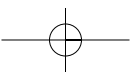
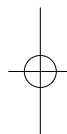
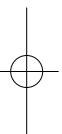
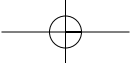


# Part I

## **Structures and biosynthesis of glycans**



# Concepts of glycobiology

## LEARNING OBJECTIVES

By the end of this chapter students should understand

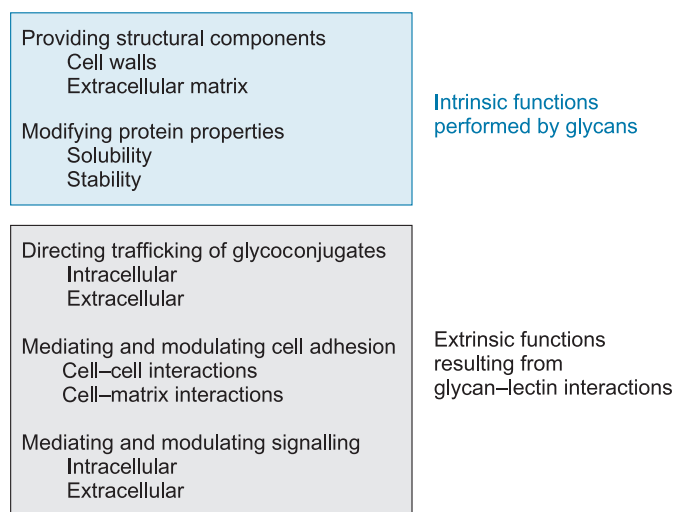
- 1 the classes of glycoconjugates and their locations in eukaryotic cells,
- 2 structures of the sugar monomers found in glycoproteins and glycolipids,
- 3 the nature of glycosidic linkages and the way they are formed,
- 4 the types of function served by glycosylation.

The challenge of glycobiology is to define the biological functions of sugars attached to proteins and membranes and to determine how these functions are carried out. Although this challenge is ongoing, the principles that describe how protein- and lipid-linked sugars mediate biological processes are becoming clear. The goal of this text is to illustrate the themes that underlie the functions of glycoproteins and glycolipids by highlighting well-understood examples of how the carbohydrate portions of these molecules work.

## 1.1 The field of glycobiology encompasses the multiple functions of sugars attached to proteins and lipids

Most biochemists initially encounter carbohydrates in the context of energy metabolism in cells. The pathways through which energy is abstracted from the breakdown of glucose and glycogen are familiar from introductory biochemistry textbooks. Glycogen is even linked to a protein core and could thus be considered a glycoprotein. However, the function of sugar molecules as forms of energy for storage and transport is generally considered to fall outside the field of glycobiology. This distinction is arbitrary but useful, because it allows us to focus on other functions of sugars, many of which are less well understood.

**Glycoconjugates** are formed when mono-, oligo-, or polysaccharides are attached to proteins or lipids. The sugar-containing portions of the resulting **glycoproteins** and **glycolipids** are generally complex heteropolymers rather than repeating



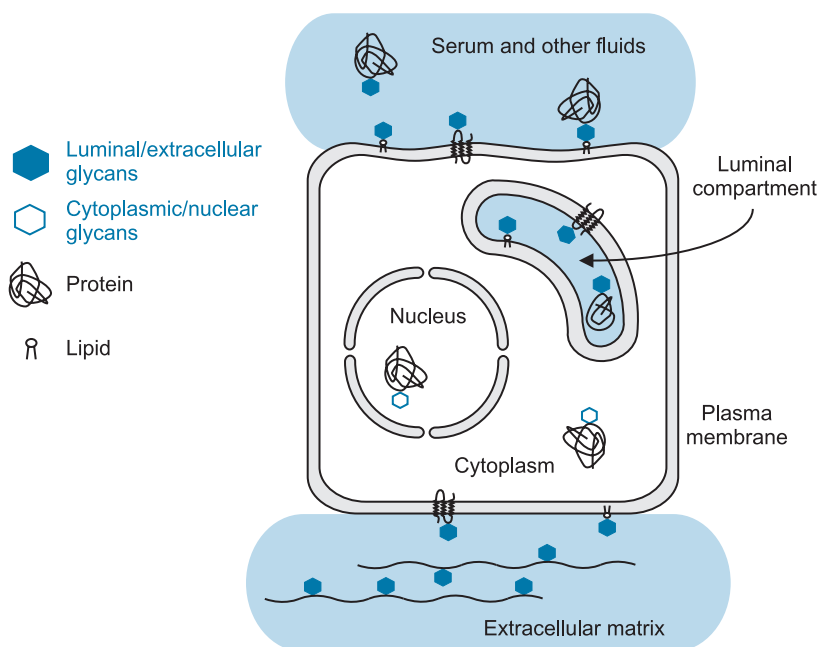
**Figure 1.1** Summary of some of the functions of glycans.

homopolymers such as the storage polysaccharides glycogen and amylose. It is common to refer to the sugars of glycoproteins and glycolipids as **glycans**. Glycans can be built from some of the same building blocks that serve as energy stores, such as glucose, but they also include other monosaccharide units.

There is no single predominant function of protein- and lipid-linked glycans. In this respect they are similar to proteins themselves, which serve diverse functions as enzymes, hormones, transporters, and structural elements. The functions of glycans fall into at least five broad categories that are summarized in this chapter and discussed in detail in subsequent parts of this book (Figure 1.1). The structural roles of glycoconjugates reflect the physical properties of glycans themselves, so the proteins and lipids to which structural glycans are attached can be viewed as scaffolds that serve an organizational role. Glycans can also affect the intrinsic properties of proteins to which they are attached. In order to function in trafficking, adhesion, and signalling, glycans usually must interact with protein receptors that are known as **lectins**. The field of glycobiology encompasses the study of glycoconjugates, the enzymes that catalyse their biosynthesis, and the lectins that recognize them.

## 1.2 There are three major classes of glycoconjugates

The majority of glycans to be considered in this book fall into three groups: those attached to lipids, and those attached to proteins either through a nitrogen atom (**N-linked**) or through an oxygen atom (**O-linked**). Both glycoproteins and glycolipids are found at the extracellular surface of the plasma membrane. In addition, glycoproteins are secreted into biological fluids, such as serum, and they also make up the insoluble extracellular matrix that surrounds cells (Figure 1.2). All of these groups



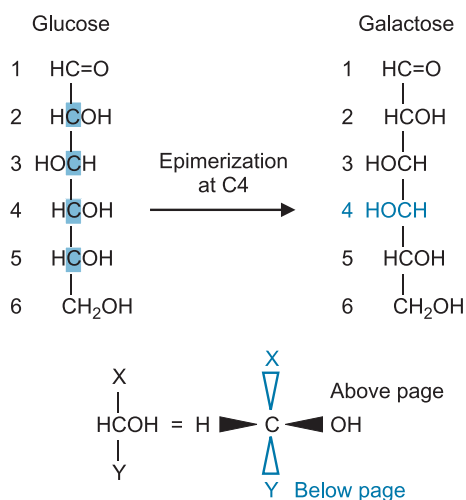
**Figure 1.2** Localization of glycoconjugates in intracellular and extracellular compartments.

of molecules have been studied intensively, but the N-linked glycans attached to soluble, secreted proteins are understood best. The emphasis on glycoproteins bearing N-linked oligosaccharides reflects the historical availability of serum glycoproteins as targets for investigation. This group of glycoproteins will be used as a point of reference throughout this book. Before they arrive at the cell surface, glycoconjugates must be made inside the cell. This process occurs in the lumen of the endoplasmic reticulum and the Golgi apparatus. Indeed, the steps of glycosylation form an integral part of the secretory machinery of the cell. Thus, most glycoconjugates are separated from the cytoplasm by a membrane. However, partitioning of glycoconjugates into the extracytoplasmic compartments of the cell is not absolute, because there are also cytoplasmic and nuclear forms of glycosylation.

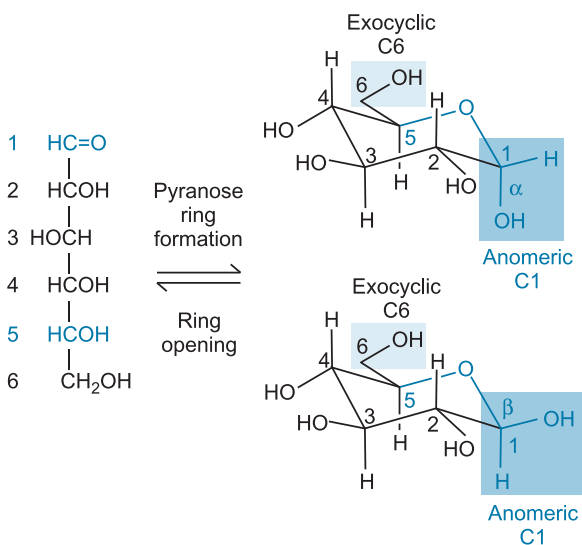
### 1.3 Glycans are composed of monosaccharides with related chemical structures

The most common constituents of glycans are **hexoses**. Four of the six carbon atoms in a hexose are chiral centres, because each of carbon atoms 2, 3, 4, and 5 is bonded to four chemically distinct structures (Figure 1.3). The substituents around each of these carbon atoms can be arranged in two stereochemically different ways. Because there are four chiral centres, each of which can exist in two configurations, there are a total of 16 possible hexoses. A series of eight common names with the

## 6 Concepts of glycobiology



**Figure 1.3** Stereochemistry of glucose and galactose. The four asymmetric carbon atoms in glucose are shaded and the convention for showing the stereochemistry is indicated at the bottom.



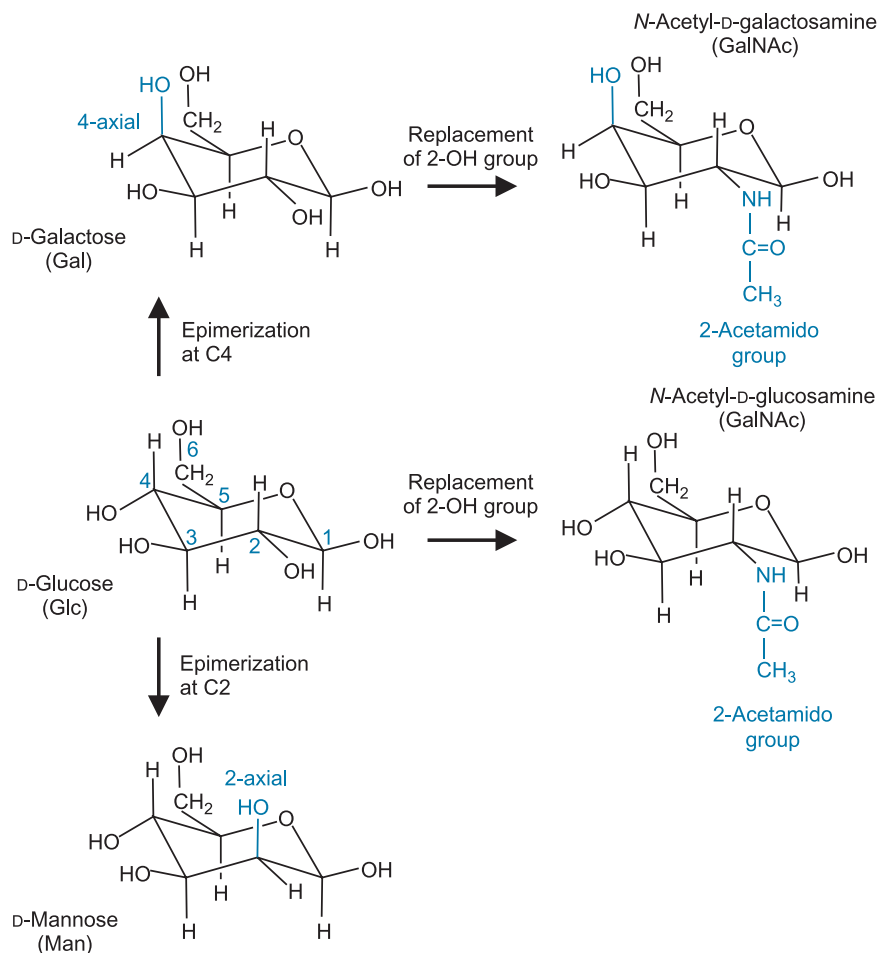
**Figure 1.4** Two ways to form a pyranose (six-member) ring from glucose, in which the anomeric C1 is in either the  $\alpha$  or  $\beta$  configuration.

prefixes *D* or *L* are used to denote these 16 hexoses. For example, **D-glucose** and **D-galactose** differ in the configuration of carbon 4 (or C4) only. A change in the stereochemical configuration of a single carbon atom in a sugar is referred to as **epimerization**, so *D*-glucose and *D*-galactose are **epimers**. In contrast, *D*-glucose and *L*-glucose are mirror images of each other, because the configuration of each of

the asymmetric carbons is reversed in these two hexoses. Such pairs of mirror images are called **enantiomers**.

The hexoses in glycoconjugates are normally found in a six-member ring form known as the **pyranose configuration** (Figure 1.4). The ring is created by reaction of the 5-hydroxyl group with the 1-aldehyde group to create a hemiacetal. The pyranose configuration can be explicitly indicated in the name and abbreviation for a hexose, as in **D-glucofuranose** and **D-Glcp**. However, it is common to make this designation implicit. In the ring form, the C1 atom is now linked to four chemically distinct atoms, so it is a chiral centre and can exist in either of two stereochemical configurations. C1 is referred to as the **anomeric carbon** and the two configurations or **anomers** are distinguished by the designations  $\alpha$  and  $\beta$ . C6 is located outside the ring and is thus **exocyclic**.

Only a few of the possible hexoses are commonly found in glycoconjugates (Figure 1.5). **Glucose (Glc)** is a convenient reference compound when comparing

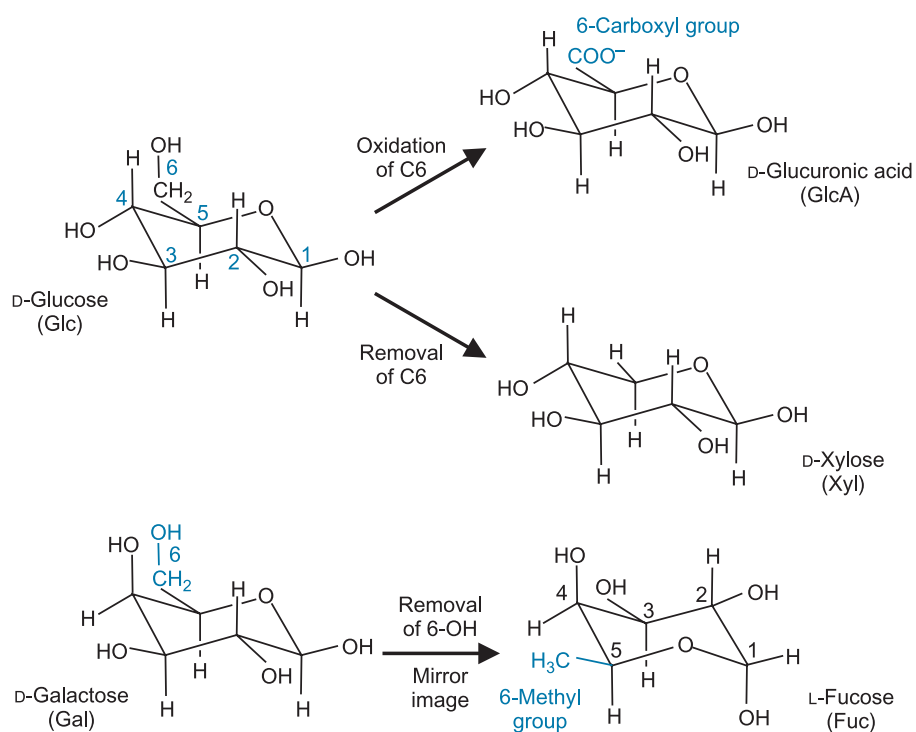


**Figure 1.5** Relationships between the common hexoses and N-acetylhexosamines.

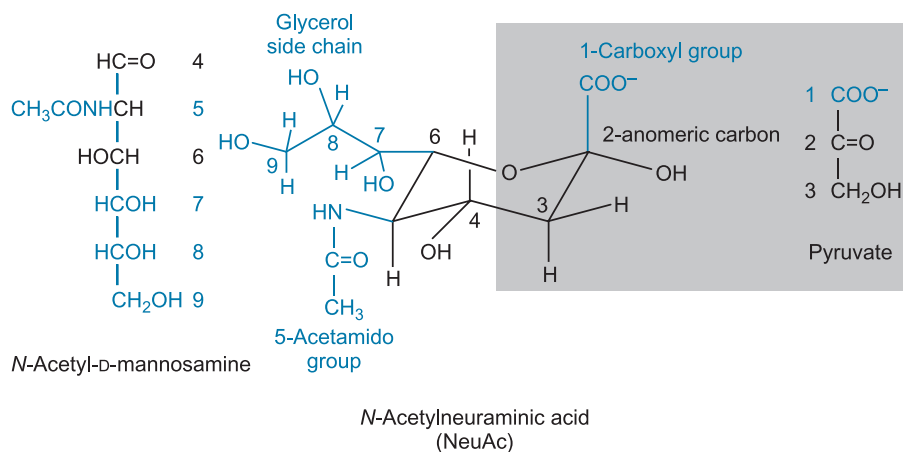
these structures, as many of them can be derived from glucose by a single epimerization or substitution. For example, epimerization at the 2 position yields **mannose (Man)**, whereas epimerization at the 4 position yields **galactose (Gal)**. Substitution of the 2-hydroxyl group of glucose or galactose with an acetylated amino group yields **N-acetylglucosamine (GlcNAc)** or **N-acetylgalactosamine (GalNAc)**. All of these hexoses are normally found in the D configuration, so it is common to make this implicit when describing glycan structures.

Additional modified forms of the simple hexoses appear in glycoconjugates, often resulting from changes in C6 (Figure 1.6). For example, oxidation of C6 to a carboxyl group creates a **sugar acid** such as **glucuronic acid (GlcA)**. Loss of C6 from glucose leads to the generation of **xylose (Xyl)**, which is a **pentose**. The form of **fucose (Fuc)** commonly found in mammalian glycoproteins is an interesting case. It is related to galactose by loss of the 6-hydroxyl group, but it is in the L rather than the D configuration. Thus, it is the mirror image of 6-deoxy-D-galactose.

The term **sialic acid** encompasses a large family of sugars. One member of this family, **N-acetylneuraminic acid (NeuAc)**, is the form that is most commonly found in mammalian glycoconjugates (Figure 1.7). NeuAc is a nine-carbon sugar acid formed by condensation of pyruvate with **N-acetylmannosamine**. It is usually found in a six-member ring configuration that is formed by joining the carbonyl group at



**Figure 1.6** Structures of some common derivatives of the hexoses.



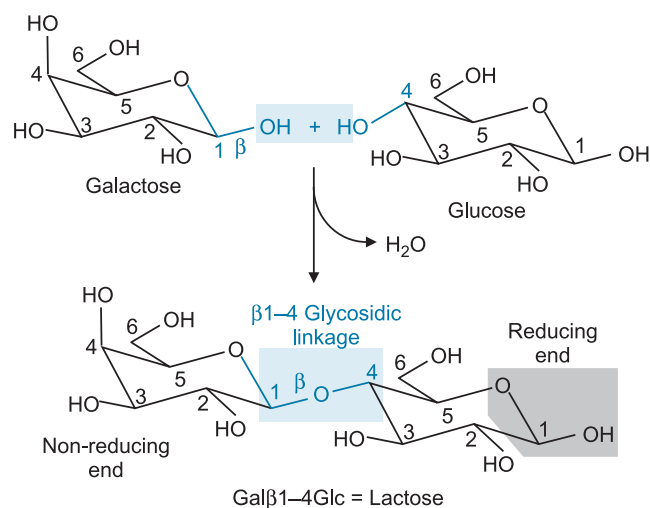
**Figure 1.7** Structure of N-acetylneuraminic acid, which is the most common form of sialic acid. C1–3 are derived from pyruvate and C4–9 are derived from N-acetylmannosamine.

C2 and the 6-hydroxyl group in a hemiketal. There are several unique substituents extending from the ring, including the carboxyl group at position 1, the acetylated amino group attached to C5, and the three-carbon chain consisting of C7, C8, and C9. Because each of C7, C8, and C9 is hydroxylated, this substituent is often referred to as the glycerol side chain.

#### 1.4 Glycosidic linkages between monosaccharides exist in multiple configurations

Reaction of a monosaccharide at the hemiacetal of the ring conformation with an hydroxyl group of another monosaccharide results in condensation to form an acetal, with concomitant elimination of a water molecule (Figure 1.8). The resulting structure is a **glycosidic linkage**. As in the free sugars, the anomeric carbon, C1, is linked to four chemically distinct atoms and can exist in  $\alpha$  or  $\beta$  configurations. The structure of a disaccharide can be indicated unambiguously by giving the structures of the constituent monosaccharides and the nature of the linkages. For example, lactose is Gal $\beta$ 1–4Glc. A glycosidic linkage has been formed between the C1 of galactose and the 4-hydroxyl group of glucose, and the anomeric C1 of galactose is in the  $\beta$  configuration. The glucose residue retains the aldehyde function in the form of a hemiacetal. Because of the ability of this group to reduce inorganic ions such as Cu<sup>2+</sup>, it is referred to as the **reducing end** of the disaccharide. The galactose residue constitutes the **non-reducing end**.

The complete chemical structure of a typical N-linked glycan that might be found in a glycoprotein can be quite cumbersome (Figure 1.9). A description of the monosaccharides and linkages suffices to convey the same information. In spite of the simplifications provided by this nomenclature, it is occasionally useful to provide a still simpler pictorial representation of the structures. By assigning shapes to the



**Figure 1.8** Formation of a glycosidic linkage. The glycosidic bond is formed between the reducing end of one monosaccharide, galactose in this case, and one of the other hydroxyl groups of a second monosaccharide, such as the 4-OH group of glucose shown here, by abstraction of a water molecule. After formation of the bond, the only reducing group remaining is at the 1 position of the glucose, which is defined as the reducing end of the disaccharide.

various monosaccharides, much of the structural information can be presented in condensed form. Although some information, such as the nature of the linkages, is lost in this representation, this symbol convention is particularly convenient when a structure has been described in detail and is being repeated in modified form.

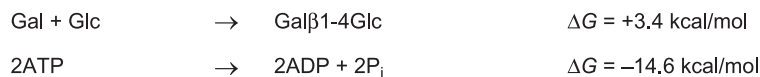
### 1.5 Formation of glycosidic linkages requires energy and is catalysed by specific enzymes

Creation of a disaccharide such as **lactose**, by formation of a glycosidic linkage between the two monosaccharides galactose and glucose, is an energetically unfavourable process (Figure 1.10). As with many other biochemical processes, the free energy needed to form the bond is generated by coupling the reaction to the energetically favourable hydrolysis of phosphate anhydride bonds. This energy coupling takes place in two stages. The energy of hydrolysis of two high-energy phosphate bonds in adenosine triphosphate (ATP) is first used to drive formation of a **nucleotide sugar donor**, uridine diphosphate (UDP)-galactose (Figure 1.11). UDP-galactose is then used to make the glycosidic linkage with glucose.

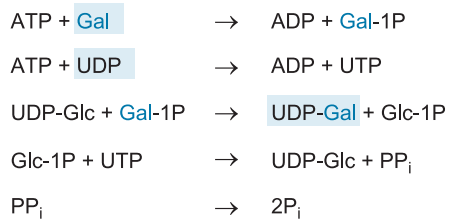
Other nucleotide sugar donors can be generated in a similar manner. A **glycosyl-transferase** catalyses the transfer of sugar from the donor and is thus responsible for the formation of the glycosidic linkage. A glycosyltransferase has specificity for a nucleotide sugar donor and an acceptor. In the synthesis of lactose, glucose serves



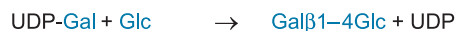
Overall energetics of glycosidic bond formation



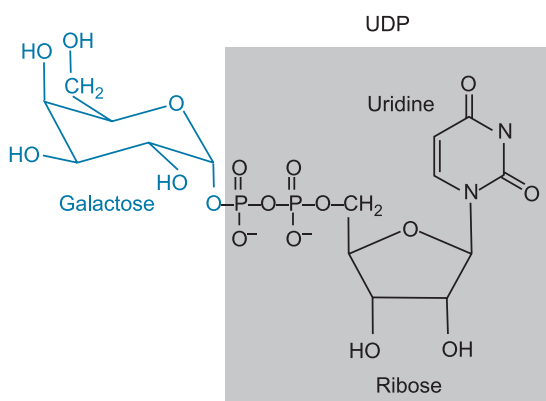
Synthesis of nucleotide sugar donor



Creation of glycosidic bond



**Figure 1.10** Energetics of formation for a glycosidic bond. The energetically unfavourable creation of the glycosidic bond is ultimately driven by the hydrolysis of two phosphate linkages in ATP. The reactions are linked by synthesis of UDP-galactose. Energy originally derived from ATP is released when the nucleotide sugar serves as the sugar donor during formation of the glycosidic bond. ADP, adenosine diphosphate; ATP, adenosine triphosphate; UDP, uridine diphosphate; UTP, uridine triphosphate; PP<sub>i</sub>, pyrophosphate.



**Figure 1.11** Structure of a nucleotide sugar that can serve as a sugar donor in a glycosyltransferase reaction. UDP, uridine diphosphate.

each recognize more than just the mannose acceptor residue (see Chapter 3). In contrast, the two galactose and sialic acid residues can be added by a single type of galactosyltransferase and sialyltransferase. Although each glycosyltransferase is highly specific, there are often several different glycosyltransferases with similar or overlapping specificities. For example, in humans there are six sialyltransferases that can each add sialic acid in 2-3 linkage to galactose.

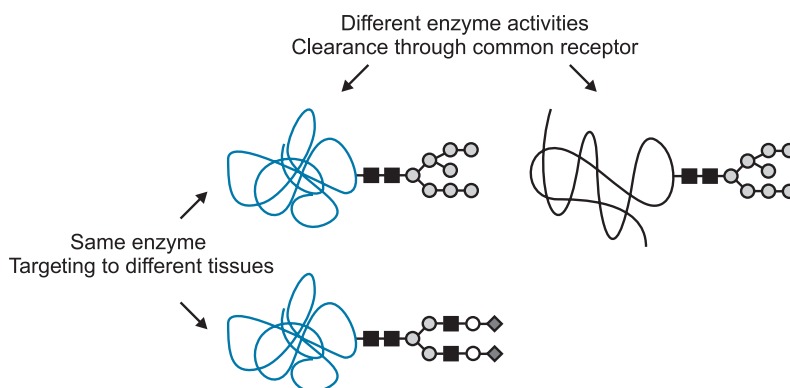
Breaking of glycosidic linkages also requires specific enzymes, but does not require the input of energy. **Glycosidases** catalyse the energetically favourable hydrolysis of glycosidic linkages. Like glycosyltransferases, glycosidases are specific, with each enzyme only catalysing hydrolysis of glycosidic linkages involving a particular sugar. For example, a **sialidase (neuraminidase)** can catalyse release of a sialic acid residue (NeuAc) from the non-reducing terminus of an oligosaccharide such as the one shown in Figure 1.9. Glycosidases can also be specific for particular linkages of a sugar. Taking sialidase as the example again, some sialidases will only hydrolyse  $\alpha$ 2–3 linkages of NeuAc, whereas others do not discriminate between  $\alpha$ 2–3- and  $\alpha$ 2–6-linked NeuAc.

---

## 1.6 Understanding structure–function relationships for glycans can be more difficult than for other classes of biopolymer

Comparison with proteins serves as a good starting point for thinking about how glycans carry out useful biological functions. In spite of their diverse biological roles, proteins share two common features that unify the study of their properties: each protein is synthesized as an identical copy by translation of an mRNA template that is encoded in the genome and the activity of a protein results from formation of a precisely folded three-dimensional structure. In contrast, glycans are assembled without a template through a series of individually catalysed reactions. The resulting structures are not unique, because many different proteins are modified with a common set of glycan structures and different copies of a single polypeptide backbone can be modified with scores of distinct glycans. In addition, glycans often appear to lack a discrete, folded structure. All of these features make it difficult to establish structure–function relationships for glycans and some novel principles must be defined to describe the ways that glycans function.

One explanation for the lack of simple rules about which specific glycans are attached to specific proteins is that the functions of the protein and glycan portions of many glycoproteins can be independent of each other. That is, all copies of a particular protein perform the same function regardless of what glycans are attached and all copies of a particular glycan perform the same function, although they are attached to different proteins (Figure 1.12). This arrangement can be particularly effective when glycans serve as tags that can be recognized and used to direct glycoprotein trafficking. For example, glycoproteins in the secretory pathway of eukaryotic cells are subjected to a series of quality-control checks. The common glycans attached



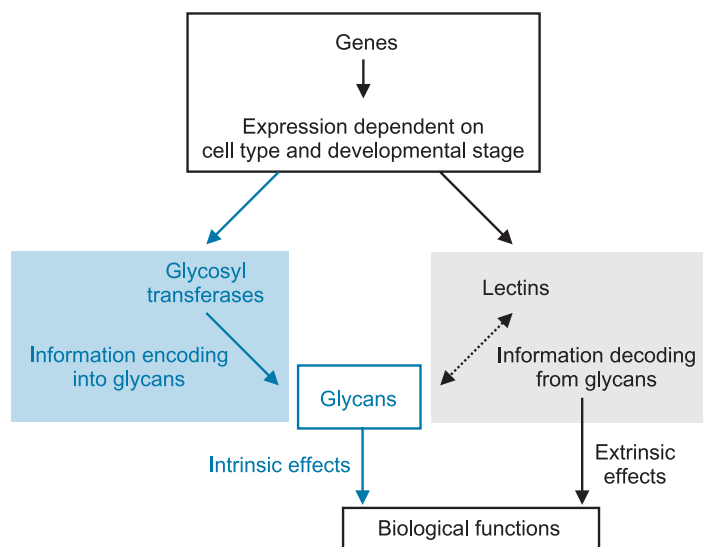
**Figure 1.12** Potential independent functions of polypeptide and glycan portions of glycoproteins. Different proteins can bear identical glycans and different copies of one protein can be heterogeneously glycosylated with multiple glycans. Thus, targeting to different sites directed by the glycans may be independent of the enzymatic or other functions of the proteins.

to the majority of these proteins are used to hold them in the appropriate luminal compartments during this process. One set of glycans can serve this function for a multitude of different secretory glycoproteins, in spite of the fact that the proteins themselves will serve diverse functions once outside the cell. In such cases, the glycan portion of the mature glycoprotein may have essentially no role once the protein has reached the cell surface.

In other cases, the independent functions of protein core and glycan decoration may be manifest in a different way. When a particular glycan is attached to a protein or lipid at the extracellular surface of the plasma membrane, it may mediate adhesion or anti-adhesion events independently of the carrier lipid or protein to which it is attached. The presence of similar glycan structures on a variety of membrane glycoproteins and glycolipids may provide a mechanism for achieving high densities of these structures without requiring a correspondingly high density of any one type of membrane protein or lipid. Thus, in some cases, glycans and the proteins or lipids to which they are attached can be studied more or less in isolation from each other. However, in other cases the role of a particular glycan is only evident in the context of a specific glycoconjugate.

### 1.7 Glycan structures are encoded indirectly in the genome

Genomic DNA sequences dictate the structures of glycoconjugates just as they determine the structure of all cell components. The sugar structures are not encoded directly in the DNA sequences but are determined by transcription and translation of genes to generate glycosyltransferases that in turn control synthesis of the glycan portions of glycoconjugates (Figure 1.13). Thus, compared with the biosynthesis of



**Figure 1.13** Steps in encoding and decoding information in glycan structures.

proteins, there is an extra step in the decoding process. The one enzyme–one linkage rule suggests that it will eventually be possible to describe the full repertoire of glycan structures that can be made in a particular cell by determining which glycosyltransferases are expressed in this cell.

An important tool in understanding both the synthesis of glycoconjugates and their functions has been the generation of knockout mice in which glycosyltransferases have been eliminated. Results of many specific experiments of this type will be cited throughout this book. However, two important general results are worth noting at this point. First, complete elimination of any of the classes of glycoconjugates is fatal to the organism at an early stage in development, demonstrating that these molecules perform critical biological functions for the organism. Second, in almost all cases, cells lacking any of these classes of glycoconjugates are viable. Thus, although a number of important roles of glycoconjugates will be discussed in a cellular context, many of their functions are organismal rather than cellular.

## SUMMARY

Because glycan structures are created by glycosyltransferases and are recognized by lectins, glycobiology encompasses the study of these proteins as well as the glycans themselves. There are a large number of possible hexoses and related structures, but only a small fraction of these monosaccharide units are found in glycoconjugates. Similarly, the monosaccharide units could potentially be combined in very many different ways through different linkages, but the glycosyltransferases catalyse synthesis of only a limited number of the possible structures. Nevertheless, there is great diversity in the glycan portions of glycoconjugates. It is often difficult to assign functions to specific glycans

because the functions may only be evident in an organismal context and gene-knockout approaches may be required to probe these roles.

---

## KEY REFERENCES

- Alper, J. (2001) Searching for medicine's sweet spot. *Science* **291**, 2338–2343. [This issue of \*Science\* highlights the field of glycobiology with several reviews and features about recent advances.](#)
- Drickamer, K. and Taylor, M.E. (1998) Evolving views of protein glycosylation. *Trends in Biochemical Sciences* **23**, 321–324. [Evolution is used to relate the different functions of glycosylation to each other.](#)
- Dwek, R.A. (1995) Glycobiology: towards understanding the function of sugars. *Biochemical Society Transactions* **23**, 1–25. [Some key examples of how glycosylation affects protein function are reviewed.](#)
- Fukuda, M. and Hindsgaul, O. (2000) *Molecular and Cellular Glycobiology*. Oxford: Oxford University Press. [This collection of reviews covers topics of recent interest in detail.](#)
- Lis, H. and Sharon, N. (1993) Protein glycosylation: structural and functional aspects. *European Journal of Biochemistry* **218**, 1–27. [In addition to summarizing what glycans are attached to proteins, these authors speculate on some of their functions.](#)
- Roseman, S. (2001) Reflection on glycobiology. *Journal of Biological Chemistry* **276**, 41527–41542. [Reflecting on the development of carbohydrate biochemistry, a veteran of the field provides perspective on how we now think about the functions of glycosylation.](#)
- Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999) *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Press. [This was the first truly comprehensive book on glycobiology and it provides extensive detail on many topics along with a wealth of references.](#)

---

## QUESTIONS

- 1.1 Compare the structural features of proteins and oligosaccharides.
- 1.2 What are the main types of function performed by glycans attached to glycoproteins and glycolipids? Discuss why glycans might be better suited than proteins to performing some of these functions.