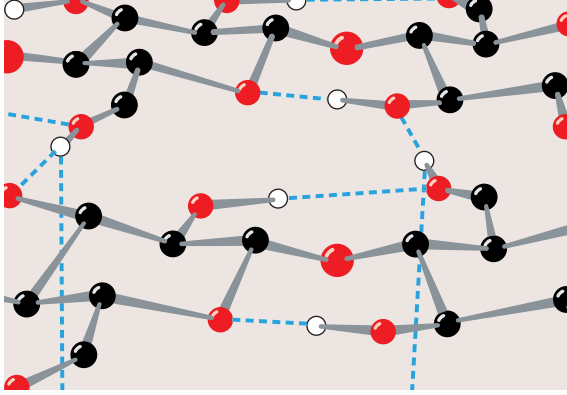


Sugars and Polysaccharides

CHAPTER 11

Carbohydrates or **saccharides** (Greek: *sakcharon,* sugar) are essential components of all living organisms and are, in fact, the most abundant class of biological molecules. The name carbohydrate, which literally means "carbon hydrate," stems from their chemical composition, which is roughly $(C \cdot H_2O)_n$, where $n \geq 3$. The basic units of carbohydrates are known as **monosaccharides.** Many of these compounds are synthesized from simpler substances in a process named **gluconeogenesis** (Section 23-1). Others (and ultimately nearly all biological molecules) are the products of **photosynthesis** (Section 24-3), the light-powered combination of CO_2 and H_2O through which plants and certain bacteria form "carbon hydrates." The metabolic breakdown of monosaccharides (Chapters 17 and 21) provides much of the energy used to power biological processes. Monosaccharides are also principal components of nucleic acids (Section 5-1A), as well as important elements of complex lipids (Section 12-1D).

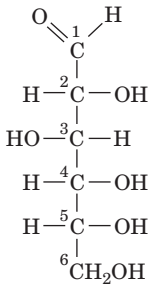
Oligosaccharides consist of a few covalently linked monosaccharide units. They are often associated with proteins **(glycoproteins)** and lipids **(glycolipids)** in which they have both structural and regulatory functions (glycoproteins and glycolipids are collectively called **glycoconjugates**). **Polysaccharides** consist of many covalently linked monosaccharide units and have molecular masses ranging well into the millions of daltons. They have indispensable structural functions in all types of organisms but are most conspicuous in plants because **cellulose,** their principal structural material, comprises up to 80% of their dry weight. Polysaccharides such as **starch** in plants and **glycogen** in animals serve as important nutritional reservoirs.

The elucidation of the structures and functions of carbohydrates has lagged well behind those of proteins and nucleic acids. This can be attributed to several factors. Carbohydrate compounds are often heterogeneous, both in size and in composition, which greatly complicates their physical and chemical characterization. They are not subject to the types of genetic analysis that have been invaluable in the study of proteins and nucleic acids because saccharide sequences are not genetically specified but are built up through the sequential actions of specific enzymes (Section 23-3B). Furthermore, it has been difficult to establish assays for the biological activities of polysaccharides because of their largely passive roles. Nevertheless, it is abundantly clear that carbohydrates are essential elements in many, if not most, biological processes.

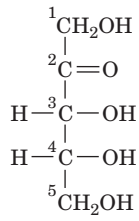
In this chapter, we explore the structures, chemistry, and, to a limited extent, the functions of carbohydrates, alone and in association with proteins. Glycolipid structures are considered in Section 12-1D. The biosynthesis of complex carbohydrates is discussed in Section 23-3.

1 MONOSACCHARIDES

Monosaccharides or **simple sugars** are aldehyde or ketone derivatives of straight-chain polyhydroxy alcohols containing at least three carbon atoms. Such substances, for example, **D-glucose** and **D-ribulose,** cannot be hydrolyzed to form simpler saccharides.



D-Glucose



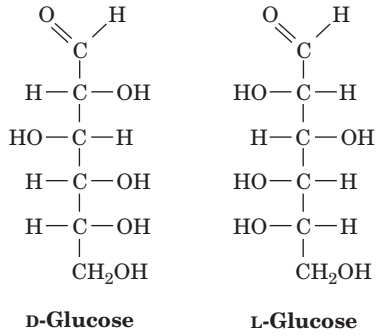
D-Ribulose

In this section, the structures of the monosaccharides and some of their biologically important derivatives are discussed.

A. *Classification*

Monosaccharides are classified according to the chemical nature of their carbonyl group and the number of their C atoms. If the carbonyl group is an aldehyde, as in glucose, the sugar is an **aldose.** If the carbonyl group is a ketone, as in ribulose, the sugar is a **ketose.** The smallest monosaccharides, those with three carbon atoms, are **trioses.** Those with four, five, six, seven, etc., C atoms are, respectively, **tetroses, pentoses, hexoses, heptoses,** etc. These terms may be combined so that, for example, glucose is an **aldohexose,** whereas ribulose is a **ketopentose.**

Examination of D-glucose's molecular formula indicates that all but two of its six C atoms—C1 and C6—are chiral centers, so that D-glucose is one of $2^4 = 16$ stereoisomers that comprise all possible aldohexoses. In general, *n*-carbon aldoses have 2^{n-2} stereoisomers. The stereochemistry and names of the D-aldoses are presented in Fig. 11-1. Emil Fischer elucidated these configurations for the aldohexoses in 1896. According to the Fischer convention (Section 4-2B), *D sugars have the same absolute configuration at the asymmetric center farthest removed from their carbonyl group as does D-glyceraldehyde.* The L sugars, in accordance with this convention, are mirror images of their D counterparts, as is shown below in Fischer projection for glucose.



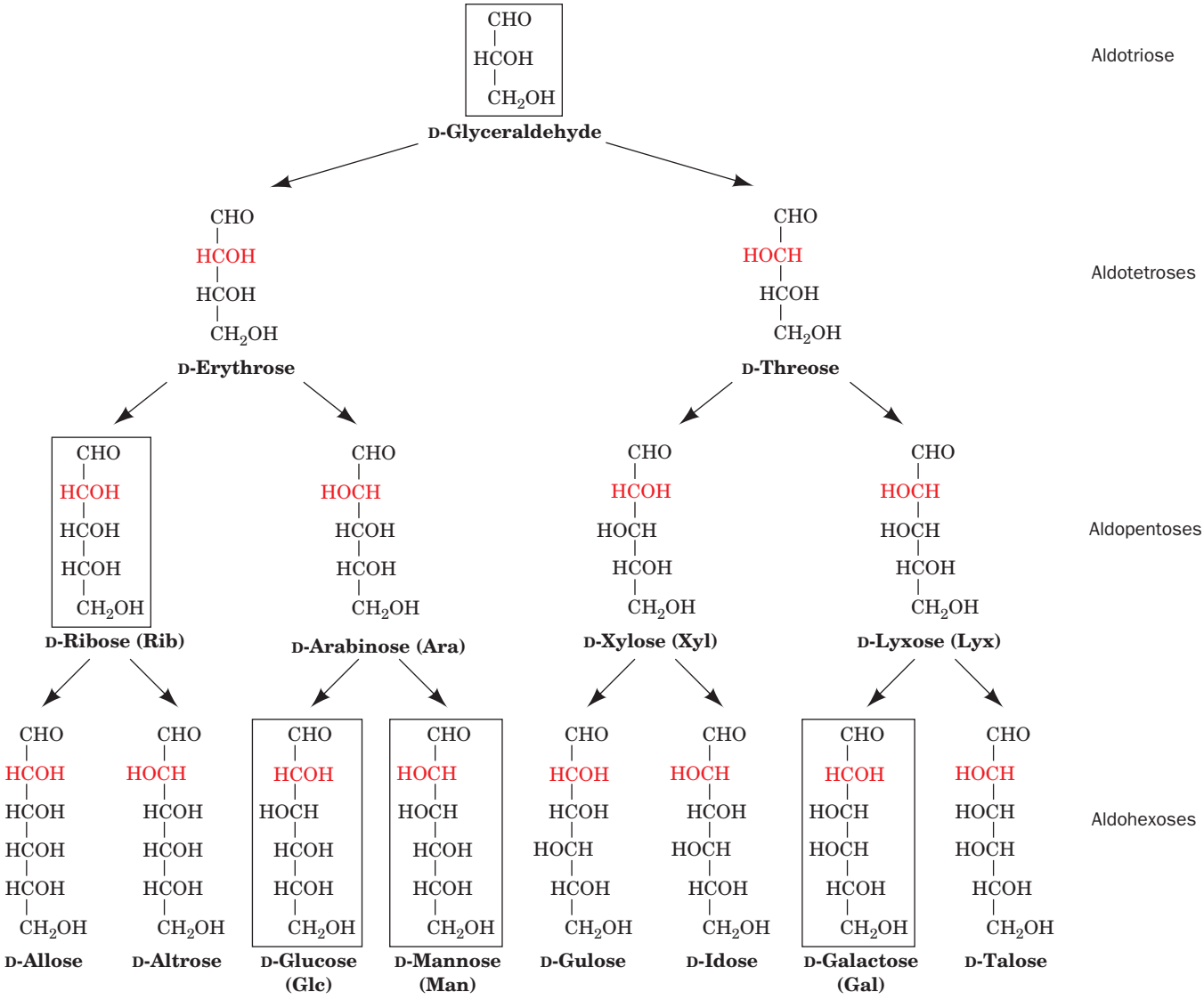


Figure 11-1 **The stereochemical relationships, shown in Fischer projection, among the D-aldoses with three to six carbon atoms.** The arrows indicate stereochemical relationships (not biosynthetic reactions). The configuration about C2 (*red*) distinguishes the members of each pair. The L- counterparts of these 15 sugars are their mirror images. The biologically most common aldoses are boxed.

Sugars that differ only by the configuration about one C atom are known as **epimers** of one another. Thus D-glucose and **D-mannose** are epimers with respect to C2, whereas D-glucose and **D-galactose** are epimers with respect to C4 (Fig. 11-1). However, D-mannose and D-galactose are not epimers of each other because they differ in configuration about two of their C atoms.

D-Glucose is the only aldose that commonly occurs in nature as a monosaccharide. However, it and several other monosaccharides including D-glyceraldehyde, D-ribose, D-mannose, and D-galactose are important components of larger biological molecules. L Sugars are biologically much less abundant than D sugars.

The position of their carbonyl group gives ketoses one less asymmetric center than their isomeric aldoses (e.g., compare D-fructose and D-glucose). *n*-Carbon ketoses therefore have 2^{n-3} stereoisomers. Those with their ketone function at C2 are the most common form (Fig. 11-2). Note that some of these ketoses are named by the insertion of *-ul-* before the suffix *-ose* in the name of the corresponding aldose; thus **D-xylulose** is the ketose corresponding to the aldose **D-xylose. Dihydroxyacetone, D-fructose,** D-ribulose, and D-xylulose are the biologically most prominent ketoses.

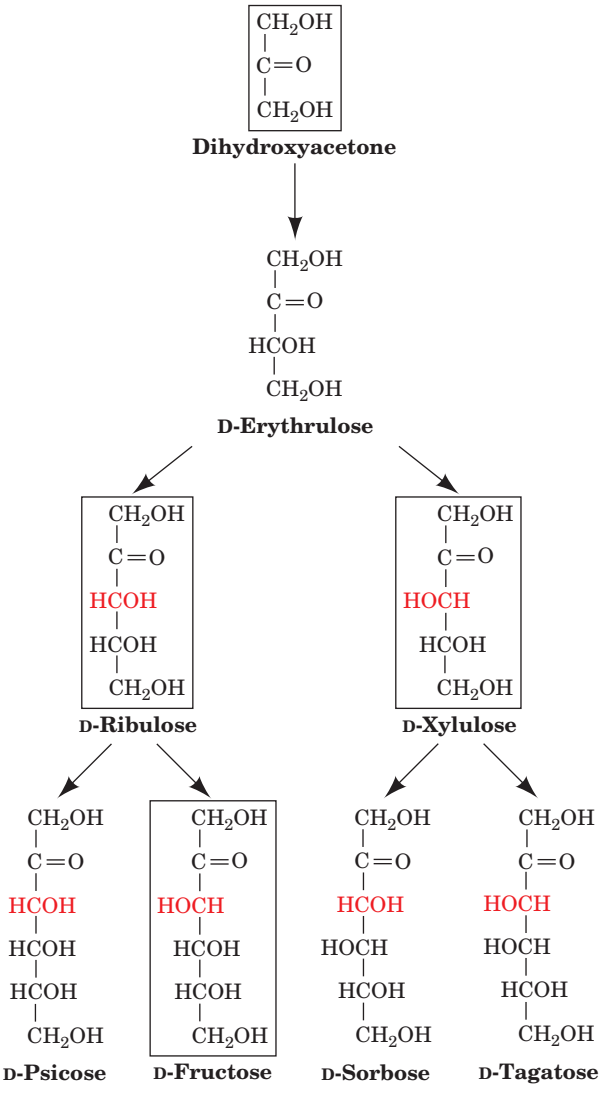
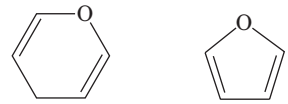


Figure 11-2 **The stereochemical relationships among the D-ketoses with three to six carbon atoms.** The configuration about C3 (*red*) distinguishes the members of each pair. The biologically most common ketoses are boxed.

B. *Configurations and Conformations*

Alcohols react with the carbonyl groups of aldehydes and ketones to form **hemiacetals** and **hemiketals,** respectively (Fig. 11-3). The hydroxyl and either the aldehyde or the ketone functions of monosaccharides can likewise react intramolecularly to form cyclic hemiacetals and hemiketals (Fig. 11-4). The configurations of the substituents to each carbon atom of these sugar rings are conveniently represented by their **Haworth projection formulas.**

A sugar with a six-membered ring is known as a **pyranose** in analogy with **pyran,** the simplest compound containing such a ring. Similarly, sugars with five-membered rings are designated **furanoses** in analogy with **furan.**



The cyclic forms of glucose and fructose with six- and five-membered rings are therefore known as **glucopyranose** and **fructofuranose,** respectively.

a. Cyclic Sugars Have Two Anomeric Forms

The Greek letters preceding the names in Fig. 11-4 still need to be explained. The cyclization of a monosaccharide renders the former carbonyl carbon asymmetric. The resulting pair of diastereomers are known as **anomers** and the hemiacetal or hemiketal carbon is referred to as the **anomeric** carbon. In the α anomer, the OH substituent to the anomeric carbon is on the opposite side of the sugar ring from the CH_2OH group at the chiral center that

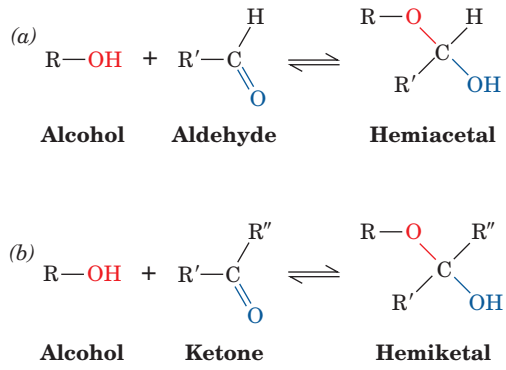


Figure 11-3 **The reactions of alcohols with (*a*) aldehydes to form hemiacetals and (*b*) ketones to form hemiketals.**

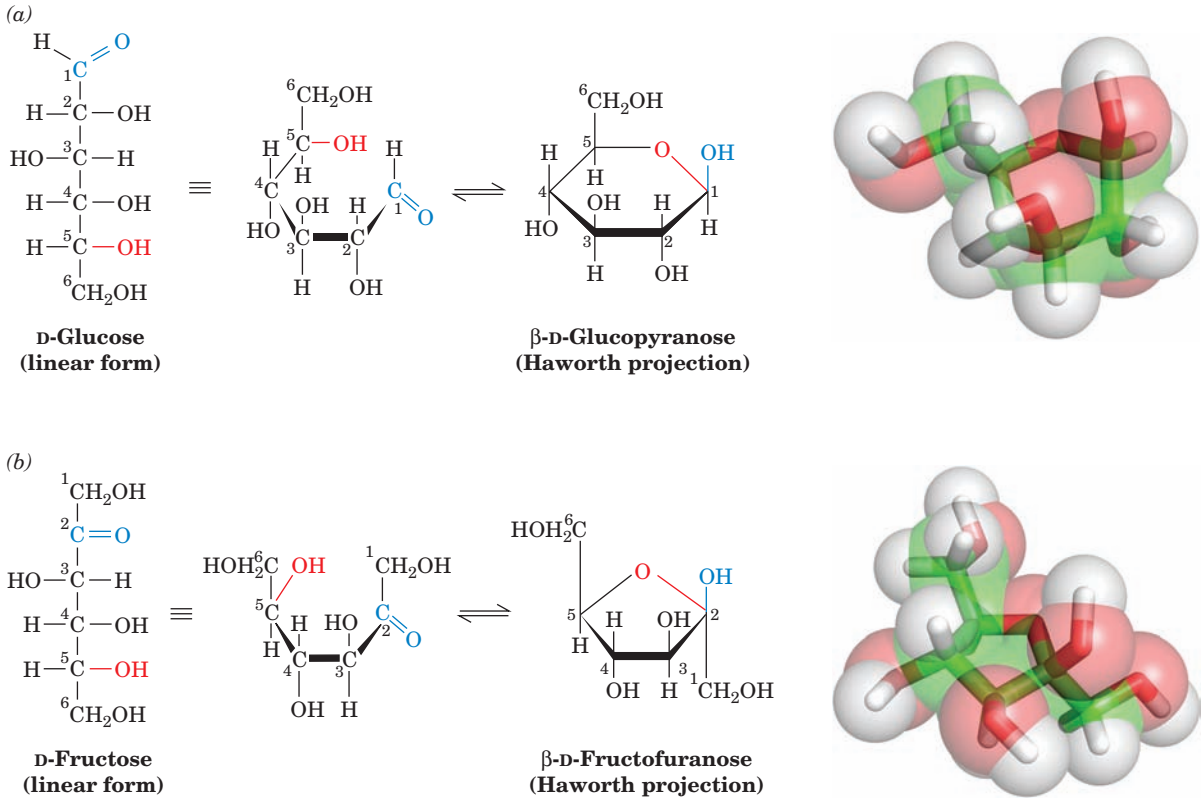


Figure 11-4 **Cyclization reactions for hexoses.** (*a*) D-Glucose in its linear form reacting to yield the cyclic hemiacetal β-D-glucopyranose and (*b*) D-fructose in its linear form reacting to yield the hemiketal β-D-fructofuranose. The cyclic sugars are shown both as Haworth projections and in stick form embedded in their semitransparent space-filling models with C green, H white, and O red.

designates the D or L configuration (C5 in hexoses). The other anomer is known as the β form (Fig. 11-5).

The two anomers of D-glucose, as any pair of diastereomers, have different physical and chemical properties. For example, the values of the specific optical rotation, $[\alpha]_D^{20}$, for α-D-glucose and β-D-glucose are, respectively, +112.2° and +18.7°. When either of these pure substances is dissolved in water, however, the specific optical rotation of the solution slowly changes until it reaches an equilibrium value of $[\alpha]_D^{20} = +52.7°$. This phenomenon is known as **mutarotation;** in glucose, it results from the formation of an equilibrium mixture consisting of 63.6% of the β anomer and 36.4% of the α anomer (the optical rotations of separate molecules in solution are independent of each other so that the optical rotation of a solution is the weighted average of the optical rotations of its components). The interconversion between these anomers occurs via the linear form of glucose (Fig. 11-5). Yet, since the linear forms of these monosaccharides are normally present in only minute amounts, these carbohydrates are

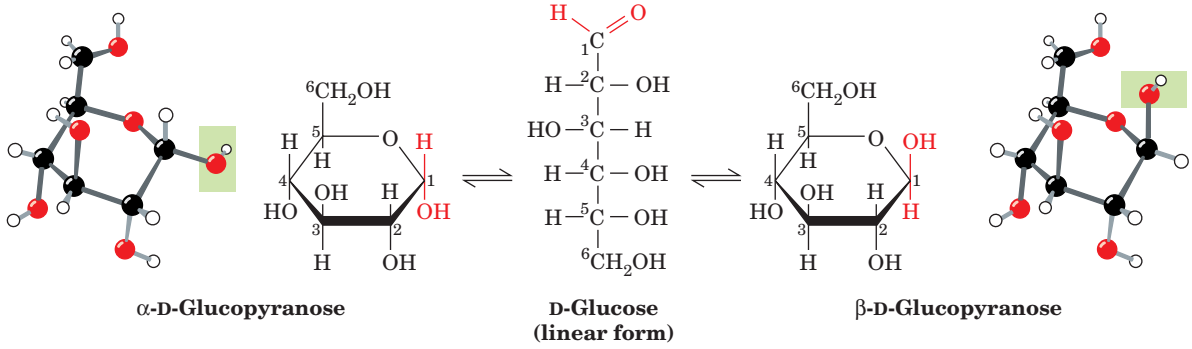


Figure 11-5 **The anomeric monosaccharides α-D-glucopyranose and β-D-glucopyranose, drawn as both Haworth projections and ball-and-stick models.** These pyranose sugars interconvert through the linear form of D-glucose and differ only by the configurations about their anomeric carbon atoms, C1.

See Kinemage Exercise 7-1

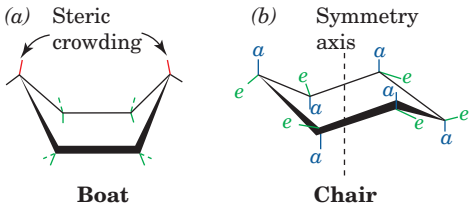


Figure 11-6 **Conformations of the cyclohexane ring.** (*a*) In the boat conformation, substituents at the "bow" and "stern" (*red*) are sterically crowded, whereas those along its sides (*green*) are eclipsed. (*b*) In the chair conformation, the substituents that extend parallel to the ring's threefold rotation axis are designated axial [*a*] and those that extend roughly outward from this symmetry axis are designated equatorial [*e*]. The equatorial substituents about the ring are staggered so that they alternately extend above and below the mean plane of the ring.

accurately described as cyclic polyhydroxy hemiacetals or hemiketals.

b. Sugars Are Conformationally Variable

Hexoses and pentoses may each assume pyranose or furanose forms. The equilibrium composition of a particular monosaccharide depends somewhat on conditions but mostly on the identity of the monosaccharide. For instance, NMR measurements indicate that whereas glucose almost exclusively assumes its pyranose form in aqueous solutions, fructose is 67% pyranose and 33% furanose, and ribose is 75% pyranose and 25% furanose (although in polysaccharides, glucose, fructose, and ribose residues are exclusively in their respective pyranose, furanose, and furanose forms). Although, in principle, hexoses and larger sugars can form rings of seven or more atoms, such rings are rarely observed because of the greater stabilities of the five- and six-membered rings that these sugars can also form. The internal strain of three- and four-membered sugar rings makes them unstable with respect to linear forms.

The use of Haworth formulas may lead to the erroneous impression that furanose and pyranose rings are planar. This cannot be the case, however, because all of the atoms in these rings are tetrahedrally (sp^3) hybridized. The pyranose ring, like the cyclohexane ring, may assume a **boat** or a **chair** conformation (Fig. 11-6). The relative stabilities of these various conformations depend on the stereochemical interactions between the substituents on the ring. The boat conformer crowds the substituents on its "bow" and "stern" and eclipses those along its sides, so that in cyclohexane it is ~25 kJ · mol^{-1} less stable than the chair conformer. The ring substituents on the chair conformer (Fig. 11-6*b*) fall into two geometrical classes: the rather close-fitting **axial** groups that extend parallel to the ring's threefold rotational axis and the staggered, and therefore minimally encumbered, **equatorial** groups. Since the axial and equatorial groups on a cyclohexane ring are conformationally interconvertible, a given ring has two alternative chair forms (Fig. 11-7); the one that predominates usually has the lesser crowding among its axial substituents. The conformational situation of a group directly affects its chemical reactivity. For example, equatorial OH groups on pyranoses esterify more readily than do axial OH groups. Note that β-D-glucose is the only D-aldohexose that can simultaneously have all five non-H substituents in the equatorial position (left side of Fig. 11-7). Perhaps this is why glucose is the most abundant naturally occurring monosaccharide. The conformational properties of furanose rings are discussed in Section 29-2Ab in relation to their effects on the conformations of nucleic acids.

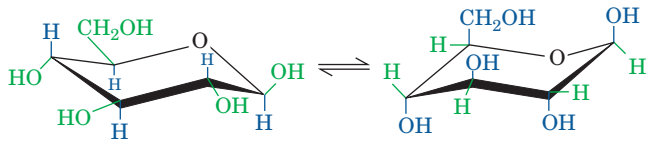


Figure 11-7 **The two alternative chair conformations of β-D-glucopyranose.** In the conformation on the left, which predominates, the relatively bulky OH and CH_2OH substituents all occupy equatorial positions, whereas in that on the right (drawn in ball-and-stick form in Fig. 11-5, *right*) they occupy the more crowded axial positions. **See Kinemage Exercise 7-1**

C. *Sugar Derivatives*

a. Polysaccharides Are Held Together by Glycosidic Bonds

The chemistry of monosaccharides is largely that of their hydroxy and carbonyl groups. For example, in an acid-catalyzed reaction, the anomeric hydroxyl of a sugar reversibly condenses with alcohols to form α- and β-**glycosides** (Greek: *glykys,* sweet) (Fig. 11-8). The bond connecting the anomeric carbon to the acetal oxygen is termed a **glycosidic**

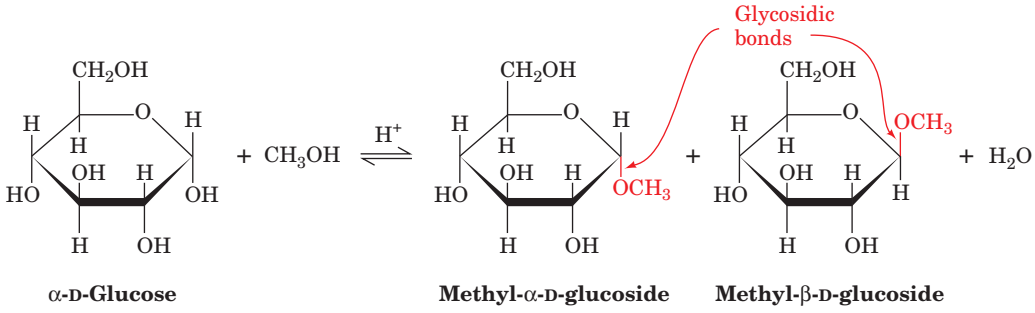


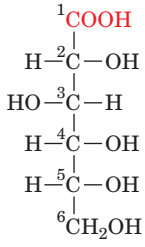
Figure 11-8 **The acid-catalyzed condensation of α-D-glucose with methanol to form an anomeric pair of methyl-D-glucosides.**

bond. *Polysaccharides are held together by glycosidic bonds between neighboring monosaccharide units.* The glycosidic bond is therefore the carbohydrate analog of the peptide bond in proteins. The bond in a nucleoside linking its ribose residue to its base is also a glycosidic bond (Section 5-1A).

The hydrolysis of glycosidic bonds is catalyzed by enzymes known as **glycosidases** that differ in specificity according to the identity and anomeric configuration of the glycoside but are often rather insensitive to the identity of the alcohol residue. Under basic or neutral conditions and in the absence of glycosidases, however, the glycosidic bond is stable, so glycosides do not undergo mutarotation as do monosaccharides. The methylation of the nonanomeric OH groups of monosaccharides requires more drastic conditions than is required for the formation of methyl glycosides, such as treatment with dimethyl sulfate.

b. Oxidation–Reduction Reactions

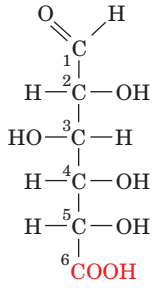
Because the cyclic and linear forms of aldoses and ketoses interconvert so readily, these sugars undergo reactions typical of aldehydes and ketones. Mild oxidation of an aldose, either chemically or enzymatically, results in the conversion of its aldehyde group to a carboxylic acid function, thereby yielding an **aldonic acid** such as **gluconic acid.** Aldonic acids are named by appending the suffix *-onic acid* to the root name of the parent aldose.



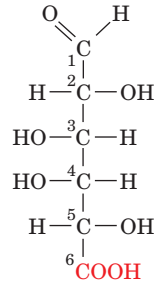
D-Gluconic acid

Saccharides bearing anomeric carbon atoms that have not formed glycosides are termed **reducing sugars** because of the facility with which the aldehyde group reduces mild oxidizing agents. A classic test for the presence of a reducing sugar is the reduction of Ag^+ in an ammonia solution **(Tollens' reagent)** to yield a metallic silver mirror lining on the inside of the reaction vessel.

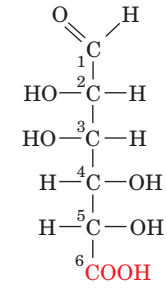
The specific oxidation of the primary alcohol group of aldoses yields **uronic acids,** which are named by appending *-uronic acid* to the root name of the parent aldose. **D-Glucuronic acid, D-galacturonic acid,** and **D-mannuronic acid** are important components of many polysaccharides.



D-Glucuronic acid



D-Galacturonic acid

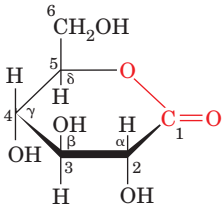


D-Mannuronic acid

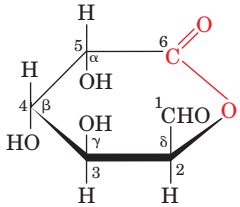
Uronic acids can assume the pyranose, furanose, and linear forms.

Both aldonic and uronic acids have a strong tendency to internally esterify so as to form five- and six-membered lactones (Fig. 11-9). **Ascorbic acid** (**vitamin C,** Fig. 11-10) is a γ-lactone that is synthesized by plants and almost all animals except primates and guinea pigs. Its prolonged deficiency in the diet of humans results in the disease known as **scurvy,** which is caused by the impairment of collagen formation (Section 8-2B). Scurvy generally results from a lack of fresh food. This is because, under physiological conditions, ascorbic acid is reversibly oxidized to **dehydroascorbic acid,** which, in turn, is irreversibly hydrolyzed to the vitamin-inactive **diketogulonic acid** (Fig. 11-10).

Aldoses and ketoses may be reduced under mild conditions, for example, by treatment with $NaBH_4$ to yield acyclic polyhydroxy alcohols known as **alditols,** which are named by appending the suffix *-itol* to the root name of the parent aldose. **Ribitol** is a component of flavin coenzymes (Section 16-2C), and **glycerol** and the cyclic polyhydroxy alcohol ***myo*-inositol** are important lipid components

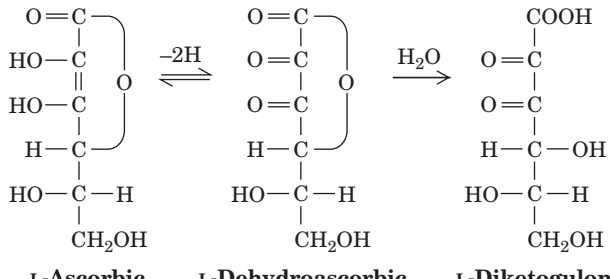


D-Glucono-δ-lactone



D-Glucurono-δ-lactone

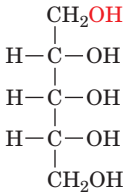
Figure 11-9 **D-Glucono-δ-lactone and D-glucurono-δ-lactone are, respectively, the lactones of D-gluconic acid and D-glucuronic acid.** The δ indicates that the O atom closing the lactone ring is also substituent to C_δ.



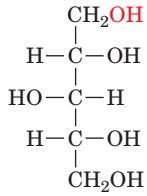
L-Ascorbic acid **L-Dehydroascorbic acid** **L-Diketogulonic acid**

Figure 11-10 **The reversible oxidation of L-ascorbic acid to L-dehydroascorbic acid.** This is followed by the physiologically irreversible hydrolysis of its lactone ring to form L-diketogulonic acid.

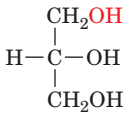
(Section 12-1). **Xylitol** is a sweetener that is used in "sugarless" gum and candies.



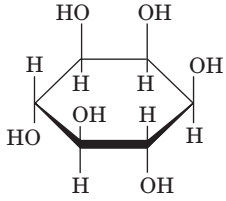
Ribitol



Xylitol



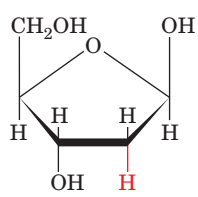
Glycerol



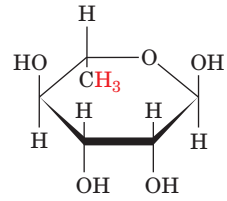
***myo*-Inositol**

c. Other Biologically Important Sugar Derivatives

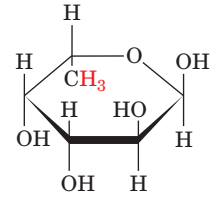
Monosaccharide units in which an OH group is replaced by H are known as **deoxy sugars.** The biologically most important of these is **β-D-2-deoxyribose,** the sugar component of DNA's sugar–phosphate backbone (Section 5-1A). **L-Rhamnose** and **L-fucose** are widely occurring polysaccharide components.



β-D-2-Deoxyribose

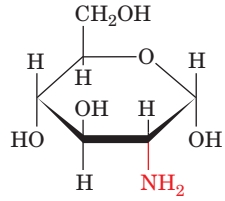


α-L-Rhamnose (6-deoxy-L-mannose)

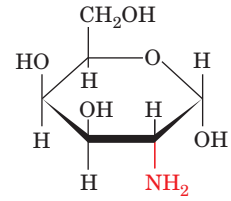


α-L-Fucose (6-deoxy-L-galactose)

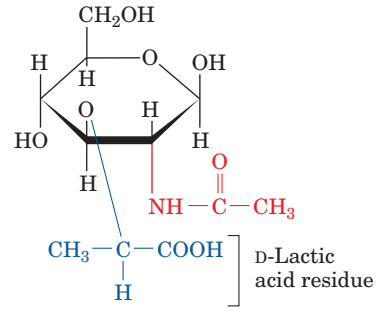
In **amino sugars,** one or more OH groups are replaced by an often acetylated amino group. **D-Glucosamine** and **D-galactosamine** are components of numerous biologically important polysaccharides.



α-D-Glucosamine (2-amino-2-deoxy-α-D-glucopyranose)

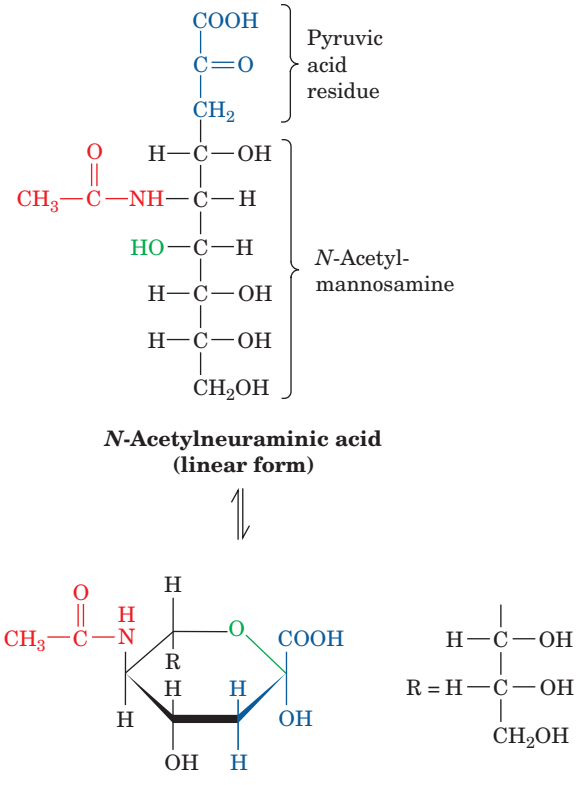


α-D-Galactosamine (2-amino-2-deoxy-α-D-galactopyranose)



***N*-Acetylmuramic acid (NAM)**

The amino sugar derivative ***N*-acetylmuramic acid,** which consists of ***N*-acetyl-D-glucosamine** in an ether linkage with **D-lactic acid,** is a prominent component of bacterial cell walls (Section 11-3Ba). *N*-Acetylneuraminic acid, which is derived from *N*-acetylmannosamine and pyruvic acid (Fig. 11-11), is an important constituent of glycoproteins (Section 11-3C) and glycolipids (Section 12-1D). *N*-Acetylneuraminic acid and its derivatives are often referred to as **sialic acids.**



***N*-Acetylneuraminic acid (linear form)**

***N*-Acetylneuraminic acid (pyranose form)**

Figure 11-11 ***N*-Acetylneuraminic acid in its linear and pyranose forms.** Note that its pyranose ring incorporates the pyruvic acid residue (*blue*) and part of the mannose moiety.

2 POLYSACCHARIDES

Polysaccharides, which are also known as **glycans,** consist of monosaccharides linked together by glycosidic bonds. They are classified as **homopolysaccharides** or **heteropolysaccharides** if they consist of one type or more than one type of monosaccharide residue. Homopolysaccharides may be further classified according to the identity of

their monomeric unit. For example, **glucans** are polymers of glucose, whereas **galactans** are polymers of galactose. Although monosaccharide sequences of heteropolysaccharides can, in principle, be as varied as those of proteins, they are usually composed of only a few types of monosaccharides that alternate in a repetitive sequence.

Polysaccharides, in contrast to proteins and nucleic acids, form branched as well as linear polymers. This is because glycosidic linkages can be made to any of the hydroxyls of a monosaccharide. Fortunately for structural biochemists, many polysaccharides are linear and those that branch tend to do so in only a few well-defined ways.

In this section, we discuss the structures of the simplest polysaccharides, the disaccharides, and then consider the structures and properties of the most abundant classes of polysaccharides. We begin by outlining how polysaccharide structures are elucidated.

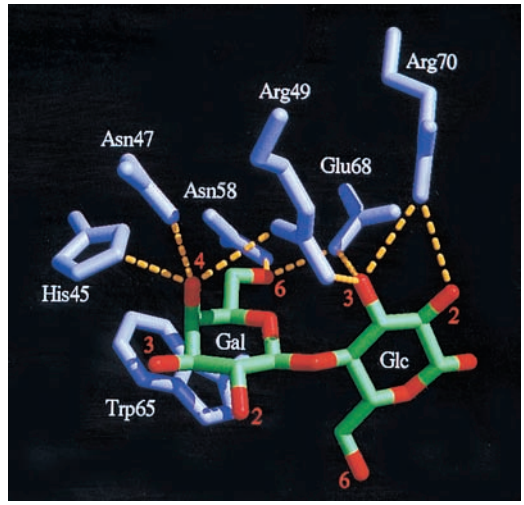


Figure 11-12 **Carbohydrate binding by a lectin in the X-ray structure of human galectin-2 in complex with the disaccharide lactose.** This lectin primarily binds β-D-galactose residues. The structure is drawn in stick form with the C and O atoms of lactose's galactose (Gal) and glucose (Glc) residues green and red, and the galectin-2 amino acid side chains violet. Hydrogen bonds between the protein side chains and the sugar residues are represented by dashed yellow lines. [Courtesy of Hakon Leffler, Lund University, Sweden. PDBid 1HLC.]

A. *Carbohydrate Analysis*

The purification of carbohydrates can, by and large, be effected by chromatographic and electrophoretic procedures similar to those used in protein purification (Sections 6-3 and 6-4), although thin layer chromatography (TLC; Section 6-3Dd) is also widely used. Affinity chromatography (Section 6-3C), using immobilized proteins known as **lectins** (Latin: *legere,* to pick or choose), is a particularly powerful technique in this regard. Lectins are sugar-binding proteins that were discovered in plants but are now known to occur in all organisms, where they participate in a wide variety of signaling, cell–cell recognition, and adhesion processes, as well as in targeting newly synthethesized proteins to specific cellular locations (Section 12-4Cg). Lectins recognize one or more specific monosaccharides with particular linkages to other sugars in oligosaccharides, usually with exquisite specificity. Their protein–carbohydrate interactions typically include multiple hydrogen bonds, which often include bridging water molecules, and the packing of hydrophobic sugar faces against aromatic side chains (Fig. 11-12). Among the best characterized lectins are jack bean **concanavalin A** (Fig. 8-40), which specifically binds α-D-glucose and α-D-mannose residues, and wheat germ **agglutinin** (so named because it causes cells to agglutinate or clump together), which specifically binds β-*N*-acetylmuramic acid and α-*N*-acetylneuraminic acid.

Characterization of an oligosaccharide requires that the identities, anomers, linkages, and order of its component monosaccharides be elucidated. The linkages of the monosaccharides may be determined by **methylation analysis** (also called **permethylation analysis),** a technique pioneered by Norman Haworth in the 1930s. *Methyl ethers not at the anomeric C atom are resistant to acid hydrolysis but glycosidic bonds are not. Consequently, if an oligosaccharide is exhaustively methylated and then hydrolyzed, the free OH groups on the resulting methylated monosaccharides mark the former positions of the glycosidic bonds.* Methylated monosaccharides are often identified by **gas–liquid chromatography** (**GLC;** a technique in which the stationary phase is an inert solid, such as diatomaceous earth, impregnated with a low-volatility liquid, such as silicone oil, and the mobile phase is an inert gas, such as He, into which the sample has been flash evaporated) combined with mass spectrometry (GLC/MS). HPLC techniques may similarly be used. Other mass spectrometric techniques for analyzing nonvolatile substances are discussed in Section 7-1I. Although all aldoses and ketoses with the same number of C atoms are isomers (Figs. 11-1 and 11-2) and hence have identical molecular masses, they have characteristic fragmentation patterns.

The sequence and anomeric configurations of the monosaccharides in an oligosaccharide can be determined through the use of specific **exoglycosidases.** These enzymes specifically hydrolyze their corresponding monosaccharides from the nonreducing ends of oligosaccharides (the ends lacking a free anomeric carbon atom) in a manner analogous to the actions of exopeptidases on proteins (Section 7-1Ab). For example, **β-galactosidase** excises the terminal β anomers of galactose, whereas **α-mannosidase** does so with the α anomers of mannose. Some of these exoglycosidases also exhibit specificity for the **aglycone,** the sugar chains to which the monosaccharide to be excised (the **glycone**) is linked. Through the use of mass spectrometry, the sequence of a polysaccharide may be deduced from the mass decrements generated by exoglycosidases. The use of **endoglycosidases** (hydrolases that cleave glycosidic bonds between nonterminal sugar residues) of varying specificities can also supply useful sequence information. The proton and ^{13}C NMR spectra of oligosaccharides can provide the complete sequence of an oligosaccharide if sufficient material is available. Moreover, two-dimensional

NMR techniques (Section 8-3Ac) can reveal oligosaccharide structures (e.g., see Section 11-2Eb).

B. *Disaccharides*

We begin our studies of polysaccharides by considering disaccharides (Fig. 11-13). **Sucrose,** the most abundant disaccharide, occurs throughout the plant kingdom and is familiar to us as common table sugar. Its structure (Fig. 11-13) was established by methylation analysis as described above and was later confirmed by its X-ray structure. To name a polysaccharide systematically, one must specify its component monosaccharides, their ring types, their anomeric forms, and how they are linked together. Sucrose is therefore *O*-α-D-glucopyranosyl-(1 → 2)-β-D-fructofuranoside, where the symbol (1 → 2) indicates that the glycosidic bond links C1 of the glucose residue to C2 of the fructose residue. Note that since these two positions are the anomeric carbon atoms of their respective monosaccharides, sucrose is not a reducing sugar (as the suffix *-ide* implies).

The hydrolysis of sucrose to D-glucose and D-fructose is accompanied by a change in optical rotation from *dextro* to *levo.* Consequently, hydrolyzed sucrose is sometimes called **invert sugar** and the enzyme that catalyzes this process, **β-D-fructofuranosidase,** is archaically named **invertase.**

Lactose [*O*-β-D-galactopyranosyl-(1 → 4)-D-glucopyranose] or milk sugar (Fig. 11-13) occurs naturally only in milk, where its concentration ranges from 0 to 7% depending on the species. The free anomeric carbon of its glucose residue makes lactose a reducing sugar.

Infants normally express the intestinal enzyme **β-D-galactosidase** or **lactase** that catalyzes the hydrolysis of lactose to its component monosaccharides for absorption into the bloodstream. Many adults, however, including most Africans and almost all Asians, have a low level of this enzyme (as do most adult mammals, since they normally do not encounter milk). Consequently, much of the lactose in any milk they drink moves through their digestive tract to the colon, where its bacterial fermentation produces large quantities of CO_2, H_2, and irritating organic acids. This results in an embarrassing and often painful digestive upset termed **lactose intolerance.** Perhaps this is why Chinese cuisine, which is noted for the wide variety of foodstuffs it employs, is devoid of milk products. However, adult members of populations with a tradition of herding cattle, mainly northern Europeans and certain African groups, continue expressing the lactase gene and hence can drink milk without a problem. Modern food technology has come to the aid of milk lovers who develop lactose intolerance: Milk products in which the lactose has been hydrolyzed enzymatically and lactase-containing pills are now widely available.

There are several common glucosyl–glucose disaccharides. These include **maltose** [*O*-α-D-glucopyranosyl-(1 → 4)-D-glucopyranose], an enzymatic hydrolysis product of starch; **isomaltose,** its α(1 → 6) isomer; and **cellobiose,** its β(1 → 4) isomer, the repeating disaccharide of cellulose. Only a few tri- or higher oligosaccharides occur in significant natural abundance. Not surprisingly, they all occur in plants.

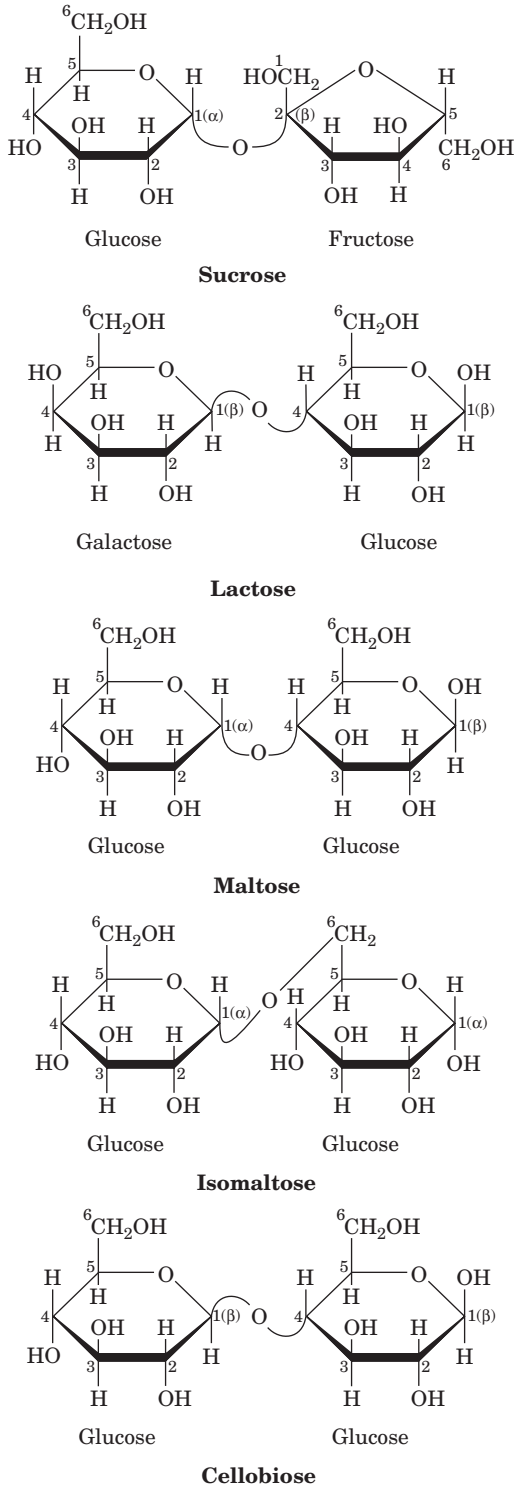


Figure 11-13 **Several common disaccharides.** **See Kinemage Exercise 7-2**

C. *Structural Polysaccharides: Cellulose and Chitin*

Plants have rigid cell walls (Fig. 1-9) that, in order to maintain their shapes, must be able to withstand osmotic pressure differences between the extracellular and intracellular spaces of up to 20 atm. In large plants, such as trees, the cell

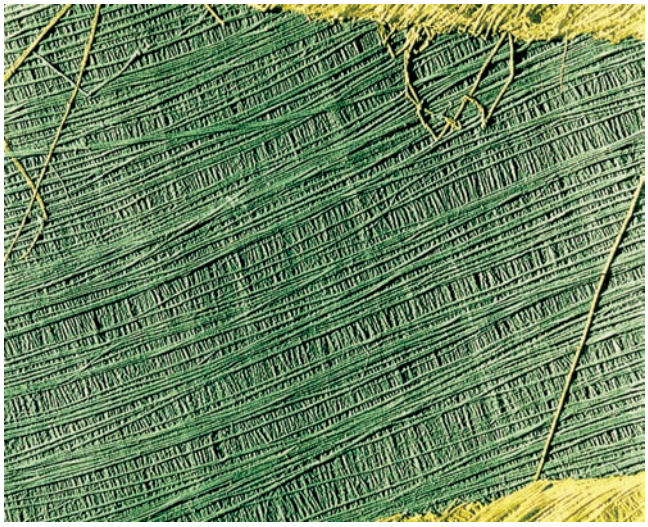
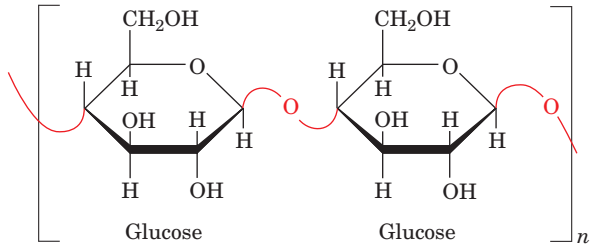


Figure 11-14 **Electron micrograph of the cellulose fibers in the cell wall of the alga *Chaetomorpha melagonium*.** Note that the cell wall consists of layers of parallel fibers. [Biophoto Associates/Photo Researchers.]



Cellulose

Figure 11-15 **The primary structure of cellulose.** Here *n* may be several thousand.

walls also have a load-bearing function. Cellulose, the primary structural component of plant cell walls (Fig. 11-14), accounts for over half of the carbon in the biosphere: $\sim 10^{15}$ kg of cellulose are estimated to be synthesized and degraded annually. Although cellulose is predominantly of vegetable origin, it also occurs in the stiff outer mantles of marine invertebrates known as **tunicates** (urochordates; Fig. 1-11).

The primary structure of cellulose was determined through methylation analysis. Cellulose is a linear polymer of up to 15,000 D-glucose residues (a glucan) linked by $\beta(1 \rightarrow 4)$ glycosidic bonds (Fig. 11-15). As is generally true of large polysaccharides, it has no defined size since, in contrast to proteins and nucleic acids, there is no genetically determined template that directs its synthesis.

X-ray studies of cellulose fibers led Anatole Sarko to tentatively propose the model diagrammed in Fig. 11-16. This highly cohesive, hydrogen bonded structure gives cellulose fibers exceptional strength and makes them water insoluble despite their hydrophilicity.

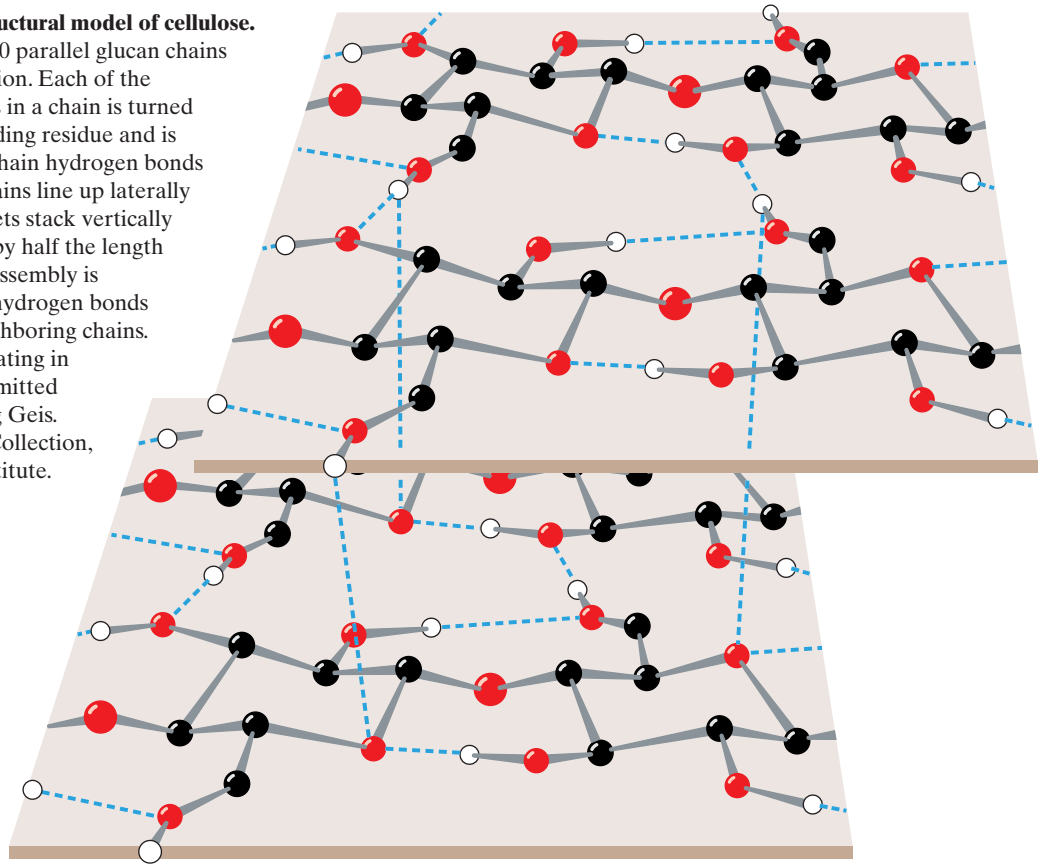
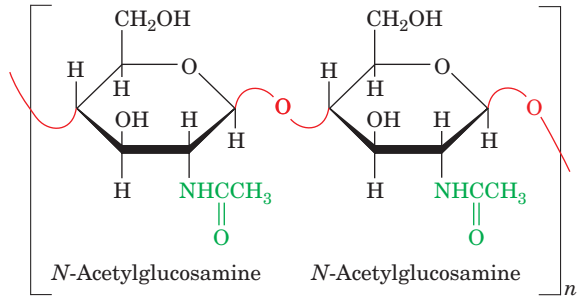


Figure 11-16 **Proposed structural model of cellulose.** Cellulose fibers consist of ~40 parallel glucan chains arranged in an extended fashion. Each of the $\beta(1 \rightarrow 4)$-linked glucose units in a chain is turned 180° with respect to its preceding residue and is held in this position by intrachain hydrogen bonds (*dashed lines*). The glucan chains line up laterally to form sheets, and these sheets stack vertically such that they are staggered by half the length of a glucose unit. The entire assembly is stabilized by intermolecular hydrogen bonds between glucose units of neighboring chains. Hydrogen atoms not participating in hydrogen bonds have been omitted for clarity. [Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission.]

In plant cell walls, the cellulose fibers are embedded in and cross-linked by a matrix of several polysaccharides that are composed of glucose as well as other monosaccharides. In wood, this cementing matrix also contains a large proportion of **lignin,** a plasticlike phenolic polymer. One has only to watch a tall tree in a high wind to realize the enormous strength of plant cell walls. In engineering terms, they are "composite materials," as is concrete reinforced by steel rods. Composite materials can withstand large stresses because the matrix evenly distributes the stresses among the reinforcing elements.

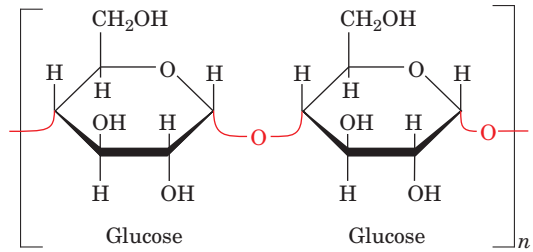
Although vertebrates themselves do not possess an enzyme capable of hydrolyzing the β(1 → 4) linkages of cellulose, the digestive tracts of herbivores contain symbiotic microorganisms that secrete a series of enzymes, collectively known as **cellulase,** that do so. The same is true of termites. Nevertheless, the degradation of cellulose is a slow process because its tightly packed and hydrogen bonded glucan chains are not easily accessible to cellulase and do not separate readily even after many of their glycosidic bonds have been hydrolyzed. The digestion of fibrous plants such as grass by herbivores is therefore a more complex and time-consuming process than is the digestion of meat by carnivores (cows, e.g., have multichambered stomachs and must chew their cud). Similarly, the decay of dead plants by fungi, bacteria, and other organisms, and the consumption of wooden houses by termites, often takes years.

Chitin is the principal structural component of the exoskeletons of invertebrates such as crustaceans, insects, and spiders and is also a major cell wall constituent of most fungi and many algae. It is estimated that ~10^{14} kg of chitin are produced annually, most of it in the oceans, and therefore that it is almost as abundant as is cellulose. Chitin is a homopolymer of β(1 → 4)-linked *N*-acetyl-D-glucosamine residue (Fig. 11-17). It differs chemically from cellulose only in that each C2-OH group is replaced by an acetamido function. X-ray analysis indicates that chitin and cellulose have similar structures.

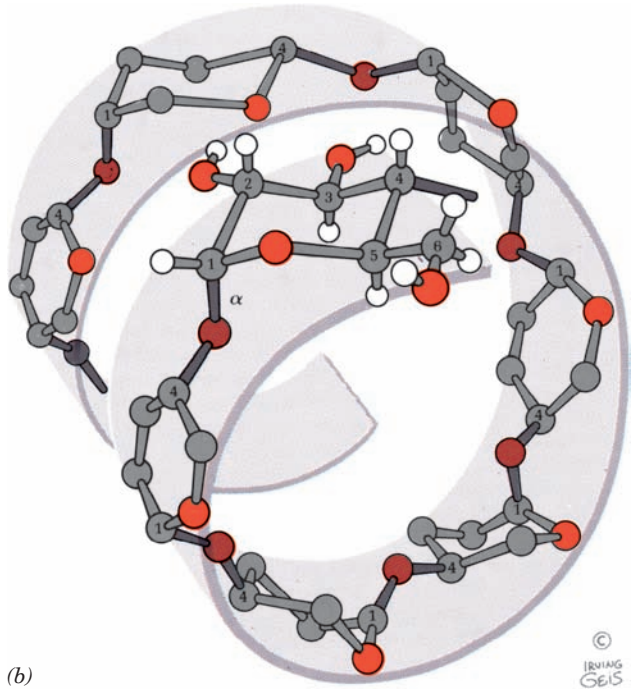


Chitin

Figure 11-17 **The primary structure of chitin.** Chitin is a β(1 → 4)-linked homopolymer of *N*-acetyl-D-glucosamine.



(a) **α-Amylose**



(b)

Figure 11-18 **α-Amylose.** (*a*) The D-glucose residues of α-amylose are linked by α(1 → 4) bonds (*red*). Here *n* is several thousand. (*b*) This regularly repeating polymer forms a left-handed helix with ~6 glucose residues per turn. Note the great differences in structure and properties that result from changing α-amylose's α(1 → 4) linkages to the β(1 → 4) linkages of cellulose (Fig. 11-16). [Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission.]

D. *Storage Polysaccharides: Starch and Glycogen*

a. Starch Is a Food Reserve in Plants and a Major Nutrient for Animals

Starch is a mixture of glucans that plants synthesize as their principal food reserve. It is deposited in the cytoplasm of plant cells as insoluble granules composed of **α-amylose** and **amylopectin.** α-Amylose is a linear polymer of several thousand glucose residues linked by α(1 → 4) bonds (Fig. 11-18*a*). Note that although α-amylose is an isomer of cellulose, it has very different structural properties. This is because cellulose's β-glycosidic linkages cause each successive glucose residue to flip 180° with respect to the preceding residue, so that the polymer assumes an easily packed, fully extended conformation (Fig. 11-16). In contrast, α-amylose's α-glycosidic bonds cause it to adopt an irregularly aggregating helically coiled conformation (Fig. 11-18*b*).

Amylopectin consists mainly of α(1 → 4)-linked glucose residues but is a branched molecule with α(1 → 6) branch

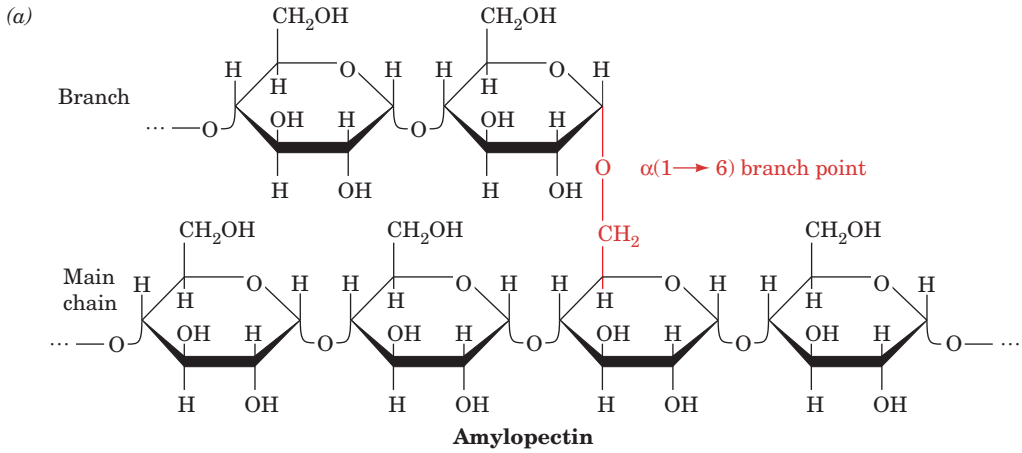


Figure 11-19 **Amylopectin.** (*a*) Its primary structure near one of its α(1 → 6) branch points (*red*). (*b*) Its bushlike structure with glucose residues at branch points indicated in red. The actual distance between branch points averages 24 to 30 glucose residues. Glycogen has a similar structure but is branched every 8 to 14 residues.

points every 24 to 30 glucose residues on average (Fig. 11-19). Amylopectin molecules contain up to 10^6 glucose residues, which makes them among the largest molecules occurring in nature. The storage of glucose as starch greatly reduces the large intracellular osmotic pressures that would result from its storage in monomeric form because osmotic pressure is proportional to the number of solute molecules in a given volume.

b. Starch Digestion Occurs in Stages

The digestion of starch, the main carbohydrate source in the human diet, begins in the mouth. Saliva contains **α-amylase,** which randomly hydrolyzes all the α(1 → 4) glucosidic bonds of starch except its outermost bonds and those next to branches. By the time thoroughly chewed food reaches the stomach, where the acidity inactivates α-amylase, the average chain length of starch has been reduced from several thousand to fewer than eight glucose units. Starch digestion continues in the small intestine under the influence of pancreatic α-amylase, which is similar to the salivary enzyme. This enzyme degrades starch to a mixture of the disaccharide maltose, the trisaccharide **maltotriose,** which contains three α(1 → 4)-linked glucose residues, and oligosaccharides known as **dextrins** that contain the α(1 → 6) branches. These oligosaccharides are hydrolyzed to their component monosaccharides by specific enzymes contained in the brush border membranes of the intestinal mucosa: an **α-glucosidase,** which removes one glucose residue at a time from oligosaccharides, an **α-dextrinase** or **debranching enzyme,** which hydrolyzes α(1 → 6) and α(1 → 4) bonds, a **sucrase,** and, at least in infants, a lactase. The resulting monosaccharides are absorbed by the intestine and transported to the bloodstream (Section 20-4A).

c. Glycogen Is "Animal Starch"

Glycogen, the storage polysaccharide of animals, is present in all cells but is most prevalent in skeletal muscle and liver, where it occurs as cytoplasmic granules (Fig. 11-20). The primary structure of glycogen resembles that of amylopectin, but glycogen is more highly branched, with branch points occurring every 8 to 14 glucose residues. Glycogen's degree of polymerization is nevertheless similar to that of amylopectin. In the cell, glycogen is degraded for metabolic use by **glycogen phosphorylase,** which phosphorolytically cleaves glycogen's α(1 → 4) bonds sequentially inward from its nonreducing ends to yield **glucose-1-phosphate.** Glycogen's highly branched structure, which has many nonreducing ends, permits the rapid mobilization of glucose in times of metabolic need. The α(1 → 6) branches of glycogen are cleaved by a debranching enzyme. These enzymes play an important role in glucose metabolism and are discussed further in Section 18-1.

E. *Glycosaminoglycans*

The extracellular spaces, particularly those of connective tissues such as cartilage, tendon, skin, and blood vessel walls, consist of collagen and elastin fibers (Section 8-2B)

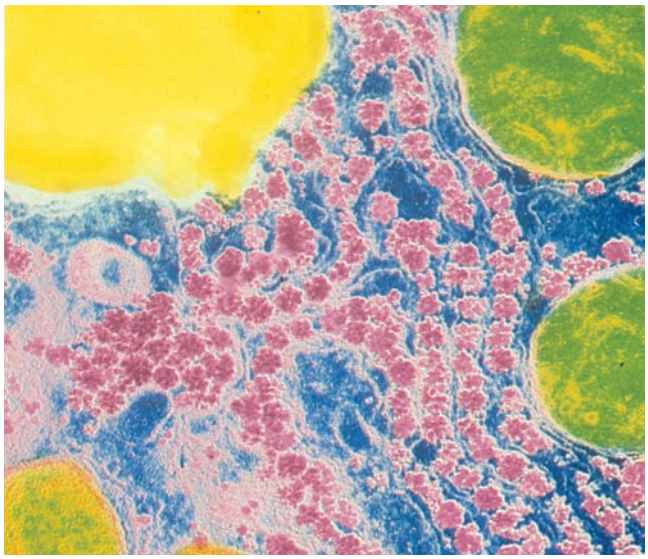
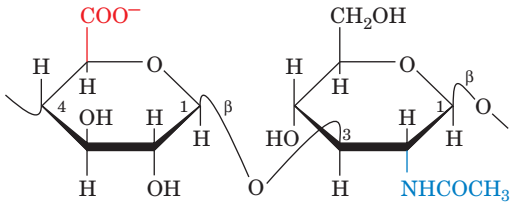


Figure 11-20 **Photomicrograph showing the glycogen granules (*pink*) in the cytoplasm of a liver cell.** The greenish objects are mitochondria and the yellow object is a fat globule. Note that the glycogen granules tend to aggregate. The glycogen content of liver may reach as high as 10% of its net weight. [CNRI/Science Photo Library/Photo Researchers, Inc.]

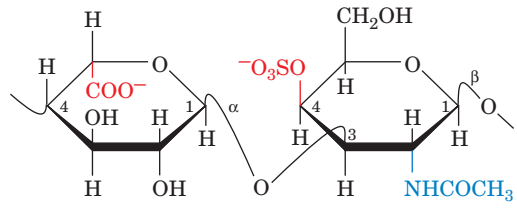
embedded in a gel-like matrix known as **ground substance.** Ground substance is composed largely of **glycosaminoglycans (GAGs;** alternatively, **mucopolysaccharides**), unbranched polysaccharides of alternating uronic acid and hexosamine residues. Solutions of GAGs have a slimy, mucuslike consistency that results from their high viscosity and elasticity. In the following paragraphs, we discuss the structural origin of these important mechanical properties.

a. Hyaluronic Acid

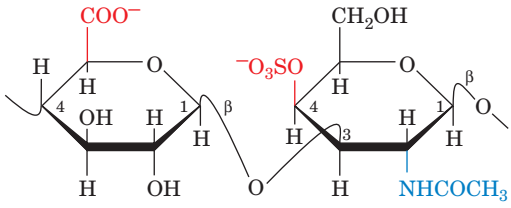
Hyaluronic acid (also called **hyaluronan**) is an important GAG component of ground substance, synovial fluid (the fluid that lubricates the joints), and the vitreous humor of the eye. It also occurs in the capsules surrounding certain, usually pathogenic, bacteria. Hyaluronic acid molecules are composed of 250 to 25,000 β(1 → 4)-linked disaccharide units that consist of D-glucuronic acid and *N*-acetyl-D-glucosamine linked by a β(1 → 3) bond (Fig. 11-21). The anionic character of its glucuronic acid residues causes hyaluronic acid to bind cations such as K^+, Na^+, and Ca^{2+} tightly. X-ray fiber analysis indicates that Ca^{2+} hyaluronate



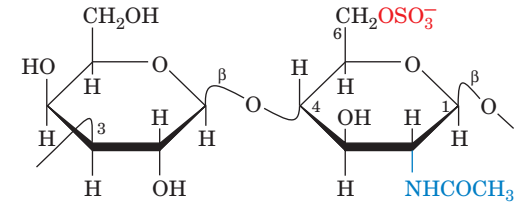
Hyaluronate



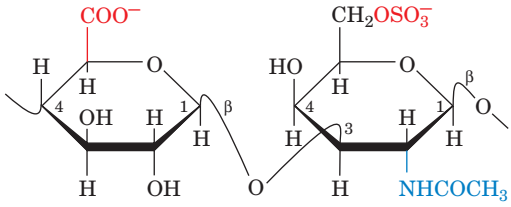
Dermatan sulfate



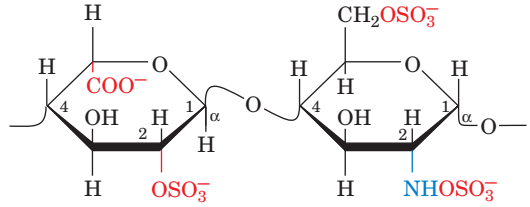
Chondroitin-4-sulfate



Keratan sulfate



Chondroitin-6-sulfate



Heparin

Figure 11-21 **The disaccharide repeating units of the common glycosaminoglycans.** The anionic groups are drawn in red and the *N*-acetylamido groups are drawn in blue. **See Kinemage Exercise 7-3**

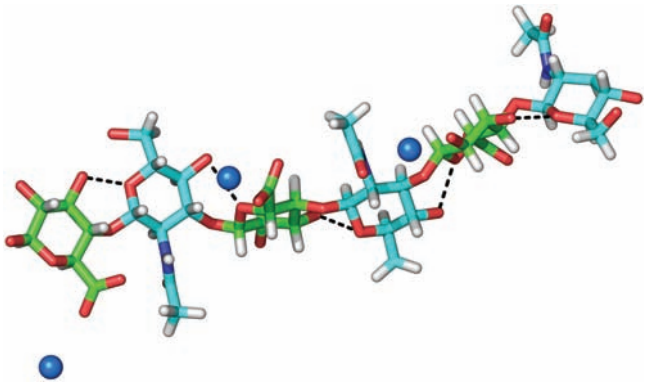


Figure 11-22 **X-ray fiber structure of Ca^{2+} hyaluronate.** Three consecutive disaccharide units of the hyaluronate fiber are drawn in stick form with atoms colored according to type with glucuronate C green, *N*-acetyl-D-glucosamine C cyan, H white, N blue, and O red. Ca^{2+} ions are represented by blue spheres. The hyaluronate polyanion forms an extended, left-handed, single-stranded helix with a pitch of 28.3 Å and ~3 disaccharide units per turn that is stabilized by intramolecular hydrogen bonds (*dashed lines*). The positions of the H atoms are inferred and hence the H atoms of the OH groups are not shown. [Based on a fiber X-ray structure by Struther Arnott, Purdue University. PDBid 4HYA.]

forms an extended, left-handed, single-stranded helix with ~3 disaccharide units per turn (Fig. 11-22).

Hyaluronate's structural features suit it to its biological function. Its high molecular mass and numerous mutually repelling anionic groups make hyaluronate an extended, rigid, and highly hydrated molecule which, in solution, occupies a volume ~1000 times that in its dry state. Hyaluronate solutions therefore have a viscosity that is shear dependent (an object under shear stress has equal and opposite forces applied across its opposite faces). At low shear rates, the hyaluronate molecules form tangled masses that greatly impede flow; that is, the solution is quite viscous. As the shear rate increases, the stiff rodlike hyaluronate molecules tend to line up with the flow and thus offer less resistance to it. This viscoelastic behavior makes hyaluronate solutions excellent biological shock absorbers and lubricants.

Hyaluronic acid and other GAGs (see below) are degraded by **hyaluronidase,** which hydrolyzes their β(1 → 4) linkages. Hyaluronidase occurs in a variety of animal tissues, in bacteria (where it presumably expedites their invasion of animal tissue), and in snake and insect toxins.

b. Other Glycosaminoglycans

Other GAG components of ground substance consist of 50 to 1000 sulfated disaccharide units which occur in proportions that are both tissue and species dependent. The most prevalent structures of these generally heterogeneous substances are (Fig. 11-21)

1. Chondroitin-4-sulfate (Greek: *chondros,* cartilage), a major component of cartilage and other connective tissue, has *N*-acetyl-D-galactosamine-4-sulfate residues in place of hyaluronate's *N*-acetyl-D-glucosamine residues.

2. Chondroitin-6-sulfate is instead sulfated at the C6 position of its *N*-acetyl-D-galactosamine residues. The two chondroitin sulfates occur separately or in mixtures depending on the tissue.

3. Dermatan sulfate (Greek: *derma,* skin), which is so named because of its prevalence in skin, differs from chondroitin-4-sulfate only by an inversion of configuration about C5 of the β-D-glucuronate residues to form α-L-iduronate. This results from the enzymatic epimerization of these residues after the formation of chondroitin. The epimerization is usually incomplete, so dermatan sulfate also contains glucuronate residues.

4. Keratan sulfate (Greek: *keras,* horn; not to be confused with the protein keratin) consists mainly of alternating β(1 → 4)-linked D-galactose and *N*-acetyl-D-glucosamine-6-sulfate residues (and hence lacks uronic acid residues). It is a component of cartilage, bone, cornea, as well as hair, nails, and horn. Keratan sulfate is the most heterogeneous of the major GAGs in that its sulfate content is variable and it contains small amounts of fucose, mannose, *N*-acetylglucosamine, and sialic acid.

5. Heparin is a variably sulfated GAG that consists predominantly of alternating α(1 → 4)-linked residues of L-iduronate-2-sulfate and *N*-sulfo-D-glucosamine-6-sulfate. It has an average of 2.5 sulfate residues per disaccharide unit, which makes it the most negatively charged polyelectrolyte in mammalian tissues (Fig. 11-23). Heparin, in contrast to the above GAGs, is not a constituent of connective tissue, but occurs almost exclusively in the intracellular granules of the **mast cells** that line arterial walls, especially in the liver, lungs, and skin. It inhibits the clotting of blood, and its release, through injury, is thought to prevent runaway clot formation (Section 35-1Ea). Heparin is therefore in wide clinical use to inhibit blood clotting, for example, in

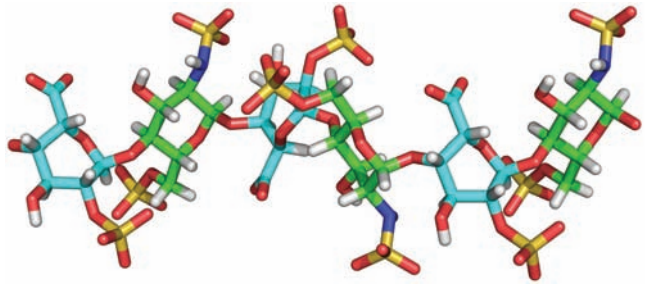


Figure 11-23 **NMR structure of heparin.** Three consecutive disaccharide units of this helical polymer are shown in stick form. Atoms are colored according to type with glucosamine C green, iduronate C cyan, H white, N blue, O red, and S yellow. The helical repeat unit is two disaccharides with a pitch of 17.5 Å. Note the high density of anionic sulfate groups. [Based on an NMR structure by Barbara Mulloy and Mark Forster, National Institute for Biological Standards and Control, Herts, U.K. PDBid 1HPN.]

postsurgical patients. **Heparan sulfate,** a ubiquitous cell-surface component as well as an extracellular substance in blood vessel walls and brain, resembles heparin but has a far more variable composition with fewer *N*- and *O*-sulfate groups and more *N*-acetyl groups.

3 GLYCOPROTEINS

Until about 1960, carbohydrates were thought to be rather dull compounds that were probably some sort of inert filler. Protein chemists therefore considered them to be a nuisance that complicated protein "purification." In fact, most eukaryotic proteins are **glycoproteins,** that is, they are covalently associated with carbohydrates. Glycoproteins vary in carbohydrate content from <1% to >90% by weight. They occur in all forms of life and have functions that span the entire spectrum of protein activities, including those of enzymes, transport proteins, receptors, hormones, and structural proteins. Their carbohydrate moieties, as we shall see, have several important biological roles, but in many cases their functions remain enigmatic.

The polypeptide chains of glycoproteins, like those of all proteins, are synthesized under genetic control. Their carbohydrate chains, in contrast, are enzymatically generated and covalently linked to the polypeptide without the rigid guidance of nucleic acid templates. The processing enzymes are generally not available in sufficient quantities to ensure the synthesis of uniform products. Glycoproteins therefore have variable carbohydrate compositions, a phenomenon known as **microheterogeneity,** that compounds the difficulties in their purification and characterization.

In this section we consider the structures and properties of glycoproteins. In particular, we shall study the glycoproteins of connective tissues, those of bacterial cell walls, and several soluble glycoproteins. We end by discussing the general principles of glycoprotein structure and function.

A. *Proteoglycans*

Proteins and glycosaminoglycans in ground substance, in **basal laminae [basement membranes;** the thin matlike extracellular matrix separating **epithelial cells** (the cells lining body cavities and free surfaces) from underlying cells], and in cell-surface membranes aggregate covalently and noncovalently to form a diverse group of macromolecules known as **proteoglycans.** *Proteoglycans consist of a* ***core protein*** *to which at least one glycosaminoglycan chain, most often keratan sulfate and/or chondroitin sulfate, is covalently linked.* Numerous types of core proteins have been characterized (Table 11-1). Proteoglycans appear to have multiple roles, most notably as organizers of tissue morphology via their interactions with molecules such as collagen; as selective filters that regulate the traffic of molecules according to their size and/or charge; and as regulators of the activities of other proteins, particularly those involved in signaling (see below).

Electron micrographs such as Fig. 11-24*a* together with reconstitution experiments indicate that proteoglycans can form huge complexes. For example, **aggrecan,** the main proteoglycan component of cartilage, has a bottlebrush-like molecular architecture (Fig. 11-24*b*), whose **proteoglycan subunit** "bristles" are noncovalently attached to a filamentous hyaluronic acid "backbone" at intervals of 200 to 300 Å. Aggrecan has three domains. Its N-terminal domain forms a globular region of 60 to 70 kD that binds noncovalently

Table 11-1 **Properties of Some Proteoglycans**

Proteoglycan	Approximate Core Protein Molecular Mass (kD)	Glycosaminoglycan Type (Number)[a]
Proteoglycans interacting with hyaluronic acid		
Aggrecan	220	CS (~100), KS (~30)
Versican	265–370	CS/DS (10–30)
Neurocan	136	CS (3–7)
Proteoglycans of the basal laminae		
Perlecan	400–467	HS/CS (3)
Agrin	250	HS (3)
Bamacan	138	CS (3)
Small leucine-rich proteoglycans		
Decorin	40	DS/CS (1)
Fibromodulin	42	KS (2–3)
Osteoglycin	35	KS (2–3)

[a]Abbreviations: CS, chondroitin sulfate; DS, dermatan sulfate; HS, heparan sulfate; KS, keratan sulfate.

Source: Iozzo, R.V., *Annu. Rev. Biochem.* **67,** 611, 626, and 624 (1998).

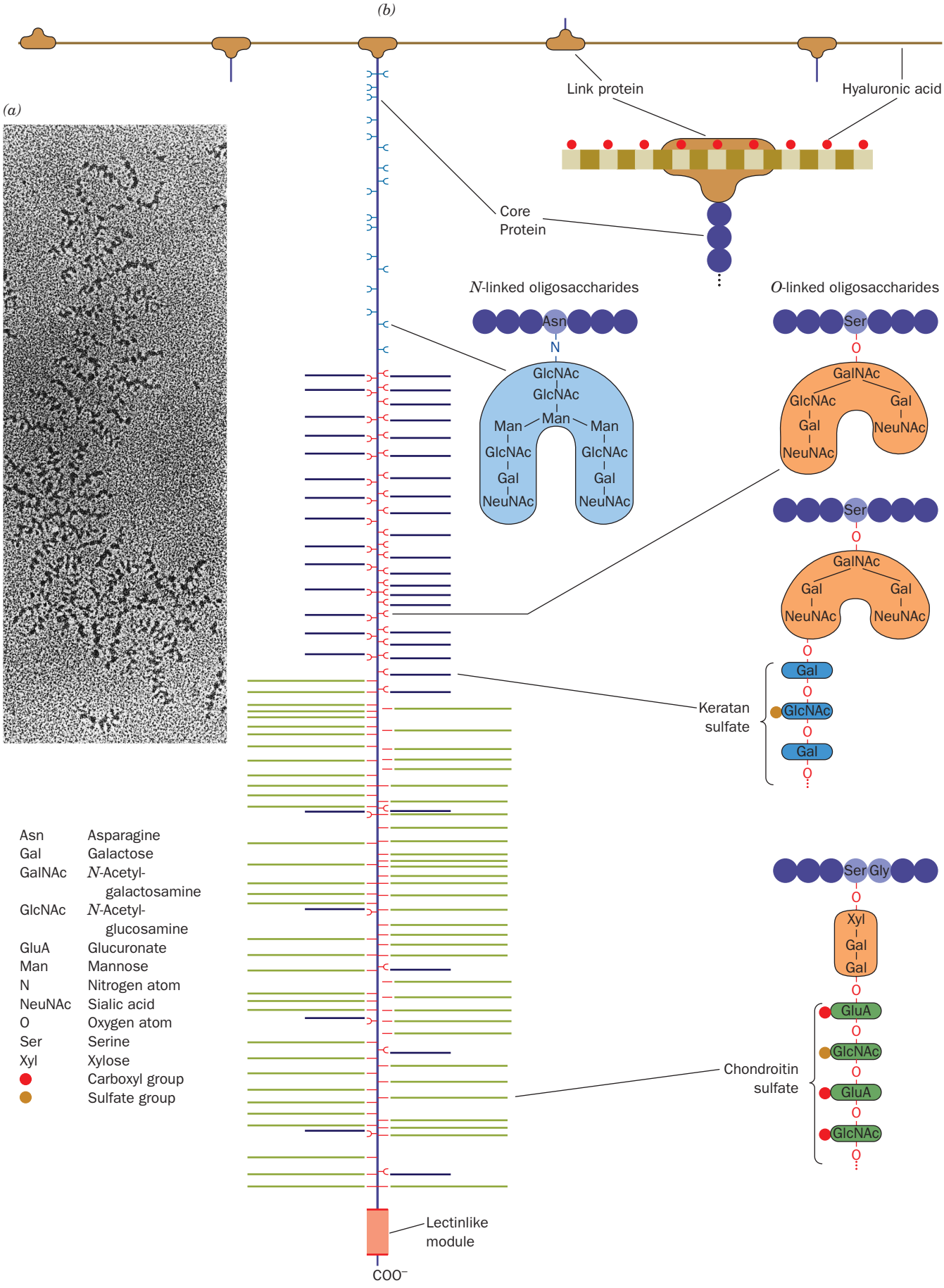

(a)
(b)
Link protein
Hyaluronic acid
Core Protein
N-linked oligosaccharides
O-linked oligosaccharides
Asn
N
Ser
O
GlcNAc
GlcNAc
Man
Man
Man
GlcNAc
GlcNAc
Gal
NeuNAc
GalNAc
Keratan sulfate
Ser
Gly
Xyl
GluA
Chondroitin sulfate
Lectinlike module
COO⁻
Asn Asparagine
Gal Galactose
GalNAc N-Acetyl-galactosamine
GlcNAc N-Acetyl-glucosamine
GluA Glucuronate
Man Mannose
N Nitrogen atom
NeuNAc Sialic acid
O Oxygen atom
Ser Serine
Xyl Xylose
Carboxyl group
Sulfate group

Figure 11-24 (*Opposite*) **Proteoglycans.** (*a*) An electron micrograph showing a central strand of hyaluronic acid, which runs down the field of view, supporting numerous projections, each of which consists of a core protein to which many bushy polysaccharide protrusions are linked. [From Caplan, A.I., *Sci. Am.* **251**(4); 87 (1984). Copyright © 1984 Scientific American, Inc. Used by permission.] (*b*) The bottlebrush model of the proteoglycan aggrecan. The core proteins, one of which is shown extending down through the middle of the diagram, project from the central hyaluronic acid strand. The core is noncovalently anchored to the hyaluronic acid via its globular N-terminal end in an association that is stabilized by link protein. The core has three saccharide-binding regions: (1) the inner region predominantly binds oligosaccharides via the side chain N atoms of Asn residues; (2) the central region binds oligosaccharides, many of which bear keratan sulfate chains, via the side chain O atoms of Ser and Thr residues; and (3) the outer region mainly binds chondroitin sulfate chains that are linked to the core protein via a galactose–galactose–xylose trisaccharide that is bonded to side chain O atoms of Ser residues in the sequence Ser-Gly. The C-terminal end of the aggrecan core protein consists of a lectinlike sequence.

to hyaluronic acid. This attachment is stabilized by the 40- to 60-kD **link protein,** which is similar in sequence to aggrecan's N-terminal domain. Aggrecan's highly extended central domain is covalently linked to a series of polysaccharides, which comprise nearly 90% of this glycoprotein's mass. They divide the central domain into three regions:

1. An N-terminal region, which overlaps the globular hyaluronic acid–binding domain, binds a relatively few carbohydrate chains. These tend to be oligosaccharides that are covalently bonded to the protein via the amide N atoms of specific Asn residues (Section 11-3Ca).

2. A region rich in oligosaccharides, many of which serve as anchor points for keratan sulfate chains. These oligosaccharides are covalently bonded to side chain O atoms of Ser and Thr residues.

3. A C-terminal region rich in chondroitin sulfate chains, which are covalently linked to the side chain O atoms of Ser residues in Ser-Gly dipeptides via galactose–galactose–xylose trisaccharides.

Aggrecan's C-terminal domain contains a lectinlike module, which binds certain monosaccharide units. Thus, aggrecan probably functions to bind together various constituents of the cell surface and the extracellular matrix (see below).

Altogether, a central strand of hyaluronic acid, which varies in length from 4000 to 40,000 Å, noncovalently binds up to 100 associated aggrecan chains, each of which covalently binds ~30 keratan sulfate chains of up to 250 disaccharide units each and ~100 chondroitin sulfate chains of up to 1000 disaccharide units each. This accounts for the enormous molecular masses of the aggrecans, which range up to 220,000 kD, and for their high degree of **polydispersity** (range of molecular masses). Note, however, that many proteoglycans do not bind to hyaluronic acid (Table 11-1) and hence function as monomers.

a. Cartilage's Mechanical Properties Are Explained by Its Molecular Structure

Cartilage consists largely of a meshwork of collagen fibrils that is filled in by proteoglycans whose chondroitin sulfate and core protein components specifically interact with the collagen. The tensile strength of cartilage and other connective tissues is, as we have seen (Section 8-2Ba), a consequence of their collagen content. Cartilage's characteristic resilience, however, results from its high proteoglycan content. The extended brushlike structure of proteoglycans, together with the polyanionic character of keratan sulfate and chondroitin sulfate, cause this complex to be highly hydrated. The application of pressure on cartilage squeezes water away from these charged regions until charge–charge repulsions prevent further compression. When the pressure is released, the water returns. Indeed, the cartilage in the joints, which lack blood vessels, is nourished by this flow of liquid brought about by body movements. This explains why long periods of inactivity cause joint cartilage to become thin and fragile.

b. Proteoglycans Modulate the Effects of Protein Growth Factors

Proteoglycans have been implicated in a great variety of cellular processes. For example, **fibroblast growth factor** (**FGF;** growth factors are proteins that function to induce their specific target cells to grow and/or differentiate; Section 19-3Aa) binds to heparin or to the heparan sulfate chains of proteoglycans and is only bound to its cell-surface receptor in complex with these glycosaminoglycans. Since the binding of FGF to heparin or heparan sulfate protects FGF from degradation, the release of this growth factor from the extracellular matrix by the proteolysis of proteoglycan core proteins or by the partial degradation of heparan sulfate probably provides an important source of active FGF–glycosaminoglycan complexes. Several other growth factors interact similarly with proteoglycans. Apparently, the abundant and ubiquitous distribution of proteoglycans limits the action of these growth factors on their target cells to short distances from the cells secreting the growth factors, a phenomenon that probably greatly influences the formation and maintenance of tissue architecture.

B. *Bacterial Cell Walls*

Bacteria are surrounded by rigid cell walls (Fig. 1-13) that give them their characteristic shapes (Fig. 1-1) and permit them to live in hypotonic (less than intracellular salt concentration) environments that would otherwise cause them to swell osmotically until their plasma (cell) membranes lysed (burst). Bacterial cell walls are of considerable medical significance because they are responsible for bacterial **virulence** (disease-evoking power). In fact, the symptoms of many bacterial diseases can be elicited in animals merely by the injection of bacterial cell walls. Furthermore, the

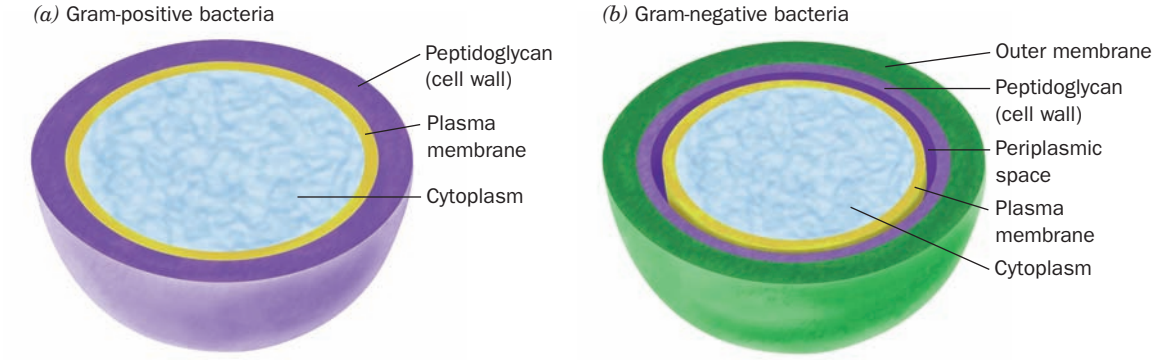


Figure 11-25 **Schematic diagram comparing the cell envelopes of (*a*) gram-positive bacteria and (*b*) gram-negative bacteria.**

characteristic **antigens** (immunological markers; Section 35-2) of bacteria are components of their cell walls and capsules, so that injection of preparations of these substances into an animal often invokes its immunity against these bacteria. Consequently, several vaccines that are based on purified bacterial polysaccharides have recently become available, including those against *Streptococcus pneumoniae,* a major cause of pneumonia, and *Neisseria meningitidis,* a major cause of meningitis.

Bacteria are classified as **gram-positive** or **gram-negative** depending on whether or not they take up gram stain (Section 1-1B). Gram-positive bacteria (Fig. 11-25*a*) have a thick (~250 Å) cell wall surrounding their plasma membrane, whereas gram-negative bacteria (Fig. 11-25*b*) have a thin (~30 Å) cell wall covered by a complex outer membrane.

a. Bacterial Cell Walls Have a Peptidoglycan Framework

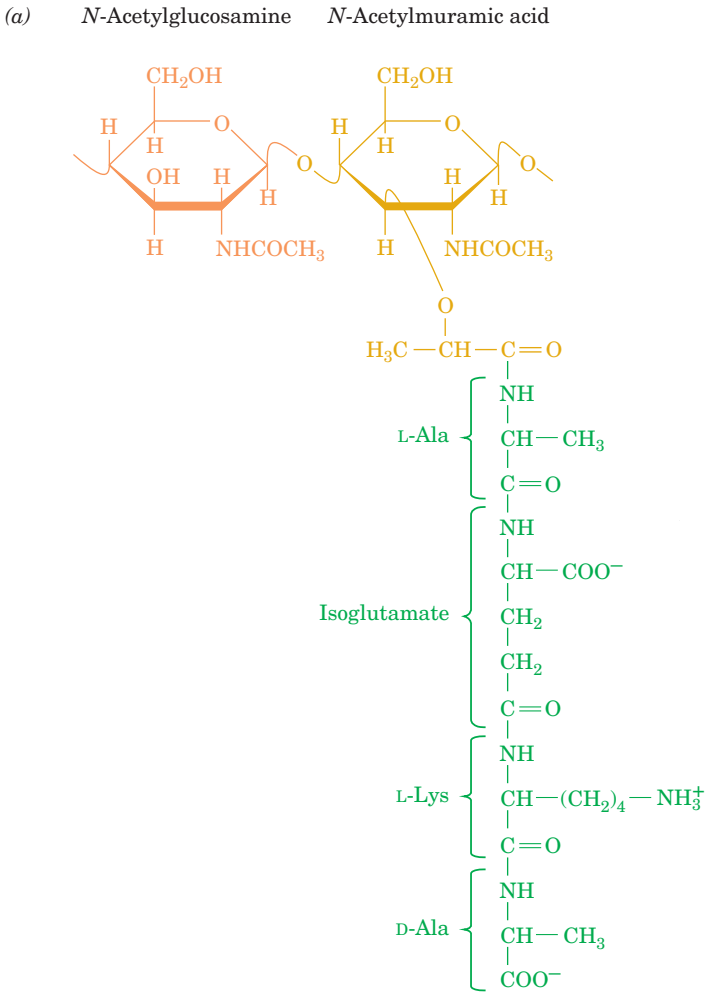
The cell walls of both gram-positive and gram-negative bacteria consist of covalently linked polysaccharide and polypeptide chains that form a framework that completely encases the cell. This substance, whose molecular structure was elucidated in large part by Jack Strominger, is known as a **peptidoglycan** or **murein** (Latin: *murus,* wall). Its polysaccharide component consists of linear chains of alternating β(1 → 4)-linked *N*-acetylglucosamine **(NAG)** and *N*-acetylmuramic acid **(NAM).** The NAM's lactic acid residue forms an amide bond with a D-amino acid–containing tetrapeptide to form the peptidoglycan repeating unit (Fig. 11-26). Neighboring parallel peptidoglycan chains are covalently cross-linked through their tetrapeptide side chains. In the gram-positive bacterium *Staphylococcus aureus,* whose tetrapeptide has the sequence L-Ala-D-isoglutamyl-L-Lys-D-Ala, this cross-link consists of a pentaglycine chain that extends from the terminal carboxyl group of one tetrapeptide to the ε-amino group of the Lys in a neighboring tetrapeptide.

Atomic force microscopy (AFM; an imaging technique that reports the variation in the force between a probe that is several nanometers in diameter and a surface of interest as the probe is scanned over the surface; its resolution is as little as several Ångstroms) was used by Simon Foster to image the cell wall of the gram-negative bacterium *Bacillus subtilis* leading to the following model (Fig. 11-27). Several glycan chains are cross-linked much as described above to form a peptidoglycan "rope," which due to its natural twist, forms an ~50-nm-diameter helical cable of up to 50 μm in length that coils around the long axis of the bacterium to form its cell wall. This structure is presumably stabilized by the formation of covalent cross-links between neighboring segments of the coil. The cell walls of gram-negative bacteria appear to be only one layer thick, whereas as those of gram-positive bacteria are postulated to consist of several such layers. How the peptidoglycan imposes cell shape is unknown.

The D-amino acids of peptidoglycans render them resistant to proteases. However, **lysozyme,** an enzyme which is present in tears, mucus, and other vertebrate body secretions, as well as in egg whites, catalyzes the hydrolysis of the β(1 → 4) glycosidic linkage between NAM and NAG. Consequently, treatment of gram-positive bacteria with lysozyme degrades their cell walls, which results in their lysis (gram-negative bacteria are resistant to lysozyme degradation). Lysozyme was discovered in 1922 by the British bacteriologist Alexander Fleming after he noticed that a bacterial culture had dissolved where mucus from a sneeze had landed. It was Fleming's hope that lysozyme would be a universal antibiotic but, unfortunately, it is clinically ineffective against pathogenic bacteria. The structure and mechanism of lysozyme are examined in detail in Section 15-2.

b. Penicillin Kills Bacteria by Inhibiting Cell Wall Biosynthesis

In 1928, Fleming noticed that the chance contamination of a bacterial culture plate with the mold *Penicillium notatum* lysed nearby bacteria (a clear demonstration of Pasteur's maxim that chance favors a prepared mind). This was caused by the presence of **penicillin** (Fig. 11-28), an antibiotic secreted by the mold. Yet the difficulties of isolating and characterizing penicillin, owing to its instability, led to the passage of over 15 years before penicillin was available for routine clinical use. Penicillin specifically binds to and inactivates enzymes that function to cross-link the peptidoglycan strands of bacterial cell walls. Since cell



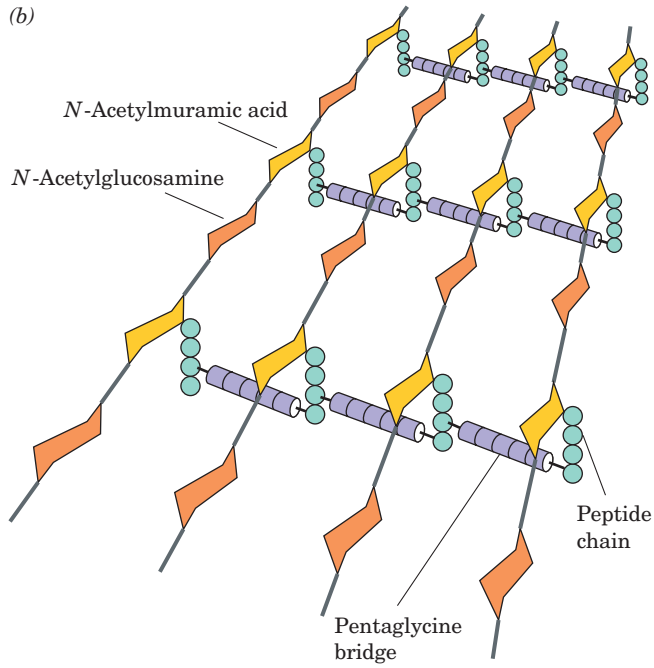


Figure 11-26 **Chemical structure of peptidoglycan.** (*a*) The repeating unit of peptidoglycan is an NAG–NAM disaccharide whose lactyl side chain forms an amide bond with a tetrapeptide. The tetrapeptide of *S. aureus* is shown. The isoglutamate is so designated because it forms an amide link via its γ-carboxyl group. In some species, its α-carboxylate group is replaced by an amide group to form D-isoglutamine and/or the L-Lys residue may have a carboxyl group appended to its C_ε to form **diaminopimelic acid.** (*b*) The *S. aureus* bacterial cell wall peptidoglycan. In other gram-positive bacteria, the Gly_5 connecting bridges shown here may contain different amino acid residues such as Ala or Ser. In gram-negative bacteria, the peptide chains are directly linked via peptide bonds.

wall expansion also requires the action of enzymes that degrade cell walls, *exposure of growing bacteria to penicillin results in their lysis;* that is, penicillin disrupts the normal balance between cell wall biosynthesis and degradation. However, since no human enzyme binds penicillin, it is of low human toxicity, a therapeutic necessity.

Penicillin-treated bacteria that are kept in a hypertonic medium remain intact, even though they have no cell wall.

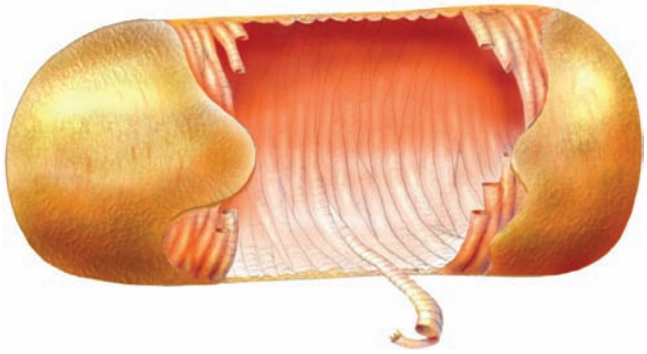
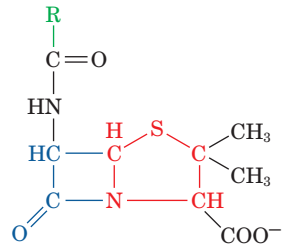


Figure 11-27 **Model of the *B. subtilis* cell wall.** The cell wall consists of a right-handed helical cable composed of several peptidoglycan strands that wraps about the bacterium's plasma membrane. The cell is ~3 μm long. [Courtesy of Simon Foster, University of Sheffield, U.K.]



Penicillin

Figure 11-28 **Structure of penicillin.** Penicillin contains a thiazolidine ring (*red*) fused to a β-lactam ring (*blue*). A variable R group is bonded to the β-lactam ring via a peptide linkage. In benzyl penicillin (penicillin G), one of several naturally occurring derivatives that are clinically effective, R is the benzyl group ($—CH_2\phi$). In **ampicillin,** a semisynthetic derivative, R is the aminobenzyl group [$—CH(NH_2)\phi$].

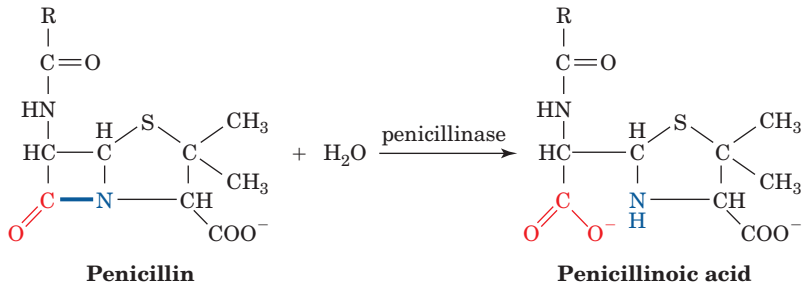
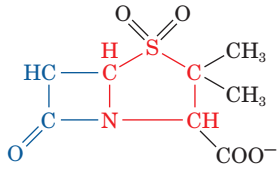


Figure 11-29 **Enzymatic inactivation of penicillin.** Penicillinase inactivates penicillin by catalyzing the hydrolysis of its β-lactam ring to form **penicillinoic acid.**

Such bacteria, which are called **protoplasts** or **spheroplasts,** are spherical and extremely fragile because they are encased by only their plasma membranes. Protoplasts immediately lyse on transfer to a normal medium.

Most bacteria that are resistant to penicillin secrete a **β-lactamase** (also known as **penicillinase**), which inactivates penicillin by hydrolytically cleaving the amide bond of its β-lactam ring (Fig. 11-29). However, the observation that penicillinase activity varies with the nature of penicillin's R group has prompted the semisynthesis of penicillins, such as **ampicillin** (Fig. 11-28), which are clinically effective against penicillin-resistant strains of bacteria. In addition, penicillins are often administered in combination with β-lactamase inhibitors such as **sulbactam.**



Sulbactam

c. Bacterial Cell Walls Are Studded with Antigenic Groups

The surfaces of gram-positive bacteria are covered by **teichoic acids** (Greek: *teichos,* city walls), which account for up to 50% of the dry weight of their cell walls. Teichoic acids are polymers of glycerol or ribitol linked by phosphodiester bridges (Fig. 11-30). The hydroxyl groups of this sugar–phosphate chain are substituted by D-Ala residues and saccharides such as glucose or NAG. Teichoic acids are anchored to the peptidoglycans via phosphodiester bonds to the C6-OH groups of their NAG residues. They often terminate in **lipopolysaccharides** (lipids that contain polysaccharides; Section 12-1).

The outer membranes of gram-negative bacteria (Fig. 11-25*b*) are composed of complex lipopolysaccharides, proteins, and phospholipids that are organized in a complicated manner. The **periplasmic space,** an aqueous compartment that lies between the plasma membrane and the peptidoglycan cell wall, contains proteins that transport sugars and other nutrients. The outer membrane functions as a barrier to exclude harmful substances (such as gram stain). This accounts for the observation that gram-negative bacteria are less affected by lysozyme and penicillin, as well as by other antibiotics, than are gram-positive bacteria.

The outer surfaces of gram-negative bacteria are coated with complex and often unusual polysaccharides known as **O-antigens** that uniquely mark each bacterial strain (Fig. 11-31). The observation that mutant strains of pathogenic bacteria lacking O-antigens are nonpathogenic suggests

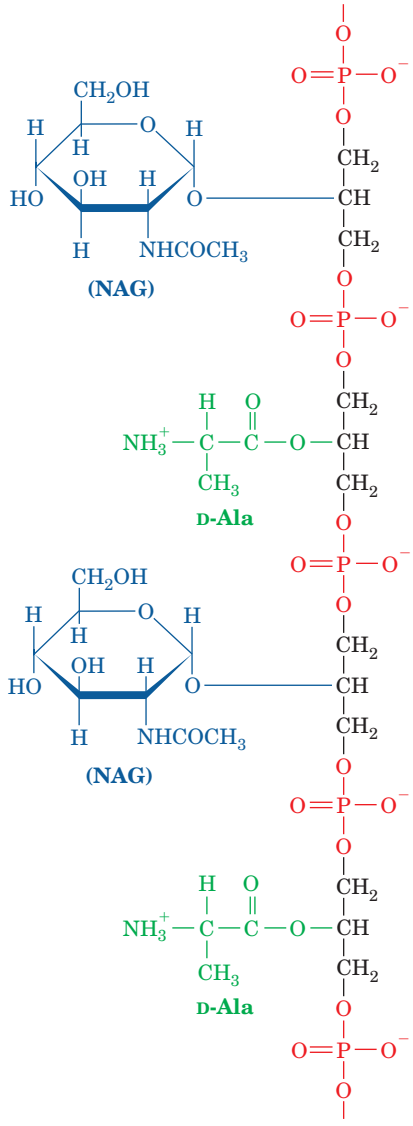
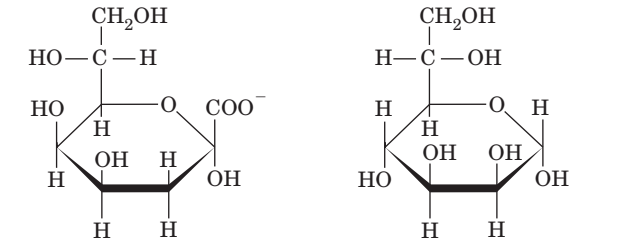


Figure 11-30 **Structure of teichoic acid.** A segment of a teichoic acid molecule with a glycerol phosphate backbone that bears alternating residues of D-Ala and NAG.



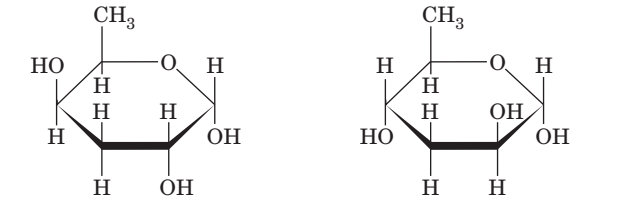


Figure 11-31 **Some of the unusual monosaccharides that occur in the O-antigens of gram-negative bacteria.** These sugars rarely occur in other organisms.

that O-antigens participate in the recognition of host cells. O-Antigens, as their name implies, are also the means by which a host's immunological defense system recognizes invading bacteria as foreign (Section 35-2A). As part of the ongoing biological warfare between pathogen and host, O-antigens are subject to rapid mutational alteration so as to generate new bacterial strains that the host does not initially recognize (the mutations are in the genes specifying the enzymes that synthesize the O-antigens).

C. *Glycoprotein Structure and Function*

a. Glycoprotein Carbohydrate Chains Are Highly Diverse

Almost all the secreted and membrane-associated proteins of eukaryotic cells are glycosylated. Indeed, protein glycosylation is more abundant than all other types of post-translational modifications combined. Oligosaccharides form two types of direct attachments to these proteins: ***N*-linked** and ***O*-linked.** Sequence analyses of glycoproteins have led to the following generalizations about these attachments.

1. *In the vast majority of N-glycosidic (N-linked) attachments, an NAG is β-linked to the amide nitrogen of an Asn in the sequence Asn-X-Ser or Asn-X-Thr, where X is any amino acid residue except Pro and only rarely Asp, Glu, Leu, or Trp* (Fig. 11-32*a*). The oligosaccharides in these linkages usually have a distinctive **core** (innermost sequence; Fig. 11-32*b*) whose peripheral mannose residues are linked to either mannose or NAG residues. These latter residues may, in turn, be linked to yet other sugar residues,

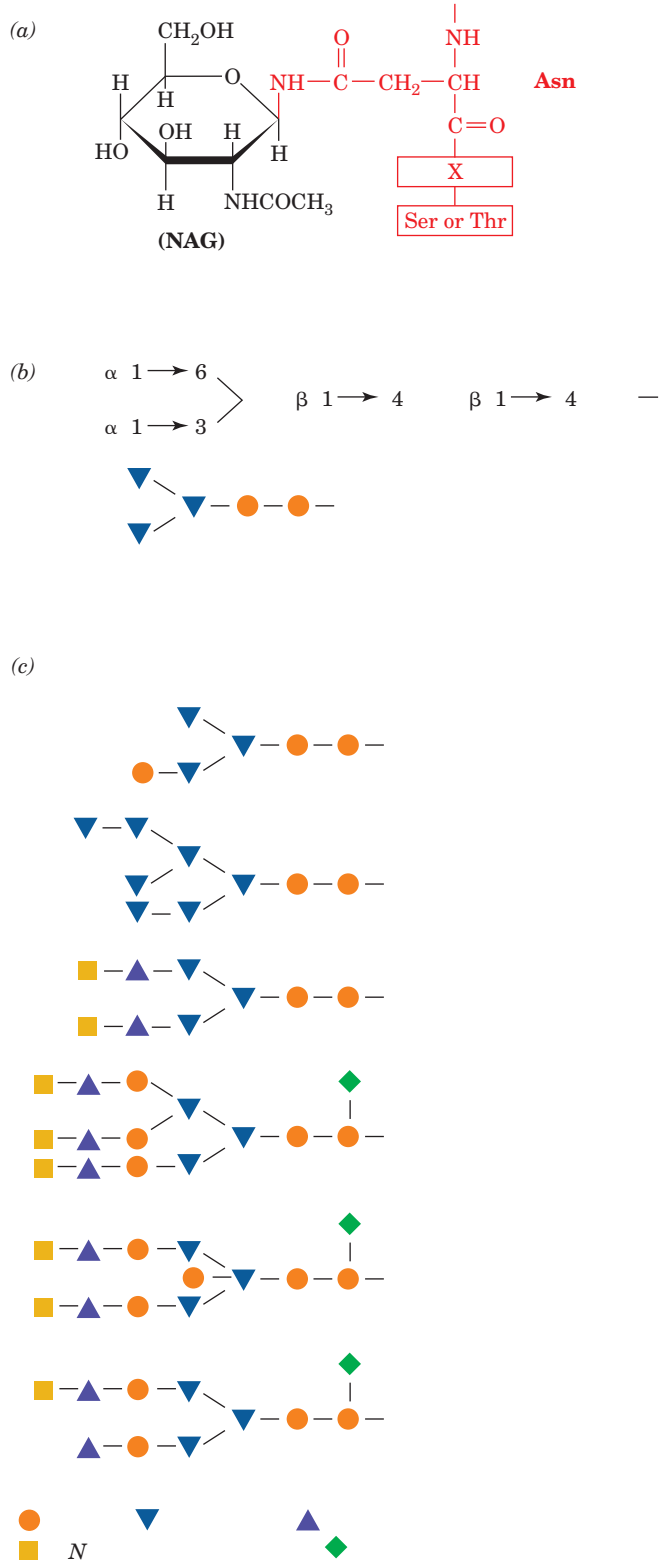


Figure 11-32 ***N*-Linked oligosaccharides.** (*a*) All *N*-glycosidic protein attachments occur through a β-*N*-acetylglucosamino–Asn bond in which the Asn occurs in the sequence Asn-X-Ser/Thr (*red*) where X is any amino acid. (*b*) *N*-Linked oligosaccharides usually have the branched $(mannose)_3(NAG)_2$ core shown. (*c*) Some examples of *N*-linked oligosaccharides. [After Sharon, N. and Lis, H., *Chem. Eng. News* **59**(13), 28 (1981).] **See Kinemage Exercise 7-4**

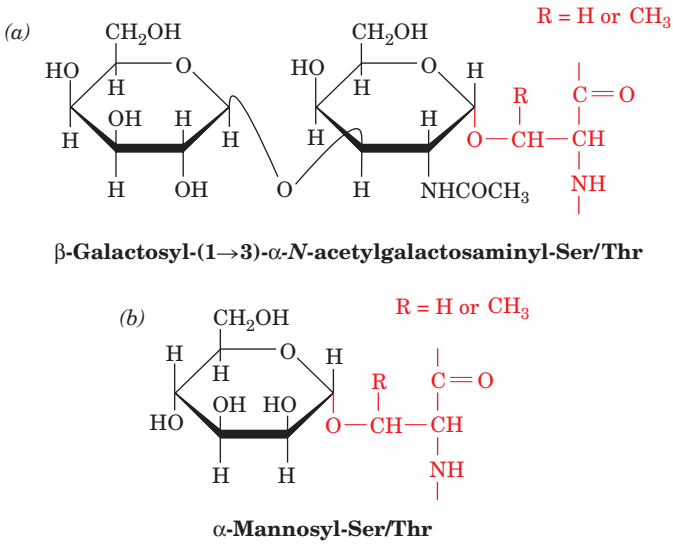


Figure 11-33 **Some common *O*-glycosidic attachments of oligosaccharides to glycoproteins (*red*).**

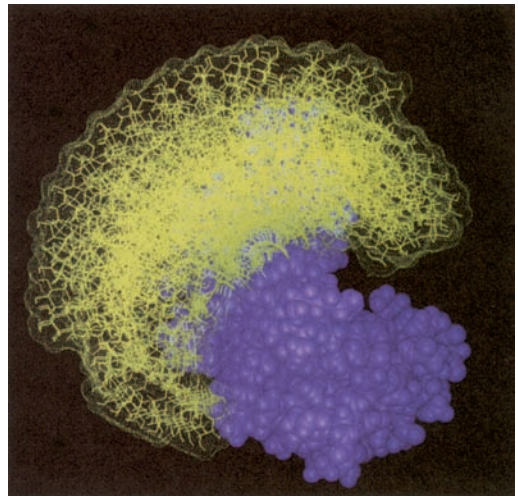


Figure 11-34 **Model of oligosaccharide dynamics in bovine pancreatic ribonuclease B (RNase B).** The allowed conformations of the $(mannose)_5(NAG)_2$ oligosaccharide (*yellow*) that is linked to a single site on the protein (*purple*) are shown in superimposed snapshots. [Courtesy of Raymond Dwek, Oxford University, U.K.]

so that an enormous diversity of *N*-linked oligosaccharides is possible (e.g., there are $\sim 10^{12}$ possible hexasaccharides, although only a small fraction of them are actually synthesized). Several *N*-linked oligosaccharides are shown in Fig. 11-32*c*.

2. *The most common O-glycosidic (O-linked) attachment involves the disaccharide core β-galactosyl-(1 → 3)-α-N-acetylgalactosamine α-linked to the OH group of either Ser or Thr* (Fig. 11-33*a*). Less commonly, glucose, galactose, mannose, and xylose form α-*O*-glycosides with Ser or Thr (Fig. 11-33*b*). All other hydroxyl-bearing amino acid side chains occasionally form *O*-glycosidic bonds: those with Tyr (e.g., in the protein **glycogenin;** Section 18-2B), 5-hydroxy-Lys (Hyl; e.g., in collagen; Section 8-2Bb), and 4-hydroxy-Pro (Hyp). However, there seem to be few, if any, additional generalizations that can be made about *O*-glycosidically linked oligosaccharides. They vary in size from a single galactose residue in collagen to chains of up to 1000 disaccharide units in proteoglycans.

N-Linked glycans are around 5-fold more common than *O*-linked glycans with only ~10% of glycoproteins having both types of attachments.

Oligosaccharides tend to attach to proteins at sequences that form β bends. Taken with their hydrophilic character, this observation suggests that *oligosaccharides extend from the surfaces of proteins rather than participate in their internal structures.* Indeed, the relatively few glycoprotein X-ray structures that have yet been reported, for example, those of **immunoglobulin G** (Section 35-2Ba) and the influenza virus **hemagglutinin** (Section 33-4Bb), are consistent with this hypothesis. This accounts for the observation that the protein structures of most glycoproteins are unaffected by the removal of their associated oligosaccharides. Both experimental and theoretical studies indicate that oligosaccharides have mobile and rapidly fluctuating conformations (Fig. 11-34; which accounts for the difficulty in crystallizing them). Thus, representations in which oligosaccharides are shown as having fixed three-dimensional structures do not tell the whole story.

b. Glycoprotein Carbohydrates Have a Variety of Functions

Cells tend to synthesize a large repertoire of a given N-linked glycoprotein, in which each variant species ***(glycoform)*** *differs somewhat in the sequences, locations, and numbers of its covalently attached oligosaccharides.* For example, one of the simplest glycoproteins, bovine pancreatic **ribonuclease B (RNase B),** differs from the well-characterized and carbohydrate-free enzyme RNase A (Section 9-1A) only by the attachment of a single *N*-glycosidically linked oligosaccharide chain. The oligosaccharide has the core sequence diagrammed in Fig. 11-35 with considerable microheterogeneity in the position of a sixth mannose residue. The oligosaccharide does not affect the native enzyme's

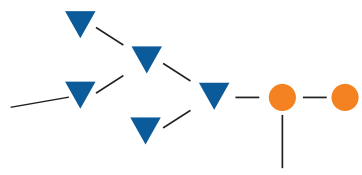


Figure 11-35 **The microheterogeneous *N*-linked oligosaccharide of RNase B has the $(mannose)_5(NAG)_2$ core shown.** A sixth mannose residue occurs at various positions on this core.

conformation, substrate specificity, or catalytic properties. However, RNase A folds to its native state more slowly than does RNase B and tends to aggregate. This suggests that the oligosaccharide functions similarly to a molecular chaperone (Section 9-2C), most likely by shielding a hydrophobic patch on the protein surface.

Human **granulocyte–macrophage colony-stimulating factor (GM-CSF),** a 127-residue protein growth factor that promotes the development, activation, and survival of the white blood cells known as **granulocytes** and **macrophages,** is variably glycosylated at two *N*-linked sites and five *O*-linked sites. Through the generation of mutant varieties of GM-CSF that lack one or both of the *N*-glycosylation sites, it was found that the lifetime of GM-CSF in the bloodstream increases with its level of glycosylation. However, GM-CSF that is produced by *E. coli* and hence is unglycosylated (bacteria rarely glycosylate the proteins they synthesize) has a 20-fold higher specific biological activity than does the naturally occurring glycoprotein.

As the foregoing examples suggest, no generalization can be made about the effects of glycosylation on protein properties; they must be experimentally determined on a case-by-case basis. Nevertheless, it is becoming increasingly evident that glycosylation can affect protein properties in many ways, including protein folding, oligomerization, physical stability, specific bioactivity, rate of clearance from the bloodstream, and protease resistance. Thus, *the species-specific and tissue-specific distribution of glycoforms that each cell synthesizes endows it with a characteristic spectrum of biological properties.*

c. *O*-Linked Glycoproteins Often Have Protective Functions

O-Linked polysaccharides tend not to be uniformly distributed along polypeptide chains. Rather, they are clustered into heavily glycosylated (65–85% carbohydrate by weight) segments in which glycosylated Ser and Thr residues comprise 25 to 40% of the sequence. The carbohydrates' hydrophilic and steric interactions cause these heavily glycosylated regions, which are also rich in Pro and other helix-breaking residues, to assume extended conformations. For example, **mucins,** the protein components of **mucus,** are *O*-linked glycoproteins that can be exceedingly large (up to $\sim 10^7$ D) and whose carbohydrate chains are often sulfated and hence mutually repelling. Mucins, which may be membrane-bound or secreted, therefore consist of stiff chains that are devoid of secondary structure and which occupy time-averaged volumes approximating those of small bacteria. Consequently, mucins, at their physiological concentrations, form intertangled networks that comprise the viscoelastic gels that protect and lubricate the mucous membranes that produced them.

Eukaryotic cells, as we shall see in Section 12-3E, have a thick and fuzzy coating of glycoproteins and **glycolipids** named the **glycocalyx** that prevents the close approach of macromolecules and other cells. How, then, can cells interact? Many cell-surface proteins, such as the receptors for various macromolecules, have relatively short and presumably stiff *O*-glycosylated regions that link these glycoproteins' membrane-bound domains to their functional domains. This arrangement is thought to extend the functional domains in a lollipop-like manner above the cell's densely packed glycocalyx, thereby permitting the functional domain to interact with extracellular macromolecules that cannot penetrate the glycocalyx.

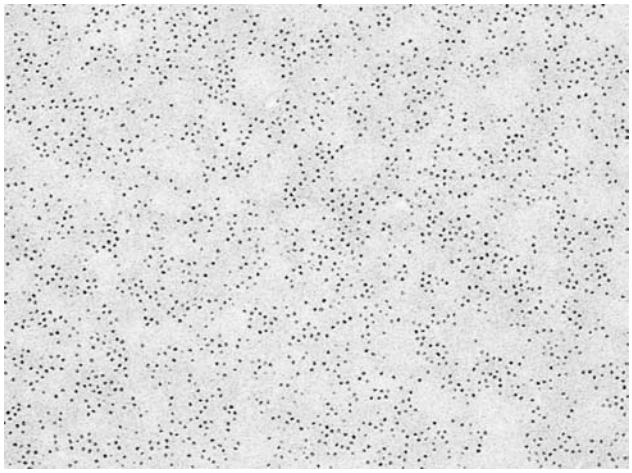
d. Oligosaccharide Markers Mediate a Variety of Intercellular Interactions

Glycoproteins are important constituents of plasma membranes (Section 12-3). The location of their carbohydrate moieties can be determined by electron microscopy. The glycoproteins are labeled with lectins that have been conjugated (covalently cross-linked) to **ferritin,** an iron-transporting protein that is readily visible in the electron microscope because of its electron-dense iron hydroxide core. Such experiments, with lectins of different specificities and with a variety of cell types, have demonstrated that *the carbohydrate groups of membrane-bound glycoproteins are, for the most part, located on the external surfaces of cell membranes.* Thus, the viability of cultured cells from multicellular organisms that have any of a large number of glycosylation mutations and the infrequent viability of whole organisms that bear such mutations indicate that oligosaccharides are important for intercellular communications but not for intracellular housekeeping functions.

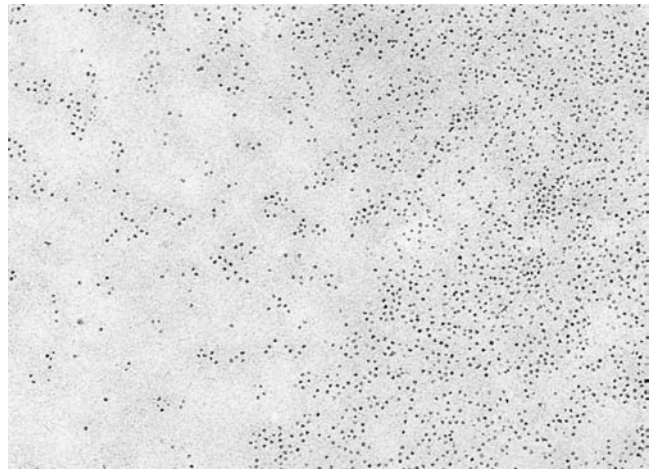
A further indication that oligosaccharides function as biological markers is the observation that the carbohydrate content of a glycoprotein often governs its metabolic fate. For example, the excision of sialic acid residues from certain radioactively labeled blood plasma glycoproteins by treatment with **sialidase** greatly increases the rate at which these glycoproteins are removed from the circulation. The glycoproteins are taken up and degraded by the liver in a process that depends on the recognition by liver cell receptors of sugar residues such as galactose and mannose, which are exposed by the sialic acid excision. A diverse series of receptors, each specific for a particular type of sugar residue, participates in removing any particular glycoprotein from the blood. A variety of glycoforms for a given glycoprotein therefore probably ensures that it has a range of lifetimes in the blood. *Similar "ticketing" mechanisms probably govern the compartmentation and degradation of glycoproteins within cells.*

The observation that cancerous cells are more susceptible to agglutination by lectins than are normal cells led to the discovery that *there are significant differences between the cell-surface carbohydrate distributions of cancerous and noncancerous cells* (Fig. 11-36). Normal cells stop growing when they touch each other, a phenomenon known as **contact inhibition.** Cancer cells, however, are under no such control and therefore form **malignant tumors** (Section 19-3B).

Carbohydrates are important mediators of cell–cell recognition and have been implicated in related processes such as fertilization, cellular differentiation, the aggregation of cells to form organs, and the infection of cells by



(a)



(b)

Figure 11-36 **The surfaces of (*a*) a normal mouse cell and (*b*) a cancerous cell as seen in the electron microscope.** Both cells were incubated with the ferritin-labeled lectin concanavalin A. The lectin is evenly dispersed on the normal cell but is aggregated into clusters on the cancerous cell. [Courtesy of Garth Nicolson, The Institute for Molecular Medicine, Huntington Beach, California.]

bacteria and viruses. For example, bacteria initiate infections by attaching to host cells (Fig. 11-37) via bacterial proteins known as **adhesins,** which each specifically bind certain host cell molecules (the adhesins' receptors). In gram-negative bacteria such as *E. coli,* adhesins are often minor components of the heteropolymeric rodlike organelles called pili (Fig. 1-3*b*). The so-called P pili that mediate the attachment of the *E. coli* strain that causes urinary tract infections in humans do so via an adhesin named **PapG** protein. This protein specifically binds to the α-D-galactopyranosyl-(1 → 4)-β-D-galactopyranose groups that are present on the surfaces of urinary tract epithelial cells. Electron microscopy studies revealed that the PapG adhesin is located at the end of the P pili's flexible tip, thereby providing this adhesin with considerable steric freedom in binding to its digalactoside receptor.

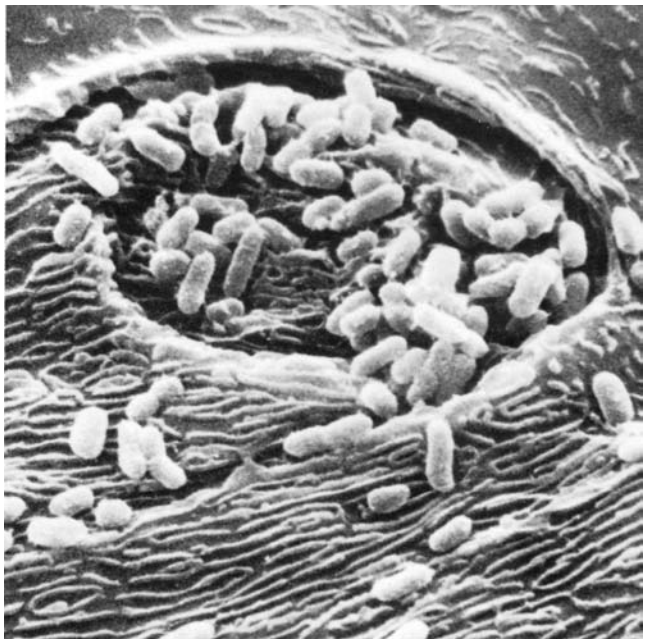


Figure 11-37 **Scanning electron micrograph of tissue from the inside of a human cheek.** The white cylindrical objects are *E. coli.* The bacteria adhere to mannose residues that are incorporated in the plasma membrane of cheek cells. This is the first step of a bacterial infection. [Courtesy of Fredric Silverblatt and Craig Kuehn, Veterans Administration Hospital, Sepulveda, California.]

In the never-ending evolutionary struggle between pathogens and their hosts, mucins have evolved to contain the target oligosaccharides of certain pathogens. These act as decoys that divert these pathogens from their target cells. This, of course, puts selective pressure on the pathogen to evolve a receptor that binds to a different cell-surface oligosaccharide.

D. *Glycomics*

Glycomics, the field of study that structurally and functionally characterizes all the carbohydrates in a given cell type, complements genomics (for DNA) and proteomics (for proteins). It is clear that the **glycome** varies with the species, cell type, developmental stage, and even environmental conditions. However, glycomics is far less developed than genomics or proteomics. There are several reasons for this:

1. The branched structures of oligosaccharides greatly increases their complexity and hence the difficulty in determining their sequences relative to those of polynucleotides and polypeptides, which are invariably linear.

2. The microheterogeneity of oligosaccharides, which often has biological significance (Section 11-3Cb), complicates their characterization relative to polynucleotides and polypeptides, which each have unique primary structures.

3. Because the biosynthesis of carbohydrates is not under direct genetic control, there is no method for amplifying them such as the polymerase chain reaction (PCR; Section 5-5F) for nucleic acids and expression systems for proteins (Section 5-5G). Thus, until recently, the only way of obtaining sufficient quantities of a particular polysaccharide was to isolate it from natural sources.

4. Methods for synthesizing specific oligosaccharides have lagged far behind methods for synthesizing polynucleotides and polypeptides (Sections 7-5 and 7-6). This is due to the branching of oligosaccharides, their large number of functional groups that must be differentially protected during elongation reactions, and the chiral nature of glycosidic bonds. However, in recent years, Peter Seeberger has developed automated, solid-phase methods for synthesizing small oligosaccharides, although these methods are, as yet, incapable of synthesizing all desired oligosaccharides, are time-consuming, and still require considerable expertise.

5. The complexity of an organism's glycome greatly exceeds that of its proteome due to the diversity the glycome's constituent carbohydrates and the number of ways they can interact with one another and with proteins.

A recent advance that has greatly accelerated glycomic research is the development of **carbohydrate microarrays** to identify the carbohydrates that specifically bind to a particular protein, RNA, or even whole cells. In this methodology, which is analogous to the use of DNA microarrays (Section 7-6B), up to several thousand different oligosaccharides are covalently or physically immobilized at specific sites on a solid surface such as a glass slide. A fluorescently labeled protein, RNA, or cell type is then incubated with the microarray, which is subsequently rinsed, and the oligosaccharides to which the protein/RNA/cell binds are identified by the fluorescence at their corresponding positions. In addition to their use in basic research, carbohydrate microarrays have been employed in such diverse applications as the identification of pathogens, the diagnosis of human diseases that are characterized by the presence of certain oligosaccharides, and the development of carbohydrate-based drugs and vaccines.

CHAPTER SUMMARY

Carbohydrates are polyhydroxy aldehydes or ketones of approximate composition $(C \cdot H_2O)_n$ that are important components of biological systems.

1 Monosaccharides The various monosaccharides, such as ribose, fructose, glucose, and mannose, differ in their number of carbon atoms, the positions of their carbonyl groups, and their diastereomeric configurations. These sugars exist almost entirely as cyclic hemiacetals and hemiketals, which, for five- and six-membered rings, are respectively known as furanoses and pyranoses. The two anomeric forms of these cyclic sugars may interconvert by mutarotation. Pyranose sugars have nonplanar rings with boat and chair conformations similar to those of substituted cyclohexanes. Polysaccharides are held together by glycosidic bonds between neighboring monosaccharide units. Glycosidic bonds do not undergo mutarotation. Monosaccharides can be oxidized to aldonic and glycuronic acids or reduced to alditols. An OH group is replaced by H in deoxy sugars and by an amino group in amino sugars.

2 Polysaccharides Carbohydrates can be purified by electrophoretic and chromatographic procedures. Affinity chromatography using lectins has been particularly useful in this regard. The sequences and linkages of polysaccharides may be determined by methylation analysis and by the use of specific exoglycosidases. Similar information may be obtained through NMR spectroscopy and/or mass spectrometric techniques. Cellulose, the structural polysaccharide of plant cell walls, is a linear polymer of $\beta(1 \rightarrow 4)$-linked D-glucose residues. It forms a fibrous hydrogen bonded structure of exceptional strength that in plant cells is embedded in an amorphous matrix. Starch, the food storage polysaccharide of plants, consists of a mixture of the linear $\alpha(1 \rightarrow 4)$-linked glucan α-amylose and the $\alpha(1 \rightarrow 6)$-branched and $\alpha(1 \rightarrow 4)$-linked glucan amylopectin. Glycogen, the animal storage polysaccharide, resembles amylopectin but is more highly branched. Digestion of starch and glycogen is initiated by α-amylase and is completed by specific membrane-bound intestinal enzymes.

3 Glycoproteins Proteoglycans of ground substance are mostly high molecular mass aggregates, many of which structurally resemble a bottlebrush. Their proteoglycan subunits consist of a core protein to which glycosaminoglycans, usually chondroitin sulfate and keratan sulfate, are covalently linked. The rigid framework of a bacterial cell wall consists of chains of alternating $\beta(1 \rightarrow 4)$-linked NAG and NAM that are cross-linked by short polypeptides to form a helical peptidoglycan cable that wraps around the bacterium. Lysozyme cleaves the glycosidic linkages between NAM and NAG of peptidoglycan. Penicillin specifically inactivates enzymes involved in the cross-linking of peptidoglycans. Both of these substances cause the lysis of susceptible bacteria. Gram-positive bacteria have teichoic acids that are linked covalently to their peptidoglycans. Gram-negative bacteria have outer membranes that bear complex and unusual polysaccharides known as O-antigens. These participate in the recognition of host cells and are important in the immunological recognition of bacteria by the host. Oligosaccharides attach to eukaryotic proteins in only a few ways. In *N*-glycosidic attachments, an NAG is invariably bound to the amide nitrogen of Asn in the sequence Asn-X-Ser(Thr). *O*-Glycosidic attachments are made to Ser or Thr in most proteins and to 5-hydroxylysine in collagen.

Oligosaccharides are located on the surfaces of glycoproteins. Glycoproteins have functions that span the entire range of protein activities, although the roles of their carbohydrate moieties are only poorly understood. For example, ribonuclease B differs from the functionally indistinguishable and carbohydrate-free ribonuclease A only by the attachment of a single oligosaccharide of somewhat variable sequence which increases the protein's rate of folding, whereas the biological properties of granulocyte–macrophage colony-stimulating factor are significantly affected by its multiple oligosaccharide chains. The

viscoelastic and hence protective properties of mucus largely result from the numerous negatively charged oligosaccharide groups of its component mucins. The carbohydrate moieties of glycoproteins in plasma membranes are invariably located on the external surfaces of the membranes. A glycoprotein's carbohydrate moieties may direct its metabolic fate by governing its uptake by certain cells or cell compartments. Glycoproteins are also important mediators of cell–cell recognition and, in many cases, are the receptors for bacterial attachment, via adhesins, in the initial stages of infection. Glycomics, the carbohydrate analog of genomics and proteomics, seeks to characterize all the carbohydrates in a particular cell type.

REFERENCES

General

Garg, H.G., Cowman, M.K., and Hales, C.A. (Eds.), *Carbohydrate Chemistry, Biology and Medical Applications,* Elsevier (2008).

Lindhorst, T.K., *Essentials of Carbohydrate Chemistry and Biochemistry* (3rd ed.), Wiley-VCH (2007).

Solomons, T.W.G. and Fryhle, C., *Organic Chemistry* (9th ed.), Chapter 22, Wiley (2008). [A general discussion of carbohydrate nomenclature and chemistry. Other comprehensive organic chemistry textbooks have similar material.]

Taylor, M.E. and Drickamer, K., *Introduction to Glycobiology* (2nd ed.), Oxford University Press (2006).

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, S., Bertozzi, C.R., Hart, G.W., and Etzler, M.E. (Eds.), *Essentials of Glycobiology* (2nd ed.), Cold Spring Harbor Laboratory Press (2009).

Oligosaccharides and Polysaccharides

Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y., Cellulose, cellulases, and cellulosomes, *Curr. Opin. Struct. Biol.* **8,** 548 (1998).

Check, E., How Africa learned to love the cow, *Nature* **444,** 994–996 (2006). [Discusses the evolution of lactose tolerance in African cattle-herding populations.]

Haslam, S.M., North, S.J., and Dell, A., Mass spectrometric analysis of *N*- and *O*-glycosylation of tissues, *Curr. Opin. Struct. Biol.* **16,** 584–591 (2006).

Seeberger, P.H. and Werz, D.B., Synthesis and medical applications of oligosaccharides, *Nature* **446,** 1046–1055 (2007).

Sharon, N. and Lis, H., History of lectins: from hemagglutinins to biological recognition molecules, *Glycobiology* **14,** 53R–62R (2004). [An historical account of lectin research and applications.]

Weis, W.I. and Drickamer, K., Structural basis of lectin–carbohydrate recognition, *Annu. Rev. Biochem.* **65,** 441–473 (1996).

Glycoproteins

Bernfield, M., Götte, M., Park, P.W., Reizes, O., Fitzgerald, M.L., Linecum, J., and Zako, M., Functions of cell surface heparan sulfate proteoglycans, *Annu. Rev. Biochem.* **68,** 729–777 (1999).

Bishop, J.R., Schuksz, M., and Esko, J.D., Heparan sulphate proteoglycans fine-tune mammalian physiology, *Nature* **446,** 1030–1037 (2007). [Reviews the multiple roles of heparan sulfate proteoglycans.]

Bülow, H.E. and Hobert, O., The molecular diversity of glycosaminoglycans shapes animal development, *Annu. Rev. Cell Dev. Biol.* **22,** 375–407 (2006).

Bush, C.A., Martin-Pastor, M., and Imberty, A., Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides, *Annu. Rev. Biophys. Biomol. Struct.* **28,** 269–293 (1999).

Chain, E., Fleming's contribution to the discovery of penicillin, *Trends Biochem. Sci.* **4,** 143–146 (1979). [An historical account by one of the biochemists who characterized penicillin.]

Drickamer, K. and Taylor, M.E., Evolving views of protein glycosylation, *Trends Biochem. Sci.* **23,** 321–324 (1998).

Esko, J.D. and Lindahl, U., Molecular diversity of heparan sulfate, *J. Clin. Invest.* **108,** 169–173 (2001).

Fukuda, M. (Ed.), Glycobiology; Glycomics; and Functional Genomics, *Methods Enzymol.* **415–417** (2006).

Handel, T.M., Johnson, Z., Crown, S.E., Lau, E.K., Sweeney, M., and Proudfoot, A.E., Regulation of protein function by glycosaminoglycans—as exemplified by chemokines, *Annu. Rev. Biochem.* **74,** 385–410 (2005).

Hayhurst, E.J., Kailas, L., Hobbs, J.K., and Foster, S.J., Cell wall peptidoglycan architecture in *Bacillus subtilis, Proc. Natl. Acad. Sci.* **105,** 14603–14608 (2008).

Iozzo, R.V., Matrix proteoglycans: From molecular design to cellular function, *Annu. Rev. Biochem.* **67,** 609–652 (1998); *and* The biology of the small leucine-rich proteoglycans, *J. Biol. Chem.* **274,** 18843–18846 (1999).

Mitra, N., Sinha, S., Ramya, T.N.C., and Surolia, A., *N*-linked oligosaccharides as outfitters for glycoprotein folding, form and function, *Trends Biochem. Sci.* **31,** 156–163 and 251 (2006). [Summarizes the ways in which oligosaccharides can influence glycoprotein structure.]

Ohtsubo, K. and Marth, J.D., Glycosylation in cellular mechanisms of health and disease, *Cell* **126,** 855–867 (2006).

Perez-Vilar, J. and Hill, R.L., The structure and assembly of secreted mucins, *J. Biol. Chem.* **274,** 31751–31754 (1999).

Rudd, P.M. and Dwek, R.A., Rapid, sensitive sequencing of oligosaccharides from glycoproteins, *Curr. Opin. Biotechnol.* **8,** 488–497 (1997).

Sasisekharan, R., Raman, R., and Prabhakar, V., Glycomics approach to structure–function relationships of glycosaminoglycans, *Annu. Rev. Biomed. Eng.* **8,** 181–231 (2006).

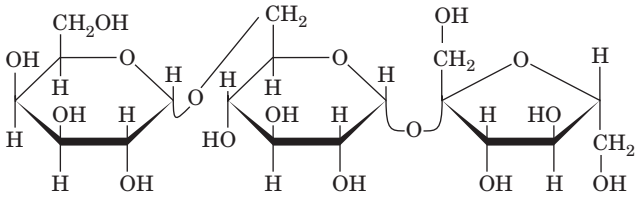
Spiro, R.G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, *Glycobiology* **12,** 43R–56R (2002). [Catalogs the various ways in which saccharides are linked to proteins, and describes the enzymes involved in glycoprotein synthesis.]

Varki, A., Nothing in glycobiology makes sense except in the light of evolution, *Cell* **126,** 841–845 (2006).

Wormald, M.R. and Dwek, R.A., Glycoproteins: Glycan presentation and protein-fold stability, *Structure* **7,** R155–R160 (1999).

PROBLEMS

1. The trisaccharide drawn below is named **raffinose.** What is its systematic name? Is it a reducing sugar?



Raffinose

2. The systematic name of **melezitose** is *O*-α-D-glucopyranosyl-(1 → 3)-*O*-β-D-fructofuranosyl-(2 → 1)-α-D-glucopyranoside. Draw its molecular formula. Is it a reducing sugar?

3. Name the linear form of D-glucose using the (*RS*) chirality nomenclature system. [See Section 4-2C. *Hint:* The branch toward C1 has higher priority than the branch toward C6.]

***4.** Draw the α-furanose form of D-talose and the β-pyranose form of L-sorbose.

5. The $NaBH_4$ reduction product of D-glucose may be named L-sorbitol or D-glucitol. Explain.

6. How many different disaccharides of D-glucopyranose are possible? How many trisaccharides?

7. A molecule of amylopectin consists of 1000 glucose residues and is branched every 25 residues. How many reducing ends does it have?

8. Most paper is made by removing the lignin from wood pulp and forming the resulting mass of largely unoriented cellulose fibers into a sheet. Untreated paper loses most of its strength when wet with water but maintains its strength when wet with oil. Explain.

***9.** Write a chemical mechanism for the acid-catalyzed mutarotation of glucose.

10. The values of the specific rotation, $[\alpha]_D^{20}$, for the α and β anomers of D-galactose are 150.7° and 52.8°, respectively. A mixture that is 20% α-D-galactose and 80% β-D-galactose is dissolved in water at 20° C. What is its initial specific rotation? After several hours, the specific rotation of this mixture reached an equilibrium value of 80.2°. What is its anomeric composition?

11. Name the epimers of D-gulose.

12. Exhaustive methylation of a trisaccharide followed by acid hydrolysis yields equimolar quantities of 2,3,4,6-tetra-*O*-methyl-D-galactose, 2,3,4-tri-*O*-methyl-D-mannose, and 2,4,6-tri-*O*-methyl-D-glucose. Treatment of the trisaccharide with β-galactosidase yields D-galactose and a disaccharide. Treatment of this disaccharide with α-mannosidase yields D-mannose and D-glucose. Draw the structure of the trisaccharide and state its systematic name.

13. The enzyme β-amylase cleaves successive maltose units from the nonreducing end of α(1 → 4) glucans. It will not cleave at glucose residues that have an α(1 → 6) bond. The end products of the exhaustive digestion of amylopectin by β-amylase are known as **limit dextrins.** Draw a schematic diagram of an amylopectin molecule and indicate what part(s) of it constitutes limit dextrins.

14. One demonstration of P.T. Barnum's maxim that there's a sucker born every minute is that new "reducing aids" regularly appear on the market. An eat-all-you-want nostrum, which was touted as a "starch blocker" [and which the Food and Drug Administration (FDA) eventually banned], contained an α-amylase-inhibiting protein extracted from beans. If this substance had really worked as advertised, which it did not, what unpleasant side effects would have resulted from its ingestion with a starch-containing meal? Discuss why this substance, which inhibits α-amylase *in vitro,* would not do so in the intestines after oral ingestion.

***15.** Treatment of a 6.0-g sample of glycogen with Tollens' reagent followed by exhaustive methylation and then hydrolysis yields 3.1 mmol of 2,3-di-*O*-methylglucose and 0.0031 mmol of 1,2,3-tri-*O*-methylgluconic acid as well as other products. (a) What fraction of glucose residues occur at (1 → 6) branch points, and what is the average number of glucose residues per branch? (b) What are the other products of the methylation–hydrolysis treatment and in what amounts are they formed? (c) What is the average molecular mass of the glycogen?

16. The lysis of a culture of *E. coli* yields a solution with mucuslike viscosity. Adding DNase to the solution greatly reduces this viscosity. What is the physical basis of the viscosity?

17. Instilling methyl-α-D-mannoside into the bladder of a mouse prevents the colonization of its urinary tract by *E. coli.* What is the reason for this effect?