Chapter 3

Cellular Organization of Glycosylation

Essentials of Glycobiology 2nd edition
Chapter 3 Cellular Organization of Glycosylation

Ajit Varki, Jeffrey D Esko, and Karen J Colley.

This chapter provides an overview of glycosylation from the perspective of a single cell, taking into account the patterns of expression, topology, and other features of the biosynthetic and degradative enzymes that are common to most cell types. The focus is mainly on eukaryotic cells, for which more information is available.

GLYCOSYLATION IS UNIVERSAL IN LIVING ORGANISMS

It is a remarkable fact that every free-living cell and every cell type within multicellular organisms is covered with a dense and complex layer of glycans. Even enveloped viruses that bud from surfaces of infected cells carry with them the glycosylation patterns of the host cell. Additionally, most secreted molecules are glycosylated and the extracellular matrices of multicellular organisms are rich in glycans and glycoconjugates. The matrices secreted by unicellular organisms when they congregate (e.g., bacterial biofilms; see Chapter 20) also contain glycans. The reason for the apparent universality of cell-surface and secreted glycosylation is not clear, but it suggests that evolution has repeatedly selected for glycans as being the most diverse and flexible molecules to position at the interface between cells and the extracellular milieu. For example, the enormous diversity, complexity, and flexibility of glycans may allow host cells to make changes to avoid pathogens, without causing major deleterious effects on cellular functions.

Like membrane proteins, secretory proteins in eukaryotic cells typically pass through an endoplasmic reticulum (ER)-Golgi pathway, the cellular system in which many major glycosylation reactions occur (see below). Perhaps for this reason, most proteins in the blood plasma of animals (with the exception of albumin) are also heavily glycosylated. Glycosylation of secreted proteins may provide solubility, hydrophilicity, and negative charge, thus reducing unwanted nonspecific intermolecular interactions in extracellular spaces and also protecting against proteolysis. Another, not mutually exclusive, hypothesis is that the glycans on secreted molecules act as decoys, binding pathogens that seek to recognize cell-surface glycans to initiate invasion.

In bacteria, archaea, and fungi, glycans have critical structural roles in forming the cell wall and in resisting large differences in osmolarity between the cytoplasm and the environment. Glycans surrounding bacteria could also have a role in defense against bacteriophages or antibiotics generated by other microorganisms in the environment.

TOPOLOGICAL ISSUES RELEVANT TO GLYCAN BIOSYNTHESIS

The ER-Golgi Pathway of Eukaryotes

Studies stemming from the classic work of George Palade and colleagues have indicated that most cell-surface and secreted molecules in eukaryotic cells originate in the ER. They then make their way via an intermediate compartment through multiple stacks of the Golgi apparatus, finally being distributed to various destinations from the trans-Golgi network. Along the way, lipids and proteins are modified by a variety of glycosylation reactions mediated by glycosyltransferases (see Chapter 5). Figure 3.1 superficially depicts some steps in the synthesis of the major glycan classes in the ER-Golgi pathway of animal cells. These pathways are discussed in the following sections and in other chapters of this book. As mentioned earlier, the ER-Golgi pathway is a universal feature of eukaryotic cells and also harbors other glycan-modifying enzymes (see below).

Not all glycans and glycoconjugates assemble within the ER-Golgi pathway. For example, many cytoplasmic and nuclear proteins contain O-GlcNAc, O-Glc, or O-Fuc, and these modifications occur in the cytoplasm (see Chapters 17 and 18). Hyaluronan and chitin assembly occurs at the plasma membrane, with direct extrusion into the extracellular matrix (see Chapter 15). In plant cells, cellulose synthesis also occurs at the plasma membrane (see Chapter 22).
Regardless of their location, most glycosylation reactions use activated forms of monosaccharides (most often nucleotide sugars) as donors for reactions that are catalyzed by enzymes called glycosyltransferases (see Chapter 4 for a listing of these enzymes and details about their biochemistry). A variety of glycan modifications are also found in nature (see Chapter 5). Of these, the most common are generated by sulfotransferases, acetyltransferases, and methyltransferases, which use activated forms of sulfate (3′phosphoadenyl-5′-phosphosulfate; PAPS), acetate (acetyl-CoA), and methyl groups (S-adenosylmethionine; AdoMet), respectively. Almost all the donors for glycosylation reactions and glycan modifications are synthesized within the cytoplasmic compartment from precursors of endogenous origin. In eukaryotes, most of these donors must be actively transported across a membrane bilayer in order to become available for reactions within the lumen of the ER-Golgi pathway.

Much effort has gone into understanding the mechanisms of glycosylation and glycan modification within the ER and the Golgi apparatus, and it is clear that a variety of interacting and competing factors determine the final outcome of the reactions. The glycosyltransferases, processing glycosidases, and sulfotransferases are well studied (see Chapter 5), and their location has helped to define various functional compartments of the ER-Golgi pathway. A popular model envisioned these enzymes as being physically lined up along this pathway in the precise sequence in which they actually work. In fact, there is considerable overlap of the enzymes across Golgi stacks and the actual distribution of a given enzyme depends on the cell type.

Some glycan chains are made on the cytoplasmic face of intracellular membranes and flipped across to the other side, but most are added to the growing chain on the inside of the ER or the Golgi (see Figure 3.1). Regardless, the portion of a molecule that faces the inside of the lumen of the ER or Golgi will ultimately face the outside of the cell or the inside of a secretory granule or lysosome. To date there are no well-documented exceptions to this topological rule. A consequence of this topological asymmetry is that many classes of glycans are optimized to be involved in cell–cell and cell–matrix interactions. Of course, these topological considerations are reversed for nuclear and cytoplasmic glycosylation (see below), because the active sites of the relevant glycosyltransferases for these reactions face the cytoplasm. Perhaps not surprisingly, the types of glycans found on the two sides of the cell membrane are generally quite distinct from each other.

**Glycosylation Pathways in Eubacteria and Archaea**

Much less is known about the topology of glycoconjugate assembly in eubacteria and Archaea (formerly grouped as prokaryotes). Bacterial cells perform most glycosylation reactions on the inner aspect of the cytoplasmic membrane, using precursors assembled in the cytoplasm (see Chapter 20). These glycan intermediates are then flipped across the cytoplasmic membrane and used to form polymeric (often large) structures in the periplasm (see Chapter 21). In gram-negative organisms, which have an outer membrane, some of the glycans or glycoconjugates must be transferred between membranes or flipped across the outer membrane. The mechanisms underlying these processes are active areas of research. Even less is known about the topology of glycosylation pathways in Archaea.

**GOLGI ENZYMES SHARE SECONDARY STRUCTURE**

Despite the lack of sequence homology among different families of glycosyltransferases and sulfotransferases, almost all Golgi enzymes share some features. Early studies of the cell biology and biochemistry of vertebrate glycosyltransferases indicated that some of these activities could be found in soluble form in secretions and body fluids; others were identified as membrane-bound activities within cells, and some exhibited both properties. Cell fractionation studies generally found cell-associated transferase activities in membrane-rich microsomal fractions, which could be liberated in soluble form with the aid of detergents. These observations implied that some transferases probably represent membrane-spanning proteins, whereas others correspond to secreted proteins. However, following the initial molecular cloning efforts that defined the sequences of a β1-4 galactosyltransferase, an α2-6 sialyltransferase, and the blood group A α1-3 N-acetylgalactosaminyltransferase, it became clear that Golgi glycosyltransferases share a common secondary structure that could account for all previous findings.

Almost all Golgi glycosyltransferases and sulfotransferases described to date have a single transmembrane domain flanked by a short amino-terminal domain and a longer carboxy-terminal domain. This structure is characteristic of
so-called type II transmembrane proteins, whose single amino-terminal membrane-spanning domain functions as a signal-anchor sequence, placing the short amino-terminal segment within the cytoplasm while directing the larger carboxy-terminal domain to the other side of the biological membrane into which the signal anchor has been inserted (Figure 3.2). For plasma-membrane-associated type II proteins, the “other side” is the extracellular surface. For glycosyltransferases, the “other side” is the lumen of the membrane-delimited compartments that constitute the ER-Golgi pathway. These include vesicles that transit from the ER to the cis cisterna of the Golgi, the cisternae of the Golgi apparatus itself, the vesiculotubular network of the trans-Golgi network, and membrane-delimited structures distal to the trans-Golgi network. This arrangement predicts that the larger carboxy-terminal domain contains the catalytic activity of the transferase, and this supposition has extensive experimental support. The intraluminal location of this domain allows it to participate in the synthesis of the growing glycans displayed by glycoproteins and glycolipids during their transit through the secretory pathway.

The type II transmembrane topology predicted by initial sequences of vertebrate glycosyltransferases has been widely confirmed experimentally. The topology may also explain reports of the expression of glycosyltransferases at the surface of mammalian cells. Interestingly, several other types of Golgi enzymes (e.g., processing glycosidases) that have been cloned share a similar topological arrangement. There are a few clear exceptions, such the UDP-GlcNAc:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) and the GlcNAc-1-phosphodiester α-N-acetylglucosaminidase that are both involved in the synthesis of the Man-6-P targeting signal of newly synthesized lysosomal hydrolases (see Chapter 30). The former is a multisubunit complex, and the latter is a type I membrane-spanning glycoprotein with its amino terminus in the lumen of the Golgi apparatus. One of the sulfotransferases involved in heparan sulfate synthesis (GlcNAc 3-O-sulfotransferase 1) is a resident soluble enzyme in the Golgi. Likewise, the protein O-fucosyltransferase (Notch FucT) appears to be a soluble enzyme in the ER.

All of the above considerations do not apply to the glycosyltransferases involved in nuclear and cytoplasmic glycan synthesis. For example, the soluble GlcNAc transferase responsible for synthesizing the O-linked GlcNAc of nuclear and cytoplasmic proteins (see Chapter 18) has no detectable homology with the Golgi GlcNAc transferases. Another variation is presented by the hyaluronan and cellulose synthases, which are multipass membrane proteins present in the plasma membrane, extruding their products directly into the extracellular space. Similarly, all of the enzymes that use dolichol-linked precursors (and, in bacteria, bactoprenol-linked precursors) have more complex multi-membrane-spanning structures and contain a motif thought to bind to the isoprenoid chain.

**GLYCOSYLTRANSFERASES CAN ALSO BE GLYCOSYLATED**

Many Golgi glycosyltransferases have consensus N-glycosylation sequences as well as serine and threonine residues that could be modified by glycosylation processes. Biochemical analyses indicate that many mammalian glycosyltransferases are indeed posttranslationally modified by glycosylation, especially N-glycosylation. Glycosylation is, in some instances, required for proper folding and/or activity, and a few studies indicate that glycosyltransferases are subject to “autoglycosylation.” There is also limited evidence that glycosyltransferases may be modified by phosphorylation. The functional relevance of such posttranslational modifications remains unknown.

**LOCALIZATION OF GLYCOSYLTRANSFERASES IN GOLGI COMPARTMENTS**

Biochemical and ultrastructural studies indicate that glycosyltransferases partially segregate into distinct compartments within the secretory pathway. Generally speaking, enzymes acting early in glycan biosynthetic pathways have been localized to cis and medial compartments of the Golgi, whereas enzymes acting later in the biosynthetic pathway tend to colocalize in the trans-Golgi cisternae and the trans-Golgi network. These observations have prompted extensive exploration of the mechanisms whereby glycosyltransferases achieve this compartmental segregation. An effort was made to find Golgi-retention sequences, by analogy with the KDEL tetrapeptide implicated in retention/retrieval of ER-associated proteins. Although some general conclusions arise from these studies, the reader should consider the following caveats:

- Observations made with one enzyme are not necessarily applicable to others.
• The Golgi-retention properties of any given glycosyltransferase may vary depending on the cell type in which localization is examined.

• Variations in the expression level of a glycosyltransferase in an experimental system can have a major influence on retention/localization properties.

• Many studies used chimeric proteins composed of segments of a glycosyltransferase fused to a reporter protein, but conclusions from such experiments have not always been verified using intact glycosyltransferases or chimeras with a different reporter protein.

• In vitro studies using intact Golgi compartments indicate some spatial and functional overlap among enzymes that were previously thought to be segregated on the basis of data from other less sensitive techniques.

Most information relevant to the retention of glycosyltransferases within specific Golgi compartments derives from experiments done with an α2-6 sialyltransferase (ST6Gal-I), a β 1-4 galactosyltransferase (GalT-I), and an N-acetylglucosaminyltransferase I (GlcNAcT-I). The former pair of enzymes tends to concentrate in the trans-Golgi compartments and the trans-Golgi network, whereas GlcNAcT-I localizes mostly to the medial-Golgi compartment. With respect to ST6Gal-I, multiple signals and mechanisms may be involved in its Golgi localization. The transmembrane domain and flanking sequences are sufficient to direct heterologous proteins to the Golgi, and it appears that it is the length of the transmembrane segment (provided it is hydrophobic in nature) and not the precise sequence of this domain that is critical. Replacement of this region with a longer unrelated hydrophobic sequence does not compromise the Golgi localization of the intact enzyme, suggesting that other sequences are involved in localization. Recent evidence suggests that the cytoplasmic tail of this protein may mediate an additional localization mechanism and that the enzyme’s luminal sequences are involved in a secondary oligomerization event that stabilizes Golgi retention.

In contrast, an examination of the Golgi-retention determinants of GalT-I points mainly to an important role for the transmembrane domain. The sequences that flank the membrane-spanning domain seem less important in Golgi retention for this enzyme. Retention of GlcNAcT-I is also dictated largely by its transmembrane domain, although its luminal sequences are also involved in an oligomerization mechanism that has a role in its localization. Considered together, the available experimental observations suggest that the localization of glycosyltransferases within specific regions of the Golgi apparatus is probably not determined by simple primary sequence motifs. Rather, this process is likely determined by several different regions of each enzyme that mediate multiple redundant localization mechanisms.

Golgi localization may also be mediated by retention (in the context of the vesicular transport model) and continuous retrieval to earlier Golgi cisternae (in the context of the cisternal maturation model). Three models have been proposed to account for the localization of glycosyltransferases to specific Golgi subcompartments. In the oligomerization/kin-recognition model, glycosylation enzymes form homooligomers or heterooligomers through interactions between their transmembrane and luminal sequences after arriving at the proper Golgi compartment. Heterooligomerization or kin recognition among enzymes in the same pathway would also presumably enhance the efficiency of the sequential glycosylation reactions. Some glycosylation enzymes have been found to form homooligomers (e.g., ST6Gal-I, GlcNAcT-1 and GlcNAcT-2), and oligomerization appears to have a role in their stable localization in the correct Golgi compartment. Experimental support for kin recognition comes from the observation that some pairs of glycosyltransferases known to catalyze sequential reactions in the same pathway colocalize to a specific Golgi compartment and are coimmunoprecipitated from cell extracts. However, this type of association has not been demonstrated for most of the enzymes.

A second model depends on partitioning the glycosyltransferases into lipid bilayers of different thicknesses. This bilayer thickness or lipid-partitioning model was proposed on the basis of observations that a cholesterol concentration gradient in the secretory pathway yields lipid bilayers of increasing thickness in the direction of cis to trans across the Golgi stack and that the transmembrane regions of glycosyltransferases are generally shorter than those of plasma membrane proteins. The model predicts that each glycosyltransferase sorts itself into the proper Golgi
location by virtue of the length of its transmembrane segment, which will retain the enzyme once it reaches the proper compartment during the enzyme’s transit through the secretory pathway. This model was formulated largely on the basis of experiments involving ST6Gal-I, where the length of the membrane-spanning domain appears to play an important part in Golgi retention. However, the general applicability of this model is not apparent because there is no consistent relationship between the length of the transmembrane segment and retention in a specific Golgi compartment for a variety of glycosyltransferases. This model may therefore help to explain the overall phenomenon of retention of glycosyltransferases in the Golgi apparatus versus the delivery of other membrane proteins to the plasma membrane.

The third mechanism is based largely on the cisternal maturation model of transport through the secretory pathway. In this model, a new Golgi cisterna containing cargo molecules forms at the cis face of the stack and it progressively “matures” as Golgi glycosylation enzymes that define each subcompartment are transported into the new cisterna from the old cisterna to modify the cisternal cargo proteins. In this model, the steady-state localization of the Golgi enzymes is maintained by continuous retrograde transport. The cytoplasmic tails of enzymes may mediate interactions with coat proteins that select the enzymes for transport in vesicles or tubules. In support of this mechanism, the cytoplasmic tails of some glycosylation enzymes, such as ST6Gal-I and FucT-I, have been shown to have a role in their Golgi localization.

Taken together, evidence suggests that glycosylation enzymes use multiple mechanisms to maintain their localization in the Golgi. The number of signals and mechanisms used by an enzyme could determine how stable its Golgi localization actually is, whether it is able to move to a later compartment, and whether it can be cleaved and secreted into the extracellular space (see below).

**PROTEOLYTIC CLEAVAGE AND SECRETION OF GOLGI GLYCOSYLTRANSFERASES**

Many Golgi enzymes are secreted by cells, sometimes in large quantities, and can be found in cell culture supernatants and various body fluids. The nature of the secreted forms of glycosyltransferases first became clear from analysis of amino-terminal peptide sequences of purified mammalian glycosyltransferases that had been isolated as soluble forms without the aid of detergents. These studies showed that the soluble forms were actually derived from their membrane-associated forms by virtue of one or more proteolytic cleavage events that occurred a short distance away from the transmembrane segment within the stem region (Figure 3.2). These proteolytic cleavage events release a catalytically active fragment of the glycosyltransferase from its transmembrane tether and allow the cell to export this fragment to the extracellular space. The existence of catalytically active fragments of glycosyltransferases that are also deficient in various portions of the stem region together with mutational analyses imply that the stem region contributes little to the catalytic function of a glycosyltransferase. Nevertheless, some experimental analyses suggest that peptide sequences within the stem region can contribute to acceptor substrate preference.

The signals within the glycosyltransferase sequence that direct proteolysis are not defined, but it appears that the proteolytic cleavages are relatively specific and are generated by proteases functioning in the trans regions of the Golgi apparatus and beyond. The production of these soluble enzymes from cell types such as hepatocytes and endothelium can also be dramatically up-regulated under certain inflammatory conditions. Because these circulating enzymes do not have access to adequate concentrations of donor nucleotide sugars (primarily located inside cells), they should be functionally incapable of performing a transfer reaction in the extracellular spaces. The biological significance of these soluble transferases therefore remains a mystery. Possibilities to consider include a lectin-like activity recognizing their acceptor substrates and/or a role in scavenging small amounts of circulating sugar nucleotides that might otherwise be available to certain microbes, such as gonococci.

**TURNOVER AND RECYCLING OF GLYCANS**

Like all components of living cells, glycans turn over constantly. Some glycoconjugates, such as transmembrane heparan sulfate proteoglycans, turn over by shedding from the cell surface through limited proteolysis. Most glycoconjugate turnover occurs by endocytosis and subsequent degradation in lysosomes (see Chapter 41). Endoglycosidases can initially cleave glycans internally, producing substrates for exoglycosidases in the lysosome.
Once broken down, individual monosaccharides are then typically exported from the lysosome into the cytoplasm, so that they can be reused (see Figure 1.8, Chapter 1). In contrast to the relatively slow turnover of glycans derived from the ER-Golgi pathway, glycans of the nucleus and cytoplasm may be more dynamic and rapidly turned over (see Chapters 17 and 18). Glycans in bacterial cells (especially those in the cell wall) also turn over during cell division when the cell wall undergoes cleavage and remodeling.

NUCLEAR AND CYTOPLASMIC GLYCOSYLATION IS VERY COMMON

Until the mid-1980s, a commonly stated dogma was that glycoconjugates such as glycoprotein and glycolipids occur exclusively on the outer surface of cells, on the internal (lumenal) surface of intracellular organelles, and on secreted molecules. As discussed above, this was in accord with information about the topology of the biosynthesis of the classes of glycans known at the time, which took place within the lumen of the Golgi-ER pathway. Thus, despite some clues to the contrary, the cytoplasm and nucleus (which are topologically semicontinuous because of the existence of nuclear pores) were assumed to be devoid of glycosylation capacity. However, it has now become clear that certain types of glycoconjugates are synthesized and reside within the cytoplasm and nucleus. Indeed, one of them (O-linked GlcNAc; see Chapter 18) may well be numerically the most common type of glycoconjugate in many cells.

GLYCOSYLATION REACTIONS AT THE PLASMA MEMBRANE

Because prokaryotic cells do not have an ER-Golgi pathway, they typically generate their cell-surface glycans at the interface of the cytoplasmic membrane and the cytoplasm or in the periplasm (see Chapter 20). Other glycans assembled at the plasma membrane include hyaluronan in vertebrate cells, chitin in invertebrate cells, and cellulose in plant cells. The topological difficulties of using cytoplasmic hydrophilic sugar nucleotides to manufacture glycans that are found on the opposite side of the cell-surface membrane are obvious. In eukaryotes, the process appears to involve enzyme molecules with multiple passes through the membrane that also act as a pore. In bacteria, membrane flippases may exist to aid in the transfer of intermediates across the various membranes. On the other hand, typical Golgi enzymes have also been claimed to be present at the cell surface in some animal cell types, with their active sites facing the extracellular space. It is not known how they could function at this location or how the nucleotide sugar donors would be delivered to this location. On the other hand, there are examples of remodeling of cell-surface glycans in animal cells, e.g., the sulf enzymes that modify heparan sulfate glycosaminoglycan (see Chapter 16) and sialidases that may remove cell-surface sialic acids (see Chapter 14).

GLYCOSYLATION IN UNEXPECTED SUBCELLULAR LOCATIONS

There are scattered reports of glycosylation in unexpected subcellular locations, for example, gangliosides in mitochondria and N-glycans in the nucleus. Many of these claims are based on incomplete evidence (see Chapter 17 for some discussion of these issues). One possibility is that there are indeed glycans in these unexpected cellular locations, but that their true structures are actually novel. Conversely, there are instances where the structural evidence is strong, but there is inadequate evidence to be certain about the topology of the claimed structures. Regardless, past experience tells us that the cell biology of glycosylation can hold many surprises, and dogmatic positions about such controversial issues are not warranted.

FURTHER READING


Figures

**FIGURE 3.1**

Initiation and maturation of the major types of eukaryotic glycoconjugates in relation to sub-cellular trafficking in the ER-Golgi-plasma membrane pathway. This illustration outlines the different mechanisms and topology for initiation, trimming, and elongation of the major glycan classes in animal cells. *Asterisks* represent the addition of outer sugars to glycans in the Golgi apparatus. N-glycans and glycosylphosphatidylinositol (GPI) anchors are initiated by the en-bloc transfer of a large preformed precursor glycan to a newly synthesized glycoprotein. O-glycans and sulfated glycosaminoglycans are initiated by the addition of a single monosaccharide, followed by extension. The most common glycosphingolipids are initiated by the addition of glucose to ceramide on the outer face of the ER-Golgi compartments, and the glycan is then flipped into the lumen to be extended. For a better understanding of the events depicted in this figure, see details in other chapters of this book: N-glycans (Chapter 8); O-glycans (Chapter 9); glycosphingolipids (Chapter 10); GPI anchors (Chapter 11); and sulfated glycosaminoglycans (Chapter 16).
FIGURE 3.2

Typical transmembrane topology and proteolytic processing of Golgi glycosyltransferases. Golgi glycosyltransferases and sulfotransferases generally have a single hydrophobic segment (TM) that functions as a signal-anchor sequence. This segment spans the lipid bilayer of the tubular and vesicular structures of the secretory pathway, including the membrane of the Golgi apparatus. This topology places the catalytic domain of a glycosyltransferase within the lumen of the Golgi apparatus and other membrane-delimited structures of the secretory pathway. The membrane-tethered form of a glycosyltransferase is susceptible to one or more proteolytic cleavage events that transect the enzyme within its “stem” region. Proteolysis can liberate a catalytically active, soluble form of the enzyme that may be released from the cell. With few exceptions, vertebrate glycosyltransferases have one or more potential asparagine-linked glycosylation sites (forked symbols). Where examined experimentally, one or more of these sites are used, indicating that most glycosyltransferases are glycoproteins.
Chapter 6

Glycosyltransferases and Glycan-processing Enzymes

Essentials of Glycobiology 3rd edition
Chapter 6 Glycosyltransferases and Glycan-processing

Enzymes

Authors: James M. Rini and Jeffrey D. Esko

<table>
<thead>
<tr>
<th>Old #</th>
<th>New #</th>
<th>Text File Name</th>
<th>Original (Words)*</th>
<th>Current (Words)*</th>
<th>% Cut</th>
<th>Original Figures</th>
<th>Current Figures</th>
<th>Original Tables</th>
<th>Current Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>3edch6PF</td>
<td>2941</td>
<td>3400</td>
<td>-</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Primary Word file name style:
3edch6PF.docx

Figure Files
- 3edch6f1NC.jpg  Fig 3 in 2nd edition
- 3edch6f2NC.jpg  Fig 4 in 2nd edition
- 3edch6f3NC.jpg  Fig 5 in 2nd edition
- 3edch6f4MOD.jpg  Modified
- 3edch6f5NC.jpg  Fig 7 in 2nd edition
- 3edch6f6NC.jpg  Fig 8 in 2nd edition
Glycosyltransferases and glycosidases are responsible for the assembly, processing, and turnover of glycans. In addition, there are a number of transferases that modify the chains by the addition of acetyl, methyl, phosphate, sulfate and other groups. This chapter covers the general characteristics of enzymes involved in glycan initiation, assembly and processing, including aspects of substrate specificity, primary sequence relationships, structures, and enzyme mechanisms.

GENERAL PROPERTIES

The biosynthesis of glycans is primarily determined by glycosyltransferases that assemble monosaccharide moieties into linear and branched glycan chains. As might be expected from the complex array of glycan structures found in nature, glycosyltransferases constitute a very large family of enzymes. In many cases they catalyze a group-transfer reaction in which the monosaccharide moiety of a simple nucleotide sugar donor substrate (e.g., UDP-Gal, GDP-Fuc, or CMP-Sia; Chapter 5) is transferred to the acceptor substrate. In some instances, the donor substrates contain a lipid moiety, such as dolichol-phosphate, linked to mannose or glucose. For other glycosyltransferases, the donor substrate is dolichol-pyrophosphate linked to an oligosaccharide, and in these cases the entire oligosaccharide is transferred to the acceptor substrate (Chapter 9). Other lipid-linked sugars serve as donor substrates for bacterial glycosyltransferases involved in the assembly of peptidoglycan (e.g., undecaprenyl-pyrophosphoryl-N-acetylmuramic acid-pentapeptide-N-acetylglucosamine), lipopolysaccharide, and capsules (Chapters 21 and 22).

Glycosyltransferases that use monosaccharides, oligosaccharides, proteins, lipids, small organic molecules and DNA as acceptor substrates have been characterized, but only glycosyltransferases involved in the biosynthesis of glycoproteins, proteoglycans, and glycolipids are discussed in this chapter. Among these enzymes, the vast majority are responsible for elongating the glycan moiety of their acceptor substrates; the remainder catalyze the initial transfer of sugar to a polypeptide or lipid acceptor. Generally speaking, the enzymes that elongate glycan chains act sequentially so that the product of one enzyme yields a preferred acceptor substrate for the subsequent action of another. The end result is a linear and/or branched structure composed of monosaccharides linked to one another. Generally, acceptor recognition does not involve the polypeptide or glycolipid moiety of the acceptor substrate if present, although several notable exceptions exist as described below.

Glycosidases that remove monosaccharides to form intermediates that are acted on by glycosyltransferases are also involved in the biosynthesis of glycans. Other glycosidases are involved in degradation of glycans, e.g. in lysosomes (Chapter 44). In addition, glycans can be modified by many other enzyme types including
sulfotransferases, phosphotransferases, O-acetyl-transferases, O-
methyltransferases, pyruvyltransferases, and phosphoethanolamine transferases.

**GLYCOSYLTRANSFERASE SPECIFICITY**

Most glycosyltransferases show a high degree of specificity for both their donor and acceptor substrates, and this lead Saul Roseman and coworkers to advance the "one enzyme–one linkage" hypothesis. The human B blood group α1–3 galactosyltransferase exemplifies this rule. This enzyme catalyzes a glycosylation reaction in which galactose is added in α linkage to the C-3 hydroxyl group of a galactose residue on the acceptor substrate (Figure 6.1). However, the enzyme only acts on galactose modified by fucose in α1–2 linkage. Prior modification by other monosaccharides, such as α2–6-linked sialic acid, yields a glycan that is not a substrate (Figure 6.1).

We now know that there are instances in which more than one glycosyltransferase can use the same acceptor to make the same linkage. The human fucosyltransferases III–VII, for example, all attach fucose in α1–3 linkage to N-acetyllactosamine moieties on glycans (Chapter 14). Other examples include the α2–3 sialyltransferases that act on galactose and the β1–4 galactosyltransferases that act on N-acetylgalactosamine. In some rare cases, a single enzyme can catalyze more than one reaction. Human fucosyltransferase III can attach fucose in either α1–3 or α1–4 linkage, and an enzyme called EXTL2 can attach either N-acetylgalactosamine or N-acetylgalactosamine in α linkage to glucuronic acid. The β1–4 galactosyltransferase involved in N-acetyllactosamine formation exhibits an unusual flexibility in specificity. When β1-4 galactosyltransferase binds α-lactalbumin (the complex is called lactose synthase), it switches its acceptor specificity from N-acetylgalactosamine to glucose, which enables lactose synthesis during milk formation. Finally, some glycosyltransferases have two separate active sites with different substrate specificities. Those that synthesize the backbones of heparan sulfate (EXT1/EXT2), for example, have one active site that catalyzes the attachment of N-acetylgalactosamine to glucuronic acid and another that attaches glucuronic acid to N-acetylgalactosamine (Chapters 16 and 17). However, the examples described above are all exceptions to the generally strict donor, acceptor and linkage specificity exhibited by most glycosyltransferases, a property that serves to define and limit the number and type of glycan structures observed in a given cell type or organism.

In contrast to those glycosyltransferases that elongate the glycan moiety of their acceptor substrates, initiation of the biosynthesis of glycoproteins and glycolipids involves glycosyltransferases that attach sugars to either a polypeptide side chain or a sphingolipid base or glycerolipid. As might be expected, these enzymes also show a high degree of specificity for their substrates as will be discussed in more detail below for those that transfer to polypeptide. Glycosyltransferases that initiate the
synthesis of glycosphingolipids transfer a monosaccharide moiety to what was originally a serine residue in the ceramide lipid precursor of sphingolipids (see Chapter 11). Because different glycolipids have different ceramide moieties, it appears that some glycosyltransferases, such as the sialyltransferases, differentially recognize their substrates based on the nature of the ceramide moiety.

GLYCOSYLTRANSFERASES THAT RECOGNIZE THE PROTEIN MOIETY OF THEIR ACCEPTOR SUBSTRATES

The glycosyltransferases that transfer sugar directly to the polypeptide chain of a protein or glycoprotein recognize their acceptor substrates in a number of different ways. All eukaryotic N-glycans are initiated by oligosaccharyltransferase, an ER-resident multi-subunit enzyme complex that cotranslationally transfers the N-glycan precursor Glc3Man9GlcNAc2 to asparagine residues in the sequence motif Asn-X-Ser/Thr (where X can be any amino acid except proline; Chapter 9). In contrast, the polypeptide GalNAc transferases responsible for the initiation of mucin-type O-glycans act after the protein has been folded and transported to the Golgi (Chapter 10). The polypeptide GalNAc transferases do not recognize a specific sequence motif but some isoform specificity has been noted. In general, they transfer GalNAc to the side chains of serine and threonine residues found in relatively unstructured regions of the folded protein. Some polypeptide O-GalNAc transferases possess a lectin domain that serves to direct the glycosyltransferase to regions of the polypeptide that already possess glycan chains. In this way, regions of polypeptide that have a high degree of carbohydrate substitution, typical of mucin structures, can be synthesized.

In addition to the O-GalNAc linkage formed by the polypeptide GalNAc transferases, a number of other glycosyltransferases can glycosylate the side chains of serine and threonine residues to generate O-GlcNAc, O-fucose, O-glucose, O-mannose, and O-xylose linkages (Chapters 13, 17 and 19). Specificity for a particular serine or threonine residue is achieved in different ways. The xylosyltransferase that O-xylosylates a serine residue in chondroitin and heparan sulfate proteoglycans, for example, has an absolute requirement for a glycine residue located carboxy terminally to the serine residue and the presence of acidic residues in the vicinity of the glycosylation site. In contrast, the O-GlcNAc transferase, OGT, responsible for adding GlcNAc to serine and threonine residues on hundreds of nuclear and cytoplasmic proteins (Chapter 19) lacks any obvious consensus sequence associated with acceptor substrate binding specificity. For amino acid–consensus sequences or glycosylation motifs used in the formation of glycopeptide bonds, see Table 6.1.

The ER-resident O-fucosyltransferases, POFUT1 and POFUT2, that specifically fucosylate epidermal growth factor-like (EGF) domain and thrombospondin type 1 repeats (TSR), respectively (Chapter 13), differ fundamentally from most other glycosyltransferases. In addition to recognizing a specific sequence motif containing
the target serine or threonine residue (Table 6.1), these enzymes only act on EGF domains and TSRs that are properly folded and disulfide-bonded. Glycosyltransferases that add O-glucose and O-GlcNAc to other serine and threonine residues in the EGF domain also exist and these enzymes also recognize a specific sequence motif and require a folded EGF domain.

A number of glycan elongating glycosyltransferases that act on glycoproteins also recognize the polypeptide moiety of their acceptor substrates. The glycoprotein hormone GalNAc transferase provides an interesting example in which modification of an N-glycan is dependent on the presence of the protein sequence motif Pro-X-Arg/Lys positioned several amino acids amino terminally to the N-glycan being modified. The motif is typically followed closely by additional positively charged residues. The X-ray crystal structure of its acceptor substrate, human chorionic gonadotropin, shows that the Pro-X-Arg/Lys motif is at the beginning of a short surface-exposed helix that also contains the additional positively charged residues (Figure 6.2). The N-acetylgalactosamine residue transferred by this enzyme subsequently undergoes a biologically important 4-O-sulfation reaction that, in the case of lutenizing hormone and follicle-stimulating hormone, generates a determinant recognized by specific liver clearance receptors that remove them from the blood (Chapter 34). Glycosyltransferases that only elongate the fucose moiety added by POFUT1 or POFUT2 also exist (Chapter 13). The specificity shown by these enzymes stems from their ability to recognize both the fucose moiety and the EGF and TSR components of their acceptor substrates. Additional examples of enzymes that elongate glycans on specific glycoprotein substrates include the polysialyltransferases specific for neural cell adhesion molecule (N-CAM) (Chapter 15) and by EXTL3, an N-acetylglucosaminyltransferase that adds GlcNAc in α1–4 linkage to glucuronic acid in the first committed step to heparan sulfate biosynthesis on proteoglycans (Chapter 17).

In a variation on this theme, GlcNAc-1-phosphotransferase is able to selectively modify the N-glycans found on a large family of lysosomal enzymes that differ in three-dimensional structure and that lack any obvious and common protein sequence motif. In this case, modification by this enzyme has been shown to be dependent on lysine residues appropriately spaced and positioned relative to the N-glycan site (Chapter 33). The ER-resident glucosyltransferase, UGGT, is also able to transfer sugar to the N-glycans on a diverse range of glycoprotein substrates (Chapter 39). In this case, the enzyme adds a glucose moiety to the N-glycans on misfolded glycoproteins rendering them substrates for the chaperones calnexin and calreticulin, which in turn recruit ERp57, an enzyme involved in promoting proper disulfide bond formation.
Approximately 1% of genes in the mammalian genome are involved in the production or modification of glycans. Almost 200,000 glycosyltransferase sequences are known across all kingdoms, and they comprise approximately 97 glycosyltransferase families as defined by primary sequence. These are described in the carbohydrate-active enzymes (CAZy) database (Chapter 8). Although the absence of significant sequence similarity between members of one family and another constitutes the basis for this classification, short sequence motifs common to the members of more than one family have been identified. These sequence elements are typically found among glycosyltransferases with a given donor substrate specificity; the sialyl motifs common to eukaryotic sialyltransferases are a good example (Figure 6.3). Sequence motifs common to galactosyltransferases, fucosyltransferases, and N-acetylglucosaminyltransferases have also been identified. In contrast, the so-called “DXD” motif (Asp–any residue–Asp) is not associated with any particular substrate specificity; this motif is involved in metal ion binding and catalysis as discussed in more detail below.

Despite the large number of sequence families that have been defined, structural analysis has shown that glycosyltransferases possess a limited number of fold types. To date, structures for 41 members of the 97 families have been determined by X-ray crystallography, and of these all but a few possess what have been termed the GT-A or GT-B folds (Figure 6.4). The first approximately 120 amino acids of the GT-A fold catalytic domain form a structure similar to that termed the Rossmann fold, found in proteins that bind nucleotides, and in the GT-A glycosyltransferases this region interacts with the nucleotide sugar donor substrate. With only a few exceptions, the GT-A enzymes have been found to possess a DXD motif and are metal-ion-dependent glycosyltransferases. The enzymes containing a GT-B-fold possess two distinct domains, and although the carboxy-terminal domain is primarily responsible for binding the nucleotide sugar donor substrate, both domains possess elements similar to the Rossmann fold. Unlike enzymes that contain the GT-A fold, the GT-B glycosyltransferases are metal-ion independent and do not possess a DXD motif. All of the GT-A and GT-B glycosyltransferases use nucleotide sugar donor substrates and, to date, the structures of only a few enzymes that use lipid-linked sugar donors have been determined. Of these, two are bacterial peptidoglycan glycosyltransferases shown to possess a lambda lysozyme-like fold and two are prokaryotic N-glycan oligosaccharyltransferases (PglB and AglB) that possess both an α-helical transmembrane region and a mixed α/β-fold periplasmic domain (Figure 6.4). PglB and AglB have been classified as GT-C glycosyltransferases and because they are evolutionarily related to the core subunit of the eukaryotic oligosaccharyltransferase, they provide a model for how an enzyme co-translationally mediates N-glycosylation in eukaryotes.
CATALYTIC AND KINETIC MECHANISMS

Glycosyltransferases catalyze their reactions with either inversion or retention of stereochemistry at the anomeric carbon atom of the donor substrate (Figure 6.5). For example, β1–4 galactosyltransferase, an inverting glycosyltransferase, transfers galactose from UDP-α-Gal to generate a β1-4-linked galactose-containing product. Inversion of stereochemistry follows from the fact that the enzyme uses an $S_{N}2$ (substitution nucleophilic bimolecular) reaction mechanism where an acceptor hydroxyl group attacks the anomeric carbon atom of UDP-Gal from one side and UDP leaves from the other (Figure 6.5a). Typically, enzymes of this type possess an aspartate or glutamate residue whose side chain serves to partially deprotonate the incoming acceptor hydroxyl group, rendering it a better nucleophile (as shown in Figure 6.6 for the β1–4 galactosyltransferase). In addition, these enzymes possess features that help to promote leaving-group departure. In the GT-A enzymes, a metal ion, bound by the DXD motif, is typically positioned to interact with the diphosphate moiety. The positively charged metal ion serves to electrostatically stabilize the additional negative charge that develops on the terminal phosphate moiety of the UDP leaving group during breakage of the sugar-phosphate bond of the donor substrate (Figure 6.6). In one of the GT-A enzymes that is not metal ion dependent, positively charged side chains stabilize the leaving group, a strategy also used by some of the GT-B-fold enzymes.

Although the mechanism used by retaining glycosyltransferases is not well understood, insight into how they might work is provided by our knowledge of glycosidase mechanism. On the basis of much structural and enzyme kinetic analysis, it is well established that inverting glycosidases proceed via a single $S_{N}2$ displacement mechanism (Figure 6.5a), whereas retaining glycosidases use a double displacement mechanism involving a covalent glycosyl-enzyme intermediate (Figure 6.5b). In the double displacement mechanism, an aspartic acid or glutamic acid side chain in the enzyme active site makes the first attack and inversion, followed by a second water-mediated attack (on the glycosyl-enzyme intermediate) and inversion, to give overall retention of stereochemistry. In fact, using mechanism-based inhibitors, the glycosyl-enzyme intermediate has been trapped and studied by X-ray crystallography for a number of glycosidases. However, similar attempts to trap and study glycosyltransferase reaction intermediates have not yet been successful. An alternate mechanism, also proposed for glycogen phosphorylase, is that of the $S_{N}1$ mechanism. In this case, the incoming nucleophile attacks from the same side as the leaving group, leading to retention of configuration.

Many glycosyltransferases have been shown to possess a Bi Bi sequential kinetic mechanism in which the donor substrate binds before the acceptor substrate, and the glycosylated acceptor is released before the nucleoside monophosphate or diphosphate, depending on the reaction. Such kinetics are readily explained by a structural model in which the active site represents a deep pocket, with the
nucleotide sugar substrate at the bottom and the acceptor substrate stacked on top. If the acceptor substrate was to bind first, it would sterically preclude donor substrate binding and, as such, catalysis would not occur. Owing to the stacked arrangement, it also follows that release of the glycosylated product must precede release of the nucleoside phosphate. Although largely consistent with such a model, the X-ray crystal structures of glycosyltransferase-substrate complexes also show that substrate-dependent ordering of flexible loops is a feature common to glycosyltransferases. Typically, donor substrate binding orders a loop(s) that in turn facilitates acceptor substrate binding. Moreover, loop ordering almost certainly serves to exclude bulk water from the active site, a strategy used by most enzymes to create an active site environment that reduces the energy of the transition state and promotes catalysis.

GLYCOSIDASES

Glycosides are a very large family of enzymes with 135 members currently listed in the CAZy database. Unlike the glycosyltransferases, however, members of this family have evolved independently many times, a fact reflected in the diverse array of three-dimensional structures observed for these enzymes. Glycosidases play important roles in the degradation of glycan structures for uptake and metabolism of sugars and for the turnover of glycoconjugates in various cellular processes. Glycosidases are also involved in the formation of intermediates that are used as substrates for glycosyltransferases in the biosynthesis of glycans. The use of glycosidases in this way is particularly important in the biosynthesis of N-glycan containing glycoproteins in higher eukaryotes and is thought to be associated with the acquisition of complex N-glycans during the evolution of multicellular organisms. In this case, the nascent glycoprotein glycan, Glc$\beta$Man$_9$GlcNAc$_2$-Asn, is trimmed by glucosidases and mannosidases in the ER and Golgi to generate substrates for the glycosyltransferases that lead to complex and hybrid-type N-glycans (Chapter 9). Glucosidase II, one of the two ER-resident glucosidases involved, also works in conjunction with UGGT during glycoprotein folding to allow repeated deglucosylation/reglucosylation during what has been termed the calnexin/calreticulin quality control cycle (Chapter 39). An interplay between glycosylation and deglycosylation is also found to occur with the O-GlcNAc moiety found on many nuclear and cytoplasmic proteins (transferred by O-GlcNAc transferase; Chapter 19). In this case, removal of the GlcNAc moiety by the glycosidase, O-GlcNAcase, provides a means of dynamically regulating the extent of O-GlcNAcylation and the diverse processes that it mediates.

SULFATION AND OTHER MODIFICATIONS

The sulfotransferases are a large family of enzymes found in both the cytoplasm and the Golgi. Sulfotransferases play a particularly important role in the production of glycosaminoglycans (Chapter 17) and the formation of L-selectin ligands, glycans required for the trafficking of lymphocytes across high endothelial venules in lymph
nodes (Chapter 34). All sulfotransferases use 3′-phospho-adenosine-5′-phosphosulfate (PAPS) as the sulfate donor (Chapter 5). Although the sequence similarity among sulfotransferases can be very low, they all possess conserved sequence motifs responsible for binding the 5′ and 3′ phosphate groups of PAPS. Moreover, structural analysis has shown that all of the sulfotransferases solved, to date, possess the same basic structure. Mechanistically, the enzyme proceeds via an $S_N 2$-like reaction, with the hydroxyl group of the acceptor making an in-line attack on the sulfate group. Structural studies and mutagenesis have shown that a histidine residue serves to activate the hydroxyl nucleophile and a lysine assists in stabilizing the PAP leaving group. Interestingly, a conserved serine residue seems to be involved in modulating the activity of these enzymes to prevent PAPS hydrolysis in the absence of acceptor substrate.

Phosphorylation of sugar residues by ATP dependent kinases also occurs, e.g. at C-2 of 0-xylose in proteoglycans (Chapter 17) and at the C-6 position of 0-mannose in α-dystroglycan, after the mannose has been further glycosylated (Chapter 13). Phosphoglycosylation, a process in which a sugar phosphate is transferred from a nucleotide sugar donor directly to a serine residue on a protein (e.g., GlcNAc-P-serine), occurs in Dictyostelium. In eukaryotic cells, mannose-6-phosphate on the N-glycans of lysosomal enzymes occurs by a two-step process involving UDP-GlcNAc as the phosphate donor. In the first step, GlcNAc-1-phosphotransferase uses UDP-GlcNAc to generate GlcNAcα1-P-6-Manα1-(N-glycan), and in the second step the GlcNAc moiety is removed by a second enzyme to give P-6-Manα1-(N-glycan) (Chapter 33).

0-Acetylation occurs in bacteria and in the modification of sialic acids, but little information is available on the chemistry of these reactions. N-Deacetylation of GlcNAc residues occurs during heparin/heparan sulfate formation (Chapter 17), lipopolysaccharide assembly (Chapters 21 and 22), and GPI-anchor synthesis (Chapter 12). The bacterial enzyme is zinc dependent, but in-depth studies of the vertebrate N-deacetylases have not been performed. N-Deacetylation of N-acetyleneuraminic acid (the most common sialic acid) has also been reported (Chapter 15).

Finally, glycans can be modified in many other ways, including pyruvlation (e.g., in the formation of N-acetylmuramic acid; Chapter 21 and 22), the addition of ethanolamine phosphate (e.g., during GPI-anchor synthesis; Chapter 12), and alkylation, deoxygenation, and halogenation in microbial glycans. All of these reactions are catalyzed by unique transferases or oxidoreductases and represent areas of active research.
ACKNOWLEDGEMENTS

The authors appreciate helpful comments and suggestions from Lei Feng, Tetsuya Hirata and Christopher Saeul.

FURTHER READING


FIGURE 6.1. The strict acceptor substrate specificity of glycosyltransferases is illustrated by the human B blood group $\alpha 1-3$ galactosyltransferase. The B transferase adds galactose in $\alpha 1-3$ linkage to the H antigen (top). This enzyme requires the $\alpha 1-2$-linked fucose modification of the H antigen for activity because the B transferase does not add to an unmodified type-2 precursor (middle), or precursors modified by sialyl residues (bottom) or other monosaccharides (not shown). (For the monosaccharide symbol code, see Figure 1.5, which is reproduced on the inside front cover.)
FIGURE 6.2. Human chorionic gonadotropin showing the determinants of recognition used by glycoprotein hormone $N$-acetylgalactosaminyltransferase. Ribbon diagram of a fragment (residues 34–58) of human chorionic gonadotropin (PDB [protein data bank] ID 1HRP). The Pro40-Leu41-Arg42 tripeptide and residues Lys44 and Lys45 correspond to residues essential for recognition by glycoprotein hormone $N$-acetylgalactosaminyltransferase. Residues Asn52-Val53-Thr54 correspond to the $N$-glycosylation sequence motif of the $N$-glycan (at Asn52) modified by the glycoprotein hormone $N$-acetylgalactosaminyltransferase. Only the chitobiose core (GlcNAcβ1-4GlcNAc) of the acceptor $N$-glycan is shown.

FIGURE 6.3. Sialyl motifs. Domain structure of a typical sialyltransferase, showing the sialyl motifs shared by this family of enzymes. The sialyl L motif of 48–49 amino acids shares significant similarity among members and may be up to 65% identical in amino acid sequence. The sialyl S motif is smaller (~23 amino acids) and diverges more among members of the family, with only two absolutely conserved residues. In both cases, identical residues are indicated and residues showing similarity are
denoted by parentheses. (Asterisk) Position of a highly conserved sequence H-X₄-E. Additional conserved motifs may exist as well.

FIGURE 6.4. Ribbon diagrams of representative GT-A, GT-B, GT-C and lysozyme-type fold glycosyltransferases. The GT-A and GT-B structures correspond to those of rabbit β1-2 N-acetylglcosaminyltransferase I (PDB ID 1FOA) and T4 phage β-glucosyltransferase (PDB ID 1J39), respectively. In both cases, the bound nucleotide sugar donor substrate is shown in stick representation. The GT-C structure is that of *Campylobacter lari* oligosaccharyltransferase, PglB (PDB ID 3RCE), and the lysozyme-type structure is that of the *Staphylococcus aureus* bacterial peptidoglycan glycosyltransferase (PDB ID 2OLV) in complex with moenomycin (stick representation).
FIGURE 6.5. Schematic representation of (a) inverting and (b) retaining catalytic mechanisms. (a) S$_2$N$_2$-like attack of the acceptor leads to inversion of stereochemistry at C1. For a glycosidase reaction, R2 would correspond to a proton and R1 would be the remainder of the glycan. A and B label general acid and base groups in the catalytic site of the enzyme. For a glycosyltransferase reaction, R2 would correspond to the remainder of the acceptor substrate and R1 would typically be the nucleoside monophosphate or diphosphate moiety of the donor substrate. (b) This mechanism has only been established for glycosidases. Two successive S$_2$N$_2$-like reactions separated by a glycosyl-enzyme intermediate lead to retention of the configuration at C1. R1 corresponds to the remainder of the glycan and R2 to a proton.

FIGURE 6.6. Catalytic site of bovine β1–4 galactosyltransferase. Composite figure shows selected residues/atoms of the superimposition of the donor complex (PDB ID 1TW1) on the acceptor complex (PDB ID 1TW5). O4 designates the C4 hydroxyl group of the GlcNAcβ1–4GlcNAc acceptor substrate positioned for in-line S$_2$N$_2$ attack
(arrow) on C1 of the UDP-Gal donor substrate. The base form of D318 serves to partially deprotonate the C4 hydroxyl group rendering it a better nucleophile. The positively charged Mg$^{++}$ ion coordinates the two phosphates of the UDP leaving group, promoting cleavage of the C1-P$_{\beta}$ bond by stabilizing the additional negative charge that develops on P$_{\beta}$ of the leaving group. D252-V253-D254 corresponds to the DXD motif in bovine β1-4 galactosyltransferase.
### TABLE 6.1 Amino acid–consensus sequences or glycosylation motifs for the formation of glycopeptide bonds

<table>
<thead>
<tr>
<th>Glycopeptide bond</th>
<th>Consensus sequence or peptide motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAc-β-Asn</td>
<td>Asn-X-Ser/Thr (X = any amino acid except Pro)</td>
</tr>
<tr>
<td>Glc-β-Asn</td>
<td>Asn-X-Ser/Thr</td>
</tr>
<tr>
<td>GalNAc-α-Ser/Thr</td>
<td>repeat domains rich in Ser, Thr, Pro, Gly, Ala with no specific sequence</td>
</tr>
<tr>
<td>GlcNAc-α-Thr</td>
<td>Thr-rich domain near Pro residues</td>
</tr>
<tr>
<td>GlcNAc-β-Ser/Thr</td>
<td>Ser/Thr-rich domains near Pro, Val, Ala, Gly</td>
</tr>
<tr>
<td>Man-α-Ser/Thr α</td>
<td>Ser/Thr-rich domains</td>
</tr>
<tr>
<td>Fuc-α-Ser/Thr</td>
<td>EGF modules (Cys-X-X-X-Ser/Thr-Cys)</td>
</tr>
<tr>
<td>Glc-β-Ser</td>
<td>EGF modules (Cys-X-Ser-X-Pro/Ala-Cys)</td>
</tr>
<tr>
<td>Xyl-β-Ser</td>
<td>Ser-Gly (in the vicinity of one or more acidic residues)</td>
</tr>
<tr>
<td>Glc/GlcNAc-Thr</td>
<td>Rho: Thr-37; Ras, Rac; Cdc42: Thr-35</td>
</tr>
<tr>
<td>Gal-Thr</td>
<td>Gly-X-Thr (X = Ala, Arg, Pro, Hyp, Ser) (vent worm)</td>
</tr>
<tr>
<td>Gal-β-Hyl</td>
<td>collagen repeats (X-Hyl-Gly)</td>
</tr>
<tr>
<td>Ara-α-Hyp</td>
<td>repetitive Hyp-rich domains (e.g., Lys-Pro-Hyp-Hyp-Val)</td>
</tr>
<tr>
<td>GlcNAc-Hyp</td>
<td>Skp1: Hyp-143</td>
</tr>
<tr>
<td>Glc-α-Tyr</td>
<td>glycogenin: Tyr-194</td>
</tr>
<tr>
<td>GlcNAc-α-1-P-Ser</td>
<td>Ser-rich domains (e.g., Ala-Ser-Ser-Ala)</td>
</tr>
<tr>
<td>Man-α-1-P-Ser</td>
<td>Ser-rich repeat domains</td>
</tr>
<tr>
<td>Man-α-C-Trp</td>
<td>Trp-X-X-Trp</td>
</tr>
<tr>
<td>Man-6-P-ethanolamine-protein</td>
<td>GPI attached after cleavage of carboxy-terminal peptide</td>
</tr>
</tbody>
</table>

*Based on Spiro R.G. 2002. Glycobiology 12:43R–56R; see also Table 1.7. (EGF) Epidermal growth factor; (TSR) thrombospondin repeats; (Hyp) hydroxyproline; (Hyl) hydroxylysine; (GPI) glycosylphosphatidylinositol; (Ara) arabinose.*