating Und-PP-sugar transferases, occur in at least two functional classes that distinguish among Und-PP-bound *N*-acetyl hexosamines or hexoses. However, it is not clear if this is due to specific recognition of the initiating sugar in the context of the phosphoisoprenoid lipid, an interaction between Wzx and the corresponding initiating transferase, or a combination of both.

6.10.2.2.1(ii) O antigen polymerase and the regulator of O antigen chain length The Wzy protein required for the polymerization of Und-PP-linked O subunits at the periplasmic face of the cytoplasmic membrane. The polymerization reaction involves transfer of nascent polymer from its Und-PP carrier to the reducing end of the new Und-PP-linked O subunit.^{136,137} Mutants with defects in the *wzy* gene produce LPS consisting of lipid A-core bound to a single O unit.¹³⁸ The released Und-PP is recycled to the active monophosphoryl form by dephosphorylation.²⁴ Wzy proteins appear to be integral membrane proteins with 11–13 predicted TM domains, and they exhibit little primary sequence similarity.^{139,140} In contrast to Wzx, Wzy proteins from different O types are not interchangeable and display specificity for the cognate O subunit or for structures containing the same linkage between O antigen subunits.^{4,21,141} Several Wzy proteins examined with robust computer programs that predict topology appear to possess a relatively large periplasmic loop that may be important in the recognition of the O antigen subunit. This topology has been experimentally demonstrated in at least one Wzy protein.¹⁴⁰ However, the enzymatic mechanism of Wzy has not been resolved, in part because

of the absence of distinguishing features in Wzy proteins, but also due to difficulties in expressing sufficient amounts for *in vitro* studies, which complicates the identification of catalytic and binding residues in Wzy.

The third component in the Wzy/Wzx-dependent pathway is the Wzz protein. Wzz generates the strain-specific chain length distribution of *O polysaccharide* chain lengths as reflected in characteristic clusters of bands following gel electrophoresis of LPS samples.¹⁴² The function of the Wzz protein is not required for bacterial growth in the laboratory setting, as polymerization can proceed in the absence of this protein. However, the distribution of the *O polysaccharide* chain length is critical for virulence.^{143–149}

Wzz proteins are in the plasma membrane, and all have two TM helices flanking a periplasmic loop with a predicted coiled-coil structure.¹⁵⁰ Wzz belongs to a family of proteins called ‘polysaccharide copolymerases’.¹⁵⁰ Some of its members occur in the synthesis of capsules and they have a larger cytosolic C-terminal region that contains ATP-binding sites and several tyrosine residues that can become phosphorylated.^{151–155} Phosphorylation and dephosphorylation of these proteins are important for the export of capsular polysaccharides (reviewed in Whitfield⁵⁶), but the specific mechanism of export is still not well understood.

More than one *wzz* gene has been observed in some microorganisms like *P. aeruginosa*¹⁵⁶ and *S. flexneri*.¹⁵⁷ It is not clear if the presence of additional Wzz activities would have an additive effect in the overall length of the *O polysaccharides* or alternatively, they would be differentially required under varying physiological conditions.

Several models have been proposed to explain the modality in the polymerization process. Wzz is hypothesized to act as a timing clock, interacting with the Wzy polymerase and modulating its activity between two states that favor either chain elongation or chain termination caused by transfer of the O polymer to the ligase.¹⁵⁸ An alternative model implicates Wzz as a molecular chaperone to assemble a complex consisting of Wzy, the WaaL ligase, and Und-PP-linked O-specific polysaccharide.⁵⁵ The specific modality would then be determined by different kinetics resulting from a Wzz-dependent ratio of Wzy relative to WaaL. However, it has been shown that ligation is not required for modality.^{133,159} The coiled-coil domains were proposed to be important for interactions of Wzz with Wzy, WaaL, or both. However, only Wzz oligomers have been identified by chemical cross-linking experiments,⁵⁴ while a definitive evidence of cross-linking of Wzz to either Wzy or WaaL is lacking. The function of Wzz proteins appears not to be specific for a given O-repeat subunit structure but the regions of the protein required for Wzz modality are not well defined.^{54,160} Recent work has shown that the periplasmic loop of the *E. coli* K-12 Wzz has an extended conformation, and mutagenesis experiments suggest that the regions predicted as coiled coils are important for Wzz function by maintaining the native conformation of the protein, but do not support the existence of coiled coils *per se*.¹⁶¹ The elucidation of the crystal structures of the periplasmic domains of three Wzz homologues that impart substantially different chain length distributions to surface polysaccharides shows that they share a common protomer structure with a long extended central α -helix.¹⁶² The protomers self-assemble into bell-shaped oligomers of variable sizes, with a large internal cavity and functional studies suggest that the top of the PCP oligomers is an important region for determining polysaccharide modal length. These observations led to a new model for Wzz function in which the oligomers would organize the polymerization such that the size of the oligomer would determine the number of associated Wzy molecules and ultimately the length of the polymer.¹⁶² A difficulty with this model is that Wzy is very poorly expressed,^{140,163} which would affect the stoichiometry of putative macromolecular complex. Also, recent work using cryoelectron microscopy and reconstitution of Wzz proteins suggest that all of them form hexameric complexes.¹⁶⁴ Regardless of the actual mechanism a theme is emerging whereby a periplasmic protein (like Wzz or Wzc in capsule export) assemble into oligomeric structures that extend into the periplasmic space.⁵⁷ In the case of capsule export it is proposed that these complexes interact with an outer membrane protein channel required for the surface assembly of the polymer. For O antigen synthesis, the putative complex may deliver the Und-PP-linked polysaccharide to the ligation step within the periplasmic space, but it still remains unclear how this process occurs, as efforts by several laboratories to demonstrate a complex have been unsuccessful to date (see Section 6.10.2.2.1(iii)).

6.10.2.2.1(iii) Are the components of the Wzy-dependent pathway in a membrane complex? Several authors have suggested that the proteins of the Wzy-dependent pathway function as multiprotein complexes.^{54,55,159,165,166} Also, it is possible that protein components for the assembly of the ECA, which is similar to the Wzy-dependent pathway, exist together in the plasma membrane as a complex.¹⁶⁷ Direct evidence exists for oligomerization *in vivo* of at least one of these proteins, Wzz, in *S. flexneri*,⁵⁴ *E. coli*

K12/O16,¹⁶⁸ and *P. aeruginosa*.¹⁵⁹ However, efforts to provide biochemical evidence for the existence of a complex involving other proteins have been unsuccessful (R. Morona, personal communication; C.L. Marolda and M.A. Valvano, unpublished; Contreras, personal communication). Compelling genetic data support the notion that Wzy, Wzz, and Wzx work in concert as a functional complex.¹²⁷ This evidence comes from reconstitution of O antigen synthesis studies in *E. coli* K-12 where it has been shown that *wzx_E* gene (encoding the translocase for ECA) can fully complement a *wzx_{O16}* translocase deletion mutant only if the majority of the ECA gene cluster is deleted. At the same time, the introduction of plasmids expressing either the Wzy_E polymerase or the Wzz_E chain length regulator proteins from the ECA cluster drastically reduces the O16 LPS complementing activity of Wzx_E. Similar results were observed with the O7 system and Wzx_{O7} can cross complement translocase defects in the O16 and O7 antigen clusters only in the absence of their corresponding Wzz and Wzy proteins. These genetic data strongly suggest that translocation of O antigen across the plasma membrane and the subsequent assembly of periplasmic Und-PP-linked *O polysaccharide* depends on interactions among Wzx, Wzz, and Wzy, which presumably form a multiprotein complex. Therefore, it is possible that multiprotein complexes at the plasma membrane exist for the translocation and assembly of the O antigen. Another evidence for the possible existence of complexes is that WecA, the initiating Und-PP-sugar transferase is located in discrete regions of the plasma membrane.⁸⁴ Early work in *Salmonella* has shown that new O antigen LPS molecules appear on the cell surface at a limited number of sites ('adhesion zones'),^{169,170} and more recent work has provided experimental evidence to support the existence of multiprotein complexes for the assembly of capsular polysaccharides serving as a molecular 'scaffold' across the periplasm.^{171,172} Fractionation experiments also show that WecA is not only in the low-density plasma membrane fraction but also in a fraction of intermediate density, which contains markers of both outer and plasma membrane proteins (Tatar and Valvano, unpublished). These membrane fractions were shown to contain newly synthesized material that is exported to the outer membrane and are considered to be the biochemical equivalent of the 'adhesion zones'.^{173–176}

6.10.2.2.1(iv) Parallels of the Wzy-dependent pathway and N-linked protein glycosylation The N-linked protein glycosylation pathway in eukaryotes has remarkable parallels with the biogenesis of Wzx/Wzy-dependent O-specific polysaccharides (Figure 2) (see Chapter 6.12). As in bacteria, the process can be divided into similar steps involving the assembly of a lipid-linked OS (analogous to the O subunit) on the cytosolic side of the ER, the translocation of this molecule across the ER membrane (analogous to the flipping reaction mediated by Wzx), and the transfer of the OS from its lipid anchor to selected asparagines of nascent glycoproteins (analogous to the ligation reaction of the O-specific polymer with the core lipid A molecules²²). The initiation reaction that results in the biosynthesis of Dol-PP-GlcNAc is followed by subsequent reactions on the cytosolic side of the ER membrane that involves the addition of another GlcNAc and three mannose residues. These reactions are mediated by specific glycosyltransferases and result in the formation of a heptasaccharide-lipid linked intermediate, Dol-PP-GlcNAc₂Man₅. These reactions are analogous to those taking place on the cytosolic side of the bacterial plasma membrane, which result in the formation of the O antigen subunits as well as other precursor molecules for peptidoglycan and cell surface polysaccharides in general. The Rft1 protein¹⁷⁷ carries out the translocation of the lipid-linked heptasaccharide intermediate across the ER membrane in *Saccharomyces cerevisiae*. A mutation in the *rft1* gene is lethal in yeast and its phenotype is associated with a protein glycosylation defect.¹⁷⁸ Rft1 is an integral membrane protein with 12 predicted TM domains, which in contrast to other lipid flippases,¹⁷⁹ lacks any motifs indicative of ABC-type transporters, and its gene is highly conserved in eukaryotic genomes. Interestingly, an *rft1* gene homologue is absent in *Plasmodium falciparum*, an organism that apparently lacks N-linked protein glycosylation.¹⁸⁰ The function of Rft1 is analogous to that of bacterial Wzx. Both proteins share similarities in size and hydrophobicity plots, and both also lack identifiable motifs. However, the two families do not share any obvious similarities in their primary amino acid sequences. Whether these two protein families share a similar translocation mechanism or their structural differences reflect the nature of the different substrates it is presently unknown. The role of Rft1 in the flipping of Dol-PP-GlcNAc₂Man₅ has recently been challenged based strictly on biochemical reconstitution experiments,¹⁸¹ which unfortunately suffer from the same shortcomings described above concerning the loss of specificity in this type of assays (see Section 6.10.2.2.1(i)).

6.10.2.2.2 ABC transporter-dependent pathway

In this pathway, the O-specific polysaccharide is completed at the inner face of the cytoplasmic membrane and the export of the polymer to the outer face for ligation requires an ABC transporter (**Figure 2(b)**). The biosynthesis of *E. coli* O8 and O9 has been used as a model system for the ABC-dependent transporter pathway.⁷ Both O8 and O9 antigens are homopolymers of mannose and their structures and antigenicity depends on the different linkages formed between the mannose residues. One salient feature of the *O polysaccharides* formed by this pathway is the participation of a primer Und-PP-GlcNAc intermediate followed by the addition of a sugar adaptor molecule. Although these *O polysaccharides* are all initiated by the activity of WecA, they differ from Wzy/Wzx *O polysaccharides* in that the GlcNAc residue transferred to lipid A-core OS during ligation occurs only once per chain, and thus it is not found within the repeat unit structure itself.^{63,182} Next, an *O polysaccharide*-specific glycosyltransferase adds an adaptor sugar residue between the und-PP-GlcNAc primer and the repeat subunit domain, and this reaction also occurs only once per chain. Different enzymes are involved in adding the adaptor. In *E. coli* O9, adaptor formation involves the addition of a single mannose residue by WbdC.¹⁸³ In the other cases, like in some serotypes of *K. pneumoniae* LPS,¹⁸² the adaptor is added by the bifunctional galactosyltransferase, WbbO, which also participates in subsequent chain extension reactions.^{74,184} The *O polysaccharide* is assembled in the chain extension phase, by the processive transfer of residues to the nonreducing terminus of the Und-PP-linked acceptor,^{183–185} which may be mediated by either monofunctional or multifunctional transferases.^{183,186,187}

Owing to the processive nature of the polymerization, an intriguing aspect about the polymers assembled by the ABC transport-dependent pathway is their mode of termination. In the case of the Wzy/Wzx-exported *O polysaccharides*, this process results from the interactions that involve the Wzy and Wzz proteins. Despite that the ABC transport-dependent pathway does not involve a Wzz protein the O-specific polysaccharides formed by this pathway display strain-specific chain length (modal) distributions.⁵³ The chain length of these polymers is controlled by the WbdD protein, which modifies the nonreducing end of the polymer,¹⁸⁸ causing the termination of the polymerization. The nature of the termination reaction varies, depending on the specific type of O antigen. The WbdD protein for *E. coli* O8 adds a 3-O-methyl group donated from S-adenosylmethionine, while the corresponding protein from the O9 serotype acts both as a kinase and a methyltransferase.¹⁸⁸ The termination reactions at the nonreducing end of the polymer are not only important for the completion of the polymerization but also serve to couple termination to polymer export, which is mediated by an ABC-2 subfamily of ATP-binding cassette transporters.¹⁸⁹

ABC-2 transporters consist of an integral membrane protein, Wzm, with an average of six TM domains, and a hydrophilic protein containing the nucleotide-binding domain, Wzt. Genes encoding these two components are present within the *O polysaccharide* biosynthesis clusters. As with other ABC transporters involved in transmembrane export, Wzm homologues for *O polysaccharide* biosynthesis display very little primary sequence identity, but Wzt homologues are much more highly conserved, especially in the nucleotide-binding region. However, Wzm proteins are functionally interchangeable between different O antigens, while Wzt proteins are not.¹⁹⁰ The nonconserved C-terminal region of Wzt determines the specificity¹⁹⁰ and recent structural data revealed that this domain forms a β sandwich with an immunoglobulin-like fold that contains the *O polysaccharide*-binding pocket.¹⁹¹ Presumably, binding of the polysaccharide to this region would promote a conformational change in the nucleotide-binding domain driving ATP hydrolysis, and promoting the interactions with Wzm resulting in the export of the Und-PP-linked *O polysaccharide*.

6.10.2.2.3 Synthase-dependent pathway

The plasmid-encoded O:54 antigen of *S. enterica* serovar Borreze is the only known example of a synthase-dependent *O polysaccharide*.^{52,76,192} The O:54-specific polysaccharide is a homopolymer made of N-acetylmannosamine (ManNAc). In a similar fashion to the ABC transport-dependent pathway, WecA⁷⁶ initiates the synthesis of the O:54 subunit and the first ManNAc residue is transferred to the Und-PP-GlcNAc primer by the nonprocessive ManNAc transferase WbbE.¹⁹³ The second transferase, WbbF, belongs to the HasA (hyaluronan synthase) family of glycosaminoglycan glycosyltransferases,¹⁹⁴ and it is proposed that this enzyme performs the chain-extension steps (**Figure 2(c)**). Synthases are integral membrane proteins,^{194,195} which appear to catalyze a vectorial polymerization reaction by a processive mechanism resulting in the extension of the polysaccharide chain with the simultaneous extrusion of the nascent polymer across the plasma

membrane.¹⁹⁵ Although the exported polymer is presumably Und-PP-linked, there is very little information on the exact mechanism of export mediated by WbbF as well as in the process leading to chain termination. The synthase family has other members including the enzymes involved in biosynthesis of cellulose, chitin, and hyaluronan,^{194,195} and the type 3 capsules of *S. pneumoniae*.^{106,196} Two conserved domains, one likely involved in the glycosyl transfer reaction and the other implicated in the translocation of the nascent polymer, characterize these enzymes.

6.10.2.3 Ligation Reaction

Irrespective of the export and polymerization modes of the saccharide molecules, nascent Und-PP-linked O antigens are ligated to terminal sugar residues of the lipid A-core OS in a reaction mediated by *waaL* gene product (Figure 1). WaaL is an integral membrane protein, and currently the only protein presumed to be required for ligation. The *waaL* gene maps within the *waa* gene cluster that also encodes other enzymes for the biosynthesis and assembly of core OS.^{7,20} The ligation reaction occurs at the periplasmic face of the cytoplasmic membrane.¹¹⁰ Mutant strains devoid of a functional *waaL* gene cannot ligate O antigen molecules to lipid A-core OS resulting in the production a 'rough' LPS lacking O antigen polysaccharide and accumulation of intracellular Und-PP-linked O antigen molecules.^{110,111} Remarkably, ligase mutants are viable, in contrast to mutants in the *wzx* translocase^{126,127} (Marolda and Valvano, unpublished). It is not clear why the accumulation of Und-PP-linked O antigen precursors would be lethal if it takes place in the cytosolic side of the membrane but not when it happens on the other side. It may be possible that the periplasmic accumulation of unprocessed Und-PP-saccharides somehow provides a signal that results in a downregulation of O antigen biosynthesis, but this has not been systematically investigated.

The mechanism of ligation is still unresolved. Although the ligase catalyzes the formation of a glycosidic bond WaaL proteins share no similarities with any of the glycosyltransferases that use sugar nucleotides. A requirement for a specific lipid A-core OS acceptor structure has been established in several model systems,^{197–200} which has led to the generalized notion that the specific WaaL protein can recognize a specific lipid A-core OS terminal structure. For example, in *E. coli* there are five chemically distinct core OS types, K-12, R1, R2, R3, and R4,²⁰ while only two types are found in *S. enterica*.^{20,201,202} Both *Salmonella* and *E. coli* K12, and presumably all Gram-negative bacteria, can ligate any number of Und-PP-linked recombinant O antigens^{7,20,202} (McGarry and Valvano, unpublished). However, it is not clear how WaaL recognizes the Und-PP-linked O antigen and, in particular, which part of this molecule participates in the enzymatic reaction.

To date, studies on O antigen ligases have been mainly limited to establishing the topology of the protein using hydrophobicity plot analyses,^{197,199,203} and protein fusions to topology probes like alkaline phosphatase and β -galactosidase.²⁰⁰ Unfortunately, WaaL proteins show significant divergence in their primary amino acid sequence, even for members from the same species.⁶ The extremely low sequence conservation among O antigen ligases makes comparative analyses difficult. Therefore, a detailed knowledge of the residues involved in ligase activity or the chemical characteristics of the ligation reaction are unknown. WaaL proteins show significant divergence in their primary amino acid sequence, even for members from the same species.⁶ The poor sequence conservation among O antigen ligases makes comparative analyses difficult. It is also difficult to establish relationships in WaaL proteins based on potential core OS acceptor structures. For example, the *E. coli* R2 and *S. enterica* WaaL proteins share ~80% amino acid sequence similarity and are functionally interchangeable, as both link the O antigen polysaccharide to a terminal glucose in the core OS that has an α -1,2-linked *N*-acetylglucosamine.¹⁹⁹ In contrast, *E. coli* R3 WaaL is ~66% similar to the *Salmonella* protein but links the O antigen polysaccharide to a different site of attachment in the core OS that resembles a similar site in the K12 core OS. However, the R3 WaaL shares very little identity with the *E. coli* K12 ligase.¹⁹ More recent evidence suggests that the specificity of the ligation reaction for a particular lipid A-core OS structure does not solely depend on the WaaL protein, but presumed additional factor or factors have not been identified.²⁰² Recently, Abeyrathne and Lam²⁰⁴ reported that highly purified WaaL from *P. aeruginosa* has ATPase activity and ATP hydrolysis is required for the *in vitro* ligation reaction. This is an intriguing finding since ATP is not present in the periplasmic space.²⁰⁵ Also, an extensive mutagenesis analysis of amino acid motifs in WaaL putatively involved in ATP binding or hydrolysis did not afford WaaL-defective proteins.²⁰⁶

Conceivably, WaaL activity requires amino acids exposed to the periplasmic space where they could interact with the donor and acceptor molecules. A critical His residue, which is somewhat conserved in many WaaL proteins, was identified in a periplasmic loop of the *Vibrio cholerae* WaaL,²⁰⁰ and a potentially common motif is emerging not only in WaaL proteins but also in proteins that ligate Und-PP-linked O antigen precursors to pili.²⁰⁷ Recent work in our laboratory has established a tri-dimensional structural model of the WaaL large periplasmic loop that consists of two pairs of almost perpendicular α -helices in which all the conserved residues in other WaaL proteins cluster within a putative catalytic region.²⁰⁶ The model also predicted that Arg288 and His337, two residues that are critical for WaaL function, face each other and are exposed to the solvent in a spatial arrangement that suggests interactions with substrate molecules. In addition, a conserved arginine, also critical for WaaL function, is invariably present in the short periplasmic loop preceding the large loop.²⁰⁶ These results support the notion that a positively charged region exposed to the periplasmic face of WaaL plays a critical role in either catalysis or binding of the Und-PP-linked O antigen substrate.

6.10.3 Future Prospects

A full understanding at the mechanistic level of the biosynthesis of O antigens continues to be challenging. The experimental demonstration of the existence of protein complexes involving the various components that participate in the biogenesis of *O polysaccharides* is a major goal in several laboratories. Isolation of these complexes has been difficult in part because expressing membrane proteins like for instance Wzy, Wzx, and WaaL is not trivial. The use of epitope tags has worked in some cases but failed in others (C. L. Marolda, E. Vines, and M. A. Valvano, unpublished). Several other aspects require new tools for analysis. In particular, elucidating the precise function of the O antigen translocase Wzx would benefit from devising an unequivocal assay for transmembrane flipping. Finally, the structural characterization of the O antigen assembly proteins at the atomic level would greatly help understanding their function. Finally, the precise role of polyisoprenoid lipids, their transmembrane movement, and in particular the recycling pathway of Und-PP would also add considerably to our current understanding of the biogenesis of O-specific polysaccharides.

Acknowledgments

Grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council, and the Canadian Cystic Fibrosis Foundation support research in the author's laboratory. The author holds a Canada Research Chair in Infectious Diseases and Microbial Pathogenesis.

References

1. B. Beutler; K. Hoebe; X. Du; R. J. Ulevitch, *J. Leukocyt. Biol.* **2003**, *74* (4), 479–485.
2. Y. Kamio; H. Nikaido, *Biochemistry* **1976**, *15* (12), 2561–2570.
3. H. Nikaido, Outer Membrane. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F. C. Neidhardt, R. Curtiss, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, Eds.; American Society for Microbiology Press: Washington, DC, 1996; pp 29–47.
4. C. Whitfield; M. A. Valvano, *Adv. Microb. Physiol.* **1993**, *35*, 135–246.
5. C. R. H. Raetz, Bacterial Lipopolysaccharides: A Remarkable Family of Bioactive Molecules. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed.; F. C. Neidhardt, R. Curtiss, III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, Eds.; American Society for Microbiology Press: Washington, DC, 1996; pp 1035–1063.
6. C. R. Raetz; C. M. Reynolds; M. S. Trent; R. E. Bishop, *Annu. Rev. Biochem.* **2007**, *76*, 295–329.
7. C. R. H. Raetz; C. Whitfield, *Annu. Rev. Biochem.* **2002**, *71* (1), 635–700.
8. M. A. Valvano; P. Messner; P. Kosma, *Microbiology* **2002**, *148*, 1979–1989.
9. H. Nikaido, *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 593–656.
10. H. Nikaido; M. Vaara, *Microbiol. Rev.* **1985**, *49* (1), 1–32.
11. H. Nikaido, *Science* **1994**, *264* (5157), 382–388.

12. P. E. Jansson, The Chemistry of O-polysaccharide Chains in Bacterial Lipopolysaccharides. In *Endotoxin in Health and Disease*; H. Brade, D. C. Morrison, S. Vogel, S. Opal, Eds.; Marcel Dekker, Inc.: New York, 1999; pp 155–178.
13. R. Stenutz; A. Weintraub; G. Widmalm, *FEMS Microbiol. Rev.* **2006**, *30* (3), 382–403.
14. G. Pluschke; A. Mercer; B. Kusecek; A. Pohl; M. Achtman, *Infect. Immun.* **1983**, *39*, 599–608.
15. G. Pluschke; M. Achtman, *Infect. Immun.* **1984**, *43* (2), 684–692.
16. K. A. Joiner, *Annu. Rev. Microbiol.* **1988**, *42*, 201–230.
17. S. Samuel; P. Reeves, *Carbohydr. Res.* **2003**, *338*, 2503–2519.
18. M. A. Valvano, *Front. Biosci.* **2003**, *8*, s452–s471.
19. D. E. Heinrichs; J. A. Yethon; C. Whitfield, *Mol. Microbiol.* **1998**, *30* (2), 221–232.
20. D. E. Heinrichs; M. A. Valvano; C. Whitfield, Biosynthesis and Genetics of Lipopolysaccharide Core. In *Endotoxin in Health and Disease*; H. Brade, D. C. Morrison, S. Vogel, S. Opal, Eds.; Marcel Dekker, Inc.: New York, NY, 1999; pp 305–330.
21. C. Whitfield, *Trends Microbiol.* **1995**, *3* (5), 178–185.
22. P. Burda; M. Aebi, *Biochim Biophys Acta* **1999**, *1426*, 239–257.
23. F. Fernandez; J. S. Rush; D. A. Toke; G. Han; J. E. Quinn; G. M. Carman; J.-Y. Choi; D. R. Voelker; M. Aebi; C. J. Waechter, *J. Biol. Chem.* **2001**, *276* (44), 41455–41464.
24. M. A. Valvano, *Mol. Microbiol.* **2008**, *67* (2), 232–235.
25. M. El Ghachi; A. Bouhss; D. Blanot; D. Mengin-Lecreulx, *J. Biol. Chem.* **2004**, *279* (29), 30106–30113.
26. M. El Ghachi; A. Derbise; A. Bouhss; D. Mengin-Lecreulx, *J. Biol. Chem.* **2005**, *280* (19), 18689–18695.
27. L. D. Tatar; C. L. Marolda; A. N. Polischuk; D. van Leeuwen; M. A. Valvano, *Microbiology* **2007**, *153*, 2518–2529.
28. T. Touze; A. X. Tran; J. V. Hankins; D. Mengin-Lecreulx; M. S. Trent, *Mol. Microbiol.* **2008**, *67* (2), 264–277.
29. P. D. Rick; R. P. Silver, Enterobacterial Common Antigen and Capsular Polysaccharides. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F. C. Neidhardt, R. Curtiss, III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, Eds.; American Society for Microbiology Press: Washington, DC, 1996; pp 104–122.
30. J. S. Brooke; M. A. Valvano, *J. Biol. Chem.* **1996**, *271* (7), 3608–3614.
31. M. A. Valvano, *J. Endotoxin. Res.* **1999**, *5*, 90–95.
32. M. A. Valvano; C. L. Marolda; M. Bittner; M. Glaskin-Clay; T. L. Simon; J. D. Klena, *J. Bacteriol.* **2000**, *182*, 488–497.
33. C. A. Schnaitman; J. D. Klena, *Microbiol. Rev.* **1993**, *57* (3), 655–682.
34. W. J. Keenleyside; C. Whitfield, Genetics and Biosynthesis of Lipopolysaccharide O-antigens. In *Endotoxin in Health and Disease*, H. Brade, D. C. Morrison, S. Vogel, S. Opal, Eds.; Marcel Dekker, Inc.: New York, 1999; pp 331–358.
35. P. R. Reeves; M. Hobbs; M. A. Valvano; M. Skurnik; C. Whitfield; D. Coplin; N. Kido; J. Klena; D. Maskell; C. R. H. Raetz; P. D. Rick, *Trends Microbiol.* **1996**, *4*, 495–503.
36. S. Tamaki; T. Sato, *J. Bacteriol.* **1971**, *105*, 968–975.
37. R. Curtiss; J. Charamella; D. R. Stallions; J. A. Mays, *Bacteriol. Rev.* **1968**, *32*, 320–348.
38. L. M. Havekes; B. J. J. Lugtenberg; W. P. M. Hoekstra, *Mol. Gen. Genet.* **1976**, *146*, 43–50.
39. C. Sherburne; D. E. Taylor, *J. Bacteriol.* **1997**, *179*, 952–955.
40. W. van Alphen; B. Lugtenberg; W. Berendsen, *Mol. Gen. Genet.* **1976**, *147* (3), 263–269.
41. H. Vakharia; R. Misra, *Mol. Microbiol.* **1996**, *19* (4), 881–889.
42. A. J. Verkleij; E. J. Lugtenberg; P. H. Ververgaert, *Biochim. Biophys. Acta* **1976**, *426* (3), 581–586.
43. M. E. Bayer; J. Koplou; H. Goldfine, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72* (12), 5145–5149.
44. J. Koplou; H. Goldfine, *J. Bacteriol.* **1974**, *117* (2), 527–543.
45. J. A. Fralick; L. L. Burns-Keliher, *J. Bacteriol.* **1994**, *176*, 6404–6406.
46. V. Koronakis; E. Koronakis; J. Li; K. Stauffer, *Mol. Microbiol.* **1997**, *23*, 617–626.
47. J. A. Yethon; E. Vinogradov; M. B. Perry; C. Whitfield, *J. Bacteriol.* **2000**, *182* (19), 5620–5623.
48. J. A. Yethon; C. Whitfield, *J. Biol. Chem.* **2001**, *276* (8), 5498–5504.
49. L. W. Riley; L. N. Junio; L. B. Libaek; G. K. Schoolnik, *Infect. Immun.* **1987**, *55*, 2052–2056.
50. D. J. Kopecko; O. Washington; S. B. Formal, *Infect. Immun.* **1980**, *29*, 207–214.
51. S. Sturm; B. Jann; K. Jann; P. Fortnagel; K. N. Timmis, *Microb. Pathog.* **1986**, *1* (3), 299–306.
52. W. J. Keenleyside; C. Whitfield, *J. Bacteriol.* **1995**, *177* (18), 5247–5253.
53. C. Whitfield; P. A. Amor; R. Koplou, *Mol. Microbiol.* **1997**, *23* (4), 629–638.
54. C. Daniels; R. Morona, *Mol. Microbiol.* **1999**, *34* (1), 181–194.
55. R. Morona; L. van den Bosch; P. A. Manning, *J. Bacteriol.* **1995**, *177* (4), 1059–1068.
56. C. Whitfield, *Annu. Rev. Biochem.* **2006**, *75*, 39–68.
57. C. Whitfield; J. Naismith, *Curr. Opin. Struct. Biol.* **2008**, *18* (4), 466–474.
58. M. S. Anderson; S. S. Eveland; N. P. Price, *FEMS Microbiol. Lett.* **2000**, *191* (2), 169–175.
59. A. R. Dal Nogare; M. A. Lehrman, *Glycobiology* **1988**, *8*, 625–632.
60. M. A. Lehrman, *Glycobiology* **1994**, *4* (6), 768–771.
61. J. S. Rush; P. D. Rick; C. J. Waechter, *Glycobiology* **1997**, *7* (2), 315–322.
62. J. F. Pennock; F. W. Hemming; R. A. Morton, *Nature* **1960**, *186*, 470–472.
63. P. D. Rick; G. L. Hubbard; K. Barr, *J. Bacteriol.* **1994**, *176* (10), 2877–2884.
64. D. C. Alexander; M. A. Valvano, *J. Bacteriol.* **1994**, *176* (22), 7079–7084.
65. A. Heifetz; R. W. Keenan; A. D. Elbein, *Biochemistry* **1979**, *18*, 2186–2192.
66. P. Brandish; K. Kimura; M. Inukai; R. Southgate; J. Lonsdale; T. Bugg, *Antimicrob. Agents Chemother.* **1996**, *40* (7), 1640–1644.
67. W. C. Mahoney; D. Duksin, *J. Biol. Chem.* **1979**, *254* (14), 6572–6576.
68. U. Meier-Dieter; R. Starman; K. Barr; H. Mayer; P. D. Rick, *J. Biol. Chem.* **1990**, *265*, 13490–13497.
69. J. D. Klena; C. A. Schnaitman, *Mol. Microbiol.* **1993**, *9* (2), 393–402.
70. Z. Yao; M. A. Valvano, *J. Bacteriol.* **1994**, *176*, 4133–4143.
71. L. Wang; S. Huskic; A. Cisterne; D. Rothermund; P. R. Reeves, *J. Bacteriol.* **2002**, *184* (10), 2620–2625.
72. L. Zhang; J. Radziejewska-Lebrecht; D. Krajewska-Pietrasik; P. Toivanen; M. Skurnik, *Mol. Microbiol.* **1997**, *23* (1), 63–76.

73. L. Wang; P. R. Reeves, *Infect. Immun.* **1998**, *66*, 3545–3551.
74. B. R. Clarke; D. Bronner; W. J. Keenleyside; W. B. Severn; J. C. Richards; C. Whitfield, *J. Bacteriol.* **1995**, *177* (19), 5411–5418.
75. M. Szabo; D. Bronner; C. Whitfield, *J. Bacteriol.* **1995**, *177* (6), 1544–1553.
76. W. J. Keenleyside; M. Perry; L. Maclean; C. Poppe; C. Whitfield, *Mol. Microbiol.* **1994**, *11* (3), 437–448.
77. H. L. Rocchetta; L. L. Burrows; J. C. Pacan; J. S. Lam, *Mol. Microbiol.* **1998**, *28* (6), 1103–1119.
78. M. Skurnik, Molecular Genetics of *Yersinia* Lipopolysaccharide. In *Genetics of Bacterial Polysaccharides*, J. B. Goldberg, Ed.; CRC Press: Boca Raton, FL, 1999; pp 23–51.
79. M. Ikeda; M. Wachi; H. K. Jung; F. Ishino; M. Matsuhashi, *J. Bacteriol.* **1991**, *173* (3), 1021–1026.
80. A. Bouhss; D. Mengin-Lecreux; D. Le Beller; J. Van Heijenoort, *Mol. Microbiol.* **1999**, *34* (3), 576–585.
81. A. O. Amer; M. A. Valvano, *Microbiology* **2002**, *148*, 571–582.
82. A. O. Amer; M. A. Valvano, *J. Bacteriol.* **2000**, *182*, 498–503.
83. A. O. Amer; M. A. Valvano, *Microbiology* **2001**, *147*, 3015–3025.
84. J. Lehrer; K. A. Vigeant; L. D. Tatar; M. A. Valvano, *J. Bacteriol.* **2007**, *189*, 2618–2628.
85. J. S. Allingham; P. A. Pribil; D. B. Haniford, *J. Mol. Biol.* **1999**, *289* (5), 1195–1206.
86. S. Sekine; A. Shimada; O. Nureki; J. Cavarelli; D. Moras; D. G. Vassylyev; S. Yokoyama, *J. Biol. Chem.* **2001**, *276* (6), 3723–3726.
87. C. F. Albright; P. Orlean; P. W. Robbins, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (19), 7366–7369.
88. J. R. Scocca; S. S. Krag, *J. Biol. Chem.* **1990**, *265* (33), 20621–20626.
89. X. Y. Zhu; M. A. Lehrman, *J. Biol. Chem.* **1990**, *265* (24), 14250–14255.
90. M. A. Lehrman, *Glycobiology* **1991**, *1* (6), 553–562.
91. A. K. Datta; M. A. Lehrman, *J. Biol. Chem.* **1993**, *268* (17), 12663–12668.
92. F. A. Troy, II, *Glycobiology* **1992**, *2* (1), 5–23.
93. G. P. Zhou; F. A. Troy, II, *Glycobiology* **2003**, *13* (2), 51–71.
94. G. P. Zhou; F. A. Troy, *Glycobiology* **2005**, *15* (4), 347–359.
95. G. P. Zhou; F. A. Troy, II, *Curr. Protein Pept. Sci.* **2005**, *6* (5), 399–411.
96. X. M. Jiang; B. Neal; F. Santiago; S. J. Lee; L. K. Romana; P. R. Reeves, *Mol. Microbiol.* **1991**, *5* (3), 695–713.
97. Y. Arakawa; R. Wacharotayankun; T. Nagatsuka; H. Ito; N. Kato; M. Ohta, *J. Bacteriol.* **1995**, *177* (7), 1788–1796.
98. P. Bugert; K. Geider, *Mol. Microbiol.* **1995**, *15*, 917–933.
99. G. Stevenson; K. Andrianopoulos; M. Hobbs; P. R. Reeves, *J. Bacteriol.* **1996**, *178* (16), 4885–4893.
100. F. Katzen; D. U. Ferreira; C. G. Oddo; M. V. Ielmini; A. Becker; A. Puhler; L. Ielpi, *J. Bacteriol.* **1998**, *180* (7), 1607–1617.
101. M. S. Saldías; K. Patel; C. L. Marolda; M. Bittner; I. Contreras; M. A. Valvano, *Microbiology* **2008**, *154* (Pt 2), 440–453.
102. T. L. Reuber; G. C. Walker, *Cell* **1993**, *74* (2), 269–280.
103. L. Wang; D. Liu; P. R. Reeves, *J. Bacteriol.* **1996**, *178* (9), 2598–2604.
104. K. Steiner; R. Novotny; K. Patel; E. Vinogradov; C. Whitfield; M. A. Valvano; P. Messner; C. Schaffer, *J. Bacteriol.* **2007**, *189*, 2590–2598.
105. B. Xayarath; J. Yother, *J. Bacteriol.* **2007**, *189* (9), 3369–3381.
106. R. T. Cartee; W. T. Forsee; M. H. Bender; K. D. Ambrose; J. Yother, *J. Bacteriol.* **2005**, *187* (21), 7425–7433.
107. J. A. Campbell; G. J. Davies; V. Bulone; B. A. Henrissat, *Biochem. J.* **1997**, *326*, 929–939.
108. P. M. Coutinho; E. Deleury; G. J. Davies; B. Henrissat, *J. Mol. Biol.* **2003**, *328* (2), 307–317.
109. L. L. Lairson; B. Henrissat; G. J. Davies; S. G. Withers, *Annu. Rev. Biochem.* **2008**, *77*, 521–555.
110. B. C. McGrath; M. J. Osborn, *J. Bacteriol.* **1991**, *173* (2), 649–654.
111. C. A. Mulford; M. J. Osborn, *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80* (5), 1159–1163.
112. P. A. Marino; B. C. McGrath; M. J. Osborn, *J. Bacteriol.* **1991**, *173* (10), 3128–3133.
113. C. Whitfield; I. S. Roberts, *Mol. Microbiol.* **1999**, *31*, 1307–1319.
114. D. Bronner; B. R. Clarke; C. Whitfield, *Mol. Microbiol.* **1994**, *14* (3), 505–519.
115. B. R. Clarke; C. Whitfield, *J. Bacteriol.* **1992**, *174* (14), 4614–4621.
116. T. D. Bugg; P. E. Brandish, *FEMS Microbiol. Lett.* **1994**, *119*, 255–262.
117. C. B. Hirschberg; M. D. Snider, *Annu. Rev. Biochem.* **1987**, *54*, 63–87.
118. M. A. McCloskey; F. A. Troy, *Biochemistry* **1980**, *19*, 2061–2066.
119. J. A. Hanover; W. J. Lennarz, *J. Biol. Chem.* **1978**, *254*, 9237–9246.
120. J. S. Rush; C. J. Waechter, *J. Cell. Biol.* **1995**, *130*, 529–536.
121. J. S. Rush; K. van Leyen; O. Ouerfelli; B. Wolucka; C. J. Waechter, *Glycobiology* **1998**, *8*, 1195–1205.
122. D. Liu; R. A. Cole; P. R. Reeves, *J. Bacteriol.* **1996**, *178* (7), 2102–2107.
123. I. T. Paulsen; A. M. Beness; M. H. Saier, *Microbiology* **1997**, *143*, 2685–2699.
124. M. M. Cunneen; P. R. Reeves, *FEMS Microbiol. Lett.* **2008**, *287* (1), 76–84.
125. C. L. Marolda; M. F. Feldman; M. A. Valvano, *Microbiology* **1999**, *145*, 2485–2496.
126. P. D. Rick; K. Barr; K. Sankaran; J. Kajimura; J. S. Rush; C. J. Waechter, *J. Biol. Chem.* **2003**, *278* (19), 16534–16542.
127. C. L. Marolda; L. D. Tatar; C. Alaimo; M. Aebi; M. A. Valvano, *J. Bacteriol.* **2006**, *188*, 5124–5135.
128. M. A. Kol; A. I. de Kroon; D. T. Rijkers; J. A. Killian; B. de Kruijff, *Biochemistry* **2001**, *40* (35), 10500–10506.
129. M. A. Kol; A. van Dalen; A. I. de Kroon; B. De Kruijff, *J. Biol. Chem.* **2003**, *278* (27), 24586–24593.
130. M. A. Kol; A. N. van Laak; D. T. Rijkers; J. A. Killian; A. I. de Kroon; B. De Kruijff, *Biochemistry* **2003**, *42* (1), 231–237.
131. C. Alaimo; I. Catrein; L. Morf; C. L. Marolda; N. Callewaert; M. A. Valvano; M. F. Feldman; M. Aebi, *EMBO J.* **2006**, *25*, 967–976.
132. M. Wacker; D. Linton; P. G. Hitchen; M. Nita-Lazar; S. M. Haslam; S. J. North; M. Panico; H. R. Morris; A. Dell; B. Wren; M. Aebi, *Science* **2002**, *298* (5599), 1790–1793.
133. M. F. Feldman; M. Wacker; M. Hernandez; P. G. Hitchen; C. L. Marolda; M. Kowarik; H. R. Morris; A. Dell; M. A. Valvano; M. Aebi, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (8), 3016–3021.
134. M. F. Feldman; C. L. Marolda; M. A. Monteiro; M. B. Perry; A. J. Parodi; M. A. Valvano, *J. Biol. Chem.* **1999**, *274*, 35129–35138.
135. C. L. Marolda; J. Vicarioli; M. A. Valvano, *Microbiology* **2004**, *150*, 4095–4105.
136. D. Bray; P. W. Robbins, *Biochem. Biophys. Res. Commun.* **1967**, *28* (3), 334–339.
137. P. W. Robbins; D. Bray; B. M. Dankert; A. Wright, *Science* **1967**, *158*, 1536–1542.

138. L. V. Collins; S. Attridge; J. Hackett, *Infect. Immun.* **1991**, 59 (3), 1079–1085.
139. R. Morona; M. Mavris; A. Fallarino; P. A. Manning, *J. Bacteriol.* **1994**, 176 (3), 733–747.
140. C. Daniels; C. Vindurampulle; R. Morona, *Mol. Microbiol.* **1998**, 28, 1211–1222.
141. W. Yi; L. Zhu; H. Guo; M. Li; J. Li; P. G. Wang, *Carbohydr. Res.* **2006**, 341 (13), 2254–2260.
142. R. A. Batchelor; G. E. Haraguchi; R. A. Hull; S. I. Hull, *J. Bacteriol.* **1991**, 173 (18), 5699–5704.
143. G. L. Murray; S. R. Attridge; R. Morona, *Mol. Microbiol.* **2003**, 47 (5), 1395–1406.
144. G. L. Murray; S. R. Attridge; R. Morona, *Microbes Infect.* **2005**, 7 (13), 1296–1304.
145. G. L. Murray; S. R. Attridge; R. Morona, *J. Bacteriol.* **2006**, 188 (7), 2735–2739.
146. J. A. Bengoechea; L. Zhang; P. Toivanen; M. Skurnik, *Mol. Microbiol.* **2002**, 44 (4), 1045–1062.
147. A. Hoare; M. Bittner; J. Carter; S. Alvarez; M. Zaldívar; D. Bravo; M. A. Valvano; I. Contreras, *Infect. Immun.* **2006**, 74, 1555–1564.
148. N. Jimenez; R. Canals; M. T. Salo; S. Vilches; S. Merino; J. M. Tomas, *J. Bacteriol.* **2008**, 190 (12), 4198–4209.
149. E. Kintz; J. M. Scarff; A. DiGiandomenico; J. B. Goldberg, *J. Bacteriol.* **2008**, 190 (8), 2709–2716.
150. R. Morona; L. Van Den Bosch; C. Daniels, *Microbiology* **2000**, 146 (1), 1–4.
151. C. Vincent; P. Doublet; C. Grangeasse; E. Vaganay; A. J. Cozzone; B. Duclos, *J. Bacteriol.* **1999**, 181 (11), 3472–3477.
152. O. Ilan; Y. Bloch; G. Frankel; H. Ullrich; K. Geider; I. Rosenshine, *EMBO J.* **1999**, 18 (12), 3241–3248.
153. A. J. Cozzone; C. Grangeasse; P. Doublet; B. Duclos, *Arch. Microbiol.* **2004**, 181 (3), 171–181.
154. P. Doublet; C. Grangeasse; B. Obadia; E. Vaganay; A. J. Cozzone, *J. Biol. Chem.* **2002**, 277 (40), 37339–37348.
155. P. Doublet; C. Vincent; C. Grangeasse; A. J. Cozzone; B. Duclos, *FEBS Lett.* **1999**, 445 (1), 137–143.
156. H. L. Rocchetta; L. L. Burrows; J. S. Lam, *Microbiol. Mol. Biol. Rev.* **1999**, 63 (3), 523–553.
157. G. Stevenson; A. Kessler; P. R. Reeves, *FEMS Microbiol. Lett.* **1995**, 125 (1), 23–30.
158. D. A. Bastin; G. Stevenson; P. K. Brown; A. Haase; P. R. Reeves, *Mol. Microbiol.* **1993**, 7 (5), 725–734.
159. C. Daniels; C. Griffiths; B. Cowles; J. S. Lam, *Environ. Microbiol.* **2002**, 4, 883–897.
160. A. V. Franco; D. Liu; P. R. Reeves, *J. Bacteriol.* **1998**, 180 (10), 2670–2675.
161. C. L. Marolda; E. R. Haggerty; M. Lung; M. A. Valvano, *J. Bacteriol.* **2008**, 190, 2128–2137.
162. A. Tocilj; C. Munger; A. Proteau; R. Morona; L. Purins; E. Ajamian; J. Wagner; M. Papadopoulos; L. Van Den Bosch; J. L. Rubinstein; J. Fethiere; A. Matte; M. Cygler, *Nat. Struct. Mol. Biol.* **2008**, 15 (2), 130–138.
163. S. Lukomski; R. A. Hull; S. I. Hull, *J. Bacteriol.* **1996**, 178 (1), 240–247.
164. K. Larve; M. S. Kimber; R. Ford; C. Whitfield, *J. Biol. Chem.* **2009**, 284 (11), 7395–7403.
165. J. A. Bengoechea; E. Pinta; T. Salminen; C. Oertelt; O. Holst; J. Radziejewska-Lebrecht; Z. Piotrowska-Seget; R. Venho; M. Skurnik, *J. Bacteriol.* **2002**, 184 (15), 4277–4287.
166. J. A. Gaspar; J. A. Thomas; C. L. Marolda; M. A. Valvano, *Mol. Microbiol.* **2000**, 38, 262–275.
167. J. Kajimura; A. Rahman; P. D. Rick, *J. Bacteriol.* **2005**, 187, 6917–6927.
168. F. Stenberg; P. Chovanec; S. L. Maslen; C. V. Robinson; L. Ilag; G. von Heijne; D. O. Daley, *J. Biol. Chem.* **2005**, 280, 34409–34419.
169. C. F. Kulpa, Jr.; L. Leive, *J. Bacteriol.* **1976**, 126 (1), 467–477.
170. P. F. Muhlradt; J. Menzel; J. R. Golecki; V. Speth, *Eur. J. Biochem.* **1973**, 35 (3), 471–481.
171. R. F. Collins; K. Beis; B. R. Clarke; R. C. Ford; M. Hulley; J. H. Naismith; C. Whitfield, *J. Biol. Chem.* **2006**, 281 (4), 2144–2150.
172. C. McNulty; J. Thompson; B. Barrett; L. Lord; C. Andersen; I. S. Roberts, *Mol. Microbiol.* **2006**, 59, 907–922.
173. K. Ishidate; E. S. Creeger; J. Zrike; S. Deb; B. Glauner; T. J. MacAlister; L. I. Rothfield, *J. Biol. Chem.* **1986**, 261 (1), 428–443.
174. E. Bouveret; R. Derouiche; A. Rigal; R. Lloubes; C. Lazdunski; H. Benedetti, *J. Biol. Chem.* **1995**, 270 (19), 11071–11077.
175. E. Cascales; M. Gavioli; J. N. Sturgis; R. Lloubés, *Mol. Microbiol.* **2000**, 38, 904–915.
176. G. Guihard; P. Boulanger; H. Benedetti; R. Lloubes; M. Besnard; L. Letellier, *J. Biol. Chem.* **1994**, 269 (8), 5874–5880.
177. J. Helenius; D. T. Ng; C. L. Marolda; P. Walter; M. A. Valvano; M. Aebi, *Nature* **2002**, 415 (6870), 447–450.
178. D. T. W. Ng; E. D. Spear; P. Walter, *J. Cell Biol.* **2000**, 150 (1), 77–88.
179. H. Sprong; P. van der Sluijs; G. van Meer, *Nat. Rev. Mol. Cell. Biol.* **2001**, 2, 504–513.
180. E. A. Davidson; D. C. Gowda, *Biochimie* **2001**, 83, 601–604.
181. C. G. Frank; S. Sanyal; J. S. Rush; C. J. Waechter; A. K. Menon, *Nature* **2008**, 454 (7204), E3–E4.
182. M. Süsskind; L. Brade; H. Brade; O. Holst, *J. Biol. Chem.* **1998**, 273 (12), 7006–7017.
183. N. Kido; V. I. Torgov; T. Sugiyama; K. Uchiya; H. Sugihara; T. Komatsu; N. Kato; K. Jann, *J. Bacteriol.* **1995**, 177 (8), 2178–2187.
184. S. Guan; A. J. Clarke; C. Whitfield, *J. Bacteriol.* **2001**, 183 (11), 3318–3327.
185. C. Weisgerber; K. Jann, *Eur. J. Biochem.* **1982**, 127 (1), 165–168.
186. N. Kido; H. Kobayashi, *J. Bacteriol.* **2000**, 182 (9), 2567–2573.
187. N. Kido; T. Sugiyama; T. Yokochi; H. Kobayashi; Y. Okawa, *Mol. Microbiol.* **1998**, 27 (6), 1213–1221.
188. B. Clarke; L. Cuthbertson; C. Whitfield, *J. Biol. Chem.* **2004**, 279 (34), 35709–35718.
189. J. Reizer; A. Reizer; M. H. Saier, Jr., *Protein. Sci.* **1992**, 1 (10), 1326–1332.
190. L. Cuthbertson; J. Powers; C. Whitfield, *J. Biol. Chem.* **2005**, 280 (34), 30310–30319.
191. L. Cuthbertson; M. S. Kimber; C. Whitfield, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104 (49), 19529–19534.
192. W. J. Keenleyside; C. Whitfield, *J. Biol. Chem.* **1996**, 271 (45), 28581–28592.
193. W. J. Keenleyside; A. J. Clarke; C. Whitfield, *J. Bacteriol.* **2001**, 183 (1), 77–85.
194. P. L. DeAngelis, *Glycobiology* **2002**, 12 (1), 9R–16R.
195. P. L. DeAngelis, *Cell. Mol. Life. Sci.* **1999**, 56 (7–8), 670–682.
196. W. T. Forsee; R. T. Cartee; J. Yother, *J. Biol. Chem.* **2000**, 275 (34), 25972–25978.
197. P. Abeyrathne; C. Daniels; K. K. Poon; M. J. Matewish; J. Lam, *J. Bacteriol.* **2005**, 187 (9), 3002–3012.
198. D. E. Heinrichs; J. A. Yethon; P. A. Amor; C. Whitfield, *J. Biol. Chem.* **1998**, 273 (45), 29497–29505.
199. D. E. Heinrichs; M. A. Monteiro; M. B. Perry; C. Whitfield, *J. Biol. Chem.* **1998**, 273 (15), 8849–8859.
200. S. Schild; A. K. Lamprecht; J. Reidl, *J. Biol. Chem.* **2005**, 280 (27), 25936–25947.
201. M. M. Olsthoorn; B. O. Petersen; S. Schlecht; J. Haverkamp; K. Bock; J. E. Thomas-Oates; O. Holst, *J. Biol. Chem.* **1998**, 273 (7), 3817–3829.

202. N. A. Kaniuk; E. Vinogradov; C. Whitfield, *J. Biol. Chem.* **2004**, 279 (35), 36470–36480.
203. J. Nesper; A. Kraiss; S. Schild; J. Blass; K. E. Klose; J. Bockemuhl; J. Reidl, *Infect. Immun.* **2002**, 70 (5), 2419–2433.
204. P. Abeyrathne; J. Lam, *Mol. Microbiol.* **2007**, 65 (5), 1345–1359.
205. A. P. Pugsley, *Microbiol. Rev.* **1993**, 57 (1), 50–108.
206. J. M. Pérez; M. A. McGarry; C. L. Marolda; M. A. Valvano, **2008**, 70 (6), 1424–1440.
207. M. Qutyan; M. Paliotti; P. Castric, *Mol. Microbiol.* **2007**, 66 (6), 1444–1458.

Biographical Sketch



Professor Miguel A. Valvano is a professor and chair of the Department of Microbiology and Immunology, University of Western Ontario, professor of Medicine, University of Western Ontario, and a Canada Research Chair in Infectious Diseases and Microbial Pathogenesis. He earned his M.D. degree from the University of Buenos Aires, Argentina and acquired further specialization in pediatrics and infectious diseases. Before moving to Canada to begin his independent career at the University of Western Ontario, Professor Valvano completed a fellowship in molecular biology in the Department of Microbiology and Immunology, Oregon Health Sciences University. His laboratory studies enzymes and other proteins involved in lipopolysaccharide synthesis, specifically the synthesis of the O antigen. Special emphasis is placed on studying various membrane proteins that are critical for O antigen synthesis, polymerization, and translocation across the membrane. Professor Valvano also studies the role of the lipopolysaccharide in the stability and permeability properties of *Burkholderia cepacia*. This microorganism is intrinsically highly resistant to antibiotics and a major health risk for children and young adults with the genetic disease cystic fibrosis.

6.11 Biosynthesis of Complex Mucin-Type O-Glycans

Inka Brockhausen, Queen's University, Kingston, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.11.1 | Introduction | 315 |
| 6.11.2 | Mucins | 316 |
| 6.11.3 | Mucin-Type O-Glycans | 318 |
| 6.11.4 | Functions of O-Glycans and Changes of O-Glycan Structures in Disease | 320 |
| 6.11.5 | Biosynthesis of O-Glycans | 322 |
| 6.11.6 | Structures and Mechanisms of Glycosyltransferases | 323 |
| 6.11.7 | Initiation of O-Glycosylation | 324 |
| 6.11.7.1 | Structures of Polypeptide GalNAc-Transferases | 325 |
| 6.11.7.2 | Site Specificity of Initial O-Glycosylation | 326 |
| 6.11.8 | Synthesis of Core 1 | 326 |
| 6.11.9 | Synthesis of Core 2 | 327 |
| 6.11.9.1 | Structure and Specificity of C2GnT1 | 328 |
| 6.11.9.2 | Role of Core 2 Structures | 330 |
| 6.11.10 | Synthesis of Core 3 | 331 |
| 6.11.11 | Synthesis of Core 4 | 332 |
| 6.11.12 | Synthesis of Minor Core Structures | 333 |
| 6.11.13 | Extension and Branching Reactions | 333 |
| 6.11.13.1 | β 3-GlcNAc-Transferases | 334 |
| 6.11.13.2 | β 6-GlcNAc-Transferases | 334 |
| 6.11.13.3 | Extension GalNAc-Transferases | 334 |
| 6.11.14 | The β 3- and β 4-Gal-Transferase Families | 335 |
| 6.11.14.1 | β 3-Gal-Transferases | 335 |
| 6.11.14.2 | β 4-Gal-Transferases | 335 |
| 6.11.15 | Sialyltransferases | 336 |
| 6.11.15.1 | α 3-Sialyltransferases | 337 |
| 6.11.15.2 | α 6-Sialyltransferases | 338 |
| 6.11.15.3 | α 8-Sialyltransferases | 339 |
| 6.11.16 | Fucosyltransferases | 339 |
| 6.11.16.1 | α 2-Fuc-Transferases | 339 |
| 6.11.16.2 | α 3/4-Fuc-Transferases | 340 |
| 6.11.17 | Blood Group Transferases | 341 |
| 6.11.17.1 | Blood Group A and B Transferases | 341 |
| 6.11.17.2 | α 3-Gal-Transferase That Synthesizes the Linear B Determinant | 342 |
| 6.11.17.3 | Cad β 4-GalNAc-Transferase | 342 |
| 6.11.17.4 | α 4-GlcNAc-Transferase | 342 |
| 6.11.18 | Sulfotransferases | 342 |
| 6.11.18.1 | Gal-3-O-Sulfotransferases | 343 |
| 6.11.18.2 | GlcNAc-6-O-Sulfotransferases | 343 |
| 6.11.19 | Future Needs and Directions | 343 |
| References | | 344 |

6.11.1 Introduction

Glycoproteins commonly carry covalently attached sugars at Asn residues (N-linked oligosaccharides) within the Asn–X–Ser/Thr motif, and at Ser or Thr residues (O-linked oligosaccharides, O-glycans). Several different types of Ser/Thr–O-linked oligosaccharides exist. The core of the linkage region of proteoglycans consists of

glucuronic acid/ β 1-3Gal/ β 1-3-Gal/ β 1-4-Xyl/ β -O-Ser oligosaccharides.¹ Many nuclear and cytoplasmic proteins have a single GlcNAc/ β -O- linked to Ser/Thr. In collagens, short oligosaccharides, Glc α 1-2Gal/ β -, are O-linked to hydroxy-Lys.² Human dystroglycan has sialyl α 2-3Gal/ β 1-4GlcNAc/ β 1-2 Man α - chains linked to Ser/Thr residues,³ which resemble the nonreducing termini of *N*-glycans. Similarly, sialyl α 2-6Gal/ β 1-4GlcNAc/ β 1-3 chains can be O-linked to Fuc α -O-Ser/Thr in epidermal growth factor (EGF) repeats of glycoproteins.⁴ The rare *O*-glycan Glc/ β 1-3Fuc α -Ser/Thr has been found in thrombospondin.⁵ EGF domains can also carry Xyl/ β 1-3Xyl/ β 1-3Glc/ β -O-Ser oligosaccharides. Mucins are particularly rich in a great variety of GalNAc α -O-Ser/Thr-based oligosaccharides, which have been named mucin-type *O*-glycans. This review discusses the biosynthesis of these mucin-type *O*-glycans.

6.11.2 Mucins

Mucins are highly O-glycosylated proteins that consist of approximately 50–80% carbohydrate by weight and are produced in epithelial cells and mucin-secreting goblet cells. The hallmark of a mucin is its variable number of tandem repeat (VNTR) region, which consists of Thr/Ser/Pro-rich peptides that are repeated many times within the peptide backbone. The lengths and sequences of VNTR and numbers of repeats vary among the different mucins (Table 1). Because of the extensive O-glycosylation of mucin peptides, molecules form an extended conformation resembling a bottle brush^{6–8} (Figure 1). Distinct peptide domains have been identified in mucins, for example, EGF domains are found near the membrane domain in membrane-bound mucins, phosphorylation sites in the cytoplasmic tail of membrane-bound mucins, proteolytic cleavage sites, Cys-rich domains, and D (dimerization) domains also found in von Willebrand factor. The latter domains have roles in polymerization. Some secreted mucins can be relatively small and monomeric such as MUC7 found in saliva (Table 1). Other secreted mucins are large, polymeric, gel forming, and are the essential constituents of the mucus gel that covers internal epithelia in the gastrointestinal tract (e.g., MUC2 in colonic mucosa) as well as in the genital tract (e.g., MUC5B) and tracheobronchial tissue (MUC5AC). These gel-forming mucins can have a size of 10 million daltons and are extremely hydrophilic, bind salts and microbes, and have protective functions. Mucins may have a transmembrane region and are found as type I membrane proteins with the O-glycosylated domain exposed to the outside of the cell (e.g., MUC1).^{9–12} These membrane-bound mucins are involved in the regulation of cell adhesion and other cell surface functions, and can associate with other cell surface molecules. In addition to classical mucins, cell surface glycoproteins with Ser/Thr/Pro-rich O-glycosylated domains are mucin-like and display mucin-type *O*-glycan structures that can interact with lectins; for example, the P-selectin glycoprotein ligand (PSGL-1) interacts with selectins, which is important for the homing of lymphocytes and in inflammation.¹³ Other glycoproteins may also carry one or more *O*-glycans at specific sites or domains.

Table 1 O-glycosylated domains of selected human mucins

| Name | TM | Secreted | Gel | VNTR | Tissue |
|--------|----|----------|-----|-------------------------------|----------------------------|
| MUC1 | + | | | GSTAPPAHGVT <u>SAPDTR</u> PAP | Breast |
| MUC2 | | + | + | PTTTPITTTTIVTPTPTGTQT | Colon, airway |
| MUC3A | + | | | HSTPSFTSSITTTETTS | Breast, trachea, intestine |
| MUC5AC | | + | + | TTSTTSAP | Stomach, breast, trachea |
| MUC5B | | + | + | SSTPGTWWILTELTTATTTESTGSTATP | Tracheobronchial, breast |
| MUC7 | | + | | TTAAPPTPSATTQAPPSSAPPE | Salivary |
| MUC9 | | + | | GGETMTTVGNQSVTP ^a | Hamster oviduct |
| MUC12 | + | | | EESTTYHRSPGSTPTTHFP | Gastrointestinal |

^a The sequence is from the VNTR of the golden hamster oviductin. Selected human transmembrane-bound (TM) mucins and secreted mucins are listed, including those that form a mucus gel (Gel). The highly O-glycosylated Ser/Thr/Pro-rich variable number of tandem repeats (VNTR) show a representative amino acid sequence in one of the repeats. The underlined sequence in MUC1 is a peptide epitope recognized by anti-MUC1 antibody SM3. Some of the tissues expressing these mucins are listed but most mucins are widely expressed and found in many other tissues and cell types.

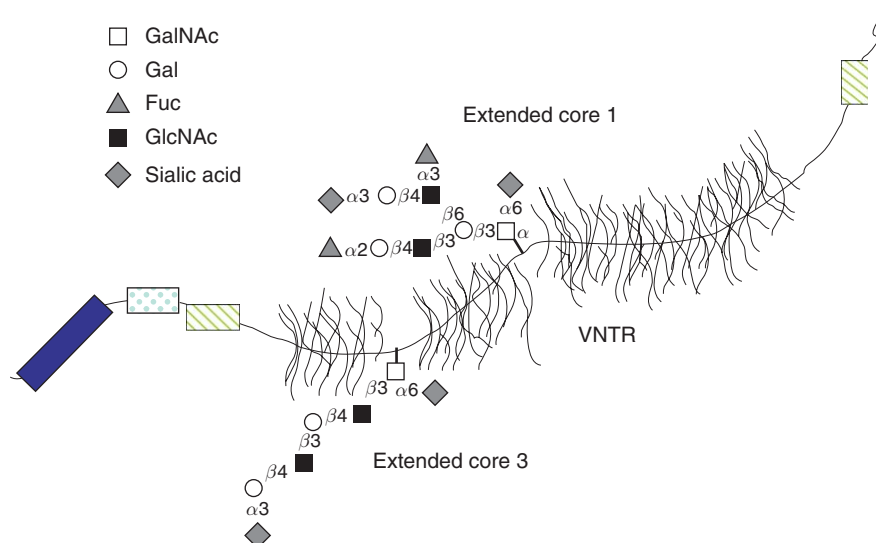


Figure 1 Model of a gel-forming mucin in the intestinal tract. A large secreted mucin is depicted, such as MUC2 in the colon. The VNTR (variable number of tandem repeat) region is rich in Ser, Thr, and Pro and is highly O-glycosylated; therefore, the peptide assumes an extended 'bottle brush'-like conformation. The majority of colonic mucin O-glycans appear to have extended core 3 structures but extended core 1 and other O-glycan structures are also found. Both ends of the molecule have Cys-rich regions and other domains that may be involved in the polymerization to form large molecules of several million daltons.

More than 20 mucin genes have been cloned to date. Mucins are found in all mammals and also in other species; for example, frog eggs are extremely rich in gel-forming mucins. The carbohydrate chains of mucins are usually heterogeneous in structures (Table 2). Some mucins, such as bronchial mucins, probably have hundreds of oligosaccharides with different structures attached to them, and may be fully O-glycosylated at every Ser/Thr residue within the VNTR. This makes it very difficult to assign functions to each of these glycans.

Many different types of carbohydrate antigens can be found on mucins (Table 2). These include the common blood group antigens ABH, the Lewis antigens, and cancer-associated Tn and T antigens. Although it is possible to predict O-glycosylation sites based on enzyme specificities and established glycosylation sites in glycoproteins (e.g., using NetOGlyc or OGPET databases), it is very difficult to assign individual O-glycan structures to specific Thr and Ser residues. Owing to the protease resistance of highly O-glycosylated peptide regions, it is virtually impossible to produce mucin glycopeptides with intact O-glycans for analyses of the specificity of O-glycosylation. However, some structural assignments have been achieved by chemical methods.¹⁴ For structural analysis, O-glycans can be cleaved off the protein by alkaline treatment under reducing conditions (β -elimination), and individual sugars can be removed by glycosidases, while acid treatment can cleave sialic acids and Fuc residues.

The functions of the large gel-forming mucins include lubrication of the epithelial surfaces and protection from chemical and physical impact and microbial damage. Bacteria, viruses, and other microbes bind to mucin-type O-glycans and are trapped by the viscous mucus layer. This may prevent their migration toward the epithelial cell surface and support their removal by ciliary action. Bacteria can also use the carbohydrate moieties of mucins as nutrients.

Cell surface-bound mucins can be involved in cell adhesion and also in the prevention of cell adhesion. For example, mucins participate in the control of cell adhesion mediated by integrin and E-cadherin.¹⁵ Some of these mucins are shed from the cell surface and are found in the bloodstream where they can play a role in the control of the immune system. For example, MUC1 from breast cancer cells is shed and can be isolated from the serum.¹⁶ Depending on the nature of the mucins, they can block natural killer cell-mediated cell lysis and the action of cytotoxic lymphocytes.¹⁷ In cancer and other diseases affecting the epithelium, mucin gene expression is often altered. Especially, MUC1 mucin is highly expressed on tumorigenic ductal epithelial cells,

Table 2 Structures of mucin-type O-glycans and alterations in cancer

| O-glycan | Structure | Alterations in cancer |
|---|---|-----------------------|
| Tn antigen | GalNAc α -Ser/Thr | ↑ |
| Sialyl-Tn (STn) antigen | Sialyl α 2-6GalNAc α -Ser/Thr | ↑ |
| Core 1, T antigen | Gal β 1-3GalNAc α - | ↑ |
| Sialyl-T antigens | Sialyl α 2-3Gal β 1-3GalNAc α - | ↑ |
| | Sialyl α 2-6(Gal β 1-3)GalNAc α - | ↑ |
| | Sialyl α 2-6(Sialyl α 2-3Gal β 1-3)GalNAc α - | ↑ |
| Polysialic acid epitope | Sialyl α 2-8sialyl α 2-3Gal β 1-3(\pm sialyl α 2-6)GalNAc α - | |
| Core 2 | GlcNAc β 1-6(Gal β 1-3)GalNAc α - | ↑↓ |
| Core 3 | GlcNAc β 1-3GalNAc α - | ↓ |
| Core 4 | GlcNAc β 1-6(Gal β 1-3)GalNAc α - | ↓ |
| Core 5 | GalNAc α 1-3GalNAc α - | |
| Core 6 | GlcNAc β 1-6GalNAc α - | |
| Core 7 | GalNAc α 1-6GalNAc α - | |
| Core 8 | Gal α 1-3GalNAc α - | |
| Type 1 chain | [GlcNAc β 1-3 Gal β 1-3] | ↓ |
| Type 2 chain | [GlcNAc β 1-3 Gal β 1-4], poly- <i>N</i> -acetylactosamines | ↑ |
| Blood group O, H | Fuc α 1-2Gal- | ↑↓ |
| Blood group A | GalNAc α 1-3(Fuc α 1-2)Gal- | ↑↓ |
| Blood group B | Gal α 1-3(Fuc α 1-2)Gal- | ↑↓ |
| Linear B | Gal α 1-3Gal- | |
| Blood group i | Gal β 1-4GlcNAc β 1-3Gal- | |
| Blood group I | Gal β 1-4GlcNAc β 1-6(Gal β 1-4GlcNAc β 1-3)Gal- | |
| Blood group Sd(a), Cad | GalNAc β 1-4(sialyl α 2-3)Gal- | ↓ |
| Lewis ^a | Gal β 1-3(Fuc α 1-4)GlcNAc β 1-3Gal- | ↑↓ |
| Lewis ^b | Fuc α 1-2Gal β 1-3(Fuc α 1-4)GlcNAc β 1-3Gal- | ↑↓ |
| Lewis ^x | Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal- | ↑↓ |
| Lewis ^y | Fuc α 1-2Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal- | ↑↓ |
| Sialyl-Lewis ^x (SLe ^x) | Sialyl α 2-3Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal- | ↑↓ |
| Sialyl-dimeric Lewis ^x | Sialyl α 2-3Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal- | ↑↓ |
| 6-Sulfo-sialyl-Lewis ^x | Sialyl α 2-3Gal β 1-4(Fuc α 1-3)(6-sulfo)GlcNAc β 1-3Gal- | |
| 6'-Sulfo-sialyl-Lewis ^x | Sialyl α 2-3(6-sulfo)Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal- | |
| 3'-Sulfo-Lewis ^x | 3-Sulfo-Gal β 1-4(Fuc α 1-3)GlcNAc- | |
| MECA-79 epitope | Gal β 1-4(6-sulfo)GlcNAc β 1-3Gal β 1-3GalNAc- | |
| HNK-1 epitope | 3-Sulfo-GlcA β 1-3Gal β 1-4GlcNAc β - | |

Fuc, fucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; GlcA, glucuronic acid; Sialyl, sialic acid.

and both peptide and carbohydrate antigens of MUC1 are often abnormal in cancer.¹⁸ This could be exploited in the diagnosis and prognosis and for the monitoring of the disease process and for use of mucin epitopes as a cancer vaccine.¹⁹ MUC1 has also been found on T and B cells^{9,10} with a possible role in autoimmunity.

The transmembrane mucin MUC4 forms a complex with the cell surface EGF-receptor ErbB2 on the cell surfaces of mammary cells and regulates cell signaling;¹¹ it promotes the phosphorylation and activation of ErbB2.^{20,21} The interaction between MUC4 and ErbB3 is also important for receptor activation in signet ring carcinoma cells.²² The expression of MUC4 is significantly increased in pancreatic cancer and is related to cell proliferation and invasive properties of cancer cells.¹²

6.11.3 Mucin-Type O-Glycans

All mucin-type O-glycans have GalNAc as the first sugar α -linked to Ser or Thr of the mucin or glycoprotein backbone (Table 2), and in mammals can contain GlcNAc, Gal, Fuc, sialic acid, GalNAc, and sulfate esters. If no further additions occur, GalNAc is exposed and has been named the Tn antigen; this occurs in some cancer cells and other conditions (Table 3), but in normal tissues very few unmodified GalNAc residues are found.

Table 3 Abnormalities of O-glycosylation enzymes associated with pathology

| Transferase | Species | Change | Phenotype | Reference(s) |
|------------------------------|---------|--------|---|--------------|
| ppGalNAc-T3 | Human | ↓ | Subcutaneous familial tumoral calcinosis | 23 |
| | | ↑ | Colon cancer | 24 |
| Core 1 β 3-Gal-T | Mouse | ↓ | Brain hemorrhage | 25 |
| | Human | ↓ | Tn syndrome, missing Cosmc | 26 |
| | Human | ↓ | IgA nephropathy, missing Cosmc | 26 |
| Core 2 β 6-GlcNAc-T | Mouse | ↓ | Reduced E- and P-selectin-dependent leukocyte rolling | 27 |
| | Human | ↑ | Leukemia | 28 |
| | Human | ↑ | Metastasis, tumor aggressiveness | 29–31 |
| | Human | ↑ | Wiscott–Aldrich syndrome | 32,33 |
| | Rat | ↑ | Diabetes heart | 34 |
| Core 3 β 3-GlcNAc-T | Human | ↓ | Gastrointestinal cancer | 35 |
| | Mouse | ↓ | Susceptibility to colitis | 36 |
| β 3GnT8 | Human | ↑ | Colon cancer | 37 |
| β 4GalT1 | Mouse | ↓ | Reduced P-selectin binding | 38 |
| | Human | ↓ | Impaired inflammation | 39 |
| Cad β 4-GalNAc-T | Human | ↓ | CDG type IIId, failure to thrive | 40 |
| | Human | ↓ | Cancer | 41 |
| | Human | ↓ | Gastrointestinal cancer | 42 |
| ST3Gal-I | Human | ↑ | Breast cancer | 43 |
| | Human | ↑ | Ovarian cancer | 44 |
| ST3Gal-IV | Mouse | ↓ | Bleeding disorders, thrombocytopenia | 45 |
| | Human | ↑ | Colon cancer | 46 |
| ST6GalNAc-II | Human | ↑ | Colon cancer | 46 |
| FUT1 | Human | ↑ | Colon cancer | 47 |
| FUT4 | Mouse | ↓ | Selectin binding | 48 |
| FUT7 | Mouse | ↓ | Defect in leukocyte rolling and neutrophil mobilization, selectin binding | 48 |
| GlcNAc6ST-1 + GlcNAc6ST-2 | Mouse | ↓ | Reduced lymphocyte homing and adherence to HEV | 49 |
| C-GlcNAc6ST | Human | ↓ | Macular corneal dystrophy type I, II | 50 |
| I-GlcNAc6ST-3 Gal3ST-2 | Human | ↓ | Loss of vision | 51 |
| | Human | ↓ | Adenocarcinoma | 52 |

T, transferase.

GalNAc can be modified by sialic acid in α 2-6 linkage to form the sialyl-Tn antigen, which is found in a restricted number of tissues such as salivary glands. Sialyl-Tn is commonly found in intestinal and several other cancer types and has been associated with a poor prognosis.⁵³

The addition of Gal in β 1-3 linkage to GalNAc forms O-glycan core 1 (Figure 2), also known as the cancer-associated T antigen. In most normal tissues, the T antigen is modified by sialic acids or blood group structures (Table 2). Core 1 can also be extended by alternating GlcNAc and Gal residues, which form the common extended backbone structures of O-glycans, N-glycans, and glycosphingolipids. O-glycan core 2 has an additional branch at the GalNAc residue (Table 2). Both the Gal and GalNAc arms of the core 2 structure can be extended and carry various types of sugars and epitopes. The branched core 2 structures are complex and bulky, compared to linear core 1 structures, and due to the rotation of the GlcNAc1-6 arm have the ability to occupy a large space and thus mask the underlying peptide structures.^{54,55} In a restricted number of tissues such as the intestines, GalNAc can be modified by GlcNAc to form the core 3 structure, which is common in colonic mucins.^{56,57} Core 4 structures are branched O-glycans that are also restricted in their occurrence and are found in intestinal and bronchial mucins. Core 5, having the GalNAc α 1-3GalNAc- linkage, has been reported in

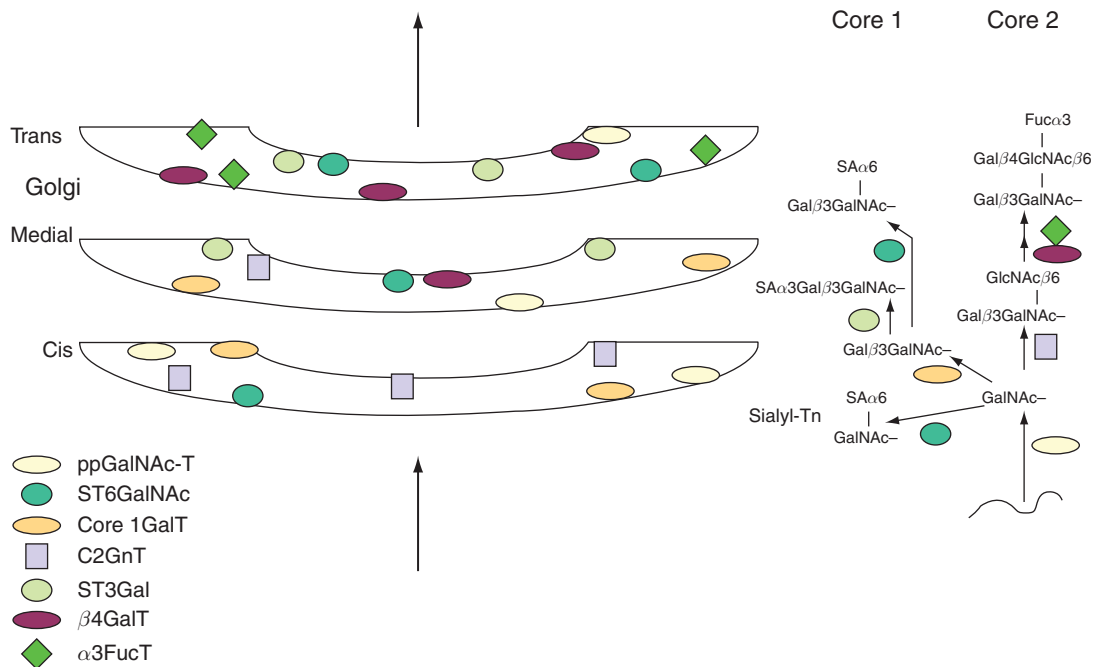


Figure 2 Pathway in the synthesis of O-glycan core 1 and core 2 structures with Golgi organization of the enzymes involved. The figure shows the relative distribution of some of the enzymes involved in O-glycosylation that have been localized to specific Golgi compartments. On the right, the O-glycan core 1 and 2 structures are depicted that are expected to be synthesized in the various compartments. Most enzymes have a fairly broad distribution; thus structures may also be synthesized in adjacent compartments. ppGalNAc-T, polypeptide GalNAc-transferase; ST6GalNAc, α 6-sialyltransferase acting on GalNAc-R; core 1 GalT, core 1 β 3-Gal-transferase; C2GnT, core 2 β 6-GlcNAc-transferase; ST3Gal, α 3-sialyltransferase; β 4GalT, β 4-Gal-transferase; α 3FucT, α 3-Fuc-transferase.

intestinal and fetal mucins as well as in adenocarcinoma.^{58,59} These O-glycans may therefore be onco-fetal structures. Core 6, GlcNAc β 1-6GalNAc-, occurs in human mucins, for example, those from ovarian cyst.⁶⁰ The unique core 7 structure, GalNAc α 1-6GalNAc-, has been found in the neutral fraction of bovine submaxillary mucin oligosaccharides.⁶¹ Core 8 is also extremely rare in mammals and has been found as sialyl α 2-6(Gal α 1-3)GalNAc- in respiratory mucins from patients with chronic bronchitis.⁶²

Blood group and tissue antigens are common at the nonreducing termini of O-glycans (Table 2). Some unusual antigens are found in duodenal and ovarian cyst mucins. In mammalian mucins, α -linked sugars are usually not substituted, with the exception of sialic acids and GalNAc-Ser/Thr. Thus the β -linked GlcNAc residue of core 6 can be extended, but the terminal α -linked sugars at the nonreducing end of core structures 5, 7, and 8 are not substituted. Mucin-type O-glycans from other species such as frogs⁶³ and *Caenorhabditis elegans*,⁶⁴ however, have additional interesting core structures as well as those resembling the mammalian types, and unusual extensions of O-glycans.⁶⁵

6.11.4 Functions of O-Glycans and Changes of O-Glycan Structures in Disease

Mucin functions have been attributed to the presence of O-glycans, which determine protein folding and conformation and the exposure of peptide and carbohydrate epitopes. The high proportion of O-glycans confers hydrophilic, adhesive, and viscous properties to mucins, and O-glycans protect a protein from proteolytic degradation. Adhesive properties of O-glycans are important in the immune system^{66,67} and may account for the binding of lectins to cell surfaces. For example, ligands for selectin interactions are found on O-glycans. In particular, O-glycan core 2 displays ligands for PSGL-1,¹³ although extended core 1 structures can also carry

these ligands⁶⁸ (Figure 1). Selectin-mediated interactions are important in the homing of lymphocytes and recruitment of leukocytes during the inflammatory response. Viruses and bacteria also bind to specific receptors found on O-glycans. For example, some *Escherichia coli* bacteria, *Plasmodium falciparum* parasites, as well as influenza virus all bind to sialyl α 2-3Gal structures present in the mucus gel or on cell surfaces. Extended chains of O-glycans are receptors for lung pathogens such as *Pseudomonas aeruginosa* and *Burkholderia cepacia*.⁶⁹

O-glycans undergo changes during development, and specific structures are required for normal development.¹ However, the roles of specific structures of O-glycans on mucins are generally not yet known, and a further insight into these highly conserved posttranslational modifications is needed. The functions of only a few specific structures have been determined. The core structures and the backbone are critical in determining the overall properties of O-glycans, their bulkiness and size, and the terminal epitopes. Type 1 or type 2 chains allow the addition of specific Lewis antigens. Sialyl-Lewis^x and Lewis^y determinants can only be found on the type 2 chain, while Lewis^a and Lewis^b occur on type 1 chains.

A number of structural abnormalities of mucins have been observed in disease.^{41,70–73} In cancer, the size and complexity of O-glycans are often reduced; thus, intermediate or truncated structures such as the Tn and T antigens appear. In breast cancer, complex, branched O-glycans are replaced by short and sialylated chains, which allows the peptide moiety of MUC1 to be exposed.⁴³ The expression of terminal antigens such as Lewis antigens can also be altered in cancer mucins and glycoproteins, and abnormal blood group antigens can be found.^{35,41,71,72,74}

In cystic fibrosis (CF), the function of the chloride channel (CFTR) is abnormal, causing disturbances in water and salt transport. As a consequence, the mucus gel in lungs, bronchi, and other tissues is highly viscous and harbors specific strains of bacteria (*P. aeruginosa* and *B. cepacia*), leading to lung infections and inflammation. Structural abnormalities of O-glycans have been reported in CF bronchial mucins and especially sialylation and sulfation are affected.⁷⁵ Similar structural alterations can be mimicked *in vitro* upon treatment of bronchial tissues with the inflammatory cytokine tumor necrosis factor (TNF)- α .⁷⁶ A number of cell types respond to TNF- α with alterations in the expression or the activities of glycosyltransferases.^{77–81} Brockhausen *et al.*⁸² have shown that the CFTR mutation is not directly related to altered O-glycan biosynthesis in the CF mouse model as well as in several cultured cells. In addition, Leir *et al.*⁸³ confirmed that mutated CFTR did not induce alterations of mucin glycosylation and sulfation in cultured airway epithelial cells. Therefore, the structural changes in CF mucins are probably due to the influence of inflammation in CF tissues.

O-glycans are found on mucins and glycoproteins involved in fertilization. Early experiments showed that O-glycans of the mammalian egg zona pellucida were involved in sperm–egg binding during fertilization.⁷⁴ However, mice having deletions in the candidate genes demonstrate that these interactions are not exclusively responsible for the process of fertilization.⁸⁴

A number of model systems have been developed to study O-glycan functions. The O-glycan extension inhibitor GalNAc α -benzyl (or GalNAc α -*p*-nitrophenyl) is a substrate for the enzymes that synthesize the core structures 1 and 3, and thus by competition blocks the extension of GalNAc– on endogenous glycoproteins. GalNAc α -benzyl also prevents sialylation of O-glycans, probably by providing alternative substrates for α 3-sialyltransferase. Thus, in HT29 cell cultures treated with GalNAc α -benzyl, sialylated and branched O-glycans attached to benzyl are found in the cell medium.⁸⁵ GalNAc α -benzyl reduces mucin O-glycosylation in cancer cells and decreases cell binding to E-selectin and E-selectin-expressing endothelial cells, which indicates that the binding is based on the recognition of O-glycans.⁸⁶ A UDP–GalNAc analogue has been described as an inhibitor of the first step in O-glycosylation.⁸⁷ However, the paper has been withdrawn in 2008 since it may be possible that the biological effects observed, such as the induction of apoptosis, were due to a metabolite of the relatively unstable UDP–GalNAc derivative.

Abnormalities in glycosyltransferase activities, nucleotide sugar synthesis and transport within the Golgi, and Golgi proteins involved in protein transport can be associated with pathology⁸⁸ (Table 3). Deficiencies in Fuc-transferases^{89,90} can lead to nonpathological structural changes in O-glycans. However, when fucosylation is generally deficient due to lack of GDP–Fuc transport into the Golgi, serious problems occur in the immune system, which are most probably due to the inability of selectins to function normally.⁹¹ When CMP–sialic acid transport into the Golgi is defective (e.g., in congenital disease of glycosylation, CDG–II_f), patients may suffer from myopathy,⁹² developmental problems, hemorrhage, and infections, and die in infant stage.⁹³ Multiple pathological changes are also seen when subunits of the conserved oligomeric Golgi (COG) complex are

mutated.⁹⁴ For example, COG8 mutations in CDG-IIIh appear to specifically affect a galactosyltransferase and result in developmental problems, retardation, and seizures.⁹⁵ Deficiency of COG7 protein⁹⁶ is associated with mislocalization of a number of glycosyltransferases, as well as prenatal growth retardation, multisystemic disorders, and death in infancy.

6.11.5 Biosynthesis of O-Glycans

In contrast to N-glycosylation, which requires the presence of a peptide sequon and preassembly of a complex oligosaccharide on a dolichol lipid carrier, mucin-type O-glycans are assembled by the sequential addition of individual sugars transferred from nucleotide sugars by glycosyltransferases in the Golgi.^{41,73,97–99} There are no known trimming reactions involved catalyzed by glycosidases, comparable to those in the N-glycosylation pathway, and no peptide sequons have been established as O-glycosylation signals. The known glycosyltransferases of the O-glycosylation pathways (Table 4) have type II membrane topology with their catalytic domain in the Golgi lumen. The nucleotide sugar donor substrates are transported into the Golgi by specific transport

Table 4 Mammalian glycosyltransferases involved in the synthesis of mucin-type O-glycans

| Enzyme | Short name | EC | CAZy GT | Features |
|--|----------------------|-----------|---------|--------------------|
| <i>Transferases specific for O-glycans</i> | | | | |
| Polypeptide GalNAc-T | ppGalNAc-T | 2.4.1.41 | 27 | Lectin domain, DxH |
| Core 1 β 3Gal-T | C1GalT | 2.4.1.122 | 31 | DxD |
| Core 3 β 3-GlcNAc-T | C3 β 3GnT6 | 2.4.1.147 | 31 | DxD |
| Core 2 β 6-GlcNAc-T (L) | C2GnT1 | 2.4.1.102 | 14 | DxD |
| | C2GnT3 | 2.4.1.102 | 14 | DxD |
| Core 2/core 4 β 6-GlcNAc-T (M) | C2GnT2 | 2.4.1.148 | 14 | DxD |
| α 3-Sialyl-T | ST3Gal-I, ST3Gal-IV | 2.4.99.3 | 29 | Sialylmotifs |
| α 6-Sialyl-T | ST6GalNAc-I | 2.4.99.3 | 29 | Sialylmotifs |
| | ST6GalNAc-II | 2.4.99.3 | 29 | Sialylmotifs |
| | ST6GalNAc-III | 2.4.99.7 | 29 | Sialylmotifs |
| | ST6GalNAc-IV | 2.4.99.7 | 29 | Sialylmotifs |
| Elongation β 3-GlcNAc-T | β 3GnT3 | 2.4.1.146 | 31 | DxD |
| Gal-3-O-sulfo-T | Gal3ST-4 | 2.8.2.11 | | |
| <i>Transferases with broader specificity for different glycans</i> | | | | |
| β 4Gal-T | β 4GalT | 2.4.1.22 | 7 | DxD |
| | | 2.4.1.38 | | |
| | | 2.4.1.90 | | |
| β 3Gal-T | β 3GalT | | 31 | DxD |
| β 3-GlcNAc-T | iGnT, β 3GnT | 2.4.1.149 | 31 | DxD |
| β 6-GlcNAc-T | IGnT | 2.4.1.150 | 14 | DxD |
| α 2Fuc-T | FUT1, FUT2 | 2.4.1.69 | 11 | |
| α 3/4Fuc-T | FUT3 | 2.4.1.65 | 10 | DxD |
| α 3Fuc-T | FUT4-9 | 2.4.1.152 | 10 | |
| Blood group A α 3GalNAc-T | A transferase | 2.4.1.40 | 6 | |
| Blood group B α 3Gal-T | B transferase | 2.4.1.37 | 6 | DxD |
| Linear B α 3Gal-T | Linear B transferase | 2.4.1.151 | 6 | DxD |
| Cad β 4-GalNAc-T | | | 12 | DxD |
| β 3-GalNAc-T | | | 31 | DxD |
| β 4-GalNAc-T | | | | DxD |
| α 4-GlcNAc-T | | | | DxD |
| α 3-Sialyl-T | ST3Gal | 2.4.99.3 | 29 | Sialylmotifs |
| GlcNAc-6-sulfotransferase | GlcNAc6ST | 2.8.2.21 | | |
| Gal-3-sulfotransferase | Gal3ST | 2.8.2.11 | | |
| Gal-6-sulfotransferase | Gal6ST | 2.8.2.21 | | |

-T, -transferase.

systems. Sulfotransferases that add terminal sulfate ester groups to Gal or GlcNAc residues are similarly localized in the Golgi as type II membrane proteins and utilize 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as the sulfate donor substrate.¹⁰⁰

Glycosyltransferases have been classified into 91 families, based on structural and mechanistic characteristics (<http://afmb.cnrs-mrs.fr/CAZy>). Retaining glycosyltransferases retain the anomeric sugar linkage of the nucleotide sugar donor in their enzyme product, while inverting transferases invert the anomeric linkage. The glycosyltransferases involved in O-glycosylation are stereo- and linkage specific (Table 4). Unlike protein synthesis, several possible attachment points can form a linkage between sugars creating a vast theoretical number of potential sugar linkages. However, owing to the distinct specificities of transferases for their acceptor substrates, only a few hundreds of different structures are found in nature. Enzyme products often form the substrate for the next step unless a sugar is added that will block further conversions. For example, sialic acid residues can block the extension of chains and thus keep the chain lengths short. Different enzymes may compete for the same substrate; thus the relative activities dictate the relative distribution of structures. The Golgi organization may prevent or allow this competition, according to the localization of enzymes in the assembly line (Figure 2). Few glycosyltransferases are absolutely specific for one acceptor substrate; usually, there is some overlap in specificity and enzymes can synthesize linkages on various cores and extended structures, and can even act on N-glycans and glycolipids as well as O-glycans (Table 4).

Among the critical control factors for the biosynthesis are tissue-specific transcription of specific exons and the use of promoters, as well as the influence of cytokines that can significantly alter the expression levels and activities of individual enzymes.⁸¹ It is likely that glycosyltransferases in the Golgi form complexes with other proteins or form dimers or oligomers, which can affect their activities.^{101–103} Core 1 β 3-Gal-transferase binds to the specific chaperone Cosmc in the endoplasmic reticulum, which stabilizes the protein. Another specific glycosyltransferase-binding protein is α -lactalbumin in the mammary gland, which binds to β 4-Gal-transferase and changes its acceptor specificity upon binding. Metal ions, membrane components, and posttranslational modifications of enzymes have also been shown to regulate activities.⁷³

All of the glycosyltransferases that assemble O-glycans have a short cytoplasmic tail at the N terminus that protrudes into the cytoplasm, a transmembrane domain, and the globular catalytic domain in the lumen of the Golgi. Both the cytoplasmic tail and the transmembrane domain of transferases appear to contain the most important factors determining localization. Since no common sequences can be detected in these domains,¹⁰⁴ the localization signals may be three-dimensional or depend on the hydrophobic and other properties of peptides. Glycosyltransferases are generally arranged in an assembly line fashion with early-acting enzymes in the cis Golgi and late-acting enzymes in trans Golgi compartments, with the glycoproteins flowing from the cis to the trans Golgi network. Generally, glycosyltransferases involved in chain extension and termination localize to the trans Golgi and trans Golgi network (Figure 2). However, some enzymes appear to be broadly distributed. The localization of individual enzymes may depend on the cell type and may be altered during cell differentiation and in cancer cells.¹⁰⁵ The initiation of O-glycosylation by polypeptide GalNAc-transferase was shown to occur in the cis Golgi compartments.¹⁰⁶ However, different polypeptide GalNAc-transferase isoenzymes have been found to be distributed throughout the Golgi in cervical cancer HeLa cells.¹⁰⁷ This phenomenon may explain the heterogeneity of structures in mucin oligosaccharides. If the first sugar is added in a late compartment, then early localized enzymes cannot extend the chain, and as a result underprocessed, 'truncated' structures are found. Consistent with an assembly line, the extension enzyme β 4-Gal-transferase localizes to late (trans) compartments, together with sialyltransferases and Fuc-transferases.^{108,109} Factors that may disturb the assembly line of glycosyltransferases are expected to cause alterations in the final structures. For example, alterations in the Golgi pH have been suggested to cause a redistribution of glycosyltransferases.¹¹⁰

6.11.6 Structures and Mechanisms of Glycosyltransferases

The mammalian glycosyltransferases can be grouped into two structural superfamilies, GT-A or GT-B, based on their expected folding patterns.¹¹¹ All of the inverting and retaining glycosyltransferases involved in O-glycosylation that have been crystallized to date exhibit a type A fold consisting of an $\alpha/\beta/\alpha$ sandwich

with the central β -sheet having several strands. The sugar–nucleotide and acceptor substrate-binding sites as well as the bound divalent metal ion are in close proximity, and nucleotide-binding sites appear to have a similar fold. Enzymes within a GT family may have a similar catalytic mechanism. The inverting transferases appear to act by a single displacement mechanism with the acceptor serving as the nucleophile. *In vitro* kinetic studies showed that glycosyltransferases act mostly with a sequential mechanism where the nucleotide sugar binds first, then the acceptor, and after transfer of the sugar UDP is released last. This mechanism has been supported in a limited number of enzymes crystallized with and without the nucleotide sugar.

A common motif found in many glycosyltransferases is the DxD sequon, which can act as a nucleophile and catalytic base. Transferases often require divalent metal ions for activity, which may have the role of complexing the negatively charged phosphate groups of the donor substrate or may interact with the DxD motif. However, basic groups in the protein may also serve to complex the negative charge of the donor substrate in those transferases not requiring divalent metal ions. Crystal structures revealed that a conformational change is induced in the enzyme upon binding of the donor substrate. This causes the movement of a flexible disordered loop over the acceptor-binding site that allows the acceptor substrate to bind. After the sugar transfer reaction has taken place, the nucleotide is released.

Glycosyltransferases also appear to have extended acceptor-binding sites that allow a variety of different substrates to bind, provided that they have specific structural features. Many of the glycosyltransferases also have posttranslational modifications such as N-glycosylation, which may be required for their stability, secretion, and activity.¹¹² The first enzyme in the pathway, polypeptide GalNAc-transferase, is unique in that it has a separate lectin-like domain, which may have a role in determining the affinity to already glycosylated acceptor substrates.

6.11.7 Initiation of O-Glycosylation

O-glycosylation is initiated in the Golgi apparatus by a family of highly conserved polypeptide α -GalNAc-transferases (ppGalNAc-T, EC 2.4.1.41) that belong to the GT27 family.^{113–115} The transfer of a GalNAc residue to the hydroxyl of Ser or Thr within suitable amino acid sequences is always the first step of O-glycan biosynthesis. The single GalNAc residue linked to Ser/Thr is named the Tn antigen, which occurs especially in advanced tumors, and is a marker for poorly differentiated carcinomas associated with invasive properties and a poor clinical outcome.⁵³ The enzyme in porcine and bovine submaxillary glands was shown by immunochemistry to localize to the cis Golgi.¹⁰⁶ Later studies showed that in HeLa cells, ppGalNAc-T2 and T3 preferentially localize to the medial and trans Golgi compartments,¹⁰⁷ which may explain the cancer-specific appearance of Tn antigens.

More than 15 human ppGalNAc-Ts have been described. Members of the ppGalNAc-T family are probably found in all eukaryotic cells with a cell type-specific expression. Orthologous genes have been described in *Drosophila melanogaster*, *C. elegans*, *Toxoplasma gondii*,¹¹⁶ and other species. The gene expression of the various members of the ppGalNAc-T family is regulated in a developmental- and tissue-specific fashion in *Drosophila*.¹¹⁷ Mutagenesis and complementation of specific members of the family showed that the enzymes support development and viability in *Drosophila*.^{114,115,117} However, a deletion of several of the ppGalNAc-T in mice did not cause distinct phenotypes.^{113,118,119} This suggests that the apparent redundancy of these enzymes and O-glycosylation is essential for survival.

Several disease states have been reported with abnormal expression of ppGalNAc-T. In colorectal carcinoma tissues, the expression levels of ppGalNAc-T1, T2, and in particular of T3, have been shown to be increased.^{24,41,73} T3 may be an important enzyme for the homeostasis of tissues. In well-differentiated colorectal tumors, the levels of ppGalNAc-T3 enzyme protein are indicators of a good prognosis.¹²⁰ In a human condition called familial tumoral calcinosis characterized by a massive accumulation of calcium phosphate crystals, several mutations of ppGalNAc-T3 have been identified.²³ T3 has an unusual ability to add GalNAc to Thr178 of fibroblast growth factor-23 (FGF23) near a protease cleavage site. Thus the absence of T3 activity affects the O-glycosylation of FGF23, and leads to a decrease in its secretion.¹²¹ Using immunocytochemistry with confocal scanning laser microscopy, ppGalNAc-T1 and T2 were found to be expressed in all of the breast

cancer cell lines examined.¹²² In contrast, T3 and T6, which were only weakly expressed in benign breast epithelial cells, were readily detectable in malignant breast cancer cells. This indicates a differential expression of specific members of the ppGalNAc-T family in breast cancer.

6.11.7.1 Structures of Polypeptide GalNAc-Transferases

ppGalNAc-T are unique transferases that have a ricin B chain-like lectin domain at the C terminus that is separate from the catalytic domain.^{123,124} These lectin-like repeat sequences (α,β,γ) form a trefoil structure and may bind as many as three sugar residues.^{125–127} Amino acids conserved in ppGalNAc-Ts include a DxH motif in the first half of the catalytic domain, and Cys residues. Cys212 and 214 are conserved among T1, T2, T3, T4, and T6. In T1 these Cys residues have free sulfhydryl groups and may interact with bound UDP.^{113,125}

The crystal structures of ppGalNAc-T1, T2, and T10 show that the catalytic and lectin domains are distinct. It has been possible to model the enzyme containing a glycopeptide derived from the MUC1 VNTR (**Table 1**) with 20 amino acids and GalNAc at Thr14 that extends from the acceptor-binding site to an extended binding site in the lectin domain.¹²³ Mouse ppGalNAc-T1 has been crystallized in a Mn^{2+} complex as a soluble protein where the N-terminal transmembrane domain and cytoplasmic tail were replaced with a maltose-binding protein. The structure showed the Mn^{2+} ion complexed between the lectin and the catalytic domains but UDP-GalNAc was not seen. The protein has a central eight-stranded β -sheet flanked by α -helices (GT-A fold). The nucleotide-binding domain (SGC domain) is separate from the lectin domain.^{123,125–127} A deletion of individual lectin domains abolished ppGalNAc-T activity toward deglycosylated bovine submaxillary mucin indicating that the lectin is an essential part of the enzyme.^{125,126}

The catalytic domain of ppGalNAc-T1 is folded through several disulfide bonds, conferring stability to the protein. Four conserved Cys residues form disulfide bonds, while free sulfhydryl groups of Cys212 and Cys214 appear to be involved in UDP-GalNAc binding and/or catalysis.^{125,126} The conserved DxH motif (Asp209 and His211 in T1) is found in a deep UDP-GalNAc-binding pocket, while the conserved His344 is involved in complexing the Mn^{2+} ion. Mutagenesis showed that these residues are essential for the activity of the bovine T1.¹²⁸ Although the crystal structure did not include UDP-GalNAc or acceptor, Glu127 could possibly be involved in binding the phosphate oxygens of UDP-GalNAc through intervening water molecules.¹²³ Mutations of aromatic residues Trp316 and highly conserved Trp328 in rat ppGalNAc-T1 led to a complete inactivation, although replacement of Trp316 with another aromatic amino acid retained low activity. Since the binding of UDP-GalNAc was not affected, these Trp residues are essential for the binding of the acceptor substrate.¹²⁹

Human ppGalNAc-T2 has a similar structure to that of ppGalNAc-T1.¹³⁰ The enzyme has been crystallized with UDP and the acceptor peptide PTTDSTTPAPTTK having seven potential O-glycosylation sites, derived from the VNTR of rat submandibular mucin. In the presence of bound UDP, peptides of up to nine amino acid residues in linear conformation can be accommodated in the extended acceptor-binding site.¹²³ In this acceptor peptide, Thr7 is likely the site of O-glycosylation.¹³¹ A comparison of the crystal structures with and without acceptor shows that a lid is formed over the UDP-binding site in the acceptor-bound form. In the absence of the acceptor, the UDP-binding cleft is open, which enables UDP to leave and UDP-GalNAc to bind. The DxH motif and His359 are involved in binding Mn^{2+} as well as the phosphate groups of UDP. The catalytic mechanism as suggested by X-ray crystallography supports the kinetic mechanism of T1¹³² as a random ordered sequential mechanism.

Crystals of a soluble form of human T10 were also grown in the presence of Mn^{2+} and UDP-GalNAc.¹³³ While the molecular shape of T1 is relatively straight, T10 assumes an L shape. UDP-GalNAc was apparently hydrolyzed and free UDP and GalNAc were seen bound to the enzyme. Mn^{2+} coordinates the phosphate groups of UDP as well as Asp237, His239, and His370. The conformations of the two enzyme forms, grown with and without UDP-GalNAc, differ. The cleft linking catalytic and lectin domains can accommodate the acceptor substrate. An exchange of this linker region with that of T1 resulted in activity toward nonglycosylated peptides and revealed that this linker is important in regulating the acceptor specificity.¹³³

6.11.7.2 Site Specificity of Initial O-Glycosylation

There is no specific sequon required for O-glycosylation but the presence of Pro in specific positions near Thr or Ser appears to favor the addition of GalNAc. The rate of O-glycosylation probably depends on the exposure of Thr and Ser residues, but also on the amino acids and existing glycosylation in the vicinity of Thr/Ser. The different isoenzymes of ppGalNAc-T differ slightly in their acceptor specificities and may be complementary in their actions. Since usually several ppGalNAc-Ts are expressed in a particular cell, their overlapping specificities ensure near-complete O-glycosylation in mucins and other glycoproteins.

For example, T4, T6, and T10 show a distinct preference for glycopeptide acceptors, while T1 and T2 also act on nonglycosylated peptides.¹³⁴ IgA is O-glycosylated in the hinge region. ppGalNAc-T1, T2, T3, T4, T6, and T9 are expressed in human IgA-producing B cells but not T7, T8, T10, and T12.¹³⁵ Compared to the other transferases, T2 has a preference for the hinge region of IgA and can glycosylate eight of the nine Ser/Thr residues and form multiply O-glycosylated IgA molecules.

Rat ppGalNAc-T6 is expressed mainly in the salivary gland and mucin-secreting intestinal tissues. T6 was found to have insignificant activity toward unglycosylated MUC5AC-derived peptides, but it has activity toward GalNAc-glycopeptides.¹³⁴ ppGalNAc-T4 is widely expressed, especially in mucin-secreting tissues such as salivary glands and intestine. The enzyme has a more restricted acceptor specificity compared to T1 and prefers peptide derived from rat submandibular gland mucin as the acceptor substrate.¹³⁶ A ppGalNAc-T4 construct lacking the lectin domain was found to be active although it had a higher K_M toward glycopeptide substrates. The presence of high concentrations of GalNAc reduced the ability to act on glycopeptides, indicating that the lectin domain regulates the enzyme specificity toward glycopeptide substrates.¹²⁷

The acceptor specificity of bovine ppGalNAc-T1 toward short peptides and glycopeptides has been extensively studied¹³⁷ (Brockhausen *et al.*, unpublished). The amino acids and the presence and the structure of oligosaccharides in the vicinity of the glycosylation site in the acceptor substrate were found to have a significant influence on the rate of GalNAc transfer. Existing glycosylation generally decreases the T1 activity. The position of Pro in the acceptor is crucial, and the activity decreases by 90% if Pro in the +3 position is substituted with Ala.

Thus, in order to understand O-glycosylation accurately, more information on the substrate recognition by the different ppGalNAc-Ts has to be obtained, together with knowledge of the amino acid sequence, glycosylation, and conformation of acceptor substrates. We also need to know the ratio of expression of isoenzymes having different specificities in the cell type producing the O-glycosylated protein.

6.11.8 Synthesis of Core 1

Most mammalian cell types can modify the GalNAc residue by the addition of Gal, to form core 1 (T antigen, **Table 2**), which in the unmodified form is not normally found in glycoproteins, with the exception of antarctic antifreeze glycoprotein. Core 1 synthesis is catalyzed by core 1 β 3-Gal-transferase (C1GalT, EC 2.4.1.122), an inverting transferase and member of the GT31 and β 3Gal-T family (β 3Gal-T7). The enzyme has been purified from rat liver and is unusually stable;¹³⁸ it has a DxD motif and requires Mn^{2+} for activity. The rat liver enzyme exists mainly as a monomer but also as a homodimer, stabilized by disulfide bonds, and both forms are active.¹³⁹

In normal cells, core 1 is further processed by the addition of sialic acid and blood group determinants, or by extension or branching reactions (**Figure 2**). The T antigen is often abnormally expressed on cancer cells and on intestinal cells from patients with inflammatory bowel disease, and binds Peanut lectin, which has been shown to stimulate cell proliferation.¹⁴⁰ The overexpression of T antigen may be due to the lack of enzymes that further process the T antigen.

Specificity studies of the rat liver C1GalT have shown that the enzyme is specific for GalNAc acceptor substrates and acts on GalNAc linked to peptide or hydrophobic aglycone groups. The enzyme requires the 2-acetamido group as well as the 3- and 4-hydroxyls but not the 6-hydroxyl of GalNAc, and can act on GlcNAc β 1-6GalNAc-substrates.¹³⁸ This means that the core 6 structures can be converted to core 2 by C1GalT. However, C1GalT does not act on sialyl α 2-6GalNAc-mucin and cannot convert the sialyl-Tn

antigen to the sialyl-T antigen, sialyl α 2-6(Gal β 1-3)GalNAc α -Ser/Thr. Using a series of synthetic glycopeptides as substrates, we have shown that C1GalT acts in a site-directed manner. Thus, the activity of C1GalT toward glycopeptide acceptor substrates is influenced by the types of amino acids near the O-glycosylation site of the acceptor, as well as by the proximity of other O-glycans, depending on their structure, size, and numbers.^{141,142} This suggests that the synthesis of both Tn and T antigens is controlled by the structure of the peptide acceptor near the O-glycosylation site as well as by existing glycosylation. The site specificity of di- and trisaccharides in the VNTR of blood group A negative (core 1-containing) porcine submaxillary mucin has been determined by Gerken *et al.*¹⁴ Chemical degradation could distinguish between substituted and nonsubstituted GalNAc. They found that the occurrence of core 1 is dependent on the peptide environment and is inversely correlated to the density of GalNAc residues. In contrast, α 2-Fuc substitution does not appear to depend on GalNAc density.

Because GalNAc α -benzyl is an excellent substrate for C1GalT and penetrates into cells in culture, it can serve as an alternative substrate for C1GalT in the Golgi, and by competition reduces core 1 synthesis on natural glycoproteins. A number of sugars can be added to the benzyl primer, and sialylated complex O-glycans can be built up, which reduces the sialylation and the overall glycosylation of mucins.^{85,86} This inhibitor has been used to show that O-glycans have a role in carrying selectin ligands, which regulate the adhesive properties of endometrial adenocarcinoma cells,¹⁴³ as well as the invasive and metastatic behavior of human colon cancer cells.¹⁴⁴

The gene encoding C1GalT activity is widely expressed in mammalian tissues, *D. melanogaster*, *C. elegans*, and other species. Cosmc is an unusual chaperone that associates with C1GalT when coexpressed in the endoplasmic reticulum, prevents it from degradation, and ensures its Golgi localization and activity. The *cosmc* gene has been identified in humans and mice¹⁴⁵ but *D. melanogaster* does not appear to have a *cosmc* gene.¹⁴⁶ A deletion of the C1GalT gene in mice leads to brain hemorrhage and a chaotic microvascular network in embryos, which died at day 14.²⁵ This suggests that C1GalT is essential for the development of the vascular system and may play a crucial role in angiogenesis. A new mouse model was established for studying core 1 function by *N*-ethyl-*N*-nitrosourea treatment of mice that led to a severe reduction in C1GalT activity.¹⁴⁷ Mice were bred as *plt1/plt1* mice and the C1GalT deficiency was transmitted in a recessive mode. In the mutant C1GalT, Tyr321 was replaced with Asn, leading to a significant reduction in activity. The mutant enzyme was still bound to Cosmc and depended on this interaction for residual activity. Thrombocytopenia and kidney disease were associated with the deficiency in core 1 synthesis in these mice. The most affected glycoproteins were platelet glycoprotein gp1b α and mucin-like podocalyxin expressed in kidney, which suggests an important role of core 1 in platelet production and kidney function.

Oocyte-specific deletion of C1GalT in the mouse led to an increase in fertility, and showed that core 1 O-glycans may have a role in the suppression of fertilization and regulation of follicular integrity.¹⁴⁸ The binding of C1GalT-deficient eggs to sperm was not affected,⁸⁴ showing that zona pellucida does not require its core 1-based O-glycans for sperm-egg binding.

There are several other human conditions associated with the lack of core 1 O-glycans. In IgA nephropathy and Henoch-Schoenlein purpura nephritis, core 1 is lacking in the hinge region of IgA, and this is associated with immune complex formation, kidney disease, and tonsillitis.^{149,150} In IgA nephropathy, Cosmc is expressed in B lymphocytes in very low amounts. Human T-lymphoblastoid cell line Jurkat and the human colon cancer cell line LSC also express truncated O-glycans and lack C1GalT activity.¹⁵¹ Both cell types express C1GalT mRNA, and mutations in the *cosmc* genes account for the loss of C1GalT activity in these cells.¹⁴⁵ The missing C1GalT activity can be restored with a functional *cosmc* gene.¹⁵² Another human condition, the Tn syndrome, is associated with somatic mutations in the *cosmc* gene.²⁶ The Tn expression in colonic tumors may also be due to a reduction in Cosmc function.¹⁵²

6.11.9 Synthesis of Core 2

Unmodified core 1 structures can be converted to branched core 2 structures (Table 1, Figure 2) by core 2 β 6-GlcNAc-transferase (C2GnT) (EC 2.4.1.102). Although the enzyme is expressed in only specific cell types, it is present in most mucin-secreting cells such as those in the intestine and bronchi. In nonmucin-producing

cell types such as leukocytes, endothelial cells, and bone cells, the activity is relatively low. The core 2 branch provides an additional backbone structure for the attachment of antigenic determinants and cell adhesion ligands such as sialyl-Lewis^x and blood group epitopes. Because of the flexibility of the chain attached to the β 1-6-linked GlcNAc residue and the additional branch, core structures are bulky and can mask the underlying peptide antigens.^{54,153} In some cultured mammary carcinoma cells, C2GnT activity is lacking, and this correlates with a reduction in O-glycan size and the exposure of MUC1 mucin epitopes. Therefore, C2GnT has an important function in determining the overall properties of O-glycans.

Three different C2GnT enzymes catalyze the synthesis of core 2 in mammalian tissues (Table 4). The enzyme described in leukocytes (C2GnT1, the L-type enzyme) is expressed most commonly and synthesizes only core 2. Another L-type enzyme, C2GnT3, is related, has similar specificity, and is expressed primarily in the thymus, with very low expression in the intestine and peripheral blood leukocytes.¹⁵⁴ A third enzyme, C2GnT2, is found mainly in mucin-expressing tissues.¹⁵⁵ C2GnT2 has a less restricted specificity; it can synthesize core 2 from core 1, core 4 from core 3, and can also add other GlcNAc β 1-6 branches to O-glycans.^{156,157} A kidney-specific form of C2GnT1 encoded by the *gsl5* gene was identified in mice; the enzyme may function in the synthesis of GlcNAc β 1-6(Gal β 1-3)GalNAc- structures in glycolipids.¹⁵⁸ C2GnT1 is encoded by one exon; there are differences in the 5'-region between the mRNA from different tissues, suggesting a tissue-specific alternate use of promoters.¹⁵⁹ C2GnT1 expression appears to be regulated by the transcription factor Sp1.¹⁶⁰

The activity of C2GnT was first detected in dog submaxillary glands,¹⁶¹ and does not require the presence of divalent cations. The activity is highly regulated during T-cell activation.¹⁶² Leukemia cells from patients with chronic and acute leukemia were shown to have high activities of the enzyme, whereas in normal granulocytes the activity is relatively low.²⁸ Activation of lymphocytes from patients with Wiscott–Aldrich syndrome, however, led to a decreased C2GnT activity likely due to a failure of normal lymphocyte maturation. Thus, lymphocytic surface sialoglycoprotein CD43 carries more core 2 structures in these patients due to a hyperactive C2GnT.^{32,33}

C2GnT1 localizes to the cis and medial Golgi compartment in mammary cells¹⁶³ where it competes with the α 3-sialyltransferase that also acts on core 1 substrate. A deletion of specific domains of C2GnT1 shows that in Chinese hamster ovary (CHO) cells the cytoplasmic tail and the membrane anchor regions are sufficient for the cis Golgi localization of C2GnT1.¹⁶⁴ By insertion of the membrane anchor of the trans Golgi enzyme α 6-sialyltransferase, C2GnT1 can be localized to the trans Golgi in CHO cells.¹⁶⁵

6.11.9.1 Structure and Specificity of C2GnT1

The specificities of C2GnT1 and 2 have been studied using core 1 derivatives as acceptor substrates that lack specific hydroxyl groups of both the Gal and GalNAc residues.^{166,167} The results show that human C2GnT1 has an absolute requirement for the 4- and 6-hydroxyls of Gal and GalNAc as well as the 2-acetamido group of GalNAc, while C2GnT2 from rat colon has a broader specificity and has an absolute requirement only for the 4- and 6-hydroxyls of GalNAc and the 6-hydroxyl of Gal. This helps to explain why C2GnT2 can use core 1, core 3, or GlcNAc β 1-3Gal- as acceptor substrates and thus can synthesize core 2, core 4, and GlcNAc β 1-6(GlcNAc β 1-3)Gal- (the big I antigen branch), respectively. In contrast, unsubstituted core 1 is the exclusive substrate for C2GnT1^{166,168} (Figure 3).

The structure of soluble mouse C2GnT1 has been determined in complex with and without the acceptor substrate Gal β 1-3GalNAc-R.¹⁶⁹ The crystallized protein was a dimer, linked through disulfide bonds formed by Cys235. However, the recombinant soluble active enzyme expressed in insect cells¹⁶⁷ and in CHO cells is mainly monomeric, while the membrane-bound form occurs as a dimer inside CHO cells.¹⁰¹

C2GnT1 is a member of the CAZy GT14 family. The catalytic domain of mouse C2GnT1 consists of an α / β / α sandwich (GT-A fold) with a central six-stranded mixed β -sheet with 3,2,1,4,6,5 topology.¹⁶⁹ The stem region of C2GnT1 contains α -helices and its two conserved N-glycosylation sites. The two N-glycans were shown to be important for the stability of the human enzyme expressed in insect cells.¹¹² In addition, the N-glycan at Asn95 was shown to contribute to the cis/medial Golgi localization of the full-length human enzyme.¹⁷⁰ The N-glycans of the mouse enzyme were difficult to determine in the crystal structure due to their flexibility.¹⁶⁹

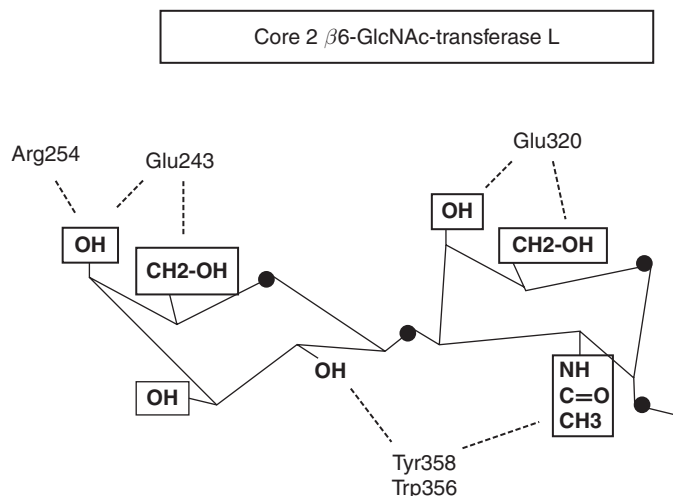


Figure 3 Specificity of C2GnT1. The recognition of the core 1 acceptor substrate by human C2GnT1 (core 2 β 6-GlcNAc-transferase L type) is shown, based on specificity studies of the human enzyme using acceptor analogues, as well as the crystal structure of the mouse C2GnT with the acceptor Gal β 1-3GalNAc-. The bold black boxes indicate a strict requirement for the 4- and 6-hydroxyls of both Gal and GalNAc, and the *N*-acetyl group of GalNAc. These sugar substituents appear to have close contact with amino acids Glu320, Glu243, and Arg254 in the mouse enzyme acceptor-binding site, as indicated. The 3-hydroxyl of Gal is only partially required but the 2-hydroxyl of Gal may be more important since it seems to be in contact with Tyr358 of the conserved WxY sequence in C2GnT1. The M type C2GnT2 has a broader specificity and the enzyme can accommodate GlcNAc β 1-3Gal- as well as GlcNAc β 1-3GalNAc-. The M enzyme from rat colon also has a strict requirement for the 4- and 6-hydroxyls of GalNAc and the 4-hydroxyl of Gal, and a partial requirement for other substituents.

The activity of C2GnT1 does not require divalent metal ions. Thus, the proposed UDP-GlcNAc-binding site in the crystal structure suggests that the pyrophosphate group of UDP-GlcNAc is not complexed with a metal ion but instead may be coordinated with basic conserved amino acids Arg378 and Lys401.

Human C2GnT1 has nine conserved Cys residues, and mutational studies showed that eight of these are essential for the activity.¹⁶⁷ The ninth residue, Cys217, is not essential and is present in the sulfhydryl form in the active enzyme. This may explain why enzyme preparations stored for extended periods of time have low activities but can be activated with reducing agents such as mercaptoethanol. The presence of UDP-GlcNAc prevented enzyme inactivation by sulfhydryl reagents, which suggests that Cys217 is near the active site. Mouse C2GnT1 was shown to be similar to the human enzyme; it forms four disulfide bonds with a free thiol group at Cys217.¹⁷¹ The crystal structure clearly showed the four disulfide bonds as well as Cys217 in the sulfhydryl form.¹⁶⁹

The mouse C2GnT1 structures with and without acceptor substrate did not reveal a significant conformational difference, which may be expected since UDP-GlcNAc is likely to bind first, inducing a conformational change that facilitates acceptor substrate binding. Nevertheless, the Gal β 1-3GalNAc disaccharide fits well into the acceptor-binding site; therefore, peptide or other aglycone groups linked to the GalNAc residue are likely to protrude into the solution. The crystal structure of mouse C2GnT1 shows that Glu320 of the invariant SPDE sequence bonds to both O4 and O6 of GalNAc and may activate the 6-hydroxyl to induce a nucleophilic attack on the C1 of the GlcNAc moiety of UDP-GlcNAc. This confirms the specificity studies and the absolute requirement for the 4- and 6-hydroxyls of both Gal and GalNAc residues and the 2-acetamido group of GalNAc of human C2GnT.¹⁶⁶ While Glu320 may be the critical catalytic base, Arg254 forms a hydrogen bond with the O4 of Gal and Glu243 binds to O4 and O6 of the Gal residue.¹⁶⁹ The *N*-acetyl group of GalNAc as well as O2 of Gal are in contact with Tyr358 and Trp356 in C2GnT1. These contacts are expected to be different in C2GnT2, which has a broader acceptor specificity and can accommodate GlcNAc-Gal- and GlcNAc-GalNAc- instead of Gal-GalNAc- (Figure 3).

6.11.9.2 Role of Core 2 Structures

Changes in the expression and activity of C2GnT1 occur upon activation of lymphocytes with interleukin-2,¹⁶² and after enterocytic differentiation of Caco-2 cells.¹⁷² This indicates a central role of C2GnT in the biological responses of cells.

The human natural killer antigen (HNK-1), 3-sulfo-glucuronic acid/ β 1-3Gal/ β 1-4GlcNAc/ β -R, is an auto-antigen in peripheral demyelinating neuropathy and has a role in cell migration. HNK-1 has been shown to be attached to O-glycans with core 2 structures in the muscle-specific Ser/Thr/Pro-rich domain of neural cell adhesion molecules.¹⁷³ The core 2 structure on mucin-like molecules on endothelial cells and leukocytes is a scaffold for selectin ligands and C2GnT plays an important role in the synthesis of selectin ligands. The cell surface mucin-like glycoprotein PSGL-1 carries high-affinity ligands for selectins, attached to core 2.^{174,175} CHO cells expressing PSGL-1 normally do not display SLe^x, but upon transfection with C2GnT1, these cells bind to P-selectin. Tyr-sulfate also participates in the P-selectin–PSGL-1 interactions.¹⁷⁵ In a mouse leukemia WEHI-3 cell line, the core 2 structure carrying the sialyl-Lewis^x determinant has been found to be selectively expressed on PSGL-1.¹⁷⁶ Cancer cells also have the ability to bind to selectins via their O-glycans, and this interaction may contribute to the migration of cancer cells into a secondary tissue site.

Mice having a deletion in the C2GnT1 gene exhibit reduced amounts of selectin ligands as well as reduced interactions between leukocytes and the endothelium via selectins, such as leukocyte rolling along the endothelium, leukocyte recruitment into the peritoneum, and lymphocyte migration to peripheral lymph nodes.²⁷ Mainly the B-cell trafficking was affected in C2GnT1 (–/–) mice.¹⁷⁷ This indicates the importance of C2GnT1 in the lymphocyte homing process and in inflammation, and possibly also in cancer metastasis. The sulfation of selectin ligands in 6-sulfo-sialyl-Lewis^x has been shown to lead to increased affinity for L-selectin. Both C2GnT1 and GlcNAc6ST contribute to the synthesis of high-affinity L-selectin ligands. Thus, deficiency in mice of both C2GnT1 and sulfotransferase GlcNAc6ST further reduces lymphocyte homing to peripheral lymph nodes.¹⁷⁸

Maturation of dendritic cells is associated with a decreased expression of C2GnT1, a loss of sialyl-Lewis^x expression, and a concomitant increase in the expression of sialyltransferases ST3Gal-I and ST6GalNAc-II.¹⁷⁹ Thus the expression of sialyl-Lewis^x as well as the migration of DC appears to be regulated by C2GnT1. C2GnT1 also seems to regulate apoptosis.¹⁸⁰ In the thymus, immature lymphocytes that differentiate and move from the thymic cortex into the medulla lose the expression of C2GnT.⁶⁷ Since core 2 structures can bind the apoptosis-inducing endogenous lectin, galectin-1, the absence of core 2 provides a survival mechanism for maturing lymphocytes. Galectin-1 induces apoptosis in prostate-specific antigen (PSA)+ prostate cancer cells LNCaP but not in a galectin-1-resistant PSA– cell line. C2GnT1 is downregulated in the galectin-1-resistant PSA– subclone, indicating that core 2 structures may be responsible for the galectin-induced apoptosis.¹⁸¹

Several conditions are associated with altered C2GnT1 activity or expression. C2GnT1 activity is increased specifically in the heart tissues of diabetic or hyperglycemic rats.³⁴ An increase in C2GnT1 expression in polymorphonuclear leukocytes was found in patients with type 1 and 2 diabetes.¹⁸² The high C2GnT1 activity may possibly promote leukocyte–endothelial cell adhesion and capillary occlusions. The mechanisms of these alterations remain to be shown.

In some cancer cells, C2GnT1 activity has been found to be increased, while in other cells the activity may be decreased. The invasive character of cancer cells frequently corresponds to a high expression of C2GnT1 and the ability of cancer cells to interact with the endothelium via selectins. Thus, the expression of C2GnT1 is associated with a poor prognosis and malignant potential in colon, lung, and prostate cancer.²⁹ C2GnT1 expression correlates with vessel invasion and tumor depth in colon cancer.³⁰ The increase in C2GnT1 expression in lung adenocarcinoma has been correlated with malignant potential and lymph node metastasis.²⁹ In human prostate cancer tissues, C2GnT1 expression correlates with disease progression and its expression is a useful prognostic marker.³¹ Using prostate cancer cells LNCaP, Hagiwara *et al.*³¹ showed that C2GnT1 facilitated the adhesion of cancer cells to extracellular matrix glycoproteins, which may contribute to the formation of invasive and aggressive tumors.

The activity of C2GnT1 is decreased in relapsing or progressing multiple sclerosis,¹⁸³ and this may possibly be related to the autoimmune process and the influence of inflammatory cytokines that have the potential to regulate glycosylation. The activity normalized upon treatment with the anti-inflammatory

IFN- β 1a. The proinflammatory cytokine TNF- α significantly decreases C2GnT1 mRNA expression in human umbilical vein endothelial cells.⁸¹ Apoptotic and TNF- α -treated aortic porcine endothelial cells also have a very low activity of C2GnT1.⁷⁷ In contrast, C1GalT activity⁷⁷ and mRNA expression⁸¹ are increased upon stimulation of endothelial cells with TNF- α . The expression of C2GnT1 and selectin ligands in human T cells can be stimulated by cytokines such as IL-2 and other interleukins.¹⁸⁴ Thus, cell activation resulting in cytokine changes can have an impact on the adhesiveness of cells and their ability to undergo the homing process.

This suggests that TNF- α , which is associated with an inflammatory environment, promotes the appearance of core 1 structures in endothelial cells. Since the core 1 elongation enzyme (β 3GnT3) shows a low expression in human endothelial cells (see Section 6.11.13.1), core 2 is expected to be the main carrier of selectin ligands. A reduction in core 2 may be a negative feedback mechanism by which the infiltration of leukocytes to the site of inflammation is reduced.

In contrast to endothelial cells, TNF- α increased the activity of C2GnT1 in cultured primary bovine synoviocytes that normally have no detectable levels of the activity.⁷⁸ Human cells derived from bone and cartilage, but not bovine chondrocytes, have activities of the enzyme that synthesizes core 2.^{79,80} It remains to be shown if the core 2 structures in bone glycoproteins also have adhesive roles.

In breast cancer cells, the synthesis of O-glycan core 1 and 2 structures occurs mainly in the cis–medial Golgi compartment with some overlap toward the medial Golgi. The sialyltransferase ST3Gal-I that acts on core 1 substrate has been shown to be localized by immunocytochemistry mainly to the medial and trans Golgi compartments in mammary cells. Its localization therefore partially overlaps with that of C2GnT1. Transfection experiments prove that the competition between the two enzymes indeed takes place. Thus the relative expression levels and activities of these two enzymes control the relative distribution of either highly sialylated or complex chains with core 2 structures of MUC1 mucin.⁵⁵

The α 6-sialyltransferase ST6GalNAc-I that acts on GalNAc-peptide or Gal β 1-3GalNAc-peptide is broadly distributed in the Golgi, and by synthesizing sialyl α 2-6 GalNAc- or sialyl α 2-6(Gal β 1-3)GalNAc- can turn off the synthesis of core 2. The transmembrane domain of ST6GalNAc-I appears to be responsible for the localization of the enzyme. Thus, a mutant construct of C2GnT1 containing the transmembrane domain of ST6GalNAc-I localized to the trans Golgi.¹⁶⁵ Since core 1, the substrate for C2GnT1, is normally processed before it reaches the trans Golgi, the new localization of C2GnT1 resulted in inefficient core 2 synthesis. This suggests the importance of a functional assembly line in the control of O-glycan biosynthesis.

6.11.10 Synthesis of Core 3

O-glycan core 3 structures have been found in mucins from colon and bronchial tissues and salivary glands. The activity of core 3 β 3-GlcNAc-transferase (C3 β 3GnT6, **Figure 4**) that synthesizes core 3 has been discovered in colonic tissues.¹⁵⁶ Although colon tissues from human and rat are the richest sources of C3 β 3GnT6, the activity is low *in vitro*, relative to that of C2GnT1 or C2GnT2. However, rat and human colonic mucins have mostly core 3 and fewer core 1 and core 2 structures. Therefore, factors activating the enzyme may be absent or destroyed *in vitro*. It is possible that the organization of the Golgi membrane or specific protein complexes maintain the activity *in vivo*.

C3 β 3GnT6 (EC 2.4.1.147) is a member of the β 3-GlcNAc-transferase family and has been classified as CAZy GT31. The enzyme has a conserved DxD motif and requires divalent cations for activity. It has five Cys residues that are conserved among β 3-GlcNAc-transferases and one conserved N-glycosylation site.¹⁸⁵ C3 β 3GnT6 does not act on core 6 to form core 4¹⁵⁶ and also cannot act on α 6-sialylated GalNAc-R acceptors. Sialylated core 3 structures such as those found in colonic mucins therefore must be synthesized by forming the GlcNAc–GalNAc backbone first, followed by sialylation.

The mRNA of C3 β 3GnT6 is expressed at high levels in the foveolar epithelium of the stomach, in colon, and small intestine, but at low levels in other tissues. In colon cancer tissues, the activity is greatly reduced compared to normal colon, which favors the synthesis of core 1, and might explain the prevalence of the T antigen (core 1) in colon cancer.^{35,36} The mRNA expression of C3 β 3GnT6 is also reduced in gastric cancer. Consistent with the loss of the enzyme in cancer tissue, the activity is not detectable in many cultured colon

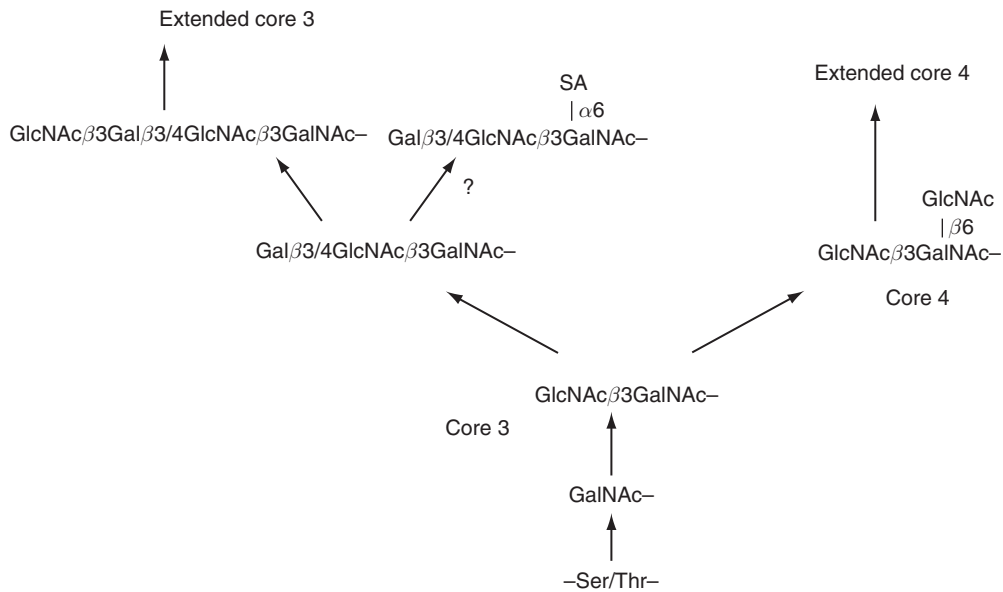


Figure 4 Pathways in the synthesis of core 3 and core 4. The scheme shows the pathways in the biosynthesis of O-glycans with core 3 and 4 structures. Polypeptide GalNAc-transferase adds GalNAc to peptide, followed by core 3 β 3-GlcNAc-transferase (C3 β 3GnT6) to form core 3. Core 3 can be converted to core 4 by C2GnT2. Alternatively, core 3 can be further extended, which prevents the conversion to core 4, but may allow the addition of sialic acid to GalNAc. Both core 3 and core 4 structures can be extended to form complex chains carrying a multitude of sugar determinants. The enzymes that synthesize the structures are active in colon tissues.

cancer cells.^{36,172,186,187} Tissue from polyposis coli often develops into cancer. The mRNA expression of C3 β 3GnT6 gradually decreases during the course of familial adenomatous polyposis.³⁶ In addition, polyposis coli cells in culture during progression to tumorigenic cells lose the activity of C3 β 3GnT6.¹⁸⁶ When human fibrosarcoma HT1080 FP-10 cells were transfected with the C3 β 3GnT6 gene, cell migration was decreased. Similar data were obtained with two human adenocarcinoma cell lines HCT-15 and HT29LMM. The ability of HT1080 FP-10 cells to form lung metastases *in vivo* in mice was also drastically reduced.³⁶ Thus the enzyme is a marker of healthy colon tissue and may be related to the suppression of the metastatic phenotype in cancer cells.

Deletion of the C3 β 3GnT6 gene in mice leads to a reduction in colonic MUC2 mucin and increased intestinal permeability. In addition, knockout mice were more susceptible than normal mice to experimental colitis induced by dextran sodium sulfate, and showed increased colorectal tumors after azoxymethane and dextran sodium sulfate treatment.³⁷ Thus, C3 β 3GnT6 has a role in the prevention of colonic tumors. C3 β 3GnT6 competes with C1GalT for the common GalNAc substrate in the mouse colon;⁸² however, it seems that core 3 synthesis prevails. In cell types that do not express core 3, the absence of core 1 synthesis leads to Tn antigen expression.¹⁵² However, in the C3 β 3GnT6 knockout mouse model, the absence of core 3 synthesis increases the exposure of the Tn antigen even in the presence of core 1 synthesis. In the CFTR-deficient mouse that shows extensive intestinal inflammation, C3 β 3GnT6 has no detectable activity in colonic tissues.⁸² Thus, core 3 synthesis is regulated by many factors that can change in disease and this remains to be further examined.

6.11.11 Synthesis of Core 4

O-glycan core 4 is synthesized from core 3 by the M type C2GnT (C2GnT2, EC 2.4.1.148).^{156,166} The C2GnT2 enzyme from rat colon is less specific toward its acceptor substrate than C2GnT1 (Figure 3) and thus can synthesize core 2, core 4, and the branch of the I antigen GlcNAc β 1-6(GlcNAc β 1-3)Gal-.¹⁵⁵⁻¹⁵⁷ C2GnT2

competes with $\alpha 6$ -sialyltransferases for the GlcNAc $\beta 1$ -3GalNAc substrate. When GlcNAc of core 3 is extended by a Gal residue, core 4 can no longer be formed by the M type C2GnT2 from rat colon.¹⁸⁸

The C2GnT2 activity is present in selected mucin-secreting cells and tissues, with colonic mucosa having a high activity. The gene is expressed in the stomach, colon, kidney, trachea, and testes, and at lower levels in the small intestine and pancreas^{155,189} and in a number of cancer cell lines including A549 lung cancer cells.^{155,172,186,187} Many colon-derived cell lines maintain this high activity, but a number of cells have lost C2GnT2, or both C2GnT2 and C2GnT1.^{155,186,187} It appears that there is a switch from C2GnT2 to C2GnT1 in colon cancer.¹⁹⁰ The expression of C2GnT2 in colon cancer cells HCT116 suppressed cell growth, cell adhesion, motility, and invasive properties. C2GnT2 also suppressed tumor growth *in vivo* in xenografts of nude mice.¹⁹⁰

C2GnT2 has been highly purified from bovine tracheal tissue¹⁹¹ and the mechanism was shown to be ordered sequentially in which UDP–GlcNAc binds to the enzyme first and UDP is released last. C2GnT2 is a member of the $\beta 6$ -GlcNAc-transferase family and has been classified into the GT14 family of inverting transferases. Although the enzyme has a DxD sequon, it does not require divalent metal ions, and its catalytic mechanism may not be related to the DxD motif; thus it may complex the pyrophosphate group of UDP–GlcNAc with cationic groups in the protein.

A related C2GnT2 gene has been found in bovine herpesvirus type 4. The gene may have been acquired from the African buffalo. The function of the gene is not known since it is not essential for virus replication.¹⁹²

C2GnT2 occurs in some mucin-producing reproductive tissues. The mRNA and activity levels of C2GnT2 are regulated in the hamster oviduct during the estrous cycle in the Golden hamster.¹⁹³ During the time of fertilization, C2GnT2 expression is high, and this corresponds to a high amount of posttranslational processing of the mucin oviductin (MUC9, **Table 1**). Changes in O-glycosylation have also been observed in human endocervical mucus.¹⁹⁴ The core 2 structures appear to be more abundant during ovulation, while sialylation is decreased.

The expression of C2GnT2 is also upregulated in H292 human airway lung cancer cells¹⁹⁵ by retinoic acid and by T-helper cytokines IL-4 and IL-13. In contrast, EGF downregulates C2GnT2 activity, shifting core 2 structures on recombinant MUC1 to core 1 structures. At the same time, EGF upregulates MUC5A expression in H292 cells. It appears that both the enzymes that synthesize core 3 and core 4 have a protective function in colon and lung tissues.

6.11.12 Synthesis of Minor Core Structures

The enzymes that add Gal or GalNAc in α -linkage to GalNAc and synthesize minor core structures (**Table 2**) have not yet been characterized. The O-glycan core 5 structure occurs in mucins from meconium and adenocarcinoma of the colon, and is thus an oncofetal structure. An $\alpha 3$ -GalNAc-T involved in the synthesis of core 5 has been described in a patient with adenocarcinoma.⁷¹ The enzyme uses asialo-bovine submaxillary mucin containing many GalNAc–mucin linkages as the acceptor substrate. The synthesis of core 6 has been reported by a GlcNAc-transferase in ovarian tissue.¹⁹⁶ The activity used GalNAc α -benzyl as an acceptor substrate but has not been characterized yet. It remains to be shown if core 6 structures in human mucins are the result of *de novo* synthesis rather than degradation of the core 2 structure. The genes responsible for the synthesis of minor core structures remain to be identified and cloned.

6.11.13 Extension and Branching Reactions

O-glycan core structures 1–4, and 6 can be extended by repeating GlcNAc–Gal residues. These can be branched, form the backbone structures that are antigenic in themselves, and can carry a number of terminal epitopes including blood group antigens. A family of $\beta 3$ -GlcNAc-transferases adds the GlcNAc $\beta 1$ -3 residue to terminal Gal, which then forms the i antigens (poly-N-acetyllactosamines, **Table 2**). GlcNAc $\beta 1$ -6 branches can be synthesized by $\beta 6$ -GlcNAc-transferases (IGnT) to form the I antigens. The GlcNAc residues can be extended by a family of $\beta 4$ -Gal-transferases to form type 2 chains (**Table 2**), and by $\beta 3$ -Gal-transferases to form type 1 chains. Some mucins are very rich in these extensions, for example, ovarian cyst and colonic mucins. Similar extensions can be found on N-glycans and on glycosphingolipids.

6.11.13.1 β 3-GlcNAc-Transferases

The β 3-GlcNAc-transferases (β 3GnT) that synthesize the i antigen are members of the CAZy GT31 family (EC 2.4.1.149) (Table 4). Several of these β 3GnT enzymes are involved in the extension of O-glycans. The β 3GnT activity is generally low *in vitro* but is ubiquitously found and at least one member of the β 3GnT family is expressed in most tissues. The enzymes require divalent metal ions and have DxD sequences. Low-molecular-weight oligosaccharides can form good acceptor substrates. β 3GlcNAcT2 is ubiquitous and highly expressed in human endothelial cells (HUVECs), and the expression levels are increased upon TNF- α stimulation.⁸¹ However, the enzyme that specifically elongates the Gal β 1-3 residue of core 1 and core 2 structures and was discovered in pig gastric mucosa¹⁹⁷ (β 3GnT3, EC 2.4.1.146) showed a low expression in human endothelial cells, although the enzyme is highly expressed in human placenta, colon, small intestine, neutrophils, lymphocytes, and high endothelial venules (HEVs).⁶⁸ Leukocyte adhesion to selectins is significantly reduced in C2GnT1 (–/–) mice but HEV adhesion to lymphocytes is only marginally reduced and HEVs still display the 6-sulfo-Lewis^x MECA-79 epitope (Table 2) with the structure Gal β 1-4(6-sulfo-)GlcNAc β 1-3Gal β 1-3GalNAc-, which is recognized by L-selectin.^{68,198} Transfection of CHO cells that express PSGL-1 and Fuc-transferase VII (FUT7) with either C2GnT1 or β 3GnT3 caused the appearance of L-selectin ligand sialyl-Lewis^x.¹⁹⁸ This suggests that β 3GnT3 from HEV is responsible for the synthesis of extended O-glycans that display selectin ligands and can carry the MECA-79 epitope.

β 3GnT1, β 3GnT2, and β 3GnT4 can synthesize type 2 chains of O-glycans and act on Gal β 1-4GlcNAc termini. T1 and T4 have low activities toward Gal β 1-3 termini and T2 has a lower activity toward type 1 chains. The β 3GnT activities are significantly increased when T2 and T8 enzymes are mixed together in the assay, suggesting that they form active complexes *in vivo*.¹⁹⁹ T4 has a restricted and low expression in brain tissues, kidney, colon, and esophagus. β 3GnT7 is mainly involved in the synthesis of keratan sulfate (KS) chains, which are highly sulfated variants of the poly-N-acetylglucosamine extensions of N- and O-glycans.²⁰⁰ T7 is expressed in placenta, colon, and lung tissues in the mouse. Treatment of lung cancer cells with antisense DNA of β 3GnT7 increased cell motility and invasive properties. This suggests that the enzyme may have antiadhesive effects and that the extended chains are involved in blocking the invasion of tumor cells. β 3GnT8 acts on N-acetylglucosamine termini as well as O-glycan core 2. In colon cancer, the expression of β 3GnT8 is significantly upregulated.³⁸

6.11.13.2 β 6-GlcNAc-Transferases

The Gal residues of elongated O-glycan chains can be branched by a family of IGnT, which form the GlcNAc β 1-6Gal- branch of I antigens. The branching β 6-GlcNAc-transferases are closely related to C2GnT and are members of the CAZy GT14 family (EC 2.4.1.150) (Table 4). While C2GnT2 can synthesize the equivalent branch on penultimate Gal residues in O-glycans, the three IGnT isoenzymes act on internal Gal residues of extended O- and N-glycans to synthesize GlcNAc β 1-6(R-GlcNAc β 1-3)Gal branches.

In humans and mice, three splicing variants of the same IGnT gene are expressed in a tissue-specific fashion. IGnT1 is expressed in brain, thymus, small intestine, colon, placenta, and other cell types.^{201,202} While IGnT2 is ubiquitous, IGnT3 has a more restricted expression and is found at low levels in heart, bone marrow, stomach, small intestine, and kidney. During development in humans, the i antigen is replaced by the I antigen, probably due to the increased expression of IGnT. The i adult phenotype associated with congenital cataracts is caused by a mutation in the IGnT exon that is common to all three forms of the IGnT gene.²⁰³ Mice with a nonfunctional common exon of IGnT do not express the I antigen, with the exception of the small intestine, which expresses C2GnT2 capable of synthesizing subterminal GlcNAc β 1-6 branches.^{155,157} The lack of I antigen in bone marrow cells in these mice was associated with a low number of peripheral blood lymphocytes as well as renal abnormalities.

6.11.13.3 Extension GalNAc-Transferases

Less common extensions of O-glycans can be synthesized by β 4-GalNAc-transferases, β 4GalNAc-T3 and T4, that add GalNAc to terminal GlcNAc residues to form LacdiNAc (GalNAc β 1-4GlcNAc β -) structures.²⁰⁴

Human β 4GalNAc-T3 is expressed in the intestines and testes, while β 4GalNAc-T4 is highly expressed in ovarian and fetal tissues and in a number of cancer cell lines.^{205,206} β 4GalNAc-T3 and T4 both act on GlcNAc β -benzyl substrates as well as on core 2, 3, and 6 derivatives.

Another class of extension enzymes are the β 3-GalNAc-transferases (β 3GalNAc-T2).²⁰⁷ β 3GalNAc-T2 has high sequence similarity to enzymes of the β 3-Gal-transferase and β 3-GlcNAc-transferase families. The enzyme synthesizes the unusual GalNAc β 1-3GlcNAc linkage and is highly expressed in heart, skeletal muscle, adipose tissue, ovary, prostate, and especially in the testes.

6.11.14 The β 3- and β 4-Gal-Transferase Families

Both Gal β 1-4GlcNAc and Gal β 1-3GlcNAc extensions are found in the backbone structures of O-glycans. Two families of Gal-transferases are involved in the synthesis of these extensions.^{204,208} β 4-Gal-transferases (β 4GalT) (EC 2.4.1.38, EC 2.4.1.90, EC 2.4.1.22) are the most common and highly active Gal-transferases, and five of these (β 4GalT1, T2, T3, T4, and T5) act on terminal GlcNAc residues of mucin-type O-glycans to synthesize type 2 chains. In addition, four β 3-Gal-transferases (β 3GalT) act on O-glycans (β 3GalT1, T2, T3, and T5) that synthesize type 1 chains. These inverting transferases have a DxD sequon and require divalent cations such as Mn²⁺ for activity.

6.11.14.1 β 3-Gal-Transferases

β 3-Gal-transferases belong to the CAZy GT31 family and are variably expressed in different tissues. Sheares and Carlson²⁰⁹ purified a β 3GalT from pig trachea that acts on the GlcNAc residue of O-glycan core 3, which may have been β 3GalT5.²¹⁰ Later cloning efforts identified β 3GalT1, T2, T3, and T5 that may be involved in the extension of O-glycans.²¹¹ β 3GalT5 appears to have a major role in the synthesis of sialyl-Lewis^a antigen and type 1 chains on O-glycans.^{40,212} Pancreatic carcinoma cells BxPC3 normally secrete glycoproteins with sialyl-Lewis^a into the cell medium. Suppression of β 3GalT5 by antisense DNA in pancreatic carcinoma cells BxPC3 yields glycoproteins with type 2 chains and sialyl-Lewis^x determinants, instead of type 1 chains and the sialyl-Lewis^a antigens. Both antigens were suppressed with GalNAc-benzyl, indicating that they were attached to O-glycans.^{213,214}

6.11.14.2 β 4-Gal-Transferases

Enzymes of the β 4-Gal-transferase family are classified as inverting transferases of the CAZy GT7 family with a GT-A fold^{208,215,216} (Table 4). The binding of α -lactalbumin, synthesized in the mammary gland, to β 4GalT1 and T2 isoenzymes changes the conformation of the enzymes, resulting in a change in the acceptor-binding site. Thus, these transferases become lactose synthases and preferentially bind glucose instead of GlcNAc.²¹⁷ The specificities of β 3GalTs are not affected by α -lactalbumin. β 4GalT1 is mainly expressed in the mammary gland. An overexpression of β 4GalT1 in mice, however, leads to altered morphogenesis of the mammary gland and impaired lactation.

The ubiquitous β 4GalT1 is the major Gal-transferase in human and animal tissues. The amino-terminal, transmembrane, and stem domains of β 4GalT1 all contribute to the proper anchoring of the enzyme to trans Golgi membranes.²¹⁸ Owing to two initiation start sites, a long and a short form of the enzyme exist. The long form only appears to have N-terminal amino acids required for targeting a small proportion of the enzyme to the cell surface.^{219,220}

Cells derived from bone and cartilage have a high activity of Gal-transferase.^{78–80} In bovine synoviocytes and human chondrocytes, TNF- α causes an increase in β 4GalT activity, but in human osteoblasts, it is transforming growth factor (TGF) β that increases β 4GalT activity. TNF- α causes an increase in the expression of β 4GalT1 and T5 in human endothelial cells.⁸¹

A deletion of β 4GalT1 in mice causes a variety of problems, possibly due to the reduced ability of these mice to produce sialyl-Lewis^x and related structures. In the absence of the enzyme, glycan structures shift from type 2 to type 1, with a drastic reduction in P-selectin binding. Delayed skin wound healing, reduced neutrophil and

macrophage recruitment to the wound site, and reduced inflammatory responses including contact hypersensitivity and delayed type hypersensitivity in these mice indicate that the enzyme is a crucial factor in the regulation of the immune system.^{216,221–223} Per-*O*-acetyl-Gal β 1-4GlcNAc β -naphthalenemethanol was developed as an inhibitor of sialyl-Lewis^x synthesis. When given systemically in mice, it can serve as a primer for the assembly of sialyl-Lewis^x, and thus reduces cancer metastases of lung carcinoma and melanoma cells mediated by sialyl-Lewis^x.²²⁴

The congenital disorder of glycosylation (CDG)-IIId has been identified as a deficiency of galactosylation.⁴⁰ Patients present with major neurological and blood clotting disorders and a lack of Gal, as well as terminal sugars attached to Gal residues, including sialic acid. Mutations in the COG8 gene affect galactosyltransferase localization and activity *in vivo* and result in severe developmental problems and neurological symptoms of CDG-IIh.⁹⁵

β 4-Gal-transferase expression is often upregulated and β 3-Gal-transferase expression downregulated in cancer cells with an increased production of type 2 chains.²²⁵ The transcription factor Sp1 appears to play a role in the upregulation of β 4GalT5.²²⁶

Bovine and human β 4GalT1 have been crystallized in many different complexes with donor and acceptor substrate analogues and Mn²⁺ ions.^{215,227} The bovine β 4GalT1 binds UDP-Gal and the Mn²⁺ ion in a deep binding pocket through Asp252 and 318, Gly292 and 315, and Glu317 residues.^{215,228} The DxD motif is in contact with the Mn²⁺ ion and the Gal residue of UDP-Gal. All of the hydroxyls of Gal interact with the protein. The binding of UDP-Gal induces a conformational change in the enzyme, which closes a loop and creates the acceptor-binding pocket. A variety of GlcNAc-terminating acceptor substrates with hydrophobic groups and peptides as aglycones can bind to the enzyme.²²⁹ Core 2, 3, 4, and 6 attached to glycopeptides derived from the VNTR of MUC2 are all good acceptors for β 4GalT1 from bovine milk. The α -linked GlcNAc derivatives are poor acceptors, which can be explained by the steric hindrance of Tyr286.²²⁸ The extended acceptor-binding site can also accommodate large GlcNAc-terminating oligosaccharides. Several hydrophobic amino acids (Phe280 and 360, Tyr286, Arg259, and Ile363) have been identified that participate in donor and acceptor substrate binding. Trp314 appears to be essential for the activity since it interacts with both donor and acceptor in the UDP-Gal-bound state and its mutation abolishes the activity.²³⁰ The amino acids of the GlcNAc-binding site are highly conserved, while there are variations in the preference for the underlying structures among members of the Gal-transferase family.

6.11.15 Sialyltransferases

Several classes of sialyltransferases act on *O*-glycans and add sialic acid in α 2-3 linkage to the Gal β 1-3-linked residue of cores 1 and 2 and to Gal residues on the extensions of core structures (ST3Gal, EC 2.4.99.6, EC 2.4.99.4, EC 2.4.99.3). The α 6-sialyltransferases form the sialyl α 1-6GalNAc linkage of the sialyl-Tn and sialyl-T antigens (ST6GalNAc, EC 2.4.99.3, EC 2.4.99.7). Extended *O*-glycans and core 1, 3, and 5 structures can also be modified by sialic acid linked to GalNAc but sialyl α 2-6Gal linkages have not been found on mammalian mucin-type *O*-glycans. Sialyl α 2-8sialic acid linkages have also been described in several glycoproteins.^{16,231}

Mammalian sialyltransferases (**Table 4**) are inverting transferases and belong to the CAZy GT29 family; they have a number of common sequences (sialylmotifs L, S, VS, and III).²³² The large (L) sialylmotif is found in the middle of the catalytic domain where it contributes to the binding of the nucleotide sugar substrate CMP β -sialic acid. A short (S) motif is located near the C terminus and contributes to the binding of both donor and acceptor substrates. Subfamilies of sialyltransferases have characteristic additional sequence motifs of 6–20 residues that may be related to their linkage-specific action.¹⁰⁴

Generally, sialylation is a terminal event and stops chain growth and branching. However, the enzymes that synthesize blood group ABO and Cad determinants, as well as α 2- and α 3-Fuc-transferases and α 8-sialyltransferases, can act on sialylated substrates. Sialic acid-containing structures have many direct and indirect functions. They can mask the underlying antigens or are recognized directly as determinants for cell adhesion molecules. In the human colon, sialic acids may be *O*-acetylated, which masks the sialyl-epitopes.²²³

O-acetylation occurs at the level of CMP-sialic acid, and is regulated during development and in cancer. A number of defects in the metabolism of sialic acids have been identified, and a loss of CMP-sialic acid transport affects the biosynthesis of O-glycans.⁹²

6.11.15.1 α 3-Sialyltransferases

Five or more members of the α 3-sialyltransferase (ST3Gal) family (Table 4) may be involved in sialylating the Gal residues of O-glycans.⁹⁷ ST3Gal-I and ST3Gal-IV have a preference for the Gal β 1-3 residue of core 1 and 2 substrates and are responsible for the synthesis of the sialyl-T antigen. The acceptor substrates can be linked to hydrophobic groups such as benzyl, *p*-nitrophenyl, or peptide. ST3Gal-II, III, and VI prefer the *N*-acetylglucosamine extensions as substrates and are involved in sialyl-Lewis^x or sialyl-Lewis^a synthesis, and are not O-glycan specific (Figure 5).^{233,234} In monocyte-derived dendritic cells, ST3Gal-I, II, and IV are regulated during differentiation by a number of growth factors and cytokines, including TNF- α .²³⁵ Inflammatory conditions can increase the expression of ST3Gal-I, III, and IV. In human bronchial mucosal tissue explants, treatment with TNF- α increased α 3-sialyltransferase activity toward *N*-acetylglucosamine, and increased the expression of ST3Gal-III and IV.⁷⁶ In HUVECs, TNF- α also significantly increases the mRNA expression of ST3Gal-I and VI.⁸¹

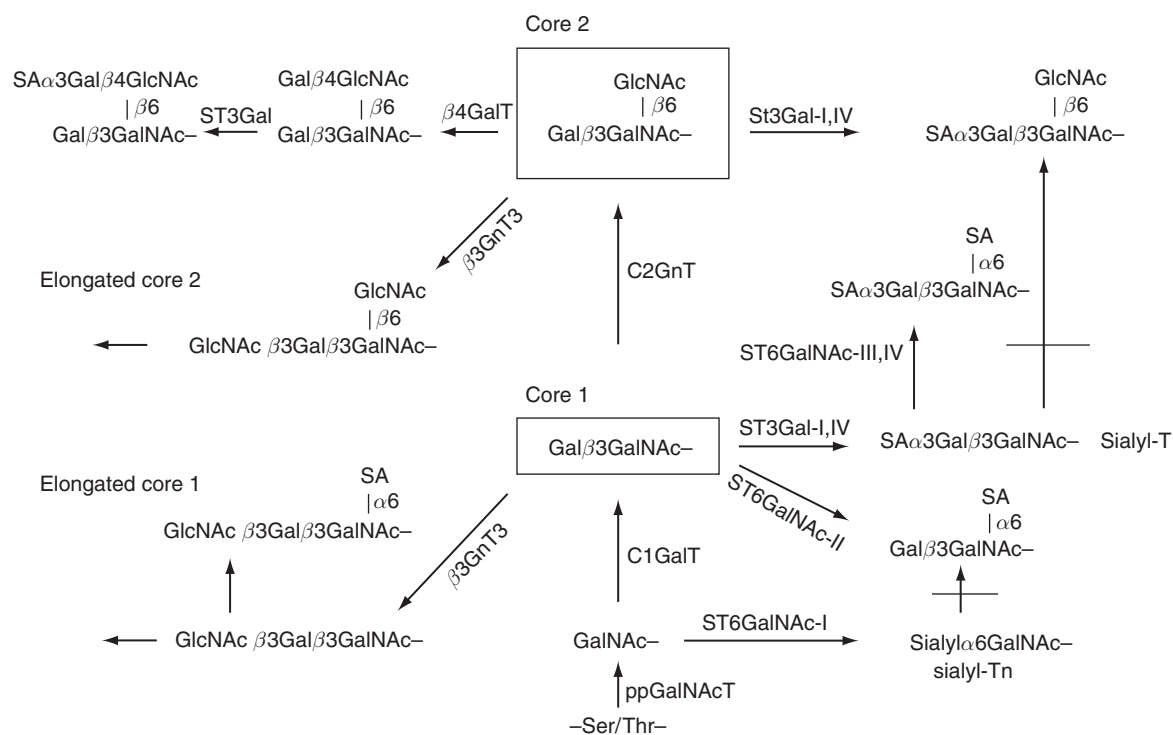


Figure 5 Sialylation pathways for O-glycan cores 1 and 2. The figure shows the conversion pathways in the biosynthesis of common sialylated O-glycans with core 1 and 2 structures. Polypeptide GalNAc-transferase (ppGalNAc-T) adds GalNAc to peptide, and core 1 β 3-Gal-transferase (C1GalT) synthesizes core 1, which can be converted to core 2 by core 2 β 6-GlcNAc-transferase (C2GnT). Cores 1 and 2 can also be elongated by β 3-GlcNAc-transferase (β 3GnT3). α 3-Sialyltransferases ST3Gal-I and IV can add sialic acid (SA) to the terminal Gal residues of cores 1 and 2. α 3-Sialylation of core 1 prevents its conversion to core 2 (indicated by a blocked arrow). GlcNAc residues can be modified by β 4-Gal-transferases (β 4GalT). The Gal residue attached to GlcNAc can then be modified by other α 3-sialyltransferases (ST3Gal). The α 6-sialyltransferase ST6GalNAc-I synthesizes the sialyl-Tn antigen from GalNAc-peptide, while ST6GalNAc-II synthesizes an α 6-sialyl-T antigen from Gal β 1-3GalNAc-peptide. This α 6-sialylated core 1 may be converted to blood group ABO but otherwise cannot be elongated. After the synthesis of the α 3-sialyl-T antigen, ST6GalNAc-III or IV can add another sialic acid to GalNAc to form the disialylated T antigen. In cancer or inflammatory bowel disease, the core 1 remains unmodified and exposes the T antigen (core 1).

The minimal protein size required for the catalytic activity of human ST3Gal-I was studied by Vallejo-Ruiz *et al.*²³⁶ Recombinant enzymes expressed in COS-7 cells were found to be N-glycosylated at each of the five N-glycosylation sites. Deletion mutants of the N-terminal region showed that deletion of 25 and 56 amino acids produced active enzymes that were secreted but further deletion of 76 and 105 amino acids resulted in inactive enzyme proteins that were not secreted. Mutagenesis showed that His316 of the VS sialylmotif, and His299 and Tyr300 of the sequence between the sialylmotifs VS and S, is crucial for the enzyme activity. In particular, the aromatic ring of Tyr300 appears to play a role in the conformation of the active enzyme.²³⁷

Mice lacking the ST3Gal-I gene develop normally but express more unmodified Gal β 1-3GlcNAc termini on CD4+ lymphocytes. These ST3Gal-I knockout mice show defective CD8+ T lymphocyte function and altered apoptosis.²³⁸ A deletion of the ST3Gal-IV gene in mice causes bleeding disorders, thrombocytopenia, and a reduction of von Willebrand factor in the serum.⁴⁵ In these mice, ST3Gal-IV is essential for L-selectin-mediated leukocyte interactions in inflammation mediated by PSGL-1.²³⁹

Sialylation is often altered in cancer cells, and the metastatic phenotype is associated with a higher degree of sialylation that may possibly be a survival mechanism for metastatic cells in the bloodstream. The expression levels of both ST3Gal-I and ST3Gal-IV were found to be increased in colorectal carcinomas⁴⁶ while especially ST3Gal-I expression was found to be increased in ovarian serous carcinoma.⁴⁴ Breast cancer tissues and cells, compared to normal controls, also exhibit increased α 3-sialyltransferase activity toward core 1 substrates as well as a higher level of ST3Gal-I mRNA.^{43,153} The increase in α 3-sialylation was associated with smaller, highly sialylated O-glycans with core 1 structures and a decrease in complex and large core 2 structures. This switch can explain the increased reactivity of SM3 antimucin antibodies in breast cancer cells (**Table 1**). ST3Gal and C2GnT both act on core 1 substrates, and compete when present in the same compartment. ST3Gal-I has been localized mainly to the medial and trans Golgi compartments of T47D breast cancer cells and normal mammary cells, while C2GnT1 was localized mainly to cis and medial Golgi compartments. This partial overlap allows the two enzymes to effectively compete for the synthesis of mucin O-glycans. Transfection studies demonstrated that an overexpression of ST3Gal-I can reduce core 2 formation and peptide epitope exposure by increasing sialylation, while an overexpression of C2GnT leads to the masking of peptide epitopes.⁵⁵ The expression of ST3Gal-I is related to tumor grade in a mouse model.²⁴⁰ In addition, tumors expressing sialylated core 1 grew significantly faster than those expressing core 2 structures. This indicates that ST3Gal-I controls the O-glycosylation and biological properties of murine and human breast cancer cells.

6.11.15.2 α 6-Sialyltransferases

Of the four O-glycan-specific α 6-sialyltransferases, ST6GalNAc-I, II, III, and IV,²³³ ST6GalNAc-I is primarily responsible for the synthesis of the sialyl-Tn antigen in mucins. The enzyme requires GalNAc-peptide as an acceptor but is less specific with respect to the substitution of GalNAc, and acts on GalNAc, core 1, and sialyl α 2-3Gal β 1-3GalNAc α - chains. ST6GalNAc-II acts preferably on acceptor substrates having core 1-linked to peptide and synthesizes the sialyl-T antigen sialyl α 2-6(Gal β 1-3)GalNAc-.²⁴¹ The enzyme does not act on GalNAc-ovine submaxillary mucin, and therefore cannot synthesize the sialyl-Tn antigen. ST6GalNAc-II is active with bovine submaxillary mucin, which carries a mixture of O-glycans. However, it remains to be shown exactly which sialyltransferases are responsible for sialylation of GalNAc residues of cores 3 and 5 in the colon. ST6GalNAc-II is expressed mainly in heart, kidney, and skeletal muscle, and is highly expressed in many breast, hepatic, and colon cancer cell lines.²⁴²

ST6GalNAc-I, III, and IV can synthesize the disialylated T-antigen found on mucins and other glycoproteins (**Figure 5**). Both ST6GalNAc-III and IV are specific for sialyl α 2-3Gal β 1-3GalNAc- structures but do not require peptide in the substrate; thus, peptide can be replaced by a hydrophobic group. Human ST6GalNAc-III has a restricted tissue expression, mainly in kidney and brain.²⁴³ TNF- α treatment of human endothelial cells upregulates the expression of ST6GalNAc-III.⁸¹ Human ST6GalNAc-IV has been cloned from HepG2 cells.²³³ It is widely distributed among tissues with the highest expression in liver and kidney. The enzyme is also expressed in many different cancer cell lines including BT20, Caco-2, HT29, and HL60 cells but not in T47D human breast cancer cells.

ST6GalNAc-I expression correlates with sialyl-Tn expression in gastric cancer cells.²⁴⁴ Overexpression of the enzyme in mouse mammary adenocarcinoma-derived cells leads to a conversion to the sialyl-Tn phenotype with decreased peanut agglutinin staining, altered morphology, and reduced migration and proliferation.²⁴⁵ The occurrence of the tumor-specific antigen sialyl-Tn is associated with an unfavorable prognosis and formation of metastatic cancer and correlates with the progression of intestinal disease to malignancy.⁵³

In breast tumors, the sialyl-Tn antigen also appears to be synthesized by ST6GalNAc-I. ST6GalNAc-I is broadly distributed in T47D breast cancer cells. This distribution promotes the synthesis of the sialyl-Tn antigen while blocking the synthesis of core 1.²⁴⁶ The expression levels of sialyl-Tn antigen as well as that of ST6GalNAc-I were found to be associated with decreased cell adhesion and increased cell migration of human breast cancer cells MDA-MB-231.^{247,248} The sialyl-Tn-positive clones exhibited enhanced tumor growth in SCID mice. The expression of ST6GalNAc-II is significantly increased in colorectal cancer, and its mRNA expression has been correlated with poor survival and invasive potential of colorectal cancer.⁴⁶

6.11.15.3 α 8-Sialyltransferases

Many glycoproteins, especially those in brain and the nervous system, have sialyl α 2-8sialic acid and polysialic acid structures. These structures have antiadhesive properties and regulate cell–cell interactions. Recently, the disialo-linkage has been found on *O*-glycans of a glycoprotein CD36 from human milk.²³¹ In addition, MUC1 from the serum of breast cancer patients was shown to have core 1 structures with the sialyl α 2-8sialic acid linkage¹⁶ (Table 2). A family of polysialic acid α 2,8-sialyltransferases (ST8Sia) exists. The α 8-sialyltransferase responsible for the synthesis of the disialo-linkage on *O*-glycans (human ST8Sia VI) has been cloned from human breast cancer cells MCF-7. The expression of the enzyme has been found to be very low in human tissues, but was present in breast cancer cells MCF-7, T47D, the megakaryocyte cell line Dami that showed a strong expression, and other cancer cell lines. The enzyme acts specifically on *O*-glycans of fetuin as well as bovine and ovine submaxillary mucins, and appears to preferentially act on sialylated core 1, sialyl α 2-3Gal/ β 1-3GalNAc–Thr/Ser, as the acceptor substrate.²⁴⁹ Five N-glycosylation sites are present in the enzyme.

6.11.16 Fucosyltransferases

The termini of mucin *O*-glycans often have Fuc residues in either α 1-2 linkage to Gal or in α 1-3 and α 1-4 linkage to GlcNAc that play important roles in the immune system, as antigens and epitopes on mucins and as cell surface ligands for cell–cell interactions via selectins. Several subfamilies of Fuc-transferases are involved in *O*-glycosylation (Table 4). α 2-Fuc-transferases (FUT1, FUT2, EC 2.4.1.69) synthesize the blood group O (H antigen) and also modify the Lewis^a and Lewis^x antigens to form Lewis^b and Lewis^y, respectively (Table 2). α 3-Fuc-transferases IV, V, VI, VII, and IX (FUT4, 5, 6, 7, and 9, EC 2.4.1.152) synthesize Lewis^a and Lewis^x and their variants, potentially attached to *O*-glycans. FUT3 (EC 2.4.1.65) has an unusual dual specificity and can synthesize both α 1-3 and α 1-4 linkages. These enzymes are expressed in a cell type-specific fashion. Mucin-secreting tissues abundantly express FUT2 and a number of α 3-Fuc-transferases. FUT4 and 7 contribute to the synthesis of selectin ligands but the two Fuc-transferases appear to have different roles.¹⁷⁴

Deficiencies in Fuc-transferases, for example, FUT1, FUT2,⁹⁰ or FUT3,⁸⁹ lead to nonpathological structural changes in *O*-glycans. However, when fucosylation is generally deficient (e.g., in GDP-Fuc transport deficiency), serious problems in the immune system arise, likely due to the inability to synthesize selectin ligands. These fucosylation defects are seen in CDG-IIc or leukocyte adhesion deficiency LADII.^{91,250,251}

6.11.16.1 α 2-Fuc-Transferases

α 2-Fuc-transferases are members of the CAZy GT11 family. FUT1 is encoded by the H gene in hematopoietic tissues²⁵² and cell types. FUT2 is the main α 2-Fuc-transferase in epithelial and mucus-secreting tissues and is encoded by the Se (secretory) gene.²⁵³ Both FUT1 and 2 act on the Gal residue of *N*-acetylglucosamines. FUT2 has a preference for the core 1 structure and is responsible for the synthesis of the common Fuc α 1-2Gal/ β 1-3(sialyl α 2-6)GalNAc structures found in mucins from porcine submaxillary gland and other tissues. The

enzyme is also involved in the synthesis of Lewis^b or Lewis^y determinants of mucins. The Fuc α 1-2 residue is added to the *N*-acetylglucosamine terminus before a Fuc α 1-3 or Fuc α 1-4 residue is added.

In the rare Bombay and para-Bombay phenotypes, the blood group H determinant is absent in red cells due to defects in the FUT1 gene, resulting in the absence of the ABO blood groups.²⁵⁴ The activity of FUT1 toward terminal Gal and *N*-acetylglucosamine termini is increased in colonic tumors and is associated with malignant progression.⁴⁷ Transfection of FUT1 into rat colon carcinoma REGb cells led to a higher aggressiveness and tumorigenicity in syngeneic rats. FUT1 expression is also related to resistance to apoptosis.²⁵⁵ In mice, overexpression of FUT1 is associated with colitis and lymphopenia along with aberrant T-cell markers. This shows that FUT1 is critically involved in T-cell development and inflammation, and plays a role in the regulation of apoptosis. FUT1 expression is increased in HUVECs⁸¹ upon stimulation with TNF- α .

The gene expression of FUT2 can be induced specifically in the intestine of germ-free mice by oral fecal microbes. This suggests that host–microbe interactions are major tissue-specific factors regulating *O*-glycan synthesis.²⁵⁶ The gene sequences of α 2-Fuc-transferases are closely related to those of blood group A and B transferases.²⁵⁷

6.11.16.2 α 3/4-Fuc-Transferases

Several genes encode α 3-Fuc-transferases III–VII and IX (FUT3–7, FUT9) and participate in the synthesis of Lewis determinants on *O*-glycans in a cell type-specific fashion. These enzymes are retaining transferases and belong to the CAZY GT10 family; they are ubiquitous and have been found in mammals, *Drosophila*, *C. elegans*, and bacteria such as *Helicobacter pylori* and many other species. FUT4 and FUT7 are expressed in myeloid cells and are the major α 3-Fuc-transferases involved in the synthesis of sialyl- and/or sulfo-Lewis^x ligands for selectins.¹³

Depending on the existing type of glycan extension, α 3/4-Fuc-transferase III (FUT3) transfers Fuc in either α 1-3 or α 1-4 linkage from GDP-Fuc to GlcNAc, to form either Lewis^x or Lewis^a structures (Table 2). Site-directed mutagenesis showed that only one amino acid (Trp111) appears to control the acceptor specificity of FUT3.²⁵⁸ Thus, replacement of Trp111 with Arg led to a change in specificity, while replacement of Asp112 by Glu resulted in a low enzyme activity. As expected of a terminally acting transferase, FUT3 localizes to the trans Golgi and trans Golgi network in baby hamster kidney cells.²⁵⁹ Enzyme lacking the cytoplasmic tail localized to earlier Golgi compartments, and this was associated with a reduction in the synthesis of Lewis^a determinants although the *in vitro* activity of the truncated enzyme was normal. This emphasizes the critical role of the proper localization of the enzyme in glycan biosynthesis. FUT3 and FUT4 are expressed in many tissues, including mucosa, and are likely to be involved in mucin synthesis. FUT3 is highly expressed throughout the gastrointestinal tract, lung, and kidneys. Upon treatment of bronchial mucosal explants with TNF- α , the activities and mRNA expression of FUT3 and FUT4 were increased.⁷⁶ Similarly, the inflammatory cytokines IL-6 and IL-8 stimulate the expression of FUT3 in bronchial explants, together with MUC4 and a number of other glycosyl- and sulfotransferases.²⁶⁰ The expression of FUT4 is also increased when HUVECs are stimulated with TNF- α .⁸¹ FUT4 undergoes cyclic changes in the human endometrium during the menstrual cycle and is responsible for the expression of selectin ligands at the time of implantation in the endometrium.²⁶¹ A suppression of FUT1 and FUT4 expression by short interfering RNA was found to decrease the expression of Lewis^y and to inhibit tumor growth of epidermoid carcinoma A431 cells in mice.²⁶²

FUT6 regulates sialyl-Lewis^x expression in breast tumors.²⁶³ Stimulation of hepatocellular carcinoma Huh-7 cells with IL-1 β caused a specific increase in the expression of FUT6, as well as that of sialyltransferase ST3Gal-IV.²⁶⁴ The mRNA expression preceded the increased expression of sialyl-Lewis^x. In leukocytes such as neutrophils and HEVs, FUT7 is the main Fuc-transferase involved in the synthesis of the leukocyte cell surface E-selectin ligands required for the inflammatory response and homing of leukocytes. Mice having a deletion of the FUT7 gene are severely deficient in leukocyte rolling and neutrophil mobilization due to the inability of leukocytes to bind to selectins. Interestingly, the decrease in P-selectin ligands in FUT7 (–/–) mice was associated with a reduction in atherosclerotic lesions.⁴⁸

The α 3-Fuc-transferases have interesting differences in specificity toward their acceptor substrates.⁷³ FUT3, 4, 5, and 6 act preferably on internal GlcNAc residues of *N*-acetylglucosamines as the substrate and are thus involved in the synthesis of internal or dimeric Lewis^x antigens.²²⁴ FUT3, 5, and 6 act on both neutral

and sialylated substrates and synthesize neutral or sialylated Lewis^x determinants. However, FUT7 requires a specific sequence of sugar addition in the assembly of sialyl-Lewis^x and acts exclusively on sialylated substrates. In contrast, human FUT9 acts only on neutral but not on sialylated substrates and has a preference for the subterminal GlcNAc residue at the nonreducing end. When FUT9 was overexpressed in HeLa cells, the enzyme was found to be localized to the trans Golgi and trans Golgi network. A deletion of the cytoplasmic domain caused the enzyme to be localized to the cis Golgi. This suggests that the cytoplasmic domain directs the Golgi localization of this enzyme.²⁶⁵

6.11.17 Blood Group Transferases

Human ABO blood groups and other carbohydrate epitopes can be found as terminal structures on red cell surface glycoproteins, but they are also commonly expressed on mucins. Animals may have blood groups O and A, or the linear B Gal α 1-3Gal epitope lacking Fuc. The ABO blood group epitopes as well as the activities of transferases synthesizing these antigens have been found to be altered in colonic and other tumor tissues and cells, and may be associated with different susceptibilities to disease.⁷⁴ Upon transplantation of animal tissues into humans, the linear B determinant acts as a xenotransplantation antigen leading to tissue rejection. The problem may be overcome by overexpression of α 2-Fuc-transferase that would compete with the linear B transferase, and thus reduce the linear B determinant.

The GalNAc α 1-3 residue of blood group A is added to the H determinant by blood group A α 3-GalNAc-transferase (EC 2.4.1.40), and the blood group B determinant similarly by blood group B α 3-Gal-transferase (EC 2.4.1.37). A related linear B α 3-Gal-transferase (EC 2.4.1.151) from nonprimates synthesizes the linear B determinant on terminal Gal residues without the Fuc α 1-2 substitution. These three enzymes are retaining transferases of the CAZy GT6 family with a GT-A fold (Table 4).

6.11.17.1 Blood Group A and B Transferases

The A and B transferases act on the same acceptor substrates, but differ in the specificity for the donor substrate, which is UDP-GalNAc for the A transferase and UDP-Gal for the B transferase. The amino acid sequences of these enzymes differ by only four amino acid residues.

The crystal structures of both A and B transferases have been determined as complexes with UDP and the Fuc α 1-2Gal-R acceptor.²⁶⁶ The transferases have a typical nucleotide sugar-binding site, and exist in several conformations.²⁶⁷ A study of chimeric A and B transferases shows that in the absence of substrates, nine specific C-terminal residues are disordered in an open form of the enzyme. A semiclosed form is seen bound to UDP, and a closed form bound to UDP and acceptor substrate. The basic groups (Glu211–Glu213) of the Dx₂D motif are coordinated to a Mn²⁺ ion that is required for activity. Mutations of the Met214 residue of the B transferase adjacent to the Dx₂D motif indicated that Met214 is essential in determining the proper conformation of Met266, required for activity, although the Dx₂D–Mn²⁺ coordination was not disturbed.²⁶⁸ The Met214Thr and Met214Val mutants exhibited both A and B transferase activities. Two amino acids appear to be in contact with the sugar moiety of the nucleotide sugar, but only one of these controls the donor specificity. In the A transferase, Leu266 has contact with GalNAc and allows the binding of the hydrophobic 2-*N*-acetyl group of GalNAc. In the B transferase, the larger Met266 binds to the 2-hydroxyl of Gal. Gly268 in the A transferase also contributes to donor specificity.²⁶⁹ The larger Ala268 in the B transferase contributes to the exclusion of the *N*-acetyl group of GalNAc. Glu303 is a possible nucleophile in the catalytic region responsible for the transfer of the donor sugar residue with inversion of configuration. A second nucleophilic attack²⁶⁶ with another inversion of configuration then may result in the retention of the α -linkage in the enzyme product. A mutant B transferase with a new specificity toward UDP-Glc was designed by theoretical modeling based on known carbohydrate–protein complex structures.²⁷⁰ Replacement of Ser185 with Asn, and Ser185 with Cys resulted in increased catalytic efficiency toward UDP-Glc and novel nucleotide sugar specificity.

6.11.17.2 α 3-Gal-Transferase That Synthesizes the Linear B Determinant

Mammals with the exception of humans and Old World monkeys synthesize the linear B determinant, which resembles the blood group B structure (**Table 2**). The α 3-Gal-transferase responsible for linear B determinant synthesis is related to AB transferases but in humans a pseudogene is found.²⁷¹ Bovine α 3-Gal-transferase has been crystallized with UMP, UDP, or UDP-Gal as donor substrate analogues and Glc, Gal, lactose, or Gal β 1-4GlcNAc as acceptor substrate analogues. The UDP-Gal complexes of the enzyme²⁷² suggest that the highly conserved Glu317 in the catalytic pocket may be the catalytic base, while Gln247 contributes to the catalysis.²⁷³ His280 is crucial in determining the UDP-Gal binding specificity. The Mn²⁺ ion coordinates the phosphate groups of UDP-Gal as well as the Asp residues of the Dx D motif (Asp225–Asp227).

6.11.17.3 Cad β 4-GalNAc-Transferase

The blood group Cad or Sd^a determinants found at the nonreducing termini of O-glycans are synthesized by members of the β 4-GalNAc-transferases that act on sialyl α 2-3Gal termini.^{274,275} Several transcripts of the Cad β 4-GalNAc-transferase (β 4GalNAc-T2) gene are expressed in Caco-2 cells and in colonic mucosa, as well as in kidney, stomach, intestine, and lymphocytes. The β 4-GalNAc-transferase activity is regulated during differentiation of cells, during development, and is often drastically reduced in cancer tissues, together with a loss of the Sd^a determinant. Restoration of the β 4GalNAc-T2 in gastrointestinal cancer cells led to an increase in the Sd^a determinant. In addition, the expression of selectin ligands was reduced, likely due to a competition for the synthesis of terminal epitopes, which was associated with decreased cancer cell adhesion to the endothelium and a suppressed metastatic potential.⁴²

6.11.17.4 α 4-GlcNAc-Transferase

Unusual O-glycan structures with GlcNAc α 1-4Gal– linkages occur in mucins from Brunner's gland and are attached to extended core 2 structures. Accessory glands of pancreaticobiliary tract and pancreatic ducts appear to express glycoproteins with these GlcNAc α 1-4Gal linkages. The α 4GlcNAc epitope was also found in a number of normal and neoplastic intestinal tissues.²⁷⁶ The structure may have protective and antimicrobial properties against *H. pylori*.²⁷⁷

A gene encoding an α 4-GlcNAc-transferase has been cloned from human stomach,²⁷⁸ and was found to be expressed in pancreatic and gastric tissues. The enzyme has a Dx D motif and acts on the galactosylated core 2 structure as an acceptor substrate. Since the mRNA levels of α 4-GlcNAc-transferase are high in the mononuclear fraction of peripheral blood in three out of four pancreatic cancer patients, the enzyme or the GlcNAc α 1-4Gal– epitope could be useful as a diagnostic tool for pancreatic cancer.²⁷⁹

6.11.18 Sulfotransferases

Mucin sulfation has a significant impact on the properties of O-glycans and their biological functions. The commonly found sulfate esters in O-glycans are Gal-3-sulfate and GlcNAc-6-sulfate. KSs attached to O-glycans can also have 6-sulfate esters of Gal. Lewis determinants with 6-sulfated Gal residues can form high-affinity ligands for selectins. Sulfotransferases synthesize sulfate ester groups from PAPS onto specific positions of Gal or GlcNAc residues of mucin O-glycans (**Table 4**). Like glycosyltransferases, sulfotransferases are type II membrane proteins in the Golgi and are specific for both sulfate donor and acceptor substrates.^{13,71,100,280}

Sulfotransferase activities are regulated during inflammatory stimulation or in pathological conditions. For example, the expression and activities of sulfotransferases may be decreased in tumor cells.^{100,186,281} Both Gal and GlcNAc sulfation may be aberrant in cancer cells (**Table 3**). The decrease in sulfation often leads to an increase in sialomucins with a high sialic acid content, relative to sulfomucins in cancer. Mucins isolated from bronchial, gastrointestinal, reproductive, and other tissues are sulfated. Bronchial mucins from CF patients are characterized by a high carbohydrate and sulfate content.^{74,82} Xia *et al.*⁷⁵ reported an increase in both sulfate and sialic acid in CF airway mucins, which could possibly result from hyperactive sulfotransferases and

sialyltransferases, due to the inflammatory stimuli. Another possible mechanism is a stimulation of glycosyltransferases that provide the backbone for the addition of sulfate, sialic acid, and other terminal structures. In the CF mouse model, no CFTR-related changes in O-glycosylation pathways could be detected, which suggests that alterations in CF, including mucin sulfation, are a consequence of the pathology associated with mutant CFTR.⁸² This is supported by the increases in sulfotransferase expression upon treatment of bronchial explants with inflammatory cytokines.^{76,260}

6.11.18.1 Gal-3-O-Sulfotransferases

The Gal-3-O-sulfotransferase family (EC 2.8.2.11) transfers sulfate to Gal residues and Gal3ST-2, Gal3ST-3, and Gal3ST-4 contribute to sulfation of O-glycans.^{13,100} Gal3ST-4 is ubiquitously expressed and appears to be the main enzyme responsible for mucin sulfation.²⁸² It prefers O-glycan core 1 and 2 substrates but also acts on type 1 and 2 chains. Gal3ST-2 acts preferably on type 2 chains and is responsible for the synthesis of 3'-sulfo-Lewis^x (Table 2),²⁸⁰ while Gal3ST-3 prefers N-acetylglucosamine chains and core 2, but not core 1, as a substrate. Gal3ST-2 can also utilize Gal β 1-3GlcNAc β 1-3Gal β 1-4Glc as an acceptor substrate. The mRNA expression of Gal3ST-2 is decreased in nonmucinous adenocarcinoma.⁵²

Sulfation at specific positions of O-glycans may block further reactions, and therefore regulates O-glycosylation pathways. For example, 3-O-sulfation of the Gal residue of core 1 blocks the addition of GlcNAc to GalNAc and the conversion to core 2.¹⁶⁶ However, a sulfate ester at the 6-position of GlcNAc allows the addition of a Gal β 1-4 residue to GlcNAc.

A Gal-6-sulfotransferase (Gal6ST) adds a sulfate ester to the 6-position of Gal.²⁸³ The enzyme is expressed in brain and cornea, and is involved in the synthesis of KSs as well as high-affinity selectin ligands.

6.11.18.2 GlcNAc-6-O-Sulfotransferases

GlcNAc-6-O-sulfotransferases (EC 2.8.2.21) acting on O-glycans include GlcNAc6ST-1–5.¹⁰⁰ The enzymes have different substrate specificities and expression patterns. GlcNAc6ST-1 is involved in the synthesis of the sulfo-sialyl-Lewis^x determinants on core 2 O-glycans in HEV that interact with selectins during the inflammatory response and homing of lymphocytes. GlcNAc6ST-2 can also synthesize the selectin ligand sulfo-sialyl-Lewis^x and is expressed in HEV. The enzyme acts on the GlcNAc residue of O-glycan core 2 and 3 structures. Deletion of both the GlcNAc6ST-1 and GlcNAc6ST-2 genes in mice⁴⁹ is associated with a deficiency of 6-sulfo-Lewis^x determinants in HEV, and consequently reduces homing and adhesion of lymphocytes to HEV.

A human intestinal GlcNAc6ST (I-GlcNAc6ST-3) acts on terminal GlcNAc residues of N-acetylglucosamine chains and core 2, but does not act on core 3 or KS. The expression of I-GlcNAc6ST-3 is decreased in adenocarcinoma, compared to normal mucosa, and this enzyme may be the major enzyme synthesizing gastrointestinal sulfomucins.⁵¹ An enzyme closely related to GlcNAc6ST-3 is human C-GlcNAc6ST, which is expressed in human brain and in corneal tissue and it synthesizes the sulfate esters of corneal KS. Patients with the hereditary condition macular corneal dystrophy type I lack sulfated KS in the serum and have mutations in C-GlcNAc6ST.⁵⁰ The condition is associated with bilateral loss of vision. Another sulfotransferase (GlcNAc6ST-5) that acts on core 2 and GlcNAc β -benzyl is also involved in KS synthesis. None of these enzymes are absolutely specific for O-glycans but have overlapping specificities.

6.11.19 Future Needs and Directions

The heterogeneous mixture of many O-glycan structures in mucins is a result of glycosyltransferase and sulfotransferase activities in the Golgi. The enzymes act in a linkage-specific fashion and their substrate specificities determine the structures of O-glycans. The sequential assembly of O-glycans is also controlled by many other factors including the localization of enzymes, proteins, and metal ion complexes, the expression levels, and posttranslational modifications. Cytokines, hormones, and agents altering cell differentiation and proliferation are also important regulatory factors of O-glycosylation. The mechanisms by which these factors

control glycosylation are extremely complex and still poorly understood. A less well-known phenomenon is the reorganization of the Golgi during cell differentiation, development, cell activation, apoptosis, and in disease. Since O-glycans are involved in critical aspects of the immune system, in inflammation, cancer, metastasis, and many other diseases, glycosyltransferases and sulfotransferases represent important targets for new therapeutic approaches. The focus of future research should be on the gene regulation of transferases, further characterization of enzyme protein structures and functions, and the organization of the site of O-glycan synthesis, the Golgi apparatus.

Acknowledgments

This work was supported by the Canadian Cystic Fibrosis Foundation. The author is grateful for the invaluable help of John Schutzbach.

Abbreviations

| | |
|----------------------------------|--------------------------------------|
| C1GalT | core 1 β 3-Gal-transferase |
| C2GnT | core 2 β 6-GlcNAc-transferase |
| C3β3GnT6 | core 3 β 3-GlcNAc-transferase |
| CAZy | carbohydrate active enzyme |
| CDG | congenital disorder of glycosylation |
| Fuc | fucose |
| FUT | Fuc-transferase |
| Gal | galactose |
| GalNAc | N-acetylgalactosamine |
| GalT | Gal-transferase |
| GlcNAc | N-acetylglucosamine |
| GlcNAc6ST | 6-sulfotransferase |
| GnT | GlcNAc-transferase |
| GT | glycosyltransferase |
| HEV | high endothelial venule |
| IGnT | I β 6-GlcNAc-transferase |
| KS | keratan sulfate |
| PAPS | 3'phosphoadenosine-5'-phosphosulfate |
| ppGalNAc-T | polypeptide GalNAc-transferase |
| ST3Gal | α 3-sialyltransferase |
| ST6GalNAc | α 6-sialyltransferase |
| STn | sialyl-Tn |
| T | transferase |
| TNF | tumor necrosis factor |
| VNTR | variable number of tandem repeat |

References

1. R. S. Haltiwanger; J. B. Lowe, *Annu. Rev. Biochem.* **2004**, *73*, 491.
2. I. Brockhausen; T. Anastassiades, *Expert Rev. Clin. Immunol.* **2008**.
3. T. Endo, *Glycoconj. J.* **2004**, *21*, 3.
4. R. J. Harris; M. W. Spellman, *Glycobiology* **1993**, *3*, 219.
5. J. Hofsteenge; K. G. Huwiler; B. Macek; D. Hess; J. Lawler; D. F. Mosher; J. Peter-Katalinic, *J. Biol. Chem.* **2001**, *276*, 6485.
6. M. A. Hollingsworth; B. J. Swanson, *Nat. Rev.* **2004**, *4*, 45.
7. Y. S. Kim; J. Gum; I. Brockhausen, *Glycoconj. J.* **1996**, *13*, 693.

8. T. A. Gerken, *Crit. Rev. Oral Biol. Med.* **1993**, *4*, 261.
9. F. Chang; H. L. Zhao; J. Phillips; G. Greenburg, *Cell. Immunol.* **2000**, *201*, 83.
10. I. Correa; T. Plunkett; A. Vlad; A. Mungul; J. Candelora-Kettel; J. M. Burchell; J. Taylor-Papadimitriou; O. J. Finn, *Immunology* **2003**, *108*, 32.
11. K. L. Carraway; S. A. Price-Schiavi; M. Komatsu; S. Jepson; A. Perez; C. A. Carraway, *J. Mammary Gland Biol. Neoplasia* **2001**, *6*, 323.
12. P. Chaturvedi; A. P. Singh; N. Moniaux; S. Senapati; S. Chakraborty; J. L. Meza; S. K. Batra, *Mol. Cancer Res.* **2007**, *5*, 309.
13. J. B. Lowe, *Curr. Opin. Cell Biol.* **2003**, *15*, 531.
14. T. A. Gerken; M. Gilmore; J. Zhang, *J. Biol. Chem.* **2002**, *277*, 7736.
15. Z. Yuan; S. Wong; A. Borrelli; M. A. Chung, *Biochem. Biophys. Res. Commun.* **2007**, *362*, 740.
16. S. J. Storr; L. Royle; C. J. Chapman; U. M. Hamid; J. F. Robertson; A. Murray; R. A. Dwek; P. M. Rudd, *Glycobiology* **2008**, *18*, 456.
17. K. Zhang; R. Sikut; G. C. Hansson, *Cell. Immunol.* **1997**, *176*, 158.
18. C. J. Jones; M. E. Ortiz; H. B. Croxatto; A. Manzur; G. Slevin; J. D. Aplin, *Biol. Reprod.* **2001**, *64*, 1535.
19. M. A. Tarp; H. Clausen, *Biochim. Biophys. Acta* **2008**, *1780*, 546.
20. M. Funes; J. K. Miller; C. Lai; K. L. Carraway, III; C. Sweeney, *J. Biol. Chem.* **2006**, *281*, 19310.
21. V. P. Ramsauer; C. A. C. Carraway; P. J. I. Salas; K. L. Carraway, *J. Biol. Chem.* **2003**, *278*, 30142.
22. A. Yokoyama; B. H. Shi; T. Kawai; H. Konishi; R. Andoh; H. Tachikawa; S. Ihara; Y. Fukui, *Biochem. Biophys. Res. Commun.* **2007**, *355*, 200.
23. O. Topaz; D. L. Shurman; R. Bergman; M. Indelman; P. Ratajczak; M. Mizrahi; Z. Khamaysi; D. Behar; D. Petronius; V. Friedman; I. Zelikovic; S. Raimer; A. Metzker; G. Richard; E. Sprecher, *Nat. Genet.* **2004**, *36*, 579.
24. K. A. Landers; M. J. Burger; M. A. Tebay; D. M. Puride; B. Scells; H. Samaratunga; M. F. Lavin; R. A. Gardiner, *Int. J. Cancer* **2005**, *114*, 950.
25. L. Xia; T. Ju; A. Westmuckett; G. An; L. Ivanciu; J. M. McDaniel; F. Lupu; R. D. Cummings; R. P. McEver, *J. Cell Biol.* **2004**, *164*, 451.
26. T. Ju; R. D. Cummings, *Nature* **2005**, *437*, 1252.
27. L. G. Ellies; S. Tsuboi; B. Petryniak; J. B. Lowe; M. Fukuda; J. D. Marth, *Immunity* **1998**, *9*, 881.
28. I. Brockhausen; W. Kuhns; H. Schachter; K. L. Matta; R. Sutherland; M. A. Baker, *Cancer Res.* **1991**, *51*, 1257.
29. E. Machida; J. Nakayama; J. Amano; M. Fukuda, *Cancer Res.* **2001**, *61*, 2226.
30. K. Shimodaira; J. Nakayama; N. Nakamura; O. Hasebe; T. Katsuyama; M. Fukuda, *Cancer Res.* **1997**, *57*, 5201.
31. S. Hagsiawa; C. Ohyama; T. Takahashi; M. Endoh; T. Moriya; J. Nakayama; Y. Arai; M. Fukuda, *Glycobiology* **2005**, *15*, 1016.
32. E. A. Higgins; K. A. Siminovitich; D. L. Zhuang; I. Brockhausen; K. Siminovitich; J. W. Dennis, *J. Biol. Chem.* **1991**, *266*, 6280.
33. F. Piller; F. Le Deist; K. I. Weinberg; R. Parkman; M. Fukuda, *J. Exp. Med.* **1991**, *173*, 1501.
34. Y. Nishio; C. E. Warren; J. A. Buczek-Thomas; J. Rulfs; D. Koya; L. P. Aiello; E. P. Feener; T. B. Miller, Jr.; J. W. Dennis; G. L. King, *J. Clin. Invest.* **1995**, *96*, 1759.
35. J. M. Yang; J. C. Byrd; B. B. Siddiki; Y. S. Chung; M. Okuno; M. Sowa; Y. S. Kim; K. L. Matta; I. Brockhausen, *Glycobiology* **1994**, *4*, 873.
36. T. Iwai; T. Kudo; R. Kawamoto; T. Kubota; A. Togayachi; T. Hiruma; T. Okada; T. Kawamoto; K. Morozumi; H. Narimatsu, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4572.
37. G. An; B. Wei; B. Xia; J. M. McDaniel; T. Ju; R. D. Cummings; J. Braun; L. Xia, *J. Exp. Med.* **2007**, *204*, 1417.
38. H. Ishida; A. Togayachi; T. Sakai; T. Iwai; T. Hiruma; T. Sato; R. Okubo; N. Inaba; T. Kudo; M. Gotoh; J. Shoda; N. Tanaka; H. Narimatsu, *FEBS Lett.* **2005**, *579*, 71.
39. M. Asano; S. Nakae; N. Kotani; N. Shirafuji; A. Nambu; N. Hashimoto; H. Kawashima; M. Hirose; M. Miyasaka; S. Takasaki; Y. Iwakura, *Blood* **2003**, *102*, 1678.
40. B. Hansske; C. Thiel; T. Lubke; M. Hasilik; S. Honing; V. Peters; P. H. Heidemann; G. F. Hoffmann; E. G. Berger; K. von Figura; C. Korner, *J. Clin. Invest.* **2002**, *109*, 725.
41. I. Brockhausen, *EMBO Rep.* **2006**, *7*, 1.
42. T. Dohi; Y. I. Kawamura, *Biochim. Biophys. Acta* **2008**, *1780*, 467.
43. J. M. Burchell; A. Mungul; J. Taylor-Papadimitriou, *J. Mammary Gland Biol. Neoplasia* **2001**, *6*, 355.
44. P. H. Wang; W. L. Lee; C. M. Juang; Y. H. Yang; W. H. Lo; C. R. Lai; S. L. Hsieh; C. C. Yuan, *Gynecol. Oncol.* **2005**, *99*, 631.
45. L. G. Ellies; D. Ditto; G. G. Levy; M. Wahrenbrock; D. Ginsburg; A. Varki; D. T. Le; J. D. Marth, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10042.
46. F. Schneider; W. Kemmer; W. Haensch; G. Franke; S. Gretschel; U. Karsten; P. M. Schlag, *Cancer Res.* **2001**, *61*, 4605.
47. J. Sun; J. Thurin; H. S. Cooper; P. Wang; M. Mackiewicz; Z. Steplewski; M. Blaszczyk-Thurin, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5724.
48. J. W. Homeister; A. Daugherty; J. B. Lowe, *Arterioscler. Thromb. Vasc. Biol.* **2004**, *24*, 1897.
49. K. Uchimura; J. M. Gauguier; M. S. Singer; D. Tsay; R. Kannagi; T. Muramatsu; U. H. von Andrian; S. D. Rosen, *Nat. Immunol.* **2005**, *6*, 1105.
50. T. O. Akama; K. Nishida; J. Nakayama; H. Watanabe; K. Ozaki; T. Nakamura; A. Dota; S. Kawasaki; Y. Inoue; N. Maeda; S. Yamamoto; T. Fujiwara; E. J. Thonar; Y. Shimomura; S. Kinoshita; A. Tanigami; M. N. Fukuda, *Nat. Genet.* **2000**, *26*, 237.
51. A. Seko; K. Nagata; S. Yonezawa; K. Yamashita, *Glycobiology* **2002**, *12*, 379.
52. A. Seko; K. Nagata; S. Yonezawa; K. Yamashita, *Jpn. J. Cancer Res.* **2002**, *93*, 507.
53. S. H. Itzkowitz; E. J. Bloom; T. S. Lau; Y. S. Kim, *Gut* **1992**, *33*, 518.
54. A. Pollex-Krüger; B. Meyer; R. Stuike-Prill; V. Sinnwell; K. L. Matta; I. Brockhausen, *Glycoconj. J.* **1993**, *10*, 365.
55. M. Dalziel; C. Whitehouse; I. McFarlane; I. Brockhausen; S. Gschmeissner; T. Schwientek; H. Clausen; J. M. Burchell; J. Taylor-Papadimitriou, *J. Biol. Chem.* **2001**, *276*, 11007.
56. D. K. Podolsky, *J. Biol. Chem.* **1985**, *260*, 15510.
57. C. Capon; E. Maes; J. C. Michalski; H. Leffler; Y. S. Kim, *Biochem. J.* **2001**, *358*, 657.
58. A. Kurosaka; H. Nakajima; I. Funakoshi; M. Matsuyama; T. Nagayo; I. Yamashina, *J. Biol. Chem.* **1983**, *258*, 11594.

59. E. F. Hounsell; A. M. Lawson; J. Feeney; H. C. Gooi; N. J. Pickering; M. S. Stoll; S. C. Lui; T. Feizi, *Eur. J. Biochem.* **1985**, *148*, 367.
60. J. H. Mutsaers; H. van Halbeek; J. F. Vliegthart; A. M. Wu; E. A. Kabat, *Eur. J. Biochem.* **1986**, *157*, 139.
61. W. G. Chai; E. F. Hounsell; G. C. Cashmore; J. R. Rosankiewicz; C. J. Bauer; J. Feeney; T. Feizi; A. M. Lawson, *Eur. J. Biochem.* **1992**, *203*, 257.
62. H. van Halbeek; A. M. Strang; M. Lhermitte; H. Rahmoune; G. Lamblin; P. Roussel, *Glycobiology* **1994**, *4*, 203.
63. A. Coppin; E. Maes; C. Flahaut; B. Coddeville; G. Strecker, *Eur. J. Biochem.* **1999**, *266*, 370.
64. Y. Guerardel; L. Balanzino; E. Maes; Y. Leroy; B. Coddeville; R. Oriol; G. Strecker, *Biochem. J.* **2001**, *357*, 167.
65. D. Florea; E. Maes; Y. Guérardel; A. Page; J. P. Zanetta; D. Cogalniceanu; G. Strecker, *Glycoconj. J.* **2006**, *23*, 377.
66. S. Tsuboi; M. Fukuda, *BioEssays*, **2001**, *23*, 46. 67. M. Fukuda, *Biochim. Biophys. Acta* **2002**, *1573*, 394.
68. J. C. Yeh; N. Hiraoka; B. Petryniak; J. Nakayama; L. G. Ellies; D. Rabuka; O. Hindsgaul; J. D. Marth; J. B. Lowe; M. Fukuda, *Cell* **2001**, *105*, 957.
69. P. Roussel; G. Lamblin, *Adv. Exp. Med. Biol.* **2003**, *535*, 17.
70. I. Brockhausen; J. Schutzbach; W. Kuhns, *Acta Anat.* **1998**, *161*, 36.
71. I. Brockhausen, *Biochim. Biophys. Acta* **1999**, *1473*, 67.
72. I. Brockhausen, *Adv. Exp. Med. Biol.* **2003**, *535*, 163.
73. I. Brockhausen. Biochemical Aspects: Biosynthesis of Mucin Type O-Glycans. In *Comprehensive Glycosciences: From Chemistry to Systems Biology*; J. P. Kamerling, Ed.; Acad. Press/Elsevier: Burlington, MA, 2007; pp 33–59.
74. I. Brockhausen; W. Kuhns, *Glycoproteins and Human Disease*; Medical Intelligence Unit, CRC Press and Mosby Year Book, Chapman & Hall: New York, 1997.
75. B. Xia; J. A. Royall; G. Damera; G. P. Sachdev; R. D. Cummings, *Glycobiology* **2005**, *15*, 747.
76. P. Delmotte; S. Degroote; J. J. Lafitte; G. Lamblin; J. M. Perini; P. Roussel, *J. Biol. Chem.* **2002**, *277*, 424.
77. I. Brockhausen; M. Lehotay; J. Yang; W. Qin; D. Young; J. Lucien; J. Coles; H. Paulsen, *Glycobiology* **2002**, *12*, 33.
78. X. Yang; M. Lehotay; T. Anastassiades; M. Harrison; I. Brockhausen, *Biochem. Cell Biol.* **2004**, *82*, 559.
79. X. Yang; J. Yip; T. Anastassiades; M. Harrison; I. Brockhausen, *Biochim. Biophys. Acta* **2006**, *1773*, 264.
80. X. Yang; J. Yip; M. Harrison; I. Brockhausen, *Internat. J. Biochem. Cell Biol.* **2007**, *40*, 471.
81. J. J. Garcia-Vallejo; W. Van Dijk; B. Van Het Hof; I. Van Die; M. A. Engelse; V. W. Van Hinsbergh; S. I. Gringhuis, *J. Cell. Physiol.* **2006**, *206*, 203.
82. I. Brockhausen; F. Vavasseur; X. Yang, *Glycoconj. J.* **2001**, *18*, 685.
83. S.-H. Leir; S. Parry; T. Palmi-Pallag; J. Evans; H. R. Morris; A. Dell; A. Harris, *Am. J. Respir. Cell Mol. Biol.* **2005**, *32*, 453.
84. S. A. Williams; L. Xia; R. D. Cummings; R. P. McEver; P. Stanley, *J. Cell Sci.* **2007**, *120*, 1341.
85. P. Delannoy; I. Kim; N. Emery; C. De Bolos; A. Verbert; P. Degand; G. Huet, *Glycoconj. J.* **1996**, *13*, 717.
86. N. Kojima; K. Handa; W. Newman; S. Hakomori, *Biochem. Biophys. Res. Commun.* **1992**, *182*, 1288.
87. E. Tian; K. G. Hagen; L. Shum; H. C. Hang; Y. Imbert; W. W. Young, Jr.; C. R. Bertozzi; L. A. Tabak, *J. Biol. Chem.* **2004**, *279*, 50382.
88. S. Wopereis; D. J. Lefeber; R. A. Wevers, *Clin. Chem.* **2006**, *52*, 574.
89. Y. Koda; H. Kimura; E. Mekada, *Blood* **1993**, *82*, 2915.
90. Y. Koda; M. Soejima; P. H. Johnson; E. Smart; H. Kimura, *Biochem. Biophys. Res. Commun.* **1997**, *238*, 21.
91. H. Schachter, *Cell. Mol. Life Sci.* **2001**, *58*, 1085.
92. J. Jaeken; G. Matthijs, *Annu. Rev. Genomics Hum. Genet.* **2007**, *8*, 261.
93. I. Martinez-Duncker; T. Dupré; V. Piller; F. Piller; J. J. Candelier; C. Trichet; G. Tchernia; R. Oriol; R. Mollicone, *Blood* **2005**, *105*, 2671.
94. R. Zeevaert; F. Foulquier; J. Jaeken; G. Matthijs, *Mol. Gen. Metab.* **2008**, *93*, 15.
95. C. Kranz; B. G. Ng; L. Sun; V. Sharma; E. A. Eklund; Y. Miura; D. Ungar; V. Lupashin; R. D. Winkel; J. F. Cipollo; C. E. Costello; E. Loh; W. Hong; H. H. Freeze, *Hum. Mol. Genet.* **2007**, *16*, 731.
96. X. Wu; R. A. Steet; O. Bohorov; J. Bakker; J. Newell; M. Krieger; L. Spaapen; S. Kornfeld; H. H. Freeze, *Nat. Med.* **2004**, *10*, 518.
97. I. Brockhausen. Glycosyltransferases Specific for the Mucin-Type O Glycans. In *Glycobiology*; C. Sansom, O. Markman, Eds.; Scion Publ Ltd: Woodbury, NY, 2007; pp 217–234.
98. I. Brockhausen; H. Schachter, Glycosyltransferases Involved in N- and O-Glycan Biosynthesis. In *Glycosciences: Status and Perspectives*; H. J. Gabius, S. Gabius, Eds; Chapman & Hall: Weinheim, 1997; pp 78–113.
99. I. Brockhausen, The Biosynthesis of O-Glycosylproteins. In *New Comprehensive Biochemistry*, J. Montreuil, J. Vliegthart, H. Schachter, Eds.; Elsevier: New York, NY, 1995; Vol. 29A, Glycoproteins, pp 201–259.
100. I. Brockhausen, *Biochem Soc. Trans.* **2003**, *31*, 318.
101. A. El-Battari; M. Prorok; K. Angata; S. Mathieu; M. Zerfaoui; E. Ong; M. Suzuki; D. Lombardo; M. Fukuda, *Glycobiology* **2003**, *13*, 941.
102. E. Bieberich; S. MacKinnon; J. Silva; D. D. Li; T. Tencomnao; L. Irwin; D. Kapitonov; R. K. Yu, *Biochemistry* **2002**, *41*, 11479.
103. F. H. Fenteany; K. J. Colley, *J. Biol. Chem.* **2005**, *280*, 5423.
104. R. Y. Patel; P. V. Balaji, *Glycobiology* **2006**, *16*, 108.
105. G. Egea; C. Francó; G. Gambús; T. Lesuffleur; A. Zweibaum; F. X. Real, *J. Cell Sci.* **1993**, *105*, 819.
106. J. Roth; Y. Wang; A. E. Eckhardt; R. L. Hill, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8935.
107. S. Röttger; J. White; H. H. Wandall; J. C. Olivo; A. Stark; E. P. Bennett; C. Whitehouse; E. G. Berger; H. Clausen; T. Nilsson, *J. Cell Sci.* **1998**, *111*, 45.
108. N. Yamaguchi; M. N. Fukuda, *J. Biol. Chem.* **1995**, *270*, 12170.
109. T. Nilsson; M. H. Hoe; P. Slusarewicz; C. Rabouille; R. Watson; F. Hunte; G. Watzele; E. G. Berger; G. Warren, *EMBO J.* **1994**, *13*, 562.
110. M. A. B. Axelsson; N. G. Karlsson; D. M. Steel; J. Ouwendijk; T. Nilsson; G. C. Hansson, *Glycobiology* **2001**, *11*, 633.
111. C. Breton; L. Snajdrova; C. Jeanneau; J. Koca; A. Imberty, *Glycobiology* **2006**, *16*, 29R.
112. D. Toki; M. Sarkar; B. Yip; F. Reck; D. Joziassse; M. Fukuda; H. Schachter; I. Brockhausen, *Biochem. J.* **1997**, *325*, 63.
113. K. G. Ten Hagen; T. A. Fritz; L. A. Tabak, *Glycobiology* **2003**, *13*, 1R.

114. K. G. Ten Hagen; D. T. Tran; T. A. Gerken; D. S. Stein; Z. Zhang, *J. Biol. Chem.* **2003**, *278*, 35039.
115. T. Schwientek; E. P. Bennett; C. Flores; J. Thacker; M. Hollmann; C. A. Reis; J. Behrens; U. Mandel; B. Keck; M. A. Schäfer; K. Haselmann; R. Zubarev; P. Roepstorff; J. M. Burchell; J. Taylor-Papadimitriou; M. A. Hollingsworth; H. Clausen, *J. Biol. Chem.* **2002**, *277*, 22623.
116. B. S. Wojczyk; M. M. Stwora-Wojczyk; F. K. Hagen; B. Striepen; H. C. Hang; C. R. Bertozzi; D. S. Roos; S. L. Spitalnik, *Mol. Biochem. Parasitol.* **2003**, *131*, 93.
117. E. Tian; K. G. Ten Hagen, *Glycobiology* **2006**, *16*, 83.
118. T. Hennet; F. K. Hagen; L. A. Tabak; J. D. Marth, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12070.
119. Y. Zhang; H. Iwasaki; H. Wang; T. Kudo; T. B. Kalka; T. Hennet; T. Kubota; L. Cheng; N. Inaba; M. Gotoh; A. Togayachi; J. Guo; H. Hisatomi; K. Nakajima; S. Nishihara; M. Nakamura; J. D. Marth; H. Narimatsu, *J. Biol. Chem.* **2003**, *278*, 573.
120. K. Shibao; H. Izumi; Y. Nakayama; R. Ohta; N. Nagata; M. Nomoto; K. Matsuo; Y. Yamada; K. Kitazato; H. Itoh; K. Kohno, *Cancer* **2002**, *94*, 1939.
121. K. Kato; C. Jeanneau; M. A. Tarp; A. Benet-Pagès; B. Lorenz-Depiereux; E. P. Bennett; U. Mandel; T. M. Strom; H. Clausen, *J. Biol. Chem.* **2006**, *281*, 18370.
122. S. A. Brooks; T. M. Carter; E. P. Bennett; H. Clausen; U. Mandel, *Acta Histochem.* **2007**, *109*, 273.
123. T. A. Fritz; J. H. Hurlley; L. B. Trinh; J. Shiloach; L. A. Tabak, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15307.
124. H. H. Wandall; F. Irazoqui; M. A. Tarp; E. P. Bennett; U. Mandel; H. Takeuchi; K. Kato; T. Irimura; G. Suryanarayanan; M. A. Hollingsworth; H. Clausen, *Glycobiology* **2007**, *17*, 374.
125. M. Tenno; S. Toba; F. J. Kézdy; A. P. Elhammer; A. Kurosaka, *Eur. J. Biochem.* **2002**, *269*, 4308.
126. M. Tenno; A. Saeki; F. J. Kézdy; A. P. Elhammer; A. Kurosaka, *J. Biol. Chem.* **2002**, *277*, 47088.
127. H. Hassan; C. A. Reis; E. P. Bennett; E. Mirgorodskaya; P. Roepstorff; M. A. Hollingsworth; J. Burchell; J. Taylor-Papadimitriou; H. Clausen, *J. Biol. Chem.* **2000**, *275*, 38197.
128. S. Wragg; F. K. Hagen; L. A. Tabak, *Biochem. J.* **1997**, *328*, 193.
129. M. Tenno; A. Saeki; A. P. Elhammer; A. Kurosaka, *FEBS J.* **2007**, *274*, 6037.
130. T. A. Fritz; A. Raman; L. A. Tabak, *J. Biol. Chem.* **2006**, *281*, 8613.
131. M. R. Pratt; H. C. Hang; K. G. Ten Hagen; J. Rarick; T. A. Gerken; L. A. Tabak; C. R. Bertozzi, *Chem. Biol.* **2004**, *11*, 1009.
132. S. Wragg; F. K. Hagen; L. A. Tabak, *J. Biol. Chem.* **1995**, *270*, 16947.
133. T. Kubota; T. Shiba; S. Sugioka; S. Furukawa; H. Sawaki; R. Kato; S. Wakatsuki; H. Narimatsu, *J. Mol. Biol.* **2006**, *359*, 708.
134. K. G. Ten Hagen; D. Tetaert; F. K. Hagen; C. Richet; T. M. Beres; J. Gagnon; M. Balys; B. VanWuyckhuysse; G. S. Bedi; P. Degand; L. A. Tabak, *J. Biol. Chem.* **1999**, *274*, 27867.
135. H. Iwasaki; Y. Zhang; K. Tachibana; M. Gotoh; N. Kikuchi; Y. D. Kwon; A. Togayachi; T. Kudo; T. Kubota; H. Narimatsu, *J. Biol. Chem.* **2003**, *278*, 5613.
136. F. K. Hagen; K. G. Ten Hagen; T. M. Beres; M. M. Balys; B. C. Van Wuyckhuysse; L. A. Tabak, *J. Biol. Chem.* **1997**, *272*, 13843.
137. I. Brockhausen; D. Toki; J. Brockhausen; S. Peters; T. Biefeldt; A. Kleen; H. Paulsen; M. Meldal; F. Hagen; L. Tabak, *Glycoconj. J.* **1996**, *13*, 849.
138. I. Brockhausen; G. Möller; A. Pollex-Krüger; V. Rutz; H. Paulsen; K. L. Matta, *Biochem. Cell Biol.* **1992**, *70*, 99.
139. T. Ju; R. D. Cummings; W. M. Canfield, *J. Biol. Chem.* **2002**, *277*, 169.
140. R. Singh; S. Subramanian; J. M. Rhodes; B. J. Campbell, *Glycobiology* **2006**, *16*, 594.
141. M. Granovsky; T. Biefeldt; S. Peters; H. Paulsen; M. Meldal; J. Brockhausen; I. Brockhausen, *Eur. J. Biochem.* **1994**, *221*, 1039.
142. I. Brockhausen; G. Möller; G. Merz; K. Adermann; H. Paulsen, *Biochemistry* **1990**, *29*, 10206.
143. H. Porowska; A. Paszkiewicz-Gadek; T. Anchim; S. Wolczynski; A. Gindzienski, *Int. J. Mol. Med.* **2004**, *13*, 459.
144. W. H. Yoon; H. D. Park; K. Lim; B. D. Hwang, *Biochem. Biophys. Res. Commun.* **1996**, *222*, 694.
145. T. Ju; R. D. Cummings, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16613.
146. R. Mueller; A. J. Huelsmeier; F. Altman; K. Ten Hagen; M. Tiemeyer; T. Hennet, *FEBS J.* **2005**, *272*, 4295.
147. W. S. Alexander; E. M. Viney; J.-G. Zhang; D. Metcalf; M. Kauppi; C. Hyland; M. R. Carpinelli; W. Stevenson; B. A. Croker; A. A. Hilton; S. Ellis; C. Selan; H. H. Nandurkar; C. C. Goodnow; B. T. Kile; N. A. Nicola; A. W. Roberts; D. J. Hilton, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16442.
148. S. A. Williams; P. Stanley, *FASEB J.* **2008**, *22*, 2273.
149. J. Novak; Z. Moldoveanu; M. B. Renfrow; T. Yanagihara; H. Suzuki; M. Raska; S. Hall; R. Brown; W. Q. Huang; A. Goepfert; M. Kilian; K. Poulsen; M. Tomana; R. J. Wyatt; B. A. Julian; J. Mestecky, *Contrib. Nephrol.* **2007**, *157*, 134.
150. T. Inoue; H. Sugiyama; Y. Kikumoto; N. Fukuoka; Y. Maeshima; H. Hattori; K. Fukushima; K. Nishizaki; Y. Hiki; H. Makino, *Contrib. Nephrol.* **2007**, *157*, 120.
151. I. Brockhausen; J. Yang; N. Dickinson; S. Ogata; S. Itzkowitz, *Glycoconj. J.* **1998**, *15*, 595.
152. T. Ju; G. S. Lanneau; T. Gautam; Y. Wang; B. Xia; S. R. Stowell; M. T. Willard; W. Wang; J. Y. Xia; R. E. Zuna; Z. Laszik; D. M. Benbrook; M. H. Hanigan; R. D. Cummings, *Cancer Res.* **2008**, *68*, 1636.
153. I. Brockhausen; J. Yang; J. Burchell; C. Whitehouse; J. Taylor-Papadimitriou, *Eur. J. Biochem.* **1995**, *233*, 607.
154. T. Schwientek; J. C. Yeh; S. B. Levery; B. Keck; G. Merckx; A. G. van Kessel; M. Fukuda; H. Clausen, *J. Biol. Chem.* **2000**, *275*, 11106.
155. T. Schwientek; M. Nomoto; S. B. Levery; G. Merckx; A. G. van Kessel; E. P. Bennett; M. A. Hollingsworth; H. Clausen, *J. Biol. Chem.* **1999**, *274*, 4504.
156. I. Brockhausen; K. L. Matta; J. Orr; H. Schachter, *Biochemistry* **1985**, *24*, 1866.
157. I. Brockhausen; K. L. Matta; J. Orr; H. Schachter; A. H. L. Koenderman; D. H. van den Eijnden, *Eur. J. Biochem.* **1986**, *157*, 463.
158. A. Suzuki; S. Yoshioka; M. Sekine; H. Yonekawa; M. Takenaka; R. Kannagi, *Glycoconj. J.* **2004**, *20*, 151.
159. V. R. Falkenberg; K. Alvarez; C. Roman; N. Fregien, *Glycobiology* **2003**, *13*, 411.
160. V. R. Falkenberg; N. Fregien, *Glycoconj. J.* **2007**, *24*, 511.
161. D. Williams; H. Schachter, *J. Biol. Chem.* **1980**, *255*, 11247.
162. F. Piller; V. Piller; R. I. Fox; M. Fukuda, *J. Biol. Chem.* **1988**, *263*, 15146.
163. C. Whitehouse; J. Burchell; S. Gschmeissner; I. Brockhausen; K. Lloyd; J. Taylor-Papadimitriou, *J. Cell Biol.* **1997**, *137*, 1229.
164. D. Skrincosky; R. Kain; A. El-Battari; M. Exner; D. Kerjaschki; M. Fukuda, *J. Biol. Chem.* **1997**, *272*, 22695.

165. M. Zerfaoui; M. Fukuda; C. Langlet; S. Mathieu; M. Suzuki; D. Lombardo; A. El-Battari, *Glycobiology* **2002**, *12*, 15.
166. W. Kuhns; V. Rutz; H. Paulsen; K. L. Matta; M. A. Baker; M. Barner; M. Granovsky; I. Brockhausen, *Glycoconj. J.* **1993**, *10*, 381.
167. X. Yang; W. Qin; M. Lehotay; D. Toki; P. Dennis; J. S. Schutzbach; I. Brockhausen, *Biochim. Biophys. Acta* **2003**, *1648*, 62.
168. W. Kuhns; R. Jain; K. L. Matta; H. Paulsen; M. A. Baker; R. Geyer; I. Brockhausen, *Glycobiology* **1995**, *5*, 689.
169. J. E. Pak; P. Arnoux; S. Zhou; P. Sivarajah; M. Satkunarajah; X. Xing; J. M. Rini, *J. Biol. Chem.* **2006**, *281*, 26693.
170. M. Prorok-Hamon; F. Notel; S. Mathieu; C. Langlet; M. Fukuda; A. El-Battari, *Biochem. J.* **2005**, *391*, 491.
171. T.-Y. Yen; B. A. Macher; S. Bryson; X. Chang; I. Tvaroska; R. Tse; S. Takeshita; A. M. Lew; A. Datti, *J. Biol. Chem.* **2003**, *278*, 45864.
172. I. Brockhausen; P. Romero; A. Herscovics, *Cancer Res.* **1991**, *5*, 3136.
173. E. Ong; M. Suzuki; F. Berlot; Y. C. Yeh; I. Franceschini; K. Angata; O. Hindsgaul; M. Fukuda, *J. Biol. Chem.* **2002**, *277*, 18182.
174. M. Martinez; M. Joffraud; S. Giraud; B. Baisse; M. P. Bernimoulin; M. Schapira; O. Spertini, *J. Biol. Chem.* **2005**, *280*, 5378.
175. M. P. Bernimoulin; X.-L. Zeng; C. Abbal; S. Giraud; M. Martinez; O. Michielin; M. Schapira; O. Spertini, *J. Biol. Chem.* **2003**, *278*, 37.
176. Z. S. Kawai; T. K. Johnson; S. Natunen; J. B. Lowe; R. D. Cummings, *Glycobiology* **2008**, *18*, 441.
177. J. M. Gauguier; S. D. Rosen; J. D. Marth; U. H. von Andrian, *Blood* **2004**, *104*, 4104.
178. N. Hiraoka; H. Kawashima; B. Petryniak; J. Nakayama; J. Mitoma; J. D. Marth; J. B. Lowe; M. Fukuda, *J. Biol. Chem.* **2004**, *279*, 3058.
179. S. Julien; M. J. Grimshaw; M. Sutton-Smith; J. Coleman; H. R. Morris; A. Dell; J. Taylor-Papadimitriou; J. M. Burchell, *J. Immunol.* **2007**, *179*, 5701.
180. M. Galvan; S. Tsuboi; M. Fukuda; L. G. Baum, *J. Biol. Chem.* **2000**, *275*, 16730.
181. H. F. Valenzuela; K. E. Pace; P. V. Cabrera; R. White; K. Porvari; H. Kaija; P. Vihko; L. G. Baum, *Cancer Res.* **2007**, *67*, 6155.
182. R. Chibber; B. M. Mahmud; D. Coppini; E. Christ; E. M. Kohner, *Diabetes* **2000**, *17*, 24.
183. A. Orlacchio; P. Sarchielli; V. Gallai; A. Datti; C. Saccardi; C. A. Palmerini, *J. Neurol. Sci.* **1997**, *151*, 177.
184. D. A. Carlow; S. Y. Corbel; M. J. Williams; H. J. Ziltener, *J. Immunol.* **2001**, *167*, 6841.
185. T. Iwai; N. Inaba; A. Naundorf; Y. Zhang; M. Gotoh; H. Iwasaki; T. Kudo; A. Togayachi; Y. Ishizuka; H. Nakanishi; H. Narimatsu, *J. Biol. Chem.* **2002**, *277*, 12802.
186. F. Vavasseur; K. Dole; J. Yang; K. L. Matta; N. Myerscough; A. Corfield; C. Paraskeva; I. Brockhausen, *Eur. J. Biochem.* **1994**, *222*, 415.
187. F. Vavasseur; J. Yang; K. Dole; H. Paulsen; I. Brockhausen, *Glycobiology* **1995**, *5*, 351.
188. I. Brockhausen; J. Yang; M. Lehotay; S. Ogata; S. Itzkowitz, *Biol. Chem.* **2001**, *382*, 219.
189. J. C. Yeh; E. Ong; M. Fukuda, *J. Biol. Chem.* **1999**, *274*, 3215.
190. M. C. Huang; H. Y. Chen; H. C. Huang; J. Huang; J. T. Liang; T. L. Shen; N. Y. Lin; C. C. Ho; I. M. Cho; S. M. Hsu, *Oncogene* **2006**, *25*, 3267.
191. P. A. Ropp; M. R. Little; P. W. Cheng, *J. Biol. Chem.* **1991**, *266*, 23863.
192. N. Markine-Goriaynoff; L. Gillet; O. A. Karlsen; L. Haarr; F. Minner; P. P. Pastoret; M. Fukuda; A. Vanderplasschen, *J. Gen. Virol.* **2004**, *85* (Pt. 2), 355.
193. D. S. McBride; I. Brockhausen; F. W. Kan, *Biochim. Biophys. Acta* **2005**, *1721*, 107.
194. Y. Andersch-Bjoerkman; K. A. Thomsson; J. M. Holmen-Larsson; E. Ekerhovd; G. C. Hansson, *Mol. Cell. Proteomics* **2007**, *6*, 708.
195. P. V. Beum; D. R. Bastola; P. W. Cheng, *Am. J. Respir. Cell Mol. Biol.* **2003**, *29*, 48.
196. S. Yazawa; S. A. Abbas; R. Madiyalakan; J. J. Barlow; K. L. Matta, *Carbohydr. Res.* **1986**, *149*, 241.
197. I. Brockhausen; D. Williams; K. L. Matta; J. Orr; H. Schachter, *Can. J. Biochem. Cell Biol.* **1983**, *61*, 1322.
198. J. Mitoma; B. Petryniak; N. Hiraoka; J. C. Yeh; J. B. Lowe; M. Fukuda, *J. Biol. Chem.* **2003**, *278*, 9953.
199. A. Seko; K. Yamashita, *Glycobiology* **2005**, *15*, 943.
200. A. Seko; K. Yamashita, *FEBS Lett.* **2004**, *556*, 216.
201. M. F. Bierhuizen; M. G. Mattei; M. Fukuda, *Genes Dev.* **1993**, *7*, 468.
202. N. Inaba; T. Hiruma; A. Togayachi; H. Iwasaki; X. H. Wang; Y. Furukawa; R. Sumi; T. Kudo; K. Fujimura; T. Iwai; M. Gotoh; M. Nakamura; H. Narimatsu, *Blood* **2003**, *101*, 2870.
203. G. Y. Chen; H. Muramatsu; M. Kondo; N. Kurosawa; Y. Miyake; N. Takeda; T. Muramatsu, *Mol. Cell. Biol.* **2005**, *25*, 7828.
204. D. Zhou, *Curr. Protein Pept. Sci.* **2003**, *4*, 1.
205. M. Gotoh; T. Sato; K. Kiyohara; A. Kameyama; N. Kikuchi; Y. D. Kwon; Y. Ishizuka; T. Iwai; H. Nakanishi; H. Narimatsu, *FEBS Lett.* **2004**, *562*, 134.
206. T. Sato; M. Gotoh; K. Kiyohara; A. Kameyama; T. Kubota; N. Kikuchi; Y. Ishizuka; H. Iwasaki; A. Togayachi; T. Kudo; T. Ohkura; H. Nakanishi; H. Narimatsu, *J. Biol. Chem.* **2003**, *278*, 47534.
207. T. Hiruma; A. Togayachi; K. Okamura; T. Sato; N. Kikuchi; Y. D. Kwon; A. Nakamura; K. Fujimura; M. Gotoh; K. Tachibana; Y. Ishizuka; T. Noce; H. Nakanishi; H. Narimatsu, *J. Biol. Chem.* **2004**, *279*, 14087.
208. M. Amado; R. Almeida; T. Schwientek; H. Clausen, *Biochim. Biophys. Acta* **1999**, *1473*, 35.
209. B. Y. Sheares; D. M. Carlson, *J. Biol. Chem.* **1983**, *258*, 9893.
210. D. Zhou; E. G. Berger; T. Hennet, *Eur. J. Biochem.* **1999**, *263*, 571.
211. M. Amado; R. Almeida; F. Carneiro; S. B. Levery; E. H. Holmes; M. Nomoto; M. A. Hollingsworth; H. Hassan; T. Schwientek; P. A. Nielsen; E. P. Bennett; H. Clausen, *J. Biol. Chem.* **1998**, *273*, 12770.
212. J. Holgersson; J. Löfling, *Glycobiology* **2006**, *16*, 584.
213. L. Mare; M. Trinchera, *Eur. J. Biochem.* **2004**, *271*, 186.
214. S. Isshiki; A. Togayachi; T. Kudo; S. Nishihara; M. Watanabe; T. Kubota; M. Kitajima; N. Shiraiishi; K. Sasaki; T. Andoh; H. Narimatsu, *J. Biol. Chem.* **1999**, *274*, 12499.
215. P. K. Qasba; B. Ramakrishnan; E. Boeggeman, *Trends Biochem. Sci.* **2005**, *30*, 53.
216. I. Brockhausen, *Drug News Perspect.* **2006**, *19*, 401.
217. B. Ramakrishnan; P. K. Qasba, *J. Mol. Biol.* **2001**, *310*, 205.
218. P. A. Gleeson; R. D. Teasdale; J. Burke, *Glycoconj. J.* **1994**, *11*, 381.

219. J. Hathaway, *J. Mammary Gland Biol. Neoplasia* **2003**, *8*, 421.
220. P. K. Qasba; B. Ramakrishnan; E. Boeggeman, *Curr. Drug Targets* **2008**, *9*, 292.
221. R. Mori; T. Kondo; T. Nishie; T. Ohshima; M. Asano, *Am. J. Pathol.* **2004**, *164*, 1303.
222. N. Kotani; M. Asano; N. Inoue; Y. Iwakura; S. Takasaki, *Arch. Biochem. Biophys.* **2004**, *426*, 258–265.
223. Y. Shen; J. Tiralongo; G. Kohla; R. Schauer, *Biol. Chem.* **2004**, *385*, 145.
224. J. R. Brown; M. M. Fuster; R. Li; N. Varki; C. A. Glass; J. D. Esko, *Clin. Cancer Res.* **2006**, *12*, 2894.
225. T. Kudo; Y. Ikehara; A. Togayachi; K. Morozumi; M. Watanabe; M. Nakamura; S. Nishihara; H. Narimatsu, *Lab. Invest.* **1998**, *78*, 797.
226. T. Sato; K. Furukawa, *J. Biol. Chem.* **2004**, *279*, 39574.
227. L. N. Gastinel; C. Cambillau; Y. Bourne, *EMBO J.* **1999**, *18*, 3546.
228. B. Ramakrishnan; P. V. Balaji; P. K. Qasba, *J. Mol. Biol.* **2002**, *318*, 491.
229. I. Brockhausen; M. Benn; S. Bhat; S. Marone; J. G. Riley; P. Montoya-Peleaz; J. Z. Vlahakis; H. Paulsen; J. S. Schutzbach; W. A. Szarek, *Glycoconj. J.* **2006**, *23*, 523.
230. D. Aoki; H. E. Appert; D. Johnson; S. S. Wong; M. N. Fukuda, *EMBO J.* **1990**, *9*, 3171.
231. U. Yabe; C. Sato; T. Matsuda; K. Kitajima, *J. Biol. Chem.* **2003**, *278*, 13875.
232. A. K. Datta; J. C. Paulson, *Indian J. Biochem. Biophys.* **1997**, *34*, 157.
233. A. Harduin-Lepers; D. C. Stokes; W. F. Steelant; B. Samyn-Petit; M. A. Krzewinski-Recchi; V. Vallejo-Ruiz; J. P. Zanetta; C. Auge; P. Delannoy, *Biochem. J.* **2000**, *352* (Pt. 1), 37.
234. A. Harduin-Lepers; R. Mollicone; P. Delannoy; R. Oriol, *Glycobiology* **2005**, *15*, 805.
235. P. A. Videira; I. F. Amado; H. J. Crespo; M. C. Algueró; F. Dall'olio; M. G. Cabral; H. Trindade, *Glycoconj. J.* **2008**, *25*, 259.
236. V. Vallejo-Ruiz; R. Haque; A.-M. Mir; T. Schwientek; U. Mandel; R. Cacan; P. Delannoy; A. Harduin-Lepers, *Biochim. Biophys. Acta* **2001**, *1549*, 161.
237. C. Jeanneau; V. Chazalet; C. Augé; D. M. Soumpasis; A. Harduin-Lepers; P. Delannoy; A. Imberty; C. Breton, *J. Biol. Chem.* **2004**, *279*, 13461.
238. J. J. Priatel; D. Chui; N. Hiraoka; C. J. Simmons; K. B. Richardson; D. M. Page; M. Fukuda; N. M. Varki; J. D. Marth, *Immunity* **2000**, *12*, 273.
239. M. Sperandio; D. Frommhold; I. Babushkina; L. G. Ellies; T. S. Olson; M. L. Smith; B. Fritzsching; E. Pauly; D. F. Smith; R. Nobiling; O. Linderkamp; J. D. Marth; K. Ley, *Eur. J. Immunol.* **2006**, *36*, 3207.
240. A. Mungul; L. Cooper; I. Brockhausen; K. Ryder; U. Mandel; H. Clausen; A. Rughetti; D. W. Miles; J. Taylor-Papadimitriou; J. M. Burchell, *Internat. J. Oncol.* **2004**, *25*, 937.
241. N. Kurosawa; N. Kojima; M. Inoue; T. Hamamoto; S. Tsuji, *J. Biol. Chem.* **1994**, *269*, 19048.
242. B. Samyn-Petit; M. A. Krzewinski-Recchi; W. F. Steelant; P. Delannoy; A. Harduin-Lepers, *Biochim. Biophys. Acta* **2000**, *1474*, 201.
243. A. Tsuchida; M. Ogiso; Y. Nakamura; M. Kiso; K. Furukawa; K. Furukawa, *J. Biochem.* **2005**, *138*, 237.
244. Y. Ikehara; N. Kojima; N. Kurosawa; T. Kudo; M. Kono; S. Nishihara; S. Issiki; K. Morozumi; S. Itzkowitz; T. Tsuda; S. I. Nishimura; S. Tsuji; H. Narimatsu, *Glycobiology* **1999**, *9*, 1213.
245. M. Clement; J. Rocher; G. Loirand; J. Le Pendu, *J. Cell Sci.* **2004**, *117*, 5059.
246. R. Sewell; M. Baeckstroem; M. Dalziel; S. Gschmeissner; H. Karlsson; T. Noll; J. Gaetgens; H. Clausen; G. Hansson; J. Burchell; J. Taylor-Papadimitriou, *J. Biol. Chem.* **2006**, *281*, 3586.
247. S. Julien; C. Lagadec; M. Krzewinski-Recchi; G. Courtand; X. Le Bourhis; P. Delannoy, *Breast Cancer Res. Treat.* **2005**, *90*, 77.
248. S. Julien; E. Adriaenssens; K. Ottenberg; A. Furlan; G. Courtand; A. S. Vercoutter-Edouart; F. G. Hanisch; P. Delannoy; X. Le Bourhis, *Glycobiology* **2006**, *16*, 54.
249. M. Teinturier-Lelievre; S. Julien; S. Juliant; Y. Guerardel; M. Duonor-Cerutti; P. Delannoy; A. Harduin-Lepers, *Biochem. J.* **2005**, *392*, 665.
250. T. Lübke; T. Marquardt; A. Etzioni; E. Hartmann; K. von Figura; C. Körner, *Nat. Genet.* **2001**, *28*, 73.
251. K. Lühn; M. K. Wild; M. Eckhardt; R. Gerardy-Schahn; D. Vestweber, *Nat. Genet.* **2001**, *28*, 69.
252. R. D. Larsen; L. K. Ernst; R. P. Nair; J. B. Lowe, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6674.
253. R. J. Kelly; S. Rouquier; D. Giorgi; G. G. Lennon; J. B. Lowe, *J. Biol. Chem.* **1995**, *270*, 4640.
254. M. Kaneko; S. Nishihara; N. Shinya; T. Kudo; H. Iwasaki; T. Seno; Y. Okubo; H. Narimatsu, *Blood* **1997**, *90*, 839.
255. C. Goupille; S. Marionneau; V. Bureau; F. Hallouin; M. Meichenin; J. Rocher; J. Le Pendu, *Glycobiology* **2000**, *10*, 375.
256. B. Lin; Y. Hayashi; M. Saito; Y. Sakakibara; M. Yanagisawa; M. Iwamori, *Biochim. Biophys. Acta* **2000**, *1487*, 275.
257. F. I. Yamamoto; H. Clausen; T. White; J. Marken; S.-I. Hakomori, *Nature* **1990**, *345*, 229.
258. F. Dupuy; J. M. Petit; R. Mollicone; R. Oriol; R. Julien; A. Maftah, *J. Biol. Chem.* **1999**, *274*, 12257.
259. V. L. Sousa; C. Brito; J. Costa, *Biochim. Biophys. Acta* **2004**, *1675*, 95.
260. S. Groux-Degroote; M. A. Krzewinski-Recchi; A. Cazet; A. Vincent; S. Lehoux; J. J. Lafitte; I. Van Seuningen; P. Delannoy, *Biochem. J.* **2008**, *410*, 213.
261. A. Ponnampalam; P. Rogers, *Reproduction* **2008**, *136*, 117.
262. Z. Zhang; P. Sun; J. Liu; L. Fu; J. Yan; Y. Liu; L. Yu; X. Wang; Q. Yan, *Biochim. Biophys. Acta* **2008**, *1783*, 287.
263. N. Matsuura; T. Narita; N. Hiraiwa; M. Hiraiwa; H. Murai; T. Iwase; T. Funahashi; T. Imai; H. Takagi; R. Kannagi, *Int. J. Oncol.* **1998**, *12*, 1157.
264. K. Higai; N. Miyazaki; Y. Azuma; K. Matsumoto, *FEBS Lett.* **2006**, *580*, 6069.
265. C. Brito; S. Kandzia; T. Graça; H. S. Conradt; J. Costa, *Biochimie* **2008**, *90*, 1279.
266. S. I. Patenaude; N. O. Seto; S. N. Borisova; A. Szpacenko; S. L. Marcus; M. M. Palcic; S. V. Evans, *Nat. Struct. Biol.* **2002**, *9*, 685.
267. J. A. Alfaro; R. B. Zheng; M. Persson; J. A. Letts; R. Polakowski; Y. Bai; S. N. Borisova; N. O. Seto; T. L. Lowary; M. M. Palcic; S. V. Evans, *J. Biol. Chem.* **2008**, *283*, 10097.
268. M. Persson; J. A. Letts; B. Hosseini-Maaf; S. N. Borisova; M. M. Palcic; S. V. Evans; M. L. Olsson, *J. Biol. Chem.* **2007**, *282*, 9564.
269. F. Yamamoto; P. D. McNeill, *J. Biol. Chem.* **1996**, *271*, 10515.
270. T. Nakahara; O. Hindsgaul; M. M. Palcic; S. Nishimura, *Protein Eng. Des. Sel.* **2006**, *19*, 571.

271. D. H. Joziassse; J. H. Shaper; E. W. Jabs; N. L. Shaper, *J. Biol. Chem.* **1991**, 266, 6991.
272. L. N. Gastinel; C. Bignon; A. K. Misra; O. Hindsgaul; J. H. Shaper; D. H. Joziassse, *EMBO J.* **2001**, 20, 638.
273. Y. Zhang; G. J. Swaminathan; A. Deshpande; E. Boix; R. Natesh; Z. Xie; K. R. Acharya; K. Brew, *Biochemistry* **2003**, 42, 13512.
274. L. Lo Presti; E. Cabuy; M. Chiricolo; F. Dall'Olio, *J. Biochem.* **2003**, 134, 675.
275. M. D. Montiel; M. A. Krzewinski-Recchi; P. Delannoy; A. Harduin-Lepers, *Biochem. J.* **2003**, 373 (Pt. 2), 369.
276. N. Nakamura; H. Ota; T. Katsuyama; T. Akamatsu; K. Ishihara; M. Kurihara; K. Hotta, *J. Histochem. Cytochem.* **1998**, 46, 793.
277. M. Kawakubo; Y. Ito; Y. Okimura; M. Kobayashi; K. Sakura; S. Kasama; M. N. Fukuda; M. Fukuda; T. Katsuyama; J. Nakayama, *Science* **2004**, 305, 1003.
278. J. Nakayama; J. C. Yeh; A. K. Misra; S. Ito; T. Katsuyama; M. Fukuda, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 8991.
279. S. Ishizone; K. Yamauchi; S. Kawa; T. Suzuki; F. Shimizu; O. Harada; A. Sugiyama; S. Miyagawa; M. Fukuda; J. Nakayama, *Cancer Sci.* **2006**, 96, 119.
280. K. Uchimura; S. D. Rosen, *Trends Immunol.* **2006**, 27, 559.
281. E. V. Chandrasekaran; J. Xue; C. Piskorz; R. D. Locke; K. Toth; H. K. Slocum; K. L. Matta, *J. Cancer Res. Clin. Oncol.* **2007**, 133, 599.
282. A. Seko; S. Hara-Kuge; K. Yamashita, *J. Biol. Chem.* **2001**, 276, 25697.
283. A. Bistrup; S. Bhakta; J. K. Lee; Y. Y. Belov; M. D. Gunn; F. R. Zuo; C. C. Huang; R. Kannagi; S. D. Rosen; S. Hemmerich, *J. Cell Biol.* **1999**, 145, 899.

Biographical Sketch



Inka Brockhausen obtained degrees in biochemistry and nutrition and food sciences from the University of Toronto. Under the guidance of Harry Schachter and with a studentship of the Cystic Fibrosis Foundation she studied novel pathways in the biosynthesis of mucin-type O-glycans and received her Ph.D. in biochemistry in 1985. She then received a postdoctoral fellowship from the Medical Research Council to examine N-glycosylation pathway and structures with Jeremy Carver. She characterized glycosyltransferases related to cancer and cystic fibrosis at the Hospital for Sick Children, Toronto and received a Scientist Award from the Medical Research Council. In 1999 the Brockhausen laboratory was established at the Department of Medicine, Division of Rheumatology, Queen's University, Kingston, supported by a Research Scientist Award from the Arthritis Society and a Premiere's Research Excellence Award. Professor Brockhausen is associated with the Human Mobility Research Centre as well as the Gastrointestinal Disease Research Unit at Queen's University, and her research program focuses on the glycobiology of inflammation, cystic fibrosis, and arthritis, as well as O antigen synthesis in Gram-negative bacteria. A major effort is made in the development of anti-inflammatory and antibacterial strategies and the design of glycosyltransferase inhibitors that have potential as therapeutic agents.

6.12 Bacterial Protein Glycosylation

Amirreza Faridmayer and Mario F. Feldman, University of Alberta, Edmonton, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.12.1 | Introduction | 351 |
| 6.12.2 | Protein N-Glycosylation and O-Glycosylation at a Glance | 352 |
| 6.12.3 | General Protein N-Glycosylation System in <i>Campylobacter jejuni</i> | 353 |
| 6.12.4 | Protein O-Glycosylation in Gram-Negative Bacteria | 355 |
| 6.12.4.1 | Flagellar O-Glycosylation in <i>Campylobacter</i> | 355 |
| 6.12.4.2 | O-Glycosylation Systems in <i>Pseudomonas aeruginosa</i> | 358 |
| 6.12.4.2.1 | Flagellin glycosylation | 358 |
| 6.12.4.2.2 | Type IV pilin glycosylation | 360 |
| 6.12.4.3 | O-Glycosylation of Type IV Pilin in <i>Neisseria</i> | 363 |
| 6.12.4.4 | Glycosylation of Autotransporters in <i>Escherichia coli</i> | 366 |
| 6.12.4.5 | Glycosylation in <i>Porphyromonas gingivalis</i> | 367 |
| 6.12.5 | Protein O-Glycosylation in Gram-Positive Bacteria | 367 |
| 6.12.6 | Protein O-Mannosylation in Mycobacteria | 368 |
| 6.12.7 | Challenges in the Study of Bacterial Protein Glycosylation | 369 |
| 6.12.8 | Glycoengineering in Bacteria | 369 |
| 6.12.8.1 | Glycoengineering in <i>Escherichia coli</i> Using the Bacterial N-Glycosylation System | 370 |
| 6.12.8.2 | Glycoengineering in <i>Escherichia coli</i> Using Bacterial O-Glycosylation Systems | 371 |
| 6.12.8.3 | The Future of Bacterial Glycoengineering | 374 |
| 6.12.9 | Concluding Remarks | 374 |
| References | | 376 |

6.12.1 Introduction

Almost 70 years ago the first glycoprotein, ovalbumin, was identified by Albert Neuberger.¹ Four decades later, in the middle of the 1970s, the first surface layer (S-layer) glycoproteins were discovered in an archaea, *Halobacterium salinarium*.^{2,3} This brought an end to the dogma that protein glycosylation is limited to eukaryotes. Studies have demonstrated that several Gram-positive (i.e., *Clostridium tyrobutyricum*,⁴ *Listeria monocytogenes*,⁵ and *Streptococcus gordonii*⁶) and Gram-negative bacteria (i.e., *Campylobacter jejuni*,⁷ *Neisseria gonorrhoeae*,⁸ *N. meningitidis*,⁹ *Pseudomonas aeruginosa*,^{10,11} *Porphyromonas gingivalis*,¹² *Bacteroides fragilis*,¹³ and enterotoxigenic *Escherichia coli*¹⁴) also have protein glycosylation systems.

Recent advances in analytical techniques, particularly mass spectrometry (MS) instrumentation, has increased the number of identified glycoproteins in bacteria. These developments have made it possible to detect glycoproteins at the subpicomole level.¹⁵ In addition, the determination of genome sequences of different bacteria in conjunction with bioinformatics tools have led to the identification of putative key enzymes (e.g., glycosyltransferases (GTs) and oligosaccharyltransferases (OTases)) involved in protein glycosylation pathways. In many cases, this process in bacteria is controlled by a group of genes located in the protein/pilin glycosylation locus (*pgl*). This locus usually is distinct from other loci involved in the biosynthesis of polysaccharides, such as lipopolysaccharides (LPS) and capsular polysaccharides, although cross talk between pathways has been reported.¹⁶ Several protein glycosylation pathways in important pathogenic bacteria such as *C. jejuni*, *Mycobacterium tuberculosis*, *P. aeruginosa*, *N. meningitidis*, and *N. gonorrhoeae* have been partially or completely elucidated.

At this stage, the exact function of protein glycosylation in bacteria is not well understood; however, addition of carbohydrates seems to be a common modification of secreted and cell-surface-associated proteins in bacteria. In addition, many bacterial appendages such as flagella and pili, are composed of glycoproteins especially among pathogenic bacteria. Flagella and pili are both important virulence factors and the existence of glycoproteins with these appendages points toward an involvement of glycoproteins in pathogenicity.

At the current time, a wealth of knowledge for protein glycosylation has been obtained from numerous studies conducted in eukaryotes rather than in bacteria. However, the simplicity of the bacterial systems has made them a great model for studying different basic aspects entailed in protein glycosylation. Moreover, recent studies identified some unique protein glycosylation mechanisms in bacteria such as *Neisseria* and *Pseudomonas* that have not been described for any other eukaryotes before.

In this chapter we relate recent findings of protein glycosylation in different Gram-negative and Gram-positive bacteria with the emphasis on pathways and enzymes involved in these systems. Also, the potential application of bacterial protein glycosylation systems for production of novel functional glycoproteins, termed bacterial glycoengineering, are discussed.

6.12.2 Protein N-Glycosylation and O-Glycosylation at a Glance

Protein glycosylation is categorized into different classes based on the glycosidic linkages between the amino acid and sugar residues.¹⁷ Although other linkages are known to occur,¹⁷ N-linked (sugars attached to Asn residues), and O-linked (sugars attached to Ser or Thr residues), are the most frequent and best-studied protein glycosylation processes.

N-Glycosylation is the most frequent type among eukaryotes, and it is predicted that more than 50% of proteins undergo this type of modification.¹⁸ Sugar donors, protein acceptor motifs, and enzymes involved are highly conserved in eukaryotic N-glycosylation. The core oligosaccharides are assembled onto a lipid carrier, dolichol-pyrophosphate (Dol-PP), in a process that begins by the sequential transfer of monosaccharides onto Dol-phosphate at the cytoplasmic side of the endoplasmic reticulum (ER) membrane. These oligosaccharides are further processed in the lumen of the ER, and subsequently transferred *en bloc* to the acceptor proteins by the OTase complex.^{19,20} In this process an N-glycosidic linkage is formed between the amide group of an asparagine (Asn) residue located in the 'glycosylation sequon', Asn-X-Ser/Thr (which X can be any amino acid except proline),^{21,22} and with a GlcNAc at the reducing end (Figure 1). The protein-linked oligosaccharides are further modified in the Golgi. One of the main roles of the N-linked protein glycosylation system is to assist proteins in acquiring the active conformation.^{19,20} In this process, the oligosaccharide indicates the state of folding and provides a signal for interaction of proteins with lectin chaperones, or directs misfolded protein for disposal in the ER.^{19,23} Unlike eukaryotes and archaea, N-glycosylation is not a common posttranslational modification in bacteria. To date, a general N-glycosylation pathway has only been identified and studied in detail for *C. jejuni* (see Section 6.12.3). However, putative protein glycosylation loci have been identified in other proteobacteria. For example, a locus containing some genes with significant similarity with the *pgl* genes of *C. jejuni* is present in the *Wolinella succinogenes* genome.²⁴ Recently, Santos-Silva *et al.*²⁵ resolved a crystal structure of a cytochrome from *Desulfovibrio gigas*, a Gram-negative sulfate-reducing bacterium. It was shown

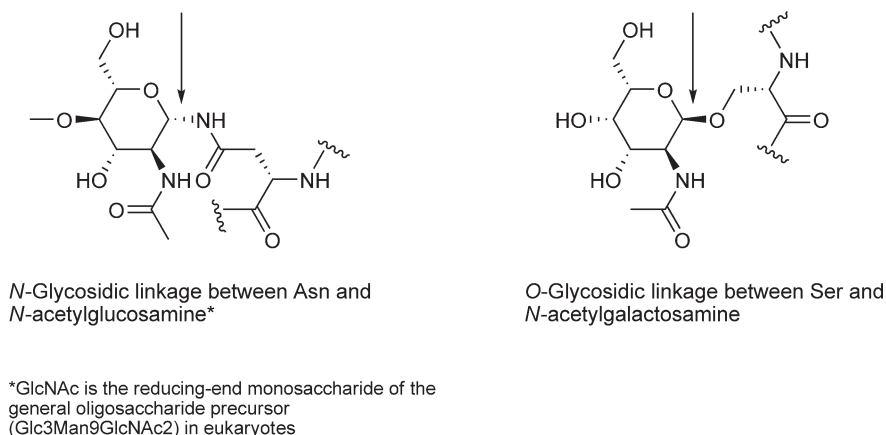


Figure 1 N- and O-glycosidic linkages.

that an Asn residue located within a typical N-glycosylation sequon is modified with a trisaccharide. In this bacterium, the genes presumably responsible for glycosylation are spread in the genome and not located in a single locus, as in all the *Campylobacter* species sequenced.²⁶

In contrast to N-glycosylation, O-glycosylation can occur in different cell compartments such as the ER, Golgi, and the cytoplasm of the eukaryotic cell.^{17,27,28} A large family of polypeptide monosaccharide-GTs catalyze the initial step of O-glycosylation by attaching a monosaccharide to the hydroxyl group of selected serine (Ser) and threonine (Thr) residues (Figure 1). Unlike in N-glycosylation, there is not a well-defined consensus sequence or motif identified for O-glycosylation, and the initiator GTs are fairly specific for both the protein acceptor and the sugar donor. The sugar donor is usually a nucleotide-activated monosaccharide, except for the ER O-mannosylation in which a mannose linked to a dolichyl-phosphate serves as a sugar precursor.²⁹ In many cases, the O-glycans are extended from a monosaccharide to an oligosaccharide, in a step-wise manner by different GTs. Interestingly, O-glycosylation is more prevalent than N-glycosylation among bacteria, and it is frequently found in both Gram-positive and Gram-negative bacteria. Recent studies have identified a completely novel mechanism of O-glycosylation present in *P. aeruginosa*,¹⁰ *N. gonorrhoea*,³⁰ and *N. meningitidis*,^{31,32} which seems to be exclusive to bacteria and is described in more detail in Sections 6.12.4.2.2 and 6.12.4.3.

Regardless of the type of glycosylation, adding sugars to proteins is important for several essential biological functions in eukaryotes, that occur both inside and outside of the cells; for example, protein quality control in the ER,^{19,20,23} protein trafficking,³³ enhancing protein stability by increasing the resistance to proteolysis,²⁰ cell-cell interactions, and communications,³⁴ specific receptors for viruses, bacteria, and parasites,³⁵ and signals for the immune system to differentiate normal cells from abnormal cells (e.g., cancer and/or foreign cells), which usually contain abnormal glycans.^{36,37} Some of these functions have also been suggested for microbial glycoproteins;²⁶ however, there is no concrete data available to assign any general functions similar to those described for eukaryotes.

6.12.3 General Protein N-Glycosylation System in *Campylobacter jejuni*

Campylobacter jejuni is a food-borne pathogen and is the leading cause of bacterial gastroenteritis worldwide. In 1998, Fry *et al.*³⁸ cloned and expressed in *E. coli* a *C. jejuni* locus that contained several genes with homology to enzymes involved in LPS biosynthesis in other bacteria. This resulted in the addition of extra glycan chains to the host LPS. On the basis of these observations, the authors proposed that these genes were involved in LPS biosynthesis in *C. jejuni*.³⁸ However, Szymanski *et al.*³⁹ showed that this locus was actually implicated in the glycosylation of multiple proteins, and named the locus *pgl* for protein glycosylation. Figure 2 illustrates the genetic organization of the *pgl* locus in *C. jejuni*. *C. jejuni pgl* mutants impaired in protein glycosylation showed reduced adhesion and invasion to eukaryotic cells in culture and poor colonization of intestinal tracts of mice and chickens. However, as the mutation affects the glycosylation of multiple proteins, the actual role of the glycans has not been established.³⁹ It also has been shown that a functional glycosylation machinery is required for the full competence of the type IV secretion system.⁴⁰

Wacker *et al.*⁴¹ showed that coexpression of the *pgl* locus with a *C. jejuni* protein, AcrA, in *E. coli* resulted in N-glycosylation of this protein. The functional transfer of the *C. jejuni* glycosylation system into *E. coli* accelerated the study of the proteins involved in this process through the exploitation of the genetic tools available in this bacterium. Remarkably, *pglB*, one of the genes contained in the *pgl* locus, encodes a homologue to Stt3, the catalytic component of the eukaryotic OTase complex. Inactivation of PglB through the

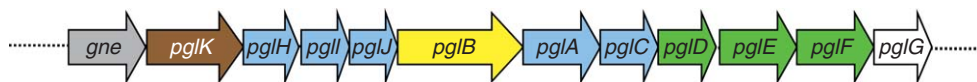


Figure 2 Genetic organization of protein N-glycosylation locus (*pgl*) of *C. jejuni*. The color coded arrows represent the genes encoding enzymes involved in N-glycosylation pathway. Blue, genes encoding glycosyltransferases that are involved in assembly of the oligosaccharide; green, genes encoding enzymes involved in biosynthesis of DATDH; brown, *pglK* encoding an ABC transporter; yellow, *pglB* that encodes the OTase; grey, *gne* is a bifunctional epimerase that is able of converting UDP-Glc to UDP-Gal, and UDP-GlcNAc to UDP-GalNAc.

mutagenesis of a conserved domain between PglB and Stt3 abolished glycosylation.⁴¹ It was later shown that in the presence of a suitable lipid-linked donor, PglB is sufficient for glycosylation.⁴² The remaining genes in the *pgl* cluster code for sugar modifying enzymes, GTs, and a putative flippase. Young *et al.*⁷ elucidated the structure of the *C. jejuni* N-glycan, and showed that the N-glycan is invariably a heptasaccharide composed of 2,4-diacetamido-2,4,6-trideoxyhexose (DATDH), GalNAc, and Glc. The glycosylation process starts at the cytoplasmic side of the inner membrane with the activity of PglC. PglC is a member of the polyprenol phosphate: N-acetylhexosamine-1-phosphate transferase family. This family of enzymes is involved in the assembly of diverse bacterial components, catalyzing reactions implicating a membrane-associated polyprenol phosphate acceptor substrate, and a cytoplasmically located (uridine diphosphate) UDP-D-amino sugar donor.⁴³ PglC transfers a DATDH residue to the undecaprenyl-phosphate (Und-P) carrier, to generate undecaprenyl pyrophosphate (Und-PP)-DATDH. When the *pgl* cluster is expressed in *E. coli*, GlcNAc replaces DATDH through the activity of the endogenous WecA GT, without affecting glycosylation.⁴⁴ Interestingly, DATDH is also the first sugar in some *Neisseria* pilin O-glycans (see Section 6.12.4.3). In fact, three genes, *pglDEF*, involved in the synthesis of UDP-DATDH from UDP-GlcNAc, have orthologues in *Neisseria*. The biosynthesis of this sugar proceeds from UDP-GlcNAc through the formation of a 4-keto-6-deoxy compound by PglF, which is converted into QuiNAc4N by PglE. This amino sugar is subsequently acetylated by the acetyltransferase PglD along with acetyl-CoA.⁴⁵ Different GTs complete the heptasaccharide. PglA and PglJ add the second and third GalNAc residue and remarkably, a single GT PglH is responsible for the addition of the three terminal GalNAc residues.^{44,46} PglI then transfers the Glc branch residue. Once the heptasaccharide is complete, the PglK flippase translocates the lipid-linked oligosaccharide into the periplasm,⁴⁷ where PglB transfers the glycan *en bloc* to the target proteins. **Figure 3** describes the N-glycosylation pathway in *C. jejuni*.

Campylobacter jejuni PglB, as its eukaryotic counterpart Stt3, recognizes the sequon N-X-S/T. However, PglB also requires a negatively charged amino acid at position minus 2 from the N, and therefore the sequon recognized by the bacterial enzyme can be represented as D/E-X-N-Z-S/T, where X and Z can be any amino acid except proline.^{48,49} The discovery of the bacterial-extended sequon opened the possibility to glycosylate other proteins. For example, recombinant cholera toxin B was engineered to contain the extended N-glycosylation acceptor sequence and was glycosylated *in vivo* in *E. coli*.⁴⁸ Nevertheless, this sequon is necessary but not sufficient for glycosylation. This feature can be a consequence of the mode of action of PglB. It has been shown that PglB can glycosylate fully folded proteins.⁵⁰ A peptide derived from AcrA was placed in a green fluorescent protein (GFP) loop known to accept insertions without affecting GFP folding. This hybrid protein was fluorescent indicating that the protein was folded. Incubation of this protein with an

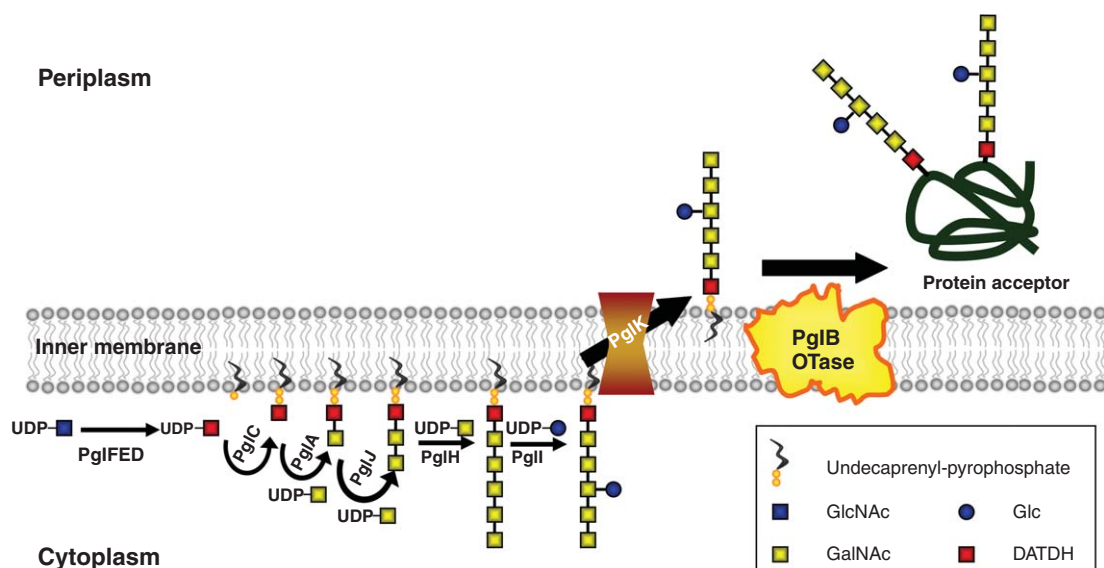


Figure 3 Proposed pathway for protein N-glycosylation in *Campylobacter jejuni*.

extract containing the lipid-linked oligosaccharide and purified PglB resulted in its glycosylation.⁵⁰ However, unfolding was necessary to glycosylate a eukaryotic protein, RNaseB. This protein was engineered to contain the Asp at position -2 from the original glycosylation site. Glycosylation of engineered RNase B was greatly increased upon unfolding. It has been proposed that in the *C. jejuni* system, N-glycosylation sites are located in flexible regions of folded proteins. On the contrary, the eukaryotic cotranslational glycosylation evolved to a mechanism presenting the substrate in a flexible form before folding.⁵⁰

Feldman *et al.*⁴² exploited the analogies between the protein glycosylation and LPS biosynthetic pathways to study PglB glycan specificity. It was found that PglB has relaxed glycan specificity, being able to accept a variety of lipid-linked glycans naturally employed in LPS biosynthesis as sugar donors. Importantly, although PglB transfers a heptasaccharide in the natural host, it was able to transfer fully polymerized O antigens from diverse organisms to AcrA in engineered *E. coli* cells. Therefore, PglB has relaxed glycan specificity regarding both composition and length of the glycan chain. Further work showed that indeed, PglB only requires an acetamido group at C2 position of the sugar at the reducing end for transfer.⁵¹ It was also suggested that glycans containing a sugar at the reducing end substituted at position 4 cannot be transferred by PglB.⁵² Furthermore, the ability of PglB to transfer glycans from diverse polyisoprenol lipid carriers *in vitro* has been tested. It was shown that the *cis*-double bond geometry and the unsaturation degree of the lipid substrate are important features, although PglB seems to be less stringent regarding the polyisoprene length of the carrier.⁵²

Interestingly, the PglK flippase also has relaxed glycan specificity. PglK is an ATP-binding-cassette (ABC)-type transporter homologous to MsbA, the translocase responsible for the flipping of lipid A core during LPS biosynthesis. Mutagenesis of PglK in *C. jejuni* partially suppresses glycosylation. It was speculated that other flippases could compensate for the absence of PglK. This hypothesis was based on the finding that flippases have relaxed glycan specificity enabling them to flip diverse lipid-linked glycans.⁵³ Alaimo *et al.*⁴⁷ showed that indeed, PglK and Wzx, which translocate the Und-PP-linked O antigens across the inner membrane, are interchangeable and can complement each other. This indicated that lipid-linked oligosaccharide translocation across membranes can be carried out by two distinct mechanisms.⁴⁷

Genes homologous to the ones in the *C. jejuni* *pgl* cluster are found in all the Campylobacteraceae family members, including *C. lari*, *C. coli*, and *C. upsaliensis*. The N-glycosylation genes have been also found in *Wolinella succinogenes* and *Desulfovibrio* spp.,²⁶ and they have been recently identified in a deep-sea vent ϵ -proteobacterial genome.⁵⁴ Interestingly, the N-glycosylation machinery is absent in *Helicobacter pylori*, an important human pathogen that is closely related to *C. jejuni*.

6.12.4 Protein O-Glycosylation in Gram-Negative Bacteria

Protein O-glycosylation occurs in several Gram-negative pathogenic bacteria such as *Campylobacter* spp., *P. aeruginosa*, *Neisseria* spp., *P. gingivalis*, and *E. coli*. These bacteria harbor O-glycoproteins that are important for their virulence. Interestingly, the majority of O-linked glycoproteins reported in these bacteria are either secreted or located at the surface of the cell. **Table 1** shows the summary of identified O-glycoproteins in different Gram-negative bacteria. However, at this stage there is little known about the exact mechanisms and pathways involved in protein O-glycosylation for most Gram-negative bacteria. The following sections are focused on describing well-studied pathways and mechanisms involved in O-linked glycosylation of *Campylobacter*, *Pseudomonas*, *Neisseria*, and pathogenic *E. coli*.

6.12.4.1 Flagellar O-Glycosylation in *Campylobacter*

As described in Section 6.12.3, the N-glycosylation system of *C. jejuni* is responsible for glycosylating several proteins in the periplasm. In addition to this system, *Campylobacter* O-glycosylates its flagellum. *Campylobacter* is unique among bacteria possessing both, N- and O-linked protein glycosylation systems.

Campylobacter has polar flagella, which are essential for motility. It has been shown that the motility imparted by the flagella is required for colonization of the gastrointestinal tract and invasion of intestinal epithelial cells *in vitro*.⁶⁴⁻⁶⁸ The *Campylobacter* flagellar filament is composed of two flagellin structural proteins FlaA

Table 1 Summary of protein O-glycosylation systems in Gram-negative bacteria

| Bacteria | Protein acceptor | Location of glycosylation | Number of glycosylation sites | Reference(s) |
|------------------------------------|-------------------------|---------------------------|-------------------------------|--------------|
| <i>Campylobacter jejuni</i> 81–176 | FlaA flagellin | Cytoplasm | 19 | 55, 56 |
| <i>Campylobacter coli</i> VC167 | FlaA flagellin | Cytoplasm | 16 | 55, 56 |
| <i>Neisseria meningitidis</i> | PilE pilin | Periplasm | 1 | 31, 57, 58 |
| <i>N. gonorrhoeae</i> | PilE pilin | Periplasm | 1 | 8, 30 |
| <i>Pseudomonas aeruginosa</i> PAK | FliC flagellin (a-type) | Cytoplasm | 2 | 59 |
| <i>P. aeruginosa</i> PAO1 | FliC flagellin (b-type) | Cytoplasm | 2 | 60 |
| <i>P. aeruginosa</i> 1244 | PilA pilin | Periplasm | 1 | 16, 31 |
| <i>P. aeruginosa</i> Pa5196 | PilA pilin | ND | >1 | 11 |
| <i>Escherichia coli</i> 287 | AIDA-I | Cytoplasm | 19 | 61 |
| <i>E. coli</i> H10407 | TibA | Cytoplasm | ND | 14 |
| <i>E. coli</i> | Ag43 α | Cytoplasm | 16 | 62, 63 |
| <i>Porphyromonas gingivalis</i> | Gingipains | ND | ND | 12 |

ND, not determined.

(predominant form) and FlaB.⁶⁹ Interestingly, each flagellar filament is composed of approximately 20 000 protein subunits⁷⁰ carrying several glycosylation sites. This demonstrates how extensively a single filament can be modified with carbohydrates.

Logan *et al.*⁷¹ suggested that certain Ser residues of flagellin isolated from *C. coli* were posttranslationally modified. Later, Doig *et al.*⁷² showed that flagellin isolated from several strains of *C. jejuni* and *C. coli* are sensitive to periodate oxidation and bind to *Limax flavus* agglutinin, a lectin recognizing sialic acid, and proposed that flagellins are glycosylated. Thibault *et al.*⁵⁵ performed a comprehensive structural analysis of *Campylobacter* flagellin using MS and nuclear magnetic resonance (NMR) spectrometry. Analysis of flagellins purified from *C. jejuni* 81–176, by electrospray ionization (ESI)–MS showed that FlaA exhibits a mass approximately 10% higher than the predicted mass of FlaA, indicating that FlaA is posttranslationally modified.⁵⁵ The precise assignment of the type and location of these modifications were determined by ESI–MS and MS/MS analyses of peptides produced by trypsin digestion of purified flagellin. Nineteen Ser and Thr residues of flagellin of *C. jejuni* 81–176 were shown to be glycosylated with an unusual sugar and its derivatives. The combination of MS and NMR analyses demonstrated that FlaA is glycosylated predominantly with pseudaminic acid (Pse) 5,7-diamino-3,5,7,9-tetradeoxy-L-glycero- α -L-manno-nonulosonic acid (Pse5Ac7Ac) in both *C. coli* and *C. jejuni*. Furthermore, other related sugars such as Pse5Am7Ac (5-acetamidino-7-acetamido-Pse), Pse5Ac7Ac8OAc (5,7-diacetamido-8-O-acetyl-Pse), and Pse5Pr7Pr (5,7-N-(2,3-dihydroxypropionyl)-Pse) were identified as other minor components of glycosylated flagellins in *Campylobacter* (Table 2).⁵⁵ Interestingly, Pse has a structural similarity to sialic acid and is also found as a component of LPS in other bacteria such as *P. aeruginosa*.⁷³

Table 2 Sugars that are identified on glycosylated flagellin from *Campylobacter jejuni* 81–176 and *Campylobacter coli* VC167

| Monosaccharide | <i>C. jejuni</i> 81–176 | <i>C. coli</i> VC167 |
|--|-------------------------|----------------------|
| Pse | + | + |
| PseAm | + | – |
| PseOAc | + | – |
| PseAmOGlnNAc | + | – |
| LegAm | – | + |
| MeLegAm | – | + |
| Uncharacterized (with the neutral mass of 431) | – | + |
| Uncharacterized (with the neutral mass of 432) | – | + |

(+), detected; (–) not detected; Pse, pseudaminic acid; PseAm, acetamidino form of Pse; PseOAc, acetylated form of Pse; PseAmOGlnNAc, N-acetyl glutamic acid attached to Pse; LegAm, acetamidino derivative of legionaminic acid; MeLegAm, N-methylacetimidoyl derivative of legionaminic acid.

Several studies focused on understanding the functions of enzymes involved in the *Campylobacter* flagellar glycosylation. These studies successfully identified nine genes (*pseA-I*) encoding several key enzymes involved in the biosynthesis of cytidine monophosphate (CMP)-Pse, the nucleotide-activated form of Pse.^{71,55,74–80} Functional assignments of the genes involved in the conversion of UDP-GlcNAc into CMP-Pse have been completed using traditional biochemical techniques and a new ‘metabolomic’ approach, which is based on identifying the accumulation of intermediates in a biosynthetic pathway by blocking different steps through mutation of the corresponding gene.⁷⁷ The six-step pathway describing the conversion of UDP-GlcNAc into CMP-Pse is depicted in **Figure 4**. *Campylobacter jejuni* can further modify CMP-Pse and uses Pse derivatives as the substrates for protein glycosylation. Mutation of the genes involved in this pathway in *C. jejuni* 81–176 impairs motility of the organism and leads to the accumulation of unglycosylated flagellin in the cell.⁸¹

In most *Campylobacter* strains except *C. jejuni* 81–176, flagellin is modified with an additional sugar, LegAm, an acetamidino derivative of legionaminic acid (5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-D-glycero-D-galactononulosonic acid).⁷⁶ LegAm is a 9-carbon sugar that is a metabolic product of the second set of genes called posttranslational modification (*ptm*) genes.^{56,82} In *C. coli* VC167, both LegAm and its *N*-methylacetimidoyl derivative were found attached to flagellin (**Table 2**).⁷⁶ Mutation of *pse* or *ptm* genes of *C. coli* VC167 did not impair flagellin glycosylation, but rather altered the glycosylation pattern, which ultimately did not affect the flagellum assembly process in either of the mutated strains. Mutation of *pse* genes resulted in flagellin predominantly glycosylated with LegAm and its derivative, whereas in *ptm* mutants, flagellin was exclusively glycosylated with Pse. The flagellin was unmodified and flagella assembly was abolished when both sugar biosynthesis pathways were blocked.⁸¹ These results together demonstrated that

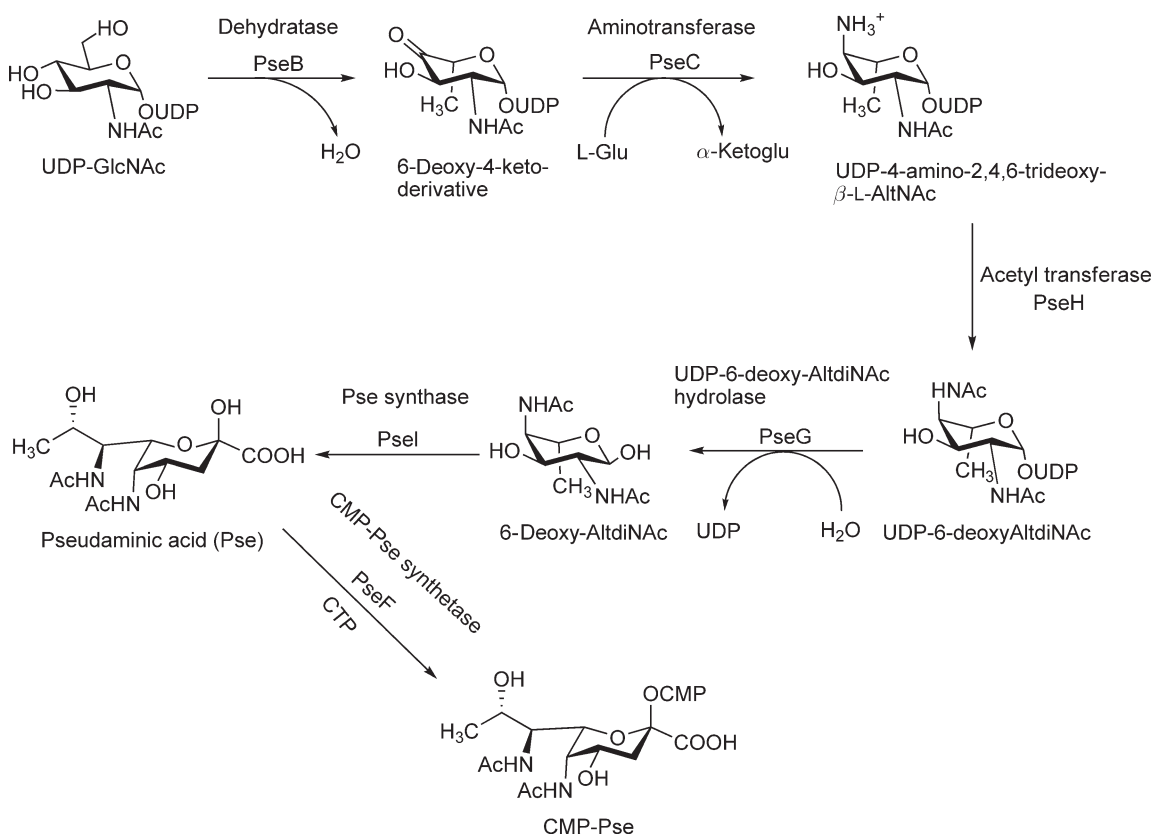


Figure 4 Proposed biosynthetic pathway of pseudaminic acid in *Campylobacter*. The predicted activity of PseF (last step) is based on the homology to the corresponding enzyme (NeuA) in the sialic acid biosynthetic pathway. Pse, pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-manno nonulosonic acid); Alt, altrose; 6-deoxy-AldiNAc, 2,4-diacetamido-2,4,6-trideoxy-L-Alt.

flagellar glycosylation in *Campylobacter* is an essential process for the formation of a normal flagellum and thereby the motility of this bacterium is dependent on protein O-glycosylation.⁸³ The same observation was reported for *H. pylori* when the sugar synthesis pathway providing the substrate of flagellar glycosylation was blocked.⁸⁴ Therefore, the enzymes involved in the synthesis of sugar donors for *Campylobacter* flagellar glycosylation system have become important targets for drug development.

6.12.4.2 O-Glycosylation Systems in *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a Gram-negative rod-shaped bacterium found in various environments. *Pseudomonas aeruginosa* is a typical example of an opportunistic pathogen in humans which almost never infects healthy individuals. Hence, it is considered an important pathogen for immunocompromised patients and those with cystic fibrosis (CF). *Pseudomonas aeruginosa* adherence to diverse surfaces is mediated by two appendages, flagella and pili. Interestingly, the main structural proteins in both of these appendages are identified to be O-glycosylated in several strains using two distinct mechanisms. The following sections describe the glycosylation systems implemented for modifying these appendages in *P. aeruginosa*.

6.12.4.2.1 Flagellin glycosylation

Motility is an important factor for pathogenicity of *P. aeruginosa*. In this bacterium motility is facilitated by either a single polar flagellum or multitrifurcated polar flagella, which are mainly composed of the flagellin structural protein named FliC. *Pseudomonas aeruginosa* flagellins are divided into two groups, a-type and b-type, based on their molecular mass and reactivity toward antisera.^{85,86} Early studies by Totten and Lory⁸⁷ showed that a-type flagellin from *P. aeruginosa* PAK has a lower electrophoretic mobility on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) than the expected molecular mass, suggesting flagellin is posttranslationally modified. However, such discrepancy in molecular mass was not observed for strains containing b-type flagellin when analyzed by SDS-PAGE. To further understand the nature of this posttranslational modification, Brimer and Montie⁸⁸ purified and analyzed flagellin from several strains bearing a-type flagellin. It was shown that a-type flagellin treated with periodate, a carbohydrate oxidizing reagent, reacted with biotin hydrazide and produced biotinylated flagellins. They also demonstrated that a-type flagellin is sensitive to treatment with trifluoromethanesulfonic acid (TFMS, removes both N- and O-linked glycans), suggesting that flagellin is glycosylated. Intact mass analysis of purified a-type flagellin from PAK strain using nano ESI-MS revealed that the mass is 3–4% higher than the predicted mass of FliC. ESI-MS/MS analysis of tryptic peptides from PAK flagellin demonstrated that Thr189 and Ser260 are O-glycosylated and the glycan moiety is composed of heterogeneous oligosaccharides composed of up to 11 monosaccharides (Figure 5). It was determined that the first monosaccharide attached to flagellin at both glycosylation sites is rhamnose (Rha). Gas chromatography (GC)-MS analysis of the carbohydrate moiety of PAK flagellin identified Rha, mannose, glucose (Glc), and 4-amino-4,6-dideoxyglucose (viosamine) as the main constituents.⁵⁹

Verma *et al.*⁶⁰ proved that the b-flagellin isolated from PAO1 is also glycosylated. Intact mass measurement of purified PAO1 flagellin with nano ESI-MS identified three peaks with higher masses than the theoretical mass of b-type flagellin, compatible with glycosylation. The type and location of glycosylation was assigned by MS/MS analysis of peptides produced by chymotrypsin or trypsin digestion of purified b-type flagellin, using a capillary liquid chromatography (LC)-nano ESI-MS.⁶⁰ This work revealed that two Ser residues at positions 191 and 195 are O-glycosylated with a complex glycan moiety with a mass of approximately 356 Da. Additional analysis has suggested that the glycan is composed of a single deoxyhexose sugar, which is further modified with an unknown molecule with a mass of 209 Da containing a phosphate group (Figure 5).⁶⁰ Further comprehensive analysis of this modified glycan is required to precisely define this complex structure.

In *P. aeruginosa* PAK, a genomic region of approximately 16 kb localized between *flgL* and *fliC* involved in protein glycosylation of a-type flagellin was identified. This region, which was named glycosylation island (GI), contains a cluster of 14 putative open reading frames (*orf*) named *orfA-N* (Figure 6).⁸⁹ Inactivation of *orfA* or *orfN* (annotated as *fgtA* for flagellar glycosyltransferase) impaired flagellin glycosylation.⁵⁹

The majority of the *orfs* located at the GI have similarities to genes that encode enzymes involved in modifying sugars of various polysaccharide biosynthetic pathways of Gram-negative bacteria. Only the functions of *orfA* (involved in nucleotide activation of 4-amino-4,6-dideoxyhexose (viosamine)) and *fgtA*

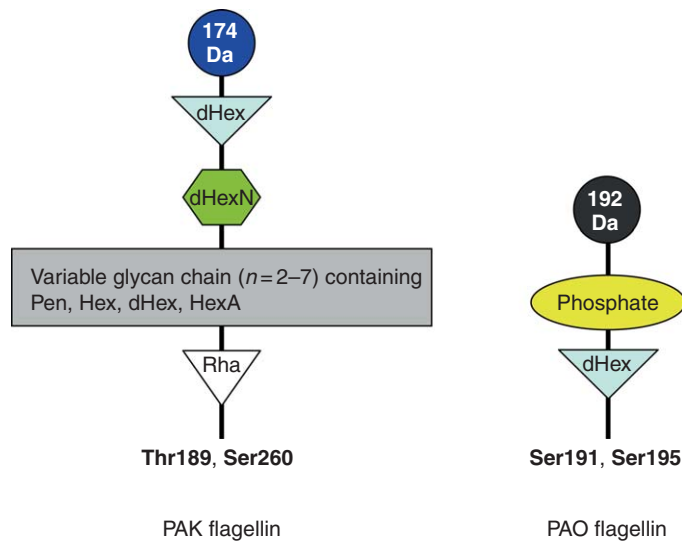


Figure 5 Glycan moieties of *Pseudomonas aeruginosa* a-type (left) and b-type (right) flagellins. Pen, pentose; Hex, hexose; dHex, deoxyhexose; HexA, hexuronic acid; dHexN, deoxyhexosamine; Rha, rhamnose. The structural assignments of the 174 and 192 Da modifications of the a-type and b-type flagellin glycan moieties are not known.

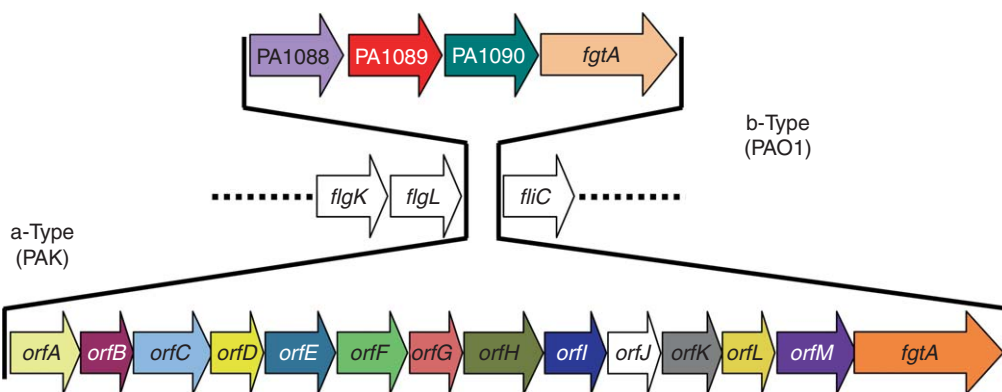


Figure 6 Genetic organization of glycosylation island (GI) genes of a-type and b-type *Pseudomonas aeruginosa* strains. The diagram illustrates the region of the chromosome, where the GI genes are located. The PAK strain (a-type) GI is constituted of 14 ORFs, whereas the PAO1 strain (b-type) consists of 4 ORFs.

((encoding a GT) belongs to family 2 which transfers Rha to flagellin) have been described.⁵⁹ The true functions of the remaining GI *orf*s are elusive.⁹⁰ In a-type flagellin glycosylation, the sugar donor for the first glycosylation step is thymidine diphosphate-rhamnose (dTDP-Rha). The locus encoding the enzymes for biosynthesis of dTDP-Rha is distinct from the GI.^{59,60} It has been shown that the rhamnose synthesizing locus (*rml*) of *P. aeruginosa* constituting of four genes, *rmlB*, *rmlD*, *rmlA*, and *rmlC*, is responsible for the production of dTDP-Rha. The Rha provided by this locus is also a precursor for synthesis of other carbohydrate components of *P. aeruginosa* LPS such as the oligosaccharide core and O antigens.⁹¹ Furthermore, it has been speculated that the other sugars involved in flagellin glycosylation are likely provided by other biosynthetic pathways.⁹⁰

It was shown that polymorphisms occur in GI of a-type strains resulting in the production of flagellin with heterogeneous sugars.^{90,92} It has been suggested that the heterogeneity of glycans on an exposed region of *P. aeruginosa* flagellin enables it to change the antigenicity,^{90,92} which can be important for escaping the host immune system.

In contrast to the GI of *P. aeruginosa* PAK (a-type flagellin strain), PAO1 (b-type strain) only has four genes in the GI (PA1088, PA1089, PA1090, and PA1091) (Figure 6). PA1091 is homologous, but not identical to *fgtA* of the a-type, and also belongs to GT-2 family. The PA1090 protein has similarity to UDP-GlcNAc pyrophosphorylase of *N. gonorrhoeae*, with a predicted nucleotidyltransferase activity, which is important for synthesizing nucleotide-activated sugar. It has been suggested that PA1089 encodes a protein similar to *Methanococcus jannaschii* phosphoserine phosphatase, which can be involved in the modification of the outermost sugar of b-flagellin glycan bearing a phosphate group. It was suggested that PA1088 encodes an enzyme, which either is involved in the biosynthesis of glycan or transfers the uncharacterized moiety to the O-linked deoxyhexose of flagellin (Figure 5). It seems that a short GI is characteristic of *P. aeruginosa* b-type flagellin strains.⁶⁰

Several studies have shown or predicted that the other species of *Pseudomonas* also harbor flagellar glycosylation system. For example, it has been demonstrated that the flagellum is glycosylated in the plant pathogen *P. syringae*.^{93–95} In other species, such as *P. fluorescense* and *P. entomophila*, the presence of the GI is only based on the homology of DNA sequences.⁹⁰ Unlike in *Campylobacter* and *H. pylori* where flagellin glycosylation is essential for flagella assembly, impairing flagellar glycosylation in *P. aeruginosa* did not affect assembly of a flagellum filament. However, *fgtA* mutants of PAK and PAO1 strains, which were defective in glycosylation, were significantly less virulent for the infection of a burn wound.⁹⁶ Similarly, a glycosylation defective *P. syringae* strain showed less infectivity in plants. Recently, Taguchi *et al.*⁹⁷ showed that *P. syringae* mutants impaired in flagella glycosylation, the swimming motility was drastically decreased in a medium with high viscosity. On the basis of these results it was suggested that flagellin glycosylation may stabilize the filament structure and likely assists the movement of the filament.⁹⁷

6.12.4.2.2 Type IV pilin glycosylation

Pseudomonas aeruginosa pili are hair-like filaments protruding from the bacterial surface and are important for attachment to different surfaces. Pili are composed of thousands of copies of a single protein, named pilin, which are assembled together to form the pilus. This bacterial appendage is particularly important for twitching motility, which is a form of solid surface translocation.⁹⁸ Pili are classified into several groups based on their morphology and their mechanism of assembly. *Pseudomonas aeruginosa* pilus is a member of the large type IV pili group. *Pseudomonas* type IV pilin (Tfp) is encoded by the *pilA* gene and is initially synthesized as a precursor (pre-pilin), which is processed by the removal of a six-residue leader sequence. This is followed by methylation of the nascent N-terminal phenylalanine to form a mature pilin prior to assembly into a pilus filament.⁹⁹ Burrows and coworkers¹⁰⁰ suggested a further classification for *P. aeruginosa* type IV pilin based on an amino acid sequence and the presence of unique accessory genes immediately downstream of *pilA*, Figure 7. In this classification type IV pilin can be categorized into five distinct phylogenetic subgroups (I–V). *Pseudomonas aeruginosa* PAO1 and PAK, the two common laboratory strains, are categorized into subgroup II, which has no

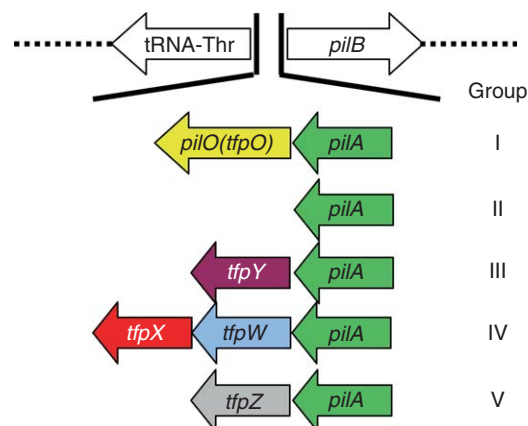


Figure 7 Proposed classification of type IV pilin alleles in *Pseudomonas aeruginosa* according to Burrows and coworkers.¹⁰⁰ Each pilin allele and associated accessory gene(s) are located in a common chromosomal locus, between the conserved *pilB* and *tRNA^{Thr}* genes.

accessory genes downstream of *pilA* and produce unmodified mature pilins.¹⁰⁰ Thus far pilin glycosylation has been reported only for strains of group I and IV.

Pseudomonas aeruginosa 1244 (group I) is the first identified strain that possesses a pilin glycosylation system. The pioneering work by Castric¹⁰ showed that in the pathogenic strain 1244 there is an ORF located downstream of *pilA*, named *pilO*, which together form an operon. Mutation in the *pilO* gene resulted in the production of pilin with lower apparent molecular mass compared to the wild type. It was also noted that purified pilin from a wild-type strain can be oxidized with periodate, which made it amenable to react with digoxigenin linked to hydrazide, whereas pilin from *pilO* defective strains remained unmodified in this process.¹⁰ On the basis of these results, it was suggested that pilin from strain 1244 is glycosylated and glycosylation is catalyzed by PilO.¹⁰ It should be noted that *P. aeruginosa* PAO1 genome sequence contains a gene called *pilO*, which is completely unrelated to *pilO* of 1244 and is not involved in pilin glycosylation. On the basis of structural prediction it was proposed that PilO contains multiple transmembrane domains and functions catalytically on the periplasmic side of the inner membrane using a lipid-linked oligosaccharide as the sugar donor.¹⁰ A recent study by Qutyan *et al.*¹⁰¹ supported the suggested preliminary model by Castric.¹⁰ However, the confirmation of the periplasmic location of PilO activity was demonstrated by Faridmoayer *et al.*,³¹ who showed that translocation of the Und-PP-linked sugar to the bacterial periplasm is required for glycosylation of pilin. Castric *et al.*¹⁶ performed a comprehensive study on the identification and characterization of the glycan moiety of pilin. They showed that glycosylated pilin from strain 1244 (O7 serotype) reacts with the monoclonal antibody (MAb) recognizing LPS from O7 strains. This antibody was unable to recognize unglycosylated pilin purified from *pilO*⁻ strain.¹⁶ ESI-MS analysis of the glycopeptides released by proteolytic digestion of pilin complemented with GC-MS and NMR analyses showed that a trisaccharide, α -5N β OHC₄7NFmPse-(2,4)- β -Xyl-(1,3)- β -FucNAc, where NFm is *N*-formyl, is β -linked to a Ser residue through an O-glycosidic linkage.¹⁶ Interestingly, this oligosaccharide is identical to the structure of a single subunit of O7 antigens, the oligosaccharide repeating unit of LPS. This suggested that the sugar donor in *P. aeruginosa* 1244 pilin glycosylation is the product of the LPS pathway.¹⁶ This hypothesis was demonstrated by DiGiandomenico *et al.*,¹⁰⁵ where two essential genes, *wbpM* and *wbpL*, of the O antigen pathway were mutated in the 1244 strain. WbpM is a bifunctional UDP-GlcNAc C6 dehydratase/C4 reductase that is required for the formation of UDP-*N*-acetyl-quinovosamine (QuiNAc, 2-acetamido-2,6-dideoxy-glucose), which is converted into UDP-FucNAc (*N*-acetylfucosamine), the first sugar donor in O7 antigen biosynthesis, by a subsequent epimerization step.^{102,103} WbpL is a GT that initiates the process of assembly of the O7 antigens by transferring FucNAc from UDP-FucNAc to the lipid carrier (Und-P).^{102,104} *Pseudomonas aeruginosa* 1244 mutants lacking *wbpM* and/or *wbpL* produced only unglycosylated pilin and O antigens were not produced.¹⁰⁵ Furthermore, the pilin glycosylation site, the C-terminal Ser148 residue, has been identified by site-directed mutagenesis.¹⁰⁶

These results collectively indicated that pilin glycosylation in *P. aeruginosa* 1244 is distinct from other protein O-glycosylation systems that normally use a nucleotide-activated monosaccharide as a sugar donor for initiating the reaction. Therefore, PilO is an OTase that uses the Und-PP-oligosaccharide as the sugar donor and transfers it *en bloc* to Ser148. Castric and coworkers¹⁰⁵ also determined glycan specificity of PilO by expression of different O antigen biosynthetic clusters in strain 1244. They demonstrated that PilO can transfer O antigens of different LPS biosynthetic pathways (i.e., *P. aeruginosa* O2, O5, O11, O13, O16, and *E. coli* O157) to pilin. The O antigen structures that are transferred by PilO vary in charge, size, configuration, linkage, and number of composing monosaccharides. However, it was noted that all of these O antigens, which are synthesized on Und-PP carrier, share some structural similarities at the reducing end monosaccharide, such as a C-1 linkage to the Und-PP carrier, a D configuration, a C-3 linkage to the adjacent sugar, and the presence of an acetamido group at position C-2.¹⁰⁷ It was noted that the variation between these O antigen subunits mainly comes from the remaining sugars. To demonstrate that the remaining sugars are not required for PilO activity, the *pilA-pilO* operon was expressed in *P. aeruginosa* PA103 (serotype O11) *wbjE* mutant, lacking the GT that assembles the second FucNAc to Und-PP-FucNAc during biosynthesis of the O11 antigens. The pilin produced in this strain was glycosylated with only one FucNAc residue. Therefore, it was postulated that the reducing sugar may contain important features to be recognized by PilO.¹⁰⁸

To further investigate the importance of glycan structure for PilO activity, Faridmoayer *et al.*³¹ functionally reconstituted the *P. aeruginosa* 1244 pilin glycosylation system in *E. coli* by coexpression of the *pilA-pilO* operon and different gene clusters for biosynthesis of various Und-PP-linked glycans. These authors showed that PilO is also able to transfer *E. coli* O7 antigens (containing GlcNAc at the reducing end) and the *C. jejuni* heptasaccharide employed in N-glycosylation, which is composed of DATDH at the reducing end (see **Table 3**). Furthermore,

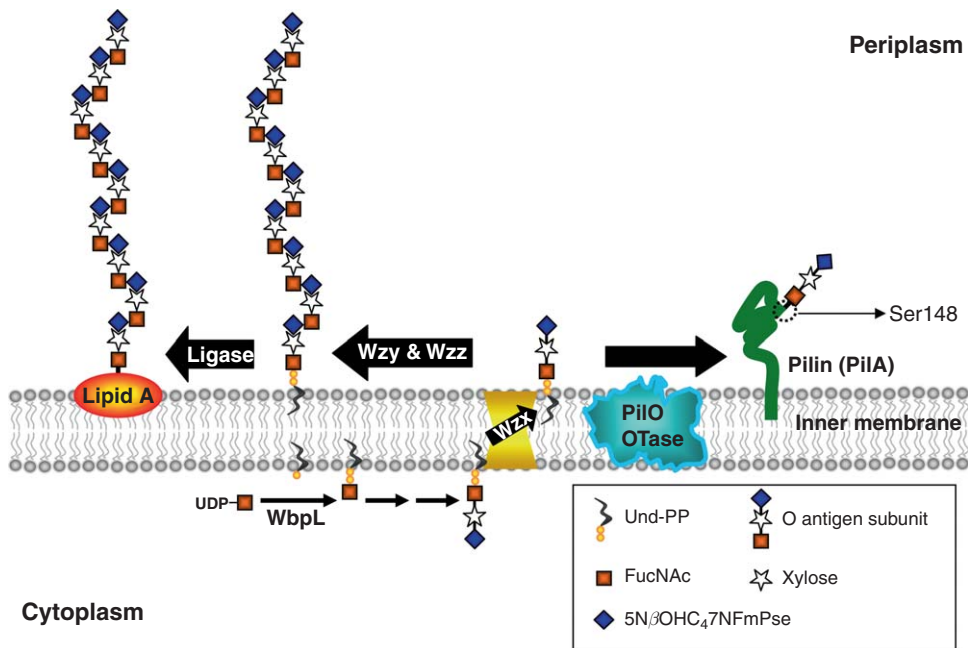


Figure 8 Proposed pathway for pilin O-glycosylation in *Pseudomonas aeruginosa* 1244. WbpL, UDP-FucNAc:Und-P FucNAc-1-P transferase; Wzx, flippase; Wzy, O antigen polymerase; Wzz, O antigen chain length regulator. 5N β OHC₄7NFmPse, 5-N-b-hydroxybutyryl-7-N-formyl-pseudaminic acid.

these results showed that PilO has relaxed glycan specificity, and that PilO activity is independent of any other *Pseudomonas* protein.³¹ Interestingly, the same study showed that PilO fails to transfer long oligosaccharide chains, even in the absence of the *P. aeruginosa* protein, suggesting that the inability to transfer polysaccharides is inherent to PilO.³¹ **Figure 8** shows the proposed pathway of pilin glycosylation in *P. aeruginosa* 1244.

Electron microscopy analysis of pili from *P. aeruginosa* strain 1244 using a glycan-specific immunogold label revealed that the glycan uniformly decorates the surface of a pilus filament.¹⁰⁹ Impairing pilin glycosylation did not affect pilus assembly and mutated strains could form normal pili and biofilms. However, strains lacking pilin glycosylation have a reduced twitching motility. A *pilO* knockout strain of *P. aeruginosa* 1244 was used to investigate the role of pilin glycosylation in pathogenicity. The competition index analysis using a mouse respiratory model, which is an approach for the evaluation of bacterial colonization in the lung, and is a sensitive indicator of pathogenicity, revealed that the survival rate of the *pilO*⁻ strain is significantly lower than the wild type.¹⁰⁹ Therefore, it was postulated that the presence of the pilin glycan contributes to increasing the survival rate in the lung environment. It was suggested that the pilin glycan is a significant virulence factor and may be involved in establishing infection in the host.¹⁰⁹

Recently, pilin glycosylation was described in another strain of *P. aeruginosa*, Pa5196 (serotype O11),¹¹ which belongs to subgroup IV (**Figure 7**).¹⁰⁰ A combination of genetic and biochemical techniques allowed the characterization of the pilin glycans of this strain. It was established that glycans are attached to one or more Thr residues. The exact position could not be determined, although positions 64 and/or 66 were suggested as potential sites. NMR characterization of the pilin glycan demonstrated that it consists of a repeating unit of α -(1,5)-linked D-arabinofuranose (D-Araf). D-Araf is uncommon among bacteria although it is the major constituent of cell wall arabinogalactan and lipoarabinomannan (LAM) polymers of mycobacteria, including *Mycobacterium tuberculosis* and *Mycobacterium leprae*. Interestingly, antibodies raised against *M. tuberculosis* LAM could identify the glycosylated pilins from Pa5196 and not 1244, indicating that the glycan is antigenically, as well as chemically, similar to those of *Mycobacterium*. The exact mechanism involved in the pilin glycosylation of Pa5196 strain remains elusive.¹¹ Further studies are required to precisely address the fundamental steps of this pathway such as the type of sugar donor (i.e., nucleotide-activated D-Araf or lipid-linked homo-oligomer of D-Araf), biosynthetic pathway of the sugar precursor and the enzyme(s) involved in the assembly of glycan to pilin.

6.12.4.3 O-Glycosylation of Type IV Pilin in *Neisseria*

Neisseria meningitidis (Nm) and *N. gonorrhoeae* (Ng) are the two most important pathogenic species of the *Neisseria* genus. *Neisseria meningitidis* (also known as meningococcus) is a causative agent of meningitis and is responsible for a considerable amount of mortality throughout the world. Gonorrhea is a sexually transmitted human disease caused by *N. gonorrhoeae* (also known as gonococcus). Nm and Ng both produce type IV pili. The process for the formation of *Neisseria* type IV pili is similar to *P. aeruginosa* (see Section 6.12.4.2.2). Expression of type IV pili mediates the colonization of human mucosal surfaces by pathogenic *Neisseria* and it is important for twitching motility, competence for natural transformation, and autoagglutination.^{110,111} Pilin, encoded by *pilE*, is the major constituent of type IV pili in *Neisseria*. A single cell of Nm and Ng is genetically capable of producing antigenically different PilE (pilin) variants. In a pilin monomer, some regions are conserved whereas the rest are highly variable. It was suggested that conserved domains have a role in pilus assembly and host cell recognition.^{57,112} Some pilin variants are not able to assemble into pili, and form soluble pilin (S-pilin) subunits instead. In this case, pilin usually lacks the first 39 amino acids of the N-terminal region, and strains producing these variants are usually less piliated.¹¹³

Early studies of pilin isolated from Ng suggested that pilin is posttranslationally modified with glycans composed of one or two galactose (Gal) residues.^{114,115} Nm pilin, which is closely related to Ng pilin, also undergoes posttranslational modification. Higher apparent mass, sensitivity to periodate oxidation, and TFMS treatment suggested that Nm pilin is also glycosylated.⁹ To understand the nature of the glycan and the glycosylation site, pilins from Nm strains were purified and digested with trypsin.⁵⁸ The peptides were fractionated by reverse-phase high-performance liquid chromatography (HPLC) and were subjected to different MS analyses (i.e., fast atom bombardment (FAB)–MS and ESI–MS), as well as gas-phase Edman sequencing.⁵⁸ The glycan moiety of pilin was also isolated by a mild alkali treatment through β -elimination, methylated, and subjected to GC–MS analysis. On the basis of comprehensive analysis of glycopeptides and glycan isolated from pilin, it was concluded that Nm pilin is O-glycosylated with a trisaccharide, Hex₂DATDH.⁵⁸ To further identify the glycan structure, the *galE* gene was mutated in Nm strain C311. GalE is a UDP Gal-4-epimerase, which is required for converting UDP-Glc into UDP-Gal, the sugar donor of lipooligosaccharide (LOS) biosynthesis in several Nm strains.⁵⁸ It was shown that the *galE* mutant of C311 strain formed a truncated LOS lacking Gal, and thus pilin in this mutant was only glycosylated with a single DATDH.⁵⁸ Further biochemical characterization of the glycan moiety from the wild-type and *galE* mutant determined the oligosaccharide structure as Gal- β -(1,4)-Gal- α -(1,3)-DATDH (Figure 9).⁵⁸

Tainer and coworkers⁸ resolved the structure of Ng pilin from strain MS11. The electron density of pilin at Ser63 suggested that this residue occupies a disaccharide composed of GlcNAc, which is attached to a Gal residue through an α -(1,3) linkage.⁸ This indicated that the glycan moiety of pilin varies by the number of sugars and differs at the reducing end monosaccharide; however, the second sugar (Gal) proximal to the reducing end is conserved (Figure 9). Further work of Marceau *et al.*⁵⁷ showed that substitution of Ser63 with Ala in pilin of Nm strain 8013 abolished pilin glycosylation, which suggested that Ser63 is the site of glycosylation and this site is conserved between Nm and Ng species. MALDI-TOF MS analysis of the tryptic peptide (Ser45–Lys73) from Nm strain 8013 pilin and the corresponding peptide from the mutated strain (PileSer63Ala) revealed that pilin is glycosylated with a disaccharide, HexHexNAc.⁵⁷ Surprisingly, this demonstrated that the glycan moiety of Nm strain 8013 is similar to Ng strain MS11, which is composed of a HexHexNAc and not Hex₂DATDH as also reported for Nm C311. Other studies identified a disaccharide composed of HexDATDH on pilins of Ng strain (Figure 9).¹¹⁶ This further indicates that the oligosaccharide structure of the glycan moiety of pilin can even vary between strains of the same species.

It has been shown that glycosylation is not important for pili assembly; although, it was noted that impairing pilin glycosylation had an adverse effect on the production of S-pilin in Nm, but not in Ng strains.^{57,117} The mechanism involved in this process is not clearly understood.

Several genes involved in pilin glycosylation have been identified. Genetic organization of *pilin glycosylation genes* is illustrated in Figure 10.¹¹⁸ Some genes involved in the glycosylation process are located in different loci in *Neisseria*. *pglA* was the first gene among *pgls* identified in Nm by Jennings *et al.*¹¹⁹ PglA has a homology to GT-1 GTs and is speculated to catalyze the transfer of the Gal, the second monosaccharide of pilin glycan, to the

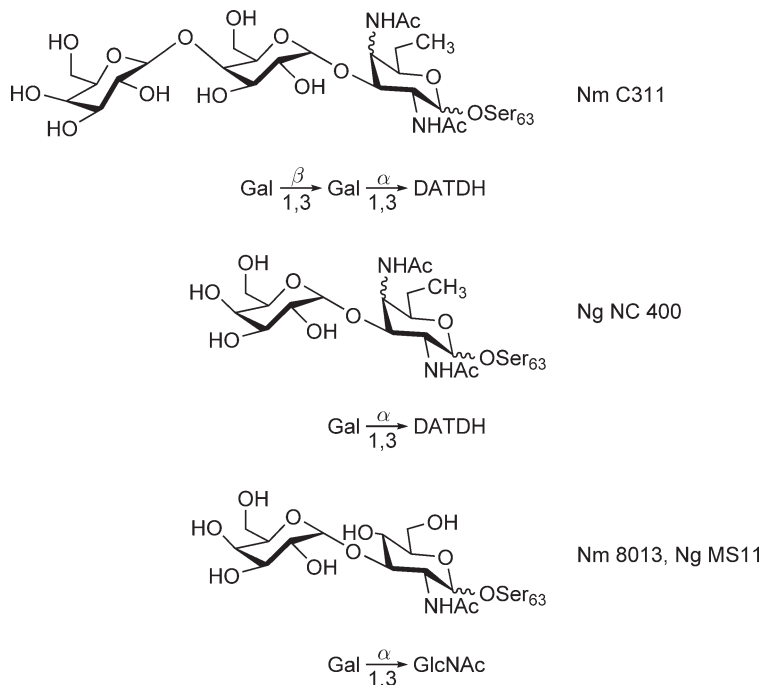


Figure 9 Glycan structures that have been identified on pilin from different *Neisseria* species and strains. Nm, *N. meningitidis*; Ng, *N. gonorrhoeae*; DATDH, diacetamido-trideoxy-hexose.

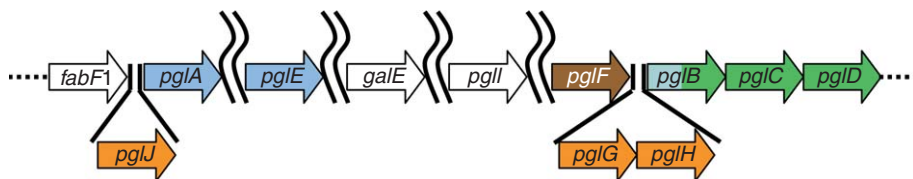


Figure 10 Genetic organization of the genes involved in *Neisseria* pilin glycosylation. Blue represents genes encoding GTs that are involved in the assembly of oligosaccharide; green indicates genes expressing enzymes involved in the biosynthesis of DATDH; brown represents *pglF* encoding a putative flippase; orange shows polymorphisms identified in *Neisseria* species (the exact function of these genes are elusive). Double line between genes indicates discontinuation.

Und-PP-DATDH through an α -(1,3) linkage (see [Figure 9](#)).¹¹⁹ Moreover, it has been suggested that PglA can also attach a Gal residue to Und-PP-GlcNAc through the same α -(1,3) linkage ([Figure 9](#)).¹¹⁹ Mutation of *pglA* resulted in the glycosylation of pilin with a similar truncated glycan as observed for the strains lacking *galE* (encoding Gal epimerase).^{119,120} Banerjee *et al.*¹²⁰ identified the Ng galactosyltransferase orthologue of PglA, which was named PgtA. Several other *pgl* genes located at the same locus were identified in different *Neisseria* strains and it was shown that this locus is polymorphic.^{121,122} In Nm strain C311, this locus contains four genes *pglF*, B, C, and D. It was speculated that *PglB*, C, D are involved in the biosynthesis of DATDH. Power and Jennings¹¹⁸ noted that these enzymes have similarity with genes in the *pgl* locus of *C. jejuni*, encoding enzymes also involved in the biosynthesis of DATDH (see Section 6.12.3). It was suggested that the *Neisseria* PglB is a bifunctional enzyme, which possesses both acetyltransferase and GT activity.¹¹⁸ Therefore, it was postulated by Aas *et al.*³⁰ that *Neisseria* PglB has a combined function of *C. jejuni* PglD (an acetyltransferase catalyzing the last step in the synthesis of UDP-DATDH)⁴⁵ and *C. jejuni* PglC (a GT that transfers the DATDH from its nucleotide-activated carrier to the Und-P).^{44,123} *Neisseria* PglC and PglD were predicted to have aminotransferase and dehydratase activity, respectively.¹¹⁸ The predicted function of the *Neisseria* PglC and PglD is in

accordance with the roles of their counterparts in *C. jejuni*, PglE (aminotransferase) and PglF (dehydratase).⁸⁰ A recent study by Aas *et al.*³⁰ showed that mutation in *pglC* and *pglD* of *Neisseria* resulted in the complete loss of the glycan, indicating the importance of these genes in the synthesis of DATDH. The *pglF* gene encodes a hypothetical protein, which was predicted to be a membrane-spanning protein,¹¹⁸ and exhibits some structural similarities to members of the Wzc/Wzx protein family.^{30,32,118} This is consistent with the role of *C. jejuni* PglK, the flippase of the N-glycosylation pathway (see Section 6.12.3).⁴⁷ Recent studies showed that the mutation of *pglF* significantly hampered pilin glycosylation in *Neisseria*.^{30,118}

Two polymorphisms have been identified in the *Neisseria* *pglBCD* locus.¹¹⁸ One of them contains an insertion of approximately 2 kb carrying two additional genes, *pglG* and *pglH* (Figure 10), encoding putative proteins with homology to GTs. These two genes carry homopolymeric tracts of G residues, which may mediate phase variation.¹¹⁸ The other polymorphism is an insertion–deletion, which occurs in *pglB*.^{118,121,122} A recent study by Chamot-Rooke *et al.*¹²⁴ revealed that an insertion in *pglB* resulted in changing PglB acetyltransferase activity. These strains produce a glycan that has an acetamido group of DATDH substituted with a glyceramido group, presumably either at C-2 or C-4 position, which forms a glyceramido trideoxyhexose (GATDH).¹²⁴

Two other genes, *pglI* and *pglE*, are involved in the biosynthesis of the trisaccharide, and are not part of the locus described above (Figure 10). PglI has homology to *P. aeruginosa* WbpC (acetyltransferase)¹²⁵ and that it is involved in O-acetylation of the Hex (Gal) residue of the pilin glycan.³⁰ PglE is a member of GT-2 glycosyltransferases that transfers the terminal Gal through a β -(1,4) linkage (Figure 9).¹²² *pglE*, which carries heptanucleotide repeats, is responsible for the phase variation between tri and disaccharide structures identified on pilin isolated from different *Neisseria* strains.¹²² In contrast to *C. jejuni*, the majority of *pgl* genes are phase variable, which explains the significant variation in glycan structures detected on *Neisseria* pilin.

Until recently it was thought that sugars were transferred to pilin in a sequential manner by the GTs described above.²⁶ In a recent study, Power *et al.*³² identified a Nm gene presenting a domain also found in the O antigen ligases, which was named *pglL*. Ligases are enzymes that catalyze the transfer of O antigens (a polymer of repeating oligosaccharides) from the Und-PP-linked carrier to the lipid A core at the last step of LPS biosynthesis. It was reported by these authors that the homology between O antigen ligases and PglL is limited to a short region, which was also identified in PilO, the *P. aeruginosa* 1244 OTase involved in pilin glycosylation (see Section 6.12.4.2.2). It should be pointed out here that the pathway for LOS biosynthesis in *Neisseria* diverges from the LPS pathways described for other Gram-negative bacteria such as *E. coli* or *Salmonella*. In *Neisseria*, LOS lacks the typical O antigen and the oligosaccharyl moiety is synthesized by the sequential transfer of sugars to the lipid A core. This pathway is also described for other bacteria such as *C. jejuni*.¹²⁶ Therefore, LOS biosynthesis does not require enzymes such as the O antigen polymerase (e.g., *E. coli* Wzy) and ligase (e.g., *E. coli* WaaL) that work with Und-PP-linked sugars. Mutation of the *pglL* gene in an Nm strain resulted in the formation of pilin with higher electrophoretic mobility than wild type when analyzed by SDS-PAGE, whereas LOS production remained unchanged in the *pglL* mutant strain.³² Aas *et al.*³⁰ identified the PglL orthologue in Ng, which was named PglO. The in-frame deletion of conserved amino acids in the homologous domain of *pglO* resulted in the formation of unglycosylated pilin, determined by intact mass measurement of pilin isolated from these mutants using ESI-MS.³⁰ These results together suggested an OTase activity for PglL/PglO. To clearly determine if PglL is an OTase, Faridmayer *et al.*³¹ functionally reconstituted pilin glycosylation in *E. coli*. This was achieved through the coexpression of different O antigen biosynthetic clusters, together with *pglL* and *pilE* in an engineered *E. coli* strain.³¹ The OTase activity of PglL was shown by the successful assembly of O antigens on pilin in the presence of PglL. Since PglL was functional in the absence of any other Nm proteins, it was concluded that PglL is sufficient for glycosylation when the proper protein acceptor (i.e., pilin) is present. The sugar specificity of PglL was tested using glycans with diverse structures such as *C. jejuni* N-glycan, *E. coli* O7, and *P. aeruginosa* O11 antigens (Table 3). The transfer of these sugars to pilin was confirmed by immunoblot and MS analyses. These glycans have completely different structures compared to the native Nm glycan, indicating that PglL has relaxed sugar specificity. This explains the diversity of glycans such as Hex₂GATDH, Hex₂DATDH, HexDATDH, and HexHexNAc attached to pilin in the native host. It was also demonstrated that translocation of the Und-PP glycans across the inner membrane is required for pilin glycosylation *in vivo*, indicating that the activity of PglL is located in the bacterial periplasm. This finding was further supported by the ability of PglL to transfer polysaccharides that are known to be polymerized in the periplasm, such as *E. coli* O7 and *P. aeruginosa* O11.³¹

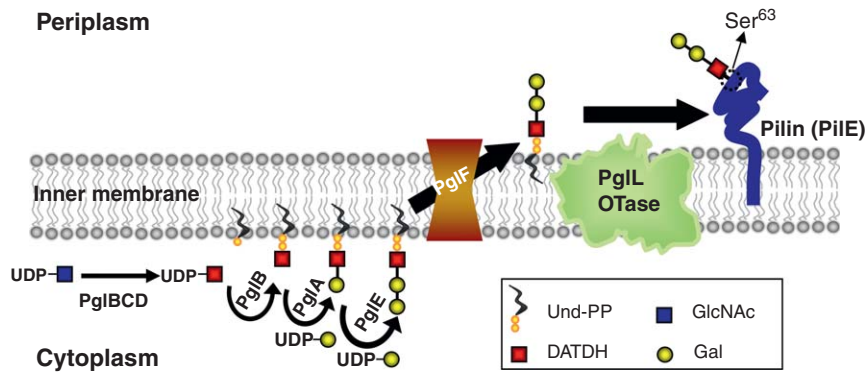


Figure 11 Proposed pathway for pilin O-glycosylation in *Neisseria*.

The results of these studies collectively suggested that during pilin O-glycosylation in *Neisseria*, the diverse monosaccharides are transferred one by one onto Und-P from nucleotide-activated sugars, forming the Und-PP-linked oligosaccharide in the cytoplasm. This process is followed by translocation of Und-PP-oligosaccharide into the periplasm where the OTase (i.e., PglL) transfers it *en bloc* to Ser⁶³ of pilin. This pathway is similar to pilin O-glycosylation in *P. aeruginosa* 1244 and protein N-glycosylation in *C. jejuni* (see Sections 6.12.3 and 6.12.4.2.2). **Figure 11** shows the proposed pilin glycosylation pathway in *Neisseria*. Furthermore, several other posttranslational modifications, in addition to glycosylation, have been identified for the *Neisseria* pilin. These studies showed that pilin could be modified at different Ser residues with phosphoglycerol,¹²⁷ phosphoethanolamine, and phosphocholine.¹¹⁶ However, exact mechanisms involved in these types of pilin modifications are not well-understood.

6.12.4.4 Glycosylation of Autotransporters in *Escherichia coli*

The autotransporters belong to a large family of secreted proteins in Gram-negative bacteria which encompass three structural motifs: a signal sequence, a passenger domain, and a translocator domain.¹²⁸ Autotransporter proteins are considered to be important virulence factors. In *E. coli*, a subgroup of autotransporter proteins consists of the AIDA (adhesion involved in diffuse adherence) adhesin from diarrhea-causing *E. coli*, the TibA (enterotoxigenic invasion locus B) adhesin/invasin associated with some enterotoxigenic *E. coli*, and the Ag43 (antigen 43) autoaggregation factor found in the majority of *E. coli* strains.¹²⁸ Interestingly, all three proteins promote bacterial aggregation and are found to be glycosylated.

AIDA-I, the autotransporter adhesin protein of the *E. coli* strain 2787 (O126:H27), facilitates diffuse adherence to cells. Benz and Schmidt⁶¹ noticed that the apparent mass of AIDA-I is higher than the mass deduced from the sequence when analyzed by SDS-PAGE. This suggested that AIDA-I is posttranslationally modified. Further analysis using GC-MS showed that this protein is exclusively modified with a heptose sugar with a ratio of 19:1 (sugar:protein); however, the exact sites of glycosylation were not identified. They also identified an ORF located upstream of the *aidA* gene encoding the heptosyltransferase responsible for the transfer of heptoses to AIDA-I, which was named *aab* (autotransporter adhesion heptosyltransferase). It was determined that AAH heptosyltransferase uses adenosine diphosphate (ADP)-glycero-mannoheptose as a sugar donor and is sufficient for glycosylation of AIDA-I.⁶¹ Niewerth *et al.*¹²⁹ examined several *E. coli* strains of the AIDA-I family and found that the *aab* gene existed in all members. It was also shown that glycosylation is important for AIDA-I-mediated adherence. The reason for loss of adherence in mutants impaired of AIDA-I glycosylation was not clear; however, inability of eukaryotic receptors to recognize unglycosylated AIDA or conformational changes of the protein were suggested.⁶¹ Further work is required to address the precise role of AIDA glycosylation.

Enterotoxigenic *E. coli* strain H10407 (O78:H11) is capable of invading epithelial cell lines derived from the human ileocecum and colon *in vitro*. Two distinct chromosomally encoded invasion loci (*tia* and *tib*) were identified in *E. coli* H10407. Lindenthal and Elsinghorst¹⁴ showed that TibA can be oxidized with periodate

which made it amenable to be labeled by hydrazide-conjugated digoxigenin, suggesting that TibA is a glycoprotein¹⁴ and functions as an adhesin.¹³⁰ They also found that the *tibC* gene upstream of *tibA* encodes a hypothetical protein that has a high degree of similarity to the heptosyltransferase encoded by the *aab* gene.¹⁴ Interestingly, it has been shown that *tibC* can complement an *aab* mutant, which resulted in restoration of AIDA glycosylation.¹³¹

Sherlock *et al.*⁶² recently identified the third member of *E. coli* autotransporters, Ag43, a self-recognizing adhesin protein that is also posttranslationally modified. Ag43 is encoded by the *flu* gene and is produced as a precursor of 1039 amino acids that is further processed into two polypeptides: the C-terminal translocator and the N-terminal passenger domain (Ag43 α).⁶² It has been noted that Ag43 exhibits approximately 25% sequence identity to the AIDA-I and TibA glycoproteins.⁶² It was shown that AAH and TibC heptosyltransferases, which glycosylate AIDA-I and TibA respectively, can also glycosylate Ag43 α .⁶² Knudsen *et al.*⁶³ identified over 16 glycosylation sites, including several Ser and Thr residues. They suggested that Ag43 α is glycosylated by a heptosyltransferase at the cytoplasm before it is transported. Sherlock *et al.*⁶² showed that glycosylation of Ag43 is not required for the self-aggregation process, but enhances bacterial binding to human cell lines. In contrast, a recent report suggested that Ag43 is not involved in direct colonization by binding to mammalian cells *in vivo*, which may rule out the importance of Ag43 glycosylation for binding to the cell receptors.¹³² Therefore, it seems that a system to invariably attach heptoses to proteins has evolved to glycosylate different adhesins in multiple *E. coli* strains, but the exact biological function of these glycans needs to be resolved.

6.12.4.5 Glycosylation in *Porphyromonas gingivalis*

Porphyromonas gingivalis, a black-pigmented, Gram-negative anaerobe, populates the subgingival crevice of the mouth. It is known to undergo a transition from its commensal status in healthy individuals to a highly invasive intracellular pathogen in human patients suffering from periodontal disease.¹³³ *Porphyromonas gingivalis* is known to produce a unique class of cysteine proteinases, termed gingipains. They consist of Arg-gingipain (Rgp) and Lys-gingipain (Kgp) that exist in both cell-associated and secreted forms which play a central role in the virulence of this organism. These enzymes are produced as large pre-proteins and are subject to elaborate, yet not fully understood, secretion, glycosylation, activation, and maturation processes.¹³⁴

A MAb raised against one of the isoforms of the Rgps purified from *P. gingivalis*, named MAb1B5, did not react with the corresponding recombinant form of the protein chain expressed in *E. coli* but was immunoreactive with a *P. gingivalis* polysaccharides.¹² Chemical deglycosylation abolished immunoreactivity with MAb1B5 and caused a reduction in the size of the membrane-associated enzymes, suggesting that the gingipains are glycosylated. Further work showed that gingipains are O-glycosylated.¹² However, the carbohydrates bound to gingipains are not fully characterized yet. Monosaccharide analysis showed that the different forms of the gingipains can contain 14–30% carbohydrate by weight of protein.¹³⁵ The sugars are attached to multiple sites on the proteins, presumably at the C-terminal region. Interestingly, differences in the monosaccharide compositions of the oligosaccharides bound to the gingipains were found, indicating that these protease isoforms are modified at different extents with diverse sugars. The variable nature of these additions may have a significant effect on the structure, stability, and immune recognition of these protease glycoproteins.¹³⁵

Unglycosylated gingipains are not catalytically active. However, the molecular basis of the inactivity is not known. Evidence suggesting that other proteins are also glycosylated in *P. gingivalis* has been presented.¹³⁶ Recently, glycosylation of several proteins has been described in a closely related bacterium, *Bacteroides fragilis*, which is a commensal in humans.¹³⁷ Further work will be needed to understand the glycosylation process in both systems, which can lead to a better understanding of the potential role of these systems in pathogenesis and commensalism.

6.12.5 Protein O-Glycosylation in Gram-Positive Bacteria

Protein glycosylation is a widespread process in Gram-positive bacteria. Several secreted and cell-surface-associated proteins of Gram-positive bacteria are modified with carbohydrates. Glycoproteins are also identified as the major components of bacterial appendages such as flagella and are abundant among S-layers. Examples of surface-associated proteins that are glycosylated in Gram-positive bacteria are the *Streptococcus*

gordonii cell-surface platelet-binding protein GspB⁶, and the *S. parasanguinis* fimbriae-associated protein Fap1.^{138,139} O-glycosylated flagellin was identified in *Butyrivibrio fibrisolvens*,¹⁴⁰ *Clostridium tyrobutyricum*,⁴ *C. acetobutylicum*,¹⁴¹ and *Listeria monocytogenes*.^{5,78} Owing to the overall similarity in glycosylation pathways between Gram-negative and Gram-positive bacteria, we briefly discuss S-layer protein glycosylation, which differs from what has been described above.

Around 30 years ago, Sleytr and Thorne¹⁴² identified that S-layer proteins of some Gram-positive bacteria such as *Clostridium thermohydrosulfuricus* (renamed *Thermoanaerobacter thermohydrosulfuricus*) and *T. thermosaccharolyticum*, are glycosylated. Eventually, S-layer glycoproteins were identified in several other Gram-positive bacteria, especially in members of the Bacillaceae family.¹⁴³

S-layer glycoproteins are usually assembled into a two-dimensional crystalline format on the cell envelope. Therefore it allows bacterial cell surfaces to be covered with the carbohydrate. This is comparable to the polysaccharide coating of Gram-negative bacteria cells by LPS.^{143,144} Several comprehensive studies have been performed to characterize S-layer glycans. Glycans involved in S-layer glycosylation are polysaccharides made of homo- or hetero-oligosaccharide units with 20–50 repeats, which is in contrast to the other systems of protein glycosylation described above. The glycans are composed of a variety of monosaccharides such as hexoses, deoxyhexoses, amino sugars, and some other rare sugars; however, in all of bacterial S-layer glycoproteins, glycans were found to be attached to Ser, Thr, and often to Tyr residues of the carrier protein through O-glycosidic linkages.^{143,144}

Genome sequences of several Gram-positive bacteria producing S-layer glycoproteins, such as *Geobacillus stearothermophilus*, have recently become available and the identified glycosylation loci in these bacteria named *slg* for S-layer glycosylations. *Geobacillus stearothermophilus* has been a preferred model system for studying the Gram-positive S-layer protein glycosylation system.¹⁴⁵ The glycan moiety of the S-layer protein in this bacterium is a homopolymer of L-Rha (composed of 13–18 homo-oligosaccharide repeating units and a linker oligosaccharide that is attached to protein through a D-Gal). It was identified that the homopolymer is O-linked to Thr590, Thr620, and Ser794 of the S-layer protein SgsE.^{145,146} A recent study by Steiner *et al.*¹⁴⁷ elucidated the role of several genes of the *slg* locus in *G. stearothermophilus* NRS 2004/3a. The function of four GTs encoded by *slg* genes *wsaC–F*, was investigated in detail by combining biochemical and NMR analysis. Three of the enzymes (WsaC, WsaD, and WsaF) are rhamnosyltransferases, whereas WsaE is a multifunctional enzyme. The N-terminal domain of WsaE was shown to catalyze methylation of the terminal α -(1,3)-linked L-Rha residue. The remaining central and C-terminal part of protein possesses rhamnosyltransferase activity, which assembles L-Rha to the linker oligosaccharide of a homopolymer to form α -(1,2) and α -(1,3) linkages during glycan chain elongation. The methylation and the glycosylation reactions occur independently. In addition to the four described genes, *wsaC–F*, there are other genes in the *slg* locus. These genes are suggested to encode enzymes involved in the biosynthesis of dTDP- β -L-Rha, ABC-transporter components, an initiating GT (WsaP) which is homologous to WbaP, a Gram-negative galactoyltransferase that transfers the Gal residue to Und-P, and a putative OTase (WsaB).¹⁴⁷ It appears that O-glycosylation of S-layer protein takes place in a similar manner to O-glycosylation of *Neisseria*, *P. aeruginosa*, and N-glycosylation of *C. jejuni*.¹⁴⁷ For further information on S-layer protein glycosylation, we recommend the review articles by Schaffer and Messner¹⁴³ and Messner *et al.*¹⁴⁴

6.12.6 Protein O-Mannosylation in Mycobacteria

Another interesting type of protein glycosylation, which has only been found in Mycobacteria, is O-mannosylation. Protein O-mannosylation, like N-glycosylation, is a highly conserved process from lower to higher eukaryotes.¹⁴⁸ A unique group of GTs, named protein O-mannosyltransferases (PMT), catalyze protein O-mannosylation in the ER.^{149,150} Interestingly, a similar protein O-mannosylation system is described in *Mycobacterium tuberculosis*, which is a causative agent of tuberculosis. Mannose polymers, composed of linear α -1,2-oligomannosides and α -1,3-oligomannosides, are linked to Thr residues in two *M. tuberculosis* glycoproteins (Apa/Rv1860 and MPB83/Mb2898).^{151,152} The mycobacterial orthologue (Rv1002c) of the *Saccharomyces cerevisiae* PMT was recently identified by VanderVen *et al.*¹⁵³ and was shown to initiate protein

O-mannosylation. The *Mycobacterial* O-mannosyltransferase is predicted to be a 55 kDa membrane protein, which shares significant features with its yeast orthologue such as 22–24% identity at the amino acid sequence levels, two conserved Arg residues located at the transmembrane domains, and an Asp-Glu motif found in the first predicted extra-cytoplasmic domain. It was also suggested that the bacterial enzyme, like its eukaryotic orthologue, uses the lipid-linked substrate as a precursor for glycosylation. VanderVen *et al.*¹⁵³ also reported that protein O-mannosylation is linked to Sec-translocation of proteins in *Mycobacteria*. This resembles the eukaryotic O-mannosylation in which the Sec61 translocon complex interacts with OTase subunits required for N-glycosylation and PMTs involved in O-mannosylation.^{154,155}

6.12.7 Challenges in the Study of Bacterial Protein Glycosylation

In most studies, the early indication that a protein was posttranslationally modified came from the observation of a higher apparent mass when analyzed by SDS-PAGE. Further evidence was then obtained through experiments such as Western-blot analysis, using glycan-specific antibodies and lectins, glycan oxidation with periodic acid, and treatment with TFMS. Although these experiments could provide preliminary information regarding the protein glycosylation, there are several drawbacks associated with these types of experiments. For instance, the level of glycoproteins produced in some bacteria is too low to be detected by Western-blot techniques. In addition, the contamination of protein with endogenous glycoconjugates such as LPS or other carbohydrates can generate false-positive results when heterogeneous samples are used. For example, despite several reports indicating that *Borrelia burgdorferi* glycosylates proteins,^{156,157} it has been shown that this organism does not synthesize glycoproteins.¹⁵⁸

The complete biochemical characterization of the glycosylprotein requires both, the determination of the glycosylation site(s) and the glycan structure. This includes identification of monosaccharides, linkages between sugars, and the configuration of each sugar, which can be carried out by MS and NMR analyses. These analyses required purified glycoprotein in quantities that are limiting for studying protein glycosylation in slow-growing bacteria. Furthermore, glycan moieties of glycoproteins are in most cases composed of complex and unusual sugars, such as those composing the *Campylobacter* flagellin glycan (see Section 6.12.4.1). The analyses are further complicated when the glycan moiety carries additional modifications that are variable, such as methylation, phosphorylation, and acetylation, just to mention a few. In many cases for example, *P. aeruginosa* flagellin, some modifications on sugars are detected but remain uncharacterized (see Section 6.12.4.2). Advances in MS instrumentations, which are becoming prime tools for glycoprotein analysis, allow researchers to analyze glycopeptides with a lower quantity. However, it should be noted that MS and subsequent MS_n analyses cannot provide sufficient information on precise structure of sugars and are generally unable to clearly identify branching sugars.

The other challenges involved in studying the biosynthetic pathways of protein glycosylation are associated with the identification of the true function of the enzymes, especially GTs that are involved in these pathways. In this regard, bioinformatics tools have provided invaluable resources for prediction of the function of these key enzymes; however, it should be taken into account that gene annotations can be misleading and therefore the precise assignment of these enzymes are only possible through biochemical analyses. Presently, the true function of enzymes involved in a few glycosylation pathways such as *C. jejuni* N-glycosylation are experimentally demonstrated (see Section 6.12.3).

6.12.8 Glycoengineering in Bacteria

Glycoproteins constitute a considerable fraction of current therapeutic agents in the market.¹⁵⁹ Addition of sugars to proteins can improve the stability of proteins and in many cases enhances the pharmacokinetic properties of the administrated protein. For example, adding two more N-glycans to erythropoietin (rHuEPO) resulted in the production of a hyperglycosylated analogue, darbepoetin, with an elevated *in vivo* activity and increased serum half-life.¹⁶⁰ Other examples are the Mpl ligand^{161,162} and leptin¹⁶³ proteins, in which attaching carbohydrates to protein was associated with improved activity and stability.

Therefore glycoengineering, an approach for the synthesis of custom-designed glycoproteins *in vivo*, is an invaluable tool for the production of a new generation of therapeutic recombinant glycoproteins. In recent years, this tool has been successfully developed in yeast for the production of both N-^{164,165} and O-linked¹⁶⁶ recombinant glycoproteins, carrying glycans similar to human sugars. The *E. coli* system was thought to be unsuitable for the production of recombinant glycoproteins due to lack of general N- and O-linked protein glycosylation systems.¹⁶⁷ However, this has changed recently through the ability of engineering newly identified bacterial N- and O-linked protein glycosylation pathways in *E. coli*.^{31,41,42} These findings have opened a new era in glycoengineering using bacterial protein glycosylation systems. Unlike eukaryotes, protein glycosylation is not a vital process in bacteria and they can tolerate the addition and manipulation of novel glycan biosynthetic pathways. This concept has been proven by merging different clusters of genes encoding enzymes that are required for the production of diverse glycans (i.e., O antigens) and protein glycosylation pathways. These glycans are used as sugar donors for protein glycosylation in engineered *E. coli* cells. **Figure 12** illustrates the general approaches that have been proposed for engineering *E. coli* cells for the production of recombinant glycoproteins, which are the topics of Sections 6.12.8.1–6.12.8.3.

6.12.8.1 Glycoengineering in *Escherichia coli* Using the Bacterial N-Glycosylation System

As described in Section 6.12.3, the first steps in *C. jejuni* and eukaryotic N-glycosylation are homologous processes. *C. jejuni* PglB was identified as the key enzyme (OTase) for this process and is sufficient for N-glycosylation. Feldman *et al.*⁴² noted that there are some similarities between LPS biosynthesis and N-glycosylation, particularly in the assembly of Und-PP-glycan in the cytoplasm and its translocation across a membrane. Furthermore, they demonstrated that PglB has relaxed glycan specificity, which is an important feature for the production of custom-made N-linked glycoproteins.⁴² Thus, it was possible to produce recombinant glycoproteins N-glycosylated with well-defined O antigen polymers using a combination of

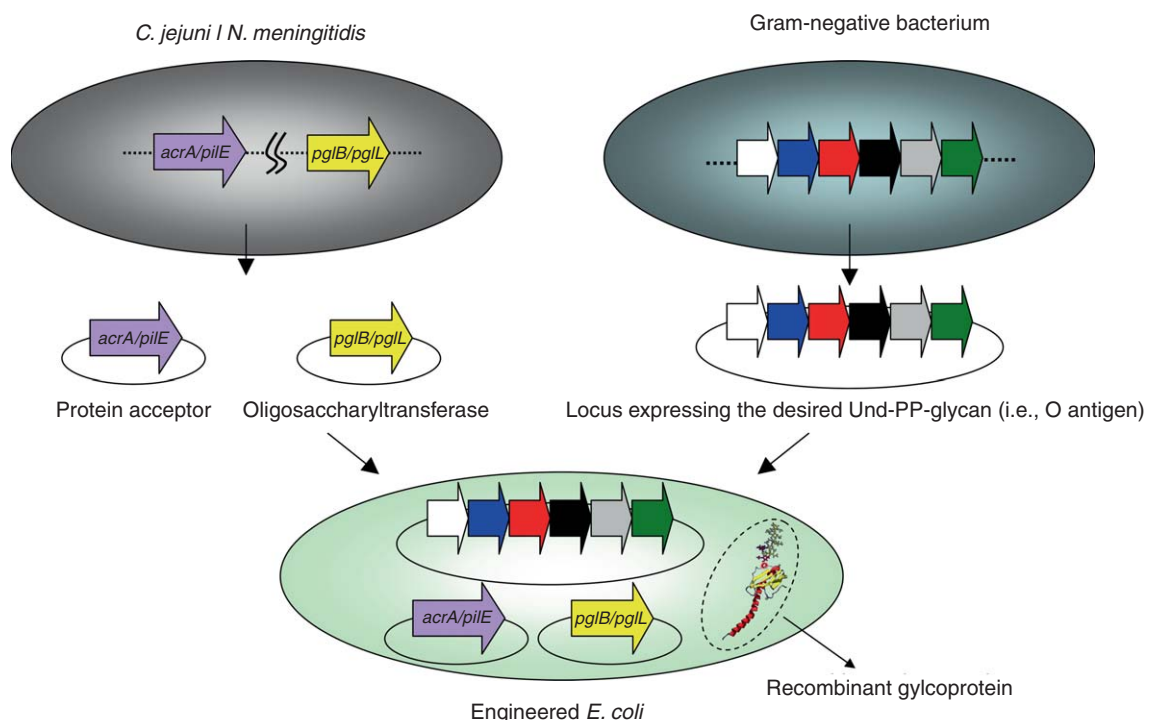


Figure 12 Glycoengineering in *Escherichia coli* using *Campylobacter jejuni* protein N-glycosylation and *N. meningitidis* protein O-glycosylation systems. *E. coli* cells are transformed with a gene encoding a protein acceptor (purple), an OTase (yellow) and a gene cluster encoding the enzymes required for the synthesis of the glycan of interest.

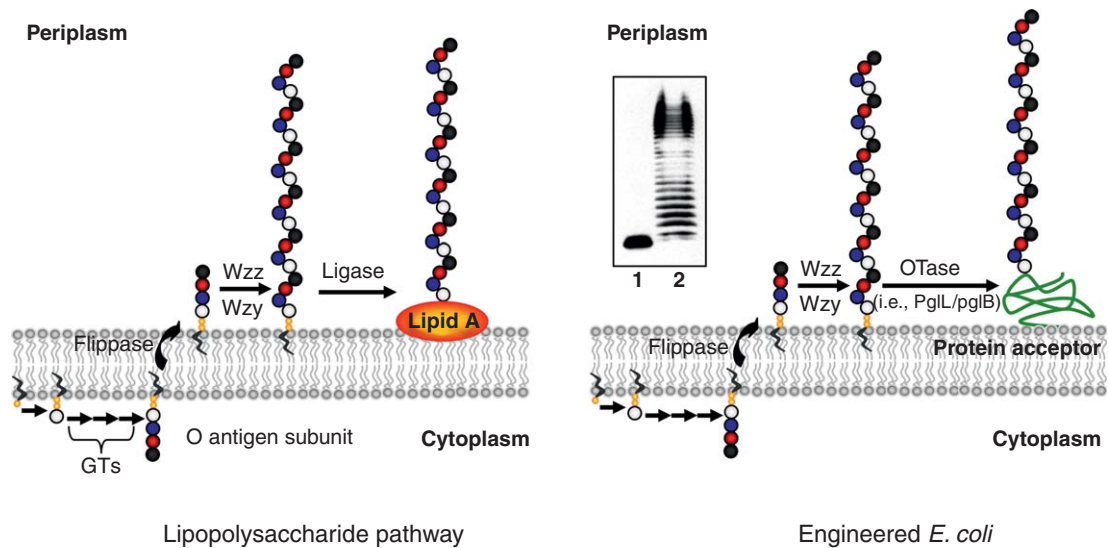


Figure 13 Combination of the LPS pathway and a protein glycosylation system in the engineered *Escherichia coli*. Left panel shows a polymerase(Wzy)-dependent LPS pathway and the right panel is an engineered *E. coli* expressing exogenous polysaccharides, a protein acceptor (e.g., pilin or AcrA), and an OTase. Inset in the right panel shows Western-blot analysis of unglycosylated protein acceptor (lane 1) and glycosylated protein with an O antigen polysaccharide (lane 2). The ladder-like pattern in lane 2 is the result of O antigens with different degrees of polymerization. GT, glycosyltransferase.

three components: LPS biosynthetic genes, the PglB OTase, and a protein acceptor in an engineered *E. coli* host (Figure 13). On the basis of this approach they demonstrated that different O antigens such as *P. aeruginosa* O11, *E. coli* O7, O9, O16, and O157 can act as the glycan donors for the glycosylation process (Table 3).^{42,51} Since PglB requires the acetamido group at position C-2 of the reducing end monosaccharide, it was not able to transfer O antigens of *Salmonella* and *E. coli* K30 capsular polysaccharides, both of which contain a hexose at the reducing end.⁵¹ Faridmoayer and Feldman also observed that PglB cannot transfer the O antigens of *E. coli* O2 (Table 3) carrying a β -(1,4) glycosidic linkage between the monosaccharide at the reducing end and the next sugar (unpublished data). Moreover, it was previously shown that it is plausible to engineer a glycosylation sequon (D/E-Z-N-X-S/T) into different proteins to make them serve as acceptors (see Section 6.12.3).⁴⁸ Collectively, these findings indicate that engineered *E. coli* cells capable of N-glycosylation are promising tools for the production of a wide range of custom-made recombinant glycoproteins, which could have a use as vaccines and therapeutics against various diseases.

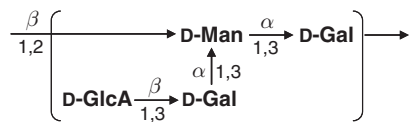
6.12.8.2 Glycoengineering in *Escherichia coli* Using Bacterial O-Glycosylation Systems

Studies on protein O-glycosylation systems in Gram-negative bacteria have revealed a distinct O-glycosylation mechanism employed for the modification of type IV pilin (as described in Sections 6.12.4.2.2 and 6.12.4.3). The Nm PglL and *P. aeruginosa* 1244 PilO were identified as OTases that use Und-PP-linked glycans as the sugar donor and they are both sufficient for glycosylation of the pilin monomer. We successfully reconstituted the Nm and *P. aeruginosa* 1244 pilin O-glycosylation systems in *E. coli*.³¹ It was shown that PglL and PilO exhibit relaxed glycan specificity and can transfer a variety of glycans to the pilin acceptor in the engineered *E. coli*; however, our observations suggested that PilO is stringent toward the glycan chain length and is able to transfer only short oligosaccharides (i.e., glycan composed of up to two repeating O antigen subunits).³¹ In contrast, the glycan chain length did not have an impact on PglL activity and it was shown that this OTase can use polysaccharides as well as oligosaccharides as precursors.³¹

We further demonstrated that PglL can transfer O antigens similar to those used for the PglB study (Table 3). It was also noticed that PglL can transfer *Salmonella* O antigens, *E. coli* K30 capsular polysaccharides, and *E. coli* O2 antigens (carrying a β -(1,4) linkage at the reducing end) to pilin (A. Faridmoayer and

Table 3 Summary of glycans that are transferred by PglL and PglB OTases in a glycoengineered *Escherichia coli*

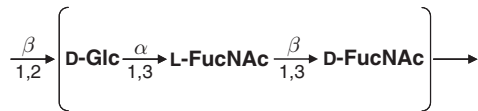
| Glycan structure | Organism derived from | PglB | PglL |
|--|-----------------------------|------|------|
| $\text{GalNAc} \xrightarrow[1,4]{\alpha} \text{GalNAc} \xrightarrow[1,4]{\alpha} \text{GalNAc} \xrightarrow[1,4]{\alpha} \text{GalNAc} \xrightarrow[1,4]{\alpha} \text{GalNAc} \xrightarrow[1,3]{\alpha} \text{DATDH} \longrightarrow$ $\begin{array}{c} \uparrow \\ \beta \\ \text{Glc} \end{array}$ | <i>Campylobacter jejuni</i> | + | + |
| $\begin{array}{c} \text{L-Fuc3NAc} \\ \downarrow \alpha 1,2 \\ \left[\text{L-Rha} \xrightarrow[1,2]{\alpha} \text{L-Rha} \xrightarrow[1,3]{\beta} \text{L-Rha} \xrightarrow[1,4]{\beta} \text{D-GlcNAc} \right] \longrightarrow \\ \uparrow \alpha 1,3 \\ \text{L-Rha} \end{array}$ | <i>E. coli</i> O2 | - | + |
| $\begin{array}{c} \text{L-Rha} \\ \downarrow \alpha 1,3 \\ \left[\text{D-VioNAc} \xrightarrow[1,2]{\beta} \text{D-Man} \xrightarrow[1,4]{\alpha} \text{D-Gal} \xrightarrow[1,3]{\beta} \text{D-GlcNAc} \right] \longrightarrow \\ \uparrow \alpha 1,3 \\ \text{D-Man} \end{array}$ | <i>E. coli</i> O7 | + | + |
| $\text{CH}_3 \xrightarrow[3]{\alpha} \text{D-Man} \xrightarrow[1,3]{\alpha} \left[\text{D-Man} \xrightarrow[1,2]{\alpha} \text{D-Man} \xrightarrow[1,2]{\alpha} \text{D-Man} \xrightarrow[1,3]{\alpha} \text{D-Man} \right]_n \xrightarrow[1,3]{\alpha} \text{D-Man} \xrightarrow[1,3]{\alpha} \text{D-GlcNAc} \longrightarrow$ | <i>E. coli</i> O9 | + | + |
| $\begin{array}{c} \text{OAc} \quad \text{D-Glc} \\ \uparrow \quad \uparrow \\ \left[\text{D-Galf} \xrightarrow[1,6]{\beta} \text{D-Glc} \xrightarrow[1,3]{\alpha} \text{L-Rha} \xrightarrow[1,3]{\alpha} \text{D-GlcNAc} \right] \longrightarrow \\ \uparrow \alpha 1,6 \\ \text{D-Glc} \end{array}$ | <i>E. coli</i> O16 | + | + |



E. coli K30 (K30 capsule)

–

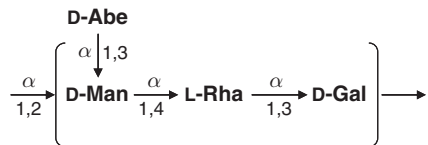
+



Pseudomonas aeruginosa O11

+

+



Salmonella typhimurium LT2

–

+

(+), transferred; (–), not transferred; Abe (abequose), 3-deoxy-D-fucose; DATDH, 2,4-diacetamido-2,4,6-trideoxyhexose; Fuc, fucose; FucNAc, N-acetylfucosamine (4-acetamido-4,6-dideoxygalactose); Gal, galactose; Galf, galactofuranose; GalNAc, N-acetylgalactosamine; Glc, glucose; GlcA, glucuronic acid; GlcNAc, N-acetylglucosamine; Man, mannose; Rha, rhamnose; VioNAc (N-acetylviosamine), 4-acetamido-4,6-dideoxyglucose.

M. F. Feldman, unpublished data). Therefore, it was concluded that PglL has extreme relaxed glycan specificity. However, the current drawback of using engineered *E. coli* for O-glycosylation is the protein acceptor, which is currently limited to pilin. Nonetheless, Horzempa *et al.*¹⁰⁸ partially identified the pilin structural requirement for the glycosylation process catalyzed by PilO and successfully engineered *P. aeruginosa* PA103 pilin, which is not a substrate for PilO, to be an acceptor for the glycosylation process. This limitation will be solved in the future, once the factors that determine the glycosylation site in pilin are completely understood.

6.12.8.3 The Future of Bacterial Glycoengineering

LPS and capsular polysaccharides are among the major bacterial surface components that are considered as important virulence factors. There have been several reports demonstrating that immunity directed toward these structures can confer protection against disease.^{168–170} Therefore, these polysaccharides have been a target for vaccine development. However, using polysaccharides alone is not an effective way to obtain a long-term immune response, and thereby conjugating these polysaccharides to appropriate protein carriers provides more effective protection against diseases.¹⁷¹ Recently, Castric and coworkers¹⁷² have shown that glycosylated pilin of *P. aeruginosa* 1244 can stimulate a protective response directed against O antigens. Therefore, glycoengineering in bacteria constitutes a robust and novel approach for the production of a wide range of glycoconjugate vaccines. The glycoproteins produced using these systems possess several advantages. For example, they have the defined sugar moieties that are reproducibly attached to genetically determined glycosylation sites, their production does not require the culture and large amount purification from pathogenic or slow-growing bacteria, and chemical cross-linking steps necessary for the attachment of polysaccharides to proteins are eliminated.

At this stage, production of humanized glycoproteins through bacterial glycoengineering techniques has not been explored in depth. There are some difficulties that may hamper this endeavor, such as the expression of properly folded human proteins in the bacterial periplasm, as well as the intricacy inherent to the biosynthesis of complex human oligosaccharides. This goal maybe achieved in the future when complete details of bacterial protein glycosylation pathways are revealed.

6.12.9 Concluding Remarks

Protein glycosylation is found as a frequent posttranslational modification in all domains of life. N- and O-linked glycosylation, the major forms of eukaryotic glycosylation, have been identified in several Gram-negative and Gram-positive bacteria. O-glycosylation seems to be more broadly distributed in bacteria than N-glycosylation, where the latter appears to be restricted to a few bacterial species. Several appendages such as flagella and pili are found to be decorated with carbohydrates, which are covalently attached through an O-glycosidic linkage to the protein. In addition, several cell surface and secreted bacterial proteins were found to be glycosylated. Two protein glycosylation systems with a remarkable similarity to eukaryotic systems have been identified in *C. jejuni* and *M. tuberculosis*. *C. jejuni*, possesses a general N-linked protein glycosylation system that modifies multiple periplasmic and membrane proteins, whereas *M. tuberculosis* have a still not well-characterized O-mannosylation system homologous to the one found in eukaryotes. In contrast, a unique protein O-glycosylation system controlled by OTases rather than GTs was identified in both Gram-negative and Gram-positive bacteria. Depending on the mechanism, glycosylation can take place in the periplasm or the cytoplasm of a bacterial cell. If the process is localized in the periplasm, the sugar donor is a lipid-linked glycan, and the glycosylation process is generally catalyzed by an OTase; cytoplasmic glycosylation is usually O-linked, the sugar precursors are nucleotide-activated monosaccharides, and the reaction is mediated by individual GTs.

Protein glycosylation is found in both, pathogenic and environmental bacteria. Although in some cases, glycosylation seems to be essential for pathogenesis, the actual role of the glycan moieties remains, in most of the cases, unknown. Extensive research is required to understand why bacteria have evolved to contain these diverse, sometimes complex, glycosylation mechanisms. Nevertheless, although their physiological roles are not yet clear, there has been significant progress toward the exploitation of these systems for biotechnological

purposes. Different protein glycosylation pathways could be functionally reconstituted into *E. coli*, and it has been shown that these cells can tolerate the manipulation of the glycosylation pathways to generate novel glycoconjugates containing diverse carbohydrates. Thus, glycoengineering in bacteria is an innovative and promising technology for the design and production of new glycan-based therapeutic agents that may have future applications in the treatment of bacterial infections, cancer, autoimmunity, and other diseases.

Abbreviations

| | |
|---|--|
| 5NβOHC₄7NFmPse | 5- <i>N</i> - β -hydroxybutyryl-7- <i>N</i> -formyl-pseudaminic acid |
| AAH | autotransporter adhesin heptosyltransferase |
| ABC | ATP-binding cassette |
| Abe | abequose (3-deoxy-D-fucose) |
| Ag43 | antigen 43 |
| AIDA | adhesion involved in diffuse adherence |
| Alt | altrose |
| Araf | arabinofuranose |
| Asn | asparagine |
| CF | cystic fibrosis |
| CMP | cytidine monophosphate |
| CoA | coenzyme A |
| Da | Dalton |
| DATDH | diacetamido-trideoxyhexose |
| Dol | dolichol |
| Dol-PP | dolichol-pyrophosphate |
| dTDP | thymidine diphosphate |
| ER | endoplasmic reticulum |
| ESI | electrospray ionization |
| FAB | fast atom bombardment |
| Fap | fimbriae-associated protein |
| Fuc | fucose |
| FucNAc | <i>N</i> -acetylfucosamine |
| Gal | galactose |
| Galf | galactofuranose |
| GATDH | glyceramido trideoxyhexose |
| GC | gas chromatography |
| GFP | green fluorescent protein |
| GI | glycosylation island |
| Glc | glucose |
| GlcA | glucuronic acid |
| GlcNAc | <i>N</i> -acetylglucosamine |
| Glu | glutamic acid |
| GT | glycosyltransferase |
| HPLC | high-performance liquid chromatography |
| kb | kilobase pair |
| Kgp | lys-gingipain |
| LAM | lipoarabinomannan |
| LC | liquid chromatography |
| Leg | legionaminic acid |
| LOS | lipooligosaccharide |
| LPS | lipopolysaccharide |
| MAb | monoclonal antibody |

| | |
|-----------------|---|
| MS | mass spectrometry |
| Ng | <i>Neisseria gonorrhoeae</i> |
| Nm | <i>Neisseria meningitidis</i> |
| NMR | nuclear magnetic resonance |
| orf | open reading frame |
| OTase | Oligosaccharyltransferase |
| pgl | protein/pilin glycosylation locus |
| PMT | protein O-mannosyltransferase |
| Pse | pseudaminic acid |
| ptm | posttranslational modification genes |
| QuiNAc4N | <i>N</i> -acetyl-quinovosamine |
| Rgp | Arg-gingipain |
| Rha | rhamnose |
| rml | rhamnose synthesizing locus |
| SDS-PAGE | sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| Ser | serine |
| S-layer | surface layer |
| slg | surface layer glycosylation |
| TFMS | trifluoromethanesulfonic acid |
| Thr | threonine |
| Tyr | tyrosine |
| UDP | uridine diphosphate |
| Und-P | undecaprenyl-phosphate |
| Und-PP | undecaprenyl-pyrophosphate |
| Vio | viosamine |
| VioNAc | <i>N</i> -acetylviosamine |
| Xyl | xylose |

References

1. A. Neuberger, *Biochem. J.* **1938**, *32*, 1435–1451.
2. M. F. Mescher; J.L. Strominger; S. W. Watson, *J. Bacteriol.* **1974**, *120*, 945–954.
3. M. F. Mescher; J. L. Strominger, *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2687–2691.
4. L. Bedouet; F. Arnold; G. Robreau; P. Batina; F. Talbot; A. Binet, *Microbios* **1998**, *94*, 183–192.
5. M. Schirm; M. Kalmokoff; A. Aubry; P. Thibault; M. Sandoz; S. M. Logan, *J. Bacteriol.* **2004**, *186*, 6721–6727.
6. B. A. Bensing; B. W. Gibson; P. M. Sullam, *J. Bacteriol.* **2004**, *186*, 638–645.
7. N. M. Young; J. R. Brisson; J. Kelly; D. C. Watson; L. Tessier; P. H. Lanthier; H. C. Jarrell; N. Cadotte; F. St Michael; E. Aberg; C. M. Szymanski, *J. Biol. Chem.* **2002**, *277*, 42530–42539.
8. H. E. Parge; K. T. Forest; M. J. Hickey; D. A. Christensen; E. D. Getzoff; J. A. Tainer, *Nature* **1995**, *378*, 32–38.
9. M. Virji; J. R. Saunders; G. Sims; K. Makepeace; D. Maskell; D. J. Ferguson, *Mol. Microbiol.* **1993**, *10*, 1013–1028.
10. P. Castric, *Microbiology* **1995**, *141* (Pt. 5), 1247–1254.
11. S. Voisin; J. V. Kus; S. Houliston; F. St-Michael; D. Watson; D. G. Cvitkovitch; J. Kelly; J. R. Brisson; L. L. Burrows, *J. Bacteriol.* **2007**, *189*, 151–159.
12. M. A. Curtis; A. Thickett; J. M. Slaney; M. Rangarajan; J. Aduse-Opoku; P. Shepherd; N. Paramonov; E. F. Hounsell, *Infect. Immun.* **1999**, *67*, 3816–3823.
13. M. J. Coyne; B. Reinap; M. M. Lee; L. E. Comstock, *Science* **2005**, *307*, 1778–1781.
14. C. Lindenthal; E. A. Elsinghorst, *Infect. Immun.* **1999**, *67*, 4084–4091.
15. P. G. Hitchen; A. Dell, *Microbiology* **2006**, *152*, 1575–1580.
16. P. Castric; F. J. Cassels; R. W. Carlson, *J. Biol. Chem.* **2001**, *276*, 26479–26485.
17. R. G. Spiro, *Glycobiology* **2002**, *12*, 43R–56R.
18. R. Apweiler; H. Hermjakob; N. Sharon, *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
19. A. Helenius; M. Aebi, *Science* **2001**, *291*, 2364–2379.
20. A. Helenius; M. Aebi, *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049.
21. E. Bause, *Biochem. J.* **1983**, *209*, 331–336.
22. C. Ronin; S. Bouchilloux; C. Granier; J. van Rietschoten, *FEBS Lett.* **1978**, *96*, 179–182.

23. L. Ellgaard; A. Helenius, *Curr. Opin. Cell Biol.* **2001**, *13*, 431–437.
24. C. Baar; M. Eppinger; G. Raddatz; J. Simon; C. Lanz; O. Klimmek; R. Nandakumar; R. Gross; A. Rosinus; H. Keller; P. Jagtap; B. Linke; F. Meyer; H. Lederer; S. C. Schuster, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11690–11695.
25. T. Santos-Silva; J. M. Dias; A. Dolla; M. C. Durand; L. L. Goncalves; J. Lampreia; I. Moura; M. J. Romao, *J. Mol. Biol.* **2007**, *370*, 659–673.
26. C. M. Szymanski; B. W. Wren, *Nat. Rev. Microbiol.* **2005**, *3*, 225–237.
27. M. Goto, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 1415–1427.
28. J. Peter-Katalinic, *Meth. Enzymol.* **2005**, *405*, 139–171.
29. W. Tanner; L. Lehle, *Biochim. Biophys. Acta* **1987**, *906*, 81–99.
30. F. E. Aas; A. Vik; J. Vedde; M. Koomey; W. Egge-Jacobsen, *Mol. Microbiol.* **2007**, *65*, 607–624.
31. A. Faridmoayer; M. A. Fentabil; D. C. Mills; J. S. Klassen; M. F. Feldman, *J. Bacteriol.* **2007**, *189*, 8088–8098.
32. P. M. Power; K. L. Seib; M. P. Jennings, *Biochem. Biophys. Res. Commun.* **2006**, *347*, 904–908.
33. N. Jentoft, *Trends Biochem. Sci.* **1990**, *15*, 291–294.
34. J. C. Paulson, *Trends Biochem. Sci.* **1989**, *14*, 272–276.
35. A. Varki, *Glycobiology* **1993**, *3*, 97–130.
36. Y. J. Kim; A. Varki, *Glycoconj. J.* **1997**, *14*, 569–576.
37. J. Taylorpapadimitriou; A. A. Epenetos, *Trends Biotechnol.* **1994**, *12*, 227–233.
38. B. N. Fry; V. Korolik; J. A. ten Brinke; M. T. Pennings; R. Zalm; B. J. Teunis; P. J. Coloe; B. A. van der Zeijst, *Microbiology* **1998**, *144* (Pt. 8), 2049–2061.
39. C. M. Szymanski; D. H. Burr; P. Guerry, *Infect. Immun.* **2002**, *70*, 2242–2244.
40. J. C. Larsen; C. Szymanski; P. Guerry, *J. Bacteriol.* **2004**, *186*, 6508–6514.
41. M. Wacker; D. Linton; P. G. Hitchen; M. Nita-Lazar; S. M. Haslam; S. J. North; M. Panico; H. R. Morris; A. Dell; B. W. Wren; M. Aebi, *Science* **2002**, *298*, 1790–1793.
42. M. F. Feldman; M. Wacker; M. Hernandez; P. G. Hitchen; C. L. Marolda; M. Kowarik; H. R. Morris; A. Dell; M. A. Valvano; M. Aebi, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3016–3021.
43. M. S. Anderson; S. S. Eveland; N. P. Price, *FEMS Microbiol. Lett.* **2000**, *191*, 169–175.
44. D. Linton; N. Dorrell; P. G. Hitchen; S. Amber; A. V. Karlyshev; H. R. Morris; A. Dell; M. A. Valvano; M. Aebi; B. W. Wren, *Mol. Microbiol.* **2005**, *55*, 1695–1703.
45. N. B. Olivier; M. M. Chen; J. R. Behr; B. Imperiali, *Biochemistry* **2006**, *45*, 13659–13669.
46. K. J. Glover; E. Weerapana; S. Numao; B. Imperiali, *Chem. Biol.* **2005**, *12*, 1311–1315.
47. C. Alaimo; I. Catrein; L. Morf; C. L. Marolda; N. Callewaert; M. A. Valvano; M. F. Feldman; M. Aebi, *EMBO J.* **2006**, *25*, 967–976.
48. M. Kowarik; N. M. Young; S. Numao; B. L. Schulz; I. Hug; N. Callewaert; D. C. Mills; D. C. Watson; M. Hernandez; J. F. Kelly; M. Wacker; M. Aebi, *EMBO J.* **2006**, *25*, 1957–1966.
49. M. Nita-Lazar; M. Wacker; B. Schegg; S. Amber; M. Aebi, *Glycobiology* **2005**, *15*, 361–367.
50. M. Kowarik; S. Numao; M. F. Feldman; B. L. Schulz; N. Callewaert; E. Kiermaier; I. Catrein; M. Aebi, *Science* **2006**, *314*, 1148–1150.
51. M. Wacker; M. F. Feldman; N. Callewaert; M. Kowarik; B. R. Clarke; N. L. Pohl; M. Hernandez; E. D. Vines; M. A. Valvano; C. Whitfield; M. Aebi, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7088–7093.
52. M. M. Chen; E. Weerapana; E. Ciepchal; J. Stupak; C. W. Reid; E. Swiezewska; B. Imperiali, *Biochemistry* **2007**, *46*, 14342–14348.
53. M. F. Feldman; C. L. Marolda; M. A. Monteiro; M. B. Perry; A. J. Parodi; M. A. Valvano, *J. Biol. Chem.* **1999**, *274*, 35129–35138.
54. S. Nakagawa; Y. Takaki; S. Shimamura; A. L. Reysenbach; K. Takai; K. Horikoshi, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12146–12150.
55. P. Thibault; S. M. Logan; J. F. Kelly; J. R. Brisson; C. P. Ewing; T. J. Trust; P. Guerry, *J. Biol. Chem.* **2001**, *276*, 34862–34870.
56. S. M. Logan; J. F. Kelly; P. Thibault; C. P. Ewing; P. Guerry, *Mol. Microbiol.* **2002**, *46*, 587–597.
57. M. Marceau; K. Forest; J. Beretti; J. Tainer; X. Nassif, *Mol. Microbiol.* **1998**, *27*, 705–715.
58. E. Stimson; M. Virji; K. Makepeace; A. Dell; H. R. Morris; G. Payne; J. R. Saunders; M. P. Jennings; S. Barker; M. Panico; I. Blench; E. R. Moxon, *Mol. Microbiol.* **1995**, *17*, 1201–1214.
59. M. Schirm; S. K. Arora; A. Verma; E. Vinogradov; P. Thibault; R. Ramphal; S. M. Logan, *J. Bacteriol.* **2004**, *186*, 2523–2531.
60. A. Verma; M. Schirm; S. K. Arora; P. Thibault; S. M. Logan; R. Ramphal, *J. Bacteriol.* **2006**, *188*, 4395–4403.
61. I. Benz; M. A. Schmidt, *Mol. Microbiol.* **2001**, *40*, 1403–1413.
62. O. Sherlock; U. Dobrindt; J. B. Jensen; R. Munk Vejborg; P. Klemm, *J. Bacteriol.* **2006**, *188*, 1798–1807.
63. S. K. Knudsen; A. Stensballe; M. Franzmann; U. B. Westergaard; D. E. Otzen, *Biochem. J.* **2008**, *412*, 563–577.
64. T. Morooka; A. Umeda; K. Amako, *J. Gen. Microbiol.* **1985**, *131*, 1973–1980.
65. I. Nachamkin; X. H. Yang; N. J. Stern, *Appl. Environ. Microbiol.* **1993**, *59*, 1269–1273.
66. O. R. Pavlovskis; D. M. Rollins; R. L. Haberberger, Jr.; A. E. Green; L. Habash; S. Strocko; R. I. Walker, *Infect. Immun.* **1991**, *59*, 2259–2264.
67. T. Takata; S. Fujimoto; K. Amako, *Infect. Immun.* **1992**, *60*, 3596–3600.
68. T. M. Wassenaar; N. M. Bleumink-Pluym; B. A. van der Zeijst, *EMBO J.* **1991**, *10*, 2055–2061.
69. P. Guerry; R. A. Alm; M. E. Power; S. M. Logan; T. J. Trust, *J. Bacteriol.* **1991**, *173*, 4757–4764.
70. R. M. Macnab, *Annu. Rev. Microbiol.* **2003**, *57*, 77–100.
71. S. M. Logan; T. J. Trust; P. Guerry, *J. Bacteriol.* **1989**, *171*, 3031–3038.
72. P. Doig; N. Kinsella; P. Guerry; T. J. Trust, *Mol. Microbiol.* **1996**, *19*, 379–387.
73. Y. A. Knirel; N. A. Kocharova; A. S. Shashkov; B. A. Dmitriev; N. K. Kochetkov; E. S. Stanislavsky; G. M. Mashilova, *Eur. J. Biochem.* **1987**, *163*, 639–652.
74. W. K. Chou; S. Dick; W. W. Wakarchuk; M. E. Tanner, *J. Biol. Chem.* **2005**, *280*, 35922–35928.
75. F. Liu; M. E. Tanner, *J. Biol. Chem.* **2006**, *281*, 20902–20909.
76. D. J. McNally; A. J. Aubry; J. P. M. Hui; N. H. Khieu; D. Whitfield; C. P. Ewing; P. Guerry; J. R. Brisson; S. M. Logan; E. C. Soo, *J. Biol. Chem.* **2007**, *282*, 14463–14475.

77. D. J. McNally; J. P. Hui; A. J. Aubry; K. K. Mui; P. Guerry; J. R. Brisson; S. M. Logan; E. C. Soo, *J. Biol. Chem.* **2006**, *281*, 18489–18498.
78. M. Schirm; I. C. Schoenhofen; S. M. Logan; K. C. Waldron; P. Thibault, *Anal. Chem.* **2005**, *77*, 7774–7782.
79. I. C. Schoenhofen; V. V. Lunin; J. P. Julien; Y. Li; E. Ajamian; A. Matte; M. Cygler; J. R. Brisson; A. Aubry; S. M. Logan; S. Bhatia; W. W. Wakarchuk; N. M. Young, *J. Biol. Chem.* **2006**, *281*, 8907–8916.
80. I. C. Schoenhofen; D. J. McNally; E. Vinogradov; D. Whitfield; N. M. Young; S. Dick; W. W. Wakarchuk; J. R. Brisson; S. M. Logan, *J. Biol. Chem.* **2006**, *281*, 723–732.
81. S. Goon; J. F. Kelly; S. M. Logan; C. P. Ewing; P. Guerry, *Mol. Microbiol.* **2003**, *50*, 659–671.
82. P. Guerry; P. Doig; R. A. Alm; D. H. Burr; N. Kinsella; T. J. Trust, *Mol. Microbiol.* **1996**, *19*, 369–378.
83. P. Guerry, *Trends Microbiol.* **2007**, *15*, 456–461.
84. M. Schirm; E. C. Soo; A. J. Aubry; J. Austin; P. Thibault; S. M. Logan, *Mol. Microbiol.* **2003**, *48*, 1579–1592.
85. J. S. Allison; M. Dawson; D. Drake; T. C. Montie, *Infect. Immun.* **1985**, *49*, 770–774.
86. B. Lanyi, *Acta Microbiol. Acad. Sci. Hung.* **1970**, *17*, 35–40.
87. P. A. Totten; S. Lory, *J. Bacteriol.* **1990**, *172*, 7188–7199.
88. C. D. Brimer; T. C. Montie, *J. Bacteriol.* **1998**, *180*, 3209–3217.
89. S. K. Arora; M. Bangerla; S. Lory; R. Ramphal, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9342–9347.
90. A. Verma; R. Ramphal, Glycosylation Islands of *Pseudomonas* Species. In *Pseudomonas: A Model System in Biology*; J. L. Ramos, A. Filloux, Eds.; Springer: Netherlands, 2007; Vol. 5, pp 31–56.
91. R. Rahim; L. L. Burrows; M. A. Monteiro; M. B. Perry; J. S. Lam, *Microbiology* **2000**, *146* (Pt. 11), 2803–2814.
92. S. K. Arora; M. C. Wolfgang; S. Lory; R. Ramphal, *J. Bacteriol.* **2004**, *186*, 2115–2122.
93. F. Taguchi; K. Takeuchi; E. Katoh; K. Murata; T. Suzuki; M. Marutani; T. Kawasaki; M. Eguchi; S. Katoh; H. Kaku; C. Yasuda; Y. Inagaki; K. Toyoda; T. Shiraishi; Y. Ichinose, *Cell Microbiol.* **2006**, *8*, 923–938.
94. K. Takeuchi; H. Ono; M. Yoshida; T. Ishii; E. Katoh; F. Taguchi; R. Miki; K. Murata; H. Kaku; Y. Ichinose, *J. Bacteriol.* **2007**, *189*, 6945–6956.
95. K. Takeuchi; F. Taguchi; Y. Inagaki; K. Toyoda; T. Shiraishi; Y. Ichinose, *J. Bacteriol.* **2003**, *185*, 6658–6665.
96. S. K. Arora; A. N. Neely; B. Blair; S. Lory; R. Ramphal, *Infect. Immun.* **2005**, *73*, 4395–4398.
97. F. Taguchi; S. Shibata; T. Suzuki; Y. Ogawa; S. Aizawa; K. Takeuchi; Y. Ichinose, *J. Bacteriol.* **2008**, *190*, 764–768.
98. A. B. Semmler; C. B. Whitchurch; J. S. Mattick, *Microbiology* **1999**, *145* (Pt. 10), 2863–2873.
99. M. S. Strom; S. Lory, *Annu. Rev. Microbiol.* **1993**, *47*, 565–596.
100. J. V. Kus; E. Tullis; D. G. Cvitkovitch; L. L. Burrows, *Microbiology* **2004**, *150*, 1315–1326.
101. M. Qutyan; M. Paliotti; P. Castric, *Mol. Microbiol.* **2007**, *66*, 1444–1458.
102. L. L. Burrows; D. F. Charter; J. S. Lam, *Mol. Microbiol.* **1996**, *22*, 481–495.
103. C. Creuzenet; J. S. Lam, *Mol. Microbiol.* **2001**, *41*, 1295–1310.
104. H. L. Rocchetta; L. L. Burrows; J. S. Lam, *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 523–553.
105. A. DiGandomenico; M. J. Mawehish; A. Bisailon; J. R. Stehle; J. S. Lam; P. Castric, *Mol. Microbiol.* **2002**, *46*, 519–530.
106. J. E. Comer; M. A. Marshall; V. J. Blanch; C. D. Deal; P. Castric, *Infect. Immun.* **2002**, *70*, 2837–2845.
107. J. Horzempa; J. E. Comer; S. A. Davis; P. Castric, *J. Biol. Chem.* **2006**, *281*, 1128–1136.
108. J. Horzempa; C. R. Dean; J. B. Goldberg; P. Castric, *J. Bacteriol.* **2006**, *188*, 4244–4252.
109. J. G. Smedley 3rd; E. Jewell; J. Roguskie; J. Horzempa; A. Syboldt; D. B. Stolz; P. Castric, *Infect. Immun.* **2005**, *73*, 7922–7931.
110. H. S. Park; M. Wolfgang; M. Koomey, *Infect. Immun.* **2002**, *70*, 3891–3903.
111. L. Plant; V. Asp; L. Lovkvist; J. Sundqvist; A. B. Jonsson, *Cell Microbiol.* **2004**, *6*, 663–670.
112. C. D. Long; R. N. Madraswala; H. S. Seifert, *Infect. Immun.* **1998**, *66*, 1918–1927.
113. R. Haas; H. Schwarz; T. F. Meyer, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 9079–9083.
114. E. R. Gubish, Jr.; K. C. Chen; T. M. Buchanan, *Infect. Immun.* **1982**, *37*, 189–194.
115. J. N. Robertson; P. Vincent; M. E. Ward, *J. Gen. Microbiol.* **1977**, *102*, 169–177.
116. F. T. Hegge; P. G. Hitchen; F. E. Aas; H. Kristiansen; C. Lovold; W. Egge-Jacobsen; M. Panico; W. Y. Leong; V. Bull; M. Virji; H. R. Morris; A. Dell; M. Koomey, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10798–10803.
117. M. Marceau; X. Nassif, *J. Bacteriol.* **1999**, *181*, 656–661.
118. P. M. Power; M. P. Jennings, *FEMS Microbiol. Lett.* **2003**, *218*, 211–222.
119. M. P. Jennings; M. Virji; D. Evans; V. Foster; Y. N. Srikhanta; L. Steeghs; P. van der Ley; E. R. Moxon, *Mol. Microbiol.* **1998**, *29*, 975–984.
120. A. Banerjee; R. Wang; S. L. Supernavage; S. K. Ghosh; J. Parker; N. F. Ganesh; P. G. Wang; S. Gulati; P. A. Rice, *J. Exp. Med.* **2002**, *196*, 147–162.
121. D. M. Kahler; W. Davis, *J. Bone Joint Surg. Am.* **2001**, *83A*, 142–143.
122. P. M. Power; L. F. Roddam; K. Rutter; S. Z. Fitzpatrick; Y. N. Srikhanta; M. P. Jennings, *Mol. Microbiol.* **2003**, *49*, 833–847.
123. K. J. Glover; E. Weerapana; M. M. Chen; B. Imperiali, *Biochemistry* **2006**, *45*, 5343–5350.
124. J. Chamot-Rooke; B. Rousseau; F. Lanternier; G. Mikaty; E. Mairey; C. Malosse; G. Bouchoux; V. Pelicic; L. Camoin; X. Nassif; G. Dumenil, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14783–14788.
125. M. J. Warren; L. F. Roddam; P. M. Power; T. D. Terry; M. P. Jennings, *FEMS Immunol. Med. Microbiol.* **2004**, *41*, 43–50.
126. C. M. Kahler; D. S. Stephens, *Crit. Rev. Microbiol.* **1998**, *24*, 281–334.
127. E. Stimson; M. Virji; S. Barker; M. Panico; I. Blench; J. Saunders; G. Payne; E. R. Moxon; A. Dell; H. R. Morris, *Biochem. J.* **1996**, *316* (Pt. 1), 29–33.
128. P. Klemm; R. M. Vejborg; O. Sherlock, *Int. J. Med. Microbiol.* **2006**, *296*, 187–195.
129. U. Niewerth; A. Frey; T. Voss; C. Le Bouguenec; G. Baljer; S. Franke; M. A. Schmidt, *Clin. Diagn. Lab. Immunol.* **2001**, *8*, 143–149.
130. C. Lindenthal; E. A. Elsinghorst, *Infect. Immun.* **2001**, *69*, 52–57.
131. C. Moormann; I. Benz; M. A. Schmidt, *Infect. Immun.* **2002**, *70*, 2264–2270.
132. M. G. de Luna; A. Scott-Tucker; M. Desvaux; P. Ferguson; N. P. Morin; E. G. Dudley; S. Turner; J. P. Nataro; P. Owen; I. R. Henderson, *FEMS Microbiol. Lett.* **2008**, *284*, 237–246.

133. E. Andrian; D. Grenier; M. Rouabhia, *J. Dent. Res.* **2006**, *85*, 392–403.
134. J. Potempa; A. Sroka; T. Imamura; J. Travis, *Curr. Protein Pept. Sci.* **2003**, *4*, 397–407.
135. A. Gallagher; J. Aduse-Opoku; M. Rangarajan; J. M. Slaney; M. A. Curtis, *Curr. Protein Pept. Sci.* **2003**, *4*, 427–441.
136. R. Nakao; Y. Tashiro; N. Nomura; S. Kosono; K. Ochiai; H. Yonezawa; H. Watanabe; H. Senpuku, *Biochem. Biophys. Res. Commun.* **2008**, *365*, 784–789.
137. C. M. Fletcher; M. J. Coyne; D. L. Bentley; O. F. Villa; L. E. Comstock, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2413–2418.
138. Z. Peng; H. Wu; T. Ruiz; Q. Chen; M. Zhou; B. Sun; P. Fives-Taylor, *Oral Microbiol. Immunol.* **2008**, *23*, 70–78.
139. Z. X. Peng; P. Fives-Taylor; T. Ruiz; M. Zhou; B. M. Sun; Q. Chen; H. Wu, *BMC Microbiol.* **2008**, *8*, 52.
140. M. L. Kalmokoff; J. W. Austin; M. F. Whitford; R. M. Teather, *Can. J. Microbiol.* **2000**, *46*, 295–303.
141. M. Lyrstis; Z. L. Boynton; D. Petersen; Z. Kan; G. N. Bennett; F. B. Rudolph, *Anaerobe* **2000**, *6*, 69–79.
142. U. B. Sleytr; K. J. Thorne, *J. Bacteriol.* **1976**, *126*, 377–383.
143. C. Schaffer; P. Messner, *Glycobiology* **2004**, *14*, 31R–42R.
144. P. Messner; K. Steiner; K. Zarschler; C. Schaffer, *Carbohydr. Res.* **2008**, *343*, 1934–1951.
145. C. Schaffer; T. Wugeditsch; H. Kahlig; A. Scheberl; S. Zayni; P. Messner, *J. Biol. Chem.* **2002**, *277*, 6230–6239.
146. K. Steiner; G. Pohlentz; K. Dreisewerd; S. Berkenkamp; P. Messner; J. Peter-Katalinic; C. Schaffer, *J. Bacteriol.* **2006**, *188*, 7914–7921.
147. K. Steiner; R. Novotny; D. B. Werz; K. Zarschler; P. H. Seeberger; A. Hofinger; P. Kosma; C. Schaffer; P. Messner, *J. Biol. Chem.* **2008**, *283*, 21120–21133.
148. L. Lehle; S. Strahl; W. Tanner, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 6802–6818.
149. L. A. Jurado; A. Coloma; J. Cruces, *Genomics* **1999**, *58*, 171–180.
150. S. Strahl-Bolsinger; A. Scheinost, *J. Biol. Chem.* **1999**, *274*, 9068–9075.
151. K. M. Dobos; K. H. Khoo; K. M. Swiderek; P. J. Brennan; J. T. Belisle, *J. Bacteriol.* **1996**, *178*, 2498–2506.
152. S. L. Michell; A. O. Whelan; P. R. Wheeler; M. Panico; R. L. Easton; A. T. Etienne; S. M. Haslam; A. Dell; H. R. Morris; S. A. Reason; J. L. Herrmann; D. B. Young; R. G. Hewinson, *J. Biol. Chem.* **2003**, *278*, 16423–16432.
153. B. C. VanderVen; J. D. Harder; D. C. Crick; J. T. Belisle, *Science* **2005**, *309*, 941–943.
154. C. Harty; S. Strahl; K. Romisch, *Mol. Biol. Cell* **2001**, *12*, 1093–1101.
155. I. M. Nilsson; G. von Heijne, *J. Biol. Chem.* **1993**, *268*, 5798–5801.
156. Y. Ge; C. Li; L. Corum; C. A. Slaughtner; N. W. Charon, *J. Bacteriol.* **1998**, *180*, 2418–2425.
157. V. Sambri; C. Stefanelli; R. Cevenini, *Arch. Microbiol.* **1992**, *157*, 205–208.
158. J. Sterba; M. Vancova; N. Rudenko; M. Golovchenko; T. L. Tremblay; J. F. Kelly; C. R. MacKenzie; S. M. Logan; L. Grubhoffer, *J. Bacteriol.* **2008**, *190*, 2619–2623.
159. G. Walsh, *Nat. Biotechnol.* **2003**, *21*, 865–870.
160. S. Elliott; T. Lorenzini; S. Asher; K. Aoki; D. Brankow; L. Buck; L. Busse; D. Chang; J. Fuller; J. Grant; N. Hernday; M. Hokum; S. Hu; A. Knudten; N. Levin; R. Komorowski; F. Martin; R. Navarro; T. Osslund; G. Rogers; N. Rogers; G. Trail; J. Egrie, *Nat. Biotechnol.* **2003**, *21*, 414–421.
161. T. D. Bartley; J. Bogenberger; P. Hunt; Y.-S. Li; H. S. Lu; F. Martin; M.-S. Chang; B. Samal; J. L. Nichol; S. Swift; M. J. Johnson; R.-Y. Hsu; V. P. Parker; S. Suggs; J. D. Skrine; L. A. Merewether; C. Clogston; E. Hsu; M. M. Hokom; A. Hornkoh; E. Choi; M. Pangelinan; Y. Sun; V. Mar; J. McNinch; L. Simonet; F. Jacobsen; C. Xie; J. Shutter; H. Chute; R. Basu; L. Selander; D. Trollinger; L. Sieu; D. Padilla; G. Trail; G. Elliott; R. Izumi; T. Covey; J. Crouse; A. Garcia; W. Xu; J. Del Castillo; J. Biron; S. Cole; M. C.-T. Hu; R. Pacifici; I. Ponting; C. Saris; D. Wen; Y. P. Yung; H. Lin; R. A. Rosselman, *Cell* **1994**, *77*, 1117–1124.
162. P. Hunt; Y. S. Li; J. L. Nichol; M. M. Hokom; J. M. Bogenberger; S. E. Swift; J. D. Skrine; A. C. Hornkoh; H. Lu; C. Clogston; L. A. Merewether; M. J. Johnson; V. Parker; A. Knudten; A. Farese; R. Y. Hsu; A. Garcia; R. Stead; R. A. Bosselman; T. D. Bartley, *Blood* **1995**, *86*, 540–547.
163. Y. Zhang; R. Proenca; M. Maffei; M. Barone; L. Leopold; J. M. Friedman, *Nature* **1994**, *372*, 425–432.
164. P. Bobrowicz; R. C. Davidson; H. Li; T. I. Potgieter; J. H. Nett; S. R. Hamilton; T. A. Stadheim; R. G. Miele; B. Bobrowicz; T. Mitchell; S. Rausch; E. Renfer; S. Wildt, *Glycobiology* **2004**, *14*, 757–766.
165. S. R. Hamilton; R. C. Davidson; N. Sethuraman; J. H. Nett; Y. Jiang; S. Rios; P. Bobrowicz; T. A. Stadheim; H. Li; B. K. Choi; D. Hopkins; H. Wischnewski; J. Roser; T. Mitchell; R. R. Strawbridge; J. Hoopes; S. Wildt; T. U. Gerngross, *Science* **2006**, *313*, 1441–1443.
166. K. Amano; Y. Chiba; Y. Kasahara; Y. Kato; M. K. Kaneko; A. Kuno; H. Ito; K. Kobayashi; J. Hirabayashi; Y. Jigami; H. Narimatsu, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3232–3237.
167. T. U. Gerngross, *Nat. Biotechnol.* **2004**, *22*, 1409–1414.
168. A. Clements; A. W. Jenney; J. L. Farn; L. E. Brown; G. Deliyannis; E. L. Hartland; M. J. Pearse; M. B. Maloney; S. L. Wesselingh; O. L. Wijburg; R. A. Strugnell, *Vaccine* **2008**, *26*, 5649–5653.
169. C. Grandjean; A. Boutonnier; B. Dassy; L. A. Mulard; J. M. Fournier, *Glycoconj. J.* **2009**, *26*, 41–55.
170. G. Nagy; T. Pal, *Biol. Chem.* **2008**, *389*, 513–520.
171. A. Weintraub, *Carbohydr. Res.* **2003**, *338*, 2539–2547.
172. J. Horzempa; T. K. Held; A. S. Cross; D. Furst; M. Qutyan; A. N. Neely; P. Castric, *Clin. Vaccine Immunol.* **2008**, *15*, 590–597.

Biographical Sketches

Amirreza Faridmoayer obtained his bachelor's degree from the Isfahan University of Technology and master's degree from the Shahid Beheshti University in Iran. He pursued his doctoral studies at the University of British Columbia, Vancouver, Canada. During his Ph.D. studies Dr. Faridmoayer worked on different aspects of protein glycosylation in yeast, particularly the processing of alpha-glucosidases, the glycosyl hydrolases that are involved in the quality control of newly formed N-linked glycoproteins. Thereafter, he joined the Alberta Ingenuity Centre for Carbohydrate Science at the University of Alberta, Edmonton, Canada to pursue postdoctoral research in the field of bacterial protein glycosylation with Dr. Mario Feldman. His current research is focused on the elucidation of mechanisms that are involved in bacterial protein O-glycosylation and their applications in glycoengineering.



Mario F. Feldman obtained his bachelor's degree in biotechnology from the University of Rosario, Argentina. He later obtained his Ph.D. degree from the University of Buenos Aires, Argentina working under the supervision of Dr. Armando Parodi. His Ph.D. thesis focused on the study of the flippases involved in LPS biosynthesis in *Escherichia coli*. Dr. Feldman carried out postdoctoral work on the chaperones involved in type III secretion systems with Dr. Guy Cornelis at the University of Louvain, Brussels, Belgium, and at the Biozentrum, Basel, Switzerland. Dr. Feldman obtained additional postdoctoral experience with Dr. Markus Aebi at the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. In association with Dr. Aebi, he worked on the characterization of PglB, the key enzyme of the *Campylobacter jejuni* N-glycosylation system. In 2006 he became an assistant professor at the Department of Biological Sciences at the University of Alberta and principal investigator at the Alberta Ingenuity Centre for Carbohydrate Science. His current research is focused on the study of the synthesis of LPS and glycoproteins in diverse bacteria, and their applications for the design of novel vaccines and glycan-based therapeutics through glycoengineering.

6.13 Structure and Biosynthesis of the Mycobacterial Cell Wall

Dean C. Crick, Delphi Chatterjee, Michael S. Scherman, and Michael R. McNeil, Colorado State University, Fort Collins, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.13.1 | Introduction | 381 |
| 6.13.2 | Peptidoglycan | 382 |
| 6.13.3 | Arabinogalactan | 382 |
| 6.13.4 | Phosphatidylinositol Mannoside, Lipomannan, Lipoarabinomannan, and Arabinomannan Structure | 383 |
| 6.13.5 | Peptidoglycan Synthesis | 385 |
| 6.13.6 | Biosynthesis of Arabinogalactan | 388 |
| 6.13.7 | Biosynthesis of the Phosphatidylinositol Containing Phosphatidylinositol Mannosides, Lipomannans, and Lipoarabinomannans | 395 |
| 6.13.8 | Mycobacterial Cell Envelope Ultrastructure | 397 |
| References | | 401 |

6.13.1 Introduction

Mycobacterium tuberculosis, the etiological agent of tuberculosis (TB), is able to persist inside the host for long periods of time without clinical symptoms, a condition that may exist through the host's life. It is estimated that more than one-third of the world's population is infected with tubercle bacilli¹ and 5–10% of these individuals will develop TB at some point in their lifetime with a death toll of approximately 1.7 million people in 2007.²

Mycobacteria are classified as Gram-positive organisms; however, they have features of both Gram-positive and Gram-negative bacteria and the compositional complexity of the mycobacterial cell envelope differentiates *Mycobacterium* species from most other prokaryotes. The cell envelope of *M. tuberculosis* is made up of three major components: a plasma membrane; a covalently linked mycolic acid, arabinogalactan, and peptidoglycan complex (MAPc); and a polysaccharide-rich capsule-like material. This envelope also contains a variety of extractable lipids and pore-forming proteins. The overall composition of the mycobacterial cell envelope is thought to provide an extraordinary permeability barrier, which renders mycobacteria resistant to many drugs.^{3–5}

It was shown, many years ago, that the major cell wall polysaccharide of mycobacteria, approximately 35% of the mass, was a branched-chain arabinogalactan (AG) with the arabinose (Ara) residues forming the nonreducing termini. This AG was subsequently shown to be covalently attached to muramic acid (Mur) residues through a phosphodiester linkage.⁶ More recently, it was shown that the AG is unique in that the Ara and galactose (Gal) residues are in the furanose configuration;⁷ hereafter abbreviated as Araf, and Galf. In addition, the cell envelope of *Mycobacterium* spp. contains a number of other unusual carbohydrate residues including α -L-rhamnose (Rha), succinylated Araf, and N-glycolylmuramic acid (MurNGlyc). The bulk of these unusual sugars are found in three polymers: the MAPc, lipoarabinomannan (LAM), and arabinomannan (AM).

The mycobacterial plasma membrane is quite similar to the plasma membrane found in most bacteria. The peptidoglycan (PG) is found in a phosphodiester linkage to a linear D-galactofuran through a 'linker region' composed of N-acetylglucosamine (GlcNAc) and Rha, to which are attached several strands of a highly branched D-Araf. Linked to the nonreducing termini of the arabinan are mycolic acids, which are oriented perpendicular to the plane of the membrane⁸ and are thought to be the basis of the permeability barrier mentioned above. In addition to the mycolic acids are extractable (noncovalently linked) lipids and other components of the capsule-like layer. This material is composed of free lipids, proteins, and carbohydrates including phosphatidylinositol mannosides (PIMs), lipomannan (LM), LAM, trehalose dimycolate (TDM or cord factor), sulfolipids, trehalose monomycolate (TMM), diacyl- and polyacyl-trehaloses,

phthiocerol-containing lipids, glucan, mannan, and AM. The lipids are generally thought to intercalate with the α -branch and meromycolate chains of the mycolic acids and cell-wall-associated proteins, while the carbohydrate polymers appear to be rather loosely associated with the outer surface of the envelope.

Much of the early structural definition of the cell wall of *Mycobacterium* spp.^{6,9–16} was conducted in the 1960s and 1970s, followed by a period of inactivity. However, developments in genomics, bioinformatics, proteomics, and analytical techniques have resulted in a more detailed and thorough understanding of the structure of the mycobacterial cell wall core and its biosynthesis. The goal of this chapter is to describe the structure and biosynthesis of the MAPc, LAM, and AM, and the ultrastructure of the mycobacterial cell envelope, in broad terms.

6.13.2 Peptidoglycan

PG is a polymer that forms a rigid layer outside of the plasma membrane, which provides cellular shape and resistance to osmotic pressure. PG from *M. tuberculosis* has been classified as A1 γ according to the classification system of Schleifer and Kandler¹⁷ or ‘chemotype IV’ cell wall¹⁸ and is analogous to that found in other bacteria such as *Escherichia coli* and *Bacillus* spp. The polymer is composed of linear chains of alternating 1–4-linked *N*-acetyl- α -D-glucosamine (GlcNAc) and modified Mur substituted with peptide side chains that may be cross-linked to those of other glycan strands. The tetrapeptide side chains of PG consist of L-alanyl-D-isoglutaminyl-*meso*-diaminopimelyl-D-alanine (L-Ala-D-Glu-DAP-D-Ala), with the Glu and DAP being further amidated.^{6,9,10,13,16} In mycobacteria, the overall degree of cross-linking is 70–80%¹²; thus, the peptide chains are heavily cross-linked, presumably providing added structural integrity to the bacterium. In comparison, the PG of *E. coli* has 50% of the peptides cross-linked.¹⁹ About 66% of the peptide cross-links found in *M. tuberculosis* PG are between the carboxyl group of a terminal D-Ala and the amino group at the D-center of DAP resulting in a D–D bond,¹⁵ and approximately 33% of the peptide cross-links occur between the carboxyl group of the L-center of one DAP residue and the amino group of the D-center of another DAP residue forming a L,D-cross-link.¹⁵ L,D-cross-links are relatively rare, but recently have also been reported in the PG of *Streptomyces* spp., *Clostridium perfringens*,²⁰ and stationary phase *E. coli*.^{21,22} *Mycobacterium leprae* is unusual in that it has a Gly residue substituted for the L-Ala residue found in the stem peptide.²³

Studies of the biosynthetic precursor UDP-*N*-acylmuramyl-L-Ala-D-Glu-DAP from *M. tuberculosis* and *Mycobacterium phlei*^{24,25} determined that all the Mur residues were, apparently, completely NGlyc substituted. Since the UDP-MurNGlyc-L-Ala-D-Glu-DAP is the immediate precursor of the UDP-MurNGlyc-L-Ala-D-Glu-DAP-D-Ala-D-Ala, it was assumed that all of the UDP-*N*-acylmuramyl-pentapeptides would be NGlyc substituted^{10,23,26} and that PG contains MurNGlyc exclusively, as has been reported.^{10,11,23} However, recent analysis found that the Mur residues in mature PG of *M. tuberculosis* and *Mycobacterium smegmatis* consist of a mixture of the NGlyc and NAc derivatives,²⁷ indicating that the precursors also have some *N*-acetyl groups on the Mur, and that PG from *M. leprae* contains MurNAc exclusively. In addition, the hydroxyl moiety of C-6 of some of the Mur residues participates in phosphodiester bonds to C-1 of GlcNAc, which in turn is (1 \rightarrow 3) linked to a Rha residue providing the ‘linker unit’ between the galactan of AG and PG.²⁸

6.13.3 Arabinogalactan

A branched-chain AG, with the Ara residues forming the nonreducing termini, was recognized as the major cell-wall polysaccharide in mycobacteria in the middle of the last century and numerous excellent reviews have been published describing the structure of the mycobacterial cell wall.^{3,26,29–35} With the exception of the GlcNAc and L-Rha making up the linker unit and the rare galactosamine (GalNH₂) residues, which are found only in slow-growing mycobacteria,³⁶ AG is composed entirely of Araf and Galf (Figure 1). Unlike most bacterial polysaccharides, AG lacks repeating units³⁷ and is comprised of a few distinct structural motifs.^{38,39} In the AG of *M. tuberculosis*, the galactan regions consist of linear alternating 5- and 6-linked β -D-Galf residues, believed to be approximately 30 residues long. The arabinan chains are attached to C-5 of some of the 6-linked Galf residues of the galactan core and the latest results suggest that three arabinan chains each are attached to

galactan chain, likely on galactosyl residues 8, 10, and 12.⁴⁰ This structural determination was done using *Corynebacterium glutamicum* but is probably applicable to *M. tuberculosis*, as it is consistent with earlier work showing long stretches of linear galactofuran in that species.⁴¹ The majority of the arabinan chains consist of approximately 30 5-linked α -D-Araf residues with branching introduced by the presence of 3,5- α -D-Araf residues. The nonreducing termini of the AG arabinan consist of the structural motif [β -D-Araf(1 \rightarrow 2)- α -D-Araf]₂-3,5- α -D-Araf-(1 \rightarrow 5)- α -D-Araf. Mycolic acids are located in clusters of four on the terminal [β -D-Araf(1 \rightarrow 2)- α -D-Araf]₂-3,5- α -D-Araf-(1 \rightarrow 5)- α -D-Araf hexaarabinofuranoside, but only two-thirds of these are mycolated. The mycolates are esterified to position 5 of both the nonreducing terminal β -D-Araf units and the penultimate 2-linked α -D-Araf residues.²⁸ In some cases, some of the nonreducing arms of the arabinan are fully substituted with mycolates and some are fully unsubstituted;²⁸ indeed, this all-or-nothing mycolation of a given arabinan chain is supported by recent data.⁴² There are some variations seen in other mycobacteria. For example, in *M. leprae*, *Mycobacterium bovis*, and *M. smegmatis*, only 50% of the arabinan termini are mycolylated.⁴³ Recent mass spectrometric analysis suggests that the arabinan chains released from the cell wall by an endogenous arabinase are also approximately 31 residues long⁴² and indicates that GalNH₂ residues in AG isolated from the *M. tuberculosis* CSU20 strain are found on the C-2 position of some of the internal 3,5-branched Araf residues.⁴⁴ In addition, the C-2 positions of other internal 3,5-branched Araf residues are succinylated.⁴² Approximately one-third of the arabinan chains in AG contain a GalNH₂ residue and one-third are succinylated. Thus, a model of the complete primary structure of the *M. tuberculosis* MAPc is now available. These structural features are summarized in **Figure 1**.

6.13.4 Phosphatidylinositol Mannoside, Lipomannan, Lipoarabinomannan, and Arabinomannan Structure

PIMs, LM, LAM, and AM appear to be structurally (**Figure 2**) and, hence, biosynthetically related to each other, in that LM appears to be mannosylated PIM, LAM appears to be arabinosylated LM, and AM appears to be LAM minus the lipid anchor. Thus, the fully mature *M. tuberculosis* LAM is most easily viewed as composite of four regions: a lipidated reducing end (in essence a phosphatidylinositol (PI) anchor), a mannan core (which together with the anchor makes up the LM region), and the arabinofuran with its mannose (Man) caps, and a 5-deoxy-5-methylthio-xylofuranose (MTX) or 5-deoxy-5-methylsulfoxy-xylofuranose (MTS). The lipid anchor (reducing end) is well understood, consisting of a PI unit^{45–47} with a Man residue linked to the 2-position of the Ino moiety. A number of acyl forms of the PIMs and, therefore, the lipid anchor exist.^{46,48–57} In all cases, the glycerol moiety is acylated on positions 1 and 2 as expected. Generally, these compounds are named according to the number of fatty acids esterified to the molecule; however, this nomenclature is confusing as some authors include the two fatty acids found on the glycerol moiety in the count, while others consider them to be understood and do not include them. Here we use a numbering system that includes the two acyl groups on the glycerol moiety when specific molecules are indicated; in other cases the more generic terms PIM and acyl-PIM are used where fatty acids are only found on the glycerol moiety or where more than two fatty acids found on the glycerol are present but the number and/or position are not defined.

The PIMs are one of several classes of phospholipids, including cardiolipin, diphosphatidylglycerol, phosphatidylinositol, and phosphatidylethanolamine, found in *M. tuberculosis*.⁵⁷ Of these, the PIMs are the most unusual and characteristic. The PIMs are substituted with up to six Man_p residues and are identified by the number of Man residues on the molecule; thus, PI with one Man_p residue is PIM₁. Although it is abundant in all mycobacteria the relative distribution varies from species to species.

Structurally, PIM₁ has a Man_p substituted on the 2-position of the *myo*-inositol (Ino), and PIM₂ contains a second Man on C-6 of the Ino.⁵⁸ PIM₃ has an additional Man_p linked to the C-6 position of the 6-linked Man_p residue. PIM₅ and PIM₆ contain α (1 \rightarrow 2)-linked Man_p (**Figure 2**) and therefore differ in structure from the other PIMs, LM, and LAM.^{50,59} The extent of acylation of PIM₁ and PIM₂ varies from the diacyl form consisting of palmitate and tuberculostearic acids esterified to the glycerol moiety to the triacyl and tetraacyl forms that contain additional palmitate moieties. Generally, the third acyl group is added to the C-6 OH of the Man_p linked to the 2-position of the Ino.^{48,50,60} In the case of tetraacylated form (Ac₄PIM₂), it has been

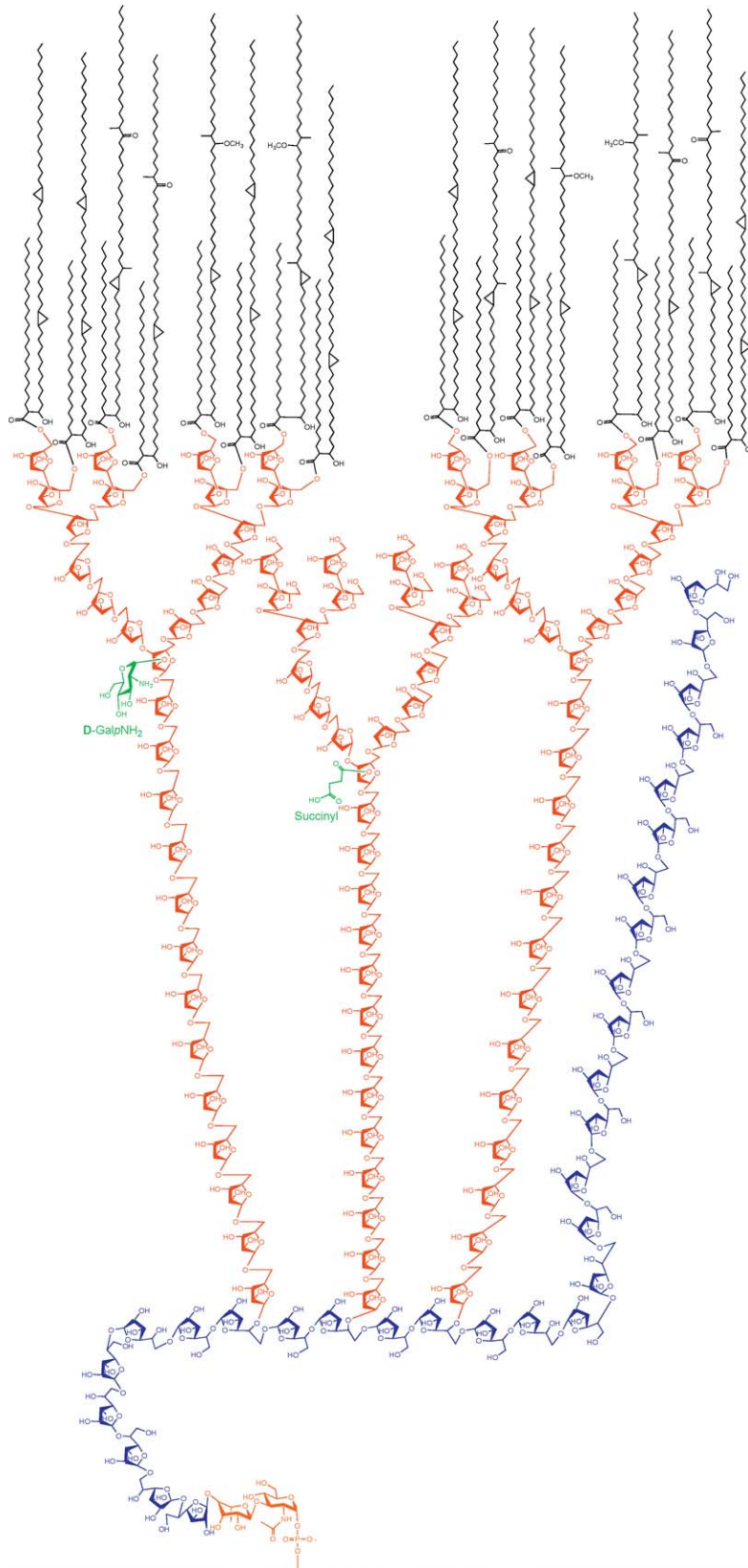


Figure 1 (continued)

suggested that there may be acylation of the inositol ring since, logically the addition of a fatty acyl group to the second Man of PIM₂ would result in termination of the glycosidic chain preventing synthesis of PIM₃–PIM₆.⁵⁰

The mannan backbone of LM and LAM is attached to the Ino of PIM at C-6, and consists of linear 6-linked α -D-Man_p units, to which are attached t- α -D-Man_p side chains on position 2,^{45,46,61} with a total of 20–30 Man_p residues found in *M. tuberculosis* LM. The arabinan side chains are attached to the mannan backbone at an unknown position. It was originally thought that LM contains an α -1,5 di (or tri) arabinoside attached to the mannan backbone, but recent work indicates that this is not the case.^{62,63} In addition, data strongly suggest that only a single arabinan chain is present on LAM⁶³ in contrast to the two or three arabinan side chains present on AG. This fact, along with the partial acid hydrolysis experiments,⁶¹ linkage analysis,^{61,64} and enzymatic degradation of LAM,⁶⁴ suggests that, as in AG, the majority of the arabinan chain consists of 5-linked α -D-Araf residues with branching introduced by the presence of 3,5- α -D-Araf residues. The arabinosides can be terminated by addition of β -D-Araf-(1 \rightarrow 2)- α -D-Araf at position 3 of the last 5-linked α -Araf unit or extended from the branch by a linear (1 \rightarrow 5)- α -D-Araf chain of six to seven residues followed by termination with the β -D-Araf-(1 \rightarrow 2)- α -D-Araf motif in *M. smegmatis* and probably *M. tuberculosis*.⁶⁵ Thus, the arabinan moiety of LAM is composed of an AG-like branched core and linear terminal stretches. The reducing end of the arabinan may be capped with a single Man_p, a dimannoside, or a trimannoside in *M. tuberculosis*, *M. leprae*, *Mycobacterium avium*, and *Mycobacterium kansasii*.^{53,55,66,67} In addition, a MTX or MTS residue may be attached to a Man_p residue through α -(1 \rightarrow 4) linkage.^{68–72} Analysis has also identified the presence of succinate residues, whose exact location remains unresolved^{66,73} and two-dimensional gel electrophoresis experiments indicate that LAM can be resolved into isoforms that are independent of capping or acylation status,⁷⁴ an observation that has yet to be explained structurally. The structure of the extracellular AM appears to be identical to that of LAM except for the loss of the lipid anchor.^{75,76}

6.13.5 Peptidoglycan Synthesis

The chromosomal arrangement of genes responsible for PG synthesis in *M. tuberculosis* is similar to that in other bacteria,⁷⁷ and based on the similarity of the basic structure, comparison of genome sequence, and direct investigations,^{10,24,25,27,33,77–79} it can reasonably be assumed that the mycobacterial PG biosynthetic pathway is similar to that of *E. coli* with the exception of additional enzymatic steps, which introduce the distinguishing features noted above (**Figure 3**). There are excellent reviews on this subject in other organisms^{80–84} and these should be applicable to mycobacteria.

Briefly, in *E. coli* the biosynthesis of PG initiates in cytosol with the formation of UDP-*N*-acetylmuramic acid (UDP-MurNAc) from UDP-GlcNAc by a two-step reaction catalyzed by MurA and MurB. UDP-MurNAc is subsequently transformed into UDP-MurNAc-L-Ala-D-Glu-*meso*-DAP-D-Ala-D-Ala (UDP-MurNAc-pentapeptide (UDP-MurNAc-PP) or Park's nucleotide) by sequential addition of L-Ala, D-Glu, *meso*-DAP, and D-Ala-D-Ala catalyzed by MurC, D, E, and F, respectively.

The final cytosolic precursor of PG biosynthesis, UDP-MurNAc-PP, is transferred to an undecaprenyl phosphate, decaprenyl phosphate (DP) in mycobacteria,⁷⁹ carrier by MraY forming MurNAc-PP-diphosphorylundecaprenol or lipid I. In a subsequent reaction catalyzed by MurG, a GlcNAc residue is transferred to lipid I forming lipid II. Lipid II, which is the final membrane-associated intermediate of PG biosynthesis, is also believed to be the substrate of a putative 'flippase', that translocates lipid II from the cytosolic face of the membrane, which is the site of synthesis to the outer layer of plasma membrane or the site of utilization; although

Figure 1 A schematic showing the structural features of *Mycobacterium tuberculosis* AG (as previously published⁴²). Arabinosyl residues are shown in red, the galactosyl residues in blue, and the linker in orange. Note that three arabinan chains of approximately 30 residues each are attached to a galactan also of about 30 residues. Some arabinan chains are substituted at the exterior branch point with succinyl residues, some with GalNH₂ residues and some are unsubstituted. As shown in the figure about two-thirds of the arabinan chains are substituted with eight mycolic acids each. The mycolic acids are shown in black with somewhat small bond lengths for presentation purposes; they are actually larger than shown with respect to AG.

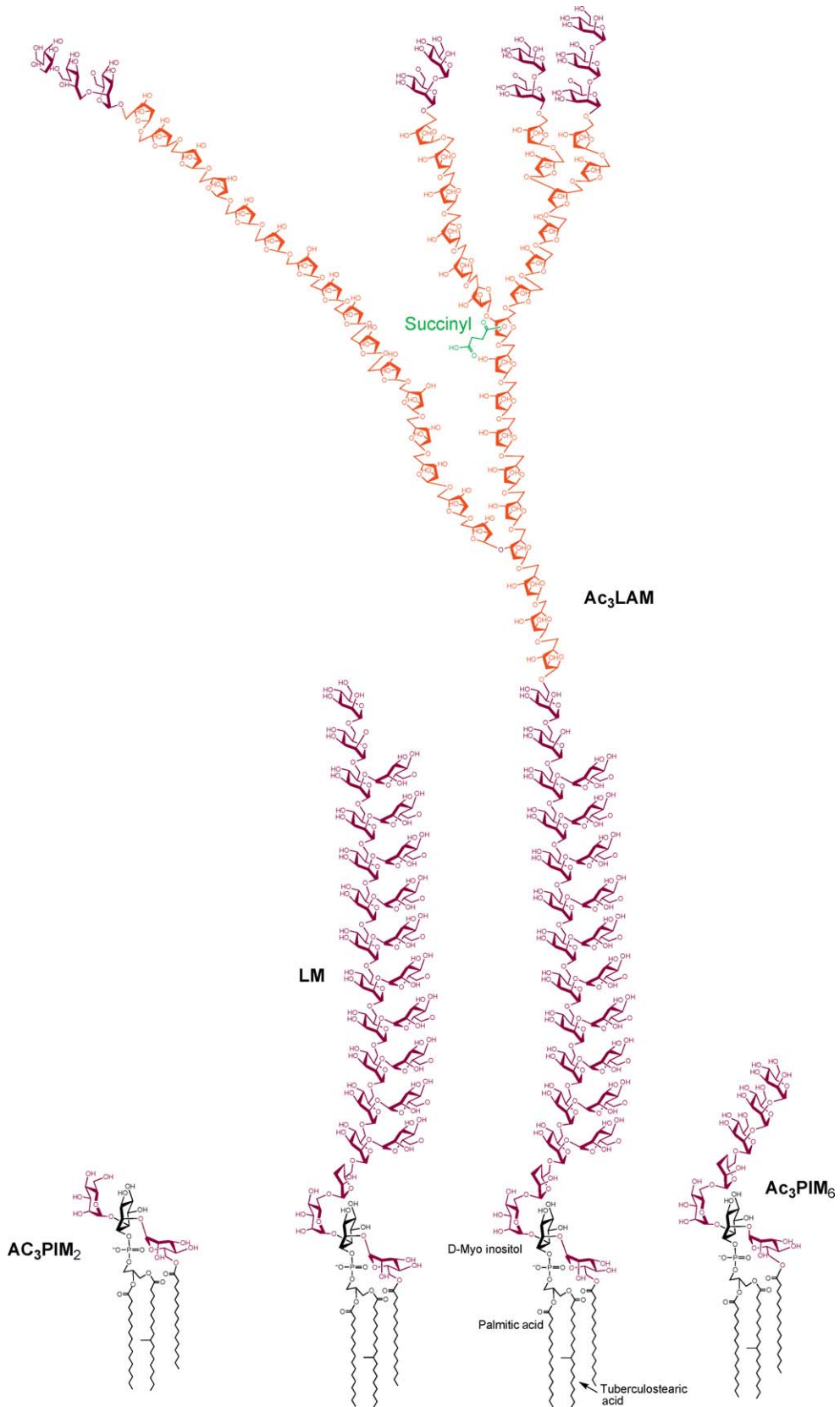


Figure 2 (continued)

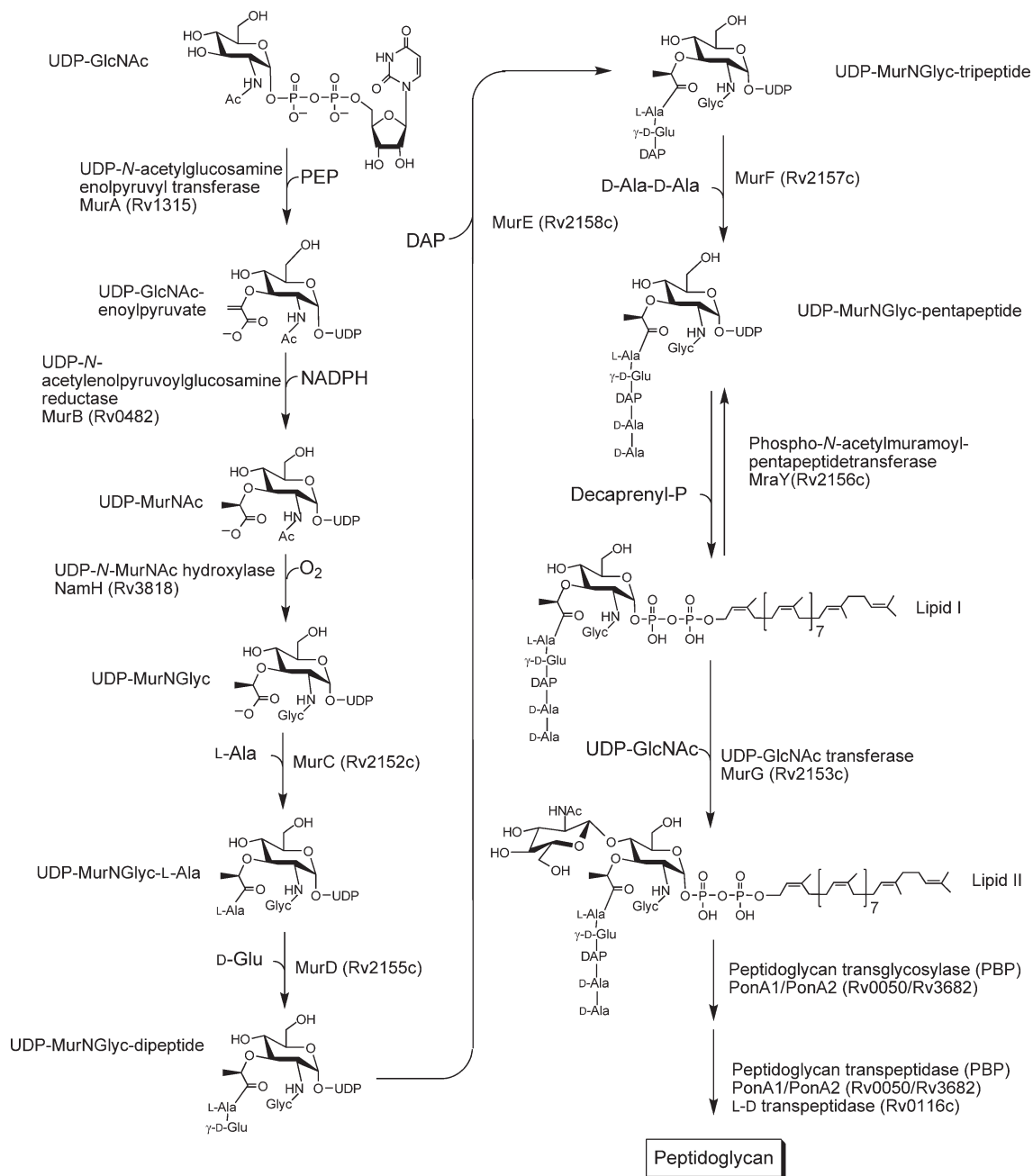


Figure 3 PG synthesis in *Mycobacterium tuberculosis*. Gene names and Rv identifiers are indicated where known. The pathway is quite similar to that found in other bacteria except for the oxidation of the acetyl residues by NamH.

Figure 2 A schematic diagram of selected PIM, LM, and LAM molecules. The arabinosyl residues are shown in red, the mannosyl residues in dark purple, and the lipid anchors in black. The actual point of attachment of the arabinan to the mannan in LAM is unknown; therefore, the position shown was chosen arbitrarily. The arabinan is much more heterogeneous in LAM than in AG with the branching less clearly defined and the structure shown is just one of many possible structures likely to be present in LAM. In contrast the structures of the PIMs are exactly known; LM is heterogeneous in size and the exact placement of nonbranched mannosyl residues is not clear.

the mechanism of the transbilayer movement of PG intermediates is unknown, a good deal of information is available regarding the transpeptidation and transglycosylation reactions required to form mature PG.

Thus, the final step of PG biosynthesis involves the formation of linear glycan chain by the polymerization of the disaccharide units of lipid II and subsequent cross-linking of the peptide side chains. These reactions are catalyzed by mono- and bifunctional enzymes known as penicillin-binding proteins (PBPs).

Nascent glycan chains are cross-linked by the formation of peptide cross-bridges. The main cross-linkage in *E. coli* is between the penultimate D-Ala and the DAP residues of an adjacent peptide from another glycan chain. The reaction is catalyzed by PBPs and involves the cleavage of the D-alanyl-D-alanine bond of the donor peptide, which provides the energy to drive the reaction.⁸⁵ The situation is similar in mycobacteria and mycobacterial PBPs are the subject of an extensive review.²¹ The PBPs are classified primarily according to their molecular mass and have penicilloyl serine transferase activity that catalyzes the cleavage of the cyclic amide bond of penicillin. The high molecular mass (HMM) PBPs are multimodular enzymes with both nonpenicillin-binding and penicillin-binding modules, which are subdivided into class A and class B based on their primary structures. The HMM class A PBPs function as transglycosylases and transpeptidases. Chambers *et al.*⁸⁶ detected four PBPs, including three of HMM in *M. tuberculosis* membranes; however, the transglycosylase or transpeptidase activities of these putative enzymes have not been demonstrated.

Mycobacterium smegmatis contains five PBPs,^{87,88} two of which have been disrupted by transposon mutagenesis. One strain, with an insertion in the gene encoding a protein that was most similar to Rv0050 (annotated as ponA1 in *M. tuberculosis*), grew slowly and had increased sensitivity to β -lactam antibiotics and greater permeability.⁸⁸ The other strain had an insertion in the gene encoding a protein most similar to Rv3682 (annotated as ponA2 in *M. tuberculosis*), which was impaired in stationary-phase survival.⁸⁹ In addition, it has been shown that PBP from *M. smegmatis* (with greatest similarity to Rv0016c in *M. tuberculosis*) plays an important role in cell division and maintenance of cell shape; signal transduction mediated by PknB and PstP appears to determine the positioning of this PBP at the septum, regulating septal PG biosynthesis.⁹⁰ The protein responsible, at least in stationary phase, for L,D-transpeptidation in *M. tuberculosis* has recently been identified as Rv0116c (LD_{TMeI}).⁹¹ Other unique issues for mycobacterial PG formation include the conversion of the *N*-acetyl group on the Mur into *N*-glycolyl and the esterification and amidation of the COOH groups of D-Glu and *meso*-DAP. Studies in the authors' laboratories suggest that the *N*-glycolyl group is generated after the formation of UDP-MurNAc and before the formation of UDP-Mur-pentapeptide. The oxidation of the *N*-acetyl group to *N*-glycolyl is catalyzed by an enzyme designated as Rv3818 or NamH.⁹² Although there is no indication that this modification is essential for *M. tuberculosis* survival, *M. smegmatis* mutants devoid of *namH* show increased sensitivity to β -lactams and lysozyme.⁹² In addition, *M. leprae* does not have a functional copy of *namH*^{92,93} in keeping with the lack of MurNGlyc in that species.

As noted above, the free carboxylic acid groups of the *meso*-DAP or D-isoglutamic acid residues of mycobacterial PG may be amidated, and some of the free carboxylic groups of the D-isoglutamic acid residues may also be modified by the addition of a glycine residue in peptide linkage.¹⁰ These modifications appear to occur at the lipid II level as lipid II, but not lipid I molecules, isolated from *M. smegmatis* showed various modifications including the occurrence of amidated D-Glu, *meso*-DAP, and terminal D-Ala.⁷⁹ The PG of *Mycobacterium* spp. also contains amidated *meso*-DAP and D-Glu,¹⁰ and a comparative analysis of the PG structure of *M. tuberculosis*, *M. leprae*, and *M. smegmatis*^{27,93} shows the presence of amidated D-Glu and *meso*-DAP in all three species. Therefore, the amidation of the PG appears to be widespread among members of *Mycobacterium* spp., including *M. leprae*. However, the responsible enzymes have yet to be identified.

6.13.6 Biosynthesis of Arabinogalactan

Biosynthesis of polysaccharides requires activated sugar donor molecules, most typically sugar nucleotides but occasionally polyprenylphosphoryl sugars. In both cases the leaving group is a substituted phosphate, which supplies the energy necessary for the glycosyl reaction to occur. For AG, the precursor molecules are UDP-GlcNAc, TDP-Rha, UDP-Galf, and decaprenylphosphoryl-D-arabinose (DPA), (Figure 4). The biosynthesis of UDP-GlcNAc is shown in Figure 5(a). This pathway differs from that of eukaryotes in that in prokaryotes the first step is to generate glucosamine-1-phosphate (GlcNH₂-1-P) from

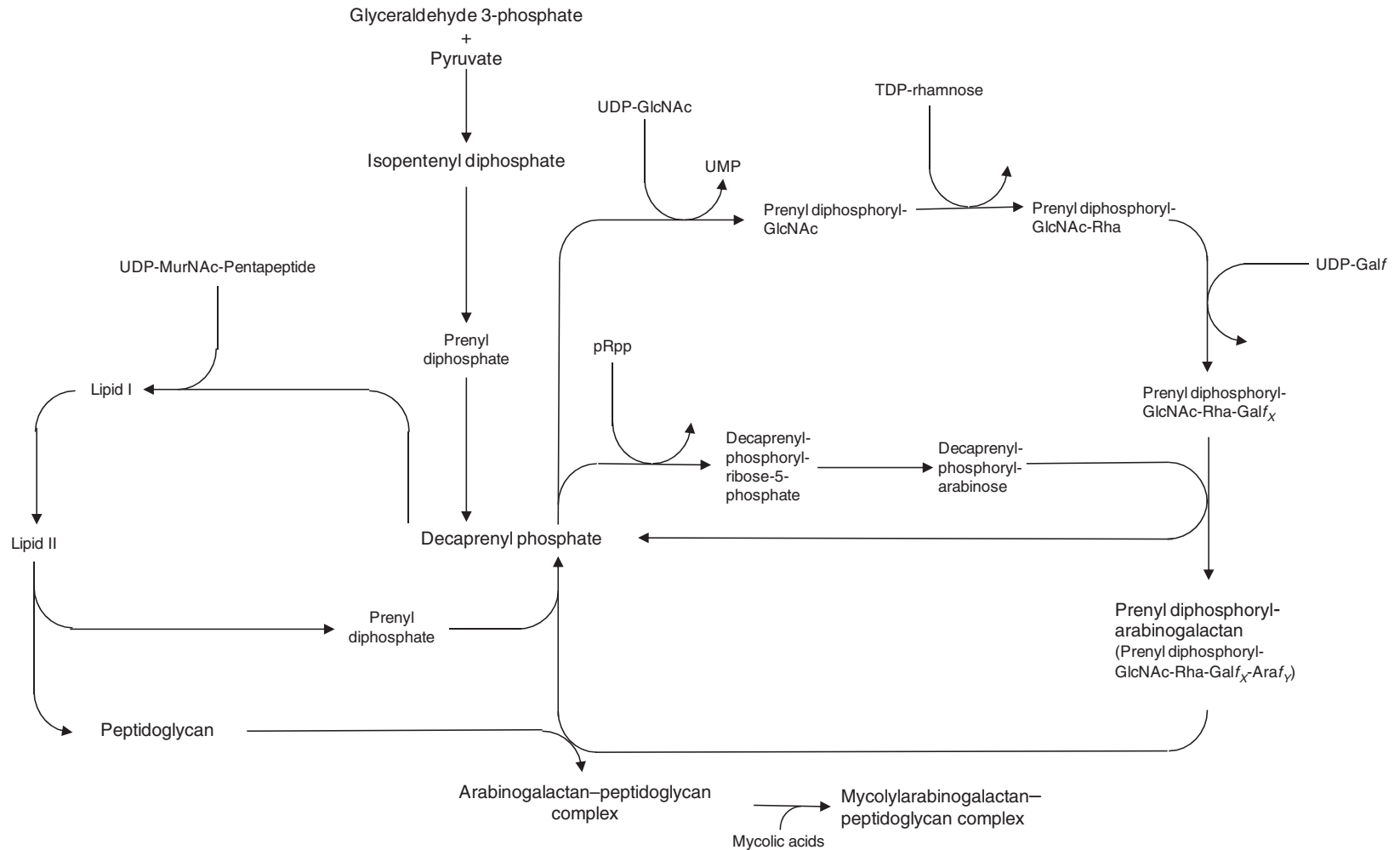


Figure 4 A schematic showing the overall biosynthesis of the MAPc in *Mycobacterium tuberculosis*. More detailed portions of the biosynthesis are presented in [Figures 5–7](#). The central role of DP is emphasized as an activated sugar donor for AG and PG syntheses.

glucosamine-6-phosphate (GlcNH₂-6-P). In contrast, in eukaryotes GlcNH₂-6-phosphate is N-acetylated and then the phosphate is moved. Also note that in *M. tuberculosis* (and other prokaryotes), the enzyme GlmU catalyzes both the N-acetylation reaction and the transferase reaction. The former is catalyzed at the C-terminal region of the protein, whereas the latter reaction takes place at the N-terminal region.^{94,95} GlmU has been shown to be essential for the growth of *M. smegmatis* and as its activity diminishes gross morphological changes in the bacteria occur as visualized by EM.⁹⁶

TDP-Rha is formed from TDP-glucose by four enzymes as shown in **Figure 5(c)**. All of these enzymes are essential for the growth of *M. tuberculosis* according to insertion mutagenesis.⁹⁷ In addition RmlB,⁹⁸ RmlC,⁹⁸ and RmlD⁹⁹ have been shown to be essential in *M. smegmatis* by knocking the genes out in the presence of a rescue plasmid with a temperature-sensitive origin of replication. Crystal structures have been obtained for RmlA–D in various species of bacteria^{100–108} including RmlC from *M. tuberculosis* in the presence and absence of the substrate analog TDP-Rha.¹⁰⁵

UDP-Galf is produced from UDP-Galp by the enzyme UDP-galactopyranose mutase (Glf) **Figure 5(b)**. This essential enzyme¹⁰⁹ uses FADH as a cofactor and evidence has been published that a covalent FADH₂-Gal intermediate is formed from the UDP-Gal substrate that undergoes pyranose/furanose equilibrium and is then reattached to UDP.¹¹⁰ Glf has been crystallized from several species including *M. tuberculosis*^{111,112} but not with substrate or substrate mimics in the active site.

Most polyprenyl phosphoryl sugars are biosynthesized from polyprenyl phosphate and a sugar nucleotide. However, extensive efforts by many laboratories to identify a sugar nucleotide of Ara in *M. tuberculosis* extracts were unsuccessful. After recognition that the Ara carbon skeleton arises from the pentose shunt,^{113,114} it was found that phosphoribose diphosphate (pRpp) is the sugar donor that reacts with DP.¹¹⁵ The enzyme catalyzing this reaction has been cloned and actively expressed using radioactive pRpp to demonstrate the formation of its product decaprenylphosphoryl-D-5-phosphoribose.¹¹⁶ The rest of the pathway has been deciphered (**Figure 5(d)**) and it was shown that the actual epimerization from ribofuranose to Ara^f took place at the lipid-linked level.¹¹⁷ That is, decaprenylphosphoryl-D-ribose (DPR) is epimerized to DPA and the potential intermediate phosphoarabinose diphosphate was not involved in DPA biosynthesis. Two enzymes are required for the epimerization (Rv3790 and Rv3791); both of these proteins are homologous to oxidoreductases. An intermediate tentatively characterized as decaprenylphosphoryl-2-keto ribose (decaprenylphosphoryl-2-keto-D-erythro-pentose) was identified.¹¹⁷ It is currently believed that Rv3790 is responsible for the oxidation of DPR and Rv3791 is responsible for the reduction of the resulting keto sugar to DPA.¹¹⁷

The biosynthesis of AG, using the sugar donors described above, is becoming reasonably well understood, although the fundamental gaps in our knowledge exist, particularly regarding the attachment of AG to PG and control of the arabinan polymerization. The biosynthesis of AG occurs on a carrier DP.¹¹⁸ The first reaction (**Figure 6**) is the transfer of GlcNAc-1-P to DP. The enzyme catalyzing this reaction has not been successfully expressed or characterized but is expected to be encoded by Rv1302 by analogy to a similar enzyme involved in LPS biosynthesis in *E. coli*. The activity of this enzyme generates GlcNAc-P-P-decaprenol. The next enzyme, a rhamnosyl transferase (WbbL), attaches an α -L-Rha residue to position 3 of the GlcNAc residue and is fairly well characterized. Its product has been identified by high-resolution mass spectrometry as Rha-GlcNAc-P-P-decaprenol.¹¹⁹ The next enzyme is the first of two Galf transferases, both of which appear to be bifunctional and attach Galf residues to fundamentally different acceptors. Thus, the protein expressed from Rv3782^{120,121} uses UDP-Galf to attach the first Galf residue to the 4-position of the rhamnosyl residue and presumably uses a different active site to attach the second Galf residue to the 5-position of the first Galf residue (see **Figure 6**). The second enzyme, GlfT (encoded by Rv3708c^{122,123}), uses the product of the Galf transferase encoded by Rv3782 to complete the galactofuran polymerization. This enzyme appears to have two activities, one of which attaches a Galf residue to the 5-position of a Galf residue that is itself attached to the 6-position of a Galf residue. The second activity attaches a Galf residue to the 6-position of a Galf residue that is itself attached to the 5-position of a Galf residue. How the final length of the galactofuran is controlled has not been determined.

The arabinan moiety of AG is an intriguing and complex molecule. The structure of the nonreducing 17 arabinosyl residues is tightly controlled and not heterogeneous. The interior α -1,5 arabinan is known to be as long as 14 arabinosyl residues; its homogeneity or heterogeneity is not known. The understanding of the biosynthesis of this polymer has been greatly aided by genetic knockouts and modifications in mycobacteria^{124–128} and related corynebacteria where AG is not essential for bacterial survival.^{40,129–132}

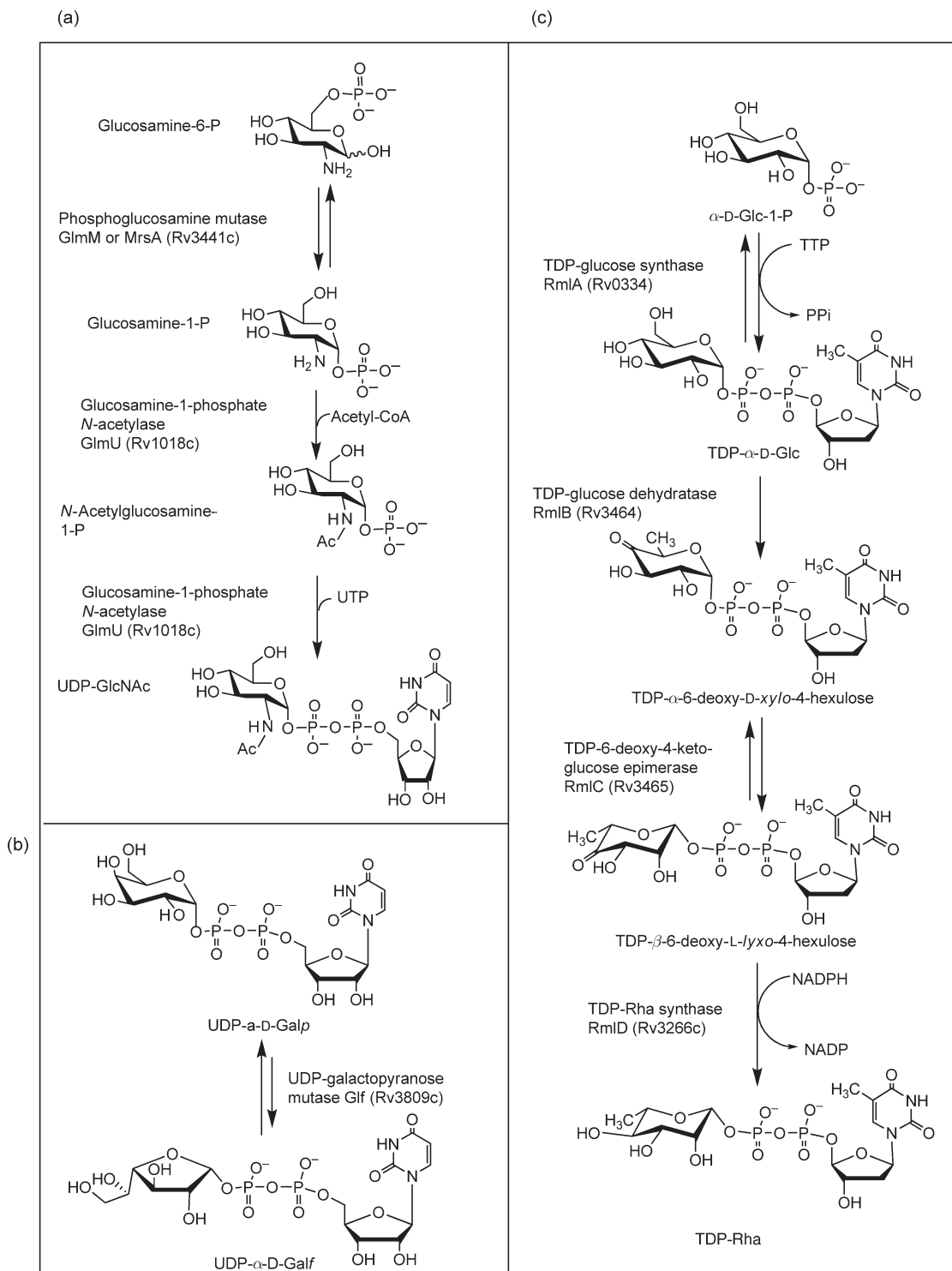


Figure 5 (Continued)

A significant issue in unraveling the nature of the *Araf* transferases responsible for mycobacterial arabinan biosynthesis is determining which enzymes are specific for AG arabinan, which are specific for LAM arabinan, and which might work on both arabinans given the structural similarities described above. In some cases, the

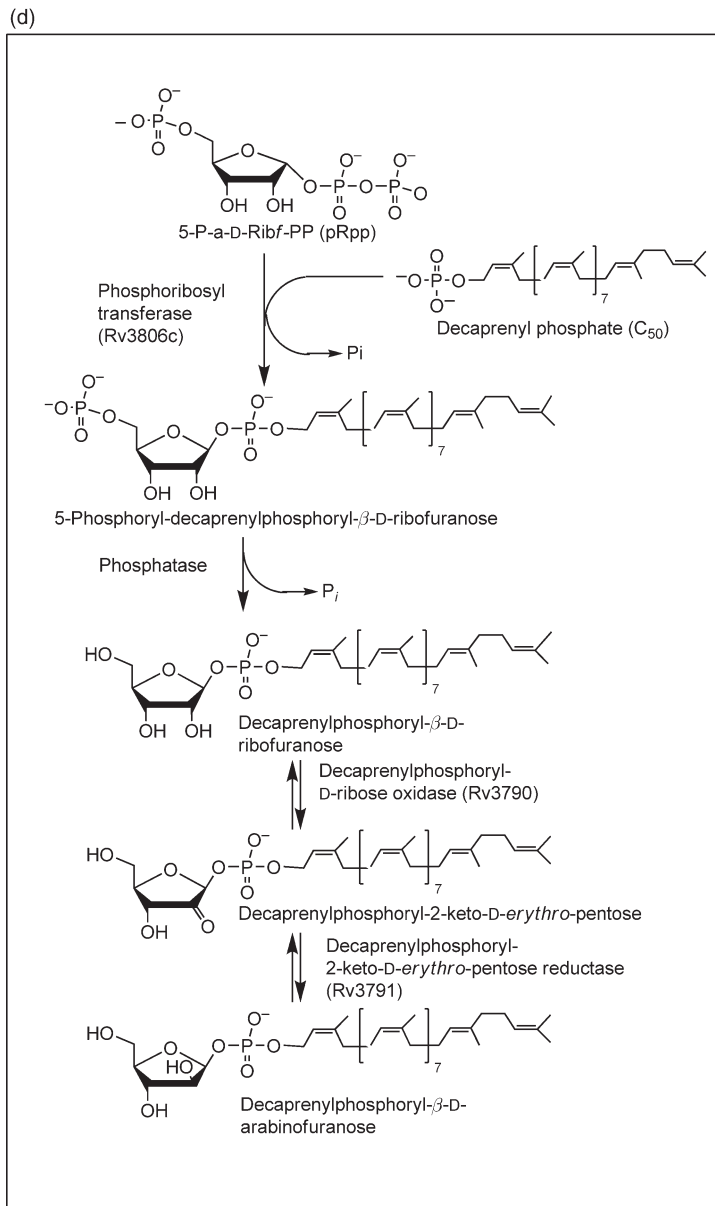


Figure 5 Biosynthesis of activated sugar donors in *Mycobacterium tuberculosis*. Panel A: synthesis of UDP-GlcNAc. Panel B: synthesis of UDP-Galf. Panel C: Synthesis of TDP-Rha. Panel D: synthesis of DPA. Gene names and Rv identifiers are indicated where known.

results are clear; in others more information is needed. Hypothetically speaking, at least seven arabinosyl-transferases should be present to conform to different linkages involved if one assumes that each enzyme performs a single function, an assumption that does not appear to be valid in all cases. A number of arabinosyltransferases have been identified to date, and there are clear indications that a few more could exist. A hypothesis concerning how arabinan synthesis could occur is presented below.

The first AftA is specific for AG as it adds three single arabinosyl residues to three positions on the galactan chain¹³⁰ (Figure 1). This enzyme has been shown to be essential for *M. smegmatis*¹³³ and is incapable of transferring an arabinosyl residue to a mannan acceptor.¹³³ The next Araf residues are attached by the Emb proteins. This group of proteins were first discovered as the target of ethambutol,¹³⁴ a first-line TB drug. The

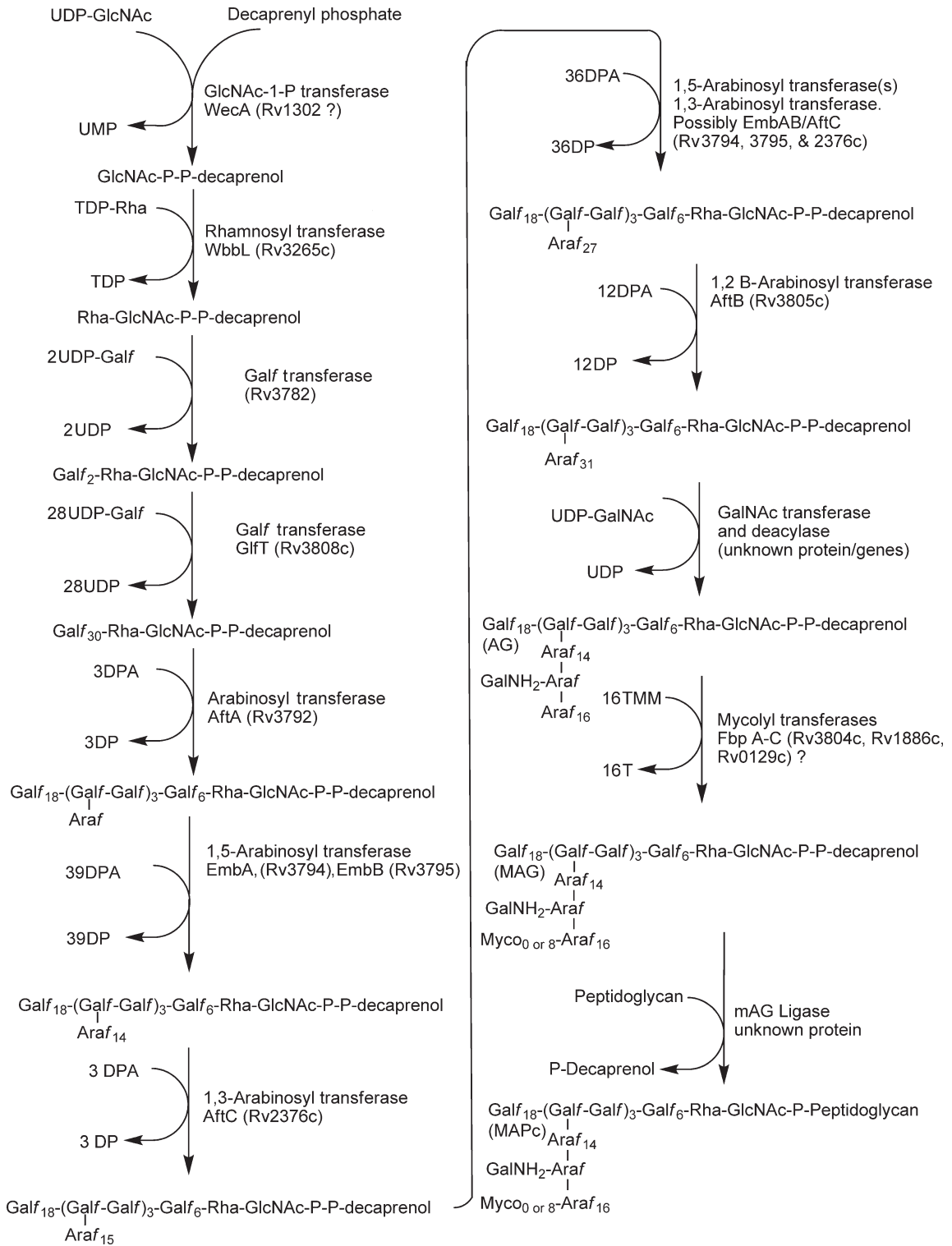


Figure 6 Detailed synthetic steps involved in AG synthesis. Gene names and Rv identifiers are indicated where known.

study of these proteins suffers from the fact that the Emb proteins are very large multitransmembrane proteins and have been refractory to expression and purification. Three very similar Emb proteins are found in mycobacteria and are designated as EmbA, EmbB, and EmbC. When *EmbC* is knocked out in *M. smegmatis*, LAM is no longer produced showing, definitively, that this protein is involved in LAM production (see the biosynthesis of LAM section).¹²⁸ When either *EmbA* or *EmbB* is knocked out in *M. smegmatis*, the structure of AG is changed¹²⁶ (discussed below) showing that both these proteins are involved in AG biosynthesis. In *C. glutamicum* only a single Emb protein is encoded for in the genome by *Cg-emb*.⁴⁰ This is consistent with the fact that *C. glutamicum* does not produce an arabinan attached to mannan (LAM), but rather, only arabinan attached to galactan (AG). When *Cg-emb* is knocked out, the corynebacteria AG contains only three nonreducing terminal arabinosyl residues attached to three Gal residues of the galactan,⁴⁰ which are the products of the activity of AftA.¹³⁰ Thus, genetic evidence strongly supports the hypothesis that Emb is an α -1,5 transferase that adds approximately 15 α -Araf residues to the single Araf 'primer.' This result correlates well with enzyme assays where the production of α -1,5 arabinosyl residues *in vitro* is inhibited by ethambutol.¹³⁵ An emerging picture consistent with these data is that the Emb proteins are present as multimers. These would be homomultimers in the case of the single Emb *C. glutamicum* protein (responsible for *C. glutamicum* AG arabinan) and in the case of mycobacterial EmbC (responsible for LAM arabinan synthesis). They could be heteromultimers for mycobacterial EmbA/EmbB (responsible for mycobacterial AG arabinan synthesis).

The next enzyme in arabinan biosynthesis is a branching enzyme. Recently, Besra and colleagues have identified Rv2673 as an α -1,3 Araf transferase (AftC).¹³⁵ Unlike most Araf transferase genes, this gene could be knocked out in *M. smegmatis*, and the resulting mutant contained AG that was mycolylated (although the amount of mycolic acids was severely reduced). Most interestingly, methylation analysis of the AG in the mutant revealed a complete lack of 3,5-linked Araf¹³⁵ leading to the possibility that this enzyme is responsible for generating both the internal and external branches. Results were confirmed with a linear α -1,5-Araf acceptor, which yielded extension at the nonreducing terminal with an α -Araf unit attached to the 3-position (yielding 3-linked Araf upon methylation analysis) with wild-type but not *aftC* knockout strains. Chatterjee *et al.* (unpublished) have found that rather than attachment at the nonreducing end, branching occurs at an internal 5-linked Araf residue to form a 3,5-linked Araf residue when using similar acceptors. The reason for the variation in results is not clear; the enzymes may behave differently *in vitro* under slightly different conditions and whether AftC attaches an Araf to the nonreducing end or to an internal Araf (or both) *in vivo* is not clear.

After the extension of the arabinan as α -1,5-linked residues by the EmbA/EmbB proteins and the addition of an α -Araf unit to the 3-position of 15th (approximately) Araf unit, the nonreducing terminal is synthesized. Thus, exactly three linear α -5-linked Araf units are attached to each arm of the branched α -3,5-linked Araf. Then the fourth Araf is branched with a β -Araf-(1 \rightarrow 2)- α -Araf disaccharide at positions 3 and 5. With the recognition of the 'capping' enzyme, AftB,¹³⁶ which attaches the terminal β -unit, arabinosyl transferases with all the required activities have been identified: α -1,5 transferase (EmbA/EmbB), α -1,3 transferase (AftC), and β -1,2 transferase (AftB). A reasonable hypothesis is that these are all of the enzymes that construct the entire AG arabinan including the nonreducing terminal. It is possible that this occurs through an enzyme complex of the three mentioned enzymes. The data supporting this contention is as follows: (1) An *embB* or an *embA* knockout in *M. smegmatis* results in alteration of the conversion of the nonreducing hexasaccharide unit into a tetrasaccharide, which does not have the β -Araf-(1 \rightarrow 2)- α -Araf disaccharide at position 3 of what, in complete AG, is a branched α -3,5-linked Araf residue;¹²⁶ yet as described above there is strong evidence that EmbA/EmbB are an α -1,5 arabinosyl transferase. *In vitro* evidence confirms the requirement of both EmbB and EmbA be present in order to add the β -Araf-(1 \rightarrow 2)- α -Araf disaccharide to a synthetic acceptor.¹²⁷ (2) Alteration of the C-terminal region of EmbC (the active site is in the N-terminal) affects the size of the arabinans on LAM⁶⁵ as might be expected if the C-terminal is involved in binding other proteins. (3) An *aftC* knockout does not extend the α -1,5 arabinan to its normal length but only adds three α -1,5-Araf residues before capping takes place with the β -Araf residue.¹³⁵ This suggests that absence of AftC affects the α -1,5 transferase activity of EmbA/EmbB. (4) Homologs of AftC and AftB have not been identified. (5) EmbC is specific to LAM biosynthesis and EmbA/EmbB to AG.¹²⁸ Thus, the entire AG arabinan would be synthesized by the EmbA/EmbB, AftC, and AftB complex, in unknown stoichiometric ratios and likely with other 'helper' proteins. By analogy, an EmbC complex with AftB and AftC could synthesize LAM. Essentially, nothing is known about the succinylation and galactosaminylation of the arabinan biosynthetically.

It is generally agreed that trehalose mono- and/or dimycolate is the donor of mycolic acids to the nonreducing end of the arabinan, and that these reactions are catalyzed by the proteins known by various names including Antigen 85 and fibronectin-binding protein. The best evidence of this is genetic knockouts in *M. tuberculosis*¹³⁷ and *C. glutamicum*,^{138,139} where mycolylation is strongly decreased when these proteins are knocked out. In both species there are multiple genes encoding related mycolyl transferases so a complete loss of mycolic acids in the wall does not occur. Biochemically, Antigen 85 has been shown to catalyze the transfer of mycolic acids from TMM to trehalose,¹⁴⁰ but the transfer from trehalose mono- or dimycolate to cell wall has yet to be shown biochemically.

6.13.7 Biosynthesis of the Phosphatidylinositol Containing Phosphatidylinositol Mannosides, Lipomannans, and Lipoarabinomannans

The biosynthesis of PI in mycobacteria appears to be similar to that of the eukaryotic cells.¹⁴¹ The *de novo* synthesis of Ino typically involves cyclization of glucose-6-PO₄, a reaction catalyzed by Ino-1-phosphate synthase (INO1, Rv0046c).¹⁴² Ino-1-phosphate is then dephosphorylated by an Ino monophosphate phosphatase.¹⁴³ PI is then formed by the transfer of diacylglycerol phosphate to Ino from CDP-diacylglycerol.¹⁴⁴ The *M. tuberculosis* gene *pgsA* (*Rv2612c*) is reported to encode a PI synthase (PIS), as overexpression in *M. smegmatis* resulted in elevated PI synthesis.¹⁴⁵ The PIS in mycobacteria is thought to be essential, as *pgsA* in *M. smegmatis* could only be deleted in the presence of an episomal copy of PIS and was predicted to be essential in *M. tuberculosis*.⁹⁷ Overall, the current data clearly indicates that the synthesis of PI in mycobacteria is very similar to or, perhaps, identical to that of eukaryotes.

Subsequently, PI is mannosylated to generate PIMs, LM, and LAM,^{48,146} a generalized and hypothetical biosynthetic pathway shown in **Figure 7**. It was originally demonstrated that incubation cell-free

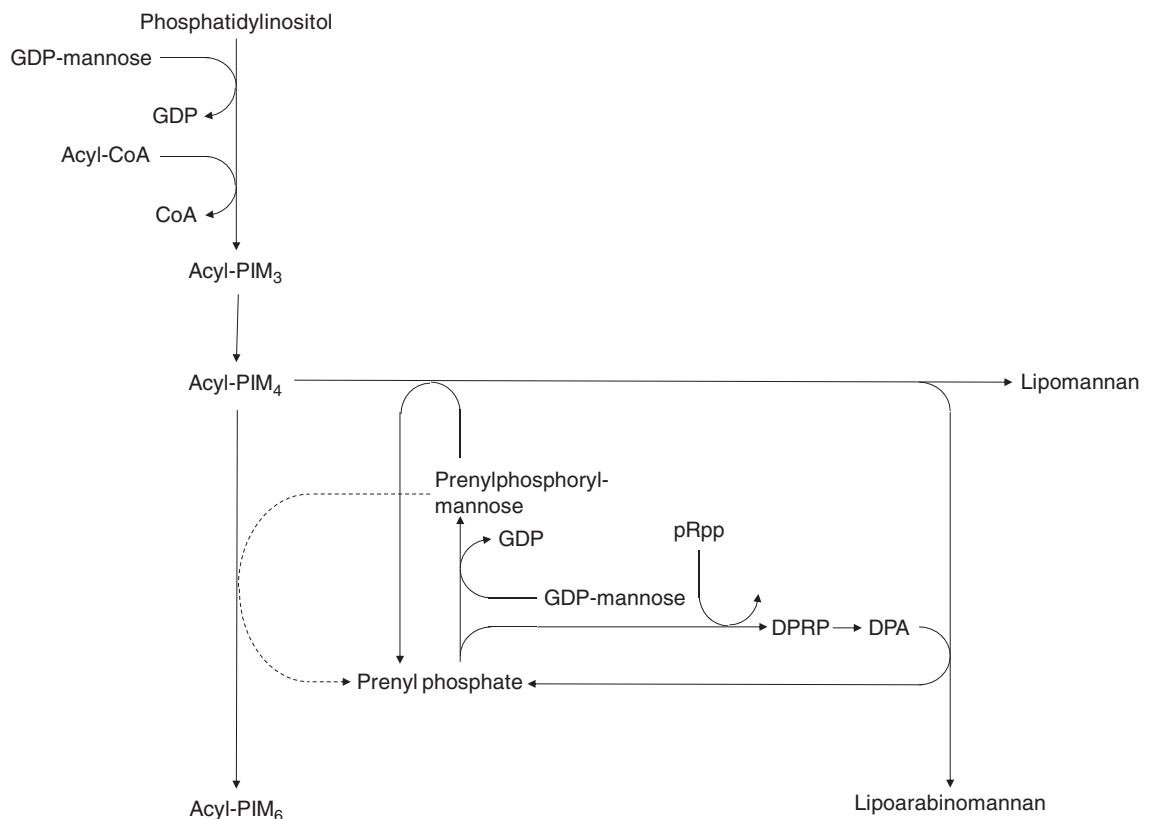


Figure 7 A hypothetical biosynthetic scheme showing the putative relationships between PIMs, LM, and LAM.

preparations from *M. pblei* and *M. tuberculosis* with tritiated GDP-Man resulted in the synthesis of two labeled PIM₂ species (PI substituted with two Man residues as explained previously), differing in their degree of acylation. Early evidence indicated that the PI molecules of mycobacteria are acylated through a system that required coenzyme A and ATP,⁶⁰ but it was 35 years before an enzyme (Rv2611c) responsible for the conversion of PIM₁ into acyl-PIM₁ was described.¹⁴⁷ A GDP-Man-dependent mannosyltransferase (*pimA*, Rv2610c) in *M. tuberculosis*¹⁴⁸ is responsible for the addition of the first Man residue to the hydroxyl group at the 2-position of the Ino of PI to form PIM₁ thus catalyzing the first committed step in PIM biosynthesis. PimA is essential for growth of *M. smegmatis*, as inactivation of *pimA* results in cell death and is, presumably essential in *M. tuberculosis* although it was not predicted to be so by high-density transposon mutagenesis.⁹⁷ It is thought that PI and PIM synthesis may be coregulated in mycobacteria as *pimA* and *pgsA* are colocalized on the chromosome.

A putative mannosyltransferase, encoded by *Rv0557* in *M. tuberculosis*, with significant primary sequence homology to PimA, was thought to be responsible for the formation of acyl-PIM₂ and was named PimB,¹⁴⁹ as the protein, in *in vitro* assays, appeared to transfer Man from GDP-Man to acyl-PIM₁.¹⁴⁹ This result was puzzling as *M. leprae*, which also produces LAM and PIMs, lacks an ortholog of Rv0557.¹⁵⁰ More recently, Rv2188c in *M. tuberculosis* has been shown to encode the mannosyltransferase involved in the conversion of acyl-PIM₁ into acyl-PIM₂ and this step is essential for the synthesis of LAM.¹⁵¹ Disruption of the homologous gene (NCgl2106) in *C. glutamicum* did not prevent the synthesis of LM, indicating that the LAM and LM in this strain may have a different glycolipid anchor or that LM and LAM are synthesized from different pools of PIM₁. A third mannosyltransferase, encoded by *pimC*, has been identified in *M. tuberculosis* CDC 1551.¹⁵² Overexpression of this gene in *M. smegmatis* led to the formation of acyl-PIM₃ in an *in vitro* assay. However, the *M. tuberculosis* H37Rv genome lacks a homolog of PimC suggesting that synthesis of acyl-PIM₃ may be carried out by a second protein in this strain.¹⁵² An enzyme designated as PimE (Rv1159) was recently identified as a probable polyrenylphosphorylmannose (Pol-P-Man)-dependent mannosyltransferase responsible for the synthesis of PIM₅ from PIM₄.¹⁵³ However, this enzyme transfers a Man residue to the 2-position of an existing Man residue and thus is not considered to play a role in LM or LAM synthesis.

None of these mannosyltransferases contain predicted transmembrane domains, despite the fact that the substrates are likely associated with the plasma membrane. Interestingly, many of the glycosyltransferases involved in *E. coli* LPS O-antigen and core assembly also lack transmembrane domains but are detected in the membrane reactions.¹⁵⁴ This is in contrast to the findings in the eukaryotic system, where most of the glycosyltransferases involved in the glycolipid synthesis are integral membrane proteins.^{155,156}

In a cell-free assay, it was demonstrated that acyl-PIM₂ is specifically extended by the addition of Man_p residues from Pol-P-Man to form higher PIMs (at least acylated PIM₃ and PIM₄)¹⁵² and linear LM possessing an (α1 → 6)-linked Man_p backbone.⁵⁹ The Man_p donors (Pol-P-Man) are synthesized from GDP-Man and polyprenyl phosphate by polyprenylphosphorylmannose synthase (*ppm1*, Rv2051c).¹⁵⁷ It is proposed that PIM₆ (derived from PIM₅, the product of PimE) is a terminal product, as it is abundant and contains two (α1 → 2)-linked Man_p residues, which are not found in LM or in the LAM mannan core (Figure 7). Thus, PIM₄ appears to be the branch point at which PIM intermediates diverge to form LM and LAM. The linear (α1 → 6)-linked Man_p backbone LM is thought to undergo further mannosylation, resulting in LM with mature branched mannan; this mannan is subsequently arabinosylated to form LAM. Definition of the biosynthetic steps leading to the formation of the higher PIMs (LM and LAM precursors), LM, and LAM is being unraveled.^{62,63,158} Early studies have suggested that the synthesis of the mannan chain of LAM involved Pol-P-Man donor species.^{159,160} More recent studies indicate that the synthesis of PIM₃ and higher PIMs all utilize Pol-P-Man as a donor substrate.¹⁶¹ Bioinformatics approaches identified putative integral membrane proteins, Rv2181 in *M. tuberculosis* and MSMEG4250 in *M. smegmatis*, with predicted 10 transmembrane domains and a GT motif, features that are common to eukaryotic mannosyltransferase of the GT-C superfamily that rely on polyprenyl-linked, rather than nucleotide-linked, sugar donors. Inactivation of MSMEG4250 led to generation of an LM-depleted strain with the accumulation of modified LAM with a mannan core that was unbranched.⁶² These data and detailed structural analyses of the products formed suggested that Rv2181 and MSMEG4250 are responsible for the addition of α(1 → 2) Man_p branches to the mannan core of LM/LAM.⁶² More recently, structural analyses of the LM and LAM variants produced by an *M. tuberculosis* Rv2181 knockout mutant revealed complete absence of α(1 → 2)-linked Man_p branching on the mannan backbones of LM and LAM.¹⁵⁸ In addition, the nonreducing end of LAM was affected in that only single Man_p residue on the nonreducing arabinan termini of

LAM was found. Coexpression of Rv2181 and Rv1635c (the latter responsible for the addition of the Man residue linked to the arabinan of Man-capped LAM¹⁶²) in *M. smegmatis*, resulted in synthesis of LAM with the arabinan termini capped with $\alpha(1 \rightarrow 2)$ Man_p di- and tri-Man_p, confirming Rv2181 in the dual role of Man capping and mannan-core branching.

As noted above, LAM arabinan is elaborate and less structured than AG arabinan, and has extended linear β -D-Araf-(1 \rightarrow 2)- α -D-Araf-(1 \rightarrow 5)- α -D-Araf-(1 \rightarrow 5)- α -D-Araf-(Ara₄) chains that are not found in AG. These nonreducing ends are structurally well defined; however, the assembly process of the polymeric arabinan on the LM core is essentially unknown and is, perhaps related to AG arabinan synthesis as described above. Thus, the *embA* and *embB* gene products are responsible for the formation of the 3-linked branch of the Ara₆ motif, whereas the *embC* is involved in the formation of the arabinan domain of LAM. The emerging biochemical data also indicates that despite overall structural similarity, the arabinans of LAM and AG are distinguished by virtue of the additional presence of extended linear termini in LAM, the synthesis of which entails yet some unknown feature of EmbC protein for the proper synthesis.⁶⁵ Beyond this, very little is known about LAM arabinan biosynthesis.

6.13.8 Mycobacterial Cell Envelope Ultrastructure

As noted in the introduction, the envelope of mycobacteria consists of a plasma membrane, a cell wall, and a capsule-like outermost layer. Although considerable information regarding the chemical composition and immunological properties of structures found within the cell envelope of mycobacteria has been reported, relatively little is known about the manner in which these components are spatially arranged. Until quite recently, EM using freeze-substitution methods provided the best data. These studies indicated that there are at least four layers that have significantly different properties, including an inner plasma membrane, an electron-dense layer (EDL), an electron-transparent zone (ETZ), and an outer, electron-dense layer (OL). The plasma membrane is a typical bilayer and is approximately 4–4.5 nm thick. The EDL is located outside the plasma membrane and likely contains the PG layer.¹⁶³ Importantly, results indicated that the EDL was closely apposed to the plasma membrane (Figure 8), providing space that could be imagined to contain the AG (assuming that AG would be less electron dense than PG after fixing). The ETZ is 9–10 nm thick¹⁶⁴ and appears to be hydrophobic due to the inability of water-soluble dyes to penetrate the region. Therefore, it is assumed that the ETZ is primarily composed of mycolic acids. The OL seems to vary in thickness, electron density, appearance among species, growth conditions, and preparation methods for microscopy.^{3,165,166} The OL from *M. tuberculosis* consists of lipids, protein, glucans, mannans, AMs, and other molecules,¹⁶⁷ and it has been postulated that LAM and LM are also associated with this layer^{168,169} and the plasma membrane.^{47,170} In addition, freeze-fracture experiments show that the mycobacteria cell envelope has two planes of weakness one associated with the plasma membrane, as expected, and another associated with the outer part of the envelope¹⁷¹ although EM using freeze-substitution methods did not indicate the presence of an outer membrane (OM).

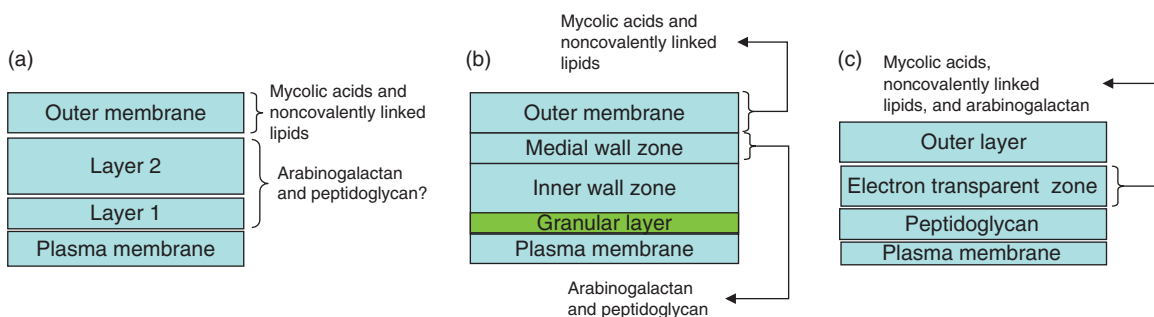


Figure 8 Models of the mycobacterial cell wall ultrastructure as determined by cryoelectron tomography, Panels a¹⁷² and b,¹⁷³ and freeze-substitution, Panel c.¹⁶³ The models presented here are reproductions based on data from the original manuscripts and attempt to provide a perspective of the relative scales. The putative location of the PG and AG is indicated.

More recently cryoelectron tomography of frozen-hydrated sections has been utilized to visualize the mycobacterial cell envelope and a structure analogous to a Gram-negative OM in particular.^{172,173} It should be recalled that cryoelectron tomography of frozen-hydrated sections requires the use of 15% sucrose or 20% dextran; the latter is used in both papers described below as a cryoprotectant and for vitrification of the sample. Since contrast is proportional to density in frozen-hydrated samples,¹⁷⁴ and the cryoprotectant provides an external mass density that approaches the density of the O-side chains of LPS making them invisible¹⁷⁵ it seems likely that visualization of the AG, LM, and LAM will be difficult using this technique.

In one set of data, four layers are reported.¹⁷² These are the plasma membrane, two layers designated as L1 and L2, and an OM, which shows features very similar to the plasma membrane. The layers designated as L1 and L2 could not be assigned due to structural appearances alone but were hypothesized to be related to the MAPc network. In the model proposed, L1 is closely apposed to the plasma membrane and could be analogous to the EDL identified by freeze substitution. In the second set of data, five structures were reported.¹⁷³ These also included a plasma membrane and an OM showing features remarkably similar to the plasma membrane. In this case, three other structures were reported. These included a granular layer (GL), which is closely associated with the plasma membrane and was considered to lie within the inner wall zone (IWZ) and a medial wall zone (MWZ), which is closely associated with the OM. These researchers hypothesized that the IWZ occupies the position shown to be a periplasmic space in other Gram-positive bacilli; the MWZ was formed, at least in part, by PG; and the OM is made up of molecules specific to mycobacteria.¹⁷³ Overall, all three models of the mycobacterial cell envelope are quite similar, with the major difference being the location of the PG (Figure 8). In two of the models, the PG is thought to be close to the plasma membrane, providing significant space for packing of AG. In the third model, the PG is thought to be closely apposed to the OM, providing minimal space for the AG.

Thus, it remains difficult to visualize exactly how the plasma membrane, OM, PG, and AG are arranged. The imaging techniques available for such a structure continue to improve but fall short of the type of data available from crystallography structures. What the latest EM studies show most clearly is a distance of 30–35 nmol l⁻¹ between the centers of the two membranes depending on the species.^{172,173} Between these two membranes lie the PG and AG, with the AG covalently linked to both the PG and the inner mycolate layer of the OM. The structures of the two polymers yield some insights. First, the ratio of the mass of AG to PG is about 1.3; that is, the mass of AG exceeds the mass of PG. Second, the structure of AG would allow the nonreducing terminal Ara_f units to be between 30 and 35 nmol l⁻¹ from the PG molecule. Thus, AG could, in theory, span the entire distance between the two membranes. It is more difficult to estimate the thickness of compacted AG since oligosaccharides rarely have a single rigid structure,¹⁷⁶ furanosides are even more flexible than pyranosides,¹⁷⁷ and the AG structure does not lend itself to forming defined structures such as rods or fibers. If one compares the thickness of a single PG layer at 2.5 nmol l⁻¹ (Vollmer *et al.*¹⁷⁸) and assumes a similar packing for AG (without any data), then one can estimate 3.4 nmol l⁻¹ (1.35 × 2.5 nmol l⁻¹) as a minimum thickness. The size of the PG + AG layer in one of the recent EM studies¹⁷³ is 6.3–7.3 nm not much more than the minimum size predicted by PG (2.5 nmol l⁻¹) + AG compacted (3.4 nmol l⁻¹) or 5.9 nmol l⁻¹. This hypothetical calculation also assumes the presence of a single layer of PG that may be reasonable. The fact that the Gal and Ara residues are all in the furanose form and connected primarily through 1,5 and 1,6 linkages suggests that the polysaccharides may be optimized for flexibility in order to allow packing of the mycolic acids.²⁹ The size and flexibility of the arabinan chain could suggest that, within limits, it can accommodate a variation in the size of the PG layer and the OM, perhaps providing an explanation for the reported variable positioning of PG in EM images.^{163,172,173}

A model of the mycobacterial cell envelope proposed in 1982 was based primarily on the chemical structure of the mycolic acids.¹⁷⁹ In this model, the plasma membrane is surrounded by PG, which in turn is surrounded by AG. The mycolic acids are packed in a monolayer, parallel to each other, and oriented perpendicularly to the plasma membrane. It was further hypothesized that noncovalently linked lipids would intercalate into the outer portion of the mycolic acids layer due to the asymmetry of the two arms of the mycolate. Subsequently, a second model was proposed¹⁶⁹ that was similar, except that the noncovalently linked lipids were proposed to form a monolayer that did not intercalate with the mycolic acids, providing a model that more nearly approximated a cell envelope with two bilayers. Subsequently, experiments showed that the mycolic acids were, in fact, aligned perpendicularly to the cell surface and formed tight crystalline arrays,⁸ providing support for the basis of both

hypotheses. Until recently, there was no indication in electron micrographs that a second lipid bilayer existed in the mycobacterial cell envelope, and there were no conclusive data pointing to which model was correct; either model could explain the existence of two freeze-fracture planes in the cell envelope and both models predict the existence of an asymmetrical bilayer that would be significantly thicker than the plasma membrane. Recent cryoelectron tomography studies identified structures that are consistent with the existence of an OM in mycobacteria;^{172,173} in both cases the structure identified as an outer bilayer appears very similar to the plasma membrane. That is, in both cases the OM appears as a symmetrical bilayer that is only slightly thicker than the plasma membrane. These observations have resulted in three new models of the mycobacterial OM. (1) The meromycolate chains of the mycolic acids span the entire hydrophobic region and are covered by fatty acids from extractable lipids on the outer leaflet. Extractable lipids also exist in the inner leaflet of the OM.¹⁷² (2) The mycolic acid layer only contributes to the inner leaflet, which contains extractable lipids, through the extended branches of the meromycolate chains. In this model, the major portion of the mycolic acid (including the entire α -chain) is located below the OM proper, somewhat like a third leaflet.¹⁷² (3) The inner leaflet of the OM consists of mycolic acids where the meromycolate chains are folded upon themselves to create a compact structure and the extractable lipids facing the AG-bound mycolates are intercalated between the mycolic acid chains.¹⁷³ Thus, both the Minnikin and Rastogi models appear not to fit the data as they both predict a thick, asymmetrical OM and each of the three new models has aspects that are not satisfying. Models 1 and 2 suggest that the inner leaflet of the OM contains extractable lipids in addition to mycolic acids, a hypothesis that is difficult to reconcile with the data suggesting that the mycolic acids form tight crystalline arrays⁸ without proposing mycolic acid islands. This is particularly true of model 1, and model 2 would suggest an additional electron-transparent layer 'below' the OM. Model 3 is based on recent analysis of mycolic acid behavior in Langmuir monolayers and molecular dynamics modeling^{180–182} and EM studies.¹⁷³ The folding of the mycolic acids upon themselves clearly explains the thickness of the OM; however, as yet, there are little data to indicate that mycolic acids at a water/air interface behave in a manner analogous to that seen in the mycobacterial OM.

Figure 9 is a schematic diagram that attempts to utilize the recent cryoelectron tomography studies, which is similar to model 3 above, and tries to include the placement of the AG. It is as consistent with both the ultrastructure and current chemical structure data as possible but, like the three models discussed above, does not fit perfectly with all of the data. The mycolates are depicted in a folded conformation, as a meromycolate chain drawn in extended conformation is longer than the thickness of the entire observed OM; therefore, this fact needs to be accommodated by any model proposed. In this figure, the PG and AG layers are drawn closely apposed to the OM to correspond to the MWZ seen in some cryotomography electron micrographs.¹⁷³ This choice was somewhat arbitrary and was made as it requires that the AG be found in a compacted configuration. The glycan chains of the AG and LAM are drawn to facilitate identification of the various structural features. However, given the inherent flexibility of oligosaccharides in general and oligofuranosides in particular, this portrayal is almost certainly misleading. In some previous reviews, models of the cell envelope were drawn in which LAM was depicted as extending through the mycolic acid layer. The observations that subcellular fractionation experiments indicated that LAM was associated with the plasma membrane^{47,170} and that monoclonal antibodies raised against LAM recognize intact mycobacteria¹⁸³ gave rise to this depiction as the latter observation indicates that the epitope is accessible to the external milieu. If the saccharide residues are modeled in an extended conformation, it seems possible that the molecule could extend from the plasma membrane to the outside of the cell envelope. However, this hypothesis is unlikely given the flexibility of the structure and the permeability barrier presented by the mycolic acid layer.¹⁸⁴ Thus, in the current model, LAM is depicted anchored in both the plasma membrane with the carbohydrate chains between the plasma membrane and the OM, where it is likely synthesized, and anchored in the OM in a position suggested recently¹⁶⁸ and in the older literature.¹⁶⁹ It seems probable that the LAM anchored in the plasma membrane is not exposed to the surface of the cell and that the reactivity of the whole bacterium to monoclonal LAM antibodies is generated by molecules anchored in the mycolic acid layer and/or by structurally similar AM molecules⁷⁵ located outside of the OM. However, the tentative nature of this model needs to be stressed. A recent paper reported studies using atomic force microscopy, chemical force microscopy, and immunogold detection, which indicated that the surface of *M. bovis* BCG cells is very hydrophobic, a result consistent with the presence of an external layer of mycolic acids but not with a surface of exposed polar head groups of associated lipids intercalated with mycolic acids.¹⁸⁵ Evidence was also presented indicating that LAM is not abundant (if present

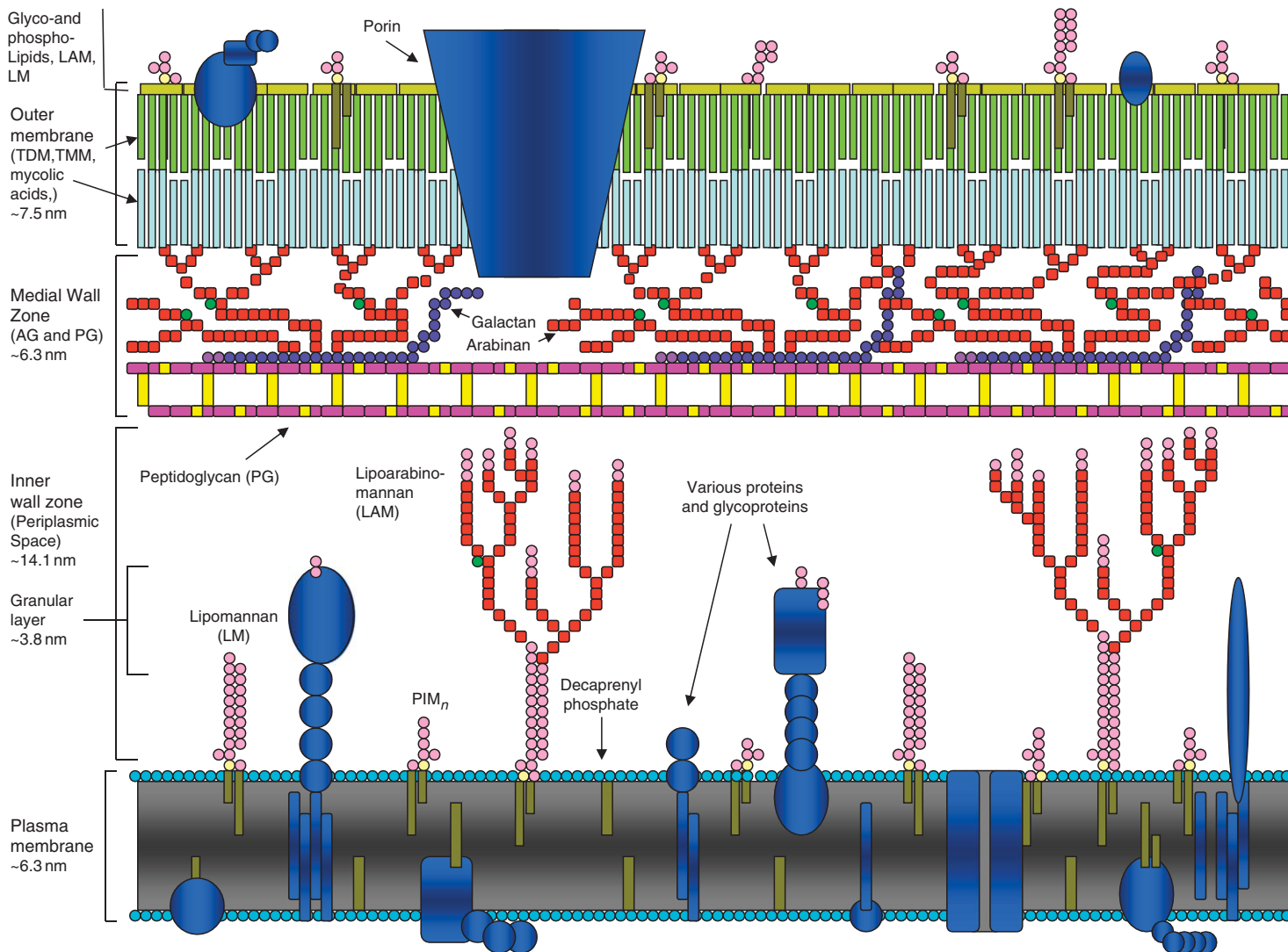


Figure 9 A schematic depiction of the overall arrangement of the cell-envelope of mycobacteria. The focus of this particular model is to arrange the plasma membrane, PG, AG, and OM in a fashion consistent with published data. The representations of the various macromolecules are schematic, for example, mycolic acids are not shown attached to the α -2-Araf residues (which they are in reality) and LAM is shown extended for visual purposes. In addition, the mycolic acids are shown in a nonextended conformation to be consistent with the observed thickness of the OM layer (see text for an explanation).

at all) on the surface of the bacilli and unlikely to be anchored in the OM. Thus, this view of the cell surface is profoundly different than that presented in **Figure 9**.

In conclusion, it can be seen that the primary structure of the unusual oligosaccharides of the cell envelope of mycobacteria is reasonably well understood, although a number of questions need to be resolved. The biosynthesis of these polymers is much less well defined but current studies are rapidly closing the gap. Oddly, it is the definition of the ultrastructure of the mycobacterial cell envelope and reconciling the data with the known chemistry that seems to present the greatest challenge in understanding the mycobacterial cell envelope.

Acknowledgment

Research in the author's laboratories is funded by NIH grants AI-37139, AI-57836, AI-49151, and AI-65357.

References

1. A. M. Ginsberg; M. Spigelman, *Nat. Med.* **2007**, *13*, 290–294.
2. Anonymous, *Nat. Med.* **2007**, *13*, 263.
3. P. J. Brennan; H. Nikaido, *Annu. Rev. Biochem.* **1995**, *64*, 29–63.
4. V. Jarlier; H. Nikaido, *FEMS Microbiol. Lett.* **1994**, *123*, 11–18.
5. H. Nikaido; V. Jarlier, *Res. Microbiol.* **1991**, *142*, 437–443.
6. E. Lederer; A. Adam; R. Ciorbaru; J. F. Petit; J. Wietzerbin, *Mol. Cell Biochem.* **1975**, *7*, 87–104.
7. M. McNeil; S. J. Wallner; S. W. Hunter; P. J. Brennan, *Carbohydr. Res.* **1987**, *166*, 299–308.
8. H. Nikaido; S. H. Kim; E. Y. Rosenberg, *Mol. Microbiol.* **1993**, *8*, 1025–1030.
9. A. Adam; J. F. Petit; J. Wietzerbin-Falszpan; P. Sinay; D. W. Thomas; E. Lederer, *FEBS Lett.* **1969**, *4*, 87–92.
10. S. Kotani; I. Yanagida; K. Kato; T. Matsuda, *Biken. J.* **1970**, *13*, 249–275.
11. E. Lederer, *Pure Appl. Chem.* **1971**, *25*, 135–165.
12. M. Matsuhashi, *Tanpakushitsu Kakusan Koso* **1966**, *11*, 875–886.
13. J. F. Petit; A. Adam; J. Wietzerbin-Falszpan; E. Lederer; J. M. Ghuyesen, *Biochem. Biophys. Res. Commun.* **1969**, *35*, 478–485.
14. J. F. Petit; E. Lederer, The Structure of the Mycobacterial Cell Wall. In *The Mycobacteria, a Source Book*; G. P. Kubica, L. G. Wayne, Eds.; Marcel Dekker: New York, 1984; pp 301–322.
15. J. Wietzerbin; B. C. Das; J. F. Petit; E. Lederer; M. Leyh-Bouille; J. M. Ghuyesen, *Biochemistry* **1974**, *13*, 3471–3476.
16. J. Wietzerbin-Falszpan; B. C. Das; I. Azuma; A. Adam; J. F. Petit; E. Lederer, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 57–63.
17. K. H. Schleifer; O. Kandler, *Bacteriol. Rev.* **1972**, *36*, 407–477.
18. J. M. Ghuyesen, *Bacteriol. Rev.* **1968**, *32*, 425–464.
19. W. Vollmer; J. V. Holtje, *J. Bacteriol.* **2004**, *186*, 5978–5987.
20. M. Leyh-Bouille; R. Bonaly; J. M. Ghuyesen; R. Tinelli; D. Tipper, *Biochemistry* **1970**, *9*, 2944–2952.
21. C. Goffin; J. M. Ghuyesen, *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 702–738.
22. M. F. Templin; A. Ursinus; J. V. Holtje, *EMBO J.* **1999**, *18*, 4108–4117.
23. P. Draper; O. Kandler; A. Darbre, *J. Gen. Microbiol.* **1987**, *133*, 1187–1194.
24. J. F. Petit; A. Adam; J. Wietzerbin-Falszpan, *FEBS Lett.* **1970**, *6*, 55–57.
25. K. Takayama; H. L. David; L. Wang; D. S. Goldman, *Biochem. Biophys. Res. Commun.* **1970**, *39*, 7–12.
26. D. C. Crick; S. Mahapatra; P. J. Brennan, *Glycobiology* **2001**, *11*, 107R–118R.
27. S. Mahapatra; H. Scherman; P. J. Brennan; D. C. Crick, *J. Bacteriol.* **2005**, *187*, 2341–2347.
28. M. McNeil; M. Daffe; P. J. Brennan, *J. Biol. Chem.* **1991**, *266*, 13217–13223.
29. A. R. Baulard; G. S. Besra; P. J. Brennan, The Cell-Wall Core of *Mycobacterium*: Structure, Biogenesis and Genetics. In *Mycobacteria: Molecular Biology and Virulence*; C. Ratledge, J. Dale, Eds.; Blackwell Science Ltd: London, 1999; pp 240–259.
30. G. S. Besra; P. J. Brennan, *Biochem. Soc. Trans.* **1997**, *25*, 845–850.
31. L. G. Dover; A. M. Cerdano-Tarraga; M. J. Pallen; J. Parkhill; G. S. Besra, *FEMS Microbiol. Rev.* **2004**, *28*, 225–250.
32. R. E. Lee; P. J. Brennan; G. S. Besra, *Curr. Top. Microbiol. Immunol.* **1996**, *215*, 1–27.
33. S. Mahapatra; J. Basu; P. J. Brennan; D. C. Crick, Structure, Biosynthesis and Genetics of the Mycolic Acid-Arabinogalactan-Peptidoglycan Complex. In *Tuberculosis and the Tubercle Bacillus*; S. T. Cole, K. D. Eisenach, D. N. McMurray, W. R. Jacobs, Eds.; ASM Press: Washington, DC, 2005; pp 275–285.
34. M. McNeil; G. S. Besra; P. J. Brennan, Chemistry of the Mycobacterial Cell Wall. In *Tuberculosis*; W. N. Rom, S. M. Garay, Eds.; Little, Brown and Company: Boston, 1996; pp 171–185.
35. R. P. Tripathi; N. Tewari; N. Dwivedi; V. K. Tiwari, *Med. Res. Rev.* **2005**, *25*, 93–131.
36. P. Draper; K. H. Khoo; D. Chatterjee; A. Dell; H. R. Morris, *Biochem. J.* **1997**, *327*, 519–525.
37. L. Anderson; F. M. Unger, Bacterial Liposaccharides. In *ACS Symposium Series 231*; L. Anderson, F. M. Unger, Eds.; ACS Symposium Series 231; American Chemical Society: Washington, DC, 1983.
38. M. Daffe; P. J. Brennan; M. McNeil, *J. Biol. Chem.* **1990**, *265*, 6734–6743.
39. E. Vilkas; C. Amar; J. Markovits; J. Vliegenthart; J. Kamerling, *Biochim. Biophys. Acta* **1973**, *297*, 423–435.
40. L. J. Alderwick; E. Radmacher; M. Seidel; R. Gande; P. G. Hitchen; H. R. Morris; A. Dell; H. Sahn; L. Eggeling; G. S. Besra, *J. Biol. Chem.* **2005**, *280*, 32362–32371.
41. G. S. Besra; K. H. Khoo; M. R. McNeil; A. Dell; H. R. Morris; P. J. Brennan, *Biochemistry* **1995**, *34*, 4257–4266.

42. S. Bhamidi; M. S. Scherman; C. D. Rithner; J. E. Prenni; D. Chatterjee; K. H. Khoo; M. R. McNeil, *J. Biol. Chem.* **2008**, *283*, 12992–13000, doi: 10.1074/jbc.M800222200.
43. M. Daffe; M. McNeil; P. J. Brennan, *Carbohydr. Res.* **1993**, *249*, 383–398.
44. A. Lee; S. W. Wu; M. S. Scherman; J. B. Torrelles; D. Chatterjee; M. R. McNeil; K. H. Khoo, *Biochemistry* **2006**, *45*, 15817–15828, doi: 10.1021/bi060688d.
45. V. Briken; S. A. Porcelli; G. S. Besra; L. Kremer, *Mol. Microbiol.* **2004**, *53*, 391–403.
46. D. Chatterjee; K. H. Khoo, *Glycobiology* **1998**, *8*, 113–120.
47. S. W. Hunter; P. J. Brennan, *J. Biol. Chem.* **1990**, *265*, 9272–9279.
48. P. Brennan; C. E. Ballou, *J. Biol. Chem.* **1967**, *242*, 3046–3056.
49. M. Gilleron; L. Bala; T. Brando; A. Vercellone; G. Puzo, *J. Biol. Chem.* **2000**, *275*, 677–684.
50. K. H. Khoo; A. Dell; H. R. Morris; P. J. Brennan; D. Chatterjee, *Glycobiology* **1995**, *5*, 117–127.
51. K. Leopold; W. Fischer, *Anal. Biochem.* **1993**, *208*, 57–64.
52. J. Nigou; M. Gilleron; T. Brando; A. Vercellone; G. Puzo, *Glycoconj. J.* **1999**, *16*, 257–264.
53. J. Nigou; M. Gilleron; B. Cahuzac; J. D. Bounery; M. Herold; M. Thurnher; G. Puzo, *J. Biol. Chem.* **1997**, *272*, 23094–23103.
54. J. Nigou; M. Gilleron; G. Puzo, *Biochem. J.* **1999**, *337*, 453–460.
55. J. Nigou; M. Gilleron; G. Puzo, *Biochimie* **2003**, *85*, 153–166.
56. J. Nigou; M. Gilleron; M. Rojas; L. F. Garcia; M. Thurnher; G. Puzo, *Microbes Infect.* **2002**, *4*, 945–953, doi: PII S1286-4579(02)01621-0.
57. M. C. Pangborn; J. A. McKinney, *J. Lipid Res.* **1966**, *7*, 627–633.
58. Y. C. Lee; C. E. Ballou, *Biochemistry* **1965**, *4*, 1395–1404.
59. G. S. Besra; C. B. Morehouse; C. M. Rittner; C. J. Waechter; P. J. Brennan, *J. Biol. Chem.* **1997**, *272*, 18460–18466.
60. P. Brennan; C. E. Ballou, *J. Biol. Chem.* **1968**, *243*, 2975–2984.
61. D. Chatterjee; S. W. Hunter; M. McNeil; P. J. Brennan, *J. Biol. Chem.* **1992**, *267*, 6228–6233.
62. D. Kaur; S. Berg; P. Dinadayala; B. Gicquel; D. Chatterjee; M. R. McNeil; V. D. Vissa; D. C. Crick; M. Jackson; P. J. Brennan, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13664–13669.
63. D. Kaur; M. R. McNeil; K. H. Khoo; D. Chatterjee; D. C. Crick; M. Jackson; P. J. Brennan, *J. Biol. Chem.* **2007**, *282*, 27133–27140, doi: 10.1074/jbc.M703389200.
64. K. H. Khoo; A. Dell; H. R. Morris; P. J. Brennan; D. Chatterjee, *J. Biol. Chem.* **1995**, *270*, 12380–12389.
65. L. B. Shi; S. Berg; A. Lee; J. S. Spencer; J. Zhang; V. Vissa; M. R. McNeil; K. H. Khoo; D. Chatterjee, *J. Biol. Chem.* **2006**, *281*, 19512–19526, doi: 10.1074/jbc.M513846200.
66. Y. Guerardel; E. Maes; V. Briken; F. Chirat; Y. Leroy; C. Loch; G. Strecker; L. Kremer, *J. Biol. Chem.* **2003**, *278*, 36637–36651.
67. H. H. Khoo; J. B. Tang; D. Chatterjee, *J. Biol. Chem.* **2001**, *276*, 3863–3871.
68. M. Joe; D. Sun; H. Taha; G. C. Completo; J. E. Croudace; D. A. Lammas; G. S. Besra; T. L. Lowary, *J. Am. Chem. Soc.* **2006**, *128*, 5059–5072, doi: 10.1021/ja057373q.
69. P. Ludwiczak; M. Gilleron; Y. Bordat; C. Martin; B. Gicquel; G. Puzo, *Microbiology* **2002**, *148*, 3029–3037.
70. S. A. Stalford; M. A. Fascione; S. J. Sasindran; D. Chatterjee; S. Dhandayuthapani; W. B. Turnbull, *Chem. Commun.* **2009**, 110–112.
71. A. Treumann; X. D. Feng; L. McDonnell; P. J. Derrick; A. E. Ashcroft; D. Chatterjee; S. W. Homans, *J. Mol. Biol.* **2002**, *316*, 89–100.
72. W. B. Turnbull; K. H. Shimizu; D. Chatterjee; S. W. Homans; A. Treumann, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 3918–3922, doi: 10.1002/anie.200454119.
73. C. Delmas; M. Gilleron; T. Brando; A. Vercellone; M. Gheorghiu; M. Riviere; G. Puzo, *Glycobiology* **1997**, *7*, 811–817.
74. J. B. Torrelles; K. H. Khoo; P. A. Sieling; R. L. Modlin; N. N. Zhang; A. M. Marques; A. Treumann; C. D. Rithner; P. J. Brennan; D. Chatterjee, *J. Biol. Chem.* **2004**, *279*, 41227–41239.
75. A. Lemassu; M. Daffe, *Biochem. J.* **1994**, *297*, 351–357.
76. E. Maes; B. Coddeville; L. Kremer; Y. Guerardel, *Glycoconj. J.* **2007**, *24*, 439–448, doi: 10.1007/s10719-007-9036-1.
77. S. Mahapatra; D. C. Crick; P. J. Brennan, *J. Bacteriol.* **2000**, *182*, 6827–6830.
78. K. A. De Smet; K. E. Kempell; A. Gallagher; K. Duncan; D. B. Young, *Microbiology* **1999**, *145*, 3177–3184.
79. S. Mahapatra; T. Yagi; J. T. Belisle; B. J. Espinosa; P. J. Hill; M. R. McNeil; P. J. Brennan; D. C. Crick, *J. Bacteriol.* **2005**, *187*, 2747–2757.
80. J. van Heijenoort, Biosynthesis of Bacterial Peptidoglycan Unit. In *Bacterial Cell Wall*; J. M. Ghuyssen, R. Hakenbeck, Eds.; Elsevier Medical Press: Amsterdam, 1994; pp 39–54.
81. J. van Heijenoort, Murein Synthesis. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed.; F. C. Neidhardt, Ed.; ASM Press, Washington, DC, 1996; Vol. 1, pp 1025–1034.
82. J. van Heijenoort, *Cell. Mol. Life Sci.* **1998**, *54*, 300–304.
83. J. van Heijenoort, *Glycobiology* **2001**, *11*, 25R–36R.
84. J. van Heijenoort, *Nat. Prod. Rep.* **2001**, *18*, 503–519.
85. J. M. Ghuyssen, *Annu. Rev. Microbiol.* **1991**, *45*, 37–67.
86. H. F. Chambers; D. Moreau; D. Yajko; C. Miick; C. Wagner; C. Hackbarth; S. Kocagoz; E. Rosenberg; W. K. Hadley; H. Nikaido, *Antimicrob. Agents Chemother.* **1995**, *39*, 2620–2624.
87. J. Basu; R. Chattopadhyay; M. Kundu; P. Chakrabarti, *J. Bacteriol.* **1992**, *174*, 4829–4832.
88. H. Billman-Jacobe; R. E. Haites; R. L. Coppel, *Antimicrob. Agents Chemother.* **1999**, *43*, 3011–3013.
89. J. Keer; M. J. Smeulders; K. M. Gray; H. D. Williams, *Microbiology* **2000**, *146*, 2209–2217.
90. A. Dasgupta; P. Datta; M. Kundu; J. Basu, *Microbiology* **2006**, *152*, 493–504.
91. M. Lavollay; M. Arthur; M. Fourgeaud; L. Dubost; A. Marie; N. Veziris; D. Blanot; L. Gutmann; J. L. Mainardi, *J. Bacteriol.* **2008**, *190*, 4360–4366, doi: 10.1128/JB.00239-08.
92. J. B. Raymond; S. Mahapatra; D. C. Crick; M. S. Pavelka, *J. Biol. Chem.* **2005**, *280*, 326–333.
93. S. Mahapatra; D. C. Crick; M. R. McNeil; P. J. Brennan, *J. Bacteriol.* **2008**, *190*, 655–661, doi: 10.1128/JB.00982-07.
94. D. Menginlecreulx; J. Vanheijenoort, *J. Bacteriol.* **1994**, *176*, 5788–5795.

95. L. R. Olsen; S. L. Roderick, *Biochemistry* **2001**, *40*, 1913–1921.
96. W. L. Zhang; V. C. Jones; M. S. Scherman; S. Mahapatra; D. Crick; S. Bhamidi; Y. Xin; M. R. McNeil; Y. F. Ma, *Int. J. Biochem. Cell Biol.* **2008**, *40*, 2560–2571.
97. C. M. Sasseti; D. H. Boyd; E. J. Rubin, *Mol. Microbiol.* **2003**, *48*, 77–84.
98. W. Li; Y. Xin; M. R. McNeil; Y. Ma, *Biochem. Biophys. Res. Commun.* **2006**, *342*, 170–178.
99. Y. Ma; F. Pan; M. R. McNeil, *J. Bacteriol.* **2002**, *184*, 3392–3395.
100. S. T. M. Allard; M. F. Giraud; C. Whitfield; M. Graninger; P. Messner; J. H. Naismith, *J. Mol. Biol.* **2001**, *307*, 283–295.
101. S. T. M. Allard; M.-F. Giraud; C. Whitfield; P. Messner; J. H. Naismith, *Acta Crystallogr. D Biol. Crystallogr.* **2000**, *56* (Pt. 2), 222–225.
102. W. Blankenfeldt; M. Asuncion; J. S. Lam; J. H. Naismith, *EMBO J.* **2000**, *19*, 6652–6663.
103. W. Blankenfeldt; M. F. Giraud; G. Leonard; R. Rahim; C. Creuzenet; J. S. Lam; J. H. Naismith, *Acta Crystallogr. D Biol. Crystallogr.* **2000**, *56*, 1501–1504.
104. W. Blankenfeldt; I. D. Kerr; M. F. Giraud; H. J. McMiken; G. Leonard; C. Whitfield; P. Messner; M. Graninger; J. H. Naismith, *Structure (Camb.)* **2002**, *10*, 773–786.
105. C. Dong; L. L. Major; V. Srikanthasasan; J. C. Errey; M. F. Giraud; J. S. Lam; M. Graninger; P. Messner; M. R. McNeil; R. A. Field; C. Whitfield; J. H. Naismith, *J. Mol. Biol.* **2007**, *365*, 146–159.
106. M.-F. Giraud; F. M. Gordon; C. Whitfield; P. Messner; S. A. McMahon; J. H. Naismith, *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55*, 706–708.
107. M.-F. Giraud; G. A. Leonard; R. A. Field; C. Berlind; J. H. Naismith, *Nat. Struct. Biol.* **2000**, *7*, 398–402.
108. M.-F. Giraud; H. J. McMiken; G. A. Leonard; P. Messner; C. Whitfield; J. H. Naismith, *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55* (Pt. 12), 2043–2046.
109. F. Pan; M. Jackson; Y. Ma; M. R. McNeil, *J. Bacteriol.* **2001**, *183*, 3991–3998.
110. M. Soltero-Higgin; E. E. Carlson; T. D. Gruber; L. L. Kiessling, *Nat. Struct. Mol. Biol.* **2004**, *11*, 539–543.
111. K. Beis; V. Srikanthasasan; H. Liu; S. W. B. Fullerton; V. A. Bamford; D. A. R. Sanders; C. Whitfield; M. R. McNeil; J. H. Naismith, *J. Mol. Biol.* **2005**, *348*, 971–982.
112. D. R. Sanders; A. G. Steins; S. A. McMahon; M. R. McNeil; C. Whitfield; J. H. Naismith, *Nat. Struct. Biol.* **2001**, *8*, 858–863.
113. J. S. Klutts; K. Hatanaka; Y. T. Pan; A. D. Elbein, *Arch. Biochem. Biophys.* **2002**, *398*, 229–239.
114. M. Scherman; A. Weston; K. Duncan; A. Whittington; R. Upton; L. Deng; R. Comber; J. D. Friedrich; M. McNeil, *J. Bacteriol.* **1995**, *177*, 7125–7130.
115. M. S. Scherman; L. KalbeBournonville; D. Bush; L. Y. Deng; M. McNeil, *J. Biol. Chem.* **1996**, *271*, 29652–29658.
116. H. Huang; M. S. Scherman; W. D’Haeze; D. Vereecke; M. Holsters; D. C. Crick; M. R. McNeil, *J. Biol. Chem.* **2005**, *280*, 24539–24543.
117. K. Mikusova; H. Huang; T. Yagi; M. Holsters; D. Vereecke; W. D’Haeze; M. S. Scherman; P. J. Brennan; M. R. McNeil; D. C. Crick, *J. Bacteriol.* **2005**, *187*, 8020–8025.
118. K. Mikusova; M. Mikus; G. Besra; I. Hancock; P. J. Brennan, *J. Biol. Chem.* **1996**, *271*, 7820–7828.
119. A. E. Grzegorzewicz; Y. Ma; V. Jones; D. Crick; A. Liav; M. R. McNeil, *Microbiology* **2008**, *154*, 3724–3730.
120. L. J. Alderwick; L. G. Dover; N. Veerapen; S. S. Gurucha; L. Kremer; D. L. Roper; A. K. Pathak; R. C. Reynolds; G. S. Besra, *Protein Expr. Purif.* **2008**, *58*, 332–341.
121. K. Mikusova; M. Belanova; J. Kordulakova; K. Honda; M. R. McNeil; S. Mahapatra; D. C. Crick; P. J. Brennan, *J. Bacteriol.* **2006**, *118*, 6592–6598.
122. L. Kremer; L. G. Dover; C. Morehouse; P. Hitchin; M. Everett; H. R. Morris; A. Dell; P. J. Brennan; M. R. McNeil; C. Flaherty; K. Duncan; G. S. Besra, *J. Biol. Chem.* **2001**, *276*, 26430–26440.
123. K. Mikusova; T. Yagi; R. Stern; M. R. McNeil; G. S. Besra; D. C. Crick; P. J. Brennan, *J. Biol. Chem.* **2000**, *275*, 33890–33897.
124. A. G. Amin; R. Goude; L. Shi; J. Zhang; D. Chatterjee; T. Parish, *Microbiology* **2008**, *154*, 240–248.
125. S. Berg; J. Starbuck; J. B. Torrelles; V. D. Vissa; D. C. Crick; D. Chatterjee; P. J. Brennan, *J. Biol. Chem.* **2005**, *280*, 5651–5663.
126. V. E. Escuyer; M. A. Lety; J. B. Torrelles; K. H. Khoo; J. B. Tang; C. D. Rithner; C. Frehel; M. R. McNeil; P. J. Brennan; D. Chatterjee, *J. Biol. Chem.* **2001**, *276*, 48854–48862.
127. S. Khasnobis; J. Zhang; S. K. Angala; A. G. Amin; M. R. McNeil; D. C. Crick; D. Chatterjee, *Chem. Biol.* **2006**, *13*, 787–795.
128. N. Zhang; J. B. Torrelles; M. R. McNeil; V. E. Escuyer; K. H. Khoo; P. J. Brennan; D. Chatterjee, *Mol. Microbiol.* **2003**, *50*, 69–76.
129. L. J. Alderwick; L. G. Dover; M. Seidel; R. Gande; H. Sahn; L. Eggeling; G. S. Besra, *Glycobiology* **2006**, *16*, 1073–1081, doi: 10.1093/glycob/cwl030.
130. L. J. Alderwick; M. Seidel; H. Sahn; G. S. Besra; L. Eggeling, *J. Biol. Chem.* **2006**, *281*, 15653–15661.
131. H. L. Birch; L. J. Alderwick; A. Bhatt; D. Rittmann; K. Krumbach; A. Singh; Y. Bai; T. L. Lowary; L. Eggeling; G. S. Besra, *Mol. Microbiol.* **2008**, *69*, 1191–1206, doi: 10.1111/j.1365-2958.2008.06354.x.
132. M. Seidel; L. J. Alderwick; H. L. Birch; H. Sahn; L. Eggeling; G. S. Besra, *J. Biol. Chem.* **2007**, *282*, 14729–14740.
133. L. Shi; R. Zhou; Z. Liu; T. L. Lowary; P. H. Seeberger; B. L. Stocker; D. C. Crick; K. H. Khoo; D. Chatterjee, *J. Bacteriol.* **2008**, *190*, 5248–5255.
134. A. E. Belanger; G. S. Besra; M. E. Ford; K. Mikusova; J. Belisle; P. J. Brennan; J. M. Inamine, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11919–11924.
135. H. L. Birch; L. J. Alderwick; A. Bhatt; D. Rittmann; K. Krumbach; A. Singh; Y. Bai; T. L. Lowary; L. Eggeling; G. S. Besra, *Mol. Microbiol.* **2008**, *69*, 1191–1206.
136. M. Seidel; L. J. Alderwick; H. L. Birch; H. Sahn; L. Eggeling; G. S. Besra, *J. Biol. Chem.* **2007**, *282*, 14729–14740.
137. M. Jackson; C. Raynaud; M. A. Laneelle; C. Gullhot; C. Laurent-Winter; D. Ensergueix; B. Gicquel; M. Daffe, *Mol. Microbiol.* **1999**, *31*, 1573–1587.
138. V. Puech; N. Bayan; K. Salim; G. Leblon; M. Daffe, *Mol. Microbiol.* **2000**, *35*, 1026–1041.
139. C. Sousa-D’Auria; R. Kacem; V. Puech; M. Tropis; G. Leblon; C. Houssin; M. Daffe, *FEMS Microbiol. Lett.* **2003**, *224*, 35–44.
140. J. T. Belisle; V. D. Vissa; T. Sievert; K. Takayama; P. J. Brennan; G. S. Besra, *Science* **1997**, *276*, 1420–1422.
141. M. Salman; J. T. Lonsdale; G. S. Besra; P. J. Brennan, *Biochim. Biophys. Acta* **1999**, *1436*, 437–450.
142. N. Bachhawat; S. C. Mande; J. Mol. Biol. **1999**, *291*, 531–536.
143. T. Parish; J. Liu; H. Nikaido; N. G. Stoker, *J. Bacteriol.* **1997**, *179*, 7827–7833.

144. H. Paulus; E. P. Kennedy, *J. Am. Chem. Soc.* **1959**, *80*, 6689–6690.
145. M. Jackson; D. C. Crick; P. J. Brennan, *J. Biol. Chem.* **2000**, *275*, 30092–30099.
146. K. Takayama; D. S. Goldman, *Biochim. Biophys. Acta* **1969**, *176*, 196–198.
147. J. Kordulakova; M. Gilleron; G. Puzo; P. J. Brennan; B. Gicquel; K. Mikusova; M. Jackson, *J. Biol. Chem.* **2003**, *278*, 36285–36295, doi: 10.1074/jbc.M303639200.
148. J. Kordulakova; M. Gilleron; K. Mikusova; G. Puzo; P. J. Brennan; B. Gicquel; M. Jackson, *J. Biol. Chem.* **2002**, *277*, 31335–31344.
149. M. L. Schaeffer; K. H. Khoo; G. S. Besra; D. Chatterjee; P. J. Brennan; J. T. Belisle; J. M. Inamine, *J. Biol. Chem.* **1999**, *274*, 31625–31631.
150. S. T. Cole; K. Eiglmeier; J. Parkhill; K. D. James; N. R. Thomson; P. R. Wheeler; N. Honore; T. Garnier; C. Churcher; D. Harris; K. Mungall; D. Basham; D. Brown; T. Chillingworth; R. Connor; R. M. Davies; K. Devlin; S. Duthoy; T. Feltwell; A. Fraser; N. Hamlin; S. Holroyd; T. Hornsby; K. Jagels; C. Lacroix; J. Maclean; S. Moule; L. Murphy; K. Oliver; M. A. Quail; M. A. Rajandream; K. M. Rutherford; S. Rutter; K. Seeger; S. Simon; M. Simmonds; J. Skelton; R. Squares; S. Squares; K. Stevens; K. Taylor; S. Whitehead; J. R. Woodward; B. G. Barrell, *Nature* **2001**, *409*, 1007–1011.
151. D. J. Lea-Smith; K. L. Martin; J. S. Pyke; D. Tull; M. J. McConville; R. L. Coppel; P. K. Crellin, *J. Biol. Chem.* **2008**, *283*, 6773–6782, doi: 10.1074/jbc.M707139200.
152. L. Kremer; S. S. Gurcha; P. Bifani; P. G. Hitchen; A. Baulard; H. R. Morris; A. Dell; P. J. Brennan; G. S. Besra, *Biochem. J.* **2002**, *363*, 437–447.
153. Y. S. Morita; C. B. C. Sena; R. F. Waller; K. Kurokawa; M. F. Sernee; F. Nakatani; R. E. Haites; H. Billman-Jacobe; M. J. McConville; Y. Maeda; T. Kinoshita, *J. Biol. Chem.* **2006**, *281*, 25143–25155, doi: 10.1074/jbc.M604214200.
154. C. R. H. Raetz; C. Whitfield, *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
155. E. G. Berger, *Glycobiology* **2002**, *12*, 29R–36R.
156. K. J. Colley, *Glycobiology* **1997**, *7*, 1–13.
157. S. S. Gurcha; A. R. Baulard; L. Kremer; C. Locht; D. B. Moody; W. Muhlecker; C. E. Costello; D. C. Crick; P. J. Brennan; G. S. Besra, *Biochem. J.* **2002**, *365*, 441–450.
158. D. Kaur; A. Obregon-Henao; H. Pham; D. Chatterjee; P. J. Brennan; M. Jackson, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17973–17977.
159. J. C. Schultz; K. Takayama, *Biochim. Biophys. Acta* **1975**, *381*, 175–184.
160. K. Yokoyama; C. E. Ballou, *J. Biol. Chem.* **1989**, *264*, 21621–21628.
161. Y. S. Morita; J. H. Patterson; H. Billman-Jacobe; M. J. McConville, *Biochem. J.* **2004**, *378*, 589–597.
162. P. Dinadayala; D. Kaur; S. Berg; A. G. Amin; V. D. Vissa; D. Chatterjee; P. J. Brennan; D. C. Crick, *J. Biol. Chem.* **2006**, *281*, 20027–20035, doi: 10.1074/jbc.M603395200.
163. T. R. Paul; T. J. Beveridge, *J. Bacteriol.* **1992**, *174*, 6508–6517.
164. P. J. Brennan; P. Draper, Ultrastructure of *Mycobacterium tuberculosis*. In *Tuberculosis: Pathogenesis, Protection, and Control*; B. R. Bloom, Ed.; American Society for Microbiology: Washington, DC, 1994; pp 271–284.
165. M. Daffe; P. Draper, *Adv. Microb. Physiol.* **1998**, *39*, 131–203.
166. P. Draper, *Front Biosci.* **1998**, *3*, D1253–D1261.
167. A. Ortalo-Magne; A. Lemassu; M. A. Laneelle; F. Bardou; G. Silve; P. Gounon; G. Marchal; M. Daffe, *J. Bacteriol.* **1996**, *178*, 456–461.
168. S. Pitarque; G. Larrouy-Maumus; B. Payre; M. Jackson; G. Puzo; J. Nigou, *Tuberculosis* **2008**, *88*, 560–565.
169. N. Rastogi, *Res. Microbiol.* **1991**, *142*, 464–476.
170. S. W. Hunter; H. Gaylord; P. J. Brennan, *J. Biol. Chem.* **1986**, *261*, 2345–2351.
171. L. Barksdale; K. S. Kim, *Bacteriol. Rev.* **1977**, *41*, 217–372.
172. C. Hoffmann; A. Leis; M. Niederweis; J. M. Plitzko; H. Engelhardt, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3963–3967, doi: 10.1073/pnas.0709530105.
173. B. Zuber; M. Chami; C. Houssin; J. Dubochet; G. Griffiths; M. Daffe, *J. Bacteriol.* **2008**, *190*, 5672–5680, doi: 10.1128/JB.01919-07.
174. J. Dubochet; A. W. McDowell; B. Menge; E. N. Schmid; K. G. Lickfeld, *J. Bacteriol.* **1983**, *155*, 381–390.
175. V. R. F. Matias; A. Al-Amoudi; J. Dubochet; T. J. Beveridge, *J. Bacteriol.* **2003**, *185*, 6112–6118, doi: 10.1128/JB.185.20.6112-6118.2003.
176. H. van Halbeek, *Methods Enzymol.* **1994**, *230*, 132–168.
177. C. A. Bush; M. Martin-Pastor; A. Imberty, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 269–293.
178. W. Vollmer; D. Blanot; M. A. de Pedro, *FEMS Microbiol. Rev.* **2008**, *32*, 149–167, doi: 0.1111/j.1574-6976.2007.00094.x.
179. D. E. Minnikin, Lipids: Complex Lipids, Their Chemistry, Biosynthesis and Roles. In *The Biology of Mycobacteria*; C. Ratledge, J. Stanford, Eds.; Academic Press: London, 1982; pp 95–184.
180. M. Villeneuve; M. Kawai; H. Watanabe; M. Watanabe; D. E. Minnikin; H. Nakahara, *Biochim. Biophys. Acta* **2005**, *1715*, 71–80, doi: 10.1016/j.bbamem.2005.07.005.
181. M. Villeneuve; M. Kawai; M. Watanabe; Y. Aoyagi; Y. Hitotsuyanagi; K. Takeya; H. Gouda; S. Hirono; D. E. Minnikin; H. Nakahara, *Biochim. Biophys. Acta* **2007**, *1768*, 1717–1726, doi: 10.1016/j.bbamem.2007.04.003.
182. M. Villeneuve; M. Kawai; H. Watanabe; Y. Aoyagi; Y. Hitotsuyanagi; K. Takeya; H. Gouda; S. Hirono; D. E. Minnikin; H. Nakahara, *Biochim. Biophys. Acta* **2007**, *1768*, 1717–1726, doi: 0.1016/j.bbamem.2007.08.003.
183. H. Gaylord; P. J. Brennan; D. B. Young; T. M. Buchanan, *Infect. Immun.* **1987**, *55*, 2860–2863.
184. H. Nikaido, *Semin. Cell Dev. Biol.* **2001**, *12*, 215–223.
185. D. Alsteens; C. Verbelen; E. Dague; D. Raze; A. R. Baulard; Y. F. Dufrene, *Eur. J. Physiol.* **2008**, *456*, 117–125, doi: 10.1007/s00424-007-0386-0.

Biographical Sketches



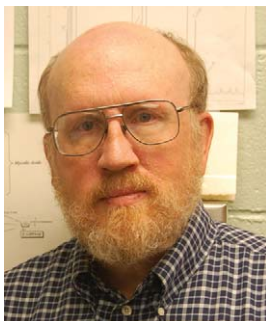
Dean C. Crick received his B.Sc. and M.Sc. degrees from the University of British Columbia and his Ph.D. (biochemistry) from the University of Western Ontario, Canada. He conducted his postdoctoral studies at the University of Kentucky under the supervision of Dr. C. J. Waechter. In 1998 he joined the faculty of the Department of Microbiology, now the Department of Microbiology, Immunology and Pathology, at Colorado State University, Fort Collins, where he pursues biochemical studies on the structure and biogenesis of the cell wall and lipids of *Mycobacterium tuberculosis*; information gained in these studies is applied to the development of new antituberculosis agents.



Delphi Chatterjee received her B.Sc. and M.Sc. degrees from Visva Bharati University, India, and her M.Sc. and Ph.D. degrees from the University of London, UK, after which she moved to Canada to work with Dr. Gerald O. Aspinall at York University. In 1984 she moved to Colorado State University where her research now focuses on elucidation of the structure–function relationships of the multisubunit, membrane-bound enzyme complexes responsible for the assembly of biologically active glycoconjugates, the structural determination of cell wall glycoconjugates, and the identification of genes encoding responsible for the synthesis of these structures.



Michael S. Scherman received his B.S. (Microbiology) and his M.S.B.A (Computer Information Systems) from Colorado State University, CO, USA. He has worked with Dr. Michael R. McNeil at the Mycobacteria Research Laboratories at Colorado State University since 1992, studying the structural aspects of the mycobacterial cell envelope, TB drug development, targeting of key biosynthetic enzymes, and computer modeling of metabolic and enzymatic pathways and structures.



Michael R. McNeil received his M.S. degree from MIT and his Ph.D. (chemistry) from the University of Colorado. Between his M.S. and Ph.D. degrees he worked for many years in the laboratory of Professor Peter Albersheim where he learnt his structural carbohydrate chemistry on plant systems. He moved to Colorado State University in 1985 and began his studies on the mycobacterial cell wall with Professor Patrick Brennan. Since that time he has been working on mycobacterial cell wall structure, cell wall biosynthesis, and the targeting of cell wall biosynthetic enzymes for TB drug development.

6.14 Structure, Biosynthesis, and Function of Glycosaminoglycans

Courtney L. Jones and Jian Liu, University of North Carolina, Chapel Hill, NC, USA

Ding Xu, University of California, San Diego, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.14.1 | Introduction | 407 |
| 6.14.2 | Heparan Sulfate and Heparin | 407 |
| 6.14.2.1 | Localization of Proteoglycans | 408 |
| 6.14.2.2 | Biosynthesis of Heparan Sulfate | 410 |
| 6.14.2.2.1 | Synthesis of building block | 410 |
| 6.14.2.2.2 | Chain initiation | 410 |
| 6.14.2.2.3 | Chain polymerization | 412 |
| 6.14.2.2.4 | Chain modification | 413 |
| 6.14.2.3 | Physiological and Pathophysiological Importance | 418 |
| 6.14.2.3.1 | Anticoagulation | 418 |
| 6.14.2.3.2 | Antiviral activity | 418 |
| 6.14.2.3.3 | Inflammation | 419 |
| 6.14.3 | Chondroitin Sulfate and Dermatan Sulfate | 419 |
| 6.14.3.1 | Localization of CSPG and DSPG | 419 |
| 6.14.3.1.1 | The lectican family | 419 |
| 6.14.3.1.2 | DS proteoglycans | 420 |
| 6.14.3.2 | Physiological Importance of Chondroitin Sulfate and Dermatan Sulfate | 420 |
| 6.14.3.2.1 | Anticoagulation | 421 |
| 6.14.3.2.2 | Cartilage function | 421 |
| 6.14.3.2.3 | Neurobiological effects | 421 |
| 6.14.3.2.4 | Pathogen receptors | 421 |
| 6.14.4 | Other Glycosaminoglycans | 421 |
| References | | 423 |

6.14.1 Introduction

Three major classes of biopolymers exist in nature: proteins, nucleic acids, and polysaccharides. Unlike proteins and nucleic acids, our understanding of polysaccharides is limited largely because of the structural complexity and the absence of methods for replicating polysaccharide molecules with high fidelity. The glycosaminoglycan (GAG) represents an important type of polysaccharide in mammals because of the wide range of biological functions encompassed by GAGs. The GAG family includes heparan sulfate (HS), chondroitin sulfate (CS), keratan sulfate (KS), and hyaluronic acid (or hyaluronan, HA). Heparin, a special form of HS, is a commonly used anticoagulant drug with a worldwide annual sale of more than \$3 billion. In addition to the anticoagulant activity, exploration of the anticancer, antiviral, and anti-inflammatory activities for therapeutic purposes has been expanded in recent years. The purpose of this chapter is to summarize our current understandings of GAGs, particularly the biosynthesis of HS and related biological functions.

6.14.2 Heparan Sulfate and Heparin

HS is considered the king of the GAG family because of its diversified biological functions and clinical applications. HS is present in both invertebrates and vertebrates in large quantities.^{1,2} The average length of HS is about 5–70 kDa, or 10–200 disaccharide repeating units.^{3,4} HS chains can be anchored by a core protein

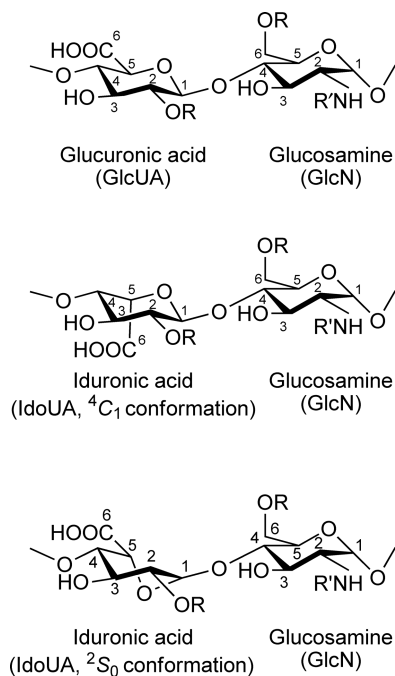


Figure 1 Disaccharide repeating units of HS. Sulfation ($R = -SO_3$) at carbon 6 (known as 6-*O*-sulfo glucosamine) of glucosamine is common. Sulfation at carbon-2 of iduronic acid (known as IdoUA2S) is common. Sulfation at carbon-3 of glucosamine (known as 3-*O*-sulfo glucosamine) is rare. Both *N*-acetyl ($R' = \text{acetyl}$, GlcNAc) and *N*-sulfo glucosamine ($R' = -SO_3$, GlcNS) are common. *N*-unsubstituted glucosamine ($R' = -H$, GlcNH₂) is a low-abundance component. IdoUA \pm 2S is present in both 4C_1 and 2S_0 conformations. Both conformations are presented.

on the surface of cells, known as proteoglycans, or pericellularly linked to the basement membrane in the extracellular matrix.³ Heparin, a commonly used anticoagulant drug, is a special form of HS. Heparin is an exclusive product of mast cells, and contains higher levels of sulfo groups and more iduronic acid. Both heparin and HS consist of a repeating disaccharide unit of glucosamine and either glucuronic or iduronic acid with various sulfations as shown in **Figure 1**.

Various sulfation patterns confer the functional selectivities of HS and heparin. These sulfations occur at the NH₂, 3-OH, and 6-OH positions of glucosamine residue. The hexuronic acid residues, including both iduronic and glucuronic acid, can include sulfo groups at the 2-OH position. To further complicate the structure, the *N*-position of the glucosamine residue can also exist in the acetylated or unsubstituted form. Different substitutions at the saccharide unit result in up to 32 variable structures for the disaccharide unit, which serves as the building block for the polysaccharide.⁵ For a polysaccharide with a size of 20 disaccharide units, the theoretical number of permutations within the chain is 10³². However, it is believed that the actual number of variable polysaccharides are less than this value because of the substrate specificity restrictions of HS biosynthetic enzymes. Nevertheless, the existing permutations lead to the structure of HS from natural sources, which exhibit vast structural heterogeneity and diversity. Dissecting the relationship of structures and functions of HS has a large impact on understanding the physiological functions of HS as well as the development of HS-based therapeutic reagents.

6.14.2.1 Localization of Proteoglycans

HS can be found in almost all animal cells.^{2,6} The lowest organisms where HS has been found are the metazoans ctenophora and cnidaria or jellyfish.⁷⁻⁹ HS chains are attached to HS proteoglycans (HSPGs) either on the cell surface or within the extracellular matrix. The amount and presence of HSPGs varies from cell to cell.⁷ An HSPG consists of a core protein to which the HS chains are attached. The core proteins are anchored to the cell

membrane via a glycosylphosphatidylinositol (GPI) anchor or a membrane-spanning domain. HSPGs found in the extracellular matrix come from the core proteins that are secreted during the biosynthesis.¹⁰

HSPGs can be classified in two ways: part-time HSPGs and full-time HSPGs. Part-time HSPGs are proteins that have the ability to carry HS side chains, but can exist without HS side chains, also known as the nonproteoglycan form. CD44, betaglycan, and testican are a few examples of these proteoglycans. Full-time HSPGs are either directly membrane bound or are linked to components within the extracellular matrix. Full-time HSPGs include syndecans, glypicans, perlecan, agrin, and type XVIII collagen.^{11–14} A total of 13 genes encode the full-time HSPGs: four syndecans (syndecan-1 to -4), six glypicans (glypican-1 to -6), and one isoform of each perlecan, agrin, and type XVIII collagen although perlecan, agrin, and type XVIII also have splice variants. The average size of these protein cores range from about 32 to almost 500 kDa (**Table 1**).

Syndecan and glypican are both cell membrane-associated proteoglycans, while perlecan, agrin, and type XVIII collagen are extracellular components (**Figure 2**). Syndecan contains an integral membrane protein and

Table 1 Summary of cell surface and extracellular matrix proteoglycans

| Proteoglycans | Size (aa) | Number of chains attached | GenBank accession numbers |
|----------------------------|-----------|---------------------------|---|
| <i>Heparan sulfate</i> | | | |
| Syndecan | 310 | 5 | CAA42851 (<i>Homo sapiens</i>) |
| Glypican | 558 | 3 | CAA38139 (<i>Homo sapiens</i>) |
| Perlecan | 4391 | 3 | P98160 (<i>Homo sapiens</i>) |
| Agrin | 2045 | 3 | NP_940978 XP_372195 (<i>Homo sapiens</i>) |
| Type XVIII collagen | 1516 | 3 | AAC39658 (<i>Homo sapiens</i>) |
| <i>Chondroitin sulfate</i> | | | |
| Aggrecan | 721 | 100–150 | AAH36445 (<i>Homo sapiens</i>) |
| Versican | 3396 | 0–30 | NP_004376 (<i>Homo sapiens</i>) |
| Brevican | 911 | | CAI16352 (<i>Homo sapiens</i>) |
| Neurocan | 1321 | | AAC80576 (<i>Homo sapiens</i>) |
| <i>Dermatan sulfate</i> | | | |
| Decorin | 359 | 1 | AAH05322 (<i>Homo sapiens</i>) |
| Biglycan | 394 | 2 | AA52287 (<i>Homo sapiens</i>) |
| <i>Keratan sulfate</i> | | | |
| Aggrecan | 721 | 15–60 | AAH36445 (<i>Homo sapiens</i>) |
| Fibromodulin | 376 | 4 | CAA53233 (<i>Homo sapiens</i>) |
| Decorin | 359 | 1 | AAH05322 (<i>Homo sapiens</i>) |
| Lumican | 338 | 2–3 | AAA91639 (<i>Homo sapiens</i>) |

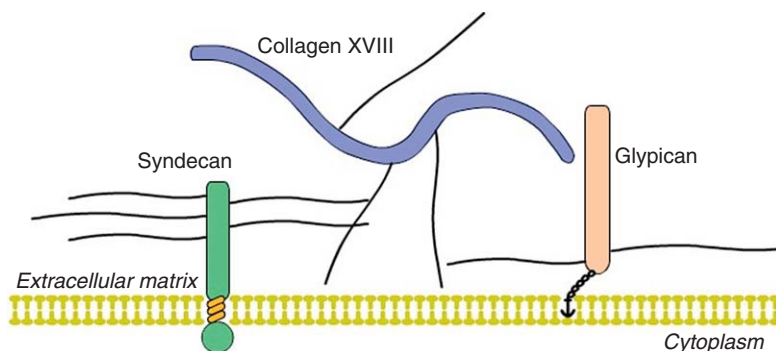


Figure 2 Schematic view of the HSPG. The core proteins are shown in colored tubes while the HS chains are shown in black lines. The core protein of syndecan is a transmembrane protein, with 3–5 attachment sites in the extracellular domain. The core protein of glypican attaches to the membrane through a glycosyl-phosphatidylinositol (GPI) anchor, with 1–3 HS attachment sites. The core protein of collagen is not directly attached to the cell membrane; it has three potential HS attachment sites.

glypican is anchored by glycosyl-phosphatidylinositol (GPI). Perlecan, agrin, and type XVIII collagen, although not directly associated with the plasma membrane, can be associated via integrins or other cell surface receptors.¹¹

6.14.2.2 Biosynthesis of Heparan Sulfate

The biosynthesis of HS has been the fundamental subject of the HS study field. A detailed view of HS biosynthesis would aid in elucidation of the structure–function relationship of HS in different biological contexts. In addition, a complete grasp of the biosynthesis mechanism would promote the development of the *in vitro* biosynthesis of HS or HS-like macromolecules for therapeutic purposes. In the past decade, as a result of the sustained effort to understand the biosynthetic pathway of HS, all enzymes involved in HS biosynthesis have been cloned and purified, which greatly advanced our knowledge of this essential biological process. It should be noted that the availability of HS biosynthetic enzymes also provided the opportunity for conducting enzyme-based synthesis of HS polysaccharides.^{15,16}

6.14.2.2.1 Synthesis of building block

HS biosynthesis takes place in the lumen of Golgi apparatus where the sugar residues are assembled and modified by the actions of multiple Golgi membrane-bound enzymes. However, the initial synthesis of building blocks occurs in the cytoplasm, not in the Golgi. These blocks are then transported into the Golgi lumen via an unknown mechanism for the subsequent modifications described below. The synthesis requires access to various activated sugar donors including UDP-xylose (UDP-Xyl), UDP-galactose (UDP-Gal), UDP-*N*-acetyl glucosamine (UDP-GlcNAc), and UDP-glucuronic acid (UDP-GlcUA) as well as the sulfo donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) for the sulfation reactions.

UDP-Xyl, UDP-Gal, and UDP-GlcUA are all derived from the activated glucose, namely UDP-glucose (UDP-Glc). UDP-Gal arises from epimerization at the C4 position of UDP-Glc; while UDP-GlcUA arises from UDP-Glc via UDP-Glc 6-dehydrogenase. UDP-Xyl is synthesized from UDP-GlcUA in the Golgi apparatus by UDP-GlcUA (Figure 3). The biosynthesis of UDP-GlcNAc is more complex. It starts from fructose 6-phosphate and requires three enzymes, a transaminase, an *N*-acetyltransferase, and a mutase, to generate GlcNAc 1-phosphate, which is then converted to UDP-GlcNAc by UDP-GlcNAc diphosphorylase. Once synthesized, these UDP-sugars are transported into the Golgi lumen through specific nucleotide sugar transporters (Figure 3).¹⁷

The sulfation of the HS chain requires the universal sulfo donor, 5'-phosphosulfate 3'-phosphoadenosine (PAPS). The biosynthesis of PAPS occurs in both the nucleus and the cytoplasm. Then PAPS is transported to the Golgi by PAPS transporter for HS biosynthesis.¹⁸

6.14.2.2.2 Chain initiation

The biosynthesis of HS is initiated by the formation of a tetrasaccharide linkage to the core proteins: GlcUA/β1-3Gal/β1-3Gal/β1-4Xyl/β1-*O*-Ser, which also serves as the linkage for the biosynthesis of CS. The formation of the tetrasaccharide linkage is catalyzed by the sequential actions of four glycosyltransferases, namely xylotransferase (XT), galactosyltransferase I (GalTI), galactosyltransferase II (GalTII), and glucuronosyltransferase I (GlcUAT1) (Figure 3).¹⁹

XT is the first enzyme that triggers the synthesis of the tetrasaccharide linkage at specific serine residues of the core proteins. Two highly homologous XTs (XTI and XTII) have been identified in vertebrates; however, only XT1 has been determined unambiguously to have XT activity.^{20,21} The full-length human XT1 has recently been expressed in mammalian cells in the active form and was shown to localize in the early *cis*-Golgi.²² Mechanisms of core protein attachment sites recognition by XT have attracted considerable attention, given the pivotal role of this enzyme in initiating synthesis of HSPG. A Ser–Gly repeating dipeptide seems to be the minimum structural requirement for the xylosylation to occur, and a Glu or Asp residue of the core protein is often present in the vicinity of the Ser–Gly sequence.²³ Nevertheless, how XT recognizes the substrate core proteins and selects the Ser residue is still an open question.

Once the xylose is transferred to the core protein, the glycan moiety is elongated by two galactosyltransferases, GalTI and GalTII. GalTI, belonging to the β1,4 GalT superfamily, transfers a galactose onto xylose to

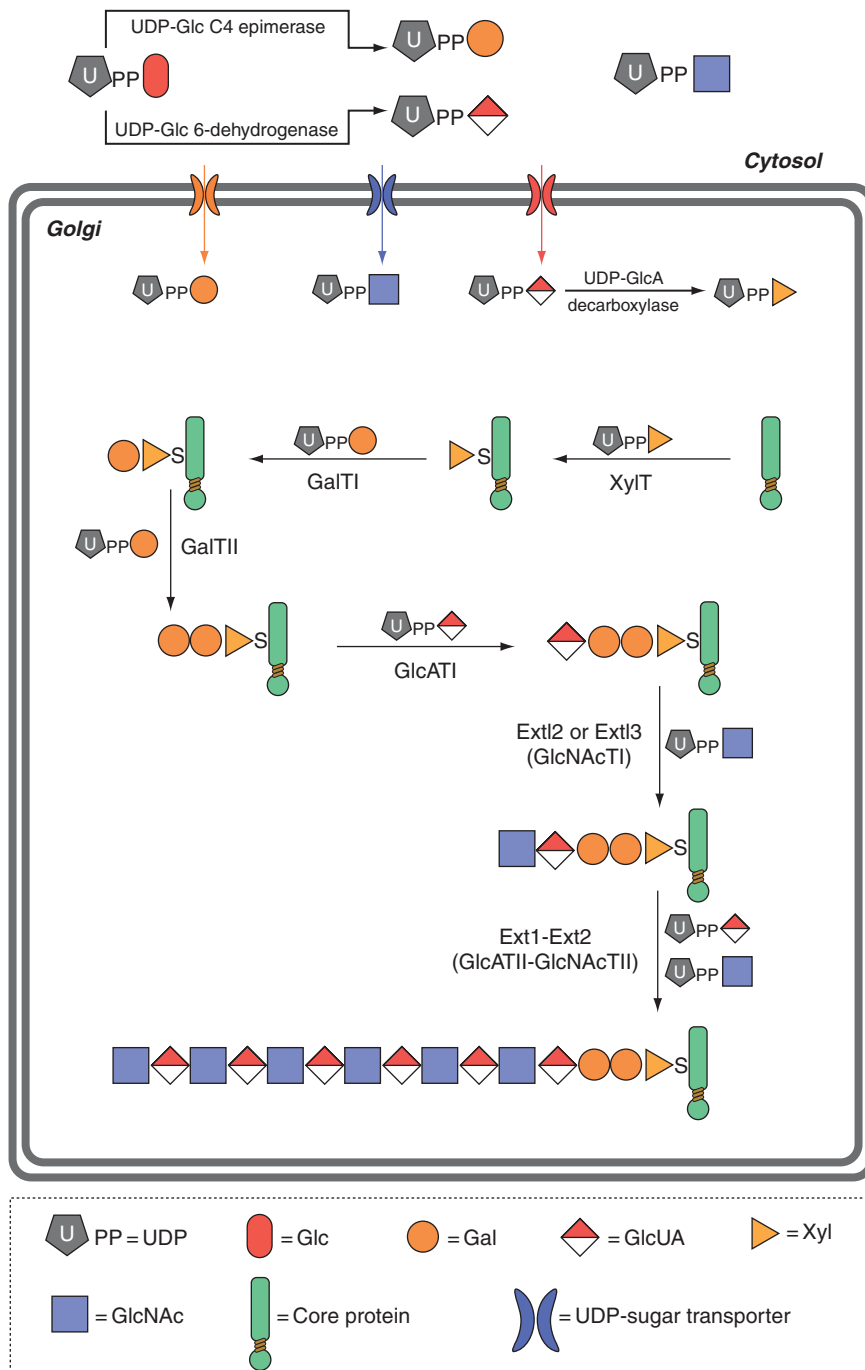


Figure 3 Schematic view of HS backbone synthesis. For simplicity, the synthesis of UDP-GlcNAc is not shown. The EXT family member EXTL1 is also absent in the figure, which was shown to possess GlcNAcTII activity.

form Gal β 1-4Xyl β 1-O-Ser, while GalTII belongs to the β 1-3 GalT superfamily, and adds the galactose to form Gal β 1-3Gal β 1-4Xyl β 1-O-Ser. Both GalTs were cloned and expressed, and their substrate specificities have been described. They have been shown to be located in the early-medial *cis*-Golgi.²⁴⁻²⁶

Synthesis of the tetrasaccharide linkage is finalized by GlcUAT1, which is localized in the medial *cis*-Golgi.²⁶ It transfers GlcUA from UDP-GlcUA to the existing glycan chain to form GlcUA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-O-Ser. The crystal structure of GlcUAT1 has been solved with the UDP, Mn²⁺, and a

substrate analogue Gal β 1-3Gal β 1-4Xyl, revealing that GlcUAT1 mainly recognizes the Gal β 1-3Gal moiety, rather than the trisaccharide as a whole.²⁷ More interestingly, a recent study demonstrated that sulfation at the Gal β 1-3Gal β 1- moiety can greatly influence the enzymatic activity of GlcUAT1, possibly providing a regulatory mechanism for proteoglycan biosynthesis.²⁸ In addition, sulfation of the galactose in the linkage region has only been seen in chondroitin sulfate proteoglycan (CSPG) but not in HSPG, suggesting it may serve as a signal to direct subsequent GAG synthesis.²⁹ Moreover, evidence suggests that the synthesis of the tetrasaccharide linkage can be carried out by GlcUAT-P, a promiscuous glucuronosyltransferase that normally works on glycoproteins, thus providing a potential redundant mechanism for the linkage region synthesis.³⁰

6.14.2.2.3 Chain polymerization

The tetrasaccharide linkage to the core protein provides a common precursor for the biosynthesis of both HS and CS polysaccharide chains. The critical step that separates polymerization of HS from that of CS is the addition of a GlcNAc α 1- unit instead of the GalNAc β 1- unit to the nonreducing end of the tetrasaccharide. This step should not be confused with the polymerization reaction, where the GlcNAc unit is also incorporated due to the action of a different enzyme.³¹ For clarity, the enzyme that adds GlcNAc to the linkage region is HS GlcNAc transferase I (GlcNAcT1), while the enzyme that adds GlcNAc to the growing HS chain is HS GlcNAc transferase II (GlcNAcTII) (**Figure 5**). Two mammalian proteins encoded by EXTL2 and EXTL3, belonging to the EXT (exostosin) gene family, have been proved to carry the GlcNAcT-I activity.^{32,33} Their activities have been demonstrated *in vitro* using a synthetic tetrasaccharide linkage analogue, GlcUA β 1-3Gal β 1-naphthalenemethanol, as the acceptor. However, the GlcNAcT-I activity could not be detected using the natural tetrasaccharide linkage GlcUA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-O-Ser as the acceptor. It has been proposed that hydrophobic residues on the core protein close to the linkage site are essential for the GlcNAcTI activity, which may attribute to the above-mentioned discrepancy, as the naphthalenemethanol group has strong hydrophobic nature.³³ Nevertheless, biochemical studies directly using HSPG core proteins as the acceptors of EXTL2 and EXTL3 are necessary to test this hypothesis. In addition, genetic studies targeting *EXTL2* and *EXTL3* are essential to establish the specific roles of these two isozymes in initiation of HS biosynthesis. It is important to note that a recent study of the boxer (*EXTL3*) mutant in zebra fish suggested that abolishing *EXTL3* function causes severe reduction of HS biosynthesis, supporting the role of *EXTL3* in initiating the synthesis of HS.³⁴

Once the GlcNAc unit is coupled to the linkage tetrasaccharide, the polymerization for the synthesis of HS is committed by adding GlcUA and GlcNAc alternatively (**Figure 3**). The enzymes encoded by the members of the EXT gene family, *EXT1* and *EXT2*, carry out the polymerization of HS backbone.^{35,36} *EXT1* and *EXT2* are both bifunctional enzymes, exhibiting GlcUATII (distinguishing from GlcUAT1 that synthesizes the linkage region) and GlcNAcTII activities (**Figure 3**). *EXT1* and *EXT2* are not redundant enzymes, instead, they form a complex and both are indispensable for the biosynthesis of HS *in vivo*.³⁷ The significance of heterocomplex of *EXT1* and *EXT2* *in vivo* are twofold. First, the complex helps direct the EXTs to the Golgi apparatus rather than the endoplasmic reticulum (ER).^{37,38} Second, mutations of either *EXT1* or *EXT2* could cause hereditary multiple exotoses, an inherited disease manifesting short stature and inequalities in limb length. Moreover, the individual EXT isoform has markedly lower HS polymerase activity compared to the *EXT1/EXT2* complex as determined by *in vitro* studies,³⁹ providing additional evidence for the importance of complex formation. It is important to note that *EXT1* and *EXT2* are not the only enzymes that possess GlcNAcT-II activity; *EXTL1* and *EXTL3* have also been demonstrated to be capable of adding GlcNAc residues to the growing HS chain *in vitro*.³³ But how the GlcNAcT-II activity of *EXTL1* and *EXTL3* correlates to that of *EXT1* and *EXT2* is an intriguing question, especially considering that they do not possess any GlcUAT-II activity. One hypothesis is that *EXTL1* and *EXTL3* may play regulatory roles in HS biosynthesis since they may compete with the *EXT1/EXT2* complex.³³ However, direct evidence supporting this hypothesis has been lacking, and the true functions of *EXTL1* and *EXTL3* in HS polymerization must wait for future genetic studies targeting these two enzymes.

The crystal structure of *EXTL2* has been solved in complex with the sugar donor UDP-GlcNAc, and in complex with UDP and the acceptor analogue GlcUA β 1-3Gal β 1-O-naphthalenemethanol, providing an excellent model for understanding the mechanism of action of EXT family members.⁴⁰ However, since *EXTL2* is the shortest isozyme of the family and it aligns to the C-terminal GlcNAcT domains of other

EXT members, its structure only helps to understand the GlcNAcT activity of EXT(L)s. Detailed biochemical studies are needed to elucidate the functions of the N-terminal domain of EXT(L)s. For some enzymes, more specifically EXT1 and EXT2, the N-terminal domain harbors the GlcUAT activity,⁴¹ and for others, EXTL1 and EXTL3, its function is still unknown.

6.14.2.2.4 Chain modification

Once polymerized, the HS backbone (or *N*-acetyl heparosan) is subjected to a series of modifications that exponentially diversify the structures of HS. The GlcUA units can be modified at two positions. The configuration at C5 position can be changed from S to R by glucuronyl C5-epimerase (C5-epi), giving rise to the IdoUA unit; and both GlcUA and IdoUA can be sulfated at the C2 position by HS 2-*O*-sulfotransferase (2-OST). The modifications of GlcNAc residues take place at three positions. The *N*-acetyl group of GlcNAc can be modified by *N*-deacetylase/*N*-sulfotransferase (NDST), a bifunctional enzyme that removes the *N*-acetyl group and installs a sulfo group; and the C3 and C6 position of glucosamine can be sulfated by HS 3-*O*-sulfotransferase (3-OST) and HS 6-*O*-sulfotransferase (6-OST), respectively (**Figure 4**).

6.14.2.2.4(i) *N*-deacetylase/*N*-sulfotransferase NDST is the enzyme that initiates modification of *N*-acetyl heparosan. *N*-sulfation of the glucosamine unit has been demonstrated to be essential for most of the subsequent modifications on HS, thus representing a critical step in HS biosynthesis.¹⁹ NDST does not modify HS uniformly, instead, the *N*-sulfated glucosamine (GlcNS) units are clustered, giving rise to the NS domain of HS, whereas the NA domain contains mostly the unreacted *N*-acetylated glucosamine (**Figure 4(b)**).¹

NDST can be divided into two domains: the N-terminal deacetylase domain (within ~600 amino acid residues of the N-terminus) and the C-terminal sulfotransferase domain (including about 280 amino acid residues). Both domains have been expressed and proved to contain the anticipated activities.^{42,43} In addition, NDST is a typical type II membrane-bound protein with the transmembrane domain at the N-terminus.

Four isoforms of NDST have been identified in human and mouse.⁴⁴⁻⁴⁷ The isoforms can be divided into two subgroups: NDST-1 and NDST-2 with 70% identities in amino acid sequence, and NDST-3 and NDST-4 with 80.4% identity. Those four NDST isoforms have distinct tissue expression patterns. NDST-1 and NDST-2 are expressed quite uniformly in different tissues, while the expression of NDST-3 and NDST-4 occurs predominantly in the brain.⁴⁵ The NDST isoforms seem to have distinct deacetylase and sulfotransferase activities, with comparable levels of each. For example, NDST-3 has good deacetylase activity but very poor sulfotransferase activity, and NDST-4 shows the opposite with poor deacetylase activity and excellent sulfotransferase activity. The biological significance of the variations of the enzymatic activities of NDST isoforms is not fully understood, but it has been proposed that it may help generate unique sulfation patterns on HS.⁴⁵

A series of genetic studies of NDST-1 and NDST-2 have provided great insights into the physiological functions of *N*-sulfation in model animals. The *Ndst-1* null mice die shortly after birth due to severe lung failure.^{48,49} In addition to prominent defects in lung development, detailed phenotypic study of *Ndst-1*-deficient mice embryos at various developmental stages revealed they display conspicuous developmental defects at the forebrain, face, and eyes,⁵⁰ reinforcing the pleiotropic effects of NDST-1 in mammalian development. The *Ndst-2* null mice have also been generated, but in contrast to the widespread effects of the *Ndst-1* mutant; the only obvious defect is in connective tissue-type mast cell.^{51,52} These cells are generally active during inflammatory processes, and they release various inflammation mediators such as heparin, histamine, and mast cell protease. However, in the mast cells of the NDST-2-deficient mice, production of heparin is aborted, the amount of histamine release is greatly reduced, and mast cell proteases are lacking. These data suggest that NDST-2 plays a predominant role in mast cells in terms of HS sulfation. Conversely, since NDST-2-deficient mice develop normally, it seems that NDST-2 is dispensable in tissues where NDST-1 may play more dominant roles. Recently, NDST-1 has been knocked out in a tissue-specific manner in endothelial cells and leukocytes, which helped demonstrate the essential roles of HS in inflammatory responses.⁵³

In comparison to the largely elucidated structure and function of the *N*-sulfotransferase domain, the function of the N-terminal region of NDSTs is apparently understudied. The N-terminal region, about 600 amino acid

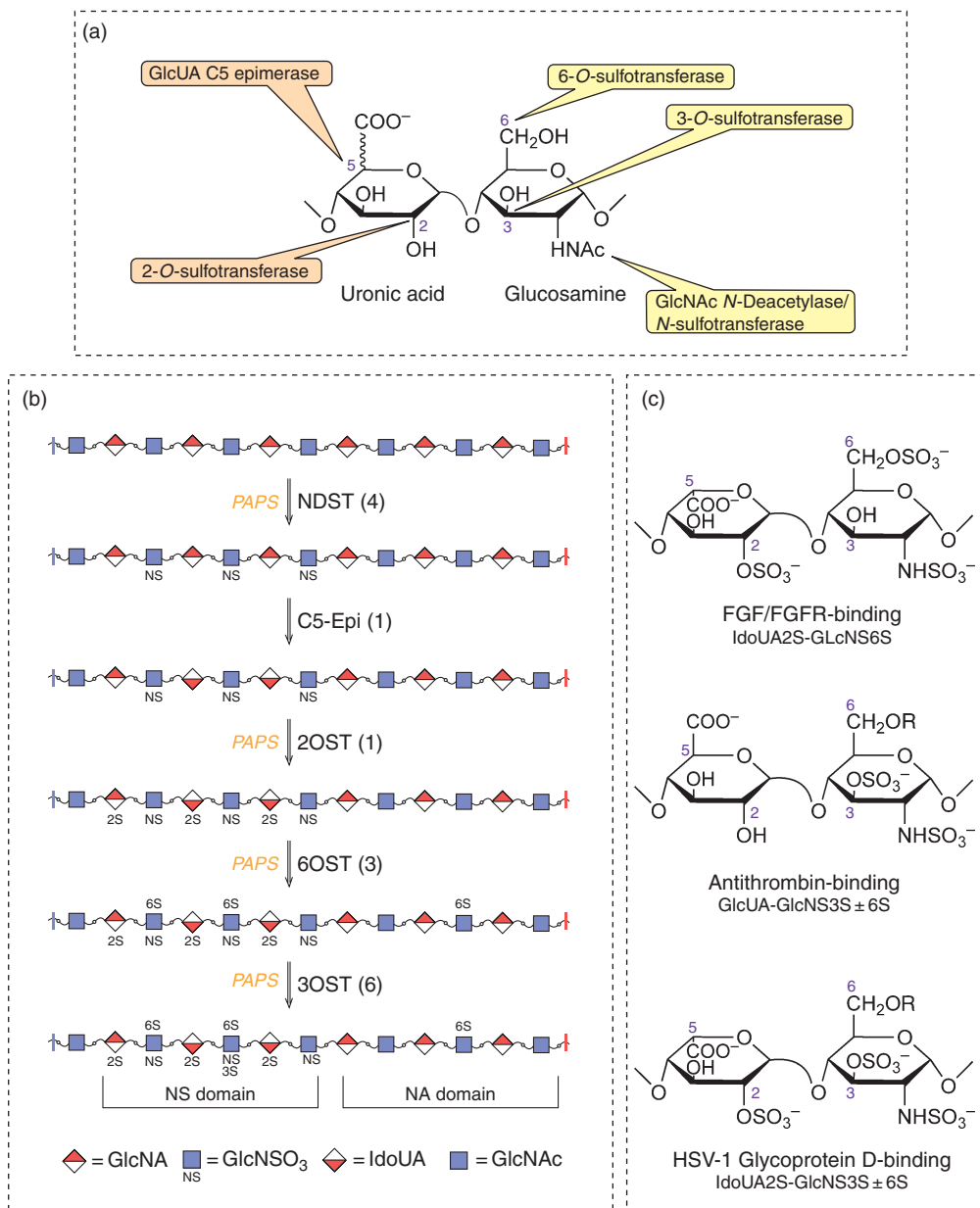


Figure 4 Schematic view of the HS modifications. (a) The HS modification enzymes and their respective modification sites are shown on the disaccharide unit. (b) HS modification pathway. The number of isoforms in human of each enzyme is shown in the parentheses after the name of the enzyme. The sequence of the modifications shown only represents a general guide, and it is important to note that the HS modification mode *in vivo* may display some variations from the main theme. (c) Representative disaccharide structures with specific biological functions.

residues in length, contains the *N*-deacetylase domain.⁴³ However, we have extremely limited knowledge of the deacetylase domain because it has no homology with the other carbohydrate deacetylases. Thus, the mechanism for deacetylation is still elusive. Understanding the deacetylation mechanism of NDSTs would be very informative, considering that varying NDST isoforms have distinct *N*-deacetylase activities. The action of different NDSTs could implant information for the subsequent enzymatic modifications, which ultimately decide the overall structure of HS.

6.14.2.2.4(ii) Glucuronyl C5-epimerase The first modification that happens on the GlcUA residue is catalyzed by HS C5-epi, which converts the configuration of proton of the C5 position, thus generating IdoUA unit (**Figure 4(a)**). A GlcNS is required at the nonreducing end of the GlcUA for the action of C5-epi, suggesting that the C5-epimerization rigorously follows *N*-sulfation (**Figure 4(b)**).¹ Switching from GlcUA to IdoUA has been suggested to give HS a more flexible structure because it is known that IdoUA adopts several pyranose ring conformations, including ¹C₄, ⁴C₁, and ²S₀, while GlcUA usually exists in ⁴C₁ conformation.⁵⁴ In addition, IdoUA is a much more favorable substrate for HS 2-OST, and the resultant IdoUA2S is essential for many biological functions.^{55–58}

Different from most other HS modification enzymes, only a single isoform of C5-epi is present in almost all species examined (with the exception of fish). C5-epi is a fairly large protein with over 600 amino acids. The epimerase catalytic domain has been tentatively assigned to the 220–230 residues of the C-terminus based on alignment to homologous proteins and biochemical studies.⁵⁹

C5-epi has been shown to form a complex with HS 2-OST *in vivo*, which is indispensable for correct Golgi localization of C5-epi.²⁶ Formation of the complex may have a substantial effect on the enzymatic activity of C5-epi. The common substrate for C5-epi and HS 2-OST, GlcUA, may be channeled through the enzyme complex to achieve a higher efficiency of modification, as it is known that IdoUA2S is no longer a substrate for C5-epi.^{26,60} The final ratio of the GlcUA/IdoUA presented in the HS may be actively regulated by this reaction.

The knockout mouse model of C5-epi has provided significant information about the physiological functions of the enzyme.⁶¹ C5-epi-deficient mice died immediately after birth due to respiratory failure, apparently as a result of poorly developed lungs. Other evident phenotypes include undeveloped kidneys, eye defects, various skeletal abnormalities such as shorter body length, and excessive mineralization. Close examination of HS from *C5-epi*^{-/-} mice shows that production of IdoUA is totally abrogated, while the proportion of GlcUA2S increased threefold and the 6-*O*-sulfation increased fivefold. The phenotypic differences between the *C5-epi*^{-/-} and *NDST-1*^{-/-} mice suggest that C5-epimerization plays unique roles in animal development, and IdoUA or IdoUA2S are specifically needed for certain cellular signaling processes.

A recent study pointed out that human C5-epi is transcriptionally regulated by the β -catenin-TCF4 transactivation complex, the terminal effector of the Wnt/ β -catenin signaling pathway.⁶² This study showed that there are two β -catenin-TCF4 *cis*-acting elements located in the enhancer region of the promoter; it also showed that ectopic expression of β -catenin and TCF4 together significantly enhances the cellular level of C5-epi transcript along with its enzymatic activity. Based on the evidence that HS is an important regulator of the interaction between Wnt and its cell surface receptor Fz (Frizzled), HS biosynthesis and Wnt signaling may be independent. It raised the possibility that a feedback loop is present between an essential signaling pathway and specific cell surface HS structures, which in turn may help achieve the final control of the animal development. Further study is needed to test the generality of this phenomenon in different cell types and in different species.

6.14.2.2.4(iii) Uronosyl 2-O-sulfotransferase 2-*O*-sulfation is the only sulfation that occurs on the uronic acid units (**Figure 4(a)**), and it is observed in HS originating from all metazoans. For instance, more than half of the IdoUA unit found in mouse embryo are 2-*O*-sulfated, while only a minute portion of the GlcUA are 2-*O*-sulfated.⁶¹ In the adult tissues of *Drosophila* and *Caenorhabditis elegans*, about 20 and 30% uronic acid units are 2-*O*-sulfated.⁶³ 2-OST is capable of reacting with both GlcUA and IdoUA units, but prefers the latter under most conditions (**Figure 4(b)**).⁶⁴

Very much like C5-epi, 2-OST has only one isoform in most species with the exception of fish. 2-OST contains a type II transmembrane domain and is localized in the Golgi apparatus. As described above, 2-OST forms a complex with C5-epi *in vivo*, which is essential for the Golgi localization of C5-epi.²⁶ However, whether complex formation is critical for the Golgi localization of 2-OST remains unclear. The complex may also influence the substrate specificities of 2-OST in HS biosynthesis *in vivo*. One can imagine that when C5-epi and 2-OST complex is formed, the majority of the resultant modification products would contain IdoUA2S. When C5-epi is dissociated from 2-OST, 2-OST may be allowed to generate GlcUA2S, and C5-epi is able to generate IdoUA without sulfation. Thus, a possible role of C5-epi is to regulate the type of uronic acid that is

subjected to 2-OST modification. Definitive demonstration of this possibility will require structural information from both 2-OST and C5-epi, and the biological significance of the interaction should be tested in animal model systems.

The HS 2-OST knockout mouse suffered from neonatal death due to renal failure.⁶⁵ The 2-OST^{-/-} mice completely fail to form kidneys, show bilateral iris coloboma, and exhibit bone shortening and overmineralization. All of these phenotypes are similar to those observed in the C5-epi^{-/-} mice, suggesting the essential roles of IdoUA2S in kidney, eye, and skeletal development. Surprisingly, 2-OST^{-/-} mice apparently develop functional lungs, a sharp contrast to the *Ndst-1* and *C5-epi* knockout mice. This difference indicates that 2-*O*-sulfation is not essential in lung development, but some IdoUA and GlcNS are required. Interestingly, close examination of the HS composition from 2-OST^{-/-} mice reveals that while 2-*O*-sulfated uronic acids are completely removed, the mice show elevated *N*-sulfation and 6-*O*-sulfation of glucosamine units. It has been proposed that the elevated sulfation at other positions of HS may attenuate the adverse effect imposed by the ablation of 2-*O*-sulfation, thus disguising some physiological functions associated with 2-*O*-sulfation.

The 2-OST null mutant has been described in *C. elegans* as well.⁶⁶ Surprisingly, 2-OST-deficient mutants are viable and fertile without any locomotory defect or embryonic paralysis as observed in perlecan-deficient worms.⁶⁷ The phenotypes of 2-OST null worms are only detectable in nervous system development by displaying axonal guidance and migration defects in a neuron-type specific manner. These findings suggest that the 2-*O*-sulfation of HS plays very specific roles in metazoan development. These roles may vary in the developmental schemes of different species.

6.14.2.2.4(iv) Glycosaminyl 6-*O*-sulfotransferase 6-*O*-sulfation of glucosamine residues is a prevalent modification of HS. The percentage of 6-*O*-sulfated disaccharide units represents 20–30% of total disaccharides in the HS from different species.^{61,63} The 6-*O*-sulfation occurs most at the NS domain, generating a GlcNS6S moiety, but it can also occur at the NA domain, generating a GlcNAc6S moiety.⁶⁸ It should be noted that 6-*O*-sulfation is the only type of sulfation that occurs at the NA domain as both 2-OST and 3-OSTs depend on the *N*-sulfation of glucosamine (**Figure 4(b)**).^{19,69}

Three isoforms of 6-OST have been identified in human and they exhibit approximately 60% identity.⁶⁸ These isoforms have highly conserved sulfotransferase domains, while varying in both N- and C-termini. Several *in vitro* substrate specificity assays suggest that 6-OST isoforms have distinct substrate preferences.^{68,70} 6-OST1 appears to favor the IdoUA-GlcNS unit significantly more than the GlcUA-GlcNS unit, while 6-OST2 prefers GlcUA-GlcNS over IdoUA-GlcNS, and 6-OST3 has almost equal activities toward both disaccharide structures. Three isoforms of 6-OST are also present in other mammals and birds as well as amphibians, while invertebrates normally have only a single isoform of 6-OST.

Unlike the previously mentioned HS modification enzymes, a mouse knockout model for 6-OST has not been reported. Nonetheless, evidence from 6-OST null *Drosophila* and *C. elegans* suggests that 6-OST plays significant roles in animal developmental processes.^{57,66,71,72} Knocking out 6-OST disrupts the development of the tracheal system; the *Drosophila* analogue of the vertebrate vascular system is lethal. Strong evidence indicates that the phenotype is directly related to the essential function of 6-*O*-sulfated HS in fibroblast growth factor (FGF) signaling.⁵⁷ Interestingly, *C. elegans* 6-OST null mutants are viable and fertile, suggesting that 6-*O*-sulfation of HS is dispensable for worm FGF signaling. However, very specific axonal guidance defects were seen in a neuron-type specific manner. The effects of worm 6-OST are believed to be related to several essential pathways in nervous system development, such as the Slit/Robo system and the Kal-1 system.^{66,72}

A morpholino-mediated functional knockdown study targeting zebrafish 6-OST, which has 58% identity to human 6-OST1 and 6-OST2, provided a glimpse of the essential functions of 6-OST in vertebrate morphogenesis.⁷³ The morpholino treatment decreased the overall 6-*O*-sulfation level by merely 20%, yet several obvious phenotypes were observed. The most notable defects of 6-OST knockdown zebrafish embryo were seen in somite development, as the embryo showed undifferentiated muscles and muscle degeneration. Additional defects caused by 6-OST knockdown were also identified in the brain and the fins. It is now known that zebrafish have at least another 6-OST isoform, which is very homologous to human 6-OST1 (81% identity), in addition to the 6-OST that is knocked down in this study, which suggests that observed phenotypes may only represent a small fraction of the real impact of 6-*O*-sulfation of HS on zebrafish embryonic development.

6.14.2.2.4(v) Glucosaminyl 3-O-sulfotransferase 3-O-sulfation of glucosamine unit is generally considered the last step in the HS biosynthetic pathway, and the rarest one. 3-O-sulfation occurs predominantly in the highly sulfated regions of HS, as 3-OST-modified disaccharide units often contain multiple sulfations (Figure 5).⁵ Modifications by 3-OST dictate unique functions of HS. For instance, 3-O-sulfation is essential for the anticoagulant activity of HS, manifesting that this single modification increases the affinity to antithrombin by 18 000-fold.⁷⁴ As a result of extensive study, 3-O-sulfation of HS has been directly associated with many other biological processes such as viral infection,⁵⁶ carcinogenesis,⁷⁵ diurnal rhythm control,⁷⁶ as well as lower organism morphogenesis.⁷¹

3-OST represents the most extended gene family among all HS sulfotransferases. Vertebrates generally contain seven isoforms, while invertebrates tend to have two. The vertebrate 3-OSTs can be divided into two subgroups based on the homology of the sulfotransferase domain. 3-OST2, 3-OST3 (including 3A and 3B), 3-OST4, and 3-OST6 form one subgroup, the so-called 3-OST3-like subgroup, with higher than 70% identity; whereas 3-OST1 and 3-OST5 form another subgroup, the so-called 3-OST1-like subgroup, with 53% of identity. The 3-OST1-like and 3-OST3-like subgroups exhibit approximately 45% identity in the sulfotransferase domain. Two 3-OST isoforms, Hs3st-A and Hs3st-B, are identified in *Drosophila*.⁷¹ Interestingly, Hs3st-A is more homologous to human 3-OST1-like subgroup, while Hs3st-B is more homologous to 3-OST3-like subgroup. Similar two 3-OST isoforms were found in *C. elegans*. These observations suggest that it is likely that 3-OST diverged into two branches early in the evolution to perform separate functions, and both branches expanded in vertebrates to carry out more diverse and specific roles.

In comparison to the promiscuous substrate specificities of 6-OSTs, members of 3-OST family have distinctive substrate specificities. 3-OST1 synthesizes a characteristic disaccharide unit with the structure of GlcUA-GlcNS3S±6S (Figure 5),⁷⁷ which is a critical modification step for synthesizing anticoagulant HS. 3-OST2, 3-OST3, 3-OST4, and 3-OST6 all generate a characteristic disaccharide unit with a structure of IdoUA2S-GlcNS (or NH₂)3S6S (Figure 5),^{78–80} which is part of the structure of HS serving as an entry receptor for herpes simplex virus 1. 3-OST5 is unique in that it enables to synthesize these two disaccharide units.⁸¹ The distinct disaccharide units generated by 3-OST confer the functions of HS.

Among seven isoforms of 3-OST, only 3-OST-1 has been knocked out in mice.⁸² The study was originally aimed to determine the effect of the level of anticoagulant HS on the blood coagulation process, provided that 3-OST-1 is a critical enzyme for synthesizing anticoagulant HS. However, the 3-OST-1^{-/-} mice exhibit a surprisingly normal anticoagulation phenotype despite the fact the level of anticoagulant HS synthesis was reduced significantly in various organs. The authors concluded that the anticoagulant HS is not a major

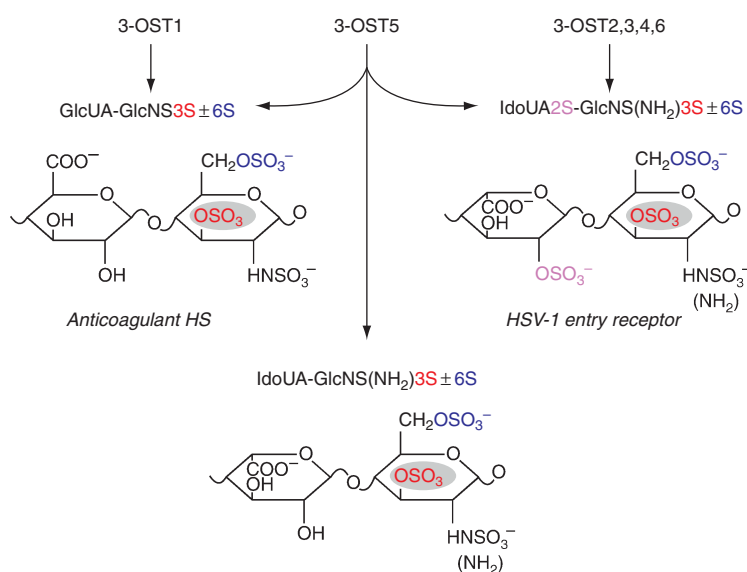


Figure 5 Substrate specificities of 3-OSTs.

hemostatic regulator regardless of the potent anticoagulant property of heparin drug. Nevertheless, 3-OST-1^{-/-} mice exhibit a genetic background-specific postnatal lethality. The cause of the lethality is unknown, yet is clearly distinct from the lethality caused by antithrombin^{-/-} mice. This finding suggests that 3-OST-1 modification of HS plays some profound roles in other biological processes although the details of this involvement remain unclear.

Studies in *Drosophila* provided the first direct evidence of the critical roles played by 3-OST in animal development.⁷¹ When *Drosophila* Hs3st-B was silenced by RNAi treatment, several morphological defects were observed. These included server rough eye phenotype, extra sensory bristles at multiple tissues, and profound defects in wing morphology such as reduced wing size, abnormal wing veins, and notching at the wing margin. The phenotypic effects of Hs3st-B silencing have been related to Notch signaling directly, as Hs3st-B was shown to influence the protein level of Notch (cell surface receptor in Notch signaling). The exact role of Hs3st-B-modified HS in the Notch signaling pathway remains to be determined.

6.14.2.3 Physiological and Pathophysiological Importance

Owing to HSPGs locales, they have been shown to be essential in many regulatory functions. Cell surface HSPGs have been shown to be coreceptors for various signaling molecules as well as play roles in cell to matrix adhesion, lipoprotein metabolism, both viral and parasitic infections, modulation of protease activity, and modulation in early embryonic development.^{83,84} There is also suggestion that binding to the HSPG scaffold can protect proteins from proteolytic degradation.⁸⁵ There are a few main pathways where HS/heparin exhibits a key regulatory function, such as the anticoagulation pathway, viral entry, and cell proliferation. HS's function is determined by its sulfation pattern and size.

6.14.2.3.1 Anticoagulation

Heparin, a special form of HS, has been used as an anticoagulant drug for decades. Heparin is an exclusive product of mast cells, while HS is expressed ubiquitously. The structures of heparin and HS are similar; however, heparin has higher contents of IdoUA and more sulfo groups per disaccharide unit. Heparin achieves the anticoagulant activity by serving as a catalyst to accelerate the interaction of antithrombin and factor IIa or factor Xa to exert anti-IIa and anti-Xa activities. Heparin and low molecular weight (LMW) heparin remain the drug of choice for the treatment of arterial and venous thrombotic disorders.⁸⁶ The major side effect of heparin drug is heparin-induced thrombocytopenia (HIT) because the binding of heparin to platelet factor 4 (PF4) leads to the aggregation of platelet.

Clinical studies have clearly demonstrated the benefits of treating cancer patients using heparin, especially LMW heparin.⁸⁷ The association between thrombotic diseases and cancer has been known to oncologists for a long time. Administration of LMW heparin has improved the survival of the patients with small cell lung cancer from 29.5 to 51.3%, while warfarin (another commonly used anticoagulant drug) did not exhibit anticancer activity.⁸⁸ In addition, treatment of advanced breast cancer patients with a combined chemotherapy with LMW heparin has reversed or stabilized the progression of the tumors.⁸⁹ Although the molecular mechanism for the link of cancer and coagulation is not completely clear, it is believed that the saccharide structures exhibiting the anticancer activity are different from those carrying anticoagulant activity as determined in a mice model.^{90,91}

6.14.2.3.2 Antiviral activity

HSPG can aid in herpes simplex viral entry (HSV-1) via the interaction with HS side chain. Thus, using a soluble HS or heparin to compete with cell surface HSPG could block HSV-1 infection. Although the exact mechanism of HSV-1 entry is not fully understood, it is known that entry via viral glycoproteins B, C, and D is due to an interaction with cell surface HS. The interaction of viral gB and gC with HS is important in cell targeting.⁹²⁻⁹⁵ Viral gD, in association with gB, gH, and gL, are involved in fusion with the cell via cell surface receptors.⁹⁶ HS modified by 3-OST-3 binds to gD and serves as an entry receptor for HSV-1. Given the fact that 3-*O*-sulfated glucosamine is rare in HS, this unique 3-*O*-sulfated HS could serve as a specific compound to inhibit HSV-1 infection. A glycoprotein D-binding octasaccharide has been characterized as $\Delta\text{UA-GlcNSIdoUA2S-GlcNAc-UA2S-GlcNS-IdoUA2S-GlcNH}_2\text{3S6S}$ and has a K_d of $19\ \mu\text{mol l}^{-1}$.⁹⁷ Further, Copeland and colleagues⁹⁸ demonstrated that the inhibition of HSV-1 entry was achieved by using

a 3-*O*-sulfated heparin octasaccharide that mimics the structure of the gD-binding octasaccharide, which demonstrated for the first time that a specific HS oligosaccharide can be used to inhibit HSV-1 infection.

6.14.2.3.3 Inflammation

In response to tissue injury or pathogen infections, the multistep process that is involved in recruiting chemokines is initiated. Endothelial cell surface HSPG has been demonstrated to play an essential role in the inflammatory response.¹³ It has been shown that chemokines bind to heparin because most chemokines are highly positively charged polypeptides, which contain patches of basic residues at heparin- and HS-binding sites.⁹⁹ The chemokine and HS interaction are believed to be essential for the presence of chemokines on the luminal surface of endothelial cells. Furthermore, HS binds to cell adhesion molecules (such as L- and P-selectins), growth factors, and growth factor receptors to regulate nearly every stage of leukocyte migrations through the blood vessel wall. The sulfation patterns of endothelial HS direct the movement of neutrophil via the interactions of L-selectin and chemokines.⁵³ The heparin-carrying 6-*O*-sulfo groups exhibit anti-inflammatory effects by blocking the binding of HS and L- and P-selectins.¹⁰⁰ It is conceivable to develop HS-based anti-inflammatory agents.¹³

6.14.3 Chondroitin Sulfate and Dermatan Sulfate

Chondroitin sulfate (CS) and dermatan sulfate (DS) comprise another class of glycosaminoglycans, which have been implicated in various biological activities, including cell signaling, coagulation, cartilage function, neurobiology, and interaction with pathogen receptors.¹⁰¹ These GAGs can be found on the cell surface or in the extracellular matrix like HS.¹⁰²

However, these polysaccharides differ from HS and heparin in both disaccharide repeating sequences and function. Structurally, CS and DS consist of a repeating disaccharide of galactosamine linked to a hexuronic residue, either iduronic or glucuronic. These linkages alternate between $\beta 1,3$ and $\beta 1,4$ as shown in **Figure 6**. CS and DS can be denoted as $(\text{GalNAc}\beta 1,4\text{GlcUA}\beta 1,3)_n$ or $(\text{GalNAc}\beta 1,4\text{IdoUA}\alpha 1,3)_m$, respectively.¹⁰³ The major distinction of CS and DS is that CS contains glucuronic residues while DS contains varying numbers of iduronic residues.

6.14.3.1 Localization of CSPG and DSPG

CS and DS exist as proteoglycans within the extracellular matrix, but the proteins to which they are linked differ. The most studied CSPGs are all members of the lectican family; whereas, dermatan sulfate proteoglycan (DSPG) protein cores vary (**Table 1**). The lectican family includes the proteins aggrecan, versican, brevican, and neurocan, which can exist with or without GAG attachment. The most studied DSPGs are biglycan and decorin.^{104,105}

6.14.3.1.1 The lectican family

The lectican family exhibits domain structural similarity. Their protein size varies between 80 and 400 kD. The N-terminal and C-terminal of all lecticans are connected by a linkage region. This region is highly homologous, and carries CS side chains. The C-terminal domain contains a C-type lectin domain as well as epidermal growth factor (EGF) and complement regulatory protein (CRP) like binding domains; while the N-terminal domain is similar to domains for the binding of HA.¹⁰⁶

Aggrecan is located within cartilage and the nervous system and is about 220 kDa. The domain structure of aggrecan consists of an N-terminal HA-binding domain, followed by a cleavage accessible point, a seemingly nonfunctional domain and the C-terminal protein-interacting domain. This proteoglycan can possess both CS and KS chains.¹⁰⁷ The number of CS chains can reach to about 100–150 while the number of KS chains is much less, around 15–50.¹⁰⁸

Versican is found within fibroblasts, the nervous system, smooth muscles, and the kidney. It is about 400 kDa in size.¹⁰⁷ This proteoglycan contains CS chains and follows the general lectican family domain scheme.¹⁰⁹ Versican can exist in four different splice variants, which vary in both size and the number of CS chains

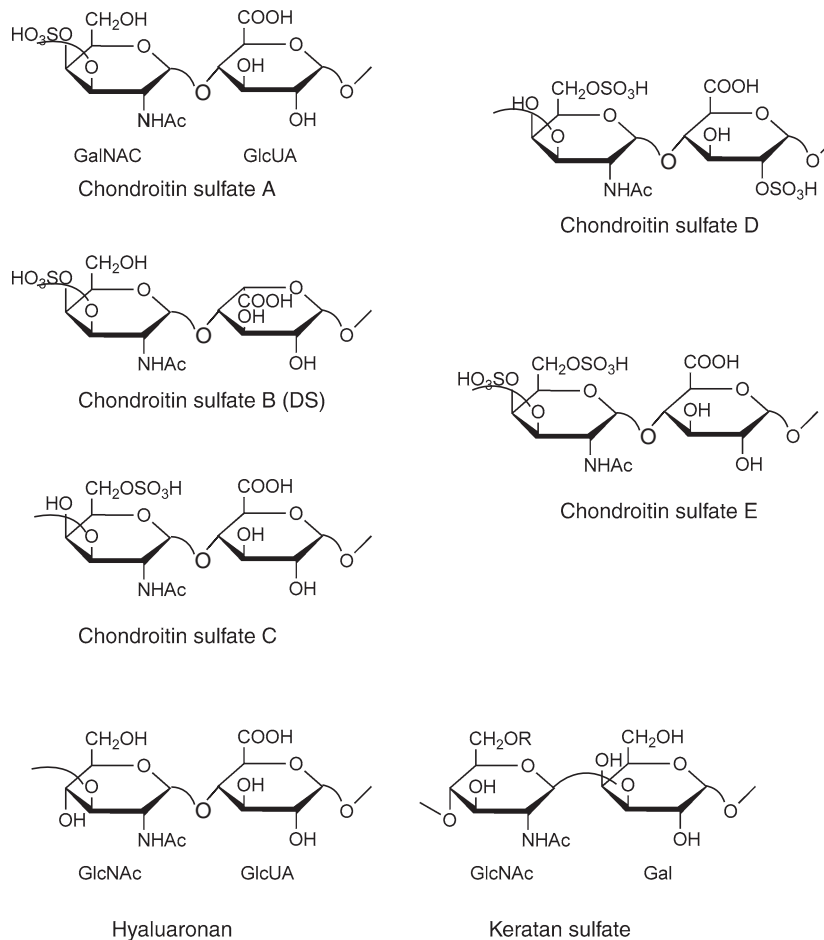


Figure 6 Structures of the disaccharide repeating units of chondroitin sulfate (CS), hyaluronan (HA), and keratan sulfate (KS).

attached. One variant however does not contain sites for CS attachment.¹¹⁰ The range for attachment sites, disregarding the non-GAG-producing chain, ranges from 10 to 30.¹¹¹

Neurocan and brevican are both found within the nervous system. Brevican has little sequence homology with the other lecticans of CSPGs and has been shown to exist as a part-time PG in the brain.^{112,113}

6.14.3.1.2 DS proteoglycans

The protein cores of DSPGs include, but are not limited to, decorin and biglycan (Table 1). These two small leucine-rich proteins are the most studied DSPGs. They are involved in ligament function and have been studied in relation to osteoarthritis.¹¹⁴ Decorin is the primary proteoglycan in ligaments.¹¹⁵ Decorin only has one site for DS attachment whereas biglycan possesses two sites.¹¹⁶ Decorin has been shown to interact with fibrils of its specific tissue, while biglycan has not been shown to have this function except under certain *in vitro* conditions.¹¹⁷

6.14.3.2 Physiological Importance of Chondroitin Sulfate and Dermatan Sulfate

Both CS and DS have been shown to exhibit multiple protein and cellular interactions, including but not limited to fibroblast growth factors, hepatocyte growth factor, tenascin-X, and heparin cofactor II.^{117,118} With all of these cellular interactions, the possible function of these proteoglycans is vast.

6.14.3.2.1 Anticoagulation

DS has been shown to modulate the activity of heparin cofactor II to inhibit thrombin without interacting with other factors of the coagulation pathway.¹¹⁸ It however has a lower affinity for HCII than heparin. DS also can bind activated protein C (APC), which can inhibit factor V.¹⁰⁴

6.14.3.2.2 Cartilage function

Both CSPGs and DSPGs have a role in cartilage and ligament function. Aggrecan, versican, decorin, and biglycan are located in the aforementioned areas and play a role in the stability of these regions.¹¹⁵

6.14.3.2.3 Neurobiological effects

All of the CSPGs can be found within the neuronal system and have been shown to work in concert in many cases. As a whole, they are known for their growth-inhibitory qualities.¹¹⁹ In normal cells, brevican has been shown to hinder cell and neurite movement.¹²⁰ However, upregulation of and increased cleavage of this PG can lead to increased glioma incursion.¹²¹

6.14.3.2.4 Pathogen receptors

CS chains have been implicated in viral entry. CS is able to bind glycoprotein C (gC) via the E unit and was shown to inhibit herpes simplex viral entry.¹²² However, in 2006, the Kitagawa group demonstrated the importance of chondroitin 4-*O*-sulfotransferase 1 (CS4ST-1) sulfation of the E disaccharide in infection by HSV.¹²³ They showed that expression of CS4ST-1 increased the size and sulfation of CS-E, which led to its increased receptiveness to HSV.¹²³

6.14.4 Other Glycosaminoglycans

HA and KS also belong to the GAG family. HA is composed of alternating $\beta 1 \rightarrow 4$ and $\beta 1 \rightarrow 3$ linkages between glucosamine and glucuronic residues. The only variability of these chains lies within their chain length, as HA does not undergo sulfation. The size of HA is between 1 and 10 MDa.¹²⁴ Along with HS, it comprises a large portion of the pericellular matrix, and has been shown to, in some cases, act as an extracellular matrix (ECM) in the absence of other ECM components.^{125,126}

HA is important for the structure and function of the pericellular matrix. It has been implicated in stability, shape, proliferation, movement, and biological responses.¹²⁷ Unlike other GAGs, HA is not synthesized as a proteoglycan, but can be attached to its synthase or other cell surface receptors, such as CD44 and RHAMM. Free HA has been shown to exhibit different conformations. In solution, HA forms largely hydrated coils.¹²⁷ HA can also form fibers. The fibrous coat can be greater than 20 μm thick. This ability to form fibers or networks aids in the functions of HA. For instance, it contributes to the cell's overall motility. It is also a molecular sieve, as while it constricts the movement of macromolecules, smaller molecules can pass through the matrix. In cartilage, the interaction of CD44 and HA is important for the mediation of homeostasis by aiding in metabolism and the survival or killing of chondrocytes.¹²⁸

HA can also associate with other PGs such as the lectican. These interactions are able to prevent the long HA chains from coiling.¹²⁷ In cartilage, this interaction is important for aggregating and immobilizing these PGs, such as aggrecan. These PGs are essential for the robustness of cartilage.¹²⁸ Within the bone, the specific function of HA is unknown, but it is known to have an influence on structural remodeling through osteoblasts and osteoclasts.

KS is a GAG that can be found in both the cornea and cartilage. Its repeating disaccharide consists of alternating $\beta 1,3$ and $\beta 1,4$ linkages between galactose and *N*-acetylglucosamine, denoted Gal($\beta 1,4$)GlcNAc($\beta 1$) and can be either N- or O-linked depending on its location (shown in **Figure 6**).^{129–131} The size of the polysaccharide chain is usually between 2 and 40 disaccharides in length depending upon the amount of sulfation. The structure of this GAG differs from the aforementioned GAGs in two ways: it does not contain an uronic residue and only has one type of sulfation, namely, sulfation at the C6 position of either or both units of a disaccharide.¹²⁹

The KSs can be divided into three distinct groups based upon their linkages to their PG. Class I (KSI) is N-linked via asparagine, class II is O-linked to serine/threonine via galactosamine, and class III is O-linked to serine via mannose.^{132,133} KS-linked proteoglycan have been found throughout the body, including the brain, bone, cartilage, and cornea. Originally, it was thought that there were only two classes of KS, I and II, and these classes were distinguished by location. KSI was said to only be in the cornea while KSII was located in the cartilage. Later, the classifications were changed to be based upon the linkage of the KS chain to its protein. Also, a third group, KSIII, was added.^{131,133}

KSI are N-linked via Asn to their core protein. These chains differ from tissue to tissue. Mainly, the KSI in body tissues is overall shorter and more sulfated than KSI of the cornea.¹³¹ KSII is solely O-linked via Ser/Thr to aggrecan and is a relatively short chain. KSIII belongs to O-linked via Ser to its core protein. KSIII differs from KSII due to the mannose unit preceding the chain.^{131,133}

In the cornea, KS is involved in hydration due to its water-binding properties.¹³² Deficiency of this GAG has also been associated with corneal transparency.¹³⁴ However, in other body tissues and organs, the hydration properties do not explain its presence.¹³² For example, KS chains are attached to fibromodulin (fibromodulin has four KS attachment sites) in bone and cartilage and aids in its mobility but also aids in the regulation of the strength of tendons.¹¹¹ Also, KSPGs within the brain have been shown to modulate the growth of mossy fibers.¹³⁵ KS can be attached to various PGs associated with other GAGs. These attachment sites can be few or many. For example, there is one KS attachment site in decorin, but 60 in aggrecan. There are also some proteins only associated with KS, which include fibromodulin (four sites) and lumican (two or three sites).¹¹¹

Acknowledgments

This work is supported by an NIH grant (AI50050). Courtney L. Jones is supported by an NIH diversity supplement grant (AI50050-07S). Ding Xu is a recipient of a predoctoral fellowship from American Heart Association, MidAtlantic Affiliate.

Abbreviations

| | |
|---------------|------------------------------------|
| 2-OST | 2-O-sulfotransferase |
| 3-OST | 3-O-sulfotransferase |
| 6-OST | 6-O-sulfotransferase |
| C5-epi | C5-epimerase |
| CS | chondroitin sulfate |
| CS4ST | chondroitin 4-O-sulfotransferase 1 |
| CS-E | chondroitin sulfate E |
| CSPG | chondroitin sulfate proteoglycan |
| DS | dermatan sulfate |
| DSPG | dermatan sulfate proteoglycan |
| EXT | extosin family |
| HS | heparan sulfate |
| HSPG | heparan sulfate proteoglycan |
| GAG | glycosaminoglycan |
| Gal | galactose |
| GalTI | galactosyltransferase I |
| GalTII | galactosyltransferase II |
| gB | HSV-1 glycoprotein B |
| gC | HSV-1 glycoprotein C |
| gD | HSV-1 glycoprotein D |
| gH | HSV-1 glycoprotein H |
| gL | HSV-1 glycoprotein L |
| GlcATI | glucuronosyltransferase I |

| | |
|-------------------|--|
| GlcNAc | <i>N</i> -acetyl glucosamine |
| GlcNAcTI | heparan sulfate <i>N</i> -acetylglucosamine transferase I |
| GlcNAcTII | heparan sulfate <i>N</i> -acetylglucosamine transferase II |
| GlcNAcT-P | promiscuous glucuronosyltransferase |
| GlcUA | glucuronic acid |
| HA | hyaluronic acid (hyaluronan) |
| IdoUA | iduronic acid |
| IdoUA2S | 2- <i>O</i> -sulfated iduronic acid |
| KS | keratan sulfate |
| KSPG | keratan sulfate proteoglycan |
| NA | <i>N</i> -acetylated domain |
| NDST | <i>N</i> -deacetylase/ <i>N</i> -sulfotransferase |
| NS | <i>N</i> -sulfated domain |
| PAPS | 3' phosphoadenosine 5' phosphosulfate |
| UDP-Gal | UDP-galactose |
| UDP-Glc | UDP-glucose |
| UDP-GlcNAc | UDP- <i>N</i> -acetyl glucosamine |
| UDP-GlcUA | UDP-glucuronic acid |
| UDP-Xyl | UDP-xylose |
| Xyl | xylose |
| XylIT/XT | xylotransferase |

References

1. J. D. Esko; U. Lindahl, *J. Clin. Invest.* **2001**, *108*, 169–173.
2. P. L. DeAngelis, *Glycobiology* **2002**, *12*, 9R–16R.
3. R. V. Iozzo; J. D. San Antonio, *J. Clin. Invest.* **2001**, *108*, 349–355.
4. R. Sasisekharan; Z. Shriver; G. Venkataraman; U. Narayanasami, *Nat. Rev. Cancer* **2002**, *2*, 521–528.
5. J. Liu; S. C. Thorp, *Med. Res. Rev.* **2002**, *22*, 1–25.
6. B. Casu; U. Lindahl, *Adv. Carbohydr. Chem. Biochem.* **2001**, *57*, 159–206.
7. H. B. Nader; S. F. Chavante; E. A. dos-Santos; F. W. Oliveira; J. F. de-Paiva; S. M. B. Jerônimo; G. F. Medeiros; L. R. D. de-Abreu; E. L. Leite; J. F. de-Sousa-Filho; R. A. B. Castro; L. Toma; I. L. S. Tersariol; M. A. Porcionatto; C. P. Dietrich, *Braz. J. Med. Biol. Res.* **1999**, *32*, 529–538.
8. G. F. Medeiros; A. Mendes; R. A. B. Castro; E. C. Baú; H. B. Nader; C. P. Dietrich, *Biochim. Biophys. Acta* **2000**, *1475*, 287–294.
9. C. W. Dunn; A. Hejnl; D. Q. Matus; K. Pang; W. E. Browne; S. A. Smith; E. Seaver; G. W. Rouse; M. Obst; G. D. Edgecombe; M. V. Sorensen; S. H. D. Haddock; A. Schmidt-Rhaesa; A. Okusu; R. M. Kristensen; W. C. Wheeler; M. Q. Martindale; G. Giribet, *Nature* **2008**, *452*, 745–749.
10. D. J. Bornemann; S. Park; S. Phin; R. Warrior, *Development* **2008**, *135* (6), 1039–1047.
11. R. V. Iozzo, *J. Clin. Invest.* **2001**, *108*, 165–167.
12. R. V. Iozzo, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 646–656.
13. C. R. Parish, *Nat. Rev. Immunol.* **2006**, *6*, 633–643.
14. A. L. W. M.M. Rops; J. van der Vlag; J. F. M. Lensen; T. J. M. Wijnhoven; L. P. W. J. van den Heuvel; T. H. van Kuppevelt; J. H. M. Berden, *Kidney Int.* **2004**, *65*, 768–785.
15. J. Chen; C. L. Jones; J. Liu, *Chem. Biol.* **2007**, *14*, 986–993.
16. J. Liu; L. C. Pedersen, *Appl. Microbiol. Biotechnol.* **2007**, *74*, 263–272.
17. P. M. Bernisone; C. B. Hirschberg, *Curr. Opin. Struct. Biol.* **2000**, *10*, 542–547.
18. H. Fuda; C. Shimizu; Y. C. Lee; H. Akita; C. A. Strott, *Biochem. J.* **2002**, *365*, 497–504.
19. J. D. Esko; S. B. Selleck, *Ann. Rev. Biochem.* **2002**, *71*, 435–471.
20. C. Gotting; J. Kuhn; R. Zahn; T. Brinkmann; K. Kleesiek, *J. Mol. Biol.* **2000**, *304*, 517–528.
21. J. Kuhn; C. Gotting; M. Scholzer; T. Kempf; T. Brinkmann; K. Kleesiek, *J. Biol. Chem.* **2001**, *276*, 4940–4947.
22. S. Schön; C. Prante; C. Bahr; J. Kuhn; K. Kleesiek; C. Götting, *J. Biol. Chem.* **2006**, *281*, 14224–14231.
23. L. Zhang; G. David; J. D. Esko, *J. Biol. Chem.* **1995**, *270*, 27127–27135.
24. R. Almeida; S. B. Lavery; U. Mandel; H. Kresseparallel; T. Schwientek; E. P. Bennett; H. Clausen, *J. Biol. Chem.* **1999**, *274*, 26165–26171.
25. X. Bai; D. Zhou; J. R. Brown; B. E. Crawford; T. Hennen; J. D. Esko, *J. Biol. Chem.* **2001**, *276*, 48189–48195.
26. M. A. Pinhal; B. Smith; S. Olson; J. Aikawa; K. Kimata; J. D. Esko, *Proc. Natl. Acad. Sci.* **2001**, *98*, 12984–12989.

27. L. C. Pedersen; K. Tsuchida; H. Kitagawa; K. D. Sugahara; T. A. Darden; M. Negishi, *J. Biol. Chem.* **2000**, *275*, 34580–34585.
28. S. Gulberti; V. Lattard; M. Fondeur; J.-C. Jacquinet; G. Mulliert; P. Netter; J. Magdalou; M. Ouzzine; S. Fournel-Gigleux, *J. Biol. Chem.* **2005**, *280*, 1417–1425.
29. K. Sugahara; H. Kitagawa, *Curr. Opin. Struct. Biol.* **2000**, *10*, 518–527.
30. G. Wei; X. Bai; A. K. Sarkar; J. D. Esko, *J. Biol. Chem.* **1999**, *274*, 7857–7864.
31. J. D. Esko; L. Zhang, *Curr. Opin. Struct. Biol.* **1996**, *10*, 542–547.
32. H. Kitagawa; H. Shimakawa; K. Sugahara, *J. Biol. Chem.* **1999**, *274*, 13933–13937.
33. B. T. Kim; H. Kitagawa; J. Tamura; T. Saito; M. Kusche-Gullberg; U. Lindahl; K. Sugahara, *Proc. Natl. Acad. Sci.* **2001**, *98*, 7176–7181.
34. J. S. Lee; S. von der Hardt; M. A. Rusch; S. E. Stringer; H. L. Stickney; W. S. Talbot; R. Geisler; C. Nüsslein-Volhard; S. B. Selleck; C. B. Chien; H. Roehl, *Neuron* **2004**, *44*, 947–960.
35. C. McCormick; Y. Leduc; D. Martindale; K. Mattison; L. E. Esford; A. P. Dyer; F. Tufaro, *Nat. Genet.* **1998**, *19*, 158–161.
36. T. Lind; F. Tufaro; C. McCormick; U. Lindahl; K. Lidholt, *J. Biol. Chem.* **1998**, *273*, 26265–26268.
37. C. McCormick; G. Duncan; T. Goutsos; F. Tufaro, *Proc. Natl. Acad. Sci.* **2000**, *97*, 668–673.
38. S. Kobayashi; K. Morimoto; T. Shimizu; M. Takahashi; H. Kurosawa; T. Shirasawa, *Biochem. Biophys. Res. Commun.* **2000**, *268*, 860–867.
39. B.-T. Kim; H. Kitagawa; J. Tanaka; J.-I. Tamura; K. Sugahara, *J. Biol. Chem.* **2003**, *278*, 41618–41623.
40. L. C. Pedersen; J. Dong; F. Taniguchi; H. Kitagawa; J. M. Krahn; L. G. Pedersen; K. Sugahara; N. Masahiko, *J. Biol. Chem.* **2003**, *278*, 14420–14428.
41. G. Duncan; C. McCormick; F. Tufaro, *J. Clin. Invest.* **2001**, *108*, 511–516.
42. Y. Kakuta; T. Sueyoshi; M. Negishi; L. C. Pedersen, *J. Biol. Chem.* **1999**, *274*, 10673–10676.
43. M. B. Duncan; M. Liu; C. Fox; J. Liu, *Biochem. Biophys. Res. Commun.* **2006**, *339*, 1232–1237.
44. J. Aikawa; J. D. Esko, *J. Biol. Chem.* **1999**, *274*, 2690–2695.
45. J.-i. Aikawa; K. Grobe; M. Tsujimoto; J. D. Esko, *J. Biol. Chem.* **2001**, *276*, 5876–5882.
46. I. Eriksson; D. Sanback; B. Ek; U. Lindahl; L. Kjellen, *J. Biol. Chem.* **1994**, *269*, 10438–10443.
47. Y. Hashimoto; A. Orellana; G. Gil; C. B. Hirschberg, *J. Biol. Chem.* **1992**, *267*, 15744–15750.
48. M. Ringvall; J. Ledin; K. Holmborn; T. van Kuppevelt; F. Ellin; I. Eriksson; A. M. Olofsson; L. Kjellen; E. Forsberg, *J. Biol. Chem.* **2000**, *275*, 25926–25930.
49. G. Fan; L. Xiao; L. Cheng; X. Wang; B. Sun; G. Hu, *FEBS Lett.* **2000**, *467*, 7–11.
50. K. Grobe; M. Inatani; S. R. Pallerla; J. Castagnola; Y. Yamaguchi; J. D. Esko, *Development* **2005**, *132*, 3777–3786.
51. E. Forsberg; G. Pejler; M. Ringvall; C. Lunderius; B. Tomasini-Johansson; M. Kusche-Gullberg; I. Eriksson; J. Ledin; L. Hellman; L. Kjellen, *Nature* **1999**, *400*, 773–776.
52. D. E. Humphries; G. W. Wong; D. S. Friend; M. F. Gurish; W. T. Qiu; C. Huang; A. H. Sharpe; R. L. Stevens, *Nature* **1999**, *400*, 769–772.
53. L. Wang; M. Fuster; P. Sriramarao; J. D. Esko, *Nat. Immunol.* **2005**, *6*, 902–910.
54. B. Mulloy; M. Forster, *Glycobiology* **2000**, *10*, 1147–1156.
55. J. Kreuger; M. Salmivirta; L. Sturiale; G. Gimenez-Gallego; U. Lindahl, *J. Biol. Chem.* **2001**, *276*, 30744–30752.
56. D. Shukla; J. Liu; P. Blaiklock; N. W. Shworak; X. Bai; J. D. Esko; G. H. Cohen; R. J. Eisenberg; R. D. Rosenberg; P. G. Spear, *Cell* **1999**, *99*, 13–22.
57. K. Kamimura; T. Koyama; H. Habuchi; R. Ueda; M. Masayuki; K. Kimata; H. Nakato, *J. Cell Biol.* **2006**, *174*, 773–778.
58. T. Kinnunen; Z. Huang; J. Townsend; M. M. Gatdula; J. R. Brown; J. D. Esko; J. E. Turnbull, *Proc. Natl. Acad. Sci.* **2005**, *102*, 1507–1512.
59. S. J. Charnock; I. E. Brown; J. P. Turkenburg; G. W. Black; G. J. Davies, *Proc. Natl. Acad. Sci.* **2002**, *99*, 12067–12072.
60. A. Hagner-McWhirter; H. H. Hannesson; P. Campbell; J. Westley; L. Rodén; U. Lindahl; J. P. Li, *Glycobiology* **2000**, *10*, 159–171.
61. J.-P. Li; F. Gong; A. Hagner-McWhirter; E. Forsberg; M. Åbrink; R. Kisilevsky; X. Zhang; U. Lindahl, *J. Biol. Chem.* **2003**, *278*, 28363–28366.
62. G. Ghiselli; A. Agrawal, *Biochem. J.* **2005**, *390*, 493–499.
63. H. Toyoda; A. Kinoshita-Toyoda; S. B. Selleck, *J. Biol. Chem.* **2000**, *275*, 2269–2275.
64. J. Rong; H. Habuchi; K. Kimata; U. Lindahl; M. Kusche-Gullberg, *Biochemistry* **2001**, *40*, 5548–5555.
65. S. L. Bullock; J. M. Fletcher; R. S. Beddington; V. A. Wilson, *Genes Dev.* **1998**, *12*, 1894–1906.
66. H. E. Bulow; O. Hobert, *Neuron* **2004**, *41*, 723–736.
67. T. M. Rogalski; B. D. Williams; G. P. Mullen; D. G. Moerman, *Genes Dev.* **1993**, *7*, 1471–1484.
68. H. Habuchi; M. Tanaka; O. Habuchi; K. Yoshida; H. Suzuki; K. Ban; K. Kimata, *J. Biol. Chem.* **2000**, *275*, 2859–2868.
69. J. Liu; R. D. Rosenberg, Heparan Sulfate α -glucosaminyl 3-O-sulfotransferase. In *Handbook of Glycosyltransferases and Their Related Genes*; N. Taniguchi, M. Fukuda, Eds.; Springer-Verlag: Tokyo, 2002; pp 475–483.
70. H. Habuchi; G. Miyake; K. Nogami; A. Kuroiwa; Y. Matsuda; M. Kusche-Gullberg; O. Habuchi; M. Tanaka; K. Kimata, *Biochem. J.* **2003**, *371*, 131–142.
71. K. Kamimura; J. M. Rhodes; R. Ueda; M. McNeely; D. Shukla; K. Kimata; P. G. Spear; N. W. Shworak; H. Nakato, *J. Cell Biol.* **2004**, *166*, 1069–1079.
72. H. E. Bulow; K. L. Berry; L. H. Topper; E. Peles; O. Hobert, *Proc. Natl. Acad. Sci.* **2002**, *99*, 6346–6351.
73. R. J. Bink; H. Habuchi; Z. Lele; E. Dolk; J. Joore; G.-J. Rauch; R. Geisler; S. W. Wilson; J. den Hertog; K. Kimata; D. Zivkovic, *J. Biol. Chem.* **2003**, *278*, 31118–31127.
74. D. H. Atha; J.-C. Lormeau; M. Petitou; R. D. Rosenberg; J. Choay, *Biochemistry* **1985**, *24*, 6723–6729.
75. K. Miyamoto; K. Asada; T. Fukutomi; E. Okochi; Y. Yagi; T. Hasegawa; T. Asahara; T. Sugimura; T. Ushijima, *Oncogene* **2003**, *22*, 274–280.
76. J. Borjigin; J. Deng; X. Sun; M. Jesus; T. Liu; M. W. Wang, *J. Biol. Chem.* **2003**, *278*, 16315–16319.
77. J. Liu; N. W. Shworak; L. M. S. Fritze; J. M. Edelberg; R. D. Rosenberg, *J. Biol. Chem.* **1996**, *271*, 27072–27082.
78. J. Liu; N. W. Shworak; P. Sinaý; J. J. Schwartz; L. Zhang; L. M. S. Fritze; R. D. Rosenberg, *J. Biol. Chem.* **1999**, *274*, 5185–5192.
79. D. Xu; V. Tiwari; G. Xia; C. Clement; D. Shukla; J. Liu, *Biochem. J.* **2005**, *385*, 451–459.

80. Z. L. Wu; M. Lech; D. L. Beeler; R. D. Rosenberg, *J. Biol. Chem.* **2004**, *279*, 1861–1866.
81. G. Xia; J. Chen; V. Tiwari; W. Ju; J.-P. Li; A. Malmström; D. Shukla; J. Liu, *J. Biol. Chem.* **2002**, *277*, 37912–37919.
82. S. HajMohammadi; K. Enjyoi; M. Princivalle; P. Christ; M. Lech; D. L. Beeler; H. Rayburn; J. J. Schwartz; S. Barzegar; A. I. de Agostini; M. J. Post; R. D. Rosenberg; N. W. Shworak, *J. Clin. Invest.* **2003**, *111*, 989–999.
83. P. W. Park; O. Reizes; M. Bernfield, *J. Biol. Chem.* **2000**, *275*, 29923–29926.
84. J. T. Gallagher, *J. Clin. Invest.* **2001**, *108*, 357–361.
85. U. Lindahl; M. Kusche-Gullberg; L. Kjellen, *J. Biol. Chem.* **1998**, *273*, 24979–24982.
86. R. C. Becker, *J. Thromb. Thrombolysis* **2004**, *18*, 55–58.
87. V. Tagalakis; M. Blostein; C. Robinson-Cohen; S. R. Kahn, *Cancer Treat. Rev.* **2007**, *33*, 358–368.
88. M. Altinbas; H. S. Coskun; O. Er; M. Ozkan; B. Eser; A. Unal; M. Cetin; S. Soyuer, *J. Thromb. Haemost.* **2004**, *2*, 1266–1271.
89. N. Seeholzer; B. Thurlimann; D. Koberle; D. Hess; W. Korte, *Blood Coagul. Fibrinolysis* **2007**, *18*, 415–423.
90. S. A. Mousa; R. J. Linhardt; J. L. Francis; A. Amirhosravi, *Thromb. Haemost.* **2006**, *96*, 816–821.
91. D. Liu; Z. Shriver; G. Venkataraman; Y. E. Shabrawi; R. Sasisekharan, *Proc. Natl. Acad. Sci.* **2002**, *99*, 568–573.
92. B. Herold; S. I. Gerber; B. J. Belval; A. M. Siston; N. Shulman, *J. Virol.* **1996**, *70*, 3461–3469.
93. M.-T. Shieh; D. WuDunn; R. I. Montgomery; J. D. Esko; P. G. Spear, *J. Cell Biol.* **1992**, *116*, 1273–1281.
94. D. Shukla; P. G. Spear, *J. Clin. Invest.* **2001**, *108*, 503–510.
95. D. WuDunn; P. G. Spear, *J. Virol.* **1989**, *63*, 52–58.
96. M. Yoon; A. Zago; D. Shukla; P. G. Spear, *J. Virol.* **2003**, *77*, 9221–9231.
97. J. Liu; Z. Shriver; R. M. Pope; S. C. Thorp; M. B. Duncan; R. J. Copeland; C. S. Raska; K. Yoshida; R. J. Eisenberg; G. Cohen; R. J. Linhardt; R. Sasisekharan, *J. Biol. Chem.* **2002**, *277*, 33456–33467.
98. R. J. Copeland; A. Balasubramaniam; V. Tiwari; F. Zhang; A. Bridges; R. J. Linhardt; D. Shukla; J. Liu, *Biochemistry* **2008**, *47*, 5774–5783.
99. A. E. Proudfoot; T. M. Handel; Z. Johnson; E. K. Lau; P. LiWang; I. Clark-Lewis; F. Borlat; T. N. Wells; M. H. Kosco-Vilbois, *Proc. Natl. Acad. Sci.* **2003**, *100*, 1885–1890.
100. L. Wang; J. R. Brown; A. Varki; J. D. Esko, *J. Clin. Invest.* **2002**, *110*, 127–136.
101. V. Prabhakar; R. Sasisekharan, *Adv. Pharmacol.* **2006**, *53*, 69–115.
102. H. Kitagawa; K. Tsutsumi; M. Ujikawa; F. Goto; J. Tamura; K. W. Neumann; T. Ogawa; K. Sugahara, *Glycobiology* **1997**, *7*, 531–537.
103. Y. Miura; H. H. Freeze, *Glycobiology* **1998**, *8*, 813–819.
104. J. M. Trowbridge; R. L. Gallo, *Glycobiology* **2002**, *12*, 117R–125R.
105. A. Aspberg; R. Miura; S. Bourdoulous; M. Shimonaka; D. Heinegard; M. Schachner; E. Ruoslahti; Y. Yamaguchi, *Proc. Natl. Acad. Sci.* **1997**, *94*, 10116–10121.
106. E. Ruoslahti, *Glycobiology* **1996**, *6*, 489–492.
107. Y. Yamaguchi, *Cell. Mol. Life Sci.* **2000**, *57*, 276–289.
108. K. J. Doege; M. Sasaki; T. Kimura; Y. Yamada, *J. Biol. Chem.* **1991**, *266*, 894–902.
109. D. R. Zimmermann; E. Ruoslahti, *EMBO J.* **1989**, *8*, 2975–2981.
110. T. N. Wight; M. J. Merrilees, *Cir. Res.* **2004**, *94*, 1158–1167.
111. J. H. Yoon; J. Halper, *J. Musculoskelet. Neuronal. Interact.* **2005**, *5*, 22–34.
112. U. Rauch; L. Karthikeyan; P. Maurel; R. U. Margolis; R. K. Margolis, *J. Biol. Chem.* **1992**, *267*, 19536–19547.
113. H. Yamada; K. Watanabe; M. Shimonaka; Y. Yamaguchi, *J. Biol. Chem.* **1994**, *269*, 10119–10126.
114. S. Momohara; N. Okada; K. Ikari; S. Mizuno; H. Okamoto, *Mod. Rheumatol.* **2007**, *17*, 301–305.
115. T. J. Lujan; C. J. Underwood; H. B. Henniger; B. M. Thompson; J. A. Weiss, *J. Orthop. Res.* **2007**, *25*, 894–903.
116. P. J. Roughley; R. J. White, *Biochem. J.* **1989**, *262*, 823–827.
117. P. J. Roughley, *Eur. Cell Mater.* **2006**, *12*, 92–101.
118. S. A. Osborne; R. A. Daniel; K. Desilva; R. B. Seymour, *Glycobiology* **2007**, *18*, 225–234.
119. R. Schäfer; D. Dehn; G. J. Burbach; T. Deller, *Neurosci. Lett.* **2008**, *439*, 61–65.
120. M. S. Viapiano; R. T. Matthews, *Trends Mol. Med.* **2006**, *12*, 488–496.
121. M. S. Viapiano; S. Hockfield; R. T. Matthews, *J. Neurooncol.* **2008**, *88*, 261–272.
122. K. Bergefall; E. Trybala; M. Johansson; T. Uyama; S. Naito; S. Yamada; H. Kitagawa; K. Sugahara; T. Bergstrom, *J. Biol. Chem.* **2005**, *280*, 32193–32199.
123. T. Uyama; M. Ishida; T. Izumikawa; E. Trybala; F. Tufaro; T. Bergstrom; K. Sugahara; H. Kitagawa, *J. Biol. Chem.* **2006**, *281*, 38668–38674.
124. P. H. Weigel; P. L. DeAngelis, *J. Biol. Chem.* **2007**, *282*, 36777–36781.
125. K. Hedman; M. Kurkinen; K. Alitalo; A. Vaheri; S. Johansson; M. Hook, *J. Cell Biol.* **1979**, *81*, 83–91.
126. A. Almond, *Cell. Mol. Life Sci.* **2007**, *64*, 1591–1596.
127. S. P. Evanko; M. I. Tammi; R. H. Tammi; T. N. Wight, *Adv. Drug Del. Rev.* **2007**, *59*, 1351–1365.
128. E. R. Bastow; S. Byers; S. B. Golub; C. E. Clarkin; A. A. Pitsillides; A. J. Fosang, *Cell. Mol. Life Sci.* **2008**, *65*, 395–413.
129. K. Meyer; A. Linker; E. A. Davidson; B. Weissmann, *J. Biol. Chem.* **1953**, *206*, 611–616.
130. A. H. K. Plaas; L. A. West; R. J. Midura, *Glycobiology* **2001**, *11*, 779–790.
131. J. L. Funderburgh, *IUBMB Life* **2002**, *54*, 187–194.
132. J. L. Funderburgh, *Glycobiology* **2000**, *10*, 951–958.
133. T. Krusius; J. Finne; R. K. Margolis; R. U. Margolis, *J. Biol. Chem.* **1986**, *261*, 8237–8242.
134. D. Lewis; Y. Davies; I. A. Nieduszynski; F. Lawrence; A. J. Quantock; R. Bonshek; N. J. Fullwood, *Glycobiology* **2000**, *10*, 305–312.
135. N. W. Shworak; L. M. S. Fritze; J. Liu; L. D. Butler; R. D. Rosenberg, *J. Biol. Chem.* **1996**, *271*, 27063–27071.

Biographical Sketches



Courtney L. Jones received her B.S. in chemistry in 2004 from the University of North Carolina at Chapel Hill. She is currently a graduate student in the Division of Medicinal Chemistry and Natural Products at UNC-CH where she is working under the guidance of Dr. Jian Liu studying the enzymatic synthesis of the anticoagulant heparin drug.



Ding Xu received his B.S. in biochemistry in 1999 from Central China Normal University. He received his Ph.D. in pharmaceutical sciences from the University of North Carolina at Chapel Hill. He studied the structure and mechanism of action of heparan sulfate biosynthetic enzymes at UNC-CH under the guidance of Dr. Jian Liu. He is currently a postdoctoral fellow in the Department of Cellular and Molecular Medicine at the University of California at San Diego.



Jian Liu received his B.S. and M.S. in chemistry and biochemistry in 1984 and 1987, respectively, from Nankai University in China. He received his Ph.D. in medicinal chemistry

and natural products from the University of Iowa in 1993. He then became a postdoctoral fellow studying the mechanism of the biosynthesis of heparan sulfate at Massachusetts Institute of Technology. Dr. Jian is now an associate professor in medicinal chemistry and natural products at the University of North Carolina at Chapel Hill. His research interest is the study of the structure and function of glycosaminoglycans.

6.15 Biochemistry and Molecular Biology of Glycogen Synthesis in Bacteria and Mammals and Starch Synthesis in Plants

Jack Preiss, Michigan State University, East Lansing, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|---------------|---|-----|
| 6.15.1 | Introduction | 430 |
| 6.15.2 | The Role of Glycogen and Starch | 431 |
| 6.15.2.1 | Bacterial Glycogen | 431 |
| 6.15.2.2 | Plant Starch | 431 |
| 6.15.2.3 | Animal Glycogen | 432 |
| 6.15.3 | Synthesis of Bacterial Glycogen and Plant Starch | 432 |
| 6.15.3.1 | Glycogen Synthesis in Bacteria | 432 |
| 6.15.3.2 | Starch Synthesis in Plants and Algae | 433 |
| 6.15.3.3 | Glycogen Synthesis in Mammals | 434 |
| 6.15.4 | Properties of the Bacterial and Plant Enzymes Involved in the Synthesis of Glycogen and Starch | 434 |
| 6.15.4.1 | ADP-Glucose Pyrophosphorylases: Structure and Properties | 434 |
| 6.15.4.1.1 | Molecular weight and subunit structure | 434 |
| 6.15.4.1.2 | Reaction mechanism | 435 |
| 6.15.5 | Regulation of the Synthesis of Bacterial Glycogen and Starch | 436 |
| 6.15.5.1 | Introduction | 436 |
| 6.15.5.2 | Activators and Inhibitors of ADP-Glc PPase | 436 |
| 6.15.5.3 | Overlapping Specificity of the ADP-Glc PPase Allosteric Activator Site | 439 |
| 6.15.5.4 | Effect of Activators and Inhibitors on ADP-Glc PPase Kinetics | 439 |
| 6.15.5.5 | Experimental Evidence Supporting the Role of ADP-Glc PPase in the Regulation of the Biosynthesis of Bacterial Glycogen and Plant Starch | 440 |
| 6.15.5.6 | Plant ADP-Glc PPases Can Be Activated by Thioredoxin | 443 |
| 6.15.5.7 | Identification of Substrate and Effector Sites and Catalytic Residues | 444 |
| 6.15.5.7.1 | Bacterial systems | 444 |
| 6.15.5.7.2 | Plant, algal, and cyanobacterial systems | 445 |
| 6.15.5.7.3 | Properties and functions of the small and large subunits of higher plant ADP-Glc PPases | 447 |
| 6.15.5.7.4 | Phylogenetic analysis of the large and small subunits | 451 |
| 6.15.6 | Subcellular Localization of ADP-Glc PPase in Plants | 451 |
| 6.15.7 | Crystal Structure of Potato Tuber ADP-Glc PPase | 451 |
| 6.15.8 | Bacterial Glycogen Synthase | 452 |
| 6.15.8.1 | Enzyme structure and Properties | 452 |
| 6.15.8.2 | Substrate and Catalytic Sites | 452 |
| 6.15.8.3 | Reversibility of the Glycogen Synthase Reaction | 454 |
| 6.15.9 | Starch Synthases | 455 |
| 6.15.9.1 | Characterization of the Starch Synthases | 455 |
| 6.15.9.1.1 | Soluble starch synthases | 455 |
| 6.15.9.1.2 | Soluble starch synthase I | 455 |
| 6.15.9.1.3 | Starch synthase II | 456 |
| 6.15.9.1.4 | Starch synthase III | 458 |
| 6.15.9.1.5 | Starch synthase IV | 458 |
| 6.15.9.1.6 | Double mutants of the soluble starch synthases | 459 |
| 6.15.9.1.7 | Starch synthases bound to the starch granule | 460 |

| | | |
|-------------------|--|-----|
| 6.15.9.1.8 | Isolation of the waxy protein structural gene | 461 |
| 6.15.9.1.9 | Further studies of GBSS and isoforms; their involvement in both amylopectin and amylose synthesis | 462 |
| 6.15.10 | Branching Enzyme | 464 |
| 6.15.10.1 | Branching Enzyme Assays | 464 |
| 6.15.10.2 | Bacterial Branching Enzymes | 464 |
| 6.15.10.3 | Plant and Algal Branching Enzymes; Characterization of Isozymes | 464 |
| 6.15.10.4 | Genetic Studies on Branching Enzyme-Deficient Mutants | 466 |
| 6.15.10.5 | Isolation of cDNA Clones Encoding the Branching Enzyme Isozyme Genes | 467 |
| 6.15.10.6 | Reserve Tissue Branching Enzyme Is Localized in the Plastid | 468 |
| 6.15.10.7 | Branching Enzyme Belongs to the α -Amylase Family | 468 |
| 6.15.10.8 | Amino Acid Residues that are Functional in Branching Enzyme Catalysis | 469 |
| 6.15.11 | Other Enzymes Involved in Starch Synthesis | 469 |
| 6.15.11.1 | Isoamylase | 469 |
| 6.15.11.2 | α -1,4 Glucanotransferase | 470 |
| 6.15.12 | Genetic Regulation of Bacterial Glycogen Synthesis | 470 |
| 6.15.13 | Properties of the Glycogen Biosynthetic Enzymes of Mammals | 472 |
| 6.15.13.1 | UDP-Glucose Pyrophosphorylase | 472 |
| 6.15.13.2 | Glycogen Synthase | 473 |
| 6.15.13.3 | Branching Enzyme | 473 |
| 6.15.13.4 | Glycogenin | 474 |
| 6.15.13.4.1 | Genetic evidence indicating that glycogenin is required for glycogen synthesis | 475 |
| 6.15.14 | Regulation of Mammalian Glycogen Synthesis | 476 |
| 6.15.14.1 | General Considerations | 476 |
| 6.15.14.2 | Regulation of Glycogen Synthase by Phosphorylation – Dephosphorylation | 476 |
| 6.15.14.2.1 | The effect of phosphorylation on glycogen synthase activity and relative effects of phosphorylation on different sites | 478 |
| 6.15.14.3 | Mechanism of Stimulation of Mammalian Skeletal Muscle Glycogen Synthesis by Insulin | 481 |
| 6.15.14.3.1 | Protein phosphatase 1 | 481 |
| 6.15.14.3.2 | Inactivation of glycogen synthase kinase-3 | 481 |
| References | | 482 |

6.15.1 Introduction

The biosynthesis of α -1,4-glucans such as starch in plants and glycogen in bacteria and mammals is a process where living organisms accumulate an energy reserve for utilization when energy and carbon sources are not available from their immediate surrounding. The advantage in having polysaccharides as storage reserves is that they have little effect on the internal osmotic pressure in the cell due to their high molecular weights and other physical properties.

Starch is a mixture of two polyglucans, amylose, and amylopectin. Amylose is a relatively small (about 1000 residues) α -1,4-glucan that is largely linear, although containing occasional α -1,6 branches. Amylopectin is a much larger molecule with a degree of polymerization of 10^4 – 10^5 residues and having frequent α -1,6 branch points. The chemical and physical aspects of the structure of the starch granule and its components, amylose and amylopectin, have been discussed in some excellent reviews;^{1,2} for the structure of animal glycogen, see the review by Manners.³ Relatively less is known on the structures of bacterial and fungal glycogen. They have not been studied to the same extent, but the available data indicate that the bacterial and fungal glycogen structures are quite similar to that of mammalian glycogen.⁴ Muscle contains spherical β -particles 1500–4000 nm in diameter, while liver contains α -particles or rosettes that appear as aggregates of β -particles. A β -particle can contain as many as 6×10^4 glucose residues, joined every 8–10 residues in straight α -1,4 chains with α -1,6 linkages.

The linkages of amylopectin and glycogen provide the cell with a glucose polysaccharide that upon action by degradative enzymes, such as amylases, isoamylases, debranching enzymes, or phosphorylases are easily converted into glucose and or α -Glc-1-P.

Although the structures of bacterial and mammalian glycogen are relatively similar, their similarity does not extend to the mechanism of their biosynthesis. Most importantly, the glucose donor for elongation of the primer glucan is different.

In vitro synthesis of a α -glucan was first demonstrated by Carl and Gerti Cori using phosphorylase and Glc-1-P.⁵ This was an accepted route for the biosynthesis of polyglucans until 1957. With the discovery of glycogen synthase in 1957,⁶ it became clear that the biosynthesis and degradation of glycogen occurs by different pathways, with phosphorylase being a degradative enzyme.

Regulation of mammalian glycogen synthesis remains an intense study not only by those seeking to understand glycogen synthesis itself, but also as a model for studying the effect of hormonal action on cellular regulation. A concise description of this important research area of cellular control is provided in this chapter, including *in vitro* experiments relevant to processes involved in physiological control.

Some recent reviews on bacterial glycogen synthesis^{7–11} and on starch biosynthesis^{12–19} discuss in more detail some of the areas presented in this chapter. For the regulation of mammalian glycogen synthesis see Cohen^{20–21} and Roach.^{22–23}

6.15.2 The Role of Glycogen and Starch

6.15.2.1 Bacterial Glycogen

Glycogen occurs in many bacteria, and accumulates under environmental conditions where growth is limited and there is an excess carbon supply.^{7,8,10,11,24–26} Glycogen accumulation has been shown to occur in the stationary phase of the growth cycle due to limitations in the source of sulfur, nitrogen, or phosphate. Glycogen is not required for bacterial growth, and mutants deficient in glycogen may grow as well as their wild-type strains. Biological functions of bacterial glycogen have been reviewed²⁵ and under conditions of no available carbon source, glycogen is probably utilized to preserve cell integrity. Bacteria require energy for maintenance under nongrowing conditions and this 'energy of maintenance' is the energy needed for processes such as maintenance of motility and intracellular pH, chemotactic response, turnover of proteins and RNA, and osmotic regulation. In media devoid of carbon source, *Escherichia coli* and *Enterobacter aerogenes* having glycogen do not degrade their RNA and protein components. In contrast, the glycogen-deficient bacteria release NH_3 from their nitrogen-containing components.²⁶ Glycogen-containing *E. aerogenes*, *E. coli*, and *Streptococcus mitis* also survive better than these organisms with no glycogen. Another function for glycogen is suggested for various *Clostridia* species. *Clostridia* can accumulate glycogen up to 60% of their dry weight and prior to spore formation, the glycogen is rapidly degraded.²⁷ *Clostridial* glycogen-deficient strains are poor spore formers, suggesting that glycogen serves as a source of carbon and energy for spore formation and maturation. Although these studies suggest that glycogen plays a role in bacterial survival, glycogen-rich *Sarcina lutea* cells die faster when starved in phosphate buffer than cells with no polysaccharide.²⁸

6.15.2.2 Plant Starch

Starch is present in many types of plant tissues and organs of almost all green plants; for example, leaves, roots, shoots, fruits, tubers, grains, and stems. As first shown in the nineteenth century starch disappears from leaves exposed to low light or left in the dark for a long time (24–48 h).²⁹ The leaf present in bright light can have a reforming of the starch granules in the chloroplast. This is seen by staining the leaf with iodine³⁰ or by either light or electron microscopy.³¹ Carbon fixation during photosynthesis in the light leads to starch formation, and the resultant starch is degraded in the dark. The products of starch degradation are used as an energy source and are also converted into sucrose, and transported to other organs of the plant. The biosynthesis and degradation of starch in the leaf is a dynamic process, with starch contents fluctuating during the day.

In fruits, storage organs, or seeds, the synthesis of starch usually occurs during the development and maturation of the tissue. Starch degradation in these tissues occurs at the time of sprouting or germination of

the seed or tuber, or in ripening of the fruit, where it is used as a source of both carbon and energy. The degradative and biosynthetic processes in the storage tissues are thus temporally separated.

6.15.2.3 Animal Glycogen

Glycogen is a major energy source in animal cells. In humans, glycogen represents about 1% wet weight of the skeletal muscle and 5% of the liver. Glycogen reserves would total some 90 g in muscle and 350 g in the liver of a 70-kg man. The storage of glucose as glycogen is important for mammalian homeostasis, and the glycogen reserves in skeletal muscle and the liver have different specific functions.

Hepatic glycogen is accumulated when excess glucose is available in the diet and is used to maintain a steady level of blood glucose. Liver glycogen is usually not a supply of energy for the liver. A 70-kg man utilizes 180 g of glucose per day for tissues that can only utilize carbohydrate as an energy source. Approximately one-half of this glucose is obtained from hepatic glycogen. Hepatic glycogen synthesis and mobilization is thus dictated by blood glucose levels and is controlled by regulatory hormones, primarily glucagon, insulin, and glucocorticoids. In contrast, all other mammalian extra hepatic tissues, utilize their glycogen reserves for specific functions of the tissue.

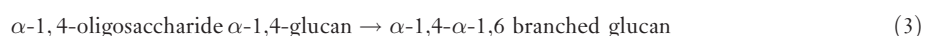
Although the glycogen content of skeletal muscle is large, compared to that of liver, it is not directly available as a source of blood glucose. During exercise, lactate is formed from skeletal muscle by the degradation of glycogen and glycolysis. At rest, the lactate is converted into glucose mainly by the liver and to some extent by the kidney via gluconeogenesis. This glucose can serve as a source of about 10–20% of the total blood glucose. In summary muscle glycogen primarily serves as an energy source, broken down to provide energy for muscle contraction. Replenishment of muscle glycogen occurs when there is high blood glucose (hyperglycemia) due to diet and stimulation by insulin of glucose transport into the muscle cells. Some of the imported glucose would be used for glycogen synthesis.

6.15.3 Synthesis of Bacterial Glycogen and Plant Starch

After the discovery of UDP-Glc and other sugar nucleotides by Leloir and associates, it was then shown that particulate or soluble cell fractions of many sources, prokaryotic and eukaryotic, could carry out glycosyl transfers from sugar nucleotides to suitable acceptor molecules. Sugar nucleotides are formed from sugar 1-phosphates and nucleoside triphosphates. The synthesis of α -1,4, α -1,6 branched polysaccharides can be divided into several steps: (1) the synthesis of the donor sugar nucleotide from sugar 1-phosphates, (2) transfer of the glucosyl unit from the sugar nucleotide to the glucan primer to form the α -1,4-glucosidic linkage, and (3) the cleavage of α -1,4 chains that are reinserted into the polysaccharide as branched chains linked by α -1,6 linkages.

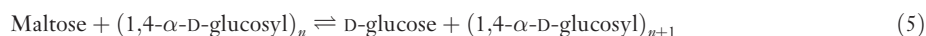
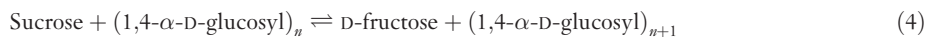
6.15.3.1 Glycogen Synthesis in Bacteria

In 1964, it was apparent that the glucose donor for glycogen synthesis in mammals was UDP-Glc, and for synthesis of starch in plants, ADP-Glc.³² Several mutants of *E. coli* deficient in UDP-glucose pyrophosphorylase (UDP-Glc PPase) accumulated normal amounts of glycogen during growth in limiting nitrogen media.³³ Thus, UDP-Glc was not the glucosyl donor for glycogen synthesis. However it was noted that in bacteria, the synthesis of ADP-Glc (Reaction (1)) was catalyzed by ADP-glucose pyrophosphorylase (ADP-Glc-PPase) (2.7.7.27; ATP: α -D-Glc-1-P adenylyltransferase^{34,35}). In 1964 it was also reported that extracts of several bacteria contained large activities of an ADP-Glc: α -1,4-D-glucan-4- α -glucosyl transferase, also known as the bacterial glycogen synthase (Reaction (2)).^{36,37} These same bacterial extracts also contained ADP-Glc PPase (Reaction (1)).



About 10% of the linkages in bacterial glycogen are α -1,6. The formation of these linkages is catalyzed by an α -1,4-glucan branching enzyme (EC 2.4.1.18; 1,4- α -D-glucan 6- α -(1,4- α glucano)-transferase) in Reaction (3). Branching enzyme activity has been detected in *E. coli*,^{38,39} in *Arthrobacter globiformis*,⁴⁰ in *Salmonella typhimurium*,⁴¹ and in *Streptococcus mitis*.⁴² The branching enzyme genes from *E. coli*,⁴³ from *Streptomyces aureofaciens*,⁴⁴ *Bacillus stearothermophilus*,^{45,46} *Bacillus caldolyticus*,⁴⁷ and from cyanobacteria^{48,49} have been cloned.

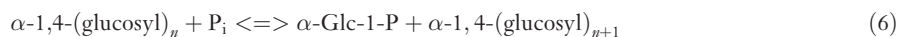
There are alternative pathways leading to the formation of α -glucans in bacteria. In certain bacteria, a glycogen-like α -glucan can be synthesized either directly from sucrose (Reaction (4)) or from maltose (Reaction (5)) or from Glc-1-P via the phosphorylase reaction (Reaction (6)).⁵⁰



Amylosucrase catalyzes Reaction (4), and is found in sucrose-grown *Neisseria* strains that accumulate large amounts of a polysaccharide similar to glycogen.^{51,52} However, amylosucrase is found in few bacterial species and is present only when sucrose is in the growth culture. *Neisseria* can metabolize exogenous sucrose but cannot synthesize sucrose. Therefore, the observed accumulation of glycogen observed in *Neisseria* and in other microorganisms grown on carbon sources other than sucrose cannot be due to amylosucrase.

Amylomaltase, the enzyme catalyzing Reaction (5), is induced along with a number of other enzymes when several strains of *E. coli*, *Aerobacter aerogenes*, *Streptococcus mutans*,^{53,54} *S. mitis*,⁵⁵ *Diplococcus pneumoniae*,⁵⁶ and *Pseudomonas stutzeri*⁵⁷ are grown on maltose or maltodextrins. The synthesis of amyломaltase, however, is repressed by glucose⁵⁸ and its activity therefore cannot account for the synthesis of glycogen in organisms grown on glucose as a carbon source.

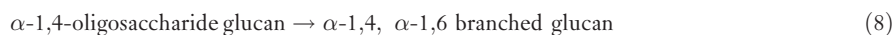
Maltodextrin phosphorylase and glycogen phosphorylase (Reaction (6)) occur in many bacteria and catalyzes the synthesis as well as the phosphorolysis of α -4-glucosidic linkages present in glycogen or starch. Maltodextrin phosphorylase however is only induced in the presence of maltodextrins, and for the microorganisms studied, glycogen phosphorylase activity is insufficient to account for their rate of glycogen accumulation.⁵⁹⁻⁶¹ Moreover, *E. coli* mutants deficient in maltodextrin phosphorylase accumulate maltodextrins, suggesting that the phosphorylase is involved in the degradation (phosphorolysis), and not in synthesis, of maltodextrins or α -1,4-glucans.⁶²



Thus, bacteria accumulating glycogen do so via the ADP-Glc pathway. Glycogen deficient or glycogen-excess mutants of *E. coli* and of *Salmonella typhimurium* (reviewed elsewhere^{7-11,13,63}) have been isolated and it has been shown that they are affected either in ADP-Glc PPase activity or in glycogen synthase activity or in both. A list of bacteria containing glycogen and/or the glycogen biosynthetic enzymes have been compiled in past reviews.^{26,63} The lists shows that glycogen accumulation is not restricted to any class of bacteria, being present in Gram-negative or Gram-positive types and even in archaeobacteria.⁶⁴

6.15.3.2 Starch Synthesis in Plants and Algae

The reactions of starch synthesis 1, 7, and 8 are essentially similar to those of glycogen synthesis in bacteria but the structure of the ultimate polysaccharides formed are different.



Reaction (7) is catalyzed by starch synthases (EC 2.4.1.21; ADP-Glc; 1,4- α -D-glucan 4- α -glucosyltransferase), the same reaction as the one catalyzed by the bacterial glycogen synthase (Reaction (2)). Here a different reaction number is given to stress that the final products of starch, amylose, and amylopectin, are different in structure. Isozymic forms of plant starch synthases^{9,12,14-16,65-69} and of branching enzymes^{9,12,14-16,65,66,70-76} have been reported. The starch synthase isozymes are different gene products and they play different roles in the synthesis of amylose and amylopectin. Many plants as well as *Cblamydomonas reinhardtii* have a granule-bound starch synthase (GBSS) that

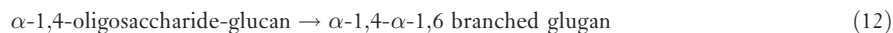
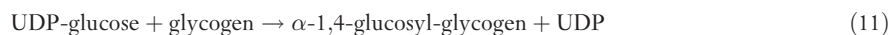
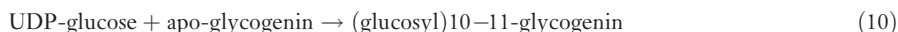
is involved in the synthesis of amylose.^{77–83} Mutants deficient in this enzyme are known as waxy mutants, and produce starch granules only containing amylopectin. Reaction (7) was first reported by Leloir *et al.*⁸⁴ with UDP-Glc as the glycosyl donor, but it was later shown that with ADP-Glc as the glucosyl donor, the enzyme had a higher V_{\max} and a much lower K_m (higher apparent affinity).³² Thus, the GBSS in reserve tissues do have some activity with UDP-Glc, although much lower than that seen with ADP-Glc. In contrast, leaf starch synthases and the soluble starch synthases of reserve tissues are specific for ADP-Glc.

Reaction (8) is catalyzed by branching enzyme, being a similar reaction to the one seen in bacteria (Reaction (3)). However, the branch chains in amylopectin are longer (about 20–24 glucose units long) and there is less branching in amylopectin (~5% of the glucosidic linkages are α -1,6) than in glycogen (10–13 glucose units long and 10% of linkages are α -1,6). Thus, most likely, the starch branching enzymes have different specificities with respect to chain transfer than those that branch glycogen, or perhaps the interaction of the starch branching enzymes with the starch synthases may be different from the interaction between the bacterial branching enzymes and glycogen synthases.

Another enzyme, debranching enzyme, is also involved in synthesis of the starch granule and its polysaccharide components amylose and amylopectin^{85–87} and this will be discussed in Section 6.15.11.1.

6.15.3.3 Glycogen Synthesis in Mammals

The glycogen synthetic pathway in mammals, is seen in Reactions (9)–(12).



The synthesis of UDP-glucose, the glucosyl donor for glycogen synthesis, is catalyzed by the enzyme UDP-Glc PPase (EC 2.7.7.9; UTP: α -D-glucose-1 phosphate uridylyltransferase; Reaction (9)). The equilibrium constant of the reaction toward UDP-Glc synthesis is less than 1 but pyrophosphate, the by-product of the reaction, is cleaved by the inorganic pyrophosphatase. Thus, the formation of UDP-Glc is essentially irreversible. In Reaction (10), the protein glycogenin accepts a glucose residue from the sugar nucleotide. Glycogenin is currently considered the first acceptor of glucose units in the initiation of glycogen synthesis.^{88–90} Rabbit muscle glycogenin is a 37-kDa self-glycosylating protein of 332 amino acids.⁹¹ Glycosylation occurs on the hydroxyl group of its tyrosine residue 194^{92–94} and Mn^{2+} is a required cofactor for the reaction. After the initial glucosylation further self-glycosylation of the first glucose unit occurs up to an oligosaccharide chain containing 7–11 glucosyl units linked as α -1,4 glucosyl linkages. The glucosylated-glycogenin can now serve as a primer for the glycogen synthase reaction (EC 2.4.1.11; UDP-glucose-glycogen 4- α glucosyl transferase; Reaction (11)). Also, glycogen and lower molecular weight maltodextrins can also serve as primers for the glycogen synthase reaction. Chain elongation of the glucosylated glycogenin continues in the glycogen synthase reaction. Subsequently, branching enzyme (EC 2.4.1.18; 1,4- α -D-glucan 6- α -(1,4 α -glucano)-transferase) catalyzes the formation of α -1,6 branched points (Reaction (12)). The actual details of events between formation of the oligosaccharide glycogenin primer and formation of the glycogen of a molecular size of 10⁷ kDa are not known.

6.15.4 Properties of the Bacterial and Plant Enzymes Involved in the Synthesis of Glycogen and Starch

6.15.4.1 ADP-Glucose Pyrophosphorylases: Structure and Properties

6.15.4.1.1 Molecular weight and subunit structure

The molecular sizes of the pure enzymes from *E. coli* B,^{95,96} *Rhodospirillum rubrum*,⁹⁷ and spinach leaf⁹⁸ have been estimated using sedimentation equilibrium ultracentrifugation, and the molecular sizes of purified and partially purified enzymes from *Aeromonas hydrophila*, *Rhodospirillum molischanum*, *Rhodospirillum tenue*, *Rhodospirillum fulvum*,

Rhodobacter spheroides, *R. gelatinosa*, *R. viridis*, *R. acidophila*, *R. globiformis*, *Salmonella typhimurium*, and *Serratia marcescens*, have been estimated using ultracentrifugation on sucrose gradients.⁹⁹ A study of the subunit molecular weights of four bacterial enzymes show that the native enzyme from bacterial sources are tetramers of similar subunits of molecular mass 45–51 kDa. In the case of the *S. marcescens* enzyme, two molecular weight species are seen, 96 and 186 kDa, respectively.¹⁰⁰ The enzyme thus exists in homotetrameric and homo dimeric forms in equilibrium presuming that the subunit molecular weight of the *S. marcescens* enzyme is similar to the *E. coli* enzyme.

The ADP-Glc PPase gene from *E. coli* has been isolated, expressed,¹⁰¹ and sequenced⁴³ and its calculated protein molecular weight as deduced from the nucleotide sequence is 48 762. This is in good agreement with the determined approximate molecular weight of 50 000.^{95,96}

The plant ADP-Glc PPases studied in most detail with respect to kinetic properties^{102–104} and structure¹⁰³ is that from spinach leaf and potato tuber.^{105,106}

The spinach leaf ADP-Glc PPase has been purified to homogeneity by preparative disc gel electrophoresis⁹⁸ and by hydrophobic chromatography.¹⁰³ The spinach leaf enzyme has a molecular mass of 206 000 and is composed of two different subunits of molecular masses of 51 and 54 kDa.^{107–109} These subunits can be distinguished not only by differences in their molecular mass but also with respect to amino acid composition, N-terminal sequences, peptide patterns of the tryptic digests on high-performance liquid chromatography (HPLC), and antigenic properties. The two subunits are quite different and the products of two genes. In contrast, and as discussed above, bacterial ADP-Glc PPases including the cyanobacterial enzymes are homotetrameric, that is, composed of four identical subunits of 50–55 kDa in mass depending on the species.^{7,10,25,26}

Other plant ADP-Glc PPases have been studied in detail and they also have been shown to be composed of two dissimilar subunits. ADP-Glc PPase from potato tuber is composed of two different subunits, 50 and 51 kDa, with $\alpha_2\beta_2$ heterotetrameric subunit structure.^{110,111} The small subunit of many higher plant ADP-Glc PPases is highly conserved among plants with 85–95% identity.¹¹² The maize endosperm ADP-Glc PPase, of molecular mass of 230 kDa, reacts with the antibody prepared against the native spinach leaf enzyme in immunoblot experiments.¹¹³ The enzyme is composed of subunits of 55 and 60 kDa, which would correspond, respectively, to the spinach leaf 51 and 54 kDa.¹¹⁴

The studies of the maize endosperm mutants, *sbrunken 2* (*sb 2*) and *brittle 2* (*bt 2*), which are deficient in ADP-Glc PPase activity are also relevant. In immunoblotting experiments and using antibodies raised against the native (holoenzyme) and against each subunit of the spinach leaf enzyme, it was found that the mutant *bt 2* endosperm lacks the 55 kDa subunit and the mutant *sb 2* endosperm lacks the 60 kDa subunit. These results¹¹³ strongly suggest that the maize endosperm ADP-Glc PPase is composed of two immunologically distinctive subunits and that the *sb 2* and *bt 2* mutations cause reduction in ADP-Glc PPase activity through the lack of one of the subunits. The *sb 2* gene would be the structural gene for the 60 kDa protein where as the *bt 2* gene would be the structural gene for the 55 kDa protein. The isolation of an ADP-Glc PPase cDNA clone from a maize endosperm library¹¹⁴ that hybridized with the small subunit cDNA clone from rice¹¹⁵ is consistent with this hypothesis. This maize ADP-Glc PPase cDNA clone was found to hybridize to a transcript present in maize endosperm but absent in *bt 2* endosperm. Thus, *bt 2* would be a mutation of the structural gene of the 55 kDa subunit of the ADP-Glc PPase.

In summary, data available so far indicate that leaf and seed ADP-Glc PPases are heterotetramers composed of two different subunits, and that, on the basis of immunoreactivity and sequence data,¹¹² there is corresponding homology between the subunits in the leaf enzyme and with the subunits of reserve tissue enzyme.

6.15.4.1.2 Reaction mechanism

In the presence of activator, pyruvate, the substrate saturation curves of the *R. rubrum* ADP-Glc PPase are hyperbolic at low temperatures. Using kinetic studies its reaction mechanism was studied.¹¹⁶ The product inhibition patterns eliminated all known sequential mechanisms except the ordered BiBi or Theorell–Chance mechanisms.¹¹⁷ Small intercept effects suggested the existence of significant concentrations of central transitory complexes. Kinetic constants obtained in the study also favored the ordered BiBi mechanism. In addition studies using ATP-[³²P]-pyrophosphate isotope exchange at equilibrium supported a sequential-ordered mechanism, which indicated that ATP is the first substrate to bind and that ADP-Glc is the last product to

dissociate from the enzyme.¹¹⁶ In the pyrophosphorolysis direction ADP-Glc the substrate would bind first and the product, ATP would dissociate last. The kinetic mechanism of the leaf barley ADP-Glc PPase has also been studied and the results also indicate that the plant enzyme follows an ordered BiBi kinetic mechanism with the ATP binding first in the synthesis direction.¹¹⁸

Binding of substrates and effectors of the *E. coli* B ADP-Glc PPase enzyme was also studied.¹¹⁹ Equilibrium dialysis showed that in the presence of 5 mmol l⁻¹ MgCl₂ and 1.5 mmol l⁻¹ fructose-1,6-bisphosphate, Glc-1-P does not bind to the enzyme. However, ATP does bind, suggesting that the reaction mechanism of the *E. coli* B enzyme is similar to the *R. rubrum* enzyme in that there is ordered binding, with MgATP binding first and Glc-1-P binding next. Chromium adenosine triphosphate (CrATP), a potent inhibitor of many enzymes which utilizes MgATP as a substrate,^{120,121} is a potent competitive inhibitor of the *E. coli* ADP-Glc PPase.¹¹⁹ When this inactive substrate analogue is present in the reaction mixture, one mole of Glc-1-P binds per mole of ADP-Glc PPase subunit.

Only two moles of MgATP or CrATP bind to the tetrameric protein in the absence of Glc-1-P. Thus, MgATP sites appear to exhibit a half-site reactivity.^{119,122} But when Glc-1-P is present, 4 moles of CrATP bind to the tetrameric protein.¹¹⁹ Thus it appears that in the ADP-Glc PPase reaction mechanism in the synthesis direction, 2 moles of MgATP initially bind to the tetrameric protein; this permits the binding of Glc-1-P to the four binding sites on the tetrameric protein. Further binding of the next 2 moles of MgATP may then follow, with concomitant catalysis occurring. A mechanism for enzyme catalysis was proposed that explains some of the kinetic and binding properties in terms of an asymmetry model in the distribution of the conformational states of the four identical subunits.¹¹⁹

6.15.5 Regulation of the Synthesis of Bacterial Glycogen and Starch

6.15.5.1 Introduction

In bacteria and in plants there is only one physiological function for ADP-Glc that is to be a donor of glucose for synthesis of α -1,4-glucosyl linkages in bacterial glycogen and for plant or algal starch. It would therefore be advantageous to conserve the ATP utilized for synthesis of the sugar nucleotide by regulating α -1,4-glucan synthesis at the level of ADP-Glc formation.

Over 50 ADP-Glc PPases (mainly bacterial but also plant) have been studied with respect to their regulatory properties and, in almost all cases, glycolytic intermediates activate ADP-Glc synthesis, while AMP, ADP, and/or P_i are inhibitors. Glycolytic intermediates in the cell can be considered as signals of carbon excess and therefore, under conditions of limited growth with excess carbon in the media, accumulation of glycolytic intermediates would be indicative for the activation of ADP-Glc synthesis. Thus, the enzyme seems to be affected by the availability of ATP in the cell and the presence of glycolytic intermediates.

6.15.5.2 Activators and Inhibitors of ADP-Glc PPase

On the basis of the differences observed in the metabolic intermediates that activate the ADP-Glc PPases studied so far, the bacterial, algal, and plant enzymes may be classified into nine groups. These groups are listed in **Table 1**.

A group of ADP-Glc PPases comprises the enteric bacteria (*Citrobacter freundii*, *Edwardsiella tarda*, *Escherichia coli*, *E. aureescens*, *Enterobacter aerogenes*, *E. cloacae*, *Klebsiella pneumoniae*, *Salmonella enteritidis*, *S. typhimurium*,⁴¹ and *Shigella dysenteriae*). These ADP-Glc PPases are activated by fructose-1,6-bisphosphate, NADPH, and pyridoxal phosphate.^{8-10,13} These bacteria have glycolysis as their main carbon disposition pathway.

Other bacteria such as *Aeromonads*, Gram-negative facultative anaerobes, or the Gram-positive aerobic organisms *Micrococcus luteus* and *Mycobacterium smegmatis*,^{8,10,13,25,26} utilizing glycolysis as their major sugar utilization pathway contain ADP-Glc PPases activated not only by FBP but also by F6P and inhibited by AMP and ADP and these enzymes are categorized in class II.

Those ADP-Glc PPases having no activator yet still inhibited by AMP are listed in class III (**Table 1**). These are enteric organisms from the genus *Serratia* (*S. liquefaciens* and *S. marcescens*)¹⁰⁰ and from *Enterobacter hafniae*. Another ADP-Glc PPase with no apparent activator is that isolated from *Clostridium pasteurianum*.²⁷

Table 1 Relationship between carbon assimilation pathway and regulatory properties of ADP-Glc PPase of different organisms

| Organism | Main carbon Utilization pathway | ADP-Glc PPase | | |
|--|---|----------------------|--|---------------------|
| | | Allosteric effectors | | |
| | | Class | Activator(s) | Inhibitor(s) |
| Prokaryotes | | | | |
| <i>Escherichia coli</i> | Glycolysis | I | Fru-1,6-bis-P | AMP |
| <i>Salmonella typhimurium</i> | | | | |
| <i>Enterobacter aerogenes</i> | | | | |
| <i>Aeromonas formicans</i> | Glycolysis | II | Fru-1,6-bis-P, Fru-6-P | AMP, ADP |
| <i>Micrococcus luteus</i> | | | | |
| <i>Mycobacterium smegmatis</i> | | | | |
| <i>Rhodopseudomonas viridis</i> | (+ Reductive carboxylic acid cycle) | | | |
| <i>Serratia marcescens</i> | Glycolysis | III | None | AMP |
| <i>Enterobacter hafniae</i> | | | | |
| <i>Clostridium pasteurianum</i> | | | | |
| <i>Agrobacterium tumefaciens</i> | Entner–Doudoroff pathway | IV | Pyruvate, Fru-6-P | AMP, ADP |
| <i>Arthrobacter viscosus</i> | | | | |
| <i>Chromatium vinosum</i> | | | | |
| <i>Rhodobacter capsulata</i> | | | | |
| <i>Rhodomicrobium vaneilli</i> | | | | |
| <i>Rhodobacter gelatinosa</i> | Glycolysis, Enter–Doudoroff pathway, and reductive carboxylic acid cycle | V | Pyruvate, Fru-6-P, and Fru-1,6-bis-P | AMP, P _i |
| <i>R. globiformis</i> | | | | |
| <i>R. sphaeroides</i> | | | | |
| <i>Rhodocyclus purpureus</i> | | | | |
| <i>Rhodospirillum rubrum</i> | Reductive carboxylic acid cycle | VI | Pyruvate | None |
| <i>Rhodospirillum tenue</i> | | | | |
| <i>Bacillus subtilis</i> | TCA cycle during sporulation | VII | None | None |
| <i>B. stearothermophilus</i> | | | | |
| Cyanobacteria | | | | |
| <i>Synechococcus</i> PCC 6301 | Oxygenic Photosynthesis | VIII | 3-PGA | P _i |
| <i>Synechocystis</i> Pcc 6803 | (Calvin cycle) | | | |
| <i>Anabaena</i> PCC 7120 | | | | |
| Eukaryotes | | | | |
| Green algae | | | | |
| <i>Chlorella fusca</i> | Oxygenic Photosynthesis | VIII | 3-PGA | P _i |
| <i>C. vulgaris</i> | (Calvin cycle) | | | |
| <i>Chlamydomonas reinhardtii</i> | | | | |
| Higher plants | | | | |
| Photosynthetic tissues | Oxygenic Photosynthesis (Calvin cycle or Hatch–Slack pathway) | VIII | 3-PGA | P _i |
| Plant leaves; e.g.; spinach wheat, pea, <i>Arabidopsis</i> , maize, rice, etc. | | | | |
| Nonphotosynthetic tissues | Sucrose catabolism and gluconeogenesis | VIII | 3-PGA | P _i |
| Potato tuber, maize endosperm | | | | |
| Barley and wheat endosperm | Sucrose catabolism and gluconeogenesis | IX | 3-PGA and Fru-6-P reverses P _i inhibition | P _i |

The enzymes placed in class IV are from bacteria mainly utilizing the Entner–Doudoroff pathway, and the ADP-Glc PPases are distinctively activated by Fru 6-P and pyruvate, with ADP, AMP, and P_i behaving as inhibitors.^{123,124} Included in this group are *Agrobacterium tumefaciens* and *Arthrobacter viscosus*. Photosynthetic

organisms capable of utilizing the anaerobic photosynthetic reductive carboxylic acid pathway, *Chromatium vinosum*, *Rhodobacter capsulata*, and *Rhodospirillum rubrum* are also in this group. They can metabolize glucose and presumably utilize the Entner–Doudoroff pathway rather than glycolysis.^{8,10,13,25}

There are ADP-Glc PPases from organisms capable of having the Embden–Meyerhoff and Entner–Doudoroff pathways as well as anaerobic photosynthetic reductive carboxylic acid pathway. These ADP-Glc PPases are activated by three effectors: Fru 1,6-bisP, Fru 6-P, and pyruvate. These three activators place them in class V (**Table 1**).^{125,126}

In class VI are ADP-Glc PPases from anaerobic bacteria *Rhodospirillum rubrum*, capable of growing in either heterotrophic conditions in the dark or autotrophic conditions in the light under anoxygenic photosynthesis (**Table 1**). These organisms cannot catabolize glucose but grow very well on pyruvate and tricarboxylic acid cycle (TCA) intermediates. ADP-Glc PPases from class VI are specifically regulated by pyruvate (**Table 1**).^{97,127}

ADP-Glc PPases grouped as class VII are enzymes from sporulating bacilli (**Table 1**). These microorganisms synthesize glycogen during sporulation, a process for survival to hostile environments. Under those conditions, the TCA cycle is the main pathway for carbon utilization fully metabolizing by-products of glycolysis.¹²⁸ The different recombinant enzymes from *B. stearothermophilus* were insensitive to regulation by different metabolites typically affecting the activity of other bacterial ADP-Glc PPases.¹²⁹ Thus, the enzymes grouped in class VII in **Table 1** are quite distinct from other ADP-Glc PPases. They are apparently unregulated enzymes and the only bacterial ADP-Glc PPases having a heterotetrameric structure of the type $\alpha_2\beta_2$.

Another group of bacterial ADP-Glc PPases are those from cyanobacteria, prokaryotes that fix CO₂ via oxygenic photosynthetic pathway similar to plants (class VIII, **Table 1**). These enzymes have 3PGA as the main activator and P_i, the inhibitor.^{130–132} The specificity for allosteric regulators of the cyanobacterial ADP-Glc PPase is identical to those of the eukaryotic photosynthesizers, the green algae, and higher plants that are also classified in class VIII (**Table 1**).¹³¹ The photosynthetic organisms use either the reductive pentose phosphate pathway (Calvin cycle) or Hatch–Slack pathway converting CO₂ into 3PGA as the first intermediate product. P_i, the inhibitor of the photosynthetic ADP-Glc PPases, decreases in concentrations under light conditions as it is utilized to synthesize ATP through photophosphorylation. Therefore, class VIII ADP-Glc PPases are regulated by the 3PGA/P_i ratio under physiological conditions.^{16,107,133–135}

The nonphotosynthetic tissue ADP-Glc PPases of higher plants can be seen as two discrepant types (**Table 1**). The potato tuber ADP-Glc PPases from reserve tissues is typically activated by 3PGA and inhibited by P_i and thus grouped as class VIII (**Table 1**).^{136,137} The ADP-Glc PPases from reserve tissues of cereals have been shown to exhibit distinctive regulatory properties, such as a lower or no sensitivity to activators.^{118,138–142} A detailed characterization of the purified wheat endosperm ADP-Glc PPase has shown that the enzyme is regulated in a different manner by metabolites.¹³⁸ The wheat endosperm enzyme is allosterically inhibited by P_i, ADP, and Fru 1,6-bisP. These inhibitions can be reversed by 3PGA and F6P. But 3PGA and F6P have no effect on enzyme activity in the absence of the inhibitors. The wheat endosperm ADP-Glc PPase thus has distinctive regulatory properties and is placed in class IX (**Table 1**).

Some higher plant enzymes have been shown to be highly insensitive to activation by 3PGA and P_i inhibition. Partially purified barley endosperm ADP-Glc PPase was shown to have low sensitivity to the regulators 3-PGA and P_i.¹¹⁸ However, 3-PGA lowered up to threefold the S_{0.5} for ATP and the Hill coefficient. At 0.1 mmol l⁻¹ ATP the activation by 3-PGA was around fourfold and phosphate 2.5 mmol l⁻¹ reversed the effect.¹⁴³ A recombinant enzyme with a (His)₆-tag from barley endosperm was expressed using the baculovirus insect cell system.¹⁴⁴ It shows no sensitivity to regulation by 3-PGA and P_i. However, the enzyme was assayed at saturating concentration of substrates and only in the pyrophosphorolysis direction. For ADP-Glc PPases the synthetic direction is more sensitive to activation. When the recombinant enzyme without the (His)₆-tag is expressed in insect cells, the heterotetrameric form still was not activated by 3-PGA nor inhibited by P_i at saturating levels of substrates.¹³⁹ Whether 3-PGA had any effect on the affinity for the substrates as shown in the enzyme purified from the endosperm was not reported. Of interest is that the small (catalytic) subunit when expressed alone is very responsive to the allosteric effectors.¹³⁹ This would suggest that the large subunit in barley endosperm desensitizes the small subunit to activation by 3-PGA and inhibition by P_i and that is the opposite seen for large subunits of potato tuber^{106,136} and *Arabidopsis*.¹⁴⁵

ADP-Glc PPase from pea developing embryos was purified to apparent homogeneity (56.5 U mg⁻¹) and was found to be activated up to 2.4-fold by 1 mmol l⁻¹ 3-PGA in the ADP-Glc synthesis direction.¹⁴⁰ In

pyrophosphorolysis, $1 \text{ mmol l}^{-1} \text{ P}_i$ inhibited the enzyme 50% and 3-PGA reversed this effect. The effect of 3-PGA or P_i on the $S_{0.5}$ for ATP was not analyzed.

The cyanobacterial ADP-Glc PPase is homotetrameric in quaternary structure as seen for the ADP-Glc PPases from other bacteria. However, it is regulated in a similar manner, activation by 3-PGA and inhibition by P_i , as the plant enzymes and is more related immunologically, to the plant enzymes.¹³¹ The main difference between the cyanobacterial and plant ADP-Glc PPases is the quaternary structure of α_4 for bacterial enzymes and $\alpha_2\beta_2$ for plants.¹³¹

6.15.5.3 Overlapping Specificity of the ADP-Glc PPase Allosteric Activator Site

In the seven regulatory groups the dominant activators are pyruvate, fructose-6-phosphate, 3-PGA, and fructose-1,6-bisphosphate. All the ADP-Glc PPases isolated from the photosynthetic anaerobic bacteria (except for *R. viridis*) are activated by pyruvate. A number of them are also activated by fructose-6 phosphate, and a few are activated by a third metabolite, fructose-1,6-bisphosphate. Fructose-1,6 bisphosphate is an activator of the enzyme from enteric organisms, as well as those from the *Aeromonads*, *M. luteus*, and *M. smegmatis*, but fructose-6-phosphate, an effective activator of the ADP-Glc PPases in these organisms is not an activator for the enteric enzymes. Also, 3-PGA, a highly effective activator for the enzyme from the cyanobacteria, green algae, and plant tissues, is a poor activator for the enteric enzymes. Conversely, fructose-1,6-bisphosphate activates the plant leaf enzymes, but less effectively than 3-PGA. Whereas $28 \mu\text{mol l}^{-1}$ fructose-1,6-bisphosphate is required for 50% of maximal stimulation of the spinach leaf enzyme at pH 8.5, only $10 \mu\text{mol l}^{-1}$ 3-PGA is required for the same effect.¹⁰⁴ Moreover, the maximum stimulation of V_{max} effected by fructose-1,6-bisphosphate (16-fold) is considerably less than that observed for 3-PGA (58-fold) at pH 8.5. In contrast, fructose-1,6-bisphosphate stimulates 30-fold, the rate of ADP-Glc synthesis catalyzed by *E. coli* B ADP-Glc PPase, while the same concentration of 3-PGA gives only a 1.5-fold stimulation.¹⁴⁴

This overlapping of specificity for the activators in the various ADP-Glc PPase classes suggests that the activator sites for the different groups are similar or related.¹¹² Indeed, looking at the deduced amino acid sequences of many of the ADP-Glc PPases there is much similarity of amino acid sequences, particularly at the allosteric binding sites as well as the substrate binding sites. One can speculate that, during evolution, mutation of the gene in the activator binding site region of the ADP-Glc PPase occurred thus modifying the activator specificity. The pressure for change may have come from a need of compatibility between the metabolite activator and the major carbon assimilation and dissimilation pathways of the organism.

Metabolites associated with energy metabolism, AMP, ADP, or P_i , are inhibitors of the ADP-Glc PPases. The enteric ADP-Glc PPases including those found in the genus *Serratia*¹⁰⁰ and *E. bafniae*, are very sensitive to AMP inhibition. The plant, algal, and cyanobacterial enzymes however, are highly sensitive to P_i . Other ADP-Glc PPases are sensitive either to P_i , ADP, or AMP. Thus, ADP-Glc and glycogen synthesis only proceeds when ATP availability in the cell is high. However, ADP-Glc PPases from *Aeromonads*,¹⁴⁶ *M. smegmatis*,¹⁴⁷ *R. rubrum*,⁹⁷ and *R. molischanum*¹⁴⁸ are not inhibited by P_i , ADP, and AMP, at concentrations of 5 mmol l^{-1} or less.

6.15.5.4 Effect of Activators and Inhibitors on ADP-Glc PPase Kinetics

Kinetic studies of several ADP-Glc PPases show that the presence of the activator in reaction mixtures usually lowers the concentration of the substrates, ATP, Glc-1-P, pyrophosphate ADP-Glc, and the cationic activator Mg^{2+} (or Mn^{2+}), required for 50% of maximal velocity (K_m or $S_{0.5}$) about 2–15-fold. The apparent affinity of the enzyme for the substrate is thus increased in the presence of the activator. The activator also increases maximal velocity and K_{cat} , 2- to 60-fold depending on the pH and the particular ADP-Glc PPase studied. The prime function of the allosteric activator may be, however, to reverse the sensitivity of the enzyme to inhibition by P_i , AMP, or ADP, which are usually noncompetitive with the substrates. Indeed, for many ADP-Glc PPases, relatively high concentrations of the activator can completely reverse the inhibition caused by AMP, P_i , or ADP. A well-studied system is the *E. coli* B enzyme, where fructose-1,6-bisphosphate modulates the sensitivity to AMP inhibition.^{144,149,150} The $I_{0.5}$ (concentration giving 50% inhibition) for 5'-AMP is about $75 \mu\text{mol l}^{-1}$ at 1.7 mmol l^{-1} fructose-1,6-bisphosphate. With lower concentrations of the activator, however, the

$I_{0.5}$ for AMP is lower and, for example, at $60 \mu\text{mol l}^{-1}$ fructose-1,6-bisphosphate, only $3.4 \mu\text{mol l}^{-1}$ AMP is needed for 50% inhibition. For the *Rhodospirillum tenue* ADP-Glc PPase, pyruvate the activator, increases the reaction rate threefold and half-maximal stimulation ($A_{0.5}$) occurs at $28 \mu\text{mol l}^{-1}$.¹⁴⁸ At low concentrations of activator ($15 \mu\text{mol l}^{-1}$), 0.5 mmol l^{-1} AMP inhibits by 90%, but this inhibition is almost completely negated by increasing concentrations of pyruvate. At 1 mmol l^{-1} pyruvate, inhibition is absent. AMP, at $62 \mu\text{mol l}^{-1}$, inhibits the enzyme 50% but in the presence of pyruvate at $50 \mu\text{mol l}^{-1}$, a concentration giving 60% of the maximal velocity, the AMP concentration required for 50% inhibition, $I_{0.5}$, was $260 \mu\text{mol l}^{-1}$. AMP has two effects on the enzyme; it increases the $A_{0.5}$ value of pyruvate (from 28 to $140 \mu\text{mol l}^{-1}$ in the presence of 0.5 mmol l^{-1} AMP) and increases the sigmoidicity of the activation curve. The Hill coefficient increases from 1.0 in the absence of AMP to 1.8 in the presence of 0.5 mmol l^{-1} AMP.

With respect to the potato tuber ADP-Glc PPase similar interactions are seen that also shows that relatively small changes in the concentrations of P_i and 3-PGA can lead to large effects on the rate of ADP-Glc synthesis, particularly at low concentrations of 3-PGA where the activation is minimal in the presence of P_i .¹³⁵ At 1.2 mmol l^{-1} P_i and 0.2 mmol l^{-1} 3-PGA, ADP-Glc synthesis is inhibited over 95%. However, if the P_i concentration decreases by 33% to 0.8 mmol l^{-1} and the 3-PGA concentration increases by 50% to 0.3 mmol l^{-1} there is a 8.5-fold increase in the rate of ADP-Glc synthesis. Conversely, at 0.4 mmol l^{-1} 3-PGA and 0.8 mmol l^{-1} P_i , the rate of ADP-Glc synthesis is 7.5 nmol per 10 min; this is reduced to 2.2 nmol per 10 min (70% decrease) if the 3-PGA concentration decreases by only 50% to 0.2 mmol l^{-1} . If the P_i concentration is also increased to 1.2 mmol l^{-1} (a 50% increase) then the synthetic rate falls to 0.65 nmol that is a reduction of ADP-Glc synthesis of 91%. The reason for the small changes in the effector concentrations giving such large effects in the synthetic rate is due to the sigmoidal nature of the curves particularly at the low concentrations of 3-PGA. In the next section evidence is presented that the ratio of activator/inhibitor modulates the activity of ADP-Glc PPase not only *in vitro*, but also *in vivo* in bacteria and in plants, thus regulating the synthesis of α -1,4 glucans in these systems.

6.15.5.5 Experimental Evidence Supporting the Role of ADP-Glc PPase in the Regulation of the Biosynthesis of Bacterial Glycogen and Plant Starch

The leaf ADP-Glc PPase is highly sensitive to 3-PGA, the primary product of CO_2 fixation by photosynthesis, and to P_i as seen in **Table 2**. Thus, it has been suggested that these compounds play a significant role *in vivo* in regulating starch biosynthesis in higher plants and algae,^{9,12,16,65,102,133–135,145} and in regulating glycogen synthesis in oxygenic photosynthetic microorganisms.^{8–13,25,26,63,65,151}

Table 2 shows the kinetic constant of both substrates ATP and Glc-1-P, and the allosteric effectors, 3-PGA and P_i for the spinach leaf and potato tuber ADP-Glc PPases. For both enzymes there is a great increase of

Table 2 Kinetic constants of ADP-Glc PPase from spinach leaf and potato Tuber

| Enzyme source | Effector/substrate (mmol l^{-1}) | $S_{0.5}/A_{0.5}$ ($\mu\text{mol l}^{-1}$) | $I_{0.5}$ ($\mu\text{mol l}^{-1}$) | Hill's constant, n | Activation fold | Reference |
|---------------|---|--|--------------------------------------|----------------------|-----------------|-----------|
| Spinach leaf | 3-PGA | 0.051 | | 1.0 | 20 | 103 |
| | P_i | | 0.045 | 1.1 | | |
| | P_i (+3-PGA, 1 mmol l^{-1}) | | 0.97 | 3.7 | | |
| | ATP | 0.38 | | 0.9 | | |
| | ATP (+3-PGA) | 0.062 | | 0.9 | | |
| | Glc-1-P | 0.12 | | 0.9 | | |
| | Glc-1-P (+3-PGA) | 0.035 | | 1.0 | | |
| Potato tuber | 3PGA | 0.16 | | 1.0 | 30 | 105 |
| | P_i (–3-PGA) | | 0.04 | NR | | |
| | P_i (+3-PGA, 3 mmol l^{-1}) | | 0.63 | NR | | |
| | ATP (+3-PGA) | 0.076 | | 1.6 | | |
| | Glc-1-P (+3-PGA) | 0.057 | | 1.1 | | |

NR, not reported.

V_{\max} by the activator, 3-PGA. Also, with spinach leaf enzyme 3-PGA increases the apparent affinity of the substrates, ATP and Glc-1-P. This is not so for the potato tuber enzyme. The most important function for 3-PGA most probably is to reverse or overcome P_i inhibition. In the presence of 1 mmol l^{-1} 3-PGA the spinach leaf ADP-Glc PPase is inhibited 50% by $0.97 \text{ mmol l}^{-1} P_i$, a concentration that is about 22 times higher than the value needed for 50% inhibition in the absence of 3-PGA while for the potato tuber enzyme that value is about 15.8.

In plants, the concentration of ATP rises in the light, leading to the formation of sugar phosphates from 3-PGA. At the same time, concentrations of P_i decreases and that of 3-PGA and other glycolytic intermediates increase, thus increasing the activity of the ADP-Glc PPase and starch synthesis. Conversely, in the dark, phosphate concentration increases, while the concentrations of 3-PGA, glycolytic intermediates, and ATP decrease, leading to inhibition of ADP-Glc synthesis and therefore of starch synthesis.

Data showing the correlation between changes in the cellular concentrations of P_i and/or 3-PGA and starch content or rates of starch synthesis have been presented and discussed in previous reviews^{16,107,133–135,152–154} and suggest that *in vivo*, 3-PGA and P_i levels affect starch synthetic rates via modulation of ADP-Glc PPase activity. Since those reviews, evidence of a different nature has been obtained indicating that the regulatory effects seen for the plant and algal ADP-Glc PPase are important in the determination of the rate of starch synthesis *in vivo*. The Kacser–Burns control analysis method^{155,156} alters the enzyme activity, either by using mutants deficient in that enzyme or by varying the physiological conditions and correlating the effect of these changes on the rate of a metabolic process (e.g., starch synthesis). If the enzyme activity is rate limiting or important in controlling the metabolic process, then a large effect on that process should be seen. However, if there is no or little effect, then the activity of that particular enzyme is not considered to be rate limiting for the overall metabolic process being measured. The influence on an enzyme is quantified as a flux coefficient ratio. If the variation of enzyme activity determines a commensurate change in the rate of the process measured, a correlation ratio close to one should be observed.

The Kacser–Burns approach has been used in the analysis of starch synthesis by mutants of *Arabidopsis thaliana*^{157,158} that studied leaf ADP-Glc PPase, and the significance of 3-PGA regulation *in vivo*.¹⁵⁷ *A. thaliana* mutant strains containing only 7% of the wild-type activity of ADP-Glc PPase; and a hybrid of mutant and wild type with 50% activity, had 10% and 61% of the wild type's starch synthetic rate, respectively, at high light intensity.¹⁵⁷ This is a fairly good correlation between the activity of the ADP-Glc PPase and the rate of synthesis of starch. The flux control coefficient was determined to be 0.64.

Despite the fairly high value seen for the ADP-Glc PPase flux control coefficient, it is highly possible that it is underestimated because of the allosteric properties of the enzyme. For flux control analysis, the value of the maximal enzyme activity measured is used in the calculations. In the case of an allosteric enzyme the potential maximal enzyme activity may not be as important as the allosteric effector concentrations determining the actual enzyme activity, *in situ*. With the plant ADP-Glc PPases, activation by 3-PGA can be anywhere from 10- to 100-fold. Moreover, inhibition by the allosteric inhibitor, P_i and variations in the $[3\text{PGA}]/[P_i]$ ratio could cause greater fluctuations in the potential maximal activity. Thus, flux coefficient control values based on only the enzyme's maximal activities of the *A. thaliana* mutants and normal ADP-Glc PPases are likely to underestimate the regulatory potential of the ADP-Glc PPase reaction by the 3-PGA/ P_i ratio. Indeed, the rate of starch synthesis is usually only 10–15% the maximal rate observed with the ADP-Glc PPase activity observed in plant extracts. This is highly suggestive that *in vivo*, the ADP-Glc PPase activity is mainly in an inhibited state.

Important evidence indicating that the *in vitro* activation of the ADP-Glc PPase is truly relevant *in vivo* comes from isolation of a class of mutants where the mutation directly affects the allosteric properties of the ADP-Glc PPase. Such mutants were found easily for the bacteria *E. coli*^{25,159–167} and *S. typhimurium*.^{25,41} **Table 3** compares the maximum amounts of glycogen accumulated by four glycogen-excess mutants of *E. coli* B with the parent strain and the maximum amounts of glycogen accumulated by two *S. typhimurium* LT-2 glycogen-excess mutants with their parent strain. In minimal media having excess glucose, the rate of glycogen accumulated in SG5, CL1136, and 618 is 2-, 3.5-, and 3.7-fold greater, respectively, than that found in *E. coli* B. The *S. typhimurium* LT-2 mutants, JP51 and JP23, accumulate 67 and 25% more glycogen than the parent strain.

Those mutants with a higher accumulation rate as well as having a higher maximal accumulation of glycogen, have ADP-Glc PPases with higher affinity for the activator, fructose 1,6-bis-P (lower $A_{0.5}$) and lower affinity for the inhibitor (higher $I_{0.5}$), AMP.

Table 3 Allosteric kinetic constants of wild-type *E. coli* and *S. typhimurium* LT-2 and their mutant ADP-Glc PPases affected in their allosteric constants

| Strain | Maximal glycogen accumulation ^a mg (g cell) ⁻¹ | Fructose 1,6-bisP ^b A _{0.5} (μmol l ⁻¹) | AMP ^c I _{0.5} (μmol l ⁻¹) | Mutation |
|-----------------------|--|---|---|----------------------|
| <i>E. coli</i> B | 20 | 68 | 75 | |
| Mutant SG14 | 8.4 | 820 | 500 | A44T ¹⁶⁸ |
| Mutant SG5 | 35 | 22 | 170 | P295C ¹⁶⁵ |
| Mutant 618 | 70 | 15 | 860 | G336D ¹⁶⁹ |
| Mutant CL1136 | 74 | 5 | 680 | R67T ¹⁶⁴ |
| <i>S. typhimurium</i> | | | | |
| LT-2 | 12 | 95 | 110 | |
| Mutant JP23 | 15 | No activation | 250 | |
| Mutant JP51 | 20 | 84 | 490 | |

^a The bacterial strains were grown in minimal media with 0.75% glucose and the data are expressed as maximal milligram of anhydroglucose units per gram (wet wt.) of cells in stationary phase.

^b A_{0.5} is the fructose 1,6-bisphosphate giving 50% of maximal activation.

^c I_{0.5} is the AMP concentration required for 50% inhibition.

Also an *E. coli* mutant, SG14, having only 29% the rate of synthesis of glycogen and accumulating only 40% of the maximal amount of glycogen as the parent, *E. coli* B, had an ADP-Glc PPase having about a 12-fold higher A_{0.5} (lower affinity) for fructose 1,6-bisphosphate.¹⁶⁰ The affinity for the inhibitor AMP, was also low (Table 3) and the SG14 mutant had only 25% the apparent affinity for the substrate, ATP (1.6 mmol l⁻¹) as the parent strain.¹⁶⁰ However, the concentration of ATP in *E. coli* is about 3 mmol l⁻¹ allowing the enzyme to have sufficient activity to support glycogen synthesis almost 1/3 the rate seen for the wild-type *E. coli*.

It should be noted that the other glycogen biosynthetic enzymes, glycogen synthase and branching enzyme in the *E. coli* mutant strains were at the same level of activity as observed for the wild-type *E. coli* B and thus could not account for the greater accumulation observed in the glycogen-excess mutants.

The *S. typhimurium* ADP-Glc PPase mutants are of some interest because their altered kinetic properties are different than seen for the *E. coli* allosteric mutants. As seen in Table 3, both activator and inhibitor constants (A_{0.5} and I_{0.5}) are affected by the mutation in *E. coli* mutants SG5, 618, and CL1136. In contrast, the *S. typhimurium* LT-2 mutant JP51 enzyme has an A_{0.5} value for fructose-1,6 bisphosphate similar to that of the parent strain ADP-Glc PPase.⁴¹ However, the JP51 enzyme has about a fivefold higher I_{0.5} value for AMP than the parent enzyme, thus suggesting that the lesser sensitivity to AMP inhibition is the reason for higher accumulation of glycogen in mutant JP51. Mutant JP23 enzyme, in contrast to the parent strain enzyme, is not activated by the fructose bis-P; that is, it is not dependent on fructose-1,6-bisphosphate for full activity.⁴¹ The mutant enzyme also is less sensitive than the parent strain enzyme to AMP inhibition (Table 3) and in addition cannot be inhibited more than 60% of its original activity at higher concentrations of AMP either in the presence or absence of fructose-1,6-bisphosphate. If the levels of AMP in cells are assumed to be 0.15 ± 0.05 mmol l⁻¹, then the JP23 ADP-Glc PPase activity could range from 50 to 60% of its total activity. Under the same conditions in the presence of 1 mmol l⁻¹ fructose-1,6-bisphosphate and range of AMP concentration indicated above, the *S. typhimurium* LT-2 enzyme would exhibit anywhere from 5 to 50% of its total activity.⁴¹ Since JP23 crude extracts have 50% of the activity of the extracts of LT-2, the JP23 enzyme activity would be 0.7–4.9 times as active as the LT-2 enzyme over this concentration range of AMP. Mutant JP23 could therefore accumulate more glycogen than LT-2, based solely on the alteration of the ADP-Glc PPase sensitivity AMP inhibition.

Allosteric mutants have been found for *C. reinhardtii* and for maize endosperm. A significant finding was made by Ball *et al.*,¹⁷⁰ who isolated a starch-deficient mutant of *C. reinhardtii* having an ADP-Glc PPase, which could not be effectively activated by 3-PGA. The inhibition by P_i was similar to the wild type.¹⁷¹ The starch deficiency was observed in the mutant whether the organism was grown photoautotrophically with CO₂ or in the dark with acetate as the carbon source. Thus, the allosteric mechanism seems to be operative for photosynthetic or nonphotosynthetic starch biosynthesis.

Another putative ADP-Glc PPase allosteric mutant from maize endosperm, which has 15% more dry weight (in addition to starch) than the normal endosperm, has been isolated and described by Giroux *et al.*¹⁷² The allosteric mutant ADP-Glc PPase was less sensitive to P_i inhibition than the normal enzyme.

The *Chlamydomonas* starch-deficient mutant and higher dry weight maize endosperm mutant studies strongly suggest that the *in vitro* regulatory effects observed with the photosynthetic and nonphotosynthetic plant ADP-Glc PPases are highly functional *in vivo* and that ADP-Glc synthesis is rate limiting for starch synthesis.

Evidence that the ADP-Glc PPase activity is rate limiting is shown by transformation of various plants with a mutant *E. coli* ADP-Glc PPase much less sensitive to inhibition by AMP.¹⁷³ A chloroplast transit peptide and a promoter was added to the ADP-Glc PPase gene to enable it to be directed to the chloroplast or amyloplast. In tobacco calli, the *E. coli* ADP-Glc PPase activity was detected and starch content increase by 1.7–8.7 times over the controls lacking the gene product.¹⁷³ Similar results were obtained with the transformation of Russet Burbank potato tubers having the ADP-Glc PPase gene containing a transit peptide under the control of a tuber-specific patatin promoter. In this case starch content was increased from 25 to 60% over the controls not containing the bacterial enzyme or the transit peptide.¹⁷³ Of interest is that, when the wild-type ADP-Glc PPase was used no increase in starch content was observed. Thus increasing the starch content required the use of a bacterial gene having an ADP-Glc PPase having allosteric properties where activation was optimized and inhibition was minimized.

Also, in wheat, it has been shown that wheat transformed with allosteric mutant of maize ADP-Glc PPase gave rise to increased starch and to plant biomass by 31%.¹⁷⁴ This was also found for rice as well as for cassava root. Rice was transformed with the allosteric mutant of maize ADP-Glc PPase and this increased biomass.¹⁷⁵ Also, cassava root was transformed with an *E. coli* ADP-Glc PPase allosteric mutant with high affinity for fructose 1,6-bisphosphate¹⁶² and had up to a 2.6-fold increase in tuberous root biomass.¹⁷⁶

Thus, the current data show the importance of the plant ADP-Glc PPase in regulating starch synthesis. Moreover, the allosteric effectors, 3-PGA and P_i , are important *in vivo* effectors in photosynthetic as well as in nonphotosynthetic cells for regulating starch synthesis.

6.15.5.6 Plant ADP-Glc PPases Can Be Activated by Thioredoxin

ADP-Glc PPase from potato tuber has an intermolecular disulfide bridge linking the two small subunits by the Cys12 residue and can be activated by reduction of the Cys12 disulfide linkage.^{177,178} At low concentrations ($10 \mu\text{mol l}^{-1}$) of 3-PGA, both spinach leaf-reduced thioredoxin *f* and *m* reduce and activate the enzyme. Fifty percent activation was seen for 4.5 and $8.7 \mu\text{mol l}^{-1}$ reduced thioredoxin *f* and *m*.¹⁷⁹ The activation was reversed by oxidized thioredoxin. Cys12 is conserved in the ADP-Glc PPases from plant leaves and other tissues except for the monocot endosperm enzymes. In photosynthetic tissues, this reduction may also be physiologically pertinent in the fine regulation of the ADP-Glc PPase.

Table 4 shows the effects of thioredoxin reduction on the kinetic constants of the allosteric effectors and the substrates. The activator constant has a twofold higher apparent affinity (twofold lower $A_{0.5}$) for the reduced enzyme form. The 3-PGA activation curve of the oxidized enzyme was slightly sigmoidal (Hill $n = 1.5$) while the reduced enzyme activation curve for 3-PGA was essentially hyperbolic (Hill $n = 0.8$). Thus the activity at

Table 4 Kinetic constants of the reduced and oxidized forms of potato tuber ADP-Glc PPase¹⁷⁷

| Effector | Redox state of ADP-Glc PPase | |
|------------------|------------------------------|-------------------------------------|
| | Oxidized | Reduced |
| 3-PGA | 84 | $A_{0.5}, \mu\text{mol l}^{-1}$ 40 |
| P_i | 120 | $I_{0.5}, \mu\text{mol l}^{-1}$ 140 |
| ATP | 90 | $S_{0.5}, \mu\text{mol l}^{-1}$ 56 |
| Glc-1-P | 32 | 34 |
| Mg^{2+} | 1830 | 1640 |

The concentration of 3-PGA was $10 \mu\text{mol l}^{-1}$ when the kinetic constants for the substrates were determined. Thioredoxin *f* at $12 \mu\text{mol l}^{-1}$ was used to reduce the enzyme.

lower concentrations of 3-PGA was much greater for the thioredoxin-reduced enzyme. The inhibition by P_i for the reduced and oxidized forms of the enzyme was the same.

Both potato tuber and potato leaf ADP-Glc PPases are plastidic; the leaf enzyme in the chloroplast and the tuber enzyme in the amyloplast.¹⁸⁰ The ferredoxin–thioredoxin system is located in the chloroplast and thus, with photosynthesis, reduced thioredoxin is formed and would activate the leaf ADP-Glc PPase. At night, oxidized thioredoxin is formed and would oxidize and inactivate the ADP-Glc PPase. This activation/inactivation process during the light/dark cycle allows a ‘fine tuning’ and dynamic regulation of starch synthesis in the chloroplasts. Thioredoxin isoforms is present in many different subcellular locations of plant tissues: cytosol, mitochondria, chloroplasts, even nuclei¹⁸¹ as well as in amyloplasts.¹⁸²

It has been shown that potato tubers from growing plants show inhibition of starch synthesis within 24 h after detachment¹⁸³ despite having high *in vitro* ADP-Glc PPase activity plus high levels of substrates, ATP, and Glc 1-P as well as an increased 3-PGA/ P_i ratio. In the detached tubers, the small subunit in nonreducing SDS–PAGE is solely in dimeric form and relatively inactive in contrast to the enzyme form of growing tubers where it was composed as a mixture of monomers and dimers. The detached tuber enzyme had a great decrease in affinity for the substrates as well as for the activator. Treatment of tuber slices with either DTT or sucrose reduced the dimerization of the ADP-Glc PPase small subunit and stimulated starch synthesis *in vivo*. These results indicate that reductive activation, observed *in vitro* of the tuber ADP-Glc PPase^{177,178} is important for regulating starch synthesis.¹⁸³ A strong correlation between sucrose content in the tuber and the reduced/activated ADP-Glc PPase was noted.

6.15.5.7 Identification of Substrate and Effector Sites and Catalytic Residues

6.15.5.7.1 Bacterial systems

Chemical modification studies have been used to obtain information on the catalytic mechanism and on the catalytic sites of the enzyme. Chemical modification studies on the ADP-Glc PPases have involved the use of the following affinity labels: (1) pyridoxal-5-phosphate (PLP), an analogue of either 3-PGA and of fructose 1,6 bisphosphate and (2) the photoaffinity substrate analogues, 8-azido-ATP and 8-azido-ADP-Glc. When 8-azido compounds are irradiated with UV light (257 nm), a nitrene radical is formed which can then react with electron-rich amino acid residues at the catalytic sites inactivating the enzyme^{184,185} and phenylglyoxal, for the identification of arginine residues, perhaps in binding of the anionic substrates.

Also, site-directed mutagenesis studies of residues suspected to be involved in catalytic mechanisms or in substrate binding found from chemical modification or structure prediction^{186,187} and modeling studies.¹⁰⁶

The above studies have provided information on the catalytic and regulatory sites of the spinach ADP-Glc PPase and on the role of the large and small subunits. In ADP-Glc PPase from *E. coli*, Lys residue 195 has been identified as the binding site for the phosphate of Glc-1-P or the β phosphate of ADP-Glc¹⁸⁸ and tyrosine residue 114 has been identified as involved in the binding of the adenosine portion of the other substrate, ATP.¹⁸⁴ Chemical modification and site-directed mutagenesis studies of the *E. coli* ADP-Glc, PPase have provided evidence for the location of the activator-binding site,¹⁸⁹ the inhibitor-binding site,^{190,191} and the substrate binding sites.^{185,188,192} In these experiments pyridoxal-P was used as an analogue for the activator, fructose 1,6-bisphosphate, and for the substrate, Glc-1-P or ADP-Glc. The photoaffinity reagent 8-azido-ATP (8N₃ATP), an ATP analogue, was shown to be a substrate^{184,185} and 8-azido-AMP (8N₃AMP), an effective inhibitor analogue.^{190,191} Since the amino acid sequence of the *E. coli* ADP-Glc PPase gene, *glgC*, is known, the identification of the amino acid sequence around the modified residue helped locate the modified residue within the primary structure of the enzyme. The amino acid residue involved in binding the activator was Lys39¹⁸⁹ and the amino acid involved in binding the adenine portion of the substrates (ADP-Glc and ATP) was Tyr114.¹⁸⁴ Tyr114 was also the major binding site for the adenine ring of the inhibitor, AMP.^{190,191} Because Lys195 was protected from reductive phosphopyridoxylation when the substrate, ADPGlc, was present, it was proposed that Lys195 is also a part of the substrate site.¹⁸⁸ Site-directed mutagenesis of Lys195 with other amino acids showed differences of 100- to 10 000-fold in K_m for Glc-1-P.¹⁸⁸ It was postulated that the epsilon amino group of the Lys195 interacted with the negative phosphate charge of Glc-1-P.¹⁸⁸ Indeed, in recent studies a homology model of the *E. coli* ADP-Glc PPase was developed in order to determine whether those

Table 5 Conservation in plant ADP-Glc PPases of the Glc-1-P¹⁸⁸ and of the ATP binding sites present in *E. coli* ADP-Glc PPase⁶³

| Source | Glc-1-P site | ATP site |
|------------------------|--------------|-----------|
| Prokaryotes | 195 | 114 |
| <i>E. coli</i> | IIEFVEKP-AN | WYRGTADAV |
| <i>S. typhimurium</i> | **D*****_** | ***** |
| <i>Anabaena</i> | V*D*S***KGE | *FQ***** |
| <i>Synechocystis</i> | *TD*S***QGE | *FQ***** |
| Plant small subunit | | |
| <i>A. thaliana</i> | ****A***KGE | *FQ***** |
| Maize endosperm 54 kDa | ****A***KGE | *FQ***** |
| Potato tuber 50 kDa | ****A***QGE | *FQ***** |
| Rice seed | *V**A***KGE | *FQ***** |
| Spinach leaf 51 kDa | ****A***KGE | *FQ***** |
| Wheat endosperm | ****A***KGE | *FQ***** |
| Plant large subunit | | |
| Maize endosperm 60 kDa | VLQ*F***KGA | *FQ****SI |
| Maize embryo | VIQ*S***KGA | *FQ***** |
| Potato tuber 51 kDa | VVQ*A***KGF | *FQ***** |
| Spinach leaf 54 kDa | VLS*S***KGD | *FQ***** |
| Wheat endosperm | VVQ*S*Q*KGD | *FR*****W |

References to these sequences for the plant ADP-Glc PPases are in Smith-White and Preiss.¹¹² The sequences for the *Anabaena* enzyme is in Charng *et al.*¹⁹³ for the *Synechocystis* enzyme in Kakefuda *et al.*¹⁹⁴ and for the wheat endosperm small subunit, in Ainsworth *et al.*¹⁹⁵ The numbers 195 and 114 correspond to Lys195 and Trp114 of the *E. coli* enzyme and * signifies the same amino acid as in the *E. coli* enzyme.

conserved amino acids in the model that are also at the Glc-1-P binding site in 11 bacterial ADP-Glc PPases and four plant and algal ADP-Glc PPases, were important for binding of Glc-1-P.¹⁹² The amino acids Glu194, Ser212, Tyr216, Asp239, Phe240, Trp274, and Asp276 were studied by site-directed mutagenesis and kinetic characterization.¹⁹² The model predicted that Glu194 carboxyl groups hydrogen bonded with the 2 and 3 hydroxyl groups of ADP-Glc while Ser212 binds O-3 and O-4 of the sugar moiety. The Glu194 mutants displayed the greatest changes in the K_m of Glc-1-P of 85- to 388-fold and the Ser212 showed changes in K_m of 14- to 377-fold.¹⁹² These were the greatest changes noted with respect to the other above amino acids mutagenized. Of importance the model overlaps with the Glc-1-P site of other sugar nucleotide pyrophosphorylases such as *Pseudomonas aeruginosa* dTDP-Glc PPase and *Salmonella typhi* CDP-Glc PPase.¹⁹² Thus the study may have implications for other sugar nucleotide pyrophosphorylases.

Table 5 shows a number of the sequences obtained for both prokaryotic and plant ADP-Glc PPases that show great conservation with the proposed Glc-1-P and ATP sites of *E. coli*.

The amino acid sequence for the enzyme from *S. typhimurium* has been deduced from the nucleotide sequence of the cloned gene.¹⁹⁶ Comparison of the nucleotide sequences of the *E. coli* and *S. typhimurium* *glgC* genes shows a 80% identity, and 90% similarity for the amino acid sequence. Most of the changes are conservative and the amino acids known to be involved in the binding of substrates and allosteric effectors, and those involved in maintaining allosteric function in the *E. coli* enzyme are all conserved in *S. typhimurium* as observed in **Table 5**. There is less conservation seen for the Glc-1-P site in the cyanobacterial sequence. However, the Lys residue and the two surrounding amino acids, Glu and Pro are conserved. Moreover, the ATP-binding site sequence is highly conserved both in the enterics and cyanobacteria. The *E. coli* enzyme has been crystallized¹⁹⁷ but the crystals were of poor diffraction quality.

6.15.5.7.2 Plant, algal, and cyanobacterial systems

6.15.5.7.2(i) Substrate sites Since the plant native ADP-Glc PPases are tetrameric and composed of two different subunits it is of interest to know if the two subunits have specific functions and if they are required for optimal catalytic activity. The enzyme contains ligand-binding sites for the activator, 3-PGA, and inhibitor,

P_i , as well as catalytic sites for the two substrates, ATP and Glc-1-P. It is possible that these sites could be located on different subunits. The overall amino acid sequence identity of the *E. coli* enzyme when aligned with the plant and cyanobacterial ADP-Glc PPases ranges from 30 to 33%.¹¹² However, as indicated in **Table 5** there is greater sequence identity when the *E. coli* ATP and Glc-1-P binding sites are compared with the corresponding sequences of the plant indicating that those sequences are also important for the plant enzyme and having the same function. Indeed, when the potato tuber ADP-Glc PPase is expressed in an *E. coli* mutant, deficient in ADP-Glc PPase activity, and a glutamate residue is substituted for lysine residue K198 (equivalent to the *E. coli* ADP-Glc PPase residue K195) via site-directed mutagenesis in the 50 kDa subunit, the K_m for Glc-1-P is increased from $57 \mu\text{mol l}^{-1}$ to over 31 mmol l^{-1} without any change on the K_m or K_a for the other substrates, Mg^{2+} , ATP, or for activator, 3-PGA.¹⁹⁸ These results indicate that Lys residue 198 of the plant ADP-Glc PPase is involved in the binding of Glc-1-P. In the case of the proposed ATP binding site instead of tyrosine there is a phenylalanine residue in the corresponding sequences of the plant and cyanobacterial enzymes. Future site-directed mutagenesis and chemical modification studies are needed to show the functionality of the WFQGTADAV region of the plant enzyme for the binding of ATP.

6.15.5.7.2(ii) Activator sites The covalent modification site of the small subunit by pyridoxal phosphate (PLP) was a lysine residue close to the C-terminus.¹⁹⁹ This may be important for binding of the activator, 3-PGA since 3-PGA (or P_i) inhibited or prevented the covalent modification.¹⁹⁹ When PLP is covalently bound, the plant ADP-Glc PPase is active and no longer requires 3-PGA for further activation. These results showing that the modified enzyme no longer requires an activator for maximal activity and that the covalent modification is prevented by the presence of the allosteric effectors strongly indicate that the activator analogue, PLP, is bound at the activator site. Three lysine residues of the spinach leaf large subunit were identified to be close to the binding site of PLP and, presumably, of the activator, 3-PGA.²⁰⁰ The chemical modifications of these Lys residues by PLP was inhibited by the presence of 3-PGA during the reductive phospho-pyridoxylation process. These three Lys residues are highly conserved in the enzymes from higher plants and cyanobacteria and as shown in **Table 6** similar sequences are seen for activator site 1 of the small subunit and activator sites 1 and 2 of the large subunit.

Similar results were obtained with the *Anabaena* ADP-Glc PPase.¹⁵¹ Chemical modification of the enzyme with PLP caused the cyanobacterial enzyme to have maximal activity independent of activator. Modification by PLP was once again prevented by 3-PGA and P_i . The modified Lys residue was identified as Lys419 and the sequence adjacent to that residue is similar to that observed for the higher plant activator sites 1 and 2 sequences in the large subunit (**Table 6**). Site-directed mutagenesis of Lys419 either to Arg, Ala, Gln, or Glu produced mutant enzymes having 25- to 150-fold lower apparent affinities for activator than that of wild-type enzyme. No other kinetic constants such as K_m for substrates and the inhibitor, P_i , were affected, nor was the catalytic efficiency of the enzyme affected. These mutant enzymes, however, were still activated to a great extent at higher concentrations of 3-PGA, suggesting the presence of an additional binding site for the activator. The Lys419Arg mutant was also chemically modified with the activator analogue, PLP, causing a dramatic alteration in the allosteric properties of the enzyme that could be prevented by having 3-PGA or P_i present during the chemical modification process. Lys382 was the residue modified and for this reason it was concluded that it is the additional site involved in the binding of the activator. As shown in **Table 6** the sequence around Lys382 in the *Anabaena* enzyme is very similar to that seen for the higher plants small and large subunit sites 1.

Site 1 corresponds to the lysyl residue near the C-terminus, Lys440, that is phosphopyridoxylated in the spinach leaf small subunit,¹⁹⁹ and corresponds to Lys441 in the potato tuber ADP-Glc PPase small subunit and to Lys468 in the rice seed small subunit. Site 2 is also situated close to the C-terminus, equivalent to Lys382 in the *Anabaena* ADP-Glc PPase and Lys404 of the potato tuber large subunit. **Table 6** also shows that the amino acid sequence of the spinach leaf small subunit peptide containing the modified lysyl residue of site 1 is highly conserved in the barley,²⁰¹ maize,²⁰² rice seeds, potato tuber and wheat leaf small subunits,¹⁹⁵ and *Anabaena*¹⁹³ and *Synechocystis*¹⁹⁴ ADP-Glc PPase subunits. Similarly, the amino acid sequence of site 2 of the spinach leaf large subunit is highly conserved in the large subunits of the potato tuber, maize,²⁰³ barley endosperm,²⁰⁴ and wheat endosperm ADP-Glc PPases.²⁰⁵

Phenylglyoxal inactivation of the enzyme can be prevented by 3-PGA or by P_i , evidence that one or more arginine residues are present in the allosteric sites of the spinach leaf enzyme, and both subunits were labeled

Table 6 Some higher plant and cyanobacterial ADP-Glc PPase activator binding sites

| Accession no. | Cyanobacteria | Activator site 1 | Activator site 2 |
|----------------------|---------------------------------------|-----------------------------------|-----------------------------------|
| | | 382 | 419 |
| Z11539 | <i>Anabaena</i> | DTIIRRAIIDKNARIG | IVVVLKNAVITDGTII |
| M83556 | <i>Synechocystis</i> | GTTIRRAIIDKNARIG | IVVVIKNVTIADGTVI |
| | | Activator site 1 Small subunit | Activator site 2 Small subunit |
| Higher plants | | | |
| x83550 | Spinach leaf | NSHIKRAIIDKNARIG | IVTVIKDALIPSGTVI |
| X66080 | Wheat leaf | NSHIKRAIIDKNARIG | IVTVIKDALLPSGTVI |
| Z48563 | Barley leaf | NSHIKRAIIDKNARIG | IVTVIKDALLPSGTVI |
| L33648 | Potato tuber, R. Burbank | NCHIKRAIIDKNARIG | IVTVIKDALIPSGIVI |
| | <i>Zea mays</i> , brittle-2 endosperm | NSCIRRAIIDKNARIG | IVTVIKDALLPSGTVI |
| Z48562 | Barley endosperm | NSHIKRAIIDKNARIG | IVTVIKDALLPSGTVI |
| J04960 | Rice endosperm | NCHIRRAIIDKNARIG | IVTVIKDALLLAEQLYE |
| | | Activator site 1 Large subunit | Activator site 2 Large subunit |
| Higher plants | | | |
| | Spinach leaf | IKDAIIDKNAR | ITVIFKNATIKDGVV |
| u66876 | Barley leaf | NTSIQNCIIDKNARIG | ITVVLKNSVIADGLVI |
| x61187 | Potato tuber, R. Burbank | NTKIRKCIIDKNAKIG | IIIILEKATIRDGTVI |
| s48563 | <i>Z. mays</i> , <i>Shrunken 2</i> | NTKIRNCIIDMNARIG | IVVILKNATINECLVI |
| u66041 | Rice endosperm | NTKIRNCIIDMNARIG | IVVILKNATNATIKHGTVI |
| x67151 | Barley endosperm | NTKISNCIIDMNARIG | IVVIQKNATIKDGTVV |
| x14350 | Wheat endosperm | NTKISNCIIDMNARIG | IVVIQKNATIKDGTVV |

The sequences are listed in one letter code and were taken from Ballicora *et al.*¹⁰⁶ and Smith-White and Preiss.¹¹² The Lys residues in bold are those covalently modified by pyridoxal-P in the *Anabaena* and spinach leaf enzymes and are highly conserved in other ADP-Glc PPases. In the case of the potato tuber enzyme the Lys residue was identified via site-directed mutagenesis experiments. The numbers 382 and 419 correspond to the Lys residues in the *Anabaena* ADP-Glc PPase subunit. The activator sites 1 and 2 are present in the small and large subunits of the plant ADP-Glc PPase.

when [14C]-phenylglyoxal was used.²⁰⁶ Where the Arg residue(s) are located in the sequence is presently unknown but there is a possibility it may be close to the Lys residue at activator site 1 of the small and large subunits.

The cDNA clones that encode the putative mature forms of the large and small subunits of the potato tuber ADP-Glc PPase have been expressed together, using compatible vectors, in an *E. coli* mutant devoid of ADP-Glc PPase activity.^{136,137} The ADP-Glc PPase activity expressed was high and the enzyme displayed catalytic and allosteric kinetic properties very similar to the ADP-Glc PPase purified from potato tuber.¹³⁶ Moreover, the enzyme activity was neutralized by antibody prepared against potato tuber and not by antibody prepared against the *E. coli* ADP-Glc PPase.¹³⁷ This expression system is a useful tool to perform site-directed mutagenesis and allow characterization of the allosteric function of the lysyl residues previously identified via chemical modification of the spinach enzyme with PLP. Site-directed mutagenesis of Lys441 of the potato ADP-Glc PPase small subunit to Ala and Glu results in mutant enzymes with lower affinities for 3-PGA in the synthesis direction, 30- to 83-fold, respectively.¹⁰⁶

6.15.5.7.3 Properties and functions of the small and large subunits of higher plant ADP-Glc PPases

The ability to express cDNA clones representing the potato tuber small and large subunits together in *E. coli*¹³⁷ to obtain a highly active enzyme also allows the expression of the subunits separately to determine their specific functions. The potato tuber small subunit, when expressed alone, had high catalytic activity if the 3-PGA concentrations were increased to 20 mmol l⁻¹.¹³⁶ The saturating 3-PGA concentration for the expressed transgenic or natural potato tuber heterotetrameric enzyme is about 3 mmol l⁻¹. The K_a of the transgenic

heterotetrameric enzyme in ADP-Glc synthesis is 0.16 mmol l^{-1} , while for the small subunit alone it is 2.4 mmol l^{-1} . Thus, the small subunit by itself has a 15-fold lower apparent affinity for the activator. The small subunit is more sensitive to P_i inhibition than the transgenic heterotetrameric enzyme having a eightfold lower K_i . The kinetics of 3-PGA activation and the P_i inhibition were the main kinetic differences between the homotetrameric small subunit and the recombinant heterotetrameric ADP-Glc PPase. These results are consistent with those obtained for the *Arabidopsis thaliana* mutant ADP-Glc PPase lacking the large subunit where the enzyme had lower affinity for the activator and higher sensitivity toward P_i inhibition than the heterotetrameric normal enzyme.²⁰⁷

The potato tuber large subunit expressed alone had negligible activity. It seems that the dominant function of the small subunit is catalysis while the dominant function of the large subunit in the cloned potato tuber enzyme is to modulate the sensitivity of the small subunit to the allosteric activation and inhibition.

It is interesting to note that the similarity of the small and large subunits, ~50–60% identity suggests a common origin.¹¹² The small subunit has catalytic activity in both sink and source tissues. Recently, catalytic activity has been observed for the large subunits that reside in the leaf and not in the sink large subunits.²⁰⁸ Gene duplication and divergence has probably led to different and functional roles catalytic and regulatory for the subunits. The ancestor of small and large subunits likely is a bacterial subunit having both catalytic as well as regulatory function in the same subunit. This is supported by the similarity between the two plant subunits with many active bacterial ADP-Glc PPases.^{8,9}

The large subunit from the potato (*Solanum tuberosum* L.) tuber ADP-Glc PPase was shown to bind substrates.¹⁰⁵ The plant heterotetramers therefore, as well as bacterial homotetramers, bind four ADP-[¹⁴C]glucose molecules.^{105,119} It can be hypothesized that the large subunit maintained its structure required for substrate binding but catalytic ability was eliminated by mutations of the residues essential for catalysis. For testing of this hypothesis, it was attempted to create a large subunit with significant catalytic activity by mutating as few residues as possible.²⁰⁹

Thus, sequence alignments of ADP-Glc PPase large and small subunits with reported activity were aligned to identify critical missing residues for catalytic activity in the large subunit.²⁰⁹ The subset of amino acids missing in the large subunit was of special interest. Lys44 and Thr54 in the large subunit of potato tuber were selected as the best candidates to study because the homologous residues, Arg33 and Lys43 in the small catalytic subunit, were completely conserved in the active bacterial and plant small subunits. Lys44 and Thr54 are also in a highly conserved region of ADP-Glc PPases (Table 7).

Table 7 Sequence comparison ADP-Glc PPase subunits in a conserved region with critical amino acids for catalysis and Glc-1-P binding²⁰⁹

| | | |
|-----|----------------|---|
| (a) | | |
| | <i>E. coli</i> | 26LAGGRGTRLKDLTNKRAKPAVH47 |
| | PSS | 26LGGGAGTRLYPLTKKRAKPAVP47 |
| | APS1 | 96LGGGAGTRLYPLTKKRAKPAVP117 |
| | APL1 | 96LGGGAGTRLFPLTKRRAKPAVP117 |
| | APL2 | 91LGGGAGTRLFPLTSKRAKPAVP112 |
| | APL3 | 95LGGGDGAKLFPLTKRAATPAVP116 |
| | APL4 | 97LGGGNGAKLFPLTMRAATPAVP118 |
| | PLS | 38LGGGEGTKLFPLTSRTATPAVP59 |
| (b) | | |
| | PSS | 188EEGRIIIEFAEK PQ GEQLQAMKVD T 211 |
| | APS1 | 257EEGRIIIEFAEK PK GEHLKAMKVD T 270 |
| | APL1 | 261DKGRV ISFSEK PKGDDLKAMAVD T 284 |
| | APL2 | 257QSGK IIQFSEK PKGDDLKAMQVD T 280 |

Sequence comparison of the catalytic site (A) and Glc-1-P binding region (B) of ADP-Glc PPases from *E. coli*, *Arabidopsis*, and potato. Sequences and their accession numbers are: *E. coli*, P00584; PSS, potato tuber small subunit, CAA88449; APS1, P55228; APL1, P55229; APL2, P55230; APL3, P55231; APL4, Q9SIK1; PLS, potato tuber large subunit, Q00081. The conserved arginines and lysines are indicated in bold in (a). The critical K residues involved in Glc-1-P binding are indicated in bold in (b).

Table 8 Activity of small and large subunit potato tuber ADP-Glc PPase mutants

| Subunits | | Units (mg) | |
|----------|-----------|-------------------|---------------|
| Small | Large | ADP-Glc synthesis | ATP synthesis |
| WT | WT | 32 ± 1 | 49 ± 2 |
| D145N | WT | 0.017 ± 0.001 | 0.037 ± 0.002 |
| D145N | K44R | 0.031 ± 0.001 | 0.033 ± 0.002 |
| D145N | T54K | 0.92 ± 0.08 | 0.56 ± 0.03 |
| D145N | K44R/T54K | 3.2 ± 0.2 | 9.0 ± 0.7 |
| R33K | WT | 3.6 ± 0.2 | 4.1 ± 0.1 |
| K43T | WT | 0.32 ± 0.1 | 0.28 ± 0.01 |

The enzyme activities of purified co-expressed small and large subunits were measured for ADP-Glc synthetic activity or ATP synthesis by pyrophosphorolysis of ADP-Glc. For ADP-Glc synthesis, 4 mmol l⁻¹ 3-PGA (activator), 2 mmol l⁻¹ ATP and 0.5 mmol l⁻¹ Glc-1-P were used. For pyrophosphorolysis, 4 mmol l⁻¹ 3-PGA, 1.5 mmol l⁻¹ ADP-Glc, and 1.4 mmol l⁻¹ P_i were used

Lys44 and Thr54 of the large subunit were mutagenized to Arg44 and Lys54, respectively. The mutant, LargeK44R/T54K, expressed in the absence of the small subunit had no activity, probably due to failure of the large subunit to form a tetramer by itself without the presence of the small subunit. This has been previously observed with the *Arabidopsis* ADP-Glc PPase.¹⁴⁵ Because wild-type small subunit has intrinsic activity, the activity of the large subunit mutants cannot be tested when coexpressed with active small subunit. Thus, the large subunit mutants were coexpressed with an inactive form of small subunit mutagenized at the catalytic residue Asp145, to D145N²⁰⁹ reducing by more than three orders of magnitude small subunit activity (Table 8). Coexpression of the large subunit double mutant K44R/T54K with small D145N generated an enzyme having 10 and 18% of the wild-type enzyme in the ADP-glucose synthetic and pyrophosphorolytic directions, respectively (Table 8). Single mutations of K44R or T54K provided enzymes with no significant activity. Therefore, only two residues Arg44 and Lys54 were needed for restoring catalytic activity to the large subunit. Consistent with this was the finding that replacement of the homologous two residues with Lys and Thr in the small subunit (mutations R33K and K43T) decreased the activity by one and two orders of magnitude, confirming the essential function of Arg and Lys in those residues for catalytic activity (Table 8). The mutant enzymes were still activated by 3-PGA and inhibited by orthophosphate (P_i) as seen for the wild-type enzyme. The wild-type enzyme and smallD145NlargeK44R/T54K had very similar kinetic properties indicating that the substrate site domain was not altered. The apparent affinities for the substrates and the allosteric properties of small subunit D145NlargeK44R/T54K were similar to those of the wild type.²⁰⁹ The new form has a similar sensitivity to P_i inhibition and the activation–inhibition interaction were the same. The L subunit restored enzyme activity due to only two mutations is evidence that the large and small subunits are derived from the same ancestor. The smallD145NlargeK44R/T54K mutant was disrupted in each subunit at their Glc-1-P site and their kinetic properties compared. With wild-type enzyme replacement of Lys198 in the small subunit of the wild-type enzyme, decreased the Glc-1-P affinity, whereas disruption of the homologous residue Lys213 in the large subunit had much less effect.¹⁷⁷ In smallD145NlargeK44R/T54K mutant, the K213R mutation of large subunit severely decreased the apparent affinity for Glc-1-P, whereas the K198R mutation on the small subunit did not indicate the large subunit double mutant, and not smallD145N, was the catalytic subunit.²⁰⁹ In the wild-type enzyme, Lys213 does not seem to play any important role, but smallD145NlargeK44R/T54K recovered its ancestral ability for the enzyme to have a low physiological *K_m* for Glc-1-P. Previous results showed that in wild type, Asp145 of small subunit is essential for catalysis, but homologous Asp160 in the large subunit is not.²¹⁰ However, mutation of D160 to N or E in the active large subunit, LK44R/T54K caused loss of activity. This confirms that catalysis as well as binding of substrates of the mutant enzyme, smallD145Nlarge K44R/T54K occurs at the large subunit.

A model of LK44R/T54K shows the predicted role of Arg44 and Lys54 in catalysis. Asp160 of large subunit, that is homologous to Asp145 catalytic residue in the small subunit and to catalytic Asp142 in the *E. coli* ADP-Glc PPase,^{210,211} interacts with Lys54. This interaction (Lys54-Asp160) has also been observed in crystal structures of enzymes catalyzing similar reactions, (dTDP-glucose pyrophosphorylase (dTDP-Glc PPase),

UDP-n-acetylglucosamine pyrophosphorylase (UDP-GlcNAc PPase)), and is postulated to be important for catalysis by correctly orienting the catalytic aspartate residue.^{210–212} Lys54 interacts with the oxygen bridging the α - and β -phosphates as it has been observed in the crystal structure of *E. coli* dTDP-Glc PPase.²¹⁴ The interaction would neutralize a negative charge density stabilizing the transition state and making PP_i a better leaving group. Arg44 interacts in the model with the β - and γ -phosphates of ATP, which correspond to the PP_i by-product. Similarly, Arg15 in the *E. coli* dTDP-Glc PPase was postulated to contribute to the departure of PP_i.¹⁵⁵ and kinetic data agree with interaction of PP_i with Arg44 in the model. A Lys44, in both the catalytic large subunit mutant and the small subunit, decreased the apparent affinity for PP_i at least by 20-fold.²⁰⁹ In wild-type large subunit, Lys44 and Thr54 cannot interact as does Arg44 and Lys54.

It is interesting to note that the *Arabidopsis* leaf large subunits are catalytically active as well as the small subunits.²⁰⁸ As seen in **Table 7** the leaf large subunits have Arg and Lys residues in the equivalent residues as seen for the catalytically active small residues. In contrast, the catalytically inactive large subunits found in reserve or sink tissues have a Lys and Thr in those respective positions (**Table 7**).

In plants there is usually only one conserved small (catalytic) subunit. However there may be several large (regulatory), subunits that can be distributed in different parts of the plant.^{145,215} This is of physiological significance as expression of different large subunits in different plant tissues may confer distinct allosteric properties to the ADP-Glc PPase needed for the plant tissue's distinct need for starch. Coexpression of *Arabidopsis* ADP-Glc PPase small subunit, APS1, with the different *Arabidopsis* large subunits, APL1, APL2, APL3, and APL4, produced heterotetramers having different regulatory and kinetic properties¹⁴⁵ (**Table 9**). The heterotetramer of small subunit APS1 with APL1, the predominant leaf large subunit, had highest sensitivity to the allosteric effectors, 3-PGA and P_i, as well as the highest apparent affinity for the substrates ATP and Glc-1-P.¹⁴⁵ The heterotetrameric pairs of APS1 with either APL3 or APL4, large subunits, prevalent in sink or storage tissues,²¹⁵ had intermediate sensitivity to the allosteric effectors and intermediate affinity for the substrates ATP and Glc-1-P.¹⁴⁵ APL2 also present mainly in sink tissues had low affinity for either 3-PGA or P_i.¹⁴⁵ Thus, differences on the regulatory properties conferred by the *Arabidopsis* large subunits are found *in vitro*. These differences noted for source and sink large subunit proteins strongly suggest that starch synthesis is modulated in a tissue-specific manner in response to 3-PGA and P_i, as well as to the substrate levels. APS1 and APL1 would be finely regulated in source tissues by both effectors and substrates. In sink tissues, the heterotetramers of APS1 with APL2, APL3, or APL4 would have lower sensitivity with respect to effectors and substrates and would be modulated more by the supply of substrates.

In addition, ApS1 is the main small subunit or catalytic isoform responsible for ADP-Glc PPase activity in all tissues of the plant based on mRNA expression.²¹⁵ APL1 is the main large subunit in leaf tissues, whereas APL3 and APL4 are the main isoforms present in reserve tissues. Also, sugar regulation of ADP-Glc PPase genes is restricted to APL3 and APL4 in leaves.²¹⁵ Sucrose induction of APL3 and APL4 transcription in leaves permits formation of heterotetramers that are less sensitive to the allosteric effectors, simulating the situation in sink tissues.

Table 9 Kinetic parameters for 3-PGA of the *A. thaliana* recombinant ADP-Glc PPases in the pyrophosphorolysis direction

| 3-PGA | Control | nH | 0.2 mmol l ⁻¹ P _i | | 2 mmol l ⁻¹ P _i | |
|-----------|---------------------------------------|-----|---|-----|---------------------------------------|-----|
| | A _{0.5} mmol l ⁻¹ | | A _{0.5} mmol l ⁻¹ | nH | A _{0.5} mmol l ⁻¹ | nH |
| APS1 | 1.2 ± 0.09 | 1.8 | 5.6 ± 0.29 | 2.5 | N.D. | |
| APS1/APL1 | 0.0017 ± 0.0005 | 1.0 | 0.019 ± 0.0019 | 1.9 | 0.48 ± 0.005 | 2.7 |
| APS1/APL2 | 0.219 ± 0.024 | 0.8 | 0.820 ± 0.12 | 0.9 | 6.95 ± 1.59 | 1.5 |
| APS1/APL3 | 0.029 ± 0.009 | 0.6 | 0.105 ± 0.025 | 0.8 | 0.29 ± 0.048 | 1.0 |
| APS1/APL4 | 0.030 ± 0.003 | 0.9 | 0.110 ± 0.008 | 1.0 | 0.80 ± 0.167 | 1.2 |

The kinetic parameters were calculated without inhibitor (P_i), and in the presence of inhibitor at 0.2 mmol l⁻¹ or 2 mmol l⁻¹. The deviation in the 3-PGA A_{0.5} data is the difference between duplicate experiments. N.D. indicates not determined. APS1 is the *Arabidopsis* small (catalytic) subunit and APL1, APL2, APL3, and APL4, the *Arabidopsis* large subunits.

6.15.5.7.4 Phylogenetic analysis of the large and small subunits

A phylogenetic tree of the ADP-Glc pyrophosphorylases present in photosynthetic eukaryotes may also shed more information about the origin of the two subunits. The tree shows that plant small and large subunits can be divided into two and four distinct groups, respectively.²⁰⁹ The two main groups of S subunits are from monocot and dicot plants. Large subunit groups however gives better correlation with their documented tissue expression. The first large subunit group, group I, is generally expressed in photosynthetic tissues²⁰⁹ and comprises large subunits from both dicots and monocots. These subunits recently have been shown to have catalytic activity and have in their sequences, Arg and Lys, as seen in the equivalent residues of 102 and 112 of *A. thaliana* large subunit, APL1.²⁰⁸ Group II displays a broader pattern of expression, and groups III and IV are expressed in storage organs (roots, stems, tubers, and seeds). Subunits from group III are only from dicot plants, and group IV are seed-specific subunits from monocots. The last two groups stem from the same branch of the phylogenetic tree and separated before monocot and dicot separation. These subunits are probably inactive in catalytic activity as they lack Arg and Lys in the homologous residues seen in *A. thaliana* APL1 and APL2.²⁰⁸

6.15.6 Subcellular Localization of ADP-Glc PPase in Plants

In green algae and in leaf cells of higher plants ADP-Glc PPase has been demonstrated to reside in the chloroplast.³⁰ More recently, using plastids isolated from maize and barley endosperm, the existence of two ADP-Glc PPases, a plastidial form and a major cytosolic form were found.^{216–218} Subsequently, cytosolic forms of ADP-Glc PPase have been found in wheat^{138,219} and rice.²²⁰ Since starch synthesis occurs in plastids, it was further proposed that in cereal endosperms, synthesis of ADP-Glc in the cytosol requires the involvement of an ADP-Glc carrier in the amyloplast envelope.²¹⁸ Subsequently, characterization, of the ADP-Glc transporter has been reported for maize,^{221,222} barley,²²³ and wheat endosperm.²²⁴

Although the major activity of ADP-Glc PPase of the endosperm of cereals is predominantly in the cytosol, however, there still is significant activity found in the endosperm plastids. If certainly the cytosolic form of ADP-Glc PPase is needed for starch synthesis²¹⁸ what is the function of the plastidial ADP-Glc PPase? An interesting report is the study of the ADP-Glc PPase small and large subunit isoforms in rice.²²⁵ The rice ADP-Glc PPase family was shown to consist of two small subunit genes and four large subunit genes.²²⁵ Subcellular localization studies using green fluorescent protein fusion constructs showed that one small subunit of the ADP-Glc PPase was localized in the plastid with three of the ADP-Glc PPase large subunits while a small subunit and one large subunit were localized in the endosperm cytosol. A lesion of either subunit localized in the endosperm cytosol led to a great reduction in endosperm starch.²²⁵ Also, a model was proposed showing the various locations of the large and small subunits in the cell.²²⁵ In the model a specific small subunits s2a and a large subunit, L3 are confined to the leaf and present in the chloroplast. In the early stage of rice development, S1, the small subunit and L1 the large subunit, were confined to the endosperm amyloplast and had a major role in starch synthesis, while a small subunit s2b and a large subunit L2 were present in the cytosol with a minor role in starch synthesis. During the middle-to-late stages of the developing rice endosperm the amyloplast S1 and L1 subunits played a minor role in starch synthesis while the cytosolic small, s2b and large, L2 subunits now played a major role with respect to ADP-Glc and starch synthesis.

6.15.7 Crystal Structure of Potato Tuber ADP-Glc PPase

The crystal structure of potato tuber homotetrameric small (catalytic) subunit ADP-Glc PPase has been determined to 2.1 Å resolution.²²⁶ The structures of the enzyme in complex with ATP and ADP-Glc were determined to 2.6 and 2.2 Å resolution, respectively. Ammonium sulfate was used in the crystallization process and was found tightly bound to the crystalline enzyme. It was also shown that the small subunit homotetrameric potato tuber ADP-Glc PPase was also inhibited by inorganic sulfate with the $I_{0.5}$ value of 2.8 mmol l⁻¹ in the presence of 6 mmol l⁻¹ 3-PGA.²²⁶ Sulfate is considered as an analogue of phosphate, the allosteric inhibitor of plant ADP-Glc PPases. Thus, the atomic resolution structure of the ADP-Glc PPase probably presents a conformation of the allosteric enzyme in its inhibited state. The crystal structure of the potato tuber ADP-Glc

PPase²²⁶ allows one to determine the location of the activator and substrate sites in the three-dimensional structure and their relation to the catalytic residue, Asp145. The structure also provides insights into the mechanism of allosteric regulation.

6.15.8 Bacterial Glycogen Synthase

6.15.8.1 Enzyme structure and Properties

The *E. coli* B²²⁷ and *E. coli* K12²²⁸ glycogen synthases have been purified to homogeneity. The enzyme from *E. coli* B has a specific activity of about 115 μmol of glucose transferred to glycogen per milligram of protein per minute.²²⁷ The subunit molecular weight was determined by SDS-gel electrophoresis to be about 49 000. Sucrose density-gradient centrifugation showed aggregated forms of the enzyme of 98 000, 135 000, and 185 000.²²⁷ Thus, *E. coli* B glycogen synthase can exist as dimers, trimers, and tetramers.

The bacterial glycogen synthase differs from the mammalian glycogen synthase in two major aspects. The bacterial enzyme has no regulatory properties, and does not exist in either inactive or active form as does the mammalian glycogen synthase. There is no evidence for either phosphorylation or dephosphorylation of the bacterial enzyme or for any other enzyme-catalyzed modification. Neither is the bacterial enzyme activated by Gcl-6-P or other glycolytic intermediates.^{229–232} Moreover, the bacterial enzyme uses ADP-Glc as its physiological sugar nucleotide glucosyl donor³⁶ rather than UDP-Glc, the physiological glucosyl donor of the mammalian glycogen synthase reaction. In one reports²³³ the sugar nucleotides UDP-Glc, CDP-Glc, GDP-Glc, and TDP-Glc had less than 1% of the activity measured for ADP-Glc. Deoxy-ADP-glucose, a nonphysiological analogue of ADP-Glc, has been found to be active to the same extent as ADP-Glc with the *Arthobacter viscosus* glycogen synthase²²⁹ as well as with the *E. coli* B enzyme.²³⁰

The K_m values reported for ADP-Glc range from 20 $\mu\text{mol l}^{-1}$ for the glycogen synthase from *M. smegmatis*²³² and 25–35 $\mu\text{mol l}^{-1}$ for the enzyme from *E. coli* B²²⁷ to 0.4 mmol l^{-1} for the *Pasteurella pseudotuberculosis* glycogen synthase.²³¹ The K_m for deoxy-ADP-glucose is 38 $\mu\text{mol l}^{-1}$ for *E. coli* B glycogen synthase and 27 $\mu\text{mol l}^{-1}$ for the *A. viscosus* enzyme. It has been shown that the bacterial glycogen synthase is very sensitive to inhibition by *p*-hydroxymercuribenzoate, and is also inhibited by ADP that competes with ADP-Glc, with a K_i (for the *E. coli* system) of 15 $\mu\text{mol l}^{-1}$.²²⁹

Several α -glucans can serve as effective primers for the glycogen synthase, for example, glycogen from animal and bacterial sources, amylose, and amylopectin. Lower molecular weight maltodextrins can also serve as primers; maltotriose is quite effective as a primer, and maltotetraose has been identified as the immediate product. Maltose is effective as a primer only at high concentrations, and maltotriose is the immediate product. Glucose is not active as an acceptor of glucose in the bacterial glycogen synthase reaction.

6.15.8.2 Substrate and Catalytic Sites

As indicated above, *E. coli* and *S. typhimurium* glycogen synthases are specific for the sugar nucleotide, ADP-Glc. Thus, an affinity analogue of ADP-Glc, adenosine diphosphopyridoxal (ADP-pyridoxal), was used to identify the ADP-Glc-binding site. Incubation of the enzyme with the analogue plus sodium borohydride led to an inactivated enzyme with the degree of inactivation correlating with the incorporation of about one mol of analogue per mol of enzyme subunit for 100% inactivation. After tryptic hydrolysis one labeled peptide was isolated and the modified Lys residue was identified as Lys15.²³⁴ The sequence, Lys-X-Gly-Gly, where lysine is the amino acid modified by ADP-pyridoxal, has also been found to be present in the mammalian glycogen synthase^{235,236} and the plant starch synthases.²³⁷

Substitution of other amino acids for Lys at residue 15 suggested that the Lys residue is mainly involved in binding the phosphate residue adjacent to the glycosidic linkage of the ADP-Glc and not in catalysis. The major effect on the kinetics of the mutants at residue 15 was the elevation of the K_m of ADP-Glc, of about 30- to 50-fold when either Gln or Glu were the substituted amino acids. Substitution of Ala for Gly at residue 17 decreased the catalytic rate constant, K_{cat} , about three orders of magnitude compared to the wild-type enzyme. Substitution of Ala for Gly18 only decreased the rate constant by 3.2-fold. The K_m effect on the substrates, glycogen, and ADP-Glc, were minimal. It was postulated²³⁸ that the two glycyl residues in the conserved

Lys-X-Gly-Gly sequence participated in the catalysis by assisting in maintaining the correct conformational change of the active site or by stabilizing the transition state.

The structural gene for the *E. coli* and *S. typhimurium* glycogen synthases, *glgA*, has been cloned^{101,196} and the nucleotide sequence of the *E. coli glgA* gene has been determined.²³⁹ It consists of 1431 bp specifying a protein of 477 amino acids with a molecular weight of 52 412.

Since there is still binding of the ADP-Glc and appreciable catalytic activity of the Lys15Gln mutant, the ADP-pyridoxal modification was repeated and in this instance about 30 times higher ADP-pyridoxal concentration was needed for inactivation of the enzyme.²⁴⁰ The enzyme was maximally inhibited about 80% and tryptic analysis of the modified enzyme yielded one peptide containing the affinity analogue and with the sequence, Ala-Glu-Asn-modified Lys-Arg. The modified Lys was identified as Lys277. Site-directed mutagenesis of Lys277 to form a Gln mutant was done and the K_m for ADP-Glc was essentially unchanged but K_{cat} was decreased 140-fold. It was concluded that Lys residue 277 was more involved in the catalytic reaction than in substrate binding.

Chemical modification studies²⁴¹ have shown that two distinct sulfhydryl groups are necessary for enzyme activity and they were protected by primer, glycogen, and by substrate, ADP-Glc, respectively. The reactive sulfhydryl residues were reasoned to be at or near the binding sites for the substrates, glycogen, and ADP-Glc. To determine the responsible residue, all cysteines present in the enzyme, Cys7, Cys379, and Cys408, were mutated with Ser. 5,5'-dithiobis(2-nitrobenzoic acid) modified and inactivated the enzyme only when Cys379 was present.²⁴² The inactivation was prevented by the substrate ADP-Glc. Mutations C379S and C379A decreased the specific activity, 5.8- and 4.3-fold, respectively, and increased the $S_{0.5}$ for ADP-Glc, 40- and 77-fold. Inhibition studies by Glc-1-P and AMP suggested that Cys379 was involved in the interaction of the enzyme with the phosphoglucose moiety of ADP-Glc. Other mutations, C379T, C379D, and C379L, showed that the site is intolerant for bulkier side chains.

Cys379 is in a conserved region. Thus, other residues were scanned by mutagenesis. Substitution of Glu377 by Ala and Gln decreased V_{max} more than 10 000-fold but did not affect the apparent affinity for ADP-Glc and glycogen binding (Table 10). Mutation of Glu377 by Asp decreased V_{max} by only 57-fold showing that the carboxyl group of Glu377 is essential for catalysis. The mutant, E377C, in an enzyme form having no other Cys residues, was inactive (Table 10). However, when carboxymethylated by iodoacetate the mutant enzyme was reactivated (Table 10). This also indicated a carboxyl residue was necessary for activity. No appreciable effect on activity was observed when another conserved residue, Ser374, in the region was mutagenized. However another conserved amino acid, Gln383, when mutagenized to alanine caused a 23-fold decrease in V_{max} but no decrease in the affinity for the substrates.²⁴²

Glycogen synthases are glucosyl transferases that retain the anomeric configuration of glucosyl linkage of ADP-Glc to the nonreducing glucosyl linkage of glycogen.²⁴³ The *E. coli* glycogen synthase was therefore modeled based on three other crystallized glycosyl linkage retaining glycosyltransferases having a GT-B fold.²⁴⁴ Comparison of the model with the structures of the active site of retaining GT-B glycosyltransferases

Table 10 Kinetic constants of *E. coli* glycogen synthase mutants

| Organism | V_{max} | | ADP-Glc | Glycogen |
|-------------|--|---------------|---------------------------------|-------------------------------|
| | $\mu\text{mol min}^{-1}\text{mg}^{-1}$ | Fold decrease | K_m $\mu\text{mol l}^{-1}$ | K_m μgml^{-1} |
| Wild type | 500 | 1 | 18 | 770 |
| E377Q | 0.02 | 25 000 | 61 | 14 |
| E377A | 0.05 | 10 000 | 70 | 27 |
| E377D | 10 | 57 | 1500 | 26 |
| E377C | 0.02 | 23 000 | 213 | 24 |
| E377C + IAA | 50 | 10 | 810 | 70 |
| D137A | 0.07 | 8140 | 21 | 91 |
| H161A | 0.8 | 710 | 200 | 52 |
| R300A | 0.22 | 2590 | 150 | 71 |
| K305A | 0.46 | 1240 | 32 | 39 |

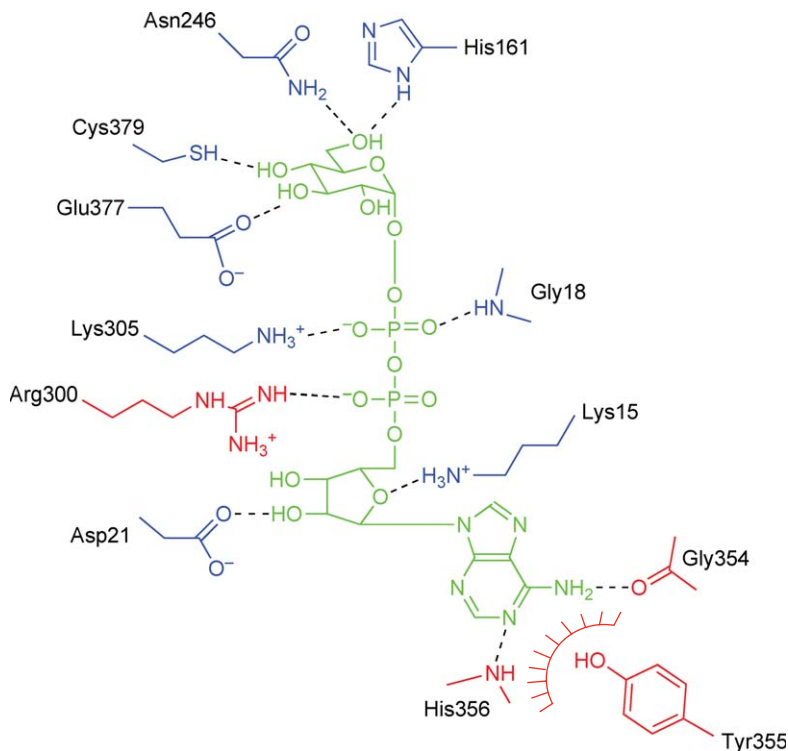


Figure 1 The proposed ADP-Glc binding domain of the glycogen synthase.²⁴⁵ The amino acids interacting with AMP portion of the sugar nucleotide are in red. Those binding with the Glc-1-P portion are in blue.

showed conserved residues in glycogen synthase with the same geometrical orientation. These residues in the *E. coli* glycogen synthase were mutagenized with respect to their importance as predicted by the model. Mutations D137A, R300A, K305A, and H161A decreased the specific activity 8100-, 2600-, 1200-, and 710-fold, respectively. These mutations did not affect the K_m for glycogen and only H161A and R300A had a higher K_m values for ADP-Glc of 11- and 8-fold, respectively (Table 10). These residues were essential, validating the model that shows a strong similarity between the active site of *E. coli* glycogen synthase and the other retaining GT-B glycosyltransferases known to date.^{242,244,245}

Recently, the glycogen synthase crystal structure has been elucidated²⁴⁶ that confirmed the correctness of our proposed model.^{242,244,245} Figure 1 presents a diagrammatic scheme of the proposed binding site for ADP-Glc in the glycogen synthase.²⁴⁵

6.15.8.3 Reversibility of the Glycogen Synthase Reaction

The glycogen synthase reaction has been shown to be reversible.²²⁷ Labeled ADP-Glc occurs from incubation of [14C]-ADP with unlabeled glycogen or from [14C]-glycogen plus ADP. The ratio of ADP to ADP-Glc at equilibrium at 37° has been shown to vary threefold in the pH range 5.27–6.82, suggesting that in the formation of a new α -1,4-glycosidic bond, a proton is liberated. Since the pK_a of ADP^{2-} ionizing to $ADP^{3-} + H^+$ is about 6.4, a proton would be liberated in varying amounts in the range of pH 5.4–7.4, and in stoichiometric amounts above pH 7.4. If the pK_a of ADP^{2-} is assumed to be 6.4,²²⁷ then the ratio of ADP^{2-} to ADP-Glc can be calculated, and it is constant; in the range of pH 5.27–6.82 and was 45.8 ± 4.5 . The constancy of this equilibrium ratio suggests that ADP^{2-} is the reactive species in the reaction.

6.15.9 Starch Synthases

Starch synthases catalyze the transfer of the glucosyl moiety of the sugar nucleotide, ADP-Glc either to a maltosaccharide or glycogen or the starch polymers, amylose and amylopectin, forming a new α -1,4-glucosidic linkage. Since the glucosyl anomeric linkage in ADP-Glc is α and the newly formed glucosidic bond is also α , the starch synthase as the glycogen synthase is considered to be a retaining GT-B glycosyltransferase according to the nomenclature of Henrissat *et al.*²⁴³

There are some differences between the bacterial and plant glycogen/starch synthase enzymes. In bacteria such as *E. coli*, there is only one glycogen synthase, encoded by one glycogen synthase gene.²³⁹ However, since 1971 it has been known that in every plant tissue studied, more than one form of starch synthase can be identified. This has been summarized in a number of references.^{12,14,16,65,247,248} The starch synthases are encoded by more than just one gene. Some starch synthases are bound to the starch granule and are designated as starch GBSSs. They may be solubilized by α -amylase digestion of the granule, while others, designated as soluble starch synthases can be found in the soluble portion of the plastid fraction.

6.15.9.1 Characterization of the Starch Synthases

A phylogenetic tree based on the various deduced amino acid sequences of the starch synthases from crop plants and including the green alga, *Chlamydomonas*, have identified five subfamilies of starch synthases.⁶⁵ These are classified as GBSS, starch synthase I (SSI), starch synthase II (SSII), starch synthase III (SSIII), and starch synthase IV (SSIV). The SSII class may have even diverged further to classes SSIIa and SSIIb.²⁴⁹ Indeed, the GBSS family may have also diverged into another class, GBSSI, GBSSIIb, or GBSSII.^{250–252}

6.15.9.1.1 Soluble starch synthases

Multiple forms of soluble starch synthases are observed in barley endosperm,²⁵³ pea seeds,^{82,254} wheat endosperm,^{255–257} maize endosperm,^{249,258–261} potato tuber,^{262–264} *Arabidopsis*,²⁶⁵ rice seed,^{266,267} *Chlamydomonas reinhardtii*,²⁶⁸ as well as in sorghum seeds, teosinte seeds, and spinach leaf, that are reviewed elsewhere.^{16,65,154,247}

6.15.9.1.2 Soluble starch synthase I

Owing to the multiplicity of starch synthase isozymes, it is unclear what specific functions these isoforms would have in starch granule synthesis. Considerable effort to isolate mutant plants specifically deficient in one of the isoforms of the starch synthase has been done in order to determine their individual functions and some insight has been provided into the possible functions of the starch synthases, either soluble or granule bound, in synthesis of both the amylose and amylopectin. In the case of SSI, a mutant of *Arabidopsis* has been isolated by T-DNA insertion that is deficient in SSI activity and having an altered amylopectin structure.²⁶⁹ Other starch metabolizing enzymes, ADP-Glc PPase, branching enzyme, α - or β -amylase, maltase, α -glucanotransferase, pullulanase, phosphorylase, or GBSS1 were unaffected by the mutation. The soluble starch synthase activity was reduced by 56–72% in crude extracts and the remaining activity presumably was associated with the other soluble starch synthases than SSI. The starch content in the mutant was reduced to 77% of the normal and had slightly more amylose. The mutant amylopectin fraction was completely debranched with isoamylase and pullulanase and the chain length distribution compared to the wild-type *Arabidopsis* leaf starch. Whereas the wild-type starch showed the typical polymodal distribution of chain length with the maximum chain length of glucose units being 11 and 12, the mutant, labeled as Atss-1-1, had a unimodal chain length distribution with the most abundant chain lengths being 13–15 glucose units long. Further comparisons showed a marked reduction of chain lengths of 8–12 glucose units in the mutant and a significant increase in chain lengths of 17–20 glucose units.²⁶⁹ Debranching of the β -amylase limit digest of the mutant and wild-type starches showed that the mutant starch had more chains having less than 10 glucose units and less chains containing 11–24 glucose units. The conclusion was that SSI is involved in incorporating glucose units into small chains ‘filling up the cluster structure’ but is not involved in making longer interior chains.²⁶⁹ These results are consistent with previous results²⁷⁰ showing that recombinant maize SSI had a reduced catalytic activity of three- to fourfold when the primer glycogen side chains had been lengthened by Glc-1-P and phosphorylase action

before subjection to starch synthase action. Also, most of the label incorporated into glycogen *in vitro* by SSI was into chains less than 10 glucose residues in length.²⁷⁰ In view of these results it appears that SSI has a role in extending the shorter A and B chains to a length of about 14 glucose units or less. Further extension or synthesis of larger chains would have to be done by the other starch synthases. Thus, SSI may be mainly involved in the synthesis of the exterior A and B chains of amylopectin. These conclusions are also consistent with the finding that the maximal velocity of maize SSI is about 1.5 to 6.5-fold greater with glycogen than with amylopectin^{271,272} and is due to the shorter chains present in glycogen than in amylopectin. However, the K_m for amylopectin was 13-fold lower than for glycogen suggesting a greater affinity of maize SSI for amylopectin.²⁷¹ The K_m for ADP-Glc was 0.1 mmol l^{-1} and of interest was that citrate at 0.5 mol l^{-1} stimulated SSI activity to form a polymer in the absence of primer.²⁷¹ This has been confirmed in later studies²⁵⁸ that also showed the SSI enzyme being a monomer of 76 kDa. Another study also showed that maize SSI was a 77 kDa monomer. The mature *Arabidopsis* SSI enzyme subunit was determined to be 67 kDa²⁶⁹ and the rice-soluble SSI was estimated to be 55 kDa.²⁶⁶ These molecular sizes are similar to what has been found for pea embryo SSI (77 kDa),⁶⁶ wheat endosperm (77 kDa),²⁵⁶ potato tuber (79 kDa),²⁶² and potato leaf (70.6 kDa).²⁶⁴ The expression levels of SSI in potato tubers was different from those found for other starch synthases, GBSSI, SSII, and SSIII²⁶⁴ having drastically lower expression levels during development. The expression level was appreciably higher in sink and source leaves suggesting a minor role in starch synthesis in storage tissue. Moreover, application of antisense technique could reduce the SSI activity to nondetectable levels in tubers. There was no great effect in the starch content or in the structure of amylopectin. Potato tuber SSI activity is minor compared to the activities seen for SSII and SSIII and thus may play a minor role in tuber starch synthesis while playing a more important role in leaves.

The gene coding for full-length maize SSI (SSI-1) and the genes coding for N-terminally truncated SSI (SSI-2 and SSI-3) have been individually expressed in *E. coli*²⁷³ and purified to homogeneity. The expressed enzymes had essentially the same properties noted for other SSI.

In summary, outside of the study in *Arabidopsis*²⁶⁹ the function of soluble SSI in other plants remains to be determined.

6.15.9.1.3 Starch synthase II

Starch synthases II have been characterized in several plants including *Cblamydomonas*. A maize endosperm mutant termed dull1 (du1), due to the dull appearance of its mature kernels,²⁷⁴ has about 15% of the starch as a slightly altered amylopectin structure known as 'intermediate material,' distinguished between amylopectin and amylose on the basis of its starch-I₂ complex.²⁷⁵⁻²⁷⁷ The du1 starch had the highest degree of branching among many normal and mutant kernel starches studied.^{275,276} Of interest was that if the mutant du1 allele was combined with other starch mutants alterations in starch structure were seen more severe than those seen in the single mutants.^{277,278}

The first enzymatic studies done on the du1 mutant were done in 1981 and both SSII and starch branching enzyme IIa (SBEIIa) were found to have reduced activity in the endosperm compared to the normal maize endosperm.²⁷⁹ SSII was shown to be different from SSI.^{272,280} SSII requires a primer for activity, could not catalyze an unprimed reaction even in the presence of 0.5 mol l^{-1} citrate and has less affinity for amylopectin than SSI. However, 0.5 mol l^{-1} citrate lowered the K_m for amylopectin, 17-fold. The activity with glycogen as a primer is 1/2 observed with amylopectin. So glycogen is not as effective as a primer as is amylopectin and that is different than what was observed for SSI. Both maize endosperm SSI and SSII had a K_m for ADP-Glc of 0.1 mmol l^{-1} .^{271,280}

The maize du1 gene has been extensively characterized.²⁵⁹ A part of the du1 locus was cloned by transposon tagging and an almost complete *DUI* cDNA sequence was determined. It was found that the gene coded for a predicted 1674 amino acid peptide with two regions 51 and 73% similar, respectively with corresponding regions in SSIII of potato tuber.^{263,281} The deduced amino acid sequence of the *DUI* cDNA codes for a putative starch synthase with a predicted molecular size of 188 kDa. It is the C-terminal portion beyond amino acid residue 1226 that is most probably where the glucosyl transferase activity resides as the 450 residues of DU1 is similar to the well-known sequences of bacterial glycogen and plant starch synthases. Gao *et al.*²⁵⁹ conclude that the starch synthase coded by *DUI* may account for the soluble isoform identified earlier as SSII.^{279,280} The only discrepancy is that the deduced molecular size of the DU1 is 188 kDa while the SSII studied by Ozburn *et al.* was 95 kDa.²⁸⁰ However the deduced molecular size does match the 180 kDa molecular mass maize of SSII as

reported by Mu *et al.*²⁵⁸ The *DU1* gene product contains two repeated regions in its unique amino terminus and one of the regions has a sequence identical to a conserved segment of starch branching enzymes.²⁶¹ This may be related to the observation that the *du1* mutation also reduces starch branching enzyme IIa (SBEIIa) activity. The reduction of SBEIIa activity may be a secondary effect because of lack of interaction between SSII with SBEIIa in the mutant or that expression of SBEIIa is inhibited due to lack of SSII.

In other experiments²⁶⁰ it was shown that DU1 was one of two major soluble starch synthases and when the C-terminal 450 residues of DU1 was expressed in *E. coli* it was shown to have SS activity. Of interest was that antisera prepared against DU1 detected a soluble protein in endosperm extracts of molecular size of greater than 200 kDa and this protein was absent in *du1*⁻ mutants.²⁶⁰ The antisera reduced starch synthase activity by 20–30% in kernal extracts. In the same study antisera prepared on SSI reduced starch synthase activity by 60%. In *du1*⁻ mutants antisera prepared against SSI reduced the SS activity to essentially zero suggesting that SSI and SSII were the only maize endosperm soluble starch synthases. Because of the high similarity in sequence of the DU1 starch synthase II to the potato SSIII and that both are exclusively soluble, it is argued that DU1 is the evolutionary counterpart of potato SSIII.^{262,263,281} It is proposed that DU1 or maize SSII should be known as maize SSIII.²⁵⁹

In pea, mutants at the *rug5* locus were isolated after chemical mutagenesis.²⁸² The mutation caused the peas to become wrinkled seeded, a sign that they were starch deficient and the three mutant alleles were shown to have 30–40% less starch.²⁵⁴ Both mature and developing pea embryos were shown to be devoid of SSII as determined in polyacrylamide gel electrophoresis and immunoblotting with antiserum raised with SSII.²⁵⁴ Analysis of other enzymes involved in starch synthesis such as ADP-Glc PPase, branching enzyme, and SSIII did not reveal any significant changes between wild type and *rug5*. However, there was a 1.7-fold increase in the specific activity in the GBSS1 activity. Other enzymes such as P-glucomutase, UDP-Glc pyrophosphorylase, or sucrose synthase were not affected by the mutation.

Analysis of the amylopectin of the *rug5* mutant showed that there were two forms. One co-eluted with amylopectin on a sepharose column while another was of much lower molecular mass eluting later than amylose. Analysis of the mutant amylopectin fractions showed that it had more or less long chains than wild-type amylopectin. Debranching of the amylopectin fractions with isoamylase and size-exclusion chromatography showed that *rug5* amylopectin, had a far higher proportion of very long (B3 chains) than did wild type and the ratio of short chains (A and B1) to long chains (B2 plus B3) was 4:1 in wild type as compared to 8:1 in the mutant, *rug5*. Also the average DP of A chains in *rug5* was lower than the wild-type A chains. Using HPLC, the wild-type amylopectin showed a maximum of chains between 12 and 15 glucose units while the *rug5* amylopectin had a broad range of chain size between 7 and 13 glucose units long.²⁵⁴ Whereas very short chains of 7–9 glucose units are rare, however in wild-type amylopectin they are in great number in the SSII mutant. Thus lack of SSII activity in the pea embryo has a pronounced effect on the amylopectin structure that other starch synthases could not substitute for or replace the function of SSII. It is proposed that SSII is specifically required for the synthesis of B2 and B3 chains but another effect may be that the failure to produce longer chains in the amylopectin affects the other activities of the other starch synthase isoenzymes.

A SSIIa activity has been shown to be lacking in barley *sex6* mutants.²⁵³ The mutation located at the *sex6* locus on chromosome 7H and the mutant contained less than 20% of the amylopectin levels seen in wild-type barley. Analysis showed that a 90 kDa band was missing-attributable to SSIIa either in the starch granule or in the cytosol. Also there was an alteration of the distribution of other starch synthases, SSI as well as the branching enzymes, BEIIa and BEIIb. In the wild-type barley grain, they are distributed both in the granule as well as in the cytosol. In the mutant, they were only present in the cytosol. However, there was no alteration in their activity or in SSIII activity that was also present in the cytosol.²⁵³ The mutant had a decrease in amylose content of about 35% but the amylopectin reduction was more drastic, 84–91%. Thus, this mutant may be considered a high amylose mutant. Analysis of the mutant amylopectin showed that it had a higher percentage of chains, 6–11 glucose units in length than the wild type and a lower percent of intermediate chain length of 12–30 resulting in a starch with a reduced gelatinization temperature. The granule morphology was also altered. It is believed that all these effects are mainly due to the complete loss of SSIIa activity, but the alteration of the distribution of the other starch synthases and the branching enzymes may also play a role in the synthesis of the altered amylopectin.

This mutation also presents another approach in obtaining a high amylose starch. As will be described later, mutations of branching enzyme can not only lead to the production of more amylose but also amylopectins with lesser and longer chains and having properties similar to amylose.

Mutations of SSIIa have been observed in wheat and in rice;²⁸³ and in wheat each of the three wheat genomes were mutated to entirely eliminate expression of the SSIIa gene product, Sgp-1 protein²⁸⁴ in one line that had reduced starch amounts and an altered starch structure. In rice, two classes of starch have been found. In indica rice variety the starch is of the long-chain type while in japonica it is of the short-chain variety.²⁸³ Genetic analysis showed that the mutation in japonica rice led to a loss of SSII.²⁸³ Thus in higher plants it seems that loss of SSII activity in dicots and SSIIa activity in monocots have the same results with respect to reduced starch content due to lowered amylopectin content and altered amylopectin chain size distribution. Thus these genes may have the same function in starch biosynthesis.

A mutant of starch synthase II has also been isolated in *C. reinhardtii*.²⁶⁸ This mutant, st-3, accumulated only 20–40% of amount of starch present in wild type. The enzyme lacking in the mutant was SSII, one of the two starch synthase isoforms present in *C. reinhardtii*.²⁶⁸ There was also an apparent increase in the amylose content as well as a modified form of amylopectin. The changes noted in the mutant amylopectin were, a decrease in the number of intermediate glucose branch chains of 8–50 glucose units and an increase in short-chain glucans (2–7 glucose units). The conclusion made was that this SSII was responsible for synthesis of intermediate size branch chains, a similar conclusion reached as indicated above with the function of SSII and SSIIa in higher plants. The SSII of *Chlamydomonas* is now referred to as SSIII.⁶⁵

6.15.9.1.4 Starch synthase III

In 1996 it was found that the major isoform of soluble starch synthase in potatoes was of a molecular size of 139–140 kDa and it was labeled as SSIII.^{263,281} SSIII was expressed at the same level in all developmental stages and in contrast to GBSS1 and SSII, the SSIII was highly expressed in sink and source leaves.²⁸¹ Of interest was that in leaf discs both GBSSI and SSIII expression was induced by sucrose addition under light.²⁸¹ Antisera to SSIII reduced the total starch synthase activity by about 75%, indicating that it was the major form of activity of the soluble starch synthases.²⁶³ A cDNA clone of the starch synthase was isolated^{265,283} and analyzed and its transcript predicted a protein of 1230 amino acids with a molecular size of 139 kDa.²⁶³ The N-terminus sequence of about 60 amino acids was suggestive of a transit peptide.²⁶³ The N-terminal extension showed little sequence similarity to that of SSII from either pea or potato.²⁶³ However from amino acid residue 780 to the C-terminal end the sequence was similar to either soluble or GBSSs or even bacterial glycogen synthases.²⁶³ Indeed, The KTGG motif that is involved in ADP-Glc substrate binding²³⁴ is conserved as KVGGL at residue 794, as well as the FEPCGL sequence starting at residue 1121 that has been shown to contain an essential Glu residue (see **Table 10**) for activity of the *E. coli* glycogen synthase.²⁴² Using the antisense RNA approach to SSIII it was found that SSIII transcripts were eliminated.^{263,281} There was no effect on the levels of the other starch synthases and there was no effect on the amount of starch produced or in the amylose content in these antisense plants as compared to wild type.^{263,281} However what was noted in both reports was a drastic alteration in the starch granule morphology.^{263,281} The conclusions made were that SSIII was a major factor for synthesis of starch with normal morphology. But the exact role it played in the synthesis of starch remains obscure.

Subsequently, the endosperm of hexaploid wheat, *Triticum aestivum* [L.], was shown to have a SSIII gene.²⁵⁷ A cDNA was isolated and contained an open reading frame for a 1629 amino acid polypeptide. The N-terminal region started with the transit peptide region of 67 amino acids the N-terminal region of 656 residues, the SSIII-specific region of 470 amino acids and then at the C-terminal region, the 436 amino acid catalytic domain. These domains were compared to those seen in SSIII from maize DU1 protein,²⁶¹ potato,^{263,281} cowpea (GenBank accession No AJ225088), and *Arabidopsis* (GenBank accession number AC007296).

6.15.9.1.5 Starch synthase IV

As of 2003, a novel class of starch synthase amino acid sequences designated as SSIV was noted in expressed sequence tag (EST) databases from several species including *Arabidopsis*, *Chlamydomonas*, wheat, and cow pea.⁶⁵

However, the SSIV gene product had not been isolated or characterized.⁶⁵ The starch synthases show high similarity with each other. Using *A. thaliana* SSIV as 100%, the cowpea SSIV is 71% identical (accession number AJ006752), the wheat SSIV has 58% identity (accession number AY044844), the rice SSIV-1, 57% identity and rice SSIV-2, 58% identity (accession numbers AY373257 and AY373258, respectively).

In 2004 however, BLAST analysis of the rice genome using the conserved catalytic C-terminus of starch synthase amino acid sequences showed not only two SSIII homologous genes but also two genes homologous to wheat SSIV.²⁶⁷ cDNA clones of SSIV-1 and SSIV-2 were isolated and expressed in *E. coli*²⁸⁵ and were expressed 2.7- and 2.4-fold, respectively, in activity over the baseline level of the *E. coli* glycogen synthase suggesting that, the SSIV genes encoded starch synthase activity. SSIV-1 was expressed mainly in endosperm and weakly in leaves while SSIV-2 was expressed mainly in leaves and weakly in endosperm.

The rice SSIV enzymes contained three distinct regions. A putative transit peptide region of 78 amino acids for SSIV-1 and 33 amino acids for SSIV-2, a region homologous to Smc/myosin tail and then a C-terminal catalytic domain region resembling the bacterial glycogen synthase.²⁸⁶ As seen with starch synthase III, the KTGGL motif that is involved in ADP-Glc substrate binding²³⁴ is conserved as KXGGL. The FEPCGL sequence that has been shown to contain the Glu residue important for catalysis for the *E. coli* glycogen synthase is present in the C-terminal region. The SSIV gene encodes an enzyme with a predicted molecular size of about 118 kDa including the transit peptide.

The pattern of expression of SSIV in rice was also studied and compared to expression of the other starch synthases.²⁸⁷ Ten genes were identified to be of the starch synthase gene family.²⁸⁷ They were grouped on the basis of sequence analysis, as SSI, SSII, SSIII, SSIV, and GBSS.²⁸⁷ Of interest were the different patterns of temporal expression for the various starch synthases during seed development. The early expressers were SSII-2, SSIII-1, and GBSSII and were expressed in the early stage of grain filling. Those expressed in the mid- to later stage of grain filling, the late expressers, were SSII-3, SSIII-2, and the third group, GBSSI, SSI, SSII-1, SSIV-1, and SSIV-2 were expressed relatively constantly during the whole stage of grain filling.²⁶⁷

Recently, a mutant of *A. thaliana* deficient in SSIV has been isolated.²⁸⁶ As seen in rice,²⁸⁵ the SSIV gene contains 16 exons separated by 15 introns.^{285,286} Western blots of wild-type *A. thaliana* showed a starch synthase of 112 kDa but the two mutant alleles showed absence of the SSIV in Western blots.²⁸⁶ The mutant alleles showed lower growth rates under a 16-h day/8-h night photoregime when compared to wild type. There was also a decrease in leaf starch of 35–40%. If the SSIV protein was restored by transformation to the mutant then both growth rate and starch levels were restored indicating that the mutant alterations were due to SSIV deficiency. Normal growth of the mutant could also be restored by growing with continuous light. Because of lowered starch synthetic rate, the levels of sucrose and fructose and glucose were higher in the mutant. The mutant alleles had no decrease in total starch synthase activity. Nor did it have any significant decrease in other starch metabolizing enzymes. Upon analysis of the mutant starch it was found that there was no change in the amylose/amylopectin ratio, and there were minor effects on the amylopectin structure with a slight decrease in the number of chains from 7 to 10 glucose units long. However, there were significant alterations in the morphology of the starch granules with respect to size and number of granules in the chloroplast. Whereas normal chloroplasts would have 4–5 starch granules per chloroplast the SSIV mutants had only one. The mutant starch granule was considerably larger than what was observed in wild type. Curiously, not only the starch synthetic rate was reduced but also its degradation rate. This was attributed to having only one granule although with a large surface having a chloroplast with less overall starch granule surface for having both the synthetic and degradation processes to proceed at normal rates.²⁸⁶ Thus, it was proposed that the function of SSIV was to establish an initial structure for starch synthesis. The interaction of SSIV with other starch synthase isoforms would be of interest to study their specific involvement in starch granule synthesis. As will be discussed below there is evidence that the GBSS may also be involved in amylopectin synthesis in addition to amylose synthesis. Thus the mode of amylopectin synthesis by both soluble and GBSSs is still unknown and remains an active area for research.

6.15.9.1.6 Double mutants of the soluble starch synthases

Various alterations have been observed in amylopectin structure in the mutants of various starch synthase isoforms. SSI is involved in extending the shorter A and B chains to a length of 14 glucose units or less.^{269,270} SSII is involved in synthesis of B2 and B3 chains.^{253,254,268,283,284} SSIII is a major factor for synthesis of starch

with normal morphology. The exact role it plays however, in the synthesis of starch still remains obscure.^{265,283} SSIV is proposed to form an initial structure in starch, that is, initiate starch synthesis. However, the role or sequence or protocol in the synthesis of the starch granule is unknown for these individual starch synthases. Unknown is their specific contributions to the structure of amylopectin. Their role in starch synthesis may depend on its activity at particular times *in vivo*, and on the relative activities of other starch synthases and branching enzyme isozymes and their particular interaction with these enzymes and the growing amylopectin structure. Some efforts to understand the starch synthase pathway of amylopectin synthesis was attempted by making transgenic potatoes where the total SSII and SSIII activities were reduced 31–80%²⁸⁸ or 36–91%²⁸⁹ by the antisense technique. The SSII and SSIII constitute greater than 90% of the total soluble starch synthase activity in the tuber.^{263,281} The GBSS and branching enzyme activities were not affected. In one report²⁸⁸ the starch granules of double mutant were compared with the wild type and the single SS mutants. First, starch granule morphology of the SSII mutants were essentially similar but the SSIII mutant²⁶³ and the SSII/SSIII double mutants had abnormal morphology.²⁸⁸ The starch granule morphology in the SSII/SSIII mutant differed from the single SSIII mutant in that scanning electron microscopy showed holes through the granule center that was not observed in the single SSIII mutant. An analysis of the amylopectin branch chains of the mutants showed that all had a greater concentration of shorter chains and a lower amount of longer chains than the wild type. However, the patterns of long- and short-chain amounts were quite different. SSII and SSIII showed great enrichment of glucose oligosaccharides 9 and 6 units long while the SSII/SSIII had enrichments of glucose units of 7 and 8 and 12 and 13. There was no relationship between the chain pattern observed for SSII/SSIII and the cumulative pattern obtained for the mutant SSII and SSIII chains.²⁸⁸ In addition, the gelatinization behavior of the mutant starches were all different signifying different alterations in structure.²⁸⁸

These results strongly support the early view that different isozymes of starch synthases have unique functions in the synthesis of amylopectin.²⁹⁰ As indicated before, the SSII and SSIII mutants do not affect the rate of starch synthesis, but the mutants have different effects on starch structure and the double mutant SSII/SSIII act in a synergistic manner in the synthesis of amylopectin and not in an independent manner. This is based on the starch granule seen in the double mutant which is unlike that of the single SSII and SSIII mutants with respect to morphology and amylopectin chain lengths. Moreover, this can be explained that in the single mutant SSII or SSIII the function observed for the remaining active enzyme is different in wild type than in the double mutant.²⁸⁸ This is most probably due to the activity of the starch synthase dependent on the type of growing glucan acceptor it is presented with. Since other starch synthase isoforms are present, the glucan polymer that is synthesized in the absence of one of the starch synthase may be different and would modify the activity of other starch synthases and even branching enzyme activity.²⁹¹

Thus the synthesis of amylopectin or starch is not the sum of the independent actions of various starch synthases and branching enzyme isoforms but is due to a complex combination of the activities of various isoforms that vary during seed and leaf development due to varying expression of the different isozymes. Essentially, the same results and conclusions were obtained in the other report.²⁹¹

6.15.9.1.7 Starch synthases bound to the starch granule

Starch synthase activity was first discovered by Leloir's group and the activity was associated with the starch granule.^{290–292} These starch synthases are designated as GBSSs to distinguish them from the starch synthases mainly in the soluble phase of the chloroplast or amyloplast. The original characterization of GBSS,²⁹⁰ was made using UDP-Glc as the glucosyl donor but subsequently ADP-Glc was found to be a superior glucosyl donor reacting at a 10-fold faster rate than UDP-Glc at 12.5 mmol l^{-1} .^{272,292} The maize GBSS was released from the starch granule by incubating ground maize starch granules with α -amylase and glucoamylase.²⁹³ The solubilized starch synthase activity was partially purified and two peaks of activity were obtained with 80% of the activity residing with GBSSI that eluted from the DEAE-cellulose column at a lower salt concentration than the GBSSII fraction.

The solubilized granule-bound enzymes showed K_m values for ADP-Glc about 10-fold lower than that measured before solubilization (0.96 mmol l^{-1} for the intact granule activity; unpublished results). The granule-bound enzyme is also active with UDP-Glc²⁹³ with 1 mmol l^{-1} UDP-Glc having about 7% of the activity rate seen with 1 mmol l^{-1} ADP-Glc. If the concentration of UDP-Glc is raised to 20 mmol l^{-1} , then the rate of activity is about 73% of that of ADP-Glc. Upon solubilization with α -amylase and glucoamylase, activity with

Table 11 Kinetic properties of the solubilized maize endosperm GBSS²⁷²

| Property | Solubilized GBSS | |
|---|------------------|------|
| | SSI | SSII |
| K_m , ADP-Glc (mmol l^{-1}) | 0.14 | 0.11 |
| K_m , Amylopectin (mg ml^{-1}) | 1.2 | 0.26 |
| Relative activity: Amylopectin = 1.0 | | |
| Rabbit liver glycogen | 1.6 | 0.8 |
| Maltose, 1 mol l^{-1} | 2.9 | 0.51 |
| Maltotriose, 0.1 mol l^{-1} | 1.6 | 1.4 |

UDP-Glc essentially vanished. Either the activity observed with UDP-Glc was not solubilized, suggesting that a different enzyme catalyzed the UDP-Glc activity, or became inactive during the amylase procedure. Alternatively, the ability of the starch-bound starch synthase to utilize UDP-Glc is dependent on the close association of the enzyme with the starch granule. In other words, the conformation of the GBSS is altered in the presence of starch to allow UDP-Glc to be catalytically active.

The kinetic properties of solubilized GBSS are seen in **Table 11**. The solubilized enzymes, now starch-free, require a primer for activity. The GBSSII, in contrast to the SSSII, has a higher apparent affinity (lower K_m) for amylopectin than the granule type I enzyme. The soluble GBSSI has a higher activity with the rabbit liver glycogen than with amylopectin whereas the soluble GBSSII has less activity with the rabbit liver glycogen than with amylopectin. The solubilized GBSSI and II can utilize oligosaccharide primers as do the soluble starch synthases and **Table 11** shows the activities with respect to maltose and maltotriose. Other maltosaccharides are also used as primers. The products of the reaction observed with maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, and maltononaose are the maltosaccharides with an additional glucosyl unit (e.g., the primer maltotetraose when glucosylated yields maltopentaose).

The molecular size of the native enzymes determined by using sucrose density gradients, or size-exclusion chromatography showed that the major granule SSI, has a mass of 60 kDa, while the solubilized GBSSII, was about 93 kDa.²⁷²

Antibody prepared against the SSSI effectively neutralized SSSI activity but has no effect on either the activities of GBSSI or GBSSII or even on SSSII. Alternatively, antibody prepared against the starch granule-bound proteins effectively inhibits GBSSI activity but has very little effect on the soluble starch synthases.²⁷²

6.15.9.1.8 Isolation of the waxy protein structural gene

Amylose content determines the degree of translucency of the endosperm (hence the name 'waxy'), and it affects the cooking and eating qualities of the grains and the industrial properties of the starch extracted from those grains.

It is widely accepted that GBSS activity is a function of the protein coded by the waxy gene due to extensive genetic evidence. The waxy locus gene product is a protein of molecular weight 58 kDa associated with the starch granule and similar to that found for the solubilized maize endosperm GBSSI.²⁷² This protein was extracted by heating the starch with SDS or by incubation at 37 °C with 9 mol l⁻¹ urea but starch synthase would become inactive under those conditions. In mutants containing the wx allele there is virtually no amylose, GBSS activity is very low^{294–296} and the waxy protein is missing.

Shure *et al.*²⁹⁷ prepared cDNA clones homologous to Wx mRNA. Later, restriction endonuclease fragments containing part of the Wx locus were cloned from strains carrying the ac wx-M9, wx-M9, and wx-M6 alleles to further characterize the controlling insertion elements activator (ac) and dissociation (ds).²⁹⁶ Excision of the ds element from certain wx alleles produces two new alleles (S5 and S9) encoding the wx proteins having altered starch synthase activities.²⁹⁸ Two of these, S9 and S5, had 53 and 32% of the starch synthase activity, respectively, seen in the normal endosperm. Mutant S9, with the higher starch synthase activity, had 36% of the amylose content observed in the nonmutant endosperm, while mutant S5 with an even lower starch

synthase activity, 32%, had only 21% of the nonmutant maize amylose content. The correlation between the amount of GBSS activity and amylose amount provided additional evidence that the waxy protein is involved in amylose synthesis.

The DNA sequence of the Waxy locus of *Zea mays* was determined by analysis of both a genomic and an almost full-length cDNA clone.²⁹⁹ Also, barley Waxy locus has been cloned and its DNA sequenced.³⁰⁰ The deduced amino acid sequences of the maize and barley clones can be compared with the amino acid sequence of the *E. coli* ADP-Glc-specific glycogen synthase.²³⁹ Both clones had the sequence seen in the bacterial enzyme starting at residue Lys15, ...KTGGL... As indicated before, the lysine in the bacterial glycogen synthase has been implicated in the binding of the substrate, ADP-Glc.^{234,242} Moreover, the finding of similarity of sequences between the bacterial glycogen synthase and the putative plant starch synthases provides more and strong suggestive evidence that the waxy gene is indeed the structural gene for the GBSS. Developing pea embryo starch contains starch synthase activity that is associated with the waxy protein. The molecular weight of the pea starch synthase is about 59 kDa, as determined by ultracentrifugation in sucrose density gradients. The starch synthase activity was solubilized by amylase treatment and then partially purified.³⁰¹ The solubilized pea GBSS preparation displayed a relatively high specific activity (over 10 μ mol glucose incorporated per minute per milligram protein). This enzyme fraction was subjected to SDS polyacrylamide and gel electrophoresis. Protein staining or immunoblotting with maize Wx antibody, showed only the Wx protein.³⁰¹ Thus, the biochemical examination of starch synthase present in starch granules from two species, maize and pea, strengthens the genetic evidence supporting the role of the Wx protein as a GBSS with a major role in the determination of amylose content of starch.

6.15.9.1.9 Further studies of GBSS and isoforms; their involvement in both amylopectin and amylose synthesis

GBSSs have been studied in a number of plants and in many cases two isoforms have been identified designated as GBSSI and GBSSII. In pea embryos there are two forms that are associated with starch, GBSSI and SSII. However, SSII also contributes much of the soluble starch synthase activity in the embryo³⁰² and its properties were discussed as a soluble starch synthase in Section 6.15.9.1.2. Clones of the two isoforms of GBSSI and GBSSII from pea embryo⁸⁰ and potato tubers were isolated^{80,82} and characterized. Whereas the GBSSI of both potato and pea embryo were very similar in deduced amino acid sequence of the waxy proteins of maize and barley, the clones of GBSSII from potato and pea were different in sequence but similar to each other. The different feature from GBSSI was the extra N-terminal domain of 203 amino acids that is hydrophilic, having basic amino acids to give it a net positive charge and being serine rich.⁸² Both GBSS's had the KTGGL-ADP-Glc polyphosphate binding domain²³⁴ and the important SRFPCG-residue (E) domain for activity.²⁴² These clones were used to determine the temporal levels of expression during development. Pea GBSSII is highly expressed earlier in development than GBSSI and the expression is much lower in other organs of pea besides the embryo. In developing potato tubers GBSSI increased in expression during development similar to patatin while the GBSSII expression was highest in very young tubers and declined in larger potatoes.³⁰² cDNAs of GBSSI and GBSSII from wheat have also been isolated and characterized.²⁵³ GBSSII was expressed in leaf, culm, and pericarp tissue but not in endosperm where the expression of GBSSI transcripts were high. Thus, expression of the two isoforms maybe tissue- or organ-specific. The wheat GBSSI and GBSSII were 66% identical in sequence.

The pea embryo and potato tuber GBSSIs have been studied *in vitro* to determine the mechanism of how it may synthesize amylose.³⁰³ Amylose was synthesized by the granules but this was prevented by preincubating the granules with α -glucosidase suggesting the precursor was a soluble oligosaccharide. Amylose was then synthesized only when malto-oligosaccharides were added to reaction mixtures.³⁰³ GBSSI had higher affinity for the malto-oligosaccharides than GBSSII (or SSSII) and that transfer of the glucosyl moiety from ADP-Glc was different for the two starch synthases.³⁰⁴ A series of malto-oligosaccharides were synthesized by GBSSI suggesting that the transfer was not stepwise and more than one glucosyl unit was added before the oligosaccharide dissociated from the enzyme.³⁰⁴ With GBSSII the product seen was with just one glucose unit added to the malto-oligosaccharide substrate.

In further studies both potato tuber GBSSI and pea embryo SSII were expressed in *E. coli* in order to obtain them in soluble form.³⁰⁵ It was immediately recognized that GBSSI within the granules had different properties

than the soluble form. The affinity for maltotriose to form amylose was less for GBSSI than its soluble form. Whereas the K_m value for maltotriose was 0.1 mmol l^{-1} for GBSSII the soluble enzyme form was not even saturated at 1 mmol l^{-1} . The processive order of glucosyl addition was also lost. Of interest was the interaction of amylopectin with potato tuber GBSSI.³⁰⁵ It acted not only as a substrate but also as an effector. Amylopectin was an effector at lower concentrations where it was a substrate. It increased the affinity for another substrate, maltotriose as well as the rate of reaction with the trisaccharide over threefold. The reaction mode for maltotriose reverted to processive in the presence of amylopectin. It is postulated that the amylopectin induced a conformational change in the GBSSI reverting it to a form it had when it was granule-bound.³⁰⁵

The GBSS enzyme in leaves can be differentiated from those observed in storage tissue. As has been shown in storage tissue, the amylose content of starch can vary from 11 to 37%.²⁷⁸ In leaf or transitory starch the amylose content is less than 15%. In pea, leaf amylose is different from storage amylose with respect to molecular size being of a larger size. This was based on gel filtration chromatography where leaf amylose had a molecular size, based on dextran standards, of 655 kDa as compared to pea embryo amylose of 470 kDa.²⁵⁰ The cDNAs of GBSSIIa from embryo and GBSSIIb from leaves were isolated and compared. The GBSSIIb predicted amino acid sequence was of 613 amino acids with a molecular size of 67.6 kDa with a 68.8% identity and 75.6% similarity with GBSSIIa. Mature GBSSIIb after transit peptide cleavage is 58.4 kDa while mature GBSSIIa is 58.3 kDa. Both cDNAs were transformed in *E. coli* and their properties characterized. Unfortunately, GBSSIIa from embryo was inactive while GBSSIIb was active.²⁵⁰ The GBSSIIb activity was compared to potato tuber GBSSI from potatoes also expressed in *E. coli* and was found to be similar in properties in respect to K_m values for ADP-Glc and amylopectin. Both GBSSIIa and Ib activities were highly activated when reincorporated into starch granules.²⁵⁰ The two isoforms synthesized distinct isoforms of amylose with the Ia type forming a shorter type of amylose than the Ib form. These results explain why the Ia mutation of pea embryos lacking GBSSIIb still synthesizes about 4–10% of the starch content as amylose and minor amounts of Ib are present in the embryo.²⁸²

In *Arabidopsis* leaves, the mechanism of amylose synthesis was also studied with the view of identifying possible primers for amylose synthesis.³⁰⁶ Malto-oligosaccharides such as maltotriose when added to *Arabidopsis* leaf starch granules and incubated with ADP-[¹⁴C] Glc stimulated the synthesis of labeled amylose. These types of experiments had also been done with starch granules from pea and potato³⁰³ and from *C. reinhardtii*.³⁰⁷ Also a malto-oligosaccharide accumulating *Arabidopsis* mutant, dpe1, synthesized more amylose compared to the wild type.³⁰⁶ This strongly suggested that malto-oligosaccharides were indeed the primers for amylose synthesis. This raises the question of the origin of the malto-oligosaccharides in the plant.

Finally, the contributions of studies of starch synthases made in *C. reinhardtii* should be noted. As indicated before SSSII now known as SSIII may be involved in the synthesis of the intermediate size chains of amylopectin.²⁶⁸ Also the first evidence that GBSSI was involved in synthesis of amylopectin in addition to amylose synthesis was obtained in *C. reinhardtii*.^{308,309} It was first noted that mutants, sta2, were deficient in GBSSI and had in addition to the loss of amylose a fraction of altered amylopectin structure with longer chains.³¹⁰ It has also been shown in potatoes that GBSSI can elongate amylopectin chains.³⁰³ Further experiments in *C. reinhardtii* showed that in double mutants lacking both SSIII and GBSSI that an altered starch structure intermediate between glycogen and starch was formed.³⁰⁹

The cloning of the *C. reinhardtii* GBSSI has been achieved and it codes for a 69-kDa protein with an extra 11.4 kDa extension at the C-terminus.³¹⁰ The *C. reinhardtii* sta2 mutants transformed with the GBSSI cDNA could now synthesize amylose. An *in vitro* synthesis of amylose was studied by a starch granule isolated in either wild-type log phase or in a growth arrested (nitrogen limited) ADP-Glc PPase mutant.³¹⁰ The GBSSI-specific activity in this system is 10–50 times higher than what is seen in higher plants and enabled the researchers to increase *in vitro*, the amount of glucan in the granule by about 1.6-fold. The percentage of amylose increased from 13 to 45% in the granule. X-ray diffraction studies showed an increase in the appearance of crystalline material.³¹⁰ A-type crystals remained constant and the total amount of B-type crystals increased from 7 to 33% of the total starch suggesting that GBSSI induces *de novo* synthesis of B-type crystals.

Amylose synthesis depends on the concentration of ADP-Glc, as GBSSI has a high K_m for the substrate as compared to the soluble starch synthases.³¹¹ The phosphoglucomutase (PGM)-deficient mutants do make starch but are deficient in amylose, even though GBSS is present. The PGM-deficient mutants can make amylopectin but not amylose as shown by detailed structure studies of the starch accumulated by the algal

mutant.³¹² A similar structure effect can be seen when the algae has a defective ADP-Glc PPase.³¹¹ This indicates that the relatively high K_m value seen for ADP-Glc for GBSSI *in vitro*, most likely is also the *in vivo* K_m value for GBSS. Other studies also show that a diminution of ADP-Glc levels in the cell causes a lowering of amylose to very low or nonexistence levels.^{313,314}

6.15.10 Branching Enzyme

6.15.10.1 Branching Enzyme Assays

Branching enzyme (BE) can be assayed in numerous ways. The iodine assay is based on the decrease in absorbance of the polysaccharide–iodine complex resulting from the branching of either amylose or amylopectin. Aliquots are taken at intervals during the incubation of the amylose or amylopectin with BE, and iodine reagent is added.^{73,315,316} The decrease of absorbance is measured at 660 nm for amylose and, for amylopectin, at 530 nm. A unit of activity is defined as decrease in absorbance of 1.0 per min at 30 °C at the defined wavelength.

The phosphorylase stimulation assay,^{315,317–319} is based on the stimulation of the ‘unprimed’ (in other words, without added glucan) phosphorylase activity of the rabbit muscle phosphorylase α . This is due to the presence of BE in the assay mixture, increases the number of nonreducing ends available to the phosphorylase for elongation. A unit is defined as 1 μ mol transferred from Glc-1-P per min at 30 °C.

The branch linkage assay³¹⁹ is an assay that measures the number of branch chains formed by branching enzyme catalysis, rather than an indirect effect of its action as in the two assays described above. The enzyme fraction is incubated with the substrate, NaBH₄-reduced amylose. The reaction is then ended by boiling and the product is incubated for debranching, with pure *Pseudomonas* isoamylase. Finally, the reducing power of the oligosaccharide chains transferred by the enzyme is measured by a reducing sugar assay. Amylose reduced with borohydride is used (rather than amylose itself) to lower the initial reducing power seen with amylose.

The branch linkage assay is the most quantitative assay for BE, but the presence of impurities such as amylolytic activity, interferes the most with this assay. The phosphorylase stimulation assay is the most sensitive. The I₂ assay is not very sensitive but allows the testing of BE specificity with various maltodextrins, providing information on the possible role of the different BE isoforms. It may be best to employ all three assays when studying the properties of the BEs, but, above all, if reliable information is being sought, the BEs must be purified to the extent that all degradative enzymes are eliminated before studying its properties.

6.15.10.2 Bacterial Branching Enzymes

The structural gene of various BEs have been cloned from many bacteria.^{45,47–49,320,324} The nucleotide and deduced amino acid sequences of the *E. coli* *glg B* gene consisted of 2181 bp specifying a protein of 727 amino acids and with a molecular weight of 84 231.³²⁴

6.15.10.3 Plant and Algal Branching Enzymes; Characterization of Isozymes

Maize endosperm has three BE isoforms.^{72,318,325,326} BEI, IIa, and IIb from maize kernels,^{72,315,326} were purified to the point they no longer contained amylolytic activity.^{72,326} Molecular weights were 82 kDa for BEI and 80 kDa for BEs IIa and IIb.^{315,317}

Table 12 summarizes the properties of the various maize endosperm BE isozymes from the studies of Takeda *et al.*³¹⁹ and Guan and Preiss.⁷² Of the three isoforms, BEI had the highest activity in the iodine assay in branching amylose and its rate of branching amylopectin was about 3% observed with amylose. The BEIIa and BEIIb isozymes branched amylopectin at twice the rate they branched amylose and catalyzed the branching of amylopectin at 2.5 to 6-times the rate observed with BEI. Takeda *et al.*³¹⁹ analyzed the branched products made *in vitro* from amylose by the BE isoforms. Isoamylase was used to debranch the products of each reaction followed by chromatography. BEIIa and BEIIb were very similar in their affinity for amylose and the size of chain transferred. When presented with amyloses of different average chain length (CL), the BEs had higher activity with the longer chain amylose. But while BEI could still catalyze the branching of an amylose of average CL of 197 with 89% of the activity shown with a CL of 405, the activity of BEII dropped sharply with

Table 12 Specific activities (units⁻¹mg protein) of maize endosperm BE isoforms

| Branching enzyme assay | | BEI | BEIIa | BEIIb |
|-------------------------------|------------------|------|-------|-------|
| Phosphorylase stimulation (A) | | 1196 | 795 | 994 |
| Branching linkage assay (B) | | 2.6 | 0.32 | 0.14 |
| Iodine stain assay (C) | | | | |
| | Amylose (C1) | 800 | 29.5 | 39 |
| | Amylopectin (C2) | 24 | 59 | 63 |
| Ratio of assay | A/B | 460 | 2484 | 7100 |
| Activities | A/C1 | 1.5 | 27 | 25 |
| | A/C2 | 49.8 | 13.5 | 15.8 |
| | C2/C1 | 0.03 | 2 | 1.6 |

The units are defined for phosphorylase stimulation and branching linkage assays as $\mu\text{mol}^{-1} \text{min}^{-1}$. For the iodine staining assay, it is defined as a decrease of one absorbance unit at 660 nm per min for amylose (C1) and one absorbance unit at 530 nm (C2).

this change in chain length. A study of the *in vitro* reaction products indicated that, the action of BEIIa and BEIIb resulted in the transfer of shorter chains than those transferred by BEI. Thus, BEI catalyzes the transfer of longer branched chains and that BEIIa and IIB catalyze the transfer of shorter chains. This may suggest that BEI produces a slightly branched polysaccharide that serves as a substrate for enzyme complexes of starch synthases and BEII isoforms to synthesize amylopectin. BEII isoforms may also play a predominant role in forming the short chains present in amylopectin. Moreover, BEI may be more involved in producing the more interior B-chains of the amylopectin while BEIIa and BEIIb would be involved in forming the exterior (A) chains.

Vos-Scheperkeuter *et al.*³²⁷ purified a single form of branching activity of 79 kDa molecular mass from potato tubers. The antibodies prepared against the native potato enzyme was found to react strongly with maize BEI and weakly with maize BEIIb. The antiserum inhibited the activities of both potato tuber BE and maize BEI in neutralization tests. Potato BE thus shows a high degree of similarity to maize BEI and a lesser extent with the maize BEII isozymes.

Up to four cDNA clones have been isolated for potato BE; coding for proteins, 91–99 kDa.^{325,328,329} All these allelic clones have sequences similar to the BEI type. The *sbeIc* allele, codes for a mature enzyme of 830 amino acids and a molecular weight of 95 180. The *sbeIc* BE protein product, expressed in *E. coli*, migrates as a 103-kDa protein.³²⁵

It is noteworthy that BE isolated from other plants, bacteria, and mammals have molecular masses ranging from 75 to about 85 kDa. These molecular masses are consistent with the molecular weights obtained from deduced amino acid sequences obtained from isolated genes or cDNA clones.

Potato BEII was first characterized and found to be a granule-bound protein in tuber starch.³³⁰ In potato, BEII appears to be less abundant than BEI. Both potato tuber BEI and BEII were cloned and expressed in *E. coli*^{331,332} and the properties of the potato tuber BE isoforms were compared.³³² As seen with the maize branching enzymes, potato BEI was more active on amylose than BEII and BEII was more active on amylopectin than BEI.

Mizuno *et al.*⁷⁰ reported four forms of branching enzyme from immature rice seeds. BE1 and BE2 (composed of BE2a and BE2b) were the major forms while BE3 and BE4 were minor forms, being less than 10% of the total branching enzyme activity. The molecular weight of the branching enzymes were: BE1, 82 kDa; BE2a, 85 kDa; BE2b, 82 kDa; BE3, 87 kDa; BE4a, 93 kDa; and BE4b, 83 kDa. However, BE1, BE2a, and BE2b were similar immunologically in their reaction toward maize endosperm BEI antibody. The rice seed BE1, BE2a, and BE2b were very similar in their N-terminal amino acid sequences. The three BE N-terminal sequences were either TMVXVVEVDHLPIT or VXVVEVDHLPITDL. The latter sequence is the same as the first but lacking the first two N-terminal amino acids. Although these activities came out in separate fractions from an anion exchange column, most likely they are the same protein on the basis of immunology and N-terminal sequences. BE2a however, is 3 kDa larger. It is possible that BE2a may be the less proteolyzed form. Antibody raised against BE3 reacted strongly against BE3 but not against, BE1, BE2a, and BE2b. Thus, rice endosperm, as noted for maize endosperm, has essentially two different isoforms of BE.

Because of the many isoforms existing for rice seed branching enzymes, Yamanouchi and Nakamura³³³ studied and compared the BEs from rice endosperm, leaf blade, leaf sheath, culm, and root. BE activity could be resolved into two fractions, BE1 and BE2. Both fractions were found in all tissues studied in different ratios of activity. The specific activity of the endosperm enzyme, either on the basis of fresh weight or protein, was 100–1000-fold greater than that of the enzyme from other tissues studied. On gel electrophoresis, rice endosperm BE2 could be resolved into two fractions, BE2a and BE2b. Of interest was that electrophoresis of the other tissue BE2 forms revealed only BE2b. BE2a was only detected in the endosperm tissue. In rice, it seems, there are tissue-specific isoforms of BE.

Three forms of branching enzyme from developing hexaploid wheat (*Triticum aestivum*) endosperm have been partially purified and characterized.³³⁴ Two forms are immunologically related to maize BEI and one form with maize BEII. The N-terminal sequences are consistent with these relationships. The wheat BEIB gene is located on chromosome 7B while the wheat BEIAD peptide genes are located on chromosomes 7A and 7D. The BE classes in wheat are differentially expressed during endosperm development in that BEII is constitutively expressed throughout the whole cycle while BEIB and BEIAD are expressed in late endosperm development.

In this respect, McCue *et al.*³³⁵ showed that in wheat (*Triticum aestivum* cv Cheyenne) the isolation of cDNAs encoding BEI and BEII that shared extensive identity with BE sequences reported for wheat and other plants. Using the cDNAs they studied the steady state RNA levels of the BEs during development. For BE2 its RNA was detectable 5 days postanthesis (DPA) and reached a maximum 10 days post DPA. BE1 steady state levels started rising at 10 DPA peaking at 15 DPA. Levels of all messages declined at 20–25 DPA.

In barley, four BE isozymes were identified by separation on FPLC and three were partially purified.³³⁶ Subsequently, two cDNA clones encoding the barley BEIIa and IIb were isolated.³³⁷ The major structural difference between the two enzymes was the presence of a 94-amino acid N-terminal extension in the BEIIb precursor.

6.15.10.4 Genetic Studies on Branching Enzyme-Deficient Mutants

There are some maize endosperm mutants that appear to increase the amylose contents of starch granules. The normal maize starch granule contains about 25% of the polysaccharide as amylose, with the rest as amylopectin. In contrast, amylose extender (ae) mutants may have as much as 55–70% of the polysaccharide as amylose and may have an amylopectin fraction with fewer branch points and with the branch chains longer in length compared to those of normal amylopectin. Results with the recessive maize endosperm mutant, ae, originally suggested that ae is the structural gene for either branching enzyme BEIIa or BEIIb^{279,317,338} as the activity of BEI was not affected by the mutation. In gene dosage experiments, Hedman and Boyer³³⁹ reported a near-linear relationship between increased dosage of the dominant ae allele and BEIIb activity. Since the separation of form IIa from IIb was not very clear, it is possible that the ae locus was also affecting the level of IIa.

Singh and Preiss³⁴⁰ concluded that, although some homology exists between the three starch BEs, there are major differences in the structure of BEI when compared with BEIIa and BEIIb as shown by its different reactivity with some monoclonal antibodies and differences in amino acid composition and proteolytic digest maps.

Recent studies by Fisher *et al.*^{341,342} in analyzing 16 isogenic lines having independent alleles of the maize ae locus, suggest that BEIIa and BEIIb are encoded by separate genes and the BEIIb enzyme is encoded by the ae gene. They isolated a cDNA clone labeled Sbe2b, which had a cDNA predicted amino acid sequence at residues 58–65 exactly the same as the N-terminal sequence of the maize BEIIb that they had purified.³⁴¹ Moreover, they did not detect in ae endosperm extracts any mRNA with the Sbe2b cDNA clone. There was some BE activity in the ae extracts but on ion-exchange chromatography behaved similarly with BEIIa.

The finding that the enzyme defect in the ae mutant is BEIIb is consistent with the finding that, *in vitro*, BEII is involved in transfer of small chains. The ae mutant has not only an increased amount in amylose content but also, the amylopectin structure is altered in having fewer and longer chains and a lesser number of total chains. In other words, there are very few short chains.

Wrinkled pea has a reduced starch level; 66–75% of that seen in the normal round seed. The amylose content is about 33% in the wild-type round form, but is 60–70% in the wrinkled pea seed. Edwards *et al.*³⁴³ measured several enzyme activities involved in starch metabolism at four different developmental stages of wrinkled pea and found that branching enzyme activity at its highest level was only 14% of that observed for

the round seed variety. The other starch biosynthetic enzymes as well as phosphorylase had similar activities in wrinkled and round seeds. Confirmation of these results were by Smith,³⁴⁴ who also showed that the r(rugosus) lesion in the wrinkled pea of genotype rr, was associated with the absence of one isoform of BE. Edwards *et al.*³⁴³ postulated that the reduction in starch content observed in the BE-deficient mutant seeds is an indirect effect of the reduced BE activity through an effect on starch synthase activity. The authors suggested that, in the absence of BE activity, starch synthase forms an α -(1 \rightarrow 4)-glucosyl elongated chain which is a poor glucosyl-acceptor (primer) for the starch synthase substrate, ADP-Glc, therefore decreasing the rate of α -(1 \rightarrow 4)-glucan synthesis. This had been shown in a similar system, in a study of rabbit muscle glycogen synthase³⁴⁵ where it was found that continual elongation of the outer chains of glycogen caused it to become a less efficient primer with a higher K_m , thus decreasing the apparent activity of the glycogen synthase. The finding that ADP-Glc in the wrinkled pea accumulated to a higher concentration than in normal pea was considered evidence that the *in vivo* activity of the starch synthase was restricted. Under *in vitro* conditions, where a suitable primer such as amylopectin or glycogen is added, the starch synthase activity in the wrinkled pea was equivalent to that found for the wild type.

Ae mutants have been found in rice.³⁴⁶ The alteration of the starch structure is very similar to that reported for the maize endosperm ae mutants. The defect is in the BE3 isozyme, BE3 of rice being more similar in amino acid sequence to maize BEII than to BEI.^{74,346} Thus, rice BE3 may catalyze the transfer of small chains, rather than long chains. Of interest was the use of gene silencing through RNA interference (RNAi) technology to silence the expression of wheat BEIIa and BEIIb that resulted in a high-amylose (>70%) phenotype.³⁴⁷ The critical aspect was to suppress expression of BEIIa. Suppression of BEIIb alone had no effect. Suppression of BEIIa decreased markedly the proportion of glucose CLs of 4–12 with a corresponding increase of CLs greater than 12. When this high amylose starch was fed to rats in a diet as a wholemeal, several indices of large bowel function, including short-chain fatty acids, were improved relative to a standard wholemeal wheat. The results indicated that the high-amylose wheat has a significant potential to improve human health through the high amylose starch content resistant to digestion.³⁴⁷

6.15.10.5 Isolation of cDNA Clones Encoding the Branching Enzyme Isozyme Genes

The r locus of pea seed was cloned using an antibody toward one of the pea BE isoforms and screening a cDNA library.³⁴⁸ However, the BE gene in the wrinkled pea contained an 800 bp insertion. Thus expression of the cDNA yielded an inactive BE. The sequence of the 2.7 kB clone had more than 50% homology to the glycogen-BE of *E. coli*³⁴⁹ and thus it was proposed that the cloned cDNA corresponded to the starch BE gene of pea seed. The *glg B* gene has been sequenced for a cyanobacterium,⁴⁹ and its deduced amino acid sequence is 62% identical extensive to the amino acid sequence in the middle area of the *E. coli* protein. This is the portion of the sequence that contains the amino acid residues critical for catalysis. Therefore, BEs in nature have extensive homology irrespective of the degree of branching of their products, that is higher (~10% α -1,6 linkages) in glycogen, the storage polysaccharide in mammals, enteric bacteria, and in cyanobacteria, than in the amylopectin (~5% α -1,6 linkages) present in higher plants. It would be of interest in determining the difference in catalysis between the BEs that are involved in synthesis of amylopectin and those involved in the catalysis of bacterial or mammalian glycogen. Can it be due to different interactions of BE with the starch synthases or glycogen synthases or are the differences inherent in the BE catalysis? The answer to this question could also resolve, in part as to why the starch granules in different plants are unique in their granule structure.

cDNA clones of genes representing different isoforms of BE of different plants have been isolated from potato tuber,^{325,328,329} maize kernel,^{72,73,341,342} cassava,³⁵⁰ and rice seeds (BEI,^{70,351} BEIII³⁴⁶). The cDNA clones of maize BEI and BEII have been overexpressed in *E. coli* and purified.^{72,73} The transgenic enzymes had the same properties as seen with the natural maize endosperm BEs with respect to specific activity and specificity toward amylose and amylopectin.⁷⁴ It is important to note that *in vitro*, maize BEI and BEII have different properties. BEI had a higher rate of branching amylose than branching amylopectin and preferentially transferred longer chains. In contrast, maize BEII had a lower rate of branching amylose than branching amylopectin and preferentially transferred shorter chains.^{72,319}

6.15.10.6 Reserve Tissue Branching Enzyme Is Localized in the Plastid

The localization of BE within the plastid has been determined in potato³⁵² using antibodies raised against potato BE and immunogold electron microscopy. The enzyme (i.e., the equivalent of the BEI isoform of maize, as indicated above) was found in the amyloplast, concentrated at the interface between the stroma and the starch granule, rather than throughout the stroma, as it is the case with the ADP-Glc PPase.¹⁸⁰ This would explain how amylose synthesis is possible when the enzyme responsible for its formation, that is, the Wx protein, GBSS, is capable of elongating both linear and branched glucans. The spatial separation of BE isoforms from the GBSSs, even if only partial, would allow the formation of amylose without it being subsequently branched by the BE. However, even if spatial separation did not exist, starch crystallization may have the same effect, that is, prevent further branching. The unpublished experiments by M. Morell and J. Preiss found that about 5% of the maize endosperm total BE activity was associated with the starch granule after amylase digestion. Whether this BE was similar to the soluble BEs was not determined, but Preiss and Sivak¹⁶ found among the proteins present in maize and pea starch, a polypeptide of about 80 kDa that reacted with antibodies raised against maize BE1. However, it is not worthy that small amounts of BE are expected to sediment with the starch granules because of its affinity for the polysaccharide. The results were confirmed by Mu-Forster *et al.*³⁵¹

6.15.10.7 Branching Enzyme Belongs to the α -Amylase Family

Amino acid sequence relationships between that of BE and amylolytic enzymes such as α -amylase, pullulanase, glucosyltransferase, and cyclodextrin glucanotransferase, especially at those amino acids believed to be contacts between the substrate and the amylase family enzymes, was first reported by Romeo *et al.*³⁵⁴ Baba *et al.*³⁵⁵ reported that there was a marked conservation in the amino acid sequence of the four catalytic regions of amylolytic enzymes in maize endosperm BEI. As shown in **Table 13**, four regions that putatively constitute the catalytic regions of the amylolytic enzymes are conserved in the starch branching isoenzymes of maize endosperm, rice seed, and potato tuber and the glycogen BE of *E. coli*. A very extensive analysis of this high

Table 13 Comparison of primary structures of various BEs with the four most conserved regions of the α -amylase family

| | <i>Region 1</i> | | <i>Region 2</i> | |
|--------------------------------------|-----------------|----------------------------|-----------------|-----------------------------|
| Maize endosperm BEI | 277 | DVVHSH | 347 | GFRFDGVT S |
| Maize endosperm BEII | 315 | DVVHSH | 382 | GFRFDGVT S |
| Potato tuber BE | 355 | DVVHSH | 424 | GFRFDG I T S |
| Rice seed BE1 | 271 | DVVHSH | 341 | GFRFDGVT S |
| Rice seed BE3 | 337 | DVVHSH | 404 | GFRFDGVT S |
| <i>E. coli</i> glycogen BE | 335 | DWVPGH | 400 | ALRV D AVAS |
| <i>B. subtilis</i> α -amylase | 100 | DAVINH | 171 | GFR R DAAKH |
| <i>B. sphaericus</i> cyclodextrinase | 238 | DAVFNH | 323 | GW R LDVANE |
| <i>P. amyloclavata</i> isoamylase | 291 | DVVYNH | 370 | GFR R DLASV |
| | <i>Region 3</i> | | <i>Region 4</i> | |
| Maize endosperm BEI | 402 | TVVA E DVS | 470 | CIAYA E SHD |
| Maize endosperm BEII | 437 | VTIG E DVS | 501 | CVTYA E SHD |
| Potato tuber BE | 453 | VTMA E EST | 545 | CVTYA E SHD |
| Rice seed BE1 | 396 | TIVA E DVS | 461 | CVTYA E SHD |
| Rice seed BE3 | 459 | ITIG E DVS | 524 | CVTYA E SHD |
| <i>E. coli</i> glycogen BE | 453 | VTMA E EST | 517 | NVFLPLN H D |
| <i>B. subtilis</i> α -amylase | 204 | FQY G EILQ | 261 | LVTW V ESH D |
| <i>B. sphaericus</i> cyclodextrinase | 350 | IIV G EV W H | 414 | SFNL L G S HD |
| <i>P. amyloclavata</i> isoamylase | 412 | RIL R E F TV | 499 | SIN F ID V HD |

The sequences have been derived from references referred in the text. Three examples of enzymes from the α -amylase family are compared. Svensson's group,^{356,357} compare over 40 enzymes ranging from amylases, glucosidases, various α -1,6-debranching enzymes as well as branching enzymes. The invariant amino acid residues believed to be involved in catalysis are in bold letters.

conservation in the α -amylase family has been reported by Svensson *et al.*^{356,357} with respect to sequence homology and also in the prediction of the (β/α)-barrel structural domains with a highly symmetrical fold of eight inner, parallel β -strands surrounded by eight helices in the various groups of enzymes in the family. The (β/α)-barrel structural domain is based on crystal structures of some α -amylases and cyclodextrin glucanotransferases.

The conservation of the putative catalytic sites of the α -amylase family in the glycogen- and starch-branching enzymes may be anticipated as BE catalyzes two continuous reactions; cleavage of an 1,4- α -glucosidic linkage in an 1,4- α -glucan chain yielding an oligosaccharide chain which is transferred to an O-6 of the same chain, or to another 1,4- α -glucan chain, with synthesis of a new (1-6)- α -glucosidic linkage.

6.15.10.8 Amino Acid Residues that are Functional in Branching Enzyme Catalysis

As indicated in **Table 13**, four regions which constitute the catalytic regions of the amylolytic enzymes are conserved in the starch branching isoenzymes of maize endosperm, rice seed, and potato tuber and the glycogen BEs of *E. coli*.^{356,357} It would be of interest to know whether the seven highly conserved amino acid residues of the α -amylase family listed in bold letters in **Table 13** are also functional in branching enzyme catalysis. Experiments such as chemical modification and analysis of the three-dimensional structure of the BEs have been carried out to determine the nature of its catalytic residues and mechanism.

The seven highly conserved amino acid residues of the α -amylase family appear to be also functional in BE catalysis. A series of experiments,³⁵⁸ in which amino acids were replaced by site-directed mutagenesis, do suggest that the conserved Asp residues of regions two and four and the Glu residue of region 3 (**Table 13**, in bold letters) are important for BEII catalysis. Arginine residues, are also important, as suggested by chemical modification with phenylglyoxal,³⁵⁹ as well as histidine residues as suggested by chemical modification studies with diethyl pyrocarbonate.³⁶⁰ Also, the regions of the C-terminus and N-terminus that are dissimilar in sequence and size in the various branching isoenzymes, BEI and BEII, of maize endosperm may be of importance in catalysis with respect to substrate preference (amylose or amylopectin) or size of chain transferred during catalysis.

Indeed, studies by Kuriki *et al.*³⁶¹ with respect to the different N- and C-termini indicate that these amino acid sequence regions are important with respect to BE specificity with respect to substrate preference (amylose or amylopectin) as well as in size of chain transferred and extent of branching. The C-terminal was functional with respect to substrate preference while the N-terminal was functional with respect to the size of chain transferred in the BE catalysis. Furthermore, truncation of 113 amino acids of the N-terminal of the *E. coli* BE, causes it to branch longer branch chains than the wild-type enzyme³⁶²⁻³⁶⁴ further indicating that the N-terminal region was involved in specifying the size of chain transferred. Recently, the crystal structure of a truncated form of the *E. coli* branching enzyme has been elucidated.³⁶⁵

6.15.11 Other Enzymes Involved in Starch Synthesis

In addition to the three starch biosynthetic enzymes, other enzymes have been shown to have some effect on starch structure. These enzymes will be briefly reviewed.

6.15.11.1 Isoamylase

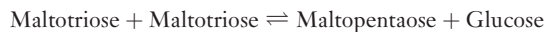
As previously indicated, a *su1* mutation in maize, which causes a deficiency of isoamylase, an enzyme normally considered to be mainly involved in starch degradation in plants or in *C. reinhardtii*, results in accumulation of a water-soluble polysaccharide, phytoglycogen instead of starch.^{85-87,366} In *C. reinhardtii*, the mutation results in complete loss of starch but in higher plants the lesser amount of amylopectin seen in the mutant plant may be due to the severity of the enzyme deficiency.^{367,368}

It was proposed that 'maturation' or trimming of the precursor of amylopectin, 'preamylopectin', was required in order for the amylopectin to aggregate into an insoluble granule structure.^{65,369} An alternative proposal is that during amylopectin synthesis there is a competition between polysaccharide aggregation into

granule starch and formation of water-soluble polysaccharide and that the water-soluble polysaccharide, phytoglycogen, is consistently degraded by isoamylase.³⁷⁰ If isoamylase activity is deficient, then the phytoglycogen accumulates and competes with amylopectin for the starch biosynthetic enzymes. The amount of phytoglycogen accumulation is dependent on the degree of loss of isoamylase activity.³⁷⁰ This competition concept between amylopectin aggregation and phytoglycogen accumulation can also explain why in certain isoamylase mutations the two polysaccharides, phytoglycogen and amylopectin, may be simultaneously present.³⁷⁰ Although there is no question that isoamylase deficiency is the cause of phytoglycogen accumulation and lower starch accumulation in the mutant plant, the mechanism for phytoglycogen accumulation is still not completely understood.

6.15.11.2 α -1,4 Glucanotransferase

Some starch-deficient mutants of *Arabidopsis* and *Chlamydomonas* have been shown to be defective in α -1,4 glucanotransferase activity. The enzyme is also known as D-enzyme and the reaction it catalyzes is shown below.



Other oligosaccharides can also act as substrates and in the reaction glucose is formed. The transglucosylase of *Arabidopsis* leaf disproportionates maltotriose and forms higher maltodextrins, much greater in comparison to maltooctaose.³⁷¹

A *C. reinhardtii* mutant lacking D-enzyme activity has been characterized and has been shown to have significantly lower levels of starch.³⁷² Other enzymes involved in starch metabolism such as ADP-Glc PPase, granule-bound and soluble starch synthase, BE, phosphorylase, α -glucosidase, amylases, and debranching enzyme activities were not affected.³⁷² The starch content in the mutant was about 6–13% of wild type and there was an excessive accumulation of maltooligosaccharides up to a polymer size of 16 glucose units.

At present it is not clear how D-enzyme deficiency causes a lowering of starch accumulation in *C. reinhardtii*. Actually an earlier report showed that a 98% reduction in D-enzyme activity of *Arabidopsis* had no effect on starch synthesis.³⁷³ Also in *Arabidopsis* it has been shown that mutants of D-enzyme overproduced starch and the overproduction of maltodextrins occurred only during the process of starch degradation and decreased during starch synthesis.³⁷⁴ This was different than what was observed for *Chlamydomonas* where maltodextrin accumulation occurred during starch synthesis and decreased during starch degradation. Whether the D-enzyme plays a role both in starch synthesis as well as in degradation is an intriguing question. It is hard to imagine that this enzyme would have different roles in oxygenic photosynthetic organisms.

In summary, much information in the past 25 years has been obtained on the enzymes involved in starch synthesis. However, a number of problems remain. There is no question that the major allosteric regulation of starch synthesis occurs at the ADP-Glc PPase-catalyzed reaction and is dependent on the ratio of 3-P-glycerate to P_i . The detailed process of how that ratio is regulated in storage and other plant nonphotosynthetic tissue remains to be elucidated. Certainly, the details of how the increase in sucrose concentration stimulates the reductive activation of the ADP-Glc PPase in potato remains to be uncovered. Also how the different isoforms of up to five or six different starch synthases coordinate their activities with the branching enzyme isoforms in synthesizing both amylose and amylopectin is still a formidable problem that requires much more detail. Of recent interest are the findings that enzymes such as isoamylase, D-enzyme and an isoform of a pastidial phosphorylase,³⁷⁵ normally considered as degradative enzymes, also play some role in the synthesis of starch. Their precise roles remain yet to be elucidated.

6.15.12 Genetic Regulation of Bacterial Glycogen Synthesis

The activity of the glycogen biosynthetic enzymes in *E. coli* increases as cultures enter the stationary phase.^{7,11,25,376} When cells are grown in an enriched medium containing yeast extract and 1% glucose and as cultures enter the stationary phase the specific activities of ADP-Glc PPase and glycogen synthase increase

11- to 12-fold, and BE increases fivefold. In defined media, BE is fully induced in the exponential phase, with only about a twofold increase in specific activity of the ADP-Glc PPase and glycogen synthase when cells reach the stationary phase. The same phenomena are also seen with the glycogen biosynthetic enzymes in *S. typhimurium*.⁴¹ Possibly the gene encoding the BE is regulated differently from the genes for ADP-Glc PPase and glycogen synthase. Cattaneo *et al.*³⁷⁶ showed that the addition of inhibitors of RNA or protein synthesis to prestationary phase cultures prevented the enhancement of glycogen synthesis in the stationary phase, suggesting that the pathway is under transcriptional control.

The structural genes for glycogen biosynthesis are clustered in two adjacent operons, also containing genes for glycogen catabolism. The structural genes for glycogen synthesis are located at approximately 75 min on the *E. coli* K-12 chromosome, and the gene order at this location, as established by transduction, is *glgA-glgC-glgB-asd*.³⁷⁷ These genes encode the enzymes glycogen synthase, ADP-Glc PPase, glycogen branching enzyme, respectively, and are close to *asd*, the structural gene for the enzyme aspartate semialdehyde dehydrogenase (EC1.2.1.11).

The molecular cloning of the *E. coli glg* structural genes¹⁰¹ greatly facilitated the study of the genetic regulation of bacterial glycogen biosynthesis; they were cloned into pBR322 via selection with the closely linked essential gene *asd*. Among the several *asd*⁺ plasmid clones that were isolated, pOP12 was found to contain a 10.5-kb PstI fragment encoding the structural genes *glgC*, *glgA*, and *glgB*.

The arrangement of genes encoded by pOP12 has also been determined by deletion mapping experiments,¹⁰¹ and the nucleotide sequence of the entire *glg* gene cluster is known.^{43,169,239,324,378} The genetic and physical map of the *E. coli* K-12 *glg* gene cluster is as follows:

Asd glgB GlgX GlgC GlgA GlgP GlpD

The continuous nucleotide sequence of over 15 kb of this region of the genome has been determined and includes the sequences of the flanking genes *asd*³⁷⁹ and *glpD* (glycerol phosphate dehydrogenase).³⁸⁰ This region of the *E. coli* K-12 chromosome is located at 4140 kb on the physical map of Kohara *et al.*³⁸¹ and is situated within the region 3584–3594 kb on version 6 of the physical map of Rudd *et al.*³⁸²

Analysis of the nucleotide sequence showed that, in addition to the *glgC*, *glgA*, and *glgB* genes, pOP12 contains an open reading frame, *glgX*, situated between *glgB* and *glgC*, and a second ORF, *glgP*, originally referred to as *glgY*, located downstream from *glgA*.³⁸³ The deduced amino acid sequence of *glgX* shows significant similarities to regions of the α -glucanases and transferases, including α -amylases, pullulanase, cyclodextrin glucanotransferase, glycogen branching enzyme, and so on. The homologous regions include residues reported to be involved in substrate binding and cleavage by α -amylases and the amylase family.^{356,357} A recent report has shown that the *glgX* gene when expressed has glycogen debranching enzyme activity,³⁸⁴ with about 16-times more activity on a phosphorylase-limit dextrin product of glycogen than on native glycogen, confirming an earlier observation on the specificity of *E. coli* debranching enzyme.³⁸⁵

The *glgP* gene was identified through its homology with the phosphorylase from the rabbit muscle glycogen^{354,386} and the through its expression and the characterization of its gene product it was confirmed that it codes for a phosphorylase.³⁸⁶ Neither *glgX* or *glgP* is needed for glycogen synthesis, suggesting that both may be more involved in glycogen catabolism.³⁵⁶

The organization of the gene cluster suggests that the *glg* genes may be transcribed as two randomly arranged operons, *glgBX* and *glgCAP*.¹¹ The *glgB* and *glgX* coding regions overlap by one base pair, *glgC* and *glgA* are separated by two base pairs, and genes *glgA* and *glgP* are separated by 18 base pairs. The close proximity of these genes would suggest translational coupling within the two proposed operons. However, a noncoding region of approximately 500 bp separates *glgB* and *glgC*. Transcriptions initiating upstream of *glgC* have been analyzed by S1 nuclease mapping³⁸⁷ and studies of the regulation of the *glg* structural genes, using *lacZ* translational fusions and other approaches, are consistent with a two-operon arrangement for the *glg* gene cluster, in which the *glgCAP* and *glgBX* operons may be preceded by growth phase-regulated promoters.

Addition of exogenous cAMP to *E. coli* W4597(K) results in a modest enhancement in the rate of *in vivo* glycogen biosynthesis.^{388,389} It was observed that the genes *cya*, encoding adenylate cyclase (EC 4.6.1.1), and *crp*, encoding cAMP-receptor protein (CRP), are required for optimal synthesis of glycogen, and that exogenous cAMP can restore glycogen synthesis in a *cya* strain but not in a *crp* mutant.³⁹⁰

Both cAMP and CRP are strong positive regulators of the expression of the *glgC* and *glgA* genes, but do not affect *glgB* expression.³⁸⁷ Addition of cAMP and CRP to S-30 extracts with *in vitro* coupled transcription–translation reactions and pOP12 as the genetic template, resulted in about 25- and 10-fold increase in the expression of *glgC* and *glgA*, respectively, without affecting *glgB* expression. In reactions of completely defined composition, using the dipeptide synthesis assay, cAMP and CRP also enhanced the expression of *glgC* and *glgA* when encoded by either plasmids or restriction fragments.³⁹¹ The dipeptide synthesis assay measures protein expression by quantifying the formation of the first dipeptide of a specified gene product directed by a DNA template.³⁹¹ A restriction fragment containing *glgC* and a 0.5 kb of DNA upstream noncoding region of *glgC*, was sufficient to permit cAMP-CRP regulated expression in the dipeptide synthesis assay, suggesting that the *glgC* gene contains its own cAMP-regulated promoter(s). Gel retardation analyses demonstrated a CRP-binding site on a 243 bp restriction fragment from the upstream region of *glgC*.³⁸⁷ Potential consensus CRP-binding sequences within the *glgC* upstream region preceding both the *E. coli*³⁸⁷ and the *S. typhimurium*³⁹² *glgC* genes have been identified.

More details on the genetic regulation of glycogen synthesis are described in other reviews.^{7,11}

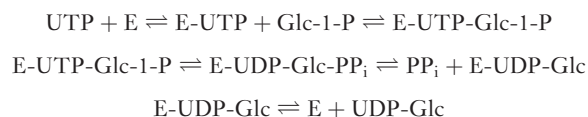
6.15.13 Properties of the Glycogen Biosynthetic Enzymes of Mammals

The main regulatory site for bacterial glycogen synthesis and for plant and algal starch synthesis is different from that of the mammalian glycogen synthesis. The differences in the mode of regulation for the whole pathway are probably connected to the fact that the glucosyl donors are different. ADP-Glc, the glucosyl donor in the bacterial and plant α -glucan systems and that UDP-Glc, the glucosyl donor for the mammalian glycogen synthesis, is also utilized for the synthesis of other sugar nucleotides, mainly UDP-galactose and UDP-glucuronate, precursors for the synthesis of several cellular constituents. Thus, the first unique reaction for the mammalian glycogen synthesis is the glycogen synthase step, and both allosteric and covalent modification regulations are exerted that are under hormonal-mediated control.

6.15.13.1 UDP-Glucose Pyrophosphorylase

The UDP-Glc PPase, which catalyzes the synthesis of UDP-Glc, seems to be ubiquitous in nature. It was first demonstrated in yeast³⁹³ and the enzyme has been isolated and characterized from bacteria, plants, and mammals.³⁹⁴ UDP-Glc PPases have been highly purified from calf liver, human liver, lamb, goat, and rabbit livers, and human erythrocytes.³⁹⁴ A molecular mass of 480 kDa was reported for the calf liver enzyme, with subunit of 60 kDa; the native enzyme is therefore an octamer of eight identical subunits. Other mammalian enzymes appear to have similar molecular masses. UDP-Glc PPases have an absolute requirement for a divalent cation and magnesium is the best cation for activity; Mn²⁺, Ca²⁺, and Ni²⁺ have activity to some extent. The optimum pH is in the range of 7–9 and the equilibrium constant in the direction of UDP-Glc formation ranges from 0.15 to 0.34 for the animal, plant, and bacterial enzymes studied.³⁹⁴ Although highly specific for UDP-Glc, the calf and human liver enzymes can also catalyze the pyrophosphorolysis of TDP-Glc, CDP-Glc, GDP-Glc, UDP-galactose, UDP-xylose, and UDP-mannose to a small extent, with 0.1–2.2% of the rate shown with UDP-Glc.³⁹⁵

The reaction mechanism of UDP-Glc PPase has been studied with the enzyme from liver, erythrocytes, and *Acanthamoeba castellanii*³⁹⁴ and is an ordered BiBi mechanism. The nucleoside phosphate is both the first substrate to be added and the last product to be released in the mechanism.



UDP-Glc is the most potent inhibitor of the animal UDP-Glc PPase. Thus, its concentration may have some regulation of the enzyme. The inhibition is competitive with UTP and it is suggested that the concentration ratio of UDP-Glc to UTP may be the most important determinant of UDP-Glc PPase activity.³⁹⁶ No other

regulatory phenomenon has been associated with the mammalian enzyme and because UDP-Glc functions not only in the synthesis of glycogen but also in the synthesis of other sugar nucleotides (e.g., UDP-galactose and UDP-glucuronic acid), it would be expected that the dominant regulation of glycogen synthesis would occur at its unique step and that is the glycogen synthase reaction.

6.15.13.2 Glycogen Synthase

Mammalian glycogen synthase has been purified from several sources, for example, skeletal and cardiac muscle, liver, adipose, and kidney.^{397–402} Also, cDNA representing the structural gene of the enzyme has been isolated (human muscle;⁴⁰³ rabbit muscle;⁴⁰⁴ rat liver;⁴⁰⁵ human liver;⁴⁰⁶ yeast^{407,408}) and the rabbit skeletal muscle cDNA has been expressed both in bacterial⁴⁰⁹ and in COS cells.^{410–412} The deduced amino acid sequences of the rabbit and human muscle glycogen synthases have 97% identity, while the rat and human liver enzymes are 92% similar. The two yeast glycogen synthases are 80% identical but they have only a 50% overall identity to the muscle glycogen synthases. The amino acid similarity between the enzymes from muscle and liver in humans is only of 60%, with the lowest identity in the N- and C-terminal regions of the proteins. The human and rat liver enzymes are truncated by 32–34 amino acids compared to the rabbit and human muscle glycogen synthases. Glycogen synthases from all of these sources seem to be composed of identical (or very similar) subunits of molecular weight 80 000–85 000. The native forms from liver or adipose tissue are aggregates of two identical subunits, whereas that from muscle contains four. Glycogen synthase exists in at least two forms; a phosphorylated form, arising from the covalent modification of serine residues by ATP; and a dephosphorylated form, which can be obtained using phosphatase on the phosphorylated form (**Figure 2**). These two forms were originally named as the ‘a’ (or I) (unphosphorylated) and ‘b’ (or D) (phosphorylated) forms; the b-form was dependent on Glc-6-P for activity, whereas the a-form was active in the absence of Glc-6-P. The a- and b-forms are also distinguished on the basis of K_m for the substrate, UDP-Glc; the b-form usually has a higher K_m (lower apparent affinity) than the a-form. It is apparent that the a-form is the physiologically active form of the enzyme while the b-form, phosphorylated, is an inactive form of the glycogen synthase.

6.15.13.3 Branching Enzyme

The BEs from rat liver^{413–415} and rabbit skeletal muscle^{416–419} have been studied in some detail. Lerner⁴¹⁶ showed that the enzyme catalyzed the formation of new (α -1-6)- α -D-glucosidic bonds with glycogen containing average chain lengths of 11–21 glucosyl residues. Using the rat liver⁴¹⁵ and the rabbit skeletal muscle⁴¹⁸ systems, it was found that the enzymes cleaved linear chains of six or more glucosyl residues from the terminal portion of the outer chains of the α -1-4- α -D-glucan substrate, and then transfer and reattach the cleaved oligosaccharide portion in a 1,6- α -D-glucosidic linkage to the outer portions of the α -D-glucan. The BE enzyme from rabbit skeletal muscle preferentially catalyzed the transfer of segments 7 glucosyl units long when it acted on polysaccharides elongated by rabbit muscle glycogen synthase action *in vitro*.⁴¹⁷ The liver and

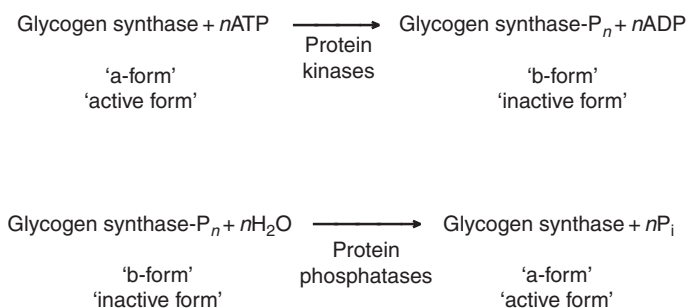


Figure 2 Covalent modification of glycogen synthase by phosphorylation by different protein kinases and dephosphorylation by protein phosphatase 1.

muscle enzymes are also active on amylose and amylopectin molecules.^{413,416} The rabbit muscle enzyme also catalyzed the formation of new 1–6- α -D-linkages in products formed by the reaction of phosphorylase with Glc-1-P and also greatly stimulated the ‘unprimed’ phosphorylase reaction.⁴²⁰ BE in a combined system with the mammalian glycogen synthase stimulates the rate of transfer of glucose units in the presence of a minimal quantity of glycogen primer.⁴¹⁶ This stimulation is very similar to that observed for the bacterial BE on bacterial glycogen synthase activity, and for the stimulation of plant BE on starch synthase activity. There is some information about the minimal size of oligosaccharide chain that can be transferred by the action of BE; the relevant sections discuss this aspect of the action of the plant and bacterial enzymes. What is not known for any BE is the amino acid residues involved in determining the specificity of size of oligosaccharide cleaved and transferred, and the interbranch distance between formation of the new branch points. This important information remains to be elucidated with respect to the mechanism of BE catalysis.

Human glycogen storage disease (‘type IV’) is due to the absence of BE.^{420,421} The glucan product isolated from the liver of these patients is an amylopectin-type polysaccharide. Thus, some branches have been formed, suggesting that some branching activity is present at low levels. Purified rabbit muscle BE can cause further branching of this α -D-glucan.⁴¹⁸ The rabbit skeletal muscle enzyme has been purified to near homogeneity.⁴¹⁸

The apparent molecular weight of the mammalian BE is about 92 kDa, as determined by sucrose density-gradient centrifugation. Thus, it is similar to the molecular weights of the plant and bacterial enzymes. The enzyme has a broad pH optimum in citrate buffer, 6.8–7.8, and is stimulated about twofold in 0.5 mol l⁻¹ sodium citrate at pH 7.0. The partially purified liver enzyme was also activated 2-fold by sodium citrate and ~1.7-fold by sodium borate.⁴¹³

BE genes have been isolated from yeast (*Saccharomyces cerevisiae*)⁴²² and from a human hepatoma cell line.⁴²³ The deduced amino acid sequence from the yeast *GLC 3* gene was compared to the bacterial BEs from *E. coli*,³¹⁹ *Synechococcus* and *Bacillus stearothermophilus*⁴⁹ and there was only 8% identity and 42% similarity of the yeast BE with the prokaryotic sequences. The cDNA encoding the human BE could complement the yeast BE mutant, *glc 3* and had a 67% identical amino acid sequence with the yeast BE.⁴²³ The human gene was located on chromosome 3.

6.15.13.4 Glycogenin

Two reviews describing the discovery and characterization of glycogenin are available.^{424,425} This protein, about 37 kDa in size, was first found to be covalently bound to glycogen⁹² and also associated with glycogen synthase even after extensive purification of the rabbit muscle glycogen synthase.⁴²⁶ Most of the glycogenin in rabbit muscle is considered to be linked covalently to glycogen. In liver, however, most of the glycogen and glycogenin are in free form and not associated with each other.⁴²⁷ The first step in converting the apoglycogenin into a primer for glycogen synthesis is an autoglucosylation of the glycogenin tyrosine residue 194 by UDP-Glc. The reaction absolutely requires either Mn²⁺ or Mg²⁺ for the autoglucosylation reactions. Up to 7–11 glucosyl units can be attached to the glucosyl-tyrosine residue. The complete sequence of the rabbit muscle glycogenin has been elucidated.⁹¹

Glycogenin, when isolated and purified, contains a glycosylated tyrosine residue and so there was always a question whether the first glucosyl unit was due to autoglucosylation by glycogenin or whether there was another enzyme responsible for the first glycosylation. The rabbit muscle glycogenin was expressed in *E. coli*^{428,429} and purified. The Tyr194 residue was already glucosylated and contained from 1 to 8 residues of glucose. The glycogenin could incorporate another 5 mol of glucose per mol of glycogenin if supplied with UDP-Glc, the K_m was 4.5 mmol l⁻¹.

Isoamylase can remove the oligosaccharide chain from the tyrosine residue⁴³⁰ and pretreatment of the glucosylated glycogenin enhanced incorporation of labeled glucose from UDP-Glc. This suggested that glycogenin can self-glucosylate its tyrosine 194 residue. More direct evidence was obtained by expressing the glycogenin in an *E. coli* mutant deficient in UDP-Glc PPase activity,⁴³¹ resulting in the production of a carbohydrate-free glycogenin, apoglycogenin. When UDP-xylose + Mn²⁺ was incubated with the glycogenin, one mol of xylose was incorporated per mol of glycogenin. With UDP-Glc, an average of 8 glucose chains are

added per glycogenin. However, upon release of the glucose chains by isoamylase, the size of chains varied with the predominant chains being in the 7–11 glucose units range. The production of a carbohydrate-free apoglycogenin and its ability to self-glucosylate eliminates the need to invoke a separate enzyme for the addition of the first glucose residue to Tyr194.

Thus, glycogenin can catalyze two different glycosylation reactions; first, the glycosylation of a tyrosine hydroxyl group and then further glycosylations to form α -1,4-glucosidic linkages. UDP and UTP were found to be effective inhibitors.⁴³² Other pyrimidine base sugar nucleotides could substitute for UDP-Glc as glucosyl donors.⁴³³ The rate of glycosylation using CDP-Glc and TDP-Glc were 71 and 33%, respectively, of the rate with UDP-Glc; both ADP-Glc and GDP-Glc were inactive. UDP-xylose could also serve as a glucosyl donor with only one xylose molecule being transferred to glycogenin itself.⁴³⁴ No further chain growth could occur with either UDP-xylose or UDP-Glc. Other reactions were also catalyzed by glycogenin. The glycogenin could transfer glucose from UDP-Glc to exogenous substrates such as *p*-nitrophenyl-linked malto-oligosaccharides,⁴³⁵ tetradecyl- β -D-maltoside, octyl- β -D-maltoside, and dodecyl- β -D-maltoside.⁴³⁶

If the recombinant glycogenin is mutated from Tyr to Phe or Thr at residue 194, the enzyme loses its ability to self-glucosylate.^{429,432,433} However, the Phe194 and Thr194 mutants are still able to glycosylate with UDP-Glc, dodecyl- β -D-maltoside,⁴³³ and *p*-nitrophenyl- β -linked malto oligosaccharides.⁴²⁹ Also noted was the ability of the mutant and normal glycogenins to hydrolyze UDP-Glc at rates similar to self-glucosylation rates of the normal enzyme. This hydrolysis is competitive with the glucose transfer to *p*-nitrophenyl-linked maltoside.⁴²⁹ The self-glucosylation, glycosylation of other acceptors, and hydrolysis all appear to be catalyzed by the same active center.

Glycogenin and the mutant proteins, Phe194 and Thr194, could also transfer glucose from UDP-Glc to maltose to form maltotriose.⁴³⁷ However, no further conversion to a higher oligosaccharide occurred. Analysis of the crystal structure by X-ray diffraction indicated that glycogenin existed similar to dimers.⁴³⁷

There have been some reports of a glycogenin in *E. coli*, similar to the mammalian glycogenin, involved in the initiation of glycogen synthesis. Barengo *et al.*⁴³⁸ reported the formation of a labeled TCA insoluble fraction upon incubation of extracts of *E. coli* with UDP-Glc-14C. The radioactivity was solubilized by α -amylase suggesting the label was an α -1,4-glucosyl oligosaccharide attached to a protein. Evidence was also presented to suggest that this labeled fraction was an intermediate in glycogen synthesis. However, the glycogenin of bacteria has not been well-characterized. Indeed it is now believed that the bacterial glycogen synthase is able to initiate *de novo* synthesis of glycogen without a primer and only from ADP-Glc.⁴³⁹

6.15.13.4.1 Genetic evidence indicating that glycogenin is required for glycogen synthesis

In *S. cerevisiae*, two genes, Glg1p and Glg2p coding for self-glucosylating proteins encode proteins of 618 and 380 amino acids, respectively and have 55% sequence identity over their N-terminal 258 amino acids. The two proteins, Glg1p and Glg2p, have amino acid sequences that are 33 and 34% identical, respectively, to that of rabbit muscle glycogenin in the N-terminal region of 258 amino acids.⁴⁴⁰ Thus, they are larger than the muscle glycogenin that has 332 amino acids. The COOH termini of Glg1p and Glg2p is largely nonidentical in sequence except for two small segments of sequence similarity. Each contains a Tyr residue in correspondence with the rabbit muscle Tyr194, the residue in Glg1p and in Glg2p is Tyr232.

When the Glg1p and Glg2p genes were disrupted separately by homologous recombination, there was little effect on glycogen accumulation,⁴⁴⁰ but loss of both genes caused almost the complete loss of glycogen. Glycogen synthase activity was normal in this double mutant so this was not the reason for the lack of glycogen synthesis. Glycogen synthesis was almost completely restored when the rabbit muscle glycogenin was expressed in the double Glg1p, Glg2p mutant, that is, the mammalian glycogenin could complement the double mutant deficiency. These data indicate that the Glg1p and Glg2p genes were involved in the initiation of glycogen synthesis.⁴⁴⁰ This report thus provides *in vivo* evidence of the requirement of glycogenin in the biosynthesis of glycogen in eukaryotes.

6.15.14 Regulation of Mammalian Glycogen Synthesis

6.15.14.1 General Considerations

As indicated, the site of regulation of glycogen synthesis in bacteria and starch synthesis in plants is at the ADP-Glc PPase step, and is different from the site of regulation in mammals. In mammalian systems, the regulatory enzyme is glycogen synthase. The difference in regulatory sites of the various systems may be linked to the difference in specificity for the glucosyl donor (ADP-Glc for the bacterial and plant α -D-glucan systems, UDP-Glc for mammals). UDP-Glc is utilized for the synthesis of other intermediaries required for the synthesis of many cellular constituents. The first unique reaction for mammalian glycogen synthesis, therefore, after synthesis of the glucosylated acceptor protein glycogenin, is the glycogen synthase step, where both allosteric control and covalent modification control are exerted. In contrast, in bacteria and plants, the only known function for ADP-Glc is the synthesis of α -1,4-D-glucosyl bonds in bacterial glycogen and starch. Thus, the prokaryote and plant cells regulate α -glucan synthesis at the level of ADP-Glc formation so as to conserve ATP utilized for synthesis of the sugar nucleotide.

The findings that regulation occurs at the glycogen synthase step in the mammalian systems and at the ADP-Glc PPase step in bacteria and plants is consistent with the concept that major regulation of a biosynthetic pathway does occur at the first unique step of the pathway. One can also recognize the need in mammalian systems for a type of regulation that involves an efficient, rapid on-off type of control of glycogen synthesis that permits synthesis of glycogen when carbohydrate or carbon is plentiful in the diet, but prevents synthesis and permits degradation of glycogen to occur during muscular contraction or during starvation. Such a mechanism involves covalent modification of the enzyme catalyzing the limiting reaction of the process in order to produce either inactive or active forms of the enzyme.

6.15.14.2 Regulation of Glycogen Synthase by Phosphorylation – Dephosphorylation

The ratio of the activity of the glycogen synthase in absence of Glc-6-P or the activity in presence of Glc-6-P, has been used as a measure of the state of phosphorylation of the enzyme. The a-form was recognized as the primary active species within the cell, especially as modulated by hormonal control. The cAMP-dependent protein kinase was shown to catalyze phosphorylation and, consequently, inactivation of this enzyme. Subsequent studies also indicated that on each subunit of the glycogen synthase there were multiple and unique phosphorylation sites. Originally, Smith *et al.*⁴⁴¹ reported that, if partially purified enzyme was incubated extensively with ATP and subsequently purified, the inactive glycogen synthase contained 6 mol of phosphate per subunit. Although this information did not initially gain general acceptance, Picton *et al.*^{442,443} showed there are seven serine phosphorylation sites on the rabbit skeletal muscle glycogen synthase. Currently, the number of potential phosphorylation sites found *in vivo*, are nine and more than ten sites can be phosphorylated *in vitro*.²³ Table 14 shows the various *in vivo* phosphorylation sites and the major protein kinases that are involved in the phosphorylation of those sites as shown *in vitro*.

Phosphorylation of these sites is catalyzed by more than six kinases.^{23,444,445} Phosphorylation at the different sites synergistically inactivates the enzyme; however, the effects observed may vary depending on the conditions used to assay the glycogen synthase. The effects of multisite phosphorylation would depend on the concentration of effectors used, such as Glc-6-P, or the substrate, UDP-Glc. For example, with preparations of enzyme containing 0.27–3.49 mol of alkali-labile phosphate per mol glycogen synthase subunit, Roach and Lerner⁴⁴⁴ reported that the $A_{0.5}$ (concentration of activator, Glc-6-P, required for 50% of maximal activation) and $S_{0.5}$ (concentration of substrate, UDP-Glc, attaining 50% of maximal velocity) varied with phosphate content, from 3.3 $\mu\text{mol l}^{-1}$ to 2.7 mmol l^{-1} and from 0.75 mmol l^{-1} to at least 60 mmol l^{-1} , respectively. Both parameters increased with phosphate content. The greatest absolute change occurred at values greater than 2 mol of phosphate bound per enzyme subunit. Plots of activity versus Glc-6-P concentration became more sigmoidal with increasing enzyme phosphate content. Activation by Glc-6-P was related primarily to modulation of UDP-Glc affinity. Several inhibitors such as ATP, ADP, AMP, UDP, and P_i had increasing effects with enzyme of increasing alkali-labile phosphate content. These investigators presented a scheme in which glycogen synthase activity is sigmoidally and inversely dependent on the state of

Table 14 *In vivo* phosphorylation sites in rabbit muscle glycogen synthase

| <i>Phosphorylation site</i> | | |
|-----------------------------|-------------|---|
| <i>Residue</i> | <i>Name</i> | <i>Protein kinases</i> |
| 7 | 2 | cAMP-dependent protein kinase Calmodulin-dependent protein kinase II Phosphorylase kinase Protein kinase C |
| 10 | | Casein kinase I |
| 640 | 3a | Glycogen synthase kinase 3 cAMP-dependent protein kinase |
| 644 | 3b | Glycogen synthase kinase 3 |
| 648 | 3c | Glycogen synthase kinase 3 |
| 652 | 4 | Glycogen synthase kinase 3 cAMP-dependent protein kinase |
| 656 | 5 | Casein kinase II (Glycogen synthase kinase 5) |
| 697 | 1a | cAMP-dependent protein kinase Protein kinase C |
| 710 | 1b | cAMP-dependent protein kinase Calmodulin-dependent protein kinase II |

Modified from Roach.²³ The sites listed in this table are only those that were shown to be labeled *in vivo*. The protein kinases listed for the sites are those that phosphorylate the enzyme, *in vitro*.

phosphorylation. The hormonal effects of insulin were counteractive to those of epinephrine, glucagon, and so on, respectively decreasing or increasing the extent of phosphorylation. Inhibitors such as UDP, ATP, and AMP, accentuate the inhibition by phosphorylation, whereas the activator, Glc-6-P and substrate, UDP-Glc diminished the extent of such inhibition.

As seen in **Table 14**, there are two phosphorylation sites at the N-terminal region of the rabbit skeletal muscle glycogen synthase. The remaining seven are situated at the C-terminal region. The cAMP-dependent protein kinase preferentially phosphorylates three sites, 1a, 1b, and 2.⁴⁴⁶ The cAMP-dependent kinase can also phosphorylate sites 3a and 4 but at a much slower rate⁴⁴⁷ and thus is not considered to be as important as glycogen synthase kinase 3 for those sites. There are overlapping specificities among the different protein kinases for site 2 as phosphorylase kinase, calmodulin-dependent protein kinase and protein kinase C also can phosphorylate this site.²²

Of interest is the phosphorylation of site 5 catalyzed by casein kinase II. Picton *et al.*⁴⁴³ showed that dephosphorylation at site 5 did not alter the regulatory kinetics of rabbit muscle glycogen synthase, nor did rephosphorylation of site 5 by casein kinase II (also called glycogen synthase kinase-5) affect the activity of the glycogen synthase. In other words, the site 5 phosphorylated glycogen synthase did not depend to any extent on Glc-6-P for maximal activity. However, the presence of phosphate at site 5 was necessary for the phosphorylation of sites 3a, 3b, and 3c by glycogen synthase kinase 3,⁴⁴⁸ which did increase the dependency of the glycogen synthase activity for Glc-6-P. The phosphate at site 5 appears to be highly stable as it is resistant to dephosphorylation by the rabbit muscle protein phosphatases but can be removed by potato acid phosphatase. Phosphorylation of sites 1a, 1b, or 2 did not require the presence of phosphate at site 5.

This observation was confirmed and extended by Fiol *et al.*^{449,450} They synthesized a peptide corresponding to the rabbit muscle glycogen synthase amino acid sequence containing sites 3a, 3b, 3c, 4, and 5. Synergism was observed between casein kinase II phosphorylation of site 5 and phosphorylation of sites, 3a, 3b, 3c, and also another site, 4, by GSK-3, of the synthesized peptide. Indeed, phosphorylation of site 5 was obligatory for the phosphorylation of the four sites at 3 and 4 by glycogen synthase kinase-3. As seen in **Table 15**, the GSK-3 sites were regularly spaced every fourth residue in the motif, SXXXS(P). It was also found that the phosphorylations by GSK-3 were ordered: first, site 4 was phosphorylated, followed by site 3c, 3b, and finally 3a.⁴⁵¹ This was clearly shown by synthesizing a series of peptides where the sites 3a, 3b, 3c, and 4 were replaced, one at a time,

Table 15 Amino acid sequences of the rabbit muscle glycogen synthase having the serine sites that are phosphorylated

(a) N-terminal phosphorylation sites

7 10
 PLSRTL**SV**SSLPGL-----
 (site 2)

(b) C-terminal phosphorylation sites

640 644 648 652 656
 RYPRP**ASVPPSP**SL**SRHSSPH**CSEDEEEPRDGLPEEDGERYDEDEEAAKD
 (sites 3a 3b 3c 4 5)

697 710
 RRNIRAPQWPRRAS**CTSSSGGSKRSNS**VDTS**SL**STPSEP-----
 (sites 1a 1b)

The serine groups that are phosphorylated are in bold.

with alanine. The alanine at site 4 peptide could not be phosphorylated at the site 3 even though it was phosphorylated at site 5 by GSK-3. Also it was observed that GSK-3 would not phosphorylate the serine residue at sites, N-terminal of the site containing the alanine residue. With alanine replacing serine at site 3b, only sites 4 and 3c were phosphorylated. With alanine substituted at site 3c, only site 4 was phosphorylated. Thus, the multiple phosphorylation by GSK-3 was of an obligate order, first 4 then 3c, 3b, and then 3a. Most probably GSK-3 recognizes the motif SXXXS(P) and this may explain the need for GSK-5 (casein kinase II) to phosphorylate site 5. The sequential formation of new recognition sequences, SXXXS(P) at the sites 5, 4, 3c, 3a, and 3b, would explain the ordered phosphorylation.

This interdependency of GSK-3 with casein kinase II has been defined as hierarchal phosphorylation.²³ Another example of the hierarchal phosphorylation is seen with cyclic-AMP-dependent protein kinase and casein kinase I. The cAMP-dependent protein kinase enhances phosphorylation of the glycogen synthase by casein kinase I.⁴⁵² The phosphorylation by casein kinase I was serine residue 10.⁴⁵¹ Synthetic peptides based on the four phosphorylated regions in the muscle glycogen sequence (residues 694–707, 706–733, 1–14, and 636–662) were synthesized and phosphorylated. Casein kinase I could not phosphorylate the unphosphorylated peptides but, if cAMP-dependent kinase phosphorylated peptides, 694–707, 706–733, and 1–14, all three peptides were easily phosphorylated by the casein kinase I.⁴⁵¹ The greatest stimulation was seen with peptide 1–14. In the case of peptide 1–14, the phosphorylation site was at Ser10 and in the case of the peptides 694–707 and 706–733, the phosphorylated residue was Thr 713. However, the rate of phosphorylation was 20- to 40-times greater at Ser10 than Thr713. Moreover, Ser10 was demonstrated to be phosphorylated *in vivo*, but Thr713 was not.²³ Thus, the physiologically important site for phosphorylation by casein kinase I is considered to be Ser10 and this phosphorylation is considered to be dependent on an initial phosphorylation of Ser7 (site 2) by the cAMP-dependent protein kinase.

The phosphorylation of peptide 636–662, the peptide encompassing glycogen synthase phosphorylation sites 3a, 3b, 3c, 4, and 5⁴⁵¹ is of great interest. Using the peptides where the alanine residue was substituted for the serine residue at the different phosphorylation sites, it was shown that serine residues 646 and 651 (Table 15) were phosphorylated by casein kinase I and this phosphorylation was significantly enhanced by prior phosphorylation of the sites 3a, 3b, 3c, 4, and 5.⁴⁵² However, it is still not clear whether these sites are phosphorylated *in vivo* and whether their phosphorylation by casein kinase I substantially affects the glycogen synthase activity. These studies mainly by Roach's group also indicate that the recognition of the serine phosphorylation site by casein kinase I is -S(P)-XXS- and for GSK-3 is SXXX-S(P).

6.15.14.2.1 The effect of phosphorylation on glycogen synthase activity and relative effects of phosphorylation on different sites

6.15.14.2.1(i) In vitro studies Phosphorylation of the glycogen synthase activity leads to decreased activity but phosphorylations at different sites have different effects. Little or no inactivation is seen with phosphorylation at sites 5, 1a, and 1b while phosphorylation of site 2 gives moderate inactivation. A most potent inactivation

is seen with phosphorylation by GSK-3 at sites 3a, 3b, and 3c.^{22,23} Although, site 5 phosphorylation does not cause any change in activity it is functional in regulation as sites 3 cannot be phosphorylated unless there is an initial phosphorylation of site 5. Phosphorylation of sites 1a, 1b, and 2 by cAMP-dependent protein kinase leads only to a partial inactivation of the rabbit muscle glycogen synthase, and phosphorylation of site 2 did not decrease activity in rabbit⁴⁵³ or rat liver glycogen synthase.⁴⁵⁴ However, casein kinase I phosphorylation of Ser-10, proceeding after phosphorylation of sites 1a, 1b, and 2 in rabbit muscle glycogen synthase and 2 in liver glycogen synthase, causes a total inactivation of the muscle⁴⁵³ and liver enzymes.⁴⁵⁴ It should be pointed out that sites 1a and 1b are absent in rat liver^{405,453,455} and human liver glycogen synthase.⁴⁰⁶

Thus, the secondary phosphorylations by casein kinase I of Ser residue 10, and by GSK-3 of sites 3a, 3b, 3c, and 4 have greater effects on the activity of glycogen synthase than the primary phosphorylations by cAMP-dependent kinase (sites 2, 1a, and 1b) and glycogen synthase kinase-5 and there may be two different routes to inactivating glycogen synthase, the GSK-3 sites and the casein kinase I site.

It should also be mentioned that the above studies were done with a recombinant glycogen synthase expressed in *E. coli*⁴⁰⁹ and the same results were seen with phosphorylation site peptide analogues. Namely, the dependency of phosphorylation by GSK-3 on the prior phosphorylation by casein kinase II, the potent inactivation by GSK-3, the partial inactivation by phosphorylation by cAMP-dependent kinase, the stimulation of phosphorylation by casein kinase I by prior phosphorylation by phosphorylase kinase, and the greater inactivation of glycogen synthase after the combined phosphorylation by casein kinase I and cAMP-dependent protein kinase.

Further studies of the recombinant glycogen synthase expressed in *E. coli* allowed Wang and Roach⁴⁵⁶ to generate mutant forms of the rabbit muscle glycogen synthase at the GSK-3 phosphorylation sites, S640A, S644A, and S648A (sites 3a, 3b, and 3c respectively). All three mutants had high \pm Glc-6-P ratios of activity (0.8–0.9). Phosphorylation of the mutants were done with GSK-3 and casein kinase II. The mutants phosphorylated at sites 5 and 4 (mutant S648A) and at sites 3c, 4, and 5 (mutant S644A) had full activity. When sites 3b, 3c, 4, and 5 (mutant S640A) were phosphorylated the activity ratio decreased modestly to about 0.6–0.7. When all sites were phosphorylated in the recombinant enzyme, the activity ratio decreased to 0.1. The results of this study demonstrated that phosphorylation site 3a and to a lesser extent 3b, correlated with the inactivation of the glycogen synthase. The apparent affinity constant for the activator, Glc-6-P, was 2.4 mmol l^{-1} and only increased appreciably, to 24 mmol l^{-1} , when site 3a was phosphorylated with the other four sites.

6.15.14.2.1(ii) In vivo studies The studies *in vitro* discussed above clearly showed that for mammalian glycogen synthase (or at least for the rabbit muscle enzyme) the important phosphorylation sites for inactivation were site 2 and (2a Ser10), sites 3a and 3b. However, it is of interest to know which are the important phosphorylation sites under hormonal control *in vivo*. In rabbit muscle enzyme, intravenous insulin administration doubles the \pm Glc-6-P activity ratio, with a decrease in phosphorylation of all the sites.^{22,457,458} Epinephrine, which increases the phosphate content of the glycogen synthase, increases the phosphorylation at practically all the phosphorylation sites.^{457,459–461}

With respect to the liver glycogen synthase, the hormones glucagon, vasopressin, and epinephrine all increased the phosphorylation of the peptide regions containing the sites 2 and 3^{22,23} and Akatsuka *et al.*⁴⁶² showed that glucagon promoted phosphorylation of the casein kinase I site now referred to as site 2a or the equivalent liver Ser residue to muscle glycogen synthase residue, Ser10.

To study in detail the role of individual phosphorylation sites in the regulation of the rabbit muscle glycogen synthase, the enzyme was overexpressed in COS M9 cells.^{410,411} The activity ratio of \pm Glc-6-P was found to be very low, ~ 0.01 , indicative of a high level of phosphorylation. Ser to Ala mutations were introduced singly, or in combinations, at the nine known phosphorylation sites and it was found that no single Ser to Ala mutation caused a substantial increase in the activity ratio.⁴¹² It was shown that simultaneous mutations were needed at both regions of site 2, the N-terminal region, and site 3, the C-terminal region. The most effective combinations were mutations at site 3a (Ser640) or site 3b (Ser644) together with site 2 (Ser7). Double mutants, Ser640Ala Ser7Ala, Ser644Ala–Ser7Ala, and Ser10Ala (site 2a)–Ser640Ala, gave activity ratios of 0.59, 0.25, and 0.21, respectively. These results were consistent with site 2 phosphorylation being a prerequisite for phosphorylation of site 2a. In contrast with the results obtained *in vitro*, the mutation of site 5 (Ser656), although affecting phosphorylation at sites 3 and 4, did not result in an increase in the activity ratio;⁴¹² the authors proposed that in

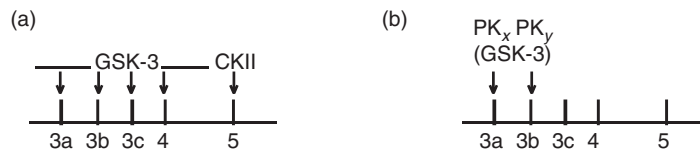


Figure 3 Phosphorylation of glycogen synthase by glycogen synthase kinase 3 (GSK-3). (a) *In vitro*, phosphorylation occurs first at site 5 by casein kinase II. Then GSK-3 phosphorylates in sequence site 4, then 3c, 3b, and last, 3a. (b) Phosphorylation of sites 3a and 3b by different protein kinases, PK_x and PK_y. *In vivo*, these kinases can phosphorylate sites 3a and 3b independent of prior phosphorylation of sites 3c, 4, and 5. PAS kinase has been shown to phosphorylate site 3a (S640).²²⁵

COS cells sites 3a and 3b may be phosphorylated by an alternative pathway independent of phosphorylation of site 5. Nevertheless, the COS cell data did show that the most important sites in the regulation of glycogen synthase were sites 2, 2a, 3a, and 3b, a conclusion consistent with the *in vitro* data.

This system was studied further and it was shown that the phosphorylation of sites 3a and 3b occurred even when mutations were made at sites 5, 4, and 3c;⁴⁶³ thus, phosphorylation of sites 3a and 3b may occur via other protein kinases other than GSK-3. Evidence supporting this view has recently been obtained by Skurat and Roach,⁴⁶⁴ who mutagenized amino acid residues close to the phosphorylation sites, 3a and 3b, that may be important for a protein kinase to recognize and phosphorylate sites 3a and 3b; that is, arginine residue 637 and proline residue 645. The mutants made, were Arg637Gln and Pro645Ala, a double mutant, R637Q, S644A (site 3b) and two triple mutants, S7A (site 2), R637Q, S644A and S7A, S644A, P645A. In addition, the serine residue of sites 3c, 4, and 5 of these mutants were mutagenized to alanine to avoid possible phosphorylation of sites 3a and 3b by GSK-3. Mutation of Arg637 to Gln eliminated phosphorylation of site 3a, suggesting that Arg637 may be important for another protein kinase to recognize site 3a. The mutant Pro645Ala also eliminated phosphorylation of site 3b suggesting a possible involvement of a 'proline-directed protein kinase.' Either mutation alone did not substantially increase the activation ratio meaning that phosphorylation at either site plus phosphorylation at sites 2 and 2a produced a totally inactive enzyme; the triple mutant, S7A, R637Q, S644A, however, was active with an activity ratio of 0.62 while S7A, S644A, P645A had an activity ratio of 0.21.⁴⁶⁴ The results also point out that in the COS cells, sites 2, 2a, 3a, and 3b are all important for regulation of glycogen synthase and most significantly, suggest that sites 3a and 3b can be phosphorylated independently of each other by distinct protein kinases. Thus, the existence of three protein kinases is proposed for the phosphorylation of sites 3a and 3b⁴⁶⁴ and are shown in **Figure 3**.

Thus, in the cell, multiple mechanisms involving at least three protein kinases in the regulation of sites 3 of glycogen synthase and two protein kinases for regulation of sites 2 may exist. The reason for redundant mechanisms of inactivation of glycogen synthase is at present, unknown, but could be the way for integrating messages of a number of hormonal and signal transduction pathways. It should be pointed out, however, that the *in vivo* studies were done in COS cells and whether the same phenomena, particularly the unidentified protein kinases, PK_x and PK_y, are relevant to control glycogen synthase in skeletal muscle, remains to be established.

Indeed, a kinase, a human PAS kinase, has been shown to phosphorylate rabbit muscle glycogen synthase specifically at serine residue 640.⁴⁶⁵ The PAS kinase in order to phosphorylate glycogen synthase requires its PAS domain for binding to the glycogen synthase as truncation of the PAS kinase where the PAS domain is missing the truncated kinase cannot phosphorylate the glycogen synthase. Glycogen negatively regulates the PAS kinase glycogen synthase interaction.⁴⁶⁵ So it is believed that the PAS kinase glycogen synthase interaction is regulated by the metabolic status of the cell.

Other interesting experiments done in the COS cell system indicated that overexpression of the glycogen biosynthetic enzymes, glycogen synthase, glycogenin, BE, and UDP-Glc PPase, alone, did not lead to increased or overaccumulation of glycogen.⁴⁶⁴ If however, the glycogen synthase mutant S7A, S640A was overexpressed, then there was about 2- to 2.5-fold increase in glycogen levels, suggesting that glycogen synthase activity was rate limiting. The co-overexpression of the mutant glycogen synthase with the UDP-Glc PPase or with glycogenin led to even greater glycogen accumulation, that is, about a 34–70% increase. Thus, with overexpression of the hyperactive glycogen synthase, the synthesis of UDP-Glc or synthesis of the glycosylated acceptor protein may become the rate-limiting reaction.

6.15.14.3 Mechanism of Stimulation of Mammalian Skeletal Muscle Glycogen Synthesis by Insulin

6.15.14.3.1 Protein phosphatase 1

Because the glycogen synthase is phosphorylated at multiple sites per subunit, it would be of interest to know whether one or several protein phosphatases catalyze dephosphorylation of these distinct sites. Four protein phosphatases are known; for background on the structures and biochemical properties of these phosphatases.^{466–468} Of these phosphatases, protein phosphatase 1, which hydrolyzes mainly the phosphate of the b-subunit of phosphorylase kinase, has strong activity on the phosphate sites 1a, 2, 2a, 3a, 3b, 3c, and 4 of glycogen synthase and is the principle enzyme in dephosphorylating glycogen synthase since both the phosphatase and glycogen synthase are usually bound to the glycogen particle. Protein phosphatase 2A, which has greater activity on the a-subunit phosphates of phosphorylase kinase, than protein phosphatase 1, also has activity on the above glycogen synthase phosphate sites.

Of interest is that protein phosphatase 1, of 37 kDa, is associated with the glycogen particles *in vivo* when it is in a complex with a 160-kDa protein referred to as the G subunit.^{469,470} When the protein phosphatase 1 is associated with the G subunit, it binds to the glycogen particle and is far more enzymatically active under physiological conditions, in dephosphorylating glycogen synthase, glycogen phosphorylase and phosphorylase kinase (enzymes that also bind to the glycogen particle) than when it is free.⁴⁷¹ The G subunit is phosphorylated by cAMP-dependent protein kinase *in vitro*^{469,472} and in response to epinephrine *in vivo*.^{473,474} Phosphorylation occurs at two serine site, site 1 and 2, separated by 18 residues and the phosphorylation of both sites causes a dissociation of the phosphatase from the G subunit.⁴⁷² The phosphatase is about five- to eightfold less active than the complex in dephosphorylating glycogen synthase and phosphorylase present in the glycogen particle and this lowering of the protein phosphatase activity is one way in which epinephrine stimulates glycogen breakdown and inhibits glycogen synthesis.

The dissociation of subunit G from the phosphatase correlates with the phosphorylation of site 2, and not with that of site 1;⁴⁷² and reassociation of G subunit with the phosphatase occurs with dephosphorylation of site 2 by protein phosphatase 2A under conditions where site 1 still retains the phosphate residue as it is more resistant to dephosphorylation by protein phosphatase 2A. Thus, inactivation of the protein phosphatase is due to phosphorylation of site 2 and not site 1.

Parker *et al.*⁴⁵⁷ showed that, in response to insulin administration, the phosphate released from glycogen synthase is mainly from sites 3a, 3b, and 3c, suggesting that insulin caused the inhibition of GSK-3 or activated protein phosphatase 1. As will be shown, both phenomena, activation of protein phosphatase 1 and inhibition of GSK-3, do occur due to insulin.

Dent *et al.*⁴⁷⁵ showed that the G protein phosphatase complex with phosphate mainly at site 1, had protein phosphatase activity which was now associated with the glycogen particle and about 2.5- to 3-fold higher activity than the dephosphorylated protein phosphatase. Moreover, a protein kinase was isolated from rabbit skeletal muscle that phosphorylated the G subunit at site 1 but not at site 2. This protein kinase also phosphorylates a ribosomal protein, S6 *in vitro*. The activity of this protein kinase was increased about twofold within 15 min after insulin administration, the same time frame for the increase in glycogen synthase activity.⁴⁵⁷ The phosphorylation of the G subunit increased phosphatase activity about 2.8-fold on glycogen synthase and phosphorylase kinase. It was also demonstrated that insulin administration stimulated *in vivo* phosphorylation of site 1 and not site 2. Dent *et al.*⁴⁷⁵ proposed that the interaction of insulin with its membrane-bound receptor activated its tyrosine protein kinase, leading to the activation of a protein serine/threonine kinase-kinase which, in turn, phosphorylates and makes active a kinase labeled as the insulin-stimulated protein kinase (ISPK). This ISPK then phosphorylates site 1 of the G subunit making the protein phosphatase 1-G complex more active. The active protein phosphatase 1 then dephosphorylates the phosphate residues off sites 3a, 3b, and 3c of glycogen synthase thus activating glycogen synthesis.

6.15.14.3.2 Inactivation of glycogen synthase kinase-3

Evidence has also accumulated indicating that insulin can induce inactivation of glycogen synthase kinase-3 in many different cells.^{476–478} The inactivation appears to be a phosphorylation catalyzed by protein kinase, also known as Akt/RAC,^{479–481} which is regulated by an activated phosphatidylinositol (PI)-3 kinase.⁴⁸²

Inactivation of the GSK-3 by protein kinase B is accompanied by phosphorylation of serine residue 9 of the GSK-3b and serine-21 of the GSK-3a isozymes *in vitro* or *in vivo*.⁴⁸¹ Although other protein kinases are known to phosphorylate and inactivate GSK-3,⁴⁸¹ it is believed that the activation by insulin of glycogen synthase via the inactivation of GSK-3 is due to the activation of protein kinase B. Insulin stimulation of L6 myotube cells caused a 10-fold increase in protein kinase B activity and decreased GSK-3 activity by 40–50%. The half time for activation of protein kinase B was 1 min, slightly faster than inhibition of GSK-3 that was 2 min; the inhibition of GSK-3 was reversed by incubation with protein phosphatase 2A.

Other studies⁴⁸³ on whole cells also indicate the importance of GSK-3 in regulating glycogen synthase activity; they made mutants of GSK-3 at serine-9 (Ser9Ala and Ser9Glu). These mutants could not be phosphorylated or their GSK-3 activity made inactive by phosphorylation. The wild type, normal, and mutant enzymes were expressed in 293 cells and their activity was determined. Cells expressing the S9A mutant, wild-type, and S9E mutant GSK-3 had 2.6-, 1.8-, and 2.0-fold higher GSK-3 activity, respectively, as compared with control cells. The higher activity of the S9A mutant suggested Ser9 as a key regulatory site for GSK-3 inactivation. However, substitution of glutamic acid for serine did not mimic the inactivation caused by the negative ion, phosphate, on the Ser9 residue. The effects of expressing the wild type and mutant GSK-3 mutants in the 293 cells on glycogen synthase – Glc-6-P/+ Glc-6-P activity ratio was analyzed. A 50% reduction in the activity ratio was seen for the cells having the S9A mutant while a 20–30% decrease was observed in cells having the wild type and S9E mutant. Thus, evidence was obtained that activation of GSK-3 is sufficient to inhibit glycogen synthase in intact cells and for supporting a physiological role for GSK-3 in regulating glycogen synthase.

Three lines of transgenic mice were generated where the rabbit skeletal muscle glycogen synthase was overexpressed in mouse skeletal muscle.⁴⁸⁴ The glycogen synthase expressed was the glycogen synthase sites 2 and 3a mutant (Ser7Ala and Ser640Ala) so that the overexpression of the synthase could not be inactivated by phosphorylation. The glycogen synthase activity was expressed by as much as 10-fold with concomitant increase of up to fivefold in glycogen content. The levels of UDP-Glc decreased markedly, consistent with the increase in glycogen synthase activity. Levels of the glycogen degradative enzyme, phosphorylase, increased up to threefold. But the activity of the insulin-sensitive glucose transporter either remained unchanged or decreased. Thus, increasing active glycogen synthase caused an increase in glycogen content, supporting the conclusion that activation of glycogen synthase contributes to the increased accumulation of glycogen observed in response to insulin.

In summary, insulin stimulates glycogen synthesis by activating protein phosphatase 1 and by inactivating glycogen synthase kinase-3. Thus the inactivated glycogen synthase becomes active and at the same time phosphorylase a and phosphorylase kinase are inactivated due to their dephosphorylation by activated protein phosphatase 1 resulting in an overall increase in glycogen synthesis and decreased glycogen degradation. A recent review by Philip Cohen²¹ summarizes in detail the events leading to activation of glycogen synthase by insulin in mammalian skeletal muscle. There is also a detailed review on GSK-3 and its involvement in the regulation of other physiological events besides regulation of glycogen synthase.⁴⁸⁵

References

1. W. R. Morrison; J. Karkalas, Starch. In *Methods in Plant Biochemistry*; P. M. Dey, Ed.; Academic Press: London, 1990; pp 323–352.
2. S. Hizukuri, Starch, Analytical Aspects. In *Carbohydrates in Food*; A. C. Eliasson, Ed.; Marcel Dekker, Inc.: New York, 1995; pp 347–429.
3. D. J. Manners, *Carbohydr. Polym.* **1991**, *16*, 37–82.
4. D. J. Manners. In *The Yeasts*; A. H. Rose, J. S. Harrison, Eds.; Academic Press: New York, 1971, Vol. 2, pp 419–439.
5. C. F. Cori; G. T. Cori, *J. Biol. Chem.* **1939**, *131*, 397–398.
6. L. F. Leloir; C. E. Cardini, *J. Am. Chem. Soc.* **1957**, *79*, 6340–6341.
7. J. Preiss; T. Romeo, Molecular Biology and Regulation of Bacterial Glycogen Biosynthesis In *Progress in Nucleic Acids Research and Molecular Biology*; K. Moldave, W. E. Cohn, Eds.; Academic Press: New York City, NY, 1994, Vol. 47, pp 299–329.
8. M. A. Ballicora; A. A. Iglesias; J. Preiss, *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 213–225.
9. M. A. Ballicora; A. A. Iglesias; J. Preiss, *Photosyn. Res.* **2004**, *79*, 1–24.

10. J. Preiss, Bacterial Glycogen Inclusions: Enzymology and Regulation of Synthesis In *Microbiology Monograph*; J. M. Shively, Ed.; Springer: Heidelberg, Germany, 2006; Vol. 1, pp 71–108.
11. J. Preiss, Glycogen: Biosynthesis and Regulation. In *EcoSal – Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 3rd ed.; A. Böck, R. Curtiss, III, J. B. Kaper, F. C. Neidhardt, T. Nyström, K. E. Rudd, C. L. Squires, Eds.; ASM Press: Washington, DC, 2009; Chapter 4.74, <http://www.ecosal.org>.
12. J. Preiss, Biochemistry and Molecular Biology of Starch Biosynthesis In *Starch: Chemistry and Technology*, 3rd ed., R. L. Whistler; J. BeMiller, Eds.; Elsevier, Inc.: Oxford: UK, 2009; pp 83–147.
13. J. Preiss, Glycogen Biosynthesis In *The Encyclopedia of Microbiology*, 3rd ed., M. Schaechter, Ed.; Oxford: Elsevier, UK, 2009; pp 145–158.
14. C. Martin; A. M. Smith, *Plant Cell* **1995**, *7*, 971–985.
15. J. Preiss, Modulation of Starch Synthesis In *Engineering Improved Carbon and Nitrogen Resource Use Efficiency in Higher Plant*; C. Foyer, P. Quick, Eds.; Taylor and Francis, Publishers: London and Washington, DC, 1997; pp 81–104.
16. J. Preiss; M. N. Sivak, Starch Synthesis in Sinks and Sources In *Photoassimilate Distribution in Plants and Crops: Source-Sink Relationships*; E Zamski, A. A. Schaffer, Eds.; Marcel Dekker Inc.: New York, 1996; pp 63–96.
17. A. M. Smith; M. Stitt, *Plant Cell Environ.* **2007**, *39*, 1126–1149.
18. S. C. Zeeman; S. M. Smith; A. M. Smith, *Biochem. J.* **2007**, *401*, 13–28.
19. A. M. Smith; K. Denyer; C. Martin, *Plant Physiol.* **1995**, *107*, 673–677.
20. P. Cohen, Muscle Glycogen Synthas. In *The Enzyme*, 3rd ed.; P. D. Boyer, E. G. Krebs, Eds.; Academic Press: San Diego, CA, 1986; Vol. XVII, Part A, pp 461–497.
21. P. Cohen. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 867–873.
22. P. J. Roach, Liver Glycogen Synthas. In *The Enzyme*, 3rd ed.; P. D. Boyer, E. G. Krebs, Eds.; Academic Press: San Diego, CA, 1986; Vol. XVII, Part A, pp 499–539.
23. P. J. Roach, *FASEB J.* **1990**, *4*, 2961–2968.
24. E. A. Dawes; P. J. Senior, *Adv. Microbiol. Physiol.* **1973**, *10*, 135–266.
25. J. Preiss, *Annu. Rev. Microbiol.* **1984**, *38*, 419–458.
26. J. Preiss, Chemistry and Metabolism of Intracellular Reserve. In *Bacteria in Nature*; E. Leadbetter, J. S. Poindexter, Eds.; Plenum Publishing Corp.: New York, 1989; Vol. 3 (A treatise on the interactions of bacteria and their habitats), pp 189–258.
27. B. M. Mackey; J. G. Morris, *J. Gen Microbiol.* **1971**, *66*, 1–13.
28. I. G. Burleigh; E. A. Dawes, *Biochem. J.* **1967**, *102*, 236–250.
29. J. Sachs. In *Lectures of the Physiology of Plants* (translated by H. M. Ward), Clarendon Press: Oxford, 1887; pp 304–325.
30. G. Edwards; D. A. Walker, Plastids and Intracellular Transport. In *C3,C4: Mechanisms of Cellular and Environmental Regulation, Photosynthesis*. University of California Press: Berkeley, CA, 1983; pp 204–207.
31. I. P. Badenhuizen. In *The Biogenesis of Starch Granules in Higher Plants*; Appleton-Century Crofts: New York, 1999.
32. E. Recondo; L. F. Leloir, *Biochem. Biophys. Res. Commun.* **1961**, *6*, 85–88.
33. N. Sigal; J. Cattaneo; I. H. Segel, *Arch. Biochem. Biophys.* **1964**, *108*, 440–451.
34. L. Shen; J. Preiss, *Biochem. Biophys. Res. Commun.* **1964**, *17*, 424–429.
35. L. Shen; J. Preiss, *J. Biol. Chem.* **1965**, *240*, 2334–2340.
36. E. Greenberg; J. Preiss, *J. Biol. Chem.* **1964**, *239*, 4314–4315.
37. L. Shen; H. P. Ghosh; E. Greenberg; J. Preiss, *Biochim. Biophys. Acta* **1964**, *89*, 370–372.
38. N. Sigal; J. Cattaneo; J. P. Chambost; A. Favard, *Biochem. Biophys. Res. Commun.* **1965**, *20*, 616–620.
39. C. Boyer; J. Preiss, *Biochemistry* **1977**, *16*, 3693–3699.
40. L. P. T. M. Zevenhuizen, *Biochim. Biophys. Acta* **1964**, *81*, 608–611.
41. K. E. Steiner; J. Preiss, *J. Bacteriol.* **1977**, *129*, 246–253.
42. G. L. Walker; J. E. Builder, *Eur. J. Biochem.* **1971**, *20*, 14–21.
43. P. A. Baecker; C. E. Furlong; J. Preiss, *J. Biol. Chem.* **1983**, *258*, 5084–5088.
44. D. Homerova; J. Kormanec, *Biochim. Biophys. Acta* **1994**, *1200*, 334–336.
45. J. A. K. W. Kiel; J. M. Boels; G. Beldman; G. Venema, *Mol. Gen. Genet.* **1991**, *230*, 136–144.
46. H. Takata; T. Takaha; T. Kuriki; S. Okada; M. Takagi; I. Imanaka, *Appl. Environ. Microbiol.* **1994**, *60*, 3096–3104.
47. J. A. K. W. Kiel; J. M. Boels; G. Beldman; G. Venema, *DNA Seq* **1992**, *3*, 221–232.
48. J. A. K. W. Kiel; H. S. A. Elgersma; G. Beldman; J. P. M. J. Vossen; G. Venema, *Gene* **1989**, *78*, 9–17.
49. J. A. K. W. Kiel; J. M. Boels; G. Beldman; G. Venema, *Gene* **1990**, *89*, 77–84.
50. S. Hestrin, Synthesis of Polymeric Homosaccharides In *The Bacteri*; I. Gunsalus, R. Y. Stanier, Eds.; Academic Press: New York, 1960; Vol. 3, pp 373–388.
51. E. J. Hehre, *Adv. Enzymol.* **1951**, *11*, 297–337.
52. G. Okada; E. J. Hehre, *J. Biol. Chem.* **1974**, *249*, 126–135.
53. T. N. Palmer; G. Wöber; W. J. Whelan, *Eur. J. Biochem.* **1973**, *39*, 601–612.
54. J. Monod; A. M. Torriani, *C. R. Acad. Sci. (Paris)* **1948**, *227*, 240–242.
55. G. J. Walker, *Biochem. J.* **1966**, *101*, 861–872.
56. S. Lacks, *Genetics* **1968**, *60*, 685–706.
57. G. Wober, *Hoppe-Seyler's Z. Physiol. Chem.* **1973**, *354*, 75–82.
58. J. Chao; C. J. Weathersbee, *J. Bacteriol.* **1974**, *117*, 181–188.
59. G. S. Chen; I. H. Segel, *Arch. Biochem. Biophys.* **1968**, *127*, 164–174.
60. G. S. Chen; I. H. Segel, *Arch. Biochem. Biophys.* **1968**, *127*, 175–186.
61. R. L. Khandelwal; T. N. Spearman; I. R. Hamilton, *Arch. Biochem. Biophys.* **1973**, *154*, 295–305.
62. M. Schwartz, *J. Bacteriol.* **1966**, *92*, 1083–1089.
63. J. Preiss; T. Romeo, *Adv. Microb. Physiol.* **1989**, *30*, 183–238.
64. J. H. Ko; C. H. Kim; D. S. Lee; Y. S. Kim, *Biochem. J.* **1996**, *319*, 977–983.
65. S. G. Ball; M. K. Morell, *Annu. Rev. Plant Biol.* **2003**, *54*, 207–233.
66. K. Denyer; A. M. Smith, *Planta* **1992**, *186*, 609–617.

67. K. Denyer; C. Sidebottom; C. M. Hylton; A. M. Smith, *Plant J.* **1993**, *4*, 191–198.
68. K. Denyer; C. M. Hylton; C. F. Jenner; A. M. Smith, *Planta* **1995**, *196*, 256–265.
69. C. M. Hylton; K. Denyer; P. L. Keeling; M.-T. Chang; A. M. Smith, *Planta* **1996**, *198*, 230–237.
70. K. Mizuno; K. Kimura; Y. Arai; T. Kawasaki; H. Shimada; T. Baba, *J. Biochem.* **1992**, *112*, 643–651.
71. M. Bhattacharyya; C. Martin; A. M. Smith, *Plant Mol. Biol.* **1993**, *22*, 525–531.
72. H. P. Guan; J. Preiss, *Plant Physiol.* **1993**, *102*, 1269–1273.
73. H. P. Guan; T. Baba; J. Preiss, *Plant Physiol.* **1994**, *104*, 1449–1453.
74. H. P. Guan; T. Baba; J. Preiss, *Cell. Mol. Biol.* **1994**, *40*, 981–988.
75. H. Guan; T. Kuriki; M. Sivak; J. Preiss, *Proc. Nat. Acad. Sci. U.S.A.* **1995**, *92*, 964–967.
76. K. Burton; R. A. Bewley; J. D. Smith; A. M. Bhattacharyya; H. Tatge; S. Ring; V. Bull; D. O. Hamilton; C. Martin, *Plant J.* **1995**, *7*, 3–15.
77. O. E. Nelson; P. S. Chourey; M. T. Chang, *Plant Physiol.* **1978**, *62*, 383–386.
78. M. Shure; S. Wessler; N. Federoff, *Cell* **1983**, *35*, 225–233.
79. Y. Sano, *Theor. Appl. Genet.* **1984**, *68*, 467–473.
80. F. R. Van der Leij; R. F. G. Visser; A. S. Ponstein; E. Jacobsen; W. J. Feenstra, *Mol. Gen. Genet.* **1991**, *228*, 240–248.
81. R. F. G. Visser; I. Somhorst; G. J. Kuipers; N. J. Ruys; W. J. Feenstra; E. Jacobsen, *Mol. Gen. Genet.* **1991**, *225*, 289–296.
82. I. Dry; A. Smith; Edwards; M. Bhattacharyya; P. Dunn; C. Martin, *Plant J.* **1992**, *2*, 193–202.
83. B. Delrue; T. Fontaine; F. Routier; A. Decq; J.-M. Wieruszkeski; N. van den Koornhuys; M. L. Maddelein; B. Fournet; S. Ball, *J. Bacteriol.* **1992**, *174*, 3612–3620.
84. L. F. Leloir; M. A. R. deFekete; C. E. Cardini, *J. Biol. Chem.* **1961**, *236*, 636–641.
85. D. Pan; O. E. Nelson, *Plant Physiol.* **1984**, *74*, 324–328.
86. M. G. James; D. S. Robertson; A. M. Meyers, *Plant Cell* **1995**, *7*, 417–429.
87. Y. Nakamura; T. Umemoto; Y. Takahata; K. Komae; E. Amano; H. Satoh, *Physiol. Plant* **1996**, *97*, 491–498.
88. C. R. Krisman; R. Barengo, *Eur. J. Biochem.* **1975**, *52*, 117–123.
89. J. Lomako; W. M. Lomako; W. J. Whelan, *FASEB J.* **1988**, *2*, 3097–3103.
90. J. Pitcher; C. Smythe; P. Cohen, *Eur. J. Biochem.* **1987**, *169*, 497–502.
91. D. G. Campbell; P. Cohen, *Eur. J. Biochem.* **1989**, *185*, 119–125.
92. I. R. Rodriguez; W. J. Whelan, *Biochem. Biophys. Res. Commun.* **1985**, *132*, 829–836.
93. C. Smythe; F. B. Caudwell; M. Ferguson; P. Cohen, *EMBO J.* **1988**, *7*, 2681–2686.
94. J. Lomako; W. M. Lomako; W. J. Whelan; R. S. Dombro; J. T. Neary; M. D. Norenberg, *FASEB J.* **1993**, *7*, 1386–1393.
95. T. H. Haugen; A. Ishaque; A. K. Chatterjee; J. Preiss, *FEBS Lett.* **1974**, *42*, 205–208.
96. T. H. Haugen; A. Ishaque; J. Preiss, *J. Biol. Chem.* **1976**, *251*, 7880–7885.
97. C. E. Furlong; J. Preiss, *J. Biol. Chem.* **1969**, *244*, 2539–2548.
98. G. Ribereau-Gayon; J. Preiss, *Meth. Enzymol.* **1971**, *23*, 618–629.
99. J. Preiss; D. A. Walsh, *The Comparative Biochemistry of Glycogen and Starch Metabolism and Their Regulation In Biology of Carbohydrate*; V. Ginsburg, Ed.; John Wiley & Sons, Inc.: New York, 1981; Vol. 1, pp 199–314.
100. J. Preiss; K. Crawford; J. Downey; C. Lammel; E. Greenberg, *J. Bacteriol.* **1976**, *127*, 193–203.
101. T. W. Okita; R. L. Rodriguez; J. Preiss, *J. Biol. Chem.* **1981**, *256*, 6944–6952.
102. J. Preiss; H. P. Ghosh; J. Wittkop, *Regulation of the Biosynthesis of Starch in Spinach Leaf Chloroplasts In Biochemistry of Chloroplast*; T. W. Goodwin, Ed.; Academic Press: New York, 1967; Vol. 2, pp 131–153.
103. L. Copeland; J. Preiss, *Plant Physiol.* **1981**, *68*, 996–1002.
104. H. P. Ghosh; J. Preiss, *J. Biol. Chem.* **1966**, *241*, 4491–4505.
105. Y. Fu; M. A. Ballicora; J. Preiss, *Plant Physiol.* **1998**, *117*, 989–996.
106. M. A. Ballicora; Y. Fu; N. M. Nesbitt; J. Preiss, *Plant Physiol.* **1998**, *118*, 265–274.
107. J. Preiss; M. Morell; B. Bloom; V. L. Knowles; T. P. Lin, *Starch Synthesis and its Regulation In Progress in Photosynthesis Research*; J. Biggins, Ed.; Nijhoff: The Hague, 1987; Vol. III, pp 693–700.
108. M. K. Morell; M. Bloom; V. Knowles; J. Preiss, *Plant Physiol.* **1987**, *85*, 185–187.
109. M. Morell; M. Bloom; R. Larsen; T. W. Okita; J. Preiss, *Biochemistry and Molecular Biology of Starch Synthesis In Plant Gene Systems and Their Biolog*; J. L. Key, L. McIntosh, Eds.; Alan R. Liss, Inc: New York, 1987; pp 227–242.
110. T. W. Okita; P. A. Nakata; J. M. Anderson; J. Sowokinos; M. Morell; J. Preiss, *Plant Physiol.* **1990**, *93*, 785–790.
111. P. A. Nakata; T. W. Greene; J. M. Anderson; B. J. Smith-White; T. W. Okita; J. Preiss, *Plant Mol. Biol.* **1991**, *17*, 1089–1093.
112. B. J. Smith-White; J. Preiss, *J. Mol. Evol.* **1992**, *34*, 449–464.
113. J. Preiss; S. Danner; P. S. Summers; M. Morell; C. R. Barton; L. Yang; M. Nieder, *Plant Physiol.* **1990**, *92*, 881–885.
114. C. Barton; L. Yang; M. Galvin; C. Sengupta-Gopalan; T. Borelli, *Isolation of the SHRUNKEN-2 and BRITTLE-2 genes from Maize In Regulation of Carbon and Nitrogen Reduction and Utilization in Maiz*; J. C. Shannon, D. P. Knievel, C. D. Boyer, Eds.; American Society of Plant Physiologists: Rockville MD, 1986; pp 363–365.
115. J. M. Anderson; J. Hnilo; R. Larson; T. W. Okita; M. Morell; J. Preiss, *J. Biol. Chem.* **1989**, *264*, 12238–12242.
116. M. R. Paule; J. Preiss, *J. Biol. Chem.* **1971**, *246*, 4602–4609.
117. W. W. Cleland, *Steady State Kinetics In The Enzymes*, 3rd ed.; P. D. Boyer, Ed.; Academic Press: San Diego, CA, 1970; Vol. II, pp 1–65.
118. L. A. Kleczkowski; P. Villand; J. Preiss; O.-A. Olsen, *J. Biol. Chem.* **1993**, *268*, 6228–6233.
119. T. Haugen; J. Preiss, *J. Biol. Chem.* **1979**, *254*, 127–136.
120. M. L. DePamphilis; W. W. Cleland, *Biochemistry* **1973**, *122*, 3714–3724.
121. C. A. Janson; W. W. Cleland, *J. Biol. Chem.* **1974**, *249*, 2572–2574.
122. N. Lazdunski, *Curr. Top. Cell. Regul.* **1972**, *6*, 267–310.
123. L. Eidels; P. L. Edelman; J. Preiss, *Arch. Biochem. Biophys.* **1970**, *140*, 60–74.
124. A. D. Uttaro; R. A. Ugalde; J. Preiss; A. A. Iglesias, *Arch. Biochem. Biophys.* **1998**, *357*, 13–21.
125. E. Greenberg; J. E. Preiss; M. Van Boldrick; J. Preiss, *Arch. Biochem. Biophys.* **1983**, *220*, 594–604.
126. R. Y. Igarashi; C. R. Meyer, *Arch. Biochem. Biophys.* **2000**, *376*, 47–58.
127. S. G. Yung; J. Preiss, *J. Bacteriol.* **1981**, *147*, 101–109.

128. K. Matsuno; T. Blais; A. W. Serio; T. Conway; T. M. Henkin; A. L. Sonnenshein, *J. Bacteriol.* **1999**, *181*, 3382–3391.
129. H. Takata; T. Takaha; S. Okada; M. Takagi; T. Imanaka, *J. Bacteriol.* **1997**, *179*, 4689–4698.
130. Y. Y. Charnig; G. Kakefuda; A. A. Iglesias; W. J. Buikema; J. Preiss, *Plant Mol. Biol.* **1992**, *20*, 37–47.
131. A. A. Iglesias; G. Kakefuda; J. Preiss, *Plant Physiol.* **1991**, *97*, 1187–1195.
132. A. A. Iglesias; G. Kakefuda; J. Preiss, *J. Protein Chem.* **1992**, *11*, 119–128.
133. J. Preiss, Biosynthesis of Starch and its Regulation In *The Biochemistry of Plants: Vol. 14, Carbohydrates, Structure and Function*; J. Preiss, Ed.; Academic Press: New York, 1988; pp 184–249.
134. J. Preiss, Biology and Molecular Biology of Starch Biosynthesis and its Regulation In *Oxford Survey of Plant Molecular and Cellular Biology*; B. J. Mifflin, Ed.; Oxford University Press: Oxford, 1991; Vol. 7, pp 59–114.
135. M. N. Sivak; J. Preiss, Starch: Basic Science to Biotechnology In *Advances in Food and Nutrition Research*; S. L. Taylor, Ed.; Academic Press: San Diego, USA, 1998; Vol. 41, pp 1–199.
136. M. A. Ballicora; M. J. Laughlin; Y. Fu; T. W. Okita; G. F. Barry; J. Preiss, *Plant Physiol.* **1995**, *109*, 245–251.
137. A. A. Iglesias; G. F. Barry; C. Meyer; L. Bloksberg; P. A. Nakata; T. Greene; M. J. Laughlin; T. W. Okita; G. M. Kishore; J. Preiss, *J. Biol. Chem.* **1993**, *268*, 1081–1086.
138. D. F. Gomez-Casati; A. A. Iglesias, *Planta* **2002**, *214*, 428–434.
139. H. Rudi; D. N. P. Doan; O. A. Olsen, *FEBS Lett.* **1997**, *419*, 124–130.
140. C. Hylton; A. M. Smith, *Plant Physiol.* **1992**, *99*, 1626–1634.
141. W. C. Plaxton; J. Preiss, *Plant Physiol.* **1987**, *83*, 105–112.
142. H. Weber; U. Heim; L. Borisjuk; U. Wobus, *Planta* **1995**, *195*, 352–361.
143. L. A. Kleczkowski; P. Villand; E. Lüthi; O.-A. Olsen; J. Preiss, *Plant Physiol.* **1993**, *101*, 179–186.
144. J. Preiss; L. Shen; E. Greenberg; N. Gentner, *Biochem.* **1966**, *5*, 1833–1845.
145. P. Crevillén; M. A. Ballicora; A. Mérida; J. Preiss; J. Romero, *J. Biol. Chem.* **2003**, *278*, 28508–28515.
146. S. G. Yung; M. Paule; R. Beggs; E. Greenberg; J. Preiss, *Arch. Microbiol.* **1984**, *138*, 1–8.
147. D. Lapp; A. D. Elbein, *J. Bacteriol.* **1972**, *112*, 327–336.
148. J. Preiss; E. Greenberg, *J. Bacteriol.* **1981**, *147*, 711–719.
149. N. Gentner; J. Preiss, *J. Biol. Chem.* **1968**, *243*, 5882–5891.
150. N. Gentner; J. Preiss, *Biophys. Res. Commun.* **1967**, *27*, 417–423.
151. Y. Y. Charnig; A. A. Iglesias; J. Preiss, *J. Biol. Chem.* **1994**, *269*, 24107–24113.
152. J. Preiss; M. N. Sivak, Biosynthesis and Regulation of Plant Starch and Bacterial and Mammalian Glycogen Synthesis In *Comprehensive Natural Products Chemistry: Vol. 3, Carbohydrates and Their Derivatives Including Tannins, Cellulose and Related Lignin*; B. M. Pinto, Ed.; Pergamon Press: Oxford, UK, 1999; pp 441–495.
153. J. Preiss, Regulation of ADP-Glucose Pyrophosphorylase In *Advances in Enzymology and Related Areas of Molecular Biology*; A. Meister, Ed.; John Wiley and Sons, Inc.: New York, NY, 1978, Vol. 46, pp 317–381.
154. J. Preiss; C. Levi, Starch Biosynthesis and Degradation In *The Biochemistry of Plant*; J. Preiss, Ed.; Academic Press: New York, 1980; Vol. 3, pp 371–423.
155. H. Kacser; J. A. Burns, *Symp. Soc. Exp. Biol.* **1973**, *27*, 65–107.
156. H. Kacser, Control of Metabolism In *The Biochemistry of Plant*; D. D. Davies, Ed.; Academic Press: New York, 1987; Vol. 11, pp 39–67.
157. H. E. Neuhaus; A. L. Kruckeberg; R. Feil; M. Stitt, *Planta* **1989**, *178*, 110–122.
158. H. E. Neuhaus; M. Stitt, *Planta* **1990**, *182*, 445–454.
159. S. Govons; N. Gentner; E. Greenberg; J. Preiss, *J. Biol. Chem.* **1973**, *248*, 1731–1740.
160. J. Preiss; E. Greenberg; A. Sabraw, *J. Biol. Chem.* **1975**, *250*, 7631–7638.
161. J. Preiss; C. Lammel; E. Greenberg, *Arch. Biochem. Biophys.* **1976**, *174*, 105–119.
162. W. K. Kappel; J. Preiss, *Arch. Biochem. Biophys.* **1981**, *209*, 15–28.
163. P. Leung; Y. M. Lee; E. Greenberg; K. Esch; S. Boylan; J. Preiss, *J. Bacteriol.* **1986**, *167*, 82–88.
164. P. Ghosh; C. Meyer; E. Remy; D. Peterson; J. Preiss, *Arch. Biochem. Biophys.* **1992**, *296*, 122–128.
165. C. R. Meyer; P. Ghosh; E. Remy; J. Preiss, *J. Bacteriol.* **1992**, *174*, 4509–4512.
166. C. R. Meyer; J. Yirsa; B. Gott; J. Preiss, *Arch. Biochem. Biophys.* **1998**, *352*, 247–254.
167. C. R. Meyer; J. A. Bork; S. Nadler; J. Yirsa; J. Preiss, *Arch. Biochem. Biophys.* **1998**, *353*, 152–159.
168. C. R. Meyer; P. Ghosh; S. Nadler; J. Preiss, *Arch. Biochem. Biophys.* **1993**, *302*, 64–71.
169. A. Kumar; P. Ghosh; Y. M. Lee; M. A. Hill; J. Preiss, *J. Biol. Chem.* **1989**, *264*, 10464–10471.
170. S. Ball; T. Marianne; L. Dirick; M. Fresnoy; B. Delrue; A. Decq, *Planta* **1991**, *185*, 17–26.
171. A. A. Iglesias; Y. Y. Charnig; S. Ball; J. Preiss, *J. Plant Physiol.* **1994**, *104*, 1287–1294.
172. M. J. Giroux; J. Shaw; G. Barry; B. J. Cobb; T. Greene; T. Okita; L. C. Hannah, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5824–5829.
173. D. M. Stark; K. P. Timmerman; G. F. Barry; J. Preiss; G. M. Kishore, *Science* **1992**, *258*, 287–292.
174. E. D. Smidansky; M. Clancydagger; F. D. Meyer; S. P. Lanning; N. K. Blake; L. E. Talbert; M. J. Giroux, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1724–1729.
175. E. D. Smidansky; J. M. Martin; L. C. Hannah; A. M. Fischer; M. J. Giroux, *Planta* **2003**, *216*, 656–664.
176. I. Uzoma; D. Arias-Garzon; S. Lawrence; R. Sayre, *Plant Biotech. J.* **2006**, *4*, 453–465.
177. Y. Fu; M. A. Ballicora; J. F. Leykam; J. Preiss, *J. Biol. Chem.* **1998**, *273*, 25045–25052.
178. M. A. Ballicora; Y. Fu; J. B. Frueauf; J. Preiss, *Biochem. Biophys. Res. Commun.* **1999**, *257*, 782–786.
179. M. A. Ballicora; J. B. Frueauf; Y. Fu; P. Schürmann; J. Preiss, *J. Biol. Chem.* **2000**, *275*, 1315–1320.
180. W. T. Kim; V. R. Francheschi; T. W. Okita; N. L. Robinson; M. Morell; J. Preiss, *Plant Physiol.* **1989**, *91*, 217–220.
181. J. P. Jacquot; J. M. Lancelin; Y. Meyer, *New Phytol.* **1997**, *136*, 543–570.
182. Y. Balmer; W. H. Vensel; N. Cai; W. Manieri; P. Schurmann; W. J. Hurkman; B. B. Buchanan, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2988–2993.
183. A. Tiessen; J. H. Hendriks; M. Stitt; A. Branscheid; Y. Gibon; E. M. Farre; P. Geigenberger, *Plant Cell* **2002**, *14*, 2191–2213.
184. Y. M. Lee; J. Preiss, *J. Biol. Chem.* **1986**, *261*, 1058–1064.

185. Y. M. Lee; S. Mukerhjee; J. Preiss, *Arch. Biochem. Biophys.* **1986**, *244*, 585–595.
186. A. Sali; T. Blundell, *J. Mol. Biol.* **1993**, *234*, 779–815.
187. R. Sanchez; A. Sali, Comparative Protein Structure Modeling: Introduction and Practical Examples with MODELLER In *Protein Structure Prediction, Methods and Protocol*; D. M. Webster, Ed.; Humana press: Totowa, NJ, 2000; pp 97–129.
188. M. A. Hill; K. Kaufmann; J. Otero; J. Preiss, *J. Biol. Chem.* **1991**, *266*, 12455–12460.
189. T. F. Parsons; J. Preiss, *J. Biol. Chem.* **1978**, *253*, 6197–6202.
190. C. E. Larsen; J. Preiss, *Biochem.* **1986**, *25*, 4371–4376.
191. C. E. Larsen; Y. M. Lee; J. Preiss, *J. Biol. Chem.* **1986**, *261*, 15402–15409.
192. C. M. Bejar; M. A. Ballicora; J. Preiss, *J. Biol. Chem.* **2006**, *281*, 40473–40484.
193. Y.-Y. Charnng; G. Kakefuda; A. A. Iglesias; W. J. Buikema; J. Preiss, *Plant Mol. Biol.* **1992**, *20*, 37–47.
194. G. Kakefuda; Y. Y. Charnng; A. A. Iglesias; L. McIntosh; J. Preiss, *Plant Physiol.* **1992**, *99*, 344–347.
195. C. Ainsworth; J. Clark; J. Balsdon, *Plant Mol. Biol.* **1993**, *22*, 67–82.
196. P. Leung; J. Preiss, *J. Bacteriol.* **1987**, *169*, 4355–4360.
197. A. M. Mulichak; E. Skrzypczak-Jankum; T. J. Rydel; A. Tulinsky; J. Preiss, *J. Biol. Chem.* **1988**, *263*, 17237–17238.
198. Y. Fu; M. A. Ballicora; J. Preiss, *Plant Physiol.* **1998**, *117*, 989–996.
199. M. Morell; M. Bloom; J. Preiss, *J. Biol. Chem.* **1988**, *263*, 633–637.
200. K. L. Ball; J. Preiss, *J. Biol. Chem.* **1994**, *269*, 24706–24711.
201. T. Thorbjørnsen; P. Villand; L. A. Kleczkowski; O.-A. Olsen, *Biochem. J.* **1996**, *313*, 149–154.
202. J. M. Bae; J. Giroux; L. C. Hannah, *Maydica* **1990**, *35*, 317–322.
203. M. R. Bhavé; S. Lawrence; C. Barton; L. C. Hannah, *Plant Cell* **1990**, *2*, 581–588.
204. P. Villand; O. A. Olsen; A. Killian; L. A. Kleczkowski, *Plant Physiol.* **1992**, *100*, 1617–1618.
205. M. R. Olive; R. J. Ellis; W. W. Schuch, *Plant Mol. Biol.* **1989**, *12*, 525–538.
206. K. L. Ball; J. Preiss, *J. Prot. Chem.* **1992**, *11*, 231–238.
207. T.-P. Lin; T. Caspar; C. R. Somerville; J. Preiss, *Plant Physiol.* **1988**, *88*, 1175–1181.
208. T. Ventriglia; M. A. Ballicora; M. Teresa Ruíz; F. Ribeiro-Pedro; F. Valverde; J. Preiss; J. M. Romero, *Plant Physiol.* **2008**, *148*, 65–76.
209. M. A. Ballicora; J. R. Dubay; C. H. Devillers; J. Preiss, *J. Biol. Chem.* **2005**, *280*, 10189–10195.
210. J. B. Freuauf; M. A. Ballicora; J. Preiss, *Plant J.* **2003**, *33*, 503–511.
211. J. B. Freuauf; M. A. Ballicora; J. Preiss, *J. Biol. Chem.* **2001**, *276*, 46319–46325.
212. W. Blankenfeldt; M. Asuncion; J. S. Lam; J. H. Naismith, *EMBO J.* **2000**, *19*, 6652–6663.
213. K. Brown; F. Pompeo; S. Dixon; D. Mengin-Lecreux; C. Cambillau; Y. Bourne, *EMBO J.* **1999**, *18*, 4096–4107.
214. J. Sivaraman; V. Sauve; A. Matte; M. Cygler, *J. Biol. Chem.* **2002**, *277*, 44214–44219.
215. P. Crevillén; T. Ventriglia; F. Pinto; A. Orea; A. Mérida; J. M. Romero, *J. Biol. Chem.* **2005**, *280*, 8143–8149.
216. K. Denyer; F. Dunlap; T. Thorbjørnsen; P. Keeling; A. M. Smith, *Plant Physiol.* **1996**, *112*, 779–785.
217. T. Thorbjørnsen; P. Villand; K. Denyer; O. A. Olsen; A. Smith, *Plant J.* **1996**, *10*, 243–250.
218. P. E. Johnson; N. J. Patron; A. R. Bottrill; J. R. Dinges; B. F. Fahy; M. L. Parker; D. N. Waite; K. Denyer, *Plant Physiol.* **2003**, *131*, 684–696.
219. I. J. Tetlow; E. J. Davies; K. A. Vardy; C. G. Bowsher; M. M. Burrell; M. J. Emes, *J. Exp. Bot.* **2003**, *54*, 715–725.
220. V. K. Sikka; S. B. Choi; I. H. Kavakli; C. Sakulsingharoj; S. Gupta; H. Ito; T. W. Okita, *Plant Sci.* **2001**, *161*, 461–468.
221. T. Möhlmann; T. Tjaden; G. Henrichs; W. P. Quick; R. Häusler; H. E. Neuhaus, *Biochem. J.* **1997**, *324*, 503–509.
222. S. Kirchberger; M. Leroch; M. A. Huynen; M. Wahl; H. E. Neuhaus; J. Tjaden, *J. Biol. Chem.* **2007**, *282*, 22481–22491.
223. N. J. Patron; B. Greber; B. F. Fahy; D. A. Laurie; M. L. Parker; K. Denyer, *Plant Physiol.* **2004**, *135*, 2088–2097.
224. C. G. Bowsher; E. F. Scrase-Field; S. Esposito; M. J. Emes; I. J. Tetlow, *J. Exp. Bot.* **2007**, *58*, 1321–1332.
225. S.-K. Lee; S.-K. Hwang; M. Hahn; J.-S. Eom; H. G. Kang; Y. Hahn; S.-B. Choi; M.-H. Cho; S. H. Bhoo; G. An; T.-R. Hahn; T. W. Okita; J.-S. Jeon, *Plant Mol. Biol.* **2007**, *65*, 531–546.
226. X. Jin; M. A. Ballicora; J. Preiss; J. H. Geiger, *EMBO J.* **2005**, *24*, 694–704.
227. J. Fox; K. Kawaguchi; E. Greenberg; J. Preiss, *Biochemistry* **1976**, *15*, 849–856.
228. J. Cattaneo; J. P. Chambost; N. Creuzat-Sigal, *Arch. Biochem. Biophys.* **1978**, *190*, 85–96.
229. E. Greenberg; J. Preiss, *J. Biol. Chem.* **1965**, *240*, 2341–2348.
230. J. Preiss; E. Greenberg, *Biochemistry* **1965**, *4*, 2328–2334.
231. D. N. Dietzler; J. L. Strominger, *J. Bacteriol.* **1973**, *113*, 946–952.
232. A. D. Elbein; M. Mitchell, *J. Bacteriol.* **1973**, *113*, 863–873.
233. E. Holmes; J. Preiss, *Arch. Biochem. Biophys.* **1979**, *196*, 436–448.
234. K. Furakawa; M. Tagaya; M. Inouye; J. Preiss; T. Fukui, *J. Biol. Chem.* **1990**, *265*, 2086–2090.
235. M. Tagaya; K. Nakano; T. Fukui, *J. Biol. Chem.* **1985**, *260*, 6670–6676.
236. A. M. Mahrenholz; Y. Wang; P. J. Roach, *J. Biol. Chem.* **1988**, *263*, 10561–10567.
237. J. Marshall; C. Sidebottom; M. Debet; C. Martin; A. Smith; A. Edwards, *Plant Cell* **1996**, *8*, 1121–1135.
238. K. Furakawa; M. Tagaya; K. Tanazawa; T. Fukui, *J. Biol. Chem.* **1993**, *268*, 23837–23842.
239. A. Kumar; C. E. Larsen; J. Preiss, *J. Biol. Chem.* **1986**, *261*, 16256–16259.
240. K. Furakawa; M. Tagaya; K. Tanazawa; T. Fukui, *J. Biol. Chem.* **1994**, *269*, 868–871.
241. E. Holmes; J. Preiss, *Arch. Biochem. Biophys.* **1982**, *216*, 736–740.
242. A. Yep; M. A. Ballicora; M. N. Sivak; J. Preiss, *J. Biol. Chem.* **2004**, *279*, 8359–8367.
243. B. Henrissat; P. M. Coutinho; G. J. Davies, *Plant Mol. Biol.* **2001**, *47*, 55–72.
244. A. Yep; M. A. Ballicora; J. Preiss, *Biochem. Biophys. Res. Commun.* **2004**, *316*, 960–966.
245. A. Yep; M. A. Ballicora; J. Preiss, *Arch. Biochem. Biophys.* **2006**, *453*, 188–196.
246. A. Buschiazzi; J. E. Ugalde; M. E. Guerin; W. Shepard; R. A. Ugalde; P. M. Alzari, *EMBO J.* **2004**, *23*, 3196–3205.
247. J. Preiss, Biosynthesis of Starch and its Regulation In *The Biochemistry of Plants: Vol. 14, Carbohydrates, Structure and Function*; J. Preiss, Ed.; Academic Press: New York, 1988; pp 184–249.

248. J. Preiss. In *Oxford Survey of Plant Molecular and Cellular Biology*; B. J. Mifflin, Ed.; Oxford University Press: Oxford, 1991, Vol. 7, pp 59–114.
249. C. Harn; M. Knight; A. Ramakrishnan; H. Guan; P. L. Keeling; B. P. Wasserman, *Plant Mol. Biol.* **1998**, *37*, 639–649.
250. A. Edwards; J. P. Vincken; L. C. J. M. Suurs; R. G. F. Visser; S. Zeeman; A. Smith; C. Martin, *Plant Cell* **2002**, *14*, 1767–1785.
251. N. Fujita; T. Taira, *Planta* **1998**, *207*, 125–132.
252. P. L. Vrinten; T. Nakamura, *Plant Physiol.* **2000**, *122*, 255–264.
253. M. K. Morell; B. Kosar-Hashemi; M. Cmiel; M. S. Samuel; P. Chandler; S. Rahman; A. Buleon; I. L. Batey; Z. Li, *Plant J.* **2003**, *34*, 173–185.
254. J. Craig; J. R. Lloyd; K. Tomlinson; L. Barber; A. Edwards; T. L. Wang; C. Martin; C. L. Hedley; A. M. Smith, *Plant Cell* **1998**, *10*, 413–426.
255. J. S. Hawker; C. F. Jenner, *Australian J. Plant Physiol.* **1993**, *20*, 197–209.
256. K. Denyer; C. M. Hylton; C. F. Jenner; A. M. Smith, *Planta* **1995**, *196*, 256–265.
257. Z. Li; G. Mouille; B. Kosar-Hashemi; S. Rahman; B. Clarke; K. R. Gale; R. Appels; M. K. Morell, *Plant Physiol.* **2000**, *123*, 613–624.
258. C. Mu; C. Harn; Y.-T. Ko; G. W. Singletary; P. L. Keeling; B. P. Wassermann, *Plant J.* **1994**, *6*, 151–159.
259. M. Gao; J. Wanat; P. S. Stinard; M. G. James; A. M. Myers, *Plant Cell* **1998**, *10*, 399–412.
260. H. Cao; J. Imparl-Radosevich; H. Guan; P. L. Keeling; M. G. James; A. M. Meyers, *Plant Physiol.* **1999**, *120*, 205–216.
261. X. Zhang; C. Colleoni; V. Ratushna; M. Sirghie-Colleoni; M. G. James; A. M. Myers, *Plant Mol. Biol.* **2004**, *54*, 865–879.
262. A. Edwards; J. Marshall; C. Sidebottom; R. G. F. Visser; A. Smith; C. Martin, *Plant J.* **1995**, *8*, 283–294.
263. J. Marshall; C. Sidebottom; M. Debet; C. Martin; A. M. Smith; A. Edwards, *Plant Cell* **1996**, *8*, 1121–1135.
264. J. Kossmann; G. J. W. Abel; F. Springer; J. R. Lloyd; L. Willmitzer, *Planta* **1999**, *208*, 503–511.
265. X. Zhang; A. M. Myers; M. G. James, *Plant Physiol.* **2005**, *138*, 663–674.
266. T. Baba; M. Nishihara; K. Mizuno; T. Kawasaki; H. Shimada; E. Kobayashi; S. Ohnishi; K. Tanaka; Y. Arai, *Plant Physiol.* **1993**, *103*, 565–573.
267. W. Dian; H. Jiang; P. Wu, *J. Exp. Bot.* **2005**, *56*, 623–632.
268. T. Fontaine; C. D'Hulst; M. L. Maddelein; F. Routier; T. M. Pepin; A. Decq; J.-M. Wieruszkeski; B. Delrue; N. Van den Koornhuys; J.-P. Bossu; B. Fournet; S. Ball, *J. Biol. Chem.* **1993**, *268*, 16223–16230.
269. D. Delvalle; S. Dumez; F. Wattedled; I. Roldan; V. Planchot; P. Berbezzy; P. Colonna; D. Vyas; M. Chatterjee; S. Ball; A. Merida; C. D'Hulst, *Plant J.* **2005**, *43*, 398–412.
270. P. D. Commuri; P. L. Keeling, *Plant J.* **2001**, *25*, 475–486.
271. C. Pollock; J. Preiss, *Arch. Biochem. Biophys.* **1980**, *204*, 578–588.
272. F. D. MacDonald; J. Preiss, *Plant Physiol.* **1985**, *78*, 849–852.
273. J. M. Imparl-Radosevich; P. Li; L. Zhang; A. L. McKean; P. L. Keeling; H. Guan, *Arch. Biochem. Biophys.* **1998**, *353*, 64–72.
274. P. C. Mangelsdorf, *Genetics* **1947**, *32*, 448–458.
275. Y.-J. Wang; P. White; L. Pollak; J.-L. Jane, *Cereal Chem.* **1993**, *70*, 521–529.
276. Y.-J. Wang; P. White; L. Pollak; J.-L. Jane, *Cereal Chem.* **1993**, *70*, 171–179.
277. O. E. Nelson; D. Pan, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1995**, *46*, 475–496.
278. J. C. Shannon; D. L. Garwood, Genetics and Physiology of Starch Development In *Starch: Chemistry and Technology*; R. L. Whistler, J. N. BeMiller, E. F. Paschall, Eds.; Academic Press: Orlando, FL, 1984; pp 25–86.
279. C. D. Boyer; J. Preiss, *Plant Physiol.* **1981**, *67*, 1141–1145.
280. J. L. Ozbun; J. S. Hawker; J. Preiss, *Plant Physiol.* **1971**, *48*, 765–769.
281. G. J. W. Abel; F. Springer; L. Willmitzer; J. Kossmann, *Plant J.* **1996**, *10*, 981–991.
282. T. L. Wang; A. Hadavizideh; A. Harwood; T. J. Welham; W. J. Harwood; R. Faulks; C. L. Hedley, *Plant Breed.* **1990**, *105*, 311–320.
283. T. Umemoto; M. Yano; H. Satoh; A. Shomura; Y. Nakamura, *Theor. Appl. Genet.* **2002**, *104*, 1–8.
284. M. Yamamori; S. Fujita; K. Hayakawa; J. Masuki; T. Yasui, *Theor. Appl. Genet.* **2000**, *101*, 21–29.
285. W. Dian; H. Jiang; P. Wu, *J. Exp. Bot.* **2005**, *56*, 623–632.
286. I. Roldan; F. Wattedled; M. M. Lucas; D. Delvalle; V. Planchot; S. Jimenez; R. Perez; S. Ball; C. D'Hulst; A. Merida, *Plant J.* **2007**, *49*, 492–504.
287. T. Hirose; T. Terao, *Planta* **2004**, *220*, 9–16.
288. A. Edwards; D. C. Fulton; C. M. Hylton; S. A. Jobling; M. Gidley; U. Rössner; C. Martin; A. M. Smith, **1999**, *Plant J.* *17*, 251–261.
289. L. F. Leloir; M. A. R. deFekete; C. E. Cardini, *J. Biol. Chem.* **1961**, *236*, 636–641.
290. M. A. R. de Fekete; L. F. Leloir; C. E. Cardini, *Nature* **1960**, *187*, 918–919.
291. J. R. Lloyd; V. Landschütze; J. Kossmann, *Biochem. J.* **1999**, *338*, 515–521.
292. E. Recondo; L. F. Leloir, *Biochem. Biophys. Res. Commun.* **1961**, *6*, 85–88.
293. F. D. MacDonald; J. Preiss, *Plant Physiol.* **1983**, *73*, 175–178.
294. O. E. Nelson; P. S. Chourey; M. T. Chang, *Plant Physiol.* **1978**, *62*, 383.
295. C. Y. Tsai, *Biochem. Genet.* **1974**, *11*, 83–96.
296. N. Federoff; S. Wessler; M. Shure, *Cell* **1983**, *35*, 235–242.
297. M. Shure; S. Wessler; N. Federoff, *Cell* **1983**, *35*, 225–233.
298. S. R. Wessler; G. Baran; M. Varagona; S. L. Dellaporta, *EMBO J.* **1986**, *5*, 2427–2432.
299. R. F. Klösgen; A. Gierl; Z. Schwarz-Sommer; H. Saedler, *Mol. Gen. Genet.* **1986**, *203*, 237–244.
300. W. Rohde; D. Becker; F. Salamini, *Nucl. Acids Res.* **1988**, *16*, 7185–7186.
301. M. N. Sivak; M. Wagner; J. Preiss, *Plant Physiol.* **1993**, *103*, 1355–1359.
302. K. Denyer; C. Sidebottom; C. M. Hylton; A. M. Smith, *Plant J.* **1993**, *4*, 191–196.
303. K. Denyer; B. Clarke; C. Hylton; H. Tatge; A. M. Smith, *Plant J.* **1996**, *10*, 1135–1143.
304. K. Denyer; D. Waite; S. Motawia; B. L. Møller; A. M. Smith, *Biochem. J.* **1999**, *340*, 183–191.
305. K. Denyer; D. Waite; A. Edwards; C. Martin; A. M. Smith, *Biochem. J.* **1999**, *342*, 647–653.
306. S. C. Zeeman; S. S. Smith; A. M. Smith, *Plant Physiol.* **2002**, *128*, 1069–1076.

307. M. Van de Wal; C. D'Hulst; J. P. Vincken; A. Buléon; R. Visser; S. Ball, *J. Biol. Chem.* **1998**, *273*, 22232–22240.
308. B. Delrue; T. Fontaine; F. Routier; A. Decq; J.-M. Wieruszkeski; N. Van den Koornhuyse; M.-L. Maddelein; S. Ball, *J. Bacteriol.* **1992**, *174*, 3612–3620.
309. M.-L. Maddelein; N. Libessart; F. Bellanger; B. Delrue; C. D'Hulst; N. Van den Koornhuyse; T. Fontaine; J.-M. Wieruszkeski; A. Decq; S. Ball, *J. Biol. Chem.* **1994**, *269*, 25150–25157.
310. F. Watteblad; A. Buléon; B. Bouchet; J.-P. Ral; L. Liénard; D. Devallé; K. Binderup; D. Dauvillée; S. Ball; C. D'Hulst, *Eur. J. Biochem.* **2002**, *269*, 3810–3820.
311. N. Van den Koornhuyse; N. Libessart; B. Delrue; C. Zabawinski; A. Decq; A. Iglesias; A. Carton; J. Preiss; S. Ball, *J. Biol. Chem.* **1996**, *271*, 16281–16287.
312. N. Libessart; M. L. Maddelein; N. Van den Koornhuyse; A. Decq; B. Delrue; S. Ball, *Plant Cell* **1995**, *7*, 1117–1127.
313. B. R. Clarke; K. Denyer; C. F. Jenner; A. M. Smith, *Planta* **1999**, *209*, 324–329.
314. J. R. Lloyd; F. Springer; A. Buléon; B. Müller-Röber; L. Willmitzer; J. Kossmann, *Planta* **1999**, *209*, 230–238.
315. C. D. Boyer; J. Preiss, *Carbohydr. Res.* **1978**, *61*, 321–334.
316. S. Hizukuri; Y. Takeda; M. Yasuda; A. Suzuki, *Carbohydr. Res.* **1981**, *94*, 205–213.
317. C. D. Boyer; J. Preiss, *Biochem. Biophys. Res. Commun.* **1978**, *80*, 169–175.
318. J. S. Hawker; J. L. Ozbun; H. Ozaki; E. Greenberg; J. Preiss, *Arch. Biochem. Biophys.* **1974**, *160*, 530–551.
319. Y. Takeda; H. P. Guan; J. Preiss, *Carbohydr. Res.* **1993**, *240*, 253–263.
320. J. A. K. W. Kiel; J. M. Boels; G. Beldman; G. Venema, *Mol. Microbiol.* **1994**, *11*, 203–218.
321. E. Rumbak; D. E. Rawlings; G. G. Lindsay; D. R. Woods, *J. Bacteriol.* **1991**, *173*, 6732–6741.
322. H. Takata; T. Takaha; S. Okada; M. Takagi; T. Imanaka, *Appl. Environ. Microbiol.* **1994**, *60*, 3096–3104.
323. H. Takata; T. Takaha; S. Okada; S. Hizukuri; M. Takagi; T. Imanaka, *Carbohydr. Res.* **1996**, *295*, 91–101.
324. P. A. Baecker; E. Greenberg; J. Preiss, *J. Biol. Chem.* **1986**, *261*, 8738–8743.
325. J. Khoshnoodi; A. Blennow; B. Ek; L. Rask; H. Larsson, *Eur. J. Biochem.* **1996**, *242*, 148–155.
326. B. K. Singh; J. Preiss, *Plant Physiol.* **1985**, *78*, 849–852.
327. G. H. Vos-Scheperkeuter; J. G. De Wit; A. S. Ponstein; W. J. Feenstra; B. Witholt, *Plant Physiol.* **1989**, *90*, 75–84.
328. J. Kossmann; R. G. F. Visser; B. T. Müller-Röber; L. Willmitzer; U. Sonnewald, *Mol. Gen. Genet.* **1991**, *230*, 39–44.
329. P. Poulsen; J. D. Kreiberg, *Plant Physiol.* **1993**, *102*, 1053–1054.
330. C.-T. Larsson; P. Hofvander; J. Khoshnoodi; B. Ek; L. Rask; H. Larsson, *Plant Sci.* **1996**, *117*, 9–16.
331. S. A. Jobling; G. P. Schwall; R. J. Westcott; C. M. Sidebottom; M. Debet; M. J. Gidley; R. Jefcoat; R. Safford, *Plant J.* **1999**, *18*, 163–171.
332. U. Rydberg; L. Andersson; R. Andersson; P. Åman; H. Larsson, *Eur. J. Biochem.* **2001**, *268*, 6140–6145.
333. H. Yamanouchi; Y. Nakamura, *Plant Cell Physiol.* **1992**, *33*, 985–991.
334. M. K. Morell; A. Blennow; B. Kosar-Hashemi; M. S. Samuel, *Plant Physiol.* **1997**, *113*, 201–208.
335. K. F. McCue; W. J. Hurkman; C. K. Tanka; O. D. Anderson, *Plant Mol. Biol. Rep.* **2002**, *20*, 191a–191m.
336. C. Sun; P. Sathish; S. Ahlandberg; A. Dieber; C. Jansson, *New Phytology* **1997**, *137*, 215–222.
337. C. Sun; P. Sathish; S. Ahlandberg; A. Dieber; C. Jansson, *Plant Physiol.* **1998**, *118*, 37–49.
338. J. Preiss; C. D. Boyer, Evidence of Independent Genetic Control of the Multiple Forms of Maize Endosperm Branching Enzymes and Starch Syntheses In *Mechanisms of Saccharide Polymerization and Depolymerization*; J. J. Marshall, Ed.; Academic Press: New York, 1980; pp 161–174.
339. K. D. Hedman; C. D. Boyer, *Biochem. Genet.* **1982**, *20*, 483–492.
340. B. K. Singh; J. Preiss, *Plant Physiol.* **1985**, *78*, 849–852.
341. D. K. Fisher; C. D. Boyer; L. C. Hannah, *Plant Physiol.* **1993**, *102*, 1045–1046.
342. D. K. Fisher; M. Gao; K.-N. Kim; C. D. Boyer; M. J. Guiltinan, *Plant Physiol.* **1996**, *110*, 611–619.
343. J. Edwards; J. H. Green; T. aP Rees, *Phytochem.* **1988**, *27*, 1615–1620.
344. A. M. Smith, *Planta* **1988**, *175*, 270–279.
345. J. Carter; E. E. Smith, *Carbohydr. Res.* **1978**, *61*, 395–406.
346. K. Mizuno; T. Kawasaki; H. Shimada; H. Satoh; E. Kobayashi; S. Okamura; Y. Arai; T. Baba, *J. Biol. Chem.* **1993**, *268*, 19084–19091.
347. A. Regina; A. Bird; D. Topping; S. Bowden; J. Freeman; T. Barsby; B. Kosar-Hashemi; Z. Li; S. Rahman; M. Morell, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 3546–3551.
348. M. K. Bhattacharyya; A. M. Smith; T. H. Ellis; C. Hedley; C. Martin, *Cell* **1990**, *60*, 115–122.
349. P. A. Baecker; E. Greenberg; J. Preiss, *J. Biol. Chem.* **1986**, *261*, 8738–8743.
350. S. N. I. M. Salehuzzaman; E. Jacobsen; R. G. F. Visser, *Plant Mol. Biol.* **1992**, *20*, 809–819.
351. Y. Nakamura; H. Yamanouchi, *Plant Physiol.* **1992**, *99*, 1265–1266.
352. A. M. Kram; G. T. Oostergetel; E. F. J. Van Bruggen, *Plant Physiol.* **1993**, *101*, 237.
353. C. Mu-Forster; R. Huang; J. R. Powers; R. W. Harriman; M. Knight; G. W. Singletary; P. Keeling; B. P. Wasserman, *Plant Physiol.* **1996**, *111*, 821–829.
354. T. Romeo; A. Kumar; J. Preiss, *Gene* **1988**, *70*, 363–376.
355. T. Baba; K. Kimura; K. Mizuno; H. Etoh; Y. Ishida; O. Shida; Y. Arai, *Biochem. Biophys. Res. Commun.* **1991**, *181*, 87–94.
356. B. Svensson, *Plant Mol. Biol.* **1994**, *25*, 141–157.
357. H. M. Jespersen; E. A. Macgregor; B. Henrissat; M. R. Sierks; B. Svensson, *J. Protein Chem.* **1993**, *12*, 791–805.
358. T. Kuriki; H. Guan; M. Sivak; J. Preiss, *J. Prot. Chem.* **1996**, *15*, 305–313.
359. H. Cao; J. Preiss, *J. Prot. Chem.* **1996**, *15*, 291.
360. K. Funane; N. Libessart; D. Stewart; T. Michishita; J. Preiss, *J. Protein Chem.* **1998**, *17*, 579–590.
361. T. Kuriki; D. C. Stewart; J. Preiss, *J. Biol. Chem.* **1997**, *272*, 28999–29004.
362. K. Binderup; R. Mikkelsen; J. Preiss, *Arch. Biochem. Biophys.* **2000**, *377*, 366–371.
363. K. Binderup; R. Mikkelsen; J. Preiss, *Arch. Biochem. Biophys.* **2002**, *397*, 279–285.
364. C. H. Devillers; M. E. Piper; M. A. Ballicora; J. Preiss, *Arch. Biochem. Biophys.* **2003**, *418*, 34–38.
365. M. C. Abad; K. Binderup; J. Rios-Steiner; J. Preiss; J. H. Geiger, *J. Biol. Chem.* **2002**, *277*, 42164–42170.

366. S. Rahman; K.-S. Wong; J. L. Jane; A. M. Myers; M. G. James, *Plant Physiol.* **1998**, *117*, 425–435.
367. J. R. Dinges; C. Colleoni; A. M. Myers; M. G. James, *Plant Physiol.* **2001**, *125*, 1406–1418.
368. Y. Nakamura; A. Kubo; T. Shimamune; T. Matsuda; K. Harada; H. Satoh, *Plant J.* **1997**, *12*, 143–153.
369. S. Ball; H.-P. Guan; M. James; A. Myers; P. Keeling; G. Mouille; A. Buléon; P. Colonna; J. Preiss, *Cell* **1996**, *86*, 349–352.
370. S. C. Zeeman; T. Umemoto; W. L. Lue; P. Auyeung; C. Martin; A. M. Smith; J. Chen, *Plant Cell* **1998**, *10*, 1699–1711.
371. T.-P. Lin; J. Preiss, *Plant Physiol.* **1988**, *86*, 260–265.
372. C. Colleoni; D. Dauvillée; G. Mouille; A. Buléon; D. Gallant; B. Bouchet; M. Morell; M. Samuel; B. Delrue; C. d'Hulst; C. Bliard; J.-M. Nuzillard; S. Ball, *Plant Physiol.* **1999**, *120*, 993–1004.
373. T. Takaha; J. Critchley; S. Okada; S. M. Smith, *Planta* **1998**, *205*, 445–451.
374. J. Critchley; S. C. Zeeman; T. Takaha; A. M. Smith; S. M. Smith, *Plant J.* **2001**, *26*, 89–100.
375. D. Dauvillée; V. Chochois; M. Steup; S. Haebel; N. Eckermann; G. Ritte; J.-P. Ral; C. Colleoni; G. Hicks; F. Wattedled; P. Deschamps; C. d'Hulst; L. Liénard; L. Cournac; J.-L. Putaux; D. Dupeyre; S. G. Ball, *Plant J.* **2006**, *48*, 274–285.
376. J. Cattaneo; M. Damotte; N. Sigal; G. Sanchez-Medina; J. Puig, *Biochem. Biophys. Res. Commun.* **1969**, *34*, 694–701.
377. M. Latile-Damotte; C. Lares, *Mol. Gen. Genet.* **1977**, *150*, 325–329.
378. F. Yu; Y. Jen; E. Takeuchi; M. Inouye; H. Nakayama; M. Tagaya; T. Fukui, *J. Biol. Chem.* **1988**, *263*, 13706–13711.
379. C. Haziza; P. Stragier; J.-C. Patte, *EMBO J.* **1982**, *1*, 379–384.
380. D. Austin; T. J. Larson, *J. Bacteriol.* **1991**, *173*, 101–107.
381. Y. Kohara; K. Akiyama; K. Isono, *Cell* **1987**, *50*, 495–508.
382. K. E. Rudd; W. Miller; J. Ostell; D. A. Benson, *Nucleic Acids Res.* **1990**, *18*, 313–321.
383. T. Romeo; J. Black; J. Preiss, *Curr. Microbiol.* **1990**, *21*, 131–137.
384. H. Yang; M. Y. Liu; T. Romeo, *J. Bacteriol.* **1996**, *178*, 1012–1017.
385. R. Jeanningros; N. Creuzat-Sigal; C. Frixon; J. Cattaneo, *Biochim. Biophys. Acta* **1975**, *438*, 186–199.
386. F. Yu; Y. Jen; E. Takeuchi; M. Inouye; H. Nakayama; M. Tagaya; T. Fukui, *J. Biol. Chem.* **1988**, *263*, 13706–13711.
387. T. Romeo; J. Preiss, *J. Bacteriol.* **1989**, *171*, 2773–2782.
388. D. N. Dietzler; M. P. Leckie; J. L. Magnani; M. J. Sughrue; P. E. Bergstein; W. L. Steinheim, *J. Biol. Chem.* **1979**, *254*, 8308–8317.
389. D. N. Dietzler; M. P. Leckie; W. L. Steinheim; J. M. Ungar; S. E. Porter, *Biochem. Biophys. Res. Commun.* **1977**, *77*, 1468–1477.
390. M. P. Leckie; R. R. Ng; S. E. Porter; D. R. Compton; D. N. Dietzler, *J. Biol. Chem.* **1983**, *258*, 3813–3824.
391. J. Urbanowski; P. Leung; H. Weissbach; J. Preiss, *J. Biol. Chem.* **1983**, *258*, 2782–2784.
392. T. Romeo; J. Moore, *Nucleic Acids Res.* **1991**, *19*, 3452.
393. A. Munch-Petersen; H. M. Kalckar; E. Cutole; E. E. B. Smith, *Nature* **1953**, *172*, 1036–1037.
394. R. L. Turnquist; R. G. Hansen, Uridinc Diphosphoryl Glucose Pyrophosphorylase In *The Enzyme*, 3rd ed., P. D. Boyer, Ed.; Academic Press: New York, 1973; Vol. 8, pp 51–71.
395. J. K. Knop; R. G. Hansen, *J. Biol. Chem.* **1970**, *245*, 2499–2504.
396. P. J. Roach; D. R. Warren; D. E. Atkinson, *Biochemistry* **1975**, *14*, 5445–5450.
397. T. R. Soderling; J. P. Hickenbottom; E. Reiman; F. L. Hunkeler; D. A. Walsh; E. G. Krebs, *J. Biol. Chem.* **1970**, *245*, 6317–6328.
398. N. E. Brown; J. Larner, *Biochim. Biophys. Acta* **1971**, *242*, 69–80.
399. C. Nakai; J. A. Thomas, *J. Biol. Chem.* **1975**, *250*, 4081–4086.
400. D. C. Lin; H. L. Segal, *J. Bio. Chem.* **1973**, *248*, 7007–7011.
401. S. D. Killilea; W. J. Whelan, *Biochemistry* **1976**, *15*, 1349–1355.
402. R. E. Miller; E. A. Miller; B. Fredholm; J. B. Yellin; R. D. Eichner; S. E. Mayer; S. Steinberg, *Biochemistry* **1975**, *14*, 2481–2488.
403. M. F. Browner; K. Nakano; A. G. Bang; R. J. Fletterick, *J. Biol. Chem.* **1989**, *86*, 1443–1447.
404. W. Zhang; M. F. Browner; R. G. Fletterick; A. A. DePaoli-Roach; P. J. Roach, *FASEB J.* **1989**, *3*, 2532–2536.
405. G. Bai; Z. Zhang; R. Werner; F. Q. Nuttall; A. W. H. Tan; E. Y. C. Lee, *J. Biol. Chem.* **1990**, *265*, 7843–7848.
406. F. Q. Nuttall; M. C. Gannon; G. Bai; E. Y. C. Lee, *Arch. Biochem. Biophys.* **1994**, *311*, 443–449.
407. I. Farkas; T. A. Hardy; A. A. DePaoli-Roach; P. J. Roach, *J. Biol. Chem.* **1990**, *265*, 20879–20886.
408. I. Farkas; T. A. Hardy; M. G. Goeb; P. J. Roach, *J. Biol. Chem.* **1991**, *266*, 15602–15607.
409. W. Zhang; A. A. DePaoli-Roach; P. J. Roach, *Arch. Biochem. Biophys.* **1993**, *304*, 219–225.
410. A. V. Skurat; Y. Cao; P. J. Roach, *J. Biol. Chem.* **1993**, *268*, 14701–14707.
411. A. V. Skurat; H.-W. Peng; H.-Y. Chang; J. F. Cannon; P. J. Roach, *Arch. Biochem. Biophys.* **1996**, *328*, 283–288.
412. A. V. Skurat; Y. Wang; P. J. Roach, *J. Biol. Chem.* **1994**, *269*, 25534–25542.
413. C. R. Krisman, *Biochim. Biophys. Acta* **1962**, *65*, 307–315.
414. J. Larner, *J. Biol. Chem.* **1953**, *202*, 491–503.
415. W. Verhue; H. G. Hers, *Biochem. J.* **1966**, *99*, 222–227.
416. B. I. Brown; D. H. Brown, α -1,4-Glucan 6-Glycosyltransferase from Mammalian Muscle In *Methods in Enzymology*; E. F. Neufeld, V. Ginsburg, Eds.; Academic Press: New York, 1966; Vol. 8, pp 395–403.
417. D. H. Brown; B. I. Brown, *Biochim. Biophys. Acta* **1966**, *130*, 263–266.
418. W. B. Gibson; B. I. Brown; D. H. Brown, *Biochemistry* **1971**, *10*, 4253–4262.
419. B. Illingworth; D. H. Brown; C. F. Cori, *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *46*, 469–478.
420. B. I. Brown; D. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *56*, 725–729.
421. B. I. Brown; D. H. Brown, Glycogen-Storage Diseases: Types I, III, IV, V, VII and Unclassified Glycogenoses In *Carbohydrate Metabolism and Its Disorder*; F. Dickens, P. J. Randle, W. J. Whelan, Eds.; Academic Press: London, 1968; Vol. 2, pp 123–150.
422. V. J. Thon; C. Vigneron-Lesens; T. Marianne-Pepin; J. Montreuil; A. Decq; C. Rachez; S. G. Ball; J. F. Cannon, *J. Biol. Chem.* **1992**, *267*, 15224–15228.
423. V. J. Thon; M. Khalil; J. F. Cannon, *J. Biol. Chem.* **1993**, *268*, 7509–7513.
424. W. J. Whelan, *BioEssays* **1986**, *5*, 136–140.
425. C. Smythe; P. Cohen, *Eur. J. Biochem.* **1991**, *200*, 625–631.
426. H. G. Nimmo; C. G. Proud; P. Cohen, *Eur. J. Biochem.* **1976**, *68*, 21–30.
427. M. C. Gannon; F. Q. Nuttall, *Trends Glycosci. Glycotechnol.* **1996**, *8*, 183–194.
428. E. Viskupic; Y. Cao; W. Zhang; C. Cheng; A. A. Depaoli-Roach; P. Roach, *J. Biol. Chem.* **1992**, *267*, 25759–25763.

429. M. D. Alonzo; J. Lomako; W. M. Lomako; W. J. Whelan, *J. Biol. Chem.* **1995**, *270*, 15315–15319.
430. J. Lomako; W. M. Lomako; W. J. Whelan, *Carbohydr. Res.* **1992**, *227*, 331–338.
431. M. D. Alonzo; J. Lomako; W. M. Lomako; W. J. Whelan; J. Preiss, *FEBS Lett.* **1994**, *352*, 222–226.
432. Y. Cao; A. M. Mahrenholz; A. A. Depaoli-Roach; P. Roach, *J. Biol. Chem.* **1993**, *268*, 14687–14693.
433. M. D. Alonzo; J. Lomako; W. M. Lomako; W. J. Whelan, *FEBS Lett.* **1994**, *342*, 38–42.
434. E. Meezan; S. Ananth; S. Manzella; P. Campbell; S. Siegal; D. J. Dillion; L. Roden, *J. Biol. Chem.* **1994**, *269*, 11503–11508.
435. J. Lomako; W. M. Lomako; W. J. Whelan, *FEBS Lett.* **1990**, *264*, 13–16.
436. S. M. Manzella; L. Rodén; E. Meezan, *Glycobiol.* **1995**, *5*, 263–271.
437. Y. Cao; L. K. Steinrauf; P. Roach, *Arch. Biochem. Biophys.* **1995**, *319*, 293–298.
438. R. Barengo; M. Flawia; C. R. Krisman, *FEBS Lett.* **1975**, *53*, 274–278.
439. J. E. Ugalde; A. J. Parodi; R. A. Ugalde, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10659–10663.
440. C. Cheng; J. Mu; I. Farkas; D. Huang; M. G. Goebel; P. J. Roach, *Mol. Cell. Biol.* **1995**, *15*, 6632–6640.
441. C. H. Smith; N. E. Brown; J. Larner, *Biochim. Biophys. Acta* **1971**, *242*, 81–88.
442. C. Pictou; A. Aitken; T. Bilham; P. Cohen, *Eur. J. Biochem.* **1982**, *124*, 37–45.
443. C. Pictou; J. Woodgett; B. Hemmings; P. Cohen, *FEBS Lett.* **1982**, *150*, 191–196.
444. P. J. Roach; J. Larner, *J. Biol. Chem.* **1976**, *251*, 1920–1925.
445. P. Cohen; D. Yellowlees; A. Aitken; A. Donella-Deana; B. A. Hemmings; P. J. Parker, *Eur. J. Biochem.* **1982**, *124*, 21–35.
446. N. Embi; P. J. Parker; P. Cohen, *Eur. J. Biochem.* **1981**, *115*, 405–413.
447. V. S. Sheroain; J. D. Corbin; T. R. Soderling, *J. Biol. Chem.* **1985**, *260*, 1567–1572.
448. D. B. Rylatt; A. Aitken; T. Bilham; G. D. Condon; N. Embi; P. Cohen, *Eur. J. Biochem.* **1980**, *107*, 529–537.
449. C. Fiol; A. M. Mahrenholz; A. Wang; R. W. Roeske; P. J. Roach, *J. Biol. Chem.* **1987**, *262*, 14042–14048.
450. C. Fiol; A. Wang; R. W. Roeske; P. J. Roach, *J. Biol. Chem.* **1990**, *265*, 6061–6065.
451. H. Flotow; P. J. Roach, *J. Biol. Chem.* **1989**, *264*, 9126–9128.
452. H. Flotow; P. R. Graves; A. Wang; C. J. Fiol; R. W. Roeske; P. J. Roach, *J. Biol. Chem.* **1990**, *265*, 14264–14269.
453. Y. Wang; A. Bell; M. A. Hermodson; P. J. Roach, *J. Biol. Chem.* **1986**, *261*, 1609–1615.
454. K. P. Haung; A. Akatsuka; T. J. Singh; K. R. Blake, *J. Biol. Chem.* **1983**, *258*, 7094–7101.
455. Y. Wang; M. Camici; F. T. Lee; Z. Ahmad; A. A. Depaoli-Roach; P. J. Roach, *Biochim. Biophys. Acta* **1986**, *888*, 225–236.
456. Y. Wang; P. J. Roach, *J. Biol. Chem.* **1993**, *268*, 23876–23880.
457. P. J. Parker; F. B. Caudwell; P. Cohen, *Eur. J. Biochem.* **1983**, *130*, 227–234.
458. J. C. Lawrence; J. N. Zhang, *J. Biol. Chem.* **1994**, *269*, 11595–11600.
459. P. J. Parker; N. Embi; F. B. Caudwell; P. Cohen, *Eur. J. Biochem.* **1982**, *124*, 47–55.
460. L. Poulter; S. C. Ang; B. W. Gibson; D. H. Williams; C. F. B. Holmes; F. B. Caudwell; P. Cohen, *Eur. J. Biochem.* **1988**, *175*, 497–510.
461. V. S. Sheroain; H. Juhl; M. Bass; T. R. Soderling, *J. Biol. Chem.* **1984**, *259*, 7024–7030.
462. A. Akatsuka; T. J. Singh; H. Nakabayashi; M. C. Lin; K. P. Huang, *J. Biol. Chem.* **1985**, *260*, 3239–3242.
463. A. V. Skurat; P. J. Roach, *J. Biol. Chem.* **1995**, *270*, 12491–12497.
464. A. V. Skurat; P. J. Roach, *Biochem. J.* **1996**, *313*, 45–50.
465. W. A. Wilson; A. V. Skurat; B. Probst; A. dePaoli-Roach; P. J. Roach; J. Rutter, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16596–16601.
466. L. M. Ballou; E. H. Fischer, Protein Phosphatases. In *The Enzyme*, 3rd ed., P. D. Boyer, E. G. Krebs, Eds.; Academic Press; San Diego, CA, 1986, Vol. XVII, Part A, pp 312–361.
467. P. Cohen, *Annu. Rev. Biochem.* **1989**, *58*, 453–508.
468. P. Cohen; P. T. Cohen, *J. Biol. Chem.* **1989**, *264*, 21435–21438.
469. P. Strålfors; A. Hiraga; P. Cohen, *Eur. J. Biochem.* **1985**, *149*, 295–303.
470. M. J. Hubbard; P. Cohen, *Eur. J. Biochem.* **1989**, *180*, 457–465.
471. M. J. Hubbard; P. Cohen, *Eur. J. Biochem.* **1989**, *186*, 711–716.
472. M. J. Hubbard; P. Cohen, *Eur. J. Biochem.* **1989**, *186*, 701–709.
473. C. Mackintosh; D. G. Campbell; A. Hiraga; P. Cohen, *FEBS Lett.* **1988**, *234*, 189–194.
474. P. Dent; D. G. Campbell; F. B. Caudwell; P. Cohen, *FEBS Lett.* **1990**, *259*, 281–285.
475. P. Dent; A. Lavoinnie; S. Nakielny; F. B. Caudwell; P. Watt; P. Cohen, *Nature* **1990**, *348*, 302–308.
476. G. I. Welsh; C. G. Proud, *Biochem. J.* **1993**, *294*, 625–629.
477. D. A. Cross; D. R. Alessi; J. R. Vandenhede; H. E. McDowell; H. S. Hundal; P. Cohen, *Biochem. J.* **1994**, *303*, 21–26.
478. G. I. Welsh; E. J. Foulstone; S. W. Young; J. M. Tavar'e; C. G. Proud, *Biochem. J.* **1994**, *303*, 15–20.
479. A. Bellacosa; J. R. Testa; S. P. Staal; P. N. Tsichlis, *Science* **1991**, *254*, 274–277.
480. P. J. Coffey; J. R. Woodgett, *Eur. J. Biochem.* **1991**, *201*, 475–481.
481. D. A. Cross; D. R. Alessi; P. Cohen; M. Andjelkovich; B. A. Hemmings, *Nature* **1995**, *378*, 785–789.
482. T. F. Franke; S. I. Yang; T. O. Chan; K. Datta; A. Kazlauskas; D. K. Morrison; D. R. Kaplan; P. N. Tsichlis, *Cell* **1995**, *81*, 727–736.
483. H. Eldar-Finkelman; G. M. Argast; O. Foord; E. H. Fischer; E. G. Krebs, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10228–10233.
484. J. Manchester; A. V. Skurat; P. Roach; S. D. Hauschka; J. C. Lawrence, *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 10707–10711.
485. S. Frame; P. Cohen, *Biochem. J.* **2001**, *359*, 1–16.

Biographical Sketch



Jack Preiss an University Distinguished Professor received his B.S. in 1953 from the City College of New York. He completed his Ph.D. in 1957 from the Duke University. He continued his postdoctoral research between 1956–58 in the same University. He was a postdoctoral fellow at the Washington University, St. Louis between 1958–59 and at the Stanford School of Medicine from 1959 to 1960. Preiss joined the National Institutes of Health as a scientist during 1960–62. Later he was a faculty at the University of California-Davis in from 1962 to 1984.

Preiss was the recipient of the Guggenheim Memorial Fellowship in 1969–70. He is a winner of several awards such as Charles Pfizer Award in 1971 from the American Chemical Society and Alexander Von Humboldt Senior U.S. Scientist Award, 1984. During 1985–89 he was the Chair, MSU Biochemistry.

In 1990 he received the Alsberg-Schoch Memorial Award by the American Association of Cereal Chemists. The Japanese Society of Starch Science Merit Award, 1992, Distinguished Faculty Award, 1994, MSU, Distinguished Faculty Award, 1996, CNS Alumni Association, Distinguished Faculty Award, 1997, Michigan Association of Governing Boards were the awards conferred on him.

In 1998 he was Loomis Lecturer at the Iowa State University. At present he is an University Distinguished Professor from 2001 at the MSU. In addition Preiss is a highly cited researcher (isihighlycited.com). He is an elected fellow of the American Association for the Advancement of Science, 2007 and of the American Society of Plant Biology, 2008.

6.16 Celluloses

Rajai H. Atalla, Cellulose Sciences International, Madison, WI, USA

Akira Isogai, University of Tokyo, Tokyo, Japan

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|-------------------|--|-----|
| 6.16.1 | Introduction | 493 |
| 6.16.2 | Structures and States of Aggregation | 494 |
| 6.16.3 | Structural Studies | 497 |
| 6.16.3.1 | Levels of Structure | 498 |
| 6.16.4 | New Spectroscopic Methods | 498 |
| 6.16.4.1 | Solid-State ^{13}C NMR Spectra and the Two Forms of Native Cellulose I_α and I_β | 498 |
| 6.16.5 | The Need for a New Paradigm | 504 |
| 6.16.5.1 | Raman Spectroscopy | 505 |
| 6.16.6 | Further Studies of Structures in Cellulose | 509 |
| 6.16.6.1 | Raman and Infrared Spectra | 509 |
| 6.16.6.2 | Solid-State ^{13}C NMR Spectra | 511 |
| 6.16.6.3 | Electron Microscopic Studies | 513 |
| 6.16.7 | Computational Modeling | 514 |
| 6.16.8 | Polymorphism in Cellulose | 516 |
| 6.16.9 | Chemical Implications of Structure | 517 |
| 6.16.10 | Cellulose Structures in Summary | 520 |
| 6.16.11 | Cellulose Chemistry | 521 |
| 6.16.11.1 | Introduction | 521 |
| 6.16.11.2 | Solvents | 521 |
| 6.16.11.3 | Derivatization | 523 |
| 6.16.11.3.1 | Esterification | 524 |
| 6.16.11.3.2 | Etherification | 526 |
| 6.16.11.3.3 | Oxidation | 529 |
| 6.16.11.4 | Degradation | 531 |
| 6.16.11.4.1 | Acid hydrolysis | 531 |
| 6.16.11.4.2 | Enzymatic degradation | 531 |
| 6.16.11.4.3 | Thermal degradation | 531 |
| 6.16.11.5 | Chemical and Enzymatic Syntheses of Cellulose | 533 |
| References | | 536 |

6.16.1 Introduction

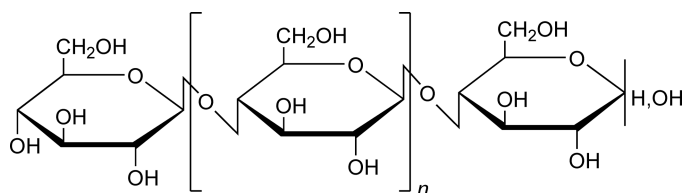
Cellulose is among the most challenging of the polysaccharides. Although it is a dominant component in the vast majority of plant forms and has a number of vital biological functions, and although cellulose-based materials have been part of our daily life for many millennia, our understanding of its nature remains incomplete. The constraints on our understanding of cellulose and its phenomenology are rooted in the complexity of its behavior and that of the structures in which it occurs in its native states as well as the different forms into which it is transformed during isolation and in the many industrial processes that are based on it as a feedstock.

These complexities are further compounded by the inadequacy of conceptual frameworks and methodologies for its characterization that have been used during the past century and to a more limited degree in the early days of the present one. Perhaps more than any other common chemical species, its chemistry and its entry into biological processes, both of biogenesis and bioconversion, are as much a function of its state of

aggregation as of its primary structure. This chapter represents a summary review and updating of earlier discussions of the phenomenology of cellulose (R. H. Atalla), together with an overview of recent developments in the chemistry of cellulose (A. Isogai). Rather than attempting to integrate the discussions, we present them in two separate sections. Section 6.16.2 will deal with the states of aggregation of celluloses and key structural issues, particularly those wherein recent developments have brought about radical changes in perspectives regarding questions of structure. Section 6.16.11 will be an updated overview of recent developments in the chemistry of cellulose, both basic and applied.

6.16.2 Structures and States of Aggregation

Since the early part of the twentieth century, cellulose has been recognized as the linear β -1,4-linked homopolymer of anhydroglucose shown schematically below. More recently, it has been noted that it is more accurate to define it as the homopolymer of anhydrocellobiose, since such definition explicitly incorporates the identity of the linkage. Given this definition, it might be anticipated that the nature and the chemical behavior of cellulose could be understood in terms of the chemistry of its monomeric constituents together with



considerations arising from its polymeric nature. The reality, however, is that the chemical or physical transformations of particular samples of cellulose are as much reflections of their sources, the processes of isolation, and their states of aggregation as they are of the chemistry of their monomeric constituents. It is for this reason that we have entitled the chapter in the plural form 'Celluloses.' The unusual nature of celluloses in their aggregated states is perhaps best represented by their relationship to water; a primary structure consisting of pyranose rings with three hydroxyl groups on each ring would, in the normal course of events, be expected to be quite soluble in water and aqueous media. This, for example, is true of the α -1,4-linked counterpart amylose that occurs in starch. But celluloses are in fact essentially insoluble, and this is but one of the many unusual patterns of behavior of cellulose aggregates. Both the oligomers of starch, the malto-oligodextrins and the Schardinger dextrins, are quite soluble in water, whereas the oligomers of cellulose, the cello-dextrins or more commonly now described as the cello-oligodextrins, are essentially insoluble in water at the octamer.

We will be concerned here with delineating the frontiers of our understanding of celluloses, in their aggregated states with particular attention to their native forms. The presentation is also very relevant to the industrial utilization of celluloses, because it addresses the nature of the native forms of many of the feedstocks used as well as the effects of processes of isolation on structure and reactivity.

The evolution of the historical perspective is included in an earlier report¹ and in references cited therein. Here the discussion of structure will focus on the results of investigations carried out during the past three decades, although some of the important contributions from earlier periods will not be neglected. Among important developments during recent decades are the introduction of spectroscopic methods that have pointed to the inadequacy of information derived solely from traditional crystallographic methods for the characterization of the aggregated states of cellulose. This is very important particularly in the context of higher plant celluloses because in many interpretations of the diffractometric data, which are very limited, assumptions concerning the symmetry of the structures have been introduced to simplify the analyses. Recent micrographic data in contrast leave little question that the assumptions are not valid for the vast majority of higher plant celluloses, which are the celluloses of most interest in current studies of celluloses. In this connection the use of solid-state ¹³C nuclear magnetic resonance (NMR) spectroscopy and Raman spectroscopy has resulted in significant new insights concerning differences in the states of aggregation of different celluloses. The application of lattice imaging methods in electron microscopic studies of algal celluloses has also been an important

development in the recent decades. Availability of these new methods and the resulting insights require the adoption of new basic paradigms as the basis for examining and assessing the phenomenology of celluloses. The focus on examination of the native states in higher plants is the key to the important new insights.

The challenge of defining structure emerges most clearly in an examination of the wide range of native celluloses that occur in nature. They occur in the cell walls of the vast majority of plant forms and are highly ordered structures that are elaborately integrated into biological structural tissues. They are also produced by a number of other organisms, including bacteria and some classes of marine life. Yet the types of order that prevail do not lend themselves at all to the description in terms of the traditional concepts that have been developed for the description of order and organization in the solid states of inorganic or inanimate organic substances. Although celluloses are often regarded as crystalline, review of observations of their many different forms will reveal that the classical definitions of crystalline order for both molecular and polymeric materials are not well suited for the characterization of celluloses. This is so in spite of the fact that experimental diffractometry may provide a basis for some useful empirical measures of the degrees of coherence of order in different celluloses. In essence, it is noted here that there is an important need for new paradigms that are explicitly cognizant of native celluloses as biological structures. These paradigms must reach beyond the traditional assumption that the phenomenology of celluloses can be reduced to fit within the context of traditional polymer theory, which has been developed on the basis of the phenomenology of synthetic polymers. Going beyond the old paradigms is essential if the understanding of celluloses and their phenomenologies is to advance to the degree needed in the twenty-first century. This is particularly so as the position of cellulose as the most abundant renewable resource emerges.

The traditional paradigms of semicrystalline polymers have created many challenges for investigators of biogenesis of celluloses. In addition, the diversity of celluloses and the difficulty in characterizing their states of aggregation have been a complicating factor in studies directed at understanding the action of different agents on cellulosic substrates whether the agents are chemical or biological. Because of the heterogeneity of the substrates, issues of accessibility arise, and these, in turn, are intimately related to the states of aggregation. The literature on biodegradation of celluloses and on the action of cellulolytic enzymes is rich with examples of the effects of sample source, history, and structure on their response. The literature on chemical modification in many ways parallels that on the action of cellulolytic enzymes. Indeed it goes further in the sense that differences in responses to the action of chemical-modifying reagents have been used to characterize the differences between the states of aggregation of cellulosic samples with different histories.

Prior to the discussion of studies that have been presented in the literature over the past four decades, it is important to clarify why the need for a new paradigm arises. That celluloses in their native states are periodic structures is beyond question. They are periodic at the level of the repeat unit in the chain, which is anhydrocellobiose and at the level of relative spatial organization of the chains relative to each other. This periodicity results in distinct characteristic diffraction patterns, particularly for the isolated celluloses. However, the native celluloses are now recognized to possess a long-period helical pattern in the chain direction. Thus, although they produce diffraction patterns, they do not possess the degree of spatial symmetry upon which classical methods of crystallographic structural analyses are based; the analysis of diffraction patterns to derive crystal structures requires the introduction of additional assumptions that are usually valid for crystalline materials, and to a more limited degree to semicrystalline polymers but which are not valid for native celluloses. These assumptions are that the crystal structures can be described as infinite lattices extending in three noncoplanar directions in Cartesian space. A further crucial assumption is that they possess infinite translational symmetry in all three of these directions. These premises underlie the possibility of constructing a reciprocal space that is the key to interpretation of diffraction patterns. That this is not true of native celluloses is perhaps best illustrated by examining electron micrographs of selected native celluloses.

In **Figure 1** we see bacterial cellulose that has been grown by Haigler² in the presence of dissolved carboxymethyl cellulose (CMC), which inhibits the aggregation of the most elementary fibrils into the ribbons that are usually observed without the CMC. The most subelementary nanofibrils in **Figure 1** are of the order of 1.5 nm in diameter. In all higher plants these most elementary fibrils aggregate into the more frequently observed nanofibrils that range between 3 and 5 nm in diameter, although in very few instances they may be as large as 6 or 7 nm. The key point of interest here is that in **Figure 1** the subelementary nanofibrils reveal an inherent characteristic tendency to curvature and a helicity in the direction of the molecular chain axes.

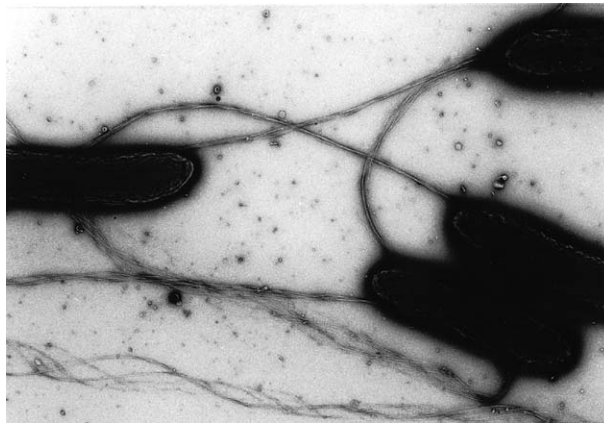


Figure 1 Nanofibrils of cellulose from *Acetobacter xylinum* grown in the presence of CMC.² Courtesy of Professor C. H. Haigler, In *Biosynthesis and Biodegradation of Cellulose*; C. H. Haigler, P. J. Weimer, Eds.; Marcel Dekker: New York, 1991; p 99.

That this tendency is inherent in the nature of aggregates of cellulose molecular chains is also observed in the structure of the cellulose fibrils derived from more primitive algae. For example in **Figure 2** we see fibrils of *Micrasterias denticulata*. **Figure 2(I)** shows the micrograph of the nanofibril, which has a 10×20 nm rectangular cross section, as it appears after dehydration during the preparation for electron micrographic studies. **Figure 2(II)** provides a rationalization by the authors of the collapse of the nanofibril to produce kinks that are spaced regularly at ~ 600 nm intervals in the electron micrograph.

It is also to be noted that **Figure 2(I)** indicates that the native helical form is easily deformed upon dehydration. This is apparent from the observation that the twisted domains are so short; if the fibril had any significant rigidity the linear segments would be much shorter. This is a matter that has not been considered in many studies of algal celluloses of relatively large fibrillar dimensions such as the 20×20 nanofibrils from *Valonia* or *Cladophora*. It is likely that the period of their helices is of the order of 2000 nm or more and thus would not be observable within the field of high-magnification electron micrographic images. It may well be also that not unlike the nanofibril of *Micrasterias*, they are deformed to fit the substrate upon which they are mounted.

Until recently, the tendency of aggregates of cellulose molecular chains to develop what has been recognized as a right-handed helical twist has not been regarded as relevant and considered a likely artifact of methods of preparation for electron microscopic observation. It was noted, however, some two decades ago by Haigler² that the larger bacterial cellulose microfibrillar aggregates have a periodicity of 600 nm. The period for *M. denticulata* shown in **Figure 2** is 1200 nm.

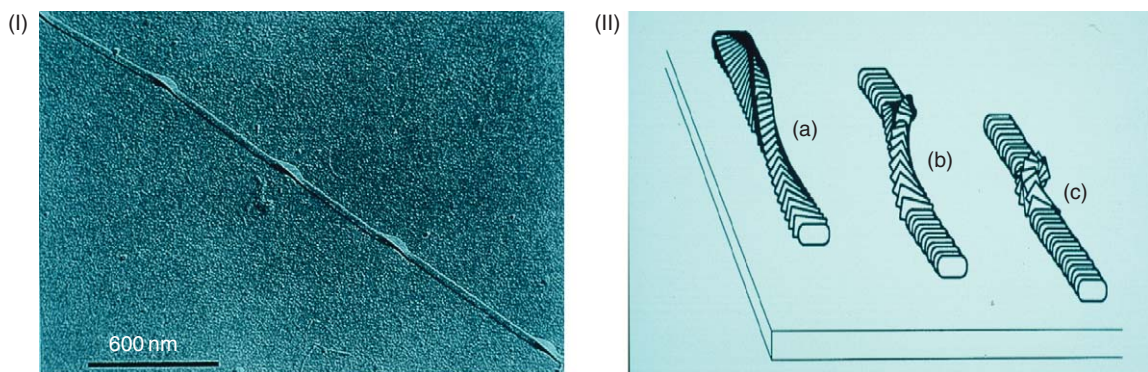


Figure 2 A fibril of *Micrasterias denticulata* as observed upon dehydration in panel (I) and a representation of the progress of dehydration suggested by the authors (II). The latter provides schematic representations of (a) the native state, (b) an intermediate stage of collapse as dehydration progresses and (c) the final appearance of the sample with the short broad domains between the more extended linear segments.³ Courtesy of Professor D. G. Gray.

The basis for recognizing and acknowledging this characteristic inherent tendency of aggregates of cellulose chains to develop a right-handed helical twist has now been established by the work of Matthews *et al.*⁴ who have demonstrated the development of the twist through molecular dynamics simulation. Although they start with the published crystal structures as initial conditions, they find that the right-handed twist evolves as the molecules are allowed to move within the context of a realistic potential field. The importance of this finding is profound since the authors had anticipated that a crystalline pattern would evolve. Instead they found that the evolution of the molecular dynamic model consistently led to a right-handed twist irrespective of the initial conditions.

Another important point that must be noted in the present context is that the basic structural units of native celluloses in higher plants are of the order of 5 nm in lateral dimensions and even those in more primitive algae or tunicates do not exceed 20–25 nm in lateral dimensions. Yet, in much of the literature of the past century, the study of the phenomenology of these celluloses has been framed within conceptual frameworks developed for macroscopic systems. In the following discussion we will note where paradigms appropriate to macroscopic systems when applied to the nanoscale phenomenology of celluloses have led to conclusions that have been misleading. We will then discuss celluloses in a context wherein the nanoscale lateral dimensions are recognized.

6.16.3 Structural Studies

Early in the twentieth century, when X-ray diffraction patterns were first observed for cellulose, it became accepted that it possesses crystallinity as an inherent element of its nature. A number of different crystalline structures were proposed and carried into the 1960s. The beginning of the past four decades of studies on the structure of cellulose was marked by the reintroduction of unit cell models based on parallel alignment of the cellulose molecular chains,^{5,6} not unlike those abandoned by Meyer and Misch⁷ in the 1930s, but also incorporating bending of the glycosidic linkage to allow the intramolecular hydrogen bond, as suggested by Hermanns.⁸ The new models were not consistent with each other, however, apart from the fact that both were based on parallel alignment of the cellulose chains. As French⁹ pointed out, they were also not strongly preferred over an antiparallel structure. In the analysis by French,⁹ it was recognized that the source of the inconsistency was not so much that the different laboratories were using different computational approaches as it was that the different diffractometric data sets were gathered from different samples and represented different intensities for similar reflections. All of these studies were undertaken before the variability of the crystalline forms of native celluloses was revealed through the high-resolution solid-state ¹³C NMR investigations.

The more recent crystallographic models also remain in question because the analyses on which they are based incorporate assumptions of symmetry in the unit cell that are inconsistent with some of the diffractometric data. Some of the reflections that are consistently observed in electron diffraction patterns and that are disallowed by the selection rules of the space group $P2_1$ ¹⁰ are consistently ignored in these crystallographic analyses.

In addition to the disallowed reflections in the electron diffraction patterns that placed the crystallographic models in question, new spectral evidence was developed pointing to the need for further refinement of structural models, particularly for native celluloses. The models derived from the crystallographic studies could not rationalize many features of the spectral data known to be quite sensitive to structural variations.

On the other hand, electron microscopic studies based on new staining techniques, specific to the reducing end groups of the polysaccharides, confirmed the parallel alignment of molecular chains within nanofibrils in algal native celluloses. These findings were further confirmed by the manifestation, at the electron microscopic level, of the action of celluloses specific to the nonreducing end group; they were clearly active at only one end of each nanofibril. While these observations were initially regarded as confirmation of crystallographic models, they only support the conclusion that the cellulose chains are organized in a parallel manner, as opposed to the antiparallel model that had been accepted since the 1930s on the basis of the work of Meyer and Misch. The remaining questions at the time were surrounding the degree to which the symmetry of space group $P2_1$, which has been the basis of many crystallographic studies, is consistent with the other structure-sensitive observations. As noted earlier, it is now clear that the assumed symmetry cannot be valid for higher plant celluloses.

6.16.3.1 Levels of Structure

It is good to revisit the issue of levels of structure at this point and clarify the levels at which the different investigative methods are most sensitive. The crystallographic models, which represent coordinates of the atoms in the unit cell, represent the most complete possible specification of structure because they include primary, secondary, and tertiary structures. And indeed, crystallographic studies of the monosaccharides and related structures provide the basis for much information concerning bond lengths and bond angles, as well as conformations in saccharide structures. The crystallographic analyses of the dimeric cellobiose and methyl β -cellobioside are among the most relevant of model saccharides. However, for polymeric systems the diffractometric data is far more limited than for a single crystal of a low molecular weight compound, so that diffraction data from a polymer must be complemented by information from other structure-sensitive methods as well as data from oligomeric structures that are based on single crystal studies. In illustration of this difficulty it is to be noted that Chu and Jeffries¹¹ in their study of cellobiose relied on over 1200 data points or reflections, whereas Ham and Williams¹² in their study of methyl β -cellobioside relied on over 1700 data points. In contrast, the maximum number of data points or reflections derived from the most recent studies of cellulose structures are of the order of 300.^{10,13}

In addition to these considerations, one must question whether a nanofibril that is 3–5 nm in diameter can be assumed to be an infinite lattice in the lateral directions, when approximately half of the molecular chains are within 1 nm from the surface and thus less constrained. The studies of cellobiose and methyl β -cellobioside were carried out on single crystals that exceeded 1 μm in all dimensions by far.

An acceptable model must not only rationalize the diffractometric data, which for cellulose is quite limited in comparison to the number of coordinates that must be specified in a definition of the unit cell, but it must also be such that it can be reconciled with information derived from other experimental measurements known to be sensitive to different levels of structure.

The new spectral evidence that must be rationalized by any acceptable structure comes from two methodologies that are particularly sensitive to structure at the secondary and tertiary levels but those were not fully available for studies of cellulose when the diffractometric studies of the 1970s were carried out. The two methodologies are Raman spectroscopy and solid-state ^{13}C NMR spectroscopy, both of which were applied to cellulosic samples for the first time during the mid- to late 1970s. The exploration of spectra measurable by these two methods can provide significant information concerning both the secondary and tertiary structures in the solid state. Because the spectral features observed are also sensitive to molecular environment, they are influenced by the degree of symmetry of the aggregated state. Hence, they provide another avenue for exploration of the validity of application of the symmetry of space group $\text{P}2_1$ to the structures of the solid state. It should be emphasized that attention to the spectral studies is not only of academic interest, but these methodologies are the ones that are now most easily applied to celluloses from higher plants for both academic and industrial investigations. And, as noted earlier, the secondary and tertiary structures of celluloses are very important to the manner in which they react to biological or chemical systems to which they are subjected.

6.16.4 New Spectroscopic Methods

The two spectroscopic techniques were first applied to studies of cellulose in the mid-1970s. As noted they are Raman spectroscopy and solid-state ^{13}C NMR. Both provided new insights concerning outstanding questions regarding the structures of cellulose. The detailed rationale for their use has been presented earlier¹ but the key issues must be reviewed here. While Raman spectroscopy led to the first questions regarding the validity of $\text{P}2_1$ symmetry and the accepted structure paradigms, it motivated the key investigations by solid-state ^{13}C NMR. In pursuit of greater clarity, however, we present the results of the application of the latter method first.

6.16.4.1 Solid-State ^{13}C NMR Spectra and the Two Forms of Native Cellulose I_α and I_β

Although applied to cellulose later than Raman spectroscopy, high-resolution solid-state ^{13}C NMR has provided perhaps the most important new insights regarding the structures of celluloses, particularly in their

native states. Earlier applications of NMR to the study of solids had been based on the exploration of dipolar interactions between pairs of magnetic nuclei and did not afford the possibility of acquisition of spectral information characteristic of molecular architecture in a manner that was parallel to that possible with liquid-state NMR spectra. The development of high-resolution solid-state NMR spectroscopy and its application to polymeric materials grew from complementary application of a number of procedures that had been developed in NMR spectroscopy. The first is proton carbon cross polarization (CP) that is used to enhance sensitivity to the low-abundance ^{13}C nucleus. This was combined with high-power proton decoupling to eliminate the strong dipolar interaction between the ^{13}C nuclei and the neighboring protons. Finally, the angular dependence of the chemical shift, or chemical shift anisotropy, is overcome by spinning the sample about an axis at a special angle to the direction of the magnetic field, commonly referred to as the magic angle, the procedure denoted by magic angle spinning (MAS). The combined application of these procedures, usually designated by CP/MAS, results in the acquisition of spectra that contain isotropic chemical shift information analogous to that obtained from liquid-state ^{13}C NMR with proton decoupling. One key difference between the two is that molecular motion in the liquid state is sufficiently rapid that chemically equivalent carbons result in single resonances, whereas in the solid-state spectra, because the molecules are immobile, the isotropic chemical shift is sensitive to the molecular environment. As a result, chemically equivalent carbons that occur in the sites that are not magnetically equivalent within an aggregated solid state may have differences between their chemical shifts.

In summary, the most important characteristic of the spectra acquired using the CP/MAS ^{13}C NMR technique is that, if they are acquired under optimal conditions, they can have sufficient resolution so that chemically equivalent carbons that occur in magnetically nonequivalent sites can be distinguished. In the present context, the corresponding carbons in different anhydroglucose units would be regarded as chemically equivalent. If they are not also symmetrically equivalent, that is, if they occur in different environments or if the anhydroglucose rings possess different conformations within the rings, at the glycosidic linkage, or at the primary alcohol group, the carbons will not have magnetically equivalent environments and will, therefore, result in distinctive resonances in the NMR spectrum. The fundamental challenge in the application of this method is to achieve a level of resolution sufficient to distinguish nonequivalences between chemically equivalent carbons. This is so because the magnetic nonequivalence can result in variations in the chemical shift that are small relative to the shifts determined by the primary chemical bonding pattern.

Another important feature of the CP/MAS ^{13}C NMR technique is that for a system such as cellulose, which consists of rather rigid hydrogen-bonded molecules and in which all carbons have directly bonded protons, the relative intensities of the resonances are expected to correspond to the proportion of the particular carbons giving rise to them. Thus, the intensities arising from each of the six carbons in the anhydroglucose ring are expected to be equal. This is an important characteristic that is central to the analysis and interpretation of the information contained within the spectra.

The first applications of the new technique to cellulose^{14,15} demonstrated the resolution of multiple resonances for some of the chemically equivalent carbons in the anhydroglucose units. It became clear that the rationalization of the spectra that were observed would provide valuable additional information concerning the structure of the celluloses investigated. The first step in such a rationalization was the assignment of the resonances that appear in the spectra.

The assignments, which have been discussed in a number of reports,^{14–20} were based on comparisons with solution spectra of cello-oligosaccharides and that of a low degree of polymerization (DP) cellulose.²¹ They are indicated in **Figure 3**, which shows a spectrum of cotton linters.²² Beginning at the upfield part of the spectrum, the region between 60 and 70 ppm is assigned to C6 of the primary alcohol group. The next cluster of resonances, between 70 and 81 ppm, is attributed to C2, C3, and C5, the ring carbons other than those anchoring the glycosidic linkage. The region between 81 and 93 ppm is associated with C4 and that between 102 and 108 ppm with C1, the anomeric carbon.

In one of the first reports on the application of the technique to studies of different celluloses, the splittings of the resonances of C4 and C1 in the spectrum of cellulose II (**Figure 4**) were regarded as confirmation of the occurrence of nonequivalent glycosidic linkages that had earlier been proposed on the basis of the comparison of the Raman spectra of cellulose II and cellobiose in the O–H stretching region.²³ These splittings were also observed in the CP/MAS spectra of the cello-oligodextrins, which aggregate in a manner very similar to that of cellulose II. The investigators who reported the spectra of the cello-oligodextrins were reluctant to depart from

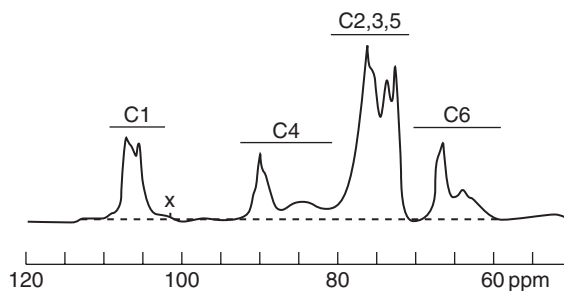


Figure 3 The ^{13}C CP/MAS spectrum of cotton linters. The horizontal bars indicate the spectral ranges of the corresponding sites in the anhydroglucose monomer unit of cellulose.

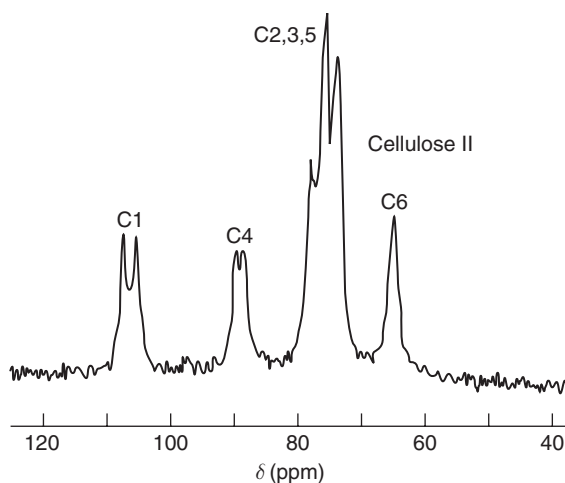


Figure 4 CP/MAS ^{13}C spectrum of high-crystallinity cellulose II recorded at relatively low resolution.

the older paradigms alluded to above and chose instead to attribute the splittings to the occurrence of nonequivalent cellulose molecules in a single unit cell.¹⁸ However, that interpretation left open the question as to why the resonances for carbons 2, 3, and 5 do not display similar splittings. If the splittings were indeed due to nonequivalent molecules it would be anticipated that those carbons nearest to the boundaries of the molecules would be the most affected. The carbons anchoring the glycosidic linkage, that is, C1 and C4, are the ones most removed from adjacent molecules, yet they also display the greatest splittings. Thus, the preferred interpretation was that the splittings of C1 and C4 were confirmation of the proposal of alternating nonequivalent glucosidic linkages in a single chain. This of course would exclude the validity of $P2_1$ as an appropriate space group for cellulose II and would lie outside the constraints of accepted structural paradigms.

Interpretation of the spectra of native celluloses presented an even more challenging task. In the spectrum of cotton linters (**Figure 3**), the two resonance regions associated with C6 and C4 include sharper resonances overlapping broader upfield wings. After excluding the possibility that the broader wings could arise entirely from the molecular mobility,^{15,16} the wings were attributed to cellulose chains in two categories of environment. The first includes all chains located at or near the surfaces of cellulose nanofibrils, which, because of their occurrence near the boundary, are less constrained with respect to the conformations they can adopt and they are not necessarily hydrogen bonded in a manner parallel to the molecules in the bulk. The surfaces are thus regarded as regions of limited two-dimensional order. The importance of this category of order had earlier been demonstrated in a study of different native celluloses undertaken by Earl and VanderHart.¹⁶ The celluloses had natural nanofibril diameters varying between 3.5 and 20 nm, and it was shown that the areas of the upfield wings of C4 and C6 declined as the surface-to-volume ratio declined.

The second category of environment contributing to the upfield wings is that of chains in regions within which the incoherence of order is not limited to two dimensions. Here, the dispersion of the frequencies at which resonances occur may arise from conformational differences, variations in bond geometries, changes in hydrogen bonding patterns, and nonuniformities in neighboring chain environments. These possibilities arise because in such regions the molecular chains are free to adopt a wider range of conformations than the ordering in a crystal lattice or its boundaries would allow.

Although the obvious upfield wings of the C4 and C6 resonances are the most direct evidence for the cellulose chains in less ordered environments, it is expected that the chains in these environments make similar contributions to the resonance regions associated with the other carbons. In the region of C1, the contribution appears to be primarily underneath the sharper resonances, although a small component appears to extend toward 104 ppm. Similarly, it is expected that the contribution from chains in the less ordered environments underlies the sharper resonances of the C2, C3, and C5 cluster.

The relative contributions of the two categories of environment to the intensities of the upfield wings were assessed in a careful analysis of the C4 wing.²² It was demonstrated that a part of the wing could be correlated with the range of the C4 resonance in amorphous cellulose prepared by ball milling. It was therefore assigned to cellulose chains occurring in the second type of environment, that is, domains wherein the incoherence of order is extended in all three dimensions. The other part of the wing was attributed to chains at the surfaces of the fibrils and, on the basis of these comparisons, it was concluded that approximately 50% of the wing is contributed by the cellulose chains in each of the two types of less ordered environments described in the preceding paragraph. Although the upfield wing of C4 is the basis of this allocation of intensities, it can be assumed that the relative contributions are similar for the upfield wing of C6 and for the component that appears to underlie the sharper resonances at C1. It is also expected that these domains contribute to the total intensity of the C2, C3, and C5 cluster between 70 and 81 ppm.

The sharper resonances in the C6 and C4 regions, centered at 66 and 90 ppm, respectively, each appear to consist of more than one resonance line even though the resolution is not sufficient to distinguish the components well. The C6 resonance seems to include at least two components, whereas the C4 resonance appears to include three closely spaced component lines. These multiplicities were interpreted as arising from carbons in the cellulose molecules within the interior of crystalline domains and therefore taken as evidence of the occurrence of chemically equivalent carbons in nonequivalent magnetic environments within the crystalline domains.

The region between 102 and 108 ppm, attributed to C1, also reveals multiplicity and sharp resonance features. Here however the upfield shoulder of the C1 resonance is very limited in contrast to the case for C4. It appears that the resonances associated with the two categories of disordered domains described above lie underneath the sharp resonances associated with the interior of the crystalline domains. It can be concluded that, in most instances, the dispersion of frequencies associated with the disorder is small relative to the shift associated with the character of the anomeric carbon C1, while that is not the case for the shifts associated with C4 and C6. One possible rationalization may be that, because of the anomeric effect, the internal coordinates surrounding C1 are much less flexible within the range of possible conformational variations than are the other internal coordinates.

In search of a rationalization of the splittings observed in the sharp resonances, CP/MAS ¹³C NMR spectra of a wide variety of samples of cellulose I were recorded. Some of these are shown in **Figure 5**.

They include ramie fibers (**Figure 5(a)**), cotton linters (**Figure 5(b)**), hydrocellulose prepared from cotton linters by acid hydrolysis (**Figure 5(c)**), a low-DP regenerated cellulose I (**Figure 5(d)**), cellulose from *Acetobacter xylinum* (**Figure 5(e)**), and cellulose from the cell wall of *Valonia ventricosa*, an alga (**Figure 5(f)**). While similar observations were reported in a number of studies^{14–20} their implications with respect to structure were more fully developed in the work of VanderHart and Atalla,^{22,24} which provides the basis for the following discussion.

All of the spectra shown in **Figures 5(a)–5(f)** are of celluloses that occur in relatively pure form in their native states and require relatively mild isolation procedures. The one exception is the spectrum (D), which was acquired from a sample regenerated from phosphoric acid at 165 °C (X). The most striking features in these spectra when viewed together are the variations in the patterns of the multiplets at C1, C4, and C6. These resonances, which are viewed as arising from chains in the interior of crystalline domains, appear to be unique to the particular celluloses; among the native forms they appear to be distinctive of the source species of the particular cellulose.

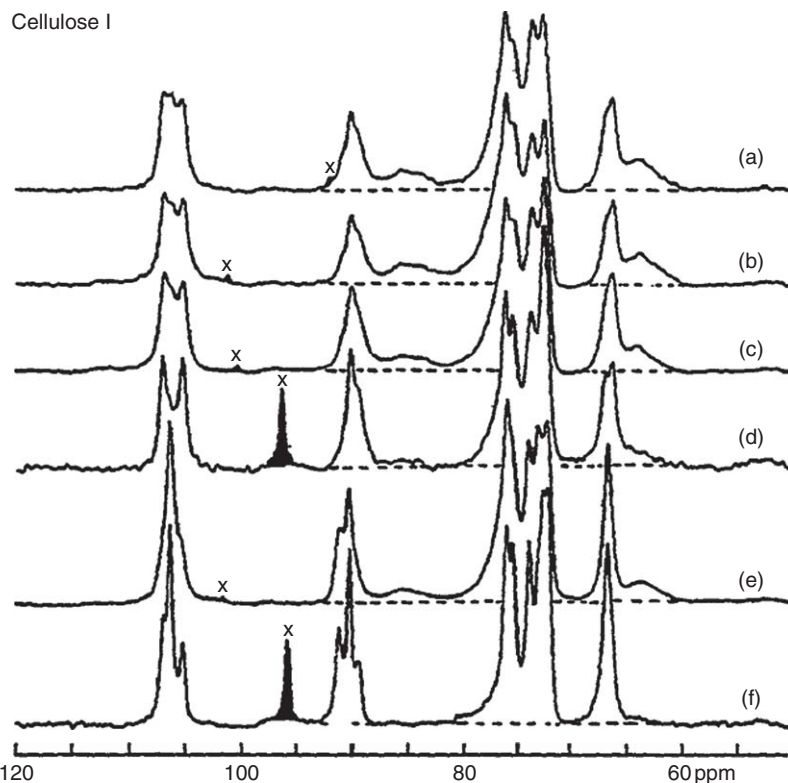


Figure 5 ^{13}C CP/MAS spectra of several cellulose I samples: (a) ramie, (b) cotton linters, (c) hydrocellulose from cotton linters, (d) low-DP regenerated cellulose I, (e) *Acetobacter xylinum* cellulose, and (f) *Valonia ventricosa* cellulose. After D. L. VanderHart; R. H. Atalla, *Macromolecular* **1984**, *17*, 1465 and R. H. Atalla; D. L. VanderHart, *Science* **1984**, *223*, 283.

The first attempt to rationalize the spectra was in terms of information that they might provide concerning a more complex structure for the unit cell of cellulose I, perhaps along the lines suggested by Honjo and Watanabe²⁵ in the 1950s. However, it soon became obvious that such a rationalization was not possible because the relative intensities within the multiplets were neither constant nor were they in ratios of small whole numbers as would be the case if the same unit cell prevailed throughout the crystalline domains. The conclusion was that the multiplicities were evidence of site heterogeneity within the crystalline domains, and that therefore native celluloses must be composites of more than one crystalline form.

Further rationalization of the spectra required a careful analysis of the multiplets at C1, C4, and C6, and the variations of the relative intensities of the lines within each multiplet among the spectra of the different celluloses. In addition to excluding a single crystal form on the basis of the considerations noted above, it was also possible to exclude the possibility of three different forms with each contributing a line to the more complex multiplets. Thus, a decomposition of the spectra on the basis of two distinct crystalline forms was pursued.

The results of the decomposition are shown as spectra in **Figures 6(b)** and **6(c)**, and were designated as the I_α and I_β forms of native cellulose. Spectrum in **Figure 6(a)** was acquired from a high-crystallinity sample of cellulose II and is included to distinguish the heterogeneity of crystalline forms occurring in the different forms of cellulose I from the long-known polymorphic variation of the crystallinity of cellulose.

Spectra in **Figures 6(b)** and **6(c)** were in fact derived from appropriate linear combinations of the spectra of the low-DP cellulose I (**Figure 5(d)**) and of the *A. xylinum* cellulose (**Figure 5(e)**). Although they represent the best approximations to the two forms of the cellulose postulated, they cannot be regarded as representative of the pure forms as they do not adequately reflect the component of the cellulose at the surfaces of the crystalline domains. Spectrum in **Figure 7(b)** does have some intensity in the upfield wings of C4 and C6, but spectrum in **Figure 7(c)** has very little evidence of such wings. There is very little question, however, that the sharp

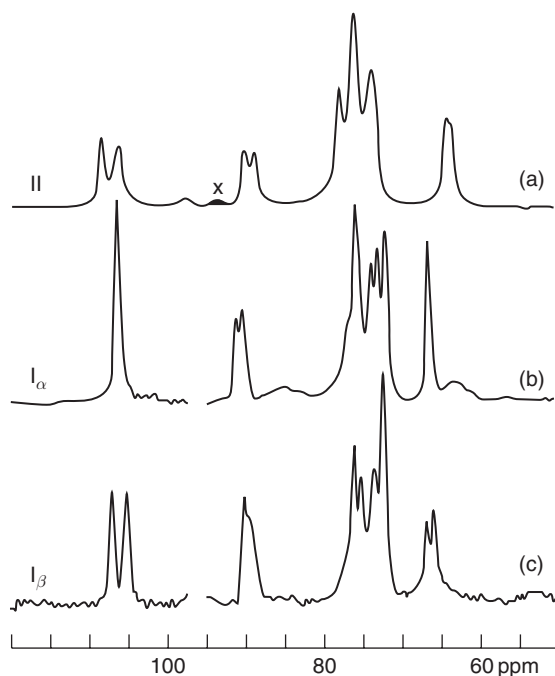


Figure 6 Comparison of the ^{13}C CP/MAS spectra (a) of a low-DP sample of cellulose II; spectra (b) and (c) corresponding to the two proposed forms of cellulose I, namely I_α and I_β , respectively. Spectra (b) and (c) were derived from linear combinations of the spectra of the low-DP regenerated cellulose I and the *Acetobacter xylinum* cellulose.

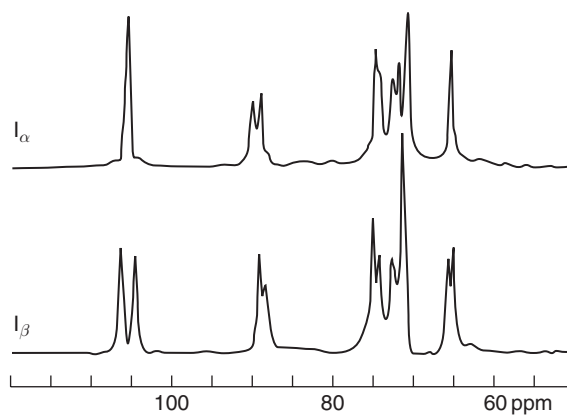


Figure 7 Alternative candidates for the spectra of I_α and I_β derived from linear combinations of the spectra of I_α -rich *Cladophora glomerata* before and after extended acid hydrolysis, which resulted in an I_β -rich cellulose.

components of spectra in **Figures 7(b)** and 7(c) include the key features in the spectra of the I_α and I_β forms. It is of interest to note here that among the distinct resonances of the I_α form at C1, C4, and C6, only the one at C4 appears to be split, whereas for the I_β form all three resonances associated with these carbons show splitting, with the one at C1 being the most pronounced.

In an effort to further validate the proposal that the I_α and I_β forms were the primary constituents of native celluloses, VanderHart and Atalla²⁶ undertook another extensive study to exclude the possibility that experimental artifacts contributed to the key spectral features assigned to the two forms. A number of possible sources of distinctive spectral features were explored. The first was the question whether the surface layers associated with crystalline domains within particular morphological features in the native celluloses could give rise to features other than those of the core crystalline domains. The second was whether variations in the anisotropic

bulk magnetic susceptibility associated with different morphologies could contribute distinctive spectral features. Exploration of the spectra of higher plant celluloses with different native morphologies revealed very little difference in the essential features of the spectra, even after the samples had been subjected to acid hydrolysis. Furthermore, it was concluded that the I_α component of higher plant celluloses was sufficiently low for which some questions were raised as to whether it occurs at all in these higher plant celluloses. In this context, it was also concluded that in the absence of the I_α component in higher plants celluloses, the line shapes of the I_β form at C4 could only be reconciled with a unit cell possessing more than four magnetically nonequivalent anhydroglucose residues per unit cell. This in turn rules out the possibility of two nonequivalent chains in a unit cell belonging to the $P2_1$ space group.

Attention was then directed to the analysis of the spectra of algal celluloses wherein the I_α component is the dominant one. Relaxation experiments confirmed that the essential spectral features identified in the two crystalline forms of cellulose were the characteristic of the core crystalline domains; when measurements were conducted such that magnetization of the surface domains was first allowed to undergo relaxation, a very little change in the spectral features was observed. The relaxation experiments suggested that the domains consisting of both the I_α and I_β forms have equal average proximity to the surface.

VanderHart and Atalla also took advantage of the spectra derived from the acid-hydrolyzed samples of the algal cellulose to generate more highly resolved representative spectra of the I_α and I_β forms. These are shown in **Figure 7** where it is clear that even in the spectrum representative of the I_α form the upfield wings of the C4 and C6 resonances are reduced to a minimum. With the completion of this study by VanderHart and Atalla, most of the questions about the possibility that the spectral features were the results of artifacts were put to rest, and the hypothesis that all native celluloses belong to one or to a combination of these forms was generally accepted.

With the above resolution of the questions concerning the nature of native celluloses in mind, it was possible to classify these celluloses with respect to the relative amounts of the I_α and I_β forms occurring in the celluloses produced by particular species. It emerged in these early studies that the celluloses from more primitive organisms such as *V. ventricosa* and *A. xylinum* are predominantly of the I_α form, whereas those from higher plants such as cotton and ramie are predominantly of the I_β form. As noted earlier, the nomenclature chosen was intended to avoid confusion with the I_A and I_B forms previously used to classify the celluloses on the basis of their infrared spectra in the O–H stretching region. In relation to that classification, the NMR spectra suggest that the I_A group has the I_α form as its dominant component, whereas the I_B group is predominantly of the I_β form.

With the recognition that all plant celluloses are composites of the I_α and I_β forms it was possible to rationalize many of the earlier difficulties in developing suitable structural models. It became clear that the efforts to reconcile the diffraction patterns in terms of a unique unit cell for native celluloses were frustrated by the reality that the celluloses were composites of two crystalline forms that were blended in different proportions in celluloses produced by different organisms.

6.16.5 The Need for a New Paradigm

It is important at this point to address the need for a new paradigm that was not recognized in the early work of Atalla and VanderHart.²³ The title of the early articles was still defined in terms of the classical approach to cellulose structure in that the two forms of cellulose, I_α and I_β , were referred to as two distinct ‘crystalline’ forms. Note was not taken at that point of the rapidly developing evidence that the lateral dimensions of most native cellulose fibrils were very limited and that cellulose nanofibrils have an inherent tendency to develop a right-handed twist when cellulose chain molecules aggregate. While this important development had shed some light on the controversies associated with many of the prior interpretations of diffractometric characterizations of native celluloses, it had not yet provided conclusive evidence that the interpretations based on the symmetry of the $P2_1$ space group for crystalline cellulose cannot be valid for native celluloses. It was the acquisition of the Raman spectra of *Tunicate* and *Valonia* celluloses that provided the conclusive evidence.

We now consider Raman spectroscopy and deal with it at greater detail because it is an approach that is less familiar to research workers in the arena of cellulose science, and also because advances in instrumentation now allow acquisition in a few minutes of the Raman spectra that previously required very tedious efforts and many hours per sample. This in turn has resulted in many laboratories acquiring the necessary instrumentation.

6.16.5.1 Raman Spectroscopy

Both Raman and infrared spectroscopies provide information about chemical functionality, molecular conformation, and hydrogen bonding. Both are based on detecting vibrational transitions from the ground states to the excited states of molecules under investigation. Raman spectroscopy, however, has some important advantages in the study of biological materials. The key advantage arises from the different bases for activity of molecular vibrations in the Raman and infrared spectra. Specifically, activity in the infrared requires finite transition moments associated with the permanent dipoles of the bonds or bond systems undergoing vibrational transitions, while activity in the Raman spectra requires finite transition moments involving the polarizabilities of the bonds. Thus, in infrared spectroscopy, the exchange of energy between the molecules and the exciting field in the far infrared is dependent on the presence of an oscillating permanent dipole. In Raman spectroscopy, in contrast, the exciting field, which is usually in the ultraviolet, visible, or near infrared range, induces a dipole moment in the molecule and the induced moment then becomes the basis for exchange of energy between the molecule and the exciting field.

It is useful in this context to view the bonds in terms of Pauling's classification along a scale between the two extremes of polar and covalent.²⁷ Bonds that are highly polar, possessing relatively high dipole moments and reduced polarizabilities, tend – when they undergo vibrational transitions – to result in bands that are intense in the infrared spectra and relatively weak in Raman spectra. Conversely, bonds that are primarily covalent in character and have relatively low permanent dipoles have high polarizabilities and hence result in vibrational bands that are intense in the Raman spectra but relatively weak in the infrared spectra. This is perhaps best illustrated by the fact that O₂ and N₂, which are homonuclear and without permanent dipoles, have very intense Raman spectra although they are inactive in the infrared absorption, whereas H₂O, which has a high permanent dipole moment, is a very strong absorber in the infrared spectra but a very weak Raman scatterer.

With respect to cellulose, the O–H groups of cellulose and those of adsorbed water are dominant in many of the spectral features in the infrared spectra. In contrast, the skeletal C–C and C–O bonds and the C–H bonds dominate the Raman spectra. A further simplification in the Raman spectra results from the circumstance that the selection rules forbidding activity of overtone and combination bands are more rigidly adhered to than is the case in the infrared spectra, so that the bands observed in the Raman spectra are usually confined to the fundamental modes of the molecules under investigation.²⁸

Raman spectroscopy possesses another important advantage in comparison to infrared spectroscopy in the study of biological systems. It is that optical heterogeneities do not present serious problems. In infrared spectroscopy, the key observable is the degree of attenuation of an incident infrared beam in comparison to a reference beam. Thus, if any process other than absorption can result in attenuation of the infrared beam, the interpretation of the spectra can be complicated. Since the refractive index of a substance can undergo large excursions into the neighborhood of strong absorption bands, the Rayleigh scattering losses will vary with frequency in the infrared absorption region of the spectrum, and they can cause anomalous features in the spectra. Furthermore, in the case of cellulosic materials the optical heterogeneities generally have dimensions in the same range as the wavelengths of the infrared absorption in the region of the fundamental vibrations. In consequence, the Rayleigh scattering of the incident infrared beam can be significant and frequency dependent. It is then difficult to discriminate attenuation of the exciting radiation by genuine absorption from that due to Rayleigh scattering.

In Raman spectroscopy, in contrast, variations in the refractive index do not present a difficulty since the excitation frequency and the frequencies of the Raman scattered photons are far removed from any absorption bands. It is, therefore, easier to record meaningful Raman spectra from samples such as cellulose, even though they may cause a high level of Rayleigh scattering of the exciting frequency.

In the context of studies on the structure of cellulose, the key advantage of Raman spectroscopy is the degree of its sensitivity to the skeletal vibrations of the cellulose molecule, with the mode of packing in the lattice having only secondary effects. This sensitivity is a consequence of the reality that most of the skeletal bonds are C–C bonds and C–O bonds, both of which have relatively high polarizabilities and hence high Raman scattering coefficients. The minimal contribution of packing effects arises from the low Raman scattering coefficients of the highly polar O–H groups, which are the functionalities that are most directly involved in intermolecular associations. The result is that intramolecular variations such as changes in internal coordinates have a significantly greater influence on the Raman spectra than do variations in intermolecular associations.

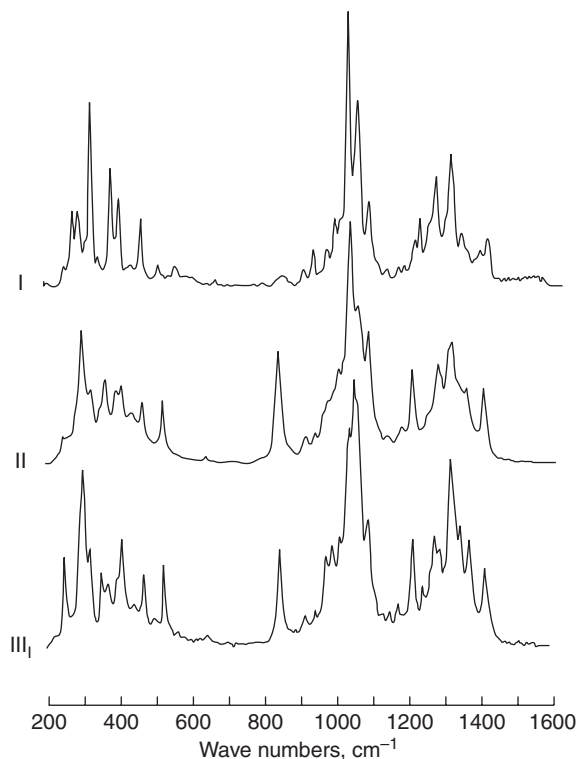


Figure 8 Raman spectra of high-crystallinity samples of celluloses I, II, and III₁.²⁹

Finally and very significantly, as the studies of the celluloses progressed, it became clear that the most dramatic differences between the vibrational spectra associated with different states of aggregation of cellulose occur in the region between 200 and 700 cm^{-1} , which is generally inaccessible with most infrared spectrometers.

This is best illustrated by viewing the spectra of highly ordered samples of the three well-established polymorphic forms of cellulose, celluloses I, II, and III₁, shown in [Figure 7](#). The designation of cellulose III as III₁ is to indicate that it was prepared from cellulose I. Indeed the sample of the spectrum which is shown in [Figure 8](#) was prepared from the sample that gave rise to the spectrum identified as I in the same figure.

To establish a basis for assessing the differences between the different cellulose polymorphs, Atalla and coworkers undertook an extensive series of studies of model compounds of increasing complexity.^{29,30-37} The model systems investigated included the 1,5-anhydropentitols, the pentitols and erythritols, the pentoses, the inositols, the hexoses, and the cello-oligodextrins including cellobiose. The studies included comprehensive normal coordinate analyses of the molecular vibrations of each of the groups of model compounds based on complementary infrared and Raman spectra. The objective of these analyses was to establish the degree to which the different classes of vibrational motions and the corresponding internal coordinates contributed to the spectral features in the different regions of the spectrum. Such a comprehensive approach was necessary because the skeletal bond systems that occur in the structures of carbohydrates are made up predominantly of C–C and C–O bonds, which possess similar reduced masses and vibrational force constants and, hence, have very similar vibrational frequencies. In consequence, a high degree of coupling occurs between the vibrations, with the result that very few of the vibrational modes are localized within specific bonds or functional groups. As a result, the traditional group frequency approach common in the assignment of infrared and Raman spectra is of very limited use except in the case of vibrations localized in the bonds of hydrogen atoms bonded to much heavier atoms such as O or C. On the other hand, the normal coordinate analyses allow identification of the degree to which the vibrations of the different types of internal coordinates contribute to each of the observed bands. Since the coupling of the vibrations is very sensitive to changes in the bond angles and the dihedral

angles associated with the bonds the vibrations of which are coupled, the normal coordinate analyses allow a detailed and systematic exploration of the effects of differences in the skeletal conformations and in the corresponding values of internal coordinates on the bands associated with particular vibrations.

With respect to the question concerning the conformations of celluloses I, II, and III, it is useful to consider first some of the pertinent information developed from the normal coordinate analyses, particularly with respect to the classes of molecular motions associated with the different spectral features. The region below 1500 cm^{-1} is the primary focus of the early exploration because the intense bands clustered at about 2900 cm^{-1} can be identified with the C–H stretching vibrations and the region beyond 3000 cm^{-1} is clearly associated with the O–H stretching vibrations and is inherently broad in a substance as highly hydrogen bonded as cellulose. In addition to the C–H and O–H stretching vibrations, the internal deformation of the methylene group on C6 is the only vibration that closely approximates a group or local mode in the usual sense implicit in discussions of assignments of vibrational spectra; the H–C–H bending vibration, sometimes alluded to as the scissors vibration, usually occurs above 1450 cm^{-1} . In all bands at frequencies below 1450 cm^{-1} , the normal coordinate analyses indicated that the vibrations are so highly coupled that, in most instances, no single internal coordinate contributes more than 20% of the potential energy change associated with any particular frequency, although in a few instances contributions were as high as 40%. Thus, as noted above, the traditional group frequency approach to the assignment of vibrational spectra, which is based on the concept of local modes, is generally not applicable in this region in the spectra of saccharides. It is necessary instead to focus on the classes of internal motions that are associated with the different frequency ranges and to interpret the spectra in terms of the influence that variations in the internal coordinates can have on the coupling between different types of vibrational deformations.

For analysis of the spectra of celluloses, it is possible to classify the groups of features in the different spectral regions in terms of the types of internal deformations that make their maximum contributions to bands in those regions. The bands between 1200 and 1450 cm^{-1} are due to modes involving considerable coupling between methine bending, methylene rocking and wagging, and C–O–H in plane bending motions; these are angle bending coordinates involving one bond to a hydrogen atom and the other to a heavy atom. However, in this region some small contributions from the skeletal stretching modes were also present. Significant contributions from ring bond stretching begin below 1200 cm^{-1} and these modes, together with C–O stretching motions, dominate between 950 and 1150 cm^{-1} . Below 950 cm^{-1} , the angle bending coordinates involving heavy atoms only (i.e., CCC, COC, OCC, OCO) begin to contribute, although ring and C–O stretches and the external bending modes of the methylene group may contribute as well. The region between 400 and 700 cm^{-1} is dominated by the heavy atom bending involving both C–O and ring bending modes, although some ring stretching coordinates still make minor contributions. In some instances, O–H out-of-plane bending motions may make minor contributions in this region as well. Between 300 and 400 cm^{-1} the ring torsions make significant contributions, and below 300 cm^{-1} they generally dominate.

In addition to the above generalized categorization concerning modes that occur in one or another of the model compound systems used in the normal coordinate analyses, the spectrum of cellulose will have contributions due to modes centered at the glycosidic linkage. The computations based on the cello-oligodextrins indicate that these modes are strongly coupled with modes involving similar coordinates in the adjacent anhydroglucose rings. The contributions of the different classes of internal coordinates to the different bands are presented in greater detail elsewhere.³⁸

As noted above and shown in **Figure 8**, the differences between the Raman spectra of celluloses I, II, and III are quite significant particularly in the region of the skeletal bending modes of vibration. In the region above 800 cm^{-1} , the differences are most obvious with respect to the relative intensities of the bands and the broadening of some of the bands upon the conversion from cellulose I to cellulose II or cellulose III. In the region below 700 cm^{-1} , in contrast, the main features are quite different in the three spectra. These differences are even more evident in the spectra of single fibers, which have been presented in considerable detail elsewhere.³⁹

In the analyses of the spectra of model compounds, changes of the magnitude indicated in **Figure 8** were associated exclusively with the occurrence of differences in conformations. It seemed very probable therefore that the differences between the spectra of celluloses I, II, and III reflect changes in the skeletal molecular conformation accompanying the transition from one form to the other. Since the basic ring structure is not

expected to change significantly,³⁸ it would appear that variations in the dihedral angles at the glucosidic linkages provide the only opportunity for conformational differences. Because of the controversy surrounding similar conclusions based on crystallographic studies carried out in the early 1960s,^{40,41} a number of experimental and theoretical avenues for validating this interpretation were also pursued.

The first consideration was whether a multiplicity of stable conformations is consistent with the results of conformational energy calculations that were available at the time.^{42,43} In both studies, the potential energy surfaces were found to possess multiple minima. When the additional constraint of a repeat length of approximately 0.515 nm per anhydroglucose unit was added, two minima representing both left-handed and right-handed departures from the twofold helix appeared to be likely loci of the stable conformations of the glucosidic linkages. It was noted in this context that these two minima were close to the positions of the dihedral angles of the glucosidic linkages in cellobiose and methyl β -cellobioside, respectively, as these were determined from crystallographic studies.^{11,12}

Next, inquiry was made into the degree to which changes in the dihedral angles about the bonds in the glucosidic linkage could influence the modes of vibration responsible for the spectral features in the different regions of the spectra. Two approaches were adopted for this purpose. The first was based on examining the Raman spectra of polysaccharide polymers and oligomers that were known to occur in different conformations. The second was a theoretical one based on an adaptation of the matrix perturbation treatment used by Wilson *et al.*⁴⁴ to discuss the effects of isotopic substitution on the infrared and Raman spectra.

The polysaccharide systems chosen for investigation were among those most closely related to cellulose in the sense that they are the α -1,4-linked polymers and oligomers of anhydroglucose. They included amylose and two of its cyclic oligomers, with primary emphasis on the latter, the α - and β -Schardinger dextrans, often also known as cyclohexa- and cyclohepta-amylose. The structures of the two oligomers differ in that the values of the dihedral angles about the bonds of the glucosidic linkages have to change to accommodate the different number of monomer units. Comparison of the Raman spectra of the cyclic dextrans showed that the differences between them were quite minor in the regions above 800 cm^{-1} , but they were quite significant in the lower frequency region dominated by the skeletal bending and torsional modes. The differences were similar in kind and distribution to the differences between celluloses I, II, and III. It was also noted that in earlier studies of the Raman spectra of amylose⁴⁵ it had been observed that forms V_a and V_h , which are very similar in conformation but had different levels of hydration, had almost identical spectra. In contrast, form B, which is known to have a distinctly different helix period, was found to have a spectrum that differs from those of forms V_a and V_h in a manner approximating the differences between the two cyclic oligodextrans. Taken together, the observations of the Raman spectra of the amyloses and the Schardinger dextrans support the interpretation of the differences between the Raman spectra of celluloses I, II, and III as pointing to differences in the chain conformations localized at the glycosidic linkages.

In the theoretical analysis, the method of Wilson *et al.*⁴⁴ was adapted to explore the consequences of variations in the dihedral angles about the bonds in the glycosidic linkage; this approach is discussed in greater detail elsewhere.¹ These considerations led to the conclusion that the skeletal bending and torsional modes are altered to a greater degree than the skeletal stretching modes when the dihedral angles associated with the glycosidic linkages undergo variations. When translated to spectral features in the Raman spectra, these observations point to major differences in the low frequency region below 700 cm^{-1} , and minor ones in the fingerprint region between 900 and 1500 cm^{-1} . These are indeed precisely the types of differences observed in comparisons of the spectra of celluloses I, II, and III.

One final consideration that was addressed is the possibility that rotations of the primary alcohol group at C6 could account for the spectral differences seen in the spectra of celluloses I, II, and III and in the spectra of the amyloses. The normal coordinate analyses of the hexoses showed that rotations about the C5–C6 bond can result in minor variations in the region below 600 cm^{-1} but that the major impact of such rotations is expected in the spectral region above 700 cm^{-1} .^{29,36} With all of the above considerations in mind, it became clear that the only plausible rationalization of the differences between the Raman spectra of celluloses I, II, and III had to be based on the possibility that differences between the skeletal conformations were the key. The key considerations have been presented elsewhere in greater detail¹ and need not be repeated here.

6.16.6 Further Studies of Structures in Cellulose

With the wide acceptance of the proposal of the two ‘crystalline’ forms (I_α and I_β) came the challenge of understanding the differences between them and their relationship to each other within the morphology of native cellulosic tissues. A number of complementary approaches were pursued by different investigators in the search for answers. Some were based on further application of solid-state ^{13}C NMR to the study of different celluloses as well as to celluloses that had been subjected to different modifying treatments. Others were based on the application of Raman and infrared spectroscopy to new classes of cellulosic samples. Others were still based on the refinement of electron microscopic and diffractometric methods. The results of these investigations will be presented in summary.

6.16.6.1 Raman and Infrared Spectra

Many studies of the vibrational spectra of the two forms have pointed to patterns of hydrogen bonding as the key difference between the two forms of native celluloses. The Raman spectra of *Valonia macrophysa* and *Halocynthia* (tunicate) celluloses reported by Atalla¹ are shown in Figure 9. These particular spectra are of interest because *V. macrophysa* is known to be approximately 65% of the I_α form, whereas *Halocynthia* is predominantly of the I_β form. Comparison of their spectra can be more rigorous than was possible in the earlier work of Wiley and Atalla³⁹ because the lateral dimensions of the fibrils of both forms are of the order of 20 nm, with the result that their spectra show equal resolution of the bands in all regions of the spectrum. It is to be noted that their spectra are essentially identical in all of the regions associated with the skeletal vibrations of all types as well as the regions associated with the vibrations involving C–H bonds, in either the bending or the stretching region. Indeed the primary differences between the two spectra are in the broad complex bands that occur in the O–H stretching region, and these differences are not unlike those noted in the earlier Raman spectral studies.

In addition, the weak band at about 840 cm^{-1} in the spectrum of *V. macrophysa* has no corresponding band in the spectrum of *Halocynthia*; this is the band attributed to the out-of-plane bending vibrations of hydrogen-bonded OH groups. There are also minor differences in the relative intensities of the methylene C–H stretches and H–C–H bending vibrations, but these are the natural consequences of different hydrogen bonding patterns for the hydroxyl group at C6. Comparison of the two spectra reinforces the interpretation presented earlier,³⁹ concluding that the only difference between the I_α and I_β forms is in the pattern of hydrogen bonding. Thus, the Raman spectral comparison of the two forms is entirely consistent with that reported for the infrared spectra of these two forms of celluloses. It must be kept in mind, of course, that the bands associated with the OH group vibrations are not expected to coincide in the Raman and infrared spectra; because of the different bases for activity in the two different spectral approaches to the measurement of vibrational frequencies,

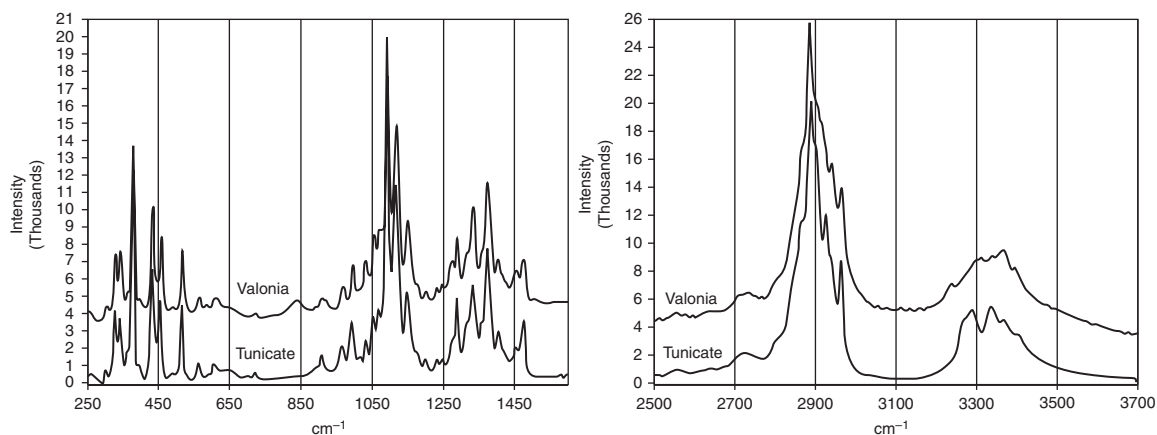


Figure 9 Raman spectra of tunicate (*Halocynthia roretzi*) and *Valonia macrophysa* celluloses in the Raman active regions.

Raman-active vibrational modes are frequently silent in the infrared and vice versa. This, of course, is true for the skeletal bands as well.

In view of the considerable variation observed in the Raman spectra of celluloses as a result of changes in molecular conformations, there can be little question that the spectra in **Figure 9** represent evidence that the conformations of the cellulose molecules in *Valonia* and *Halocynthia* are essentially identical. It is also important to note that the Raman spectra of the celluloses from *V. macrophysa* and *V. ventricosa*, both of which have been used in different studies as representatives of the I_α form, are effectively indistinguishable in all regions of the spectra. This is also true of the Raman spectra of celluloses from the algae *C. glomerata* and *Rhizoclonium heiryglypticum*, which have also been used in many studies as representative of celluloses that are predominantly of the I_α form.

In summary then, the Raman and infrared spectral studies undertaken after the discovery of the composite nature of native celluloses point to the conclusion that the only difference between the two forms is in the pattern of hydrogen bonding between chains that possess identical conformations. Yet electron diffractometric studies have been interpreted to indicate that the two forms represent two crystalline phases with different crystal habits.⁴⁶ More recently, diffractometric studies by Nishiyama *et al.*^{10,13} have also been interpreted along the same lines. It is therefore important to consider what information may be developed from the vibrational spectra with regard to this question.

The key conclusion drawn from the diffractometric data was that the I_α form represents a triclinic phase with one chain per unit cell, whereas the I_β form represents a monoclinic phase belonging to space group $P2_1$ with two nonequivalent chains per unit cell. The implication of course is that three different conformations of cellulose coexist in the different algal celluloses that are 60–70% I_α . This finding is clearly contradicted by the spectra of *Valonia* and *Halocynthia* in **Figure 9**, which are essentially identical except for the bands associated with OH groups bond stretching or bending regions. The conclusion regarding the similarity of the conformations of I_α and I_β was noted in the report by Wiley and Atalla.³⁹ While that report may have been questioned because the resolution of the ramie fiber was not as good as that of *Valonia*, the spectra of *Halocynthia* shown in **Figure 11** are of sufficiently high resolution and sufficiently similar that the comparisons can be made with confidence and it can be asserted that the two forms of native cellulose I_α and I_β have near identical conformations with the exception of differences in hydrogen bonding patterns. Thus, the crystallographic models of native celluloses need to be abandoned.

Another key issue in the contrast between the interpretation of **Figure 9** and the interpretation of the diffractometric data is that when crystal structures possess more than one molecule per unit cell and the molecules have the same vibrational frequencies, the vibrational modes of the unit cell become degenerate. Under these circumstances, couplings will arise between equivalent modes in the different molecules, and it is generally observed that such couplings result in splittings of the bands associated with the primary vibrational modes. The type of coupling that is relevant in the case of cellulose is that described as correlation field splitting.⁴⁷ This effect arises because, as a result of the coupling, the vibrations of a particular mode in the two molecules will now occur at two frequencies that are different from those of the isolated molecule; one of the two new frequencies will have the modes in the two different molecules in phase with each other, whereas the other will have the modes out of phase with each other. Such correlation field effects result in doublets with a splitting of 10–15 cm^{-1} in some fundamental vibrational modes of crystalline polyethylenes having two nonequivalent chains per unit cell. Since no evidence of such splittings occurs anywhere in the *Halocynthia* spectrum shown in **Figure 9**, it must be concluded that the I_β form cannot have more than one molecule per unit cell. It also cannot be suggested that the two molecules in a monoclinic unit cell that are nonequivalent may have modes that are at different frequencies, because the skeletal bands in the *Halocynthia* spectrum are essentially identical to those in the *Valonia* spectrum. Furthermore, this similarity was also reported in the infrared spectra observed by Sugiyama *et al.* cited earlier. Thus, it is clear that the vibrational spectra, both Raman and infrared, point to the conclusion that both the I_α and I_β forms have only one molecule per unit cell. This conclusion of course raises the question as to why the crystallographic data has been viewed for so long as pointing to a two-chain unit cell with the symmetry of space group $P2_1$. This question arises again since the recent publication of a series of studies by Nishiyama *et al.*^{10,13} based on complementary X-ray, synchrotron, and neutron scattering studies of I_α and I_β celluloses, wherein the reliance on the symmetry of space group $P2_1$ was again the foundation for the solution of the 'crystallographic' structures. Those results continue to embrace the old paradigm of crystalline cellulose.

6.16.6.2 Solid-State ^{13}C NMR Spectra

It is not surprising that the methodology that first provided the basis for understanding the composite nature of native celluloses in terms of the I_α and I_β duality has continued to be the one most often used for seeking deeper understanding of the differences between native celluloses derived from different biological sources. This has been facilitated by the greater availability of solid-state ^{13}C NMR spectrometer systems and by the relative simplicity of the procedures for acquiring the spectra from cellulosic samples.

The studies undertaken on the basis of further examination of the solid-state ^{13}C NMR spectra of celluloses are in a number of categories. The first group is focused on further examination of the spectra of different native celluloses, in part aided by mathematical procedures for deconvolution of the spectra or for resolution enhancement. Another group relies on exploring the spectral manifestations of native celluloses that have been modified in different ways. Yet a third approach is based on investigation of celluloses subjected to different but well-known procedures for inducing structural transformations in the solid-aggregated state of cellulose.

Since the approaches adopted by some groups of investigators in their acquisition and analysis of the spectra have been different, they will be presented in a manner that reflects these differences. The work of the different groups with regard to native cellulose and its response to a variety of treatments will be explored first to the extent that it illuminates questions concerning the nature of native celluloses.

The group at the Kyoto University Institute for Chemical Research carried out important studies that were complementary to those undertaken by VanderHart and Atalla.^{22,24,48} More recently, a number of other groups have made contributions. Since a number of questions concerning the nature of the I_α and I_β forms remained outstanding, it is useful to begin with an overview of the findings of different groups in this respect. These will then make it possible to view the results of studies using other methods in a clearer perspective.

The early studies by the Kyoto University group have been well summarized in a report that addresses the key points that were the focus of their investigation.⁴⁹ In a careful analysis of the chemical shifts of the C1, C4, and C6 carbons in the (CP/MAS) spectra of monosaccharides and disaccharides for which crystallographic structures were available, Horii *et al.* recognized a correlation between the chemical shifts and the dihedral angles defined by the bonds associated with these particular carbons. In particular, with respect to C6, they demonstrated a correlation between the chemical shift of the C6 resonance and the value of the dihedral angle χ defining the orientation of the OH group at C6 relative to the C4–C5 bond in the pyranose ring. This correlation is of value in the interpretation of the solid-state ^{13}C NMR spectra with respect to structure as well as the discussion of the implications of splittings of the C6 resonances observed in some of the spectra.

Of even greater interest, in light of the discussions of deviations from the twofold screw axis symmetry in some of the structures, it was observed that the chemical shifts of C1 and C4 are correlated with the dihedral angles about the glycosidic linkage. In particular, there was a correlation between the chemical shift of C1 and the dihedral angle ϕ about the C1–O bond and a correlation between the chemical shift of C4 and the dihedral angle ψ about the O–C4 bond. As the spectra published in the earliest studies did not have sufficient resolution to reveal the splittings of the resonances of C1 and C4, the possibility of occurrence of nonequivalent glucosidic linkages was not addressed at that time.

In addition to the analysis of the correlation between the chemical shifts and the dihedral angles, the Kyoto group investigated the distribution of cellulosic matter between crystalline and noncrystalline domains on the basis of measurements of the relaxation of magnetization associated with the different features of the spectra. By the measurement of the values of the spin lattice relaxation times $T_1(\text{C})$ associated with the different spectral features, they developed a quantification of the degree of crystallinity in different celluloses. They also undertook analysis of the line shapes of the different resonances, particularly that of the C4 resonance. The line shape analysis was based on deconvolution of the spectral features into combinations of Lorentzian functions centered at the assigned shifts for particular resonances. It is to be noted that the use of Lorentzian functions, which can be justified at a fundamental level in the case of spectra from molecules in solution, has no basis in any fundamental understanding of the phenomenology of acquisition of the solid-state ^{13}C NMR spectra. However, since deconvolution into Lorentzian functions has been found to be a useful tool in assessing the spectral features in the spectra of cellulose, its use has continued. The qualifications that must be kept in mind when it is used have recently been addressed by Campbell and VanderHart.⁵⁰

In the early studies by Horii *et al.*, all of the upfield wing of the C4 resonance was attributed to molecules in noncrystalline domains. On this basis, the line shape analysis of the C4 resonance of different native celluloses did not seem consistent with the model proposed by VanderHart and Atalla with respect to the composite nature of native celluloses.^{22,24} In later studies, when Horii *et al.* took note of the fact that, in the study by VanderHart and Atalla, approximately half of the upfield wing of C4 in the spectra of higher plant celluloses was attributed to the surface molecules of crystalline domains, Horii *et al.*⁵¹ indicated that their results confirm the proposal of VanderHart and Atalla. It is to be noted that in their early reports in this area, Horii *et al.* used the designations I_b and I_a to describe the different groups of celluloses in which the I_α and I_β forms were dominant. However, in their more recent studies they have adopted the I_α and I_β designations.

In pursuit of further understanding of the I_α and I_β duality, Horii *et al.* explored the effects of transformative treatments on the solid-state ¹³C NMR spectra. The first group of studies was directed at the effects of annealing, first in saturated steam⁵² and later in aqueous alkaline solutions (0.1 N NaOH) selected to avoid hydrolytic decomposition of the cellulose.^{53,54} In summary, the key findings were that the celluloses wherein the I_α form is dominant are substantially transformed into the I_β form when conditions are established so as to allow the transformation to be complete. The cellulose representative of the I_α form that was used for these studies was *V. macrophysa*. The effects of the annealing treatment are demonstrated in **Figure 10**, which shows the progression in the degree of conversion as the temperature of treatment is increased. Each of the treatments was for 30 min in the aqueous alkaline solution. These results, of course, point to the susceptibility of the I_α form to the conversion to the I_β form, suggesting that the latter is the more stable form. To test this hypothesis, a sample of tunicate cellulose, which had earlier been shown to be of the I_β form was similarly heat treated; it showed little change as a result of the annealing.

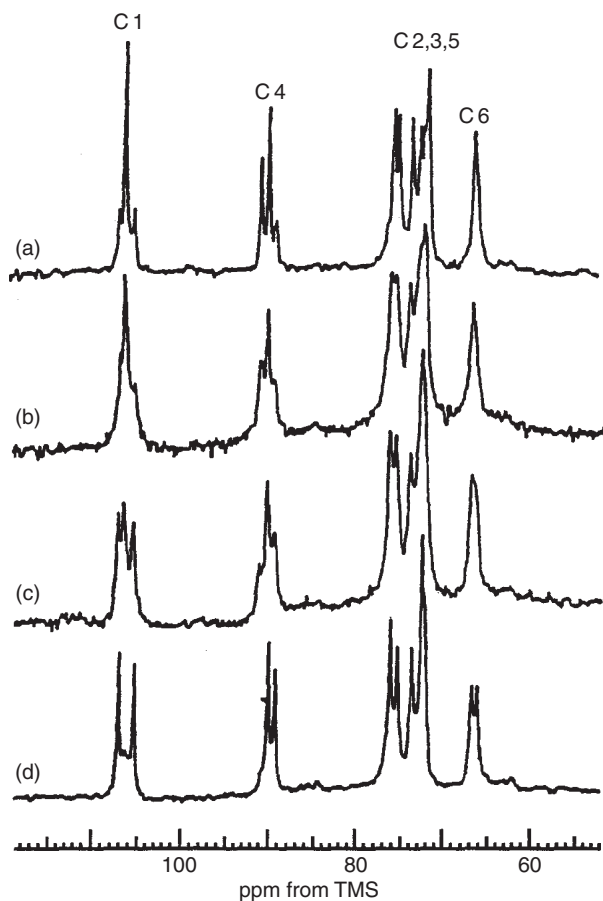


Figure 10 50 MHz CP/MAS ¹³C NMR spectra of *Valonia* cellulose annealed in 0.1 N NaOH solution: (a) original, (b) 220 °C, (c) 240 °C, and (d) 260 °C.

Additional studies by the Kyoto group relied on the solid-state ^{13}C NMR to explore the effects of different variables on the structure of cellulose.⁵⁵ It is in order, in the present context, to briefly note the results of one study in which celluloses from *A. xylinum* cultures were investigated. One of the variables explored was the temperature of the culture; it was observed that lower temperatures favored the formation of the I_α form at the expense of the I_β form. This finding raises a fundamental question regarding the possibility that the variation of the balance between the two forms is, in part, an adaptive response to changes in the environment. This in turn is a reminder that native celluloses have an important biological function and that the organisms producing them can alter the balance between the two forms of cellulose in response to changes in their environments. That similar changes may occur during the biogenesis of native celluloses in higher plants cannot be ruled out.

We follow with a commentary on the manifestations in the solid-state ^{13}C NMR spectra of a broader category of structural changes induced by different treatments known to alter the states of aggregation of cellulose. In selecting the investigations to be noted in our discussion, we will focus on studies that provide insight into the variations of the states of aggregation with the history of particular celluloses, with respect to both source and processes of isolation and transformation.

In 1990, Newman and Hemingson⁵⁶ began to combine some additional methods of processing the ^{13}C NMR spectral data with those that had been used previously such as the monitoring of the value of $T_1(\text{C})$ associated with the different spectral features. While these procedures incorporate a significant degree of empiricism, they have facilitated rationalization of the spectral features of a number of native celluloses and may therefore be valuable contributions to the repertoire of methods available for interpreting the ^{13}C NMR spectra of native celluloses. It must be noted, however, that the application of these methods has been complemented in the work of Newman and Hemingson, by a considerable degree of awareness of the complexity of the structures of both native and processed celluloses, so that their application by others needs to be approached with this awareness in mind. This work is described in the publications of Newman *et al.*^{56–61} and has been presented in overview elsewhere.¹

A different approach to mathematical analysis of the solid-state ^{13}C NMR spectra of celluloses was introduced by the group at the Swedish Forest Products Laboratory (STFI).⁶² They took advantage of statistical multivariate data analysis techniques that had been adapted for use with spectroscopic methods. Principal component analyses (PCA) were used to derive a suitable set of subspectra from the CP/MAS spectra of a set of well-characterized cellulosic samples. The relative amounts of the I_α and I_β forms and the crystallinity index for these well-characterized samples were defined in terms of the integrals of specific features in the spectra. These were then used to derive the subspectra of the principal components, which in turn were used as the basis for a partial least squares analysis of the experimental spectra. Once the subspectra of the principal components are validated by relating their features to the known measures of variability, they become the basis for analysis of the spectra of other cellulosic samples that were not included in the initial analysis. Here again the interested reader can refer to the original publications^{62–65} or the overview presented earlier.¹

6.16.6.3 Electron Microscopic Studies

The use of electron microscopy in the study of celluloses, particularly in their native state, has resulted in important advances beginning with investigations that were undertaken at the time of the introduction of the earliest electron microscopes. The early work has been ably reviewed by a number of authors.^{66,67} Of particular note among these is the coverage of the subject in the treatise by Preston.⁶⁸ Here, we will focus on studies that have been important to advancing the understanding of the structure of cellulose at the submicroscopic level.

The earliest and most significant observations, from a structural perspective, were those noted earlier by Hieta *et al.*⁶⁹ wherein they applied a staining method incorporating a chemistry that requires the presence of reducing end groups. They observed that when whole nanofibrils of *Valonia* were viewed, only one end of each microfibril was stained. This clearly indicated that the molecular chains were parallel as the reducing ends of the cellulose chains occurred together at one end of the fibrils. Soon thereafter the conclusion by Hieta *et al.* was independently confirmed by another method introduced by Chanzy and Henrissat⁷⁰ wherein the nanofibrils were subjected to the action of a cellobiohydrolase (CBH) that is specific in its action on the nonreducing ends of the cellulose chains. As noted earlier, they observed a clear narrowing of the tips of the nanofibrils to a triangular form at only one end of each nanofibril.

These early studies were focused on nanofibrils from algal celluloses that, because of their larger lateral dimensions, could be more easily visualized in detail. More recently, the technique of specific staining of reducing end groups was adapted for application to cotton nanofibrils by Maurer and Fengel.⁷¹ In addition to application of the technique to examination of native cellulose, Maurer and Fengel applied the method to the examination of nanofibrils of mercerized cellulose (cellulose II) for which they also observed staining at only one end of the nanofibrils. This last observation, which indicates a parallel chain structure in cellulose II, is very much in contrast to the crystallographic models that point to an antiparallel structure for this form of cellulose. It reinforces the view that the structure of cellulose II still has many uncertainties associated with it, in spite of many theoretical analyses that have attempted to rationalize the antiparallel form.

In yet another important set of investigations reported by Sugiyama *et al.*, at approximately the same time,^{72–74} it was demonstrated that lattice images could be recorded from the nanofibrils of *V. macrophysa*. The first images captured were based on lateral observation of the nanofibrils.^{72,73} Later, the techniques were refined to allow the acquisition of lattice images of cross sections of nanofibrils.⁷⁴ The significance of these observations was that it was now possible to demonstrate conclusively that the nanofibrils are uniform in formation and that there is no evidence that they are composed of smaller subunits that aggregate together to form the individual nanofibrils that are observed in the electron micrographs. Thus, the observations resolved some of the questions that had arisen earlier concerning the interpretation of electron micrographs of native celluloses;^{68,75} the findings of Sugiyama *et al.* were the first direct evidence that the approximately 20×20 nm cross sections were not composed of distinguishable smaller subunits. It should be noted, however, that the electron diffraction processes responsible for the formation of the lattice images are dominated by the organization of the heavy atoms in the molecular chains and would be insensitive to any nonuniformity in the hydrogen bonding patterns within the interior of the 20×20 nm fibrils. The homogeneity of the microfibrils revealed in the lattice images is an issue that needs to be revisited in the context of discussions of biogenesis, for in each instance the homogeneous crystalline domains clearly include a much larger number of cellulose chains than could possibly arise from the individual membrane complexes associated with the biogenesis of cellulose.

Later studies by Sugiyama *et al.*⁷⁶ were based on electron diffraction and were directed at addressing questions concerning the nature of the differences between the I_α and I_β forms of cellulose. In a landmark study, electron diffraction patterns were recorded from *V. macrophysa* both in its native state, wherein the I_α and I_β forms occur in their natural relative proportions, and after annealing using the process first reported by Horii and coworkers,⁷⁷ which converts the I_α form into the I_β form. The native material, which is predominantly the I_α form, was shown to produce a complex electron diffraction pattern similar to that which had earlier led Honjo and Watanabe²⁵ to propose an eight-chain unit cell. In sharp contrast, the annealed sample, which is essentially all of the I_β form, produced a more simple and symmetric pattern that could be indexed approximately in terms of a two-chain monoclinic unit cell.

On the basis of the results of electron diffraction studies by Sugiyama,⁷⁶ Nishiyama *et al.* undertook their very elaborate analyses of the diffraction patterns of the two forms of cellulose using X-ray, synchrotron, and neutron scattering. They also concluded that the two forms of cellulose have different unit cells, which imply that three different conformations coexist in the algal celluloses that are 60–70% of the I_α form. This not only contradicts the clear evidence from the Raman spectra shown in **Figure 9**, but even more importantly is in direct conflict with the results of the lattice image studies reported earlier by Sugiyama *et al.*^{72–74} that showed the nanofibrils of *V. macrophysa* consist of a uniform pattern with no substructures to it; *Valonia* is approximately 65% I_α and 35% I_β .

6.16.7 Computational Modeling

Computational modeling has become an important aid in advancing the understanding of complex molecular systems. Molecular modeling methods have allowed the exploration of many different factors and interactions at the molecular level and the degree to which they may contribute to the phenomenology of different molecular systems. So, computational modeling has found particular favor in the analyses of large molecules of biological origin. And, of course, cellulose and its oligomers have attracted attention. It is valuable to review briefly some of the efforts directed at advancing the understanding of cellulose because, in addition to providing

insights regarding the contributions of different classes of interactions, they illustrate the reality that the results of analyses can often be the consequences of assumptions and premises introduced at the outset, rather than conclusions that can provide definitive answers to questions under exploration. It has been our experience that conclusions concerning the structures that are most favored for different celluloses change when new sets of potential functions are introduced into the computational programs and as different approaches to finding the most favored structures are adopted.

The analysis by Reese and Skerret⁷⁸ was one of the first computational efforts to explore the constraints on the freedom of variation inherent in the structure of cellobiose. It relied on a potential function that is focused on van der Waals interactions in order to establish the degree to which domains within ψ/ϕ space may be excluded by hard sphere overlap. The key finding was that approximately 95% of ψ/ϕ space was indeed excluded from the accessibility on the basis of hard sphere overlaps that were unacceptable in the sense that they required particular atoms associated with the region of the glycosidic linkage to be significantly closer to each other than the sum of the van der Waals radii. And upon mapping the energy associated with allowable conformations, they found that the two regions represented energy minima close to the conformations defined by the twofold helical constraint implicit in the symmetry of space group $P2_1$. The boundaries of the acceptable region are not very far removed from the domains within the contours; the region between the two domains along the twofold helix line was not excluded by hard sphere overlap but it did represent a saddle point in the potential energy surface.

The next group of computational studies did incorporate the hydrogen bonding energies as well as the van der Waals interactions. Whether they exhibited the double minima, and the degree to which the double minima were pronounced, depended in large measure on the relative weighting given to the different types of nonbonded interactions. In many, particularly those relying on the potential energy functions incorporated in the linked atom least squares (LALS) programs, the weighting was based on fitting the potential functions to optimize the match between computed structures for small molecules for which the crystal structures were known from crystallographic studies. The results were that some showed very shallow minima off the twofold helix line;⁷⁹ the twofold helical structures associated with the space group $P2_1$ were then rationalized on the basis that the departure represented a small difference in the energies that were regarded as within the error of the computation. When the criterion for quality of fit is chosen as a global minimum of the potential energy, without attention to the fact that it may incorporate unacceptable hard sphere overlaps, the results of the computational analysis can be misleading.

Later studies did not incorporate disproportionate weighting of the different types of nonbonded interactions^{43,80} and the result is perhaps best illustrated in the mapping of the potential energy for cellobiose shown in **Figure 11** taken from the study of cello-oligomers by Henrissat *et al.*⁷⁹ In this instance, for purposes of

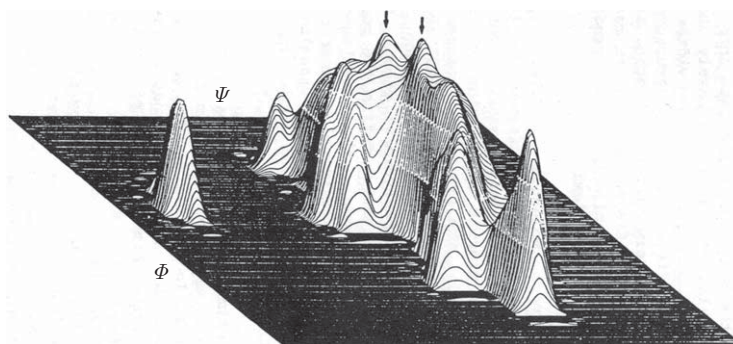


Figure 11 Perspective drawing of the three-dimensional shape of the mirror image of the conformational energy well for the full angular range for ϕ and ψ . The volume was constructed using the following scheme: $V(\phi, \psi) > 15 \text{ kcal mol}^{-1}$: $V(\phi, \psi) = 0$; $V(\phi, \psi) < 15 \text{ kcal mol}^{-1}$: $V^p(\phi, \psi) = -(V(\phi, \psi) - 15)$. V being the energy expressed^p relative to the minimum. Proceeding from top to bottom of the three-dimensional shape, we note: the very low energy region (the arrows point towards the conformations observed for crystalline cellobiose and methyl β -D-cellobioside). The 5 kcal mol^{-1} to the 10 kcal mol^{-1} energy contours correspond to the light grey region of the volume.

visualization, the ψ/ϕ map presents the mirror image of the potential energy surface computed for cellobiose. While well more than two minima are shown in this mapping, it is to be noted that only the two corresponding to the crystal structures of cellobiose and methyl β -cellobioside, marked by arrows, are within the boundaries established in the analysis by Reese and Skerret described earlier. The other minima correspond to conformations that are more favorable to hydrogen bonding, but with relatively high energies associated with the van der Waals interactions pointing to severe hard sphere overlap.

More recent computational studies have taken advantage of the vast expansion of computational capabilities in that they are able to follow the evolution of a molecular system in time under the influence of extant molecular interactions and fields of influence of the different constituent atoms. This has been the work of Brady and coworkers most recently described in the report by Matthews *et al.*⁴ noted above. It is indeed the first computational study that is consistent with the observations of long-period helical organization noted in **Figures 1 and 2**. Further studies are ongoing and promise to enhance the understanding of the phenomenology of celluloses at the nanoscale level.

6.16.8 Polymorphism in Cellulose

One of the discoveries growing out of the early diffractometric studies of cellulose was that it can occur in a number of allomorphic forms in the solid state, each producing distinctive X-ray diffractometric patterns.⁸¹ In addition to the cellulose II form, which has been discussed extensively, two other forms have been recognized; these are cellulose III and cellulose IV. It is of interest to consider them briefly because they reflect the capacity of cellulose to aggregate in a wide variety of secondary and tertiary structures and because some of the higher plant celluloses produce diffraction patterns that are not unlike those of cellulose IV. Furthermore, they reflect the tendency for some of the celluloses to retain some memory of their earlier states of aggregation in a manner not yet understood.

Cellulose III is of little interest from a biological perspective except to the extent that its behavior may reveal some of the interesting characteristics of the native celluloses from which it can be prepared. It can be prepared from either native cellulose or cellulose II by the treatment with anhydrous liquid ammonia at temperatures near -30°C . It produces distinctive X-ray patterns, Raman spectra, and solid-state ^{13}C NMR spectra. The most interesting characteristic it has is that it can be restored to the original form by treatment in boiling water. Because of this characteristic it is common to designate samples of cellulose III as either III_I or III_{II} to indicate both the source material and the form that will be recovered if the cellulose is boiled in water. In the case of native celluloses, the transformation to the III_I form and back to the I form also has the unusual effect of converting celluloses in which the I _{α} form is dominant, such as those from algal sources, into forms in which the I _{β} form is dominant. This effect, first reported by Chanzy *et al.*,⁸² is accompanied by the partitioning of the algal microfibrils into smaller ones that are closer in lateral dimensions to those characteristic of higher plants. Both the solid-state ^{13}C NMR and the Raman spectra also then appear more like those of the higher plants. No such changes have been reported for native celluloses in which the I _{β} form is dominant. These behaviors by cellulose III point to a memory effect with respect to the secondary and tertiary structures of cellulose that remains very much a mystery at the present time.

Cellulose IV is most often described as the high-temperature cellulose because it can be prepared by exposing the source cellulose to temperatures in the vicinity of 260°C while it is immersed in glycerol. In this preparation, it is reported to depend in structure on whether it is prepared from cellulose I or cellulose II; hence, the frequent designation as IV_I or IV_{II}. When prepared from cellulose I, it is first converted to the III_I form prior to the treatment at high temperature in glycerol. When prepared from cellulose II, it can be produced directly from the II form or through the III_{II} form as an intermediate. However, in the instance of cellulose IV there are no known procedures that allow restoration to the original form; the use of the different designations reflects some differences in the diffraction patterns observed from the two different forms. Furthermore, most of the reported preparations of cellulose IV from native forms of cellulose have been from higher plant celluloses wherein the I _{β} form is dominant and the lateral dimensions of the native microfibrils are quite small; it is not at all clear that treatment of microfibrils of larger lateral dimensions such as those of *Valonia* or those of *Halocynthia* will result in such changes.

In addition to its preparation by heating at 260 °C in glycerol, cellulose IV has been recovered when cellulose is regenerated from solution at elevated temperatures. This has been observed with solutions in phosphoric acid regenerated in boiling water or in ethylene glycol or glycerol at temperatures above 100 °C.⁸³ It has also been observed upon regeneration from the dimethylsulfoxide-paraformaldehyde solvent system at the elevated temperatures.⁸⁴ In yet another exploration of high-temperature effects on the aggregation of cellulose, it was found that when amorphous celluloses are prepared under anhydrous conditions and then induced to crystallize by exposure to water, the exposure at elevated temperatures results in the formation of cellulose IV rather than cellulose II, which is the form usually obtained upon crystallization at room temperature.⁸⁴

The samples of cellulose IV obtained through regeneration from solution were shown to have Raman spectra that could be represented as linear combinations of the spectra of celluloses I and II, at first suggesting that it may be a mixed lattice in which molecules with two different secondary structures coexist. However, a mixed crystal would be expected to have Raman spectra that are quite different from those of celluloses I and II. It appears more likely that cellulose IV consists of nanonuclei of celluloses I and II. Such a mixture would explain the observation that the spectra of cellulose IV are linear combinations of the spectra of celluloses I and II.

One of the complications in interpreting the observations of the occurrence of cellulose IV is that its X-ray diffraction powder pattern is very similar to that of cellulose I. The 020 reflection is nearly identical to that of cellulose I and the 110 and 1–10 reflections collapse into a single reflection approximately midway between those of cellulose I. As a result, many of the less well-ordered native celluloses produce X-ray patterns that could equally well be interpreted as indicating cellulose I, but with inadequate resolution of the 110 and 1–10 reflections, or as indicating cellulose IV. They are usually characterized as indicating cellulose I because they represent celluloses derived from native sources. Indeed, when cellulose IV was first observed, it was thought to be a less ordered form of cellulose I.

The close relationship between cellulose IV and the native state is also reflected in reports of its observation in the native state of primary cell-wall celluloses. These were the observations based on electron diffraction studies of isolated primary cell-wall celluloses.⁸⁵

6.16.9 Chemical Implications of Structure

It was noted earlier that an acceptable fit to the diffractometric data is not the ultimate objective of structural studies. Rather it is the development of a model that possesses a significant measure of validity and usefulness as the basis for organizing, explaining, and predicting the results of experimental observations. In the above sections, the new and evolving conceptual framework for describing the structures of cellulose was described in relation to spectral observations. It is also important to consider the degree to which the structural information that has been developed above may be useful as the basis for advancing the understanding of the response of celluloses to chemical reagents and enzyme systems. It is useful first to review briefly the past work directed at rationalizing the responses to such agents.

The vast majority of studies of the chemistry of cellulose have been directed at the preparation of cellulose derivatives with varying degrees of substitution depending on the desired product. Sometimes the goal is to prepare a cellulose derivative that possesses properties that differ significantly from those of the native form; some derivatives are water-soluble, others are thermoplastic, and others still are used as intermediates in the processes for the regeneration of cellulose in the form of films or fibers. At other times the objective is to introduce relatively small amounts of substitution to modify the properties of the cellulosic substrate without it losing its macroscopic identity or form such as fiber or microcrystalline powder or regenerated filament or film. All such modification processes begin with a heterogeneous reaction system, which may or may not eventually evolve into a homogeneous system as the reaction progresses. Thus, in all chemical investigations that begin with cellulose as one of the ingredients, issues associated with heterogeneous reaction systems arise. Understandably, the one that has been dominant in most investigations is the question of the accessibility of the cellulose.

A variety of methods have been developed to relate accessibility to microstructure. Almost all of them begin with the premise that the cellulose can be regarded as having a crystalline fraction and a disordered or amorphous fraction. It is then assumed that the amorphous or disordered fraction is accessible while the

crystalline fraction is not. In some instances, the portion of the crystalline domains that is at their surface is regarded as accessible and it is therefore included as part of the disordered fraction. In other instances, the particular chemistry is thought to occur only in the disordered fraction and the surfaces of crystalline domains are not included. The different approaches have been reviewed by Bertoniere and Zeronian,⁸⁶ who regard the different approaches as alternative methods for measuring the degree of crystallinity or the crystalline fraction in the particular celluloses.

A number of different chemical and physical approaches are described by Bertoniere and Zeronian. The first is based on acid hydrolysis followed by quantification of the weight loss due to dissolution of glucose, cellobiose, and the soluble oligomers.⁸⁷ This method is thought to incorporate some error in the quantification of the crystalline domains because the chain cleavage upon hydrolysis can facilitate crystallization of chain molecules that had been kept in disorder due to entanglement with other molecules. Another method is based on monitoring the degree of formylation of cellulose when reacted with formic acid to form the ester.⁸⁸ In this method, the progress of the reaction with cellulose is compared to a similar reaction with starch, which provides a measure of the possibility of formylation in a homogeneous system wherein the issue of accessibility does not arise.

In another method developed by Rowland and his coworkers^{89–92} accessible hydroxyl groups are tagged through reaction of the particular cellulose with *N,N*-diethylaziridinium chloride to produce diethylaminoethyl (DEAE)-cellulose. This is then hydrolyzed, subjected to enzyme action to remove the untagged glucose, silylated, and subjected to chromatographic analysis. This method has the added advantage that it can be used to explore the relative reactivity of the different hydroxyl groups. It is usually observed that the secondary hydroxyl group on C2 is the most reactive and the one at C3 the least reactive, with the primary hydroxyl at C6 having a reactivity approaching that of the group on C2 under some conditions. Here, of course, steric effects are also factors in these substitution reactions.

Among the physical methods discussed by Bertoniere and Zeronian⁸⁶ are the ones based on sorption and solvent exclusion. One of the earliest studies relying on the use of sorption as a measure of accessibility was the classical study by Mann and Marrinan⁹³ wherein deuterium exchange with the protons was monitored. The cellulose was exposed to D₂O vapor for a period sufficient to attain equilibrium and then the degree of exchange was measured by observation of the infrared spectra. Comparison of the band associated with the O–D stretching vibration to those associated with the O–H stretching vibration provided the measure of the relative amounts of accessible and inaccessible hydroxyl groups. Another approach to monitoring availability to adsorbed molecules is the measurement of moisture regain upon conditioning under well-defined conditions as described by Zeronian and coworkers.⁹⁴

The method of solvent exclusion has been used to explore issues of accessibility on a somewhat larger scale. An approach pioneered by Stone and Scallan⁹⁵ and Stone *et al.*⁹⁶ relied on static measurement using a series of oligomeric sugars and dextrans of increasing size to establish the distribution of pore sizes in different preparations of a variety of native celluloses.

While the methods for characterizing celluloses on the basis of their accessibility have been useful, they do not provide a basis for understanding the level of structure at which the response of a particular cellulose is determined. This follows from the rather simple categorization of the substrate cellulose into ordered and disordered fractions corresponding to the fractions that are thought to be crystalline and those that are not. This classification does not allow discrimination between effects that have their origin at the level of secondary structure and those that arise from the nature of the tertiary structure. Thus, in terms of chemical reactions, this approach does not facilitate separation of steric effects that follow from the conformation of the molecule as it is approached by a reacting species, from the effects of accessibility, which is inherently a consequence of the tertiary structure.

The possibility of advancing the understanding of the chemical implications of structure is best illustrated in the context of hydrolytic reactions. Among the patterns that emerge fairly early in any examination of the published literature on acid hydrolysis and on enzymatic degradation of cellulose are the many similarities in the response to the two classes of hydrolytic agents. In both instances, a rapid initial conversion to glucose and cellobioses is followed by a period of relatively slower conversion, the rate of conversion in the second period depending on the prior history of the cellulosic substrate. In general, the nonnative polymorphic forms are degraded more rapidly during this second phase. In addition, it is found that the most crystalline or highly

ordered of the native celluloses are particularly resistant to attack, with the most highly crystalline regions converted much more slowly than any of the other forms of cellulose.

The relationship of the patterns of hydrolytic susceptibility to the range of conformational variation discussed above can be interpreted in terms of contrast between the states of the glycosidic linkage in cellobiose and β -methylcellobioside. The differences between the states that are likely to contribute to the differences in observed reactivity are of two types. The first is the differences in the steric environment of the glycosidic linkage, particularly with respect to activity of the C6 group as a steric hindrance to, or as a potential promoter of, proton transfer reactions, depending on its orientation relative to the adjacent glycosidic linkage. The second type of difference is electronic in nature and involves readjustment of the hybridization of the bonding orbitals at the oxygen in the linkage. It is worthwhile to examine the potential contribution of each of these effects.

The steric environment emerges most simply from the examination of scale models of the cellodextrins. They reveal that when C6 is positioned in a manner approximating the structure in β -methylcellobioside, the methylene hydrogens are so disposed that they contribute significantly to the creation of a hydrophobic protective environment for the adjacent glycosidic linkage. If, however, the rotation about the C5–C6 bond is allowed, the primary hydroxyl group can come into proximity with the linkage and can provide a potential path for more rapid proton transfer.

If, as suggested earlier on the basis of spectral data, the orientation of some C6 groups in native cellulose is locked in by its participation in a bifurcated hydrogen bond to the hydroxyl group on C3, it may contribute to the higher degree of resistance to hydrolytic action. Access to the linkage oxygen would be through a relatively narrow solid angle, barely large enough to permit the entry of the hydronium ions that are the primary carriers of protons in acidic media.⁹⁷ If, on the other hand, the C6 group has greater freedom to rotate, as is likely to be the case in cellulose II, the hindrance due to the methylene hydrogens can be reduced and, in some orientations, the oxygen of the primary hydroxyl group may provide a tunneling path for the transfer of protons from the hydronium ions to the glycosidic linkage. This would result in greater susceptibility of nonnative celluloses to hydrolytic attack.

The key role of C6 in stabilizing the native cellulose structures is supported by findings concerning the mechanism of action of the dimethylsulfoxide-paraformaldehyde solvent system for cellulose, which is quite effective in solubilizing even the most crystalline of celluloses. The crucial step in the mechanism that has been established for this system is substitution of a methylol group on the primary hydroxyl at the C6 carbon.^{98,99}

The effect of conformation on the electronic structure of the linkage is also likely to be a factor with respect to its susceptibility to hydrolytic attack. Although there is no basis for anticipating the directions of this effect at this time, it is well to consider it from a qualitative perspective. First, it is clear that the hybrids of oxygen orbitals involved in the bonds to carbon must be nonequivalent because the bond distances differ to a significant degree.^{24,25} The angle of approximately 116° imposed on the linkage is likely to result in greater differences between the bonding orbitals and the lone pair orbitals than might be expected in a typical glycosidic linkage that is free from strain. Among themselves, the lone pair orbitals are likely to be different because of their different disposition with respect to the ring oxygen adjacent to C1 in the linkage; the differences may be small and subtle, but they are no less real. Given these many influences on the nature of the hybridization at the oxygen in the linkage, it seems most unlikely that they would remain unaltered by changes in the dihedral angles of the magnitude of the difference between cellobiose and β -methylcellobioside. Hence, a difference in electronic character must be expected.

At present it is not possible to estimate the magnitude of the effects discussed or to speculate concerning the direction of the change in relative reactivity of the glycosidic linkage in the two different conformations. Yet it is clear that differences can be anticipated and they may be viewed, within limits of course, as altering the chemical identity of the glycosidic linkage when its conformation changes. It remains for future studies to define the differences more precisely.

The points raised with regard to the influence of conformation on factors that determine the pathways for chemical reaction have not been the specific subjects of investigation because the methods for characterizing the secondary structure as apart from the tertiary structure have not been available. It is also true that suitable conceptual frameworks have not been available for developing the questions beyond the levels of the order – disorder duality. With the development of the approaches outlined above for exploring and distinguishing between matters of secondary structure and those of tertiary structure it is quite likely that in the years ahead it will be possible to achieve a higher level of organization of information concerning the chemistry of cellulose.

With respect to questions of tertiary structure the key issue introduced by the new structural information, and the one that has not been explored at all at this time, is whether the different hydrogen bonding patterns associated with the I_α/I_β duality have associated with them the differences between the reactivity of the hydroxyl groups involved. It is not clear at this time how experiments exploring such effects might be carried out so as to separate issues associated with the differences between the hydrogen bonding patterns from issues associated with the differences in accessibility.

6.16.10 Cellulose Structures in Summary

Before turning to the more detailed discussion of recent developments in cellulose chemistry in Section 6.16.11 of this chapter, it is helpful to summarize where studies at the nanoscale level stand at the present time and to assess the degree of confidence with which we can use their conclusions as the basis for further discussion.

From crystallographic studies, based on both X-ray and electron diffraction measurements, it can be concluded that the secondary structures of native celluloses are ribbon-like conformations approximating twofold helical structures. Their organization into crystallographic unit cells remains uncertain, however. The monoclinic space group $P2_1$, with two chains per unit cell, has been proposed both for the earlier studies prior to the discovery of the I_α/I_β duality in native forms and more recently for the I_β form. The I_α form is thought to possess a triclinic unit cell structure. Some important questions remain regarding the degree to which these are adequate representations of the organization of the crystalline domains in native celluloses. The majority of crystallographic studies also point to parallel alignment of the cellulose chains in the native celluloses, and this conclusion has been confirmed by electron micrographic observations. For cellulose II also, the structures derived from the X-ray diffraction data suggest a ribbon-like secondary structure approximating twofold helical organization and, in this instance, antiparallel alignment of the chains. For cellulose II, the antiparallel proposal has been contradicted by recent electron micrographic observations. The unit cell organization of space group $P2_1$, with two chains per unit cell, has also been suggested for cellulose II, although the degree of confidence is even less than that with respect to the structures of cellulose I.

The early Raman spectroscopic studies could not clearly be reconciled with the premise that both celluloses I and II possess the twofold helical conformations as the crystallographic studies had suggested. The Raman spectra, together with some corresponding infrared spectra, also pointed to the probability that the repeat unit of the structure of crystalline celluloses is anhydrocellobiose, so that alternating nonequivalent glycosidic linkages occur within each chain. To preserve the ribbon-like structural approximation, the different conformations of celluloses I and II were rationalized as alternate left- and right-handed departures from the twofold helical structure, with those in cellulose II representing somewhat larger departures from the twofold helical conformation than those in cellulose I.

The introduction of high-resolution solid-state ^{13}C NMR spectral analyses to the study of celluloses resulted in resolution of one of the fundamental mysteries in the variability of native celluloses by establishing that all native celluloses are composites of two forms. These were identified as the I_α and I_β forms to distinguish them from the I_A and I_B categories that had been introduced more than three decades earlier to distinguish between the celluloses produced by algae and bacteria and the celluloses from higher plants. The correspondence between the two classifications is that those in the I_A category have the I_α form as the dominant component, whereas those in the I_B category are predominantly of the I_β form. The nature of the difference between the I_α and I_β forms remains the subject of serious inquiry. Recognition of the I_α/I_β duality has facilitated a significant amount of additional research seeking to establish the balance between the two forms in a wide range of higher plant celluloses.

In later studies, the Raman spectra and corresponding infrared spectra indicated that the primary differences between the I_α and I_β forms of native cellulose were in the pattern of hydrogen bonding. Furthermore, the Raman spectra of the two forms raise questions as to whether the structures can possess more than one molecule per unit cell since there is no evidence of any correlation field splittings of any of the bands in the spectra of the two forms.

Electron microscopic studies relying on agents that act selectively either at the reducing or at the nonreducing end groups of the cellulose chains have provided convincing evidence that cellulose chains are aligned parallel to one another in native cellulose. More recently, similar evidence has been presented supporting the

view that the alignment of the chains is also parallel in cellulose II. Other electron microscopic studies using the methods of lattice imaging have been used to demonstrate that the highly ordered microfibrils derived from algal celluloses represent homogeneous lattice structures with respect to the diffraction planes defined by the organization of their heavy atoms.

Electron diffraction studies carried out on algal celluloses after the discovery of the I_α/I_β duality have been interpreted to indicate that the two forms may alternate along the length of individual microfibrils. These observations can also be interpreted as manifestations of the slow twisting about the long axis that has been observed in other studies of similar algal celluloses.

The possibility of the coexistence of the I_α and I_β forms within a superlattice structure has been suggested in the context of studies intended to mimic the conditions of biogenesis. These need to be examined in greater detail in relation to the discussions of native celluloses and of their biogenesis.

Our discussion of structure has focused so far on issues of structure at the nanoscale level, identified as corresponding to domains that are of the order of 2 nm in dimension. Organization at the next level in the hierarchy of structure, defined as the range between 2 and 50 nm, requires consideration of a number of issues that have not been dealt with adequately in the literature on structures of cellulose. These include the well-recognized departures from a linear lattice, which have been generally regarded as measures of disorder when in fact they are more appropriately regarded as indicators of the nonlinear organization in a biological structure. Another issue arising at this level is associated with the occurrence of significant fractions of the cellulose molecules at the surface of the microfibrils of most native celluloses, particularly in the case of higher plants. It is the question as to whether the nanofibrillar structure can be viewed as a separate phase in the traditional sense and whether the criteria developed for the stability of homogeneous phases in the context of classical thermodynamics can have meaning when applied to native cellulosic structures at the nanoscale level. These issues arise in relation to discussions of native celluloses and their biogenesis.

Finally, the new developments with respect to structure, which facilitate separate though complementary focus on secondary and tertiary structures, provide a basis for exploring the relationships between structure and reactivity in new ways. The framework for more detailed characterization of the states of aggregation should permit more systematic exploration of the influence of source and history of cellulosic samples on their entry into chemical and biological processes.

6.16.11 Cellulose Chemistry

6.16.11.1 Introduction

Cellulosic substances have been used in various fields from commodities to industrial materials after mechanical and chemical modifications. Especially because chemically modified celluloses have some unique functional properties, and also because of their biodegradability in most instances, the chemistry of cellulose has become one of the major areas in cellulose science. **Figure 12** illustrates the chemical structure of cellulose in terms of chemical modifications.¹⁰⁰ Three hydroxyl groups in the glucose residue, one primary and the other two secondary, are the sites for substitution reactions, which are the most common in cellulose derivatizations. The (1→4)- β -glycoside bonds and other functional groups such as carboxyls and aldehydes present in most cellulosic material as minor groups are also possible sites for chemical modifications.

6.16.11.2 Solvents

Dissolution of cellulose has three major purposes. The first is to prepare regenerated and man-made cellulose fibers or films from cellulose solutions at the industrial level. Environmentally friendly and cost-profitable systems to dissolve and regenerate cellulose are now required. The second purpose is to use cellulose solutions as homogeneous reaction media during chemical modifications, which have been investigated at the laboratory level. The last one is to analyze cellulose samples. Molecular mass and molecular mass distribution studies using cellulose solutions are included in this category. Numerous cellulose solvents have, therefore, been developed and studied for these purposes.

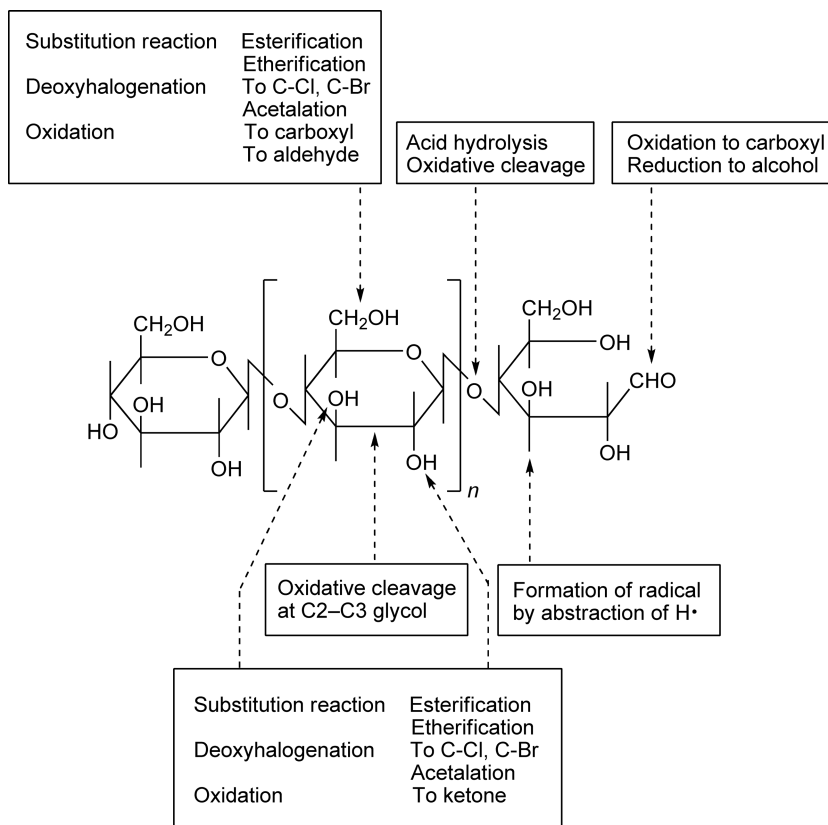


Figure 12 Positions in cellulose structure for chemical reactions.

Table 1 summarizes some representative solvents or solvent systems for cellulose. The xanthate system has been used for a long time to make viscose rayon and sponge. However, rayon production, where H_2S is released to varying amounts during the recovery system of spent solvent, is shrinking because of environmental issues. On the other hand, cellulose is soluble in aqueous NaOH alone under limited conditions. Microcrystalline cellulose with DP 200–300 is soluble in 6–9% NaOH by freezing and defrosting procedures.¹⁰¹ Pretreatments of cellulose such as steam explosion are necessary for complete dissolution in the aqueous NaOH for normal bleached wood pulp and cotton linters with DP of more than 500.¹⁰²

Many nonaqueous cellulose solvent systems have been developed in the past three decades. The solvent systems consist of reagent(s) reactive at cellulose hydroxyl groups and a polar aprotic solvent such as dimethylsulfoxide (DMSO) or *N,N*-dimethylacetamide (DMAc). Trifluoroacetic acid is the only volatile solvent to dissolve cellulose.¹⁰³ Cellulose hydroxyls form unstable derivatives or complexes with the solvent component(s) in the solutions (**Figure 13**). These nonaqueous cellulose solvents were used to prepare cellulose derivatives under homogeneous conditions mainly to control the degree of substitutions (DS) and distribution of substituents. The LiCl/DMAc system dissolves cellulose by one of the following two procedures: (1) solvent exchange of cellulose soaked in water to DMAc through ethanol, followed by stirring in 8% LiCl/DMAc at room temperature and (2) heating of cellulose/DMAc suspension at approximately 165°C for 30 min and LiCl is added to the cellulose suspension at about 100°C in the course of cooling to adjust to 8% LiCl/DMAc .¹⁰⁴ Because cellulose solutions in LiCl/DMAc are fairly stable and the powder-like LiCl is easy to handle, these solvent systems have been extensively studied as reaction media for homogeneous derivatizations of cellulose. Furthermore, the LiCl/DMAc solvent system has been used as an eluent in size-exclusion chromatographic analyses using a multiangle laser light scattering (SEC-MALLS) detector for molecular mass and molecular mass distribution of cellulose samples.¹⁰⁵

Table 1 Conventional and new cellulose solvent systems

| Category | Solvent | Remarks |
|-------------------------------|--|---|
| Acid | >52% H ₂ SO ₄ | Partial hydrolysis |
| | >85% H ₃ PO ₄ | Partial hydrolysis |
| Alkali | 6% LiOH | Needs pretreatment of cellulose |
| | 6–9% NaOH | Needs pretreatment of cellulose |
| Alkaline metal | Cu(NH ₃) ₄ (OH) ₂ (cuoxam) | Cuprammonium rayon production system |
| Complex | Cu(H ₂ NCH ₂ CH ₂ NH ₂) ₂ (OH) ₂ (cuen) | Standard solvent for DPv measurement |
| | Co(H ₂ NCH ₂ CH ₂ NH ₂) ₂ (OH) ₂ | |
| | Ni(NH ₃) ₆ (OH) ₂ | |
| | Cd(H ₂ NCH ₂ CH ₂ NH ₂) ₃ (OH) ₂ (cadoxen) | |
| | Zn(H ₂ NCH ₂ CH ₂ NH ₂) ₃ (OH) ₂ | |
| Alkaline Derivatives | Fe/3(tartaric acid)/3NaOH (EWNN) | Relatively stable |
| | CS ₂ /NaOH | Dissolves cellulose, forming xanthogenate Viscose rayon production system |
| Inorganic salt | >64% ZnCl ₂ | Dissolves cellulose by heating at 100 °C |
| | >50% Ca(SCN) ₂ | Dissolves cellulose by heating at 100 °C |
| Organic solvent Systems | Cl ₃ CHO/DMF | Dissolves cellulose, forming chloral |
| | (CH ₂ O) _x /DMSO | Hemiacetals at all cellulose-OH Dissolves cellulose, forming (poly)-methylol hemiacetals at cellulose-OH |
| | N ₂ O ₄ /DMF, N ₂ O ₄ /DMSO | Dissolves cellulose, forming nitrite ester at all cellulose-OH |
| | LiCl/DMAc, LiCl/DMI | Stable; needs pretreatment of cellulose |
| | SO ₂ /amine/DMSO | Unstable; gives stable amorphous regenerated cellulose |
| | CH ₃ NH ₂ /DMSO | Dissolves cellulose, forming complex |
| | CF ₃ COOH (TFA) | Dissolves cellulose, forming TFA ester at C6-OH; volatile solvent |
| | (Bu) ₄ N ⁺ F ⁻ ·3H ₂ O/DMSO | Dissolves cellulose with DP <650 |
| | ~80% NMMO/H ₂ O | Uses for lyocell production Dissolves cellulose by heating at 90 °C |
| | Others | NH ₄ SCN/NH ₃ /water |
| N ₂ H ₄ | | Explosive |

Bu, butyl-; DMAc, *N,N*-dimethylacetamide; DMF, *N,N*-dimethylformamide; DMI, *N,N*-dimethylimidazolidinone; DMSO, dimethylsulfoxide; DPv, degree of polymerization as measured by viscometry; NMMO, *N*-methylmorpholine-*N*-oxide; TFA, trifluoroacetic acid.

N-Methylmorpholine-*N*-oxide (NMMO) containing approximately 20% water can dissolve cellulose by melting of NMMO at about 90 °C and has recently been used to make regenerated cellulose fiber (lyocell) at the industrial level. NMMO has a tendency to be an oxidant, and therefore some antioxidant must be added to the solution before spinning in order to avoid strong coloration and oxidative degradation. It is possible to insert an air-gap stage in between spinning nozzles and the water bath for regeneration, and thus this system is called a ‘dry-wet spinning process.’ Drawing can be achieved in the air-gap stage, resulting in a higher degree of crystallinity as well as a higher degree of orientation of crystalline domains in the regenerated cellulose fiber prepared thereby (Table 2).¹⁰⁶

6.16.11.3 Derivatization

Because cellulose has three different hydroxyl groups at C2, C3, and C6 in the anhydroglucose unit (Figure 12), various chemical reactions can be applied. Some cellulose derivatives have been manufactured at an industrial level and used in various fields as functional polymers. Factors influencing the characteristics of cellulose derivatives are the following: (1) chemical structures of substituents introduced, (2) degree of substitution (DS: the amount of substituents per anhydroglucose unit), (3) distribution of substituents, (4) degree of polymerization and its distribution, and (5) impurities including the presence of chromophores and others.

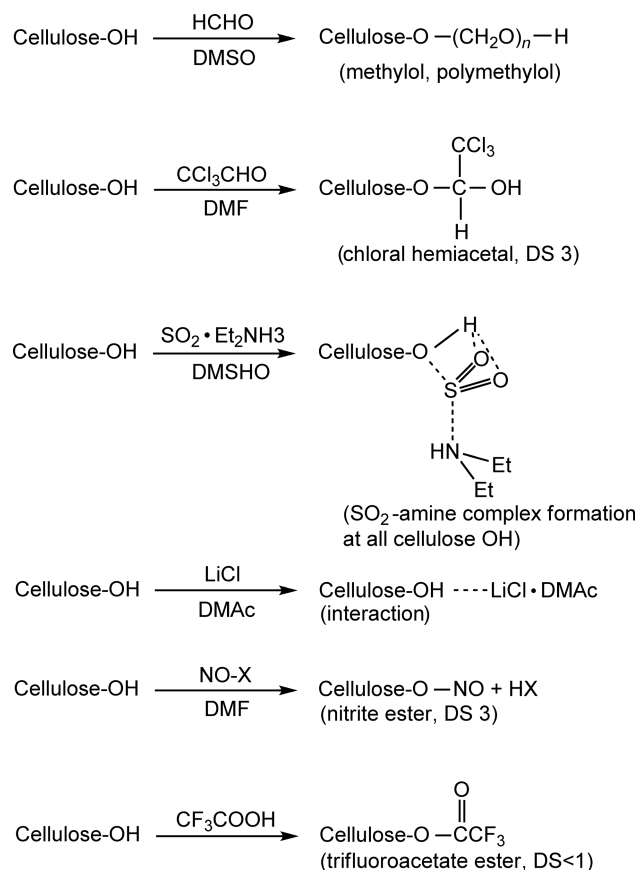


Figure 13 Dissolving states of cellulose in nonaqueous cellulose solvents.

Table 2 Crystallinity index and degree of orientation of crystals of regenerated cellulose fibers calculated from their X-ray diffraction patterns

| Cellulose solvent | Crystal structure | Crystallinity index (%) | Degree of orientation of crystals (%) |
|--|-------------------|-------------------------|---------------------------------------|
| NaOH/CS ₂ /water (viscose rayon) | Cellulose II | 24 | 85 |
| Aqueous Cu(NH ₃) ₄ (OH) ₂ (cuprammonium rayon) | Cellulose II | 41 | 90 |
| 6–9% Aqueous NaOH | Cellulose II | 46 | 75 |
| 80% NMMO/water (lyocell) | Cellulose II | 46 | 91 |

6.16.11.3.1 Esterification

Typical cellulose esters examined so far are depicted in **Figure 14**. Cellulose acetate and cellulose nitrate are the representative cellulose organic and inorganic esters, respectively, produced at the industrial level. Cellulose triacetate (DS > 2.9) where DS is the degree of substitution, is prepared by heating cellulose suspended in a mixture of acetic anhydride/acetic acid/H₂SO₄ around 60 °C. The cellulose suspension becomes a clear solution as cellulose triacetate is formed in the mixture. Cellulose triacetate is soluble in chlorinated hydrocarbons such as dichloromethane. Cast films of cellulose triacetate have characteristic optical properties that eliminate polarization of penetrating light and thus have been used as film bases for photographs and supporting films for liquid crystal displays. Cellulose diacetate (DS 2.3–2.5) is prepared consecutively from the cellulose triacetate solution by diluting with water and heating. Cellulose diacetate is soluble in acetone and can be formed into fibers or films. Cellulose diacetate films have been used for ultrafiltration to purify tap water partly in place of chlorination.

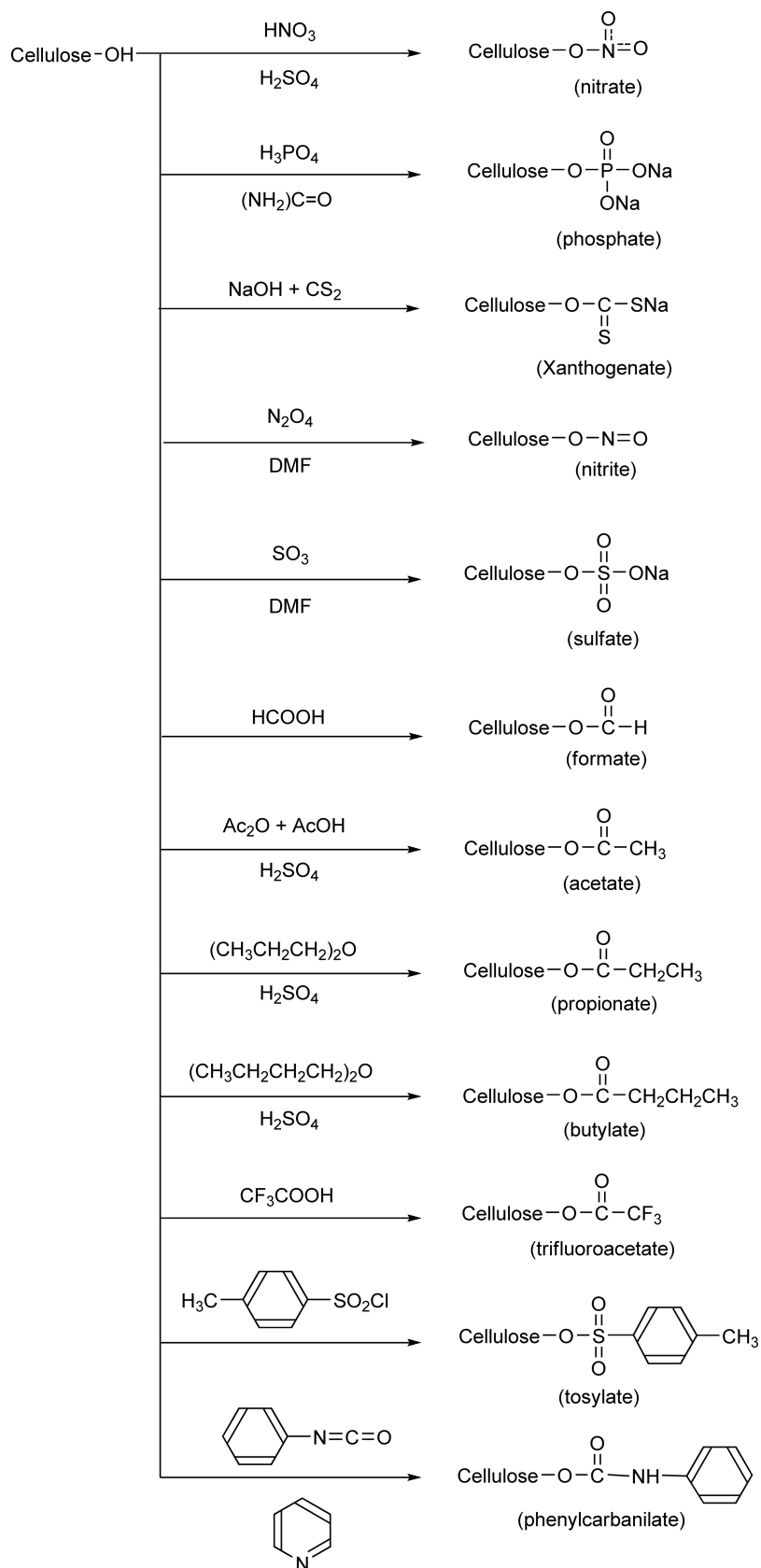


Figure 14 Typical cellulose esters.

For the manufacture of cellulose derivatives it is necessary to use high α pulps. α is a common measure of the purity of a pulp with respect to other cell wall polysaccharides, the hemicelluloses that may not have been removed during the pulping process. It is defined as the percentage of the total weight of the pulp that remains after extraction with 18% NaOH aqueous solution; it is usually above 95% for pulps intended for the production of derivatives. When pulps having lower α -cellulose contents are used, some acetone-insoluble gel fractions originating from hemicelluloses are formed from both softwood and hardwood bleached kraft pulps.¹⁰⁷ Thus, bleached sulfite pulp or at least bleached prehydrolyzed kraft pulp prepared from softwood is acceptable as the pulp resources at this point. Therefore, the ability to use normal bleached kraft pulp produced from eucalyptus without any spinning problems is one of the significant themes for cellulose diacetate production. Furthermore, one-step production of cellulose diacetate without going through the cellulose triacetate stage is also desired.

Various cellulose esters such as acetate, tosylate (*p*-toluenesulfonate), sinnamoylate, and fluorine-containing substituents have been prepared with pyridine as a base under homogeneous and nonaqueous conditions using, for example, 8% LiCl/DMAc. The DS and sometimes the distribution of substituents can be well controlled by these methods. Although positive results were obtained in laboratory studies, none of these nonaqueous cellulose solvent systems have been used as derivatization media at the industrial level. The multicomponent solvent systems, the high boiling point of solvents such as DMAc and DMSO, the necessity of pretreatments including complete drying of cellulose, and solvent-exchange processes for dissolution make it difficult to adapt the laboratory processes involving nonaqueous solvents for industrial application in the production of cellulose derivatives.

6.16.11.3.2 Etherification

Figure 15 illustrates representative cellulose ethers. In most cases in industrial operations, cellulose ethers are produced through alkal cellulose (cellulose swollen with, e.g., 18% aqueous NaOH) by reacting with etherifying reagents around 60 °C in the presence of *i*-propanol or *i*-butanol, where etherification proceeds heterogeneously to the swollen alkal cellulose without dissolution. Washing and purification can, therefore, be carried out easily by filtration. Because the presence of water in the reaction media essentially precludes the etherification, it is quite difficult to achieve DS values close to 3 by one step as long as the aqueous alkal cellulose system is adopted.

The sodium salt of carboxymethylcellulose (CMC), methylcellulose (MC), hydroxyethylcellulose (HEC), and hydroxypropylcellulose (HPC) are typical water-soluble cellulose ethers manufactured at an industrial scale and used primarily as thickeners in various fields. Commercial CMC has DS values in the range of 0.6–1.2. Repetition of carboxymethylation (at least three times containing washing stage in each) gives CMC with DS close to 3. HEC and HPC are produced from alkal cellulose by reacting with ethylene oxide and propylene oxide, respectively. In the case of these cellulose ethers, additional substitution also occurs on hydroxyl groups of the introduced substituents, like grafting, by increasing the amount of substituents, and thus molecular substitution (MS) in place of DS is used for these cellulose ethers. Water-soluble cellulose ethers in solution states have been characterized by SEC-MALLS, and the occurrence of coagulation among cellulose ether molecules in water under particular conditions has been reported.¹⁰⁸ When a small amount of long alkyl chains (C₁₂–C₂₄) is introduced into HEC, the viscosity of aqueous solutions increases to very high levels through the formation of hydrophobic interactions among HEC molecules in water.¹⁰⁹

Cellulose ethers having new substituents, those with high DS or particular DS, and those having regioselective substitution have been prepared under various conditions at laboratory scale sometimes using nonaqueous cellulose solvent systems. Nonaqueous media are advantageous to prepare cellulose ethers with high DS, which generally cannot be achieved by the conventional aqueous alkal cellulose system. About 30 kinds of cellulose ethers with DS 3 were prepared by one-step reactions with powdered NaOH and etherifying reagents using the cellulose solution in SO₂/diethylamine/DMSO.¹¹⁰ Triphenylmethyl (trityl) group can be introduced selectively at C6 primary hydroxyl of cellulose by homogeneous reaction with pyridine in 8% LiCl/DMAc. This tritylcellulose was used as an intermediate for further conversion to some regioselective cellulose ethers and esters such as 2,3-di-*O*-methylcellulose and 6-*O*-methylcellulose (**Figure 16**).¹¹¹

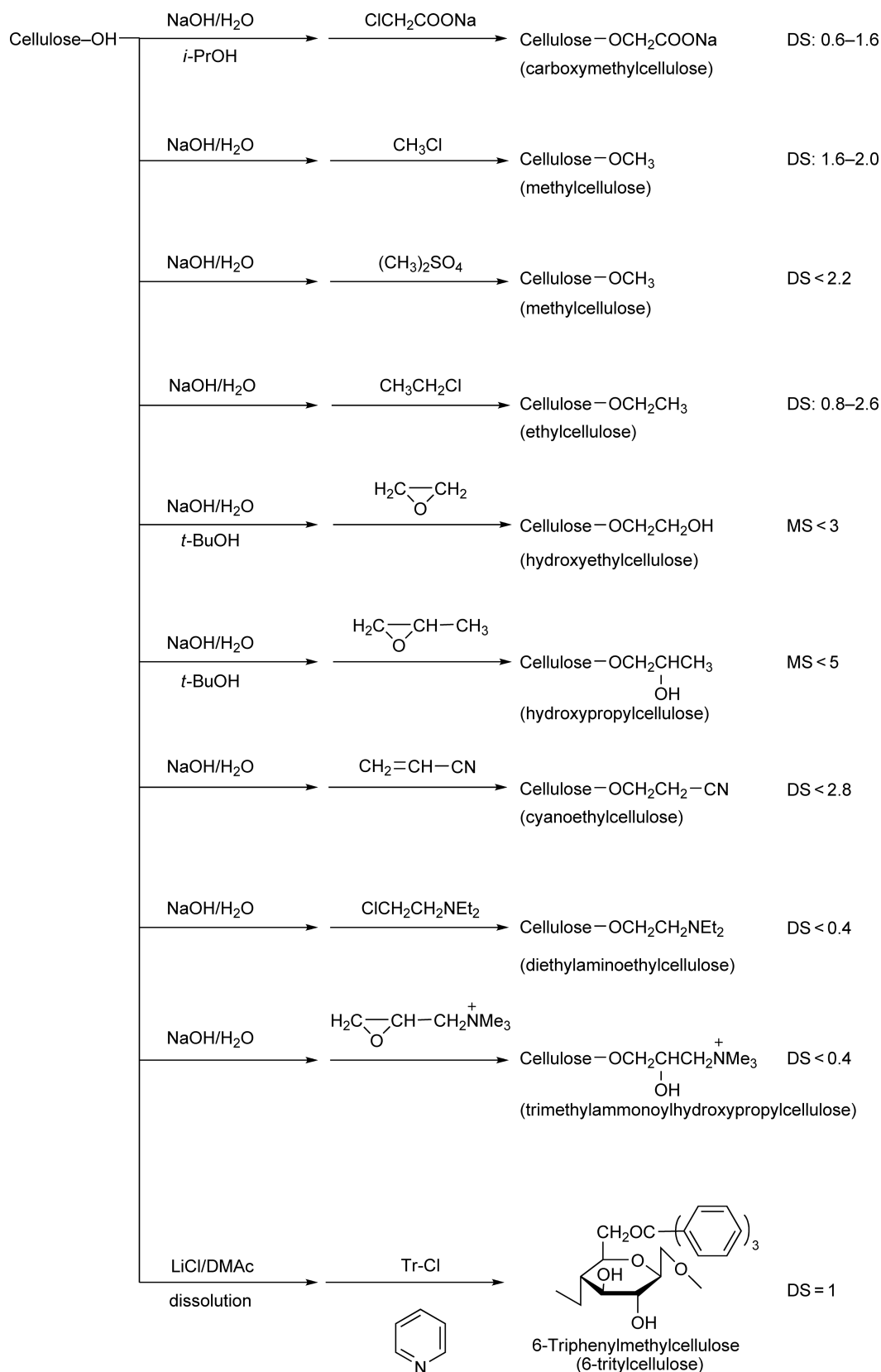


Figure 15 Typical cellulose ethers.

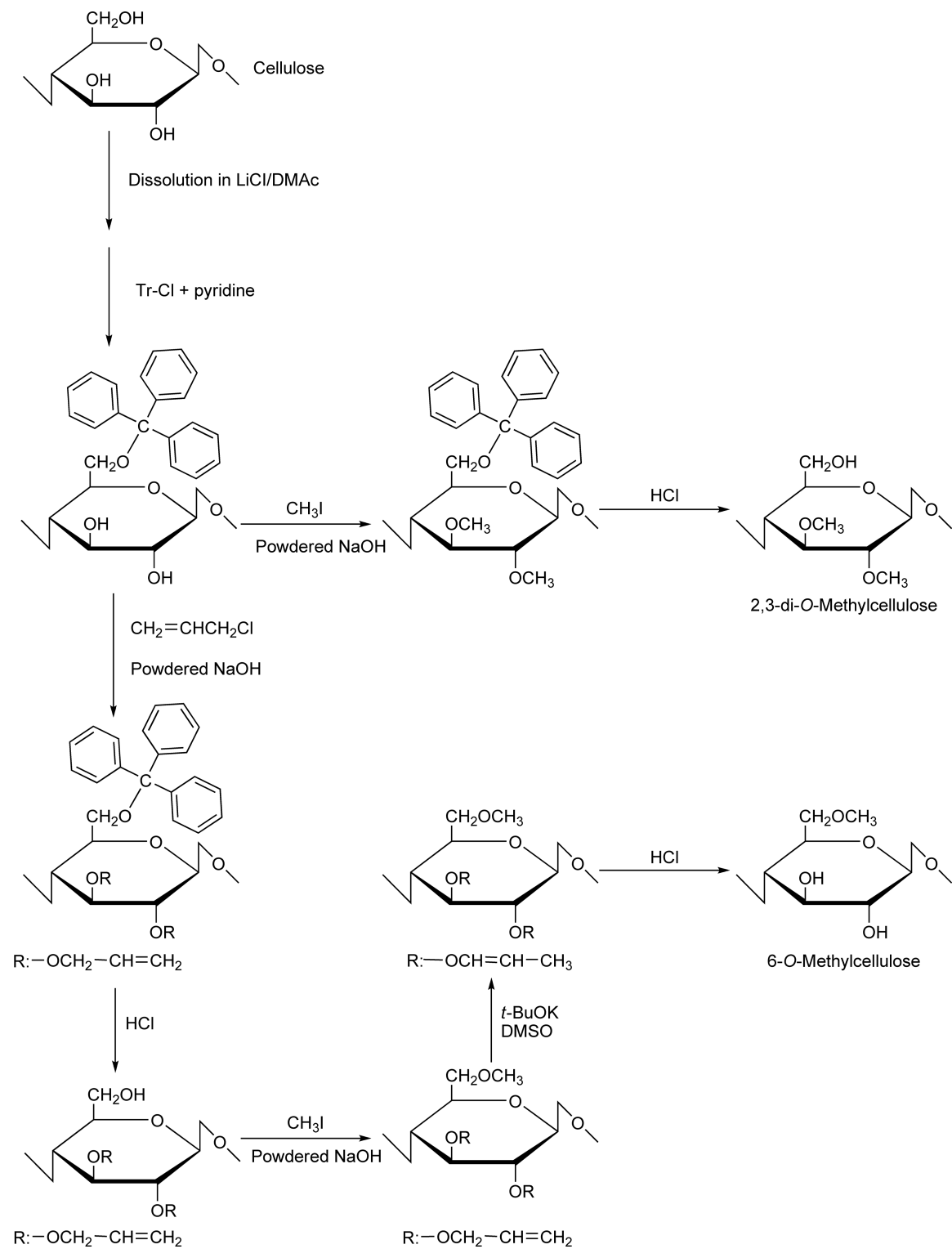


Figure 16 Preparation scheme of regioselectively substituted methylcelluloses via tritylcellulose.

6.16.11.3.3 Oxidation

There are several methods to modify the chemical structure of cellulose by oxidation. Periodate oxidation of cellulose suspended in water and N_2O_4 oxidation of cellulose suspended in chloroform are well-known conventional methods to convert cellulose to dialdehyde cellulose and C6-carboxy cellulose, respectively. Generally, however, some side reactions including depolymerization are inevitable during these oxidations, and it is difficult, therefore, to achieve completely regioselective oxidation.

TEMPO (2,2,6,6-tetramethylpiperidine-1-oxy radical) is a water-soluble and commercially available radical. When sodium hypochlorite is used as a cooxidant in the presence of catalytic amounts of NaBr and TEMPO in water, C6 primary hydroxyl groups of polysaccharides dissolved in water at pH 10–11 can selectively be converted to carboxyl groups.¹¹² When regenerated, mercerized, and also ball-milled native celluloses are suspended in water and are used as the starting materials, water-soluble products are obtained quantitatively at room temperature within 1 h (mostly within 20 min). NMR analyses revealed that these water-soluble oxidized products have homogeneous chemical structures, (1→4)- β -linked polyglucuronic acid (Figure 17).¹¹³ Thus, selective oxidation at C6 primary hydroxyl groups of cellulose can be achieved by the TEMPO-mediated oxidation in aqueous media (Figure 18), although remarkable depolymerization is inevitable.¹¹⁴ This new water-soluble polyglucuronic acid is degradable by endo- and exo-type lyases.¹¹⁵

On the other hand, when the TEMPO-mediated oxidation is applied to native celluloses, the C6 primary hydroxyls only on the surfaces of cellulose microfibrils are oxidized to carboxylate groups, maintaining the original crystallinity indices and the crystal sizes of cellulose I allomorph.^{116,117} Highly viscous and transparent gels are obtained by mild disintegration of the TEMPO-oxidized native celluloses in water, when the contents of carboxylate groups formed are higher than 1.2 mmol g^{-1} for wood celluloses. The gels consist of individual fibrils 3–5 nm in width and at least a few micrometers in length dispersed in water (Figure 19(a)).¹¹⁸ Transparent and high gas-barrier films were obtained by casting of the nanofiber gels on a plate (Figure 19(b)).¹¹⁷

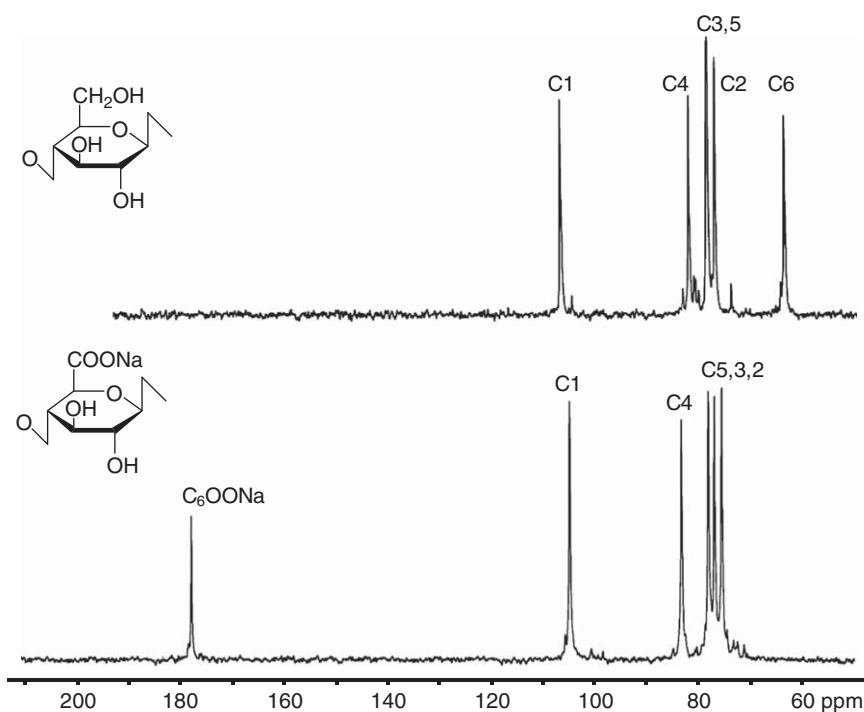


Figure 17 ^{13}C NMR spectra of cellulose oligomer (DP 7) in DMSO and cellouronic acid Na salt, (1→4)- β -linked polyglucuronate, in D_2O . Cellouronic acid Na salt was prepared from viscose rayon by the TEMPO-mediated oxidation.

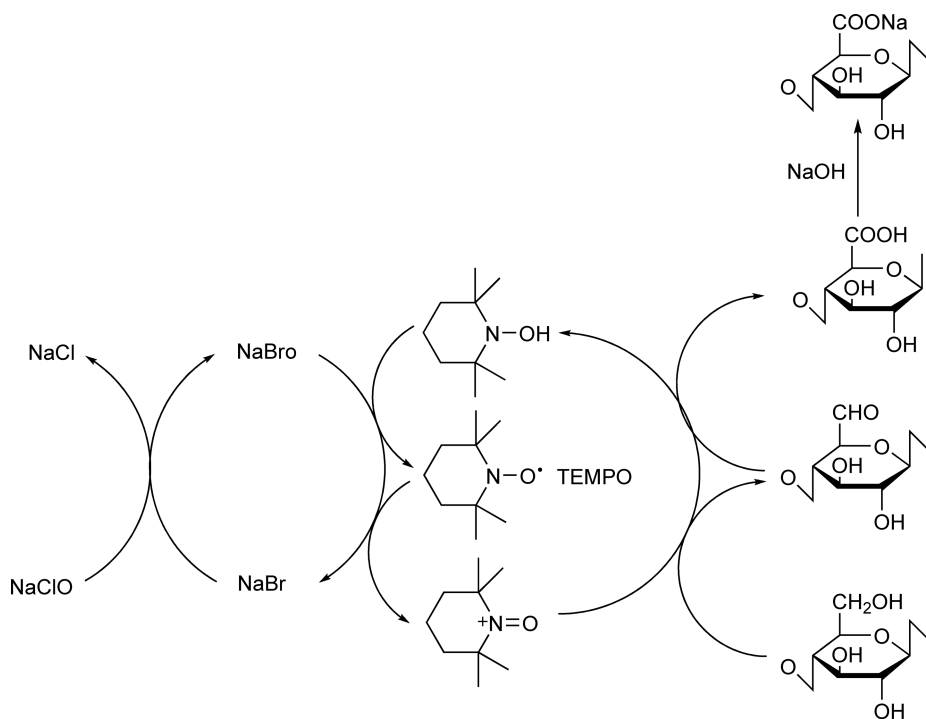


Figure 18 TEMPO-mediated oxidation of C6 primary hydroxyl group of cellulose to carboxyl group.

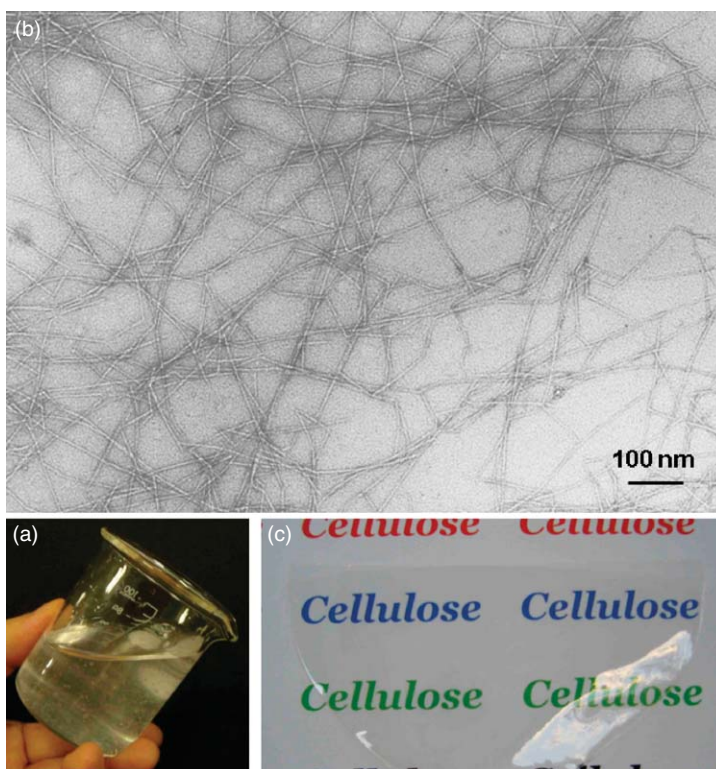


Figure 19 A gel of TEMPO-oxidized cellulose nanofibers dispersed in water (a), transmission electron micrograph of TEMPO-oxidized cellulose nanofibers (b), and a film prepared from the nanofibers (c).

6.16.11.4 Degradation

Cellulosic materials undergo numerous and sometimes harsh stimuli during manufacturing processes and under circumstances in use. Qualities of cellulosic materials are then decreased through degradation by these outside stimuli. If these degradations can be well controlled, on the other hand, useful cellulose-related compounds can be obtained.

6.16.11.4.1 Acid hydrolysis

Because (1→4)- β -glycoside bond is among the acetals, which are unstable under acid conditions, cellulose is more or less susceptible to acid hydrolysis. Degradation of acidic paper during storage is caused by the gradual hydrolysis of cellulosic fibers by residual sulfuric acid in paper. Various gas-phase neutralization treatments have, therefore, been applied to library books consisting of acidic paper. On the other hand, acid hydrolysis of cellulosic biomass can be used to produce glucose, which is then converted to ethanol by fermentation. Susceptibility of cellulosic materials to acid hydrolysis is remarkably different between their ordered and disordered regions. When native cellulosic materials such as cotton linters and bleached chemical wood pulps are heated in a dilute acid, hydrolysis of disordered regions in cellulose microfibrils precedes that of ordered regions to form the so-called 'microcrystalline cellulose' with DP 200–300. Microcrystalline cellulose has been manufactured at an industrial scale and used in medicinal tablets, additives for foods and cosmetics, and others. A part of glucose once formed from cellulose by acid hydrolysis is further degraded to hydroxymethylfurfural, levulinic acid, formic acid, and others during acid hydrolysis (Figure 20).

6.16.11.4.2 Enzymatic degradation

Cellulases hydrolyze cellulose under mild conditions compared to inorganic or organic acid. Generally, cellulases like CBH II consist of core and cellulose-binding domains and a linker that binds the two domains. The core domain contains an active center to hydrolyze cellulose in a catalytic manner and the subsites that interact with cellulose chain close to the active center. The cellulose-binding domains consist of amino acids having aromatic rings such as tyrosine or tryptophan, and these aromatic rings of the cellulose-binding domains attach to hydrophobic plains of cellulose chains through van der Waals force. The active center of the core domains of cellulases can then attack the cellulose chain. The subsites of the core domains can give some mechanical stress to the cellulose chain, and one of the glucose residues of the cellulose chain is forced to have the unstable boat form. Thus, remarkable reduction of activation energy to hydrolyze cellulose can be achieved.¹¹⁸

Various types of cellulases have been reported so far, and they have been known for some time to be of two types, exo- and endo-cellulases, depending on whether or not the cellulase can recognize the reducing ends of cellulose chains. CBH and endoglucanase (EG) are then classified into two types, respectively. However, recent studies revealed that there are no exo-type cellulases and that all cellulases are included in the endo-type cellulases. On the other hand, the following classification is now well accepted: the processive and nonprocessive cellulases on the basis of hydrolysis patterns of cellulose chains (Figure 21).¹¹⁹ The processive type can successively hydrolyze cellulose to form cellobiose without separating from the cellulose chain by sliding stepwise along the chain. In the case of nonprocessive cellulases, on the other hand, the active center of the cellulase is detached from the cellulose chain after each hydrolysis. On the basis of this new classification, the CHB should be classified as an endo-type processive enzyme, whereas EG is an endo-type nonprocessive enzyme.

6.16.11.4.3 Thermal degradation

When cellulose is analyzed by thermogravimetry under a nitrogen atmosphere, thermal decomposition starts at 200–270 °C, depending on the purity of the cellulose sample. The temperature of ignition in air is in the range of 390–420 °C, and the maximum flame temperature can reach 800 °C or more. When cellulose is heated at temperatures of more than 100 °C under reduced pressure, levoglucosan (Figure 19) is obtained at a maximum yield of 70%. When thermal carbonization is applied to cellulosic materials under nonreactive gas atmosphere, the yields of carbon are lower than the theoretical value (44%) because some of the cellulose is converted to levoglucosan. Yields of carbon can be increased by adding hydrochloric acid, which enhances dewatering rather than the formation of levoglucosan.¹²⁰ When cellulose microcrystals, which are obtained from crystalline native celluloses by acid hydrolysis, are carbonized under suitable conditions, carbon nanorods are obtained.¹²¹

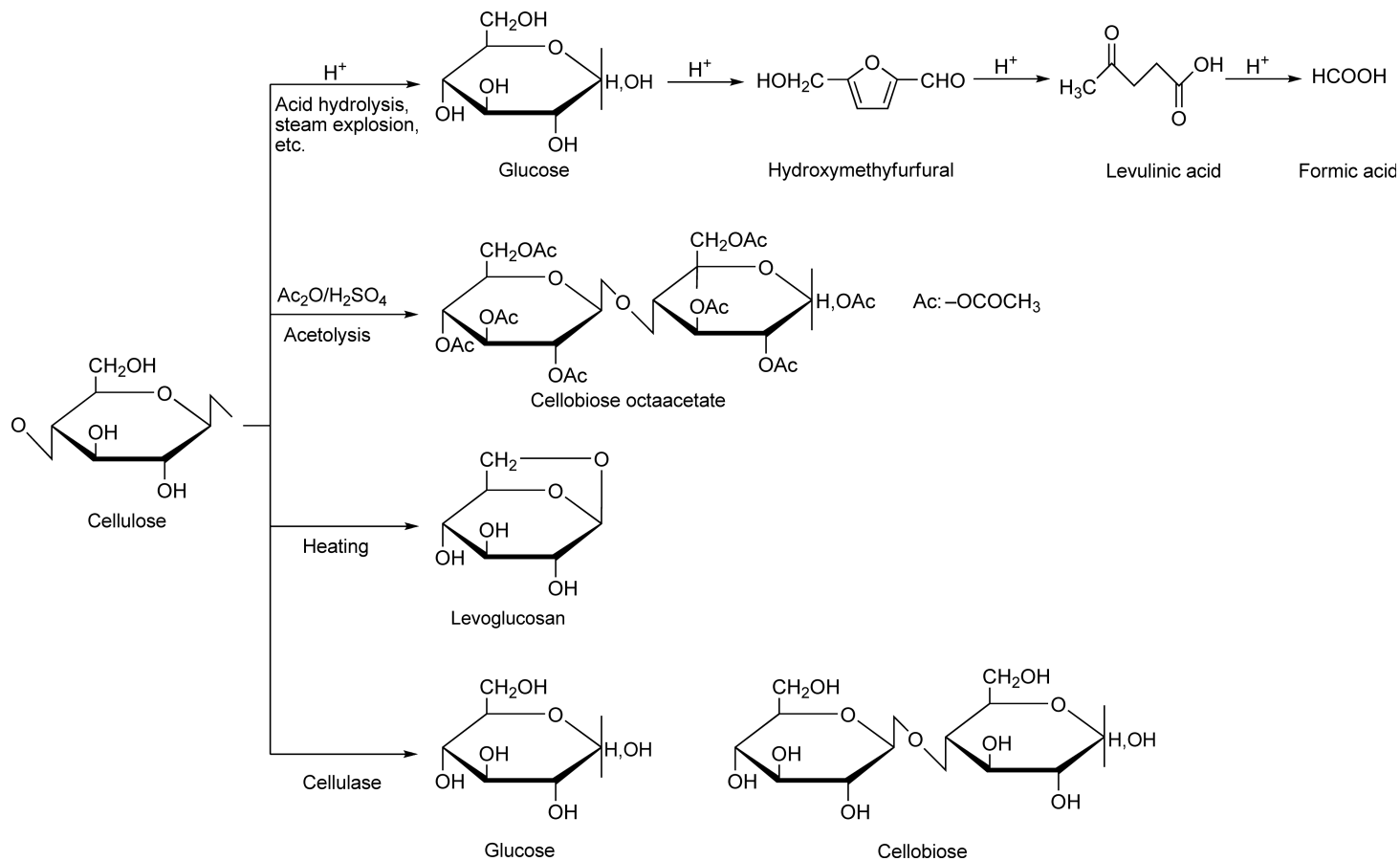


Figure 20 Some degradation products of cellulose.

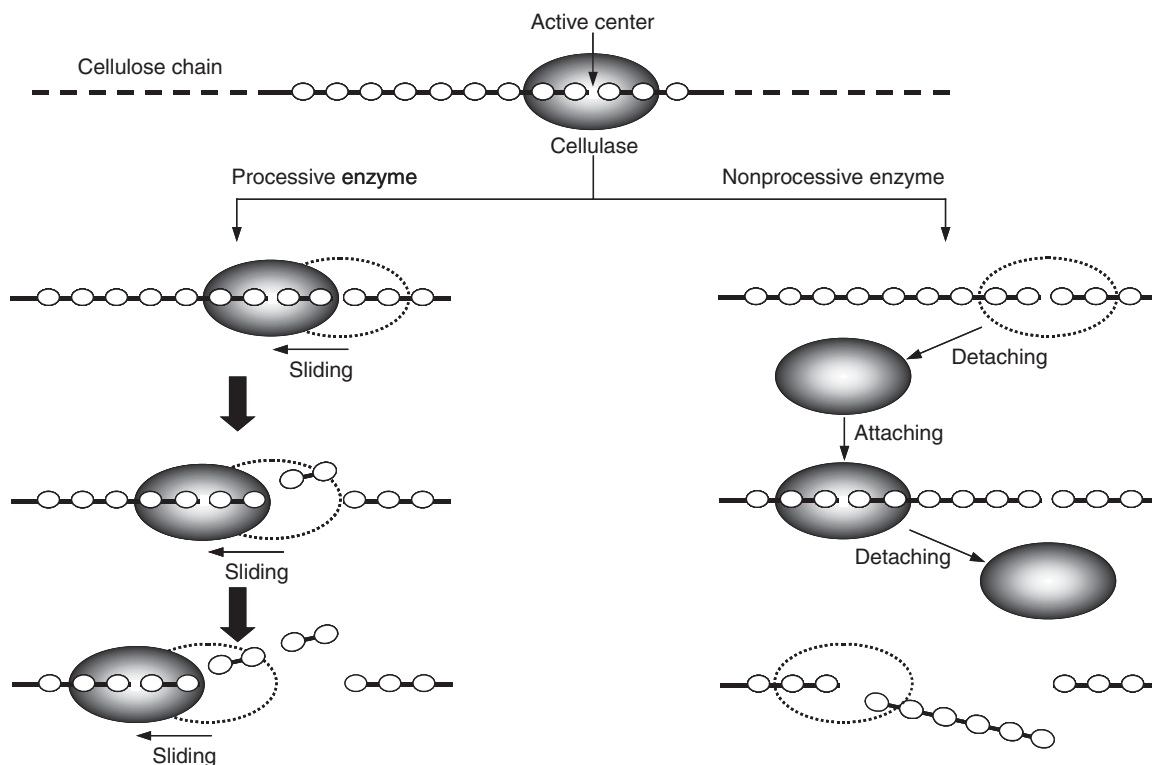


Figure 21 Classification of cellulase into processive and nonprocessive types.

Microwave treatment is one of the heating methods and provides more homogeneous heating when compared to heating that relies on the thermal conduction. Saccharification of lignocellulosics has been studied in this manner, and the maximum yield of glucose was reported to be 81% based on the theoretical value.¹²² Explosion while dispersed in a compressed fluid and laser irradiation treatments are also types of thermal treatments, and glucose and levoglucosan are obtained by these treatments. Supercritical water treatment is also one of the heating methods in the presence of acid. Dissociation of water can increase under supercritical conditions, and thus water behaves like an acid catalyst with respect to cellulose. Noncatalytic hydrolysis of cellulose can, therefore, be achieved by supercritical water treatments. The maximum yields of glucose were reported to be in the range of 32–48%.¹²³

6.16.11.5 Chemical and Enzymatic Syntheses of Cellulose

Although cellulose is one of the naturally occurring polysaccharides, artificial synthesis of cellulose from monomers or dimers has been of scientific interest for a long time. Syntheses of cellulose have recently been successful on the basis of the following two different principles.

Cellulase hydrolyzes the (1→4)- β -glycoside bonds of cellulose, and this enzymatic hydrolysis is essentially reversible. Kobayashi *et al.*¹²⁴ succeeded in cellulose synthesis using the reversible reaction of cellulase, where a particular substrate, β -cellobiosyl fluoride, was used in an aqueous buffer containing acetonitrile (Figure 22).

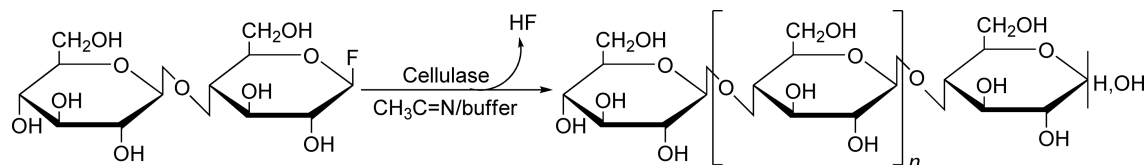


Figure 22 Enzymatic synthesis of cellulose from β -cellobiosyl fluoride in acetonitrile/aqueous buffer mixture.

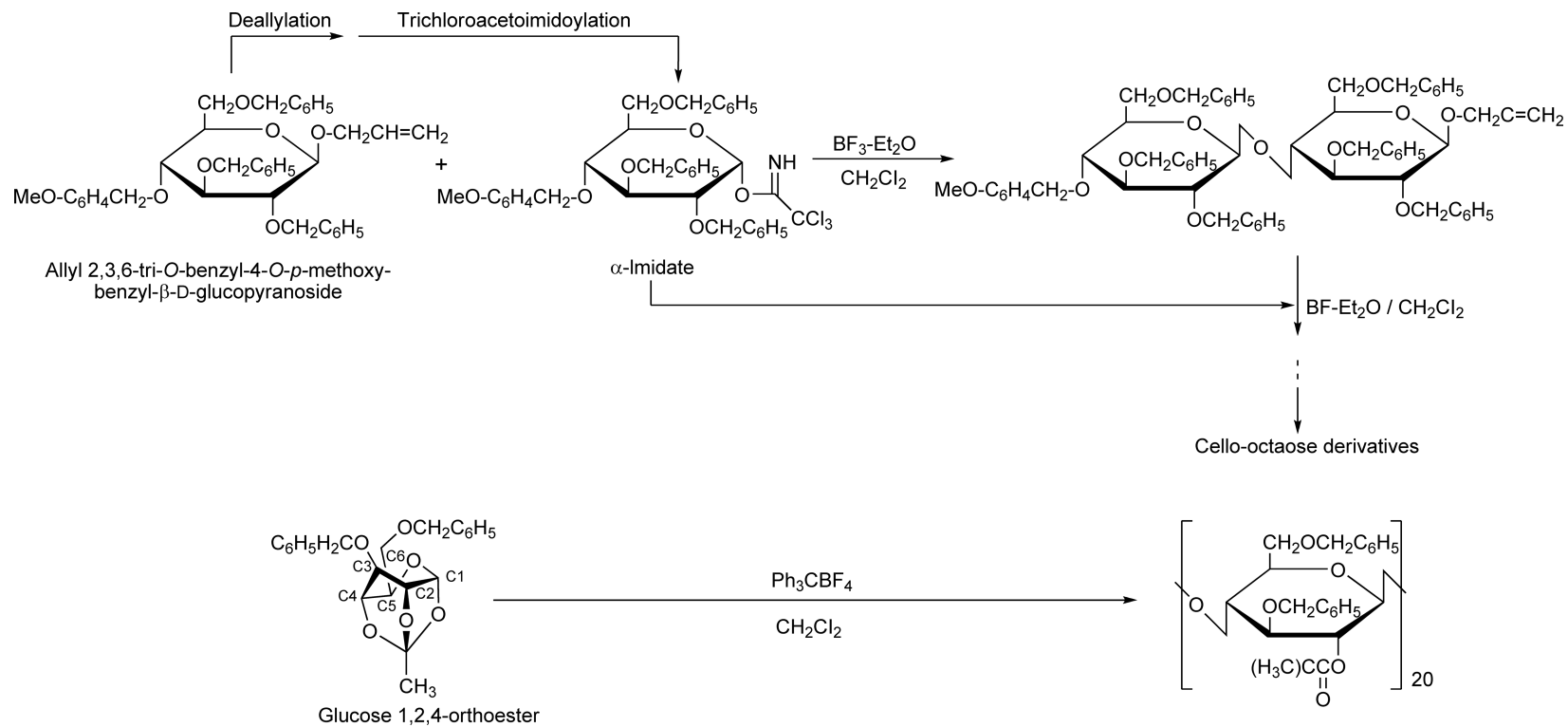


Figure 23 Chemical syntheses of cellulose.

The origin and purity of cellulase and the combination of the solvent systems influenced the yields of synthesized cellulose and its crystal structure. Cellulase from *Trichoderma viride* gave the highest yield, ~54%, of water-insoluble low-molecular-weight cellulose (DP 22), whereas β -glucosidase gave no cellulose. Enzymatically synthesized cellulose generally has the crystal structure of cellulose II, which is the same as that of regenerated or mercerized cellulose. On the other hand, cellulose having the crystal structure of cellulose I, which is the same as that of native cellulose, was synthesized *in vitro* by using a highly purified cellulase component.¹²⁵

Cellulose oligomers and low-DP cellulose have been prepared by the following three routes and the successive elimination of the protecting groups at the hydroxyl groups: (1) linear synthetic reaction using the imidate method from allyl 2,3,6-tri-*O*-benzyl-4-*O*-*p*-methoxybenzyl- β -D-glucopyranoside, (2) confluent-type reaction between cellotetraose and cellooctaose derivatives using the imidate method, and (3) cationic ring-opening polymerization of glucose 1,2,4-orthoester (Figure 23). These studies revealed the roles of particular protecting groups at particular hydroxyl positions of the starting materials in regioselective reactions and conformational restrictions for polymerization of carbohydrate monomers.¹²⁶

Glossary

ball-milled cellulose When crystalline celluloses are subjected to ball milling using, for example, a planetary ball-milling apparatus, decrystallization occurs associated with partial depolymerization. Reactivity or solubility of cellulose can be improved by decrystallization through ball-milling treatment.

cellulose microfibril Crystalline elements in native celluloses, 3–10 nm in width, depending on the origins.

distribution of substituents Distribution of substituent groups between C2, C3, and C6 hydroxyl groups of cellulose, distribution of substituent groups along one cellulose chain, and distribution of substituent groups between cellulose chains.

glycoside bond Linkage type of anhydromonosaccharide units in polysaccharides and categorized in acetals: unstable and fairly stable under acid and alkaline conditions, respectively. Cleaved by acid or enzymatic hydrolysis.

hemicellulose Noncellulosic polysaccharides in plants. Glucomannan and glucuronoxylan are representatives of hardwood and softwood hemicelluloses, respectively.

kraft pulp Chemical pulp manufactured from wood chips in aqueous NaOH/NaSH solution at 150–170 °C under alkaline conditions by a continuous system. Primarily used for paper production. Energy and most of the chemicals are recyclable.

mercerized cellulose Native celluloses treated by swelling in 15–20% NaOH followed by washing with water to remove NaOH. Cellulose I crystal structure of native celluloses is transferred to cellulose II-type structure.

microcrystalline cellulose Dilute acid-hydrolyzed native celluloses with DP 200–300.

periodate oxidation Oxidation of glycol-type C–C bonds in water containing KIO₄ under weakly acidic conditions at room temperature, forming dialdehyde groups. 2,3-Dialdehyde cellulose is formed from cellulose.

regenerated cellulose Cellulose fibers and films prepared from once-dissolved states of cellulose in solutions.

steam explosion A process for fiberization of cellulosic biomass where it is subjected to high pressure steam and then forced through an orifice into an environment at ambient pressure; the sudden expansion of the steam results in fiberization.

sulfite pulp Chemical pulp manufactured from wood chips in aqueous sulfurous acid/sulfite salt solution at 120–140 °C under weakly acidic conditions by a batch system. Primarily used for production of regenerated cellulose fibers, films, and derivatives.

thermogravimetry Relationships between weight decrease and temperature by heating under air or inert gas atmosphere.

α -cellulose content A standard value for evaluation of cellulose purity in cellulosic materials without lignin using solubility in 17.5% NaOH at room temperature.

Abbreviations

| | |
|------------------|---|
| CBH | cellobiohydrolase |
| CMC | carboxymethylcellulose |
| CP | cross polarization |
| DMAc | <i>N,N</i> -dimethylacetamide |
| DMSO | dimethylsulfoxide |
| DP | degree of polymerization |
| DS | degree of substitutions |
| EG | endoglucanase |
| HEC | hydroxyethylcellulose |
| HPC | hydroxypropylcellulose |
| LALS | linked atom least squares |
| MAS | magic angle spinning |
| MC | methylcellulose |
| MS | molecular substitution |
| NMMO | <i>N</i> -methylmorpholine- <i>N</i> -oxide |
| NMR | nuclear magnetic resonance |
| PCA | principal component analyses |
| SEC-MALLS | size-exclusion chromatography analyses using a multi-angle laser-light scattering |
| STFI | Swedish Forest Products Laboratory |
| TEMPO | 2,2,6,6-tetramethylpiperidine-1-oxy |

References

1. R. H. Atalla, In *Comprehensive Natural Products Chemistry*; B. M. Pinto, Ed.; Elsevier: NY, 1999; Vol. 3, pp 529–598.
2. C. H. Haigler, In *Biosynthesis and Biodegradation of Cellulose*; C. H. Haigler, P. J. Weimer, Eds.; Marcel Dekker: New York, 1991; p 99.
3. S. J. Hanley; J.-F. Revol; L. Godbout; D. G. Grey, *Cellulose* **1997**, *4*, 209.
4. J. F. Matthews; C. E. Skopec; P. E. Mason; P. Zuccato; R. W. Torget; J. Sugiyama; M. E. Himmel; J. W. Brady, *Carbohydr. Res.* **2006**, *341*, 138–152.
5. K. H. Gardner; J. Blackwell, *Biopolymers* **1974**, *13*, 1975.
6. J. J. Hebert; L. L. Muller; *J. Appl. Polym. Sci.* **1974**, *18*, 3373.
7. K. H. Meyer; L. Misch, *Helv. Chim. Acta* **1937**, *20*, 232.
8. P. H. Hermans, *Physics and Chemistry of Cellulose Fibers*; Elsevier: New York, 1949.
9. A. French, In *The Structures of Celluloses*; R. H. Atalla, Ed.; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987; p 15.
10. Y. Nishiyama; H. Chanzy; P. Langan, *JACS* **2002**, *124*, 9074.
11. S. S. C. Chu; G. A. Jeffrey, *Acta Crystallogr.* **1968**, *B24*, 830.
12. J. T. Ham; D. G. Williams, *Acta Crystallogr.* **1970**, *B29*, 1373.
13. Y. Nishiyama; J. Sugiyama; H. Chanzy; P. Langan, *JACS* **2003**, *125*, 14300.
14. R. H. Atalla; J. C. Gast; D. W. Sindorf; V. J. Bartuska; G. E. Maciel, *J. Am. Chem. Soc.* **1980**, *102*, 3249.
15. W. L. Earl; D. L. VanderHart, *J. Am. Chem. Soc.* **1980**, *102*, 3251.
16. W. L. Earl; D. L. VanderHart, *Macromolecular* **1981**, *14*, 570.
17. D. L. VanderHart, *Deductions about the Morphology of Wet and Wet Beaten Cellulose from Solid State ¹³C NMR*. NBSIR 82-2534; National Bureau of Standards: Washington, DC, 1982.
18. C. A. Fyfe; R. L. Dudley; P. J. Stephenson; Y. Deslandes; G. K. Hamer; R. H. Marchessault, *J. Am. Chem. Soc.* **1983**, *105*, 2469.
19. F. Horii; A. Hirai; R. Kitamaru, *Polym. Bull.* **1982**, *8*, 163.
20. G. E. Maciel; W. L. Kolodziejewski; M. S. Bertran; B. R. Dale, *Macromolecular* **1982**, *15*, 686.
21. J. C. Gast; R. H. Atalla; R. D. McKelvey, *Carbohydr. Res.* **1980**, *84*, 137.
22. D. L. VanderHart; R. H. Atalla, *Macromolecular* **1984**, *17*, 1465.
23. R. H. Atalla; D. L. VanderHart, *Science* **1984**, *223*, 283.
24. R. H. Atalla, *Adv. Chem. Ser.* **1979**, *181*, 55.
25. G. Honjo; M. Watanabe, *Nature* **1958**, *181*, 326.
26. D. L. VanderHart; R. H. Atalla, In *The Structures of Celluloses*; R. H. Atalla, Ed.; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987; p 88.
27. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press: Ithaca, NY, 1960; p 65.
28. L. A. Woodward, *Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy*; Oxford: London, 1972; p 344.

29. H. A. Wells, The Vibrational Spectra of Glucose, Galactose and Mannose. Doctoral Dissertation, The Institute of Paper chemistry: Appleton, WI, 1977.
30. L. J. Pitzner, The Vibrational Spectra of the 1,5-anhydropentitols. Doctoral Dissertation, The Institute of Paper Chemistry: Appleton, WI, 1973.
31. L. J. Pitzner; R. H. Atalla, *Spectrochim. Acta* **1975**, 31A, 911.
32. G. M. Watson, The Vibrational Spectra of the Pentitols and Erythritol. Doctoral Dissertation, The Institute of Paper Chemistry: Appleton, WI, 1974.
33. S. L. Edwards, The Vibrational Spectra of the Pentose Sugars. Doctoral Dissertation, The Institute of Paper Chemistry: Appleton, WI, 1976.
34. R. M. Williams, The Vibrational Spectra of the Inositols. Doctoral Dissertation, The Institute of Paper Chemistry: Appleton, WI, 1977.
35. R. M. Williams; R. H. Atalla, *J. Phys. Chem.* **1984**, 88, 508.
36. H. A. Wells; R. H. Atalla, *J. Mol. Struct.* **1990**, 224, 385.
37. K. P. Carlson, The Vibrational Spectra of the Cellodextrins. Doctoral Dissertation, The Institute of Paper Chemistry: Appleton, WI, 1978.
38. J. H. Wiley; R. H. Atalla, *Carbohydr. Res.* **1987**, 160, 113.
39. J. H. Wiley; R. H. Atalla, In *The Structures of Celluloses*; R. H. Atalla, Ed.; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987; p 151.
40. T. Petipas; M. Oberlin; J. Mering, *J. Polym. Sci. C* **1963**, 2, 423.
41. M. Norman, *Text. Res. J.* **1963**, 33, 711.
42. D. A. Rees; R. J. Skerret, *Carbohydr. Res.* **1968**, 7, 334.
43. S. Melberg; K. Rasmussen, *Carbohydr. Res.* **1979**, 71, 25.
44. E. B. Wilson, Jr.; J. C. Decius; P. C. Cross, *Molecular Vibrations*; McGraw-Hill: New York, 1955; p 188.
45. M. L. Nelson; R. T. O'Connor, *J. Appl. Polym. Sci.* **1964**, 8, 1311.
46. J. Sugiyama; T. Okano; H. Yamamoto; F. Horii, *Macromolecular* **1990**, 23, 3196.
47. D. I. Bower; W. F. Maddams, *The Vibrational Spectroscopy of Polymers*; Cambridge University Press: Cambridge, UK, 1989.
48. R. H. Atalla; D. L. VanderHart, In *Cellulose and Wood – Chemistry and Technology*; C. Scheurch, Ed.; Wiley-Interscience: New York, 1989; p 169.
49. F. Horri; A. Hirai; R. Kitamaru, In *The Structures of Celluloses*; R. H. Atalla, Ed.; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987; p 119.
50. G. C. Campbell; D. L. VanderHart, *J. Mag. Reson.* **1992**, 96, 69–93.
51. F. Horii; A. Hirai; R. Kitamaru, *Macromolecules* **1987**, 20 (9), 2117–2120.
52. H. Yamamoto; F. Horii, *Macromolecular* **1993**, 26, 1313.
53. H. Yamamoto; F. Horii; H. Odani, *Macromolecular* **1989**, 22, 4130.
54. H. Yamamoto; F. Horii, *Macromolecular* **1993**, 26, 1313.
55. H. Yamamoto; F. Horii, *Cellulose* **1994**, 1, 57.
56. R. H. Newman; J. A. Hemmingson, *Holzforschung* **1990**, 44, 351.
57. R. H. Newman; J. A. Hemmingson, *Cellulose* **1995**, 2, 95.
58. R. H. Newman, *J. Wood Chem. Technol.* **1994**, 14, 451.
59. R. H. Newman; M. A. Ha; L. D. Melton, *J. Agric. Food Chem.* **1994**, 42, 1402.
60. R. H. Newman; L. M. Davies; P. J. Harris, *Plant Physiol.* **1996**, 111, 474.
61. R. H. Newman, *Cellulose* **1997**, 4, 269.
62. T. Larsson; U. Westermark; T. Iversen, *Carbohydr. Res.* **1995**, 278, 339.
63. H. Lennholm; T. Larsson; T. Iversen, *Carbohydr. Res.* **1994**, 261, 119.
64. H. Lennholm; T. Larsson; T. Iversen, *Carbohydr. Res.* **1994**, 261, 119.
65. T. Larsson; K. Wickholm; T. Iversen, *Carbohydr. Res.* **1997**, 302, 19.
66. C. W. Hock, In *Cellulose and Cellulose Derivatives*, Pt. I.; E. Ott, H. M. Spurlin, M. W. Grafflin, Eds.; Interscience: New York, 1954; p 347.
67. F. F. Morehead, In *Cellulose and Cellulose Derivatives*, Pt. IV; N. M. Bikales, L. Segal, Eds.; Wiley-Interscience: New York, 1971; p 213.
68. R. D. Preston, *The Physical Biology of Plant Cell Walls*; Chapman and Hall: London, 1974.
69. K. Hieta; S. Kuga; M. Usuda, *Biopolymers* **1984**, 23, 1807.
70. H. Chanzy; B. Henrissat, *FEBS Lett.* **1985**, 184, 285.
71. A. Maurer; D. Fengel, *Holz als Roh- und Werkstoff* **1992**, 50, 493.
72. J. Sugiyama; H. Harada; Y. Fujiyoshi; N. Uyeda, *Mokuzai Gakkaishi* **1984**, 30, 98.
73. J. Sugiyama; H. Harada; Y. Fujiyoshi; N. Uyeda, *Mokuzai Gakkaishi* **1985**, 31, 61.
74. J. Sugiyama; H. Harada; Y. Fujiyoshi; N. Uyeda, *Planta* **1985**, 166, 161.
75. A. Frey-Wyssling, *The Plant Cell Wall*; Gebruder Borntrager: Berlin, 1976.
76. J. Sugiyama; T. Okano; H. Yamamoto; F. Horii, *Macromolecular* **1990**, 23, 3196.
77. H. Yamamoto; F. Horii; H. Odani, *Macromolecules* **1989**, 22, 4130.
78. D. A. Rees; R. J. Skerret, *Carbohydr. Res.* **1968**, 7, 334.
79. B. Henrissat; S. Perez; I. Tvaroska; W. T. Winters, In *The Structures of Celluloses*; R. H. Atalla, Ed.; ACS Symposium Series 340; American Chemical Society: Washington, DC, 1987; p 38.
80. A. Sarko; R. Muggli, *Macromolecular* **1974**, 7, 486.
81. O. Ellefsen; B. A. Tonessen, In *Cellulose and Cellulose Derivatives*, Pt. IV; N. M. Bikales, L. Segal, Eds.; Wiley-Interscience: New York, 1971; p 151.
82. H. Chanzy; B. Henrissat; M. Vincendon; S. Tanner; P. S. Belton, *Carbohydr. Res.* **1987**, 160, 1.
83. R. H. Atalla; B. E. Dimick; S. C. Nagel, In *Cellulose Chemistry and Technology*; J. C. Arthur, Jr., Ed.; ACS Symp. Series No. 40, 1977; p 30.

84. R. H. Atalla; J. D. Ellis; L. R. Schroeder, *J. Wood Chem. Technol.* **1984**, *4*, 465.
85. H. Chanzy; K. Imada; A. Mollard; R. Vuong; F. Barnoud, *Protoplasma* **1979**, *100*, 303.
86. N. R. Bertoniere; S. H. Zeronian, In *The Structures of Celluloses*; R. H. Atalla, Ed.; ACS Symposium Series 340, American Chemical Society: Washington, DC, 1987; p 255.
87. S. P. Rowland; E. J. Robbets, *J. Polym. Sci. A-1* **1972**, *10*, 2447.
88. R. F. Nickerson, *Text. Res. J.* **1951**, *21*, 195.
89. S. P. Rowland; E. J. Roberts; C. P. Wade, *Text. Res. J.* **1969**, *39*, 530.
90. S. P. Rowland, In *Modified Cellulosics*; R. M. Rowell, R. A. Young, Eds.; Academic Press: New York, 1978; p 147.
91. S. P. Rowland; E. J. Roberts; J. L. Bose, *J. Polym. Sci. A-1* **1971**, *9*, 1431.
92. S. P. Rowland; E. J. Roberts; J. L. Bose; C. P. Wade; *J. Polym. Sci. A-1* **1971**, *9*, 1623.
93. J. Mann; H. J. Marinar, *J. Polym. Sci.* **1958**, *32*, 357.
94. S. H. Zeronian; M. L. Coole; K. W. Alger; J. M. Chandler, *J. Appl. Polym. Sci. Appl. Polym. Symp.* **1983**, *37*, 1053.
95. J. E. Stone; A. M. Scallan, *Pulp. Pap. Mag. Can.* **1968**, *69*, 69.
96. J. E. Stone; E. Treiber; E. Abrahamson, *TAPPI* **1969**, *28*, 139.
97. R. P. Bell, *The Proton in Chemistry*; Cornell University Press: Ithaca, NY, 1973.
98. D. C. Johnson; M. D. Nicholson; F. C. Haigh, *Appl. Polym. Symp.* **1976**, *28*, 931.
99. D. C. Johnson; M. D. Nicholson, *Cellul. Chem. Technol.* **1977**, *11*, 349.
100. A. Isogai, In *Wood and Cellulosic Chemistry*; D. N.-S. Hon, N. Shiraiishi, Eds.; Mercer Dekker: New York, 2000; pp 599–625.
101. A. Isogai; R. H. Atalla, *Cellulose* **1998**, *5*, 309–319.
102. K. Kamide; K. Okajima; K. Kowsaka, *Polym. J.* **1992**, *24*, 71–86.
103. A. Isogai; R. H. Atalla, *Carbohydr. Polym.* **1992**, *19*, 25–28.
104. A. F. Turbak; A. El-Kafrawy; F. W. Snyder, Jr.; A. B. Auerbach, Solvent System for Cellulose. U. S. Patent 4,302,252, 1981.
105. T. Schult; T. Hjerde; O. I. Optun; P. J. Kleppe; S. Moe, *Cellulose* **2002**, *9*, 149–158.
106. K. Okajima; C. Yamane, *Cellulose Commun.* **1997**, *4*, 7–12.
107. S. Saka; K. Takahashi; H. Matsumura, *J. Appl. Polym. Sci.* **1998**, *69*, 1445–1449.
108. K. Jumel; S. E. Harding; J. R. Mitchell; K.-M. To; I. Hayter; J. E. O'Mullane; S. Ward-Smith, *Carbohydr. Polym.* **1996**, *29*, 105–109.
109. R. Tanaka; J. Meadows; G. O. Phillips; P. A. ad Williams, *Carbohydr. Polym.* **1990**, *12*, 443–459.
110. A. Isogai; A. Ishizu; J. Nakano, *J. Appl. Polym. Sci.* **1986**, *31*, 341–352.
111. T. Kondo; D. G. Gray, *Carbohydr. Res.* **1991**, *220*, 173–183.
112. A. E. J. de Nooy; A. C. Besemer; H. Bekkum, *Carbohydr. Res.* **1995**, *269*, 89–98.
113. A. Isogai; Y. Kato, *Cellulose* **1998**, *5*, 153–164.
114. T. Isogai; M. Yanagisawa; A. Isogai, *Cellulose* **2009**, *16*, 117–127.
115. N. Konno; A. Isogai; N. Habu; N. Iihashi, *Cellulose* **2008**, *15*, 547–553.
116. T. Saito; S. Kimura; Y. Nishiyama; A. Isogai, *Biomacromolecules* **2007**, *8*, 2492–2496.
117. H. Fukuzumi; T. Saito; Y. Kumamoto; T. Iwata; A. Isogai, *Biomacromolecules* **2009**, *10*, 162–165.
118. B. Henrissat, *Cellulose* **1994**, *1*, 169–196.
119. G. Davies; B. Henrissat, *Structure* **1995**, *3*, 853–859.
120. D. Y. Kim; Y. Nishiyama; M. Wada; S. Kuga, *Cellulose* **2001**, *8*, 29–33.
121. S. Kuga; D. Y. Kim; Y. Nishiyama; R. M. Brown, *Mol. Cryst. Liq. Cryst.* **2002**, *387*, 237–243.
122. J. Azuma; T. Asai; M. Isaka; T. Koshijima, *J. Ferment. Technol.* **1985**, *63*, 529–536.
123. S. Saka; T. Ueno, *Cellulose* **1999**, *6*, 177–191.
124. S. Kobayashi; K. Kashiwa; T. Kawasaki; S. Shoda, *J. Am. Chem. Soc.* **1991**, *113*, 3079–3084.
125. J. H. Lee; R. M. Brown; S. Kuga; S. Shoda; S. Kobayashi, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9195.
126. F. Nakatsubo; H. Kamitakahara; M. Hori, *J. Am. Chem. Soc.* **1996**, *118*, 1677–1681.

Biographical Sketches



Rajai H. Atalla received degrees from Rensselaer Polytechnic Institute (1955), and the University of Delaware (1958, 1960) in chemical engineering and chemical physics. During

8 years at the Hercules Research Center, he studied phase transitions in semicrystalline polymers, anomalous IR and NMR spectra of many compounds. He also developed the first theoretical model for photodegradation of inorganic pigment. As professor of chemical physics and engineering at The Institute of Paper Chemistry (1968–89), he pioneered the application of Raman spectroscopy and ^{13}C SS NMR spectra of native celluloses with David VanderHart of NIST. They found that all native celluloses are composites of two forms I_α and I_β (1984). With Umesh Agarwal, using a Raman microprobe, he developed the first direct evidence of orientation of lignin in secondary walls (1984). In 1989, as head of Chemistry and Pulping Research at USDA Forest Service and adjunct professor in Chemical and Biological Engineering at UW, he led the development of inorganic analogs of lignin peroxidases for use in liquid-effluent-free pulping and bleaching systems. Between 1999 and 2007, as a senior and pioneering research scientist he returned to the studies of molecular architecture in plant cell walls with emphasis on secondary walls and native celluloses. He has published over 180 papers and chapters, edited a book on *Structures of Cellulose*, and is a fellow of the International Academy of Wood Science and the Technical Association of the Pulp and Paper Industry. He received the Anselme Payen Award of the Cellulose Division of ACS and multiple USDA awards including the Forest Service Chief's Distinguished Scientist Award. Upon retiring from the Forest Service he established Cellulose Sciences International as a consulting firm and to develop new technologies for utilization of celluloses.



Akira Isogai is a professor of cellulose, pulp and paper science, Department of Biomaterial Sciences, The University of Tokyo, Japan. He received his B.S., M.S., and Ph.D. degrees from The University of Tokyo in 1980, 1982, and 1985, respectively. He was a postdoctoral fellow at The Institute of Paper Chemistry, Appleton, Wisconsin, USA, from 1985 to 1986, an assistant professor of The University of Tokyo from 1986 to 1994, and an associate professor of The University of Tokyo from 1994 to 2003. He was a visiting scientist at Forest Products Laboratory, USDA, Madison, Wisconsin, USA, from 1989 to 1990, a fellow of International Academy of Wood Science (2006–), and a board member of Cellulose Society of Japan, Japan TAPPI, Society of Japan Packaging Science, Japan Wood Research Society, Society of Japan Fiber Science and Technology, and Society of High Performance Paper. He was a member of ACS, TAPPI, Japan Polymer Science, and so on.

6.17 Vascular Plant Lignification: Biochemical/Structural Biology Considerations of Upstream Aromatic Amino Acid and Monolignol Pathways

Dhrubojyoti D. Laskar, Oliver R. A. Corea, Ann M. Patten, ChulHee Kang, Laurence B. Davin, and Norman G. Lewis, Washington State University, Pullman, WA, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.17.1 | Introduction | 542 |
| 6.17.1.1 | Lignification: A Pivotal Role in Vascular Plant Evolution | 542 |
| 6.17.1.2 | The Emergence of Vascular Land Plants | 543 |
| 6.17.2 | Biochemistry of Phenylalanine and Tyrosine Formation in Vascular Plants: The Entry Point to Phenylpropanoid Metabolism and to Lignification | 543 |
| 6.17.2.1 | The Previous Enigma of Arogenate and/or Prephenate Dehydratases in Plants: The Role of Arogenate Dehydratase | 545 |
| 6.17.2.2 | The Enigma of Prephenate and/or Arogenate Dehydrogenases in Plants | 549 |
| 6.17.2.3 | The Enigma of Prephenate and/or Phenylpyruvate/ <i>p</i> -Hydroxyphenylpyruvate Aminotransferases in Plants | 549 |
| 6.17.2.4 | Regulation of Aromatic Amino Acid Biosynthesis | 551 |
| 6.17.2.4.1 | Product feedback control | 551 |
| 6.17.2.4.2 | Transcription | 553 |
| 6.17.2.4.3 | Phenylalanine and tyrosine biochemical partitioning | 554 |
| 6.17.2.5 | Bacterial Prephenate Dehydratases and Plant Arogenate Dehydratases: A Structural Biology Comparison via Homology Modeling | 554 |
| 6.17.3 | Structural Biology Considerations for the Monolignol-Forming Biosynthetic Pathway Steps | 557 |
| 6.17.3.1 | Biochemical Mechanisms of Phenylalanine and Tyrosine Ammonia Lyases: Comparison to Histidine Ammonia Lyases | 557 |
| 6.17.3.1.1 | HAL, PAL, and TAL: Discovery of the MIO prosthetic group | 557 |
| 6.17.3.1.2 | Putative catalytic mechanisms for HAL, PAL, and TAL | 560 |
| 6.17.3.1.3 | HAL and PAL: Proposed Tyr loop-in model for 'breathing' motion for substrate access | 562 |
| 6.17.3.1.4 | The molecular basis of PAL and TAL substrate versatility | 563 |
| 6.17.3.2 | Hydroxycinnamoyl CoA:Shikimate/Quinate Hydroxycinnamoyltransferase | 564 |
| 6.17.3.3 | Cytochrome P-450 Hydroxylation Reactions (Cinnamate 4-Hydroxylase, ' <i>p</i> -Coumarate 3-Hydroxylase', and 'Ferulate 5-Hydroxylase'): Comparison to Bacterial/Mammalian P-450s | 569 |
| 6.17.3.3.1 | Subcellular localization of C4H, ' <i>p</i> C3H', and 'F5H' | 569 |
| 6.17.3.3.2 | Cinnamate 4-hydroxylase | 569 |
| 6.17.3.3.3 | <i>p</i> -Coumarate 3-hydroxylase | 571 |
| 6.17.3.3.4 | Ferulate 5-hydroxylase | 571 |
| 6.17.3.3.5 | Bacterial/mammalian cytochrome P-450 catalysis/structural studies: Relevance to C4H, ' <i>p</i> C3H', and 'F5H' | 572 |
| 6.17.3.4 | 4-Coumarate:coenzyme A Ligases: Comparison to Other Adenylate-Forming Enzymes | 576 |
| 6.17.3.4.1 | Conformational changes during catalysis: Relevance to 4CL? | 576 |
| 6.17.3.4.2 | Mode of action of catalysis | 579 |
| 6.17.3.4.3 | Protein interactions with pyrophosphate leaving group | 580 |
| 6.17.3.4.4 | Homology modeling of At4CL2 and attempted substrate-binding pocket manipulations | 581 |
| 6.17.3.5 | COMTs and CCOMTs | 581 |

| | | |
|---------------|--|-----|
| 6.17.3.5.1 | O-Methyltransferases in the monolignol pathway: Discovery of 'COMT' and 'CCOMT' proper | 581 |
| 6.17.3.5.2 | S-COMT 3D structure | 582 |
| 6.17.3.5.3 | 3D structure of CCOMT | 584 |
| 6.17.3.5.4 | Structural comparisons of S-COMT and CCOMT and catalytic mechanisms | 585 |
| 6.17.3.5.5 | 3D structure of 'COMT' | 586 |
| 6.17.3.6 | Cinnamoyl CoA Reductase | 587 |
| 6.17.3.7 | Cinnamyl Alcohol Dehydrogenase: Comparison to Horse Liver Alcohol Dehydrogenase | 590 |
| 6.17.3.7.1 | Horse liver ADH | 591 |
| 6.17.3.7.2 | CAD | 591 |
| 6.17.3.7.3 | Comparison of AtCAD5 with poplar 'SAD' and yeast 'CAD' | 592 |
| 6.17.3.7.4 | CAD catalysis and putative proton relay | 593 |
| 6.17.4 | Conclusions | 595 |
| | Acknowledgments | 595 |
| | References | 595 |

6.17.1 Introduction

The substantial progress made in understanding how aromatic amino acid metabolism and monolignol biosynthesis occur in vascular plants, leading to cell wall polymeric lignins^{1,2} (see **Figure 1**),³ as well as lignans,^{4–7} allyl/propenyl phenols,⁷ and so on, is comprehensively discussed in this contribution. It builds upon an earlier 1999 report⁸ that particularly highlighted the pioneering studies in monolignol pathway elucidation. The thrust of the current contribution focuses on two scientific areas. This includes (1) progress toward understanding how phenylalanine (1) and tyrosine (2) are formed in vascular plants, these being the precursor metabolites affording entry into the phenylpropanoid/lignin and related biochemical pathways, and (2) structural biology and kinetic considerations associated with monolignol formation. Other aspects of vascular (lignified) cell wall development have been critically discussed elsewhere. These include the current understanding of metabolic networks involved in the lignin branch of phenylpropanoid metabolism and metabolic 'cross talk', transcriptional control of lignification and assembly of specific cell wall types, biochemical and physiological consequences of manipulating the lignin-forming (monolignol) pathway *in planta*, and progress made toward establishing lignin primary structures including theoretical (computational) considerations of the various noncovalent interactions.^{1,2} Previous contributions had also described in detail monolignol composition variations in lignins from different plant species, distinct cell wall sublayers, and cell wall types.^{1,2}

6.17.1.1 Lignification: A Pivotal Role in Vascular Plant Evolution

Lignification is arguably one of the most important biochemical processes that evolved in nature.^{1,2,8} In this way, the vast majority of land-based multicellular plants developed a system whereby some of their cells (actually cell wall types) evolved to facilitate conduction of water and nutrients in their various forms of vascular apparatus (see arrowheads, **Figures 2(a)–2(f)**). This, in turn, provided the means for all vascular plants (such as horsetails, ferns, cycads, gymnosperms, and angiosperms as depicted in **Figures 2(a)–2(f)**) to pipe water from their roots to aerial tissues including, in some instances, to treetops hundreds of feet above the ground.⁹ This was generally achieved through thickening and reinforcement of specific cell wall types in the vascular apparatus (**Figures 2(g) and 2(h)**), this largely occurring through secondary processes involving lignification. In this way, the reinforced vascular apparatus afforded a physical system with the following advantages: enabled water and nutrients to be conducted throughout the plant; provided a means for vascular plants to stand upright and withstand compressive forces, as well as provided physical barriers to encroaching predators and pathogens.⁹

Interestingly, the amount of cell wall lignin deposition in the vascular apparatus can vary from a few percent to approximately 30% depending on the plant species, this in turn depending on whether they are of nonwoody or woody character.^{1,2} Lignins are, however, essentially absent from algae, true aquatic plants, and basal land

plants, such as mosses, hornworts, and liverworts. Yet, Russian scientists have reported the presence of trace amounts of putative lignins in *Fucus vesiculosus*¹⁰ and *Cystoseira barbata*.^{11,12} This question was recently reexamined with the red alga *Calliarthron cheilosporioides*,¹³ with putative lignin contents likely in the order of $\leq 0.15\%$ or so of dry weight. Presumably, in such organisms, these substances have no substantial role in cell wall reinforcement, relative to the lignified vascular plants proper.

6.17.1.2 The Emergence of Vascular Land Plants

The emergence of vascular plants on land and their subsequent diversifications were crucial to establishment of various terrestrial environments on earth. While the earliest plants to emerge (~ 460 million years ago (Mya))^{14,15} had limited resistance to desiccation, the later evolution of the vascular systems allowed them (by ~ 400 Mya)^{16,17} to move from consistently moist environments to those further from surface water.⁹ This adaptation occurred through progressive development of xylem secondary cell walls from loose helical thickened anatomical structures to that of intricately layered cell walls (primary and secondary) (Figures 2(g) and 2(h)), with these cell wall assemblies also containing both lateral and vertical water-conducting structures (including pits and end plates) (Figure 3). Together, these were crucial for long-distance transport of water and nutrients. During this progressive evolutionary development, however, the xylem cell walls incorporated phenolic compounds into their secondary cell walls for structural reinforcement (to help offset negative transpirational pressure), as well as to increase the hydrophobicity of the cell walls of the water transport conduits (thereby facilitating water movement). Plant forms thus began to attain impressive heights with evolution of reinforced lignified fiber-tracheid, tracheid, vessel, and fiber cells.⁹ For example, the great heights (up to 115 m) of giant sequoia (*Sequoiadendron giganteum*) are, in part, due to lignified fiber-tracheids, which also serve to help support their massive trunks.

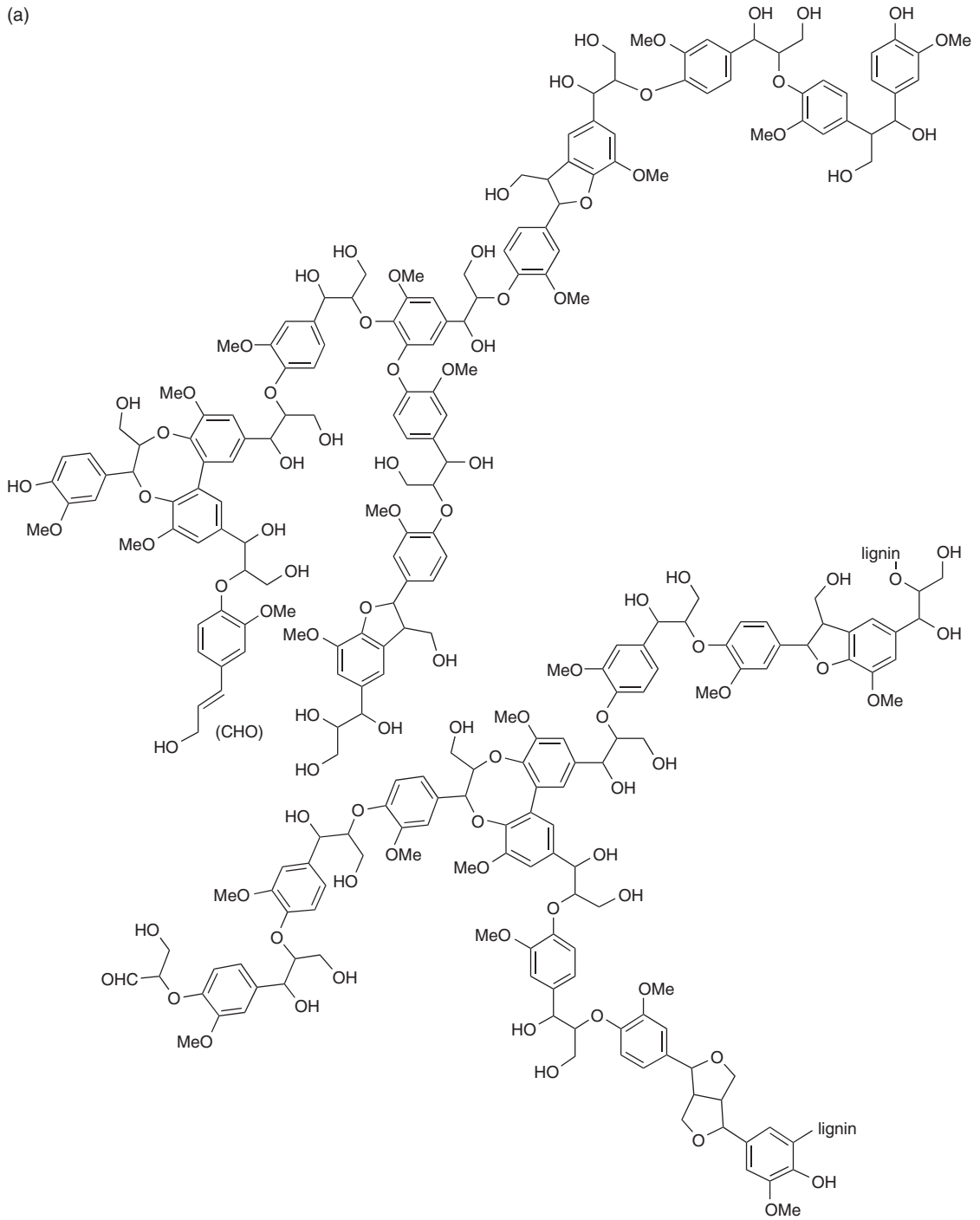
In sum, one of the most significant evolutionary developments in the plant kingdom was the transition of aquatic plant forerunners, such as algae, to a land base. To a large part, this successful transition resulted from the emergence, and further elaboration, of the phenylpropanoid pathway away from Phe (1)/Tyr (2) (Figure 4) and protein biosynthesis. This phenylpropanoid biochemical pathway enabled plants to survive in a desiccating habitat (e.g., through suberization),^{18–20} as well as the means of growing vertically upright. The latter was facilitated by a reinforcement mechanism (lignification^{1,2,8} of cell walls) counteracting the effects of compressive forces acting on, for example, plant stems during growth and development. Together, these polymeric substances also helped provide physical barriers to herbivores, pathogens, and so on. Yet, many other related pathway branches of phenylpropanoid metabolism also evolved to provide, among other functions, the means to counteract/help counteract the effects of opportunistic pathogens through formation of their often bioactive substances, for example, through the lignan,^{4–7} flavonoid,²¹ stilbene,²² and allyl/propenol phenol pathways.^{7,23–26}

The goal of this particular contribution is thus to summarize and comprehensively evaluate the progress made in both aromatic amino acid (Phe (1)/Tyr (2)) and monolignol biosynthesis pathways, with a particular emphasis on the structural biology of the same. Monolignols, such as 19, 21, 22 and 23, can serve as building blocks for the polymeric cell wall lignins,^{1,2,8} as well as precursors in lignan^{4–7} and allyl/propenyl phenol^{7, 23–26} biosynthesis.

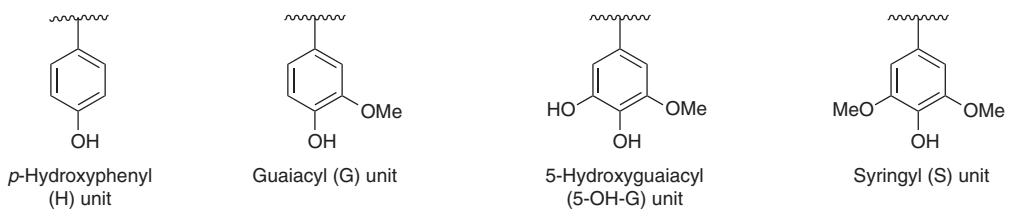
6.17.2 Biochemistry of Phenylalanine and Tyrosine Formation in Vascular Plants: The Entry Point to Phenylpropanoid Metabolism and to Lignification

The shikimate–chorismate pathway affords chorismate (37, Figure 5), which can then be converted into the aromatic amino acids Phe (1), Tyr (2), as well as Trp (43, Figure 6), with both aromatic amino acid and shikimate–chorismate pathways being localized in plants to chloroplasts/plastids.^{27–29} Interestingly, shikimate (29) was originally isolated from the vascular plant *Illicium religiosum*, shikiminoki (star anise tree) in 1885.^{30,31} The ‘main trunk’ of the shikimate–chorismate pathway from erythrose 4-phosphate (E4P, 30) and phosphoenolpyruvate (PEP, 31) to chorismate (37) apparently follows the same sequence of reactions in all organisms (bacteria, fungi, plants) examined thus far, as does the Trp (43) branch³² (Figures 5 and 6). For shikimate–chorismate metabolism, the corresponding upstream enzymes, encoding genes, and pathway intermediates have been extensively studied since the initial discovery of the involvement of shikimate (29) in aromatic amino acid biosynthesis nearly 60 years ago³³ (for reviews, see Bentley,³² Gilchrist and Kosuge,³⁴ Herrmann and Weaver,³⁵ Pittard,³⁶ Schmid and Amrhein³⁷). Yet, before discussing

(a)



(b)



what is known about Phe (1)/Tyr (2) formation in vascular plants, a brief summary of the known pathways in bacteria and/or fungi is first provided for needed historical context.

In bacteria/fungi, the conversion of chorismate (37) to Phe (1) and Tyr (2) appears to have two potential routes operative (Figure 6): Phe (1) can be synthesized via either transamination of phenylpyruvate (39) by an aromatic aminotransferase and/or dehydration/decarboxylation of aroenate (41) by aroenate dehydratase (ADT). On the other hand, Tyr (2) can be synthesized via either transamination of *p*-hydroxyphenylpyruvate (40) and/or decarboxylation of aroenate (41) by aroenate dehydrogenase. In microorganisms, various combinations of these routes are known to exist,^{38–43} while fungi seem to exclusively use transamination of phenylpyruvate (39) and *p*-hydroxyphenylpyruvate (40).^{44–47} However, the enzymology of both routes to Phe (1) and Tyr (2) can differ markedly between microbial species. For example, Gram-negative enteric bacteria often have dual-function enzymes called P-proteins and T-proteins, which have chorismate mutase (CM) domains fused to prephenate dehydratase (PDT) and prephenate dehydrogenase (PDH) domains, respectively.^{39,48} P-proteins and T-proteins can thus directly catalyze reactions from chorismate (37) to phenylpyruvate (39) and to *p*-hydroxyphenylpyruvate (40), respectively. Also common in enteric bacteria and some superfamily-B genera, such as *Pseudomonas aeruginosa*, are cyclohexadienyl dehydratases (CDTs)^{41,42} and cyclohexadienyl dehydrogenases (CDHs),^{41,43,48} which can utilize both routes to Phe (1) and Tyr (2), respectively, due to broad substrate versatility for prephenate (38) and aroenate (41). However, there is no obvious homology between CDTs and their monosubstrate-recognizing counterparts (i.e., PDTs/ADTs),^{42,43} while a very low degree of homology was noted between CDH and PDH/aroenate dehydrogenase.⁴³

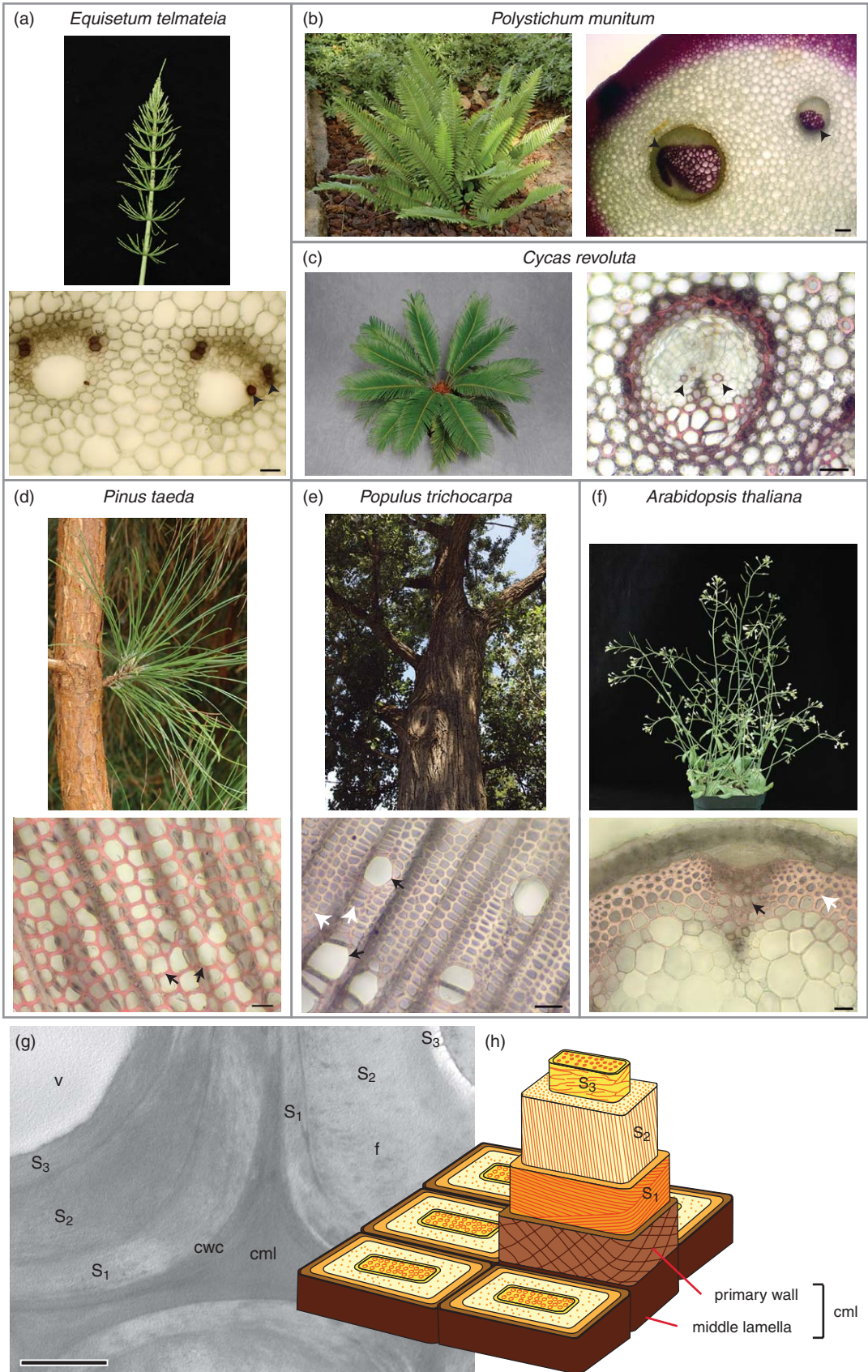
On the other hand, the pathways to both aromatic amino acids, Phe (1)/Tyr (2), in vascular plants are only beginning to be clarified, and could not readily be deduced from bacterial/fungal sequence/comparisons, for example, in terms of substrate(s). Thus the key to fully understanding the pathways to these two aromatic amino acids in plants was to identify the enzymes, as well as to ultimately establish their substrate specificities/feedback properties; that is, of the actual dehydratases, dehydrogenases, and aromatic aminotransferases involved, including how transcriptional regulation is attained.

Discussed below therefore is what is known currently about Phe (1)/Tyr (2) biosynthesis *in planta*, as it relates to the conversion of prephenic acid (38) to both aromatic amino acids 1/2 (see Figure 6), and the large scientific and technical gaps that still remain.

6.17.2.1 The Previous Enigma of Aroenate and/or Prephenate Dehydratases in Plants: The Role of Aroenate Dehydratase

The presumed primary determinant of which pathway is used for Phe (1) biosynthesis *in planta* is the substrate specificity of the dehydratase(s) for prephenate (38) or aroenate (41) conversion into phenylpyruvic acid (39) or Phe (1), respectively (Figure 6). Initially, vascular plants were assumed to have dehydratases similar to those in microorganisms, and these were therefore thought to employ a PDT (EC 4.2.1.51) to afford phenylpyruvate (39);³⁴ however, no enzyme assays had been performed to support this view. Later, biochemical assays of crude plant extracts detected ADT (EC 4.2.1.91) but not PDT activities, suggesting that aroenate (41) was the primary intermediate.^{49–51} Presumably, because of enzyme stability problems and the like, there were no subsequent reports of purification of either ADTs or PDTs from plants using conventional protein purification techniques. More recently, phenylpyruvate (39) was once again proposed as an intermediate in Phe (1) biosynthesis,⁵² with a gene encoding a putative PDT being described. Aroenate (41), however, was not examined as a substrate. Moreover, in that study, biochemical assays were carried out only using crude extracts, and no kinetic data for prephenate (38) were obtained. Thus, there was no proof that this gene encoded a PDT.

Figure 1 Artistic depiction of (a) gymnosperm lignin structure as envisaged by Brunow *et al.*³ and (b) lignin aromatic ring residues (H, G, 5-OH-G, and S). See Davin *et al.*^{1,2} for discussion and limitations of proposed lignin structures. (a) Adapted from G. Brunow; I. Kilpeläinen; J. Sipilä; K. Syrjänen; P. Karhunen; H. Setälä; P. Rummakko, *Oxidative Coupling of Phenols and the Biosynthesis of Lignin*. In *Lignin and Lignan Biosynthesis*; N. G. Lewis, S. Sarkanen, Eds.; ACS Symposium Series: Washington, DC, USA, 1998; Vol. 697, pp 131–147.



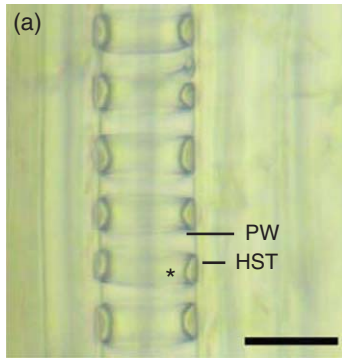
To fully establish whether aroenate (41), phenylpyruvate (39), or both were pathway intermediates to Phe (1) in vascular plants, it was essential to unambiguously identify all enzymatic processes (and encoding genes) needed for conversion of prephenate (38) into Phe (1) both *in vivo* and *in vitro*. Moreover, it was also essential to obtain rigorous enzymatic kinetic data using highly purified recombinant enzymes *in vitro* for all potential substrates, in order to compare and contrast relative efficacies/feedback inhibition properties and so forth.

It was, therefore, deemed instructive to first examine whether Phe (1) formation was occurring via aroenate (41) in *Arabidopsis thaliana*. In this regard, an initial *in silico* screening resulted in provisional identification of six putative PDT domain-containing genes in *Arabidopsis*. In terms of an experimental determination of biochemical function, the first and only detailed molecular and biochemical analysis of this multigene family was carried out using all six individual *Arabidopsis* ADT/PDT isoforms expressed in *Escherichia coli*.⁵³ This study established that all six ADT isoforms indeed predominantly used aroenate (41) as a substrate rather than prephenate (38), and could therefore be classified as ADTs. These same six genes were also briefly mentioned in a microarray analysis by Ehrling *et al.*,⁵⁴ being arbitrarily named according to provisionally annotated – but unproven and unexamined – biochemical functions as AtADT1 (At1g11790), AtADT2 (At3g07630), AtADT3 (At2g27820), AtADT4 (At3g44720), AtADT5 (At5g22630), and AtADT6 (At1g08250). In agreement with our experimental findings, it has since been shown that an *Oryza* ADT, most similar to AtADT2, also has approximately 10-fold higher catalytic efficiency with aroenate (41) than with prephenate (38),⁵⁵ again indicating that the aroenate (41) route is very likely the primary means of Phe (1) biosynthesis in this monocot species as well.

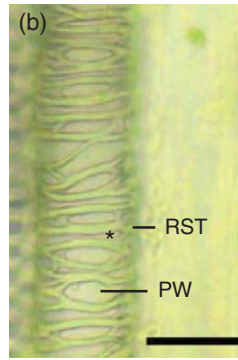
Another open question recently resolved involved the subcellular location of the pathway beyond chorismate (37) to Phe (1) and Tyr (2). Most higher plants examined thus far contain two forms of CM, CM1 being feedback-sensitive and chloroplast-localized, and CM2 being feedback-insensitive and cytosolic;^{56–59} however, it was not known whether there were cytosolic forms of the other enzymes between prephenate (38) and Phe (1)/Tyr (2). In this regard, based on computer predictions by TargetP,⁶⁰ five of the six *Arabidopsis* ADTs had signal (transit) peptides that putatively directed them to the chloroplast.⁵³ Using ADT–GFP (green fluorescent protein) fusion proteins,⁶¹ it was also recently demonstrated that all six *Arabidopsis* ADTs were indeed chloroplast-localized and absent from the cytosol. This result was thus consistent with the earlier finding that a recombinant *Oryza* ADT isozyme was imported into intact pea (*Pisum sativum*) chloroplasts, with the signal peptide being cleaved upon entry to yield the mature protein.⁵⁵ These findings therefore seem to resolve the long-standing question of whether a distinct feedback-insensitive shikimate pathway exists in the cytosol, in parallel to the known chloroplastic pathway.^{49,62} However, it remains unknown why a putative, feedback-insensitive, CM isoform exists in the cytosol, if there are no enzymes present to utilize prephenate (38).

Further studies now need to be carried out to gain a more complete biochemical understanding of ADT, including elucidation of its precise biochemical mechanism (discussed below), and for establishing both the role and regulation of the ADT multigene family in plants. Of particular interest will be how (or if) their manipulation differentially affects Phe (1)/Tyr (2) and protein biosynthesis, as well as other forms of downstream metabolism, for example, leading to (lignified) cell wall deposition/synthesis and other phenylpropanoid-related formation. It must again be emphasized however that there is no compelling/unambiguous evidence for any type of PDT activity in vascular plants, including *Arabidopsis*.

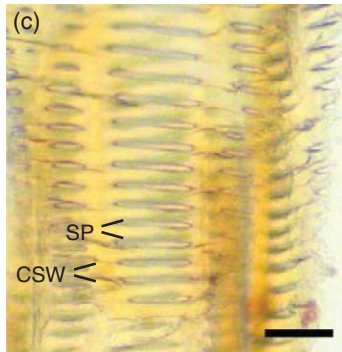
Figure 2 Elaboration of the lignification pathway during vascular plant evolution allowed for the strengthening of secondary cell walls that support water flow such as the metaxylem (arrowhead) of *Equisetum telmateia* (a), *Polystichum munitum* (b), and *Cycas revoluta* (c), or the tracheids (arrows) of *Pinus taeda* (d), the vessels (black arrows) and fibers (white arrows) of *Populus trichocarpa* (e), or the (meta)xylem (black arrows) and interfascicular fibers (white arrows) of *Arabidopsis thaliana* (f). Lignified secondary cell walls in woody plants (*Populus trichocarpa*, g) often consist of three distinct lignified layers: S₁, S₂, S₃ (g and h). These may vary in width as in the fiber (f) and vessel (v) cell-wall elements shown here, or vary in monomer (lignin) composition (data not shown). Scales: 1 μm (a), 50 μm (b and c); 25 μm (d–f). Pictures from Ann M. Patten, Washington State University (a–f). Reprinted from the American Journal of Botany, Vol. 94, A. M. Patten; M. Jourdes; E. E. Brown; M.-P. Laborie; L. B. Davin; N. G. Lewis, Reaction tissue formation and stem tensile modulus properties in wild type and *p*-coumarate-3-hydroxylase downregulated lines of alfalfa, *Medicago sativa* (Fabaceae), pp 912–925, Copyright 2007, with permission from the Botanical Society of America (g).



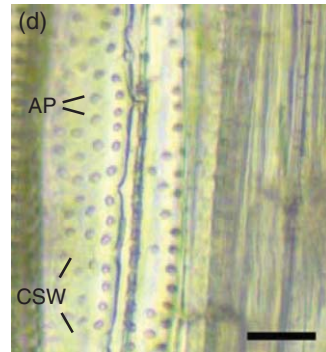
Horsetail
(*Equisetum telmateia*)



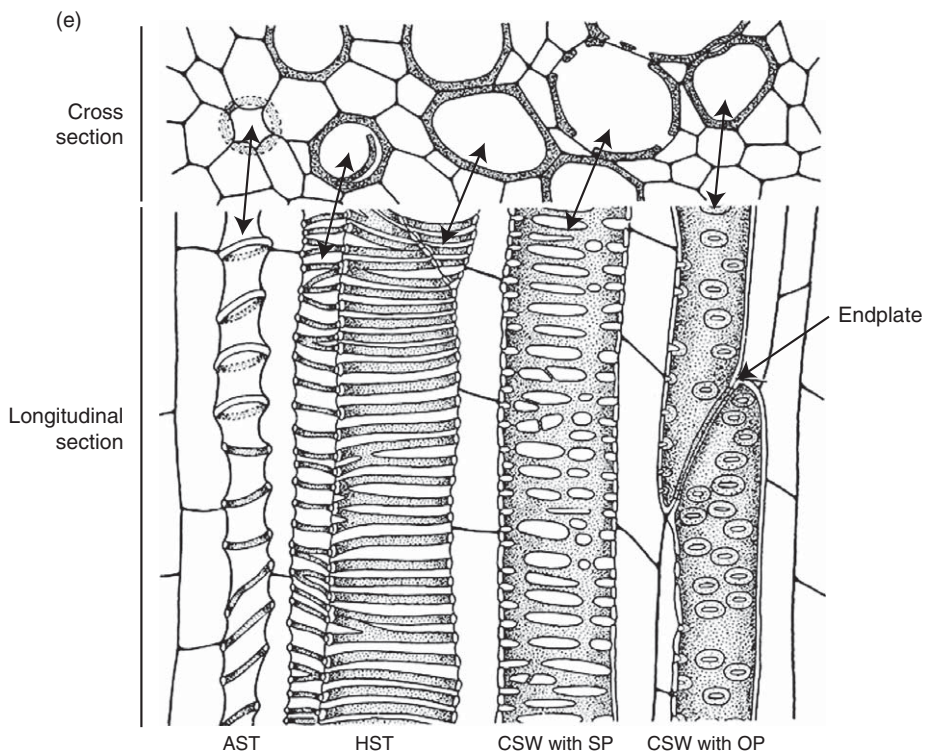
Adder's-tongue fern
(*Ophioglossum reticulatum*)



Bracken fern
(*Pteridium aquilinum*)



Alfalfa
(*Medicago sativa*)



6.17.2.2 The Enigma of Prephenate and/or Arogenate Dehydrogenases in Plants

While the possible order of reactions in Tyr (2) biosynthesis is somewhat better understood than that of Phe (1) biosynthesis, at least in *Arabidopsis*, large gaps apparently still remain. Nevertheless, two arogenate dehydrogenase genes (*TyrAAT1* and *TyrAAT2*) from *Arabidopsis* were isolated, heterologously expressed in *E. coli*, and their corresponding proteins biochemically characterized *in vitro*.^{63,64} Of these, TyrAAT1 had a strict specificity for arogenate (41) over prephenate (38), whereas TyrAAT2 also used prephenate (38), albeit with the latter being approximately 2000 times less efficient than arogenate (41) as a substrate. As for ADTs, computer predictions by TargetP⁶⁰ suggested that both arogenate dehydrogenases had transit peptides directing them to the chloroplast, and this has since been confirmed via analysis of TyrAAT1- and TyrAAT2-GFP fusion proteins.⁶¹ Additionally, for some other plant species studied so far (i.e., sorghum,⁶⁵ tobacco,⁶⁶ maize, wheat, wild oat, watergrass, green foxtail, pigweed, velvetleaf, nightshade, morning glory (D. L. Siehl, personal communication)), arogenate dehydrogenase activity was preliminarily biochemically detected in crude extracts, while PDH activity was not. On the other hand, in legumes, that is, soybean, alfalfa, and mung bean, comparable biochemical analyses resulted in dual arogenate/prephenate dehydrogenase activities being detected⁶⁷ (D. L. Siehl, personal communication). Indeed, both activities in soybean were apparently separated after hydrophobic interaction chromatography,⁶⁷ again raising the possibility of at least two pathways to Tyr (2). However, such observations contrast with bacteria, such as *P. aeruginosa* and many enteric bacteria, which have both arogenate and prephenate dehydrogenase activities in a single enzyme, CDH.^{42,68} There are thus several aspects that still need to be clarified as far as Tyr (2) biosynthesis in plants is concerned, that is, whether there are arogenate dehydrogenase, PDH, or both activities involved in different plant species. Clarification to such questions may also provide much needed insight into the biochemical partitioning of Phe (1)/Tyr (2) during grass lignin⁶⁹/rosmarinic acid^{24,70} (44) and colchicine (45)^{71–77} (Figure 7) biosynthesis (discussed later).

6.17.2.3 The Enigma of Prephenate and/or Phenylpyruvate/*p*-Hydroxyphenylpyruvate Aminotransferases in Plants

The biochemical step or steps where amino transfer occurs in the pathway is (are) also currently only very preliminarily characterized/identified for Phe (1) and Tyr (2) biosynthesis from prephenate (38) in plants. On the one hand, prephenate aminotransferases (PNTs) have been provisionally detected in several different crude plant extracts;^{78–81} however, conversely, both phenylpyruvate (39) and *p*-hydroxyphenylpyruvate (40) transaminase activities have also been noted, albeit with 10 times lower K_m affinities versus that of prephenate (38).⁸² To date, neither the corresponding proteins nor the encoding genes have been described though. Interestingly, in *Arabidopsis*, a model plant species and the first to have its genome sequenced,⁸³ there are more than 50 proteins with putative aminotransferase domains (pfam (protein family database) motif PF00155),⁸⁴ all of which have been provisionally annotated as aminotransferases (TAIR, <http://www.arabidopsis.org>). However, their precise biochemical roles are unknown. Taken together, the limited progress thus far underscores the very incomplete level of our current understanding as regards amino transfer step(s) in Phe (1)/Tyr (2) biosynthesis, and the urgent need to unambiguously clarify/resolve such questions.

Figure 3 Selected examples of secondary cell wall thickenings in vascular plants. Helical (HST; a, e) and annular (AST, e) secondary cell wall thickenings were among the first types to appear during early vascular plant evolution. AST and HST are found in extant primitive plants (e.g., *Equisetum telmateia*; a) and early growth of higher plants (e.g., gymnosperm and angiosperm spp.; e). Reticulate secondary cell wall thickenings (RST) (e.g., *Ophioglossum reticulatum*; b) and scalariform-pitted secondary cell wall thickenings (SP; e.g., *Pteridium aquilinum*; c) with lateral openings for controlled lateral water flow are shown. More continuous secondary cell wall structures (CSW; c,d) occur in mature stems of higher plants, forming rigid water systems, as in wood. Within these structures, lateral water flow is tightly controlled by pit openings that may occur as alternate (AP; d) or opposite (OP; e) patterns. Vertical water flow occurs between adjoined xylem elements, with intervening endplate structures (e) that control air bubble movement and thus lower the risk of catastrophic embolism in relatively taller stems. PW, primary wall. Scales: 20 μm (a, b), 30 μm (c, d). Reprinted from A. M. Patten, Ph. D. thesis, Washington State University, Copyright 2007, with permission (a–d) and from K. Esau, *Plant Anatomy*, Copyright 1965, with permission from John Wiley and Sons, Inc. (e).

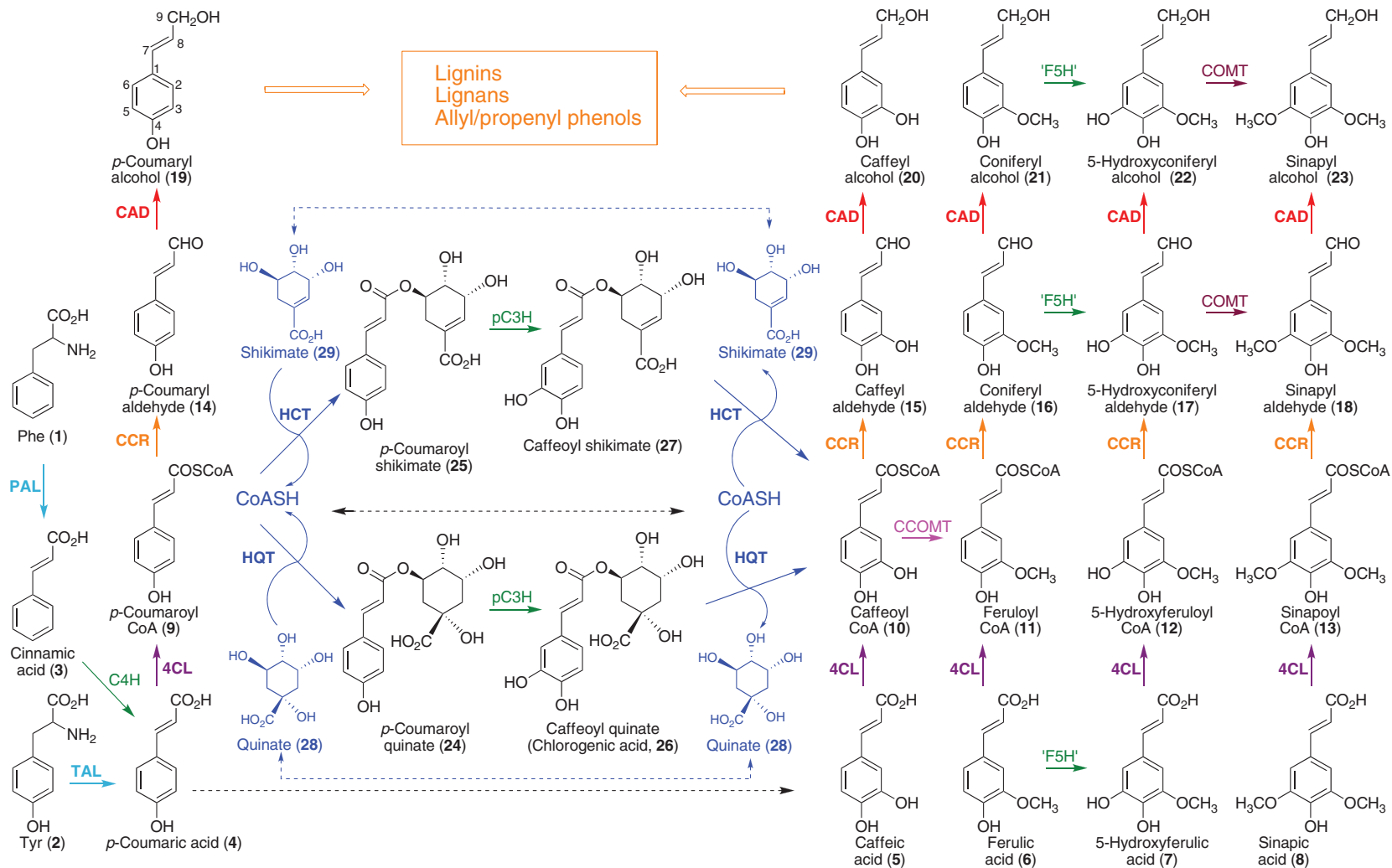


Figure 4 Current view of the phenylpropanoid pathway to the monolignols **19–23**. 4CL, 4-hydroxycinnamate:coenzyme A ligases; ‘pC3H’, ‘*p*-coumarate 3-hydroxylase’; C4H, cinnamate 4-hydroxylase; CAD, cinnamyl alcohol dehydrogenases; CCOMT, hydroxycinnamoyl CoA O-methyltransferases; CCR, cinnamoyl-CoA oxidoreductases; COMT, caffeic acid O-methyltransferases; ‘F5H’, ‘ferulate 5-hydroxylase’; HCT, hydroxycinnamoyl-CoA shikimate hydroxycinnamoyltransferase; HQT, hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase; PAL, phenylalanine ammonia lyase; TAL, tyrosine ammonia lyase.

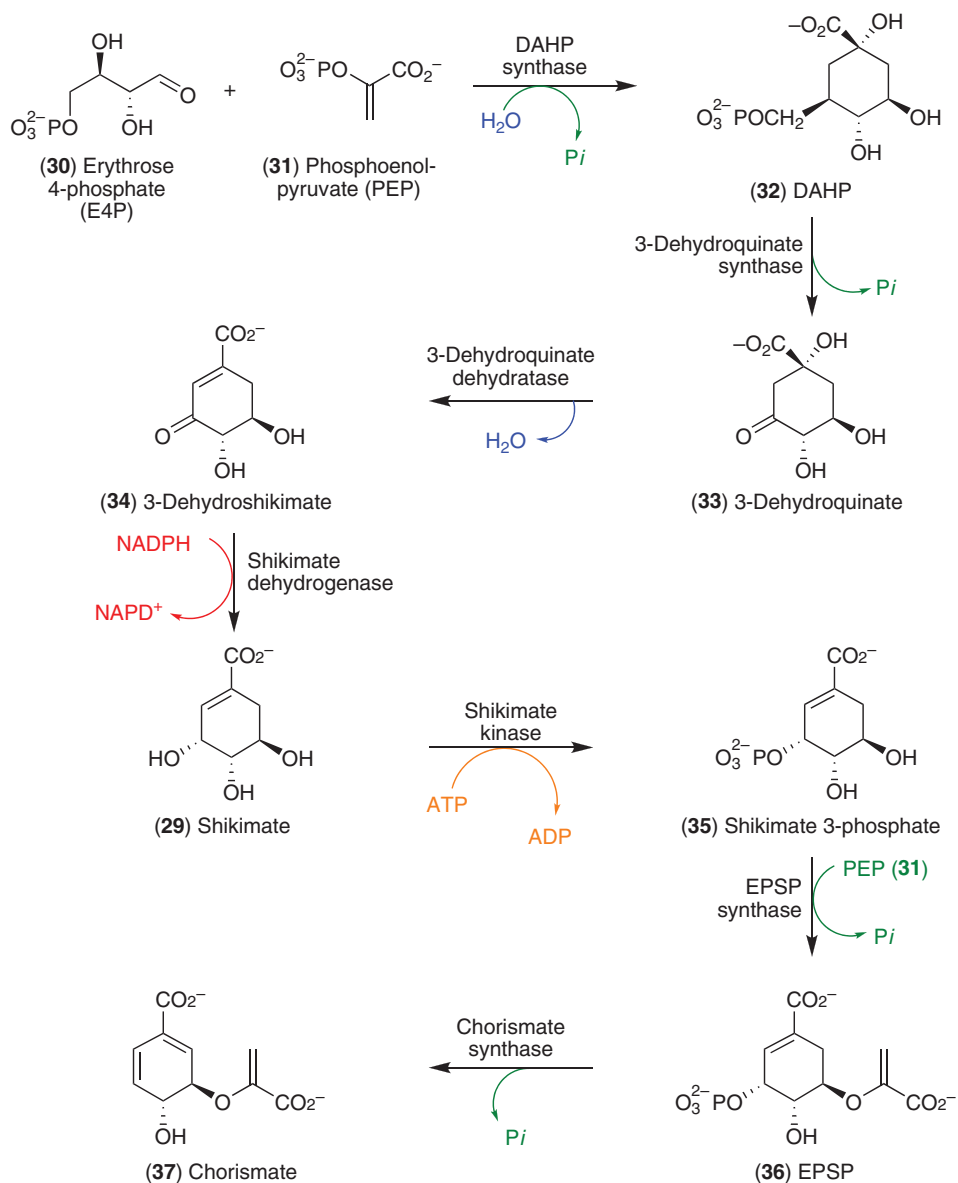


Figure 5 The shikimate–chorismate pathway. DAHP, 3-deoxy-D-arabino-2-heptulosonate-7-phosphate; EPSP, 5-enolpyruvylshikimate-3-phosphate.

6.17.2.4 Regulation of Aromatic Amino Acid Biosynthesis

For needed context, while this has been a very well-studied process – largely in bacteria – this topic is also mentioned as it pertains to the overall regulation of Phe (1)/Tyr (2) biosynthesis *in planta*. It is evident that regulation needs to be comprehensively understood simultaneously from the perspective of product feedback control, transcription, and translation, in order to obtain a comprehensive understanding of the processes involved. In addition, how these biochemical processes interface in ‘cross talk’ with downstream phenylpropanoid metabolism/nitrogen recycling also requires urgent resolution.

6.17.2.4.1 Product feedback control

Amino acid biosynthesis is often regulated by the end products of a given pathway to control the relative amounts of amino acids being produced.⁸⁵ Aromatic amino acid biosynthesis thus far is known to be feedback regulated at three points in the shikimate/chorismate pathway. The initial step is catalyzed by 3-deoxy-D-arabino-2-heptulosonate-7-

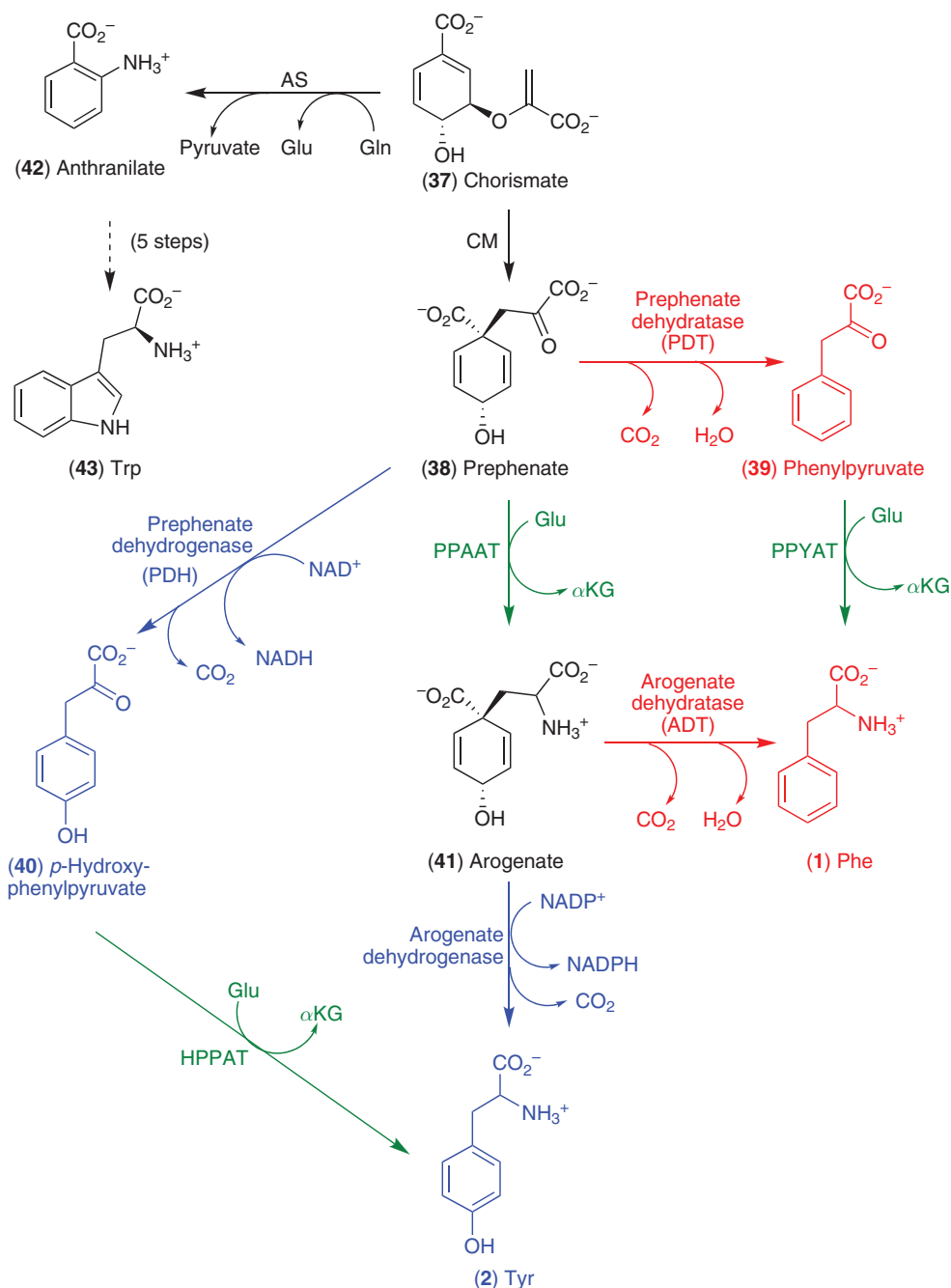
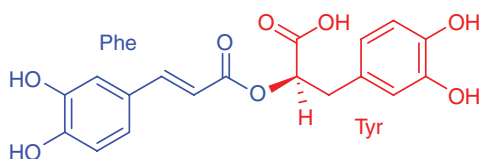


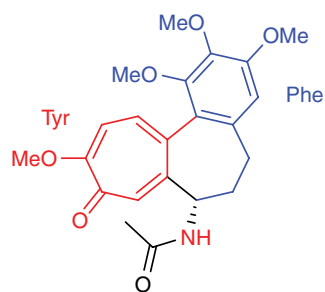
Figure 6 Proposed biosynthetic pathways from chorismate (37), prephenate (38), and aroenate (41) to Phe (1), Tyr (2), and Trp (43) in plants and microorganisms. ADT, aroenate dehydratase; AS, anthranilate synthase; CM, chorismate mutase; HPPAT, *p*-hydroxyphenylpyruvate aminotransferase; PDH, prephenate dehydrogenase; PPAAT, prephenate aminotransferase; PPYAT, phenylpyruvate aminotransferase.

phosphate (DAHP synthase, Figure 5), which is sensitive to feedback inhibition by Phe (1), Tyr (2), and Trp (43) in prokaryotes and lower eukaryotes,^{35–37} but is only apparently inhibited by aroenate (41) in plants.^{49,86,87} Inhibition thus blocks carbon flow through the pathway, thereby halting production of all three aromatic amino acids (1, 2, and 43). Conversely, a feedback-insensitive DAHP synthase isoform was detected in the cytosol of several plant species, which may be part of a feedback-insensitive shikimate pathway.^{86,87}

Enzymes catalyzing the conversions from chorismate (37), namely, CM and anthranilate synthase (AS) (Figure 6), are also subject to feedback regulation.⁸⁸ CM is inhibited by Phe (1) and Tyr (2), and activated by

*Ocimum basilicum*

(44) Rosmarinic acid

*Colchicum autumnale*

(45) Colchicine

Figure 7 Rosmarinic acid (44) found, for example, in basil (*Ocimum basilicum*) and colchicine (45) from autumn crocus (*Colchicum autumnale*). Photographs by Daniel G. Vassão and Laurence B. Davin, Washington State University.

Trp (43), while AS is inhibited by Trp (43).^{89,90,91} As such, CM and AS are thought to participate in helping control carbon flux between Phe (1)/Tyr (2) and Trp (43) pathways in the chloroplast/plastid. Again, feedback-insensitive forms have been identified for CM and AS in the cytosol^{56,92–94} but the physiological significance is unknown.

The last point of regulation controls carbon flux between Phe (1) and Tyr (2). Either prephenate (38) or aroenate (41) is the potential branch point, depending on the route used by a given species (see Figure 6). Preliminary biochemical analyses of dehydratases and dehydrogenases in both microorganisms and plants suggested that they too are sensitive to feedback regulation by Phe (1) and Tyr (2).^{38,51,63,64,88} Molecular studies of feedback regulation of the *Arabidopsis* aroenate dehydrogenases (TyrAAT1 and TyrAAT2) have also confirmed that they are sensitive to inhibition by Tyr (2) *in vitro*,⁶³ as predicted from biochemical analyses.⁵¹ Similarly, an *Oryza* ADT isozyme was shown to be inhibited by Phe (1), whereas a feedback-insensitive form of this same isozyme was found to cause an accumulation of Phe (1) as well as Trp (43).⁵⁵

Comprehensive studies are now required, however, to determine the feedback properties of all of the enzymes leading to Phe (1)/Tyr (2) biosynthesis, including the individual *Arabidopsis* ADTs and the yet to be provisionally identified aminotransferases both *in vivo* and *in vitro*. This is a particularly interesting and important question, considering that feedback-insensitive isoforms exist for the other branch point enzymes DAHP synthase, CM, and AS.

6.17.2.4.2 Transcription

There is also substantial evidence that the shikimate pathway and its branches toward all three aromatic amino acids 1, 2, and 43 are highly regulated at the transcriptional level, for example, in response to wounding,⁹⁵ fungal elicitors, or high light intensity.^{96–98} These enzymes again include DAHP synthase,^{95,98} CM,⁹⁹ and AS.^{90,100} Little is yet known, however, about transcriptional regulation of the Phe (1)/Tyr (2) branch beyond the CM reaction. This is again due to the genes encoding some of these enzymes having only recently been identified (i.e., as described above for ADT) and/or not yet being identified as for the aminotransferases. This

would thus appear to be a productive line of inquiry, that is, to examine transcriptional profiling, upstream and downstream of these enzymatic steps, with respect to, for example, cell wall synthesis and lignification.

6.17.2.4.3 Phenylalanine and tyrosine biochemical partitioning

As indicated in **Figure 4**, the phenylpropanoid pathway is largely Phe (**1**)-derived, although in some plant species, Tyr (**2**) may also have a smaller role. These amino acids can thus ultimately lead to, or be involved in, the synthesis of molecules such as cell wall structural compounds, lignins and suberins,^{1,8,18–20,101,102} lignans,^{4–7} allyl/propenol phenol pathways,^{7,23–26} as well as flavonoids/cyanidins^{103,104} and phytoalexins.¹⁰⁵ Interestingly, while all of these metabolites are generally Phe (**1**)-derived, the lignins in grasses are considered to be (at least in part) derived from Tyr (**2**); in that regard, Tyr (**2**) is sometimes considered to be partly involved in the biosynthesis of the hydroxycinnamic acid (**4**, **6**, and **8**) component of grass lignins⁶⁹ although how this putative biochemical partitioning of Phe (**1**) and Tyr (**2**) occurs during lignification (and other processes) is unknown. Another example of possible biochemical partitioning of both amino acids has been preliminarily reported in rosmarinic acid (**44**, **Figure 7**) biosynthesis, where one-half is Phe (**1**)-derived and the other half is from Tyr (**2**).^{24,70} A third example is in alkaloid formation, where both Phe (**1**) and Tyr (**2**) are differentially involved in the formation of the two halves of molecules such as colchicine (**45**, **Figure 7**).^{71–76} Currently, how such metabolic partitioning is achieved, particularly with both phenylalanine/tyrosine ammonia lyases (PAL/TAL) being cytosolic, is unknown.

6.17.2.5 Bacterial Prephenate Dehydratases and Plant Arogenate Dehydratases: A Structural Biology Comparison via Homology Modeling

Both plant ADTs and bacterial PDTs catalyze a rather unusual decarboxylative dehydration (**Figure 6**), the mechanism of which is poorly understood at the molecular level. Recently, three-dimensional (3D) structures of a *Staphylococcus aureus* PDT (Sa-PDT) in its apo form (**Figure 8(a)**), and that of a *Chlorobium tepidum* PDT (Ct-PDT) complexed with Phe (**1**), were reported, with both determined at 2.3 Å resolution (Sa-PDT, PDB ID: 2QMW; Ct-PDT, PDB ID: 2QMX).¹⁰⁶ Both Sa-PDT and Ct-PDT were found to exist in homotetrameric form being derived from two symmetric dimers (dimer-of-dimer), with each dimer resulting from two monomers situated in an asymmetric orientation. Each monomer also has two distinct domains: in Sa-PDT, for example, there is an N-terminal domain formed by two subdomains, PDTa (residues 1–81 and 168–176) and PDTb (residues 85–161) with a cleft in between involved in catalysis, and a smaller C-terminal ACT-like regulatory domain (derived from aspartokinase, chorismate mutase, and TyrA; residues 185–264) (**Figure 8(a)**).

Despite the low degree of identity (~27%) between the catalytically most active ADT (AtADT2) and Sa-PDT, however, a putative structure of AtADT2 monomer in one asymmetric unit can be modeled herein (**Figure 8(b)**) using the PHYRE protein fold web server (version 0.2). Provisionally, the corresponding catalytic N-terminal domain also has a cleft between the two putative subdomains (ADTa and ADTb), and the putative regulatory ACT domain is quite similar to that of Sa-PDT and Ct-PDT, indicating that the participating amino acid residues for substrate binding and/or catalysis, as well as regulation, might be closely related among these enzymes.

Analyses of multiple sequence alignments of PDTs, and the recently established ADTs, also indicate that all contain a highly conserved peptide Thr-Arg-Phe (TRF) motif^{53,107} located in the cleft region between the PDTa/PDTb subdomains of Sa-PDT and Ct-PDT,¹⁰⁶ and the putative cleft region between ADTa/ADTb subdomains of AtADT2.

In this context, site-directed mutagenesis of the distantly related *E. coli* P-protein harboring both CM and PDT domains had suggested that residue Thr278 of its TRF motif was critical as regards PDT (-like) activity,¹⁰⁷ although the precise biochemical role of these residues was not established. Indeed, even with the 3D structures of Sa-PDT and Ct-PDT now available, no new insights into substrate binding and catalysis have been obtained. This structural study did establish however that the TRF motifs in Sa-PDT (Thr168, Arg169, and Phe170, **Figure 8(c)**), as well as in the provisionally modeled AtADT2 (Thr268, Arg269, and Phe270) herein, were present in α -helices associated with the cleft between the two N-terminal PDT subdomains (PDTa and PDTb) and ADT subdomains (ADTa and ADTb). It will thus be instructive in future to obtain PDT and/or ADT binary complexes with prephenate (**38**) and/or arogenate (**41**), in order to establish the precise mode of binding of the substrate and reaction mechanism at the proposed catalytic site.

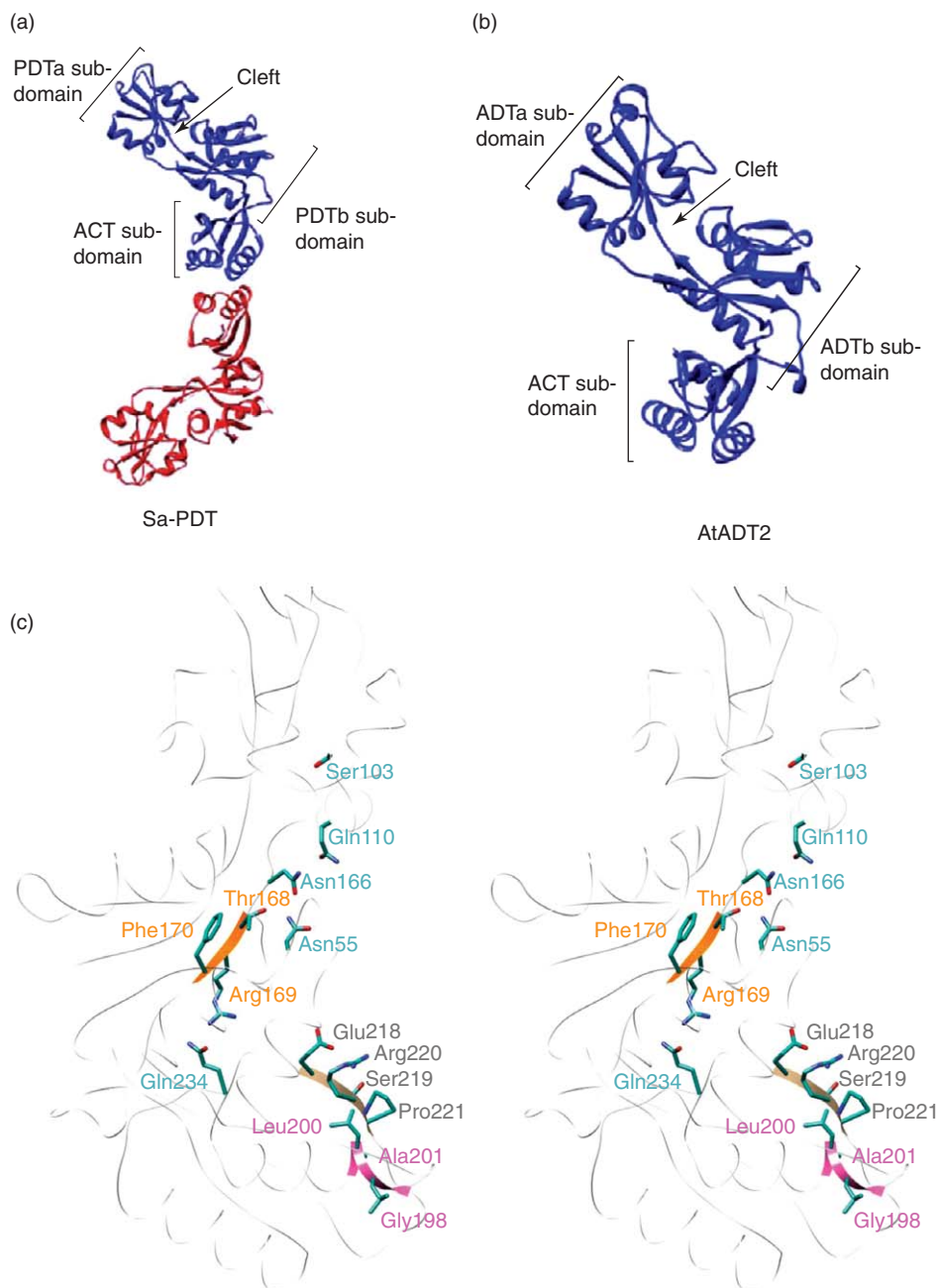


Figure 8 Bacterial prephenate dehydratase (PDT) and homology modeling of *Arabidopsis* arogenate dehydratase (AtADT2). Ribbon representation of (a) Sa-PDT dimer (PDB ID: 2QMW) and (b) homology modeled AtADT2 monomer in one asymmetric unit with the two subdomains (PDTa and PDTb or ADTa and ADTb, respectively) having a cleft in between, as well as a corresponding C-terminal ACT subdomain. (c) Stereoview of Sa-PDT active site conserved residues highlighting the tripeptide Thr168, Arg169, and Phe170 (TRF) motif (in orange), the Glu218, Ser219, Arg220, and Pro221 (ESRP) motif (in gray), and the homologous GXCX (Gly198, Leu199, Leu200, and Ala201) motif (in pink) (Leu199 is not shown) as well as Asn55, Ser103, Gln110, Asn116 and Gln234. All ribbon diagrams and stereoview of interaction of active site residues were prepared using the molecular modeling program UCSF Chimera (<http://www.cgl.ucsf.edu>), unless otherwise indicated.

Some additional insights into allosteric regulation of PDTs and possibly ADTs have also provisionally been obtained by comparison of the apo form of Sa-PDT with that of the binary complex of Ct-PDT with Phe (1).¹⁰⁶ This analysis indicated that the ACT domain interface in dimeric form is present in either a relaxed (R) or tense (T)

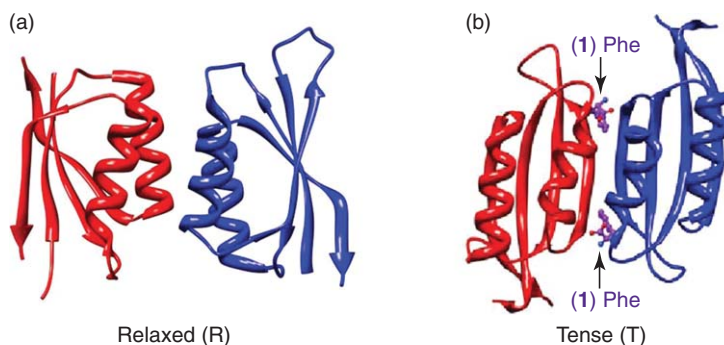


Figure 9 Proposed allosteric regulation of PDT through conformational changes from relaxed (R) state to tense (T) state. (a) Ribbon representation of ACT domain interface of Sa-PDT dimer (PDB ID: 2QMW) in an active open conformation (relaxed, R) and (b) the ACT domain interface of Ct-PDT dimer (PDB ID: 2QMX) in a closed conformation (tense, T) upon binding of L-Phe (1) (purple).

state, depending on whether Phe (1) is bound or not at the interface between the two monomeric ACT regulatory domains of the enzyme (Figure 9). Based on this interpretation, allosteric regulation can thus be considered to occur via a conformational change from an active open (R) to a closed (T) form; that is, during the process of Phe (1) binding at the ACT interface, the two PDT α subdomains of the dimer push toward each other, while the two PDT β subdomains of the dimer move away from each other. A similar consideration can also be made for AtADT2.

Two other highly conserved motifs in the proposed regulatory region of the *E. coli* P-protein were also suggested to be involved in Phe (1) binding and feedback inhibition, respectively, namely, the hydrophilic motif (Glu329-Ser330-Arg331-Pro332 or ESRP) and the hydrophobic motif (Gly309-Ala310-Leu311-Val312 or GXLX). Indeed, mutational and isothermal calorimetric analyses indicated that mutations of residues comprising the ESRP, as well as the GXLX, motifs in the P-protein reduced sensitivity to Phe (1)-mediated feedback regulation.^{108,109} Interestingly, the ESRP motifs (Glu218-Ser219-Arg220-Pro221 in Sa-PDT and Glu319-Ser320-Arg321-Pro322 in AtADT2) and the GXLX motifs in Sa-PDT (Gly198, Leu199, Leu200, Ala201) and AtADT2 (Gly299, Val300, Leu301, Phe302) are both located at the interface between the two ACT regulatory domains in the dimers (Figure 8(c)). As discussed above, in the CtPDT enzyme, the ACT domain interface (i.e., the ESRP and GXLX motifs) was shown to directly bind Phe (1), and thus to induce the T state.¹⁰⁶ Presumably, a similar situation holds for ADTs/PDTs since these also contain both ESRP and GXLX motifs, but this needs to be experimentally established.

Taken together, while the above studies are of interest, they have as yet provided no definitive insight into the overall catalytic process and/or the mechanism of either PDTs or ADTs. Indeed, it is now essential to precisely identify the catalytically active amino acid residues participating in the decarboxylation/dehydration process for both ADT and PDT, and how the overall biochemical mechanism is effectuated and regulated at the molecular level. In terms of catalysis, the overall reaction presumably proceeds via activation of either the carboxylate or the hydroxyl leaving group of the substrate, with a possible mechanism given in Figure 10. This assumes that the ADT (and PDT) facilitates orientation of the appropriate polar amino acid residues of the substrate(s) in the active site(s) to interact with either the carboxylate and/or the hydroxyl groups of the substrate for facile decarboxylative dehydration.

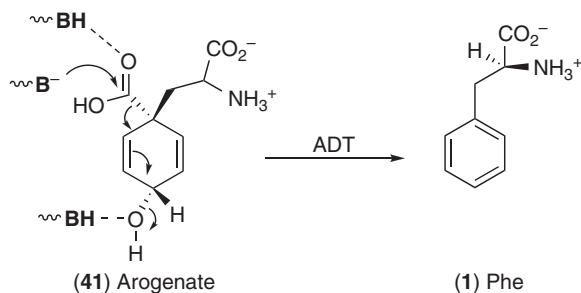


Figure 10 Proposed mechanism for decarboxylative dehydration of aroenate (41) to Phe (1) by ADT. B, base.

6.17.3 Structural Biology Considerations for the Monolignol-Forming Biosynthetic Pathway Steps

The biochemical pathways leading to the formation of the monolignols begin from Phe (1)/Tyr (2) via the action of PAL (EC 4.3.1.24) and – to a lesser extent, depending on the species – TAL (EC 4.3.1.23). The downstream biochemical steps thereafter involve various hydroxylations, esterifications/acylations, methylations, CoA ligations, and reductions. There is still much to be understood, however, as regard the regulatory mechanisms and 3D structures of all of the enzymes that participate in monolignol biosynthesis. It must be emphasized that much of what we know of these enzymes stems largely though from studies of related proteins in other biochemical systems. From a structural biology perspective, the major emphasis of current monolignol pathway enzyme studies is mainly directed toward a full understanding of the molecular basis of their substrate specificities/versatilities, including limited substrate degeneracy in some instances.

6.17.3.1 Biochemical Mechanisms of Phenylalanine and Tyrosine Ammonia Lyases: Comparison to Histidine Ammonia Lyases

Our earlier comprehensive review⁸ had summarized both the discovery and purification of PAL (TAL), as well as the central reliance of these pioneering studies in comparison with that of the related histidine ammonia lyase (HAL, EC 4.3.1.3). PAL and TAL, the most extensively studied enzymatic conversions in plant secondary metabolism, catalyze reversible eliminations of ammonium ion from Phe (1) and Tyr (2) to yield *E*-cinnamic (3) and (*E*)-*p*-coumaric (4) acids (Figure 11), respectively. However, it is important to note that PAL/TAL are cytosolic as compared to the upstream shikimate/chorismate pathways that are found in chloroplasts/plastids. That is, for monolignol pathway metabolism/lignification, both Phe (1)/Tyr (2) are transported into the cytosol. Our current level of knowledge as regards Phe (1)/Tyr (2) transport to the cytosol, with subsequent nitrogen recycling (ammonium ion transporters/and ammonium ion re-assimilation back into the chloroplast) followed by generation of the amino acid donor (Gln) for Phe (1)/Tyr (2) biosynthesis through GS/GOGAT, is however discussed elsewhere.^{110–112}

Depending on the organism, cytosolic PAL/TAL can have differing efficacies for Phe (1) and Tyr (2).^{113–120} For example, PAL from dicotyledonous plants, such as *Arabidopsis*,¹²⁰ is highly specific for Phe (1), whereas the corresponding homolog in certain purple phototrophic bacteria has enhanced specificity toward Tyr (2), and is thus classified as a TAL.¹¹⁶ (In the latter example, the *p*-coumaric acid (4) so formed is a precursor of the chromophore of photoactive yellow protein.) Conversely, PAL/TAL can have similar substrate versatilities for both Phe (1) and Tyr (2) in monocots.^{113,115} The molecular weights of PAL and TAL range from 220 to 340 kDa, with these proteins being found in tetrameric form (~55–85 kDa per monomer).¹¹⁹ Neither enzyme, however, has any cofactor or metal ion requirement for catalysis.¹¹⁹

More recent contributions have mainly focused upon PAL/TAL recombinant protein analyses and their 3D structures, as well as attempts to establish their kinetic properties and biochemical mechanisms (discussed below).^{8,117,118,120–122}

6.17.3.1.1 HAL, PAL, and TAL: Discovery of the MIO prosthetic group

PAL/TAL both belong to the L-amino acid ammonia lyase family, which catalyzes the formation of various α,β -unsaturated acids by elimination of ammonia (ammonium ion) from the corresponding L- α -amino acids. This family of proteins includes aspartate ammonia lyase (AAL), methylaspartate ammonia lyase (MAL), HAL,

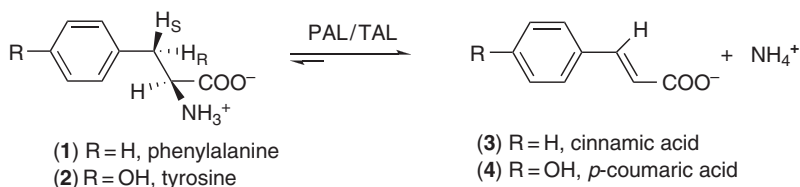


Figure 11 Reactions catalyzed by phenylalanine and tyrosine ammonia lyases.

PAL, and TAL, with the β -eliminations being mediated by proton abstraction from the β -position with concomitant removal of the amino group from the α -position of the substrate amino acids. Although originally demonstrated with HAL,^{123,124} deamination of both Phe (**1**) and Tyr (**2**) was also shown to proceed by loss of the pro-3*S*-proton from the α -carbon via trans-elimination (see **Figure 11**).^{125,126}

Elimination of ammonia from amino acids is, however, not readily achievable chemically at physiological pH. Under such conditions, the amino acids are zwitterionic in character, and treatment with base abstracts instead a proton from the α -ammonium group due to its more acidic nature. Indeed, a direct chemical-based elimination can only be effectuated, for example, by peralkylation of an amino group to afford a positively charged quaternary ammonium ion species, which can then undergo Hoffman elimination when exposed to a strong base, for example, conversion of ergothionine (**46**) into 2-mercapto-urocanate (**47**) (**Figure 12(a)**).¹²⁷

Conversely, L-amino acid ammonia lyases utilize unmodified substrates for catalysis. For example, in conversions catalyzed by AAL and MAL,^{128,129} their substrates are deaminated directly through carbanion intermediates (not shown) without the need for any organic cofactor. Furthermore, a detailed 3D structural analysis of AAL¹³⁰ suggested that it utilized the unmodified amino acid side chain of its substrate via an E1 cb-type mechanism – it was postulated that the γ -carboxyl group in L-aspartate was sufficient to activate its β -proton for abstraction. However, with HAL, PAL, and TAL, the β -proton in each substrate is less acidic than that of L-aspartate, thereby making it less favorable for abstraction. As a result, it was envisaged that these enzymes had a different mechanism to increase the relative acidity of their β -hydrogen atoms.

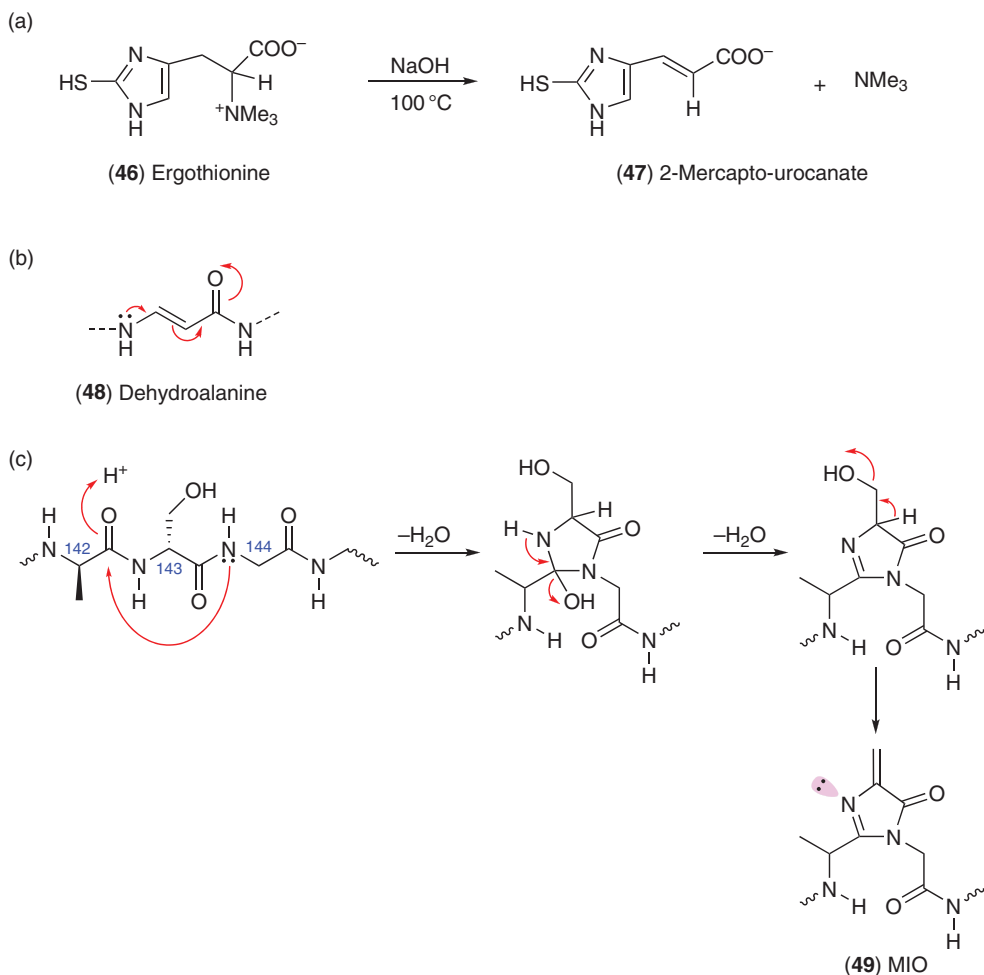


Figure 12 Toward understanding the molecular basis of HAL, PAL, and TAL catalysis. (a) Hoffman elimination of trimethylamine from α -amino acids, for example, ergothionine (**46**). (b) Putative dehydroalanine (DHA, **48**) substructure. (c) Proposed mechanism for the formation of 4-methylidene-imidazole-5-one (MIO, **49**) substructure.

Interestingly, prior to determination of HAL, PAL, and TAL 3D structures, a putative mechanism for these elimination reactions was proposed based on studies of HAL and other deaminases,^{123,124,131,132} where it was considered that they contained a dehydroalanine (DHA, **48**, **Figure (12b)**) moiety as a prosthetic group.^{123,124} This was ultimately based on the observation that all HAL, PAL, and TAL proteins contain the conserved tripeptide motif Ala-Ser-Gly, the Ser residue of which was envisaged to generate the DHA moiety via posttranslational modification.^{131,132} The DHA (**48**) moiety is known to occur naturally in peptide antibiotics (lantibiotics) such as nisin,¹³³ and can also be formed by side-chain water elimination from serine residues in proteins following base treatment.^{134,135} In apparent agreement with this proposal, site-directed mutagenesis of HAL from *Pseudomonas putida* established that the corresponding Ser143Ala mutant was essentially catalytically inactive, suggesting that this residue indeed harbored the HAL prosthetic group precursor.¹³¹ Additionally, site-directed mutagenesis studies replacing Ser202 by either alanine or threonine in PAL from parsley (*Petroselinum crispum*) also resulted in loss of catalytic activity, thus again indirectly suggesting that the putative DHA (**48**) prosthetic group was being formed posttranslationally at the Ser202 residue.¹³⁶

Next, it was envisaged that in HAL, PAL, and, by extrapolation, TAL, nucleophilic attack by the amino group of each substrate occurred on the putative DHA (**48**) prosthetic group. This was thus considered to result in covalently bonded enzyme–intermediate complexes, whose substrate β -carbon hydrogen atoms would now be sufficiently acidic for facile abstraction. However, the electrophilic nature of the putative DHA (**48**) residue could also be significantly reduced via delocalization of the lone pair of electrons in the nitrogen atom into a Michael system (**Figure 12(b)**), thereby limiting the capacity for efficient nucleophilic attack by the amino group of the substrate. This implied that perhaps some other form of chemical modification was needed in the proposed prosthetic group to account for its enhanced electrophilicity, but where delocalization of the nitrogen lone pair to the α,β -saturated system could not occur.

Analysis of HAL (2.1 Å resolution) from *P. putida* has provided insights into the mechanism of action.¹³⁷ Although this established that HAL existed as a homotetramer of D_2 symmetry, the electron density around residue 143 was inconsistent with the presence of the putative DHA (**48**) moiety; indeed, the proposed DHA moiety was not compatible with the documented presence of a carbonyl oxygen at the adjacent Ala142 residue.¹³⁷ Instead, a novel reactive electrophile, 4-methylidene-imidazole-5-one (MIO, **49**), acting as prosthetic group was discovered, this being autocatalytically formed by cyclization and dehydration of residues within the conserved Ala142-Ser143-Gly144 motif^{137,138} (**Figure 12(c)**). MIO (**49**) can, however, be regarded as a modified form of DHA (**48**) with enhanced electrophilicity, as cyclization to the imidazolone ring restricts planarization of the amide nitrogen (Gly144). That is, the polypeptide fold prevents delocalization of the nitrogen lone pair into the α,β -unsaturated system, which, in turn, results in a stronger electrophilic capacity for the prosthetic group.

Interestingly, for X-ray diffraction studies, Schwede *et al.*¹³⁹ obtained high-quality crystals of HAL from *P. putida* only after ‘crystal engineering’ to prevent nonspecific aggregation. This was achieved by specific mutation of the ‘solvent-exposed’ Cys273 residue to alanine,¹⁴⁰ which diminished undefined aggregations and resulted in well-defined crystals. The 3D structure so obtained was not only consistent with the interpretation of the observed electron density, but it also provided a detailed geometry of the refined electrophilic group, MIO (**49**).^{137,141}

Amino acid sequence homologies between PAL and HAL also initially permitted the parsley PAL 3D structure to be constructed through homology modeling (as for ADT/PDT).¹⁴² This model provisionally suggested that the active site of PAL closely resembled that of HAL, with both having the same Ala-Ser-Gly tripeptide signature and thus the prosthetic MIO group (**49**).^{137,138} Detection of MIO (**49**) in both PAL and HAL was also achieved spectroscopically by isolation of separate peptide fractions after individual pronase digestion of both enzymes.¹⁴³ Comparison of UV difference spectra from 240 to 360 nm of the HAL mutant, Ser143Ala, and the wild-type (WT) HAL, as well as the PAL mutant, Ser202Ala, and the WT PAL, measured at enzyme concentrations of 0.2, 0.4, 0.8, and 1.6 mg ml⁻¹, respectively, provided additional evidence for the presence of the MIO prosthetic group (**49**) in both WT enzymes.¹⁴³ (The UV spectrum of both WT HAL and PAL exhibited a UV maximum between 305 and 310 nm consistent with the MIO residue (**49**), whereas both mutants lacking the conjugated system did not.)

Detailed structure–function analysis of PAL from the yeast *Rhodospiridium toruloides*^{144,145} and parsley¹⁴⁶ has further established both the existence of the prosthetic MIO group (**49**) and PAL homotetrameric nature as for HAL.¹³⁷ Furthermore, for illustrative purposes, a comparison of the 3D structures of HAL, PAL, and TAL in their apo forms is depicted in **Figure 13**. The binary complex of the *R. toruloides* PAL (2.1 Å resolution)¹⁴⁴ with *E*-cinnamic acid (**3**) was shown to be a homotetramer, with each monomer having a molecular weight of

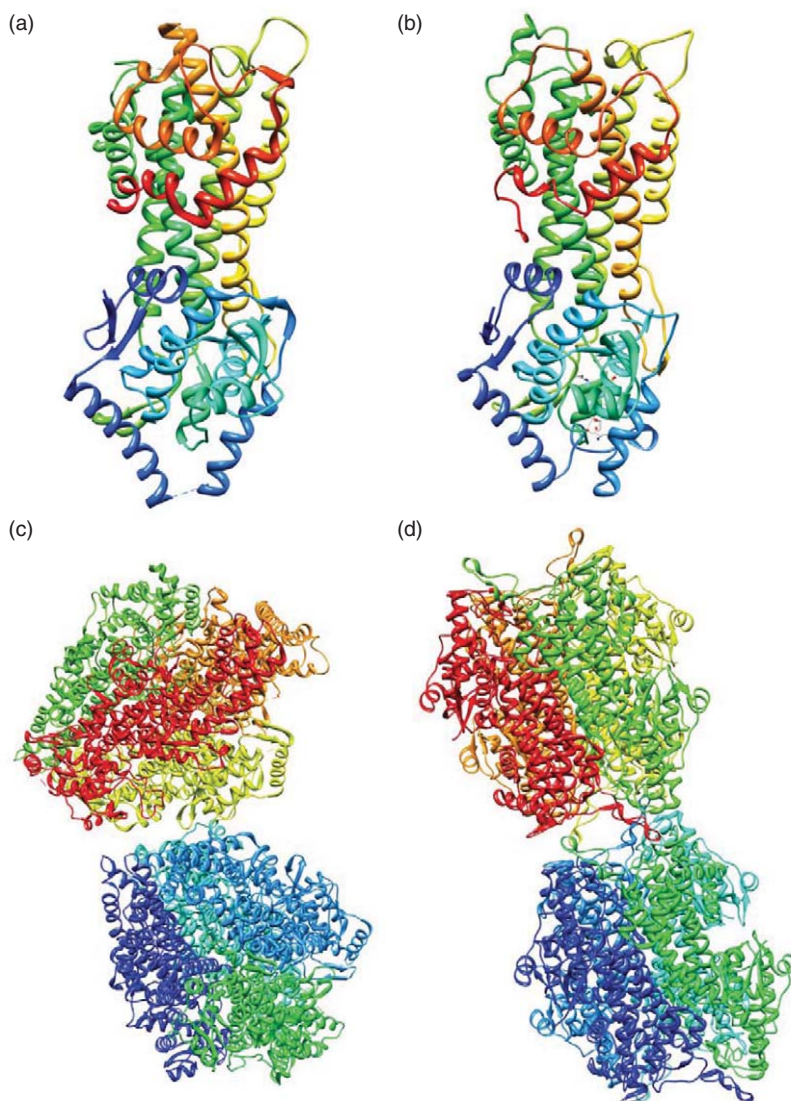


Figure 13 Ribbon representation of 3D structures of (a) monomeric form of NpPAL from *Nostoc punctiforme* (PDB ID: 2NYF), (b) monomeric form of HAL from *Pseudomonas putida* (PDB ID: 1B8F), (c) homotetrameric form of PAL from *Rhodosporidium toruloides* (PDB ID: 1T6P), and (d) homotetrameric form of TAL from *Rhodobacter sphaeroides* (PDB ID: 2O6Y).

approximately 77 kDa and containing 716 residues per subunit arranged in a 222 symmetry. The electron density observed was again consistent with the presence of the MIO (49) prosthetic group (residues 211–213), where again the PAL homotetramer contains four individual MIO (49) triad residues with its anchoring helices.¹⁴⁵ No new insights were, however, gained as regards *E*-cinnamic acid (3) binding and the amino acid residues involved. Interestingly, the eukaryotic PALs are approximately 20 kDa per subunit larger than prokaryotic HALs, as they additionally contain approximately 54 amino acid residues in the N-terminal extension as well as an inserted 122-residue multihelix region in the C-terminal domain.

6.17.3.1.2 Putative catalytic mechanisms for HAL, PAL, and TAL

Two possible mechanisms both E1cb-like^{127,138,147,148} have been proposed for HAL and PAL to help define the role of the electrophilic MIO cofactor (49) in catalyzing α -elimination of the amino groups and subsequent stereospecific β -proton abstractions from their substrates. For mechanism A¹⁴⁸ (Figure 14(a), shown only for PAL), the

conversion can be envisaged as being initiated by an electrophilic attack on the methyldene group of the MIO (49) by the $-\text{NH}_2$ group of the substrate (1), thereby generating an intermediate protein–substrate complex (50) bearing a positive charge on the nitrogen atom. The electron-withdrawing nature of the nitrogen atom facilitates stereospecific deprotonation of the β -proton to generate carbanion 51, which then undergoes elimination as shown to afford intermediate 52 and cinnamate (3). The MIO– NH_2 complex (52) is then finally protonated by a general acid, which results in release of ammonium ion and regeneration of the MIO cofactor (49).

For mechanism B^{127,138,147} (Figure 14(b), shown only for PAL), a Friedel–Crafts-type reaction is envisaged where the electron-rich aromatic ring of the substrate amino acid is attacked by the MIO electrophile (49), thereby putatively generating carbocation intermediate 53. The positive charge on the latter is in the vicinity of the β -hydrogen, thereby enhancing β -proton abstraction by the enzymatic base to afford the putative intermediate complex 54. The catalytic cycle can then be completed by fragmentation of intermediate 54, this being accompanied by elimination of ammonium ion and release of cinnamate (3) from the MIO cofactor (49).

The proposed elimination (E1 cb) mechanisms (mechanisms A and B) were the result of studies of the behavior of substrate analogs and secondary deuterium isotope effects in the PAL reaction,¹⁴⁹ solvent isotope exchange studies with HAL,¹⁵⁰ and kinetic isotope effects studies of a ring perdeuterated phenylalanine in PAL.¹⁵¹ Later, a synthetic model putatively mimicking the PAL elimination reaction (mechanism B) was also described¹⁵² (Figure 15), taking into consideration the electrophilic behavior of a Michael acceptor within a sterically appropriate distance from a Phe (1)-like substrate. Upon treatment with a Lewis acid (AlCl_3) and

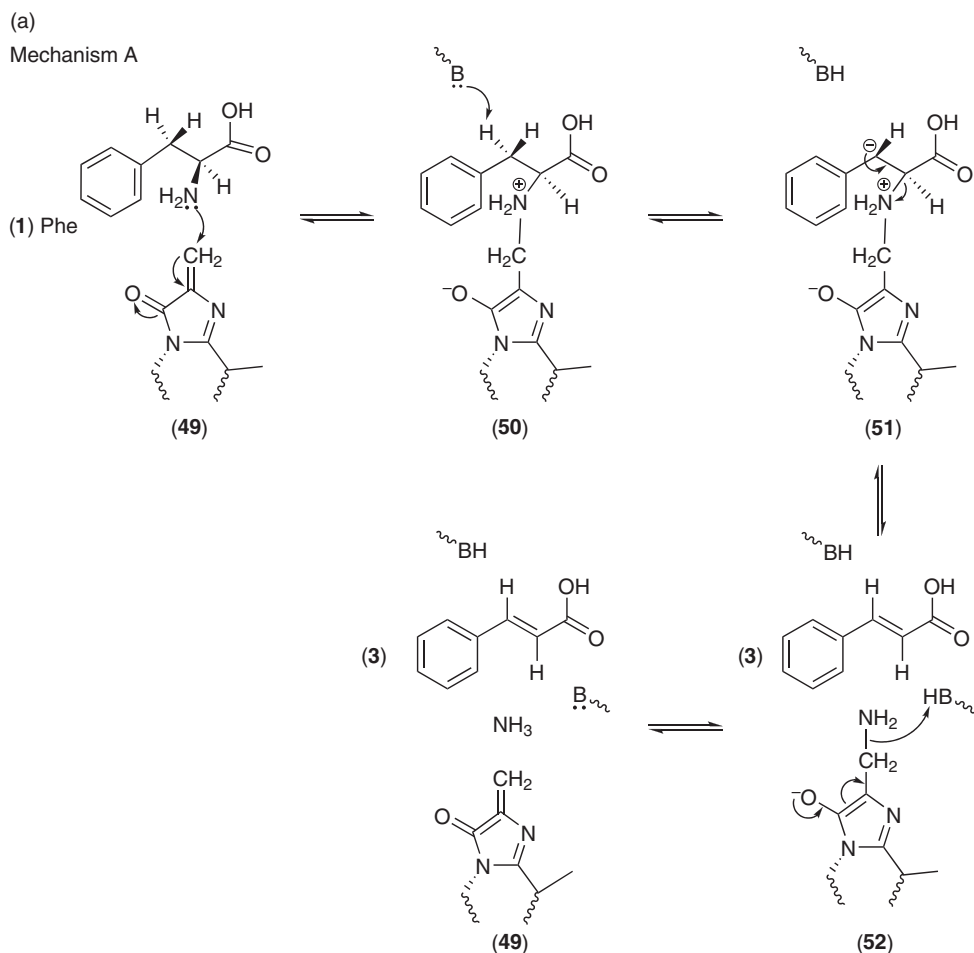


Figure 14 (Continued)

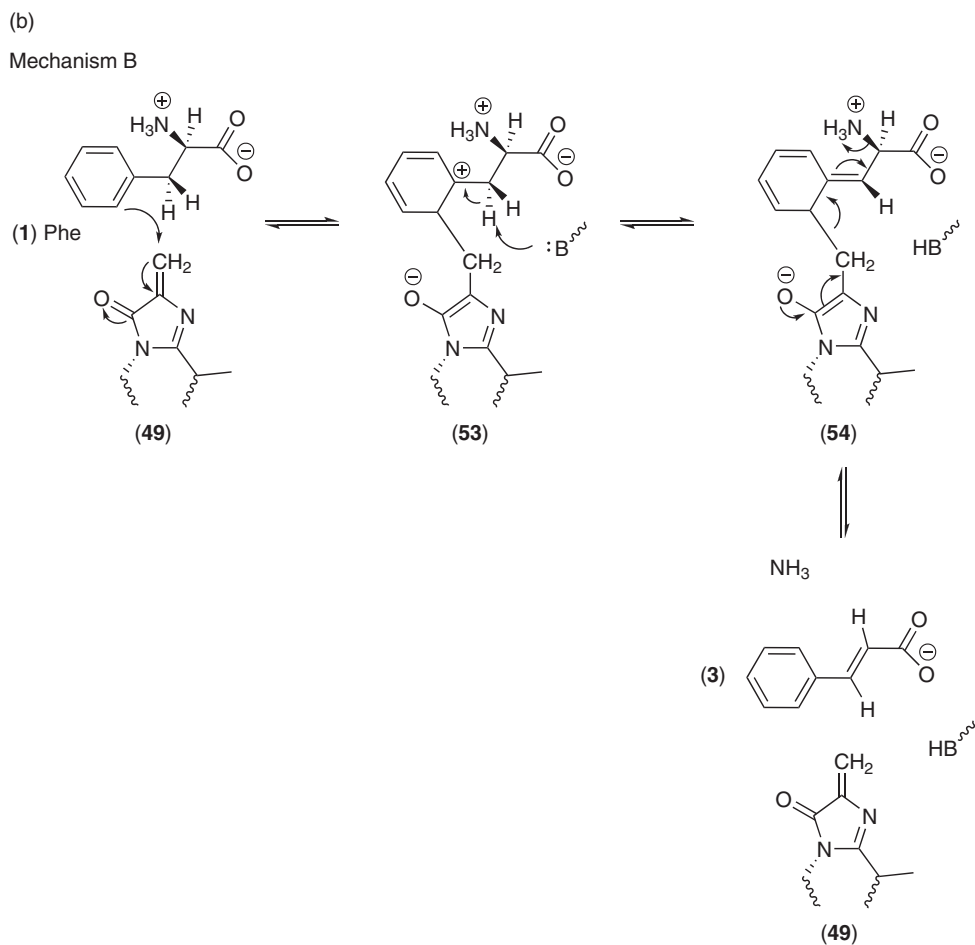


Figure 14 E1 cb-like reaction mechanisms proposed for PAL reaction: (a) mechanism A and (b) mechanism B, a Friedel-Crafts-type reaction.

subsequent hydrolysis, the model compound (\pm)- γ -[(dimethylamino)methyl]-3-methoxy-2,4,6-trimethyl- α -methylidenebenzenebutanal (**55**) (**Figure 15**) resulted in elimination of dimethylamine. However, this elimination reaction did not occur when using the less electron-rich 2,4,6-trimethyl- (**59**) and/or 2,4,6-trimethyl-3-nitrophenyl (**60**) derivatives.¹⁵² This, therefore, provided additional indirect evidence for mechanism B, which can be provisionally considered as the more biochemically favorable elimination mechanism at present.

6.17.3.1.3 HAL and PAL: Proposed Tyr loop-in model for ‘breathing’ motion for substrate access

Apart from posttranslational cyclization of the catalytic Ala-Ser-Gly tripeptide signature residues, Tyr110 (in PAL) and its equivalent Tyr53 (in HAL) are also conserved in both PAL and HAL. Furthermore, site-directed mutagenesis of Tyr53 to Phe (**1**) in HAL from *P. putida*¹⁴¹ and Tyr110 to Phe (**1**) in PAL from parsley¹⁴² resulted in an approximately 2650- and 75 000-fold decrease in k_{cat} , respectively. These residues were thus also considered as essential for catalytic activity, even though Tyr53 and Tyr110 were far removed from the exocyclic methylene C-atom center of the prosthetic MIO group (**49**). These findings led to a proposal of tyrosine-containing loops in both PAL and HAL, involving either Tyr loop-in and/or Tyr loop-out model structures.¹⁵³ Molecular dynamics studies¹⁵³ of the proposed Tyr loop structures suggested that the Tyr loop-in model resulted in a ‘breathing’ motion, thereby opening/closing entrance to the active site for substrate entrance/product release. By contrast,

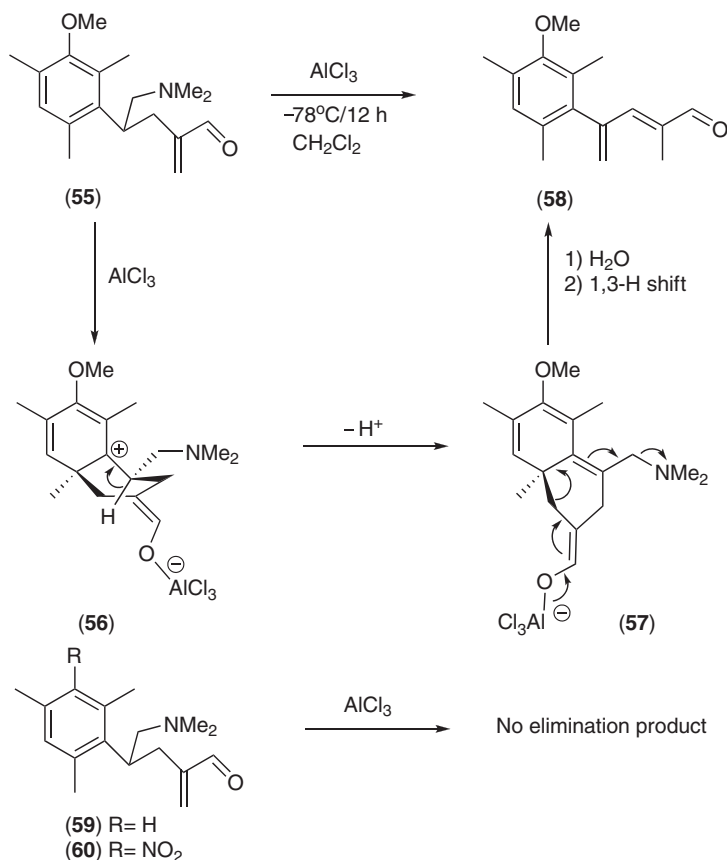
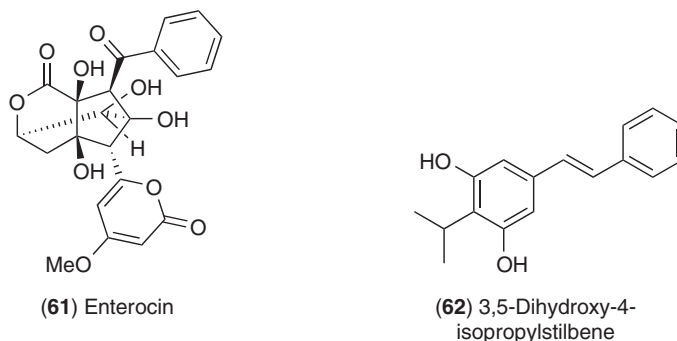


Figure 15 Synthetic models putatively mimicking PAL-catalyzed elimination reactions.



simulation of the Tyr loop-out model structure indicated near irreversible folding with weaker binding of the substrate, and therefore was presumed not to be catalytically productive. A similar loop-in motion for Tyr53 was also anticipated for HAL, although in that case there was no indication of a Tyr53 loop-out structure. Taken together, these structural studies currently envisage that the Tyr loop-in regions in both PAL and HAL are in an essential loop-in conformation, with structural folding, to facilitate catalysis.

6.17.3.1.4 The molecular basis of PAL and TAL substrate versatility

PAL and TAL are apparently represented to a very limited extent in bacteria, this presumably accounting for the general paucity of phenylpropanoids in such organisms. Isolation and characterization of PALs have, however, been described for *R. toruloides*,^{144,145} *Streptomyces maritimus*,¹⁵⁴ *Photobabddus luminescens*,¹⁵⁵ *Sorangium cellulosum*,¹⁵⁶ and *Streptomyces verticillatus*,¹⁵⁷ with the *E*-cinnamic acid (3) so-formed being considered, for

example, as an intermediate for biosynthesis of antibiotics, such as enterocin (**61**)¹⁵⁴ and 3,5-dihydroxy-4-isopropylstilbene¹⁵⁵ (**62**). Bacterial TALs are also found in organisms such as *Rhodobacter capsulatus*,¹¹⁶ *Rho. sphaeroides*,^{118,158} and *Halorhodospira halophila*.¹⁵⁹ Kinetic studies of bacterial TAL demonstrated an enhanced substrate specificity toward Tyr (**2**) (k_{cat}/K_m values $\sim 1780\,000\text{ mol}^{-1}\text{ s}^{-1}$) over Phe (**1**) (k_{cat}/K_m values $\sim 11800\text{ mol}^{-1}\text{ s}^{-1}$).¹¹⁶

The 3D structures of *Rho. sphaeroides* TAL¹¹⁸ in apo form, as well as complexed with *p*-coumaric acid (**4**) (determined at 1.5 Å resolution, **Figure 13(d)**), established the presence of two homotetramers of monomeric polypeptide chains per asymmetric unit, arranged in a 222-point symmetry. The structures of TAL were, as expected, similar to those of *P. putida* HAL¹³⁷ and *R. toruloides* PAL,^{144,145} containing the MIO prosthetic group (**49**) derived from the signature tripeptide Ala149-Ser150-Gly151, with a well-defined electron density. As for PAL,¹⁴⁵ the TAL homotetramer contained four individual MIO (**49**) triad residues, with the MIO (**49**) nitrogen atoms derived from Ser150 also being putatively stabilized by formation of hydrogen bonds with Tyr300 and the side-chain amide of a Gln436 residue. The binding of *p*-coumaric acid (**4**) is envisaged to occur through interaction of the carboxylate group of (**4**) via hydrogen bond formation with an Asn435 residue and a salt bridge with the guanido group of Arg303 residue. The phenyl ring of the bound *p*-coumaric acid (**4**) was also surrounded by the hydrophilic residues Asn432 and Gln436, as well as nonpolar amino acid residues, Leu90, Leu153, and Met405. Additionally, hydrogen bond formation of the hydroxyl group of **4** with the side chain of the His89 residue and a water molecule was documented.

The major difference between PAL from *R. toruloides* and TAL from *Rho. sphaeroides* was an His89 residue conserved in TAL but not in PAL.^{117,118} On substituting His89 with Phe in TAL, the corresponding protein was apparently effectively converted into a PAL.¹¹⁷ Indeed, structural comparison of the WT TAL with the His89Phe TAL mutant established His89 as being the putative determinant of substrate specificity, apparently due to the reduced steric hindrance as compared to the Phe89 mutant, thus conferring selectivity of L-tyrosine (**2**) as substrate in TAL.¹¹⁸

Additional 3D structures of cyanobacterial PALs in their apo form were obtained using *Anabaena variabilis* (ATCC 29413) and *Nostoc punctiforme* (ATCC 29133)¹⁶⁰ (**Figure 13(a)**), with both being approximately 77% identical in amino acid sequence with 567 (*Anabaena variabilis* PAL) and 569 residues (*Nostoc punctiforme* PAL) per subunit. The cyanobacterial PALs were also homotetramers (222 symmetry), with the electron density map again clearly defining the presence of the MIO (**49**) prosthetic group. The nitrogen atom (Ser168) of the imidazolone ring of the MIO (**49**) is hydrogen bonded to Tyr314 residue with the ring stacked against the side chain of the Phe363 residue with an interplanar separation of approximately 4 Å.

Although similar in size to prokaryotic HAL (HAL from *P. putida* has 509 residues), cyanobacterial PALs are approximately 20% smaller than eukaryotic PALs (e.g., PAL from *R. toruloides* contains 716 residues per subunit) but, nevertheless, had significantly similar amino acid sequences to the latter. Relative comparison of the superimposed polypeptide backbone chain of the structures of prokaryotic HAL and eukaryotic PAL with that of the cyanobacterial PAL thus established that the *Anabaena variabilis* homolog differs by virtue of approximately 1.4 Å rmsd for 448 equivalent residues to parsley PAL (identity, $\sim 36\%$), 1.1 Å rmsd for 463 equivalent residues to yeast PAL (identity, $\sim 34\%$), and 1.5 Å rmsd for 450 equivalent residues to HAL from *P. putida* (identity, $\sim 29\%$), indicating that cyanobacterial PALs are more closely related structurally to yeast and plant PALs than prokaryotic HAL. Therefore, cyanobacterial PAL may represent an evolutionary intermediate from which eukaryotic PALs are derived.

6.17.3.2 Hydroxycinnamoyl CoA:Shikimate/Quinate Hydroxycinnamoyltransferase

Hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyltransferase (HCT, EC 2.3.1.133) catalyzes esterification of shikimate (**29**) and quinate (**28**) with *p*-coumaroyl CoA (**9**) to afford the corresponding esters **25** and **24**, respectively. This biochemical conversion was discovered by Stöckigt and Zenk,¹⁶¹ with later studies describing its partial purification and the reversibility of the enzymatic conversion.^{162,163} Interestingly, this step has now been established as the forerunner to the second hydroxylation (at C3) to ultimately afford caffeoyl CoA (**10**) (see **Figure 4**). As for PAL/TAL, this protein is considered cytosolic, as the enzyme contains

no signal peptide¹⁶⁴ and is not colocalized to the endoplasmic reticulum (ER) as for cinnamate 4-hydroxylase (C4H)^{165,166} (discussed below).

Gang *et al.*¹⁶⁷ also preliminarily reported HCT activity with a partially purified protein from sweet basil (*Ocimum basilicum*) peltate glandular trichomes, which apparently displayed a high substrate affinity for *p*-coumaroyl CoA (**9**) and shikimic acid (**29**). Hoffmann *et al.*¹⁶⁴ later described the partial characterization of HCT from tobacco (*Nicotiana tabacum*). When heterologously expressed in *E. coli*, it preferentially utilized *p*-coumaroyl CoA (**9**) as acyl donor, for transfer of the *p*-coumarate unit to shikimic (**29**) or quinic (**28**) acid as acceptors. Moreover, recent studies on gene silencing of HCT in *Nicotiana benthamiana*,¹⁶⁸ *Medicago sativa*,¹⁶⁹ and *N. tabacum* (C. L. Cardenas *et al.*, manuscript in preparation) have since established its involvement in phenylpropanoid/lignin biosynthesis in these species.

Cloning of a cDNA encoding HCT from tobacco (*N. tabacum*, NtHCT) stems and expression of the recombinant protein in fully functional form in *E. coli* as a fusion product have also been carried out, with kinetic parameters of the purified protein determined (C. L. Cardenas *et al.*, manuscript in preparation).¹⁶⁴ The overall catalytic properties of recombinant NtHCT with various hydroxycinnamoyl CoAs (**9**–**13**) using either shikimic (**29**) or quinic (**28**) acid as substrate were quite informative (C. L. Cardenas *et al.*, manuscript in preparation). With shikimic acid (**29**) at a saturating concentration, the observed k_{cat}/K_m of recombinant NtHCT established that the dominant HCT activity was with *p*-coumaroyl CoA (**9**) (114 840 mol⁻¹ s⁻¹) over that of caffeoyl (**10**), feruloyl (**11**), or sinapoyl (**13**) CoA (48 420, 6880, and 420 mol⁻¹ s⁻¹, respectively). Recombinant NtHCT was able, however, to less efficiently transfer hydroxycinnamoyl moieties from *p*-coumaroyl CoA (**9**) and caffeoyl CoA (**10**) to quinic acid (**28**). Indeed, k_{cat}/K_m values at saturating concentration of quinic acid (**28**) were lower by approximately 7.5- and 131-fold for *p*-coumaroyl CoA (**9**) (15 310 mol⁻¹ s⁻¹) and caffeoyl CoA (**10**) (370 mol⁻¹ s⁻¹), respectively, on comparison to catalytic turnover of NtHCT with shikimic acid (**29**) (saturating substrate).

The kinetic studies of the recombinant NtHCT also further provided evidence for the reverse reaction towards formation of the hydroxycinnamoyl CoAs **9** and **10** (C. L. Cardenas *et al.*, manuscript in preparation). In the presence of coenzyme A, NtHCT catalyzed cleavage of the ester bond of the respective hydroxycinnamoyl shikimate esters **25** and **27**. The reverse reaction kinetics demonstrated that NtHCT was able to convert *p*-coumaroyl shikimate (**25**) to *p*-coumaroyl CoA (**9**) and caffeoyl shikimate (**27**) to caffeoyl CoA (**10**), with k_{cat}/K_m values of 31 000 and 19 110 mol⁻¹ s⁻¹, respectively, at saturating concentrations of CoA.

In order to establish the catalytic mechanism/substrate binding mode (B. Youn *et al.*, manuscript in preparation), NtHCT was obtained in its apo form and also as a binary complex with *p*-coumaroyl CoA (**9**) and two different ternary complexes with *p*-coumaroyl CoA(**9**)/shikimic acid (**29**) and *p*-coumaroyl shikimate (**25**)/CoA (at 2.0, 2.7, 2.3, and 2.0 Å resolution, respectively). NtHCT, a monomeric protein, was found to contain two domains, domain 1 (residues 1–185 and 377–395) and domain 2 (residues 186–376 and 396–435), with the overall structure consisting of 15 β -strands and 15 α -helices (Figure 16). Interestingly, domains 1 and 2 are interconnected to form an active site tunnel thereby providing two distinct entry sites into the same tunnel for *p*-coumaroyl CoA (**9**) and shikimic acid (**29**), at opposite faces of the enzyme.

The residues involved in substrate **9** and **29** binding can be deduced from the structural analysis of NtHCT (Figures 17(a) and 17(b)), and include interactions of the backbone nitrogen of the Asp286 residue with the hydroxyl group of the pantothenate unit of *p*-coumaroyl CoA (**9**) and the 5'-diphosphate group of the same with two arginine residues, Arg290 and Arg375. A hydrogen-bonding network from the side-chain hydroxyl group of Ser254 with the diphosphate group and the carbonyl oxygen in *p*-coumaroyl CoA (**9**) to the N_{e1} of Trp373 was also noted. The hydroxyl side chains of residues Ser38 and Tyr40 also anchor the phenolic hydroxyl of the *p*-coumaroyl moiety through an ordered solvent molecule.

The bound CoA moiety had an identical orientation in both binary and ternary complexes of NtHCT, with the glycosidic dihedral angle of the adenosine unit of *p*-coumaroyl CoA (**9**) adopting an anticoinformation in both cases. Moreover, the size and arrangement of the structural elements comprising the entry site of the *p*-coumaroyl CoA (**9**) in NtHCT is such that the adenine base moiety of the CoA cannot fully enter the active site tunnel. The binding of shikimic acid (**29**) in the active site of NtHCT apparently occurs through hydrogen bonding of the C3 and C5 hydroxyl groups with the side chain of Thr36 and Thr371 residues, respectively. In addition, the Arg358 residue accounts for the salt-bridge formation by its guanido group with the carboxyl group of the shikimic acid (**29**).

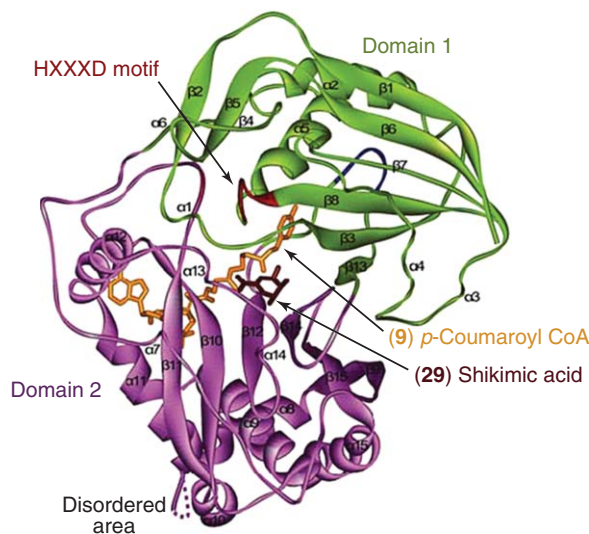


Figure 16 Ternary complex of NtHCT with *p*-coumaroyl CoA (9) and shikimic acid (29). Domains 1 and 2 are depicted in green and violet respectively, with *p*-coumaroyl CoA (9) in orange and shikimic acid (29) in dark-brown. The disordered area of domain 2, residues 250 and 251, is indicated as a dotted line. The ¹⁵³HXXXD¹⁵⁷ motif is shown in red and secondary structure elements are numbered sequentially ($\alpha 1$ – $\alpha 15$ and $\beta 1$ – $\beta 15$).

Using a Dali search in order to identify structural homologs, HCT was found to have structural and sequence motifs characteristic of the CoA-dependent acyltransferase family (B. Youn *et al.*, manuscript in preparation). The results indicate proteins that could be aligned with NtHCT: vinorine synthase from *Rauvolfia serpentina*¹⁷⁰ (PDB code: 2BGH, with 26% sequence identity and Z score of 35.3); a polyketide-associated protein A5 (PapA5) from *Mycobacterium tuberculosis*¹⁷¹ (PDB code: 1Q9J, with 18.8% sequence identity and Z score of 20.7); a nonribosomal peptide synthetase (VibH) from *Vibrio cholerae*¹⁷² (PDB code: 1L5A, with 17.7% sequence identity and Z score of 17.8); choline acetyltransferase from *Rattus norvegicus*¹⁷³ (PDB code: 1T1U, with 17.7% sequence identity and Z score of 10.7); and carnitine acetyltransferase from *Mus musculus*¹⁷⁴ (PDB code: 1NDB, with 17.6% sequence identity and Z score of 5.8).

In this context, the 3D structures of the above mentioned proteins^{170–174} share a common reaction mechanism. All of these proteins contain a conserved HXXXD motif in the active site and a catalytically active histidine residue, located at the center of the active site, with the latter considered as a general base for deprotonation of the hydroxyl group of the substrate prior to acyl group transfer (esterification). However, an apparent exception was noted with VibH – although it contained the conserved HXXXD motif, its histidine residue (His126) did not serve in an equivalent role in acyl transfer catalysis, as its mutation to alanine or glycine had little effect on catalysis.¹⁷²

In NtHCT, however, the HXXXD motif, His153-His154-Ala155-Ala156-Asp157, was located in a loop between $\beta 8$ and $\alpha 5$ forming the interface of domains 1 and 2. Site-directed mutagenesis of His153 with alanine resulted in an enzyme that did not display HCT activity (C. L. Cardenas *et al.*, manuscript in preparation), providing indirect evidence for its involvement in the catalytic process. In addition, the side-chain residue Asp157 of the HXXXD motif was oriented away from the His153 residue and apparently formed a salt bridge with another residue Arg288, thereby eliminating the possibility of any direct interaction with His153. Residues Asp157 and Arg288 are located at the apex positions of domains 1 and 2, respectively. However, mutation of the Asp157 residue by replacement with alanine resulted in an approximately 12-fold decrease in the k_{cat}/K_m values ($9620 \text{ mol}^{-1} \text{ s}^{-1}$) for *p*-coumaroyl CoA (9) and shikimic acid (29), as compared to the catalytic activity of NtHCT ($k_{cat}/K_m \sim 114840 \text{ mol}^{-1} \text{ s}^{-1}$).

Thus, it is considered that the catalytically active His153 residue acts as a general base, which is properly oriented for abstraction of a proton from the hydroxyl group of the C3 atom of shikimic acid (29)

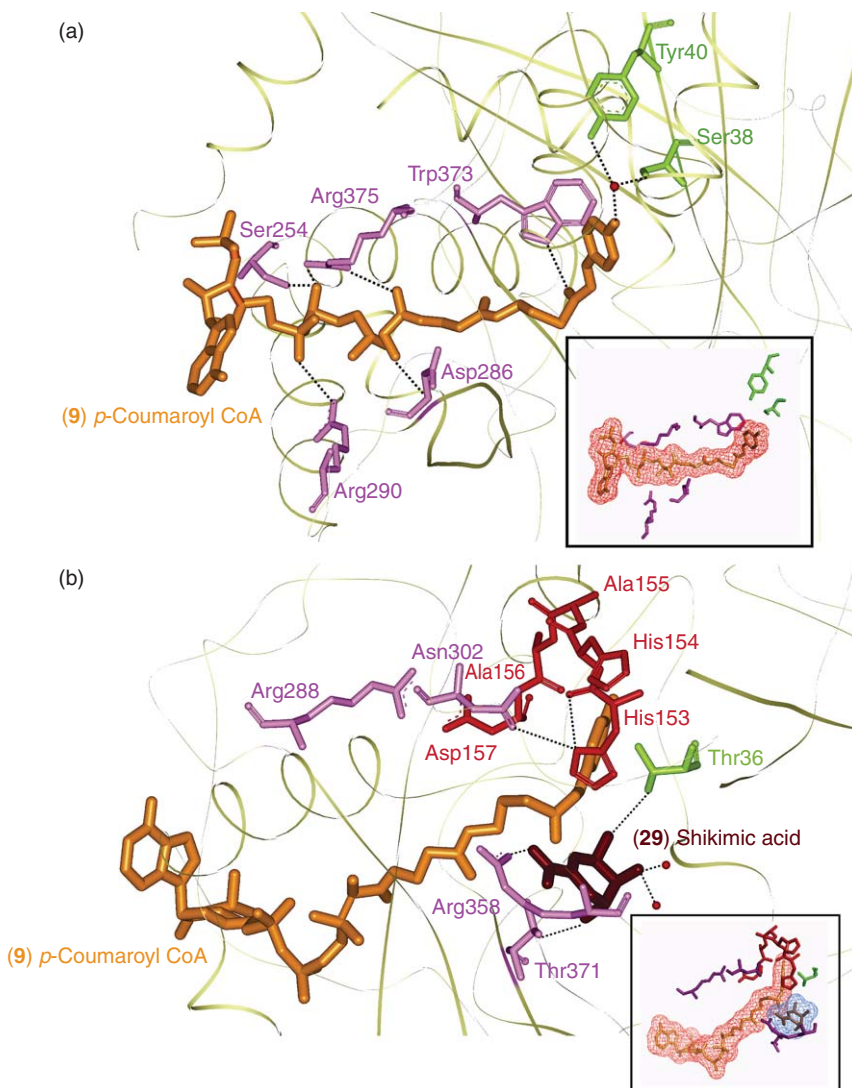


Figure 17 Putative substrate-binding pocket of NtHCT in binary (a) and ternary (b) complexes. The glycosidic bond dihedral angles of the adenosine unit of *p*-coumaroyl CoA (9) in both binary and ternary complexes adopt an *anti* conformation. In insets in the lower right corners of (a) and (b), the experimental difference Fourier maps ($|F_o| - |F_c|$) correspond to *p*-coumaroyl CoA (9) and shikimic acid (29) in binary (a) and ternary complexes (b) of NtHCT, respectively, and are shown at a contour level of 2.5σ . The participating residues for binding and catalytic reaction mechanism are shown in green (from domain 1), pink (from domain 2), and red ($^{153}\text{HXXXD}^{157}$ motif). The salt bridge and/or hydrogen bonds are represented by dotted lines. Water molecules are shown as red spheres.

(Figure 18). This results in nucleophilic attack of the shikimic acid (29) to the carbonyl oxygen of the thioester of *p*-coumaroyl CoA (9) moiety, with generation of the corresponding oxyanion in a tetrahedral transition state. This intermediate can then provisionally be stabilized by hydrogen bond formation with shikimic acid (29), as well as with the hydroxyl side chain of the Thr371 residue, thereby facilitating transfer of the *p*-coumaroyl moiety from *p*-coumaroyl CoA (9), as depicted in Figure 18. Lastly, the phenolic moieties of both *p*-coumaroyl CoA (9) and *p*-coumaroyl shikimate (25) are considered to be indirectly anchored to hydroxyl side chains of Ser38 and Tyr40 through an ordered solvent molecule (Figure 17(a)).

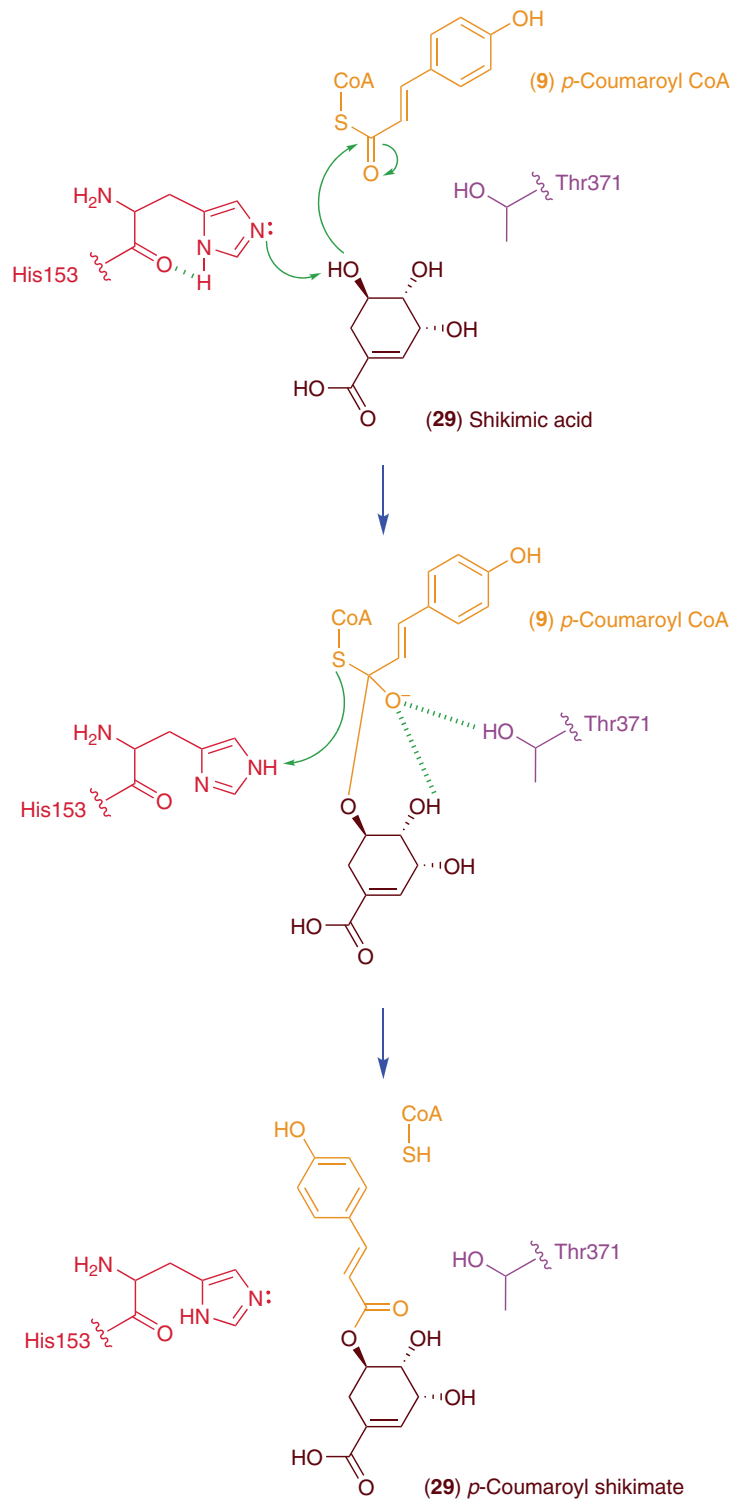


Figure 18 Proposed catalytic mechanism of NthCT.

6.17.3.3 Cytochrome P-450 Hydroxylation Reactions (Cinnamate 4-Hydroxylase, 'p-Coumarate 3-Hydroxylase', and 'Ferulate 5-Hydroxylase'): Comparison to Bacterial/Mammalian P-450s

The pioneering contributions by various researchers in this area were comprehensively discussed earlier.⁸ In this regard, the five monolignols **19–23** (Figure 4), which can ultimately lead to various lignins, lignans, allyl/propenyl phenols, and so on, differ solely in their patterns of hydroxylation/methoxylation at positions 3 and 5 of the aromatic rings. All of these O₂-requiring hydroxylation reactions are cytochrome P-450 mediated,^{175–178} albeit with each displaying a specific regioselectivity. Each P-450 also acts in conjunction with a nicotinamide adenine dinucleotide phosphate (reduced form) (NADPH)-cytochrome P-450 reductase.^{179–181}

In general, cytochrome P-450s have very characteristic UV/visible spectra due to the presence of their heme prosthetic groups, and thus display characteristic absorptions at 450 nm (the so-called Soret band) upon CO binding when reduced with dithionate.¹⁸² While this spectroscopic property has long been used to estimate P-450 contents,¹⁸³ the O₂/NADPH dependence and characteristic light-reversible CO inhibition properties of P-450s are also frequently employed to definitively identify enzymes of this class.

The first monolignol pathway regioselective cytochrome P-450 conversion, catalyzed by C4H (EC 1.14.13.11), introduces a hydroxyl group at C4 of the substrate, *E*-cinnamic acid (**3**). Additionally, if no further hydroxylation occurs on the aromatic ring, the final monolignol pathway product is *p*-coumaryl alcohol (**19**). The second hydroxylation, catalyzed by *p*-coumarate-3-hydroxylase (*p*C3H), utilizes *p*-coumaroyl shikimate (**25**) as substrate to afford the corresponding caffeoyl derivative (**27**). This is then ultimately converted into caffeoyl CoA (**10**), which, if not ring modified further, is normally metabolized to coniferyl alcohol (**21**). The third hydroxylation, catalyzed by ferulate-5-hydroxylase (F5H), introduces a hydroxyl group at the C5 position of the aromatic ring, whose final products in the monolignol pathway are either 5-hydroxyconiferyl (**22**) or sinapyl (**23**) alcohols. That is, regardless of the actual substrate, all three enzymes differ in terms of regioselectivity of hydroxylation at carbons 3, 4, and 5, respectively, on the aromatic ring.

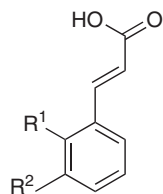
6.17.3.3.1 Subcellular localization of C4H, 'pC3H', and 'F5H'

Unlike either chloroplast/plastid (ADTs)- or cytosolic (PAL/TAL/HCT)-localized enzymes, the P-450s (C4H, 'pC3H', and 'F5H') are all membrane bound, again illustrating the different levels of subcellular organization in the monolignol-generating pathway. Moreover, bioinformatic analyses, such as with the neutral network-based tool, Target P,⁶⁰ have helped identify the signal peptides of C4H, 'pC3H', and 'F5H' that target them to the ER. The signal peptide anchors each protein to the ER membrane by a hydrophobic helix near the N-terminus, followed by a small region rich in basic amino acids and a proline-rich region ((Pro/Ile)-Pro-Gly-Pro-X-(Gly/Pro)-X-Pro) immediately after the N-terminal hydrophobic helix. In addition to being membrane bound, these three P-450s are also glycosylated. Interestingly, C4H, 'pC3H', and 'F5H' have been proposed to provide membrane attachment sites for a scaffold that holds all of the monolignol pathway enzymes from PAL to cinnamyl alcohol dehydrogenase (CAD),¹⁸⁴ although this remains to be definitively established.

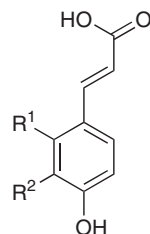
6.17.3.3.2 Cinnamate 4-hydroxylase

C4H was initially discovered in 1967 by Russell and Conn¹⁷⁵ in pea seedlings. Much later, C4H was first purified by Gabriac *et al.*¹⁸⁵ from Jerusalem artichoke (*Helianthus tuberosus*), and the gene encoding C4H (CYP73A1) was also first cloned from the same plant source¹⁶⁵ (discussed fully in Lewis *et al.*⁸)

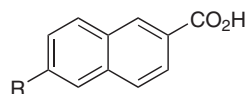
C4Hs have molecular weights ranging from approximately 50 000 to 58 000 depending on the plant species. The catalytic properties of *H. tuberosus* C4H (CYP73A1) heterologously expressed in a *Saccharomyces cerevisiae* strain (WAT11), coexpressing the *Arabidopsis* NADPH-cytochrome P-450 reductase ATR1,¹⁸⁶ established that hydroxylation occurred with *E*-cinnamic acid (**3**) as substrate.¹⁸⁷ Other known plant P-450 substrates such as terpenoids or fatty acids did not serve as substrates. However, while C4H is generally considered substrate specific for *E*-cinnamic acid (**3**), recent studies by Chen *et al.*¹⁸⁸ using the C4H from *Arabidopsis* (CYP73A5) heterologously expressed in *S. cerevisiae* (WAT11 strain) demonstrated that it was much more substrate versatile than previously understood when tested with 14 substrate analogs (**63–76**) *in vitro*. Although C4H was able to hydroxylate various analogs of *E*-cinnamic acid (**3**), that is, **70–76** to afford **77–83**, the most efficiently utilized substrate was still the natural substrate (**3**) with a $k_{\text{cat}}/K_{\text{m}}$ value of 3 430 000 mol⁻¹ s⁻¹.



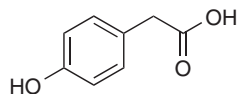
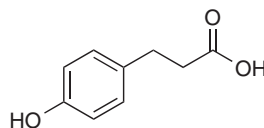
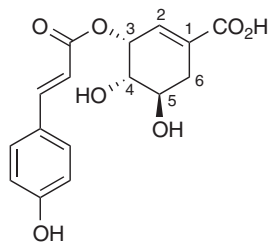
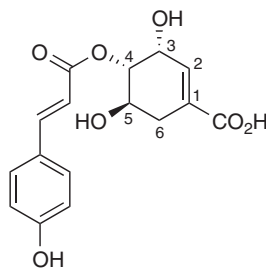
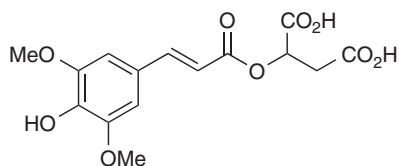
- (63) $R^1 = H, R^2 = Cl$
 (64) $R^1 = H, R^2 = NO_2$
 (65) $R^1 = H, R^2 = OMe$
 (66) $R^1 = R^2 = OMe$
 (67) $R^1 = OH, R^2 = H$
 (68) $R^1 = H, R^2 = F$
 (69) $R^1 = H, R^2 = Me$
 (70) $R^1 = H, R^2 = OH$
 (71) $R^1 = F, R^2 = H$
 (72) $R^1 = Cl, R^2 = H$
 (73) $R^1 = NO_2, R^2 = H$
 (74) $R^1 = Me, R^2 = H$
 (75) $R^1 = OMe, R^2 = H$
 (76) $R^1 = OCH_2Me, R^2 = H$



- (77) $R^1 = H, R^2 = OH$
 (78) $R^1 = F, R^2 = H$
 (79) $R^1 = Cl, R^2 = H$
 (80) $R^1 = NO_2, R^2 = H$
 (81) $R^1 = Me, R^2 = H$
 (82) $R^1 = OMe, R^2 = H$
 (83) $R^1 = OCH_2Me, R^2 = H$



- (84) $R = H, 2\text{-Naphthoic acid}$
 (85) $R = OH, 6\text{-Hydroxy-2-naphthoic acid}$

(86) *p*-Hydroxyphenylacetate(87) *p*-Hydroxyphenylpropionate(88) 3-*O*-*p*-Coumaroyl shikimate(89) 4-*O*-*p*-Coumaroyl shikimate

(90) Sinapoyl malate

Interestingly, 2-naphthoic acid (**84**) could also serve as a mimic for *E*-cinnamic acid (**3**), and this property has since been used for generic fluorometric assays to assess the catalytic activities of CYP73As as the product, 6-hydroxy-2-naphthoic acid (**85**) is strongly fluorescent.¹⁸⁹

6.17.3.3.3 *p*-Coumarate 3-hydroxylase

The C3 hydroxylation step was initially presumed to lead to direct formation of caffeic acid (**5**) from *p*-coumaric acid (**4**) in the monolignol-forming pathway. However, this became the subject of much debate in terms of the nature of the enzyme and/or the substrate(s) involved.⁸ Some researchers reported that ‘*p*C3H’ was an ascorbate-, NADPH-, or flavin adenine nucleotide (FAD)-dependent mixed function oxygenase,^{190–195} or a chloroplast lamella-bound enzyme using plastoquinone or ferredoxin as electron donor¹⁹⁶ with *p*-coumaric acid (**4**) as substrate. Others also reported that 3-hydroxylation instead occurred using substrates such as *p*-coumaroyl CoA (**9**),^{197,198} *p*-hydroxyphenylacetate (**86**), and/or *p*-hydroxyphenylpropionate (**87**),¹⁹⁹ with the enzymes presumed to catalyze this reaction being proposed as a polyphenol oxidase,¹⁹⁹ a soluble FAD-dependent hydroxylase,¹⁹⁸ and a Zn²⁺-dependent dioxygenase that was inactive at the normal cytoplasmic pH.¹⁹⁷ Ultimately, none of these reported conversions were applicable to the monolignol-forming pathway hydroxylation steps.

True identification of ‘*p*C3H’ in the monolignol pathway must be credited to the seminal work of Heller and Kühnl¹⁷⁶ and Kühnl *et al.*,¹⁷⁷ who demonstrated the involvement of a P-450-catalyzing 3-hydroxylation of *p*-coumaric acid (**3**) esters of shikimate (**29**) and quinate (**28**), leading to the formation of caffeoyl shikimate (**27**) and/or caffeoyl quinate (chlorogenic acid, **26**), respectively (see **Figure 4**). A gene encoding a ‘*p*C3H’ (CYP98A3) was later obtained from *Arabidopsis*,²⁰⁰ and the corresponding recombinant protein of molecular weight of approximately 57 800 was also demonstrated to be able to utilize *p*-coumaroyl quinate (**24**) and *p*-coumaroyl shikimate (**25**) as substrates, with the latter being approximately fourfold more efficient ($k_{\text{cat}}/K_m = 1\,457\,000$ vs. $369\,500\text{ mol}^{-1}\text{ s}^{-1}$). Neither *p*-coumaric acid (**4**), *p*-coumaroyl-CoA (**9**), *p*-coumaroyl aldehyde (**14**), *p*-coumaroyl alcohol (**19**), nor the 1-*O*-glucose ester or the 4-*O*-glucoside of *p*-coumaric acid (**4**) served as substrates for hydroxylation at the 3-position. It was further observed that CYP98A3 utilized *E*-*p*-coumaroyl shikimate (**25**) more efficiently than the 3-*O*- and/or 4-*O*-isomeric forms **88** and **89**.

More recently, Mahesh *et al.*²⁰¹ reported isolation and characterization of two other distinct CYP98 cDNAs (CYP98A35 and CYP98A36) from green beans of *Coffea canephora*. Both genes were present as a single copy in the *C. canephora* genome with the encoding proteins displaying 88% similarity to each other. Expression of both proteins individually in *S. cerevisiae* (strain WAT11) and subsequent kinetic studies demonstrated that CYP98A35 had a similar affinity for both *p*-coumaroyl shikimate (**25**) and *p*-coumaroyl quinate (**24**) (4065 and $3720\text{ mol}^{-1}\text{ s}^{-1}$, respectively), albeit with much lower k_{cat}/K_m values as compared to CYP98A3 (see above). CYP98A36 also used *p*-coumaroyl shikimate (**25**) as substrate but poorly ($k_{\text{cat}}/K_m = 2380\text{ mol}^{-1}\text{ s}^{-1}$). Indeed, although the catalytic turnover of both proteins was much lower than with *Arabidopsis* CYP98A3,²⁰⁰ these researchers raised the possibility of CYP98A35 being involved in the biosynthesis of chlorogenic acid (**26**), whereas CYP98A36 was envisaged as required for the formation of monolignols. However, semiquantitative RT-PCR analyses to determine whether both genes showed different tissue-specific expressions, as well as whether their expression patterns correlated with an accumulation of chlorogenic acid (**26**), were inconclusive. What physiological roles both enzymes actually have in *C. canephora* will be useful to unambiguously establish, given that the kinetic data are much lower than that for the *Arabidopsis* ‘*p*C3H’ (CYP98A3), that is, the possibility thus exists for a different metabolic conversion with alternate substrates.

6.17.3.3.4 Ferulate 5-hydroxylase

Ferulate 5-hydroxylase (F5H) catalyzes the third P-450 hydroxylation step at the C5 position of the phenolic ring to ultimately afford, via the monolignol pathway, 5-hydroxyconiferyl (**22**) and/or sinapyl (**23**) alcohols. This P-450 was first detected in xylem and sclerenchyma-enriched tissues from poplar (*Populus × euramericana*), with microsomal extracts able to catalyze hydroxylation of ferulic acid (**6**) to 5-hydroxyferulic acid (**7**) in the presence of NADPH with apparent K_m values of $6.3\text{ }\mu\text{mol l}^{-1}$.¹⁷⁸ This enzyme was thus characterized as F5H.

Later, an *Arabidopsis fab1* mutant was obtained which lacked sinapoyl malate (**90**) that is constitutively present in WT lines. This was considered to result from a deficiency in ‘F5H’ activity,²⁰² with the putative F5H gene (designated CYP84A1) subsequently obtained by T-DNA tagging.²⁰³ It encoded an enzyme of molecular

weight of approximately 58 800, whose Pro450 to Gly460 region was presumed to harbor the heme-binding motif. However, no precise biochemical function of the corresponding putative 'F5H' *in vitro* was established. Instead, a function was only indirectly obtained by complementation of the *fab1* mutant phenotype with the isolated F5H gene, which resulted in *Arabidopsis* plants again accumulating sinapoyl malate (90) in their leaf tissues.²⁰³ This gave no definitive insight, however, into the substrate(s) employed.

Phylogenetic analysis also established that AtF5H (CYP84A1) shared more sequence homology with an avocado ripening-related P-450 (CYP71A1, ~34% identity), as well as with two *Petunia hybrida* flavonoid-3',5'-hydroxylases A and B (CYP75A and CYP75B, respectively, ~32% identity), than with the *H. tuberosus* C4H (CYP73A1, ~29% identity).²⁰³ Interestingly, flavonoid-3',5'-hydroxylases catalyze *meta*-hydroxylation of their flavonoid substrates, thereby raising the possibility that the 'F5H' encoded a *meta*-hydroxylation enzyme.

Lewis and coworkers^{8,204} extensively discussed the uncertainties as regards the nature of the true physiological substrate for 'F5H', this resulting in large part from the findings reported by the Fukushima laboratory.^{205,206} These researchers^{205,206} had suggested that coniferyl alcohol (21) was the preferred substrate for 5-hydroxylation, rather than ferulic acid (6). This was proposed even though ferulic acid (6) had previously been reported as substrate for 5-hydroxylation,¹⁷⁸ and 5-hydroxyferulic acid (7) was a natural product in both maize (*Zea mays*) and barley (*Hordeum vulgare*).²⁰⁷ However, it was also demonstrated that administration of pentadeutero [9-²H₂, OC²H₃]-coniferyl alcohol (21) to shoots of *Magnolia kobus* resulted in its intact conversion into [9-²H₂, OC²H₃]-sinapyl alcohol (23), suggesting again that 21 might be the physiological substrate for 5-hydroxylation.²⁰⁸ (These researchers did not, however, rule out the possibility that the C5 hydroxylation might also partially occur with the corresponding aldehyde (16).)

Following these findings, the question of likely physiological substrate for 'F5H' was further examined by Humphreys *et al.*²⁰⁹ In this regard, following heterologous expression of the *Arabidopsis* 'F5H' in *S. cerevisiae* (WAT11 strain), kinetic analyses suggested that both coniferyl alcohol (21) ($k_{\text{cat}}/K_{\text{m}} \sim 290 \text{ mol}^{-1} \text{ s}^{-1}$) and coniferyl aldehyde (16) ($k_{\text{cat}}/K_{\text{m}} \sim 116 \text{ mol}^{-1} \text{ s}^{-1}$) were preferred substrates when compared to ferulic acid (6) ($k_{\text{cat}}/K_{\text{m}} \sim 0.2 \text{ mol}^{-1} \text{ s}^{-1}$). Furthermore, in a parallel study by Osakabe *et al.*,²¹⁰ it was reported that a recombinant P-450, LsM88, from sweet gum (*Liquidambar styraciflua*) with 75% amino acid sequence identity and 82% similarity to the putative *Arabidopsis* 'F5H' (CYP84A1) preferentially catalyzed the 5-hydroxylation of coniferyl aldehyde (16) ($k_{\text{cat}}/K_{\text{m}} \sim 25\,900 \text{ mol}^{-1} \text{ s}^{-1}$) over that of ferulic acid (6) ($k_{\text{cat}}/K_{\text{m}} \sim 180 \text{ mol}^{-1} \text{ s}^{-1}$).²¹⁰ Catalytic properties using coniferyl alcohol (21) as substrate for 'F5H' hydroxylation were, however, not reported.

In a more recent report, Weng *et al.*²¹¹ identified and characterized a cytochrome P-450 from *Selaginella moellendorffii*, SmF5H (CYP788A1), which was capable of catalyzing 5-hydroxylation of coniferyl aldehyde (16) ($k_{\text{cat}}/K_{\text{m}} \sim 240 \text{ mol}^{-1} \text{ s}^{-1}$) and coniferyl alcohol (21) ($k_{\text{cat}}/K_{\text{m}} \sim 100 \text{ mol}^{-1} \text{ s}^{-1}$) more efficiently than ferulic acid (6) ($k_{\text{cat}}/K_{\text{m}} \sim 5 \text{ mol}^{-1} \text{ s}^{-1}$). As to be expected, a phylogenetic analysis of SmF5H demonstrated that all of the P-450 signature motifs were conserved, although this gene had only approximately 37% sequence identity to the previously identified P-450 CYP84 and CYP75 (flavonoid hydroxylase) families. Interestingly, it does not belong to the same P-450 family as angiosperm F5H (CYP84s).²¹¹

In summary, the pioneering studies of Grand¹⁷⁸ established that the enzyme 'F5H' was a cytochrome P-450. However, there still remains some level of uncertainty about the actual substrate(s) used by putative 'F5Hs' for this conversion. This is because in the above recombinant protein studies from the Chapple and Chiang laboratories,^{209–211} using ferulic acid (6), coniferyl aldehyde (16), and coniferyl alcohol (21) as potential substrates, the kinetic data ($k_{\text{cat}}/K_{\text{m}}$) are several orders of magnitude lower than that obtained for either C4H and *p*C3H'.

6.17.3.3.5 Bacterial/mammalian cytochrome P-450 catalysis/structural studies: Relevance to C4H, 'pC3H', and 'F5H'

Three-dimensional structural studies on C4H, 'pC3H', and 'F5H' have not yet been reported, partly because plant cytochrome P-450 structural studies in general have lagged significantly behind their mammalian and bacterial counterparts.²¹² Accordingly, much of our current understanding on catalysis and structure–function relationships of plant P-450s is derived from structural studies with bacterial/mammalian homologs.^{213–216}

6.17.3.3.5.(i) Catalysis The general catalytic cycle for cytochrome P-450s was initially summarized by Segall *et al.*,²¹⁷ and was adapted for C4H, 'pC3H', and 'F5H'.⁸ Catalytic turnover begins with substrate binding

followed by consumption of reducing equivalents in the form of NAD(P)H via an electron transfer chain toward the oxygen activation by the heme iron. However, with the advent of high-resolution 3D structures, as well as systematic site-directed mutagenesis and molecular dynamics computational studies, our current understanding of oxygen activation in cytochrome P-450s now includes eight intermediates (91–98) in the catalytic cycle.^{218–223} Indeed, the overall structure and chemistries of cytochrome P-450s have recently been updated in an excellent review by Denisov *et al.*,²²⁴ and a modified catalytic cycle is depicted in **Figure 19**.

The steps involved include binding of substrate to a low-spin (LS) ferric state enzyme (91), which is then converted to the high-spin (HS) substrate-bound complex (92) through displacement of coordinated water (as the sixth ligand of the heme iron), and which has a more positive reduction potential for conversion to the ferrous state (93).²²⁴ Oxygen then binds to form the oxy-P-450 complex (94), which is capable of accepting a second electron to afford the ferric-peroxy anion (95), this being protonated to generate the ferric hydroperoxy complex (96). A second protonation next occurs at the distal oxygen resulting in subsequent heterolytic cleavage of O–O with formation of a putative ferryl (Fe(IV)=O) species (97). The latter radical cation can then attack the substrate resulting in formation of the product complex (98), which finally dissociates the hydroxylated product thereby enabling the catalytic cycle to continue.

This catalytic cycle also helps explain the multiple side reactions (shunt paths) of P-450s that can occur *in vivo*.²²⁵ In this regard, three major abortive reactions are possible: (1) auto-oxidation of the oxyferric complex

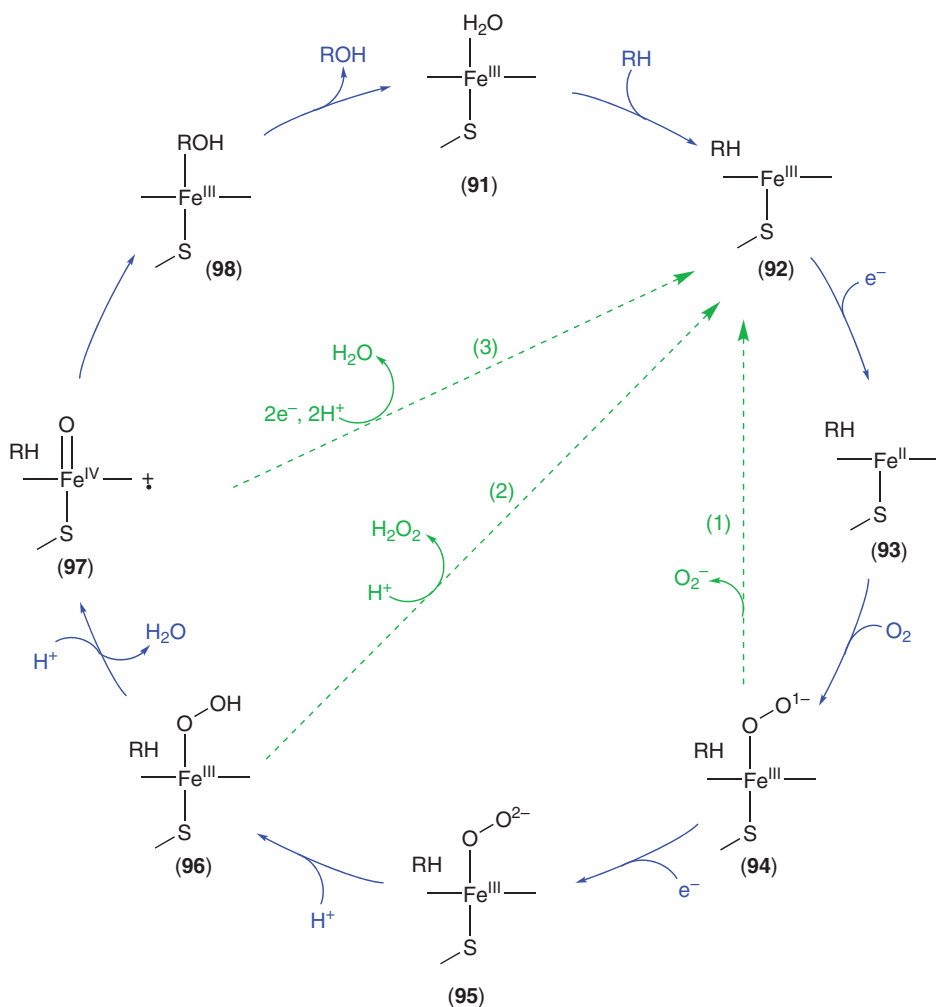


Figure 19 Proposed general catalytic cycle of cytochrome P-450s (reactions (1)–(3) are abortive). Adapted from I. G. Denisov; T. M. Makris; S. G. Sligar; I. Schlichting, *Chem. Rev.* **2005**, *105*, 2253–2277.

(94) with concomitant production of a superoxide anion, (2) dissociation of the hydroperoxide anion (96) from the iron forming hydrogen peroxide, that is, a peroxide shunt, and (3) oxidase uncoupling with oxidation of the ferryl-oxo intermediate (97) resulting in the formation of two molecules of water by a four-electron reduction of the dioxygen molecule.

6.17.3.3.5.(ii) 3D Structures Following the first structural analysis of a cytochrome P-450 from *P. putida*,²¹⁴ considerable progress has been made over the past 7 years in elucidating mammalian/fungal/bacterial P-450 3D structures. Yet, even with more than 5000 different cloned genes encoding various P-450s, structures are currently known for only 26 P-450s: This includes 18 different P-450s from bacteria/fungi and 8 different mammalian P-450s.²¹² There are also two recent reports on 3D structures of the plant P-450, allene oxide synthase (AOS) from *A. thaliana*²²⁶ and guayule (*Parthenium argentatum*).²²⁷ However, the structural studies on AOS are not considered herein for comparative analysis of putative 3D structures of C4H, 'pC3H', and 'F5H'. AOS does not function as a monooxygenase, and requires no molecular oxygen or NADPH cytochrome P-450 reductase for catalysis.

The increased understanding of structure–function relationships of nonplant cytochrome P-450s^{213–216} can also be provisionally applied to C4H, 'pC3H', and 'F5H'. Specifically, all three proteins again contain the highly conserved P-450 signature (heme-binding domain),^{228,229} a motif perfectly centered around its conserved cysteine residue. The latter then both binds and provides the thiolate chain to the heme protein (**Figure 20**). Furthermore, from the various mammalian/bacterial structural studies, it has been demonstrated that substrate recognition sites (SRSs) of P-450s are on the distal pocket to the heme, which include two highly diverse regions, the BC loop and FG loop, and which form an entrance to the pocket.^{230,231}

New insights into the active site of P-450s have been obtained through 3D structural analysis of the xenobiotic-metabolizing mammalian P-450 2B4 (1.9 Å resolution, PDB ID: 1SUO) from rabbit (*Oryctolagus cuniculus*) (**Figures 21(a) and 21(b)**).²³² The interaction of the active site residues and binding of an inhibitor, 4-(4-chlorophenyl) imidazole (CPI, 99), are depicted in the stereoview (**Figure 21(c)**). Binding of the inhibitor is envisaged to occur through interaction with the active site residues, Ile101, Phe115, and Val447, with the chlorine substituent at the 4-position of the phenyl ring of CPI (99). These residues, together with Ile114, Phe297, and Val367, form a hydrophobic region that surrounds the phenyl ring of the inhibitor 99. The imidazole ring of CPI (99) is, however, also anchored by Ala298 (not shown), Glu301, Thr302, and Ile363, with the side-chain oxygens of Glu301 and Thr302 forming a hydrogen-bonding network with the free nitrogen of the imidazole ring of CPI (99).

Structural studies of P-450 2B4 also indicate that Thr302 (**Figure 21(c)**), a highly conserved threonine residue among P-450s, is located several residues upstream of the oxygen-binding site in the heme domain. This threonine residue is suggested to be involved in catalysis, together with two intervening water molecules, via facilitating oxygen activation by transfer of a proton from the surface of the protein to the presumed ferric hydroperoxy complex (96). Taken together, the 3D structure of the P-450 2B4 discussed here can be provisionally anticipated as a reasonable working model for other cytochrome P-450s.

However, with no 3D structures for C4H, 'pC3H', and 'F5H' currently available, this limits any incisive explanation at this point as to how differing regiospecificities are controlled at the molecular level. Nevertheless,

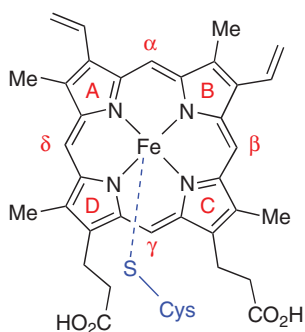


Figure 20 Cytochrome P-450 heme thiolate prosthetic group.

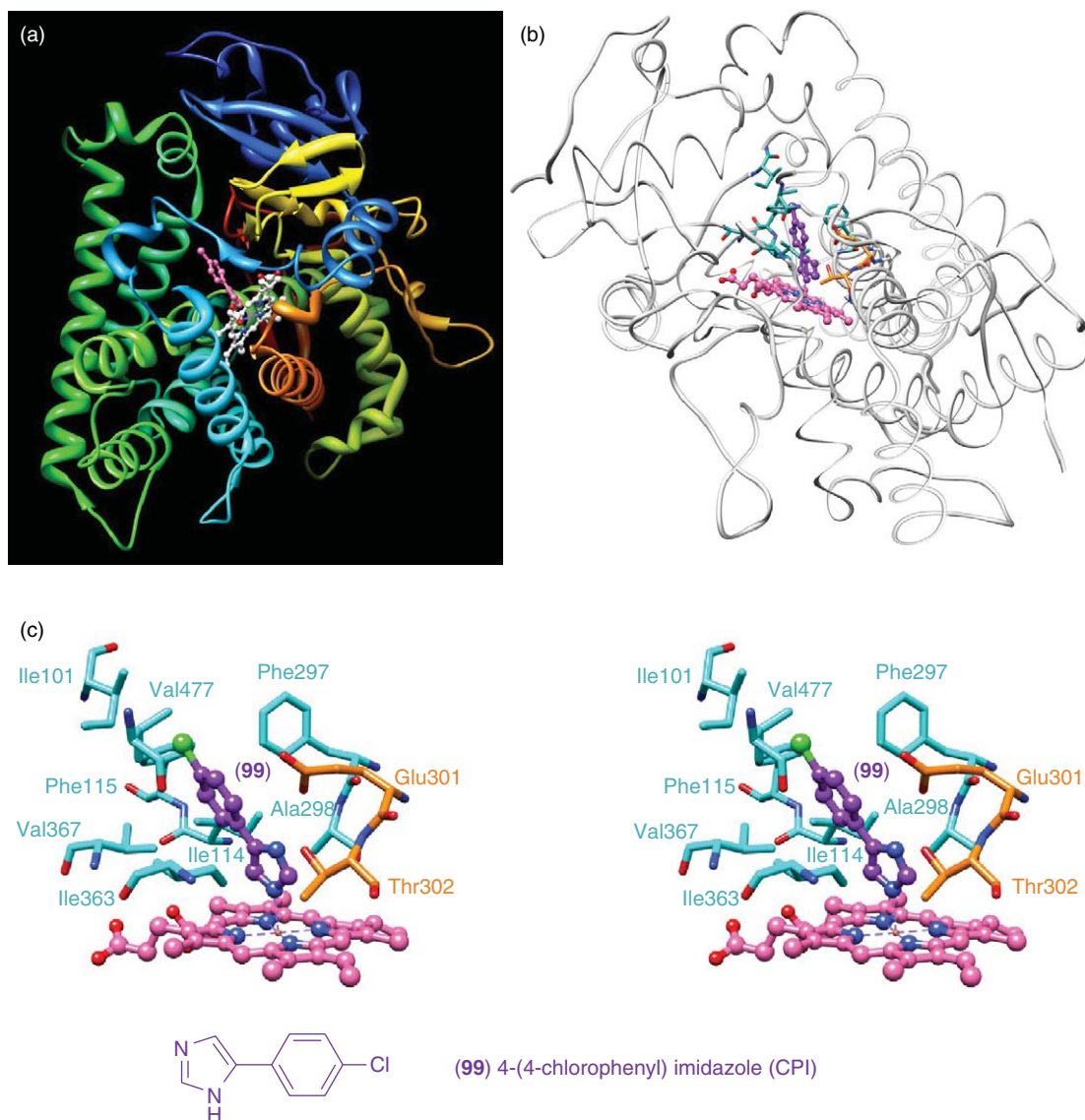


Figure 21 Mammalian cytochrome P-450 2B4. (a) Ternary complex (1.9 Å resolution). (b) Binding site of heme thiolate prosthetic group, with substrate position occupied by 4-(4-chlorophenyl) imidazole (CPI) (99, purple). (c) Stereoview of interaction of active site residues of P-450 2B4 with CPI (99), with the heme prosthetic group shown in pink.

Schalk *et al.*²³³ have attempted to establish a structure–function relationship with C4H. In that study, the presence of a proline residue highly conserved in all plant C4Hs was noted, for example, Pro448 in CYP73A1, which is adjacent to the invariant heme-binding cysteine. (By contrast, this proline residue is only very rarely conserved in the amino acid sequences of animal P-450s.) Site-directed mutagenesis of this residue by replacement with Val, Ile, and Phe (residues found in animal P-450s) mainly resulted in reduced C4H activity in comparison to WT.²³³ That is, the specific activities of WT and mutants Pro448Phe, Pro448Val, and Pro448Ile with *E*-cinnamic acid (3) as substrate were calculated as 683 ± 23 , 24 ± 11 , 199 ± 3 , and 6 ± 0.3 pkat mg^{-1} , respectively. This significant loss of activity, particularly with Pro448Phe and Pro448Ile, was also accompanied by a decrease in CO binding and rate of oxidation of NADPH, suggesting impaired dioxygen binding and reduction capacity in the active site of the heme protein. These findings thus suggest that the Pro448 mutation leads to disruption of the heme–protein interaction, resulting in destabilization of the prosthetic heme group.

More recent studies on the binding of the substrate relative to the active site of *H. tuberosus* C4H (CYP73A1) were also reported by Schoch *et al.*^{234,235} In the first report,²³⁴ replacement of the N-terminus with a peptidic amphipathic sequence PD1 was carried out in order to increase expression of a water-soluble enzyme for ¹H NMR (nuclear magnetic resonance) spectroscopic analyses, with the resulting mutant protein heterologously expressed in *S. cerevisiae* (W(R) strain²³⁶). The paramagnetic relaxation effects of CYP73A1–Fe(III) in the ¹H NMR spectra of the substrate indicated, however, that the heme iron in the active site was positioned parallel to the average initial orientation of the substrate. The second report,²³⁵ describing site-directed mutagenesis and homology-based modeling of the active site of CYP73A1 with that of four other P-450s (i.e., P-450BM-3 (2HPD),²³⁷ P-450_{CAM} (3CCP),²³⁸ P-450_{terp} (1CPT),²³⁹ and P-450eryF (1OXA)),²⁴⁰ also suggested that Asn302 and Ile371 were essential residues for substrate binding and orientation for C4H activity. Additionally, site-directed mutagenesis of the Lys484 residue with methionine (located far enough from the active site not to affect substrate binding) resulted in a 50% decrease in catalytic activity with *E*-cinnamic acid (**3**) and naphthoic acid (**84**) as substrate(s).²³⁵ Thus, from these site-directed mutagenesis and active site modeling analyses, it was postulated that residue Lys484, although not involved in initial substrate binding, significantly contributed toward the substrate reorientation thereby helping to account for the regiospecificity of the catalytic reaction with interactions via a solvent molecule.

In summary, although the three hydroxylation enzymes (C4H, ‘*p*C3H’, and ‘F5H’) have now all been identified, little is yet definitively understood as to how their distinct regiospecificities are attained. This will require more comprehensive and detailed studies than currently available, including fully establishing the molecular basis of substrate binding and catalytic turnover in each case.

6.17.3.4 4-Coumarate:coenzyme A Ligases: Comparison to Other Adenylate-Forming Enzymes

The pioneering contributions leading to both enzyme discovery and gene cloning of the same were summarized in the first edition.⁸ Unlike the upstream pathway enzymes (ADT to the P-450s), however, the cytosolic 4-coumarate:coenzyme A ligases (4CLs, EC 6.2.1.12) are more substrate versatile/substrate degenerate being able to differentially catalyze the formation of various *p*-hydroxycinnamoyl CoA esters **9–13** from the corresponding hydroxycinnamic acids, *p*-coumaric (**4**), caffeic (**5**), ferulic (**6**), 5-hydroxyferulic (**7**), and (to a much lesser extent) sinapic (**8**) acids.^{8,204,241} (Figure 22). For example, in *Arabidopsis*, there are four bona fide 4CLs (At1g51680 (At4CL1), At3g21240 (At4CL2), At1g65060 (At4CL3), and At3g21230 (At4CL5)) capable of differentially ligating the various hydroxycinnamic acids **4–8** *in vitro*, with sinapic acid (**8**) being a very poor substrate.²⁴¹ Of these isoforms, At4CL1 is catalytically most active ($k_{\text{cat}}/K_m = 1\,214\,500$ and $666\,000\text{ mol}^{-1}\text{ s}^{-1}$ for caffeic (**5**) and *p*-coumaric (**4**) acids, respectively). Additionally, 4CLs have molecular weights ranging from approximately 55 000 to 67 000 depending on the plant source. There are currently no reported 3D structures for any 4CL.

Provisional understanding of the catalytic mechanism can be deduced by comparison with other ATP-binding enzymes from both prokaryotes and eukaryotes, such as peptide synthetases, luciferases, and acetyl-CoA and acyl-CoA synthetases. These enzymes either form an adenylate intermediate, which is further esterified with CoA as for 4CLs (e.g., acetyl-CoA synthetases (Acs), and acyl-CoA synthetases), or are oxidized by molecular oxygen (luciferases), or are enzyme-bound CoA derivatives, such as 4'-phosphopantetheine (nonribosomal peptide synthetases).

The reaction mechanism of CoA ligation was initially proposed over 50 years ago by Berg for the formation of acetyl coenzyme A,²⁴² and this can also be envisaged for 4CL-catalyzed processes. In the first ‘half-reaction’ (which is ATP (**100**) and Mg²⁺ dependent) (Figure 22), specific 4CL isoforms can differentially ligate 4-hydroxycinnamic acids **4–8** with ATP (**100**) to afford 4-hydroxycinnamoyl–adenylate intermediates, such as **101–105**. In the second ‘half-reaction’, the thiol functionality of CoA (**106**) attacks the carbonyl carbon of the hydroxycinnamic acid derivative thereby displacing the AMP leaving group (**107**) to form products, such as hydroxycinnamoyl CoA-ester derivatives **9–13**.

6.17.3.4.1 Conformational changes during catalysis: Relevance to 4CL?

Although no 3D structure for any 4CL is currently available, biochemical and 3D gramicidin S synthetase 1 (PheA) structural analyses of adenylate-forming enzymes, such as firefly luciferases,²⁴³ peptide synthetases (e.g., which

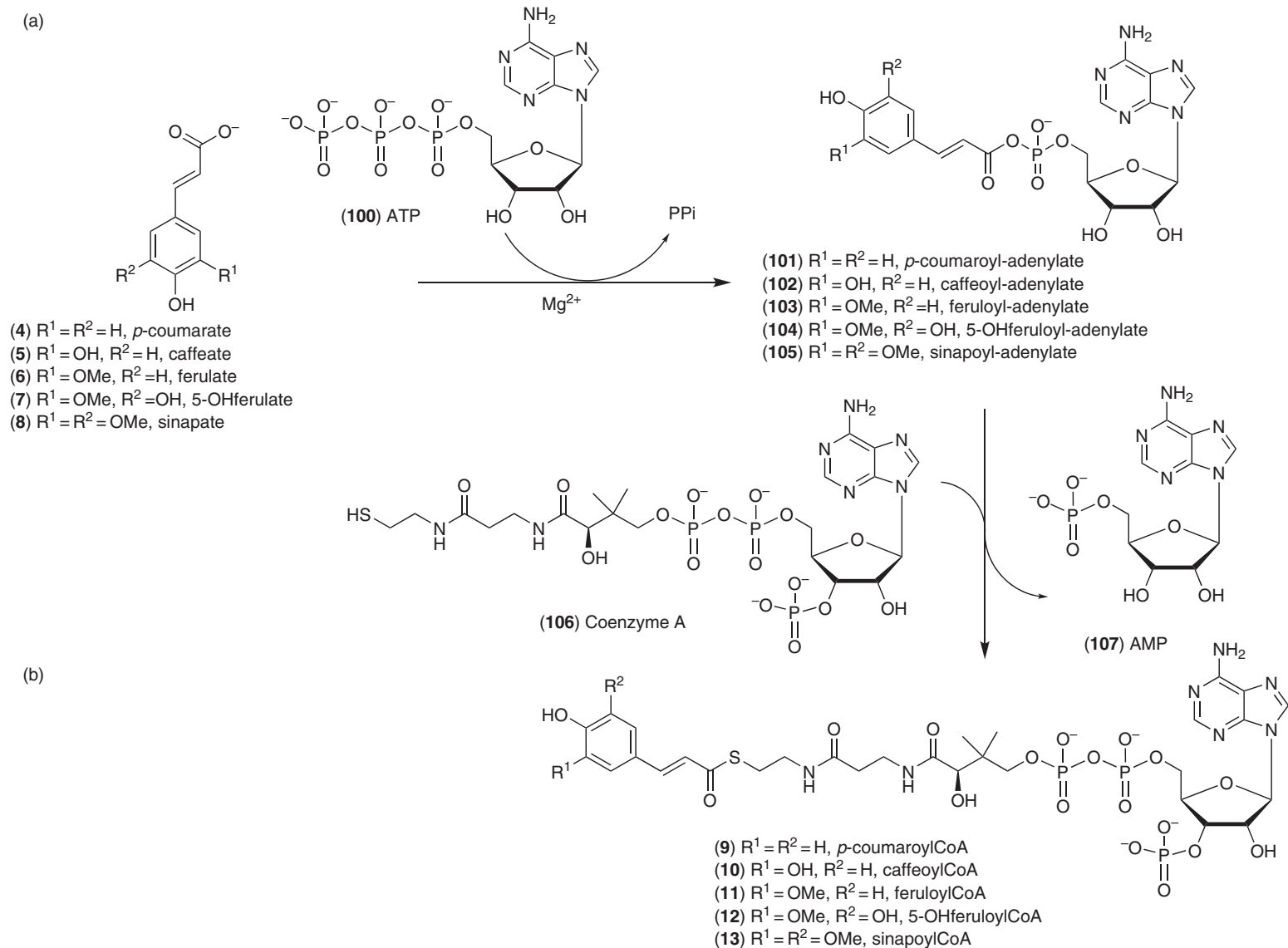
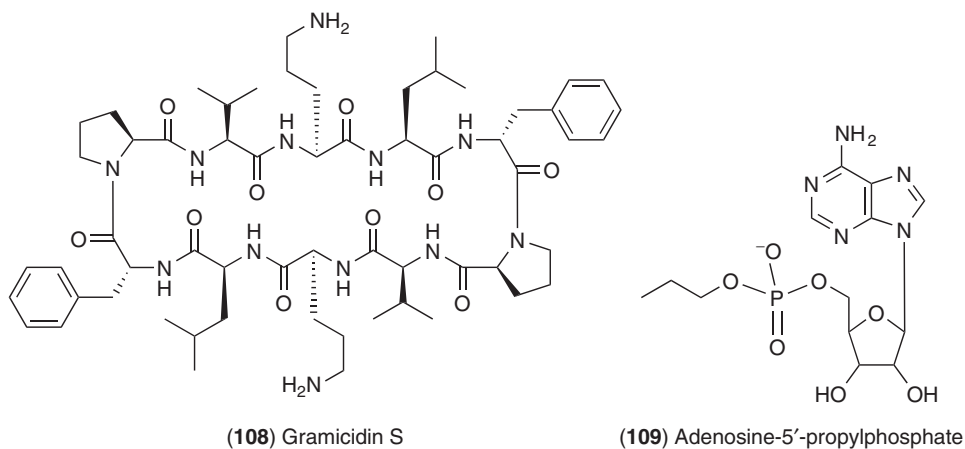


Figure 22 The two 'half-reactions' catalyzed by 4-coumarate: CoA ligase (4CL). (a) Adenylation and (b) thioesterification.



catalyzes formation of the cyclic peptide antibiotic gramicidin S (**108**),²⁴⁴ Acs,^{245,246} and 4-chlorobenzoate:CoA ligases (CBLs),^{247–249} have potentially provided very useful insights into 4CL catalysis. These enzymes, like 4CL, share a common reaction mechanism for formation of adenylate intermediates, with the catalytic processes of adenylation half-reaction and thioesterification half-reaction occurring via a ping-pong mechanism.

The 3D structural analyses of firefly luciferase (apo form, 2.0 Å resolution),²⁴³ gramicidin S synthetase 1 (PheA), complexed with AMP (**107**) and L-Phe (**1**) (1.9 Å resolution),²⁴⁴ as well as Acs complexed with adenosine-5'-propylphosphate (**109**) and CoA (**106**) (1.75 Å resolution),²⁴⁵ have established the presence of a two-domain architecture. In each case, a comparatively large N-terminal core domain is connected to a much smaller C-terminal cap domain by a solvated peptide linker, with the active site situated at the interface between the two domains. In addition, 3D structures of CBL²⁴⁸ as a binary complex with 4-chlorobenzoyl-adenylate (4-CB-AMP) (**111**, **Figure 23**) and as a ternary complex with 4-chlorophenacyl-CoA (**113**) (an inert

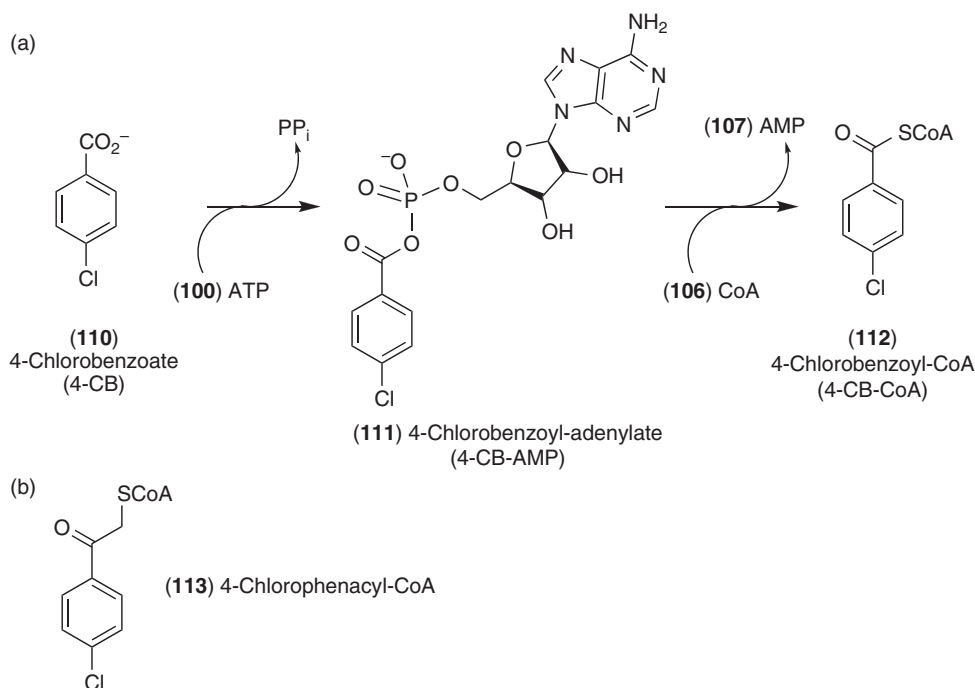


Figure 23 Biochemical conversion of 4-chlorobenzoate (4-CB, **110**) into 4-chlorobenzoyl-CoA (4-CB-CoA, **112**) catalyzed by 4-chlorobenzoate:CoA ligase (CBL) (a) and 4-chlorophenacyl-CoA (**113**), an inert analog of product **112** (b).

analog of product 4-chlorobenzoyl-CoA (4-CB-CoA, **112**) and AMP (**107**) at 2.25 and 2.0 Å resolution, respectively, have provided additional information on both conformations generated during adenylate and thioester conversions. That is, both conformations are envisaged to be generated during the course of the two ‘half-reactions’ by rotation around a hinge region within the solvated linker. Specifically, following completion of the initial adenylation step, the C-terminal domain undergoes rotation by approximately 140° to adopt a second conformation, in so doing altering the interface with the N-terminal region. In this way, the second ‘half’ of the reaction (thioesterification step) is enabled.²⁴⁸ Currently, however, this is the only example of an adenylate-forming 3D structure with binary/ternary complexes reported in both adenylate-forming and thioester-forming conformations.

Reger *et al.*²⁴⁶ had also proposed comparable conformational changes within the C-terminal domain for Acs from *Salmonella enterica*, where the conserved hinge residue (Asp517 for Acs) was considered as the pivot point for an equivalent C-terminal rotation of approximately 140°. The catalytic efficiency of Acs toward both the adenylation and thioesterification reactions was also investigated by mutating the hinge residue, Asp517, with proline and glycine, the effects of which were provisionally envisaged to increase both the rigidity and conformational flexibility of the hinge region, respectively. The kinetic data in support of this interpretation were, however, not definitive.

6.17.3.4.2 Mode of action of catalysis

Initial studies²⁵⁰ envisaged that the cysteine residue in the GEICIRG motif which is conserved in all 4CLs was required for catalysis (specifically Cys403 in *Arabidopsis* At4CL2, At3g21240). However, from our previous comprehensive analysis,⁸ it was noted that there was no experimental evidence in support of the involvement of any particular cysteine residue. Later, a kinetic analysis of the Cys403Ala mutated protein from At4CL2 eliminated its essential involvement in 4CL catalysis,²⁵¹ that is, given that the mutated enzyme still retained approximately 45% 4CL activity relative to WT (specific activity of 94 vs. 209 nkat mg⁻¹, respectively).

Conversely, mutational and structural studies on the adenylate-forming superfamily resulted in a conserved lysine residue (e.g., Lys945 in surfactin synthetase²⁵²) being proposed as involved in catalysis, this being envisaged to occur by providing favorable polar interactions for effective substrate orientation during adenylate formation. Specifically, site-directed mutagenesis of firefly luciferase Lys529 to arginine, glutamine, and alanine established that k_{cat} values of the mutated Lys529Arg, Lys529Gln, and Lys529Ala enzymes were 625-, 1250-, and 1250-fold lower than that of the WT enzyme, respectively,²⁵³ whereas the comparable mutated residue in acetyl CoA synthetase (Lys609Ala) resulted in an enzyme unable to catalyze adenylation.²⁴⁶ Additional reports for several other members of the adenylate-forming family also strongly supported the involvement of this highly conserved lysine residue in catalysis via activation of the substrate carboxylate group, as noted for propionyl-CoA synthetase,²⁵⁴ and CBL.²⁴⁹ In an analogous manner, mutation of the corresponding lysine residue (i.e., Lys540 to Asn in At4CL2) resulted in near abolition of enzymatic activity, supporting its involvement in 4CL adenylation as well.²⁵¹

Carboxylate group activation for adenylate formation has also been proposed to occur via nucleophilic attack of the negatively charged carboxylate ion onto the phosphorus atom of the negatively charged phosphoryl group in ATP (**100**).²⁴⁹ However, this biochemical process is feasible only if the participating enzyme can meet the following two criteria: (1) orientation of the positively charged amino acid residue to the carboxylate and phosphoryl functionalities for in-line attack and (2) shielding of point charges by the active site residues to avoid unfavorable charge–charge interactions at close proximity.

In this regard, structural²⁴⁸ and mutagenesis²⁴⁹ data obtained during studies of the aforementioned CBL in *Alcaligenes* sp. suggested that Lys492 (the equivalent of Lys540 in At4CL2) of the C-terminal domain and His207 and Thr307 of the N-terminal domain were the catalytic residues functioning in these two distinct ‘half-reactions’ ((**Figures 24(a)** and **24(b)**). In addition, another residue, Thr161 of the N terminal domain, was also proposed to interact with the departing PP_i in the first half-reaction, as well as with the phosphoryl group of the 4-CB-AMP (**111**) in the second half-reaction.²⁴⁹

Thus, the overall catalysis can be envisaged as follows: For the first half-reaction, the catalytic residue His207 facilitates activation of the substrate, 4-chlorobenzoate (4-CB, **110**), and the Lys492 participates in stabilization of the carboxylate group of the substrate, whereas the hydroxyl side chain of Thr307 activates the ATP (**100**) through hydrogen bonding with its negatively charged phosphoryl oxygen. Together, this enables

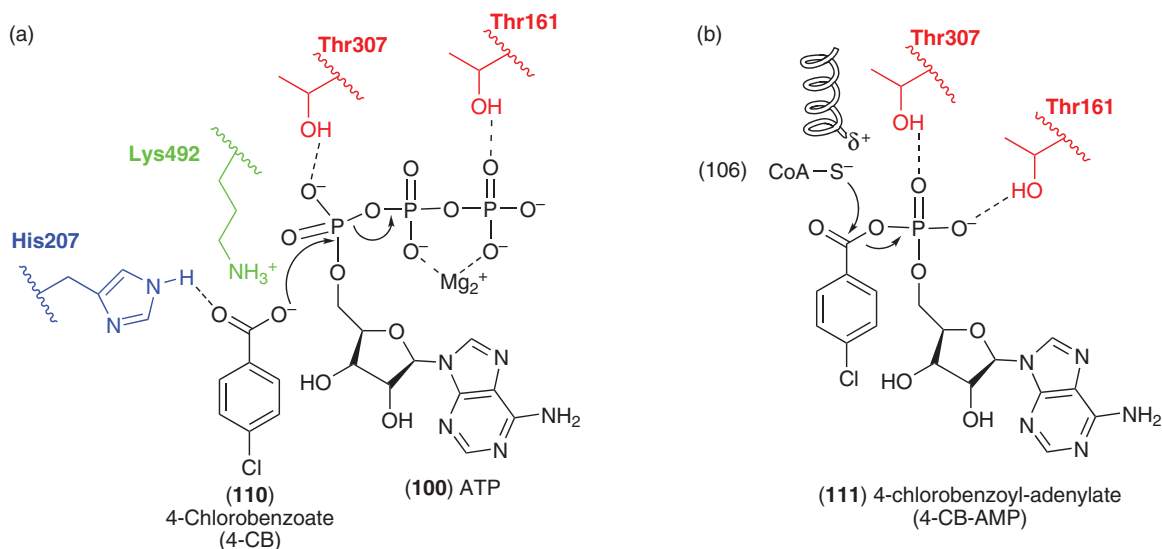


Figure 24 The proposed roles of key catalytic amino acid residues in (a) adenylation and/or (b) thioesterification during 4-chlorobenzoate:CoA (CBL) catalysis.

nucleophilic attack, resulting in formation of the 4-CB-AMP intermediate (111) with departure of the leaving group, PP_i , being facilitated by complexation to Mg^{2+} and hydrogen bonding to the hydroxyl group of Thr161 (Figure 24(a)).²⁴⁹ The second ‘half-reaction’ (thioesterification) involves an additional nucleophilic attack of the thiol of CoA (106) to the carboxyl carbon of the 4-CB-AMP intermediate (111), with displacement of AMP (107) being activated by hydrogen bond formation between the phosphoryl group oxygen atom moiety and the hydroxyl group side chains of Thr307 and Thr161, respectively (Figure 24(b)).

An analogous situation can also be anticipated for 4CL catalysis, with the equivalent residues for ‘His207, Lys492, Thr161, and Thr307’ in 4CL being employed. This will best be fully documented by establishing the active site interaction in the binary/ternary complexes of 4CL, when available, in both adenylate- and thioester-forming conformations, respectively, with further verification of the role(s) of the active site residues by site-directed mutagenesis studies.

6.17.3.4.3 Protein interactions with pyrophosphate leaving group

Based on comparison of numerous ATP-binding enzymes in both prokaryotes and eukaryotes, a highly conserved putative nucleotide-binding (AMP and/or ATP) motif, that is, (Ser/Thr) (Ser/Thr/Gly) Gly (Ser/Thr) (Ser/Thr) Gly X Pro Lys Gly, was initially described,²⁵⁵ this also being used in the classification of the superfamily of adenylate-forming enzymes.²⁵⁶ However, this so-called nucleotide-binding (AMP and/or ATP) motif was later suggested to interact with the pyrophosphate leaving group, PP_i . This was because the ternary 3D structure (complexed with AMP (107) and L-Phe (1), 1.9 Å resolution) of PheA indicated that the putative nucleotide-binding motif (residues 190–199) formed a flexible loop between two β -strands in the active site.²⁴⁴ However, the absence of electron density in the region of these residues indicated conformational flexibility. This flexible loop was thus suggested to instead accommodate the pyrophosphate leaving group during catalysis, rather than directly participating in nucleotide (ATP (100)/AMP (107)) binding.

Almost contemporaneously, the highly conserved lysine residue in this motif (e.g., Lys186 in tyrocidine synthetase 1 (TycA) from *Bacillus brevis*) was proposed as a probable candidate for the cationic locus to interact with the phosphate group of the bound nucleotide (100/107).²⁵⁷ In agreement with this, mutagenesis of the corresponding Lys186 residue with arginine and threonine gave corresponding mutated proteins with approximately 10 and 1% of Phe (1)-dependent PP_i -ATP exchange enzymatic activity relative to that of the WT enzyme.^{257,258}

In a similar manner, site-directed mutagenesis of the comparable lysine residue (Lys211) of this motif in At4CL2 with serine gave a mutated 4CL with approximately 3% of WT activity (5.6 vs. 209 nkat mg^{-1}) using

caffeic acid (5) as substrate.²⁵¹ Interestingly, the overall K_m values for the Lys211Ser mutant for both ATP (100) ($151 \mu\text{mol l}^{-1}$) and caffeic acid (5) ($14 \mu\text{mol l}^{-1}$) were not significantly affected in comparison to those for WT At4CL2 (K_m for ATP (100) $\sim 163 \mu\text{mol l}^{-1}$ and caffeic acid (5) $\sim 24 \mu\text{mol l}^{-1}$), that is, binding of ATP (100) and caffeic acid (5) was not altered by this mutation.

6.17.3.4.4 Homology modeling of At4CL2 and attempted substrate-binding pocket manipulations

Based on these similarities, a homology-based model of At4CL2 was later proposed by comparison with PheA as template.²⁵⁹ According to this model, a putative substrate-binding pocket was provisionally defined based on an estimated distance of approximately 6 Å from the center of the modeled bound substrate, caffeic acid (5). This was speculated based on all putative van der Waals interactions and hydrogen bonds between the amino acid residues in proximity to 5. Thus, a substrate-binding pocket in the modeled At4CL2 was hypothesized to be comprised of 12 amino acid residues in the N-terminal domain, that is, Ile252, Tyr253, Asn256, Met293, Lys320, Gly322, Ala323, Gly346, Gly348, Pro354, Val355, and Leu356. Attempts to ‘engineer’ the putative substrate-binding pocket were then carried out to produce At4CL2 mutants that would more facily utilize either ferulic (6) or sinapic (8) acid as substrates, conversions that were reportedly not carried out by At4CL2. However, these data differed from our own analyses of At4CL2 that demonstrated that it was able to poorly convert ferulic acid (6) into feruloyl CoA (11) ($k_{\text{cat}}/K_m = 930 \text{ mol}^{-1} \text{ l s}^{-1}$), with caffeic acid (5) though being the preferred substrate ($k_{\text{cat}}/K_m = 492\,400 \text{ mol}^{-1} \text{ l s}^{-1}$).²⁴¹ Nevertheless, a double At4CL2 mutant, Met293Pro and Lys320Leu, was shown to be able to ligate ferulic acid (6) ($k_{\text{cat}}/K_m = 530\,470 \text{ mol}^{-1} \text{ l s}^{-1}$), this being presumed to result from less steric hindrance for the methoxyl group of ferulic acid (6) in the modeled substrate-binding pocket.²⁵⁹ In a similar vein, two double mutants (i.e., Met293Pro with deletion of Val355 and Met293Pro with deletion of Leu356) were generated with the corresponding mutated proteins envisaged to provide a better substrate-binding pocket for sinapic acid (8). These were reportedly able to generate sinapoyl CoA (13), albeit very poorly as compared to *p*-coumaroyl CoA (9), that is, k_{cat}/K_m for (8) = 7590 and 11 890 $\text{mol}^{-1} \text{ l s}^{-1}$ vs. k_{cat}/K_m for (4) = 256 350 and 277 240 $\text{mol}^{-1} \text{ l s}^{-1}$.²⁵⁹

In summary, comparison of the structural and functional data obtained for other members of adenylating-forming enzymes has provided useful insights into the catalytic reaction mechanisms involving 4CLs. Nevertheless, it will be very instructive to obtain 3D X-ray structures for 4CL (apo form and binary/ternary complexes) and to determine how these compare with the other enzymes mentioned above, as well as the precise factors affecting differential substrate versatility/degeneracy.

6.17.3.5 COMTs and CCOMTs

As indicated in both earlier and more recent critical reviews, there was significant confusion as regards the enzymes involved in *O*-methylation in the monolignol pathway and the substrates utilized.^{1,2,8} However, it is now known that two classes of cytosolic *O*-methyltransferases (OMTs) are involved: caffeic acid *O*-methyltransferase (COMT, EC 2.1.1.68) and caffeoyl CoA *O*-methyltransferase (CCOMT, EC 2.1.1.104) (see Figure 4).

6.17.3.5.1 *O*-Methyltransferases in the monolignol pathway: Discovery of ‘COMT’ and ‘CCOMT’ proper

Based on the study of crude enzyme preparations *in vitro*,^{260,261} ‘COMT’ was originally considered to be a monofunctional enzyme in gymnosperms, catalyzing the regiospecific methylation of caffeic acid (5) to afford ferulic acid (6). By contrast, in angiosperms, it was considered bifunctional and able to use both caffeic (5) and 5-hydroxyferulic (7) acids to give ferulic (6) and sinapic (8) acids, respectively. This bifunctional nature of COMT in angiosperms was also purportedly supported by reports of transgenic manipulations (downregulation) of COMT in tobacco²⁶² and alfalfa;²⁶³ in these studies, it was concluded that lignin amounts were either unaffected in tobacco²⁶² or reduced (by $\sim 50\%$) in alfalfa,²⁶³ but with little to no alterations in their monomeric ratios of both coniferyl (21) and sinapyl (23) alcohol-derived moieties. Together, these reports were thus considered to provide additional evidence that COMT was bifunctional, at least in angiosperms. However,

these studies and their findings were ultimately proved incorrect, in large part because of the methodologies employed and experimental design (discussed in Davin *et al.*²).

By contrast, the group of Fritig and Legrand, who had purified the first ‘COMT’ proper from tobacco,²⁶⁴ unambiguously established ‘COMT’ to have, as its true physiological function, methylation of 5-hydroxyguaiacyl moieties to afford sinapyl alcohol (23).²⁶⁵ That is, this step involved addition of the second methyl group in monolignol pathway biosynthesis.²⁶⁵

The preceding (or first monolignol pathway) methylation step was ultimately discovered to utilize a CoA derivative, that is, caffeoyl CoA (10), to afford the corresponding feruloyl derivative (11), rather than the corresponding caffeic acid (5) to produce ferulic acid (6).^{266–270} Taken together, these studies had established that two quite distinct biochemical processes were operative for each methylation step in monolignol biosynthesis.

More importantly, the substrate versatility of both CCOMT and COMT enzyme preparations *in vitro* highlighted the significance of compartmentalization of particular substrates and enzymes *in vivo*, that is, given that COMT only methylates 5-hydroxyguaiacyl moieties *in vivo*, and thus is not ‘exposed’ to other substrates.

6.17.3.5.2 S-COMT 3D structure

Much of what was understood about (monolignol pathway) OMTs was initially derived from studies using rat liver OMT, that is, the so-called S-COMT (soluble catechol-*O*-methyltransferase, EC 2.1.1.6). This protein has a molecular weight of approximately 24 700 and exists in monomeric form.^{271,272}

In terms of biochemical mechanism, S-COMT 3D structure was determined in ternary complex form (2.0 Å resolution) with a competitive inhibitor (3,5-dinitrocatechol, 114, Figure 25(a)), Mg²⁺, and *S*-adenosyl-L-methionine (SAM, 115, Figure 25(b)),²⁷³ and consisted of eight α -helices and seven β -strands with a typical core α/β Rossmann fold. The Mg²⁺ is complexed with side-chain oxygen atoms of Asp141, Asp169, and Asn170, and when bound in this way, the adenine group of the cofactor SAM (115) is presumed in van der Waal contact/interaction with Trp143, His142, and Met91 (Figure 26). Additionally, this apparently results in the ribose ring of SAM (115)

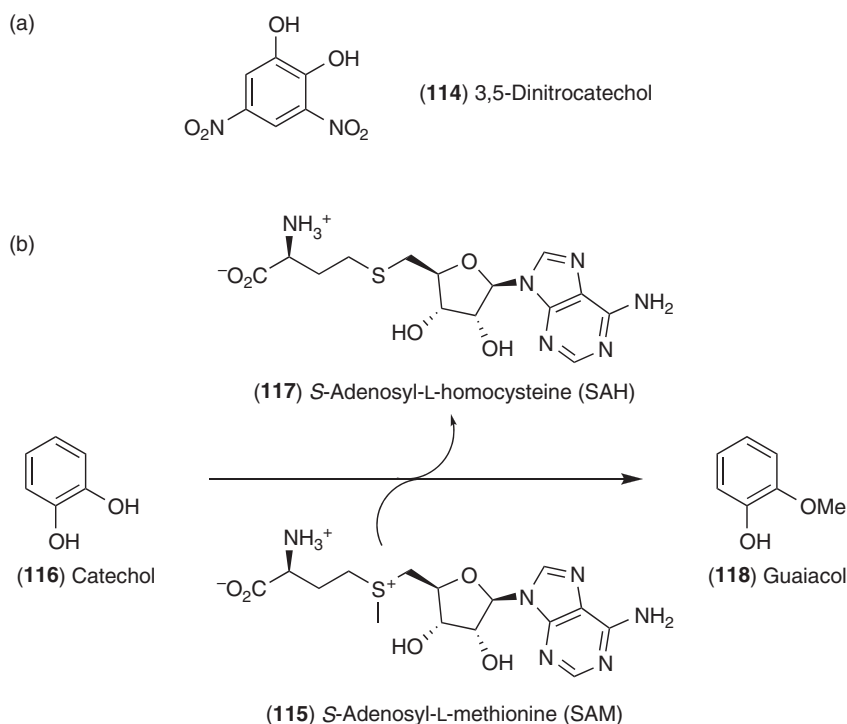


Figure 25 Soluble catechol-*O*-methyltransferase (S-COMT). (a) Inhibitor 3,5-dinitrocatechol (114). (b) Reaction catalyzed by S-COMT.

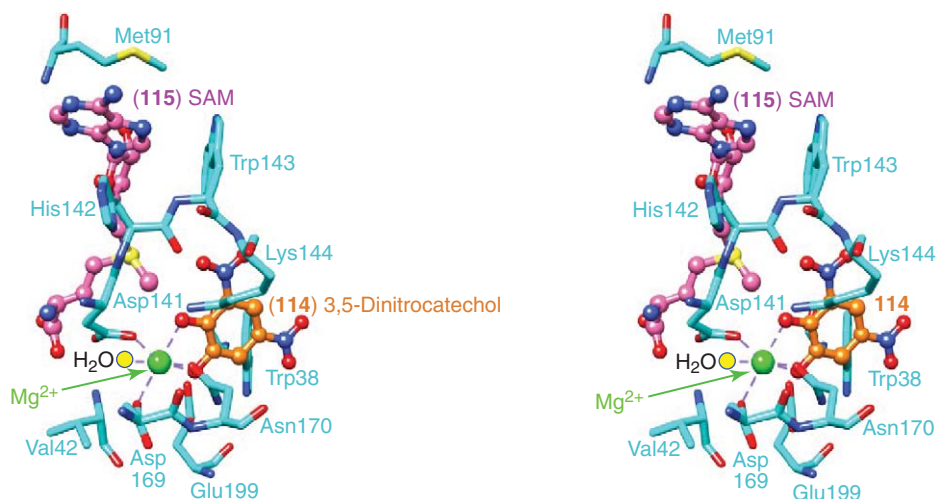


Figure 26 Stereoview of the substrate-binding pocket of S-COMT from rat liver in ternary complex (PDB ID: 1VID) with the inhibitor, 3,5-dinitrocatechol (**114**) (orange), Mg^{2+} (green), SAM (**115**) (pink) and a water molecule (yellow) coordinated to Mg^{2+} .

being oriented parallel to the Trp143 residue. A hydrogen-bonding network occurs between the backbone Val42 nitrogen (and a water molecule) and the methionine carboxyl oxygen of SAM (**115**), as well as between Asp141 and the methionine amino group of SAM (**115**) (Figure 26). Met40 (not shown) also orientates the sulphur of SAM (**115**) towards the nucleophilic hydroxyl group of **114**. In this way, the sulfur-containing electrophilic methyl group of SAM (**115**) is oriented toward the nucleophilic phenol hydroxyl group of the inhibitor **114**. Interestingly, the planar structure of the inhibitor **114** enables stacking of its aromatic ring with the Trp38 residue, which is located edge-to-face with **114**, with Trp143 also considered to have van der Waals interaction with the nitro group of the inhibitor **114**. Both phenolic hydroxyl groups of the 3,5-dinitrocatechol (**114**) then complex with Mg^{2+} , resulting in an exact juxtaposition of one of the phenolic hydroxyl groups with the donor methyl group of SAM (**115**). By contrast, the other phenolic hydroxyl group of **114** is hydrogen bonded to the carboxyl oxygen of Glu199, thereby preventing anion formation of that hydroxyl group.

Recently, Tsuji *et al.*²⁷⁴ described further aspects of the 3D structure of rat liver S-COMT by comparison of the apo and holo form (bound to inhibitor **114** and SAM (**115**)) obtained at 2.2 and 2.6 Å resolution, respectively. Although the structure of the apo form was considered to be in partially open form, it was not possible to clearly define what amino acid residues were involved in the ligand-binding site or even the cleft corresponding to the SAM (**115**)-binding site. However, with binding of cofactor (SAM (**115**)) and inhibitor **114**, a conformational change was suggested to occur from the apo form to the holo form in S-COMT, with the latter providing an effective binding site for the inhibitor/substrate and cofactor. This study thus suggested that the distinct catalytic site in S-COMT was accessible only after a conformational change in the apo form occurred after binding of the ligands in the holo form.

Previously, the enzymatic reaction was, nevertheless, envisaged to proceed by an ordered sequential kinetic mechanism with SAM (**115**) binding to S-COMT first, this being followed successively by Mg^{2+} and substrate. The reaction mechanism of S-COMT was thus proposed to involve bimolecular transfer of a methyl group from SAM (**115**) to the proximal phenoxide ion of the catechol (**116**) via an S_N2 -like transition state, thereby generating guaiacol (**118**) and *S*-adenosyl-*L*-homocysteine (SAH, **117**).²⁷⁵ Studies on the α -deuterium and carbon-13 isotope effects for S-COMT-catalyzed methylation also established that the transfer of the methyl group was rate-determining.^{276–278}

For comparative purposes, the 3D structures of rat liver S-COMT complexed with 3,5-dinitrocatechol (**114**), Mg^{2+} and SAM (**115**), ‘CCOMT’ complexed with SAH (**117**) and 5-hydroxyferuloyl CoA (**12**), and ‘COMT’ complexed with SAH (**117**) and ferulic acid (**6**) are depicted in Figure 27, and are discussed below.

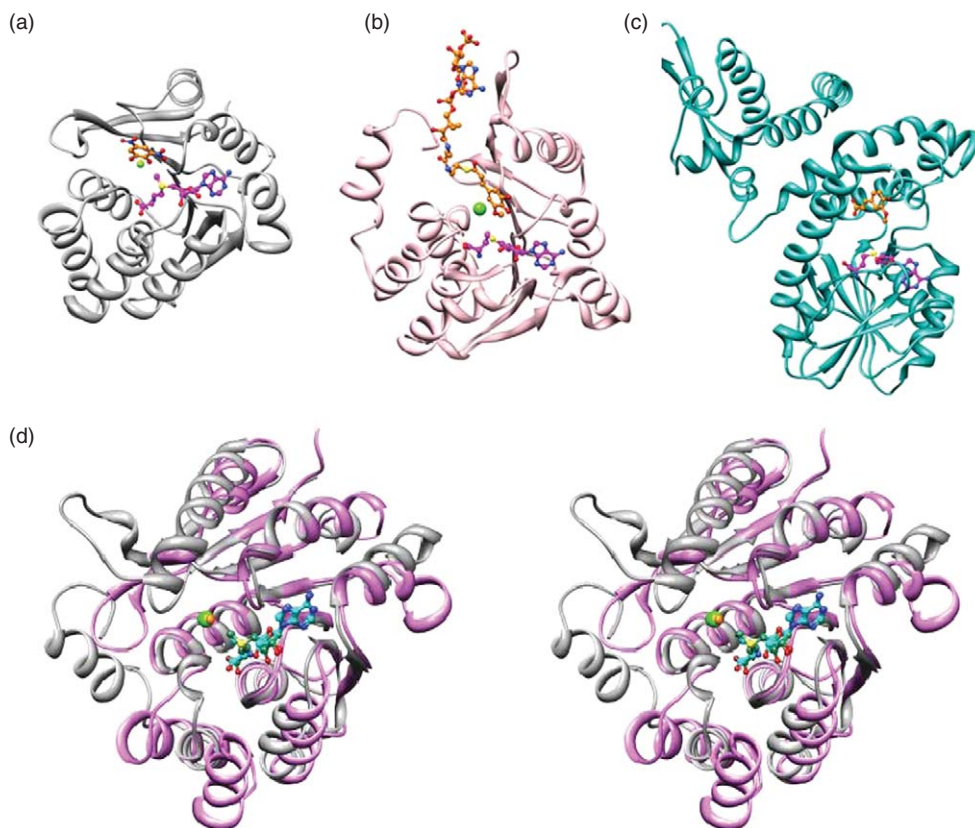


Figure 27 (a) Ribbon representation of S-COMT (PDB ID: 1VID) complexed with the competitive inhibitor 3,5-dinitrocatechol (**114**) (orange) and SAM (**115**) (purple). (b) Ribbon representation of CCOMT (PDB ID: 1SUI) complexed with 5-hydroxyferuloyl CoA (**12**) (orange) and SAH (**117**) (purple). (c) Ribbon representation of COMT (PDB ID: 1KYZ) complexed with ferulic acid (**6**) (orange) and SAH (**117**) (purple). (d) Stereoview of superimposition of the structure of CCOMT (pink) complexed with 5-hydroxyferuloyl CoA (**12**) and SAH (**117**) with that of S-COMT (grey) complexed with 3,5-dinitrocatechol (**114**) and SAM (**115**). The divalent metal cations are shown in green.

6.17.3.5.3 3D structure of CCOMT

CCOMTs are dimeric with a molecular weight of approximately 24 000–28 000 per monomer.^{268,279} The 3D structures of a CCOMT from alfalfa as ternary complexes were obtained upon incubation of CCOMT with SAM (**115**) and caffeoyl (**10**)/5-hydroxyferuloyl (**12**) CoAs (2.70 and 2.65 Å resolution (**Figures 28(a)** and **28(b)**)).²⁸⁰ Unlike rat liver monomeric S-COMT, CCOMT is present as a dimer with each monomer consisting of a single catalytic domain, and possessing a core α/β Rossmann fold harboring the SAM (**115**)/SAH (**117**)-binding site. In this regard, the adenine group of SAM (**115**)/SAH (**117**) is proposed to hydrogen bond with both Asp165 and the backbone Ala140 nitrogen (not shown), whereas the hydroxyl group of the ribose ring is considered hydrogen bonded to the carboxyl group of an Asp111 residue. In addition, it was envisaged that hydrogen bonding occurred between the terminal carboxylate of SAH (**117**) with Ser93, as well as the terminal NH_3^+ group of SAH (**117**) with Gly87 and Ser93 along with formation of a salt bridge with the Glu85 residue (**Figure 28(c)**).

CCOMT putatively contains a divalent metal-binding active site, indicative of a metal-dependent catalytic reaction, and therefore somewhat resembles S-COMT. However, in CCOMT, a divalent Ca^{2+} was detected rather than Mg^{2+} , with the former being presumed to be exchanged in the crystallization solution.²⁸⁰ Ca^{2+} was also considered chelated in the active site through interactions with side-chain oxygens of Thr63, Glu67, Asp163, Asp189, and Asn190 residues, a water molecule, and the 3-methoxy oxygen atom of feruloyl CoA (**11**) (**Figure 28(c)**), with the negatively charged 3'-phosphate group of the adenosine-3'-5'-diphosphate moiety of CoAs stabilized by Lys21 and Arg206. Aromatic residues, Tyr208, Tyr212, and Trp193, were also considered

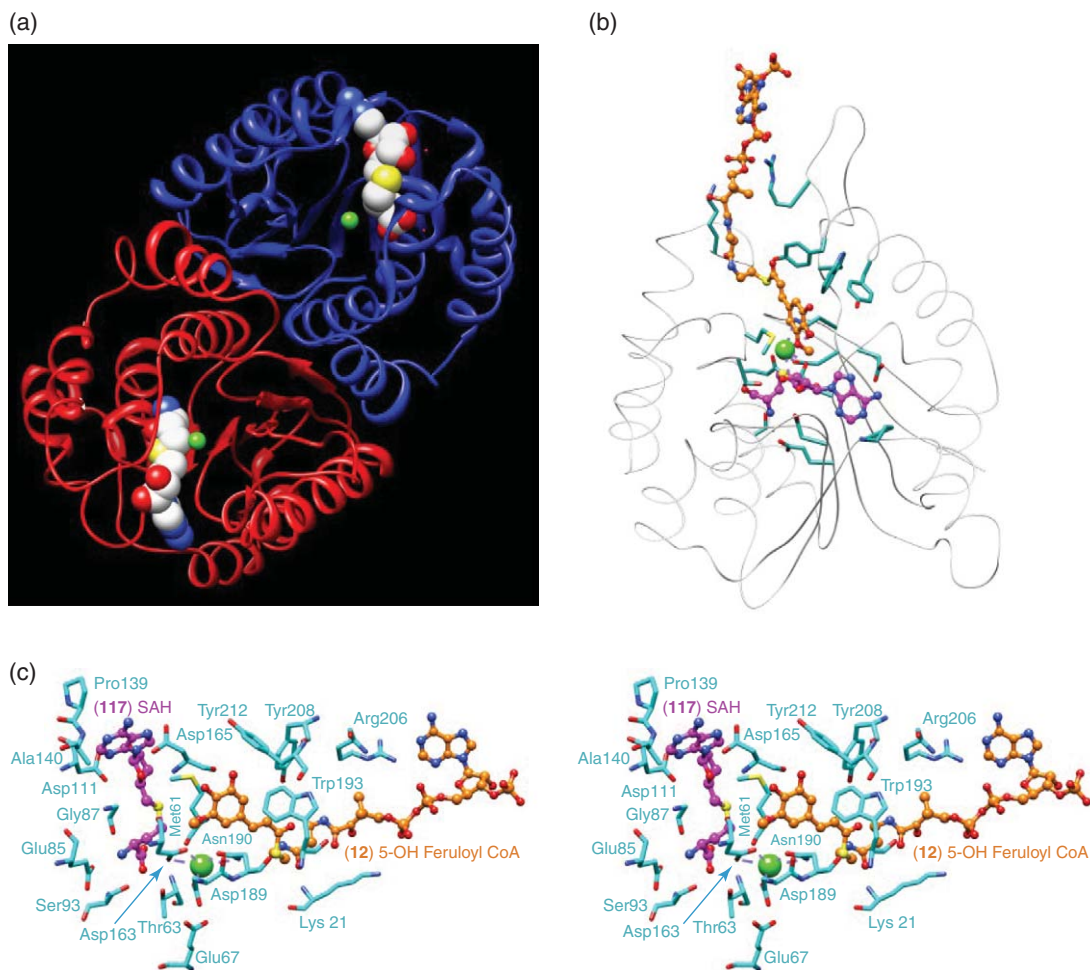


Figure 28 (a) Ribbon representation of ‘CCOMT’ (PDB ID: 1SU1) from alfalfa. (b) Binding of 5-hydroxyferuloylCoA (**12**) (orange) at the active site of ‘CCOMT’. (c) Stereoview of interaction of active site residues of ‘CCOMT’ in ternary complex with SAH (**117**) (purple) and 5-hydroxyferuloyl CoA (**12**) (orange) (PDB ID: 1KYW). Ca^{2+} is shown as a green sphere.

to sandwich the phenolic moiety of the molecule; two other residues Met61 and Asp163 postulated to reside in the active site had unknown role(s).

6.17.3.5.4 Structural comparisons of S-COMT and CCOMT and catalytic mechanisms

For S-COMT,²⁷³ juxtaposition of the hydroxyl group of the substrate, catechol (**116**), to the donor methyl group of SAM (**115**) was possible only if both phenolic hydroxyl groups were complexed to Mg^{2+} . Mg^{2+} thus has a crucial role in substrate binding, as well as being involved somewhat in catalytic activity, since it is coordinated in the active site to the side chains of Asp141, Asp169, and Asn170, a water molecule trans to Asn170, and the two hydroxyl groups of catechol (**116**) (Figure 26). The Lys144 residue in the catalytic site of S-COMT has an $-\text{NH}_2$ side chain considered to act as a catalytic base, for example, for proton abstraction of the hydroxyl group in catechol (**116**) that is in close proximity to the sulfur-containing methyl group of SAM (**115**). Following the generation of the nucleophilic phenoxide group in substrate **116**, attack of the electron-deficient methyl group of SAM (**115**) occurs, thereby facilitating methyl transfer from SAM (**115**).²⁸¹ In so doing, the Glu199 residue also hydrogen bonds with the other phenolic hydroxyl group, thereby preventing anion formation at that center.²⁷³

For CCOMT, by contrast, juxtaposition of the reactive phenolic groups of the substrate, caffeoyl CoA (**10**), to the donor methyl group of **115** was proposed to occur instead by complexation with Ca^{2+} .²⁸⁰ This differs

from S-COMT, where proton abstraction of the reactive hydroxyl group of the substrate **116** was envisaged to result via action of the catalytic base (Lys144 in S-COMT) thereby facilitating nucleophilic attack to the electron-deficient methyl group of SAM (**115**). However, for CCOMT, it was considered that the reactive phenolic group of the substrate **10** already existed as an oxyanion (phenoxide ion) in order to balance positive charges surrounded by Ca^{2+} and S-CH_3^+ of SAM (**115**) in the active site. Thus, according to these researchers, there was no equivalent of either a catalytic residue Lys144 or the stabilizing Glu199 residue in CCOMT, as previously established for S-COMT.^{8,273}

6.17.3.5.5 3D structure of 'COMT'

The 3D structures of a 'COMT' from alfalfa were obtained at 2.1 and 2.4 Å resolution, complexed with SAH (**117**) and ferulic acid (**6**)/5-hydroxyconiferaldehyde (**17**), respectively (**Figure 29**).²⁸² As for CCOMT, 'COMT' also exists as a homodimer in solution,²⁸² with a molecular weight of approximately 40 000 per monomer.^{264,282} Additionally, 'COMT' has a large catalytic C-terminal domain with a core α/β Rossmann fold involved in cofactor (SAM (**115**)/SAH (**117**)) binding. A comparatively smaller N-terminal domain was envisaged to be primarily responsible for dimer formation, although this was absent from CCOMT.

In the SAM (**115**)/SAH (**117**)-binding pocket of 'COMT', a Lys265 residue was envisaged to account for an electrostatic interaction with the carboxylate group of SAH (**117**) whereas Asp206 and a water molecule apparently stabilized the terminal amino group of SAH (**117**) through hydrogen bonding (**Figure 29(c)**). An

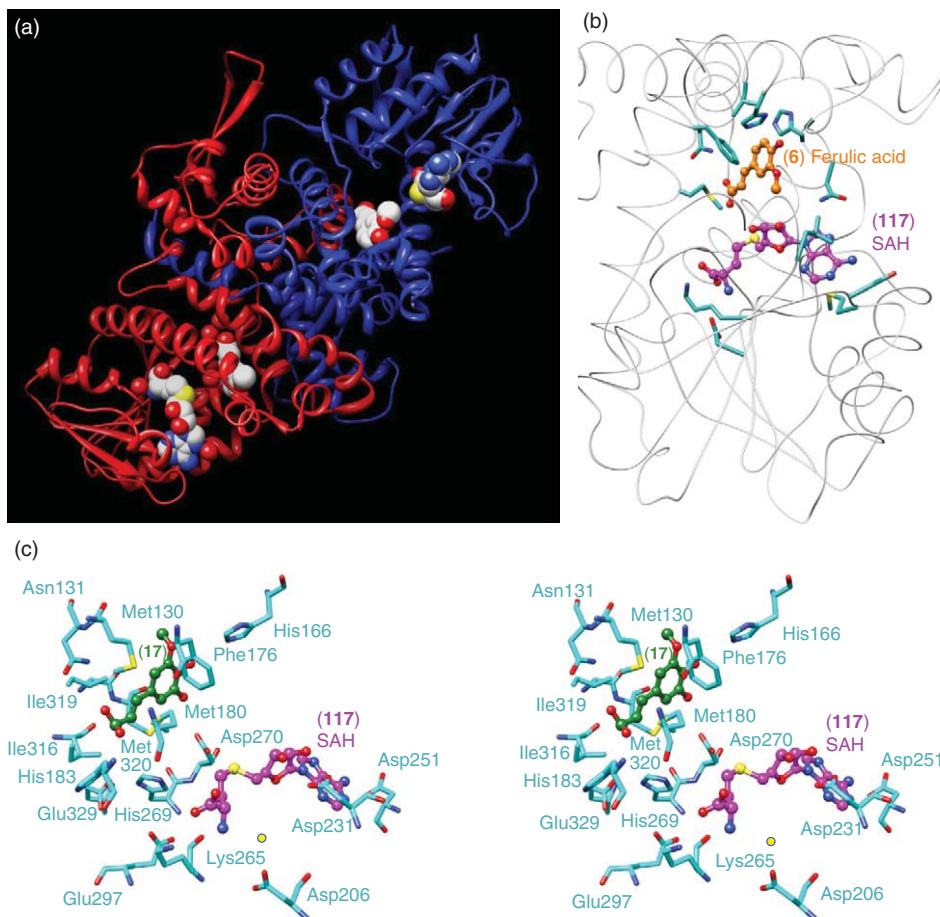


Figure 29 (a) Ribbon representation of COMT (PDB ID: 1KYZ) from alfalfa. (b) Proposed binding of ferulic acid (**6**) (orange) and SAH (**117**) (purple) at the active site of COMT. (c) Stereoview of interaction of active site residues with SAH (**117**) (purple) and 5-hydroxyconiferyl aldehyde (**17**) (green). H_2O molecule is shown as a yellow sphere.

extended hydrogen-bonding network was also proposed with the ribose hydroxyl groups to Asp231 and of the exocyclic amino group of the adenine ring with the Asp251 residue. Additionally, hydrophobic residues Phe176, Met130, Met320, and Met180 were considered to position the aromatic rings of either ferulic acid (**6**) or 5-hydroxyconiferaldehyde (**17**) thereby enabling the reactive C-phenolic group to be in close proximity to SAM (**115**). In addition, a His166 residue was considered to hydrogen bond with the 4-hydroxyl group of the substrate **17**, thereby orientating the substrate for preferential methylation at the 5-hydroxyl position. The hydrophilic residues His183 and Asn131 also apparently interacted with the C3 side chain of the substrate, with the latter being surrounded by a number of hydrophobic residues such as Met180, Met130, Ile316, and Ile319. In this way, the relative hydrophobicity in this region was also envisaged to confer preferential selectivity toward more neutral aldehydes and alcohols (i.e., 5-hydroxyconiferyl aldehyde (**17**) and 5-hydroxyconiferyl alcohol (**22**)) over that of a negatively charged substrate such as caffeic acid (**5**) in 'COMT' (**Figure 29(c)**).

Unlike S-COMT from rat liver, 'COMT' (alfalfa)²⁸² is also presumed to utilize a general base without the aid of any metal cofactor for catalysis. Specifically, methyl transfer from SAM (**115**) to caffeic acid (**5**) and/or 5-hydroxyconiferaldehyde (**17**) in 'COMT' is considered to be initiated by deprotonation of the 3- or 5-hydroxyl group via the His269 residue acting as a base, thereby generating the nucleophile that subsequently attacks the electron-deficient methyl group of SAM (**115**). The orientation of this histidine residue (His269) was considered to be most likely controlled by hydrogen bond formation with the adjacent residues Glu297 and Asp270. Moreover, Glu329 was observed to form hydrogen bonds with the general base (His269) thereby providing a catalytically productive position toward the hydroxyl group of the substrate (**6/17**) poised for transmethylation. Indeed, the same study indicated that site-directed mutagenesis of His269 by substituting with Leu, Asn, or Gln in 'COMT' (alfalfa) resulted in enzymes that did not display methyltransferase activity (data not shown).²⁸²

6.17.3.6 Cinnamoyl CoA Reductase

Cytosolic cinnamoyl CoA reductases (CCRs, EC 1.2.1.44) can differentially catalyze the substrate-versatile NADPH-dependent reductions of *p*-hydroxycinnamoyl CoA derivatives **9–13** to afford the corresponding *p*-hydroxycinnamyl aldehydes **14–18**, respectively. CCR was first detected in *Salix alba* cell-free extracts²⁸³ and purified later from soybean (*Glycine max*) cell suspension cultures,^{284,285} spruce (*Picea abies* L.) cambial sap,^{285,286} and poplar (*Populus × euramericana*) stems.²⁸⁷ Subsequent cloning and characterization of the first cDNA encoding a CCR was reported by Lacombe *et al.*²⁸⁸ from *Eucalyptus gunnii* in 1997. From the 11 genes annotated as CCRs in the *A. thaliana* genome, only AtCCR1 and AtCCR2 efficiently catalyzed reduction of **9–13** to **14–18**, with both showing higher efficacy with feruloyl (**11**) and 5-hydroxyferuloyl (**12**) CoAs (i.e., $k_{cat}/K_m = 518\,890$ and $485\,770\text{ mol}^{-1}\text{ s}^{-1}$ for AtCCR1 vs. $84\,700$ and $61\,275\text{ mol}^{-1}\text{ s}^{-1}$ for AtCCR2, respectively)²⁸⁹ (F. C. Cochrane *et al.*, manuscript in preparation). CCRs can have molecular weights ranging from approximately 36 000 to 40 000 depending on the species, and are cytosolic.

CCRs are considered members of the 3 β -hydroxysteroid dehydrogenase/dihydroflavonol-4-reductase (DFR) superfamily (short-chain dehydrogenase/reductase family), and have a conserved N-terminal region harboring the Rossmann fold NADPH binding motif.²⁸⁸ Multiple sequence alignments among various CCRs have also established the presence of a highly conserved motif, Asn-Trp-Tyr-Cys-Tyr, as involved in catalysis.²⁸⁸ There is currently no reported CCR 3D structure, however, although attempts to homology model it with the red yeast *Sporobolomyces salmonicolor* carbonyl reductase (SSCR) have been made.

In this context, Kamitori *et al.*²⁹⁰ published 3D structures of the NADPH-dependent SSCR in both apo form (PDB ID: 1ZZE) and as a binary complex with NAD(P)H (PDB ID: 1Y1P) obtained at 1.8 and 1.6 Å resolution, respectively. As a member of the short-chain dehydrogenase/reductase family, SSCR can convert ethyl-4-chloro-3-oxobutanoate (COBE, **119**) into the corresponding *S*-alcohol (**120**) (**Figure 30(a)**). The protein has a characteristic Rossmann fold NAD(P)H-binding motif (Gly19-Ala20-Asp21-Gly22-Phe23-Val24-Ala25), as well as a proposed hydrogen-bonding network from NADPH to the nitrogen atom of residues Gly22, Phe23, Val24, and Ala25.²⁹⁰ Additionally, Tyr208, Thr209, and Thr223 apparently form a hydrogen-bonding network with the carboxamide group and the Pro206 was viewed to form a stacking interaction with the nicotinamide ring of NADPH. The most interesting aspect of this structural analysis was the observed structural change in two loop regions, Ile91-Tyr101 and Pro216-Ser222 (**Figure 30(b)**), colored in violet, that occurred with NADPH binding. That is, movement of these mobile loops with NADPH binding resulted in formation of a

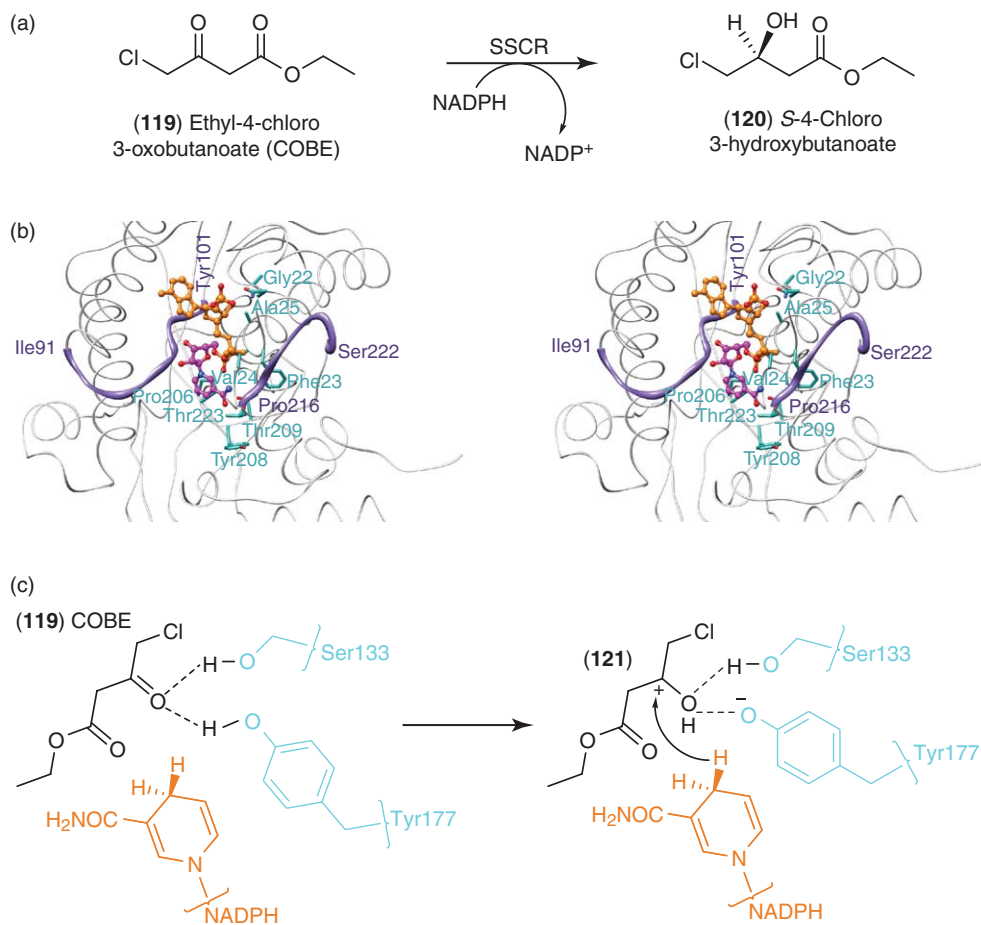


Figure 30 (a) Reaction catalyzed by *Sporobolomyces salmonicolor* carbonyl reductase (SSCR). (b) Three-dimensional view of NADPH binding in SCCR (PDB ID: 1Y1P) with the formation of hydrophobic mobile loop (colored in violet). (c) Proposed mechanism of oxygen activation and hydride transfer in SCCR with COBE (**119**).

hydrophobic channel, which not only provided access to the catalytic site by the substrate, but also facilitated binding of linear substrates, such as COBE (**119**). It was thus further considered that the catalytically active residues Tyr177 and Ser133 activated the carbonyl carbon atom of the substrate, COBE (**119**), thereby creating an intermediate carbocation (**121**) for efficient hydride transfer from the 4-carbon atom of NADPH as shown in **Figure 30(c)**.

From these data, Ma and Tian²⁹¹ also attempted to model a CCR 3D structure from wheat (*Triticum aestivum*) (TaCCR2) using SCCR as template. Although both NADPH-binding motifs (Thr-Gly-Ala-Gly-Gly-Phe-Ile) and an active site (Asn-Trp-Tyr-Cys-Tyr) were found, no further insight could, however, be gained into substrate binding, catalytic mechanism, or roles of specific residues for CCR catalysis.

Thus, although there are no reports of a 3D structure of CCR, the mechanism of reduction catalyzed by this enzyme has been discussed previously by Lewis *et al.*⁸ by comparison with 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA, **122**) reductase from *Pseudomonas mevalonii* (solved at 3.0 Å resolution).²⁹² Some insights gained from Lewis *et al.*⁸ and more recent studies may be applicable to CCR. Specifically, HMG-CoA reductase was established to be a dimer, in contrast to the monomeric CCR, and to possess two open cavities at the dimer interface embodying the active site, thus bringing together the conserved residues implicated in both binding and catalysis.²⁹² The domains Glu-X-X-X-Gly-X-X-X-X-Pro responsible for HMG-CoA (**122**) binding, Gly-X-X-Gly-Gly-X-Thr and Asp-Ala-Met-Gly-X-Asn for NAD(H) binding, as well as a Glu83 loop (for catalysis involving Glu83 and HMG-CoA (**122**) binding) and a His381 residue were

all presumed to be involved in catalysis. Specifically, three catalytic residues (Glu83, Asp283, and His381) were reportedly indispensable for catalysis by HMG-CoA reductase.

It was thus further proposed that the C-terminal residues (337–428) in HMG-CoA reductase, which contained His381 (one of the three residues for catalysis), formed a flap domain that closes the active site upon substrate binding.²⁹² Closing of the flexible flap domain and subsequent positioning of the His381 residue close to the scissile bond of HMG-CoA (**122**) were further suggested from the 3D structural analysis of the ternary complex with HMG-CoA (**122**)/NAD⁺ and mevalonate (**125**)/NADH, obtained at 2.8 Å resolution.²⁹³ Structural analysis of the ternary complexes of HMG-CoA reductase indicated that, in addition to the three previously determined catalytic residues Glu83, Asp283, and His381, another residue (Lys267), not previously implicated in the active site, was also critical for catalysis. This study thus demonstrated the roles of the participating residues in the active site in catalysis in HMG-CoA reductase, and a revised mechanism for the catalytic reduction was thus proposed (Figure 31).²⁹³

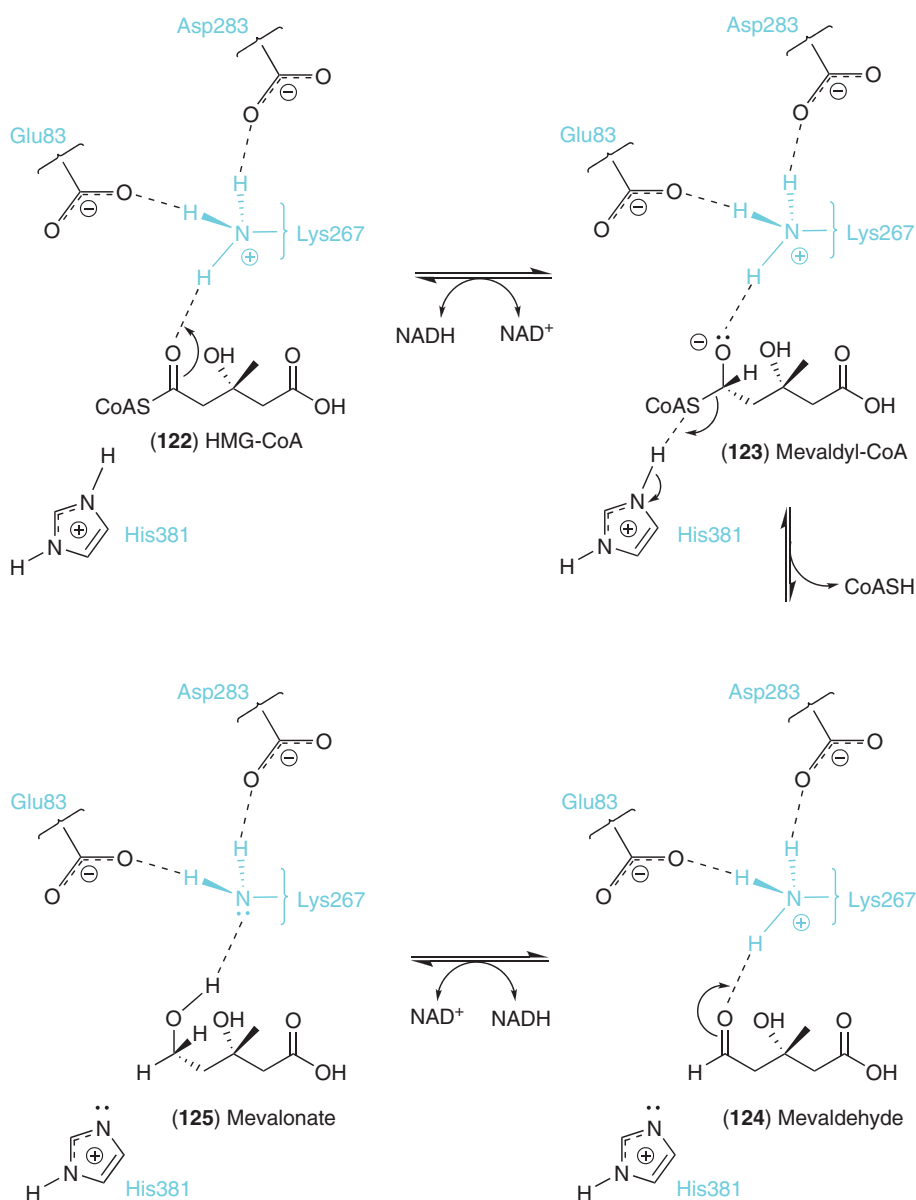


Figure 31 The proposed mechanism for the catalytic reduction by HMG-CoA reductase toward formation of mevalonate (**125**).

In the ternary complexes of the HMG-CoA reductase, with Lys267 oriented at the center, a hydrogen-bonding network was observed between Asn271, Glu83, and Asp283 in its nearest shell neighbor and this was further extended to an Arg85 residue in the neighboring second shell through hydrogen bonding to Glu83, Glu82, and Asp283. This lysine residue (Lys267) was proposed to polarize the carbonyl oxygen atom of HMG-CoA (**122**), dictating the formation of the hemi-thioacetal (mevaldyl-CoA, **123**), with the proximal His381 residue to the sulfur atom mevaldyl-CoA (**123**) subsequently participating in protonation of the leaving CoASH group, thereby facilitating generation of mevaldehyde (**124**) as depicted in **Figure 31**. Another histidine residue (His385) within the hydrogen-bonding distance to His381 was also suggested to be responsible for reprotonation of His381. The location of the electrostatic Lys267 further allowed activation of the aldehydic oxygen atom of the mevaldehyde (**124**) resulting in facile hydride transfer from another molecule of NADH, thereby completing the formation of mevalonate (**125**).

Accordingly, it is evident from the structural studies of both SSCR and HMG-CoA reductase that the electrostatic interaction of the catalytic residues (Tyr177 and Ser133 for SSCR and Lys267 for HMG-CoA reductase) with the carbonyl oxygen of the substrate can account for carbonyl group activation that facilitates efficient initial hydride transfer from NAD(P)H molecule. A similar situation can thus be anticipated for CCR with catalytic reaction commencing by activation of the thioester oxygen of the *p*-hydroxycinnamoyl CoA derivatives **9–13** through electrostatic interaction of the catalytic residues. This would induce a partial carbonium character in the thioester carbon atom, facilitating the type B hydride transfer from NAD(P)H,²⁹⁴ thereby forming the hemi-thioacetal intermediates **126–130** (**Figure 32**). Subsequent hydrolysis of the C–S bond of the hemi-thioacetal intermediates **126–130** and protonation of the released thioanion (CoAS⁻), as noted for HMG-CoA reductase,²⁹³ could result in formation of the corresponding *p*-hydroxycinnamyl aldehyde derivatives **14–18** as products, thereby completing the catalytic cycle. As before, ternary complexes of CCR are needed in order to understand fully how (differential) substrate recognition and catalysis occur.

6.17.3.7 Cinnamyl Alcohol Dehydrogenase: Comparison to Horse Liver Alcohol Dehydrogenase

CADs (EC 1.1.1.195) are a class of substrate-versatile NADPH-dependent cytosolic enzymes that differentially catalyze the reduction of various *p*-hydroxycinnamyl aldehydes **14–18** to afford the corresponding *p*-hydroxycinnamyl alcohols **19–23**. Initially discovered and characterized by Zenk and coworkers,^{283,295} the subsequent claim of cloning a CAD gene from bean (*Phaseolus vulgaris*)²⁹⁶ was incorrect. Misidentified in part due to lack of any biochemical confirmation, the putative CAD was shown later to encode malic enzyme.^{297,298} The first bona fide CAD gene, obtained from tobacco, encoded an enzyme of 357 amino acids corresponding to a

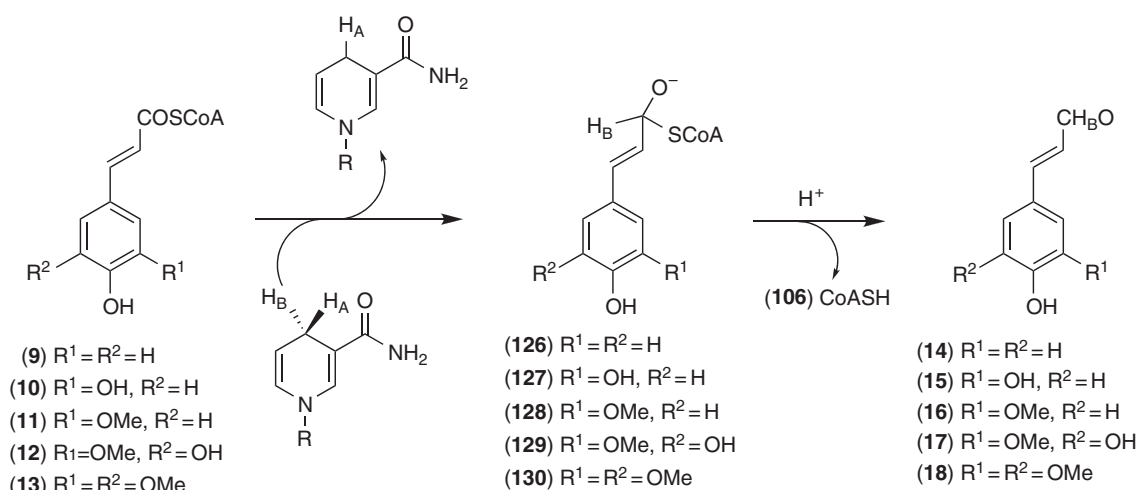


Figure 32 The proposed mechanism for CCR-mediated reaction via hemi-thioacetal intermediate **126–130** formation.

molecular weight of 38 760.²⁹⁹ In the *Arabidopsis* database, there were 17 genes annotated as CADs, 8 of which were not CADs, that is, they did not contain the characteristic Zn-binding domain of CAD and had no similarity (0.9–1.6%) to the tobacco CAD.³⁰⁰ Of the remaining nine, AtCAD5 (At4g34230) and AtCAD4 (At3g 19450) were the most catalytically active *in vitro*, having the highest homology to the tobacco CAD (~83% similarity). AtCAD5 utilized all five aldehydes **19–23** effectively, whereas AtCAD4 poorly used sinapyl aldehyde (**23**), that is, k_{cat}/K_m values ($\text{mol}^{-1} \text{s}^{-1}$) for AtCAD4 and AtCAD5 were 74 000 and 1 091 000 for **19**, 15 000 and 107 000 for **20**, 32 000 and 348 000 for **21**, 30 000 and 370 000 for **22**, and 2600 and 700 000 for **23**, respectively.³⁰⁰

CAD is a type A reductase,²⁹⁵ abstracting the proR hydride from NADPH via a two-electron hydride transfer mechanism. It belongs to the alcohol dehydrogenase (ADH) family, the members of which are zinc-dependent medium-chain dehydrogenases/reductases (MDR) with two Zn^{2+} ions per subunit. One zinc atom is thought to have a structural role, whereas the other forms the core of the catalytic site. Interestingly, medium-chain zinc-dependent ADHs exist as homotetramers as in bacteria, archaea, and yeast, or as homodimers as in plants and vertebrates.

6.17.3.7.1 Horse liver ADH

The 3D structures of ADHs have been determined from a variety of sources, with horse liver ADH (LADH) being the most well studied and where a considerable knowledge of dimeric ADH structures has now been gained.^{301,302} The enzymes in this class contain an active site Zn^{2+} , a substrate-binding pocket, and a coenzyme-binding domain (the so-called Rossmann fold). A characteristic feature of dimeric ADHs is the relative movement of the coenzyme-binding domain upon binding, which results in closure of the interdomain cleft. The extent of conformational change accompanied with such domain movements can vary within several ADHs to give rise to intermediate conformational states between fully ‘open’ and fully ‘closed’ conformations.³⁰³ Under suitable crystallization conditions, these intermediate states can be trapped in different steps of the catalytic cycle.

Apart from domain alteration, a catalytic mechanism based on a proton relay system and hydride transfer has been proposed mainly centered around the catalytic zinc site. The Zn^{2+} ion is ligated to two cysteine residues (Cys46 and Cys174 as in LADH³⁰²), a histidine (His67 as in LADH³⁰²), and/or glutamic acid (Glu155 as in human liver sorbitol dehydrogenase³⁰⁴) residues and/or a water molecule. Two alternative paths for the mode of action of the catalytic Zn^{2+} have been proposed: the first, which is the standard text book description of the ADH mechanism,³⁰⁵ involves displacement of a water molecule by the incoming substrate to leave the Zn^{2+} tetrahedrally coordinated to the substrate and the same three protein ligands as in the apo enzyme. Catalysis was then envisaged to occur via a proton relay system from the bulk solvent.

In an alternative mechanism, the substrate molecule is again coordinated to tetrahedral Zn^{2+} , with the coordinated water molecule now serving as a site for transient proton transfer, thereby generating a penta-coordinated zinc intermediate.^{306–309} Formation of this intermediate during substrate turnover was supported by time-resolved freeze-quenched X-ray absorption fine spectroscopic analysis of the thermophilic bacterium *Thermoanaerobacter brockii* alcohol dehydrogenase (TbADH).³¹⁰ These results thus provided further evidence for the dynamic alteration of the Zn^{2+} from a tetrahedral to a penta-coordinated form with detection of two new penta-coordinated intermediate states; these included the water molecule in the zinc coordination sphere during a single catalytic cycle.

6.17.3.7.2 CAD

By contrast, the 3D structural analysis of a bona fide CAD has been reported only for the *Arabidopsis* AtCAD4 and AtCAD5,³¹¹ as two other reports of putative CADs from *Saccharomyces cerevisiae*³¹² and *Populus tremuloides* (aspen)³¹³ are not currently considered as CADs proper. *Saccharomyces cerevisiae* does not produce metabolites in the phenylpropanoid pathway, thereby making it unclear as to why this specific metabolic role was being contemplated. In addition, the putative sinapyl aldehyde dehydrogenase (SAD) from aspen was not established to have the specific function claimed (see Davin *et al.*^{1,2} and Anterola and Lewis²⁰⁴ for a full discussion).

The 3D structures of dimeric AtCAD4/5 were obtained in the apo form and as a binary complex with cofactor (NADPH, **Figure 33**) at 2.0 and 2.6 Å resolution, respectively, but not as a ternary complex.³¹¹ A modeled ternary complex was, however, provisionally obtained with substrate *p*-coumaryl aldehyde (**14**) in the putative substrate-binding pocket.³¹¹

These data established that each subunit of AtCAD5 was composed of two distinct domains comprising the Rossmann fold with the nucleotide-binding domain (residues 163–301) and catalytic domain (residues 1–162

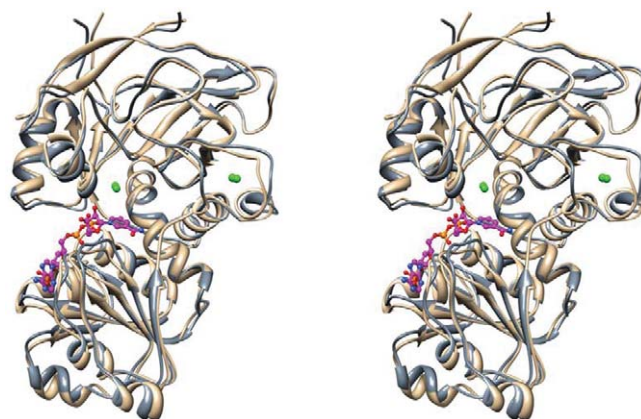


Figure 33 Stereoview of superimposed structure of AtCAD5 in its apo form (PDB ID: 2CF5) (colored in tan) and binary form (PDB ID: 2CF6) (colored in gray) complexed with NADP⁺ (pink). Structural and catalytic zinc are shown as green spheres.

and 302–357) (**Figure 34(a)**). The catalytic zinc was ligated to four amino acid residues, Cys47, His69, Glu70, and Cys163, in a tetrahedral orientation in both the apo form and the binary complex (**Figures 34(b)** and **34(c)**). More specifically, Zn²⁺ was located inside a cleft formed between the two domains positioned at the bottom of the hydrophobic substrate-binding pocket. By contrast, the structural zinc ion was tetrahedrally coordinated with four cysteine residues, Cys100, Cys103, Cys106, and Cys114, being located in a short α -helix-containing loop (residues 98–116) (**Figure 34(b)**).

Three flexible loops, residues 188–193, 211–216, and 275–286, have been suggested as involved in cofactor binding by AtCAD5. The first loop region is the highly conserved cofactor-binding domain Gly188–Leu189–Gly190–Gly191–Val192–Gly193, which participates in binding of the pyrophosphate group of NADP⁺. The second loop region (residues 211–216) contains a highly conserved Ser213 enabling preference for NADP(H) over NAD(H). Ser213 is located at the carboxy end of the β B strand, which forms a hydrogen bond with the 2'-phosphate group of NADP(H). The other residues Ser211, Ser212, and Lys216 located between β B and α C, which are also highly conserved among bona fide CADs, interact with the 2'-phosphate group of the adenine ribose with the side chain of Lys216 hydrogen bonded to the 3'-oxygen atom of the same ribose ring. The Gly275 in the third loop region (residues 275–286) located between β E and β F accounts for the interaction of the amide group of NADP⁺, thereby connecting the two domains of AtCAD5.

The substrate-binding pocket of AtCAD5 is provisionally considered to be composed of mostly hydrophobic 12 amino acid residues, with 9 residues Thr49, Gln53, Leu58, Met60, Cys95, Trp119, Val276, Phe299, and Ile300 from one subunit and 3 residues Pro286, Met289, and Leu290 from the other subunit (**Figure 34(c)**). It is interesting to note though that the entry region of the putative binding pocket of AtCAD5 is considered much larger in comparison to horse liver ADH, this being accounted for possibly due to the major deletion between β 8 and β 9 region in AtCAD5. Additionally, comparison of the structures of AtCAD4 and AtCAD5 did not indicate any notable differences in the substrate-binding site, with most residues being conserved within potential interacting distances to the substrate. However, the substrate-binding site of AtCAD4 was lined with 12 residues, which differed only in two residues from AtCAD5, that is, Cys95 and Met289 in AtCAD5 were replaced by Val96 and Ile290, respectively, in AtCAD4.

6.17.3.7.3 Comparison of AtCAD5 with poplar 'SAD' and yeast 'CAD'

Additionally, a detailed comparison of AtCAD5 and AtCAD4 with the putative yeast CAD³¹² and aspen SAD³¹³ (as discussed earlier) with respect to their amino acid sequences and putative binding site geometries was also made.³¹¹ The similarity and identity of aspen SAD to AtCAD5 (62.6/53.1%) and AtCAD4 (62.5/53.3%) were found to be relatively low, as were those for the putative yeast CAD, that is, 69.2/35.7 and 67.6/35.4%, respectively, for AtCAD5 and AtCAD4. Moreover, the substrate-binding site of AtCAD5/4 differed considerably and was considered to be smaller than that of the putative yeast CAD and aspen SAD. Furthermore, the presumed active site of aspen SAD contained only two residues (Cys98 and Ile292) identical

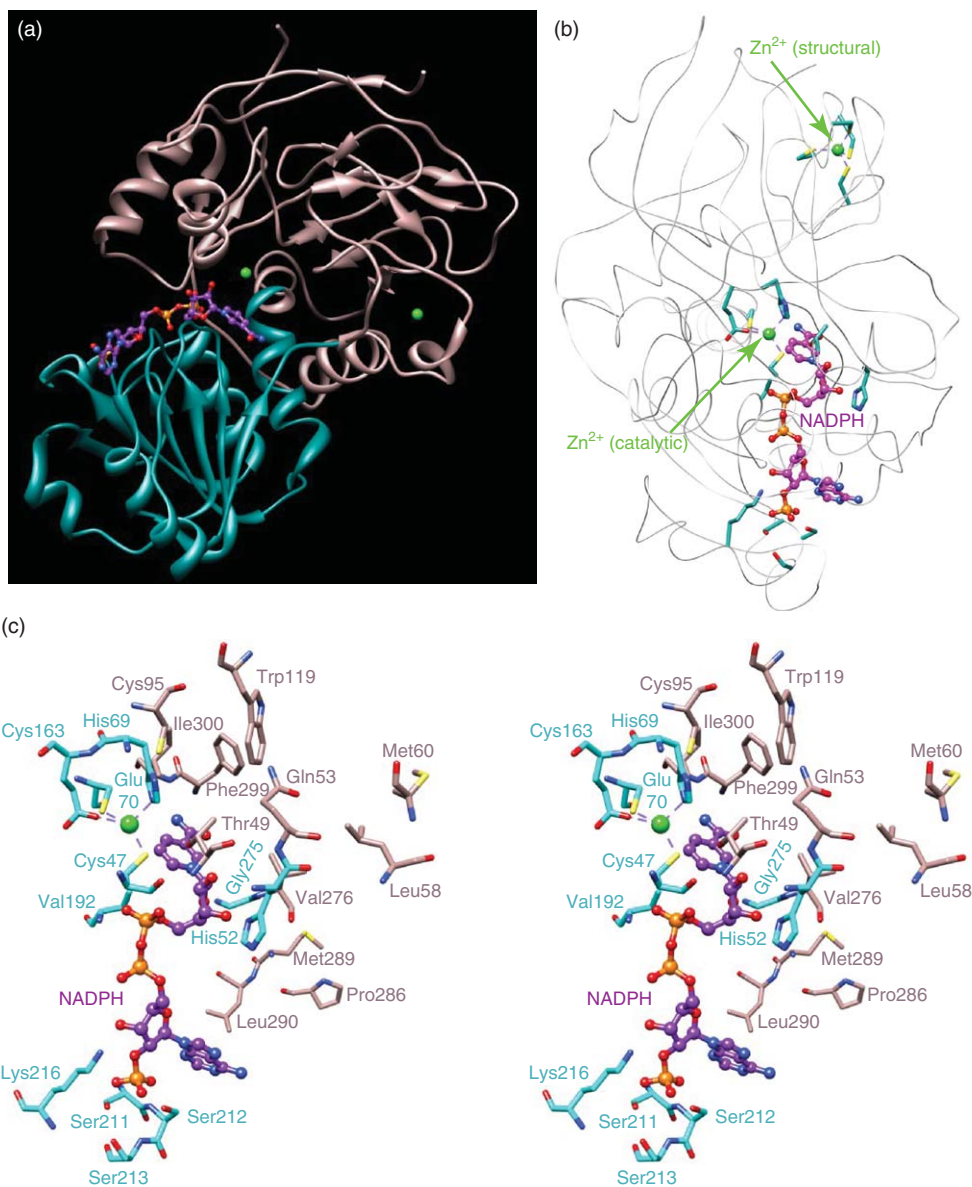


Figure 34 (a) Ribbon representation of AtCAD5 (PDB ID: 2CF6) complexed with NADP⁺ showing the nucleotide binding (cyan) and catalytic (pink) domains. (b) Structure of substrate-binding pocket of the NADP⁺ binary form showing both the structural and catalytic zinc sites. (c) Stereoview of the active site of AtCAD5, the catalytic Zn²⁺ being tetrahedrally coordinated by Cys47, His69, Cys163, and Glu70 residues and the interaction of Val192, Ser211, Ser213, Lys216, and Gly275 residues with the NADP⁺ molecule, as well as the proposed hydrophobic amino acid residues (pink) involved in the substrate binding pocket in the binary complex.

to the above mentioned 12 amino acid residues in the putative substrate binding site of AtCAD5 (Cys95 and Ile300), whereas only one residue Leu291 (equivalent residue in AtCAD5 is Leu290) was conserved for the putative yeast CAD among the 12 residues of AtCAD5 in the substrate-binding site. It will thus be of interest to establish the physiological/biochemical roles of both the yeast ‘CAD’ and the aspen ‘SAD’.

6.17.3.7.4 CAD catalysis and putative proton relay

A mechanism initially considered for AtCAD5 involves displacement of the zinc-bound Glu70 residue by the aldehyde substrate through coordination via the carbonyl oxygen thereby retaining the tetrahedral symmetry of Zn²⁺ ion in its apo form. Thus, it was initially considered that activation of the aldehyde carbonyl might

occur by coordination with Zn^{2+} to enable a facile two-electron type A hydride transfer from the cofactor (NADPH) thereby permitting a proton relay mechanism from bulk solvent through participation of the Thr49, His52, and Asp57 residues together with the 2' and 3'-oxygen atoms of the ribose ring of the NADPH cofactor. This proton relay mechanism was considered based on structural studies of the alcohol dehydrogenase from the archaeon *Sulfolobus solfataricus* (SsADH).³¹⁴ However, site-directed mutagenesis (C. Lee *et al.*, manuscript in preparation) where Thr49, His52, and Asp57 were individually replaced by Ala established that only Thr49 was involved in catalysis, that is, there is no such relay operative with these residues.

A recent report on high-resolution 3D structures of ternary complexes of LADH with NADH and inhibitors (dimethyl sulfoxide (DMSO, **131**) and isobutyramide (IBA, **132**)) at 1 Å resolution has provided evidence for a ligand coordination switch at the active site metal ion.³¹⁵ A ternary complex with a slower binding rate of inhibitor to the active metal center was designed by substitution of catalytic Zn^{2+} with Cd^{2+} (Cd being considered to have a slower rate of binding capacity than Zn) and a weak inhibitor (DMSO (**131**)). A flash-freezing protocol was then used to capture the ternary complex of the Cd-substituted LADH in the transition state with the inhibitor molecule (DMSO (**131**)) being partly in both the primary and secondary sphere of ligation of the active metal site. Structural analysis of the corresponding ternary complex (NADH and DMSO (**131**)) indicated that, in addition to coordination with amino acid residues Cys46, His67, Cys174, and inhibitor (DMSO (**131**)), the active metal ion in the transition state was also bound to a hydroxide ion (derived from solvent water molecule) in close proximity to the pyridine ring of NADH. However, with the ternary complex of LADH (retaining the active Zn site) with NADH and IBA (**132**) (strong inhibitor), no complexation with hydroxide ion in the active metal site was observed (**Figure 35(b)**). This structural comparison study suggested that during catalytic conversion of the substrate in ADHs, a water molecule (as hydroxide ion) might be

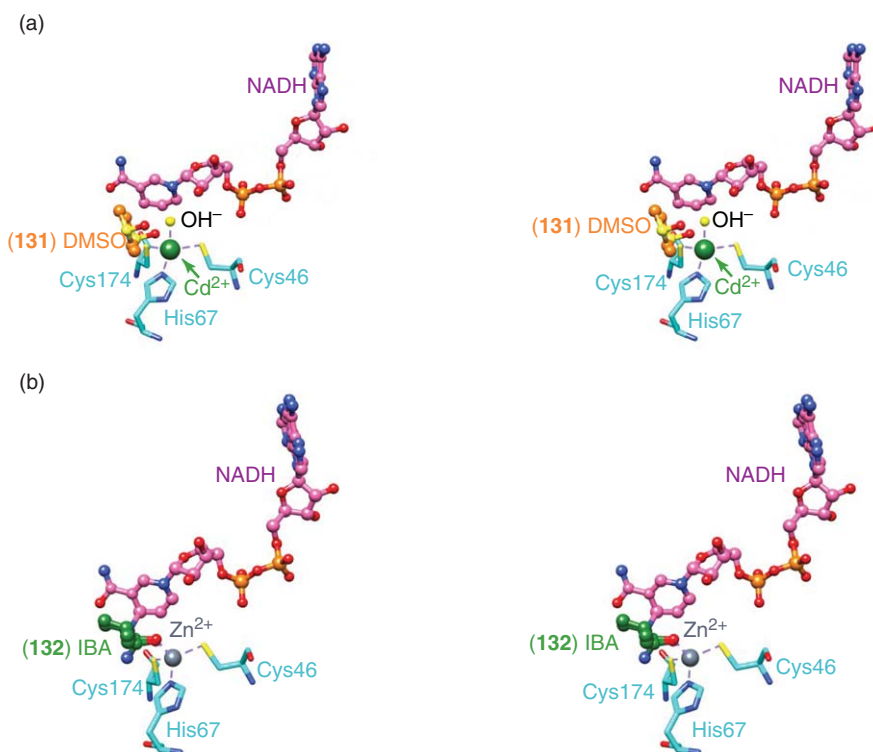


Figure 35 Comparison between the ligand-binding site (stereoview) of (a) the transient penta-coordinated cadmium (green)-substituted LADH in ternary complex with the inhibitor DMSO (**131**) (orange) and NADH (pink) (PDB ID: 2JHF), showing coordination with OH^- ion (yellow) derived from a water molecule and (b) tetrahedrally coordinated zinc (gray) in liver sorbitol dehydrogenase ternary complex with the inhibitor IBA (**132**) (green) and NADH (pink) (PDB ID: 2JHG).

involved as a transient penta-coordinated active metal center, with the hydroxide ion eliminated and/or totally removed from the catalytic site for completion of catalysis.

Furthermore, another very recent report on the high-resolution structure analysis of a series of binary and ternary complexes of glucose dehydrogenase, a zinc-dependent MDR enzyme from *Haloferox mediterranei*, has provided additional evidence for associated electronic changes at the zinc site resulting from movement of the zinc ion along with a ligated water molecule.³¹⁶ Such changes in the zinc coordination site during a single catalytic cycle have suggested that active site dynamic coordination chemistry is omnipresent in zinc metalloenzymes. Thus, it will be instructive to obtain a high-resolution 3D structure of CAD with a series of ternary complexes (cofactor and different substrates, as well as various inhibitors) to validate the mode of action of the zinc coordination site, that is, in order to provide the molecular basis for the proton relay during catalysis. In addition, this will provide useful insights into how the various substrates are differentially bound and processed.

6.17.4 Conclusions

Over the past decade, good progress has been made in understanding the biochemical mechanism and kinetic parameters of the various enzymes in the monolignol pathway, as well as how plants produce the corresponding aromatic amino acids Phe (1)/Tyr (2) for phenylpropanoid metabolism, and so on. Although monolignol pathway products represent one of the nature's greatest sinks in organic carbon (e.g., for lignin formation), much though still remains to be done. For example, it will be important to ascertain all of the factors affecting carbon allocation to the monolignol pathway, particularly those that are upstream of PAL/TAL, as well as to identify unambiguously the enzymes/genes involved. The next decade should also provide us with the means to probe further the biochemical mechanisms and to establish the factors that result in both narrow and broad substrate versatility, depending on the enzyme.

Acknowledgments

The authors thank the US Department of Energy (DE FG03-97ER20259), the US National Science Foundation (MCB-0417291), the National Institute of General Medical Sciences (5 R01 GM066173-02), the United States Department of Agriculture (Agricultural Plant Biochemistry #2006-03339), the BioEnergy Science Center, the US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science, McIntire Stennis, and the G. Thomas and Anita Hargrove Center for Plant Genomic Research for generous financial support.

References

1. L. B. Davin; A. M. Patten; M. Jourdes; N. G. Lewis, Lignins: A Twenty-First Century Challenge. In *Biomass Recalcitrance. Deconstructing the Plant Cell Wall for Bioenergy*; M. E. Himmel, Ed.; Blackwell Publishing: Oxford, UK, 2008; pp 213–305.
2. L. B. Davin; M. Jourdes; A. M. Patten; K.-W. Kim; D. G. Vassão; N. G. Lewis, *Nat. Prod. Rep.* **2008**, *25*, 1015–1090.
3. G. Brunow; I. Kilpeläinen; J. Sipilä; K. Syrjänen; P. Karhunen; H. Setälä; P. Rummakko, Oxidative Coupling of Phenols and the Biosynthesis of Lignin. In *Lignin and Lignan Biosynthesis*; N. G. Lewis, S. Sarkanen, Eds.; ACS Symposium Series: Washington, DC, USA, 1998; Vol. 697, pp 131–147.
4. L. B. Davin; N. G. Lewis, *Phytochem. Rev.* **2003**, *2*, 257–288.
5. L. B. Davin; N. G. Lewis, *Curr. Opin. Biotechnol.* **2005**, *16*, 398–406.
6. N. G. Lewis; L. B. Davin, Lignans: Biosynthesis and Function. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed.; Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Series Eds.; Elsevier: Oxford, 1999; pp 639–712.
7. D. G. Vassão; K.-W. Kim; L. B. Davin; N. G. Lewis, Lignans (Neolignans) and Allyl/Propenyl Phenols: Biogenesis, Structural Biology, and Biological/Human Health Considerations. In *Comprehensive Natural Products Chemistry II, Vol. 1: Structural Diversity I*; C. Townsend, Y. Ebizuka, Eds.; L. N. Mander, H.-W. Liu, Series Eds.; Elsevier: Oxford, UK, 2009; p 817–928.
8. N. G. Lewis; L. B. Davin; S. Sarkanen, The Nature and Function of Lignins. In *Comprehensive Natural Products Chemistry, Vol. 3.: Carbohydrates and Their Derivatives Including Tannins, Cellulose and Related Lignins*; M. B. Pinto, Ed.; Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Series Eds.; Elsevier: Oxford, 1999; pp 617–745.

9. A. M. Patten; D. G. Vassão; M. P. Wolcott; L. B. Davin; N. G. Lewis, Trees: A Remarkable Biochemical Bounty. In *Comprehensive Natural Products Chemistry II, Vol. 3: Discovery, Development and Modification of Bioactivity*; R. Verpoorte, Ed.; L. N. Mander, H.-W. Liu, Series Eds.; Elsevier: Oxford, UK, 2009; p 1174–1296.
10. V. M. Reznikov; M. F. Mikhaseva; M. A. Zil'bergleit, *Chem. Nat. Compd.* **1979**, *14*, 554–556.
11. I. V. Dovgan; E. I. Medvedeva, *Chem. Nat. Compd.* **1983**, *19*, 81–84.
12. I. V. Dovgan; E. I. Medvedeva; E. N. Yanishevskaya, *Chem. Nat. Compd.* **1983**, *19*, 84–87.
13. P. T. Martone; J. M. Estevez; F. Lu; K. Ruel; M. W. Denny; C. Somerville; J. Ralph, *Curr. Biol.* **2009**, *19*, 169–175.
14. R. M. Bateman; P. R. Crane; W. A. DiMichele; P. R. Kenrick; N. P. Rowe; T. Speck; W. E. Stein, *Annu. Rev. Ecol. Syst.* **1998**, *29*, 263–292.
15. L. E. Graham, *Origin of Land Plants*; John Wiley and Sons, Inc: New York, 1993; pp 287.
16. D. Edwards; K. L. Davies; L. Axe, *Nature* **1992**, *357*, 683–685.
17. P. Kenrick; P. R. Crane, *The Origin and Early Diversification of Land Plants. A Cladistic Study*; Smithsonian Institution Press: Washington, DC, 1997; pp 560.
18. M. A. Bernards; N. G. Lewis, *Phytochemistry* **1998**, *47*, 915–933.
19. M. A. Bernards, *Can. J. Bot.* **2002**, *80*, 227–240.
20. R. Franke; L. Schreiber, *Curr. Opin. Plant Biol.* **2007**, *10*, 252–259.
21. G. Forkmann; W. Heller, Biosynthesis of Flavonoids. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed., Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Series Eds.; Elsevier: Oxford, 1999; pp 713–748.
22. J. Schröder, The Chalcone/Stilbene Synthase-Type Family of Condensing Enzymes. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed.; Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Series Eds.; Elsevier: Oxford, 1999; pp 749–771.
23. T. Koeduka; E. Fridman; D. R. Gang; D. G. Vassão; B. L. Jackson; C. M. Kish; I. Orlova; S. M. Spassova; N. G. Lewis; J. P. Noel; T. J. Baiga; N. Dudareva; E. Pichersky, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10128–10133.
24. D. G. Vassão; D. R. Gang; T. Koeduka; B. Jackson; E. Pichersky; L. B. Davin; N. G. Lewis, *Org. Biomol. Chem.* **2006**, *4*, 2733–2744.
25. D. G. Vassão; S.-J. Kim; J. K. Milhollan; D. Eichinger; L. B. Davin; N. G. Lewis, *Arch. Biochem. Biophys.* **2007**, *465*, 209–218.
26. D. G. Vassão; L. B. Davin; N. G. Lewis, Metabolic Engineering of Plant Allyl/Propenyl Phenol and Lignin Pathways: Future Potential for Biofuels/Bioenergy, Polymer Intermediates and Specialty Chemicals? In *Advances in Plant Biochemistry Molecular Biology*; H. J. Bohnert, H. T. Nguyen, N. G. Lewis, Eds.; Elsevier: Oxford, 2008; Vol. 1, pp 385–428.
27. H. Bickel; L. Palme; G. Schultz, *Phytochemistry* **1978**, *17*, 119–124.
28. H. Bickel; G. Schultz, *Phytochemistry* **1979**, *18*, 498–499.
29. B. Buchholz; G. Schultz, *Z. Pflanzenphysiol.* **1980**, *100*, 209–215.
30. J. F. Eykman, *Recl. Trav. Chim. Pays Bas* **1885**, *4*, 32.
31. J. F. Eykman, *Ber. Dtsch. Chem. Ges.* **1891**, *24*, 1278–1303.
32. R. Bentley, *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 307–384.
33. B. D. Davis, *J. Biol. Chem.* **1951**, *191*, 315–325.
34. D. G. Gilchrist; T. Kosuge, Aromatic Amino Acid Biosynthesis and its Regulation. In *The Biochemistry of Plants. A Comprehensive Treatise*; P. K. Stumpf, E. E. Conn, Eds. in chief, Vol. 5: Amino Acids and Derivatives. B. J. Mifflin, Ed.; Academic Press: New York, 1980; pp 507–531.
35. K. M. Herrmann; L. M. Weaver, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 473–503.
36. A. J. Pittard, Biosynthesis of the Aromatic Amino Acids. In *Escherichia coli and Salmonella typhimurium – Cellular and Molecular Biology*, 2nd ed.; F. C. Neidhardt, R. Curtiss, III, J. L. Ingraham, E. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, Eds.; American Society of Microbiology: Washington, DC, 1996; Vol. 1, pp 458–484.
37. J. Schmid; N. Amrhein, *Phytochemistry* **1995**, *39*, 737–749.
38. G. S. Byng; R. J. Whitaker; C. L. Shapiro; R. A. Jensen, *Mol. Cell Biol.* **1981**, *1*, 426–438.
39. T. A. A. Dopheide; P. Crewther; B. E. Davidson, *J. Biol. Chem.* **1972**, *247*, 4447–4452.
40. S. L. Stenmark; D. L. Pierson; R. A. Jensen; G. I. Glover, *Nature* **1974**, *247*, 290–292.
41. T. Xia; S. Ahmad; G. Zhao; R. A. Jensen, *Arch. Biochem. Biophys.* **1991**, *286*, 461–465.
42. G. S. Zhao; T. H. Xia; R. S. Fischer; R. A. Jensen, *J. Biol. Chem.* **1992**, *267*, 2487–2493.
43. G. Zhao; T. Xia; L. O. Ingram; R. A. Jensen, *Eur. J. Biochem.* **1993**, *212*, 157–165.
44. G. H. Braus, *Microbiol. Rev.* **1991**, *55*, 349–370.
45. R. A. Jensen; L. Zamir; M. Saint Pierre; N. Patel; D. L. Pierson, *J. Bacteriol.* **1977**, *132*, 896–903.
46. F. Lingens; W. Goebel; H. Uesseler, *Biochem. Z.* **1966**, *346*, 357–367.
47. F. Lingens; W. Goebel; H. Uesseler, *Eur. J. Biochem.* **1967**, *1*, 363–374.
48. S. Ahmad; R. A. Jensen, *Mol. Biol. Evol.* **1988**, *5*, 282–297.
49. R. A. Jensen, Tyrosine and Phenylalanine Biosynthesis: Relationship between Alternative Pathways, Regulation and Subcellular Location. In *Recent Advances in Phytochemistry, Vol. 20: The Shikimic Acid Pathway*; E. E. Conn, Ed.; Plenum Press: New York, 1985; pp 57–81.
50. E. Jung; L. O. Zamir; R. A. Jensen, *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7231–7235.
51. D. L. Siehl; E. E. Conn, *Arch. Biochem. Biophys.* **1988**, *260*, 822–829.
52. K. M. Warpeha; S. S. Lateef; Y. Lapik; M. Anderson; B.-S. Lee; L. S. Kaufman, *Plant Physiol.* **2006**, *140*, 844–855.
53. M.-H. Cho; O. R. A. Corea; H. Yang; D. L. Bedgar; D. D. Laskar; A. M. Anterola; F. A. Moog-Anterola; R. L. Hood; S. E. Kohalmi; M. A. Bernards; C. Kang; L. B. Davin; N. G. Lewis, *J. Biol. Chem.* **2007**, *282*, 30827–30835.
54. J. Ehrling; N. Mattheus; D. S. Aeschliman; E. Li; B. Hamberger; I. F. Cullis; J. Zhuang; M. Kaneda; S. D. Mansfield; L. Samuels; K. Ritland; B. E. Ellis; J. Bohlmann; C. J. Douglas, *Plant J.* **2005**, *42*, 618–640.
55. T. Yamada; F. Matsuda; K. Kasai; S. Fukuoka; K. Kitamura; Y. Tozawa; H. Miyagawa; K. Wakasa, *Plant Cell* **2008**, *20*, 1316–1329.
56. M. Benesova; R. Bode, *Phytochemistry* **1992**, *31*, 2983–2987.

57. T. A. D'Amato; R. J. Ganson; C. G. Gaines; R. A. Jensen, *Planta* **1984**, *162*, 104–108.
58. J. Eberhard; T. T. Ehrler; P. Epple; G. Felix; H.-R. Raesecke; N. Amrhein; J. Schmid, *Plant J.* **1996**, *10*, 815–821.
59. J. Eberhard; M. Bischoff; H.-R. Raesecke; N. Amrhein; J. Schmid, *Plant Mol. Biol.* **1996**, *31*, 917–922.
60. O. Emanuelsson; H. Nielsen; S. Brunak; G. von Heijne, *J. Mol. Biol.* **2000**, *300*, 1005–1016.
61. P. Rippert; J. Puyaubert; D. Grisolle; L. Derrier; M. Matringe, *Plant Physiol.* **2009**, *149*, 1251–1260.
62. C. A. Bonner; R. A. Jensen, Upstream Metabolic Segments that Support Lignin Biosynthesis. In *Lignin and Lignan Biosynthesis*; N. G. Lewis, S. Sarkanen, Eds.; ACS Symposium Series: Washington, DC, 1998; Vol. 697, pp 29–41.
63. P. Rippert; M. Matringe, *Eur. J. Biochem.* **2002**, *269*, 4753–4761.
64. P. Rippert; M. Matringe, *Plant Mol. Biol.* **2002**, *48*, 361–368.
65. J. A. Connelly; E. E. Conn, *Z. Naturforsch. C* **1986**, *41*, 69–78.
66. C. G. Gaines; G. S. Byng; R. J. Whitaker; R. A. Jensen, *Planta* **1982**, *156*, 233–240.
67. D. L. Siehl, The Biosynthesis of Tryptophan, Tyrosine and Phenylalanine from Chorismate. In *Plant Amino Acids. Biochemistry and Biotechnology*; B. K. Singh, Ed.; Marcel Dekker: New York, 1999; pp 171–204.
68. S. Ahmad; R. A. Jensen, *FEBS Lett.* **1987**, *216*, 133–139.
69. T. Higuchi; Y. Ito; I. Kawamura, *Phytochemistry* **1967**, *6*, 875–881.
70. M. Petersen; M. S. J. Simmonds, *Phytochemistry* **2003**, *62*, 121–125.
71. A. R. Battersby; J. J. Reynolds, *Proc. Chem. Soc.* **1960**, 346–347.
72. A. R. Battersby; R. B. Herbert, *Proc. Chem. Soc.* **1964**, 260.
73. A. R. Battersby; R. Binks; D. A. Yeowell, *Proc. Chem. Soc.* **1964**, 86.
74. E. Leete; P. E. Nemeth, *J. Am. Chem. Soc.* **1960**, *82*, 6055–6057.
75. E. Leete; P. E. Nemeth, *J. Am. Chem. Soc.* **1961**, *83*, 2192–2194.
76. E. Leete, *J. Am. Chem. Soc.* **1963**, *85*, 3666–3669.
77. K. Yoshida; T. Hayashi; K. Sano, *Phytochemistry* **1988**, *27*, 1375–1378.
78. C. A. Bonner; R. A. Jensen, *Arch. Biochem. Biophys.* **1985**, *238*, 237–246.
79. W. De-Eknamkul; B. E. Ellis, *Arch. Biochem. Biophys.* **1988**, *267*, 87–94.
80. J. L. Rubin; R. A. Jensen, *Plant Physiol.* **1979**, *64*, 727–734.
81. D. L. Siehl; J. A. Connelly; E. E. Conn, *Z. Naturforsch. C* **1986**, *41*, 79–86.
82. F. Wightman; J. C. Forest, *Phytochemistry* **1978**, *17*, 1455–1471.
83. *The Arabidopsis Genome Initiative*, *Nature* **2000**, *408*, 796–815.
84. A. Bateman; L. Coin; R. Durbin; R. D. Finn; V. Hollich; S. Griffiths-Jones; A. Khanna; M. Marshall; S. Moxon; E. L. L. Sonnhammer; D. J. Studholme; C. Yeats; S. R. Eddy, *Nucleic Acids Res.* **2004**, *32*, D138–D141.
85. B. K. Singh; B. F. Matthews, *Amino Acids* **1994**, *7*, 165–174.
86. R. J. Ganson; R. A. Jensen, *Plant Physiol.* **1987**, *83*, 479–482.
87. J. L. Rubin; R. A. Jensen, *Plant Physiol.* **1985**, *79*, 711–718.
88. B. K. Singh; D. L. Siehl; J. A. Connelly, Shikimate Pathway: Why Does it Mean so Much to so Many? In *Oxford Survey of Plant Molecular and Cellular Biology*; B. J. Mifflin, Ed.; Oxford University Press: Oxford, 1991; Vol. 7, pp 143–185.
89. J. Pittard; F. Gibson, *Curr. Top. Cell. Regul.* **1970**, *2*, 29–63.
90. C. Poulsen; R. J. M. Bongaerts; R. Verpoorte, *Eur. J. Biochem.* **1993**, *212*, 431–440.
91. R. M. Romero; M. F. Roberts; J. D. Phillipson, *Phytochemistry* **1995**, *39*, 263–276.
92. J. E. Brotherton; R. M. Hauptmann; J. M. Widholm, *Planta* **1986**, *168*, 214–221.
93. D. G. Gilchrist; T. Kosuge, *Arch. Biochem. Biophys.* **1975**, *171*, 36–42.
94. S. K. Goers; R. A. Jensen, *Planta* **1984**, *162*, 117–124.
95. W. E. Dyer; J. M. Henstrand; A. K. Handa; K. M. Herrmann, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7370–7373.
96. J. Görlach; H.-R. Raesecke; G. Abel; R. Wehrli; N. Amrhein; J. Schmid, *Plant J.* **1995**, *8*, 451–456.
97. B. Keith; X. Dong; F. M. Ausubel; G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8821–8825.
98. K. F. McCue; E. E. Conn, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7374–7377.
99. G. Kuroki; E. E. Conn, *Plant Physiol.* **1988**, *86*, 895–898.
100. K. K. Niyogi; G. R. Fink, *Plant Cell* **1992**, *4*, 721–733.
101. M. A. Bernards; M. L. Lopez; J. Zajicek; N. G. Lewis, *J. Biol. Chem.* **1995**, *270*, 7382–7386.
102. L. B. Davin; N. G. Lewis, Phenylpropanoid Metabolism: Biosynthesis of Monolignols, Lignans and Neolignans, Lignins and Suberins. In *Recent Advances in Phytochemistry; Vol. 26: Phenolic Metabolism in Plants*; H. A. Stafford, R. K. Ibrahim, Eds.; Plenum Press: New York, 1992; pp 325–375.
103. F. Galvano; L. La Fauci; G. Lazzarino; V. Fogliano; A. Ritieni; S. Ciappellano; N. C. Battistini; B. Tavazzi; G. Galvano, *J. Nutr. Biochem.* **2004**, *15*, 2–11.
104. J. B. Harborne; C. A. Williams, *Phytochemistry* **2000**, *55*, 481–504.
105. R. A. Dixon, Isoflavonoids: Biochemistry, Molecular Biology and Biological Functions. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed.; Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Series Eds.; Elsevier: Oxford, 1999; pp 773–823.
106. K. Tan; H. Li; R. Zhang; M. Gu; S. T. Clancy; A. Joachimiak, *J. Struct. Biol.* **2008**, *162*, 94–107.
107. S. Zhang; D. B. Wilson; B. Ganem, *Biochemistry* **2000**, *39*, 4722–4728.
108. J. S. Liberles; M. Thórolfsson; A. Martínez, *Amino Acids* **2005**, *28*, 1–12.
109. G. Pohnert; S. Zhang; A. Husain; D. B. Wilson; B. Ganem, *Biochemistry* **1999**, *38*, 12212–12217.
110. R. A. Razal; S. Ellis; S. Singh; N. G. Lewis; G. H. N. Towers, *Phytochemistry* **1996**, *41*, 31–35.
111. P. S. van Heerden; G. H. N. Towers; N. G. Lewis, *J. Biol. Chem.* **1996**, *271*, 12350–12355.
112. G. H. N. Towers; S. Singh; P. S. van Heerden; J. Zuiches; N. G. Lewis, Integrating Nitrogen and Phenylpropanoid Metabolic Pathways in Plants and Fungi. In *Lignin and Lignan Biosynthesis*; N. G. Lewis, S. Sarkanen, Eds.; ACS Symposium Series: Washington, DC, 1998; Vol. 697, pp 42–54.
113. E. A. Havir; P. D. Reid; H. V. Marsh, Jr., *Plant Physiol.* **1971**, *48*, 130–136.
114. E. L. Camm; G. H. N. Towers, *Phytochemistry* **1973**, *12*, 961–973.

115. J. Rösler; F. Krekel; N. Amrhein; J. Schmid, *Plant Physiol.* **1997**, *113*, 175–179.
116. J. A. Kyndt; T. E. Meyer; M. A. Cusanovich; J. J. Van Beeumen, *FEBS Lett.* **2002**, *512*, 240–244.
117. K. T. Watts; B. N. Mijts; P. C. Lee; A. J. Manning; C. Schmidt-Dannert, *Chem. Biol.* **2006**, *13*, 1317–1326.
118. G. V. Louie; M. E. Bowman; M. C. Moffitt; T. J. Baiga; B. S. Moore; J. P. Noel, *Chem. Biol.* **2006**, *13*, 1327–1338.
119. K. R. Hanson; E. A. Havir, Phenylalanine Ammonia-Lyase. In *The Biochemistry of Plants. A comprehensive Treatise*, P. K. Stumpf, E. E. Conn, Eds. in Chief, Vol. 7: *Secondary Plant Products*; E. E. Conn, Ed.; Academic Press: New York, 1981; pp 577–625.
120. F. C. Cochrane; L. B. Davin; N. G. Lewis, *Phytochemistry* **2004**, *65*, 1557–1564.
121. L. Poppe; J. Rétey, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 3668–3688.
122. S. Bartsch; U. T. Bornscheuer, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 3362–3365.
123. I. L. Givot; T. A. Smith; R. H. Abeles, *J. Biol. Chem.* **1969**, *244*, 6341–6353.
124. R. B. Wickner, *J. Biol. Chem.* **1969**, *244*, 6550–6552.
125. K. R. Hanson; R. H. Wightman; J. Staunton; A. R. Battersby, *J. Chem. Soc. Chem. Commun.* **1971**, 185–186.
126. R. Ife; E. Haslam, *J. Chem. Soc. C* **1971**, 2818–2821.
127. B. Langer; M. Langer; J. Rétey, *Adv. Protein Chem.* **2001**, *58*, 175–214.
128. M. M. Jayasekera; W. Shi; G. K. Farber; R. E. Viola, *Biochemistry* **1997**, *36*, 9145–9150.
129. C. W. Levy; P. A. Buckley; S. Sedelnikova; Y. Kato; Y. Asano; D. W. Rice; P. J. Baker, *Structure* **2002**, *10*, 105–113.
130. W. Shi; J. Dunbar; M. M. K. Jayasekera; R. E. Viola; G. K. Farber, *Biochemistry* **1997**, *36*, 9136–9144.
131. M. Langer; G. Reck; J. Reed; J. Rétey, *Biochemistry* **1994**, *33*, 6462–6467.
132. M. Langer; A. Lieber; J. Rétey, *Biochemistry* **1994**, *33*, 14034–14038.
133. H.-G. Sahl; R. W. Jack; G. Bierbaum, *Eur. J. Biochem.* **1995**, *230*, 827–853.
134. Y. Ohmiya; H. Hayashi; T. Kondo; Y. Kondo, *J. Biol. Chem.* **1990**, *265*, 9066–9071.
135. D. B. Volkin; A. M. Klibanov, *J. Biol. Chem.* **1987**, *262*, 2945–2950.
136. B. Schuster; J. Rétey, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8433–8437.
137. T. F. Schwede; J. Rétey; G. E. Schulz, *Biochemistry* **1999**, *38*, 5355–5361.
138. J. Rétey, *Biochim. Biophys. Acta* **2003**, *1647*, 179–184.
139. T. F. Schwede; M. Bädeker; M. Langer; J. Rétey; G. E. Schulz, *Protein Eng.* **1999**, *12*, 151–153.
140. D. Hernandez; A. T. Phillips, *Protein Expr. Purif.* **1993**, *4*, 473–478.
141. D. Röther; L. Poppe; S. Viergutz; B. Langer; J. Rétey, *Eur. J. Biochem.* **2001**, *268*, 6011–6019.
142. D. Röther; L. Poppe; G. Morlock; S. Viergutz; J. Rétey, *Eur. J. Biochem.* **2002**, *269*, 3065–3075.
143. D. Röther; D. Merkel; J. Rétey, *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 2462–2464.
144. J. C. Calabrese; D. B. Jordan; A. Boodhoo; S. Sariaslani; T. Vannelli, *Biochemistry* **2004**, *43*, 11403–11416.
145. L. Wang; N. A. Gamez; C. N. Sarkissian; M. Straub; M. G. Patch; G. W. Han; S. Striepeke; P. Fitzpatrick; C. R. Scriver; R. C. Stevens, *Mol. Genet. Metab.* **2005**, *86*, 134–140.
146. H. Ritter; G. E. Schulz, *Plant Cell* **2004**, *16*, 3426–3436.
147. J. Rétey, *Naturwissenschaften* **1996**, *83*, 439–447.
148. L. Poppe, *Curr. Opin. Chem. Biol.* **2001**, *5*, 512–524.
149. A. Gloge; B. Langer; L. Poppe; J. Rétey, *Arch. Biochem. Biophys.* **1998**, *359*, 1–7.
150. T. Furuta; H. Takahashi; Y. Kasuya, *J. Am. Chem. Soc.* **1990**, *112*, 3633–3636.
151. A. Lewandowicz; J. Jemielity; M. Kańska; J. Zoń; P. Paneth, *Arch. Biochem. Biophys.* **1999**, *370*, 216–221.
152. M. Rettig; A. Sigrist; J. Rétey, *Helv. Chim. Acta* **2000**, *83*, 2246–2265.
153. S. Pilbák; A. Tomín; J. Rétey; L. Poppe, *FEBS J.* **2006**, *273*, 1004–1019.
154. L. Xiang; B. S. Moore, *J. Bacteriol.* **2005**, *187*, 4286–4289.
155. J. S. Williams; M. Thomas; D. J. Clarke, *Microbiology* **2005**, *151*, 2543–2550.
156. A. M. Hill; B. L. Thompson; J. P. Harris; R. Segret, *J. Chem. Soc. Chem. Commun.* **2003**, 1358–1359.
157. A. V. Emes; L. C. Vining, *Can. J. Biochem.* **1970**, *48*, 613–622.
158. K. T. Watts; P. C. Lee; C. Schmidt-Dannert, *ChemBioChem* **2004**, *5*, 500–507.
159. J. A. Kyndt; F. Vanrobaeys; J. C. Fitch; B. V. Devreese; T. E. Meyer; M. A. Cusanovich; J. J. Van Beeumen, *Biochemistry* **2003**, *42*, 965–970.
160. M. C. Moffitt; G. V. Louie; M. E. Bowman; J. Pence; J. P. Noel; B. S. Moore, *Biochemistry* **2007**, *46*, 1004–1012.
161. J. Stöckigt; M. H. Zenk, *FEBS Lett.* **1974**, *42*, 131–134.
162. M. J. C. Rhodes; L. S. C. Woollorton, *Phytochemistry* **1976**, *15*, 947–951.
163. B. Ulbrich; M. H. Zenk, *Phytochemistry* **1980**, *19*, 1625–1629.
164. L. Hoffmann; S. Maury; F. Martz; P. Geoffroy; M. Legrand, *J. Biol. Chem.* **2003**, *278*, 95–103.
165. H. G. Teutsch; M. P. Hasenfratz; A. Lesot; C. Stoltz; J.-M. Garnier; J.-M. Jeltsch; F. Durst; D. Werck-Reichhart, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4102–4106.
166. M. Mizutani; E. Ward; J. DiMaio; D. Ohta; J. Ryals; R. Sato, *Biochem. Biophys. Res. Commun.* **1993**, *190*, 875–880.
167. D. R. Gang; T. Beuerle; P. Ullmann; D. Werck-Reichhart; E. Pichersky, *Plant Physiol.* **2002**, *130*, 1536–1544.
168. L. Hoffmann; S. Besseau; P. Geoffroy; C. Ritzenthaler; D. Meyer; C. Lapierre; B. Pollet; M. Legrand, *Plant Cell* **2004**, *16*, 1446–1465.
169. G. Shadle; F. Chen; M. S. S. Reddy; L. Jackson; J. Nakashima; R. A. Dixon, *Phytochemistry* **2007**, *68*, 1521–1529.
170. X. Ma; J. Koepke; S. Panjikar; G. Fritsch; J. Stöckigt, *J. Biol. Chem.* **2005**, *280*, 13576–13583.
171. J. Buglino; K. C. Onwueme; J. A. Ferreras; L. E. N. Quadri; C. D. Lima, *J. Biol. Chem.* **2004**, *279*, 30634–30642.
172. T. A. Keating; C. G. Marshall; C. T. Walsh; A. E. Keating, *Nat. Struct. Biol.* **2002**, *9*, 522–526.
173. L. Govindasamy; B. Pedersen; W. Lian; T. Kukar; Y. Gu; S. Jin; M. Agbandje-McKenna; D. Wu; R. McKenna, *J. Struct. Biol.* **2004**, *148*, 226–235.
174. G. Jogl; L. Tong, *Cell* **2003**, *112*, 113–122.
175. D. W. Russell; E. E. Conn, *Arch. Biochem. Biophys.* **1967**, *122*, 256–258.
176. W. Heller; T. Kühnl, *Arch. Biochem. Biophys.* **1985**, *241*, 453–460.

177. T. Kühnl; U. Koch; W. Heller; E. Wellmann, *Arch. Biochem. Biophys.* **1987**, *258*, 226–232.
178. C. Grand, *FEBS Lett.* **1984**, *169*, 7–11.
179. I. Benveniste; B. Gabriac; F. Durst, *Biochem. J.* **1986**, *235*, 365–373.
180. M. Mizutani; D. Ohta, *Plant Physiol.* **1998**, *116*, 357–367.
181. D. K. Ro; J. Ehlting; C. J. Douglas, *Plant Physiol.* **2002**, *130*, 1837–1851.
182. M. Klingenberg, *Arch. Biochem. Biophys.* **1958**, *75*, 376–386.
183. T. Omura; R. Sato, *J. Biol. Chem.* **1964**, *239*, 2370–2378.
184. B. S. J. Winkel, *Annu. Rev. Plant Biol.* **2004**, *55*, 85–107.
185. B. Gabriac; D. Werck-Reichhart; H. Teutsch; F. Durst, *Arch. Biochem. Biophys.* **1991**, *288*, 302–309.
186. P. Urban; C. Mignotte; M. Kazmaier; F. Delorme; D. Pompon, *J. Biol. Chem.* **1997**, *272*, 19176–19186.
187. M. A. Pierrel; Y. Batard; M. Kazmaier; C. Mignotte-Vieux; F. Durst; D. Werck-Reichhart, *Eur. J. Biochem.* **1994**, *224*, 835–844.
188. H. Chen; H. Jiang; J. A. Morgan, *Phytochemistry* **2007**, *68*, 306–311.
189. M. Schalk; Y. Batard; A. Seyer; S. Nedelkina; F. Durst; D. Werck-Reichhart, *Biochemistry* **1997**, *36*, 15253–15261.
190. J. M. Boniwell; V. S. Butt, *Z. Naturforsch. C* **1986**, *41*, 56–60.
191. M. Kojima; W. Takeuchi, *J. Biochem.* **1989**, *105*, 265–270.
192. H. A. Stafford; S. Dresler, *Plant Physiol.* **1972**, *49*, 590–595.
193. P. F. T. Vaughan; V. S. Butt, *Biochem. J.* **1969**, *113*, 109–115.
194. P. F. T. Vaughan; V. S. Butt, *Biochem. J.* **1970**, *119*, 89–94.
195. V. S. Butt; C. J. Lamb, Oxygenases and the Metabolism of Plant Products. In *The Biochemistry of Plants. A Comprehensive Treatise*, P. K. Stumpf, E. E. Conn, Eds. in Chief, Vol. 7: *Secondary Plant Products*; E. E. Conn Ed. Academic Press: New York, 1981; pp 627–665.
196. D. J. Bartlett; J. E. Poulton; V. S. Butt, *FEBS Lett.* **1972**, *23*, 265–267.
197. R. E. Kneusel; U. Matern; K. Nicolay, *Arch. Biochem. Biophys.* **1989**, *269*, 455–462.
198. J. Kamsteeg; J. van Brederode; P. M. Verschuren; G. van Nigtevecht, *Z. Pflanzenphysiol.* **1981**, *102*, 435–442.
199. Z.-X. Wang; S.-M. Li; R. Löscher; L. Heide, *Arch. Biochem. Biophys.* **1997**, *347*, 249–255.
200. G. Schoch; S. Goepfert; M. Morant; A. Hehn; D. Meyer; P. Ullmann; D. Werck-Reichhart, *J. Biol. Chem.* **2001**, *276*, 36566–36574.
201. V. Mahesh; R. Million-Rousseau; P. Ullmann; N. Chabrilange; J. Bustamante; L. Mondolot; M. Morant; M. Noirot; S. Hamon; A. de Kochko; D. Werck-Reichhart; C. Campa, *Plant Mol. Biol.* **2007**, *64*, 145–159.
202. C. C. S. Chapple; T. Vogt; B. E. Ellis; C. R. Somerville, *Plant Cell* **1992**, *4*, 1413–1424.
203. K. Meyer; J. C. Cusumano; C. Somerville; C. C. S. Chapple, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6869–6874.
204. A. M. Anterola; N. G. Lewis, *Phytochemistry* **2002**, *61*, 221–294.
205. K. Fukushima; N. Matsui; S. Taguchi; S. Yasuda, On the Roles of Monolignol Glucosides and Monolignol Pathway (Coniferyl Alcohol to 5-Hydroxyconiferyl Alcohol to Sinapyl Alcohol) in Lignin Biosynthesis of *Magnolia kobus*, *Proceedings of the 9th International Symposium on Wood and Pulping Chemistry*, Montreal, Canada, **1997**; Vol. 2, pp 28.1–28.4.
206. F. Chen; S. Yasuda; K. Fukushima, *Planta* **1999**, *207*, 597–603.
207. H. Ohashi; E. Yamamoto; N. G. Lewis; G. H. N. Towers, *Phytochemistry* **1987**, *26*, 1915–1916.
208. N. Matsui; F. Chen; S. Yasuda; K. Fukushima, *Planta* **2000**, *210*, 831–835.
209. J. M. Humphreys; M. R. Hemm; C. Chapple, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10045–10050.
210. K. Osakabe; C. C. Tsao; L. Li; J. L. Popko; T. Umezawa; D. T. Carraway; R. H. Smeltzer; C. P. Joshi; V. L. Chiang, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8955–8960.
211. J.-K. Weng; X. Li; J. Stout; C. Chapple, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7887–7892.
212. D. Nelson, <http://drnelson.utmem.edu>, accessed July 17, 2009.
213. M. Otyepka; J. Skopalik; E. Anzenbacherová; P. Anzenbacher, *Biochim. Biophys. Acta* **2007**, *1770*, 376–389.
214. T. L. Poulos; R. Raag, *FASEB J.* **1992**, *6*, 674–679.
215. R. C. Wade; P. J. Winn; I. Schlichting; Sudarko, *J. Inorg. Biochem.* **2004**, *98*, 1175–1182.
216. Y. Zhao; J. R. Halpert, *Biochim. Biophys. Acta* **2007**, *1770*, 402–412.
217. M. D. Segall; M. C. Payne; W. Ellis; G. T. Tucker; N. Boyes, *Chem. Res. Toxicol.* **1998**, *11*, 962–966.
218. B. W. Griffin; J. A. Peterson; R. W. Estabrook, Cytochrome P450: Biophysical Properties and Catalytic Function. In *Porphyrins*; D. Dolphin, Ed.; Academic Press: New York, 1979; Vol. 7, pp 333–375.
219. J. T. Groves, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3569–3574.
220. P. Hlavica, *Eur. J. Biochem.* **2004**, *271*, 4335–4360.
221. M. Newcomb; D. Aebisher; R. Shen; R. E. Chandrasena; P. F. Hollenberg; M. J. Coon, *J. Am. Chem. Soc.* **2003**, *125*, 6064–6065.
222. P. R. Ortiz de Montellano; J. J. De Voss, *Nat. Prod. Rep.* **2002**, *19*, 477–493.
223. I. Schlichting; J. Berendzen; K. Chu; A. M. Stock; S. A. Maves; D. E. Benson; R. M. Sweet; D. Ringe; G. A. Petsko; S. G. Sligar, *Science* **2000**, *287*, 1615–1622.
224. I. G. Denisov; T. M. Makris; S. G. Sligar; I. Schlichting, *Chem. Rev.* **2005**, *105*, 2253–2277.
225. R. Bernhardt, *Rev. Physiol. Biochem. Pharmacol.* **1996**, *127*, 137–221.
226. D.-S. Lee; P. Nioche; M. Hamberg; C. S. Raman, *Nature* **2008**, *455*, 363–370.
227. L. Li; Z. Chang; Z. Pan; Z.-Q. Fu; X. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13883–13888.
228. S. E. Graham-Lorence; J. A. Peterson, *Methods Enzymol.* **1996**, *272*, 315–326.
229. V. F. Kalb; J. C. Loper, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7221–7225.
230. O. Gotoh, *J. Biol. Chem.* **1992**, *267*, 83–90.
231. C. D. Stout, *Structure* **2004**, *12*, 1921–1922.
232. E. E. Scott; M. A. White; Y. A. He; E. F. Johnson; C. D. Stout; J. R. Halpert, *J. Biol. Chem.* **2004**, *279*, 27294–27301.
233. M. Schalk; S. Nedelkina; G. Schoch; Y. Batard; D. Werck-Reichhart, *Biochemistry* **1999**, *38*, 6093–6103.
234. G. A. Schoch; R. Attias; M. Belghazi; P. M. Dansette; D. Werck-Reichhart, *Plant Physiol.* **2003**, *133*, 1198–1208.
235. G. A. Schoch; R. Attias; M. Le Ret; D. Werck-Reichhart, *Eur. J. Biochem.* **2003**, *270*, 3684–3695.
236. G. Truan; C. Cullin; P. Reisdorf; P. Urban; D. Pompon, *Gene* **1993**, *125*, 49–55.

237. K. G. Ravichandran; S. S. Boddupalli; C. A. Hasemann; J. A. Peterson; J. Deisenhofer, *Science* **1993**, 261, 731–736.
238. R. Raag; T. L. Poulos, *Biochemistry* **1989**, 28, 7586–7592.
239. C. A. Hasemann; K. G. Ravichandran; J. A. Peterson; J. Deisenhofer, *J. Mol. Biol.* **1994**, 236, 1169–1185.
240. J. R. Cupp-Vickery; T. L. Poulos, *Nat. Struct. Biol.* **1995**, 2, 144–153.
241. M. A. Costa; D. L. Bedgar; S. G. A. Moinuddin; K.-W. Kim; C. L. Cardenas; F. C. Cochrane; J. M. Shockey; G. L. Helms; Y. Amakura; H. Takahashi; J. K. Millhollan; L. B. Davin; J. Browse; N. G. Lewis, *Phytochemistry* **2005**, 66, 2072–2091.
242. P. Berg, *J. Biol. Chem.* **1956**, 222, 991–1013.
243. E. Conti; N. P. Franks; P. Brick, *Structure* **1996**, 4, 287–298.
244. E. Conti; T. Stachelhaus; M. A. Marahiel; P. Brick, *EMBO J.* **1997**, 16, 4174–4183.
245. A. M. Gulick; V. J. Starai; A. R. Horswill; K. M. Homick; J. C. Escalante-Semerena, *Biochemistry* **2003**, 42, 2866–2873.
246. A. S. Reger; J. M. Carney; A. M. Gulick, *Biochemistry* **2007**, 46, 6536–6546.
247. A. M. Gulick; X. Lu; D. Dunaway-Mariano, *Biochemistry* **2004**, 43, 8670–8679.
248. A. S. Reger; R. Wu; D. Dunaway-Mariano; A. M. Gulick, *Biochemistry* **2008**, 47, 8016–8025.
249. R. Wu; J. Cao; X. Lu; A. S. Reger; A. M. Gulick; D. Dunaway-Mariano, *Biochemistry* **2008**, 47, 8026–8039.
250. M. Becker-André; P. Schulze-Lefert; K. Hahlbrock, *J. Biol. Chem.* **1991**, 266, 8551–8559.
251. H.-P. Stuitable; D. Büttner; J. Ehling; K. Hahlbrock; E. Kombrink, *FEBS Lett.* **2000**, 467, 117–122.
252. L. W. Hamoen; H. Eshuis; J. Jongbloed; G. Venema; D. van Sinderen, *Mol. Microbiol.* **1995**, 15, 55–63.
253. B. R. Branchini; M. H. Murtiashaw; R. A. Magyar; S. M. Anderson, *Biochemistry* **2000**, 39, 5433–5440.
254. A. R. Horswill; J. C. Escalante-Semerena, *Biochemistry* **2002**, 41, 2379–2387.
255. A. Bairoch, *Nucleic Acids Res.* **19**, 2241–2245 (1991).
256. M. Fulda; E. Heinz; F. P. Wolter, *Mol. Gen. Genet.* **1994**, 242, 241–249.
257. R. Dieckmann; M. Pavela-Vrancic; E. Pfeifer; H. von Döhren; H. Kleinkauf, *Eur. J. Biochem.* **1997**, 247, 1074–1082.
258. M. Gocht; M. A. Marahiel, *J. Bacteriol.* **1994**, 176, 2654–2662.
259. K. Schneider; K. Hövel; K. Witzel; B. Hamberger; D. Schomburg; E. Kombrink; H.-P. Stuitable, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 8601–8606.
260. M. Shimada, *Wood Res.* **1972**, 53, 19–65.
261. M. Shimada; H. Fushiki; T. Higuchi, *Mokuzai Gakkaishi* **1973**, 19, 13–21.
262. U. N. Dwivedi; W. H. Campbell; J. Yu; R. S. S. Datla; R. C. Bugos; V. L. Chiang; G. K. Podila, *Plant Mol. Biol.* **1994**, 26, 61–71.
263. W. Ni; N. L. Paiva; R. A. Dixon, *Transgenic Res.* **1994**, 3, 120–126.
264. C. Hermann; M. Legrand; P. Geoffroy; B. Fritig, *Arch. Biochem. Biophys.* **1987**, 253, 367–376.
265. R. Atanassova; N. Favet; F. Martz; B. Chabbert; M. -T. Tollier; B. Monties; B. Fritig; M. Legrand, *Plant J.* **1995**, 8, 465–477.
266. A.-E. Pakusch; R. E. Kneussel; U. Matern, *Arch. Biochem. Biophys.* **1989**, 271, 488–494.
267. T. Kühnl; U. Koch; W. Heller; E. Wellmann, *Plant Sci.* **1989**, 60, 21–25.
268. A. E. Pakusch; U. Matern; E. Schiltz, *Plant Physiol.* **1991**, 95, 137–143.
269. D. Schmitt; A.-E. Pakusch; U. Matern, *J. Biol. Chem.* **1991**, 266, 17416–17423.
270. Z. -H. Ye; R. E. Kneusel; U. Matern; J. E. Varner, *Plant Cell* **1994**, 6, 1427–1439.
271. M. Salminen; K. Lundström; C. Tilgmann; R. Savolainen; N. Kalkkinen; I. Ulmanen, *Gene* **1990**, 93, 241–247.
272. J. Vidgren; C. Tilgmann; K. Lundström; A. Liljas, *Proteins* **1991**, 11, 233–236.
273. J. Vidgren; L. A. Svensson; A. Liljas, *Nature* **1994**, 368, 354–358.
274. E. Tsujii; K. Okazaki; M. Isaji; K. Takeda, *J. Struct. Biol.* **2009**, 165, 133–139.
275. H. C. Gulberg; C. A. Marsden, *Pharmacol. Rev.* **1975**, 27, 135–206.
276. C. H. Gray; J. K. Coward; K. B. Schowen; R. L. Schowen, *J. Am. Chem. Soc.* **1979**, 101, 4351–4358.
277. M. F. Hegazi; R. T. Borcharid; R. L. Schowen, *J. Am. Chem. Soc.* **1979**, 101, 4359–4365.
278. J. Rodgers; D. A. Femec; R. L. Schowen, *J. Am. Chem. Soc.* **1982**, 104, 3263–3268.
279. K. Inoue; V. J. H. Sewalt; G. M. Ballance; W. Ni; C. Stürzer; R. A. Dixon, *Plant Physiol.* **1998**, 117, 761–770.
280. J.-L. Ferrer; C. Zubieta; R. A. Dixon; J. P. Noel, *Plant Physiol.* **2005**, 137, 1009–1017.
281. Y.-J. Zheng; T. C. Bruce, *J. Am. Chem. Soc.* **1997**, 119, 8137–8145.
282. C. Zubieta; P. Kota; J.-L. Ferrer; R. A. Dixon; J. P. Noel, *Plant Cell* **2002**, 14, 1265–1277.
283. G. G. Gross; J. Stöckigt; R. L. Mansell; M. H. Zenk, *FEBS Lett.* **1973**, 31, 283–286.
284. H. Wengenmayer; J. Ebel; H. Grisebach, *Eur. J. Biochem.* **1976**, 65, 529–536.
285. T. Lüderitz; H. Grisebach, *Eur. J. Biochem.* **1981**, 119, 115–124.
286. T. Lüderitz; G. Schatz; H. Grisebach, *Eur. J. Biochem.* **1982**, 123, 583–586.
287. F. Sarni; C. Grand; A. M. Boudet, *Eur. J. Biochem.* **1984**, 139, 259–265.
288. E. Lacombe; S. Hawkins; J. Van Doorselaere; J. Piquemal; D. Goffner; O. Poeydomenge; A.-M. Boudet; J. Grima-Pettenati, *Plant J.* **1997**, 11, 429–441.
289. A. M. Patten; C. L. Cardenas; F. C. Cochrane; D. D. Laskar; D. L. Bedgar; L. B. Davin; N. G. Lewis, *Phytochemistry* **2005**, 66, 2092–2107.
290. S. Kamitori; A. Iguchi; A. Ohtaki; M. Yamada; K. Kita, *J. Mol. Biol.* **2005**, 352, 551–558.
291. Q.-H. Ma; B. Tian, *Biol. Chem.* **2005**, 386, 553–560.
292. C. M. Lawrence; V. W. Rodwell; C. V. Stauffacher, *Science* **1995**, 268, 1758–1762.
293. L. Taberner; D. A. Bochar; V. W. Rodwell; C. V. Stauffacher, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 7167–7171.
294. G. G. Gross; W. Kreiten, *FEBS Lett.* **1975**, 54, 259–262.
295. R. L. Mansell; G. G. Gross; J. Stöckigt; H. Franke; M. H. Zenk, *Phytochemistry* **1974**, 13, 2427–2435.
296. M. H. Walter; J. Grima-Pettenati; C. Grand; A. M. Boudet; C. J. Lamb, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, 85, 5546–5550.
297. M. H. Walter; J. Grima-Pettenati; C. Grand; A. M. Boudet; C. J. Lamb, *Plant Mol. Biol.* **1990**, 15, 525–526.
298. M. H. Walter; J. Grima-Pettenati; C. Feuillet, *Eur. J. Biochem.* **1994**, 224, 999–1009.
299. M. E. Knight; C. Halpin; W. Schuch, *Plant Mol. Biol.* **1992**, 19, 793–801.

300. S.-J. Kim; M.-R. Kim; D. L. Bedgar; S. G. A. Moinuddin; C. L. Cardenas; L. B. Davin; C.-H. Kang; N. G. Lewis, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1455–1460.
301. R. Pietruszko, *Adv. Exp. Med. Biol.* **1975**, *56*, 1–31.
302. H. Eklund; B. Nordström; E. Zeppezauer; G. Söderlund; I. Ohlsson; T. Boiwe; B.-O. Söderberg; O. Tapia; C.-I. Brändén; Å. Åkeson, *J. Mol. Biol.* **1976**, *102*, 27–59.
303. E. S. Cedergren-Zeppezauer; I. Andersson; S. Ottonello; E. Bignetti, *Biochemistry* **1985**, *24*, 4000–4010.
304. M. Klimacek; H. Hellmer; B. Nidetzky, *Biochem. J.* **2007**, *404*, 421–429.
305. G. Pettersson; J. P. Klinman, *CRC Crit. Rev. Biochem. Mol. Biol.* **1986**, *21*, 349–389.
306. R. T. Dworschack; B. V. Plapp, *Biochemistry* **1977**, *16*, 2716–2725.
307. M. W. Makinen; W. Maret; M. B. Yim, *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 2584–2588.
308. M. W. Makinen; M. B. Yim, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 6221–6225.
309. A. J. Sytkowski; B. L. Vallee, *Biochemistry* **1978**, *17*, 2850–2857.
310. O. Kleifeld; A. Frenkel; J. M. L. Martin; I. Sagi, *Nat. Struct. Biol.* **2003**, *10*, 98–103.
311. B. Youn; R. Camacho; S. G. A. Moinuddin; C. Lee; L. B. Davin; N. G. Lewis; C. Kang, *Org. Biomol. Chem.* **2006**, *4*, 1687–1697.
312. E. Valencia; C. Larroy; W. F. Ochoa; X. Parés; I. Fita; J. A. Biosca, *J. Mol. Biol.* **2004**, *341*, 1049–1062.
313. E. K. Bomati; J. P. Noel, *Plant Cell* **2005**, *17*, 1598–1611.
314. L. Esposito; I. Bruno; F. Sica; C. A. Raia; A. Giordano; M. Rossi; L. Mazzarella; A. Zagari, *Biochemistry* **2003**, *42*, 14397–14407.
315. R. Meijers; H.-W. Adolph; Z. Dauter; K. S. Wilson; V. S. Lamzin; E. S. Cedergren-Zeppezauer, *Biochemistry* **2007**, *46*, 5446–5454.
316. P. J. Baker; K. L. Britton; M. Fisher; J. Esclapez; C. Pire; M. J. Bonete; J. Ferrer; D. W. Rice, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 779–784.

Biographical Sketches



Dhrubojyoti D. Laskar received his B.Sc. Honors in chemistry from Assam University in 1995, his M.Sc. in organic chemistry from Gauhati University, and his Ph.D. in organic chemistry from Dibrugarh University (with Dr. Dipak Prajapati, RRL, Jorhat) in 2002. He was also a recipient of the Senior Research Fellowship awarded by the Council of Scientific and Industrial Research (CSIR), India (2000). He joined Professor Lewis' group in 2003 at Washington State University as a postdoctoral fellow; since then, his research has been specifically focused upon the phenylpropanoid pathway, lignin biosynthesis, lignin macromolecular configuration, and effects of modulating lignin formation/deposition.



Oliver Corea completed his B.Sc. in genetics in 2004 at the University of Western Ontario, London, ON, Canada, followed by a M.Sc. in cellular and molecular biology with Professor Susanne Kohalmi in 2006 at the same institution. Since 2007, he has been pursuing his Ph.D. (molecular plant sciences) studies under the supervision of Professor Norman G. Lewis at Washington State University. His Ph.D. research focuses on the biochemical characterization of the arogenate dehydratase (ADT) gene family in *Arabidopsis* and the analysis of ADT knockout lines. In 2008, he was awarded best graduate student poster at the Phytochemical Society of North America annual meeting.



Ann M. Patten received her B.Sc. in microbiology (Pennsylvania State University, 1988), an M.S. in forest resources (University of Idaho, 1999), and a Ph.D. in molecular plant sciences (Washington State University, 2007) under the supervision of Professor Norman G. Lewis. Her Ph.D. studies were supported by Helen and Loyal H. Davis Fellowship and John and Edith McDougall Fellowship. Currently, Dr. Patten is carrying out postdoctoral studies in the Lewis' group; her research focuses on plant cell wall formation/lignification using various types of light and electron microscopic approaches.



Professor ChulHee Kang received his B.Sc. (physics and microbiology, 1980) and M.S. (microbiology, 1982) from Seoul National University and his Ph.D. (biophysics, 1987) from the University of California, Berkeley. He then pursued his postdoctoral studies at the Lawrence Berkeley National Laboratory (1988–89) with Dr. Sun-Hou Kim and at the Massachusetts Institute of Technology (1989–92) with Dr. Alexander Rich. After 2 years as a research scientist in the Department of Biology at MIT, he became an assistant professor (1994–99), associate professor (1999–2002), and the Edward Meier Distinguished Professor of Biophysics (2002–present) in the School of Molecular Biosciences at Washington State University.



Laurence B. Davin received her B.Sc. (1983) and Ph.D. (1987) degrees in plant biochemistry/physiology from the Université Paul Sabatier (Toulouse, France). Later, Dr. Davin studied glucosinolate biogenesis with Ted Underhill in the National Research Council of Canada's Plant Biotechnology Institute as a postdoctoral fellow. Since 1989, her research has been focused on lignan, lignin, and allyl/propenyl phenol biosynthesis at both the Virginia Polytechnic Institute and State University (Blacksburg, VA) and the Institute of Biological Chemistry (Washington State University). Dr. Davin is a member of the Editorial Board for *Recent Advances in Phytochemistry*.



Professor Norman G. Lewis completed his B.Sc. at the University of Strathclyde (Honors, chemistry, 1973) and his Ph.D. at the University of British Columbia (chemistry, 1977). He next pursued postdoctoral studies at Cambridge University (chemistry, 1978–80) with Professor Sir Alan R. Battersby. Before joining Washington State University in 1990, he held various research positions at ICI, the Pulp and Paper Research Institute of Canada, the National Research Council of Canada, and Virginia Tech. His research interests mainly involve lignan, lignin, and allyl/propenyl phenol biosynthesis. He serves on numerous scientific advisory committees, grant panels, and editorial boards, including as Regional Editor (*Phytochemistry*), Monitoring Editor (*Plant Physiology*), and as Executive Editor of *Advances in Plant Biochemistry and Molecular Biology*.

6.19 Nucleoside Analogues

Darrell R. Davis, The University of Utah, Salt Lake City, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.19.1 | Introduction | 663 |
| 6.19.2 | Stereoelectronic Factors that Affect Nucleic Acid Stability | 665 |
| 6.19.2.1 | The Relationship Between Sugar Conformation and Duplex Stability | 665 |
| 6.19.2.2 | 2'-OMe Modification Effects on RNA Duplex Stability | 665 |
| 6.19.2.3 | Base Modification Influences Stacking and RNA Stability | 666 |
| 6.19.3 | 2'-O-Methyl and 2'-Fluoro Modification | 666 |
| 6.19.3.1 | 2'-O-Methyl and 2'-Fluoro Modifications of RNA Duplexes | 666 |
| 6.19.3.2 | 2'-O-Methyl Modifications in Thermophilic tRNA | 666 |
| 6.19.3.3 | 2'-O-Methyl Modification in Ribosomal RNA | 667 |
| 6.19.4 | Pseudouridine | 668 |
| 6.19.4.1 | Fundamental Properties of ψ Modification | 668 |
| 6.19.4.2 | ψ Effects in tRNA | 669 |
| 6.19.4.3 | ψ in Small Nuclear RNA | 669 |
| 6.19.4.4 | ψ in Helix 69 of 23S rRNA | 670 |
| 6.19.5 | Base-Modified Nucleosides in tRNA | 672 |
| 6.19.5.1 | Background | 672 |
| 6.19.5.2 | 2-Thio, 5-X-Modified Uridines – s ² U, mnm ⁵ s ² U, and mcm ⁵ s ² U | 672 |
| 6.19.5.3 | Nucleosides that Form Rare Tautomers – cmo ⁵ U and τ m ⁵ U | 674 |
| 6.19.5.4 | Modified Cytidines – Ac ⁴ C, f ⁵ C, Ac ⁴ Cm, and f ⁵ Cm | 675 |
| 6.19.5.5 | The Lysine-Modified Nucleoside, Lysidine | 676 |
| 6.19.5.6 | Purine Modifications – m ¹ G, i ⁶ A, t ⁶ A, yW, and Q | 677 |
| 6.19.6 | Summary and Future Prospects | 678 |
| References | | 679 |

6.19.1 Introduction

Natural RNA molecules are post-transcriptionally modified, there being at last count 96 chemically distinct modified nucleosides.¹ With a small number of exceptions, RNA modification occurs by enzymes acting directly on the parent A, U, G, or C nucleosides present in an RNA transcript.² For all RNA classes in bacteria, and for the tRNAs in all organisms, the modification reaction is performed by sequence and structurally specific enzymes. That is, a modification enzyme recognizes a target nucleoside within a specific sequence context. The enzyme may also require specific structural motifs within which the sequence must be located, for example, the tRNA L-shape. A modification could also be made in both a tRNA and a ribosomal RNA, but the sequence context would likely be very similar in the two RNAs. For rRNAs in eukaryotes, the process of RNA 2'-O-methylation (2'-OMe) and pseudouridylation (ψ) is quite different from that of tRNA modification. In these rRNAs, the modification is carried out using an RNA 'guide' in the form of a small nucleolar RNA (snoRNA) that has sequence complementarity for the rRNA target site, and specifies the site of modification using a generic modification enzyme acting on the snoRNA/rRNA complex.^{3,4} The generic nature of the snoRNA system explains in part why eukaryotic rRNAs are more extensively modified than bacterial RNAs; the snoRNA guided system uses fewer enzymes, and relies only on RNA sequence complementarity to identify the target. This is more economical than the essentially one enzyme–one modification system used for methylation and ψ of rRNA in bacteria (see Chapter 6.20).

rRNAs and snRNAs for the most part contain relatively simple modifications, primarily 2'-OMe modifications and ψ , along with a smattering of base methylations. One should not discount the profound effects that these modest chemical changes can have on function, but this chemical repertoire is limited compared to the

diversity of modifications found in tRNA. In addition to the proliferation of 2'-OMe nucleosides and ψ throughout tRNAs, the small (~75 nt on average) tRNAs are host to most of the chemical diversity among nucleoside analogues. Conserved tRNA modifications are clustered in the functionally important regions involved in tertiary structure (e.g., the s^4U8 , m^5U54 , $\psi55$, and D19 nucleosides), and in the anticodon region directly involved in modulating translational fidelity. **Figure 1** shows a generic cloverleaf for the secondary structure of eukaryotic tRNAs. The fact that the tRNA wobble position, 34, and the conserved purine 37 adjacent to the anticodon harbor the greatest diversity of modified nucleosides is apparent in the figure. In addition to the largest number of different modifications, by far the most chemically complex modified nucleosides are found at positions 34 and 37.

In this review, we begin the discussion with two modifications that are chemically simple and widespread, the 2'-OMe modification and ψ . Within the context of the 2-OMe discussion we have also included the nonnatural 2'-F modifications. A review of 2'-F modification highlights the stereochemical similarities to 2'-OMe, and shows why 2'-OMe and 2'-F are both becoming widely used in oligonucleotide therapeutics. More exotic modifications from the tRNA anticodon region are grouped according to the parent nucleosides from which they are derived, and discussed according to how the modifications affect fundamental properties

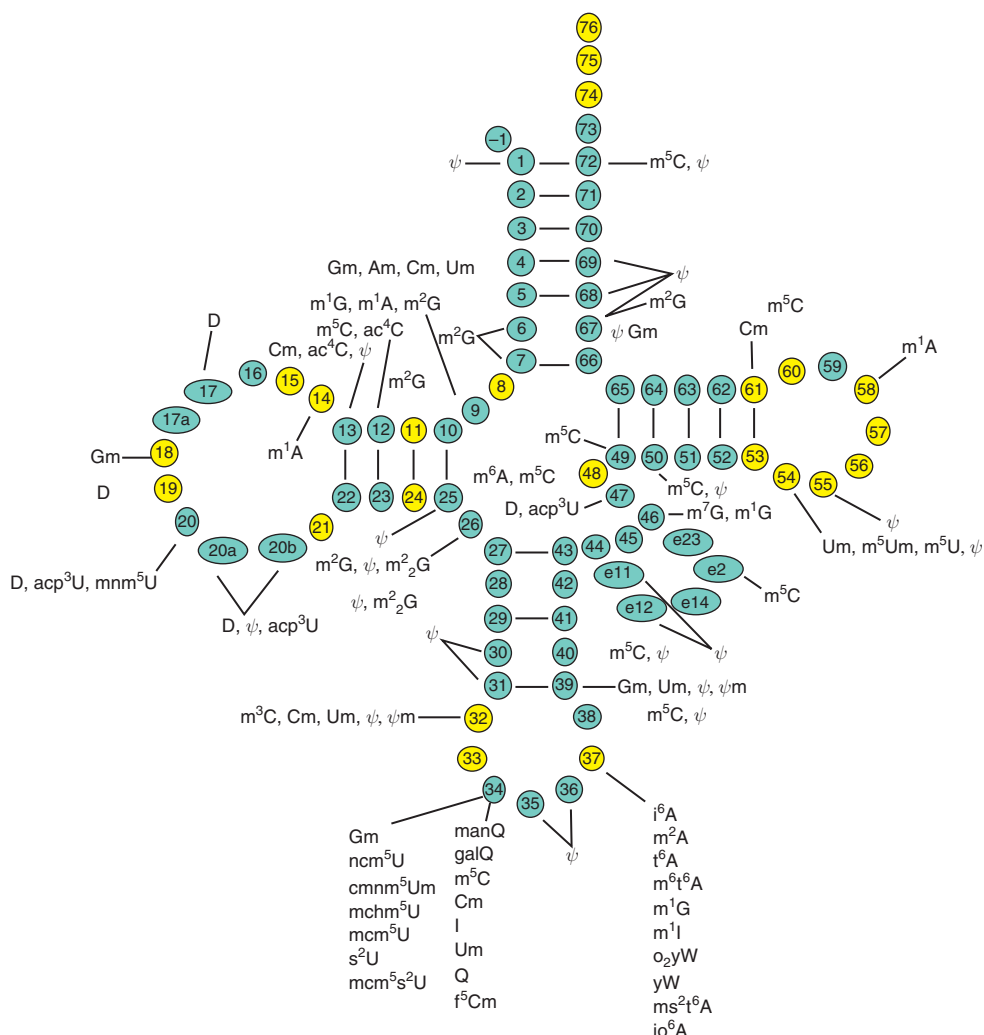


Figure 1 Summary of the identities and locations for modified nucleosides in eukaryotic tRNA.^{1,5} Conserved/invariant nucleosides are in gold and sites of common insertions are shown as ellipses.

such as nucleoside conformation and base stacking. This overview of a large and diverse field will hopefully provide an appreciation for the biological importance of nucleoside modification, and a glimpse of how much remains to be discovered.

6.19.2 Stereoelectronic Factors that Affect Nucleic Acid Stability

6.19.2.1 The Relationship Between Sugar Conformation and Duplex Stability

The thermodynamic stability of simple RNA duplexes, hairpins, and more complex RNA molecules up to the size of a tRNA can be measured by the midpoint of their thermal denaturation, the T_m .⁶ The melting temperature of both simple and complex RNA molecules is mainly a function of the stability of the component duplex regions, although single-stranded stacking and special loop elements such as tetraloops can contribute to stability.^{7,8} The major factor affecting RNA helical stability is the base pair composition, and the nearest-neighbor effect.⁹ Implicit in the nearest-neighbor approximation is that base stacking effects determine nucleic acid stability.¹⁰ Lee and Tinoco^{11,12} realized that the nucleoside sugar conformation had a direct effect on base stacking interactions. Small RNAs were seen to have an increase in the percentage of 3'-endo nucleoside conformation in direct relationship to the degree of stacking, and nucleotides in the middle of A-form RNA helices are 100% 3'-endo.

The cause and effect relationship between sugar conformation and base stacking–duplex stability suggested that nucleoside modifications that promote a 3'-endo sugar conformation at the nucleoside level would result in an increase in RNA duplex stability.^{11,13,14} To the extent that single-strand stacking contributes to thermodynamic stabilization of a folded RNA, 3'-endo-promoting modifications in single-stranded loops may also provide global stabilization. The classic examples of sugar modifications that stabilize RNA are the 2'-OMe and 2'-F modifications (Figure 2).

6.19.2.2 2'-OMe Modification Effects on RNA Duplex Stability

In nature, 2'-OMe modification is used to modulate RNA structure and stability. At the nucleoside level, it has been shown that the effect of 2'-OMe is to bias the conformational preference toward the 3'-endo conformation found in A-form helical structures.¹⁵ This conformational preference positions the base in the axial orientation and concomitantly increases RNA base stacking.^{16,17} Changing the 2'-sugar substituent alters the stereoelectronic properties of the nucleoside, and these effects can reinforce each other to favor the 3'-endo, axial conformation. The steric effect is easy to visualize, and can be demonstrated with a simple molecular model where the steric interaction between the 2' position and the base is minimized in the 3'-endo conformation. This effect is greater for pyrimidines where the 2-carbonyl clashes with a bulky 2' group if the sugar conformation is 2'-endo. Another important effect of nucleoside modification is exemplified by the nonnatural 2'-F modification where the strongly electron-withdrawing fluorine stabilizes the 3'-endo, axial conformation through the gauche effect, just as the 2'-OH of RNA promotes the 3'-endo conformation compared to deoxyribose.^{18–20}

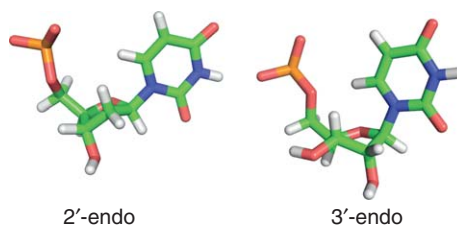


Figure 2 Sugar conformations for B-form DNA (2'-endo) and A-form RNA (3'-endo) shown for deoxyuridine and uridine, respectively. Notice that the distance between the 2'-OH and the O2 carbonyl of U is maximized by the 3'-endo sugar conformation.

6.19.2.3 Base Modification Influences Stacking and RNA Stability

Numerous base modifications also stabilize the 3'-endo sugar conformation, and may also improve base stacking by affecting dipole–dipole interactions between adjacent bases.^{15,21} Among the natural modifications, 2-thiouridine (s²U) is the champion modification for stabilizing RNA. This modification has effects on the nucleoside sugar conformation, increases base pairing strength, and has base stacking effects that all reinforce an A-form RNA structure. The more specific effects are steric interactions between the 2-thiocarbonyl and the 2'-hydroxyl that favor a 3'-endo sugar pucker, an increase in acidity for the imino proton making a stronger base pair, and polarization of the anomeric bond favoring an axial base orientation.^{21–24} 2-Thiouridine is a modification *tour de force* for such a simple chemical alteration. Other examples of base modifications that improve base stacking through polarization changes are the 5-methyl group of m⁵U and m⁵C,²⁵ and the formyl and acetyl modifications of f⁴C and ac⁴C. The latter two modifications appear to change the double bond polarization in the pyrimidine ring and may also promote the 3'-endo sugar conformation through the anomeric effect.^{26–30}

6.19.3 2'-O-Methyl and 2'-Fluoro Modification

6.19.3.1 2'-O-Methyl and 2'-Fluoro Modifications of RNA Duplexes

2'-OMe modification was first shown to stabilize RNA oligonucleotide duplexes by Ohtsuka and coworkers.¹³ In nature, a striking example of RNA duplex stabilization by 2'-OMe modification is the special three base pair duplex formed between the tRNA anticodon and the three base codon triplet. Uridine at the tRNA wobble position is almost always modified, and these modifications modulate the selectivity of the tRNA for either A or G pairing.³¹ It has been shown that promotion of the 3'-endo sugar conformation for the wobble nucleoside is effective at restricting pairing to A compared to G.^{15,32} This restriction of wobble pairing is typically accomplished with 2'-OMe modification, 2-thio modification, or some combination of these modifications along with additional base modification.

Applied to longer RNA duplexes, the increase in T_m upon 2'-OMe modification has seen wide application in antisense RNAs and more recently for siRNAs.^{33–36} The 2'-OMe modification provides stabilization of approximately 1.3 °C per modification while the 2'-F provides approximately 1.8 °C per modification. These modifications increase the potency of therapeutic antisense and siRNAs, and have the added benefit of imparting nuclease resistance to the oligonucleotide.³⁷ In the area of antisense oligonucleotides, it has been established that an important design feature is to maintain the 2'-endo conformation in the center of the oligonucleotide in order to support RNase H activity. This 2'-endo feature comes at the cost of antisense affinity for the RNA target, and is incompatible with 2'-OMe or 2'-F modification. However, the modifications can be placed in the 'wings' of the antisense oligonucleotide to increase affinity and nuclease resistance. The combination of a high-affinity flanking region and a central RNase H active region has been termed a 'gapmer'³⁸(see Chapter 2.19).

It appears that siRNAs are much more sensitive to RNA modification than antisense oligonucleotides, and only a few, simple chemical modifications are tolerated.³⁹ Since the publication by Chiu and Rana, there has been some progress in defining the limits to which sugar modifications can be incorporated into siRNAs. 2'-OMe and 2'-F can be used throughout the sense strand, while in the antisense strand there is some modulation of siRNA activity. Generally, the 3' end of the antisense strand can be extensively modified to increase stability and nuclease resistance, while the 5' end requires more care.^{33,36,40} Position-specific modification in the seed region has been used to reduce off-target effects,³⁴ suggesting also that 2'-OMe or 2'-F modification might have utility in discriminating between siRNA and micro RNA activities (Figure 3).

6.19.3.2 2'-O-Methyl Modifications in Thermophilic tRNA

The presence of modest levels of 2'-OMe modification in natural RNAs from mesophiles begged the question whether 2'-OMe modifications would be prevalent in the RNAs of thermophilic organisms. A study of the hyperthermophiles *Pyrococcus furiosus* and *Pyrodictum occultum* showed that the tRNA from these organisms had

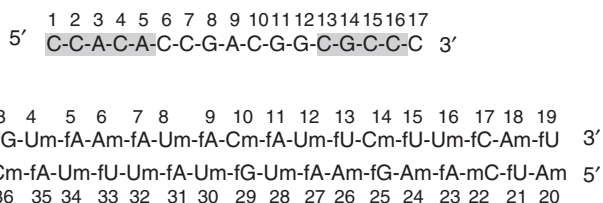


Figure 3 (Top) An antisense 2'-OMe oligonucleotide designed to use the 'gapmer' strategy.⁴¹ Shaded nucleotides in the 'wings' are 2'-OMe modified, the central region is DNA (2'-deoxy), and phosphorothioate backbone modification is used throughout. The 2'-OMe gapmer was approximately 70-fold more potent than a corresponding all-deoxy oligonucleotide. (Bottom) A modified siRNA duplex with alternating 2'-OMe and 2'-F modifications.³⁶ The modified siRNA was approximately 500-fold more potent than an unmodified siRNA.

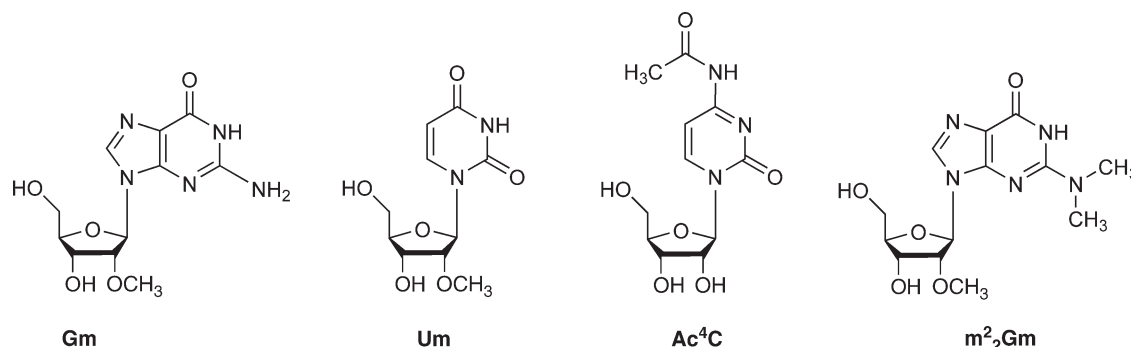


Figure 4 Representative modified nucleosides in thermophilic tRNAs. 2'-OMe G and 2'-OMe U, N4-acetyl C, and N2-dimethyl, 2'-OMe G.

T_m values of approximately 100 °C.⁴² It was perhaps not surprising that the tRNAs had such high T_m s since the optimal growth temperatures of these organisms is over 100 °C. However, studies of tRNA from other thermophilic organisms with optimal growth at 80 °C showed tRNA T_m values of only approximately 80 °C, while *Escherichia coli* tRNA has a quite similar T_m of 75 °C despite the modest optimal growth temperature of 37 °C.⁴³ McCloskey and coworkers showed that the % GC content of the hyperthermophilic tRNAs was over 80%, accounting for some of the observed stabilization but that this GC content was in fact a bit lower than for *Thermus thermophilus* tRNA^{Met} with a T_m of approximately 85 °C.⁴⁴ Furthermore, it was shown by Watanabe *et al.* that *T. thermophilus* tRNA was greatly stabilized by Mg²⁺ (35 °C increase) while the hyperthermophilic tRNA was only stabilized by about 15 °C by Mg²⁺. The mechanism of stabilization of the *P. furiosus* and *P. occultum* tRNAs was in fact due to a very high number of modified nucleosides, particularly Cm, Um, Gm, Am, Ac⁴C, Ac⁴Cm, s²C, s²U, and m₂Gm.⁴² The high percentage of 2'-OMe nucleosides, in particular, was proposed as a major factor in the stabilization above what could be obtained from increased GC content alone (Figure 4). The effects of thiolated pyrimidines and acetyl-C are discussed elsewhere in this report.

6.19.3.3 2'-O-Methyl Modification in Ribosomal RNA

2'-OMe modifications are found in all large and small subunit ribosomal RNAs investigated to date. Bacteria have very few total modifications compared to higher organisms with only 14 total modifications in the SSU RNA of *E. coli* and one 2'-OMe nucleoside in each of *E. coli* and *T. thermophilus* SSU RNA.⁴⁵ In higher organisms, yeast has 45 total modifications, 17 2'-OMe, and the human SSU rRNA has at least 40 2'-OMe nucleosides.^{3,46} While it seems obvious from this comparison that the extent of modification is not well correlated with growth temperature (and by inference RNA stability) when comparing bacteria and eukaryotic rRNA, studies of thermophilic archaea provide a different picture. *Sulfolobus solfataricus* is an archaeal hyperthermophile containing 38 total modifications in the SSU rRNA, with 22 2'-OMe nucleosides. As seen in studies of thermophilic tRNAs, it was found that the number of 2'-OMe nucleosides increased with increasing growth temperature.^{42,47}

Much remains to be learned about the role of nucleoside modification in ribosomal RNA, but the observation that modifications are clustered in the functionally important regions of the ribosome supports the hypothesis that modifications are used to fine-tune ribosomal function.³

6.19.4 Pseudouridine

6.19.4.1 Fundamental Properties of ψ Modification

Pseudouridine is the most common RNA modification, being found in tRNA,⁴⁸ snRNA,⁴⁹ and rRNA⁵⁰ throughout the three kingdoms of life.¹ Studies of tRNA function established that ψ modification in the anticodon affected translational fidelity, but a structural basis for this effect was not immediately obvious since uridine and pseudouridine could both pair with A. As with other RNA modifications,⁵¹ it was noticed some time ago that ψ modifications in rRNA tended to be clustered in the functionally important regions, notably the peptidyl transferase region of 23S RNA.⁵² Structural studies using nuclear magnetic resonance (NMR),⁵³ and an X-ray structure of tRNA^{Gln} complexed with the Gln aminoacyl synthetase,⁵⁴ established that ψ stably coordinates a water molecule between the base and the RNA backbone. This subtle structural property results in the promotion of an idealized A-form geometry around the ψ residue, and an increase in base stacking concomitant with RNA structural stabilization.⁵⁵ For RNAs with ψ modification in a base-paired region, the conversion of U to ψ results in a typical ΔG increase of approximately $0.5 \text{ kcal mol}^{-1}$, and a $3\text{--}5 \text{ }^\circ\text{C}$ increase in T_m per ψ modification.^{56,57} This effect is highly dependent on the structural environment, with ψ modifications in long duplexes providing only modest increases in thermodynamic stability,⁵⁸ while ψ in native structural contexts usually has a greater effect. In natural RNAs, ψ is typically found paired with A or G at the interface of stem and loop regions, in the loop regions proper, or in short helical regions that are A–U rich and presumably unstable in the absence of modification.⁵⁰

The stabilization of A-form RNA structure by ψ modification arises from several synergistic effects. The sequestration of a water molecule between the phosphate backbone and the ψ base orients the nucleotide such that the base is axial, causing the nucleotide to favor the 3'-endo sugar conformation. This 3'-endo nucleotide conformation for ψ is energetically favored despite the lack of an anomeric effect stabilizing the axial orientation in C-nucleosides.⁵² The importance of the water bridge for promoting the axial/3'-endo conformation is shown by the fact that at the nucleoside monomer level where there is no 5'-phosphate, ψ favors 2'-endo relative to uridine, yet even in short single-stranded RNA oligonucleotides, ψ dramatically promotes the 3'-endo conformation compared to unmodified U (Figure 5).⁵⁵

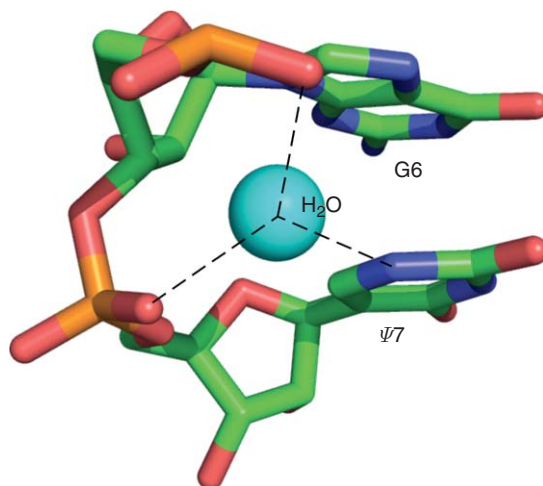


Figure 5 The crystal structure of a ψ within the U2 snRNA–branchpoint duplex showing a stable water molecule coordinated to two phosphates and to the ψ N1–H proton.⁵⁹ PDB ID:3CGR.

6.19.4.2 ψ Effects in tRNA

Pseudouridines are found throughout the tRNA cloverleaf, most notably in the eponymous T ψ C loop containing the highly conserved ψ 55 residue, and in the anticodon stem–loop domain. Modifications in the anticodon often have significant effects on tRNA function since the proximity to mRNA binding region can affect protein synthesis. In *E. coli* and *Salmonella typhimurium*, the *bisT*(TruA) gene is responsible for modifying positions 38/39/40 while in yeast the corresponding modification activity is due to Deg1/pus3.^{60–62} The tRNA anticodon stem–loop may also contain ψ modification at position 32, and any of the three anticodon positions. *HisT* mutants lacking pseudouridine have reduced growth rates, reduced protein elongation rates, and increased errors in translation.^{63–65} We have shown that the A31– ψ 39 base pair in tRNA is destabilized when the ψ is replaced with an unmodified U, that is, in a tRNA isolated from a *bisT* mutant strain.^{53,57} The A31– ψ 39 base pair in tRNA is at the interface between the anticodon stem and loop region, and the structural stabilization provided by ψ is crucial to maintaining the proper tRNA anticodon structure in tRNA^{Lys} and tRNA^{Phe}. Other *E. coli* tRNAs, such as tRNA^{His} have consecutive ψ 38/ ψ 39 residues, and in these tRNAs the stabilizing force is approximately additive, providing even greater stacking stabilization for bases on the 3' side of the tRNA anticodon.⁵⁶ An additional ψ modification formed by a separate pseudouridine synthase, RluA, is found at position 32 of many tRNAs, including *E. coli* tRNA^{Phe}.⁶⁶ Tworowska and Nikonowicz were able to introduce ψ 32 and ψ 39 into an isotopically labeled tRNA^{Phe} anticodon stem–loop to prepare a sample for heteronuclear NMR study. This powerful approach allowed them to show that ψ 32/ ψ 39 modification causes some stabilization of the tRNA structure, but that ψ modification without an additional modification at position 37 was insufficient to promote either a U-turn geometry or the bifurcated ψ 32–A38 base pair observed in native tRNA^{Phe} (Figure 6).⁶⁷

6.19.4.3 ψ in Small Nuclear RNA

In eukaryotes, the spliceosome is a large, dynamic ribonucleoprotein complex responsible for removing introns from mRNAs and generating mature, spliced mRNAs. The splicesomal RNAs are post-transcriptionally modified.⁶⁸ Here we will focus on one highly conserved snRNA, U2, to illustrate the effects of ψ modification on snRNA structure and function (Figure 7).

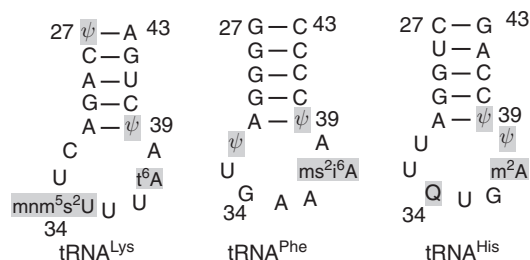


Figure 6 The anticodon stem–loop domains of *Escherichia coli* tRNA^{Lys}, tRNA^{Phe}, and tRNA^{His} contain ψ residues.

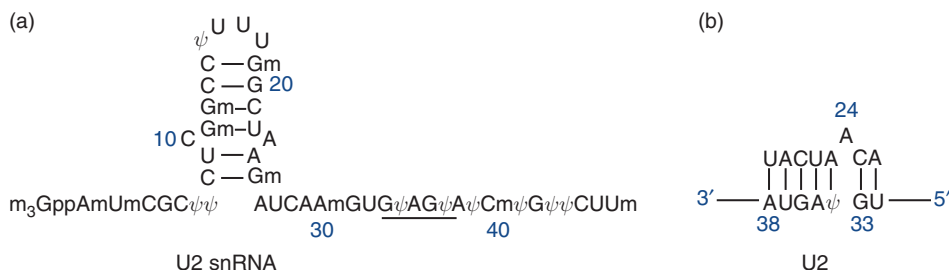


Figure 7 (a) The 5' end of U2 snRNA from vertebrates, containing extensive ψ and 2'-O-methyl modifications.^{68,69} The underlined region from 33–37 is the BSSR. (b) Secondary structure of the yeast U2 branch site showing A24 bulged due to A23– ψ 34 pairing. The numbering in (b) is from Greenbaum and coworkers.⁷⁰

U2 snRNA is a required component of the slicing machinery, and contains one of the most extensively modified regions in RNA. In vertebrates, the first 60 nucleotides contain 11 ψ s in addition to many 2'-OMe residues.⁶⁹ In yeast, the extent of modification is reduced, but ψ s are conserved in the branch site recognition region (BSRR), and these modifications have been shown to be necessary for spliceosome assembly and function.^{71,72} The mechanism of RNA splicing requires the availability of the 2' hydroxyl of a conserved adenosine at the splice site. Greenbaum and coworkers have demonstrated that ψ modification promotes a bulged conformation for A24 (yeast numbering), the branch site nucleophile. The corresponding U-containing RNA allows the As to be stacked within the helix.^{70,73} Their NMR experiments, along with modification-dependent changes detected by 2-aminopurine fluorescence, established that there was a clear difference in the structural environment of the branchpoint A between the ψ -modified and the unmodified RNAs. The NMR and fluorescence experiments suggested an A-bulge structure. The subsequent NMR solution structure supported a model where an A- ψ base pair caused the branchpoint A to be extruded from the helix.⁷³

In a recent crystallographic study, Lin and Kielkopf confirmed that ψ modification results in a structure for the branch site that in turn results in an extrahelical A, but that the selection of which two adenosines would be extrahelical depends on the sequence flanking the branch site.⁵⁹ This is consistent with the observation that either of the two consecutive adenosines across from the ψ can serve as the branch nucleophile.⁷⁴ The crystallographic study investigated three different snRNA-branchpoint duplexes in order to determine which of the two adenosines is used as the nucleophile, and they compared the sequence containing consecutive As with a branchpoint variant from higher eukaryotes where a G is paired with the ψ . The G- ψ -containing duplex was particularly interesting due to the ψ -containing wobble pair, and because of the fortuitous result that duplex BPS3 did not contain strong crystal contacts influencing the conformation of the extrahelical A. Analysis of the extrahelical A environment led Lin and Kielkopf to question one aspect of the NMR structure, that is whether the extrahelical A is positioned in the minor groove as proposed by Greenbaum and coworkers.^{59,73} The NMR structure could be problematic in that the 2'-OH is not readily accessible; however, one could imagine that interactions in the intact spliceosome could easily affect the minor structural change needed to completely swing the branchpoint A into position for splicing. The NMR and crystallography groups are in agreement regarding the importance of a ψ modification for extruding the branchpoint adenosine nucleoside.

6.19.4.4 ψ in Helix 69 of 23S rRNA

For researchers interested in the biological role of modified nucleosides in RNA, H69 in the large subunit RNA (also known as the 1915 loop of 23S RNA) has long been an intriguing subject.⁵⁰ Helix 69 is remarkable in that the loop contains three conserved ψ residues, while in human rRNA the stem has two additional ψ modifications. The seven nucleotide H69 loop has a conserved sequence for the non- ψ residues as well, and these specific nucleotides are known to be important for function.⁷⁵ Helix 69 in the bacterial large subunit (LSU) RNA also contains the only known example of the modified nucleoside $m^3\psi$.⁷⁶ In the human LSU RNA there is a regular ψ at that position, yet an adjacent 2'-OMe A residue may provide a key methyl group to the structure (Figure 8).⁵⁰

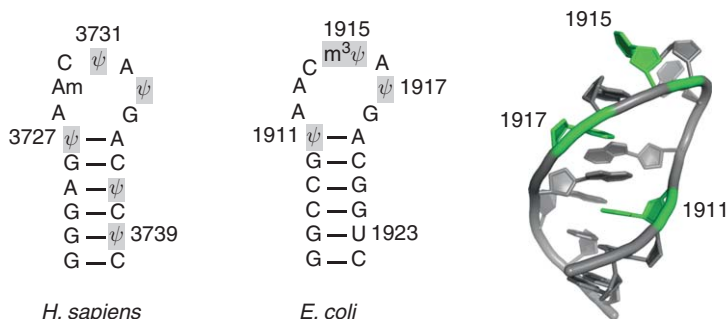


Figure 8 Helix 69 from *Homo sapiens* and *Escherichia coli* large subunit RNAs. (Right) The crystal structure of H69 from the *E. coli* 70S ribosome (ψ residues in green) suggests that ψ 1911 is unpaired in contrast to the model derived from NMR experiments. ψ 1917 may H-bond to N7 of A1912, providing an explanation for the stable imino proton seen by Chow and coworkers. The 3D structure was generated from PDB ID: 2AWB.⁷⁷

Helix 69 is postulated to be involved in several different interactions during the course of protein synthesis. If H69 participates in multiple functions, then conformational flexibility could be important for function, and structural interactions sensitive to pH changes near the physiological range could allow H69 more functional range. This speculation about structural plasticity in H69 is intriguing in light of the recent description by Chow and coworkers that ψ modification of H69 leads to a pH-sensitive structural change.⁷⁸

Residues within helix 69 are directly involved in stabilizing the 70S ribosome as methylation of bases A1912 and A1918 in the H69 loop interferes with subunit association, and deletion of key residues within the loop also results in severely impaired ribosome function.^{75,79} Ribosomal release factor (RRF) makes direct contact with $m^3\psi$ 1915 in H69 along with ψ 1917 and A1916.⁸⁰ The participation of the intersubunit bridge B2a in RRF binding could provide a functional link between the ribosomal A-site in 16S RNA and the LSU RNA that is important for mRNA release upon termination. Noller and coworkers have shown that deletion of the H69 loop resulted in a dominant lethal phenotype, but that, remarkably, protein synthesis *in vitro* remained robust. A recent structure of the 70S ribosome in complex with release factor 1 shows that H69 is directly involved in translation termination, with A1913 participating in a stacking interaction with A1492, one of the key residues for mRNA binding.⁸¹ The conclusion from Noller's work is that despite the myriad of functions ascribed to H69, two crucial roles are likely to be stabilization of 70S ribosomes in the absence of tRNA through the B2a intersubunit bridge involving H69, and stabilization of release factor 1 binding upon translation termination.^{81,82}

The pseudouridines within the H69 loop are synthesized by the RulD protein, one of the few modification enzymes in *E. coli* that is required for normal growth, that is, the mutant is severely impaired.⁸³ Consistent with the findings of Noller and coworkers, and the modification results of Maivali and Remme, biochemical evidence supports a role for pseudouridine modification where a specific ψ -dependent structure of H69 is necessary for subunit association.^{79,81} Our current understanding of the ribosome structure is insufficient to ascertain whether the ψ s are directly involved in subunit interactions, or whether modification causes a conformation change involving other conserved bases in the H69 loop. Since ψ modification of H69 is absolutely conserved and necessary for proper function, an understanding of the influence of modification on structure would be informative.

Chow and coworkers have used NMR and circular dichroism (CD) spectroscopy to investigate the H69 stem-loop structure, and have also characterized the influence of ψ modification on the thermodynamic stability of the H69 stem-loop. Besides the obvious question of the differences between modified and unmodified RNAs, they have addressed the relative roles of the stem and the loop regions, the effect of ψ modification in these two domains, and the differences between the human and *E. coli* RNAs. They synthesized model RNA hairpins corresponding to residues 1906–1924 of *E. coli* 23S RNA with either ψ 1915 or $m^3\psi$ 1915, and compared these RNAs with U-containing hairpins.^{78,84,85} Early studies showed that modifications in the stem were generally stabilizing as had been seen for tRNA systems, while loop ψ s were surprisingly destabilizing.⁸⁵ The result of full modification was an overall neutral effect with respect to global thermodynamics. In tRNAs, an NMR signal has been generally seen for ψ N1–H regardless of whether the ψ is base paired.^{55,56,67} A striking result from the *E. coli* H69 paper was that the loop ψ s did not seem to have stably bound water molecules, perhaps explaining the thermodynamic destabilization.⁸⁵

More recent work from the Chow laboratory established that ψ modification results in a pH-dependent structural change with an apparent pK_a of approximately 6.5.⁷⁸ The corresponding all-U hairpin has a similar fold to the modified RNA, but the base pairing geometry and stability is not pH sensitive. With respect to their NMR studies, at low pH the 1915 N1–H peak is dramatically increased in intensity, suggesting that a stable water bridge forms in this structure, while the 1917 N1–H peak is still not observed. For ψ 1917, despite the absence of an N1–H interaction (unless this peak is overlapped with 1915 N1–H), a stable interaction was detected for the N3–H ψ 1917 proton at a chemical shift consistent with a noncanonical pair. The pH-dependent structural changes observed by Chow and coworkers could be seen as support for a tunable system where different states of the ribosome could alter the pK_a of a key H69 residue. A ψ -dependent, pH-sensitive conformational switch would explain why the high density of ψ residues is conserved.

6.19.5 Base-Modified Nucleosides in tRNA

6.19.5.1 Background

tRNAs contain more than 70 modifications ranging in chemical complexity from the simple ψ and 2'-OME modifications discussed above, to the elaborate heterocycles and the amino acid-modified nucleosides found in the tRNA anticodon.^{1,5} The roles of many modifications are still unknown, but the function of a subset of the modifications in the anticodon region has been established due to their direct effects on translation.^{2,86} The reviews by Yokoyama and Nishimura, and by Bjork have summarized the observations establishing the importance of these modifications. In the present work, we will provide an update for a few specific base modifications where recent genetic and biochemical studies have provided considerable insight into the biochemical effects of these nucleosides. The recent literature will also be interpreted in light of structural studies from our laboratory, and others, that are beginning to correlate the intrinsic chemical properties of a few of these nucleosides with their effects on tRNA anticodon structure (Figure 9).

6.19.5.2 2-Thio, 5-X-Modified Uridines – s^2U , mnm^5s^2U , and mcm^5s^2U

tRNAs with U35 and U36 in the anticodon – tRNA^{Gln}, tRNA^{Glu}, and tRNA^{Lys} – have a modified wobble nucleoside among the family related to the mnm^5s^2U nucleoside found in *E. coli*. This modification has long been recognized as affecting aminoacyl synthetase recognition, codon affinity, and reading frame maintenance.² It is one of the few modifications where deletion of the modification gene(s) appears to be lethal, although the presumed lethality of the *mnmA/mnmE* double mutant in *E. coli* seems to be based on the inability to isolate a double mutant strain. Nevertheless, the *mnmE* mutant in *E. coli* is synthetically lethal,⁸⁷ unmodified tRNA^{Lys} does not bind to programmed ribosomes,^{88,89} and the mutants lacking s^2U or the mnm^5U side chain, individually, are severely impaired.⁹⁰

Studies of *E. coli* and *S. typhimurium* ribosomal frameshifting established that both the s^2 and mnm^5 groups in mnm^5s^2U affected the recruitment of tRNAs to the ribosomal A-site.⁶³ The effect was more significant for tRNA^{Lys} than for tRNA^{Gln} in this system, consistent with other studies showing the strong dependence of the tRNA^{Lys} anticodon structure on wobble position modification.^{91,92} For frameshifting effects in the P-site, both the wobble position modifications helped maintain the reading frame; for tRNA^{Gln}, decoding the CAA codon was rather sensitive to modification while decoding CAG was less sensitive. The conclusion was that in both the Lys and Gln tRNAs, the mnm^5s^2U wobble modification plays an important role in reducing slippage by the P-site tRNA.⁶³

The eukaryotic homologue of mnm^5s^2U is mcm^5s^2U . Bjork and coworkers⁹³ showed that a yeast double mutant lacking the ability to make both the s^2U and the mcm^5 side chains of mcm^5s^2U was not viable. However, the double mutant could be partially rescued by overexpression of unmodified tRNA^{Gln}, tRNA^{Glu}, and tRNA^{Lys}, all of which normally contain mcm^5s^2U .³⁴ Additional experiments established that tRNA^{Lys} was the most sensitive, and that the partial rescue was due to compensation for the decreased efficiency with which the tRNA read the appropriate codon as opposed to other factors such as tRNA aminoacylation or tRNA stability, that is, a more weakly bound tRNA could be compensated for by increasing the concentration.

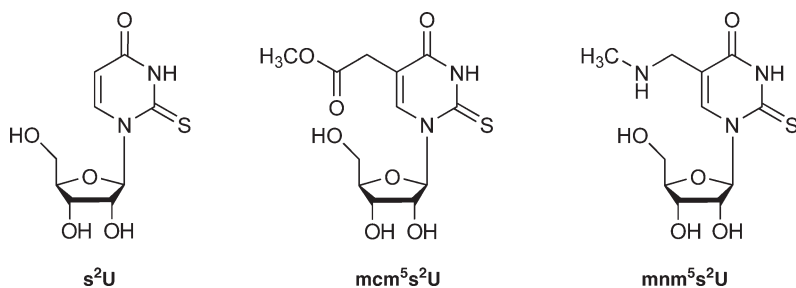


Figure 9 Structures of s^2U , mcm^5s^2U , and mnm^5s^2U . These modified nucleosides are found at the wobble position of Lys, Gln, and Glu tRNAs pairing with A or sometimes G at the third codon position.

Our laboratory has systematically studied the $\text{mnm}^5\text{s}^2\text{U}$ and $\text{mcm}^5\text{s}^2\text{U}$ wobble position modifications in order to understand how these modifications affect tRNA structure and how structural changes are related to function.^{91,92,94,95} The 2-thio modification is important for stacking stabilization, especially in the $\text{tRNA}^{\text{Lys}}\text{UUU}$ anticodon where the consecutive Us are very poor stackers. This modification alone is sufficient to provide sufficient affinity for tRNA to bind programmed ribosomes, and for the tRNA to support translation *in vivo*.^{63,96} The function of the mnm^5 and mcm^5 side chain modifications is to restrict the conformational space of the wobble nucleoside, and this steric effect organizes the anticodon sufficiently to prevent miscoding.^{63,91}

A Watson–Crick G- $\text{mnm}^5\text{s}^2\text{U}$ geometry has been proposed by Takai and Yokoyama for G recognition by $\text{mnm}^5\text{s}^2\text{U}$ -containing tRNAs.^{97,98} It was argued that the rare tautomer is promoted by the side chain, and by the increased acidity of N3–H due to sulfur modification. The mnm^5U nucleoside without a 2-thio has also been proposed to recognize G in this way during protein synthesis. However, the crystal structure of mnm^5U -containing tRNA^{Lys} bound to an AAG codon shows that the $\text{mnm}^5\text{U}34$ wobble nucleoside forms a bifurcated hydrogen bond with G.⁹⁹ This base pair geometry is between the positions that U34 occupies in a Watson–Crick pair and the position that would be necessary for a standard U–G wobble pair. Murphy *et al.* argue that the very poorly stacked $\text{tRNA}^{\text{Lys}}\text{UUU}$ anticodon would be further destabilized if the mnm^5U –G pair was in a wobble geometry because the modified U would then be completely unstacked with U35. tRNAs with a purine at position 35 would not be so destabilized since there could still be partial stacking with U34 in a wobble geometry. Agris has reviewed work from several laboratories establishing that a key function of anticodon modification is to decrease entropy in the free tRNA, and that this organizational effect can be important regardless of whether the wobble modification expands or restricts wobble recognition (Figure 10).^{100,101}

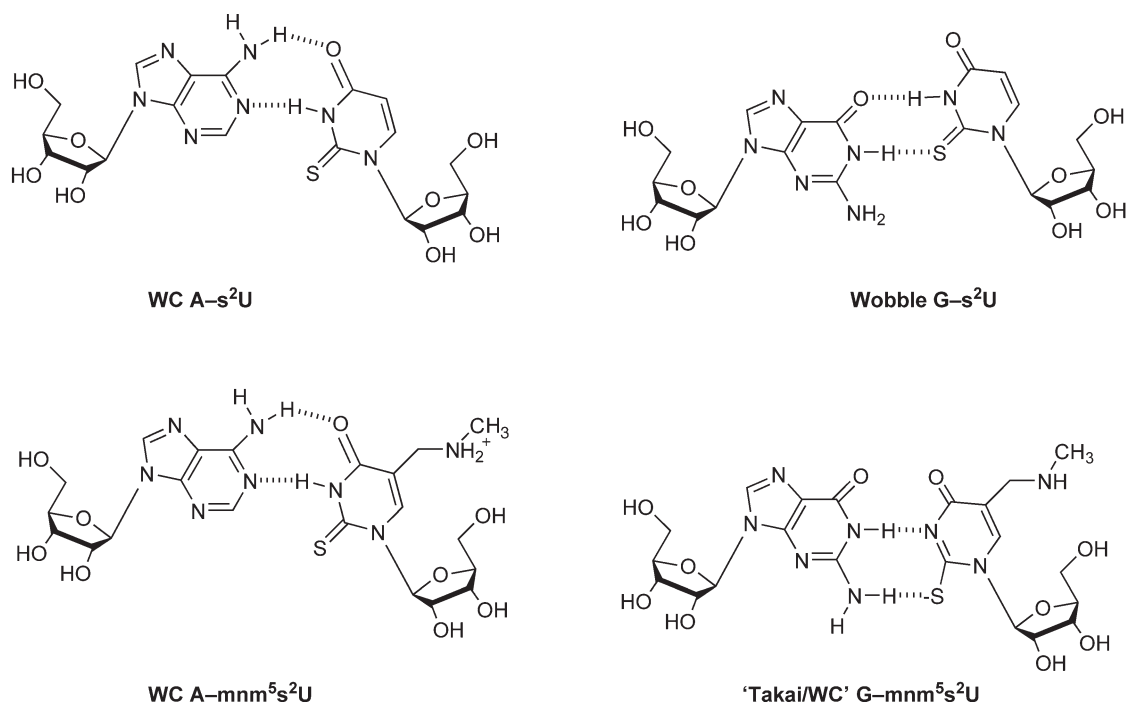


Figure 10 Base pair geometries for s^2U -modified nucleosides. ($\text{A-s}^2\text{U}$ Watson–Crick) pairs are stabilized relative to A-U pairs. The 2-thio for Watson–Crick pairs is in the nonhydrogen bonding position allowing sulfur to stabilize by increasing the acidity of the N3 proton, promoting a 3'-endo sugar conformation, and by stacking better with neighboring pairs. (Wobble $\text{G-s}^2\text{U}$) pairs are destabilized relative to G-U pairs since S2 forms a weaker H-bond than O2. (Watson–Crick $\text{A-mnm}^5\text{s}^2\text{U}$) pairs are stabilized through the sulfur effect as described for $\text{A-s}^2\text{U}$. The amine side chain is protonated at pH 7.0 and likely provides additional stabilization through charge neutralization. A (Watson–Crick $\text{G-mnm}^5\text{s}^2\text{U}$) geometry has been proposed by Takai and Yokoyama for G recognition by $\text{mnm}^5\text{s}^2\text{U}$ -containing tRNAs.^{97,98}

6.19.5.3 Nucleosides that Form Rare Tautomers – cmo^5U and $\tau\text{m}^5\text{U}$

The wobble position nucleoside uridine-5-oxyacetic acid (cmo^5U) has long been considered the counterpoint modification to the 2-thiouridine family of modifications. Whereas wobble modifications such as $\text{mm}^5\text{s}^2\text{U}$ restrict wobble pairing, the cmo^5U modification expands wobble pairing allowing $\text{cmo}^5\text{U}34$ -containing tRNAs to recognize A, G, U, and perhaps even C in the first codon position.¹⁰² For example, $\text{tRNA}^{\text{Pro}}_{\text{cmo}^5\text{UGG}}$ was shown by Bjork and coworkers to be capable of reading all four proline codons, CCA, CCG, CCU, and CCC, demonstrating that *in vivo* the modified U expanded wobble even to recognition of C. Kothe and Rodnina published a detailed kinetic analysis of tRNA^{Ala} containing the modified anticodon sequence cmo^5UGC , to investigate its codon recognition properties and to compare ribosome binding of the GC-rich Ala codon recognition compared to tRNA^{Phe} (GAA anticodon). They determined that the affinities of the two tRNAs were similar and that the binding of $\text{tRNA}^{\text{Ala}}_{\text{cmo}^5\text{UGC}}$ for the GCC alanine codon, containing the presumably unstable $\text{cmo}^5\text{U}-\text{C}$ pair, was only slightly weaker than a fully cognate pairing interaction, and dramatically more stable than a noncognate wobble interaction. When we last reviewed this modification, it was generally accepted that the mechanism of wobble expansion was by promotion of the 2'-endo sugar conformation in cmo^5U and related modified nucleosides.³⁰ Yokoyama *et al.*³² had presented a convincing mechanism for expansion of wobble pairing by the xo^5U family of nucleosides based on their demonstration that these nucleosides favored the 2'-endo conformation. The 2'-endo conformation is more flexible than 3'-endo,¹⁰ therefore, it was plausible that this property at the nucleoside level explained the wobble expansion by cmo^5U in tRNAs; a 2'-endo conformation would allow the modified wobble nucleoside the flexibility to pair with any codon nucleoside (four-way wobble).

However, recent studies by Agris and coworkers indicates that the effect of cmo^5U in the context of the tRNA anticodon is quite different than the behavior at the nucleoside level.¹⁰³ Although the 5-position modification may favor the 2'-endo conformation in the nucleoside, interactions of the side chain in the anticodon context results in a more rigid, organized, and stacked structure for modified $\text{tRNA}^{\text{Val},3}_{\text{UAC}}$. This affect is consistent with a steric mechanism for uridine 5-substituted nucleosides in general, as we described for $\text{mm}^5\text{s}^2\text{U}$, $\text{mcm}^5\text{s}^2\text{U}$, and mcm^5U .⁹¹ The effect that Vendeix *et al.* described for the free tRNA was consistent with that seen for ribosome-bound $\text{tRNA}^{\text{Val}}_{104}$. The crystal structure for the bound tRNA with each of the four codons showed that all the tRNAs were bound in a conformation that maximized stacking, and that recognition of G was facilitated by a shift to the enol tautomer, enabling a Watson-Crick Geometry. An emerging, general model for wobble recognition is that the native, modified tRNAs have preordered anticodon structures that minimize the entropic penalty of codon binding, and modulate base pair geometries to maximize stacking within the anticodon triplet (Figure 11).¹⁰⁰

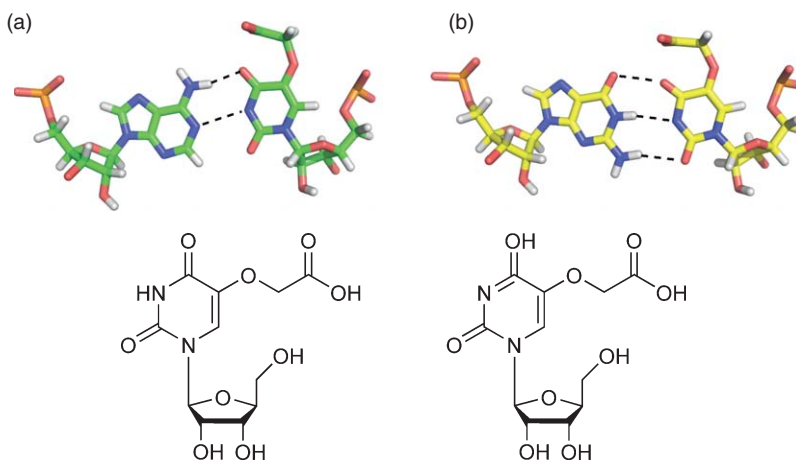


Figure 11 Base pairing in the 30S ribosome crystal structures for: (a) $\text{tRNA}^{\text{Val}}_{\text{cmo}^5\text{UAC}}$ paired with A. Adenosine pairs with cmo^5U in the keto form shown below the base pair from the crystal structure. (b) G presumably pairs with cmo^5U in the enol form shown below the base pair. Base pairs are from PDB ID: 2UUC and 2UU9.

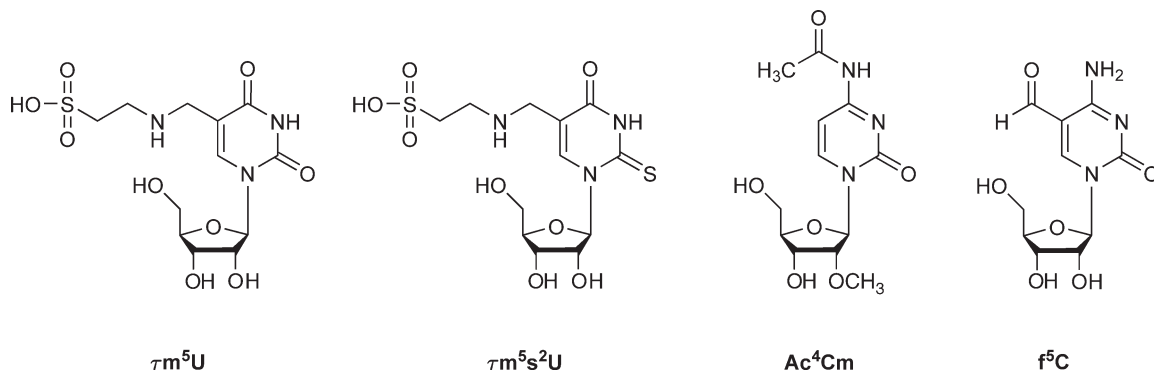


Figure 12 τm^5U , τm^5s^2U , Ac^4Cm , and f^5C structures.

Watanabe and coworkers have investigated the structure and function of human mitochondrial tRNAs containing wobble modifications derived from taurine, the nucleosides 5-taurinomethyluridine (τm^5U) and 5-taurinomethyl-2-thiouridine (τm^5s^2U).^{105,106} Mutations that affect the modification of the tRNA are associated with MERRF disease (myoclonus epilepsy associated with ragged red fibers) or MELAS, a mitochondrial myopathy.¹⁰⁷ Certain mutations known to cause these diseases result in specifically unmodified tRNA^{Lys} or tRNA^{Leu} isoacceptors that fail to properly recognize cognate codons. *In vitro* translation experiments in an *E. coli* system showed that the τm^5U modification was important for stabilizing U–G pairing.¹⁰⁵ In a result that was quite analogous to that seen for cmo^5U described above, crystal structures of a tRNA containing τm^5U bound to either UUA or UUG codons showed that the τm^5U –G pair and the τm^5U –A pair had very similar, slightly distorted Watson–Crick geometries.¹⁰⁵ This suggests that the τm^5U modification may tautomerize to the enol form as seen for cmo^5U ,¹⁰⁴ or it could stabilize the bifurcated geometry seen for mnm^5U .⁹⁹ Regardless of those details, the requirement for τm^5U in mitochondrial tRNA^{Leu} provides further support for the importance of 34–35 stacking stabilization in tRNAs having U34 and either an A–U or an U–A pair in the middle (35) position (Figure 12).

6.19.5.4 Modified Cytidines – Ac^4C , f^5C , Ac^4Cm , and f^5Cm

Cytidine in the tRNA wobble position is often modified in either the sugar or the base with functional groups that promote the 3'-endo sugar conformation at the nucleoside level. Experiments by Takemoto *et al.* suggested that f^5C modification restricts wobble pairing so that only G is recognized by tRNA^{Met}_{f⁵CAU}. However, wobble restriction by f^5C is not absolute since in cytoplasmic bovine tRNA^{Leu}_{f⁵CmAA}, both the UUC and UUA leucine codons might be recognized.^{108,109} Watanabe and coworkers suggest that bovine mitochondrial tRNA^{Met} with f^5C34 may recognize both AUG and AUA as methionine codons, but their base pairing model would require an A⁺–C pair.¹¹⁰ The effects of f^5C and f^5Cm on codon recognition are particularly interesting in light of discussions in previous sections about wobble modification, given the extreme effect that the f^5 modifications have on the nucleoside conformation. The properties of f^5C are similar to that of ac^4C modification in that the 3'-endo conformation is significantly stabilized.^{29,111} The base modifications alone stabilize the 3'-endo conformation, presumably through stereoelectronic effects at the anomeric bond and through space interactions involving the H5–H6 double bond and electron lone pairs on O4'.^{26–30} The effects of f^5C on RNA duplex stability have not been investigated.

The ac^4C nucleoside is found at the wobble position of methionine tRNA from *E. coli* as well as glutamine, glutamate, lysine, proline, and serine tRNAs in archaea.⁵ Kawai *et al.*¹¹² showed that ac^4C strongly favors the 3'-endo conformation, and that this conformational preference is reinforced by simultaneous 2'-OMe to form the ac^4Cm nucleoside found in tRNAs from hyperthermophilic archaea. The presence of N-acetylated cytidine in organisms growing above 100 °C is somewhat remarkable given the hydrolytic lability of this functionality. In light of the dramatic effects that ac^4 and f^5 modifications have on the cytidine conformational preference, these modifications could have applications for RNA duplex stabilization similar to that shown recently for s^2U .⁵⁸

6.19.5.5 The Lysine-Modified Nucleoside, Lysidine

In *E. coli*, and *Bacillus subtilis* the *tilS* gene codes for the enzyme tRNA^{Ile}-lysidine synthetase responsible for forming the lysine-modified nucleoside lysidine (k^2C) (Figure 13).^{113,114} tRNA^{Ile}_{k²CAU} specifically recognizes the rare AUA codon in a lysidine-dependent fashion;^{115,116} the ‘major’ Ile codons AUU and AUC are read by tRNA^{Ile}_{GAU}.¹¹⁷ This wobble position-modified C nucleoside is an essential modification in bacteria,¹¹⁶ and analogous genes are found in archaea, but apparently not in eukaryotes.¹¹⁸ The fact that lysidine is absent in eukaryotes, and required in bacteria has led researchers to comment on the possibility that the TilS enzyme presents a novel antibacterial target.¹¹⁸

The unmodified CAU anticodon is of course the sequence of tRNA^{Met}, and this raised the question as to how tRNA^{Met}_{CAU} and tRNA^{Ile}_{k²CAU} are distinguished by their relevant aminoacyl tRNA synthetases. For both tRNA^{Met} and tRNA^{Ile}_{k²CAU} the wobble position nucleoside is critical, serving as a positive recognition element for the two cognate synthetases. For tRNA^{Met}, C34 is a positive determinant and k^2C34 is a negative determinant; for tRNA^{Ile} the opposite is true.¹¹⁸ Additionally, a small number of residues in the anticodon, the anticodon stem, and in the acceptor stem provide further discrimination against improper charging of the two closely related tRNAs.^{118,119}

The fundamental question about how lysidine alters codon recognition at the wobble position remains to be conclusively answered. The structure and protonation state was determined by Muramatsu *et al.*¹¹³ through NMR analysis and chemical synthesis. The nucleobase is apparently charged at physiological pH, allowing for several alternate structures depending on where one places the proton. The structure in Figure 13(b) is an attractive possibility since it allows for Watson–Crick pairing with A, which would maximize stacking between the wobble k^2C and the weakly stacked U35. As we have now repeated several times, an emerging theme for modified wobble pairing is for modifications to promote Watson–Crick-like geometries for noncanonical interactions, especially in AU-rich codons (Figure 14).

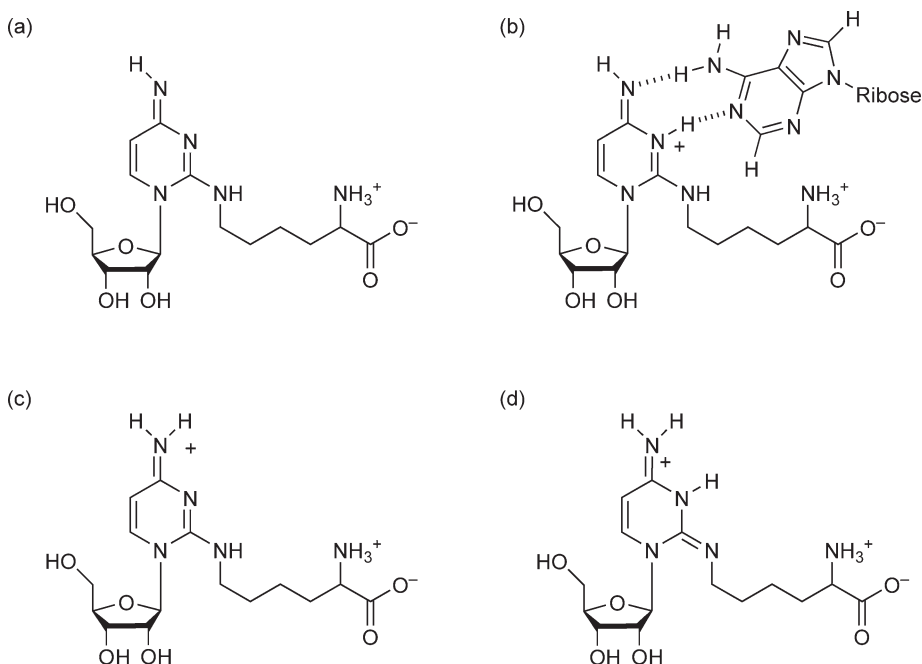


Figure 13 Structures and base pairing of the modified nucleoside lysidine (k^2C). (a) The neutral form of lysidine. The lysidine base has a pK_a of approximately 12, and therefore the base would be charged at $pH = 7$ along with a zwitterionic form for the side chain. (b) Protonated form of lysidine allowing pairing with A in a Watson–Crick geometry. (c) and (d) Alternative protonated forms of lysidine that would be less favorable for A-pairing.

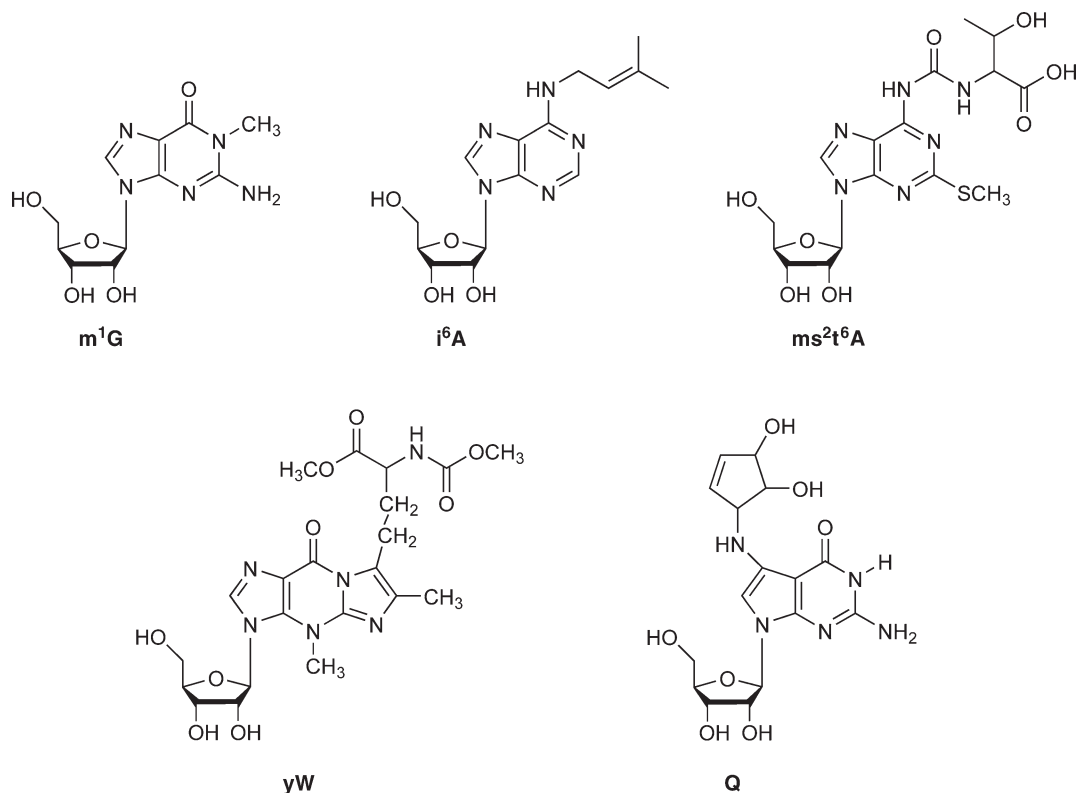


Figure 14 Chemical structures of m¹G, yW, i⁶A, ms²t⁶A, and Q. i⁶A and t⁶A are found both with and without the 2-thiomethyl group that can further increase base stacking with a nucleotide in the first codon position (cross-strand stacking).

6.19.5.6 Purine Modifications – m¹G, i⁶A, t⁶A, yW, and Q

tRNAs have a conserved purine at position 37 adjacent to the anticodon, and this purine is invariably modified (Figure 1). Crystal structures of the ribosome with tRNAs bound show that purine 37 participates in a cross-strand stacking interaction with the first codon nucleotide,⁹⁹ essentially a 3'-terminal stacking interaction, known to provide RNA duplex stability comparable to that of an additional base pair.⁷ Since purines are 'better stackers' than pyrimidines, the conservation of a purine at position 37 is logical, but the universal modification suggests that the position 37 base may be important for more than stacking stabilization, or that modifications may enhance the stacking interaction as well as participate in other structural interactions within the anticodon.⁹⁹ Indeed, NMR studies have shown that m¹G, i⁶A, and t⁶A each function to organize the tRNA anticodon structure to stabilize the appropriate stacked conformation on the 3'-side of the tRNA, and to provide an open-loop structure that minimizes cross-loop base pairing. It appears that the ideal conformation for mRNA binding is one where the position 37 nucleotide promotes 37, 38, and 39 stacking with a discontinuity of stacking with the position 36 anticodon base.^{91,92,120} At the same time, this allows for continuous stacking of 34–36 (for *Saccharomyces pombe* tRNA^{Met}; this may not be true),¹²⁰ and destabilization of pairing between U33 and the purine at position 37. For m¹G, the modification clearly blocks pairing with U33 by blocking the Watson–Crick face of G37, yW would also be blocked from Watson–Crick pairing. For t⁶A37, the conformation observed in the crystal structure, and shown by NMR to be the dominant form in solution destabilizes an across-the-loop base pair with U33.^{91,92,99,121} Nikonowicz and coworkers used NMR to show that i⁶A37 promotes an open-loop conformation in *E. coli* tRNA^{Phe}, and that the combination of i⁶A37 and divalent metals stabilizes the canonical U-turn structure observed in solution for most native tRNAs, and seen in tRNA crystal structures (Figure 15).^{122,123}

The deceptively simple m¹G modification is universally conserved and has been proposed as being a member of the minimal set of genes required for life.^{125,126} Genes for m¹G are found in all three domains of life,

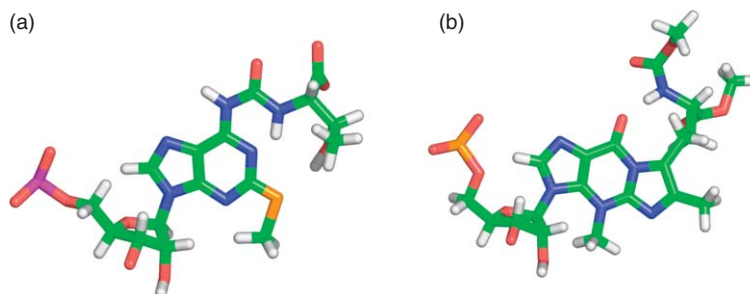


Figure 15 Structure of ms^2t^6A37 from the crystal structure of tRNA^{Lys} (PDB ID: 1XMQ) and the yW nucleoside from tRNA^{Phe} (PDB ID: 1EHZ). The tricyclic ring of yW provides an extensive stacking interface for interaction with the cross-strand codon base. The carbamoyl threonine side chain of t^6A likewise is planar with the adenosine base, a conformation stabilized by a hydrogen bond from the amide to the N1-nitrogen of A. This provides an extensive stacking interface as seen in the crystal structure of tRNA^{Lys} bound to the 30S ribosome.^{99,124}

including the genome of *Mycoplasma genitalium*, the organism with the smallest sequenced genome. Bjork *et al.*¹²⁶ demonstrated that m^1G37 in tRNA is required for normal growth in *E. coli* and in *Saccharomyces cerevisiae*.¹²⁶ Their studies indicate that m^1G plays a crucial role in reading frame maintenance by modulating the strength and therefore the fidelity of tRNA/mRNA pairing. Their genetic and biochemical results are consistent with the model of Agris where m^1G helps maintain an open-loop structure for tRNAs.^{101,127}

The hypermodified nucleosides yW and Q are notable in the extensive additional functionality present, and in the case of Q the fact that the 7-deaza base is inserted rather than modified in place.¹²⁸ Both of these nucleosides influence reading frame maintenance, particularly in cases of +1 frameshifting.^{63,129} Bjork and coworkers interpret the effects on +1 frameshifting, as opposed to little or no effect on -1 frameshifting, in the context of the simultaneous slippage model.¹²⁹ Both these modifications affect the affinity of tRNA for their cognate codons, although the Bjork model requires that these effects be manifested in the P-site where less is known regarding the structural effects of tRNA modification. Both Q and yW modification affect the frequency of frameshifting in retroviruses. The effect of yW is directly on tRNA^{Phe}, while the effect of Q undermodification is seen if the AAU codon read by Q-containing tRNA^{Asn} is immediately upstream of the shift site.^{130,131} The influence of yW can be interpreted in the context of base stacking as has been described for other position 37 modifications. To explain the effect of Q modification, one is inclined to view this similarly to the results for modified uridines in the wobble position. Q-containing tRNAs that pair with U are more sensitive to modification than those that pair with C. That is, both Q34 and G34 interact efficiently with C, but Q modification improves pairing with U in the A-site, decreasing +1 frameshifting. At present we do not know the details of the base pairing interaction, but the effect of Q modification is reminiscent of the changes seen for modified U wobble nucleosides.

6.19.6 Summary and Future Prospects

RNA biochemists studying nucleoside modifications have been plagued by the question, ‘what do they all do?’ Only recently has it been demonstrated that RNA modifications can be necessary for cell viability.⁹³ Gene sequences and more powerful genetic tools are now allowing researchers to identify modification genes and then knock out RNA modifications singly, or in combination with other genetic lesions.^{132,133} Sophisticated kinetic experiments have also shown that tRNA and rRNA modification deletions can have major effects on the fidelity of protein synthesis.¹³⁴ These approaches and the expanded appreciation for the role of RNA in cellular function will certainly lead to the identification of new roles for RNA modification. The recent characterization of inosine modification in micro RNAs is one example of the convergence of nucleoside modification and the new, expanded RNA world.^{135,136} Modification enzymes may also turn out to be important therapeutic targets. The lysidine-modified nucleoside is necessary for growth in bacteria, but is not found in eukaryotes.¹³⁷ This suggests that therapeutics inhibiting enzymes in the lysidine pathway might be effective antibiotics with few side effects.

The s^2U modification in the wobble position of tRNA has one of the greatest effects on RNA stability of any single modification. The large stabilization provided by this nucleoside and the strong A–U versus G–U discrimination of s^2U suggest the potential for application in the area of siRNA therapeutics. Nawrot and coworkers found that in 5–7 s^2U modifications in an siRNA duplex raised the T_m from near 80 °C in 15, 17, and 19 bp duplexes to over 95 °C for the modified duplexes.⁵⁸ This modification may see increased application in the siRNA field since the modifications are tolerated by the siRNA machinery,⁵⁸ and may modulate binding to other cellular sites, such as toll-like receptors.¹³⁸

Other base-modified nucleosides also have tremendous effects at the nucleoside level, although these nucleosides have not been investigated at the RNA duplex level. It would seem that f^5C and ac^4C as well as their 2'-OMe derivatives have the potential to significantly stabilize therapeutic RNAs such as miRNAs and siRNAs. The ac^4C and ac^4Cm nucleosides are overrepresented in rRNAs from hyperthermophilic archaea, suggesting the possibility for a major RNA stabilization effect.⁴⁵ This is amazing in light of the hydrolytic susceptibility of the N4-acetyl group. An investigation of these and other nucleoside analogues from thermophiles may lead to novel applications in the area of therapeutic RNAs where the thermal stability and nuclease resistance are foremost among the challenges. Nucleoside modification is nature's elegant solution to these problems; we simply need to understand what nature has done, and then create innovative applications of this toolbox.

References

1. J. Rozenski; P. F. Crain; J. A. McCloskey, *Nucleic Acids Res.* **1999**, *27*, 196–197.
2. G. R. Bjork, Biosynthesis and Function of Modified Nucleosides. In *tRNA: Structure, Biosynthesis and Function*; D. Soll, U. L. RajBhandary, Eds.; ASM Press: Washington, DC, 1995; pp 165–205.
3. W. A. Decatur; M. J. Fournier, *Trends Biochem. Sci.* **2002**, *27*, 344–351.
4. J. P. Bachellerie; J. Cavaille, *Trends Biochem. Sci.* **1997**, *22*, 257–261.
5. M. Sprinzl; K. S. Vassilenko, *Nucleic Acids Res.* **2005**, *33*, D139–D140.
6. J. D. Puglisi; I. Tinoco, *Methods Enzymol.* **1989**, *180*, 304–325.
7. M. J. Serra; D. H. Turner; S. M. Freier, *Methods Enzymol.* **1995**, *259*, 243–261.
8. T. Xia; D. H. Mathews; D. H. Turner, Thermodynamics of RNA Secondary Structure Formation. In *Comprehensive Natural Products Chemistry*; D. Soll, S. Nishimura, P. B. Moore, Eds.; Elsevier: Amsterdam, 1999; pp 21–47.
9. I. Tinoco, Jr.; O. C. Uhlenbeck; M. D. Levine, *Nature* **1971**, *230*, 362–367.
10. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag: New York, 1984.
11. C. H. Lee; I. Tinoco, *Biochemistry* **1977**, *16*, 5403–5414.
12. C. H. Lee; I. Tinoco, *Biophys. Chem.* **1980**, *11*, 283–294.
13. H. Inoue; Y. Hayase; A. Imura; S. Iwai; K. Miura; E. Ohtsuka, *Nucleic Acids Res.* **1987**, *15*, 6131–6148.
14. G. Kawai; U. Harumi; M. Yasuda; K. Sakamoto; T. Hashizume; J. McCloskey; T. Miyazawa; S. Yokoyama, *Nucleic Acids Res.* **1991**, *25*, 49–50.
15. G. Kawai; Y. Yamamoto; T. Kamimura; T. Masegi; M. Sekine; T. Hata, *Biochemistry* **1992**, *31*, 1040–1046.
16. C. Lee; I. Tinoco, *Biophys. Chem.* **1980**, *11*, 283–294.
17. C. H. Lee; I. Tinoco, Jr., *Biochemistry* **1977**, *16*, 5403–5414.
18. W. K. Olson, *J. Am. Chem. Soc.* **1982**, *104*, 278–286.
19. S. Uesugi; H. Miki; M. Ikehara; H. Iwahashi; Y. Kyogoku, *Tetrahedron Lett.* **1979**, *42*, 4073–4976.
20. W. Guschlbauer; K. Jankowski, *Nucleic Acids Res.* **1980**, *8*, 1421–1433.
21. R. K. Kumar; D. R. Davis, *Nucleic Acids Res.* **1997**, *25*, 1272–1280.
22. H. Sierzputowska-Gracz; E. Sochacka; A. Malkiewicz; K. Kuo; C. Gehrke; P. Agris, *J. Am. Chem. Soc.* **1987**, *109*, 7171–7177.
23. S. Yokoyama; Z. Yamaizumi; S. Nishimura; T. Miyazawa, *Nucleic Acids Res.* **1979**, *6*, 2611–2626.
24. R. K. Kumar; D. R. Davis, *Nucleosides Nucleotides* **1997**, *16*, 1469–1472.
25. L. C. Sowers; B. R. Shawk; W. D. Sedwick, *Biochem. Biophys. Res. Commun.* **1987**, *148*, 790–794.
26. E. Juaristi; G. Cuevas, *The Anomeric Effect*, CRC Press: Boca Raton, FL, 1995.
27. E. Egert; H. J. Lindner; W. Hillen; M. C. Bohm, *J. Am. Chem. Soc.*, **1980**, *102*, 3707–3713.
28. W. Uhl; J. Reiner; H. G. Gassen, *Nucleic Acids Res.* **1983**, *11*, 1167–1180.
29. G. Kawai; T. Hashizume; M. Yasuda; T. Miyazawa; J. McCloskey; S. Yokoyama, *Nucleosides Nucleotides* **1992**, *11*, 759–771.
30. D. R. Davis, Biophysical and Conformational Properties of Modified Nucleosides in RNA. In *Modification and Editing of RNA: The Alteration of RNA Structure and Function*; H. Grosjean, R. Benne, Eds.; ASM Press: New York, 1998; pp 85–102.
31. P. Agris, *Biochimie* **1991**, *73*, 1345–1349.
32. S. Yokoyama; T. Watanabe; K. Murao; H. Ishikura; Z. Yamaizumi; S. Nishimura; T. Miyazawa, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4905–4909.
33. P. Dande; T. P. Prakash; N. Sioufi; H. Gaus; R. Jarres; A. Berdeja; E. E. Swayze; R. H. Griffey; B. Bhat, *J. Med. Chem.* **2006**, *49*, 1624–1634.
34. A. L. Jackson; J. Burchard; D. Leake; A. Reynolds; J. Scheluter; J. Guo; J. M. Johnson; L. Lim; J. Karpilow; K. Nichols; W. Marshall; A. Khvorova; P. S. Linsley, *RNA* **2006**, *12*, 1197–1205.

35. T. P. Prakash; C. R. Allerson; P. Dande; T. A. Vickers; N. Sioufi; R. Jarres; B. F. Baker; E. E. Swayze; R. H. Griffey; B. Bhat, *J. Med. Chem.* **2005**, *48*, 4247–4253.
36. C. R. Allerson; N. Sioufi; R. Jarres; T. P. Prakash; N. Naik; A. Berdeja; L. Wanders; R. H. Griffey; E. E. Swayze, B. Bhat, *J. Med. Chem.* **2005**, *48*, 901–904.
37. T. P. Prakash; B. Bhat, *Curr. Top. Med. Chem.* **2007**, *7*, 641–649.
38. T. P. Prakash; J. F. Johnston; M. J. Graham; T. P. Condon; M. Manoharan, *Nucleic Acids Res.* **2004**, *32*, 828–833.
39. Y. L. Chiu; T. M. Rana, *RNA* **2003**, *9*, 1034–1048.
40. B. A. Kraynack; B. F. Baker, *RNA* **2006**, *12*, 163–176.
41. B. P. Monia; J. F. Johnston; H. Sasnor; L. L. Cummins, *J. Biol. Chem.* **1996**, *271*, 14533–14540.
42. J. Kowalak; J. Dalluge; J. McCloskey; K. Stetter, *Biochemistry* **1994**, *33*, 7869–7876.
43. K. Watanabe; T. Oshima; K. Iijima; Z. Yamaizumi; S. Nishimura, *J. Biochem.* **1980**, *87*, 1–13.
44. K. Watanabe; T. Oshima; S. Nishimura, *Nucleic Acids Res.* **1976**, *3*, 1703–1713.
45. J. A. McCloskey; J. Rozenski, *Nucleic Acids Res.* **2005**, *33*, D135–D138.
46. B. E. Maden, *Prog. Nucleic Acid Res. Mol. Biol.* **1990**, *39*, 241–303.
47. K. R. Noon; E. Bruenger; J. A. McCloskey, *J. Bacteriol.* **1998**, *180*, 2883–2888.
48. J. A. McCloskey; S. Nishimura, *Acc. Chem. Res.* **1977**, *10*, 403–410.
49. J. R. Patton; R. W. Padgett, *BMC Mol. Biol.* **2005**, *6*, 20.
50. J. Ofengand; A. Bakin, *J. Mol. Biol.* **1997**, *266*, 246–268.
51. B. E. Maden, *J. Mol. Biol.* **1988**, *201*, 289–314.
52. B. G. Lane; J. Ofengand; M. W. Gray, *Biochimie* **1995**, *77*, 7–15.
53. D. R. Davis; C. D. Poulter, *Biochemistry* **1991**, *30*, 4223–4231.
54. J. G. Arnez; T. A. Steitz, *Biochemistry* **1994**, *31*, 7560–7567.
55. D. R. Davis, *Nucleic Acids Res.* **1995**, *23*, 5020–5026.
56. D. R. Davis; C. A. Veltri; L. Nielsen, *J. Biomol. Struct. Dyn.* **1998**, *15*, 1121–1132.
57. P. C. Durant; D. R. Davis, *J. Mol. Biol.* **1999**, *285*, 115–131.
58. K. Sipa; E. Sochacka; J. Kazmierczak-Baranska; M. Maszewska; M. Janicka; G. Nowak; B. Nawrot, *RNA* **2007**, *13*, 1301–1316.
59. Y. Lin; C. L. Kielkopf, *Biochemistry* **2008**, *47*, 5503–5514.
60. H. O. Kammen; C. C. Marvel; L. Hardy; E. E. Penhoet, *J. Biol. Chem.* **1988**, *263*, 2255–2263.
61. F. Leconte; G. Somos; A. Sauer; E. C. Hurt; Y. Motorin; H. Grosjean, *J. Biol. Chem.* **1998**, *273*, 1316–1323.
62. C. L. Turnbough, Jr.; R. J. Neill; R. Landsberg; B. N. Ames, *J. Biol. Chem.* **1979**, *254*, 5111–5119.
63. J. Urbonavicius; Q. Qian; J. M. B. Durand; T. G. Hagervall; G. R. Bjork, *EMBO J.* **2001**, *20*, 4863–4873.
64. J. R. Roth; D. N. Anton; P. E. Hartman, *J. Mol. Biol.* **1966**, *22*, 305–323.
65. D. T. Palmer; P. H. Blum; S. W. Artz, *J. Bacteriol.* **1983**, *153*, 357–363.
66. S. Raychaudhuri; L. Niu; J. Conrad; B. G. Lane; J. Ofengand, *J. Biol. Chem.* **1999**, *274*, 18880–18886.
67. I. Tworowska; E. P. Nikonowicz, *J. Am. Chem. Soc.* **2006**, *128*, 15570–15571.
68. Y. T. Yu; E. C. Scharl; C. M. Smith; J. A. Steitz, The Growing World of Small Nuclear Ribonucleoproteins. In *The RNA World*, 2nd ed.; R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds.; CSHL Press: Cold Spring Harbor, NY, 1999; pp 487–524.
69. S. Massenet; A. Mougin; C. Branlant, Posttranscriptional Modifications in the U Small Nuclear RNAs. In *Modification and Editing of RNA*; H. Grosjean, R. Benne, Eds.; ASM Press: Washington, DC, 1998; pp 201–227.
70. M. I. Newby; N. L. Greenbaum, *RNA* **2001**, *7*, 833–845.
71. X. Zhao; Y. T. Yu, *RNA* **2004**, *10*, 681–690.
72. C. Yang; D. S. McPheeters; Y. T. Yu, *J. Biol. Chem.* **2005**, *280*, 6655–6662.
73. M. I. Newby; N. L. Greenbaum, *Nat. Struct. Biol.* **2002**, *9*, 958–965.
74. C. C. Query; M. J. Moore; P. A. Sharp, *Genes Dev.* **1994**, *8*, 587–597.
75. A. Liiv; D. Karitkina; U. Maivali; J. Remme, *BMC Mol. Biol.* **2005**, *6*, 18.
76. J. A. Kowalak; E. Bruenger; T. Hashizume; J. Peltier; J. Ofengand; J. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 688–693.
77. B. S. Schuwirth; M. A. Borovinskaya; C. W. Hau; W. Zhang; A. Vila-Sanjurjo; J. M. Holton; J. H. Cate, *Science* **2005**, *310*, 827–834.
78. S. C. Abeyirigunawardena; C. S. Chow, *RNA* **2008**, *14*, 782–792.
79. U. Maivali; J. Remme, *RNA* **2004**, *10*, 600–604.
80. R. K. Agrawal; M. R. Sharma; M. C. Kiel; G. Hirokawa; T. M. Booth; C. M. Spahn; R. A. Grassucci; A. Kaji; J. Frank, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8900–8905.
81. M. Laurberg; H. Asahara; A. Korostelev; J. Zhu; S. Trakhanov; H. F. Noller, *Nature* **2008**, *454*, 852–857.
82. I. K. Ali; L. Lancaster; J. Feinberg; S. Joseph; H. F. Noller, *Mol. Cell* **2006**, *23*, 865–874.
83. N. S. Gutsell; M. P. Deutscher; J. Ofengand, *RNA* **2005**, *11*, 1141–1152.
84. M. Sumita; J. P. Desaulniers; Y. C. Chang; H. M. Chui; L. Clos, II; C. S. Chow, *RNA* **2005**, *11*, 1420–1429.
85. M. Meroueh; P. J. Grohar; J. Qiu; J. SantaLucia, Jr.; S. A. Scaringe; C. S. Chow, *Nucleic Acids Res.* **2000**, *28*, 2075–2083.
86. S. Yokoyama; S. Nishimura, Modified Nucleosides and Codon Recognition. In *tRNA: Structure, Biosynthesis and Function*; D. Soll, U. RajBhandary, Eds.; ASM Press: Washington, DC, 1995; pp 207–224.
87. M. Martinez-Vicente; L. Yim; M. Villarroya; M. Mellado; E. Perez-Paya; G. R. Bjork; M. E. Armengod, *J. Biol. Chem.* **2005**, *280*, 30660–30670.
88. C. Yarian; M. Marszalek; E. Sochacka; A. Malkiewicz; R. Guenther; A. Miskiewicz; and P. F. Agris, *Biochemistry* **2000**, *39*, 13390–13395.
89. U. Von Ahlsen; R. Green; R. Schroeder; H. Noller, *RNA* **1997**, *3*, 49–56.
90. T. G. Hagervall; S. C. Pomerantz; J. A. McCloskey, *J. Mol. Biol.* **1998**, *284*, 33–42.
91. P. C. Durant; A. C. Bajji; M. Sundaram; R. K. Kumar; D. R. Davis, *Biochemistry* **2005**, *44*, 8078–8089.
92. M. Sundaram; P. C. Durant; D. R. Davis, *Biochemistry* **2000**, *39*, 12575–12584.
93. G. R. Bjork; B. Huang; O. P. Persson; A. S. Bystrom, *RNA* **2007**, *13*, 1245–1255.
94. A. Bajji; D. R. Davis, *Org. Lett.* **2000**, *2*, 3865–3868.

95. A. Bajji; M. Sundaram; D. G. Myszkka; D. R. Davis, *J. Am. Chem. Soc.* **2002**, *124*, 14302–14303.
96. S. Ashraf; E. Sochacka; R. Cain; R. Guenther; A. Malkiewicz; P. Agris, *RNA* **1999**, *5*, 188–194.
97. K. Takai, *Nucleic Acids Symp. Ser. (Oxf)* **2005**, *49*, 317–318.
98. K. Takai; S. Yokoyama, *Nucleic Acids Res.* **2003**, *31*, 6383–6391.
99. F. V. Murphy; V. Ramakrishnan; A. Malkiewicz; P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186–1191.
100. P. F. Agris, *EMBO Rep.* **2008**, *9*, 629–635.
101. E. M. Gustilo; F. A. Vendeix; P. F. Agris, *Curr. Opin. Microbiol.* **2008**, *11*, 134–140.
102. S. K. Mitra; F. Lustig; B. Akesson; T. Axberg; P. Elias; U. Lagerkvist, *J. Biol. Chem.* **1979**, *254*, 6397–6401.
103. F. A. Vendeix; A. Dziergowska; E. M. Gustilo; W. D. Graham; B. Sproat; A. Malkiewicz; P. F. Agris, *Biochemistry* **2008**, *47*, 6117–6129.
104. A. Weixlbaumer; F. V. t. Murphy; A. Dziergowska; A. Malkiewicz; F. A. Vendeix; P. F. Agris; V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14*, 498–502.
105. S. Kurata; A. Weixlbaumer; T. Ohtsuki; T. Shimazaki; T. Wada; Y. Kirino; K. Takai; K. Watanabe; V. Ramakrishnan; T. Suzuki, *J. Biol. Chem.* **2008**, *283*, 18801–18811.
106. T. Suzuki; T. Suzuki; T. Wada; K. Saigo; K. Watanabe, *EMBO J.* **2002**, *21*, 6581–6589.
107. Y. Kirino; T. Suzuki, *RNA Biol.* **2005**, *2*, 41–44.
108. J. P. Pais de Barros; G. Keith; C. El Adlouni; A. L. Glasser; G. Mack; G. Dirheimer; J. Desgres, *Nucleic Acids Res.* **1996**, *24*, 1489–1496.
109. C. Takemoto; T. Koike; T. Yokogawa; L. Benkowski; L. Spremulli; T. Ueda; K. Nishikawa; K. Watanabe, *Biochimie* **1995**, *77*, 104–108.
110. J. Moriya; T. Yokogawa; K. Wakita; T. Ueda; K. Nishikawa; P. Crain; T. Hashizume; S. Pomerantz; J. McCloskey; G. Kawai; N. Hayashi; S. Yokoyama; K. Watanabe, *Biochemistry* **1994**, *33*, 2234–2239.
111. G. Kawai; T. Yokogawa; K. Nishikawa; T. Ueda; T. Hashizume; J. McCloskey; S. Yokoyama; K. Watanabe, *Nucleosides Nucleotides* **1994**, *13*, 1189–1199.
112. C. Edmonds; P. Crain; R. Gupta; T. Hashizume; C. Hocart; J. Kowalak; S. Pomerantz; K. Stetter; J. McCloskey, *J. Bacteriol.* **1991**, *173*, 3138–3148.
113. T. Muramatsu; S. Yokoyama; N. Horie; A. Matsuda; T. Ueda; Z. Yamaizumi; Y. Kuchino; S. Nishimura; T. Miyazawa, *J. Biol. Chem.* **1988**, *263*, 9261–9267.
114. T. Muramatsu; K. Nishikawa; F. Nemoto; Y. Kuchino; S. Nishimura; T. Miyazawa; S. Yokoyama, *Nature* **1988**, *336*, 179–181.
115. F. Harada; S. Nishimura, *Biochemistry* **1974**, *13*, 300–307.
116. A. Soma; Y. Ikeuchi; S. Kanemasa; K. Kobayashi; N. Ogasawara; T. Ote; J. Kato; K. Watanabe; Y. Sekine; T. Suzuki, *Mol. Cell* **2003**, *12*, 689–698.
117. M. Yarus; B. G. Barrell, *Biochem. Biophys. Res. Commun.* **1971**, *43*, 729–734.
118. Y. Ikeuchi; A. Soma; T. Ote; J. Kato; Y. Sekine; T. Suzuki, *Mol. Cell* **2005**, *19*, 235–246.
119. K. Nakanishi; S. Fukai; Y. Ikeuchi; A. Soma; Y. Sekine; T. Suzuki; O. Nureki; *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7487–7492.
120. E. Lescrinier; K. Nauwelaerts; K. Zanier; K. Poesen; M. Sattler; P. Herdewijn, *Nucleic Acids Res.* **2006**, *34*, 2878–2886.
121. J. W. Stuart; Z. Gdaniec; R. H. Guenther; M. Marszalek; E. Sochacka; A. Malkiewicz; P. F. Agris, *Biochemistry* **2000**, *39*, 13396–13404.
122. J. Cabello-Villegas; I. Tworowska; E. P. Nikonowicz, *Biochemistry* **2004**, *43*, 55–66.
123. J. Cabello-Villegas; W. C. Winkler; E. P. Nikonowicz, *J. Mol. Biol.* **2002**, *319*, 1015–1034.
124. H. Shi; P. B. Moore, *RNA* **2000**, *6*, 1091–1105.
125. G. R. Björk; P. M. Wikstrom; A. S. Bystrom, *Science* **1989**, *244*, 986–989.
126. G. R. Bjork; K. Jacobsson; K. Nilsson; M. J. Johansson; A. S. Bystrom; O. P. Persson, *EMBO J.* **2001**, *20*, 231–239.
127. J. W. Stuart; K. M. Koshlap; R. Guenther; P. F. Agris, *J. Mol. Biol.* **2003**, *334*, 901–918.
128. N. Okada; S. Noguchi; H. Kasai; N. Shindo-Okada; T. Ohgi; T. Goto; S. Nishimura, *J. Biol. Chem.* **1979**, *254*, 3067–3073.
129. J. Urbonavicius; G. Stahl; J. M. Durand; S. N. Ben Salem; Q. Qian; P. J. Farabaugh; G. R. Bjork, *RNA* **2003**, *9*, 760–768.
130. B. A. Carlson; S. Y. Kwon; M. Chamorro; S. Oroszlan; D. L. Hatfield; B. J. Lee, *Virology* **1999**, *255*, 2–8.
131. B. A. Carlson; J. F. Mushinski; D. W. Henderson; S. Y. Kwon; P. F. Crain; B. J. Lee; D. L. Hatfield, *Virology* **2001**, *279*, 130–135.
132. B. Huang; J. Lu; A. S. Bystrom, *RNA* **2008**, *14*, 2183–2194.
133. D. Piekna-Przybylska; P. Przybylski; A. Baudin-Baillieu; J. P. Rousset; M. J. Fournier, *J. Biol. Chem.* **2008**, *283*, 26026–26036.
134. U. Kothe; M. V. Rodnina, *Mol. Cell* **2007**, *25*, 167–174.
135. J. W. Habig; T. Dale; B. L. Bass, *Mol. Cell* **2007**, *25*, 792–793.
136. Y. Kawahara; B. Zinshteyn; P. Sethupathy; H. Iizasa; A. G. Hatzigeorgiou; K. Nishikura, *Science* **2007**, *315*, 1137–1140.
137. H. Grosjean; G. R. Bjork, *Trends Biochem. Sci.* **2004**, *29*, 165–168.
138. K. Kariko; D. Weissman, *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 523–532.

Biographical Sketch



Darrell R. Davis received his B.S. (1982) in chemistry from the University of Puget Sound and a Ph.D. (1988) in organic chemistry from the University of Utah under the direction of C. Dale Poulter. From 1988 to 1989 he was an NIH postdoctoral fellow at the University of Washington, Department of Chemistry, in the laboratory of Brian R. Reid. In 1989 he was appointed to the Department of Medicinal Chemistry in the Departments of Medicinal Chemistry and of Biochemistry at The University of Utah where he currently holds the rank of professor. He has authored and coauthored approximately 50 publications. Current research in the Davis' laboratory involves the use of NMR spectroscopy to investigate the structure and function of RNA, particularly RNA molecules involved in viral replication.

6.19 Nucleoside Analogues

Darrell R. Davis, The University of Utah, Salt Lake City, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.19.1 | Introduction | 663 |
| 6.19.2 | Stereoelectronic Factors that Affect Nucleic Acid Stability | 665 |
| 6.19.2.1 | The Relationship Between Sugar Conformation and Duplex Stability | 665 |
| 6.19.2.2 | 2'-OMe Modification Effects on RNA Duplex Stability | 665 |
| 6.19.2.3 | Base Modification Influences Stacking and RNA Stability | 666 |
| 6.19.3 | 2'-O-Methyl and 2'-Fluoro Modification | 666 |
| 6.19.3.1 | 2'-O-Methyl and 2'-Fluoro Modifications of RNA Duplexes | 666 |
| 6.19.3.2 | 2'-O-Methyl Modifications in Thermophilic tRNA | 666 |
| 6.19.3.3 | 2'-O-Methyl Modification in Ribosomal RNA | 667 |
| 6.19.4 | Pseudouridine | 668 |
| 6.19.4.1 | Fundamental Properties of ψ Modification | 668 |
| 6.19.4.2 | ψ Effects in tRNA | 669 |
| 6.19.4.3 | ψ in Small Nuclear RNA | 669 |
| 6.19.4.4 | ψ in Helix 69 of 23S rRNA | 670 |
| 6.19.5 | Base-Modified Nucleosides in tRNA | 672 |
| 6.19.5.1 | Background | 672 |
| 6.19.5.2 | 2-Thio, 5-X-Modified Uridines – s ² U, mnm ⁵ s ² U, and mcm ⁵ s ² U | 672 |
| 6.19.5.3 | Nucleosides that Form Rare Tautomers – cmo ⁵ U and τ m ⁵ U | 674 |
| 6.19.5.4 | Modified Cytidines – Ac ⁴ C, f ⁵ C, Ac ⁴ Cm, and f ⁵ Cm | 675 |
| 6.19.5.5 | The Lysine-Modified Nucleoside, Lysidine | 676 |
| 6.19.5.6 | Purine Modifications – m ¹ G, i ⁶ A, t ⁶ A, yW, and Q | 677 |
| 6.19.6 | Summary and Future Prospects | 678 |
| References | | 679 |

6.19.1 Introduction

Natural RNA molecules are post-transcriptionally modified, there being at last count 96 chemically distinct modified nucleosides.¹ With a small number of exceptions, RNA modification occurs by enzymes acting directly on the parent A, U, G, or C nucleosides present in an RNA transcript.² For all RNA classes in bacteria, and for the tRNAs in all organisms, the modification reaction is performed by sequence and structurally specific enzymes. That is, a modification enzyme recognizes a target nucleoside within a specific sequence context. The enzyme may also require specific structural motifs within which the sequence must be located, for example, the tRNA L-shape. A modification could also be made in both a tRNA and a ribosomal RNA, but the sequence context would likely be very similar in the two RNAs. For rRNAs in eukaryotes, the process of RNA 2'-O-methylation (2'-OMe) and pseudouridylation (ψ) is quite different from that of tRNA modification. In these rRNAs, the modification is carried out using an RNA 'guide' in the form of a small nucleolar RNA (snoRNA) that has sequence complementarity for the rRNA target site, and specifies the site of modification using a generic modification enzyme acting on the snoRNA/rRNA complex.^{3,4} The generic nature of the snoRNA system explains in part why eukaryotic rRNAs are more extensively modified than bacterial RNAs; the snoRNA guided system uses fewer enzymes, and relies only on RNA sequence complementarity to identify the target. This is more economical than the essentially one enzyme–one modification system used for methylation and ψ of rRNA in bacteria (see Chapter 6.20).

rRNAs and snRNAs for the most part contain relatively simple modifications, primarily 2'-OMe modifications and ψ , along with a smattering of base methylations. One should not discount the profound effects that these modest chemical changes can have on function, but this chemical repertoire is limited compared to the

diversity of modifications found in tRNA. In addition to the proliferation of 2'-OMe nucleosides and ψ throughout tRNAs, the small (~75 nt on average) tRNAs are host to most of the chemical diversity among nucleoside analogues. Conserved tRNA modifications are clustered in the functionally important regions involved in tertiary structure (e.g., the s^4U8 , m^5U54 , $\psi55$, and D19 nucleosides), and in the anticodon region directly involved in modulating translational fidelity. **Figure 1** shows a generic cloverleaf for the secondary structure of eukaryotic tRNAs. The fact that the tRNA wobble position, 34, and the conserved purine 37 adjacent to the anticodon harbor the greatest diversity of modified nucleosides is apparent in the figure. In addition to the largest number of different modifications, by far the most chemically complex modified nucleosides are found at positions 34 and 37.

In this review, we begin the discussion with two modifications that are chemically simple and widespread, the 2'-OMe modification and ψ . Within the context of the 2-OMe discussion we have also included the nonnatural 2'-F modifications. A review of 2'-F modification highlights the stereochemical similarities to 2'-OMe, and shows why 2'-OMe and 2'-F are both becoming widely used in oligonucleotide therapeutics. More exotic modifications from the tRNA anticodon region are grouped according to the parent nucleosides from which they are derived, and discussed according to how the modifications affect fundamental properties

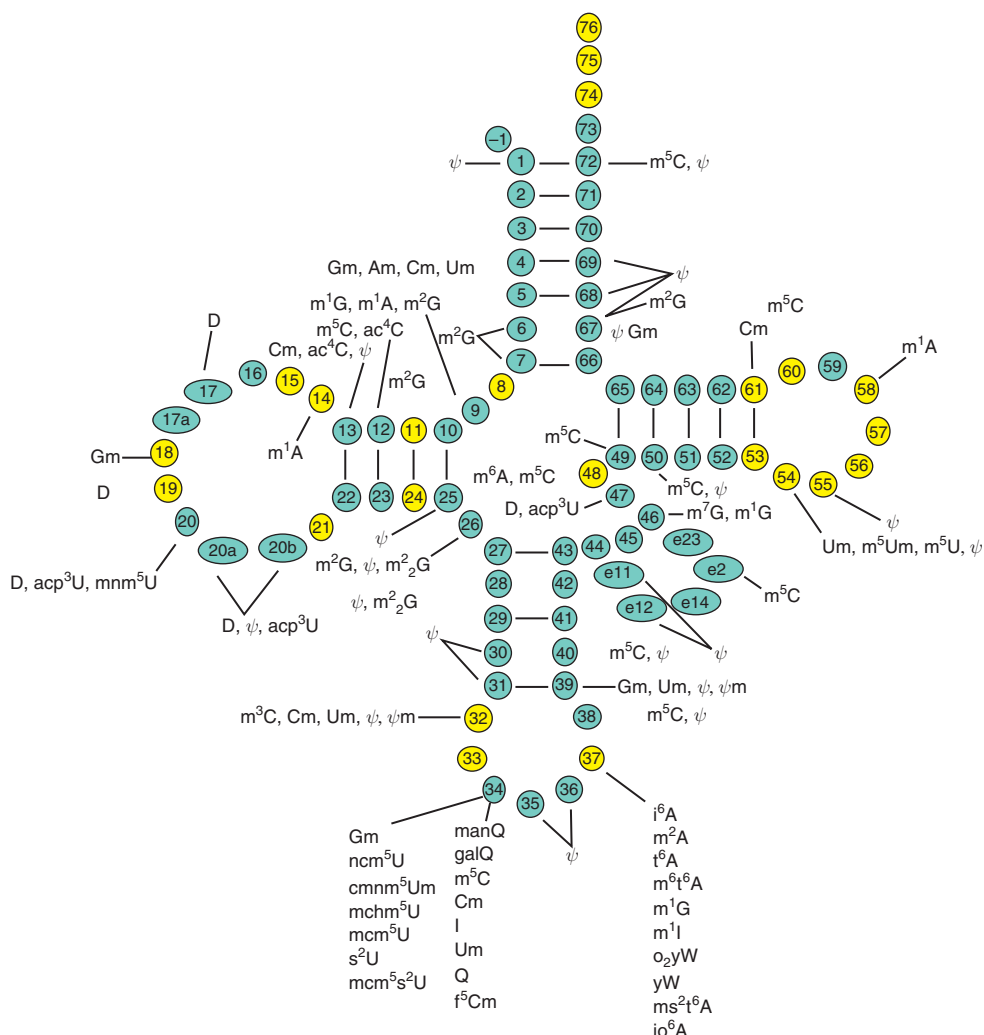


Figure 1 Summary of the identities and locations for modified nucleosides in eukaryotic tRNA.^{1,5} Conserved/invariant nucleosides are in gold and sites of common insertions are shown as ellipses.

such as nucleoside conformation and base stacking. This overview of a large and diverse field will hopefully provide an appreciation for the biological importance of nucleoside modification, and a glimpse of how much remains to be discovered.

6.19.2 Stereoelectronic Factors that Affect Nucleic Acid Stability

6.19.2.1 The Relationship Between Sugar Conformation and Duplex Stability

The thermodynamic stability of simple RNA duplexes, hairpins, and more complex RNA molecules up to the size of a tRNA can be measured by the midpoint of their thermal denaturation, the T_m .⁶ The melting temperature of both simple and complex RNA molecules is mainly a function of the stability of the component duplex regions, although single-stranded stacking and special loop elements such as tetraloops can contribute to stability.^{7,8} The major factor affecting RNA helical stability is the base pair composition, and the nearest-neighbor effect.⁹ Implicit in the nearest-neighbor approximation is that base stacking effects determine nucleic acid stability.¹⁰ Lee and Tinoco^{11,12} realized that the nucleoside sugar conformation had a direct effect on base stacking interactions. Small RNAs were seen to have an increase in the percentage of 3'-endo nucleoside conformation in direct relationship to the degree of stacking, and nucleotides in the middle of A-form RNA helices are 100% 3'-endo.

The cause and effect relationship between sugar conformation and base stacking–duplex stability suggested that nucleoside modifications that promote a 3'-endo sugar conformation at the nucleoside level would result in an increase in RNA duplex stability.^{11,13,14} To the extent that single-strand stacking contributes to thermodynamic stabilization of a folded RNA, 3'-endo-promoting modifications in single-stranded loops may also provide global stabilization. The classic examples of sugar modifications that stabilize RNA are the 2'-OMe and 2'-F modifications (Figure 2).

6.19.2.2 2'-OMe Modification Effects on RNA Duplex Stability

In nature, 2'-OMe modification is used to modulate RNA structure and stability. At the nucleoside level, it has been shown that the effect of 2'-OMe is to bias the conformational preference toward the 3'-endo conformation found in A-form helical structures.¹⁵ This conformational preference positions the base in the axial orientation and concomitantly increases RNA base stacking.^{16,17} Changing the 2'-sugar substituent alters the stereoelectronic properties of the nucleoside, and these effects can reinforce each other to favor the 3'-endo, axial conformation. The steric effect is easy to visualize, and can be demonstrated with a simple molecular model where the steric interaction between the 2' position and the base is minimized in the 3'-endo conformation. This effect is greater for pyrimidines where the 2-carbonyl clashes with a bulky 2' group if the sugar conformation is 2'-endo. Another important effect of nucleoside modification is exemplified by the nonnatural 2'-F modification where the strongly electron-withdrawing fluorine stabilizes the 3'-endo, axial conformation through the gauche effect, just as the 2'-OH of RNA promotes the 3'-endo conformation compared to deoxyribose.^{18–20}

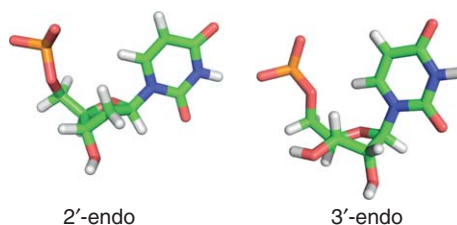


Figure 2 Sugar conformations for B-form DNA (2'-endo) and A-form RNA (3'-endo) shown for deoxyuridine and uridine, respectively. Notice that the distance between the 2'-OH and the O2 carbonyl of U is maximized by the 3'-endo sugar conformation.

6.19.2.3 Base Modification Influences Stacking and RNA Stability

Numerous base modifications also stabilize the 3'-endo sugar conformation, and may also improve base stacking by affecting dipole–dipole interactions between adjacent bases.^{15,21} Among the natural modifications, 2-thiouridine (*s*²U) is the champion modification for stabilizing RNA. This modification has effects on the nucleoside sugar conformation, increases base pairing strength, and has base stacking effects that all reinforce an A-form RNA structure. The more specific effects are steric interactions between the 2-thiocarbonyl and the 2'-hydroxyl that favor a 3'-endo sugar pucker, an increase in acidity for the imino proton making a stronger base pair, and polarization of the anomeric bond favoring an axial base orientation.^{21–24} 2-Thiouridine is a modification *tour de force* for such a simple chemical alteration. Other examples of base modifications that improve base stacking through polarization changes are the 5-methyl group of m⁵U and m⁵C,²⁵ and the formyl and acetyl modifications of f⁴C and ac⁴C. The latter two modifications appear to change the double bond polarization in the pyrimidine ring and may also promote the 3'-endo sugar conformation through the anomeric effect.^{26–30}

6.19.3 2'-O-Methyl and 2'-Fluoro Modification

6.19.3.1 2'-O-Methyl and 2'-Fluoro Modifications of RNA Duplexes

2'-OMe modification was first shown to stabilize RNA oligonucleotide duplexes by Ohtsuka and coworkers.¹³ In nature, a striking example of RNA duplex stabilization by 2'-OMe modification is the special three base pair duplex formed between the tRNA anticodon and the three base codon triplet. Uridine at the tRNA wobble position is almost always modified, and these modifications modulate the selectivity of the tRNA for either A or G pairing.³¹ It has been shown that promotion of the 3'-endo sugar conformation for the wobble nucleoside is effective at restricting pairing to A compared to G.^{15,32} This restriction of wobble pairing is typically accomplished with 2'-OMe modification, 2-thio modification, or some combination of these modifications along with additional base modification.

Applied to longer RNA duplexes, the increase in T_m upon 2'-OMe modification has seen wide application in antisense RNAs and more recently for siRNAs.^{33–36} The 2'-OMe modification provides stabilization of approximately 1.3 °C per modification while the 2'-F provides approximately 1.8 °C per modification. These modifications increase the potency of therapeutic antisense and siRNAs, and have the added benefit of imparting nuclease resistance to the oligonucleotide.³⁷ In the area of antisense oligonucleotides, it has been established that an important design feature is to maintain the 2'-endo conformation in the center of the oligonucleotide in order to support RNase H activity. This 2'-endo feature comes at the cost of antisense affinity for the RNA target, and is incompatible with 2'-OMe or 2'-F modification. However, the modifications can be placed in the 'wings' of the antisense oligonucleotide to increase affinity and nuclease resistance. The combination of a high-affinity flanking region and a central RNase H active region has been termed a 'gapmer'³⁸(see Chapter 2.19).

It appears that siRNAs are much more sensitive to RNA modification than antisense oligonucleotides, and only a few, simple chemical modifications are tolerated.³⁹ Since the publication by Chiu and Rana, there has been some progress in defining the limits to which sugar modifications can be incorporated into siRNAs. 2'-OMe and 2'-F can be used throughout the sense strand, while in the antisense strand there is some modulation of siRNA activity. Generally, the 3' end of the antisense strand can be extensively modified to increase stability and nuclease resistance, while the 5' end requires more care.^{33,36,40} Position-specific modification in the seed region has been used to reduce off-target effects,³⁴ suggesting also that 2'-OMe or 2'-F modification might have utility in discriminating between siRNA and micro RNA activities (Figure 3).

6.19.3.2 2'-O-Methyl Modifications in Thermophilic tRNA

The presence of modest levels of 2'-OMe modification in natural RNAs from mesophiles begged the question whether 2'-OMe modifications would be prevalent in the RNAs of thermophilic organisms. A study of the hyperthermophiles *Pyrococcus furiosus* and *Pyrodictum occultum* showed that the tRNA from these organisms had

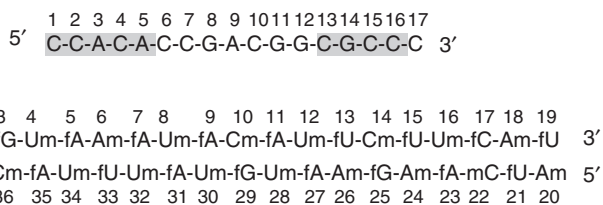


Figure 3 (Top) An antisense 2'-OMe oligonucleotide designed to use the 'gapmer' strategy.⁴¹ Shaded nucleotides in the 'wings' are 2'-OMe modified, the central region is DNA (2'-deoxy), and phosphorothioate backbone modification is used throughout. The 2'-OMe gapmer was approximately 70-fold more potent than a corresponding all-deoxy oligonucleotide. (Bottom) A modified siRNA duplex with alternating 2'-OMe and 2'-F modifications.³⁶ The modified siRNA was approximately 500-fold more potent than an unmodified siRNA.

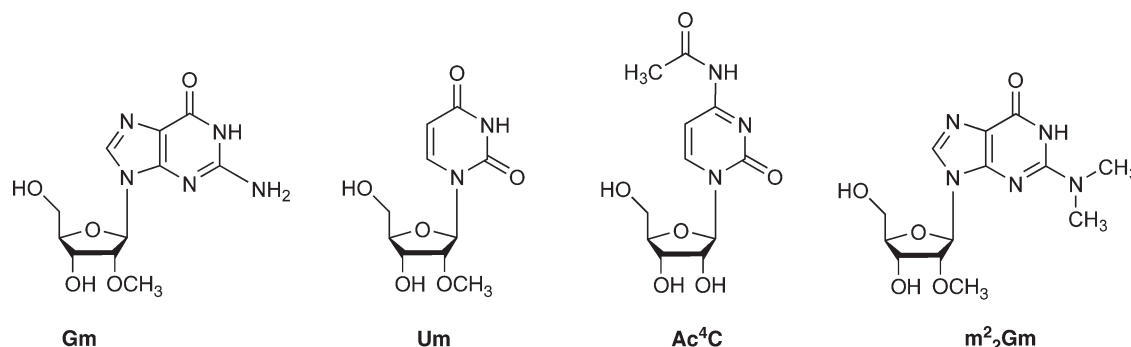


Figure 4 Representative modified nucleosides in thermophilic tRNAs. 2'-OMe G and 2'-OMe U, N4-acetyl C, and N2-dimethyl, 2'-OMe G.

T_m values of approximately 100 °C.⁴² It was perhaps not surprising that the tRNAs had such high T_m s since the optimal growth temperatures of these organisms is over 100 °C. However, studies of tRNA from other thermophilic organisms with optimal growth at 80 °C showed tRNA T_m values of only approximately 80 °C, while *Escherichia coli* tRNA has a quite similar T_m of 75 °C despite the modest optimal growth temperature of 37 °C.⁴³ McCloskey and coworkers showed that the % GC content of the hyperthermophilic tRNAs was over 80%, accounting for some of the observed stabilization but that this GC content was in fact a bit lower than for *Thermus thermophilus* tRNA^{Met} with a T_m of approximately 85 °C.⁴⁴ Furthermore, it was shown by Watanabe *et al.* that *T. thermophilus* tRNA was greatly stabilized by Mg²⁺ (35 °C increase) while the hyperthermophilic tRNA was only stabilized by about 15 °C by Mg²⁺. The mechanism of stabilization of the *P. furiosus* and *P. occultum* tRNAs was in fact due to a very high number of modified nucleosides, particularly Cm, Um, Gm, Am, Ac⁴C, Ac⁴Cm, s²C, s²U, and m₂Gm.⁴² The high percentage of 2'-OMe nucleosides, in particular, was proposed as a major factor in the stabilization above what could be obtained from increased GC content alone (Figure 4). The effects of thiolated pyrimidines and acetyl-C are discussed elsewhere in this report.

6.19.3.3 2'-O-Methyl Modification in Ribosomal RNA

2'-OMe modifications are found in all large and small subunit ribosomal RNAs investigated to date. Bacteria have very few total modifications compared to higher organisms with only 14 total modifications in the SSU RNA of *E. coli* and one 2'-OMe nucleoside in each of *E. coli* and *T. thermophilus* SSU RNA.⁴⁵ In higher organisms, yeast has 45 total modifications, 17 2'-OMe, and the human SSU rRNA has at least 40 2'-OMe nucleosides.^{3,46} While it seems obvious from this comparison that the extent of modification is not well correlated with growth temperature (and by inference RNA stability) when comparing bacteria and eukaryotic rRNA, studies of thermophilic archaea provide a different picture. *Sulfolobus solfataricus* is an archaeal hyperthermophile containing 38 total modifications in the SSU rRNA, with 22 2'-OMe nucleosides. As seen in studies of thermophilic tRNAs, it was found that the number of 2'-OMe nucleosides increased with increasing growth temperature.^{42,47}

Much remains to be learned about the role of nucleoside modification in ribosomal RNA, but the observation that modifications are clustered in the functionally important regions of the ribosome supports the hypothesis that modifications are used to fine-tune ribosomal function.³

6.19.4 Pseudouridine

6.19.4.1 Fundamental Properties of ψ Modification

Pseudouridine is the most common RNA modification, being found in tRNA,⁴⁸ snRNA,⁴⁹ and rRNA⁵⁰ throughout the three kingdoms of life.¹ Studies of tRNA function established that ψ modification in the anticodon affected translational fidelity, but a structural basis for this effect was not immediately obvious since uridine and pseudouridine could both pair with A. As with other RNA modifications,⁵¹ it was noticed some time ago that ψ modifications in rRNA tended to be clustered in the functionally important regions, notably the peptidyl transferase region of 23S RNA.⁵² Structural studies using nuclear magnetic resonance (NMR),⁵³ and an X-ray structure of tRNA^{Gln} complexed with the Gln aminoacyl synthetase,⁵⁴ established that ψ stably coordinates a water molecule between the base and the RNA backbone. This subtle structural property results in the promotion of an idealized A-form geometry around the ψ residue, and an increase in base stacking concomitant with RNA structural stabilization.⁵⁵ For RNAs with ψ modification in a base-paired region, the conversion of U to ψ results in a typical ΔG increase of approximately $0.5 \text{ kcal mol}^{-1}$, and a $3\text{--}5 \text{ }^\circ\text{C}$ increase in T_m per ψ modification.^{56,57} This effect is highly dependent on the structural environment, with ψ modifications in long duplexes providing only modest increases in thermodynamic stability,⁵⁸ while ψ in native structural contexts usually has a greater effect. In natural RNAs, ψ is typically found paired with A or G at the interface of stem and loop regions, in the loop regions proper, or in short helical regions that are A–U rich and presumably unstable in the absence of modification.⁵⁰

The stabilization of A-form RNA structure by ψ modification arises from several synergistic effects. The sequestration of a water molecule between the phosphate backbone and the ψ base orients the nucleotide such that the base is axial, causing the nucleotide to favor the 3'-endo sugar conformation. This 3'-endo nucleotide conformation for ψ is energetically favored despite the lack of an anomeric effect stabilizing the axial orientation in C-nucleosides.⁵² The importance of the water bridge for promoting the axial/3'-endo conformation is shown by the fact that at the nucleoside monomer level where there is no 5'-phosphate, ψ favors 2'-endo relative to uridine, yet even in short single-stranded RNA oligonucleotides, ψ dramatically promotes the 3'-endo conformation compared to unmodified U (Figure 5).⁵⁵

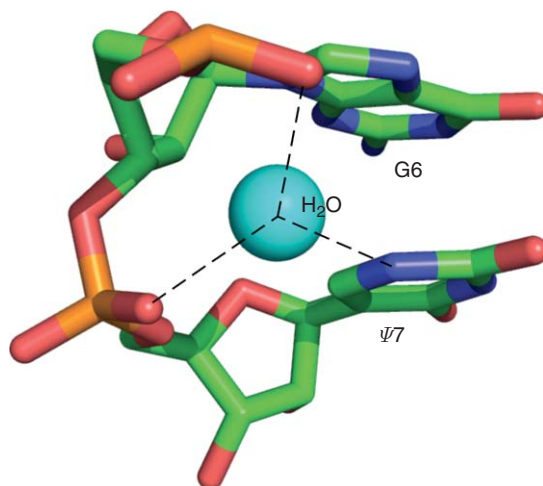


Figure 5 The crystal structure of a ψ within the U2 snRNA-branchpoint duplex showing a stable water molecule coordinated to two phosphates and to the ψ N1–H proton.⁵⁹ PDB ID:3CGR.

6.19.4.2 ψ Effects in tRNA

Pseudouridines are found throughout the tRNA cloverleaf, most notably in the eponymous T ψ C loop containing the highly conserved ψ 55 residue, and in the anticodon stem–loop domain. Modifications in the anticodon often have significant effects on tRNA function since the proximity to mRNA binding region can affect protein synthesis. In *E. coli* and *Salmonella typhimurium*, the *bisT*(TruA) gene is responsible for modifying positions 38/39/40 while in yeast the corresponding modification activity is due to Deg1/pus3.^{60–62} The tRNA anticodon stem–loop may also contain ψ modification at position 32, and any of the three anticodon positions. *HisT* mutants lacking pseudouridine have reduced growth rates, reduced protein elongation rates, and increased errors in translation.^{63–65} We have shown that the A31– ψ 39 base pair in tRNA is destabilized when the ψ is replaced with an unmodified U, that is, in a tRNA isolated from a *bisT* mutant strain.^{53,57} The A31– ψ 39 base pair in tRNA is at the interface between the anticodon stem and loop region, and the structural stabilization provided by ψ is crucial to maintaining the proper tRNA anticodon structure in tRNA^{Lys} and tRNA^{Phe}. Other *E. coli* tRNAs, such as tRNA^{His} have consecutive ψ 38/ ψ 39 residues, and in these tRNAs the stabilizing force is approximately additive, providing even greater stacking stabilization for bases on the 3' side of the tRNA anticodon.⁵⁶ An additional ψ modification formed by a separate pseudouridine synthase, RluA, is found at position 32 of many tRNAs, including *E. coli* tRNA^{Phe}.⁶⁶ Tworowska and Nikonowicz were able to introduce ψ 32 and ψ 39 into an isotopically labeled tRNA^{Phe} anticodon stem–loop to prepare a sample for heteronuclear NMR study. This powerful approach allowed them to show that ψ 32/ ψ 39 modification causes some stabilization of the tRNA structure, but that ψ modification without an additional modification at position 37 was insufficient to promote either a U-turn geometry or the bifurcated ψ 32–A38 base pair observed in native tRNA^{Phe} (Figure 6).⁶⁷

6.19.4.3 ψ in Small Nuclear RNA

In eukaryotes, the spliceosome is a large, dynamic ribonucleoprotein complex responsible for removing introns from mRNAs and generating mature, spliced mRNAs. The splicesomal RNAs are post-transcriptionally modified.⁶⁸ Here we will focus on one highly conserved snRNA, U2, to illustrate the effects of ψ modification on snRNA structure and function (Figure 7).

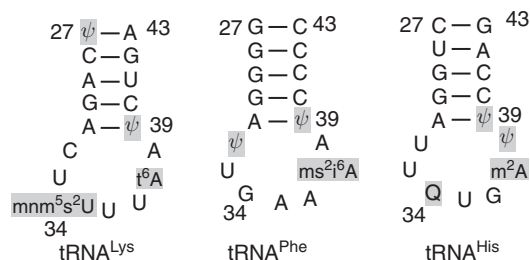


Figure 6 The anticodon stem–loop domains of *Escherichia coli* tRNA^{Lys}, tRNA^{Phe}, and tRNA^{His} contain ψ residues.

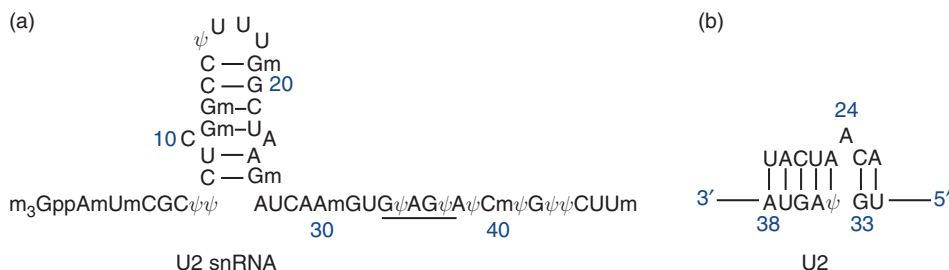


Figure 7 (a) The 5' end of U2 snRNA from vertebrates, containing extensive ψ and 2'-O-methyl modifications.^{68,69} The underlined region from 33–37 is the BSSR. (b) Secondary structure of the yeast U2 branch site showing A24 bulged due to A23– ψ 34 pairing. The numbering in (b) is from Greenbaum and coworkers.⁷⁰

U2 snRNA is a required component of the slicing machinery, and contains one of the most extensively modified regions in RNA. In vertebrates, the first 60 nucleotides contain 11 ψ s in addition to many 2'-OMe residues.⁶⁹ In yeast, the extent of modification is reduced, but ψ s are conserved in the branch site recognition region (BSRR), and these modifications have been shown to be necessary for spliceosome assembly and function.^{71,72} The mechanism of RNA splicing requires the availability of the 2' hydroxyl of a conserved adenosine at the splice site. Greenbaum and coworkers have demonstrated that ψ modification promotes a bulged conformation for A24 (yeast numbering), the branch site nucleophile. The corresponding U-containing RNA allows the As to be stacked within the helix.^{70,73} Their NMR experiments, along with modification-dependent changes detected by 2-aminopurine fluorescence, established that there was a clear difference in the structural environment of the branchpoint A between the ψ -modified and the unmodified RNAs. The NMR and fluorescence experiments suggested an A-bulge structure. The subsequent NMR solution structure supported a model where an A- ψ base pair caused the branchpoint A to be extruded from the helix.⁷³

In a recent crystallographic study, Lin and Kielkopf confirmed that ψ modification results in a structure for the branch site that in turn results in an extrahelical A, but that the selection of which two adenosines would be extrahelical depends on the sequence flanking the branch site.⁵⁹ This is consistent with the observation that either of the two consecutive adenosines across from the ψ can serve as the branch nucleophile.⁷⁴ The crystallographic study investigated three different snRNA-branchpoint duplexes in order to determine which of the two adenosines is used as the nucleophile, and they compared the sequence containing consecutive As with a branchpoint variant from higher eukaryotes where a G is paired with the ψ . The G- ψ -containing duplex was particularly interesting due to the ψ -containing wobble pair, and because of the fortuitous result that duplex BPS3 did not contain strong crystal contacts influencing the conformation of the extrahelical A. Analysis of the extrahelical A environment led Lin and Kielkopf to question one aspect of the NMR structure, that is whether the extrahelical A is positioned in the minor groove as proposed by Greenbaum and coworkers.^{59,73} The NMR structure could be problematic in that the 2'-OH is not readily accessible; however, one could imagine that interactions in the intact spliceosome could easily affect the minor structural change needed to completely swing the branchpoint A into position for splicing. The NMR and crystallography groups are in agreement regarding the importance of a ψ modification for extruding the branchpoint adenosine nucleoside.

6.19.4.4 ψ in Helix 69 of 23S rRNA

For researchers interested in the biological role of modified nucleosides in RNA, H69 in the large subunit RNA (also known as the 1915 loop of 23S RNA) has long been an intriguing subject.⁵⁰ Helix 69 is remarkable in that the loop contains three conserved ψ residues, while in human rRNA the stem has two additional ψ modifications. The seven nucleotide H69 loop has a conserved sequence for the non- ψ residues as well, and these specific nucleotides are known to be important for function.⁷⁵ Helix 69 in the bacterial large subunit (LSU) RNA also contains the only known example of the modified nucleoside $m^3\psi$.⁷⁶ In the human LSU RNA there is a regular ψ at that position, yet an adjacent 2'-OMe A residue may provide a key methyl group to the structure (Figure 8).⁵⁰

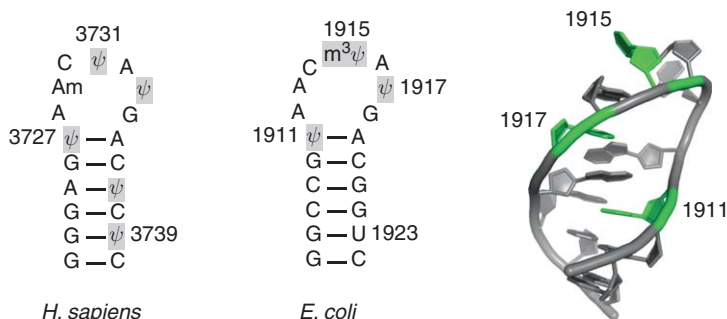


Figure 8 Helix 69 from *Homo sapiens* and *Escherichia coli* large subunit RNAs. (Right) The crystal structure of H69 from the *E. coli* 70S ribosome (ψ residues in green) suggests that ψ 1911 is unpaired in contrast to the model derived from NMR experiments. ψ 1917 may H-bond to N7 of A1912, providing an explanation for the stable imino proton seen by Chow and coworkers. The 3D structure was generated from PDB ID: 2AWB.⁷⁷

Helix 69 is postulated to be involved in several different interactions during the course of protein synthesis. If H69 participates in multiple functions, then conformational flexibility could be important for function, and structural interactions sensitive to pH changes near the physiological range could allow H69 more functional range. This speculation about structural plasticity in H69 is intriguing in light of the recent description by Chow and coworkers that ψ modification of H69 leads to a pH-sensitive structural change.⁷⁸

Residues within helix 69 are directly involved in stabilizing the 70S ribosome as methylation of bases A1912 and A1918 in the H69 loop interferes with subunit association, and deletion of key residues within the loop also results in severely impaired ribosome function.^{75,79} Ribosomal release factor (RRF) makes direct contact with m³ ψ 1915 in H69 along with ψ 1917 and A1916.⁸⁰ The participation of the intersubunit bridge B2a in RRF binding could provide a functional link between the ribosomal A-site in 16S RNA and the LSU RNA that is important for mRNA release upon termination. Noller and coworkers have shown that deletion of the H69 loop resulted in a dominant lethal phenotype, but that, remarkably, protein synthesis *in vitro* remained robust. A recent structure of the 70S ribosome in complex with release factor 1 shows that H69 is directly involved in translation termination, with A1913 participating in a stacking interaction with A1492, one of the key residues for mRNA binding.⁸¹ The conclusion from Noller's work is that despite the myriad of functions ascribed to H69, two crucial roles are likely to be stabilization of 70S ribosomes in the absence of tRNA through the B2a intersubunit bridge involving H69, and stabilization of release factor 1 binding upon translation termination.^{81,82}

The pseudouridines within the H69 loop are synthesized by the RulD protein, one of the few modification enzymes in *E. coli* that is required for normal growth, that is, the mutant is severely impaired.⁸³ Consistent with the findings of Noller and coworkers, and the modification results of Maivali and Remme, biochemical evidence supports a role for pseudouridine modification where a specific ψ -dependent structure of H69 is necessary for subunit association.^{79,81} Our current understanding of the ribosome structure is insufficient to ascertain whether the ψ s are directly involved in subunit interactions, or whether modification causes a conformation change involving other conserved bases in the H69 loop. Since ψ modification of H69 is absolutely conserved and necessary for proper function, an understanding of the influence of modification on structure would be informative.

Chow and coworkers have used NMR and circular dichroism (CD) spectroscopy to investigate the H69 stem-loop structure, and have also characterized the influence of ψ modification on the thermodynamic stability of the H69 stem-loop. Besides the obvious question of the differences between modified and unmodified RNAs, they have addressed the relative roles of the stem and the loop regions, the effect of ψ modification in these two domains, and the differences between the human and *E. coli* RNAs. They synthesized model RNA hairpins corresponding to residues 1906–1924 of *E. coli* 23S RNA with either ψ 1915 or m³ ψ 1915, and compared these RNAs with U-containing hairpins.^{78,84,85} Early studies showed that modifications in the stem were generally stabilizing as had been seen for tRNA systems, while loop ψ s were surprisingly destabilizing.⁸⁵ The result of full modification was an overall neutral effect with respect to global thermodynamics. In tRNAs, an NMR signal has been generally seen for ψ N1–H regardless of whether the ψ is base paired.^{55,56,67} A striking result from the *E. coli* H69 paper was that the loop ψ s did not seem to have stably bound water molecules, perhaps explaining the thermodynamic destabilization.⁸⁵

More recent work from the Chow laboratory established that ψ modification results in a pH-dependent structural change with an apparent pK_a of approximately 6.5.⁷⁸ The corresponding all-U hairpin has a similar fold to the modified RNA, but the base pairing geometry and stability is not pH sensitive. With respect to their NMR studies, at low pH the 1915 N1–H peak is dramatically increased in intensity, suggesting that a stable water bridge forms in this structure, while the 1917 N1–H peak is still not observed. For ψ 1917, despite the absence of an N1–H interaction (unless this peak is overlapped with 1915 N1–H), a stable interaction was detected for the N3–H ψ 1917 proton at a chemical shift consistent with a noncanonical pair. The pH-dependent structural changes observed by Chow and coworkers could be seen as support for a tunable system where different states of the ribosome could alter the pK_a of a key H69 residue. A ψ -dependent, pH-sensitive conformational switch would explain why the high density of ψ residues is conserved.

6.19.5 Base-Modified Nucleosides in tRNA

6.19.5.1 Background

tRNAs contain more than 70 modifications ranging in chemical complexity from the simple ψ and 2'-OME modifications discussed above, to the elaborate heterocycles and the amino acid-modified nucleosides found in the tRNA anticodon.^{1,5} The roles of many modifications are still unknown, but the function of a subset of the modifications in the anticodon region has been established due to their direct effects on translation.^{2,86} The reviews by Yokoyama and Nishimura, and by Bjork have summarized the observations establishing the importance of these modifications. In the present work, we will provide an update for a few specific base modifications where recent genetic and biochemical studies have provided considerable insight into the biochemical effects of these nucleosides. The recent literature will also be interpreted in light of structural studies from our laboratory, and others, that are beginning to correlate the intrinsic chemical properties of a few of these nucleosides with their effects on tRNA anticodon structure (Figure 9).

6.19.5.2 2-Thio, 5-X-Modified Uridines – s^2U , mnm^5s^2U , and mcm^5s^2U

tRNAs with U35 and U36 in the anticodon – tRNA^{Gln}, tRNA^{Glu}, and tRNA^{Lys} – have a modified wobble nucleoside among the family related to the mnm^5s^2U nucleoside found in *E. coli*. This modification has long been recognized as affecting aminoacyl synthetase recognition, codon affinity, and reading frame maintenance.² It is one of the few modifications where deletion of the modification gene(s) appears to be lethal, although the presumed lethality of the *mnmA/mnmE* double mutant in *E. coli* seems to be based on the inability to isolate a double mutant strain. Nevertheless, the *mnmE* mutant in *E. coli* is synthetically lethal,⁸⁷ unmodified tRNA^{Lys} does not bind to programmed ribosomes,^{88,89} and the mutants lacking s^2U or the mnm^5U side chain, individually, are severely impaired.⁹⁰

Studies of *E. coli* and *S. typhimurium* ribosomal frameshifting established that both the s^2 and mnm^5 groups in mnm^5s^2U affected the recruitment of tRNAs to the ribosomal A-site.⁶³ The effect was more significant for tRNA^{Lys} than for tRNA^{Gln} in this system, consistent with other studies showing the strong dependence of the tRNA^{Lys} anticodon structure on wobble position modification.^{91,92} For frameshifting effects in the P-site, both the wobble position modifications helped maintain the reading frame; for tRNA^{Gln}, decoding the CAA codon was rather sensitive to modification while decoding CAG was less sensitive. The conclusion was that in both the Lys and Gln tRNAs, the mnm^5s^2U wobble modification plays an important role in reducing slippage by the P-site tRNA.⁶³

The eukaryotic homologue of mnm^5s^2U is mcm^5s^2U . Bjork and coworkers⁹³ showed that a yeast double mutant lacking the ability to make both the s^2U and the mcm^5 side chains of mcm^5s^2U was not viable. However, the double mutant could be partially rescued by overexpression of unmodified tRNA^{Gln}, tRNA^{Glu}, and tRNA^{Lys}, all of which normally contain mcm^5s^2U . Additional experiments established that tRNA^{Lys} was the most sensitive, and that the partial rescue was due to compensation for the decreased efficiency with which the tRNA read the appropriate codon as opposed to other factors such as tRNA aminoacylation or tRNA stability, that is, a more weakly bound tRNA could be compensated for by increasing the concentration.

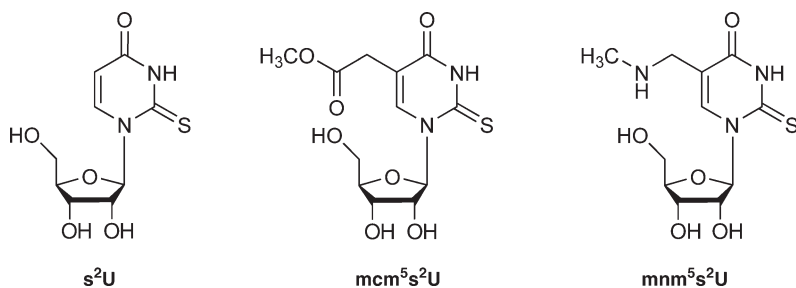


Figure 9 Structures of s^2U , mcm^5s^2U , and mnm^5s^2U . These modified nucleosides are found at the wobble position of Lys, Gln, and Glu tRNAs pairing with A or sometimes G at the third codon position.

Our laboratory has systematically studied the $\text{mnm}^5\text{s}^2\text{U}$ and $\text{mcm}^5\text{s}^2\text{U}$ wobble position modifications in order to understand how these modifications affect tRNA structure and how structural changes are related to function.^{91,92,94,95} The 2-thio modification is important for stacking stabilization, especially in the $\text{tRNA}^{\text{Lys}}\text{UUU}$ anticodon where the consecutive Us are very poor stackers. This modification alone is sufficient to provide sufficient affinity for tRNA to bind programmed ribosomes, and for the tRNA to support translation *in vivo*.^{63,96} The function of the mnm^5 and mcm^5 side chain modifications is to restrict the conformational space of the wobble nucleoside, and this steric effect organizes the anticodon sufficiently to prevent miscoding.^{63,91}

A Watson–Crick G- $\text{mnm}^5\text{s}^2\text{U}$ geometry has been proposed by Takai and Yokoyama for G recognition by $\text{mnm}^5\text{s}^2\text{U}$ -containing tRNAs.^{97,98} It was argued that the rare tautomer is promoted by the side chain, and by the increased acidity of N3–H due to sulfur modification. The mnm^5U nucleoside without a 2-thio has also been proposed to recognize G in this way during protein synthesis. However, the crystal structure of mnm^5U -containing tRNA^{Lys} bound to an AAG codon shows that the $\text{mnm}^5\text{U}34$ wobble nucleoside forms a bifurcated hydrogen bond with G.⁹⁹ This base pair geometry is between the positions that U34 occupies in a Watson–Crick pair and the position that would be necessary for a standard U–G wobble pair. Murphy *et al.* argue that the very poorly stacked $\text{tRNA}^{\text{Lys}}\text{UUU}$ anticodon would be further destabilized if the mnm^5U –G pair was in a wobble geometry because the modified U would then be completely unstacked with U35. tRNAs with a purine at position 35 would not be so destabilized since there could still be partial stacking with U34 in a wobble geometry. Agris has reviewed work from several laboratories establishing that a key function of anticodon modification is to decrease entropy in the free tRNA, and that this organizational effect can be important regardless of whether the wobble modification expands or restricts wobble recognition (Figure 10).^{100,101}

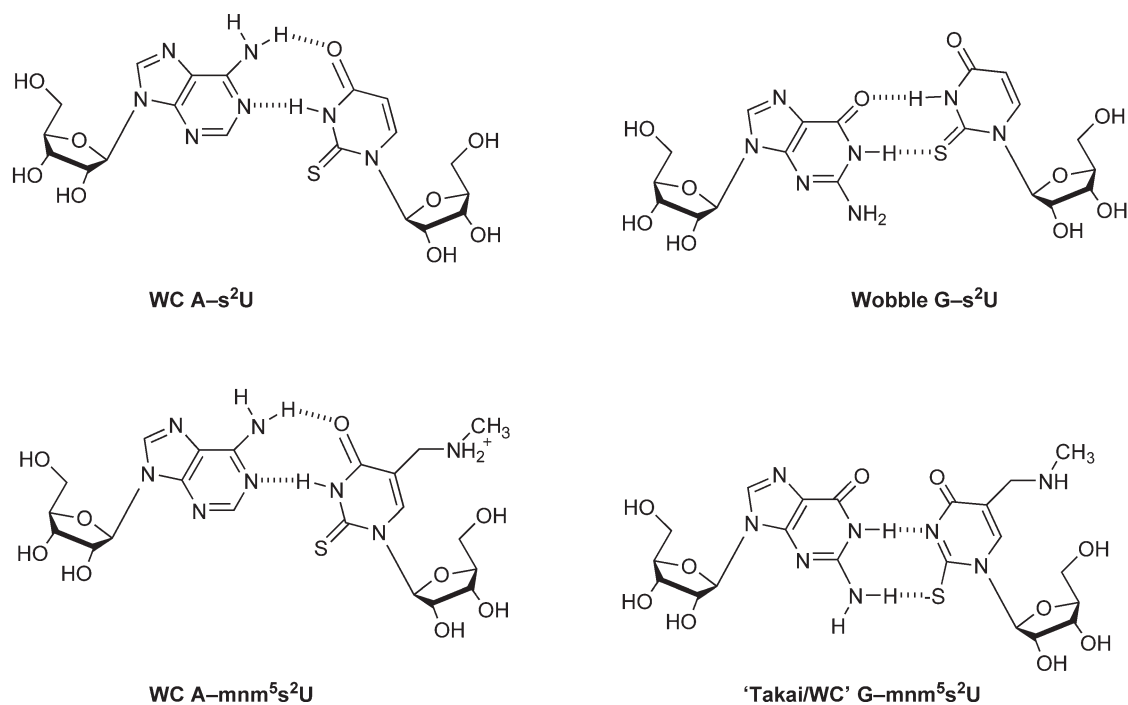


Figure 10 Base pair geometries for s^2U -modified nucleosides. ($\text{A}-\text{s}^2\text{U}$ Watson–Crick) pairs are stabilized relative to $\text{A}-\text{U}$ pairs. The 2-thio for Watson–Crick pairs is in the nonhydrogen bonding position allowing sulfur to stabilize by increasing the acidity of the N3 proton, promoting a 3'-endo sugar conformation, and by stacking better with neighboring pairs. (Wobble $\text{G}-\text{s}^2\text{U}$) pairs are destabilized relative to $\text{G}-\text{U}$ pairs since S2 forms a weaker H-bond than O2. (Watson–Crick $\text{A}-\text{mnm}^5\text{s}^2\text{U}$) pairs are stabilized through the sulfur effect as described for $\text{A}-\text{s}^2\text{U}$. The amine side chain is protonated at pH 7.0 and likely provides additional stabilization through charge neutralization. A (Watson–Crick $\text{G}-\text{mnm}^5\text{s}^2\text{U}$) geometry has been proposed by Takai and Yokoyama for G recognition by $\text{mnm}^5\text{s}^2\text{U}$ -containing tRNAs.^{97,98}

6.19.5.3 Nucleosides that Form Rare Tautomers – cmo^5U and $\tau\text{m}^5\text{U}$

The wobble position nucleoside uridine-5-oxyacetic acid (cmo^5U) has long been considered the counterpoint modification to the 2-thiouridine family of modifications. Whereas wobble modifications such as $\text{mm}^5\text{s}^2\text{U}$ restrict wobble pairing, the cmo^5U modification expands wobble pairing allowing $\text{cmo}^5\text{U}34$ -containing tRNAs to recognize A, G, U, and perhaps even C in the first codon position.¹⁰² For example, $\text{tRNA}^{\text{Pro}}_{\text{cmo}^5\text{UGG}}$ was shown by Bjork and coworkers to be capable of reading all four proline codons, CCA, CCG, CCU, and CCC, demonstrating that *in vivo* the modified U expanded wobble even to recognition of C. Kothe and Rodnina published a detailed kinetic analysis of tRNA^{Ala} containing the modified anticodon sequence cmo^5UGC , to investigate its codon recognition properties and to compare ribosome binding of the GC-rich Ala codon recognition compared to tRNA^{Phe} (GAA anticodon). They determined that the affinities of the two tRNAs were similar and that the binding of $\text{tRNA}^{\text{Ala}}_{\text{cmo}^5\text{UGC}}$ for the GCC alanine codon, containing the presumably unstable $\text{cmo}^5\text{U}-\text{C}$ pair, was only slightly weaker than a fully cognate pairing interaction, and dramatically more stable than a noncognate wobble interaction. When we last reviewed this modification, it was generally accepted that the mechanism of wobble expansion was by promotion of the 2'-endo sugar conformation in cmo^5U and related modified nucleosides.³⁰ Yokoyama *et al.*³² had presented a convincing mechanism for expansion of wobble pairing by the xo^5U family of nucleosides based on their demonstration that these nucleosides favored the 2'-endo conformation. The 2'-endo conformation is more flexible than 3'-endo,¹⁰ therefore, it was plausible that this property at the nucleoside level explained the wobble expansion by cmo^5U in tRNAs; a 2'-endo conformation would allow the modified wobble nucleoside the flexibility to pair with any codon nucleoside (four-way wobble).

However, recent studies by Agris and coworkers indicates that the effect of cmo^5U in the context of the tRNA anticodon is quite different than the behavior at the nucleoside level.¹⁰³ Although the 5-position modification may favor the 2'-endo conformation in the nucleoside, interactions of the side chain in the anticodon context results in a more rigid, organized, and stacked structure for modified $\text{tRNA}^{\text{Val},3}_{\text{UAC}}$. This affect is consistent with a steric mechanism for uridine 5-substituted nucleosides in general, as we described for $\text{mm}^5\text{s}^2\text{U}$, $\text{mcm}^5\text{s}^2\text{U}$, and mcm^5U .⁹¹ The effect that Vendeix *et al.* described for the free tRNA was consistent with that seen for ribosome-bound $\text{tRNA}^{\text{Val}}_{104}$. The crystal structure for the bound tRNA with each of the four codons showed that all the tRNAs were bound in a conformation that maximized stacking, and that recognition of G was facilitated by a shift to the enol tautomer, enabling a Watson-Crick Geometry. An emerging, general model for wobble recognition is that the native, modified tRNAs have preordered anticodon structures that minimize the entropic penalty of codon binding, and modulate base pair geometries to maximize stacking within the anticodon triplet (Figure 11).¹⁰⁰

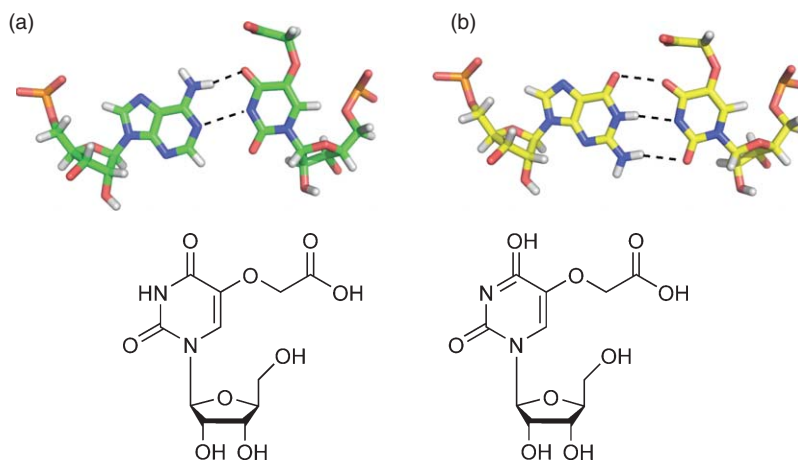


Figure 11 Base pairing in the 30S ribosome crystal structures for: (a) $\text{tRNA}^{\text{Val}}_{\text{cmo}^5\text{UAC}}$ paired with A. Adenosine pairs with cmo^5U in the keto form shown below the base pair from the crystal structure. (b) G presumably pairs with cmo^5U in the enol form shown below the base pair. Base pairs are from PDB ID: 2UUC and 2UU9.

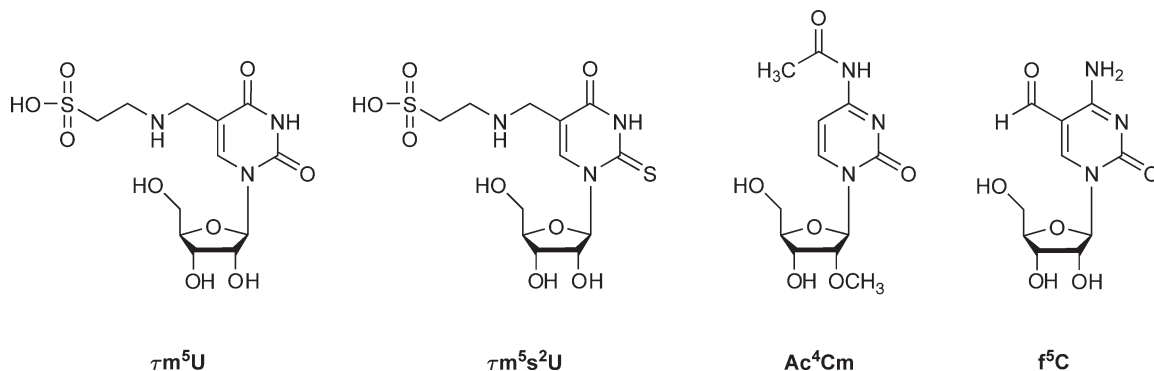


Figure 12 τm^5U , τm^5s^2U , Ac^4Cm , and f^5C structures.

Watanabe and coworkers have investigated the structure and function of human mitochondrial tRNAs containing wobble modifications derived from taurine, the nucleosides 5-taurinomethyluridine (τm^5U) and 5-taurinomethyl-2-thiouridine (τm^5s^2U).^{105,106} Mutations that affect the modification of the tRNA are associated with MERRF disease (myoclonus epilepsy associated with ragged red fibers) or MELAS, a mitochondrial myopathy.¹⁰⁷ Certain mutations known to cause these diseases result in specifically unmodified tRNA^{Lys} or tRNA^{Leu} isoacceptors that fail to properly recognize cognate codons. *In vitro* translation experiments in an *E. coli* system showed that the τm^5U modification was important for stabilizing U–G pairing.¹⁰⁵ In a result that was quite analogous to that seen for cmo^5U described above, crystal structures of a tRNA containing τm^5U bound to either UUA or UUG codons showed that the τm^5U –G pair and the τm^5U –A pair had very similar, slightly distorted Watson–Crick geometries.¹⁰⁵ This suggests that the τm^5U modification may tautomerize to the enol form as seen for cmo^5U ,¹⁰⁴ or it could stabilize the bifurcated geometry seen for mnm^5U .⁹⁹ Regardless of those details, the requirement for τm^5U in mitochondrial tRNA^{Leu} provides further support for the importance of 34–35 stacking stabilization in tRNAs having U34 and either an A–U or an U–A pair in the middle (35) position (Figure 12).

6.19.5.4 Modified Cytidines – Ac^4C , f^5C , Ac^4Cm , and f^5Cm

Cytidine in the tRNA wobble position is often modified in either the sugar or the base with functional groups that promote the 3'-endo sugar conformation at the nucleoside level. Experiments by Takemoto *et al.* suggested that f^5C modification restricts wobble pairing so that only G is recognized by tRNA^{Met}_{f⁵CAU}. However, wobble restriction by f^5C is not absolute since in cytoplasmic bovine tRNA^{Leu}_{f⁵CmAA}, both the UUC and UUA leucine codons might be recognized.^{108,109} Watanabe and coworkers suggest that bovine mitochondrial tRNA^{Met} with f^5C34 may recognize both AUG and AUA as methionine codons, but their base pairing model would require an A⁺–C pair.¹¹⁰ The effects of f^5C and f^5Cm on codon recognition are particularly interesting in light of discussions in previous sections about wobble modification, given the extreme effect that the f^5 modifications have on the nucleoside conformation. The properties of f^5C are similar to that of ac^4C modification in that the 3'-endo conformation is significantly stabilized.^{29,111} The base modifications alone stabilize the 3'-endo conformation, presumably through stereoelectronic effects at the anomeric bond and through space interactions involving the H5–H6 double bond and electron lone pairs on O4'.^{26–30} The effects of f^5C on RNA duplex stability have not been investigated.

The ac^4C nucleoside is found at the wobble position of methionine tRNA from *E. coli* as well as glutamine, glutamate, lysine, proline, and serine tRNAs in archaea.⁵ Kawai *et al.*¹¹² showed that ac^4C strongly favors the 3'-endo conformation, and that this conformational preference is reinforced by simultaneous 2'-OMe to form the ac^4Cm nucleoside found in tRNAs from hyperthermophilic archaea. The presence of N-acetylated cytidine in organisms growing above 100 °C is somewhat remarkable given the hydrolytic lability of this functionality. In light of the dramatic effects that ac^4 and f^5 modifications have on the cytidine conformational preference, these modifications could have applications for RNA duplex stabilization similar to that shown recently for s^2U .⁵⁸

6.19.5.5 The Lysine-Modified Nucleoside, Lysidine

In *E. coli*, and *Bacillus subtilis* the *tilS* gene codes for the enzyme tRNA^{Ile}-lysidine synthetase responsible for forming the lysine-modified nucleoside lysidine (k²C) (**Figure 13**).^{113,114} tRNA^{Ile}_{k²CAU} specifically recognizes the rare AUA codon in a lysidine-dependent fashion;^{115,116} the ‘major’ Ile codons AUU and AUC are read by tRNA^{Ile}_{GAU}.¹¹⁷ This wobble position-modified C nucleoside is an essential modification in bacteria,¹¹⁶ and analogous genes are found in archaea, but apparently not in eukaryotes.¹¹⁸ The fact that lysidine is absent in eukaryotes, and required in bacteria has led researchers to comment on the possibility that the TilS enzyme presents a novel antibacterial target.¹¹⁸

The unmodified CAU anticodon is of course the sequence of tRNA^{Met}, and this raised the question as to how tRNA^{Met}_{CAU} and tRNA^{Ile}_{k²CAU} are distinguished by their relevant aminoacyl tRNA synthetases. For both tRNA^{Met} and tRNA^{Ile}_{k²CAU} the wobble position nucleoside is critical, serving as a positive recognition element for the two cognate synthetases. For tRNA^{Met}, C34 is a positive determinant and k²C34 is a negative determinant; for tRNA^{Ile} the opposite is true.¹¹⁸ Additionally, a small number of residues in the anticodon, the anticodon stem, and in the acceptor stem provide further discrimination against improper charging of the two closely related tRNAs.^{118,119}

The fundamental question about how lysidine alters codon recognition at the wobble position remains to be conclusively answered. The structure and protonation state was determined by Muramatsu *et al.*¹¹³ through NMR analysis and chemical synthesis. The nucleobase is apparently charged at physiological pH, allowing for several alternate structures depending on where one places the proton. The structure in **Figure 13(b)** is an attractive possibility since it allows for Watson–Crick pairing with A, which would maximize stacking between the wobble k²C and the weakly stacked U35. As we have now repeated several times, an emerging theme for modified wobble pairing is for modifications to promote Watson–Crick-like geometries for noncanonical interactions, especially in AU-rich codons (**Figure 14**).

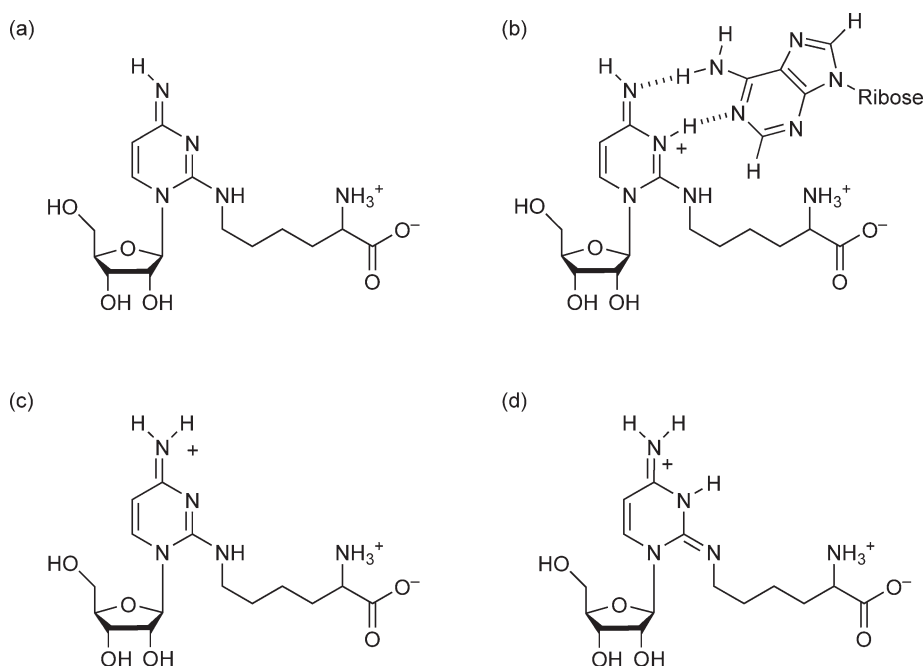


Figure 13 Structures and base pairing of the modified nucleoside lysidine (k²C). (a) The neutral form of lysidine. The lysidine base has a pK_a of approximately 12, and therefore the base would be charged at pH = 7 along with a zwitterionic form for the side chain. (b) Protonated form of lysidine allowing pairing with A in a Watson–Crick geometry. (c) and (d) Alternative protonated forms of lysidine that would be less favorable for A-pairing.

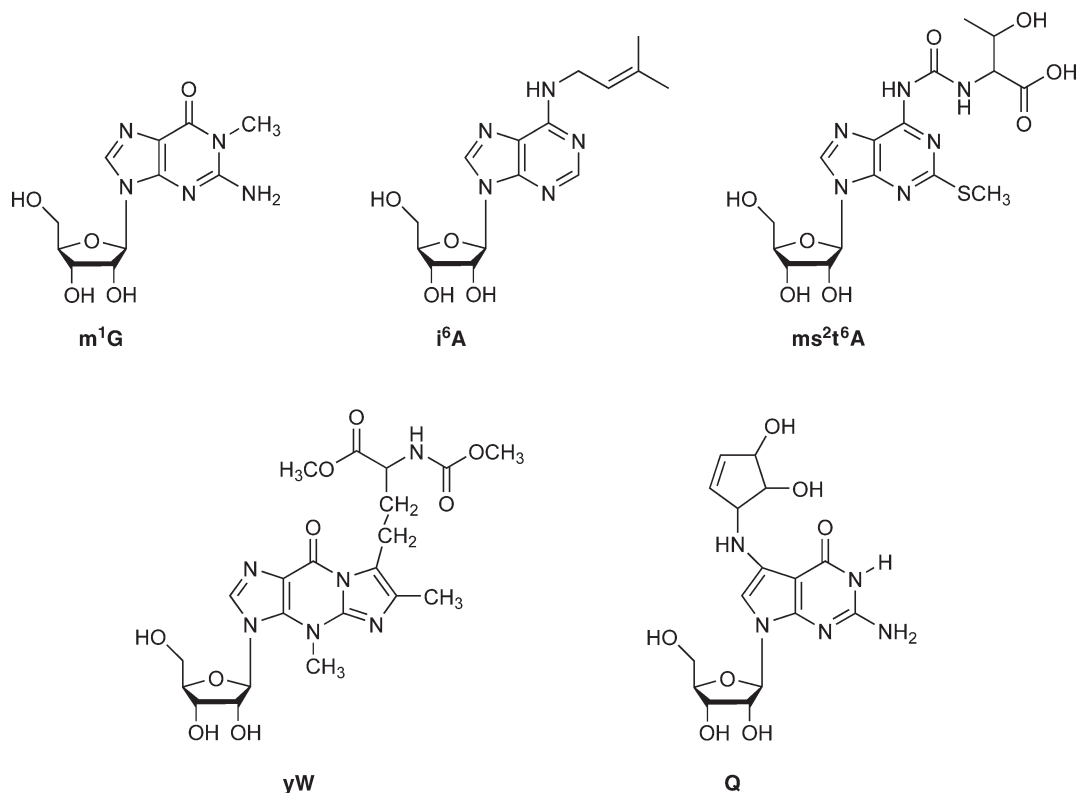


Figure 14 Chemical structures of m¹G, yW, i⁶A, ms²t⁶A, and Q. i⁶A and t⁶A are found both with and without the 2-thiomethyl group that can further increase base stacking with a nucleotide in the first codon position (cross-strand stacking).

6.19.5.6 Purine Modifications – m¹G, i⁶A, t⁶A, yW, and Q

tRNAs have a conserved purine at position 37 adjacent to the anticodon, and this purine is invariably modified (Figure 1). Crystal structures of the ribosome with tRNAs bound show that purine 37 participates in a cross-strand stacking interaction with the first codon nucleotide,⁹⁹ essentially a 3'-terminal stacking interaction, known to provide RNA duplex stability comparable to that of an additional base pair.⁷ Since purines are 'better stackers' than pyrimidines, the conservation of a purine at position 37 is logical, but the universal modification suggests that the position 37 base may be important for more than stacking stabilization, or that modifications may enhance the stacking interaction as well as participate in other structural interactions within the anticodon.⁹⁹ Indeed, NMR studies have shown that m¹G, i⁶A, and t⁶A each function to organize the tRNA anticodon structure to stabilize the appropriate stacked conformation on the 3'-side of the tRNA, and to provide an open-loop structure that minimizes cross-loop base pairing. It appears that the ideal conformation for mRNA binding is one where the position 37 nucleotide promotes 37, 38, and 39 stacking with a discontinuity of stacking with the position 36 anticodon base.^{91,92,120} At the same time, this allows for continuous stacking of 34–36 (for *Saccharomyces pombe* tRNA^{Met}; this may not be true),¹²⁰ and destabilization of pairing between U33 and the purine at position 37. For m¹G, the modification clearly blocks pairing with U33 by blocking the Watson–Crick face of G37, yW would also be blocked from Watson–Crick pairing. For t⁶A37, the conformation observed in the crystal structure, and shown by NMR to be the dominant form in solution destabilizes an across-the-loop base pair with U33.^{91,92,99,121} Nikonowicz and coworkers used NMR to show that i⁶A37 promotes an open-loop conformation in *E. coli* tRNA^{Phe}, and that the combination of i⁶A37 and divalent metals stabilizes the canonical U-turn structure observed in solution for most native tRNAs, and seen in tRNA crystal structures (Figure 15).^{122,123}

The deceptively simple m¹G modification is universally conserved and has been proposed as being a member of the minimal set of genes required for life.^{125,126} Genes for m¹G are found in all three domains of life,

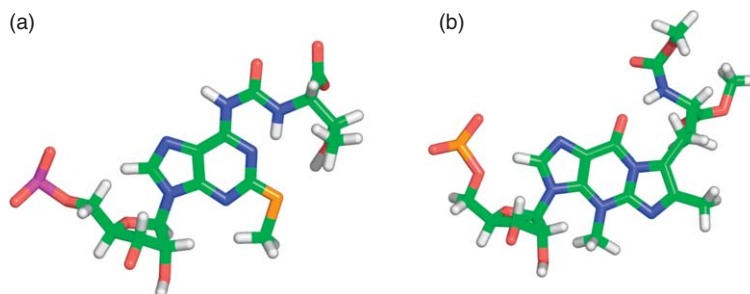


Figure 15 Structure of ms^2t^6A37 from the crystal structure of tRNA^{Lys} (PDB ID: 1XMQ) and the yW nucleoside from tRNA^{Phe} (PDB ID: 1EHZ). The tricyclic ring of yW provides an extensive stacking interface for interaction with the cross-strand codon base. The carbamoyl threonine side chain of t^6A likewise is planar with the adenosine base, a conformation stabilized by a hydrogen bond from the amide to the N1-nitrogen of A. This provides an extensive stacking interface as seen in the crystal structure of tRNA^{Lys} bound to the 30S ribosome.^{99,124}

including the genome of *Mycoplasma genitalium*, the organism with the smallest sequenced genome. Bjork *et al.*¹²⁶ demonstrated that m^1G37 in tRNA is required for normal growth in *E. coli* and in *Saccharomyces cerevisiae*.¹²⁶ Their studies indicate that m^1G plays a crucial role in reading frame maintenance by modulating the strength and therefore the fidelity of tRNA/mRNA pairing. Their genetic and biochemical results are consistent with the model of Agris where m^1G helps maintain an open-loop structure for tRNAs.^{101,127}

The hypermodified nucleosides yW and Q are notable in the extensive additional functionality present, and in the case of Q the fact that the 7-deaza base is inserted rather than modified in place.¹²⁸ Both of these nucleosides influence reading frame maintenance, particularly in cases of +1 frameshifting.^{63,129} Bjork and coworkers interpret the effects on +1 frameshifting, as opposed to little or no effect on -1 frameshifting, in the context of the simultaneous slippage model.¹²⁹ Both these modifications affect the affinity of tRNA for their cognate codons, although the Bjork model requires that these effects be manifested in the P-site where less is known regarding the structural effects of tRNA modification. Both Q and yW modification affect the frequency of frameshifting in retroviruses. The effect of yW is directly on tRNA^{Phe}, while the effect of Q undermodification is seen if the AAU codon read by Q-containing tRNA^{Asn} is immediately upstream of the shift site.^{130,131} The influence of yW can be interpreted in the context of base stacking as has been described for other position 37 modifications. To explain the effect of Q modification, one is inclined to view this similarly to the results for modified uridines in the wobble position. Q-containing tRNAs that pair with U are more sensitive to modification than those that pair with C. That is, both Q34 and G34 interact efficiently with C, but Q modification improves pairing with U in the A-site, decreasing +1 frameshifting. At present we do not know the details of the base pairing interaction, but the effect of Q modification is reminiscent of the changes seen for modified U wobble nucleosides.

6.19.6 Summary and Future Prospects

RNA biochemists studying nucleoside modifications have been plagued by the question, ‘what do they all do?’ Only recently has it been demonstrated that RNA modifications can be necessary for cell viability.⁹³ Gene sequences and more powerful genetic tools are now allowing researchers to identify modification genes and then knock out RNA modifications singly, or in combination with other genetic lesions.^{132,133} Sophisticated kinetic experiments have also shown that tRNA and rRNA modification deletions can have major effects on the fidelity of protein synthesis.¹³⁴ These approaches and the expanded appreciation for the role of RNA in cellular function will certainly lead to the identification of new roles for RNA modification. The recent characterization of inosine modification in micro RNAs is one example of the convergence of nucleoside modification and the new, expanded RNA world.^{135,136} Modification enzymes may also turn out to be important therapeutic targets. The lysidine-modified nucleoside is necessary for growth in bacteria, but is not found in eukaryotes.¹³⁷ This suggests that therapeutics inhibiting enzymes in the lysidine pathway might be effective antibiotics with few side effects.

The s^2U modification in the wobble position of tRNA has one of the greatest effects on RNA stability of any single modification. The large stabilization provided by this nucleoside and the strong A–U versus G–U discrimination of s^2U suggest the potential for application in the area of siRNA therapeutics. Nawrot and coworkers found that in 5–7 s^2U modifications in an siRNA duplex raised the T_m from near 80 °C in 15, 17, and 19 bp duplexes to over 95 °C for the modified duplexes.⁵⁸ This modification may see increased application in the siRNA field since the modifications are tolerated by the siRNA machinery,⁵⁸ and may modulate binding to other cellular sites, such as toll-like receptors.¹³⁸

Other base-modified nucleosides also have tremendous effects at the nucleoside level, although these nucleosides have not been investigated at the RNA duplex level. It would seem that f^5C and ac^4C as well as their 2'-OMe derivatives have the potential to significantly stabilize therapeutic RNAs such as miRNAs and siRNAs. The ac^4C and ac^4Cm nucleosides are overrepresented in rRNAs from hyperthermophilic archaea, suggesting the possibility for a major RNA stabilization effect.⁴⁵ This is amazing in light of the hydrolytic susceptibility of the N4-acetyl group. An investigation of these and other nucleoside analogues from thermophiles may lead to novel applications in the area of therapeutic RNAs where the thermal stability and nuclease resistance are foremost among the challenges. Nucleoside modification is nature's elegant solution to these problems; we simply need to understand what nature has done, and then create innovative applications of this toolbox.

References

1. J. Rozenski; P. F. Crain; J. A. McCloskey, *Nucleic Acids Res.* **1999**, *27*, 196–197.
2. G. R. Bjork, Biosynthesis and Function of Modified Nucleosides. In *tRNA: Structure, Biosynthesis and Function*; D. Soll, U. L. RajBhandary, Eds.; ASM Press: Washington, DC, 1995; pp 165–205.
3. W. A. Decatur; M. J. Fournier, *Trends Biochem. Sci.* **2002**, *27*, 344–351.
4. J. P. Bachellerie; J. Cavaille, *Trends Biochem. Sci.* **1997**, *22*, 257–261.
5. M. Sprinzl; K. S. Vassilenko, *Nucleic Acids Res.* **2005**, *33*, D139–D140.
6. J. D. Puglisi; I. Tinoco, *Methods Enzymol.* **1989**, *180*, 304–325.
7. M. J. Serra; D. H. Turner; S. M. Freier, *Methods Enzymol.* **1995**, *259*, 243–261.
8. T. Xia; D. H. Mathews; D. H. Turner, Thermodynamics of RNA Secondary Structure Formation. In *Comprehensive Natural Products Chemistry*; D. Soll, S. Nishimura, P. B. Moore, Eds.; Elsevier: Amsterdam, 1999; pp 21–47.
9. I. Tinoco, Jr.; O. C. Uhlenbeck; M. D. Levine, *Nature* **1971**, *230*, 362–367.
10. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag: New York, 1984.
11. C. H. Lee; I. Tinoco, *Biochemistry* **1977**, *16*, 5403–5414.
12. C. H. Lee; I. Tinoco, *Biophys. Chem.* **1980**, *11*, 283–294.
13. H. Inoue; Y. Hayase; A. Imura; S. Iwai; K. Miura; E. Ohtsuka, *Nucleic Acids Res.* **1987**, *15*, 6131–6148.
14. G. Kawai; U. Harumi; M. Yasuda; K. Sakamoto; T. Hashizume; J. McCloskey; T. Miyazawa; S. Yokoyama, *Nucleic Acids Res.* **1991**, *25*, 49–50.
15. G. Kawai; Y. Yamamoto; T. Kamimura; T. Masegi; M. Sekine; T. Hata, *Biochemistry* **1992**, *31*, 1040–1046.
16. C. Lee; I. Tinoco, *Biophys. Chem.* **1980**, *11*, 283–294.
17. C. H. Lee; I. Tinoco, Jr., *Biochemistry* **1977**, *16*, 5403–5414.
18. W. K. Olson, *J. Am. Chem. Soc.* **1982**, *104*, 278–286.
19. S. Uesugi; H. Miki; M. Ikehara; H. Iwahashi; Y. Kyogoku, *Tetrahedron Lett.* **1979**, *42*, 4073–4976.
20. W. Guschlbauer; K. Jankowski, *Nucleic Acids Res.* **1980**, *8*, 1421–1433.
21. R. K. Kumar; D. R. Davis, *Nucleic Acids Res.* **1997**, *25*, 1272–1280.
22. H. Sierzputowska-Gracz; E. Sochacka; A. Malkiewicz; K. Kuo; C. Gehrke; P. Agris, *J. Am. Chem. Soc.* **1987**, *109*, 7171–7177.
23. S. Yokoyama; Z. Yamaizumi; S. Nishimura; T. Miyazawa, *Nucleic Acids Res.* **1979**, *6*, 2611–2626.
24. R. K. Kumar; D. R. Davis, *Nucleosides Nucleotides* **1997**, *16*, 1469–1472.
25. L. C. Sowers; B. R. Shawk; W. D. Sedwick, *Biochem. Biophys. Res. Commun.* **1987**, *148*, 790–794.
26. E. Juaristi; G. Cuevas, *The Anomeric Effect*, CRC Press: Boca Raton, FL, 1995.
27. E. Egert; H. J. Lindner; W. Hillen; M. C. Bohm, *J. Am. Chem. Soc.*, **1980**, *102*, 3707–3713.
28. W. Uhl; J. Reiner; H. G. Gassen, *Nucleic Acids Res.* **1983**, *11*, 1167–1180.
29. G. Kawai; T. Hashizume; M. Yasuda; T. Miyazawa; J. McCloskey; S. Yokoyama, *Nucleosides Nucleotides* **1992**, *11*, 759–771.
30. D. R. Davis, Biophysical and Conformational Properties of Modified Nucleosides in RNA. In *Modification and Editing of RNA: The Alteration of RNA Structure and Function*; H. Grosjean, R. Benne, Eds.; ASM Press: New York, 1998; pp 85–102.
31. P. Agris, *Biochimie* **1991**, *73*, 1345–1349.
32. S. Yokoyama; T. Watanabe; K. Murao; H. Ishikura; Z. Yamaizumi; S. Nishimura; T. Miyazawa, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4905–4909.
33. P. Dande; T. P. Prakash; N. Sioufi; H. Gaus; R. Jarres; A. Berdeja; E. E. Swayze; R. H. Griffey; B. Bhat, *J. Med. Chem.* **2006**, *49*, 1624–1634.
34. A. L. Jackson; J. Burchard; D. Leake; A. Reynolds; J. Scheluter; J. Guo; J. M. Johnson; L. Lim; J. Karpilow; K. Nichols; W. Marshall; A. Khvorova; P. S. Linsley, *RNA* **2006**, *12*, 1197–1205.

35. T. P. Prakash; C. R. Allerson; P. Dande; T. A. Vickers; N. Sioufi; R. Jarres; B. F. Baker; E. E. Swayze; R. H. Griffey; B. Bhat, *J. Med. Chem.* **2005**, *48*, 4247–4253.
36. C. R. Allerson; N. Sioufi; R. Jarres; T. P. Prakash; N. Naik; A. Berdeja; L. Wanders; R. H. Griffey; E. E. Swayze, B. Bhat, *J. Med. Chem.* **2005**, *48*, 901–904.
37. T. P. Prakash; B. Bhat, *Curr. Top. Med. Chem.* **2007**, *7*, 641–649.
38. T. P. Prakash; J. F. Johnston; M. J. Graham; T. P. Condon; M. Manoharan, *Nucleic Acids Res.* **2004**, *32*, 828–833.
39. Y. L. Chiu; T. M. Rana, *RNA* **2003**, *9*, 1034–1048.
40. B. A. Kraynack; B. F. Baker, *RNA* **2006**, *12*, 163–176.
41. B. P. Monia; J. F. Johnston; H. Sasnor; L. L. Cummins, *J. Biol. Chem.* **1996**, *271*, 14533–14540.
42. J. Kowalak; J. Dalluge; J. McCloskey; K. Stetter, *Biochemistry* **1994**, *33*, 7869–7876.
43. K. Watanabe; T. Oshima; K. Iijima; Z. Yamaizumi; S. Nishimura, *J. Biochem.* **1980**, *87*, 1–13.
44. K. Watanabe; T. Oshima; S. Nishimura, *Nucleic Acids Res.* **1976**, *3*, 1703–1713.
45. J. A. McCloskey; J. Rozenski, *Nucleic Acids Res.* **2005**, *33*, D135–D138.
46. B. E. Maden, *Prog. Nucleic Acid Res. Mol. Biol.* **1990**, *39*, 241–303.
47. K. R. Noon; E. Bruenger; J. A. McCloskey, *J. Bacteriol.* **1998**, *180*, 2883–2888.
48. J. A. McCloskey; S. Nishimura, *Acc. Chem. Res.* **1977**, *10*, 403–410.
49. J. R. Patton; R. W. Padgett, *BMC Mol. Biol.* **2005**, *6*, 20.
50. J. Ofengand; A. Bakin, *J. Mol. Biol.* **1997**, *266*, 246–268.
51. B. E. Maden, *J. Mol. Biol.* **1988**, *201*, 289–314.
52. B. G. Lane; J. Ofengand; M. W. Gray, *Biochimie* **1995**, *77*, 7–15.
53. D. R. Davis; C. D. Poulter, *Biochemistry* **1991**, *30*, 4223–4231.
54. J. G. Arnez; T. A. Steitz, *Biochemistry* **1994**, *31*, 7560–7567.
55. D. R. Davis, *Nucleic Acids Res.* **1995**, *23*, 5020–5026.
56. D. R. Davis; C. A. Veltri; L. Nielsen, *J. Biomol. Struct. Dyn.* **1998**, *15*, 1121–1132.
57. P. C. Durant; D. R. Davis, *J. Mol. Biol.* **1999**, *285*, 115–131.
58. K. Sipa; E. Sochacka; J. Kazmierczak-Baranska; M. Maszewska; M. Janicka; G. Nowak; B. Nawrot, *RNA* **2007**, *13*, 1301–1316.
59. Y. Lin; C. L. Kielkopf, *Biochemistry* **2008**, *47*, 5503–5514.
60. H. O. Kammen; C. C. Marvel; L. Hardy; E. E. Penhoet, *J. Biol. Chem.* **1988**, *263*, 2255–2263.
61. F. Leconte; G. Somos; A. Sauer; E. C. Hurt; Y. Motorin; H. Grosjean, *J. Biol. Chem.* **1998**, *273*, 1316–1323.
62. C. L. Turnbough, Jr.; R. J. Neill; R. Landsberg; B. N. Ames, *J. Biol. Chem.* **1979**, *254*, 5111–5119.
63. J. Urbonavicius; Q. Qian; J. M. B. Durand; T. G. Hagervall; G. R. Bjork, *EMBO J.* **2001**, *20*, 4863–4873.
64. J. R. Roth; D. N. Anton; P. E. Hartman, *J. Mol. Biol.* **1966**, *22*, 305–323.
65. D. T. Palmer; P. H. Blum; S. W. Artz, *J. Bacteriol.* **1983**, *153*, 357–363.
66. S. Raychaudhuri; L. Niu; J. Conrad; B. G. Lane; J. Ofengand, *J. Biol. Chem.* **1999**, *274*, 18880–18886.
67. I. Tworowska; E. P. Nikonowicz, *J. Am. Chem. Soc.* **2006**, *128*, 15570–15571.
68. Y. T. Yu; E. C. Scharl; C. M. Smith; J. A. Steitz, The Growing World of Small Nuclear Ribonucleoproteins. In *The RNA World*, 2nd ed.; R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds.; CSHL Press: Cold Spring Harbor, NY, 1999; pp 487–524.
69. S. Massenet; A. Mougin; C. Branlant, Posttranscriptional Modifications in the U Small Nuclear RNAs. In *Modification and Editing of RNA*; H. Grosjean, R. Benne, Eds.; ASM Press: Washington, DC, 1998; pp 201–227.
70. M. I. Newby; N. L. Greenbaum, *RNA* **2001**, *7*, 833–845.
71. X. Zhao; Y. T. Yu, *RNA* **2004**, *10*, 681–690.
72. C. Yang; D. S. McPheeters; Y. T. Yu, *J. Biol. Chem.* **2005**, *280*, 6655–6662.
73. M. I. Newby; N. L. Greenbaum, *Nat. Struct. Biol.* **2002**, *9*, 958–965.
74. C. C. Query; M. J. Moore; P. A. Sharp, *Genes Dev.* **1994**, *8*, 587–597.
75. A. Liiv; D. Karitkina; U. Maivali; J. Remme, *BMC Mol. Biol.* **2005**, *6*, 18.
76. J. A. Kowalak; E. Bruenger; T. Hashizume; J. Peltier; J. Ofengand; J. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 688–693.
77. B. S. Schuwirth; M. A. Borovinskaya; C. W. Hau; W. Zhang; A. Vila-Sanjurjo; J. M. Holton; J. H. Cate, *Science* **2005**, *310*, 827–834.
78. S. C. Abeysirigunawardena; C. S. Chow, *RNA* **2008**, *14*, 782–792.
79. U. Maivali; J. Remme, *RNA* **2004**, *10*, 600–604.
80. R. K. Agrawal; M. R. Sharma; M. C. Kiel; G. Hirokawa; T. M. Booth; C. M. Spahn; R. A. Grassucci; A. Kaji; J. Frank, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8900–8905.
81. M. Laurberg; H. Asahara; A. Korostelev; J. Zhu; S. Trakhanov; H. F. Noller, *Nature* **2008**, *454*, 852–857.
82. I. K. Ali; L. Lancaster; J. Feinberg; S. Joseph; H. F. Noller, *Mol. Cell* **2006**, *23*, 865–874.
83. N. S. Gutsell; M. P. Deutscher; J. Ofengand, *RNA* **2005**, *11*, 1141–1152.
84. M. Sumita; J. P. Desaulniers; Y. C. Chang; H. M. Chui; L. Clos, II; C. S. Chow, *RNA* **2005**, *11*, 1420–1429.
85. M. Meroueh; P. J. Grohar; J. Qiu; J. SantaLucia, Jr.; S. A. Scaringe; C. S. Chow, *Nucleic Acids Res.* **2000**, *28*, 2075–2083.
86. S. Yokoyama; S. Nishimura, Modified Nucleosides and Codon Recognition. In *tRNA: Structure, Biosynthesis and Function*; D. Soll, U. RajBhandary, Eds.; ASM Press: Washington, DC, 1995; pp 207–224.
87. M. Martinez-Vicente; L. Yim; M. Villarroya; M. Mellado; E. Perez-Paya; G. R. Bjork; M. E. Armengod, *J. Biol. Chem.* **2005**, *280*, 30660–30670.
88. C. Yarian; M. Marszalek; E. Sochacka; A. Malkiewicz; R. Guenther; A. Miskiewicz; and P. F. Agris, *Biochemistry* **2000**, *39*, 13390–13395.
89. U. Von Ahlsen; R. Green; R. Schroeder; H. Noller, *RNA* **1997**, *3*, 49–56.
90. T. G. Hagervall; S. C. Pomerantz; J. A. McCloskey, *J. Mol. Biol.* **1998**, *284*, 33–42.
91. P. C. Durant; A. C. Bajji; M. Sundaram; R. K. Kumar; D. R. Davis, *Biochemistry* **2005**, *44*, 8078–8089.
92. M. Sundaram; P. C. Durant; D. R. Davis, *Biochemistry* **2000**, *39*, 12575–12584.
93. G. R. Bjork; B. Huang; O. P. Persson; A. S. Bystrom, *RNA* **2007**, *13*, 1245–1255.
94. A. Bajji; D. R. Davis, *Org. Lett.* **2000**, *2*, 3865–3868.

95. A. Bajji; M. Sundaram; D. G. Myszkka; D. R. Davis, *J. Am. Chem. Soc.* **2002**, *124*, 14302–14303.
96. S. Ashraf; E. Sochacka; R. Cain; R. Guenther; A. Malkiewicz; P. Agris, *RNA* **1999**, *5*, 188–194.
97. K. Takai, *Nucleic Acids Symp. Ser. (Oxf)* **2005**, *49*, 317–318.
98. K. Takai; S. Yokoyama, *Nucleic Acids Res.* **2003**, *31*, 6383–6391.
99. F. V. Murphy; V. Ramakrishnan; A. Malkiewicz; P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186–1191.
100. P. F. Agris, *EMBO Rep.* **2008**, *9*, 629–635.
101. E. M. Gustilo; F. A. Vendeix; P. F. Agris, *Curr. Opin. Microbiol.* **2008**, *11*, 134–140.
102. S. K. Mitra; F. Lustig; B. Akesson; T. Axberg; P. Elias; U. Lagerkvist, *J. Biol. Chem.* **1979**, *254*, 6397–6401.
103. F. A. Vendeix; A. Dziergowska; E. M. Gustilo; W. D. Graham; B. Sproat; A. Malkiewicz; P. F. Agris, *Biochemistry* **2008**, *47*, 6117–6129.
104. A. Weixlbaumer; F. V. t. Murphy; A. Dziergowska; A. Malkiewicz; F. A. Vendeix; P. F. Agris; V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14*, 498–502.
105. S. Kurata; A. Weixlbaumer; T. Ohtsuki; T. Shimazaki; T. Wada; Y. Kirino; K. Takai; K. Watanabe; V. Ramakrishnan; T. Suzuki, *J. Biol. Chem.* **2008**, *283*, 18801–18811.
106. T. Suzuki; T. Suzuki; T. Wada; K. Saigo; K. Watanabe, *EMBO J.* **2002**, *21*, 6581–6589.
107. Y. Kirino; T. Suzuki, *RNA Biol.* **2005**, *2*, 41–44.
108. J. P. Pais de Barros; G. Keith; C. El Adlouni; A. L. Glasser; G. Mack; G. Dirheimer; J. Desgres, *Nucleic Acids Res.* **1996**, *24*, 1489–1496.
109. C. Takemoto; T. Koike; T. Yokogawa; L. Benkowski; L. Spremulli; T. Ueda; K. Nishikawa; K. Watanabe, *Biochimie* **1995**, *77*, 104–108.
110. J. Moriya; T. Yokogawa; K. Wakita; T. Ueda; K. Nishikawa; P. Crain; T. Hashizume; S. Pomerantz; J. McCloskey; G. Kawai; N. Hayashi; S. Yokoyama; K. Watanabe, *Biochemistry* **1994**, *33*, 2234–2239.
111. G. Kawai; T. Yokogawa; K. Nishikawa; T. Ueda; T. Hashizume; J. McCloskey; S. Yokoyama; K. Watanabe, *Nucleosides Nucleotides* **1994**, *13*, 1189–1199.
112. C. Edmonds; P. Crain; R. Gupta; T. Hashizume; C. Hocart; J. Kowalak; S. Pomerantz; K. Stetter; J. McCloskey, *J. Bacteriol.* **1991**, *173*, 3138–3148.
113. T. Muramatsu; S. Yokoyama; N. Horie; A. Matsuda; T. Ueda; Z. Yamaizumi; Y. Kuchino; S. Nishimura; T. Miyazawa, *J. Biol. Chem.* **1988**, *263*, 9261–9267.
114. T. Muramatsu; K. Nishikawa; F. Nemoto; Y. Kuchino; S. Nishimura; T. Miyazawa; S. Yokoyama, *Nature* **1988**, *336*, 179–181.
115. F. Harada; S. Nishimura, *Biochemistry* **1974**, *13*, 300–307.
116. A. Soma; Y. Ikeuchi; S. Kanemasa; K. Kobayashi; N. Ogasawara; T. Ote; J. Kato; K. Watanabe; Y. Sekine; T. Suzuki, *Mol. Cell* **2003**, *12*, 689–698.
117. M. Yarus; B. G. Barrell, *Biochem. Biophys. Res. Commun.* **1971**, *43*, 729–734.
118. Y. Ikeuchi; A. Soma; T. Ote; J. Kato; Y. Sekine; T. Suzuki, *Mol. Cell* **2005**, *19*, 235–246.
119. K. Nakanishi; S. Fukai; Y. Ikeuchi; A. Soma; Y. Sekine; T. Suzuki; O. Nureki; *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7487–7492.
120. E. Lescrinier; K. Nauwelaerts; K. Zanier; K. Poesen; M. Sattler; P. Herdewijn, *Nucleic Acids Res.* **2006**, *34*, 2878–2886.
121. J. W. Stuart; Z. Gdaniec; R. H. Guenther; M. Marszalek; E. Sochacka; A. Malkiewicz; P. F. Agris, *Biochemistry* **2000**, *39*, 13396–13404.
122. J. Cabello-Villegas; I. Tworowska; E. P. Nikonowicz, *Biochemistry* **2004**, *43*, 55–66.
123. J. Cabello-Villegas; W. C. Winkler; E. P. Nikonowicz, *J. Mol. Biol.* **2002**, *319*, 1015–1034.
124. H. Shi; P. B. Moore, *RNA* **2000**, *6*, 1091–1105.
125. G. R. Björk; P. M. Wikstrom; A. S. Bystrom, *Science* **1989**, *244*, 986–989.
126. G. R. Bjork; K. Jacobsson; K. Nilsson; M. J. Johansson; A. S. Bystrom; O. P. Persson, *EMBO J.* **2001**, *20*, 231–239.
127. J. W. Stuart; K. M. Koshlap; R. Guenther; P. F. Agris, *J. Mol. Biol.* **2003**, *334*, 901–918.
128. N. Okada; S. Noguchi; H. Kasai; N. Shindo-Okada; T. Ohgi; T. Goto; S. Nishimura, *J. Biol. Chem.* **1979**, *254*, 3067–3073.
129. J. Urbonavicius; G. Stahl; J. M. Durand; S. N. Ben Salem; Q. Qian; P. J. Farabaugh; G. R. Bjork, *RNA* **2003**, *9*, 760–768.
130. B. A. Carlson; S. Y. Kwon; M. Chamorro; S. Oroszlan; D. L. Hatfield; B. J. Lee, *Virology* **1999**, *255*, 2–8.
131. B. A. Carlson; J. F. Mushinski; D. W. Henderson; S. Y. Kwon; P. F. Crain; B. J. Lee; D. L. Hatfield, *Virology* **2001**, *279*, 130–135.
132. B. Huang; J. Lu; A. S. Bystrom, *RNA* **2008**, *14*, 2183–2194.
133. D. Piekna-Przybylska; P. Przybylski; A. Baudin-Baillieu; J. P. Rousset; M. J. Fournier, *J. Biol. Chem.* **2008**, *283*, 26026–26036.
134. U. Kothe; M. V. Rodnina, *Mol. Cell* **2007**, *25*, 167–174.
135. J. W. Habig; T. Dale; B. L. Bass, *Mol. Cell* **2007**, *25*, 792–793.
136. Y. Kawahara; B. Zinshteyn; P. Sethupathy; H. Iizasa; A. G. Hatzigeorgiou; K. Nishikura, *Science* **2007**, *315*, 1137–1140.
137. H. Grosjean; G. R. Bjork, *Trends Biochem. Sci.* **2004**, *29*, 165–168.
138. K. Kariko; D. Weissman, *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 523–532.

Biographical Sketch



Darrell R. Davis received his B.S. (1982) in chemistry from the University of Puget Sound and a Ph.D. (1988) in organic chemistry from the University of Utah under the direction of C. Dale Poulter. From 1988 to 1989 he was an NIH postdoctoral fellow at the University of Washington, Department of Chemistry, in the laboratory of Brian R. Reid. In 1989 he was appointed to the Department of Medicinal Chemistry in the Departments of Medicinal Chemistry and of Biochemistry at The University of Utah where he currently holds the rank of professor. He has authored and coauthored approximately 50 publications. Current research in the Davis' laboratory involves the use of NMR spectroscopy to investigate the structure and function of RNA, particularly RNA molecules involved in viral replication.

6.20 RNA Modifying Enzymes

George A. Garcia, Julie K. Hurt, and Yi-Chen Chen, University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|--|-----|
| 6.20.1 | Introduction | 683 |
| 6.20.2 | Methylation | 684 |
| 6.20.2.1 | Methylation on Carbon | 684 |
| 6.20.2.2 | Methylation on Tertiary Nitrogen | 687 |
| 6.20.2.3 | Methylation on Primary Nitrogen | 690 |
| 6.20.2.4 | Methylation on Oxygen | 693 |
| 6.20.3 | Thiolation and Selenation | 696 |
| 6.20.3.1 | Thiolation | 696 |
| 6.20.3.2 | Selenation | 699 |
| 6.20.4 | Deamination | 700 |
| 6.20.4.1 | Adenosine Deamination (A to I) | 701 |
| 6.20.4.2 | Cytidine Deamination (C to U) | 704 |
| 6.20.5 | Alkylation | 705 |
| 6.20.5.1 | Dimethylallylation | 705 |
| 6.20.5.2 | Lysidine Formation | 708 |
| 6.20.6 | Reduction/Oxidation | 709 |
| 6.20.7 | Transglycosylation | 712 |
| 6.20.7.1 | tRNA-Guanine Transglycosylase | 712 |
| 6.20.7.2 | Pseudouridine Synthase | 718 |
| 6.20.7.3 | Other Base Modifications via Transglycosylation? | 723 |
| 6.20.8 | Complex Modifications | 723 |
| 6.20.8.1 | Queuosine and Derivatives | 723 |
| 6.20.8.2 | Wyosine and Derivatives | 727 |
| 6.20.9 | Conclusions and Perspectives | 731 |
| References | | 734 |

6.20.1 Introduction

It is widely known that nucleic acids consist of linear oligomers of the canonical nucleosides: adenosine, guanosine, cytosine, thymidine, and uridine; joined by phosphodiester linkages. What is much less appreciated is the fact that there are over 100 chemically distinct, modified nucleosides that occur in nucleic acids. The vast majority of these occur in RNA and about 80% of those occur in transfer RNA (tRNA). These modifications are widely believed to modulate (fine tune) RNA structure and function. The modification of RNA has been known since the discovery of Pseudouridine (Ψ , pseudo U) in 1957.¹ However, it was more recently discovered that certain mRNAs are subject to post-transcriptional modification that changes one canonical base into another (or its equivalent).^{2,3} This has been termed RNA editing and has required a paradigm shift in the dogma that the DNA sequence of a gene uniquely determined the sequence of the resulting protein. RNA modification and editing (enzymology, occurrence, and function) has been the subject of recent reviews and books.⁴⁻⁷

Thus far, all modifications of RNA are generated post-transcriptionally via the actions of enzymes that recognize the base in question in the context of its RNA. The chemical nature of RNA modifications varies from the relatively simple actions of single enzymes (e.g., methylations) to the very complex (e.g., queuine), which require elaborate, multienzyme pathways for their biosynthesis. According to estimates, 5% of the *Escherichia coli* genome codes for RNA modifying enzymes.⁷ The RNA modifications can be categorized by the nature of the chemical reaction that generates the modification (**Table 1**). In this chapter, we use this categorization to present a review of the enzymes involved in generating these natural products. Our focus

Table 1 RNA modifications categorized by the nature of the chemical transformation

| Type of reaction | Cofactors/Cosubstrates | Examples |
|-------------------------|--|---|
| Methylation | AdoMet (in one case folate) | m ⁵ U (rT), m ⁵ C, m ¹ A, m ¹ G, m ⁷ G, m ¹ I, m ⁶ A, m ² G,... |
| Thiolation (selenation) | Cys or β-mercaptopyruvate? | s ² U, s ⁴ U, ms ² A, s ² T,... |
| Deamination | H ₂ O | A to I, C to U,... |
| Alkylation | DMAPP, Thr, Lys, PRPP | i ⁶ A, t ⁶ A, L, pRibp,... |
| Reduction/oxidation | Flavin, NADH, NADPH | D, ms ² io ⁶ A, oY,... |
| Transglycosylation | Queuine, PreQ ₁ , PreQ ₀ | Q, G*, Ψ |
| Others | Multiple | Y,... |

will be on the chemical and mechanistic aspects of these transformations. However, recognition of the RNA and specific nucleoside to be modified is a critical factor for RNA modification. Therefore, we also discuss RNA recognition where it is particularly important.

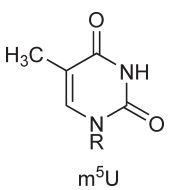
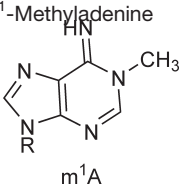
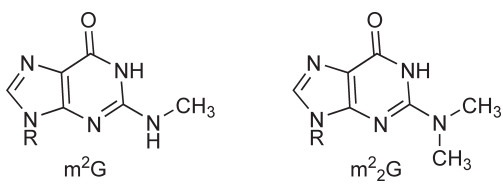
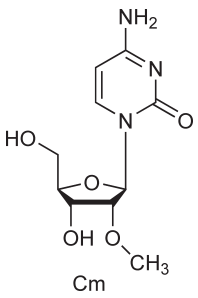
6.20.2 Methylation

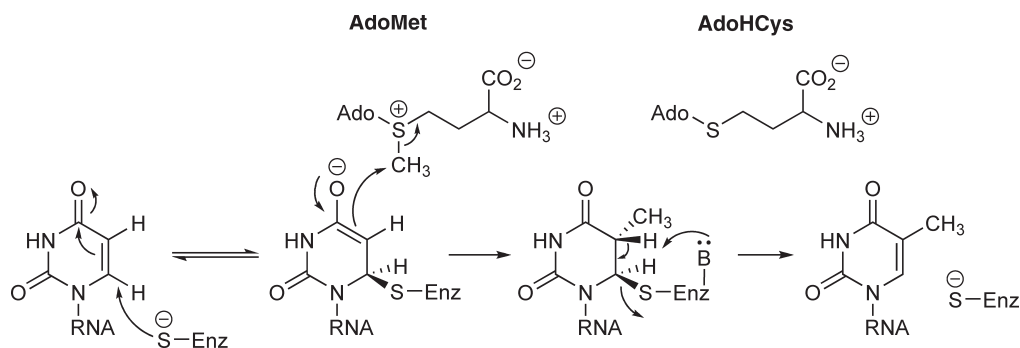
One of the most abundant, and relatively simple, modifications found in nucleic acids is the addition of a methyl group, almost exclusively from *S*-adenosylmethionine (AdoMet), to a heteroatom or carbon on either the base itself or the ribose sugar. With a variety of sites available for modification, including methylation of carbon, nitrogen, and oxygen, a diverse group of methylated nucleosides are found in nature. **Table 2** illustrates several structural examples of each type of methylation. Among the modifications, the most interesting one is that of methylated carbon, with the reaction mechanism and kinetics best characterized for the enzyme tRNA (m⁵U54)-methyltransferase (RUMT).⁸ The mechanism of this enzyme was reviewed in some detail by Garcia and Goodenough-Lashua, and is briefly discussed in this work.⁹ Historically, methyltransferases (or MTases) are structurally characterized by multidomains involved in substrate recognition. Although most of the MTases have been identified and characterized some time ago, we are gaining increased knowledge of the active site and key residues involved in interaction with substrates in many cases through crystallographic analyses. Accordingly, this structural insight will serve as the focus for this section.

6.20.2.1 Methylation on Carbon

RUMT, which catalyzes the modification m⁵U54 in tRNA, is probably the most fully characterized RNA MTase. One interesting aspect of methylation on carbon is how the enzymes activate the carbon for nucleophilic attack on AdoMet. It is presumed (reasonably) that all RNA MTases carry out the direct displacement of methyl from AdoMet. The widely accepted mechanism for RUMT is very similar to that for thymidylate synthase, where an enzymic cysteine attacks C⁶, facilitating C⁵ attack upon AdoMet (**Figure 1**). Santi and coworkers have continued their characterization of the RUMT reaction in recent years by addressing the question of accessibility of U54 in the commonly modified T-loop/-arm of tRNA (TΨC).¹⁰ Because the nucleoside is buried in the tertiary structure of tRNA between the interacting D- and TΨC-loops, the authors hypothesized that the enzyme must interrupt the interaction and thereby expose U54 for methylation. An ensemble of solution nuclear magnetic resonance (NMR) structures were analyzed using a T-arm minihelix consisting of 17 nucleotides (corresponding to positions 49–65 in substrate tRNA) and resolved with molecular dynamics. It was shown that there is a fair degree of flexibility observed in the individual residues, including U54. Experiments have also been conducted to help elucidate the structural requirements for RUMT RNA substrates in relation to other modified bases.¹¹ Characterization of hairpin substrates corresponding again to the TΨC stem and loop with pseudoU at the position corresponding to 55 was a substrate for RUMT comparable to the unmodified hairpin. However, other alterations to U54 and U55, including deoxyuridine at each position and methylated pseudoU 55, resulted in structures not active as substrates for RUMT.

Table 2 Structural examples of methylated nucleosides

| Target atom for methylation | Modification name | Example structure |
|-----------------------------|--|---|
| Carbon | 5-Methyluracil |  m ⁵ U |
| Tertiary nitrogen | N ¹ -Methyladenine |  m ¹ A |
| Primary nitrogen | N ² -Methylguanine (mono- and dimethyl shown) |  m ² G m ² ₂ G |
| Oxygen | 2'-O-Methylcytosine |  Cm |

**Figure 1** Catalytic mechanism for m⁵U methyltransferases. This mechanism was proposed by Kealey *et al.* Reproduced from J. T. Kealey; X. Gu; D. V. Santi, *Biochimie* **1994**, 76, 1133.

Although traditionally found in tRNA, m⁵U is also seen in ribosomal RNA (rRNA). The enzyme Ruma catalyzes the site-specific methylation of U1939 in rRNA, and is distinguished from enzymes in the RUMT family in that it contains sequence homology with the putative RNA-binding TRAM domain in the N-terminus.¹² With a crystal structure determined to 1.95 Å, four conserved cysteine residues were found in

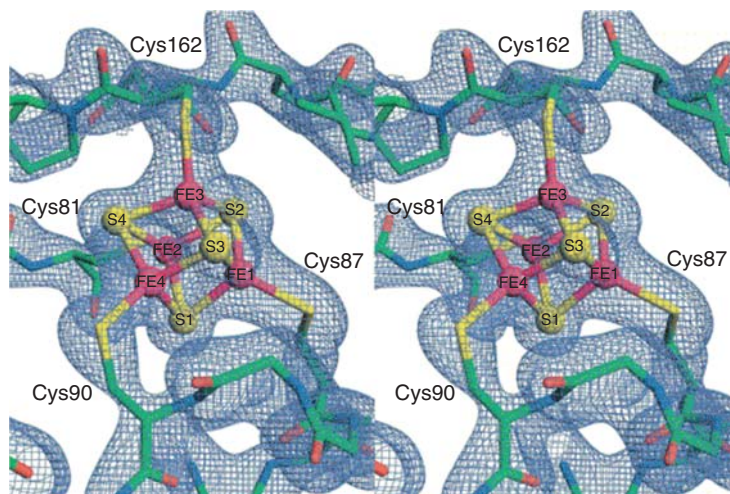


Figure 2 Electron density map of the iron–sulfur cluster and coordinating cysteine residues in Ruma. Reproduced from T. T. Lee; S. Agarwalla; R. M. Stroud, *Structure* **2004**, *12*, 397, with permission from Elsevier Limited.

the central domain of the protein coordinating an iron–sulfur [4Fe–4S] cluster, further distinguishing Ruma from RUMT as seen in **Figure 2**. Mutational analysis of each conserved cysteine to either leucine or alanine yielded insoluble protein, leading to the hypothesis that binding of the [4Fe–4S] cluster is important for Ruma folding.

Cysteine has been shown to serve an important catalytic function as well in both RNA and DNA MTases that modify carbon.¹³ The Pro-Cys dipeptide of the DNA m⁵C MTase has been previously studied.¹⁴ The cysteine nucleophilically attacks C6 in the pyrimidine (Py) ring, activating C5 for methylation in an AdoMet-dependent reaction. The bacterial protein Fmu (also known as RrmB) was the first m⁵C rRNA MTase characterized *in vitro* and found to modify 16S rRNA site-specifically at C967 with a K_M value in the high nanomolar range.¹⁵ Because the RNA and DNA MTases catalyze similar reactions, it was hypothesized that the same conserved cysteine in the Pro-Cys dipeptide was also the catalytic residue. Mutational analysis of this cysteine (Cys325Ala) in Fmu resulted in an active protein, yet mutation of a second conserved Cys375 (which aligns with a conserved nucleophile in RUMT) yielded an inactive enzyme that was unable to form the covalent intermediate with the cytosine base.¹³ As a follow up to this study, the yeast m⁵C-MTase, Nop2p, was used as a model for determining the role of both the Pro-Cys and the secondary cysteine (Thr-Cys, in this case) in the kinetic mechanism.¹⁶ Mutation of the Pro-Cys cysteine was lethal in Nop2p, leading the authors to propose a role for both cysteines in catalysis shown in **Figure 3**. They proposed that the Thr-Cys moiety is involved in covalent attack of the RNA cytosine and the Pro-Cys dipeptide is important for product release. The crystal structures of both the apo-Fmu and in complex with AdoMet were determined to 1.65 and 2.1 Å, respectively, and provided further support that Cys375 is positioned to act as the catalytic nucleophile. Cys325 was also close in proximity and believed to be involved in breakdown of the covalent intermediate.¹⁷

YebU, another site-specific rRNA MTase, was found to catalyze the m⁵C modification at position C1407 in 16S rRNA; interestingly, the enzyme modifies 16S rRNA in the 30S ribosomal subunit and not isolated 16S rRNA.¹⁸ A crystal structure of YebU (also called RsmF), determined at 2.9 Å, revealed a fold characteristic of the PUA (PseudoUridine synthase and Archaeosine transglycosylase) domain, also found in pseudoU synthase and the archaeal tRNA-guanine transglycosylase (TGT). Four of the key conserved residues among structures of PUA-domain containing proteins were found in YebU: Leu405, Val430, Val441, and Val443. Docking studies with 30S ribosomal subunit illustrated that the MTase interacts with both the 16S rRNA and ribosomal proteins, providing some evidence as to why the 30S complex maybe the most favorable substrate.¹⁹

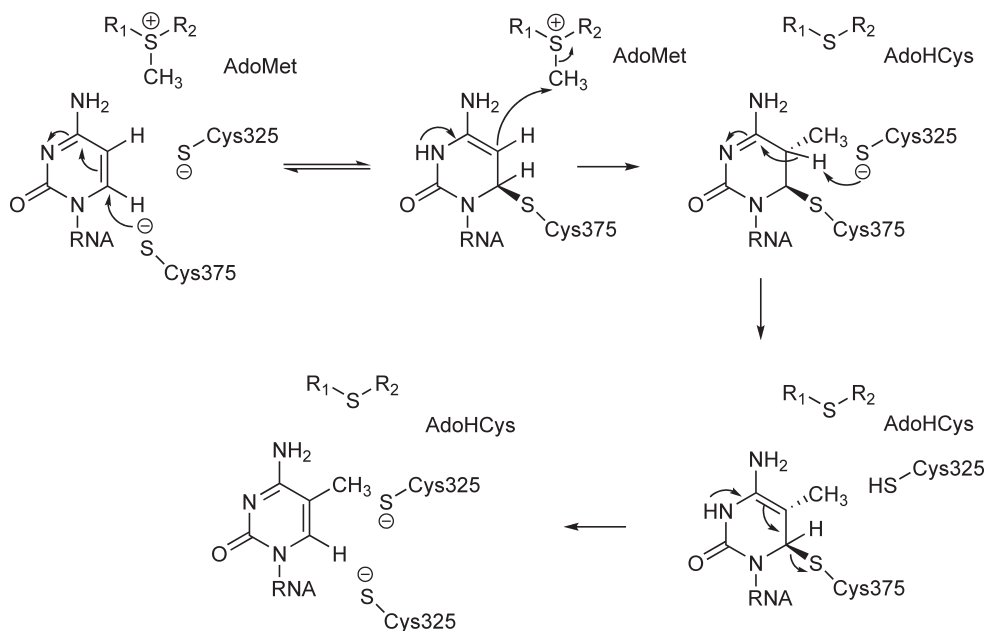


Figure 3 Proposed mechanism for the m^5C methyltransferase Fmu includes two catalytically important cysteine residues. This mechanism was proposed by Foster *et al.* Reproduced from P. G. Foster; C. R. Nunes; P. Greene; D. Moustakas; R. M. Stroud, *Structure* **2003**, *11*, 1609.

Although more fully characterized, an interesting development in the tRNA m^5C MTase literature came with the study of the yeast enzyme, YBL024w (Trm4).²⁰ Of the four m^5C modifications in yeast tRNA (C34, 40, 48, 49), [3H]-methyl was incorporated *in vitro* at each cytosine by purified Trm4, making it an example of the more recently discovered MTases that recognize multiple sites in a single RNA substrate. As a final piece of evidence for the essential role of Trm4, total tRNA prepped from a *trm4* deletion yeast strain demonstrated the absence of tRNA m^5C *in vivo*. Another interesting enzyme–substrate interaction was recently discovered when a MTase from *Drosophila* with sequence homology to known DNA MTases (Dnmt2) was shown to methylate tRNA^{Asp}.²¹ As work in this field continues to progress, we may start to see more examples of nontraditional substrates and interactions for enzymes catalyzing this diverse methylation.

6.20.2.2 Methylation on Tertiary Nitrogen

Modification at N^1 of adenosine (m^1A) has been well characterized in the T Ψ C loop of tRNA at position 58, most notably in stabilizing initiator tRNA^{Met} and also in stabilizing tRNA cloverleaf structure.^{22,23} Most bacterial methylation enzymes have been shown to contain multiple domains responsible for binding different substrates, namely AdoMet or RNA. Our understanding of the quaternary structure of these enzymes has been greatly improved with the determination of several crystal structures of m^1A -MTases.^{24–27} Using the bacterial m^1A58 MTase from *Mycobacterium tuberculosis* as an example, the X-ray crystal structure illustrated the enzyme Rv2118c as a homotetrameric complex, where each monomer consists of a small N-terminal domain and a larger AdoMet-binding C-terminal domain.²⁴ The authors describe the structure as a ‘dimer of tight dimers’, where the dimer interaction is formed by ionic bonds and the overall tetramer is held together by both ionic and hydrophobic interactions. Similar to bacterial enzymes, the archaeal species *Pyrococcus abyssi* also produces a homotetrameric m^1A MTase, but in this case the methylation is not site specific.²⁶ In addition to catalyzing the methylation at position 58, the archaeal enzyme also methylates A57, which was previously shown to be an intermediate in the formation of 1-methylinosine.²⁸ This regional specificity is unique to the archaeal enzyme and may involve an overlap with inosine biosynthesis. It should be noted that m^1I57 is specific to archaea. The

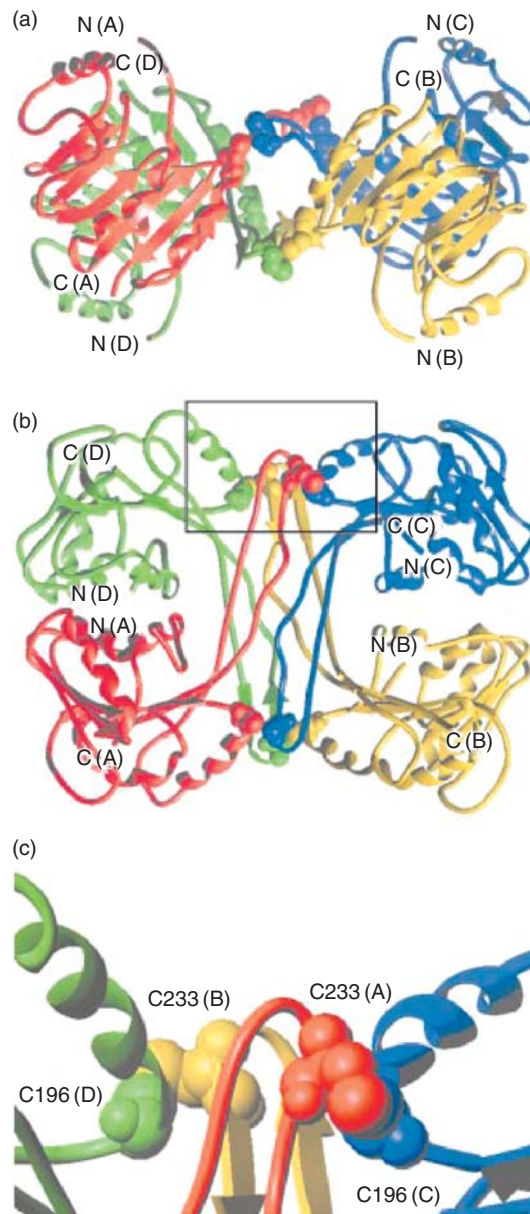


Figure 4 X-ray crystal structure of the homotetrameric archaeal m^1A methyltransferase Trm1. (a) View from the top, (b) view from the side, and (c) close-up of the stabilizing intermolecular disulfide bonds. Reproduced from M. Roovers; J. Wouters; J. M. Bujnicki; C. Tricot; V. Stalon; H. Grosjean; L. Droogmans, *Nucleic Acids Res.* **2004**, 32, 465, with permission from Oxford University Press.

efficiency of modification at A57 was higher than methylation at A58 in a $tRNA^{Asp}$ (G57A) mutant. A crystal structure determined to 2.66 Å demonstrated another key difference between the bacterial and archaeal enzyme, that is, the importance of key intermolecular disulfide bonds in tetramer formation in place of the hydrophobic interactions described previously (Figure 4). The authors demonstrated through mutational analysis of each Cys to Ser that the formation of these intermolecular interactions are particularly important for stability at higher temperatures. This was not previously predicted for thermophilic organisms but is becoming an increasingly prevalent idea due to relatively recent findings in chemical genomics analyses of archaeal species.²⁹

In contrast, the m^1A58 modification in eukarya is catalyzed by a heterodimeric complex. In *Saccharomyces cerevisiae*, methylation is believed to be catalyzed by the enzyme complex Gcd10p/Gcd14p (also known as Trm6/Trm61).³⁰ Although Gcd14p itself contains an AdoMet-binding domain, the purified protein did not display any enzymatic activity. In addition, purified Gcd10p has been shown to have RNA binding capability,³¹ which led the authors to consider that m^1A methylation in yeast could result from the activity of two subunits working in concert. The activity of purified Gcd10p/Gcd14p complex was characterized and exhibited a K_M for tRNA in the nanomolar range and for AdoMet in the micromolar range. The human homologues of the yeast enzymes, hTrm6p/hTrm61p, exist as a heterodimer when expressed in yeast, and *in vitro* assays demonstrated an activity similar to the yeast complex for A58-specific methylation.³² In a follow-up study, Ozanick *et al.*²⁷ modeled the Trm6/Trm61 complex as a heterotetramer after the structure of the *M. tuberculosis* enzyme, and performed site-directed mutagenesis with the hope of elucidating key residues involved in inter-subunit stability. A variety of alanine mutants were constructed, but none strongly affected dimer association or AdoMet binding. It was subsequently determined that the mutations affected tRNA binding, leading the authors to hypothesize that tRNA binding could be involved in the interface between the heterodimeric subunits. Such an example of subunit interactions working in concert for a fairly simplistic modification makes these MTases very intriguing for further study.

Similar to the N^1 -methyladenosine modification, enzymes catalyzing methylation of two tertiary nitrogens in guanosine, yielding m^1G and m^7G , have recently been isolated from the three kingdoms and new substrates identified. TrmD (tRNA m^1G37 MTase in bacteria) has been the most fully characterized, recognizing the tRNAs for leucine, proline, and one for arginine.³³ It has been proposed that the entire tRNA structure is recognized in binding to TrmD, and that modification at G37 occurs only when the substrate tRNA contains guanine at position 36 as well.³⁴ However, a recent example showed that tRNA from *Aquifex aeolicus* containing an A36–G37 dinucleotide could serve as a substrate for TrmD, whereas guanine preceded by either C36 and U36 did not.³⁵ A crystal structure of *E. coli* TrmD determined to 2.5 Å provided evidence for tRNA binding in the C-terminal domain, which is distinct from the AdoMet-binding region. An extensive mutagenesis study on conserved residues from every domain and participating portion of the structure helped to elucidate potential catalytic residues.³⁶ Sequence alignments of TrmD and similar proteins (including SpoU and YbeA) revealed that the MTase fold is distinct from the canonical Rossmann fold, in that the AdoMet-binding loop is not located at the N-terminus but at the C-terminus of the protein. This new class of proteins was given the name SPOUT (SpoU-TrmD family).^{37,38} The structure of bacterial TrmD has been shown to mimic that of the eukaryl m^1A MTase described above, in that the enzyme exists as a homodimer where the AdoMet-binding site is formed at the interface of the two monomers.³⁹ A crystal structure of TrmD from *Haemophilus influenzae* determined to 1.85 Å revealed that the MTase contains a trefoil knot, resulting in a bent pocket for AdoMet binding, which also adopts a bent conformation. Key residues at the dimer interface include Ser170 and Asp177, which can provide hydrogen-bonding interactions with AdoMet. In addition, the authors suggest that residue positioning in the crystal structure of TrmD makes Asp169 and Arg154 in the dimer interface the most likely contacts with G37 in substrate tRNA (Figure 5), proposing also that Asp169 may be directly involved in catalysis. Crystallographic evidence from several bacterial enzymes has aided our understanding of this interesting example of substrate binding.

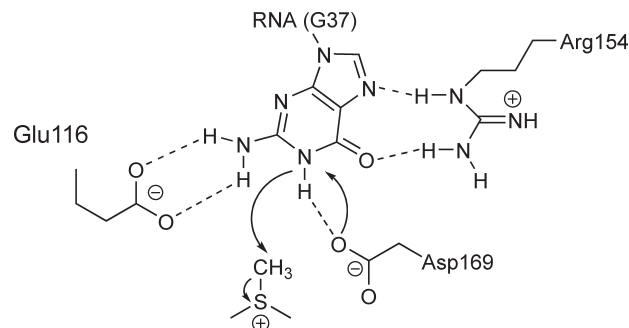


Figure 5 Orientation of G37 in the active site of the m^1G methyltransferase Trm. This mechanism was proposed by Ahn *et al.* Reproduced from H. J. Ahn; H. W. Kim; H. J. Yoon; B. I. Lee; S. W. Suh; J. K. Yang, *EMBO J.* **2003**, *22*, 2593.

Studying the monomeric m¹G37 MTase from the archaeal species *Methanocaldococcus jannaschii*, Christian *et al.*⁴⁰ determined that the sequence of Mj0883 is not related to the bacterial TrmD but is more homologous to the eukaryotic enzyme, Trm5, which catalyzes the m¹G37 modification in cytosolic and mitochondrial tRNAs.^{41,42} In addition, alanine mutation of the key GXG motif (involved in binding to the methionine of the AdoMet substrate) conserved in class I MTases abolished activity. In a later report, Christian *et al.*⁴³ expanded their mutagenesis studies to several conserved active site residues to determine their roles in catalysis. Alanine mutations of several charged and polar uncharged amino acid residues did not dramatically affect the activity of the enzyme in the majority of cases. Proline 267 is conserved in the NLP motif of the Trm5 homologues, and it was demonstrated, with steady-state kinetic analysis, that the Pro267Ala mutation had the most deleterious effect on catalysis. The most severe effect was seen in the K_M for tRNA; gel-shift assays also revealed a decreased ability to bind substrate tRNA, and the authors proposed that the effect could be in maintaining the rigidity of the backbone to stabilize enzyme–substrate complexes.

The bacterial MTases catalyzing the m⁷G modification in the variable loop of tRNA have also been well studied. In eubacteria, the m⁷G46 MTase YggH, of the classical Rossmann fold class, is responsible for modification of tRNA.^{44,45} A *yggH* knockout strain lacked m⁷G46 tRNA, indicating that this enzyme is required for catalysis. The authors also demonstrated that a critical determining factor in substrate recognition was in the length of the variable loop, and not necessarily the global tertiary structure. An interesting difference between bacterial species was illustrated with the crystal structure of the *Bacillus subtilis* m⁷G MTase TrmB, which was shown to exist as a homodimer.⁴⁶ The interface of the two subunits is primarily hydrophobic and encapsulates five water molecules and a potassium ion, which makes contacts with both main- and side-chain residues in each monomer. In contrast, the m⁷G46 modification in yeast is conferred with a heterodimeric complex, Trm8/Trm82.⁴⁷ Yeast strains with deletions in either *trm8* (the homologue of TrmB) or *trm82* did not exhibit m⁷G modification. To test the ability of AdoMet to bind to several alanine mutants, *in vitro* steady-state kinetic analyses were performed.⁴⁸ Residues predicted to be important by sequence alignment (e.g., Arg150, Asn152, Asp180, and Glu220) were probed, and alanine mutations of all but asparagine 152 resulted in a decreased activity for Trm8. The most severe point mutation, Glu220Ala, had a 10-fold reduction in k_{cat} . Matsumoto *et al.*⁴⁹ also found that the apparent conserved variable loop sequence required for m⁷G modification is A-A-G-Py-Py (positions 44–48). Unlike the case of the m¹A heterodimeric MTase, the authors suggest that Trm8 is responsible for catalysis. The role of Trm82 is not understood, but they speculate that the protein may be involved in tRNA recognition. In eukarya, the m⁷G modification is found predominantly in the 5'-methylguanosine cap structure in mRNA. Because this modification is generally classified as RNA processing as opposed to modification, it will not be discussed in this chapter (for a review see Gu and Lima⁵⁰).

6.20.2.3 Methylation on Primary Nitrogen

In contrast to MTases that modify tertiary nitrogens, much less appears to be known about the gene products that encode MTases for the modification of primary nitrogens of the exocyclic amine moieties. Characterized predominantly in eukaryotic mRNA, the N⁶-methyladenosine modification is found in both the cap region of the mRNA as well as sparsely within the open reading frame and is believed to play a role in RNA processing.⁵¹ The primary catalytic function of the human m⁶A MTase is linked to the subunit MT-A70, which shares sequence homology with the yeast protein, IME4.⁵² This protein is involved in the meiosis and sporulation pathway in *S. cerevisiae*. An IME4 deletion strain lacks the m⁶A modification in yeast mRNA, providing an interesting possible role for N⁶-methyladenosine modification in the replication cycle of the organism.

YbiN, a bacterial m⁶A MTase, was recently discovered to site-specifically methylate A1618 of 23S rRNA in *E. coli*.⁵³ However, the enzyme could not modify either a 23S protein-free rRNA or the 50S ribosomal subunit *in vitro*, suggesting that there may be other contributing protein, RNA, or cofactor interactions responsible for catalysis. The crystal structure of another 23S rRNA m⁶A2058 MTase, ErmC, from *B. subtilis* was solved in complex with AdoMet, which revealed slight conformational changes in the enzyme upon substrate binding.⁵⁴ Interestingly, binding of the rRNA substrate was studied previously, revealing that the enzyme catalyzed an N⁶,N⁶-dimethyladenosine (m⁶₂A) modification where the reaction mechanism proceeded through an N⁶-monomethylated intermediate.⁵⁵ To elucidate the mechanism for dimethyl modification, a detailed kinetic

analysis was performed with both un- and mono-methylated 23S rRNA, and the authors concluded that an overall sequential mechanism composed of two individual random bi-bi reactions was followed. Two molecules of AdoMet are required for catalysis, and because m⁶A2058 rRNA is a good substrate for the enzyme, the authors propose that following mono-methylation the RNA dissociates and the enzyme reassociates to perform the second methylation reaction.

Another interesting example of the m₂⁶A modification occurs in 16S rRNA.⁵⁶ The crystal structure of KsgA from *E. coli* was determined to 2.1 Å, the tertiary structure of which was determined to be homodimeric and structurally related to ErmC. In contrast to the 23S rRNA MTase, KsgA catalyzes a dimethyl modification of adjacent adenosines (A1518 and A1519) in a pre-formed 30S ribosomal subunit, although the extent of formation is under some debate. Residues important to AdoMet-binding were found in motifs I–IV of KsgA (i.e., Asn113–Leu114–Pro115–Tyr116, particularly Asn113 and Pro115). A later report by Desai *et al.*⁵⁷ validated that the bacterial KsgA can in fact modify a 30S ribosomal subunit but only if the complex is in an ‘inactive’ conformation. Previous studies have suggested that high concentrations of the cations NH₄⁺ and Mg²⁺ help the 30S subunit adopt an ‘active’ conformation, while the ‘inactive’ conformation is stabilized by low concentrations.^{58,59} To support the notion that the inactive 30S conformation is a substrate for the enzyme, kinetic analyses were performed in different buffers varying the concentration of both cations. The active complex conformation was determined to be unmodified. This result was supported by crystal structure evidence of the active 30S conformation from *T. thermophilus*, which reveals that the substrate adenosines are buried in the core of the complex.⁶⁰ A conformational change in the 30S subunit exposes both A1518 and A1519 to solvent, allowing for modification by KsgA. Although more has been discovered regarding the structure and substrate recognition of KsgA, the order of binding still remains unclear with respect to the 16S rRNA substrate. The enzyme may act in a manner similar to ErmC, with dissociation after each modification, or possibly dimethylate a single adenosine with each binding event, or a third possibility exists that there may be only a single RNA-binding step where all four methyl groups are transferred before KsgA dissociates. Further studies are necessary to fully probe the processivity of this interesting RNA modification enzyme.

Structural insights have been possible with the MTases responsible for the m²G modification. The protein PAB1283 from the archaeal species *P. abyssi* catalyzes the m²G10 methylation in tRNA. This was structurally characterized and contains two domains: the C-terminal Rossmann-fold, AdoMet-binding domain and the THUMP domain in the N-terminus (ThioUridine synthases, RNA Methyltransferases, and Pseudouridine synthases) predicted to be involved in RNA binding.^{61,62} The authors suggest that the THUMP domain of PAB1283, although not necessary for catalysis, is probably involved in orienting the substrate for proper interaction with the C-terminal catalytic domain. Three aspartate residues and a water molecule were predicted to stabilize the AdoMet molecule, each recognizing a separate portion of the substrate: Asp185 stabilizes the methionine side chain, Asp208 coordinates the ribose, and Asp235 interacts with the adenine ring. As with ErmC, PAB1283 has been shown to catalyze both mono- or dimethylation m²G10 with the mono-methylated species serving as an intermediate; however, in archaea the m₂²G modification is favored with increasing temperature.

Although the m²G10 modification has been studied more fully, it is not the only N²-methylguanosine in tRNA: Trm1, from the archaeal species *Pyrococcus furiosus*, catalyzes the modification N²,N²-dimethylguanosine in tRNA at position 26.⁶³ The protein was found to be monomeric and very thermostable. It was observed that the temperature and the ratio of RNA substrate to enzyme influenced mono- versus dimethylation of the guanine base. At 30 °C, the m²G26 product was more pronounced, yet at 50 °C the dimethylated product was in excess. In the case where the concentration of tRNA far exceeded that of Trm1 (400-fold excess) with incubation at 50 °C, there was a buildup of the m²G intermediate, which the authors interpret by suggesting that following addition of a single methyl group to substrate tRNA the enzyme dissociates before performing the second methylation step. The proposed kinetic mechanism for both mono- and dimethylation is illustrated in **Figure 6**. Grosjean and coworkers also determined that the major identity elements of the tRNA consist of two G–C base pairs in the D-arm (positions 10–25 and 11–24) and a 5-nucleotide variable loop, which are in agreement with the enzymes previously characterized in eukaryotes.^{64,65} In contrast to the m₂²G10 MTase, presence of these recognition elements favor the dimethyl- as opposed to the mono-methylated target guanosine, suggesting that these enzymes catalyzing similar modifications have evolved independently.⁶⁶

The N^2 -methylguanosine modification has also been found in *E. coli* in both 16S rRNA at G966, G1207, and G1516 and in 23S rRNA at G1835 and G2445.⁶⁷ The protein RsmD (*ybbF* in *E. coli*) was found to be a MTase specific for m^2G modification at G966,⁶⁸ and RsmC (and homologues) catalyzes site-specific methylation of guanine 1207 in 16S RNA.^{69,70} Both enzymes have been shown to recognize the target 16S rRNA in the 30S subunit in some intermediate assembly conformation, but not the naked rRNA.^{71,72} This case serves as another emerging example where the presence of ribosomal proteins are critical for rRNA modification, either through key interactions with the rRNA modification enzymes, or perhaps by presenting the RNA in a specific, recognizable conformation. In addition to the more fully characterized 16S rRNA MTases, the gene products responsible for two m^2G modifications in 23S rRNA have been recently identified. Consecutively published papers reported that gene deletion analyses in *E. coli* revealed that *yggO* (RlmG) is a site-specific MTase for G1835 in 23S rRNA, and *ycbY* (RlmL) selectively modifies G2445.^{73,74} Deletion of both genes was found to affect the bacterial growth rate, but as with the 16S rRNA MTases, future *in vitro* studies are required to understand the kinetics and mechanism for methylation.

6.20.2.4 Methylation on Oxygen

In addition to the variety of modifications found on the heterocyclic base discussed above, many nucleosides are also methylated at the 2'-hydroxyl of the ribose moiety and make up approximately 8% of existing tRNA modifications.⁷⁵ One interesting example of the 2'-*O*-ribose modification was studied in *S. cerevisiae*, where the yeast tRNA molecules corresponding to His, Pro, and Gly(G-C-C) contain a 2'-*O*-methylated nucleoside at position 4 in the acceptor stem.⁷⁶ A methylated cytosine is found in tRNA^{Pro} and tRNA^{Gly}, and the modified Am nucleoside is found in tRNA^{His}. Modifications in a duplex region of tRNA are very rare, yet modification at this position (#4) is conserved in eukaryotes. A yeast knock-out strain of the gene *trm13* was produced, and it was determined by HPLC and primer extension analysis that tRNAs purified from this organism did not exhibit the 2'-*O*-methyl modification at position 4.

Similarly, the *T. thermophilus* Gm18 MTase, TrmH, was characterized as a site-specific modification enzyme for guanine 18 in the D-loop of tRNA.⁷⁷ The crystal structure of the enzyme was determined at 1.85 Å, which characterized the homodimeric MTase as containing the trefoil knot structure of the SPOUT family.⁷⁸ The AdoMet-binding site is lined with predominantly hydrophobic residues surrounding the adenine ring. In addition, a hydrogen bonding network consisting of Glu124, Ser150, and Asn152 are believed to interact with the RNA substrate. In fact, alanine mutations of any of these amino acids reduces methylation activity. Also involved in the catalytic center through hydrogen bonding interaction with Ser150 and interaction with the bound phosphate ion (believed to represent the phosphate backbone moiety of the tRNA substrate) is Arg41', the basic residue in this case belonging to the opposite monomer (Figure 7). The authors propose that

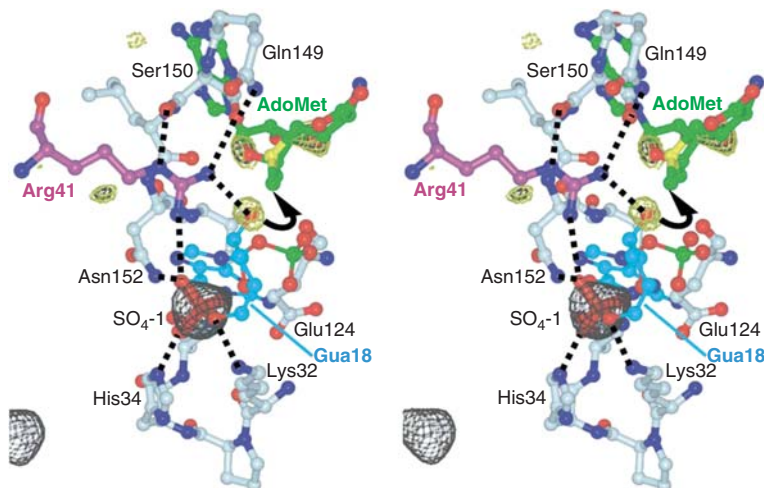


Figure 7 Active site residues with bound AdoMet in the 2'-*O*-methyltransferase TrmH. Reproduced from O. Nureki; K. Watanabe; S. Fukai; R. Ishii; Y. Endo; H. Hori; S. Yokoyama, *Structure* **2004**, *12*, 593, with permission from Elsevier Limited.

Arg41' may serve as the general base through activation by the hydrogen-bonding network, and by deprotonating the 2'-hydroxyl of the tRNA substrate, facilitate direct attack upon the methyl group of AdoMet. Interestingly, this hypothesis was supported by an Arg41Ala mutant that demonstrated no methylation activity and the fact that this arginine residue is conserved in the SpoU family of MTases. The catalytic mechanism of this enzyme is interesting, in that binding of the tRNA would alter the hydrogen-bonding network of the active site and poise Arg41' for activation of the 2'-*O*-ribose, suggesting that substrate-binding induces the proper conformation of the enzyme for catalysis.

In addition to methylated ribose moieties of tRNA, the 2'-*O*-ribose modification is found at several positions of rRNA. Mutational analysis of RrmJ, a universally conserved 2'-*O*-ribose MTase that modifies U2552 in 23S rRNA, illustrated that there are at least three catalytically important residues: Lys38, Asp124, and Lys164, which together form a catalytic triad.⁷⁹ Alanine mutants of each of the above amino acids exhibited an extreme decrease in the rate of reaction. Glutamate 199 also exists in proximity to the catalytic triad, yet the mutant did not exhibit a defective phenotype *in vivo* but was less active in the *in vitro* methylation assays. The authors propose that perhaps Glu199 is not directly involved in catalysis but in mediating attack through hydrogen bonding-induced polarization of the catalytic residues (Figure 8). The adjacent nucleoside G2251 is also modified to 2'-*O*-methylguanosine (Gm) by the enzyme RlmB, characterized in *E. coli*.⁸⁰ The crystal structure of the MTase was determined at 2.5 Å; RlmB was found to exist as a homodimer with the knotted active site structure of the SpoU family. Another interesting example is the *thyA* gene from *M. tuberculosis*, which was characterized through isolation of a capreomycin drug-resistant *thyA* mutant.⁸¹ The enzyme methylates the 2'-hydroxyl of two cytosine nucleosides in different rRNAs: C1409 in 16S rRNA and C1920 in 23S rRNA, which are brought in close spatial proximity through association of the 30S and 50S ribosomal subunits. Although the link to a new drug target and the identification of new rRNA modifications is interesting, further studies will need to be conducted to determine the mechanism for modification of each nucleoside by TlyA.

In contrast to eukaryotes and eubacteria, the 2'-*O*-ribose modification of tRNA and rRNA in archaea has been shown to occur by two different mechanisms: MTases that site-specifically modify the 2'-hydroxyl moiety and small nucleolar RNAs (snoRNAs) (also called guide RNA) that position the nucleoside for modification by forming a duplex recognized specifically by the enzyme. The latter guide mechanism has been reviewed in recent years.^{82–84} There have been a few site-specific MTases identified to modify tRNA, the most recent being Trm56 from *P. abyssi*, which is a 2'-*O*-methyltransferase specific to Cm56 in the T-loop found only in archaea.⁸⁵ Cytosine 56 forms a critical base-pairing interaction with G19 and helps to stabilize the D/T loop interaction in the tertiary structure of the tRNA. The major recognition element for the enzyme in addition to C56 is the T-loop structure, as an RNA substrate lacking the D-loop was still modified *in vitro*. The authors initially characterized the enzyme as a member of the class IV MTases (the SPOUT family). The determination of the crystal structure of Trm56 from *Pyrococcus horikoshii* helped to further elucidate the mechanism of this unique MTase.⁸⁶ The protein was found to exist as a dimer with a large interaction surface consisting predominantly of hydrophobic residues and a few key hydrogen-bonding interactions. Each monomer contains the trefoil knot structure typical of the SPOUT family, which serves as the AdoMet binding site; several basic and acidic residues interact with the AdoMet substrate in addition to polar noncharged amino acids providing hydrogen-bonding interactions. Although less common in eukaryotes, the Gm2922 modification in 27S pre-rRNA was believed to be catalyzed by a small nuclear RNA (snRNA), but it was later determined that the site-specific MTase Spb1 was responsible for the modification.⁸⁷

Modification of the 2'-hydroxyl moiety on the ribose can also occur in the anticodon loop of tRNA. Although methylation of nucleosides is widespread, 2'-*O*-ribose methylation is found occasionally in the first position of the anticodon but not the second or third. Because the modification of nucleosides in and adjacent to the anticodon loop of tRNA is commonplace, Satoh *et al.*⁸⁸ examined the decoding efficiency of tRNA^{Ser} in a cell-free expression system when either the first, second, or third nucleosides in the anticodon (either C-G-A or C-U-C) were methylated at the 2'-hydroxyl. While 2'-*O*-methylcytosine (Cm) in the first position increased the translational efficiency, methylation of both the second and third nucleoside (both the double modification as well as individual methylations at either position) resulted in dramatic reductions in activity. This study serves as a reminder that although methylation may seem simple and diverse throughout the RNA sequence, we still have much to learn about the functionality and biological importance of these modified nucleosides.

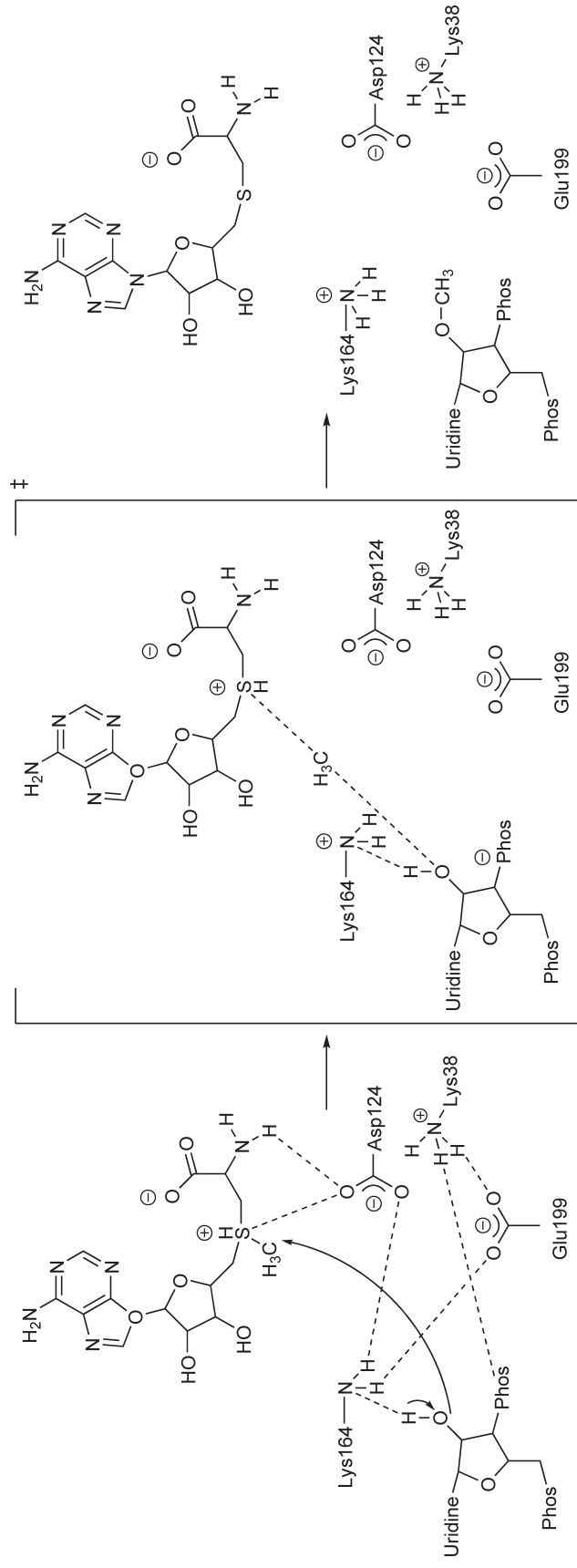


Figure 8 Proposed reaction mechanism for the rRNA 2'-O-methyltransferase RrmJ. The mechanism was proposed by Hager *et al.* Reproduced from J. Hager; B. L. Staker; H. Bugli; U. Jakob, *J. Biol. Chem.* **2002**, 277, 41978.

6.20.3 Thiolation and Selenation

6.20.3.1 Thiolation

Thiolation of nucleosides has been found in a great diversity of species across the kingdoms, especially in tRNAs. Of the approximately 10 different thiolated nucleosides that have been identified so far, thiolation at either position 2 or 4 of purine/pyrimidine rings (e.g., 2-thiocytidine (s^2C), 4-thiouridine (s^4U), 5-methyl-aminomethyl-2-thiouridine (mnm^5s^2U), and 2-methylthio- N^6 -isopentenyl-adenosine (ms^2i^6A), **Figure 9**) appears to predominate. While the sulfur is attached to an unsubstituted carbon in the case of ms^2i^6A , it is actually more often the case that a carbonyl oxygen substituent is replaced by sulfur, as seen in the other thionucleosides mentioned above. Most of the thiolated nucleosides occur in the anticodon loop of tRNAs, suggesting a significant role for thiolation in terms of the fidelity and efficiency of protein translation. Over the past decade two mechanisms for thiolation (thiolation directly on carbon vs. displacement of an oxygen substituent) have been proposed and extensively studied.⁸⁹ In this section, RNA modifying enzymes that mediate/catalyze sulfur trafficking in nucleoside thiolation will be discussed along with their corresponding mechanisms.

Biosynthesis of 4-thiouridine at position 8 of certain prokaryotic tRNAs has been demonstrated to be catalyzed by two enzymes, IscS and ThiI.^{90,91} IscS, a cysteine desulfurase, mediates the sulfur transfer from cysteine and forms a persulfide intermediate.⁹² Even though IscS has been shown to participate in the formation of [4Fe-4S] clusters as a sulfur donor,⁹³ the IscS persulfide delivers sulfur directly to ThiI without the involvement of a [4Fe-4S] cluster.⁹⁴ (However, ms^2i^6A and s^2C biosyntheses are classified as [4Fe-4S] cluster-dependent, which will be discussed later in this section.) Two possible mechanisms for 4-thiouridine generation have been proposed to include sulfur transfers via persulfide groups, as shown in **Figure 10**.^{95,96} The nucleophilic attack appears to be ATP/Mg²⁺-dependent. Briefly, IscS seizes sulfur from the substrate cysteine and subsequently transfers that S atom to one of the active site cysteine residues of ThiI. After the substrate tRNA is adenylated to make the 4-oxygen a good leaving group, the electrophilic sulfur attacks the activated uridine to generate 4-thiouridine and AMP. Site-directed mutagenesis studies have determined that Cys456 and Cys344 are crucial for the activity of *E. coli* ThiI. In comparison with wild-type ThiI, the Cys344Ala mutant showed a 2700-fold reduction in s^4U formation whereas there was no s^4U generation detected in the Cys456Ala

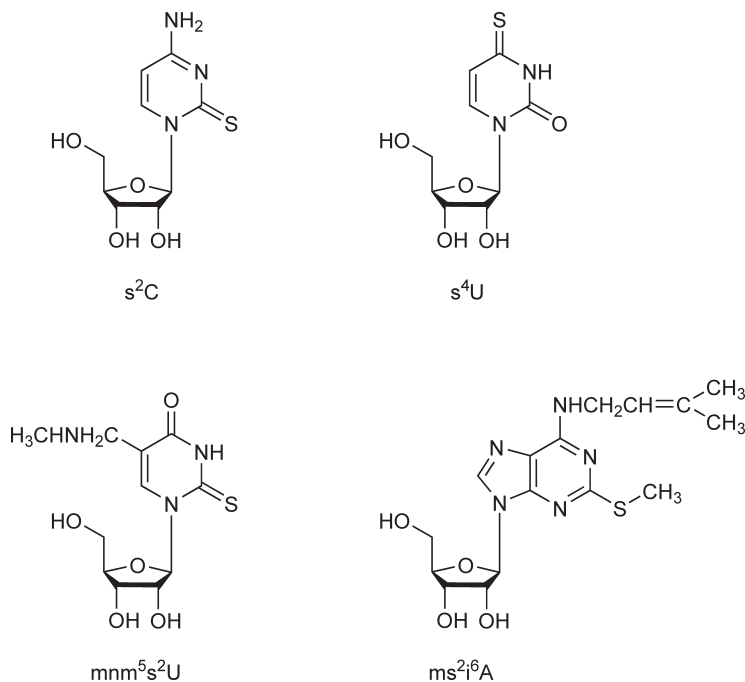


Figure 9 Examples of identified thiolated nucleosides. 2-Thiocytidine (s^2C), 4-thiouridine (s^4U), 5-methyl-aminomethyl-2-thiouridine (mnm^5s^2U), and 2-methylthio- N^6 -isopentenyl-adenosine (ms^2i^6A).

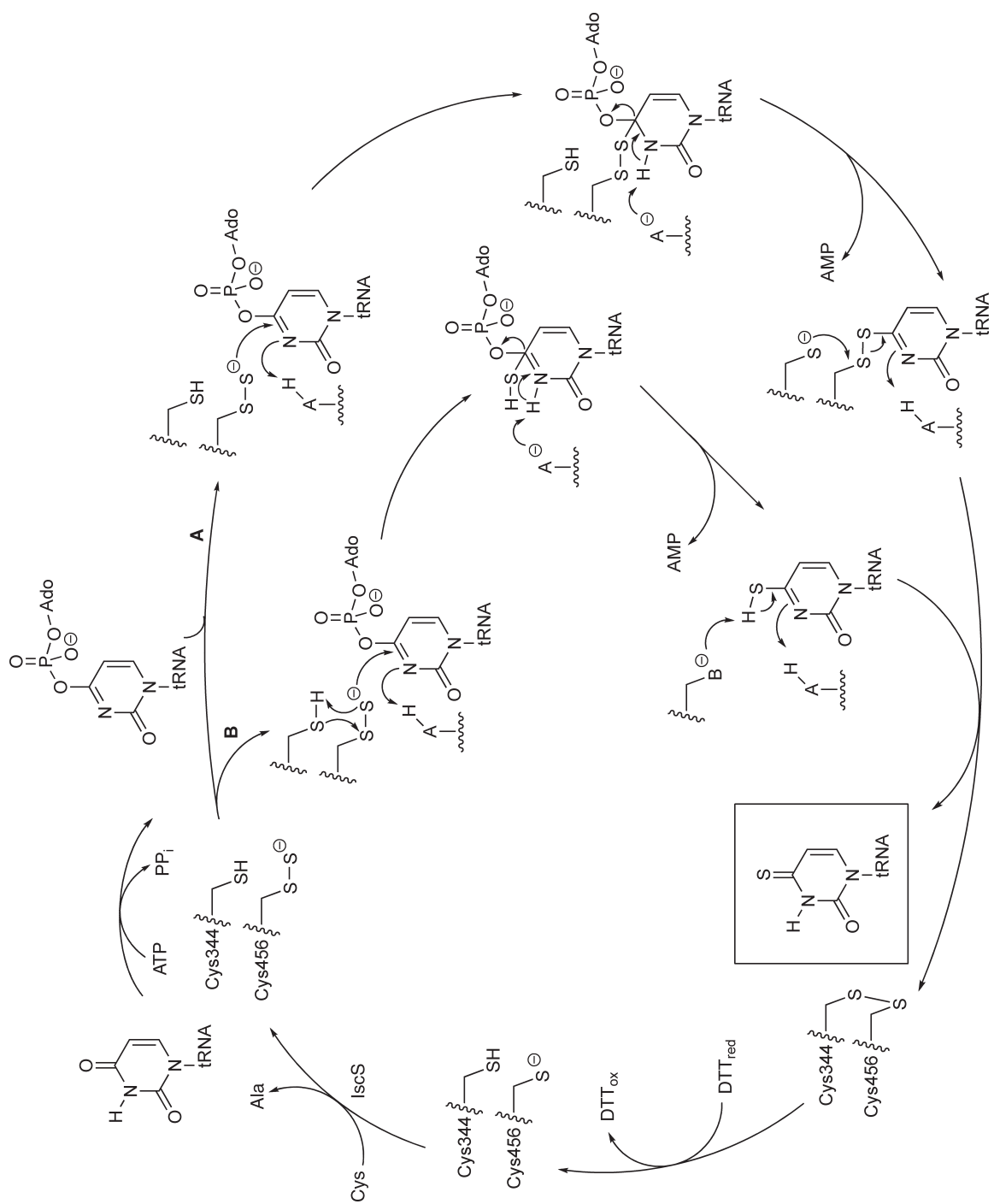


Figure 10 Proposed catalytic mechanisms for 4-thiouridine formation. The two mechanisms differ in which sulfur serves as a nucleophile, where A: the persulfide group on Cys456 attacks the adenylylated uridine, or B: that persulfide group is used as a source of bisulfide (HS⁻) that becomes a nucleophile. The mechanisms were proposed by Mueller *et al.* Reproduced from E. G. Mueller; P. M. Patenchar; C. J. Buck, *J. Biol. Chem.* **2001**, 276, 33588.

mutant.^{95,97} *In vitro* analysis with [³⁵S]-cysteine revealed that Cys456 of ThiI becomes radiolabeled over time, and the label disappears after treatment with reductant such as dithiothreitol (DTT), consistent with the formation of a persulfide.⁹⁶ In 2006, Mueller and coworkers utilized a fluorescent alkylating agent, 5-({2-[(iodoacetyl)amino]ethyl}amino)naphthalene-1-sulfonic acid (*I*-AEDANS), to chemically trap the Cys456 persulfide group. With the help of mass spectrometry, persulfide and disulfide intermediates were indeed observed during the reaction.⁹⁸ To demonstrate ATP-dependence in the reaction, it has been shown that ThiI shares a conserved sequence SGGXD(S/T), the so-called P-loop motif,⁹⁹ with the 'PP_i synthetase family', whose members are capable of adenylating and subsequently replacing carbonyl oxygen. In addition, mutation of two possible key residues with respect to adenylation, Asp189Ala and Lys321Met, have resulted in the abrogation of s⁴U biosynthesis in *E. coli*.¹⁰⁰ More recently, Waterman *et al.*,¹⁰¹ determined the crystal structure of *Bacillus anthracis* ThiI in complex with AMP (Figure 11), which further confirmed the intermediacy of adenylation during catalysis. They were also able to identify both the RNA-binding THUMP domain and P-loop domain in the structure.

Derivatives of s²U are generally found at the wobble position (U34) of the anticodon in tRNAs. Similar to s⁴U, the mechanism of this thio-modification is also thought to be [4Fe-4S] independent and to occur via persulfide intermediates. A sequence homologue of ThiI, MnmA, which recognizes anticodon stem loop structures of tRNA^{Glu} and tRNA^{Lys}, is required for *in vitro* s²U synthesis.¹⁰² However, relative to ThiI, the *in vitro* activity of MnmA was found to be quite low, suggesting that additional factors might be involved in this modification.¹⁰³ Therefore, not too surprisingly, a more complicated sulfur trafficking system has been proposed. In 2006, Suzuki and coworkers identified three other intermediate proteins, TusA, a TusBCD complex, and TusE that seem to be required for s²U biosynthesis in *E. coli*.¹⁰⁴ The flow of sulfur transfer starts at IscS, which seizes sulfur from cysteine via the formation of a persulfide group. This sulfur is transferred to TusA, which subsequently relays the sulfur to TusD in the TusBCD complex and then to TusE. At the final step, TusE delivers the sulfur to MnmA-tRNA complex and then to the final substrate, presumably adenylated uridine (Figure 12). To date, it is still not known why biosynthesis of s²U needs to be carried out by such a complicated sulfur transfer system; however, it is conceivable that those intermediate sulfur-carrying proteins may also participate in other sulfur transfer pathways which can mediate their downstream biological functions. This implies that there is a control system for cells to regulate the levels of various sulfur metabolites by sharing sulfur trafficking proteins.^{104,105}

In contrast to s⁴U and mnm⁵s²U biosyntheses, it is thought that ms²i⁶A and s²C biosyntheses follow a [4Fe-4S] cluster-dependent pathway.⁹⁴ The modification of ms²i⁶A has been found at position 37 of tRNAs and is believed to be a multistep reaction, where N⁶-isopentenylation (discussed in the alkylation section) takes place first followed by C² thiolation and methylation. It has been shown that C² thiolation and methylation in ms²i⁶A biosynthesis are catalyzed by the proteins IscS and MiaB. Simultaneously, it was reported by two distinct groups that IscS is involved in the biosyntheses of all identified thiolated nucleosides;^{106,107} however, unlike the cases of s⁴U and mnm⁵s²U mentioned earlier in this section, IscS delivers the sulfur to IscU, a scaffold protein with a [4Fe-4S] cluster, which in turn facilitates the cluster assembly.¹⁰⁸ Even though the mechanism has yet to be elucidated in detail, the [4Fe-4S] cluster can then be transported to activate the apo-form of a [4Fe-4S] protein,

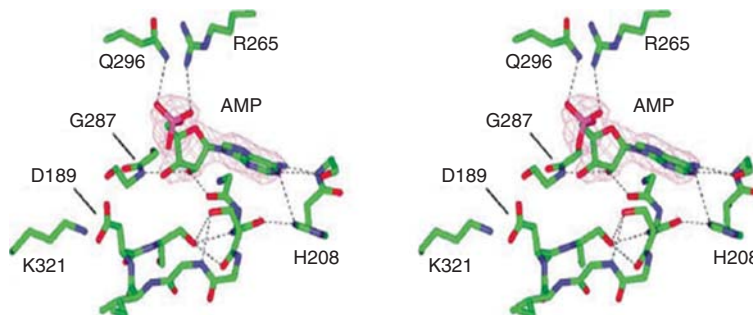


Figure 11 The active site of *Bacillus anthracis* ThiI in complex with AMP. Key residues in the active site are also presented. Reproduced from D. G. Waterman; M. Ortiz-Lombardia; M. J. Fogg; E. V. Koonin; A. A. Antson, *J. Mol. Biol.* **2006**, 356, 97, with permission from Elsevier Limited.

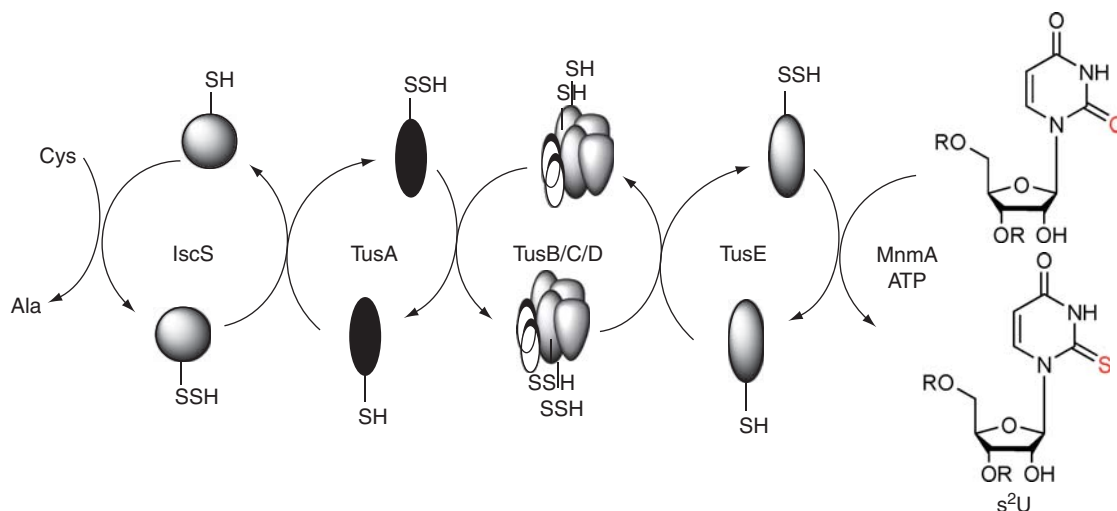


Figure 12 Proposed sulfur transfer mechanism for 2-thiouridine biosynthesis. Note: TusB/C/D forms a $\alpha_2\beta_2\gamma_2$ hexamer while sulfur is passed from TusA to TusD and then to the downstream carriers. It has yet to be elucidated why 2-thiouridine biosynthesis needs such a complex sulfur transfer system. The mechanism was proposed by Ikeuchi *et al.* Reproduced from Y. Ikeuchi; N. Shigi; J. Kato; A. Nishimura; T. Suzuki, *Mol. Cell* **2006**, *21*, 97.

MiaB.^{109,110} MiaB was first determined to play an essential role in C–S bond formation in ms^2i^6A (or ms^2io^6A) by Björk and coworkers,¹⁰⁹ where they observed only i^6A (or io^6A) present but not ms^2i^6A (or ms^2io^6A) in tRNAs from MiaB mutant strains. In 2002, Pierrel *et al.*,¹¹¹ further confirmed that MiaB is a monomeric, [4Fe–4S]-containing protein and three cysteine residues (Cys157, Cys161, and Cys164) were suggested to be involved in iron chelation. Mutagenesis studies with respect to those three cysteines revealed a deficit in terms of ms^2i^6A formation *in vivo* as well, suggesting the [4Fe–4S] cluster is required for enzyme activity. A conserved CX₃CX₂C sequence motif (corresponding to Cys157, Cys161, and Cys164 in *E. coli* MiaB) has been found via structural analyses in other enzymes that possess [4Fe–4S] clusters.¹¹² Recently, based upon various spectroscopic observations and site-directed mutagenesis, a second [4Fe–4S] cluster found in the N-terminal domain of MiaB (presumably chelated by Cys10, Cys46, and Cys79) has been proposed to serve as a sulfur donor during the reaction (**Figure 13**); however, the function of this cluster is still obscure.¹¹³ MiaB is annotated as a member of the radical-AdoMet family.¹¹⁴ Probably not coincidentally, ms^2i^6A biosynthesis has been shown to involve the formation of radical-AdoMet. **Figure 13** shows the proposed mechanism for ms^2i^6A formation. Briefly, the [4Fe–4S] cluster is chelated by three cysteine residues of MiaB and two molecules of AdoMet are required. In the first step of the reaction, the [4Fe–4S] cluster reduces one molecule of AdoMet to generate a 5'-deoxyadenosyl radical (Ado•). The radical is then transferred from Ado• to the C² position of i^6A so that $i^6A\bullet$ is formed. Subsequently, MiaB donates a sulfur atom to $i^6A\bullet$ and yields an s^2i^6A intermediate followed by the final step, where the second AdoMet mediates the methyl transfer and completes the formation of ms^2i^6A .¹¹⁵

The modified nucleoside s^2C has been found at position 32 in bacterial and archaeal tRNAs but not in eukaryotes. Even though the mechanism of s^2C biosynthesis has yet to be elucidated, the reaction appears to be IscS-dependent.^{106,107} In addition, Jäger and coworkers reported that the protein TtcA is essential for catalysis, and a conserved CX₂C motif in the TtcA protein sequence implies a possible mechanism that involves [4Fe–4S] cluster formation. This inference was supported by mutagenesis studies, where replacement of either Cys122 or Cys125 by alanine or serine resulted in the absence of s^2C *in vivo*.^{89,116}

6.20.3.2 Selenation

Selenium belongs to the same group as sulfur in the periodic table and many of its chemical properties are comparable to sulfur. Not too surprisingly, 5-methyl-aminomethyl-2-selenouridine (mnm^5Se^2U) thus has been found in the wobble position of tRNAs that code for lysine, glutamate, and glutamine.^{117,118} The biosynthetic

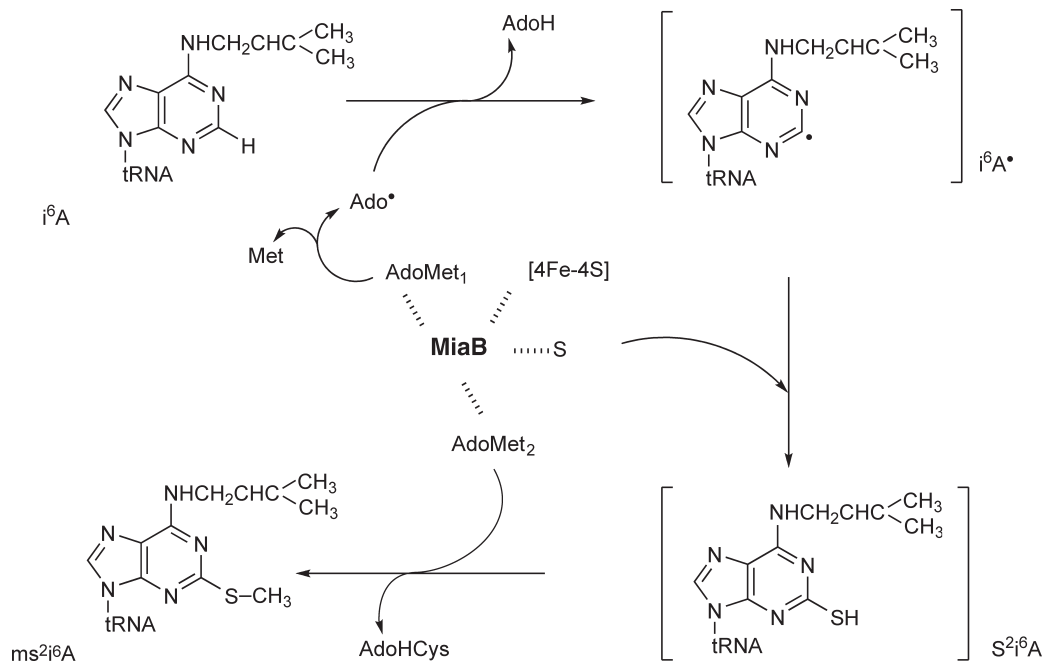


Figure 13 Proposed catalytic mechanism for ms^2i^6A synthesis by MiaB. Note: MiaB is a bifunctional (involved in both thiolation and methylation) radical-AdoMet enzyme that contains two molecules of AdoMet and at least an iron-sulfur cluster. Ado \bullet = 5'-adenosyl radical; AdoH = 5'-deoxyadenosine; AdoHcy = adenosylhomocysteine. The mechanism was proposed by Pierrel *et al.* Reproduced from F. Pierrel; T. Douki; M. Fontecave; M. Atta, *J. Biol. Chem.* **2004**, 279, 47555.

pathway for mnm^5Se^2U has been proposed, as shown in **Figure 14**, although its catalytic mechanism has yet to be fully elucidated. A comparison of the loss and formation of [^{35}S]- and [^{75}Se]-labeled tRNAs has suggested that mnm^5s^2U serves as the precursor of the mnm^5Se^2U , where selenium directly substitutes for sulfur at the C² position of uridine.¹¹⁹ In addition, [^{31}P]-NMR studies have shown this replacement requires monoseleno-phosphate, which is generated from ATP and selenide by selenophosphate synthetase (SelD).¹²⁰ Subsequently, a tRNA 2-selenouridine synthase was purified from *S. typhimurium*, and *in vitro* analysis demonstrated the enzyme to be capable of converting thiouridine into selenouridine in the presence of SelD. Interestingly, although this modification involves an 'ATP-activated' intermediate, the S to Se substitution reaction was reported not to be ATP-dependent.¹²¹ Further studies illustrated that the biosynthesis of mnm^5Se^2U as well as that of mnm^5s^2U is IscS-dependent, supporting the notion that mnm^5s^2U biosynthesis is most likely followed by the formation of mnm^5Se^2U .¹²² Larson and coworkers recently identified an *E. coli* gene, *ybbB*, that encodes a 43 kDa selenouridine synthase. *In vitro* assays showed that the purified C-terminal polyhistidine-tagged enzyme catalyzes mnm^5Se^2U synthesis in the presence of $SePO_3$. Through mutagenesis studies, they also demonstrated a conserved Cys97 (located in the so-called rhodanese sequence motif, CC⁹⁷XXG) of the selenouridine synthase is required for 2-selenouridine formation *in vivo*.¹²³

6.20.4 Deamination

RNA editing is a post-transcriptional event that changes one canonical nucleotide into another in nascent RNA transcripts, especially in mRNAs. Since this has the potential to change the amino acid identity of the pertinent codon, RNA editing has been shown to play an extremely important role in generating protein diversity and/or regulating translation specificity. Of all the RNA editing events, deamination is probably the one that has been most extensively studied. In this section, two representative deamination modifications, adenosine to inosine (A to I) and cytidine to uridine (C to U) conversions are discussed.

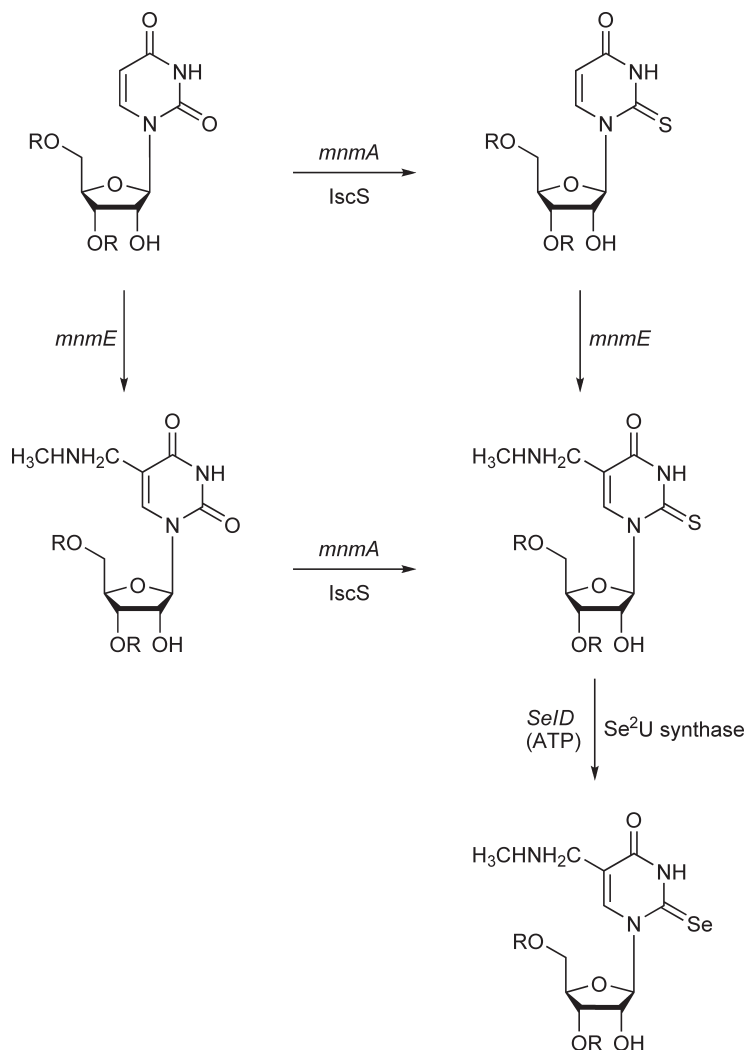


Figure 14 Proposed pathway for 2-selenouridine synthesis. Note: The $\text{mnm}^5\text{s}^2\text{U}$ to $\text{mnm}^5\text{Se}^2\text{U}$ conversion requires ATP to generate SePO_3 intermediate in the presence of *SeI/D*. The S to Se substitution is thought to be catalyzed by 2-selenouridine synthase, which is not yet fully characterized. The mechanism was proposed by Mihara *et al.* Reproduced from H. Mihara; S. Kato; G. M. Lacourciere; T. C. Stadtman; R. Kennedy; T. Kurihara; U. Tokumoto; Y. Takahashi; N. Esaki, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 6679.

6.20.4.1 Adenosine Deamination (A to I)

Adenosine deamination converts the amino group at the C⁶ carbon of adenosine into a carbonyl moiety, forming the modified nucleoside inosine (I), which is considered to play a role in increasing the genetic diversity due to its capability of base pairing with A, C, and U.¹²⁴ This base modification has been extensively observed in both tRNAs and mRNA precursors (pre-mRNAs) from various species.⁹ In eukaryotic tRNAs, adenosine deamination takes place either at the wobble position (A34) or A37 (in the latter case, it is usually followed by methylation to form N¹-methylinosine, m¹I). However, tRNA^{Arg} is the only known tRNA that is susceptible to A to I conversion at position 34 in prokaryotes.¹²⁵ Modification on pre-mRNAs requires a double-stranded RNA (dsRNA) duplex structure, which is formed by intramolecular base pairing between an exonic sequence encompassing the modified adenosine and a downstream intronic sequence.^{126–128} Owing to recognition of different substrates, the enzymes that catalyze A to I conversion have been classified as Adenosine Deaminases Acting on RNAs (ADARs) and (ADATs) Adenosine Deaminases Acting on

tRNAs.^{125,129,130} Nevertheless, sequence analysis has shown a high degree of similarity between these two families of enzymes, especially for a conserved catalytic domain (also known as the deaminase domain). One striking difference is that ADATs lack double-stranded RNA-binding motifs.^{130,131} The deaminase domain contains four key amino acid residues essential for catalysis: two cysteines, one histidine, which are responsible for Zn²⁺-chelation, and a glutamate that is believed to facilitate proton shuffling.^{132–134} Similar structural characteristics have also been extensively found in RNA cytidine deaminases (CDARs), suggesting these three groups of enzymes all belong to the same protein superfamily.¹³⁵

To date, three ADARs have been identified in mammalian cells (ADAR1, ADAR2, and ADAR3), whereas only one or two types of ADARs are found in lower metazoan organisms.¹³⁶ In addition to the ability to recognize and modify pre-mRNAs, ADAR1 and ADAR2 have also been shown to process microRNA (miRNA) precursors and thus alter their gene expression or substrate specificity. However, the significance of miRNA editing requires further investigation.^{137,138} On the basis of sequence analysis of the ADAR family, not only do these enzymes contain catalytic deaminase domains, as mentioned earlier, but they also possess dsRNA-binding domains (dsRBDs) that mediate the binding between enzymes and their substrates. Recently, Jantsch and Öhman¹³⁶ have provided a detailed overview on the molecular architecture, substrate specificity, and biological role of ADARs.

The active site conformation and catalytic mechanism of ADARs are similar to those of ADATs, which is discussed later in this section. Intriguingly, while it is shown that a duplex RNA structure is necessary for A to I conversion by ADARs, the issue of how the target nucleosides gain access to the active sites of these enzymes was unclear. To answer this question, a base-flipping mechanism has been proposed and investigated; in 1997, Hough and Bass¹³⁹ first revealed that the C-termini of ADARs exhibit a high degree of similarity to cytidine and adenosine DNA MTases, which are well known to ‘snatch’ their substrate nucleosides out of the double helix into an extrahelical position for modification.^{140,141} Later, substituting a fluorescent nucleic acid base analogue, 2-aminopurine, for adenosine at the editing site, Stephens *et al.*,¹⁴² demonstrated a significant conformational change in the presence of an adenosine deaminase, ADAR2.¹⁴³ This observation is consistent with the results obtained from two N⁶-adenosine DNA MTases, M•*TaqI* and M•*EcoRI*, which both have been shown to modify double-stranded DNA substrates following a base-flipping mechanism.^{144,145}

In 1996, ADATs were first isolated and partially purified from yeast extracts and shown to catalyze A to I conversion at the wobble position of substrate tRNAs. Meanwhile, the use of [¹⁸O]–H₂O has proved that inosine biosynthesis is conducted via a simple hydrolytic deamination instead of a base-exchange mechanism, as the oxygen in inosine originates from water.¹⁴⁶ This process does not require any cofactors and is an efficient single-step reaction. Kinetic studies have determined the *K*_M of yeast ADAT to be 2.3 nmol l⁻¹ with respect to one of its natural substrates, tRNA^{Ser}. Further studies have demonstrated that two recombinant yeast ADATs, Tad1p, and Tad2p/Tad3p (heterodimer), are essential for I37 and I34 formation, respectively.^{130,147} Subsequently, the first prokaryotic ADAT, *E. coli* TadA, was cloned and purified. TadA corresponds to the Tad2p/Tad3p heterodimer in yeast as *E. coli* does not encode a Tad3-like protein but TadA is able to form a homodimer *in vitro*.¹⁴⁸ Site-directed mutagenesis has shown that the anticodon arm of tRNA^{Arg}, or more specifically the loop sequence U-A-C-G from positions 33 to 36, is sufficient for substrate recognition. This finding was recently confirmed by a crystal structure of *S. aureus* TadA in complex with anticodon stem-loop of tRNA^{Arg}.¹⁴⁹ A momentous conformational change in tRNA^{Arg} was observed in the TadA–tRNA minihelix complex, where a splayed anticodon loop (nucleotides 33–37) is formed to maximize contacts with the enzyme surface. Most recently, crystal structures of TadAs from four other bacterial species (*A. aeolicus*, *A. tumefaciens*, *E. coli*, and *S. pyogenes*), have also been determined and the overall architectures of these enzymes are highly homologous, especially the Zn²⁺ binding sites.^{134,149–152}

These structures provide insight into the active site of TadA. A putative catalytic mechanism has been proposed using the *E. coli* enzyme as an example (Figure 15). Briefly, the active site of *E. coli* TadA (Figure 16) contains a zinc ion coordinated by His68, Cys98, Cys101, and one molecule of water, whereas Glu70 forms a hydrogen bond with the zinc-bound water acting as a general acid/base during catalysis. The reaction mechanism begins with attack at C⁶ of the purine ring by H₂O to generate a tetrahedral intermediate. Subsequently, Glu70 facilitates a proton transfer from water to the amino group on the C⁶ carbon followed by the elimination of ammonia.¹³⁴ A recent study also revealed the transition state structure of *E. coli* TadA via kinetic isotope effects (KIEs) and has provided direct evidence to support this proposed mechanism.¹⁵³

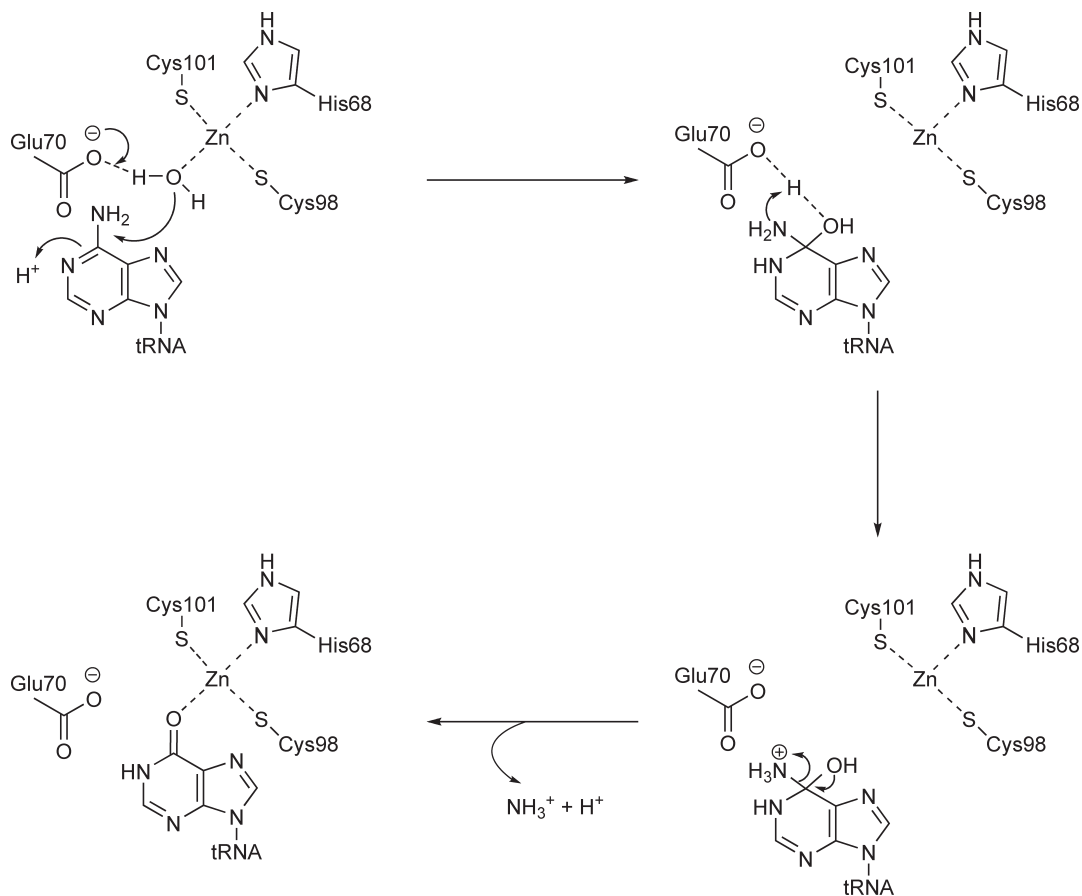


Figure 15 Proposed catalytic mechanism for adenosine deamination of tRNA by TadA. Note: *Escherichia coli* TadA numbering. The mechanism was proposed by Kim *et al.* Reproduced from J. Kim; V. Malashkevich; S. Roday; M. Lisbin; V. L. Schramm; S. C. Almo, *Biochemistry* **2006**, *45*, 6407.

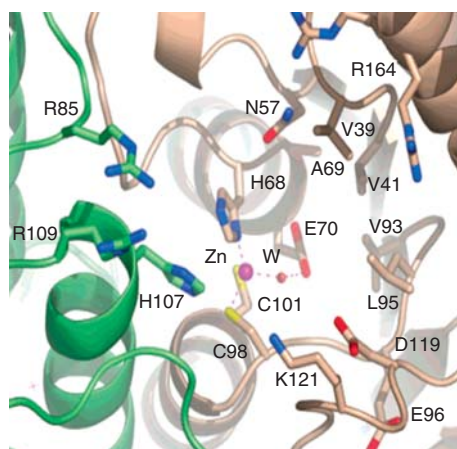


Figure 16 The active site of *Escherichia coli* TadA. The tetrahedrally coordinated zinc ion (Zn) and nucleophilic water (W) are presented as pink and red spheres, respectively. Two subunits of this homodimeric enzyme are shown in green and brown. Key residues considered responsible for catalysis and tRNA substrate recognition are also indicated. Reprinted with permission from J. Kim; V. Malashkevich; S. Roday; M. Lisbin; V. L. Schramm; S. C. Almo, *Biochemistry* **2006**, *45*, 6407. Copyright (2006) American Chemical Society.

6.20.4.2 Cytidine Deamination (C to U)

Similar to adenosine deamination, cytidine deamination replaces the amino group at the C⁴ carbon of cytidine with a carbonyl group and forms uridine (U). As with the A to I conversion, C to U conversion has been observed in both mRNA precursors and tRNAs. However, in most of the cases that have been found so far, C to U editing takes place at pre-mRNAs, as it was first identified in the mRNA precursor of apolipoprotein B (apoB).^{128,154,155} This reaction changes a glutamate codon (C-A-A) to a stop codon (U-A-A) at position C6666 of the *apoB* gene, resulting in the early termination of translation and the production of a truncated protein (termed apoB48, in contrast to the full-length protein apoB100). Earlier studies have shown that this modification is site specific. Through *in vitro* conversion assays, it was demonstrated that a conserved 26 nucleotide-long sequence (6662–6687) spanning the editing site is essential and sufficient for substrate recognition.¹⁵⁶ Site-directed mutagenesis with respect to this motif further suggested that the active site of cytidine deaminase likely binds to an A-U-rich region in an 11 nucleotide-long sequence (6671–6681), annotated as a mooring sequence, located 4–6 nucleotides (the so-called spacer element) downstream of the cytidine to be modified.^{157,158} Moreover, Hersberger and Innerarity¹⁵⁹ found that in addition to the mooring sequence and spacer element, more distal motifs (termed as efficiency elements) flanking the editing site are conserved in various mammalian species and required for maximal efficiency of C to U conversion. Unlike mRNA substrates for adenosine deamination, those for cytidine deamination do not seem to require the modified site to be within a double-stranded duplex structure, and a transient base-flipping theory thereby does not apply to the editing nucleotide in those substrates. Instead, they appear to form a stem-loop secondary structure exposing the cytidine to be modified within the loop and allow presentation of that specific cytidine to the enzyme active site.^{160,161}

The best understood RNA-modifying, cytidine deaminase, currently known as APOBEC-1 (ApoB mRNA Editing Enzyme Catalytic polypeptide 1), was first identified in 1993 from the rat small intestine cDNA library.^{162,163} Protein sequence analysis of recombinant APOBEC-1 shows homology to cytidine deaminase from *E. coli*, yeast, humans, and so on, including three essential zinc-chelating residues. In addition, upon treatment with a zinc chelating agent, the enzyme activity is abolished.¹⁶⁴ However, it has been revealed that APOBEC-1 alone is not sufficient for deamination of cytidine and requires the formation of a multisubunit complex with at least one auxiliary protein, termed ACF (APOBEC-1 Complementation Factor) or ASF (APOBEC-1 Stimulating Factor), to exhibit *in vitro* editing capability.^{165–167} ACF/ASF contains three RNA recognition motifs and has been shown to interact with apoB mRNA *in vitro* and *in vivo*, as evidenced by UV cross-linking and immunoprecipitation studies. In addition, ACF-depleted rat liver extracts showed no C to U conversion, and an ACF/APOBEC-1 complex could be coimmunoprecipitated from transfected cells.¹⁶⁶ Nevertheless, the holoprotein complex seems to contain more auxiliary components, such as ABBP-1 or ABBP-2/HEDJ, whose importance in apoB RNA editing is still under investigation.¹⁶⁸

The crystal structures of *E. coli* and yeast cytidine deaminase were solved by Betts *et al.*¹⁶⁹ and Xie *et al.*,¹⁷⁰ respectively. Homology modeling of mammalian APOBEC-1 with those two structures suggested that they all share high similarity in tertiary and quaternary structures.^{170,171} Similar to adenosine deaminases, cytidine deaminases are homodimeric zinc-containing proteins and catalyze C to U conversion via a hydrolytic mechanism as described earlier. Recently, Maris *et al.*,¹⁷² reported the NMR structure of a 31 nucleotide-long apoB mRNA stem loop. Together with *in vitro* modifying assays, the authors proposed how this secondary structure is recognized by ACF and how the substrate cytidine can be subsequently presented to the active site of APOBEC-1. They concluded that under physiological conditions and in the presence of the enzyme, the modified cytidine is formed at the 5' end of an octa-loop. The NH₂ group of cytidine is buried inside the loop and not easily accessible. Binding of ACF triggers a conformational change by 'melting' the RNA stem-loop structure, allowing APOBEC-1 to approach its target nucleotide (Figure 17). This finding was indirectly supported by Navaratnam and coworkers, where they observed that APOBEC-1 was active in the absence of ACF and had an optimal editing temperature of approximately 45 °C, while ACF could broaden the temperature range with respect to RNA editing activity, lowering the optimal temperature to physiological temperature.¹⁷³

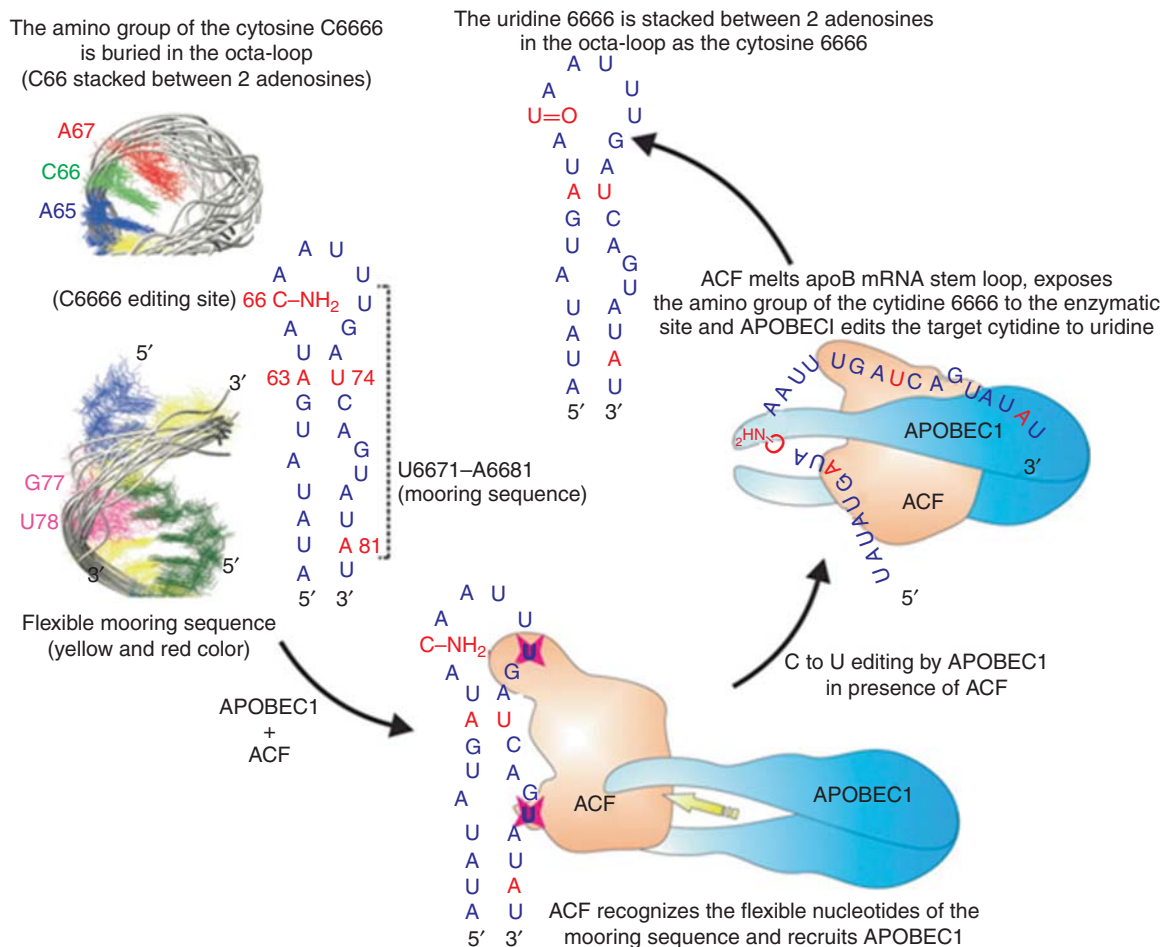


Figure 17 Proposed catalytic mechanism for cytidine deamination of ApoB mRNA by APOBEC1 and ACF. Two uridines thought to be first recognized by ACF are labeled with red stars. Reproduced from C. Maris; J. Masse; A. Chester; N. Navaratnam; F. H.-T. Allain, *RNA* **2005**, *11*, 173, with permission from the RNA Society.

6.20.5 Alkylation

At least four alkylated nucleosides have been found in tRNAs, including N^6 -(dimethylallyl)adenosine (i^6A); lysidine (L); N^6 -(N -threonylcarbonyl)adenosine (ϵ^6A) and (2'- O -ribosylphosphate)adenosine (**Figure 18**). Since the catalytic mechanisms of dimethylallylation and lysidine formation have been studied and proposed recently, these two modifications are discussed in this section. In contrast, very few further investigations on the remaining two alkylation reactions were conducted during the past decade. Therefore, readers are directed to read a previous review by Garcia and Goodenough-Lashua for an overview of threonylation and ribosylation.⁹

6.20.5.1 Dimethylallylation

Dimethylallylation of adenosine (i^6A) refers to the modification occurring at position A37 (3' adjacent to the anticodon) in both prokaryotic and eukaryotic tRNAs. In some of the early literature, this reaction was also known as isopentenylation, while i^6A was termed Δ^2 -isopentenyladenosine. Several decades ago, i^6A was first isolated from yeast.^{174,175} In many organisms, it can be further thiomethylated at C^2 to form ms^2i^6A ,^{176,177} whose biosynthetic mechanism had been discussed earlier in this chapter. i^6A biosynthesis is catalyzed by dimethylallyltransferase, which is encoded by the *miaA* gene in *E. coli*.^{178,179} and the *mod5* gene in yeast.¹⁸⁰

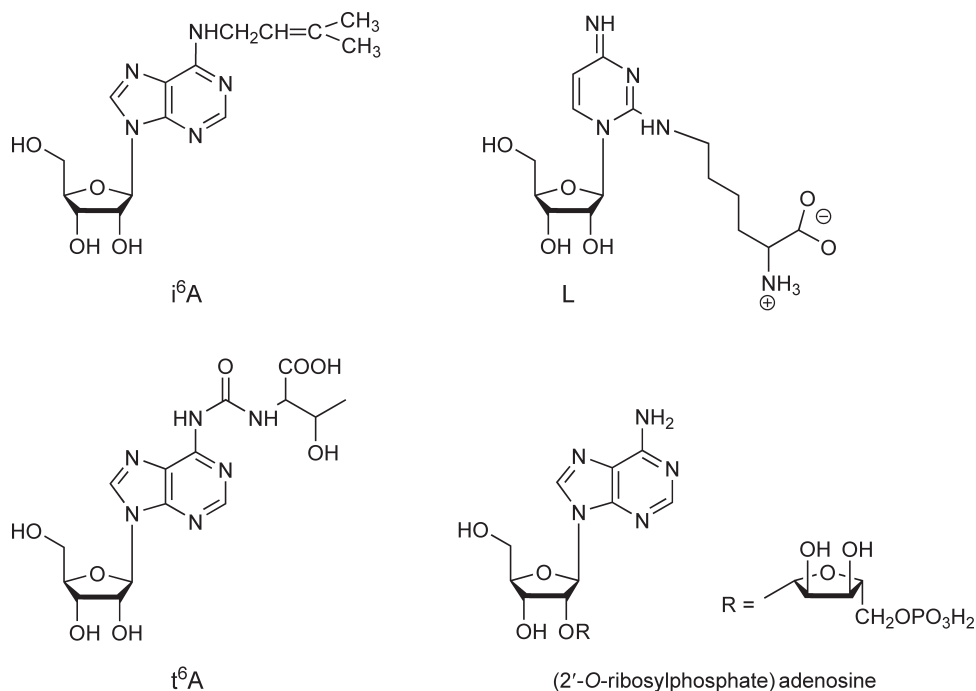


Figure 18 Examples of alkylated nucleosides. N⁶-(dimethylallyl)adenosine (i⁶A), lysidine (L), N⁶-(N-threonylcarbonyl)adenosine (t⁶A), and (2'-O-ribosylphosphate)-adenosine.

The dimethylallyl moiety is transferred from dimethylallyl pyrophosphate (DMAPP), derived from mevalonic acid, to N⁶ of A37 in tRNAs that read codons beginning with a uridine, such as tRNA^{Phe}, tRNA^{Trp}, tRNA^{Tyr}.¹⁸¹

To investigate the substrate recognition of dimethylallyltransferase, *E. coli* MiaA was cloned and purified by two separate groups in 1997.^{182,183} Leung *et al.* found that MiaA can recognize a synthetic minihelix corresponding to the anticodon stem and loop (17-mer) of tRNA^{Phe}, even though its catalytic efficiency (k_{cat}/K_M) was about 17-fold lower than that of the wild-type tRNA^{Phe}, suggesting that the minimal recognition elements reside in the anticodon arm. This observation has also been seen in other tRNA modifying enzymes such as TGT (anticodon arm) and tRNA m⁵U54 MTase (T-arm).^{184,185} Inhibition studies determined that ATP/ADP, GTP, and CTP are all able to competitively (with respect to DMAPP) inhibit MiaA activity *in vitro*,¹⁸² indicating that DMAPP may bind to a conserved ATP/GTP P-loop binding motif,¹⁸⁶ which has been found in MiaA and its homologue, Mod5.¹⁸⁷ In addition, [³H]-DMAPP incorporation assays indicate that the K_M values for tRNA^{Phe} and DMAPP were found to be 2.85 ± 0.4 and 632 ± 131 nmol l⁻¹, respectively,¹⁸² whereas they were reported by Moore and Poulter¹⁸³ as 96 ± 11 nmol l⁻¹ and 3.2 ± 0.5 μmol l⁻¹. In spite of these discrepancies, the $\leq 10^{-7}$ mol l⁻¹ K_M suggests a high affinity between MiaA and its cognate tRNA^{Phe}. Binding studies have suggested that this MiaA-mediated reaction follows an ordered sequential mechanism, where the binding of tRNA occurs before that of the dimethylallyl donor, DMAPP, evidenced by the following observations: a tight-binding interaction between MiaA and tRNA^{Phe} was seen and the dissociation constant was determined to be 5.2 ± 1.2 nmol l⁻¹.¹⁸³ In addition, DMAPP was not able to bind to the enzyme in the absence of tRNA; however, it did bind in the presence of a minihelix anticodon loop of tRNA^{Phe} with A37 replaced by inosine ($K_D = 3.4 \pm 0.6$ μmol l⁻¹).¹⁸³ Using 17 anticodon stem variants of *E. coli* tRNA^{Ser} (G-A-A) and 7 other tRNAs from *E. coli* or yeast, Grosjean and coworkers further elucidated that instead of the entire anticodon loop/arm, sequence A-A-A (corresponding to positions 36-37-38 of the tRNA anticodon loop) plays a crucial role in tRNA recognition by MiaA, which was obtained from a crude *E. coli* cell extract. It was also found that substrate recognition favors retention of the overall 3D tRNA structure and allows slight sequence variations of the proximal anticodon stem. Nevertheless, the authors concluded from the docking model that upon binding of MiaA, the tRNA anticodon arm may be distorted. This altered conformation could result in

maximal contact(s) between tRNA and enzyme active site.¹⁸⁸ Later, Soderberg and Poulter¹⁸⁹ quantitatively determined the steady-state kinetic parameters using purified recombinant MiaA and tRNAs with variations in the anticodon arm. They found that A36G and A38G mutant minihelices showed a 3000- and 40-fold decrease in catalytic efficiency (k_{cat}/K_M), respectively, relative to that of tRNA^{Phe} (wt) minihelix. Meanwhile, kinetic data also indicated the importance of the helix-loop structure in the anticodon arm, where a random coil minihelix analogue with a completely disrupted 5-base pair helix stem resulted in total abrogation of enzyme activity. Nevertheless, results are very similar to those seen with TGT.¹⁹⁰

It has been reported that MiaA is a monomer in solution but might bind its tRNA substrate as a multimer.^{182,183} It was also demonstrated that MiaA-mediated catalysis requires Mg^{2+} , which appears to be involved in stabilization and activation of the leaving pyrophosphate (PP_i) group.^{183,191} Sequence analysis of 28 genes encoding dimethylallyltransferases from various organisms has identified 27 conserved amino acid residues. Site-directed mutagenesis of the *E. coli* MiaA was performed by Soderberg and Poulter¹⁹² to gain insight into substrate binding and catalytic mechanism. They identified 10 possible key residues and proposed that Lys23, Thr24, His67, and Arg217 may be relevant to DMAPP binding while Lys56, Arg167, Arg190, and Lys280 seem to play a role in tRNA binding. In addition, Thr19 and Tyr47 are likely involved in stabilizing the development of positive charge in the electrophilic dimethylallyl group during the alkylation reaction. Recently, Huang and coworkers have published the crystal structure of dimethylallyltransferase from *Pseudomonas aeruginosa* alone and in complex with PP_i at a resolution of 1.9 Å. Interestingly, the enzyme seems to possess a tunnel spanning through the core, where tRNA and DMAPP bind. A docking model constructed based on the crystal structure has suggested that the two substrates most likely access each other from opposite sites of the tunnel due to interactions with specific residues on the protein surface, and binding of tRNA possibly leads to a conformational change in the enzyme active site to accommodate DMAPP (Figure 19).¹⁹³ Together with the results obtained from mutagenesis studies, a possible catalytic mechanism of dimethylallyltransferase has been proposed (Figure 20). Briefly, the substrate tRNA approaches from one side of the tunnel and the amino group of A37 forms a hydrogen bond with the conserved Asp37 residue, which is also assumed to act as a general base during catalysis. The allylic substrate subsequently binds and interacts with the P-loop residues (GXTXXGKT) and Mg^{2+} via its PP_i moiety. Finally, a nucleophilic attack occurs at the carbon adjacent to the PP_i group and transfers the dimethylallyl moiety to the tRNA substrate.^{192,193}

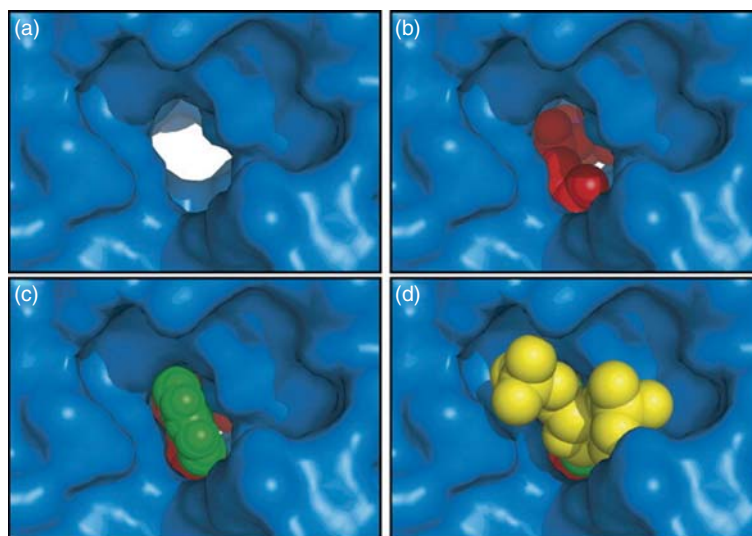


Figure 19 A model of the active site tunnel from *Pseudomonas aeruginosa* dimethylallyltransferase. (a) The binding tunnel with no substrate bound; (b) the tunnel with DMAPP, a magnesium ion and three water molecules coordinated to the magnesium ion (all in red); (c) the tunnel containing all ligands in (b) and the adenine (green); (d) the binding site containing all ligands in (b) and an adenosine including 5' and 3'-phosphates (with the ribose and phosphates in yellow). Reproduced from W. Xie; C. Zhou; R. H. Huang, *J. Mol. Biol.* **2007**, 367, 872, with permission from Elsevier.

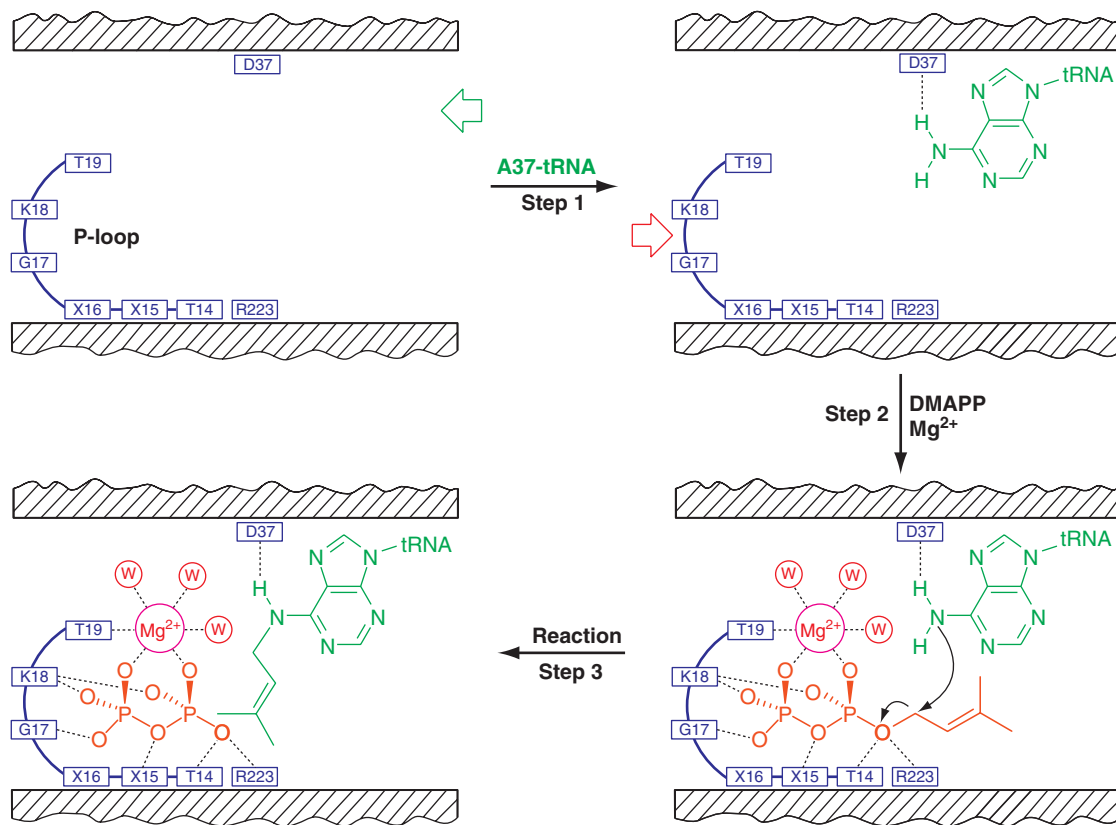


Figure 20 Proposed catalytic mechanism for dimethylallylation of tRNA. Two substrates, tRNA and DMAPP, enter the active site tunnel from opposite sites as indicated by the green and red arrows. Coordinated magnesium ion (Mg^{2+}), water (W), and key residues in the active site are also shown. Reproduced from W. Xie; C. Zhou; R. H. Huang, *J. Mol. Biol.* **2007**, 367, 872, with permission from Elsevier.

6.20.5.2 Lysidine Formation

Lysidine, a lysine-containing cytidine derivative, has been found at the wobble position (C34) of A-U-A codon-specific tRNA^{Ile} in eubacteria. Lysidine formation results from the attachment of a lysine residue via the ϵ -amino group to the C² position of cytidine. Even though it was first isolated from *E. coli* decades ago,¹⁹⁴ the chemical structure of lysidine was determined by Muramatsu *et al.*,¹⁹⁵ using proton NMR and mass spectrometry several years later. Instead of recognizing guanosine, lysidine actually forms a base-pair interaction with adenosine, converting the codon specificity of tRNA^{Ile}, with the genetically encoded anticodon C-A-U, from A-U-G (methionine) to A-U-A (L-A-U anticodon). As a consequence, this modification can also alter the translation specificity between methionine and isoleucine. Interestingly, replacement of C34 of tRNA^{Ile} with L34 leads to a significant switch from methionine-accepting activity to isoleucine-accepting activity. Kinetically, the initial isoleucylation velocity for C-A-U-bearing tRNA^{Ile} was less than 1/10 of that for L-A-U-bearing tRNA^{Ile}, while the methionylation of L-A-U-bearing tRNA^{Ile} was observed. The initial methionylation velocity was more than 1/3 of that for tRNA^{Met}.¹⁹⁶ Using a reverse genetic approach together with mass spectrometry, the enzyme responsible for lysidine formation was identified and named tRNA^{Ile}-lysidine synthetase (TilS). Recombinant TilS showed enzymatic activity with respect to lysidine formation *in vitro* in an ATP-dependent manner, consistent with the fact that a conserved SGXDS sequence (ATP binding motif, the so-called P loop) has been found in the N-terminal domain.¹⁹⁷ Intriguingly, TilS homologues are yet to be found in eukaryotic cells, since eukaryotes produce tRNAs^{Ile} with either the anticodon I-A-U or U-A-U instead of C-A-U to decode the A-U-A codon, as well as A-U-C and A-U-U codons.^{198–200}

Recently, the crystal structure of Tils from *A. aeolicus* was determined at a resolution of 2.42 Å.²⁰¹ The structure of the *E. coli* enzyme had been reported and deposited in the protein data bank (PDB) as a protein of unknown function named MesJ. Structural analysis revealed that Tils is a homodimer and each subunit is composed of an N-terminal dinucleotide-binding domain, a long α -helical linker, and a C-terminal globular domain. Not too surprisingly, the N-terminal domain shares high structural homology with the ATP-PPase domain of GMP synthetase,²⁰² suggesting that the two enzymes might exhibit a similar chemistry during catalysis. Therefore, a putative two-step reaction has been proposed to elucidate the catalytic mechanism of Tils, as shown in **Figure 21**. Briefly, C34 of tRNA^{Ile} is adenylated at position C² along with the generation of PP_i. (This is similar to the activation of carbonyl oxygens via adenylation used by thiolation enzymes discussed previously.) The second substrate lysine, which is presumably activated at a subdomain appended to the N-terminal domain, can then nucleophilically attack the adenylated C² carbon to form lysidine. In addition, this proposed mechanism implies that tRNA^{Ile} and lysine likely bind to Tils in a sequential order, while a possible lysine-binding site was identified near the ATP-binding site. Site-directed mutagenesis and *in vitro* lysidine incorporation studies have suggested several enzymological characteristics of Tils. In summary, Ser32, Ser37, Phe61, Arg113, and Tyr114 appear to be involved in ATP recognition while Glu140, Arg174, and Arg205 are responsible for recognizing C34 of tRNA^{Ile}. After adenylation, Asp191, Thr193, and Asn194 are thought to interact with α -amino and α -carboxyl groups of lysine when Asp36 and Asp137 most likely deprotonate and activate the ϵ -amino group. Furthermore, it was also shown that the C-terminal domain of *A. aeolicus* Tils plays a key role in distinguishing tRNA^{Ile} from tRNA^{Met} via a specific base C20 and the C29–G41 base pair in the anticodon stem of tRNA^{Ile}. In contrast, the *E. coli* enzyme seems to accomplish the discrimination in a different manner, by globally recognizing the anticodon stem (G27–U43) and acceptor stem (C4–G69 and C5–G68).^{201,203}

Using *E. coli* Tils and [³²P]-ATP, Suzuki and coworkers were able to verify the adenylated intermediate at C34 of tRNA^{Ile} (generated by *in vitro* run-off transcription). Not only was PP_i detected in the Tils hydrolysis reaction using [γ -³²P]-ATP, but labeled AMP was also observed prior to addition of lysine in a lysidine formation assay using [α -³²P]-ATP. The intensity of the radiolabel was reduced significantly in the presence of lysine, and a C34G mutant of tRNA^{Ile} was labeled over the course of the reaction. In terms of substrate specificity, the authors found that a minihelix anticodon stem-loop structure could not be recognized by Tils, suggesting that the whole 3D structure of tRNA might be required. Intriguingly, in contrast to *A. aeolicus* Tils, the *E. coli* enzyme has an additional C-terminal domain that presumably contributes to tRNA recognition. This observation is consistent with the fact that Tils from these two bacteria do not exhibit cross-reactivity with respect to tRNA^{Ile}, demonstrating a species-specific tRNA recognition pattern for Tils.²⁰³ More recently, Kuratani *et al.*,²⁰⁴ cocrystallized the *A. aeolicus* Tils with ATP, Mg²⁺, and lysine at a resolution of 2.5 Å and elucidated the active site architecture in more detail. They have revealed that the active site of Tils consists of a large hole for ATP binding and a narrow tunnel for lysine to settle in; however, in contrast to what was proposed previously, these two binding sites are distant prior to tRNA^{Ile} binding. It seems likely that a conformational change must occur later during the second step of catalysis and brings the lysine molecule closer to adenylated tRNA^{Ile}.²⁰⁴ In addition, the presence of Mg²⁺ in the active site provides evidence for the formation of adenylated cytidine during catalysis since the Mg²⁺-coordination usually facilitates the nucleophilic attack on the α -phosphate of ATP and the release of the PP_i group.^{191,204–206}

6.20.6 Reduction/Oxidation

One of the most abundant modified nucleosides found in tRNA is the nonaromatic base, dihydrouridine (D, **Figure 22**), found in the ‘D-loop’ (named after the presence of dihydrouridine). Incorporation of this nonplanar molecule is interesting because it has the potential to interrupt base stacking within the RNA molecule. Very little is known about the enzymes that catalyze this modification. Believed to be catalyzed by flavin-dependent enzymes utilizing NADPH as a cofactor, the reduction of uracil to dihydrouracil is thought to follow a mechanism similar to the biochemically characterized homologous dihydroorotate dehydrogenase (DHODH) and dihydropyrimidine dehydrogenase (DHPDH) enzymes.²⁰⁷ For comparison, **Figure 23** illustrates the proposed catalytic mechanism for class IA DHODHs.²⁰⁸ Recently, a role for human dihydrouridine

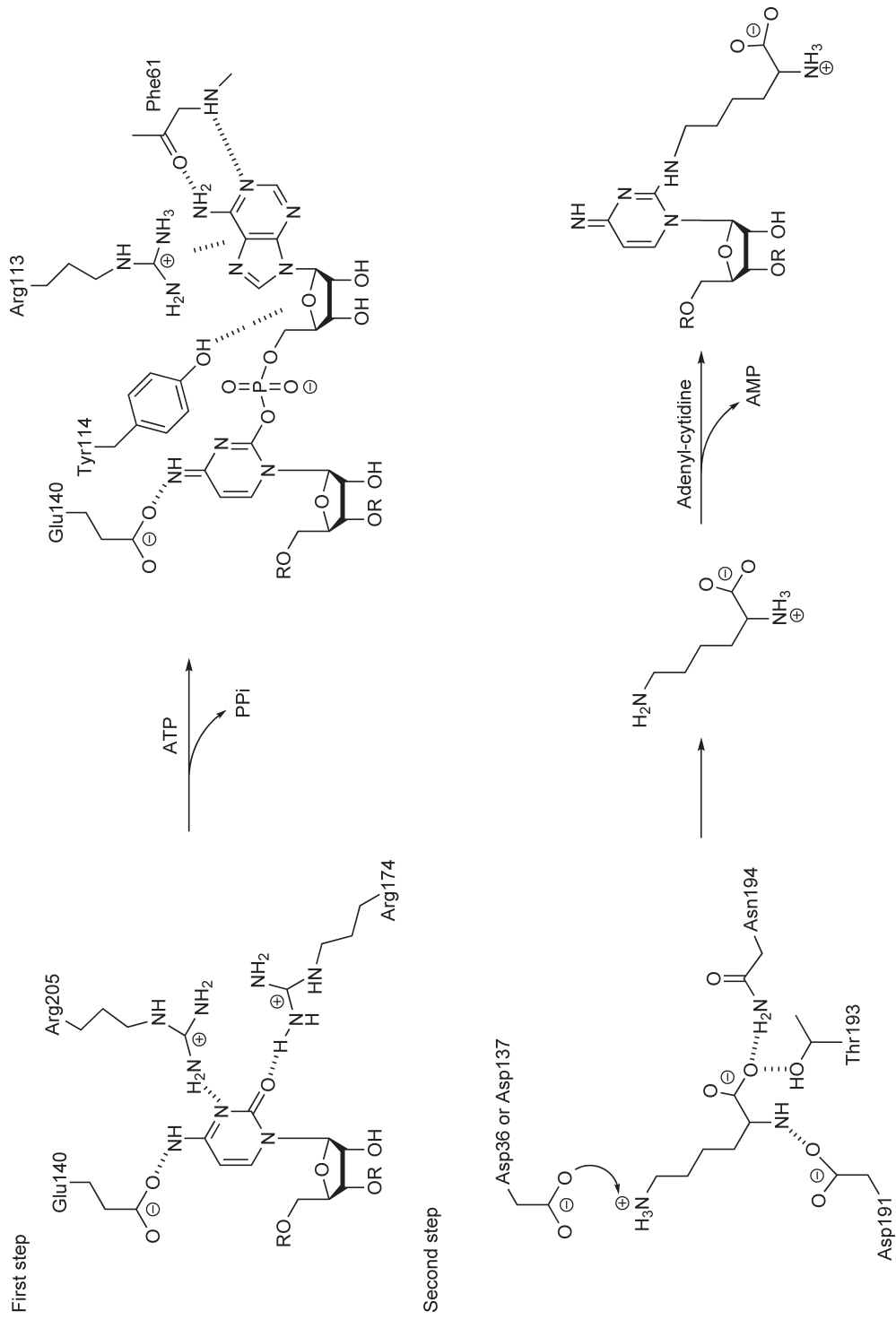


Figure 21 Putative catalytic mechanism for lysidine formation by TlS. Note: *A. aeolicus* TlS numbering. This two-step reaction is proposed to be ATP-dependent and generate an adenylylated intermediate. Key residues thought to be involved in catalysis and substrate recognition are indicated. This mechanism was proposed by Nakanishi *et al.* Reproduced from K. Nakanishi; S. Fukai; Y. Ikeuchi; A. Soma; Y. Sekine; T. Suzuki; O. Nureki, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7487.

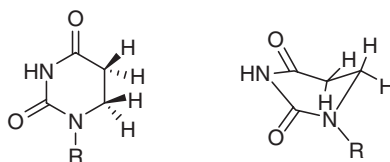


Figure 22 Structure of dihydrouridine. Note that the reduction of the 5–6 double bond removes the aromaticity and planarity of the ring.

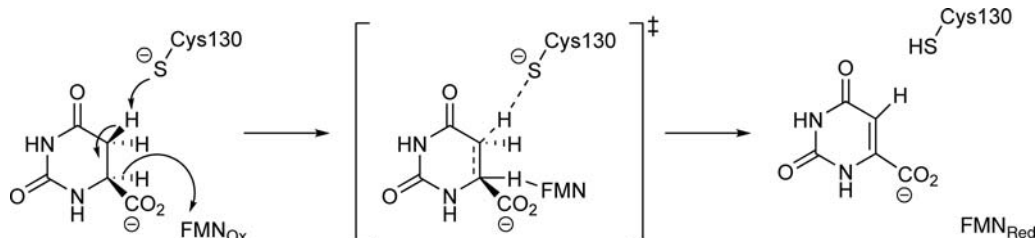


Figure 23 Proposed catalytic mechanism for the Class IA dihydrourate dehydrogenase based on initial characterization of the enzyme isolated from *L. lactis*. This mechanism was proposed by Fagan *et al.* Reproduced from R. L. Fagan; K. F. Jensen; O. Bjornberg; B. A. Palfrey, *Biochemistry* **2007**, *46*, 4028.

synthase-2 (hDUS2) has been proposed in the signaling pathway of protein kinase PKR, where the two proteins were shown to interact in a yeast two-hybrid system.²⁰⁹ There is much interest in this class of enzymes in not only studying this unusual modified nucleoside, but also the biological relevance of this relatively simple transformation.

Difficulty with enzyme identification and purification has hindered the study of these modification enzymes. As demonstrated with some of the elusive MTase genes, comparative genomics were initially utilized to determine potential dihydrouridine synthase (DUS) candidates.²⁰⁷ Three genes were identified in *E. coli* from the COG0042 family: *yjbN*, *ybdG*, and *yobI*. Single, double, and triple knockout mutant analyses revealed that although each enzyme is responsible for some portion of the total dihydrouridine content in RNA, none but the triple deletion mutant resulted in a total loss of the modification. The authors propose that each enzyme may recognize different RNA substrates; interestingly, the authors note that the COG0042 gene cluster does not contain sequence similarity with known RNA-binding motifs (RRMs), which could suggest a novel mechanism for substrate recognition. Initial catalytic information was provided for YjbN in a follow-up report addressing alanine mutations of conserved residues in putative and confirmed DUS sequences.²¹⁰ Briefly, because purification of *E. coli* DUSs proved difficult due to solubility issues, *in vivo* studies were carried out in which the triple knockout strain was complemented with plasmid-borne variants of the DUS. Although no definitive structural information is available, the authors demonstrated that catalytic activity was abolished with mutation of the following conserved amino acids: Asn111, Cys114, Lys153, Arg155, His186, and Arg188.

As a complement to the work performed in eubacteria, the genes responsible for the dihydrouridine modification have also been studied in eukaryotes. Dihydrouridine is commonly found in six positions of yeast tRNA at U16, 17, 20, 20a, 20b, and 47.²¹¹ As before, a genomic analysis in *S. cerevisiae* identified four genes as potential DUSs: YML080 (Dus1), YNR015 (Dus2), YLR401 (Dus3), and YLR405 (Dus4).²¹² The authors demonstrated that the extent of dihydrouridine formation was enhanced in yeast extracts upon addition of a mixture of NADPH/NADH, indicating that the DUS utilizes at least one of these cofactors for catalysis. Binding assays revealed initial insights into tRNA substrate specificity, where Dus1 bound pre-tRNA^{Phe} and pre-tRNA^{Tyr}, and Dus2 recognizes the pre-tRNA molecules for tyrosine and leucine. This suggests that although there may be overlap in tRNA recognition, the modification of dihydrouridine in tRNA^{Phe} is exclusive to Dus1 activity. Further, ribonuclease digestion analysis revealed that Dus1 catalyzes the D modification at U16 and U17. Additional mutational analyses revealed that Dus2 modifies position 20 in tRNA^{Leu}, whereas Dus4 converts U20a and U20b into dihydrouridine.²¹¹ Dus3, in contrast, was shown by primer extension to modify outside of the D-loop, site-specifically targeting U47 in the variable loop of tRNA^{Tyr}.

The first purified DUS was characterized recently *in vitro*.²¹³ The *Thermotoga maritima* synthase, TM0096, and Dus2 from *S. cerevisiae* were chosen for kinetic analysis, as the crystal structure of TM0096 had revealed binding of flavin mononucleotide (FMN) and demonstrated structural homology to the flavin-dependent enzyme DHODH.²¹⁴ Palfey and coworkers determined that tRNA purified from *E. coli* (Δ dus2) extracts were better substrates than tRNA transcribed *in vitro*, suggesting that the presence of other modified nucleoside(s) is important for this DUS activity. This initial *in vitro* analysis will serve as a foundation for future biochemical analysis to further elucidate the mechanism of catalysis of this interesting class of modification enzymes.

6.20.7 Transglycosylation

Of all modified bases known, three are unique in their mechanism of incorporation into RNA. Queuosine (Q), archaeosine, and pseudoU (Figure 24) are each generated by the cleavage and reformation of the glycosidic bond to the ribose 1' carbon. This occurs via a 'transglycosylation' reaction. In the cases of queuosine and archaeosine, the reaction is intermolecular, with the modified base replacing the genetically encoded guanine. In the case of pseudoU, the genetically encoded uridine is isomerized intramolecularly to pseudoU, a 'C-nucleoside' (a carbon-carbon glycosidic bond), with no evidence for incorporation of exogenous uracil. (Note: we have recently reviewed the issue of transglycosylation-based RNA modification.²¹⁵ This section reiterates and updates that review.)

Queuosine and archaeosine are known to occur only in tRNA, with queuosine in the anticodon wobble position of tRNAs with anticodons of G-U-N in eukarya and eubacteria. Archaeosine occurs in position 15 of the 'D' loop of the majority of archaeal tRNAs. Both these bases (or their precursors) are incorporated into tRNA via the action of TGT, which is found in all three kingdoms.²¹⁶ PseudoU is generated by a family of pseudoU synthases^{217,218} and is found in various positions in tRNAs as well as rRNAs, and snRNAs (small nuclear RNAs).²¹⁹

6.20.7.1 tRNA-Guanine Transglycosylase

TGT is the key enzyme involved in the biosynthesis of queuosine and archaeosine (see next section for a discussion of queuosine biosynthesis). TGT occurs in all three kingdoms of life with very few exceptions (yeast and mycoplasma). In eukarya, queuine (the base of queuosine) is a dietary factor (found in many sources²²⁰) and is incorporated directly into tRNA by TGT.²²¹ In eubacteria, TGT incorporates a queuine precursor (preQ₁, 7-aminomethyl-7-deazaguanine, note: we use the purine numbering system) into tRNA, which is further modified to queuine. In archaea, TGT incorporates preQ₀ (7-cyano-7-deazaguanine) into position 15 of substrate tRNAs²²² (Figure 24). This is subsequently converted into the amidine (archaeosine) by an unknown process. TGTs across the three kingdoms exhibit 20–30% sequence identity, whereas within each kingdom the sequence identity is higher (30–40%). Sequence identities include key active site amino acids and a structural zinc-binding motif that was originally identified in the eubacterial TGT via mutagenesis and biochemical studies²²³ and ultimately confirmed after solution of the TGT X-ray crystal structure.²²⁴

The eubacterial and archaeal TGTs are clearly the best studied in terms of structure, mechanism, and substrate recognition. The eukaryl TGTs are much less understood, although they have been isolated from a number of sources.^{221,225–228} It has been proposed that the eukaryl TGT is a heterodimer with one catalytic subunit homologous to the eubacterial TGT and a second subunit that is proposed to regulate the activity of the eukaryl TGT in response to phosphorylation by protein kinase C.²²⁹ USP14, a 60 kDa ubiquitin-specific protease, has been proposed²³⁰ to be the eukaryl TGT regulatory subunit; however, rigorous evidence supporting this hypothesis is lacking. Very little work has been carried out to characterize the eukaryl TGT, although some work has been done regarding the heterocyclic substrate and inhibitor specificity.²³¹

The chemical and kinetic mechanisms of the eubacterial TGT have been the subject of intense study. Although very few details remain unclear, the main features of the TGT catalytic mechanism are now well established. Kinetically, TGT follows a ping-pong mechanism where tRNA binds first, forming a covalent complex with the enzyme followed by the dissociation of guanine.²³² The heterocyclic substrate (preQ₁) then binds and finally preQ₁-tRNA is released.

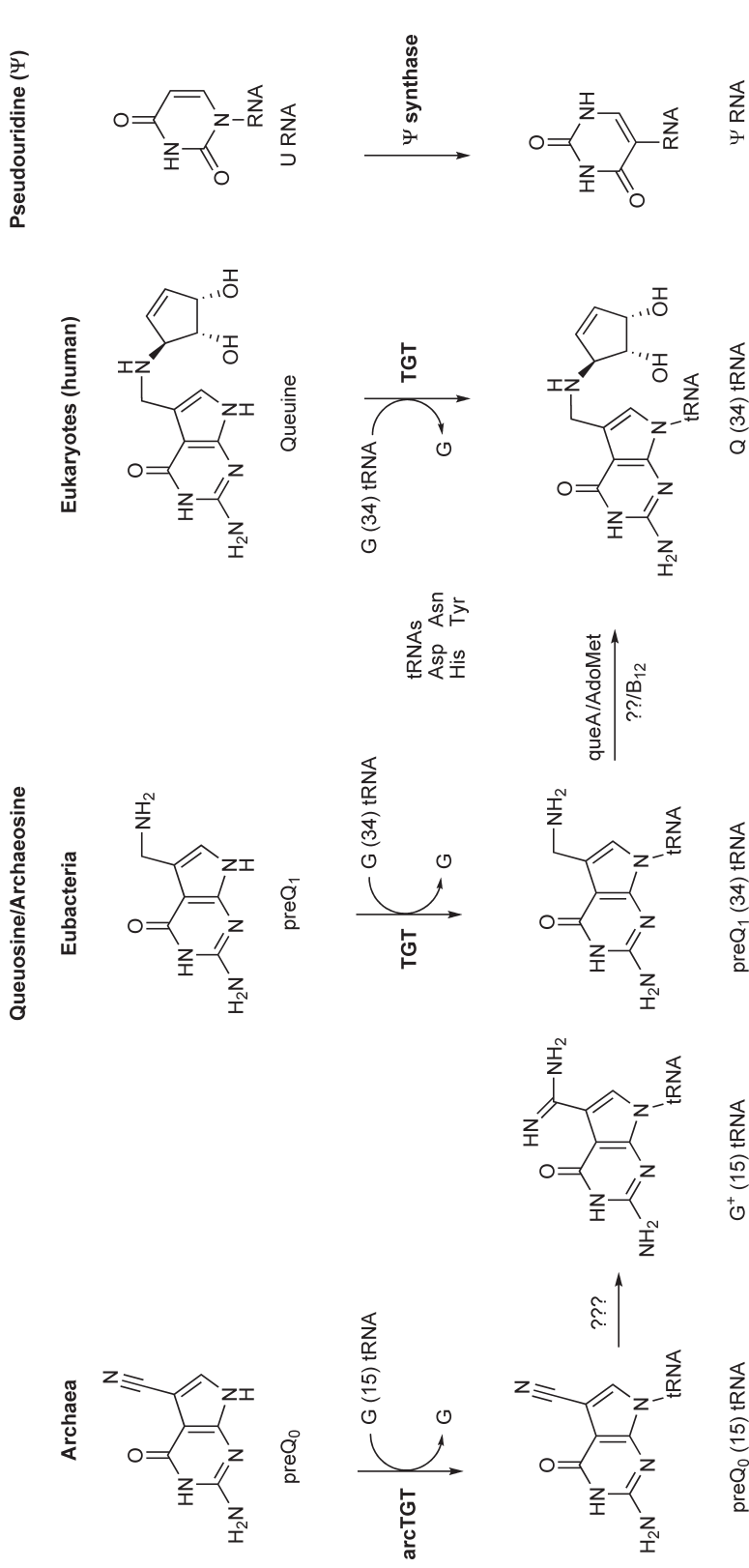


Figure 24 Transglycosylase reactions. The four known transglycosylation reactions involved in RNA modification generating: in archaea, archaeosine (G⁺); in eubacteria and eukarya, queuosine (Q); and across the three kingdoms, pseudouridine (Ψ).

The nature of covalent bond formation has been probed by a number of techniques. It was first identified by SDS–PAGE band shift analyses under mildly denaturing conditions.²³³ Incubation of TGT and tRNA, with no heterocyclic substrate, yielded a band that migrated on SDS–PAGE to approximately 70 kDa (roughly equivalent to a TGT–tRNA complex). This was only observed when the complex was denatured at room temperature, boiling the sample resulted in total loss of the complex band. There are three aspartates in the TGT active site (Asp89, Asp143, and Asp264, *E. coli* TGT numbering), which are absolutely conserved among TGTs across all kingdoms. Aspartate 143 is involved in heterocyclic substrate recognition (see below). Mutagenesis studies of aspartates 89 and 264 revealed that both are absolutely critical to the TGT reaction, with only the conservative glutamate mutants retaining detectable activity and the ability to form the covalent complex.^{233–235} Initial interpretation of the mutagenesis results presented aspartate 89 as the more likely candidate to be the enzymic nucleophile that forms the covalent bond. Subsequently, the X-ray crystal structure of the *Zymomonas mobilis* TGT in covalent complex with a minihelical RNA revealed aspartate 264 as the true enzymic nucleophile (Figure 25).²³⁶ The mutagenesis studies, in combination with the pH profile for the TGT reaction,²³⁷ are consistent with aspartate 89 acting as a general acid to protonate guanine making it a good leaving group. Asp89 also is presumed to act as a general base to deprotonate preQ₁ to facilitate nucleophilic attack and formation of the modified tRNA product. An overall scheme for the chemical mechanism is presented in Figure 26.

A mechanism-based inhibitor (7-fluoromethyl-7-deazaguanine, FMPP) was designed based on the understanding that TGT must deprotonate the incoming heterocyclic substrate prior to glycosidic bond formation and displacement of aspartate 264.²³⁸ The inhibitor was designed to be a very close analogue of the natural substrate preQ₁ to take advantage of its binding affinity. In this case, as the enzyme deprotonates N⁹ (which would normally lead to attack at the ribose C¹), fluoride is eliminated generating a Michael acceptor which then alkylates and inactivates the enzyme. All data were consistent with mechanism-based inactivation and suggest that a covalent bond may be formed between the inhibitor and cysteine 145, although direct evidence for this is lacking.

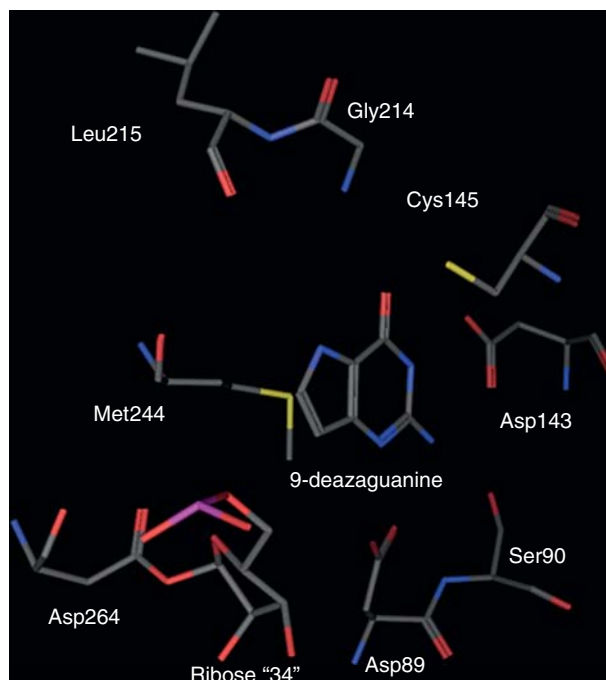


Figure 25 X-ray crystal structure of TGT covalent complex. The structure, solved in the presence of the inhibitor, 9-deazaguanine, clearly shows the covalent bond between Asp264 and the ribose of G34. Note: the figure is shown with the *Escherichia coli* TGT numbering for consistency. The figure is generated from PDB # 1Q2R. Reproduced from W. Xie; X. J. Liu; R. H. Huang, *Nat. Struct. Biol.* **2003**, *10*, 781–788.

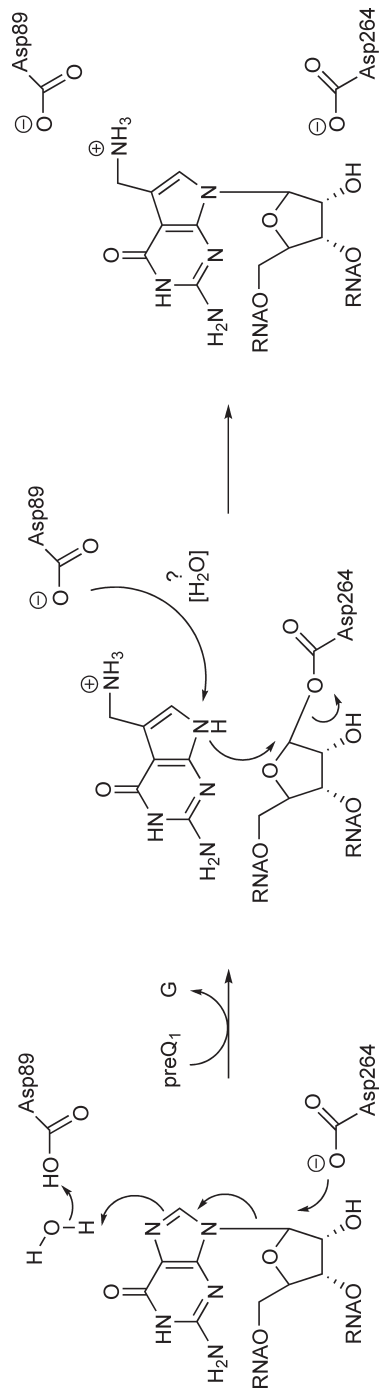


Figure 26 A chemical mechanism for TGT involving a covalent intermediate via Asp264 and acid-base catalysis by Asp89 is consistent with all known data.

The TGT reaction is actually quite slow, $k_{\text{cat}} \sim 0.1 \text{ min}^{-1}$.²³⁹ Rapid-quench kinetic studies have been performed to elucidate the rate-limiting step in the TGT reaction.²⁴⁰ (Basically, the reaction is rapidly quenched and analyzed by mildly denaturing SDS-PAGE to assess the rate of formation of the covalent complex.) These studies revealed that the apparent rate of covalent complex formation (including all steps from free enzyme and free RNA to covalent complex) is about 100 times faster than k_{cat} . The rate of covalent complex breakdown is technically more challenging to assess, but preliminary data suggest it is also much faster than k_{cat} . Preliminary results of tRNA binding and dissociation kinetics (S. M. Chervin and G. A. Garcia, unpublished) are consistent with product release being rate limiting.

TGTs across the three kingdoms are very similar and very distinct in their recognition of the heterocyclic substrate. One common factor involves the recognition of the amino pyrimidone portion of the heterocyclic substrate. This is absolutely crucial to maintain the correct Watson-Crick base pairing of the wobble nucleotide. As mentioned above, aspartate 143 is absolutely conserved across all known TGTs. Mutagenesis and computational studies have confirmed that aspartate 143 forms two hydrogen bonds with the aminopyrimidone portion of both guanine and preQ₁ (Figure 27).²⁴¹ Mutation of aspartate 143 to asparagine results in the inversion of substrate specificity (in terms of K_M) from a guanine-like substitution pattern to xanthine.²⁴²

Where the TGTs across the three kingdoms differ in their substrate recognition involves the 7-position substituent (i.e., aminomethyl vs. cyano vs. cyclopentenyldiol-aminomethyl). Romier *et al.*²⁴³ proposed a structural basis for this discrimination based on homology modeling of active sites of eukaryal and archaeal TGTs compared to the eubacterial TGT structure. In the eukaryal TGT, cysteine 145 (*E. coli* numbering) is replaced by valine, and valine 217 is replaced by glycine. These two changes (along with two others, which are less striking and more variable) result in an expansion of the active site generating a binding pocket for the cyclopentenyldiol side chain of queine that is not accommodated by the eubacterial TGT. In the case of the archaeal TGT, there are a number of amino acid changes (cysteine 145 and tyrosine 148 both to proline, valine 217 to leucine) along with a predicted movement of the main chain that result in the rigidification/constriction of the active site such that only the linear cyano side chain of preQ₀ can be accommodated. Interestingly, the eubacterial TGT recognizes guanine, preQ₀, and preQ₁ with comparable efficiency, although preQ₁ is a significantly better substrate. Recent studies have shown that glutamate 219 (235 in *Z. mobilis*) serves as a key role in mediating a conformational switch in the TGT active site that switches substrate selectivity between guanine/preQ₀ and preQ₁.²⁴⁴ Mutation of glutamate 219 to glutamine results in a much higher K_M for preQ₁.

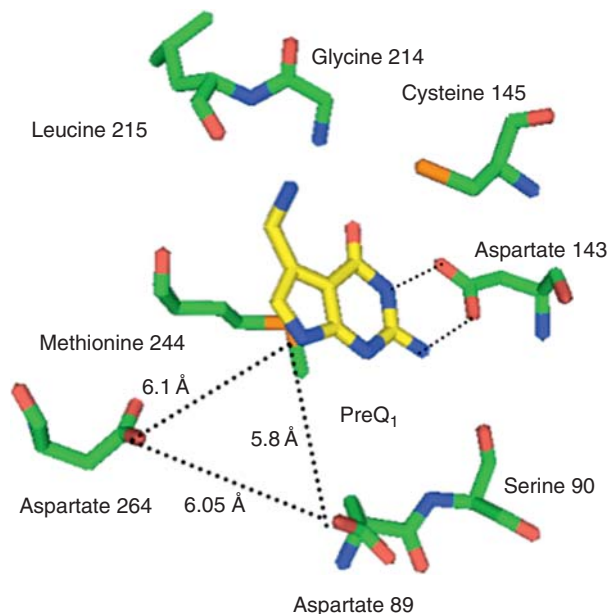


Figure 27 Hydrogen-bonding network for PreQ₁ in the TGT active site. The structure of the *Zymomonas mobilis* TGT•preQ₁ complex (PDB File: 1POE) shows that preQ₁ interacts with three aspartates in the TGT active site. Asp143 is responsible for recognition of the aminopyrimidone moiety. Aspartates 89 and 264 form a roughly equilateral triangle with the N⁹ of preQ₁.

tRNA recognition by the eubacterial and archaeal TGTs has been extensively studied. Queuine occurs in the wobble position (substituting for the genetically encoded G34) of the anticodon of tRNAs coding for aspartate, asparagine, histidine, and tyrosine. These tRNAs share the anticodon sequence $\underline{\text{G}}\text{-U-N}$, and it is the $\underline{\text{G}}$ that is exchanged for Q. Earlier it was shown that the eubacterial TGT did not require the entire tRNA molecule for recognition *in vitro*.¹⁹⁰ This is in contrast to the eukaryal TGT, which has been shown in *Xenopus* oocytes, to require the entire tRNA architecture for recognition *in vivo*.²⁴⁵ Further studies of the eubacterial TGT have identified a minimal recognition motif consisting of a U-G-U sequence (corresponding to positions 33-34-35 of the tRNA anticodon loop) in a minihelix loop.¹⁸⁴ *In vitro* studies of minihelical RNAs in which the U-G-U sequence was permuted around the loop suggest that the loop conformation influences recognition.²⁴⁶ Particularly noteworthy is the observation that shifting the U-G-U sequence one base forward in the loop, such that it now encompasses the anticodon, resulted in loss of recognition if U33 was present. In most tRNAs, U33 causes a specific conformation of the anticodon loop (the so-called 'U turn') that is likely present to enhance the ability of the anticodon to anneal to the mRNA codon. This lack of recognition effectively prevents TGT from modifying one of the threonine tRNA isoacceptors (anticodon U-G-U).

A number of RNA modifying enzymes exhibit 'multi-site' specificity, some within a single tRNA and some between tRNA and other RNA species. This issue has been addressed for the eubacterial TGT by a number of studies. In addition to recognizing a minihelical RNA, the eubacterial TGT has been shown to recognize, *in vitro*, an unusual dimeric form of tRNA^{Tyr,247} a site within an unmodified T Ψ C loop of tRNA^{Phe,248} and most recently a site within the mRNA for a pathogenic bacterial virulence gene.²⁴⁹ In fact, studies have shown that TGT will recognize a DNA minihelix as long as the thymidines are replaced with 2'-deoxyuridines.²⁵⁰ Whether these modification events occur *in vivo* and if they are physiologically relevant remains to be determined; however, these findings do indicate that TGT can modify other RNA species.

The archaeal TGT has been shown to recognize its tRNA substrate in a dramatically different conformation (' λ form', **Figure 28**).²⁵¹ Archaeosine (G⁺) occurs in position 15 of the tRNA. This position is critically involved in tertiary interactions between the D arm and the variable loop that assist the tRNA in adopting its canonical 'L-shaped' conformation. This core region of the tRNA molecule has a high number of modified bases and these modifications are assumed to contribute to the structural stability of the tRNA. Unmodified *E. coli* tRNA^{Tyr} exhibits a melting temperature 5–10° lower than the modified tRNA.¹⁸⁴ In hyperthermophiles, this difference is larger, 15–20°. The first event in the thermal unfolding of tRNA is the 'melting' of the tertiary interactions, followed by helical dissociation/unwinding, consistent with base modification stabilization of tRNA structure. Two factors could potentially contribute to the arcTGT recognition of the λ form tRNA. The

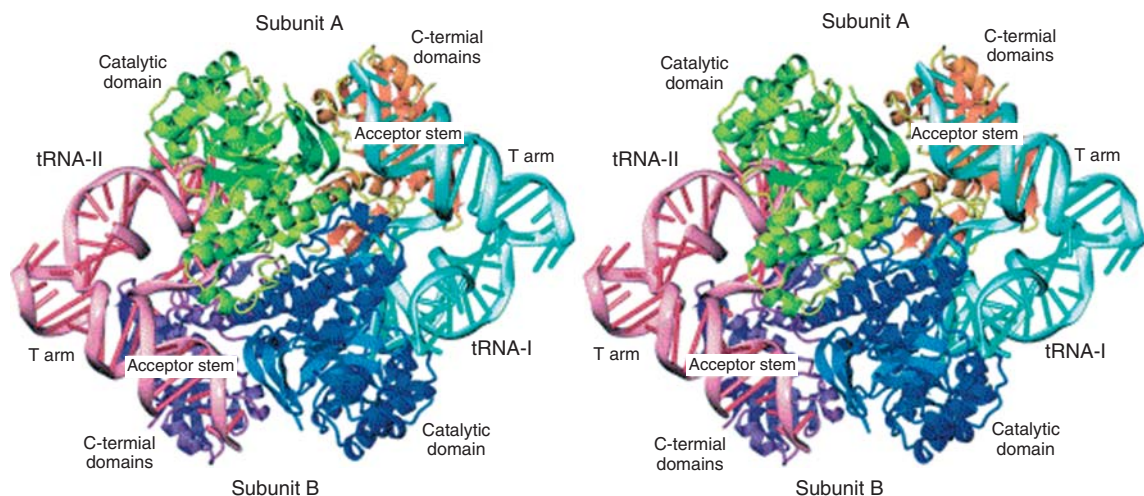


Figure 28 Structure of the archaeal TGT bound to tRNA^{Val}. Note that the D-loop of the tRNA has been pulled away from its normal interaction in the tRNA and protrudes into the enzyme active site. The tRNA is in the ' λ ' conformation. Reproduced from R. Ishitani; O. Nureki; N. Nameki; N. Okada; S. Nishimura; S. Yokoyama, *Cell* **2003**, 113, 383–394, with permission from Elsevier Limited.

first is that the unmodified tRNA is likely to have more conformational flexibility in this region. The second is that in order to gain access to G15 to carry out the transglycosylation, the arcTGT may induce the λ form conformation. The arcTGT also contains a C-terminal domain that is not present in TGTs from other kingdoms.²⁵² Interestingly, this domain is shared by pseudoU synthases and has been termed the PUA domain. While the PUA domain is clearly involved in RNA recognition, it has been shown that the PUA domain of the *Pyrococcus furiosus* arcTGT is not required for archaeosine formation *in vitro*.²⁵³

It is perhaps worthwhile to make a note about the origin of TGT. It has been proposed that TGT in eubacteria and eukarya arose via convergent evolution.²⁵⁴ However, more recent data are more consistent with divergent evolution. For example, while the overall pathways for queuosine biosynthesis are dramatically different between eubacteria and eukarya, the key transglycosylation step carried out by TGT is identical. There is a 'kingdom-based difference' in TGT tertiary structure; however, present evidence is consistent with the eubacterial TGT being a homodimer²³⁶ and the eukaryal TGT being a heterodimer.^{225–227} Analysis of the human genome (and other eukaryal genomes) has revealed a gene for TGT (the catalytic subunit of the heterodimer) that has a great degree of sequence homology with the eubacterial and archaeal TGTs (20–30% identity, including key active site aspartates and the zinc-binding motif). There are differences in heterocyclic substrate recognition; however, homology modeling has attributed these differences to specific amino acid substitutions in the active site.²⁴³ No obvious difference in tRNA recognition is seen between the eukaryal and eubacterial TGTs, as the same tRNAs are modified in the same positions. This is not the case for the archaeal TGT, although recognition of a guanine in the third position of a minihelical loop motif is conserved. Regulation by phosphorylation^{229,255} is significantly different; however, one could assume the case that this has evolved in eukarya. The presence of TGT in all three kingdoms of life and the high degree of sequence and structural (at least between eubacteria and archaea²⁵²) homology clearly indicate that TGT has existed for a very long time, predating the division of the three kingdoms. It seems fairly clear that the present forms of TGT have arisen via divergent, not convergent evolution.

6.20.7.2 Pseudouridine Synthase

Pseudouridine (Ψ , pseudoU) is the most abundant modified nucleoside in RNA and has been referred to as the 'fifth nucleoside'. PseudoU is found in a plethora of positions within tRNAs as well as in a number of positions in rRNAs and in snRNAs. It is quite likely that the true extent of pseudoU occurrence in RNA species will not be fully understood for some time. One of the difficulties in detecting RNA base modifications is the relatively low abundance of many RNA species *in vivo*. Unlike queuosine and archaeosine, where one could theoretically synthesize radiolabeled forms of the modified base and use this to probe for RNA species that are substrates for TGT, less direct and possibly less sensitive methods must be used to detect pseudoU formation.^{256–258}

To date, there have been six families of pseudoU synthases identified.²⁵⁹ The families are named for the first pseudoU synthase in each family that was cloned; TruA, TruB, TruD, RluA, RsuA, and the newest Pus10. While there is little detectable sequence homology between the families, their catalytic domains share a common structure and a catalytic aspartate that is conserved across the families. This suggests that the catalytic mechanism is conserved, but the mode of RNA recognition most likely differs. Structural similarity between members of the TruA, TruB, RluA, and RsuA families led to the conclusion that they arose via divergent evolution.²⁶⁰ TruA (and presumably its family members) is a homodimer, while the other pseudoU synthases appear to be monomeric. The monomeric pseudoU synthases tend to have a second domain that appears to be involved in RNA recognition. At least one pseudoU synthase has been found to contain a structural zinc.²⁶¹

PseudoU is generated via an isomerization of the genetically encoded uridine. Technically this reaction is an intramolecular transglycosylation. Early hypotheses regarding the chemical mechanism of pseudoU synthase were based on analogies drawn to thymidylate synthetase. It was thought that an active site cysteine might attack at C⁶ of the uracil base yielding a dihydrouridine-like intermediate similar to that leading to methylation at C⁵ (Figure 29(a), X = Cys). Chemical studies have shown that the glycosidic bond of dihydrouridine is labile²⁶² and could open, allowing for the rotation of the uracil and formation of the C-riboside. However, mutagenesis studies have clearly eliminated any key catalytic role for any of the potential cysteines in the active site.^{263,264} Subsequently, more pseudoU synthase sequences were determined, and the absolute conservation of a single aspartate prompted an alternative hypothesis involving this aspartate as the catalytic nucleophile.

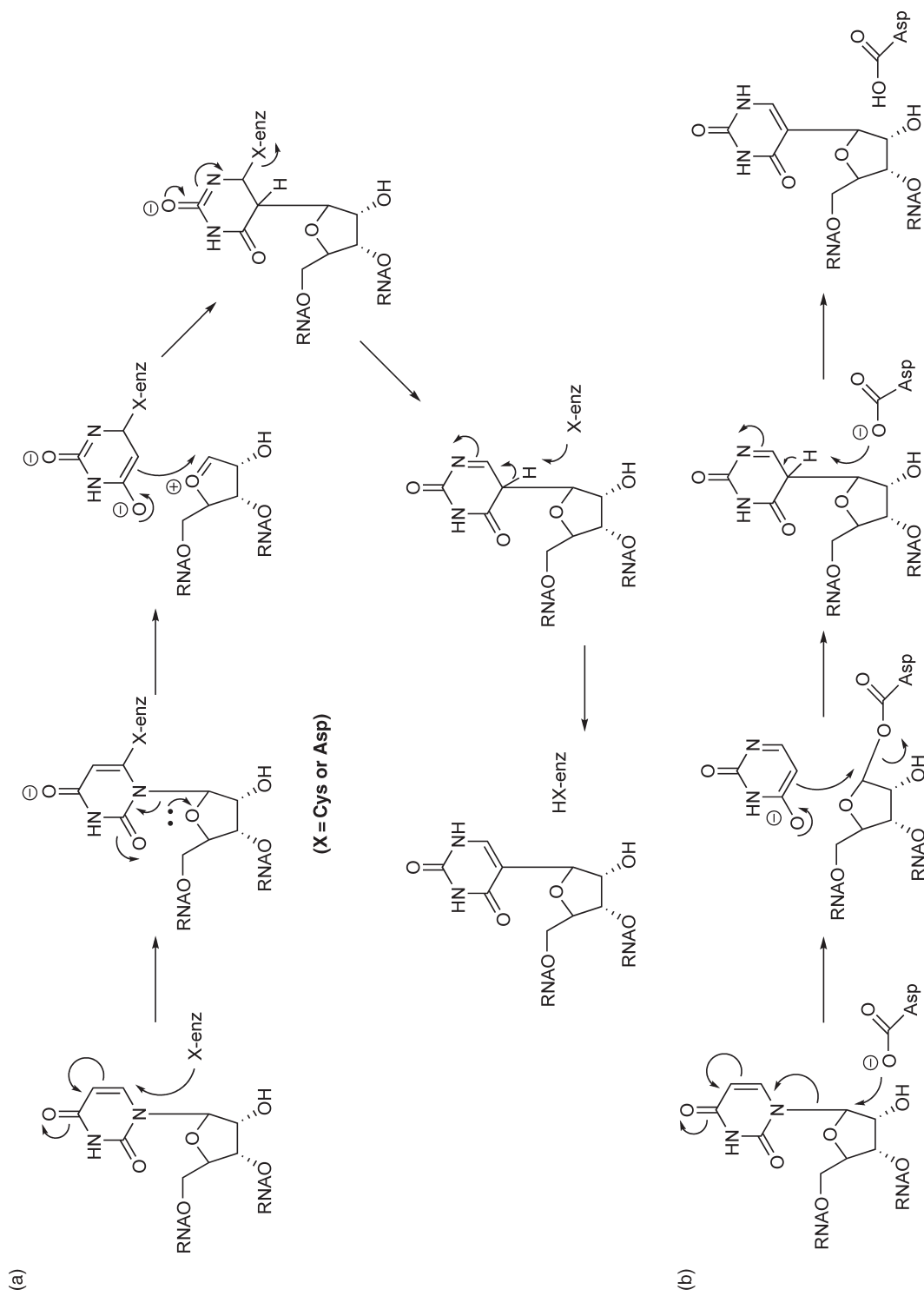


Figure 29 Postulated chemical mechanisms for pseudouridine synthase. (a) Nucleophilic attack at uracil C⁶, 'X' was initially thought to be a cysteine; however, mutagenesis studies eliminated that possibility and identified a conserved and essential aspartate. (b) Nucleophilic attack at ribose C^{1'} by the aspartate analogous to the mechanism followed by TGT.

In 1982, Frendewey *et al.*²⁶⁵ showed that 5-fluorouridine substituted RNA was a potent inhibitor of pseudoU synthase. Studies of 5-fluorouridine RNAs with pseudoU synthase TruA (formerly called pseudoU synthase I) led Santi and colleagues²⁶⁶ to propose a mechanism similar to that for TGT where the aspartate attacks C1' of the ribose, eliminating uracil and forming a covalent intermediate (**Figure 29(b)**). The uracil is then free to rotate and reattack the ribose displacing the aspartate. For the 5-fluorouridine RNA, the 5-fluorouracil presumably cannot attack in this manner, and the enzyme is inactivated. Subsequently, Santi and colleagues²⁶⁷ showed that, upon hydrolysis, the covalent adduct formed between pseudoU synthase and 5-fluorouridine RNA generates a product RNA that contains a 5-fluorouridine that is modified by the addition of water across the 5–6 double bond. This prompted the authors to modify their previous mechanistic proposal to involve aspartate attack at C⁶ of the uridine (**Figure 29(a)**, X = Asp), essentially identical to the original mechanistic proposal except that aspartate, not cysteine, acts as the nucleophile. In the case of the 5-fluorouridine analogue, that reaction is trapped at the covalent adduct, which is then susceptible to hydrolysis, leaving a net addition of water across the 5–6 double bond.

For some time, the pseudoU synthase chemical mechanism seemed firmly established, later the X-ray crystal structure of pseudoU synthase TruB in complex with 5-fluorouridine RNA was solved.²⁶⁸ Surprisingly, this structure revealed that the enzyme was not covalently linked to the RNA and that the 5-fluorouridine had isomerized to the 5 glycoside (with fluorine attached) and was hydrated at C⁶ (**Figure 30**). In 2004, Mueller and colleagues looked more deeply at the interaction between 5-fluorouridine RNA and pseudoU synthase TruB.²⁶⁹ They reasoned that if an ester between aspartate and C⁶ were hydrolyzed, then carrying out the reaction in ¹⁸O water would incorporate ¹⁸O into the aspartate. Mass spectrometry clearly indicated that the RNA and not the enzymic aspartate contained the ¹⁸O label. Mueller then returned to Santi's original mechanism (aspartate attack at ribose), where the aspartate, freed after isomerization, activates a water molecule to attack the uracil (**Figure 31**). This mechanism is compelling as it also supports the aspartate acting as a general base, deprotonating C⁵ in the normal reaction generating pseudoU. While it is possible that pseudoU synthases from different families have subtly different chemical mechanisms, it seems more likely that a common mechanism, such as that proposed by Mueller is followed by all.

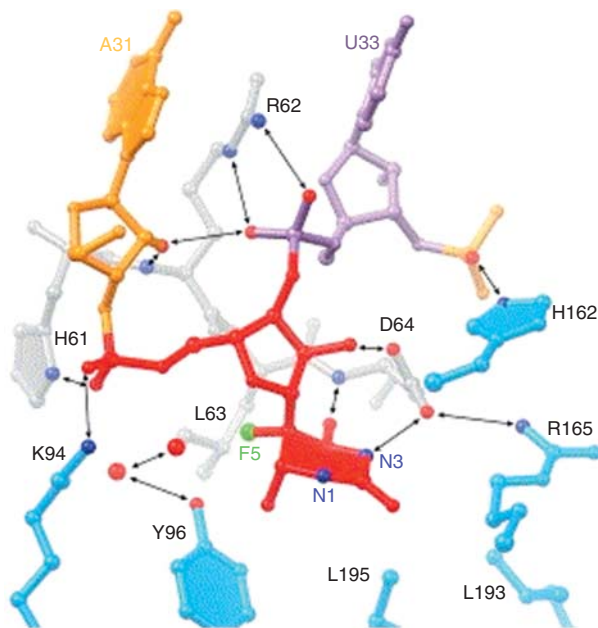


Figure 30 X-ray crystal structure of pseudoU synthase TruB with rearranged 5-fluoroU RNA bound. Note the fluorine attached to C⁶ and the tetrahedral (sp³) hybridization of C⁶. Reproduced from C. Hoang; A. R. Ferré-D'Amaré, *Cell* **2001**, 107, 929–939, with permission from Elsevier Limited.

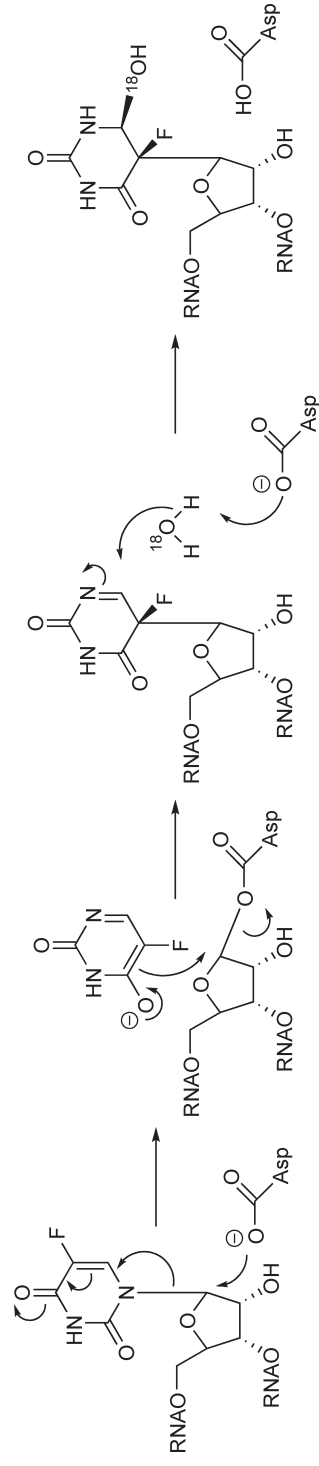


Figure 31 Postulated mechanism for rearranged 5-fluoroU product of pseudouridine synthase TruB. Consistent with H_2^{18}O labeling studies, Mueller proposed a mechanism involving $\text{C}1'$ attack by aspartate. After isomerization and release of aspartate upon formation of the C-glycosidic bond, the aspartate then activates a water molecule to add across the $\text{C}5-\text{C}6$ double bond.

While the catalytic mechanisms for pseudoU synthases are very highly conserved, the mode of RNA recognition is not. (It is beyond the scope of this chapter to fully review RNA binding by pseudoU synthases and the reader is referred to recent reviews for more in-depth discussions.^{268,270–273}) In some cases, pseudoU synthases that act upon tRNAs do so at an early stage of tRNA biogenesis and require the presence of an intron (which is subsequently removed by splicing) for recognition.^{274,275} A number of pseudoU synthases have been found to exhibit multi-site (multiple sites within an RNA) and multi-substrate (multiple RNA substrates) specificity. Examples include positions 38 and 39 in yeast tRNA²⁷⁶ and positions 32 in tRNA, and 746 in 23S rRNA.²⁷⁷

As mentioned above, TruA is a homodimer. It binds to the anti-codon loop/D loop axis of its tRNA substrate across the dimer interface (**Figure 32(a)**) in a quasi-symmetric fashion. The flexibility of the anticodon loop allows for the multisite specificity of TruA (e.g., uridines 38, 39, and 40). Many of the other pseudoU synthases contain a second, RNA recognition domain. For TruB (and the archaeal TGT), this domain has been termed the PUA domain (for PseUdouridine Synthase-Archaeosine TGT). The PUA domain is one example of an RNA-recognition motif (RRM) and is found in a wide range of RNA-binding proteins (and even some proteins including sulfate reductases and glutamate kinases that do not appear to bind RNA).²⁷⁰ Interestingly, the PUA domains in different enzymes utilize different modes of interaction with RNA (**Figure 32(b)**).

The X-ray crystal structure of the human pseudoU synthase, Pus10, has recently been solved.²⁵⁹ This is a novel pseudoU synthase and defines the sixth family of pseudoU synthases. It is responsible for the pseudouridylation at position 55 in tRNAs. The structure of Pus10 reveals a secondary domain that is homologous to the THUMP domain (introduced in the methylation section). Similarly to the PUA domain, the THUMP domain

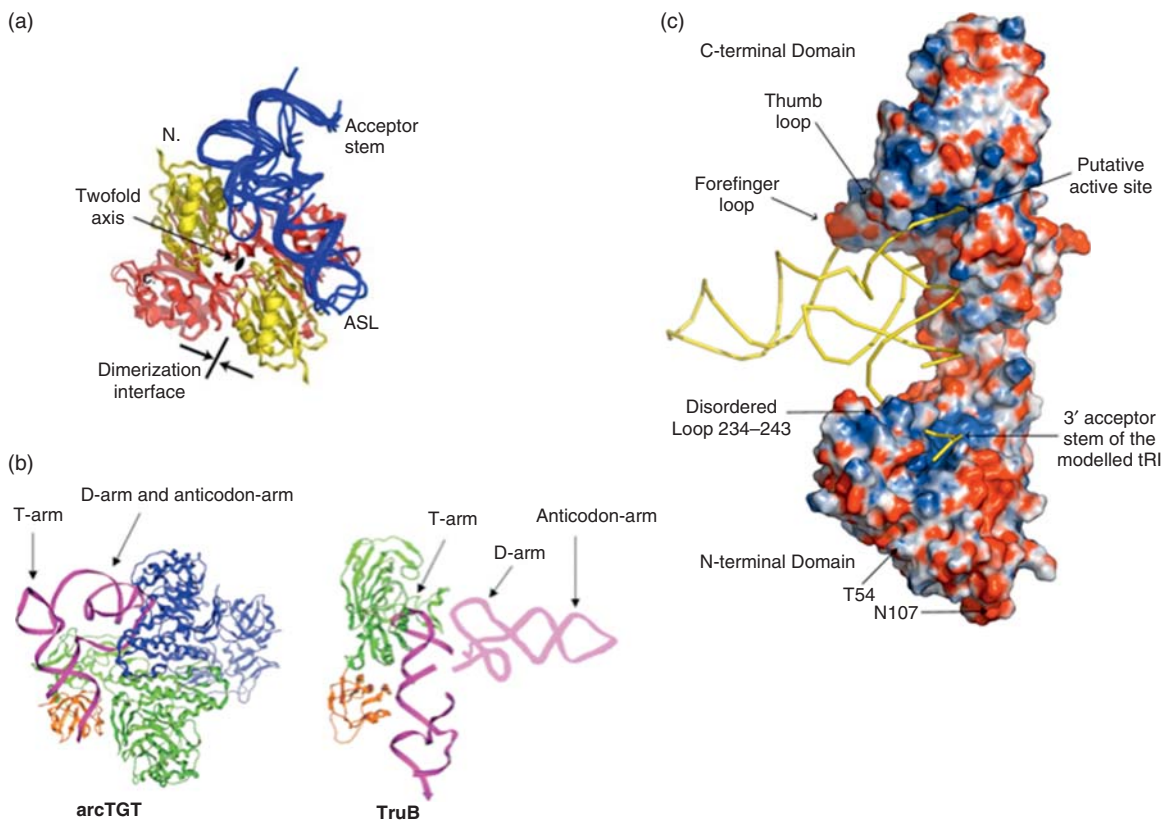


Figure 32 X-ray crystal structures of pseudoU synthases. (a) TruA (homodimer) (Reproduced from S. Hur; R. M. Stroud, *Mol. Cell* **2007**, 26, 189, with permission from Elsevier Limited.), (b) ArcTGT and TruB (PUA domains) (Reproduced from I. Perez-Arellano; J. Gallego; J. Cervera, *FEBS J.* **2007**, 274, 4972, with permission from Wiley–Blackwell Publishing.), and (c) Pus10 (THUMP domain) (Reproduced from C. J. McCleverty; M. Hornsby; G. Spraggan; A. Kreuzsch, *J. Mol. Biol.* **2007**, 373, 1243, with permission from Elsevier Limited.).

is a relatively ancient RRM and has been found in a number of RNA modifying enzymes.^{61,62,101,259} PseudoU synthase Pus10 binds its tRNA substrate across the acceptor stem/TΨC loop axis, orienting the TΨC loop in the active site (**Figure 32(c)**).

There are a number of pseudoU synthases that recognize and modify specific positions within both 23S and 16S rRNAs^{214,277–280} and snRNAs.^{219,281} In these cases, the pseudoU synthases use small, nucleolar RNAs (snoRNAs) as ‘guide RNAs’ to direct them to their substrate and site of action.^{282–285} In the case of *Saccharomyces cerevisiae*, the complete set of guide RNAs has recently been identified.²⁸⁶ This seems like a very efficient mechanism for substrate/site recognition where a single enzyme may recognize many sites. This ‘RNA-guided’ modification of RNA is also used by many RNA MTases (see earlier in this chapter) and in fact some of the same guide RNAs are used by both types of modification enzymes. In at least one case, the intron in a tRNA has been shown to direct the pseudouridylation of the tRNA.²⁸⁷

It seems plausible that pseudoU synthases existed very early in evolution based on the structural and mechanistic conservation of their catalytic domains. Evolutionary divergence has likely occurred, resulting in the low sequence homologies and the varied modes of RNA recognition involving different RRM domains. This has generated the six known families of pseudoU synthases, and it is certainly possible that more remain to be discovered.

6.20.7.3 Other Base Modifications via Transglycosylation?

Transglycosylation, as a means of RNA modification, is currently known only for queuosine, archaeosine, and pseudoU. (Transglycosylation was once thought to be the mechanism for the formation of inosine in certain tRNAs, although later work showed this to result from an adenosine deaminase-type activity.¹⁴⁶) Transglycosylation is an efficient mechanism for the introduction of an elaborately modified base (e.g., queuine) into RNA, especially if the heterocyclic core is altered. However, there are a number of RNA editing (changing from one canonical nucleoside to another canonical nucleoside) events in which the mechanism is uncharacterized and for which transglycosylation is an attractive possibility. These include a U to A conversion in α -galactosidase mRNA²⁸⁸ and a G to U conversion also in α -galactosidase mRNA.²⁸⁹ It is certainly possible, if not likely that other forms of RNA modification/editing may be found that utilize a transglycosylation mechanism.

6.20.8 Complex Modifications

Among the most complex of the known modified nucleosides are those that involve multiple enzymatic transformations, best exemplified by queuosine (Q) and wyosine (Y) and their derivatives. In contrast to some of the enzymes performing more ‘simple’ modifications, much less mechanistic information was known about these interesting modifications until recently. With advances in comparative genomics technologies, several genes/gene clusters have been identified and suggested to be involved in the generation of these complex modifications. This has helped to clarify our knowledge of the biosynthetic pathways of these interesting natural products.

6.20.8.1 Queuosine and Derivatives

Queuosine is a highly modified nucleoside that is conserved across eukaryotic and eubacterial kingdoms, although the details of the biosynthesis of queuosine-tRNA are different. Queuosine biosynthesis has been the subject of recent reviews.^{290,291} Queuosine occurs in the wobble position of the anticodon of tRNAs that code for aspartate, asparagine, histidine, and tyrosine (anticodons of G-U-N) across the two kingdoms. To date, queuosine has not been found in any archaeal species. First discovered in bovine amniotic fluid, Crain *et al.*²⁹² determined the structure of the modified nucleoside through mass spectrometry and NMR (see **Figure 24**, Section 6.20.7). In eubacteria, there is an elaborate pathway for queuosine biosynthesis, while in eukarya, queuine (the free base of queuosine) is a dietary factor. The key (and common) step in queuosine biosynthesis is the incorporation of the modified base into tRNA catalyzed by TGT (reviewed in Section 6.20.7.1).

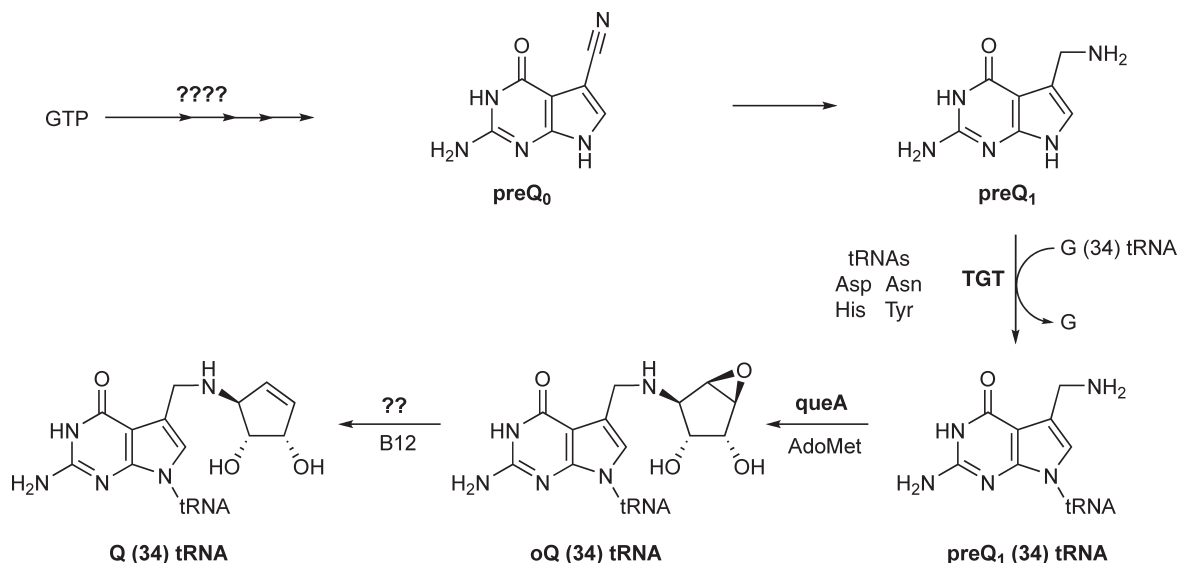


Figure 33 Previously known queuosine pathway in eubacteria. It has been known for some time that the queuosine pathway occurs at two levels: the free heterocyclic base (yielding preQ₁) and at the level of the tRNA. These two levels are linked by the key step of incorporation of preQ₁ into tRNA by TGT.

In eubacteria, the queuosine pathway can be conceptualized as occurring at two levels: the free heterocyclic base and the nucleoside attached to RNA (Figure 33). For many years, it has been known that the key intermediate, preQ₁, is derived from GTP. Owing to lack of evidence, it has long been thought that preQ₁ derives from GTP in a process analogous to that of pterin/folate biosynthesis and/or that for the antibiotic toyacamycin.²⁹³ PreQ₁ is the substrate for the eubacterial TGT which incorporates it into the substrate tRNA. Two subsequent reactions occur at the level of the tRNA in which a cyclopentyl moiety is added and then converted into the cyclopentene.

Recently the early steps of queuosine biosynthesis (before incorporation into tRNA) have been elucidated, initially via comparative genomics.²⁹⁴ Potential gene candidates were selected based on specific criteria, including predicted catalytic mechanism and occurrence of known Q-biosynthetic genes *tgt* and *queA*. *Acinetobacter calcoaceticus* was used as a model organism for constructing mutants of candidate genes, where the effect of each mutation was determined by isolating the tRNA and using HPLC to determine the presence of queuosine in the Q-cognate tRNAs. Four candidates were initially selected in the *ykvJKLM* operon, and single-gene deletions in each case (renamed *queC*, *-D*, *-E*, and *-F*, respectively) resulted in a loss of Q-tRNA. Complementation of the defects with the operon from *B. subtilis* restored the queuosine modification. The enzymatic activity for each gene was predicted by sequence homology: *queC* – ATPase, *queD* – 6-pyruvoyl-tetrahydropterin synthase-like, *queE* – radical AdoMet, *queF* – GTP cyclohydrolase I-like.

Initially, QueF was thought to be responsible for the first step in queuosine biosynthesis based upon its sequence similarity to the GTP cyclohydrolase family. However, very recent studies (D. Iwata-Reuyl, personal communication) have implicated GTP cyclohydrolase I (FolE) as being responsible for the first step in preQ₁ biosynthesis. QueC, the putative ATPase, was characterized further in a mutant strain of *E. coli* that was shown to be deficient in queuosine.²⁹⁵ By studying different mutant strains complemented with a plasmid carrying the *queC* gene, including a *queA* deletion mutant and a strain incapable of converting the intermediate preQ₀ into preQ₁, the authors were able to narrow the activity of QueC to a step in the biosynthetic pathway involved in the conversion of GTP into preQ₀. The exact roles of QueD and QueE remain unclear.

The crystal structure of QueF from *B. subtilis* was solved to 2.25 Å, the protein exhibited a quaternary structure consisting of a dimer of hexamers (a dodecamer) belonging to the ‘tunnel-fold’ superfamily, where substrate binding is believed to be at the subunit interface.²⁹⁶ To confirm the role of the enzyme *in vitro*, the genes from *B. subtilis* and *E. coli* were cloned and expressed.²⁹⁷ After determining that the predicted GTP cyclohydrolase activity was not exhibited by QueF from these two species, alternative roles were explored.

Ultimately, QueF was determined to be an NADPH-dependent preQ₀ oxidoreductase, catalyzing the first identified reduction of a nitrile in a biological system (i.e., the preQ₀ to preQ₁ step).

Further biochemical analysis of this unique biochemical reaction was carried out using the QueF protein isolated from *B. subtilis*.²⁹⁸ The enzyme was found to have maximal activity at pH 7.5 with no apparent metal dependence, although Cu²⁺ and Fe³⁺ were found to inhibit catalysis. In the kinetic mechanism, it is assumed that preQ₀ binds first, followed by two molecules of NADPH with equal affinity for the enzyme. More detailed steady-state kinetics determined K_M values for NADPH in the low micromolar range, and for preQ₀ in the high nanomolar range. QueF was inactivated by iodoacetamide treatment, presumably through chemical modification of a cysteine residue important to catalysis. In alignment studies, Cys55 had previously been suggested to interact with the nitrile moiety of preQ₀, and this finding supports that hypothesis. The authors also determined that preincubation in the presence of preQ₀ protected enzyme activity from iodoacetamide inactivation. Additionally, serine and alanine mutants of Cys55 were inactive, providing further evidence for the importance of this residue in the QueF reaction. On the basis of the above information, Iwata-Reuyl and colleagues proposed a mechanism for the QueF reaction involving a cysteine acting as a catalytic nucleophile and two NADPH (Figure 34).

Recently, the gene *cinQ* from the *cinAQ* copper-inducible operon from *Pseudomonas putida* has been shown to exhibit preQ₀ oxidoreductase activity.²⁹⁹ Although part of a copper-inducible operon, insertional disruption of the *cinQ* gene did not demonstrate a link to copper sensitivity. However, additional studies are required to further probe this link between tRNA modification and copper homeostasis.

As reviewed in Section 6.20.7.1, TGT catalyzes the incorporation of preQ₁ into tRNA. This is then followed by QueA (*S*-adenosylmethionine: tRNA ribosyltransferase isomerase), which is the enzyme responsible for the transfer and isomerization of the ribose moiety of AdoMet to preQ₁-tRNA, resulting in the intermediate epoxyqueuosine (oQ). Mechanistic studies for this reaction were first reported in 2000,³⁰⁰ where Iwata-Reuyl and coworkers sought to determine which ribose carbon of AdoMet was directly attached to the aminomethyl moiety of preQ₁. To distinguish between the possible candidates, C^{1'} and C^{4'}, the authors incubated [1'-¹³C] AdoMet in a TGT/QueA coupled assay *in vitro* to obtain a quantifiable amount of oQ following digestion of the substrate tRNA. As shown in the scheme (Figure 35) the authors provide a mechanism consistent with their result that the isotopically labeled carbon was found at position C³ of the epoxycyclopentandiol ring of oQ intermediate, providing evidence for C–N bond formation occurring at C^{4'}. This mechanism involves an unprecedented utilization of AdoMet as a ribosyl donor and is proposed to proceed via sulfur ylide chemistry.

The authors followed up this analysis with more detailed, *in vitro* kinetics of purified QueA in consecutive papers. First, a pH profile determined the optimal pH to be 8.7 for the purified QueA.³⁰¹ Divalent cations Mg²⁺ and Mn²⁺ showed no effect on enzyme activity in submillimolar concentrations, but did inhibit QueA activity in the millimolar range. A steady-state kinetic analysis under initial velocity conditions revealed K_M values for AdoMet in the micromolar range and for preQ₁-tRNA in the low micromolar range. A minihelix substrate corresponding to the anticodon arm of the substrate tRNA was also characterized, which had a faster k_{cat} when compared with full-length tRNA, but the higher K_M gave an overall lower value for k_{cat}/K_M . Further bi-substrate kinetic analysis revealed a sequential mechanism, where a ternary complex was required for catalysis.³⁰² Inhibition kinetics were conducted with *S*-adenosylhomocysteine (AdoHcys) and sinefungin, an analogue structurally similar to AdoMet; K_i values were determined for AdoHcys in the low micromolar range with respect to AdoMet, whereas sinefungin inhibited the enzyme in the micromolar range. Both inhibitors exhibited competitive inhibition patterns with respect to AdoMet, as well as uncompetitive inhibition in both cases with respect to preQ₁-tRNA. Ultimately, an ordered sequential kinetic mechanism was determined for QueA, where preQ₁-tRNA binds first and AdoMet second, followed by the products released in the order of adenine, methionine, and oQ-tRNA.

X-ray crystal structural studies have also helped our understanding of the QueA active site. The first structure was determined with QueA from *T. maritima* at 2.0 Å, revealing a protein with no structural homologues (via a DALI search).³⁰³ The structure showed mostly polar amino acids, both charged and uncharged, in the domain believed to be involved in tRNA binding. A second structure was determined from *Bacillus subtilis* at 2.9 Å, characterizing more globally the two domains of the monomer, a smaller six-strand β-barrel domain linked to a larger αβα-sandwich domain.³⁰⁴ Several conserved amino acid residues were found in the hinge region between the two domains and in the large domain. The authors speculate that the

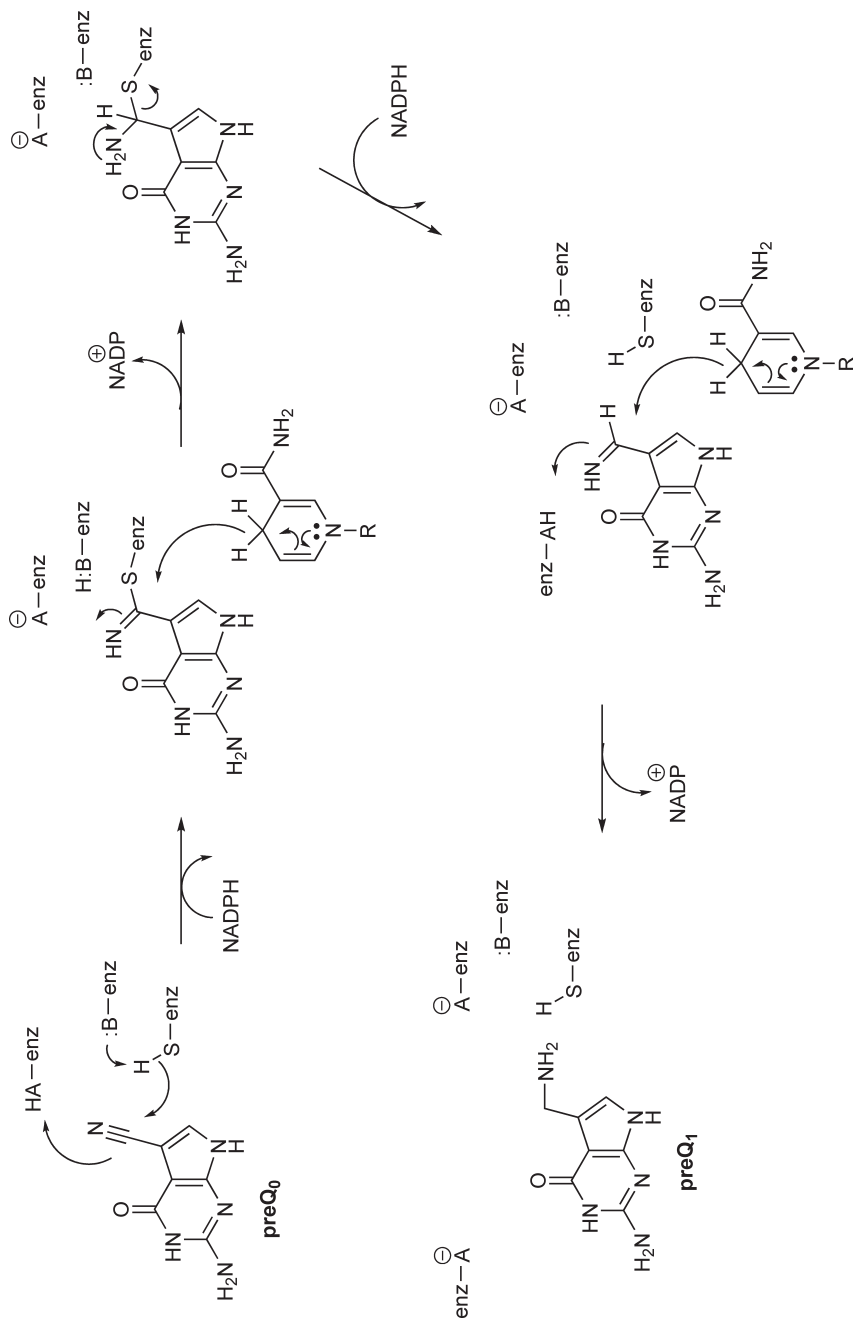


Figure 34 Proposed mechanism for PreQ₀ reduction by QueF. On the basis of their experimental data, Lee *et al.* have proposed a mechanism for QueF that involves a nucleophilic cysteine and two equivalents of NADPH. Reproduced from B. W. K. Lee; S. G. Van Lanen; D. Iwata-Reuy, *Biochemistry* **2007**, *46*, 12844.

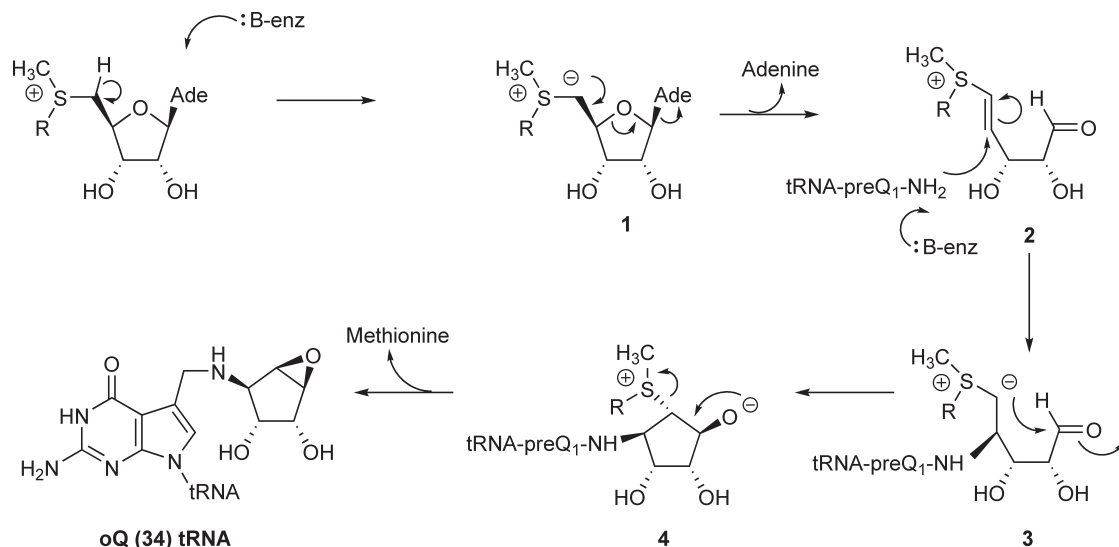


Figure 35 Proposed mechanism for the QueA reaction. The QueA reaction involves a unique utilization of AdoMet and was proposed to proceed via sulfur ylide-type chemistry. Reproduced from S. G. Van Lanen; D. Iwata-Reuyl, *Biochemistry* **2003**, 42, 5312.

catalytic center exists in the interface between the two domains, and believe the small subunit to exhibit a high degree of flexibility, based on an overlay of the two structures. Although the structural determinations have helped clarify our understanding of QueA, both authors agree that an important step will be the cocrystallization of enzyme with either substrate to provide further evidence of the key residues involved in the catalytic mechanism.

Much less is known about the final step of the pathway, specifically the B_{12} -dependent conversion of oQ into queuosine. In his review of queuosine biosynthesis, Iwata-Reuyl draws a parallel between the reduction of oQ and the reaction performed by ribonucleotide reductases and proposes a mechanism involving a thiyl radical and redox-active disulfide (Figure 36).

As mentioned above, in eukarya, queuine is a dietary factor widely dispersed in nature.²²⁰ However, in addition to utilizing free queuine, eukarya have the ability to salvage queuine from either queuine riboside or queuosine monophosphate.^{305–307} A eukaryl gene, *qtrtd1*, has been annotated as a TGT-like protein based on its sequence homology to TGT. In every instance where QTRTD1 has been found, the organism utilizes a TGT that recognizes queuine (J. R. Katze, personal communication). While the zinc-binding site, Asp143 and Asp264 (Glu in QTRTD1) are conserved, QTRTD1 differs from all other TGTs in that Asp89 does not appear to be conserved. It is possible that this change in a key catalytic residue may alter the enzyme mechanism so that it catalyzes the hydrolysis of a queuine nucleoside rather than the transglycosylation. QTRTD1 seems to be a very attractive candidate for the eukaryl queuine salvage enzyme.

While in most cases it appears that the production of queuosine tRNA completes the pathway, further modification of queuosine has also been observed (Figure 37).²⁹¹ In eukaryotes, two glycosylation events occur, mannosylation of tRNA^{Asp} and galactosylation of tRNA^{Tyr} on the diol of the queuosine side chain.^{308,309} Recently, it has been discovered that in *E. coli*, a paralogue of glutamyl-tRNA synthase (YadB) glutamylates the diol side chain of queuosine in tRNA^{Asp}.^{291,310} The enzymes responsible for the glycosylations and the physiological roles for these three modifications are unknown.

6.20.8.2 Wyosine and Derivatives

Until recently, less information was known regarding the enzymes involved in the biosynthesis of wyosine (or wyeosine), a complicated tri-cyclic modified nucleoside found in eukaryl and archaeal species. We now have a clearer understanding of how this base is constructed and many of the enzymes involved (Figure 38). It has

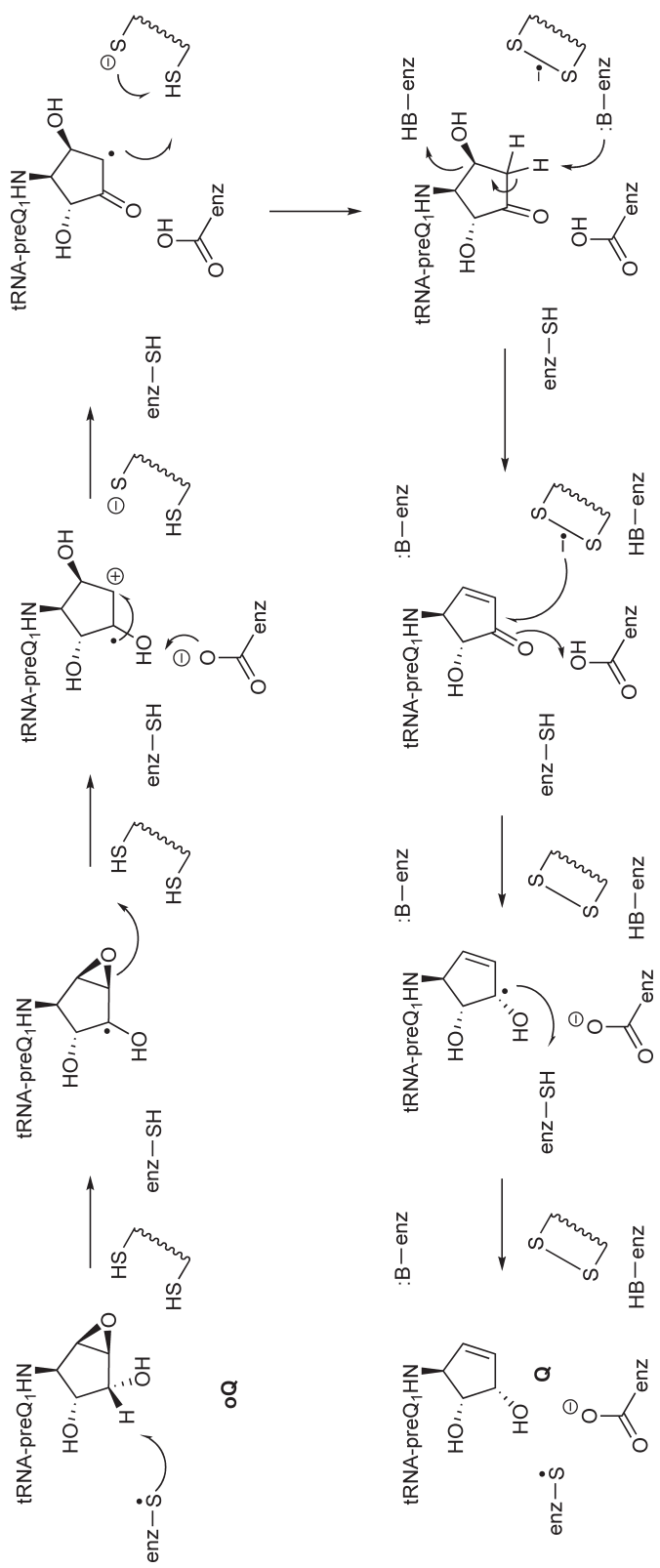


Figure 36 Proposed mechanism for oQ reduction to queuosine is based on an analogy drawn to ribonucleotide reductases and involves a thiyl radical and a redox active disulfide. This mechanism was proposed by Iwata-Reuyl. Reproduced from D. Iwata-Reuyl, *Bioorg. Chem.* **2003**, *31*, 24–43.

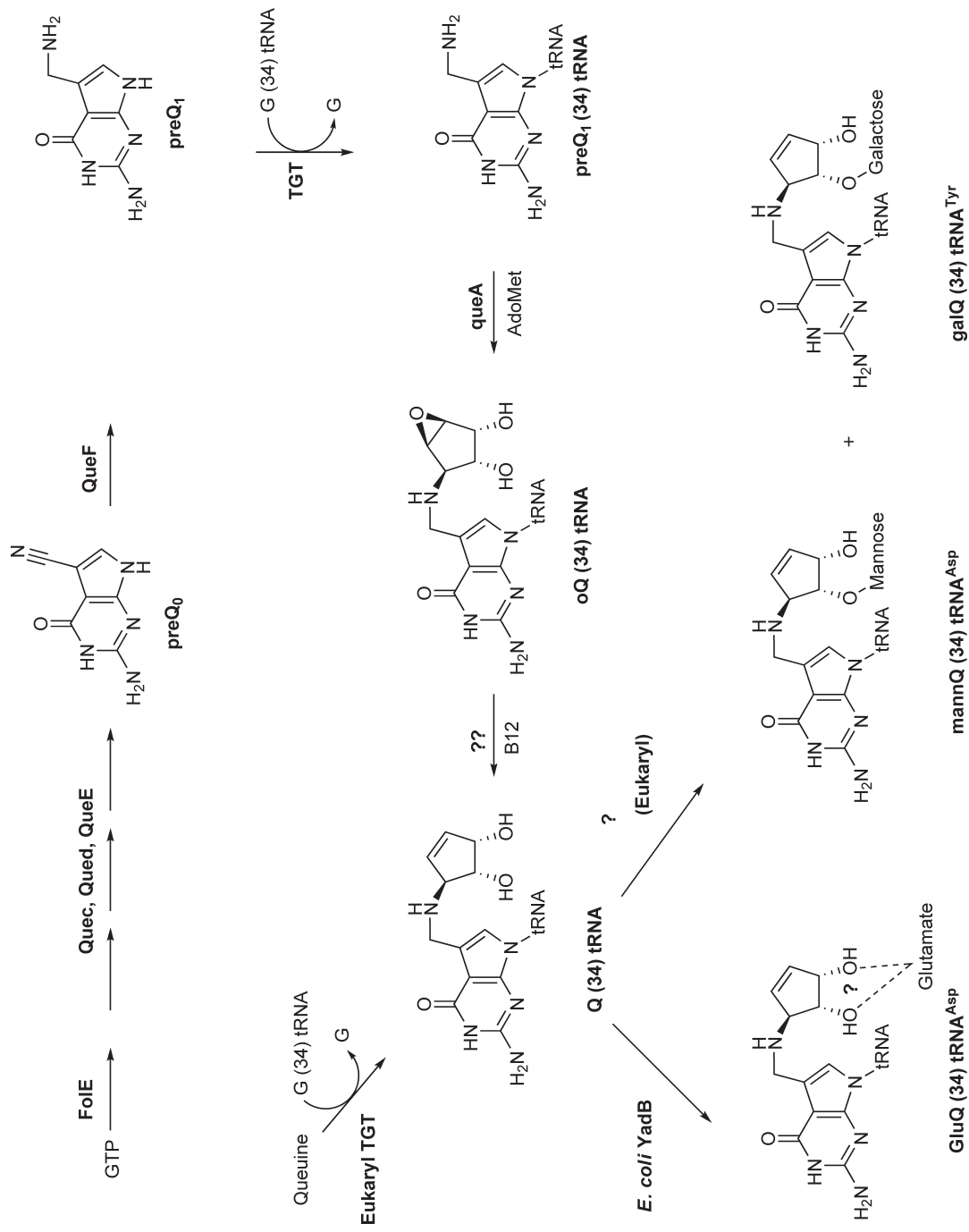


Figure 37 Current understanding of queuosine pathways in eubacteria and eukarya. In eubacteria, five enzymes are involved in the production of preQ₁. In both eubacteria and eukarya, TGT incorporates a preformed base into tRNA. In eubacteria, queuosine synthesis is completed by two additional enzymes. In both eubacteria and eukarya, queuosine can be further modified, depending on the tRNA involved, via the addition of either a carbohydrate or glutamate to the cyclopentyl diol moiety.

been understood for some time that the initial step in wyosine biosynthesis is an AdoMet-dependent methylation by Trm5 at N¹ of guanine 37 in substrate tRNA^{Phe}.^{311,312} Again, comparative genomics played a role in identification of a gene cluster, COG0731, using this MTase as a search criterion.³¹³ Knockout studies of one of the family members in *S. cerevisiae*, *YPL207w*, revealed a hypomodified tRNA^{Phe} with respect to wybutosine (yW) relative to wild type.³¹⁴ In the knockout, a buildup of a methylguanosine derivative (assumed to be m¹G) was observed by HPLC, and when complimented with a vector expressing *YPL207w*, the wyosine modification was restored. The authors hypothesized that the enzyme product of this gene must act in the biosynthesis immediately following the initial m¹G modification, but there are no reported *in vitro* experiments to confirm this.

Kalhor *et al.*³¹⁵ have provided evidence for a gene involved in the later steps of the wyosine biosynthetic pathway, *YML005w* (renamed *TRM12*). A knockout in *S. cerevisiae* led to a buildup of the intermediate ImG-14 detectable by HPLC and mass spectrometry. In addition, methylesterification activity was not observed, so the authors hypothesized that *TRM12* may be involved in this reaction as well. Using a reverse genetic approach, many of the key genes involved in wybutosine biosynthesis were elucidated.³¹⁶ Studying single gene deletions from over 350 *S. cerevisiae* strains and using mass spectrometry to analyze modified RNA intermediates present in each, the authors identified four genes that were required for wybutosine biosynthesis: *YPL207w*, *YML005w*, *YGL050w*, and *YOL141w* (also named TYW1, -2, -3, and -4, respectively). In addition, the authors found that other modified nucleosides were present in tRNA^{Phe} purified from each mutant, demonstrating that the presence of yW is not required for the formation of other modifications. It was determined that TYW1 catalyzes the cyclization of m¹G to form the tricyclic structure in yW biosynthesis, with the activity split between two domains: the N-terminus consists of a flavodoxin-1 domain, while the C-terminus contains a radical AdoMet domain. TYW2 is an AdoMet-dependent enzyme that belongs to the same family as TRM5, the initial MTase in the biosynthetic pathway. It was found that TYW4 is also an AdoMet-dependent MTase with methylesterification activity. The function of TYW3 was not elucidated through sequence homology, but was shown to be highly conserved.

The first protein crystal structure determined from the newly characterized wyosine biosynthetic pathway was TYW1, with two archaeal structures published consecutively. The homologue from *M. jannaschii* was determined at 2.4 Å³¹⁷ and the structure from *P. horikoshii* was determined at 2.2 Å.³¹⁸ Both groups concluded that TYW1 contains two [4Fe-4S] clusters, each of which is coordinated to three Cys residues in the active site. Only the Goto-Ito structure (Figure 39) had electron density that was representative of an [4Fe-4S] cluster, the Suzuki structure was compared with other similar proteins known to bind multiple [4Fe-4S] clusters for docking studies. The first [4Fe-4S] cluster was found within the radical AdoMet enzyme signature motif CX₃CX₂C.³¹⁷ The second cluster is positioned at the opposite side of the catalytic center from the AdoMet binding site, approximately 11 Å from the first [4Fe-4S] cluster, surrounded by predominantly hydrophilic residues. Although no *in vitro* characterization has been reported to date to determine the kinetics of substrate binding, gel-mobility shift assays were performed with both full-length tRNA^{Phe} and the corresponding anticodon arm minihelix, demonstrating that the anticodon stem loop is a minimal recognition element for TYW1.³¹⁸

6.20.9 Conclusions and Perspectives

Over the past 50 years, it has been known that nucleic acids contain unusual nucleosides that appear to be derivatives of the five canonical nucleosides. In all cases known to date, a relatively large number of enzymes generates these modifications post-transcriptionally. The majority of these modifications occur in RNA and currently there are slightly more than 100 such modified nucleosides. Earlier reviews and books have discussed the identification, physiological roles, and biosynthesis of many of these modifications (although much is still not understood).^{5,6,319}

During the past decade (since our last review⁹), much progress has been made. In the area of methylation, the mechanistic issues surrounding methyl transfer from AdoMet were clearly established earlier. The major recent developments involve the determination of a large number of X-ray crystal structures of MTases, which have provided more insight into substrate RNA recognition and provide a framework for deeper understanding of mechanistic issues. These developments in structural biology are a common theme for a large number of RNA modifying enzymes representing a diversity of nucleoside modifications. Again these structures have provided

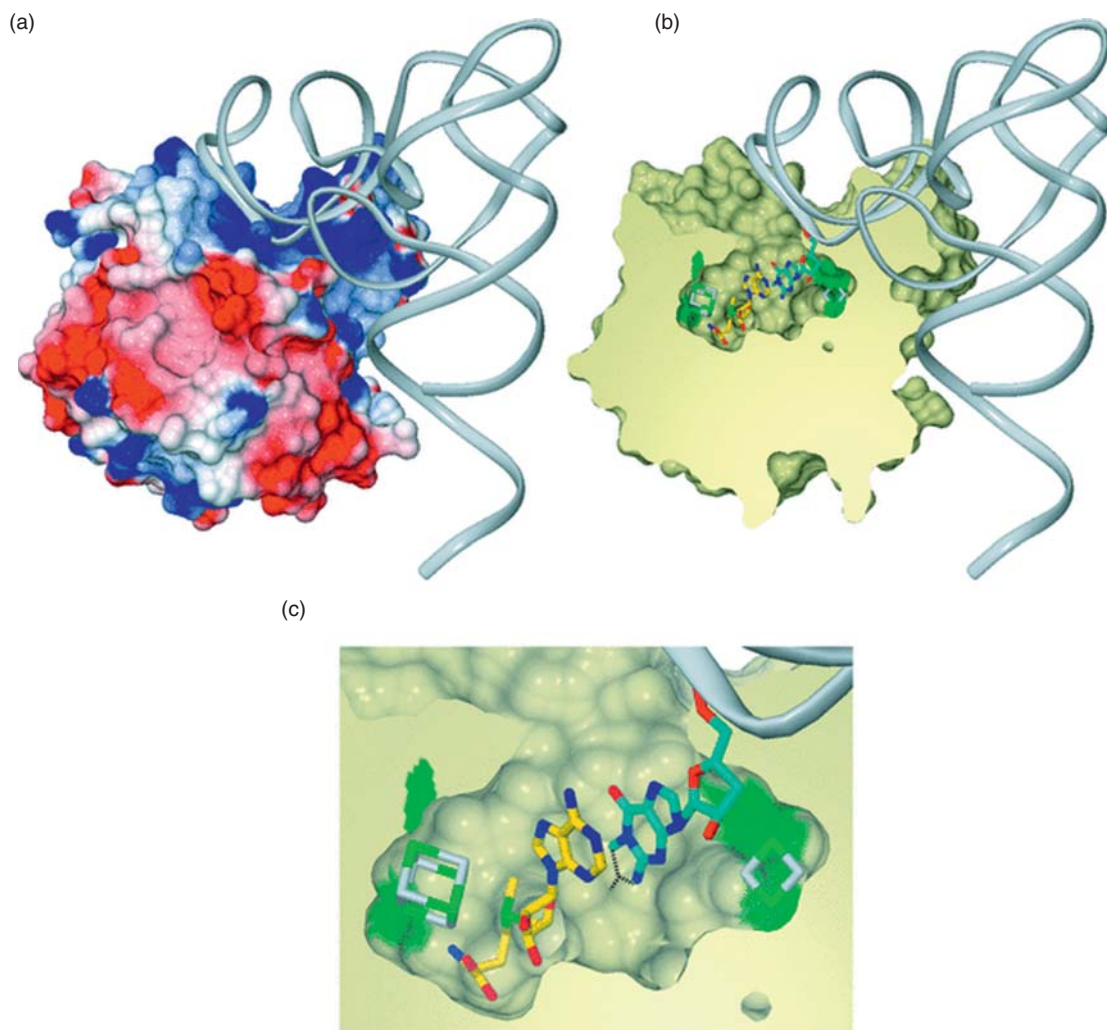


Figure 39 Docking model of *Pyrococcus horikoshii* TYW1, FeS clusters, AdoMet, and tRNA. Reproduced from S. Goto-Ito; R. Ishii; T. Ito; R. Shibata; E. Fusatomi; S. Sekine; Y. Bessho; S. Yokoyama, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 1059, with permission from IUCr (<http://journals.iucr.org/>).

more insight into these enzymes (e.g., confirmation of adenosine deaminase mechanism) and have, in some cases, allowed researchers to probe evolutionary relationships between enzyme families based on structural homologies (e.g., the pseudoU synthase families). The biosynthetic pathways for the more complex modifications (e.g., queuosine and wyosine) are more clearly understood, largely through the application of comparative genomics techniques (in concert with careful enzymological characterization) to identify biosynthetic genes. The chemical mechanisms for thiolation and some alkylation events (e.g., lysidine) are clearer now due to recent studies. One area that lacks enough data is the reduction- and oxidation-derived modifications. Recent work has been initiated in this direction (e.g., DUS) so that the process could be elucidated in the near future.

Nucleoside modifications are widely believed to fine tune RNA structure and function. While the vast majority of nucleoside modifications have been found in tRNA, an increasing number of modifications are also being found in other RNA species (e.g., rRNA). It seems very likely that, as technology improves, a large number of nucleoside modifications can be found in a more diverse set of RNA species. It has been estimated that 5% of the *E. coli* genome codes for nucleoside modification enzymes. It will become increasingly more important to characterize both the mechanism and substrate recognition for the enzymes responsible for generating these interesting natural products.

Abbreviations

| | |
|---------------------------------------|--|
| [4Fe-4S] | iron-sulfur cluster |
| ACF | APOBEC-1 complementation factor |
| ADARs | adenosine deaminases acting on RNAs |
| ADATs | adenosine deaminases acting on tRNAs |
| Ado[•] | 5'-deoxyadenosyl radical |
| AdoHCys | S-adenosylhomocysteine |
| AdoMet | S-adenosylmethionine |
| apoB | apolipoprotein B |
| APOBEC-1 | ApoB mRNA editing enzyme catalytic polypeptide 1 |
| ASF | APOBEC-1 stimulating factor |
| CDARs | RNA cytidine deaminases |
| Cm | 2'-O-methylcytosine |
| D | dihydrouridine |
| DALI | distance matrix alignment program |
| DHODH | dihydroorotate dehydrogenase |
| DHPDH | dihydropyrimidine dehydrogenase |
| DMAPP | dimethylallyl pyrophosphate |
| dsRBD | double-stranded RNA-binding domain |
| dsRNA | double-stranded RNA |
| DTT | dithiothreitol |
| DUS | dihydrouridine synthase |
| FMN | flavin mononucleotide |
| FMPP | 7-fluoromethyl-7-deazaguanine |
| Gm | 2'-O-methylguanosine |
| <i>I</i>-AEDANS | 5-({2-[(iodoacetyl)amino]ethyl}amino)naphthalene-1-sulfonic acid |
| i⁶A | N ⁶ -(dimethylallyl)adenosine |
| ImG-14 | 4-demethylwyosine |
| io⁶A | N ⁶ -(<i>cis</i> -hydroxyisopentenyl)adenosine |
| KIE | kinetic isotope effect |
| L | lysidine |
| m¹A | 1-methyladenosine |
| m¹G | 1-methylguanosine |
| m¹I | 1-methylinosine |
| m²₂G | 2,2-dimethylguanosine |
| m²G | 2-methylguanosine |
| m⁵C | 5-methylcytosine |
| m⁵U | 5-methyluracil (thymine) |
| m⁶₂A | 6,6-dimethyladenosine |
| m⁶A | 6-methyladenosine |
| m⁷G | 7-methylguanosine |
| miRNA | microRNA |
| mnm⁵s²U | 5-methyl-aminomethyl-2-thiouridine |
| mnm⁵Se²U | 5-methyl-aminomethyl-2-selenouridine |
| ms²i⁶A | 2-methylthio-N ⁶ -isopentenyl-adenosine |
| ms²io⁶A | N ⁶ -(4-hydroxyisopentenyl)-2-methylthioadenosine |
| MTase | methyltransferase |
| NAD | nicotinamide adenine dinucleotide |
| NADPH | nicotinamide adenine dinucleotide phosphate |

| | |
|-------------------------|--|
| NMR | nuclear magnetic resonance |
| oQ | epoxyqueuosine |
| PDB | protein data bank |
| PP_i | pyrophosphate |
| preQ₀ | 7-cyano-7-deazaguanine |
| preQ₁ | 7-aminomethyl-7-deazaguanine |
| pseudoU | pseudouridine |
| PUA | PseudoUridine synthase and Archaeosine transglycosylase |
| domain | |
| Py | pyrimidine |
| Q | queuosine, 7-(((4,5- <i>cis</i> -dihydroxy-2-cyclopenten-1-yl)amino)methyl)- 7-deazaguanosine |
| RRM | RNA-recognition motif |
| rRNA | ribosomal RNA |
| RUMT | tRNA (m ⁵ U54)-methyltransferase |
| s²C | 2-thiocytidine |
| s⁴U | 4-thiouridine |
| snoRNA | small nucleolar RNA |
| snRNA | small nuclear RNA |
| SPOUT | SpoU-TrmD family |
| t⁶A | N ⁶ -(N-threonylcarbonyl)adenosine |
| TGT | tRNA-guanine transglycosylase |
| THUMP | THioUridine synthases, RNA Methyltransferases, and Pseudouridine synthases |
| TRAM | RNA-binding motif originally characterized in TRM2 and miaB |
| domain | |
| tRNA | transfer RNA |
| T^ψC | commonly modified T-loop/-arm of tRNA |
| USP14 | ubiquitin-specific protease 14 |
| Y | wyosine, 4,6-dimethyl-3-β-D-ribofuranosyl-3,4-dihydro-9H-imidazo[1,2-a]purin-9-one |
| yW | wybutosine, 3H-imidazo[1,2-α]purine-7-butanolic acid, 4,9-dihydro- α-((methoxycarbonyl)amino)- 4,6-dimethyl-9-oxo- 3-beta-D-ribofuranosyl methyl ester |

References

1. F. F. Davis; F. W. Allen, *J. Biol. Chem.* **1957**, *227*, 907–915.
2. R. Benne; J. Van den Burg; J. Brakenhoff; P. Sloof; J. Van Boom; M. Tromp, *Cell* **1986**, *46*, 819–826.
3. R. Benne, *RNA Editing: The Alteration of Protein Coding Sequences of RNA*; Ellis Horwood: New York, 1993.
4. D. Iwata-Reuyl, *Curr. Opin. Chem. Biol.* **2008**, *12*, 33.
5. H. Grosjean, *Fine-Tuning of RNA Functions by Modification and Editing*; Springer-Verlag: Berlin, 2005.
6. H. Grosjean; R. Benne, *Modification and Editing of RNA: The Alteration of RNA Structure and Function*; ASM Press: Washington, DC, 1998.
7. G. R. Björk, Biosynthesis and Function of Modified Nucleosides. In *tRNA: Structure, Biosynthesis and Function*; D. Söll, U. L. RajBhandary, Eds.; ASM Press: Washington, DC, 1994.
8. J. T. Kealey; X. Gu; D. V. Santi, *Biochimie* **1994**, *76*, 1133.
9. G. A. Garcia; D. M. Goodenough-Lashua, Mechanisms of RNA-Modifying and -Editing Enzymes. In *Modification and Editing of RNA*; H. Grosjean, R. Benne, Eds.; ASM Press: Washington, DC, 1998.
10. L. J. Yao; T. L. James; J. T. Kealey; D. V. Santi; U. Schmitz, *J. Biomol. NMR* **1997**, *9*, 229.
11. R. Sengupta; S. Vainauskas; C. Yarian; E. Sochacka; A. Malkiewicz; R. H. Guenther; K. M. Koshlap; P. F. Agris, *Nucleic Acids Res.* **2000**, *28*, 1374.
12. T. T. Lee; S. Agarwalla; R. M. Stroud, *Structure* **2004**, *12*, 397.
13. Y. Q. Liu; D. V. Santi, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8263.
14. D. V. Santi; C. E. Garrett; P. J. Barr, *Cell* **1983**, *33*, 9.
15. X. R. Gu; C. Gustafsson; J. Ku; M. Yu; D. V. Santi, *Biochemistry* **1999**, *38*, 4053.
16. M. Y. King; K. L. Redman, *Biochemistry* **2002**, *41*, 11218.
17. P. G. Foster; C. R. Nunes; P. Greene; D. Moustakas; R. M. Stroud, *Structure* **2003**, *11*, 1609.

18. N. M. Andersen; S. Douthwaite, *J. Mol. Biol.* **2006**, *359*, 777.
19. B. M. Hallberg; U. B. Ericsson; K. A. Johnson; N. M. Andersen; S. Douthwaite; P. Nordlund; A. E. Beuscher; H. Erlandsen, *J. Mol. Biol.* **2006**, *360*, 774.
20. Y. Motorin; H. Grosjean, *RNA* **1999**, *5*, 1105.
21. M. G. Goll; F. Kirpekar; K. A. Maggart; J. A. Yoder; C. L. Hsieh; X. Y. Zhang; K. G. Golic; S. E. Jacobsen; T. H. Bestor, *Science* **2006**, *311*, 395.
22. G. R. Björk, *Prog. Nucleic Acid. Res. Mol. Biol.* **1995**, *50*, 263–338.
23. M. Helm; G. Attardi, *J. Mol. Biol.* **2004**, *337*, 545.
24. A. Gupta; P. H. Kumar; T. K. Dineshkumar; U. Varshney; H. S. Subramanya, *J. Mol. Biol.* **2001**, *312*, 381.
25. L. Droogmans; M. Roovers; J. M. Bujnicki; C. Tricot; T. Hartsch; V. Stalon; H. Grosjean, *Nucleic Acids Res.* **2003**, *31*, 2148.
26. M. Roovers; J. Wouters; J. M. Bujnicki; C. Tricot; V. Stalon; H. Grosjean; L. Droogmans, *Nucleic Acids Res.* **2004**, *32*, 465.
27. S. G. Ozanick; J. M. Bujnicki; D. S. Sem; J. T. Anderson, *Nucleic Acids Res.* **2007**, *35*, 6808.
28. H. Grosjean; F. Constantinesco; D. Foiret; N. Benachenhou, *Nucleic Acids Res.* **1995**, *23*, 4312.
29. P. Mallick; D. R. Boutz; D. Eisenberg; T. O. Yeates, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9679.
30. J. Anderson; L. Phan; A. G. Hinnebusch, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5173.
31. M. T. Garcia-Barrio; T. Naranda; C. R. V. Dealdana; R. Cuesta; A. G. Hinnebusch; J. W. B. Hershey; M. Tamame, *Genes Dev.* **1995**, *9*, 1781.
32. S. Ozanick; A. Krecic; J. Andersland; J. T. Anderson, *RNA* **2005**, *11*, 1281.
33. G. R. Björk, Modified Nucleotides in Positions 34 and 37 of tRNAs and Their Predicted Coding Capacities. In *Modification and Editing of RNA*; H. B. Grosjean, R. Benne, Eds.; ASM: Washington, DC, 1998.
34. M. Redlak; C. AndraosSelim; R. Giege; C. Florentz; W. M. Holmes, *Biochemistry* **1997**, *36*, 8699.
35. H. Takeda; T. Toyooka; Y. Ikeuchi; S. Yokobori; K. Okadome; F. Takano; T. Oshima; T. Suzuki; Y. Endo; H. Hori, *Genes Cells* **2006**, *11*, 1353.
36. P. A. Elkins; J. M. Watts; M. Zalacain; A. van Thiel; P. R. Vitazka; M. Redlak; C. Andraos-Selim; F. Rastinejad; W. M. Holmes, *J. Mol. Biol.* **2003**, *333*, 931.
37. V. Anantharaman; E. V. Koonin; L. Aravind, *J. Mol. Microbiol. Biotechnol.* **2002**, *4*, 71.
38. J. Y. Liu; W. R. Wang; D. H. Shin; H. Yokota; R. Kim; S. H. Kim, *Proteins* **2003**, *53*, 326.
39. H. J. Ahn; H. W. Kim; H. J. Yoon; B. I. Lee; S. W. Suh; J. K. Yang, *EMBO J.* **2003**, *22*, 2593.
40. T. Christian; C. Evilia; S. Williams; Y. M. Hou, *J. Mol. Biol.* **2004**, *339*, 707.
41. C. Lee; G. Kramer; D. E. Graham; D. R. Appling, *J. Biol. Chem.* **2007**, *282*, 27744.
42. S. Dunin-Horkawicz; A. Czerwoniec; M. J. Gajda; M. Feder; H. Grosjean; J. M. Bujnicki, *Nucleic Acids Res.* **2006**, *34*, D145.
43. T. Christian; C. Evilia; Y. M. Hou, *Biochemistry* **2006**, *45*, 7463.
44. L. G. S. De Bie; M. Roovers; Y. Oudjama; R. Wattiez; C. Tricot; V. Stalon; L. Droogmans; J. M. Bujnicki, *J. Bacteriol.* **2003**, *185*, 3238.
45. H. Okamoto; K. Watanabe; Y. Ikeuchi; T. Suzuki; Y. Endo; H. Hori, *J. Biol. Chem.* **2004**, *279*, 49151.
46. I. Zegers; D. Gigot; F. van Vliet; C. Tricot; S. Aymerich; J. M. Bujnicki; J. Kosinski; L. Droogmans, *Nucleic Acids Res.* **2006**, *34*, 1925.
47. A. Alexandrov; M. R. Martzen; E. M. Phizicky, *RNA* **2002**, *8*, 1253.
48. E. Purta; F. van Vliet; C. Tricot; L. G. De Bie; M. Feder; K. Skowronek; L. Droogmans; J. M. Bujnicki, *Proteins* **2005**, *59*, 482.
49. K. Matsumoto; T. Toyooka; C. Tomikawa; A. Ochi; Y. Takano; N. Takayanagi; Y. Endo; H. Hori, *FEBS Lett.* **2007**, *581*, 1599.
50. M. G. Gu; C. D. Lima, *Curr. Opin. Struct. Biol.* **2005**, *15*, 99.
51. S. M. Carroll; P. Narayan; F. M. Rottman, *Mol. Cell. Biol.* **1990**, *10*, 4456.
52. M. J. Clancy; M. E. Shambaugh; C. S. Timpte; J. A. Bokar, *Nucleic Acids Res.* **2002**, *30*, 4509.
53. P. V. Sergiev; M. V. Serebryakova; A. A. Bogdanov; O. A. Dontsova, *J. Mol. Biol.* **2008**, *375*, 291.
54. G. Schluckebier; P. Zhong; K. D. Stewart; T. J. Kavanaugh; C. Abad-Zapatero, *J. Mol. Biol.* **1999**, *289*, 277.
55. C. Denoya; D. Dubnau, *J. Biol. Chem.* **1989**, *264*, 2615.
56. H. C. O'Farrell; J. N. Scarsdale; J. P. Rife, *J. Mol. Biol.* **2004**, *339*, 337.
57. P. M. Desai; J. P. Rife, *Arch. Biochem. Biophys.* **2006**, *449*, 57.
58. A. Zamir; R. Miskin; D. Elson, *FEBS Lett.* **1969**, *3*, 85.
59. A. Zamir; R. Miskin; D. Elson, *J. Mol. Biol.* **1971**, *60*, 347.
60. B. T. Wimberly; D. E. Brodersen; W. M. Clemons; R. J. Morgan-Warren; A. P. Carter; C. Vornrhein; T. Hartsch; V. Ramakrishnan, *Nature* **2000**, *407*, 327.
61. J. Armengaud; J. Urbonavicius; B. Fernandez; G. Chaussinand; J. M. Bujnicki; H. Grosjean, *J. Biol. Chem.* **2004**, *279*, 37142–37152.
62. G. Gabant; S. Auxilien; I. Tuszyńska; M. Locard; M. J. Gajda; G. Chaussinand; B. Fernandez; A. Dedieu; H. Grosjean; B. Golinelli-Pimpaneau; J. M. Bujnicki; J. Armengaud, *Nucleic Acids Res.* **2006**, *34*, 2483–2494.
63. F. Constantinesco; Y. Motorin; H. Grosjean, *J. Mol. Biol.* **1999**, *291*, 375.
64. J. Edqvist; H. Grosjean; K. B. Straby, *Nucleic Acids Res.* **1992**, *20*, 6575.
65. J. Edqvist; K. B. Straby; H. Grosjean, *Biochimie* **1995**, *77*, 54.
66. J. Urbonavicius; J. Armengaud; H. Grosjean, *J. Mol. Biol.* **2006**, *357*, 387.
67. P. V. Sergiev; A. A. Bogdanov; O. A. Dontsova, *Nucleic Acids Res.* **2007**, *35*, 2295.
68. D. V. Lesnyak; J. Osipiuk; T. Skarina; P. V. Sergiev; A. A. Bogdanov; A. Edwards; A. Savchenko; A. Joachimiak; O. A. Dontsova, *J. Biol. Chem.* **2007**, *282*, 5880.
69. L. Huang; L. Hung; M. Odell; H. Yokota; R. Kim; S.-H. Kim, *J. Struct. Funct. Genomics* **2002**, *2*, 121–127.
70. J. M. Bujnicki; L. Rychlewski, *Bmc Bioinformatics* **2002**, *3*.
71. C. Weitzmann; S. J. Tumminia; M. Boublik; J. Ofengand, *Nucleic Acids Res.* **1991**, *19*, 7089.
72. J. S. Tscherne; K. Nurse; P. Popienick; J. Ofengand, *J. Biol. Chem.* **1999**, *274*, 924.
73. D. V. Lesnyak; P. V. Sergiev; A. A. Bogdanov; O. A. Dontsova, *J. Mol. Biol.* **2006**, *364*, 20.
74. P. V. Sergiev; D. V. Lesnyak; A. A. Bogdanov; O. A. Dontsova, *J. Mol. Biol.* **2006**, *364*, 26.

75. H. Grosjean; M. Sprinzl; S. Steinberg, *Biochimie* **1995**, *77*, 139–141.
76. M. L. Wilkinson; S. M. Crary; J. E. Jackman; E. J. Grayhack; E. M. Phizicky, *RNA* **2007**, *13*, 404.
77. H. Hori; T. Suzuki; K. Sugawara; Y. Inoue; T. Shibata; S. Kuramitsu; S. Yokoyama; T. Oshima; K. Watanabe, *Genes Cells* **2002**, *7*, 259.
78. O. Nureki; K. Watanabe; S. Fukai; R. Ishii; Y. Endo; H. Hori; S. Yokoyama, *Structure* **2004**, *12*, 593.
79. J. Hager; B. L. Staker; H. Bugl; U. Jakob, *J. Biol. Chem.* **2002**, *277*, 41978.
80. G. Michel; V. Sauve; R. Larocque; Y. G. Li; A. Matte; M. Cygler, *Structure* **2002**, *10*, 1303.
81. S. K. Johansen; C. E. Maus; B. B. Plikaytis; S. Douthwaite, *Mol. Cell* **2006**, *23*, 173.
82. J. P. Bachelier; J. Cavaille; A. Huttenhofer, *Biochimie* **2002**, *84*, 775.
83. B. Clouet-d'Orval; C. Gaspin; A. Mougín, *Biochimie* **2005**, *87*, 889.
84. S. L. Reichow; T. Hamma; A. R. Ferre-D'Amare; G. Varani, *Nucleic Acids Res.* **2007**, *35*, 1452.
85. M. H. Renalier; N. Joseph; C. Gaspin; P. Thebault; A. Mougín, *RNA* **2005**, *11*, 1051.
86. M. Kuratani; Y. Bessho; M. Nishimoto; H. Grosjean; S. Yokoyama, *J. Mol. Biol.* **2008**, *375*, 1064.
87. B. Lapeyre; S. K. Purushothaman, *Mol. Cell* **2004**, *16*, 663.
88. A. Satoh; K. Takai; R. Ouchi; S. Yokoyama; H. Takaku, *RNA* **2000**, *6*, 680.
89. R. Leipuviene; Q. Qian; G. R. Bjork, *J. Bacteriol.* **2004**, *186*, 758–766.
90. E. G. Mueller; C. J. Buck; P. M. Palenchar; L. E. Barnhart; J. L. Paulson, *Nucleic Acids Res.* **1998**, *26*, 2606–2610.
91. R. Kambampati; C. T. Lauhon, *Biochemistry* **1999**, *38*, 16561–16568.
92. R. Kambampati; C. T. Lauhon, *J. Biol. Chem.* **2000**, *275*, 10727–10730.
93. J. Frazzon; D. R. Dean, *Curr. Opin. Chem. Biol.* **2003**, *7*, 166–173.
94. H. K. Lundgren; G. R. Bjork, *J. Bacteriol.* **2006**, *188*, 3052–3062.
95. E. G. Mueller; P. M. Palenchar; C. J. Buck, *J. Biol. Chem.* **2001**, *276*, 33588–33595.
96. C. M. Wright; P. M. Palenchar; E. G. Mueller, *Chem. Commun.* **2002**, 2708–2709.
97. P. M. Palenchar; C. J. Buck; H. Cheng; T. J. Larson; E. G. Mueller, *J. Biol. Chem.* **2000**, *275*, 8283–8286.
98. C. M. Wright; G. D. Christman; A. M. Snellinger; M. V. Johnston; E. G. Mueller, *Chem. Commun.* **2006**, 3104–3106.
99. P. Bork; E. V. Koonin, *Proteins* **1994**, *20*, 347–355.
100. E. G. Mueller; P. M. Palenchar, *Protein Sci.* **1999**, *8*, 2424–2427.
101. D. G. Waterman; M. Ortiz-Lombardia; M. J. Fogg; E. V. Koonin; A. A. Antson, *J. Mol. Biol.* **2006**, *356*, 97–110.
102. R. Kambampati; C. T. Lauhon, *Biochemistry* **2003**, *42*, 1109–1117.
103. E. G. Mueller, *Nat. Chem. Biol.* **2006**, *2*, 185–194.
104. Y. Ikeuchi; N. Shigi; J. Kato; A. Nishimura; T. Suzuki, *Mol. Cell* **2006**, *21*, 97–108.
105. C. T. Lauhon, *Nat. Chem. Biol.* **2006**, *2*, 182–183.
106. C. T. Lauhon, *J. Bacteriol.* **2002**, *184*, 6820–6829.
107. K. Nilsson; H. K. Lundgren; T. G. Hagervall; G. R. Bjork, *J. Bacteriol.* **2002**, *184*, 6830–6835.
108. A. D. Smith; J. N. Agar; K. A. Johnson; J. Frazzon; I. J. Amster; D. R. Dean; M. K. Johnson, *J. Am. Chem. Soc.* **2001**, *123*, 11103–11104.
109. B. Esberg; H. C. E. Leung; H. C. T. Tsui; G. R. Björk; M. E. Winkler, *J. Bacteriol.* **1999**, *181*, 7256–7265.
110. D. C. Johnson; D. R. Dean; A. D. Smith; M. K. Johnson, *Annu. Rev. Biochem.* **2005**, *74*, 247–281.
111. F. Pierrel; G. R. Bjork; M. Fontecave; M. Atta, *J. Biol. Chem.* **2002**, *277*, 13367–13370.
112. K. H. Kaminska; U. Baraniak; M. Boniecki; K. Nowaczyk; A. Czerwoniec; J. M. Bujnicki, *Proteins* **2008**, *70*, 1–18.
113. H. L. Hernandez; F. Pierrel; E. Elleingand; R. Garcia-Serres; B. H. Huynh; M. K. Johnson; M. Fontecave; M. Atta, *Biochemistry* **2007**, *46*, 5140–5147.
114. H. J. Sofia; G. Chen; B. G. Hetzler; J. F. Reyes-Spindola; N. E. Miller, *Nucleic Acids Res.* **2001**, *29*, 1097–1106.
115. F. Pierrel; T. Douki; M. Fontecave; M. Atta, *J. Biol. Chem.* **2004**, *279*, 47555–47563.
116. G. Jäger; R. Leipuviene; M. G. Pollard; Q. Qian; G. R. Bjork, *J. Bacteriol.* **2004**, *186*, 750–757.
117. W. M. Ching; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 374–377.
118. A. J. Wittwer, *J. Biol. Chem.* **1983**, *258*, 8637–8641.
119. A. J. Wittwer; T. C. Stadtman, *Arch. Biochem. Biophys.* **1986**, *248*, 540–550.
120. Z. Veres; L. Tsai; T. D. Scholz; M. Politino; R. S. Balaban; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2975–2979.
121. Z. Veres; T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8092–8096.
122. H. Mihara; S. Kato; G. M. Lacourciere; T. C. Stadtman; R. Kennedy; T. Kurihara; U. Tokumoto; Y. Takahashi; N. Esaki, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6679–6683.
123. M. D. Wolfe; F. Ahmed; G. M. Lacourciere; C. T. Lauhon; T. C. Stadtman; T. J. Larson, *J. Biol. Chem.* **2004**, *279*, 1801–1809.
124. F. H. C. Crick, *J. Mol. Biol.* **1966**, *19*, 548.
125. M. Schaub; W. Keller, *Biochimie* **2002**, *84*, 791.
126. M. Higuchi; F. N. Single; M. Kohler; B. Sommer; R. Sprengel; P. H. Seeburg, *Cell* **1993**, *75*, 1361–1370.
127. H. Lomeli; J. Mosbacher; T. Melcher; T. Hoger; J. R. P. Geiger; T. Kuner; H. Monyer; M. Higuchi; A. Bach; P. H. Seeburg, *Science* **1994**, *266*, 1709.
128. A. P. Gerber; W. Keller, *Trends Biochem. Sci.* **2001**, *26*, 376.
129. B. L. Bass, *Trends Biochem. Sci.* **1997**, *22*, 157.
130. A. Gerber; H. Grosjean; T. Melcher; W. Keller, *EMBO J.* **1998**, *17*, 4780.
131. S. Maas; A. P. Gerber; A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8895.
132. S. Maas; T. Melcher; P. H. Seeburg, *Curr. Opin. Cell Biol.* **1997**, *9*, 343.
133. S. H. Liaw; Y. J. Chang; C. T. Lai; H. C. Chang; G. G. Chang, *J. Biol. Chem.* **2004**, *279*, 35479.
134. J. Kim; V. Malashkevich; S. Roday; M. Lisbin; V. L. Schramm; S. C. Almo, *Biochemistry* **2006**, *45*, 6407.
135. S. Maas; A. Rich, *Bioessays* **2000**, *22*, 790.
136. M. F. Jantsch; M. Ohman, RNA Editing by Adenosine Deaminases that Act on RNA (ADARs). In *Nucleic Acids and Molecular Biology*; H. U. Göringer, Ed.; Springer: New York/Dordrecht, 2008; pp 51–84.

137. W. D. Yang; T. P. Chendrimada; Q. D. Wang; M. Higuchi; P. H. Seeburg; R. Shiekhattar; K. Nishikura, *Nat. Struct. Mol. Biol.* **2006**, *13*, 13.
138. Y. Kawahara; B. Zinshteyn; T. P. Chendrimada; R. Shiekhattar; K. Nishikura, *EMBO Rep.* **2007**, *8*, 763.
139. R. F. Hough; B. L. Bass, *RNA* **1997**, *3*, 356.
140. S. Klimasauskas; S. Kumar; R. J. Roberts; X. D. Cheng, *Cell* **1994**, *76*, 357.
141. R. J. Roberts; X. D. Cheng, *Annu. Rev. Biochem.* **1998**, *67*, 181.
142. O. M. Stephens; H. Y. Yi-Brunozzi; P. A. Beal, *Biochemistry* **2000**, *39*, 12243.
143. H. Y. Yi-Brunozzi; O. M. Stephens; P. A. Beal, *J. Biol. Chem.* **2001**, *276*, 37827.
144. B. W. Allan; N. O. Reich, *Biochemistry* **1996**, *35*, 14757.
145. B. Holz; S. Klimasauskas; S. Serva; E. Weinhold, *Nucleic Acids Res.* **1998**, *26*, 1076.
146. S. Auxilien; P. F. Crain; R. W. Trewyn; H. Grosjean, *J. Mol. Biol.* **1996**, *262*, 437.
147. A. P. Gerber; W. Keller, *Science* **1999**, *286*, 1146.
148. J. Wolf; A. P. Gerber; W. Keller, *EMBO J.* **2002**, *21*, 3841.
149. H. C. Losey; A. J. Ruthenburg; G. L. Verdine, *Nat. Struct. Mol. Biol.* **2006**, *13*, 153.
150. M. Kuratani; R. Ishii; Y. Bessho; R. Fukunaga; T. Sengoku; M. Shirouzu; S. Sekine; S. Yokoyama, *J. Biol. Chem.* **2005**, *280*, 16002–16008.
151. Y. Elias; R. H. Huang, *Biochemistry* **2005**, *44*, 12057.
152. W. H. Lee; Y. K. Kim; K. H. Nam; A. Priyadarshi; E. H. Lee; E. E. Kim; Y. H. Jeon; C. Cheong; K. Y. Hwang, *Proteins* **2007**, *68*, 1016.
153. M. Luo; V. L. Schramm, *J. Am. Chem. Soc.* **2008**, *130*, 2649.
154. L. M. Powell; S. C. Wallis; R. J. Pease; Y. H. Edwards; T. J. Knott; J. Scott, *Cell* **1987**, *50*, 831.
155. S. H. Chen; G. Habib; C. Y. Yang; Z. W. Gu; B. R. Lee; S. A. Weng; S. R. Silberman; S. J. Cai; J. P. Deslypere; M. Rosseneu; A. M. Gotto; W. H. Li; L. Chan, *Science* **1987**, *238*, 363.
156. M. S. Davies; S. C. Wallis; D. M. Driscoll; J. K. Wynne; G. W. Williams; L. M. Powell; J. Scott, *J. Biol. Chem.* **1989**, *264*, 13395.
157. R. R. Shah; T. J. Knott; J. E. Legros; N. Navaratnam; J. C. Greeve; J. Scott, *J. Biol. Chem.* **1991**, *266*, 16301.
158. S. Anant; N. O. Davidson, *Mol. Cell. Biol.* **2000**, *20*, 1982.
159. M. Hersberger; T. L. Innerarity, *J. Biol. Chem.* **1998**, *273*, 9435.
160. N. Richardson; N. Navaratnam; J. Scott, *J. Biol. Chem.* **1998**, *273*, 31707.
161. M. Hersberger; S. Patarroyo-White; K. S. Arnold; T. L. Innerarity, *J. Biol. Chem.* **1999**, *274*, 34590.
162. B. B. Teng; C. F. Burant; N. O. Davidson, *Science* **1993**, *260*, 1816.
163. N. O. Davidson; T. L. Innerarity; J. Scott; H. Smith; D. M. Driscoll; B. Teng; L. Chan, *RNA* **1995**, *1*, 3.
164. N. Navaratnam; J. R. Morrison; S. Bhattacharya; D. Patel; T. Funahashi; F. Giannoni; B. B. Teng; N. O. Davidson; J. Scott, *J. Biol. Chem.* **1993**, *268*, 20709.
165. A. Mehta; D. M. Driscoll, *Mol. Cell. Biol.* **1998**, *18*, 4426.
166. A. Mehta; M. T. Kinter; N. E. Sherman; D. M. Driscoll, *Mol. Cell. Biol.* **2000**, *20*, 1846.
167. H. Lellek; R. Kirsten; I. Diehl; F. Apostel; F. Buck; J. Greeve, *J. Biol. Chem.* **2000**, *275*, 19848.
168. V. Blanc; N. O. Davidson, *J. Biol. Chem.* **2003**, *278*, 1395.
169. L. Betts; S. B. Xiang; S. A. Short; R. Wolfenden; C. W. Carter, *J. Mol. Biol.* **1994**, *235*, 635.
170. K. F. Xie; M. P. Sowden; G. S. C. Dance; A. T. Torelli; H. C. Smith; J. E. Wedekind, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8114.
171. N. Navaratnam; T. Fujino; J. Bayliss; A. Jarmuz; A. How; N. Richardson; A. Somasekaram; S. Bhattacharya; C. Carter; J. Scott, *J. Mol. Biol.* **1998**, *275*, 695.
172. C. Maris; J. Masse; A. Chester; N. Navaratnam; F. H. T. Allain, *RNA* **2005**, *11*, 173.
173. A. Chester; V. Weinreb; C. W. Carter; N. Navaratnam, *RNA* **2004**, *10*, 1399.
174. R. H. Hall; M. J. Robins; L. Stasiuk; R. Thedford, *J. Am. Chem. Soc.* **1966**, *88*, 2614.
175. K. Biemann; S. Tsunakawa; J. Sonnenbi; H. Feldmann; D. Dutting; H. G. Zachau, *Angew. Chem. Int. Ed.* **1966**, *5*, 590.
176. P. F. Agris; D. J. Armstrong; K. P. Schafer; D. Soll, *Nucleic Acids Res.* **1975**, *2*, 691.
177. T. H. Tsang; M. Buck; B. N. Ames, *Biochim. Biophys. Acta* **1983**, *741*, 180.
178. J. Caillet; L. Droogmans, *J. Bacteriol.* **1988**, *170*, 4147.
179. D. M. Connolly; M. E. Winkler, *J. Bacteriol.* **1989**, *171*, 3233.
180. M. E. Dihanich; D. Najarian; R. Clark; E. C. Gillman; N. C. Martin; A. K. Hopper, *Mol. Cell. Biol.* **1987**, *7*, 177.
181. M. Sprinzl; K. S. Vassilenko, *Nucleic Acids Res.* **2005**, *33*, D139.
182. H. C. E. Leung; Y. Q. Chen; M. E. Winkler, *J. Biol. Chem.* **1997**, *272*, 13073.
183. J. A. Moore; C. D. Poulter, *Biochemistry* **1997**, *36*, 604.
184. A. W. Curnow; G. A. Garcia, *J. Biol. Chem.* **1995**, *270*, 17264–17267.
185. X. R. Gu; K. M. Ivanetich; D. V. Santi, *Biochemistry* **1996**, *35*, 11652.
186. M. Saraste; P. R. Sibbald; A. Wittinghofer, *Trends Biochem. Sci.* **1990**, *15*, 430.
187. D. M. Connolly; M. E. Winkler, *J. Bacteriol.* **1991**, *173*, 1711.
188. Y. Motorin; G. Bec; R. Tewari; H. Grosjean, *RNA* **1997**, *3*, 721.
189. T. J. Soderberg; C. D. Poulter, *Abstr. Pap. Am. Chem. Soc.* **2000**, *219*, U125.
190. A. W. Curnow; F. L. Kung; K. A. Koch; G. A. Garcia, *Biochemistry* **1993**, *32*, 5239–5246.
191. M. N. Ashby; P. A. Edwards, *J. Biol. Chem.* **1990**, *265*, 13157.
192. T. Soderberg; C. D. Poulter, *Biochemistry* **2001**, *40*, 1734.
193. W. Xie; C. Zhou; R. H. Huang, *J. Mol. Biol.* **2007**, *367*, 872.
194. F. Harada; S. Nishimur, *Biochemistry* **1974**, *13*, 300.
195. T. Muramatsu; S. Yokoyama; N. Horie; A. Matsuda; T. Ueda; Z. Yamaizumi; Y. Kuchino; S. Nishimura; T. Miyazawa, *J. Biol. Chem.* **1988**, *263*, 9261.
196. T. Muramatsu; K. Nishikawa; F. Nemoto; Y. Kuchino; S. Nishimura; T. Miyazawa; S. Yokoyama, *Nature* **1988**, *336*, 179.
197. A. Soma; Y. Ikeuchi; S. Kanemasa; K. Kobayashi; N. Ogasawara; T. Ote; J. Kato; K. Watanabe; Y. Sekine; T. Suzuki, *Mol. Cell* **2003**, *12*, 689.

198. B. Senger; S. Auxilien; U. Englisch; F. Cramer; F. Fasiolo, *Biochemistry* **1997**, *36*, 8269.
199. C. Marck; H. Grosjean, *RNA* **2002**, *8*, 1189.
200. H. Grosjean; G. R. Bjork, *Trends Biochem. Sci.* **2004**, *29*, 165.
201. K. Nakanishi; S. Fukai; Y. Ikeuchi; A. Soma; Y. Sekine; T. Suzuki; O. Nureki, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7487.
202. J. G. Tesmer; T. J. Klem; M. L. Deras; V. J. Davisson; J. L. Smith, *Nat. Struct. Biol.* **1996**, *3*, 74.
203. Y. Ikeuchi; A. Soma; T. Ote; J. Kato; Y. Sekine; T. Suzuki, *Mol. Cell* **2005**, *19*, 235–246.
204. M. Kuratani; Y. Yoshikawa; Y. Bessho; K. Higashijima; T. Ishii; R. Shibata; S. Takahashi; K. Yutani; S. Yokoyama, *Structure* **2007**, *15*, 1642.
205. M. T. Miller; B. O. Bachmann; C. A. Townsend; A. C. Rosenzweig, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14752.
206. L. M. Lois; C. D. Lima, *EMBO J.* **2005**, *24*, 439.
207. A. C. Bishop; J. M. Xu; R. C. Johnson; P. Schimmel; V. de Crecy-Lagard, *J. Biol. Chem.* **2002**, *277*, 25090.
208. R. L. Fagan; K. F. Jensen; O. Bjornberg; B. A. Paley, *Biochemistry* **2007**, *46*, 4028–4036.
209. M. Mittelstadt; A. Frump; T. Khuu; V. Fowlkes; I. Handy; C. V. Patel; R. C. Patel, *Nucleic Acids Res.* **2008**, *36*, 998.
210. D. F. Savage; V. de Crecy-Lagard; A. C. Bishop, *FEBS Lett.* **2006**, *580*, 5198.
211. F. Xing; S. L. Hiley; T. R. Hughes; E. M. Phizicky, *J. Biol. Chem.* **2004**, *279*, 17850.
212. F. Xing; M. R. Martzen; E. M. Phizicky, *RNA* **2002**, *8*, 370.
213. L. W. Rider et al., *J. Biol. Chem.* in press.
214. K. Mizutani; Y. Machida; S. Unzai; S. Y. Park; J. R. H. Tame, *Biochemistry* **2004**, *43*, 4454–4463.
215. G. A. Garcia; J. D. Kittendorf, *Bioorg. Chem.* **2005**, *33*, 229–251.
216. N. Okada; S. Nishimura, *J. Biol. Chem.* **1979**, *254*, 3061–3066.
217. R. Cortese; H. O. Kammen; S. J. Spengler; B. N. Ames, *J. Biol. Chem.* **1974**, *249*, 1103–1108.
218. C. J. Green; H. O. Kammen; E. E. Penhoet, *J. Biol. Chem.* **1982**, *257*, 3045–3052.
219. J. R. Patton, *Biochemistry* **1994**, *33*, 10423–10427.
220. J. R. Katze; B. Basile; J. A. McClosky, *Science* **1982**, *216*, 55–56.
221. N. Shindo-Okada; N. Okada; T. Ohgi; T. Goto; S. Nishimura, *Biochemistry* **1980**, *19*, 395.
222. M. Watanabe; M. Matsuo; S. Tanaka; H. Akimoto; S. Asahi; S. Nishimura; J. R. Katze; T. Hashizume; P. F. Crain; J. A. McCloskey; N. Okada, *J. Biol. Chem.* **1997**, *272*, 20146–20151.
223. S. Chong; A. W. Curnow; T. J. Huston; G. A. Garcia, *Biochemistry* **1995**, *34*, 3694–3701.
224. C. Romier; K. Reuter; D. Suck; R. Ficner, *EMBO J.* **1996**, *15*, 2850–2857.
225. W. R. Farkas; K. B. Jacobson, *Insect Biochem.* **1980**, *10*, 183–188.
226. N. K. Howes; W. R. Farkas, *J. Biol. Chem.* **1978**, *253*, 9082–9087.
227. R. K. Slany; S. O. Mueller, *Eur. J. Biochem.* **1995**, *230*, 221–228.
228. T. L. Walden, Jr.; N. Howes; W. R. Farkas, *J. Biol. Chem.* **1982**, *257*, 13218–13222.
229. W. Langgut; T. Reisser, *Nucleic Acids Res.* **1995**, *23*, 2488–2491.
230. K. L. Deshpande; P. H. Seubert; D. M. Tillman; W. R. Farkas; J. R. Katze, *Arch. Biochem. Biophys.* **1996**, *326*, 1–7.
231. W. R. Farkas; K. B. Jacobson; J. R. Katze, *Biochim. Biophys. Acta* **1984**, *781*, 64–75.
232. D. M. Goodenough-Lashua; G. A. Garcia, *Bioorg. Chem.* **2003**, *31*, 331–344.
233. C. Romier; K. Reuter; D. Suck; R. Ficner, *Biochemistry* **1996**, *35*, 15734–15739.
234. J. D. Kittendorf; L. M. Barcomb; S. T. Nonekowsky; G. A. Garcia, *Biochemistry* **2001**, *40*, 14123–14133.
235. J. D. Kittendorf; T. Sgraja; K. Reuter; G. Klebe; G. A. Garcia, *J. Biol. Chem.* **2003**, *278*, 42369–42376.
236. W. Xie; X. J. Liu; R. H. Huang, *Nat. Struct. Biol.* **2003**, *10*, 781–788.
237. D. M. Goodenough-Lashua, Ph.D. Dissertation in Medicinal Chemistry, University of Michigan, Ann Arbor, 2002.
238. G. C. Hoops; L. B. Townsend; G. A. Garcia, *Biochemistry* **1995**, *34*, 15539–15544.
239. F.-L. Kung; G. A. Garcia, *FEBS Lett.* **1998**, *431*, 427–432.
240. S. M. Chervin; J. D. Kittendorf; G. A. Garcia, *Methods Enzymol.* **2007**, *425*, 121–137.
241. K. A. Todorov; X. J. Tan; S. T. Nonekowsky; G. A. Garcia; H. A. Carlson, *Biophys. J.* **2005**, *89*, 1965–1977.
242. K. A. Todorov; G. A. Garcia, *Biochemistry*, **2006**, *45*, 617–625.
243. C. Romier; J. E. W. Meyer; D. Suck, *FEBS Lett.* **1997**, *416*, 93–98.
244. N. Tidten; B. Stengl; A. Heine; G. A. Garcia; G. Klebe; K. Reuter, *J. Mol. Biol.* **2007**, *374*, 764.
245. H. Grosjean; J. Edqvist; K. B. Straby; R. Giege, *J. Mol. Biol.* **1996**, *255*, 67–85.
246. S. T. Nonekowsky; G. A. Garcia, *RNA* **2001**, *7*, 1432–1441.
247. A. W. Curnow; G. A. Garcia, *Biochimie* **1994**, *76*, 1183–1191.
248. F. L. Kung; S. Nonekowsky; G. A. Garcia, *RNA* **2000**, *6*, 233–244.
249. J. K. Hurt; S. Olgen; G. A. Garcia, *Nucleic Acids Res.* **2007**, *35*, 4905–4913.
250. S. T. Nonekowsky; F. L. Kung; G. A. Garcia, *J. Biol. Chem.* **2002**, *277*, 7178–7182.
251. R. Ishitani; O. Nureki; N. Nameki; N. Okada; S. Nishimura; S. Yokoyama, *Cell* **2003**, *113*, 383–394.
252. R. Ishitani; O. Nureki; S. Fukai; T. Kijimoto; N. Nameki; M. Watanabe; H. Kondo; M. Sekine; N. Okada; S. Nishimura; S. Yokoyama, *J. Mol. Biol.* **2002**, *318*, 665–677.
253. J. Sabina; D. Soll, *J. Biol. Chem.* **2006**, *281*, 6993–7001.
254. R. C. Morris; M. S. Elliott, *Mol. Genet. Metab.* **2001**, *74*, 147–159.
255. R. C. Morris; B. J. Brooks; P. Eriotou; D. F. Kelly; S. Sagar; K. L. Hart; M. S. Elliott, *Nucleic Acids Res.* **1995**, *23*, 2492–2498.
256. M. Hengesbach; M. Meusburger; F. Lyko; M. Helm, *RNA* **2008**, *14*, 180.
257. Q. Dai; R. Fong; M. Saikia; D. Stephenson; Y. T. Yu; T. Pan; J. A. Piccirilli, *Nucleic Acids Res.* **2007**, *35*, 6322.
258. M. Buchhaupt; C. Pelfer; K. D. Entian, *Anal. Biochem.* **2007**, *361*, 102.
259. C. J. McCleverty; M. Hornsby; G. Spraggon; A. Kreuzsch, *J. Mol. Biol.* **2007**, *373*, 1243.
260. E. G. Mueller, *Nat. Struct. Biol.* **2002**, *9*, 320–322.
261. V. Arluison; C. Hountondji; B. Robert; H. Grosjean, *Biochemistry* **1998**, *37*, 7268–7276.
262. J. J. Prior; D. V. Santi, *J. Biol. Chem.* **1984**, *259*, 2429–2434.
263. X. M. Zhao; D. A. Horne, *J. Biol. Chem.* **1997**, *272*, 1950–1955.

264. V. Ramamurthy; S. L. Swann; C. J. Spedaliere; E. G. Mueller, *Biochemistry* **1999**, *38*, 13106–13111.
265. D. A. Frendewey; D. M. Kladianos; V. G. Moore; I. I. Kaiser, *Biochim. Biophys. Acta* **1982**, *697*, 31–40.
266. L. X. Huang; M. Pookanjanatavip; X. G. Gu; D. V. Santi, *Biochemistry* **1998**, *37*, 344–351.
267. X. R. Gu; Y. Liu; D. V. Santi, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14270–14275.
268. C. Hoang; A. R. Ferre-D'Amare, *Cell* **2001**, *107*, 929–939.
269. C. J. Spedaliere; J. M. Ginter; M. V. Johnston; E. G. Mueller, *J. Am. Chem. Soc.* **2004**, *126*, 12758–12759.
270. I. Perez-Arellano; J. Gallego; J. Cervera, *FEBS J.* **2007**, *274*, 4972.
271. M. Charette; M. W. Gray, *IUBMB Life* **2000**, *49*, 341–351.
272. S. Hur; R. M. Stroud, *Mol. Cell* **2007**, *26*, 189.
273. E. V. Koonin, *Nucleic Acids Res.* **1996**, *24*, 2411–2415.
274. Y. Choffat; B. Suter; R. Behra; E. Kubli, *Mol. Cell Biol.* **1988**, *8*, 3332–3337.
275. J. Pienkowska; D. Michalowski; W. J. Krzyzosiak; Z. Szweykowska-Kulinska, *Biochim. Biophys. Acta – Gene Struct. Expression* **2002**, *1574*, 137–144.
276. F. Lecointe; G. Simos; A. Sauer; E. C. Hurt; Y. Motorin; H. Grosjean, *J. Biol. Chem.* **1998**, *273*, 1316–1323.
277. J. Wrzesinski; K. Nurse; A. Bakin; B. G. Lane; J. Ofengand, *RNA* **1995**, *1*, 437–448.
278. J. Wrzesinski; K. Nurse; A. Bakin; B. G. Lane; J. Ofengand, *Biochemistry* **1995**, *34*, 8904–8913.
279. I. Ansmant; S. Massenet; H. Grosjean; Y. Motorin; C. Branlant, *Nucleic Acids Res.* **2000**, *28*, 1941–1946.
280. C. Hoang; J. J. Chen; C. A. Vizthum; J. M. Kandel; C. S. Hamilton; E. G. Mueller; A. R. Ferre-D'Amare, *Mol. Cell* **2006**, *24*, 535.
281. I. Behm-Ansmant; A. Urban; X. J. Ma; Y. T. Yu; Y. Motorin; C. Branlant, *RNA* **2003**, *9*, 1371–1382.
282. E. Tran; J. Brown; E. S. Maxwell, *Trends Biochem. Sci.* **2004**, *29*, 343–350.
283. J. Ni; A. L. Tien; M. J. Fournier, *Cell* **1997**, *89*, 565–573.
284. T. Kiss, *EMBO J.* **2001**, *20*, 3617–3622.
285. P. Ganot; M. L. Bortolin; T. Kiss, *Cell* **1997**, *89*, 799–809.
286. C. Torchet; G. Badis; F. Devaux; G. Costanzo; M. Werner; A. Jacquier, *RNA* **2005**, *11*, 928–938.
287. Z. Szweykowska-Kulinska; B. Senger; G. Keith; F. Fasiolo; H. Grosjean, *EMBO J.* **1994**, *13*, 4636–4644.
288. F. J. Novo; D. C. Gorecki; G. Goldspink; K. D. MacDermot, *Gene Ther.* **1997**, *4*, 488–492.
289. H. Sakuraba; C. M. Eng; R. J. Desnick; D. F. Bishop, *Genomics* **1992**, *12*, 643–650.
290. D. Iwata-Reuyl, *Bioorg. Chem.* **2003**, *31*, 24–43.
291. H. Grosjean; V. de Crecy-Lagard; G. R. Bjork, *Trends Biochem. Sci.* **2004**, *29*, 519.
292. P. F. Crain; S. K. Sethi; J. R. Katze; J. A. McCloskey, *J. Biol. Chem.* **1980**, *255*, 8405–8407.
293. R. Suhadolnik; T. Uematsu, *J. Biol. Chem.* **1970**, *245*, 4365.
294. J. S. Reader; D. Metzgar; P. Schimmel; V. de Crecy-Lagard, *J. Biol. Chem.* **2004**, *279*, 6280.
295. R. Gaur; U. Varshney, *J. Bacteriol.* **2005**, *187*, 6893.
296. M. A. Swairjo; R. R. Reddy; B. Lee; S. G. Van Lanen; S. Brown; V. de Crecy-Lagard; D. Iwata-Reuyl; P. Schimmel, *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2005**, *61*, 945.
297. S. G. Van Lanen; J. S. Reader; M. A. Swairjo; V. de Crecy-Lagard; B. Lee; D. Iwata-Reuyl, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4264.
298. B. W. K. Lee; S. G. Van Lanen; D. Iwata-Reuyl, *Biochemistry* **2007**, *46*, 12844.
299. D. Quaranta; R. McCarty; V. Bandarian; C. Rensing, *J. Bacteriol.* **2007**, *189*, 5361.
300. S. D. Kinzie; B. Thern; D. Iwata-Reuyl, *Org. Lett.* **2000**, *2*, 1307.
301. S. G. Van Lanen; S. D. Kinzie; S. Matthieu; T. Link; J. Culp; D. Iwata-Reuyl, *J. Biol. Chem.* **2003**, *278*, 10491.
302. S. G. Van Lanen; D. Iwata-Reuyl, *Biochemistry* **2003**, *42*, 5312.
303. I. Mathews; R. Schwarzenbacher; D. McMullan; P. Abdubek; E. Ambing; H. Axelrod; T. Biorac; J. M. Canaves; H. J. Chiu; C. M. Deacon; M. DiDonato; M. A. Elsliger; A. Godzik; C. Grittini; S. K. Grzechnik; J. Hale; E. Hampton; G. W. Han; J. Haugen; M. Hornsby; L. Jaroszewski; H. E. Klock; E. Koesema; A. Kreuzsch; P. Kuhn; S. A. Lesley; I. Levin; M. D. Miller; K. Moy; E. Nigoghossian; O. Y. Jie; J. Paulsen; K. Quijano; R. Reyes; G. Spraggon; R. C. Stevens; H. van den Bedem; J. Velasquez; J. Vincent; A. White; G. Wolf; Q. P. Xu; K. O. Hodgson; J. Wooley; I. A. Wilson, *Proteins* **2005**, *59*, 869.
304. C. Grimm; R. Ficner; T. Sgraja; P. Haebel; G. Klebe; K. Reuter, *Biochem. Biophys. Res. Commun.* **2006**, *351*, 695.
305. U. Gunduz; J. R. Katze, *J. Biol. Chem.* **1984**, *259*, 1110.
306. G. M. Kirtland; T. D. Morris; P. H. Moore; J. J. O'Brian; C. G. Edmonds; J. A. McCloskey; J. R. Katze, *J. Bacteriol.* **1988**, *170*, 5633.
307. D. J. Vandenbergh; M. D. Grant; V. Severns, *Mol. Cell. Probes* **2003**, *17*, 319–320.
308. N. Okada; S. Nishimura, *Nucleic Acids Res.* **1977**, *4*, 2931–2937.
309. E. Haumont; L. Droogmans; H. Grosjean, *Eur. J. Biochem.* **1987**, *168*, 219.
310. J. C. Salazar; A. Ambrogelly; P. F. Crain; J. A. McCloskey; D. Soll, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7536.
311. S. Blobstein; D. Grunbergh; I. Weinstein; K. Nakanishi, *Biochemistry* **1973**, *12*, 188.
312. L. Droogmans; H. Grosjean, *EMBO J.* **1987**, *6*, 477.
313. V. de Crecy-Lagard, *Pract. Bioinf.* **2004**, *15*, 169–190.
314. W. F. Waas; V. de Crecy-Lagard; P. Schimmel, *J. Biol. Chem.* **2005**, *280*, 37616.
315. H. R. Kalhor; M. Penjwini; S. Clarke, *Biochem. Biophys. Res. Commun.* **2005**, *334*, 433.
316. A. Noma; Y. Kirino; Y. Ikeuchi; T. Suzuki, *EMBO J.* **2006**, *25*, 2142.
317. Y. Suzuki; A. Noma; T. Suzuki; M. Senda; T. Senda; R. Ishitani; O. Nureki, *J. Mol. Biol.* **2007**, *372*, 1204.
318. S. Goto-Ito; R. Ishii; T. Ito; R. Shibata; E. Fusatomi; S. Sekine; Y. Bessho; S. Yokoyama, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 1059.
319. R. Hall, *The Modified Nucleosides in Nucleic Acids*; Columbia University Press: New York, 1971.

Biographical Sketches



George A. Garcia received his Ph.D. in pharmaceutical chemistry from the University of California, San Francisco in 1988, where he worked with Professor George L. Kenyon on the enzymology of benzoylformate decarboxylase and inhibitor design for tyrosine kinase. He was an NIH-sponsored postdoctoral fellow in the lab of Professor Alan R. Fersht in Cambridge, England in 1988 and 89. There he worked on tyrosyl-tRNA synthetase. He joined the faculty of the University of Michigan, College of Pharmacy in 1990 and presently works as associate professor of medicinal chemistry. Professor Garcia's research interests are in the area of the enzymology of RNA modification and in novel antibiotic drug discovery. He has served on the NIH Bioorganic and Natural Products Chemistry Study Section. Professor Garcia's research study has been supported by grants from NIH and from NSF. He has authored approximately 45 papers, reviews, and book chapters.



Julie K. (Cutcher) Hurt received a B.A. in chemistry (biology minor) from Kalamazoo College in 2004. She is completing her Ph.D. under the guidance of Professor George A. Garcia in the Department of Medicinal Chemistry, University of Michigan. Her research involves the primary virulence factor in *Shigella flexneri*, VirF, a member of the AraC family of transcriptional activators. Ms. Hurt is studying the link between VirF expression and activity as it relates to post-transcriptional RNA modification by TGT.



Yi-Chen Chen, a native of Taiwan, received his B.S. in pharmacy and M.S. in pharmaceuticals with Professor Wen-Jen Lin (on the topic of developing novel drug carriers via biocompatible polymers) from the National Taiwan University. He is currently a Ph.D. candidate working under the guidance of Professor George A. Garcia in the Department of Medicinal Chemistry, University of Michigan. His research project is to characterize a human tRNA-modifying enzyme, TGT in order to gain further insight into the biological significance of queuosine modification.

6.21 Riboswitches

Tina M. Henkin, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

| | | |
|------------|---|-----|
| 6.21.1 | Introduction | 743 |
| 6.21.2 | Mechanisms of Gene Regulation | 744 |
| 6.21.2.1 | Transcriptional Attenuation | 744 |
| 6.21.2.2 | Translation Initiation | 745 |
| 6.21.2.3 | mRNA Stability and Processing | 746 |
| 6.21.3 | Regulatory Signals | 746 |
| 6.21.3.1 | Mebabolite-Binding Riboswitches | 748 |
| 6.21.3.2 | RNA Thermosensors | 750 |
| 6.21.3.3 | RNA-Binding Riboswitches | 751 |
| 6.21.3.4 | Composite Riboswitches | 752 |
| 6.21.4 | Structure of Riboswitch RNAs | 753 |
| 6.21.4.1 | Riboswitch Architectures | 753 |
| 6.21.4.1.1 | Single-domain riboswitches | 753 |
| 6.21.4.1.2 | Multidomain riboswitches | 754 |
| 6.21.4.2 | Structural Biology | 754 |
| 6.21.5 | Riboswitches and Gene Identification | 754 |
| 6.21.5.1 | Genes Identified Based on Association with a Riboswitch | 755 |
| 6.21.6 | Using Riboswitches to Control Gene Expression | 755 |
| 6.21.6.1 | Natural Riboswitches as Genetic Control Elements for Heterologous Gene Expression | 755 |
| 6.21.6.2 | Synthetic Riboswitches | 756 |
| 6.21.7 | Perspectives and Future Directions | 756 |
| References | | 757 |

6.21.1 Introduction

Recent studies of a variety of genetic regulatory systems in bacteria have revealed mechanisms in which an RNA transcript can directly measure a physiological parameter, and use this information to influence the expression of gene(s) encoded within that transcript.¹ Regulatory RNAs of this type, termed riboswitches, utilize a segment of RNA to monitor a specific regulatory signal related to the function of that gene. Riboswitches are commonly found at the 5' end of the regulated transcript, upstream of the start of the regulated coding sequence. This region of the RNA, designated the leader sequence or leader RNA, corresponds to the 5'-untranslated region (5'-UTR) in eukaryotic mRNAs, and contains elements responsible for signal recognition and for the effects on downstream gene expression. A few riboswitch elements have been identified in other regions of the transcript in eukaryotic genes. Most known riboswitch RNAs bind small molecules, often a substrate or end product of the biosynthetic pathway in which the regulated gene is involved. In metabolite-binding riboswitches, binding of the effector results in a structural rearrangement in the RNA that affects the fate of the transcript. Other classes of riboswitches can bind specific cellular RNAs (such as tRNA) or respond to physical parameters (such as temperature) with a similar signal-dependent RNA structural rearrangement that impacts gene expression.

Riboswitches are similar in general principles and physiological function to a number of classical regulatory systems in which RNA structural rearrangements impact gene expression.² For example, RNA-binding proteins (like *Bacillus subtilis* TRAP) can bind to a target mRNA to control transcription attenuation or translation initiation; their RNA-binding activity can respond to a regulatory signal (e.g., tryptophan

availability), allowing information about a cellular component to be transmitted to the gene expression machinery via a shift in the structure of the mRNA.³ Physiological signals such as tRNA charging (which is a measure of both the availability of a specific amino acid and the functionality of the tRNA aminoacylation machinery) or the presence of antibiotics can also be monitored by a ribosome as it attempts to translate a short peptide encoded within the leader RNA; the processivity of the translating ribosome affects the leader RNA structure, which in turn affects the fate of the transcript.^{2,4} The major difference between these systems and riboswitch RNAs is that the latter systems use the RNA transcript to directly sense and respond to the physiological signal, without a need for other factors (e.g., RNA-binding proteins or a translating ribosome) to monitor and transmit the signal to the gene expression machinery. Each riboswitch system is highly specific for its cognate signal, and the leader RNAs exhibit patterns of conserved sequence and/or structural elements that are responsible for the specificity of this response.

6.21.2 Mechanisms of Gene Regulation

Riboswitch-mediated gene regulation occurs after transcription initiation, as the regulatory element is part of the regulated transcript. In bacteria, preferred modes of regulation are at the levels of premature termination of transcription (referred to as transcriptional attenuation) and translation initiation. Rare examples identified so far in eukaryotic cells affect mRNA stability and splicing.

6.21.2.1 Transcriptional Attenuation

Regulation at the level of premature termination of transcription is most commonly found in low G + C Gram-positive bacteria, but can also be found in other groups of bacteria. The transcriptional unit of genes regulated at this level includes a transcriptional terminator (or attenuator), positioned upstream of the regulated coding sequence(s). This arrangement results in two classes of transcripts originating from a single promoter, a shorter transcript resulting from termination at the attenuator site and a longer transcript resulting from transcription elongation past the attenuator (**Figure 1(a)**). The regulated attenuator is usually an intrinsic terminator capable of functioning in the absence of dedicated termination factors such as Rho protein. Intrinsic terminators are active within the nascent RNA transcript, and they consist of a G + C-rich helix immediately followed by a U-run that forms a U-A RNA–DNA helix within RNA polymerase (RNAP). Folding of the nascent RNA into the helix, while RNAP is paused at the end of the U-run, affects the stability of the transcription elongation complex and results in termination of transcription and release of RNAP and the terminated transcript from the DNA template.

An intrinsic terminator can be sensitive to RNA rearrangements when sequences that participate in the formation of the terminator helix can be sequestered into an alternate RNA structure designated an anti-terminator. The decision between formation of the mutually exclusive terminator and antiterminator structures can be determined by binding of an extrinsic factor; in riboswitch systems, the RNA structure is modulated by direct binding of an effector molecule (such as a small molecule or RNA). Binding of effector molecules can promote the formation of the terminator (resulting in inhibition of transcription of the downstream coding sequences) or the antiterminator (resulting in increased gene expression). In systems in which the terminator form of the RNA is more stable than the antiterminator form, the expression is off in the absence of effector binding, so that the effector acts as an inducer to increase the expression of the downstream gene(s) (**Figure 1(b)**). If the antiterminator is more stable than the terminator, then expression is on in the absence of the effector, which serves as a repressor molecule (**Figure 1(c)**). In this case, effector binding often stabilizes an element that includes sequences that would otherwise form the 5' side of the antiterminator structure; this alternate element serves as an anti-antiterminator, as it competes with antiterminator formation, and stabilization of the anti-antiterminator prevents the formation of the more stable antiterminator and allows the less stable terminator helix to form (**Figure 1(d)**).

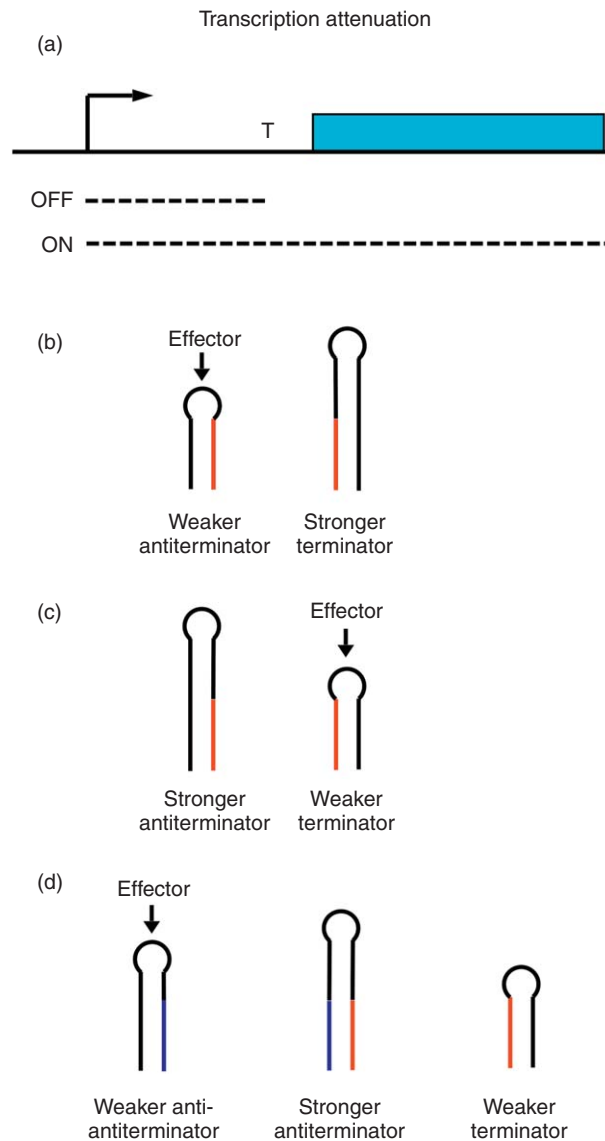


Figure 1 Regulation at the level of transcriptional attenuation. (a) A gene (or operon) that is regulated at the level of transcriptional attenuation contains a transcriptional termination signal (T) between the promoter (bent arrow) and the start of the regulated coding sequence (blue box). Constitutive transcription initiation results in synthesis of two classes of transcripts (dashed lines). Under conditions where the terminator is active, a short transcript corresponding to the leader RNA is synthesized (OFF); under conditions where the terminator is inactive, a long transcript that includes the regulated coding sequence is synthesized (ON). (b) In riboswitch RNAs in which the terminator helix is more stable than the antiterminator helix, expression is off in the absence of the effector and binding of the effector is required to stabilize the antiterminator and allow readthrough of the termination site. (c) In riboswitch RNAs in which the antiterminator is more stable than the terminator, expression is on in the absence of the effector and binding of the effector stabilizes the terminator, turning off the expression of the downstream gene. (d) Systems like those in (c) often use an additional structural element (an anti-antiterminator) that competes with the strong antiterminator; binding of the effector to the anti-antiterminator stabilizes that structure, preventing the formation of the antiterminator and allowing the formation of the terminator helix to repress downstream gene expression. The red and blue lines show sequences shared between alternate structures.

6.21.2.2 Translation Initiation

Signal-dependent RNA rearrangements also commonly regulate bacterial gene expression at the level of translation initiation. Regulation at this level is commonly found in Gram-negative bacteria and high G + C Gram-positive bacteria (notably members of the Actinobacteria). Bacterial translation initiation is dependent

on access of the translation initiation region of an mRNA to the 30S ribosomal subunit. The ribosome-binding site (RBS) of the mRNA consists of the Shine–Dalgarno (SD) sequence (for which the consensus is GGAGG, which is complementary to a sequence at the 3' end of 16S ribosomal RNA) and the start codon (usually AUG, although GUG and UUG are also functional in bacteria); spacing between these elements is generally 6–10 nt. Sequestration of the translation initiation region into a helical element prevents ribosome binding, and therefore inhibits expression. The full-length transcript is synthesized constitutively, and its activity as a template for translation is the target for regulation (**Figure 2(a)**).

Binding of regulatory molecules can determine whether the RNA folds into an inhibitory structure, usually termed an anti-SD (ASD)–SD helix, or a competing structure in which the ASD region is sequestered by pairing to an anti-ASD (AASD) sequence and is therefore unavailable to bind to the SD. The formation of the ASD–SD helix prevents translation initiation, while the formation of the competing AASD–ASD structure liberates the SD sequence and allows binding of the ribosome. In systems where the ASD–SD helix is stronger than the AASD–ASD helix, the transcript is not efficiently recognized by the ribosome and translation is repressed. The effector molecule stabilizes the weaker AASD–ASD helix, allowing translation to occur (**Figure 2(b)**). In systems where the AASD–ASD helix is more stable than the ASD–SD helix, the default state of the system is on, and binding of the effector to the ASD–SD helix is required to stabilize that helix and inhibit translation of the downstream coding sequence (**Figure 2(c)**). In systems of this type, the RNA often can form a third element (AAASD–AASD) that sequesters sequences that would otherwise form the AASD–ASD sequence, and that third element is stabilized by binding of the effector (**Figure 2(d)**); this element is formally equivalent to the anti-antiterminator described above.

Transcription termination elements must act during transcription of the 5' region of the RNA, before RNAP reaches the termination site. Transcription and translation are coupled in bacteria, so that binding of a ribosome to the mRNA and translation initiation can occur immediately upon emergence of the portion of the transcript containing the ribosome-binding site from the transcription elongation complex. Regulation at the level of translation initiation therefore can occur either as the transcript is synthesized or after the full-length transcript has been made. Transcription termination decisions are essentially irreversible, as the system is committed to either termination or antitermination once the transcription elongation complex reaches the position of the attenuator. In contrast, regulation at the level of translation initiation has the potential to allow multiple decisions within the lifetime of the transcript, as the RNA can refold into alternate structures as the effector molecule binds and releases from the transcript. The possibility of reversible switching between the on and off states is dependent on the on and off rates for effector binding and the stability of the transcript in the cell.

6.21.2.3 mRNA Stability and Processing

While regulation by transcription termination and translation initiation predominates in the riboswitch systems that have been characterized to date, there are a few examples that operate at other steps of gene expression. The *glmS* riboswitch is unique in that binding of the effector promotes transcript self-cleavage by activation of a ribozyme activity inherent in the *glmS* RNA.⁵ The cleaved transcript is then subject to degradation by RNase J1,⁶ resulting in decreased expression of the downstream gene (**Figure 3**). In systems of this type, the synthesis rate of the full-length transcript is constant. However, the steady-state level of the transcript changes in response to effector binding. Riboswitches have also been identified in plants and fungi and they have been shown to affect transcript stability and mRNA splicing; these all belong to the Thi box family, and respond to thiamine pyrophosphate (TPP). The eukaryotic riboswitches identified to date are located in the 3'-untranslated region (3'-UTR) of the transcript or are positioned within an intron.^{7,8}

6.21.3 Regulatory Signals

Riboswitches can respond to a variety of physiological signals, including small molecules, physical parameters such as temperature, and cellular RNAs such as tRNA. In each case, the signal that is monitored is directly related to the genes that are regulated, and the direction of the response (i.e., increased or decreased gene expression) is consistent with the nature of the signal and the physiological function of the regulated gene. In a number of cases,

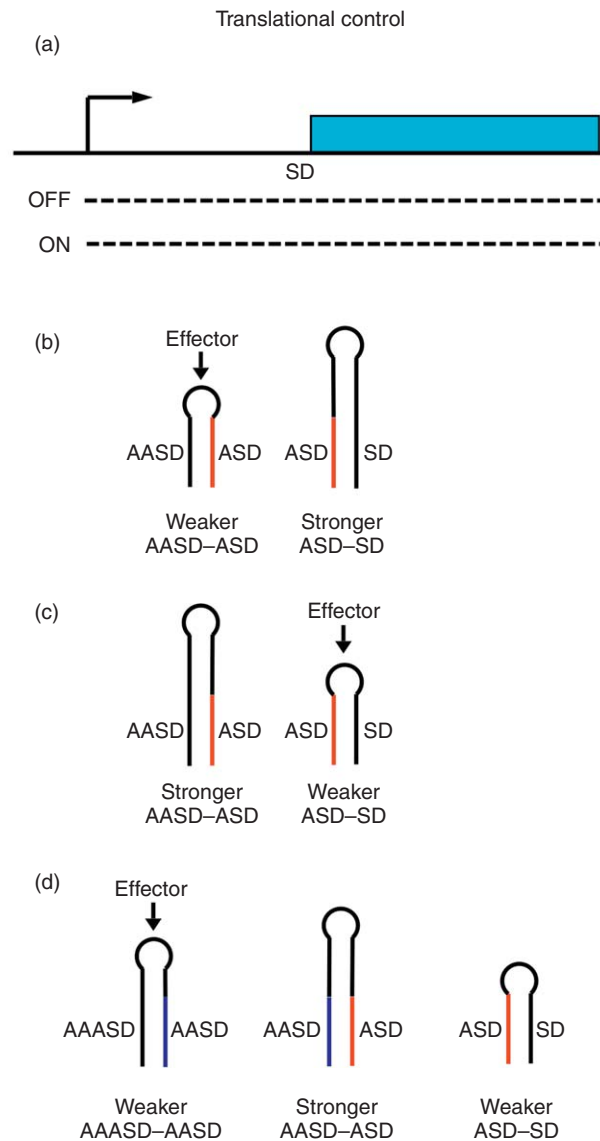


Figure 2 Regulation at the level of translation initiation. (a) Genes that are regulated at the translational level usually have promoters (bent arrow) that are always active, resulting in constitutive synthesis of the full-length transcript (dotted line). The Shine–Dalgarno (SD) sequence located at the start of the coding sequence (blue box) is the target for regulation, by sequestration into a structure in which the SD is paired with an anti-SD (ASD) sequence that prevents ribosome binding (OFF). The formation of the ASD–SD structure often competes with an alternate AASD–ASD structure that releases the SD and allows translation to occur (ON). (b) In systems where the ASD–SD structure is more stable than the AASD–ASD structure, binding of the effector can stabilize the AASD–ASD structure, allowing translation to occur. (c) In systems in which the AASD–ASD structure is more stable than the ASD–SD structure, expression is off unless the effector binds to and stabilizes the ASD–SD structure. (d) Systems like those in (c) often use an additional alternate structure that sequesters sequences (blue) that would otherwise participate in the AASD–ASD structure. This additional structure (AAASD–AASD) therefore competes with the AASD–ASD structure. The formation of the AAASD–AASD structure is stabilized by binding of the effector, and allows repression of gene expression by releasing the ASD sequence to pair with the SD sequence. Red and blue lines represent sequences that participate in alternate pairing.

the same signal-sensing element within the RNA (e.g., the ligand-binding domain for metabolite-binding riboswitches) can be used to regulate gene expression at multiple levels (e.g., by either transcription attenuation or translation initiation) through a combination of the same effector sensing domain with either terminator versus antiterminator alternate structures or ASD–SD versus AASD–ASD alternative structures.

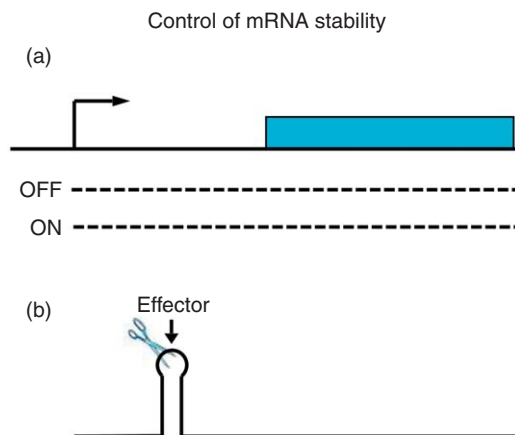


Figure 3 Regulation at the level of mRNA degradation. Genes that are regulated by differential mRNA stability usually have promoters (bent arrow) that are always active, resulting in constitutive synthesis of the full-length transcript (dotted line). (b) Binding of the effector triggers degradation of the transcript. In the *glmS* system, this occurs by effector-dependent activation of a self-cleavage (ribozyme) activity inherent in the leader RNA, resulting in degradation of the transcript by cellular RNases (OFF). The transcript is stable in the absence of effector binding, allowing expression of the downstream coding sequence (ON).

6.21.3.1 Metabolite-Binding Riboswitches

A list of known riboswitches that respond to a variety of small molecules is shown in [Table 1](#). The ligands sensed by these riboswitches range from tiny metal ions to amino acids to large, complex cofactors. There is no clear correlation between the size of the ligand and the size of the corresponding ligand-binding RNA element.⁹ Some of the smaller ligands (e.g., lysine) utilize large RNAs consisting of multiple helical elements assembled into a complex tertiary structure.^{10–13} In contrast, the glycine-binding element is small and not very complicated in the overall arrangement.¹⁴ Three completely different RNA elements have been identified that recognize *S*-adenosylmethionine (SAM),^{15–17} and crystal structures are available for the ligand-bound form of all three of these RNAs.^{18–20} The three classes of SAM riboswitches utilize completely different structural arrangements to recognize SAM, demonstrating the plasticity of RNA in forming a specific ligand recognition pocket. Even a single riboswitch type can be highly variable. For example, natural variants of the SAM-responsive S_{MK} box riboswitch range in size from ~80 to ~500 nt, with the extra sequence present in long helical elements inserted into variable regions of the riboswitch.¹⁷

All metabolite-binding riboswitches exhibit high specificity for their cognate ligand, and discriminate efficiently against related compounds. Specificity is generally correlated with the probability that a related compound would be found in the natural environment. For example, the SAM-binding RNAs all exhibit high discrimination against *S*-adenosylhomocysteine (SAH), a cellular by-product of utilization of SAM as a methyl donor in methylation reactions.^{15–17} Failure to bind SAH is essential for riboswitch function *in vivo*, as repression of genes involved in SAM biosynthesis when SAM pools are low but SAH pools are high would result in a deficiency in SAM production. The three classes of SAM riboswitch RNAs exhibit differential ability to recognize other SAM-related compounds that are unlikely to be encountered in nature, consistent with selective pressure for discrimination against SAH but not other SAM-related compounds. It is interesting to note that a different RNA element has evolved to recognize SAH but not SAM;²¹ this element is used to activate a pathway involved in recycling of SAH to methionine, and the expression of the regulated genes is induced when SAH is high (in contrast to feedback repression of genes involved in SAM biosynthesis by high SAM).²²

While most metabolite-binding riboswitches utilize complex RNA arrangements to recognize the entire molecular structure of the ligand, nucleotide-binding riboswitches, like the purine-responsive G box, also utilize the base-pairing properties of the ligand to recognize the appropriate nucleotide. G box RNAs include a single cytosine residue within the ligand-binding pocket that pairs with the guanine ligand; replacement of the cytosine with a uridine is sufficient to switch the specificity of the riboswitch, and allows binding of adenine in place of guanine.^{23,24}

Table 1 Classes of metabolite-binding riboswitch RNAs

| <i>Riboswitch</i> | <i>Ligand</i> | <i>Mechanism</i> | <i>Description</i> ^a | <i>Structure</i> ^a |
|-------------------------------|------------------------------|--|---------------------------------|-------------------------------|
| FMN | Flavin mononucleotide (FMN) | Transcription, translation | 25, 26 | 27 |
| THI box | Thiamine pyrophosphate (TPP) | Transcription, translation, mRNA stability | 26, 28, 29 | 30–32 |
| B ₁₂ | Adenosylcobalamin | Translation | 33, 34 | |
| S box (SAM-I) | S-adenosylmethionine (SAM) | Transcription (translation rare) | 15, 35, 36, 37 | 18 |
| SAM-II | S-adenosylmethionine (SAM) | Translation | 16 | 19 |
| S _{MK} box (SAM-III) | S-adenosylmethionine (SAM) | Translation | 17, 38 | 20 |
| SAH box | S-adenosylhomocysteine (SAH) | Translation | 21 | |
| L box | Lysine | Transcription, translation | 10, 11 | 12, 13 |
| Glycine | Glycine | Transcription | 14 | |
| Purine | Guanine/adenine | Transcription, translation | 39, 40 | 23, 24 |
| dG | Deoxyguanosine | Transcription | 41 | |
| Cyclic di-GMP | Cyclic di-GMP | Transcription, translation | 42 | |
| glmS | Glucosamine-6-phosphate | mRNA stability | 5 | 65, 66 |
| preQ1 | 7-Aminoethyl 7-deazaguanine | Transcription | 43 | |
| Mg | Magnesium | Transcription | 44, 45 | 45 |

^a Reference citation.

The metal-binding riboswitches represent a special class of riboswitches that recognize very small ligands (e.g., Mg ions). At least two distinct classes of Mg-responsive riboswitch RNA have been identified, and these differ not only in their overall structure but also in their ligand specificity, as the *Escherichia coli* Mg-binding RNA exhibits higher selectivity for magnesium versus manganese than that observed for the *B. subtilis* Mg-binding RNA.^{44,45} It is unclear whether the difference in selectivity observed *in vitro* reflects a difference in gene expression *in vivo*, as the ability to respond to a noncognate effector depends on the relative intracellular concentrations of both the cognate and noncognate ligands.

Most metabolite-binding riboswitches use a negative feedback response, whereby the presence of a compound that is an end product of a biosynthetic pathway signals the cell to repress synthesis of gene products required for the function of that pathway. A similar pattern is observed for transport systems, where the availability of a particular metabolite within the cell signals that further synthesis of systems to import that molecule into the cell is unnecessary. A rarer class of metabolite-binding riboswitches exhibit induction of gene expression by a molecule that serves as a substrate for that pathway (e.g., high glycine induces the expression of enzymes involved in glycine cleavage¹⁴ and high SAH induces a pathway for conversion of SAH to methionine¹⁶). In all cases for which the effector has been identified, a direct connection can be established between the regulated gene and the molecule that modulates its expression.

6.21.3.2 RNA Thermosensors

RNA thermosensors usually consist of a negative-acting RNA element that sequesters the RBS and inhibits translation at low temperature; increased temperature results in unfolding of this inhibitory helix and induction of expression (Figure 4). These are commonly found in genes involved in cellular responses to transition to higher temperature, for example in heat-shock genes.⁴⁶ A classic example is found in the *E. coli* *rpoH* gene, encoding the heat-shock sigma factor (σ^H) responsible for transcription of genes involved in the heat-shock

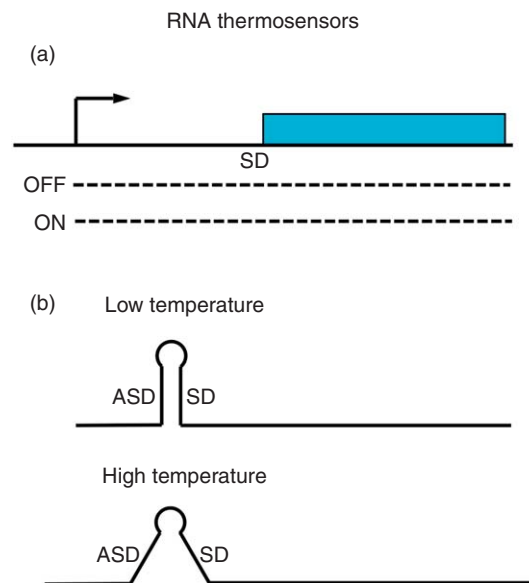


Figure 4 RNA thermosensors. (a) All known RNA thermosensors regulate gene expression at the level of translation initiation. Genes in this group usually have promoters (bent arrow) that are always active, resulting in constitutive synthesis of the full-length transcript (dotted line). The Shine–Dalgarno (SD) sequence located at the start of the coding sequence (blue box) is the target for regulation, by sequestration into a structure in which the SD is paired with an anti-SD (ASD) sequence that prevents ribosome binding (OFF). (b) In most RNA thermosensors, the ASD–SD structure forms during growth at low temperature, and gene expression is repressed under these conditions. This structure is destabilized at high temperature, releasing the SD sequence and allowing translation initiation to occur (ON). A rarer class of RNA thermosensors exhibit formation of the ASD–SD structure at high temperature, and release of that structure (and induction of expression) at lower temperatures, usually because of the formation of an alternate structure that sequesters the ASD sequence.

response.⁴⁷ During steady-state growth at normal temperature, cells maintain a basal level of *rpoH* transcript that is held in a translationally inactive state by a thermosensor element that sequesters the ribosome-binding site. A shift to high temperature unfolds the thermosensor, allowing a burst of RpoH translation and a rapid increase in RpoH protein abundance. The RpoH (σ^H) protein binds to core RNAP and directs the synthesis of a large amount of *rpoH* transcript that is immediately active in translation due to destabilization of the inhibitory helix. A shift back to normal temperature allows the formation of the inhibitory helix in all *rpoH* transcript molecules in the cell. This results in a rapid shutdown of translation of the *rpoH* transcript. Inhibition of RpoH synthesis is accompanied by other mechanisms that result in a return to the normal state by inactivation of σ^H protein. A similar mechanism is used to upregulate translation of the PrfA regulatory protein in *Listeria monocytogenes*, which uses the thermosensor in the *prfA* transcript to induce virulence gene expression in response to the transition between the free-living state and entrance into the mammalian host.⁴⁸

RNA thermosensors that turn on gene expression in response to cold shock have also been observed.⁴⁶ In systems of this type, the RNA element that forms at low temperature is predicted to compete with a structure that would otherwise inhibit expression, so that a shift to a lower growth temperature results in induction of the cold-shock response. These systems are less well characterized than the heat-responsive thermosensor RNAs.

6.21.3.3 RNA-Binding Riboswitches

The role of small RNAs (sRNAs) in gene regulation in all cell types has recently become a major area of interest.^{49,50} These RNAs usually use base pairing with their target mRNA transcripts to mediate expression, although rarer classes use different mechanisms, including titration of RNA-binding proteins or direct binding to RNAP.^{51,52} Small regulatory RNAs can be encoded on the opposite strand of their target gene (*cis*-encoded sRNAs), in which case they exhibit perfect complementarity to their target transcript, or they can be encoded elsewhere in the genome (*trans*-encoded sRNAs), in which case they exhibit only partial complementarity. In most cases, sRNAs that bind mRNAs affect the fate of that mRNA through occlusion of the ribosome-binding site or through creation of a site that triggers RNA degradation. Mechanisms of this type do not involve structural rearrangements of the target mRNA, and are therefore not considered riboswitches. A few examples have been uncovered in which binding of the sRNA to a region upstream of the ribosome-binding site results in the release of a structure that would otherwise prevent translation initiation (e.g., induction of *rpoS* translation by the DsrA sRNA⁵³); systems of this type more closely mimic the ‘riboswitch’ model, but differ in that synthesis of the sRNA represents the direct genetic response to the physiological signal, which is then transmitted to the *rpoS* target by the sRNA.

The T box regulatory mechanism represents the clearest example of an RNA-responsive riboswitch.^{54–56} In most genes in this large family, the leader RNA contains competing terminator and antiterminator structures, as described for the metabolite-binding riboswitch RNAs. The switch between the terminator and antiterminator states is determined by binding of a specific cellular tRNA to the nascent RNA transcript (Figure 5). Uncharged tRNA binds at two positions: (1) the anticodon of the tRNA pairs with a three-nucleotide sequence that represents the codon specific for that tRNA and (2) four unpaired nucleotides at the acceptor end of the tRNA pair with four nucleotides in an internal bulge within the antiterminator. The first pairing is primarily responsible for specific recognition of a single tRNA class, while the second pairing occurs only with uncharged tRNA, and allows discrimination between uncharged and charged tRNA species. Stabilization of the antiterminator by uncharged tRNA results in readthrough of the termination site and increased expression of the downstream gene(s). Aminoacylation of the tRNA does not interfere with the codon–anticodon interaction, but blocks stabilization of the antiterminator; as a result, the charged tRNA acts as a competitive inhibitor for binding of the uncharged tRNA effector.⁵⁷ Like many metabolite-binding riboswitches, the T box riboswitch can also function at the level of translation initiation, by replacing the terminator helix with a helix that sequesters the translation initiation site.^{58,59}

The T box system is unique among RNA-responsive systems in that the intracellular concentration of the regulatory RNA (the cognate tRNA) remains constant. In contrast to systems that monitor changes in effector concentration, the T box leader RNA monitors the aminoacylation state of the cognate tRNA, which varies as a consequence of changes in cellular physiology (e.g., availability of the aminoacyl-tRNA synthetase or the amino acid necessary for charging of that tRNA). Genes regulated by this mechanism are usually involved in

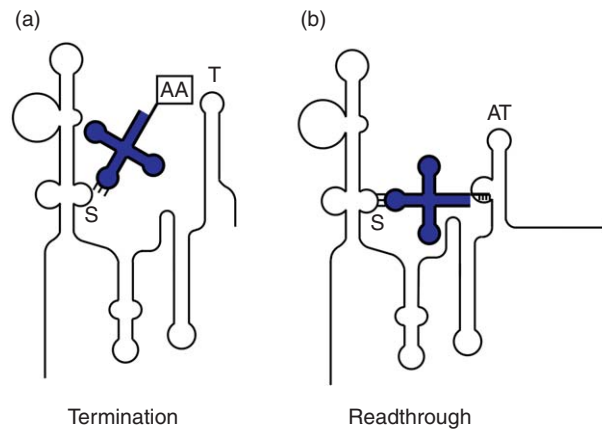


Figure 5 RNA-mediated regulation by the T box mechanism. (a) Efficient charging of a specific tRNA isoacceptor (e.g., tRNA^{Tyr} for the *tyrS* operon, which encodes tyrosyl-tRNA synthetase) results in pairing of the tRNA anticodon loop with a matching codon (a UAC tyrosine codon for *tyrS*) in the Specifier Loop (S) of the T box leader RNA. Charged tRNA is unable to participate in the second leader RNA–tRNA interaction (tRNA acceptor end pairing to the antiterminator) because of steric hindrance caused by the presence of the amino acid (AA). Failure to stabilize the antiterminator results in the formation of the terminator helix (T) and premature termination of transcription. (b) The uncharged form of the specific tRNA can carry out both anticodon loop–Specifier Loop binding and acceptor end–antiterminator binding. Binding at the antiterminator (AT) stabilizes the antiterminator and prevents the formation of the more stable terminator helix. Stabilization of the antiterminator allows transcription to continue past the termination site, resulting in the expression of the downstream coding sequence. Each gene in the T box family monitors a specific tRNA class, which matches the amino acid specificity of the regulated gene. Discrimination between uncharged and charged tRNA occurs at the antiterminator.

aminoacylation of a specific tRNA (or synthesis or acquisition of the amino acid used to aminoacylate that tRNA). This pattern is consistent with the connection between gene function and signal sensing described above for metabolite-binding riboswitches. The T box system is also unique in that two different effectors are monitored simultaneously, but with opposite effects. Uncharged tRNA stimulates antitermination, whereas charged tRNA inhibits antitermination (by competing with uncharged tRNA for binding to the nascent RNA). The system therefore monitors the ratio of uncharged to charged tRNA, rather than the absolute concentration of the uncharged tRNA. tRNA-directed antitermination and tRNA–leader RNA binding have been reproduced *in vitro*,^{55,57} supporting the model that T box family leader RNAs act as true riboswitches.

6.21.3.4 Composite Riboswitches

Several classes of riboswitches have been found in tandem arrangements. The simplest arrangement involves tandem arrays of the same riboswitch, which allows higher sensitivity to the regulatory signal. The first example of this was the triple repeat of a tRNA^{Thr}-responsive T box riboswitch upstream of the *B. subtilis* *thrZ* gene, which encodes a secondary copy of threonyl-tRNA synthetase.⁶⁰ Each T box element in the array requires binding of a separate molecule of uncharged tRNA^{Thr} to promote readthrough of the termination site. As a consequence, the *thrS* gene, which encodes the primary threonyl-tRNA synthetase and contains a single T box riboswitch, is induced first, when tRNA^{Thr} charging is reduced only slightly, whereas the secondary *thrZ* gene is expressed only upon a more significant drop in charging of tRNA^{Thr}. The use of multiple copies of the same riboswitch is also found in a TPP-responsive gene in *Bacillus anthracis*, which is preceded by tandem copies of the Thi box riboswitch.⁶¹

Tandem copies of different classes of riboswitches can also promote response to multiple signals. For example, the *Bacillus clausii* *metE* gene, which encodes a B₁₂-independent form of methionine synthase, is regulated by tandem S box and B₁₂ riboswitches.⁶² Expression is repressed by both SAM (which serves as an indicator that methionine pools are high) and B₁₂ (which represses *metE* synthesis in favor of *metH*, which encodes the more efficient B₁₂-dependent form of the enzyme). Similarly, a cysteine biosynthesis operon in

Clostridium acetobutylicum is regulated by a combination of an S box riboswitch (which responds to SAM), a T box riboswitch (which responds to tRNA^{Cys}), and an antisense RNA.⁶³ The combination of multiple signal responses can be used to integrate a diverse set of physiological signals to regulate a single transcriptional unit.

6.21.4 Structure of Riboswitch RNAs

Riboswitch structural arrangements were first predicted through phylogenetic analyses that suggested a particular arrangement of secondary structural elements (and in some cases, tertiary structural elements) that frame conserved primary sequence elements. Structural mapping was used to verify the predicted secondary structure. Recent crystal structures of the ligand-bound form of several metabolite-binding riboswitch RNAs have provided new insights into the three-dimensional structures of the ligand-binding domain of these RNAs and the basis for ligand recognition.⁹ Less information is available for the other classes of riboswitch RNAs, or for the ligand-free form.

6.21.4.1 Riboswitch Architectures

Riboswitch RNAs are found in two basic configurations. In the simpler form, a single RNA domain is responsible not only for sensing of the regulatory signal but also for mediating the effect on gene expression. The more complex riboswitches consist of two separate domains: (1) the ligand-binding domain (or aptamer), which carries out specific recognition of the ligand; and (2) the gene expression domain, which is responsible for the effects on the expression of regulated genes. Signal-dependent changes in the aptamer domain are transmitted to the gene expression domain via an RNA structural rearrangement.

6.21.4.1.1 Single-domain riboswitches

As noted above, the simplest riboswitches are the RNA thermosensors, which are helical elements that respond to changes in temperature. All known RNA thermosensors regulate at the level of translation initiation, and the helical element directly sequesters the ribosome-binding site of the regulated gene.⁴⁵ The gene expression domain is therefore an integral part of the riboswitch, and a single structural change (e.g., melting of the helix at higher temperature) is sufficient to promote induction of gene expression. Natural RNA thermosensors often consist of relatively complex helices that have evolved to respond to the appropriate temperature window, possibly through a stepwise melting of different portions of the helix; synthetic RNA thermosensors can be significantly simpler.⁶⁴

The tRNA-responsive T box riboswitches are more complex, but also integrate the gene expression domain within the ligand-binding domain. In these RNAs, a helical element near the 5' end of the riboswitch binds to the anticodon loop of the regulatory tRNA, while a second element near the 3' end of the riboswitch binds the acceptor end of the same tRNA molecule.⁵⁶ This secondary element serves as the antiterminator, as it includes sequences that would otherwise form the 5' side of the transcriptional terminator. As the terminator helix is predicted to be more stable than the competing antiterminator, base pairing of the acceptor end of the tRNA with residues within an internal bulge in the antiterminator stabilizes the antiterminator structure and prevents the formation of the terminator. In the T box mechanism, the 5' element of the riboswitch is primarily responsible for specific tRNA recognition, while the antiterminator element is responsible for discrimination between uncharged and charged species of the tRNA. Both interactions are necessary for modulation of downstream gene expression, and the antiterminator element is an intrinsic part of the ligand-binding domain.

The SAM-binding S_{MK} box riboswitch is unusual among the metabolite-binding riboswitches in that the SAM-binding domain includes the gene expression elements. In this RNA, binding of SAM results in pairing of the SD sequence with an ASD sequence; sequestration of the SD sequence prevents translation initiation. Genetic and biochemical studies demonstrated that the SD–ASD pairing is required for SAM binding;^{16,38} this differs from the S box SAM-binding riboswitch in which the SAM-binding domain is separable from the gene expression domain.^{14,17} The SD sequence in the S_{MK} box participates directly in the formation of the SAM-binding pocket, and the central position of the SD sequence (a G residue) makes specific contacts with SAM.¹⁹

6.21.4.1.2 Multidomain riboswitches

Most metabolite-binding riboswitches consist of separate ligand-binding and gene expression domains. In these RNAs, the gene expression domain is usually downstream of the ligand-binding domain and makes no contacts with the effector molecule. Ligand-dependent rearrangements of the effector-binding domain are responsible for changes in the overall architecture of the RNA, and it is these changes that affect the structure of the gene expression domain, for example, by releasing or sequestering an antiterminator or AASD–ASD helix that in turn modulates gene expression. The ligand-binding domain of the multidomain riboswitches can be isolated and shown to mediate specific recognition of the effector molecule, in some cases with a higher affinity than the intact riboswitch; this increase in affinity is likely to be due to the absence of competing RNA structural elements.

6.21.4.2 Structural Biology

The currently known riboswitch structures are shown in **Table 1**. Most efforts have focused on crystal structure analysis of the ligand-bound form of the aptamer domain of riboswitches that bind small molecules.⁹ The ligand-free form generally fails to form diffractable crystals, with the exception of the unusual *glmS* riboswitch/ribozyme in which the ligand does not affect the structure of the RNA but instead triggers the RNA cleavage event.^{65,66} Removal of the ligand by dialysis from riboswitch RNA crystals formed in the presence of ligand often results in maintenance of the ligand-bound structure, indicating that the structure is stabilized by the crystal lattice. However, in most cases, little information is available about the true ligand-free form, raising interesting questions about how the ligand is recognized and how the RNA transitions from the free to bound form, and whether binding is in fact reversible.

The T box RNA is a complex riboswitch that binds both the anticodon and acceptor ends of the cognate tRNA. The solution structure of the antiterminator domain, which is responsible for specific recognition of the acceptor end of uncharged tRNA, has been determined.⁶⁷ This domain consists of two helices separated by a 7 nt bulge that includes the residues that pair with the tRNA. The bulge region is somewhat flexible, with the key residues displayed for interaction with the tRNA. This represents a rare example of information about the ligand-free form of a riboswitch RNA, but in this case only a portion of the structure is known, and the molecular basis for stabilization of the antiterminator by tRNA binding remains unclear.

Unlike the metabolite-binding or T box riboswitches, RNA thermosensors do not bind a ligand and instead exhibit a temperature-dependent shift in structure. The ROSE (repression of heat-shock gene expression) element, responsible for induction of heat-shock gene expression in rhizobia, has been characterized by NMR.⁶⁸ This analysis allowed determination of the three-dimensional structure of the RNA in its inhibitory state (at low temperature) as well as analysis of the structural transitions that occur in the RNA as the temperature is increased.

6.21.5 Riboswitches and Gene Identification

Most riboswitches were initially identified through characterization of sets of genes with known functions and known or predictable regulatory patterns (e.g., genes for biosynthesis of a particular amino acid can be predicted to exhibit low expression when that amino acid is provided in the growth medium, and high expression when cells are starved for that amino acid). Recognition of conserved sequence elements and structural patterns (e.g., RNA helices) in the leader regions of genes in the same family, coupled with improved bioinformatics tools and an explosion of genomic sequence data, led to the identification of previously unidentified genes with the same RNA pattern, suggestive of use of the same regulatory mechanism.^{69–71} This then allows testing of specific models about possible roles for these new genes of unknown function in the same pathway as that in which the known genes with this mechanism are involved. This pattern of predictive logic followed by direct experimental testing has allowed identification of a number of new genes and pathways, and is also useful for identification of regulatory genes and transport systems that are otherwise difficult to couple to function based only on bioinformatics.

6.21.5.1 Genes Identified Based on Association with a Riboswitch

The S box system represents an early example of utility of riboswitches for gaining insight into the function of previously uncharacterized gene families. Identification of an S box riboswitch upstream of a series of genes of unknown function in *B. subtilis* led to the prediction that these genes were in some way involved in the metabolism of methionine or SAM.^{35,72} Testing of this prediction led to the discovery of a pathway for recycling of SAM (after its utilization in polyamine biosynthesis) not previously known to be present in this organism, and allowed the identification of the genes responsible for this pathway in other organisms. Similarly, the anti-TRAP regulatory protein (AT), which regulates tryptophan biosynthesis in *B. subtilis*, was identified because of the presence of a T box riboswitch predicted to respond to tryptophanyl-tRNA.^{73,74} The AT protein is not related to other known gene families, and its identity and role in the regulation of the *trp* operon was previously unknown.

Riboswitch identification also provides tools for recognition of misannotation in genome databases. This is especially evident in the case of genes encoding amino acid transporters, for which identification of the cognate substrate is often difficult. For example, a number of putative transporter genes annotated as members of the GlnP (glutamine) transporter family because of sequence similarity are likely to instead transport amino acids in the histidine family, as they are preceded by T box riboswitches predicted to respond to tRNA^{His}.⁵⁹ Recognition and correction of misannotation events of this type, which are often the result of 'transitive annotation' (where a gene is identified solely based on its similarity to another uncharacterized gene, which may also be misannotated), is crucial to the effectiveness of genome sequence databases.

Several putative riboswitches have been identified by bioinformatics analyses for which no previous genetic analyses were available to provide insight into likely function or regulatory factors.⁷⁵ These were initially designated as 'orphan' riboswitches, and several have now been characterized. It is highly likely that continuous improvement in bioinformatics tools, coupled with the growing availability of genome sequence data, will continue to provide new riboswitch candidates for further study.

6.21.6 Using Riboswitches to Control Gene Expression

The possibility of using known riboswitches to control the expression of heterologous genes is readily demonstrated by the use of reporter gene fusions to investigate riboswitch function. For example, fusions of a T box or S box leader sequence to a *lacZ* reporter provided initial information about the response of the riboswitch to different growth conditions, and the effect on reporter gene expression of mutations in sequence elements predicted to be important for the regulatory response.^{54,35} The ability to replace the coding sequence located downstream of the reporter with reporter gene constructs, and observe regulation of reporter gene expression consistent with that observed for the native gene, provides a proof-of-principle for the use of riboswitches as genetic regulatory elements for heterologous gene expression. However, there are a number of parameters that should be considered in designing systems of this type. These include the host organism and its ability to efficiently utilize a specific regulatory mechanism, the magnitude of genetic control required relative to the range of signal variability available within the cell, and the potential need for very tight repression under noninducing conditions.

6.21.6.1 Natural Riboswitches as Genetic Control Elements for Heterologous Gene Expression

The use of natural riboswitch RNAs to control heterologous gene expression has certain advantages. Use of these elements in systems from which they were derived (or closely related organisms) ensures that effective ligand concentrations for both ON and OFF states can be generated, and in general the appropriate growth conditions can be inferred from previous studies with the natural riboswitch. Furthermore, the design of these systems can employ genetic regulatory elements (such as transcriptional terminators and SD-ASD pairings) that are already known to function in the host organism. This is an important issue, as natural riboswitches tend

to operate differently in different groups of organisms. For example, most metabolite-binding riboswitches in low G + C Gram-positive bacteria regulate at the level of premature termination of transcription, whereas regulation at the level of translation initiation predominates in Gram-negative bacteria, although there are exceptions in both cases. Transfer of a riboswitch to a closely related organism often allows normal regulation (e.g., the *Enterococcus faecalis* S_{MK} box exhibits a similar regulatory pattern in *E. faecalis* and *B. subtilis*),¹⁶ while transfer to a more distant organism often fails.

6.21.6.2 Synthetic Riboswitches

Synthetic RNA aptamers that specifically recognize small molecules were uncovered through *in vitro* selection of randomized RNA pools.^{76,77} The recognition that RNA rearrangements can regulate gene expression, and that natural riboswitch RNAs couple specific ligand recognition to gene regulation, led to the recognition that combinations of synthetic aptamers with engineered gene expression modules that utilize RNA structural rearrangements could allow development of new tools for gene expression control. Early examples of this approach used synthetic RNA aptamers for compounds like tetracycline and theophylline to regulate the formation of structures that sequester the translation initiation region.^{78–80} These were developed both for bacterial and eukaryotic systems, in both cases exploiting the sensitivity of translation initiation to mRNA structure. Similar elements can be designed to control expression at the level of transcription attenuation, although most efforts have focused on translational control. This makes sense, as it is intrinsically simpler to modulate occlusion of a ribosome-binding site; regulation of transcription termination requires a more restrictive kinetic window for ligand binding and RNA rearrangement, since once RNAP has reached the termination site it is committed to either the termination or readthrough pathway and is insensitive to regulatory effects.

A key issue in regulatory mechanisms of this type is the dynamic range of the response to effector concentration. This is affected not only by the affinity of the ligand-binding domain for the ligand (which can be highly variable even for riboswitches in the same class⁸¹), but also by the relative stabilities of the ON and OFF states of the gene expression control segment of the RNA (e.g., the ASD–SD helix versus the AASD–ASD helix or terminator versus antiterminator). The system is also greatly affected by the ability to modulate the intracellular concentration of the effector molecule over a range that matches the sensitivity of the riboswitch.

Synthetic riboswitch systems have also been engineered that utilize RNA thermosensors to control translation initiation.⁶⁴ These are much simpler than natural RNA thermosensors, and have the advantage that calibration to the signal is readily adjustable by subtle changes in the melting temperature of the RNA helix that controls gene expression.

6.21.7 Perspectives and Future Directions

It has become increasingly apparent in the last few years that riboswitch RNAs play a major role in regulating a large variety of gene classes in bacteria. The majority of these genes are involved in the biosynthesis of key cellular metabolites (or cellular components for using these metabolites, like aminoacylated tRNAs). Understanding how these regulatory systems function provides tools for manipulating cell physiology, and production of important metabolites, in predictable ways. Furthermore, the regulatory elements themselves, and synthetic counterparts designed using our knowledge of the function of natural riboswitches, permit careful regulation of heterologous genes, to further increase our ability to manipulate cell physiology to produce products of our choice within bacterial cells. Certain riboswitch architectures are also functional in eukaryotic systems, or can be readily engineered to function in these systems. Riboswitch RNAs therefore provide a useful toolbox for regulating gene expression in a variety of systems. The rapid advances in this field in only a few years make it highly likely that many more RNA elements of this type remain to be uncovered, increasing both the number of genes known to be regulated by this mechanism and the variety of tools available for using riboswitch RNAs as a tool for genetic expression.

Glossary

5'-UTR 5'-untranslated region of an mRNA, located upstream of a coding sequence.

anti-Shine–Dalgarno Sequence complementary to the Shine–Dalgarno; pairing of the Shine–Dalgarno and anti-Shine–Dalgarno prevents translation initiation.

antiterminator RNA element that prevents the activity of an attenuator.

aptamer Ligand-binding domain of a riboswitch RNA.

attenuator Terminator located in the leader region of a transcriptional unit, used to regulate the expression of downstream genes.

leader region Segment of a gene between the transcription initiation site and the start of the coding sequence.

leader RNA Segment of an RNA that is upstream of the first coding sequence.

nascent RNA Newly synthesized RNA transcript, emerging from RNAP during transcription elongation.

promoter Signal for RNA polymerase to bind DNA and initiate transcription.

ribosome-binding site Translation initiation signal, consisting of Shine–Dalgarno plus start codon (usually AUG, GUG, UUG).

riboswitch Regulatory element that utilizes a segment of RNA to monitor a specific regulatory signal.

RNAP RNA polymerase.

Shine–Dalgarno sequence Binding site for the 30S ribosome on an mRNA for translation initiation.

terminator Signal for RNA polymerase to stop transcription and release the DNA template and the nascent transcript.

References

1. T. M. Henkin, *Genes Dev.* **2008**, *22*, 3383–3390.
2. F. J. Grundy; T. M. Henkin, *Crit. Rev. Biochem. Mol. Biol.* **2006**, *41*, 329–338.
3. P. Babitzke, *Curr. Opin. Microbiol.* **2004**, *7*, 132–139.
4. C. Yanofsky, *RNA* **2007**, *13*, 1141–1154.
5. W. C. Winkler; A. Nahvi; A. Roth; J. A. Collins; R. R. Breaker, *Nature* **2004**, *428*, 281–286.
6. J. A. Collins; I. Irnov; S. Baker; W. C. Winkler, *Genes Dev.* **2007**, *21*, 3356–3368.
7. T. Kubodera; M. Watanabe; K. Yoshiuchi; N. Yamashita; A. Nishimura; S. Nakai; K. Gomi; H. Hanamoto, *FEBS Lett.* **2003**, *555*, 516–520.
8. M. T. Cheah; A. Wachter; N. Sudarsan; R. R. Breaker, *Nature* **2007**, *447*, 497–500.
9. R. K. Montange; R. R. Batey, *Annu. Rev. Biophys.* **2008**, *37*, 117–133.
10. F. J. Grundy; S. C. Lehman; T. M. Henkin, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12057–12062.
11. N. Sudarsan; J. K. Wickiser; S. Nakamura; M. S. Evert; R. R. Breaker, *Genes Dev.* **2003**, *17*, 2688–2697.
12. A. D. Garst; A. Heroux; R. P. Rambo; R. T. Batey, *J. Biol. Chem.* **2008**, *283*, 22347–22351.
13. A. Serganov; L. Huang; D. J. Patel, *Nature* **2008**, *455*, 1263–1267.
14. M. Mandal; M. Lee; J. E. Barrick; Z. Weinberg; G. M. Emilsson; W. L. Ruzzo; R. R. Breaker, *Science* **2004**, *306*, 275–279.
15. B. A. M. McDaniel; F. J. Grundy; I. Artsimovitch; T. M. Henkin, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3083–3088.
16. K. A. Corbino; J. E. Barrick; J. Lim; R. Welz; B. J. Tucker; I. Puskarz; M. Mandal; N. D. Rudnick; R. R. Breaker, *Genome Biol.* **2005**, *6*, R70.
17. R. T. Fuchs; F. J. Grundy; T. M. Henkin, *Nat. Struct. Mol. Biol.* **2006**, *13*, 226–233.
18. R. K. Montange; R. T. Batey, *Nature* **2006**, *441*, 1172–1175.
19. S. D. Gilbert; R. P. Rambo; D. Van Tyne; R. T. Batey, *Nat. Struct. Mol. Biol.* **2008**, *15*, 761–772.
20. C. Lu; A. M. Smith; R. T. Fuchs; F. Ding; K. Rajashankar; T. M. Henkin; A. Ke, *Nat. Struct. Mol. Biol.* **2008**, *15*, 1076–1083.
21. J. X. Wang; E. R. Lee; D. R. Morales; J. Lim; R. R. Breaker, *Mol. Cell* **2008**, *29*, 691–702.
22. B. A. Murphy; F. J. Grundy; T. M. Henkin, *J. Bacteriol.* **2002**, *184*, 2314–2318.
23. R. T. Batey; S. D. Gilbert; R. K. Montange, *Nature* **2004**, *432*, 411–415.
24. A. Serganov; Y. R. Yuan; O. Pikofskaya; A. Polonskaia; L. Malinina; A. T. Phan; C. Hobartner; R. Micura; R. R. Breaker; D. J. Patel, *Chem. Biol.* **2004**, *11*, 1729–1741.
25. W. C. Winkler; S. Cohen-Chalamis; R. R. Breaker, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15908–15913.
26. A. S. Mironov; I. Gusarov; R. Rafikov; L. E. Lopez; K. Shatalin; R. A. Kreneva; D. A. Perumov; E. Nudler, *Cell* **2002**, *111*, 747–756.
27. A. Serganov; L. Huang; D. J. Patel, *Nature* **2009**, *458*, 233–237.
28. J. Miranda-Rios; M. Navarro; M. Soberon, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9736–9741.
29. W. C. Winkler; A. Nahvi; R. R. Breaker, *Nature* **2002**, *419*, 952–956.
30. T. E. Edwards; A. R. Ferre-D'Amare, *Structure* **2006**, *14*, 1459–1468.
31. A. Serganov; A. Polonskaia; A. T. Phan; R. R. Breaker; D. J. Patel, *Nature* **2006**, *441*, 1167–1171.
32. S. Thore; M. Leibungut; N. Ban, *Science* **2006**, *312*, 1208–1211.

33. X. Nou; R. J. Kadner, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7190–7195.
34. A. Nahvi; N. Sudarsan; M. S. Evert; X. Zou; K. L. Brown; R. R. Breaker, *Chem. Biol.* **2002**, *9*, 1043–1049.
35. F. J. Grundy; T. M. Henkin, *Mol. Microbiol.* **1998**, *30*, 737–749.
36. V. Epshtein; A. S. Mironov; E. Nudler, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5052–5056.
37. W. C. Winkler; A. Nahvi; N. Sudarsan; J. E. Barrick; R. R. Breaker, *Nat. Struct. Biol.* **2003**, *10*, 701–707.
38. R. T. Fuchs; F. J. Grundy; T. M. Henkin, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4876–4880.
39. M. Mandal; B. Boese; J. E. Barrick; W. C. Winkler; R. R. Breaker, *Cell* **2003**, *113*, 577–586.
40. M. Mandal; R. R. Breaker, *Nat. Struct. Mol. Biol.* **2004**, *11*, 29–35.
41. J. N. Kim; A. Roth; R. R. Breaker, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16092–16097.
42. N. Sudarsan; E. R. Lee; Z. Weinberg; R. H. Moy; J. N. Kim; K. H. Link; R. R. Breaker, *Science* **2008**, *321*, 411–413.
43. A. Roth; W. C. Winkler; E. E. Regulski; B. W. Lee; J. Lim; I. Jona; J. E. Barrick; A. Ritwik; J. N. Kim; R. Welz; D. Iwata-Reuyl; R. R. Breaker, *Nat. Struct. Mol. Biol.* **2007**, *14*, 308–317.
44. M. H. Cromie; Y. Shi; T. Latifi; E. A. Groisman, *Cell* **2006**, *125*, 71–84.
45. C. E. Dann, III; C. A. Wakeman; C. L. Sieling; S. C. Baker; I. Irvov; W. C. Winkler, *Cell* **2007**, *130*, 415–424.
46. F. Narberhaus; T. Waldminghaus; S. Chowdbury, *FEMS Microbiol. Rev.* **2006**, *30*, 3–16.
47. M. T. Morita; Y. Tanaka; T. T. Kodama; Y. Kyogoku; H. Yanagi; T. Yura, *Genes Dev.* **1999**, *13*, 655–665.
48. J. Johansson; P. Mandin; A. Renzoni; C. Chiaruttini; M. Springer; P. Cossart, *Cell* **2002**, *110*, 551–561.
49. S. Gottesman; C. A. McCullen; M. Guillier; C. K. Vanderpool; N. Majdalani; J. Benhammou; K. M. Thompson; P. C. FitzGerald; N. A. Sowa; D. J. FitzGerald, *Cold Spring Harb. Symp. Quant. Biol.* **2006**, *71*, 1–11.
50. G. Storz; J. A. Opdyke; K. M. Wassarman, *Cold Spring Harb. Symp. Quant. Biol.* **2006**, *71*, 269–273.
51. P. Babitzke; T. Romeo, *Curr. Opin. Microbiol.* **2007**, *10*, 156–163.
52. K. M. Wassarman, *Mol. Microbiol.* **2007**, *65*, 1425–1431.
53. N. Majdalani; C. Cuning; D. Sledjeski; T. Elliott; S. Gottesman, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12462–12467.
54. F. T. Grundy; T. M. Henkin, *Cell* **1993**, *74*, 475–482.
55. F. J. Grundy; W. C. Winkler; T. M. Henkin, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11121–11126.
56. T. M. Henkin; F. J. Grundy, *Cold Spring Harb. Symp. Quant. Biol.* **2006**, *71*, 231–237.
57. M. R. Yousef; F. J. Grundy; T. M. Henkin, *J. Mol. Biol.* **2005**, *349*, 273–287.
58. A. G. Vitreschak; A. A. Mironov; V. A. Lyubetsky; M. S. Gelfand, *RNA* **2008**, *14*, 717–735.
59. A. Gutierrez-Preciado; T. M. Henkin; C. Yanofsky; E. Merino, *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 36–61.
60. N. Gendron; H. Putzer; M. Grunberg-Manago, *J. Bacteriol.* **1994**, *176*, 486–494.
61. R. Welz; R. R. Breaker, *RNA* **2007**, *13*, 573–582.
62. N. Sudarsan; M. C. Hammond; K. F. Block; R. Welz; J. E. Barrick; A. Roth; R. R. Breaker, *Science* **2006**, *314*, 300–304.
63. G. Andre; S. Even; H. Putzer; P. Burquiere; C. Croux; A. Danchin; I. Martin-Verstraete; O. Soutourina, *Nucleic Acids Res.* **2008**, *36*, 5955–5969.
64. J. Neupert; D. Karcher; R. Bock, *Nucleic Acids Res.* **2008**, *36*, e124.
65. D. J. Klein; A. R. Ferre-D'Amare, *Science* **2006**, *313*, 1752–1756.
66. J. C. Cochrane; S. V. Lipchock; S. A. Strobel, *Chem. Biol.* **2007**, *14*, 97–105.
67. M. S. Gerdeman; T. M. Henkin; J. V. Hines, *J. Mol. Biol.* **2003**, *326*, 189–201.
68. S. Chowdhury; C. Maris; F. H. Allain; F. Narberhaus, *EMBO J.* **2006**, *25*, 2487–2497.
69. C. Abreu-Goodger; E. Merino, *Nucleic Acids Res.* **2005**, *33*, W690–W692.
70. P. Bengert; T. Dandekar, *Nucleic Acids Res.* **2004**, *32*, W154–W159.
71. Z. Weinberg; J. E. Barrick; Z. Yao; A. Roth; J. N. Kim; J. Gore; J. X. Wang; E. R. Lee; K. F. Block; N. Sudarsan; S. Neph; M. Tompa; W. L. Ruzzo; R. R. Breaker, *Nucleic Acids Res.* **2007**, *35*, 4809–4819.
72. F. J. Grundy; T. M. Henkin, Synthesis of Serine, Glycine, Cysteine and Methionine. In *In Bacillus subtilis and Its Closest Relatives: From Genes to Cells*; A. L. Sonenshein, J. A. Hoch, R. Losick, Eds.; American Society for Microbiology Press: Washington, DC, 2002; pp 245–254.
73. J. P. Sarsero; E. Merino; C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2656–2661.
74. A. Valbuzzi; C. Yanofsky, *Science* **2001**, *293*, 2057–2059.
75. J. E. Barrick; K. A. Corbino; W. C. Winkler; A. Nahvi; M. Mandal; J. Collins; M. Lee; A. Roth; N. Sudarsan; I. Jona; J. K. Wickiser; R. R. Breaker, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6421–6426.
76. L. Gold; C. Tuerck, *Science* **1990**, *249*, 505–510.
77. A. D. Ellington; J. W. Szostak, *Nature* **1990**, *346*, 818–822.
78. B. Suess; S. Hanson; C. Berens; B. Fink; R. Schroeder; W. Hillen, *Nucleic Acids Res.* **2003**, *31*, 1853–1858.
79. S. Hanson; G. Bauer; B. Fink; B. Suess, *RNA* **2005**, *11*, 503–511.
80. A. Ogawa; M. Maeda, *ChemBioChem* **2008**, *9*, 206–209.
81. J. Tomsic; B. A. McDaniel; F. J. Grundy; T. M. Henkin, *J. Bacteriol.* **2008**, *190*, 823–833.

Biographical Sketch



Tina M. Henkin, is a professor and chair of Microbiology at the Ohio State University, distinguished professor at the College of Biological Sciences, and member, Center for RNA Biology. She obtained her BA in biology from Swarthmore College, Ph.D. in genetics from the University of Wisconsin, and postdoc in molecular biology and microbiology, Tufts University Medical School. She holds the following positions: Member, American Academy of Arts and Sciences Chair, Division H, American Society for Microbiology; Fellow, American Academy of Microbiology; Fellow, American Association for the Advancement of Science; Ohio State University Distinguished Scholar Award; and National Academy of Sciences Pfizer Award in Molecular Biology. Dr. Henkin's research focuses on the analysis of genetic regulatory mechanisms in Gram-positive bacteria. Her laboratory pioneered the use of comparative sequence analysis to identify families of genes with conserved regulatory motifs that act at the RNA level to modulate gene expression in response to specific effector molecules. This work revealed that a variety of molecular signals, including tRNA and metabolites, can be recognized by nascent RNA transcripts, and that binding of the effector to the RNA causes an RNA structural rearrangement that can affect transcription termination or translation initiation. These studies provided the foundation for the identification of a growing number of 'riboswitch' systems, which have recently been shown to be widely used to control many classes of genes in many systems.