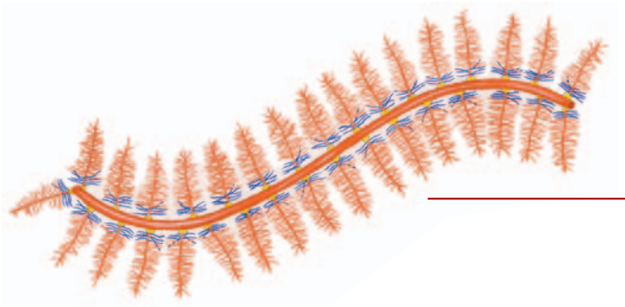


chapter 7

CARBOHYDRATES AND GLYCOBIOLOGY

Ah! sweet mystery of life . . .

—Rida Johnson Young (lyrics) and Victor Herbert (music), "Ah! Sweet Mystery of Life," 1910

I would feel more optimistic about a bright future for man if he spent less time proving that he can outwit Nature and more time tasting her sweetness and respecting her seniority.

—E. B. White, "Coon Tree," 1977

Carbohydrates are the most abundant biomolecules on Earth. Each year, photosynthesis converts more than 100 billion metric tons of CO_2 and H_2O into cellulose and other plant products. Certain carbohydrates (sugar and starch) are a dietary staple in most parts of the world, and the oxidation of carbohydrates is the central energy-yielding pathway in most nonphotosynthetic cells. Insoluble carbohydrate polymers serve as structural and protective elements in the cell walls of bacteria and plants and in the connective tissues of animals. Other carbohydrate polymers lubricate skeletal joints and participate in recognition and adhesion between cells. More complex carbohydrate polymers covalently attached to proteins or lipids act as signals that determine the intracellular location or metabolic fate of these hybrid molecules, called **glycoconjugates.** This chapter introduces the major classes of carbohydrates and glycoconjugates and provides a few examples of their many structural and functional roles.

Carbohydrates are polyhydroxy aldehydes or ketones, or substances that yield such compounds on hydrolysis. Many, but not all, carbohydrates have the empirical formula $(CH_2O)_n$; some also contain nitrogen, phosphorus, or sulfur.

There are three major size classes of carbohydrates: monosaccharides, oligosaccharides, and polysaccharides (the word "saccharide" is derived from the Greek *sakcharon*, meaning "sugar"). **Monosaccharides,** or simple sugars, consist of a single polyhydroxy aldehyde or ketone unit. The most abundant monosaccharide in nature is the six-carbon sugar D-glucose, sometimes referred to as dextrose. Monosaccharides of more than four carbons tend to have cyclic structures.

Oligosaccharides consist of short chains of monosaccharide units, or residues, joined by characteristic linkages called glycosidic bonds. The most abundant are the **disaccharides,** with two monosaccharide units. Typical is sucrose (cane sugar), which consists of the six-carbon sugars D-glucose and D-fructose. All common monosaccharides and disaccharides have names ending with the suffix "-ose." In cells, most oligosaccharides consisting of three or more units do not occur as free entities but are joined to nonsugar molecules (lipids or proteins) in glycoconjugates.

The **polysaccharides** are sugar polymers containing more than 20 or so monosaccharide units, and some have hundreds or thousands of units. Some polysaccharides, such as cellulose, are linear chains; others,

such as glycogen, are branched. Both glycogen and cellulose consist of recurring units of D-glucose, but they differ in the type of glycosidic linkage and consequently have strikingly different properties and biological roles.

7.1 Monosaccharides and Disaccharides

The simplest of the carbohydrates, the monosaccharides, are either aldehydes or ketones with two or more hydroxyl groups; the six-carbon monosaccharides glucose and fructose have five hydroxyl groups. Many of the carbon atoms to which hydroxyl groups are attached are chiral centers, which give rise to the many sugar stereoisomers found in nature. We begin by describing the families of monosaccharides with backbones of three to seven carbons—their structure and stereoisomeric forms, and the means of representing their three-dimensional structures on paper. We then discuss several chemical reactions of the carbonyl groups of monosaccharides. One such reaction, the addition of a hydroxyl group from within the same molecule, generates the cyclic forms of five- and six-carbon sugars (the forms that predominate in aqueous solution) and creates a new chiral center, adding further stereochemical complexity to this class of compounds. The nomenclature for unambiguously specifying the configuration about each carbon atom in a cyclic form and the means of representing these structures on paper are therefore described in some detail; this information will be useful as we discuss the metabolism of monosaccharides in Part II. We also introduce here some important monosaccharide derivatives encountered in later chapters.

The Two Families of Monosaccharides Are Aldoses and Ketoses

Monosaccharides are colorless, crystalline solids that are freely soluble in water but insoluble in nonpolar solvents. Most have a sweet taste. The backbones of common monosaccharide molecules are unbranched carbon chains in which all the carbon atoms are linked by single bonds. In the open-chain form, one of the carbon atoms is double-bonded to an oxygen atom to form a carbonyl group; each of the other carbon atoms has a hydroxyl group. If the carbonyl group is at an end of the carbon chain (that is, in an aldehyde group) the monosaccharide is an **aldose;** if the carbonyl group is at any other position (in a ketone group) the monosaccharide is a **ketose.** The simplest monosaccharides are the two three-carbon trioses: glyceraldehyde, an aldotriose, and dihydroxyacetone, a ketotriose (Fig. 7–1a).

Monosaccharides with four, five, six, and seven carbon atoms in their backbones are called, respectively, tetroses, pentoses, hexoses, and heptoses. There are aldoses and ketoses of each of these chain lengths:

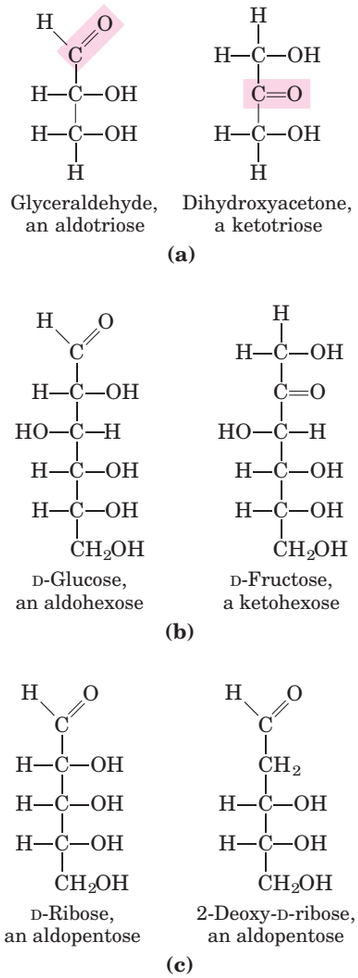
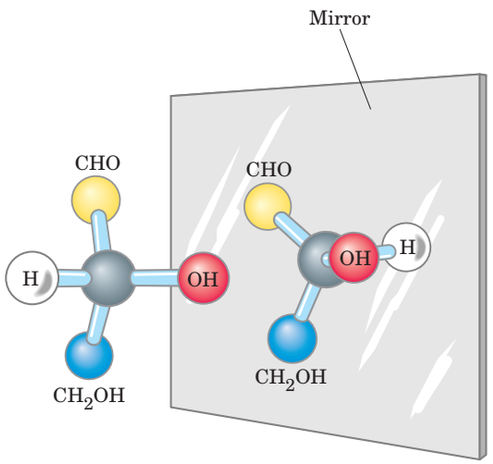


FIGURE 7–1 Representative monosaccharides. (a) Two trioses, an aldose and a ketose. The carbonyl group in each is shaded. **(b)** Two common hexoses. **(c)** The pentose components of nucleic acids. D-Ribose is a component of ribonucleic acid (RNA), and 2-deoxy-D-ribose is a component of deoxyribonucleic acid (DNA).

aldotetroses and ketotetroses, aldopentoses and ketopentoses, and so on. The hexoses, which include the aldohexose D-glucose and the ketohexose D-fructose (Fig. 7–1b), are the most common monosaccharides in nature. The aldopentoses D-ribose and 2-deoxy-D-ribose (Fig. 7–1c) are components of nucleotides and nucleic acids (Chapter 8).

Monosaccharides Have Asymmetric Centers

All the monosaccharides except dihydroxyacetone contain one or more asymmetric (chiral) carbon atoms and thus occur in optically active isomeric forms (pp. 17–19). The simplest aldose, glyceraldehyde, contains one chiral center (the middle carbon atom) and therefore has two different optical isomers, or enantiomers (Fig. 7–2).



Ball-and-stick models

Fischer projection formulas

Perspective formulas

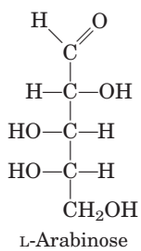
FIGURE 7–2 Three ways to represent the two stereoisomers of glyceraldehyde. The stereoisomers are mirror images of each other. Ball-and-stick models show the actual configuration of molecules. By convention, in Fischer projection formulas, horizontal bonds project out of the plane of the paper, toward the reader; vertical bonds project behind the plane of the paper, away from the reader. Recall (see Fig. 1–17) that in perspective formulas, solid wedge-shaped bonds point toward the reader, dashed wedges point away.

By convention, one of these two forms is designated the D isomer, the other the L isomer. As for other biomolecules with chiral centers, the absolute configurations of sugars are known from x-ray crystallography. To represent three-dimensional sugar structures on paper, we often use **Fischer projection formulas** (Fig. 7–2).

In general, a molecule with n chiral centers can have 2^n stereoisomers. Glyceraldehyde has $2^1 = 2$; the aldohexoses, with four chiral centers, have $2^4 = 16$ stereoisomers. The stereoisomers of monosaccharides of each carbon-chain length can be divided into two groups that differ in the configuration about the chiral center *most distant* from the carbonyl carbon. Those in which the configuration at this reference carbon is the same as that of D-glyceraldehyde are designated D isomers, and those with the same configuration as L-glyceraldehyde are L isomers. When the hydroxyl group on the reference carbon is on the right in the projection formula, the sugar is the D isomer; when on the left, it is the L isomer. Of the 16 possible aldohexoses, eight are D forms and eight are L. Most of the hexoses of living organisms are D isomers.

Figure 7–3 shows the structures of the D stereoisomers of all the aldoses and ketoses having three to six carbon atoms. The carbons of a sugar are numbered beginning at the end of the chain nearest the carbonyl group. Each of the eight D-aldohexoses, which differ in the stereochemistry at C-2, C-3, or C-4, has its own name: D-glucose, D-galactose, D-mannose, and so forth (Fig. 7–3a). The four- and five-carbon ketoses are designated by inserting "ul" into the name of a corresponding aldose; for example, D-ribulose is the ketopentose corresponding to the aldopentose D-ribose. The ketohexoses are named otherwise: for example, fructose (from the Latin *fructus,* "fruit"; fruits are rich in this sugar) and sorbose (from *Sorbus,* the genus of mountain ash, which has berries rich in the related sugar alcohol sorbitol). Two sugars that differ only in the configuration around one carbon atom are called **epimers;** D-glucose and D-mannose, which differ only in the stereochemistry at C-2, are epimers, as are D-glucose and D-galactose (which differ at C-4) (Fig. 7–4).

Some sugars occur naturally in their L form; examples are L-arabinose and the L isomers of some sugar derivatives that are common components of glycoconjugates (Section 7.3).



L-Arabinose

The Common Monosaccharides Have Cyclic Structures

For simplicity, we have thus far represented the structures of aldoses and ketoses as straight-chain molecules (Figs 7–3, 7–4). In fact, in aqueous solution, aldotetroses and all monosaccharides with five or more carbon atoms in the backbone occur predominantly as cyclic (ring) structures in which the carbonyl group has formed a covalent bond with the oxygen of a hydroxyl

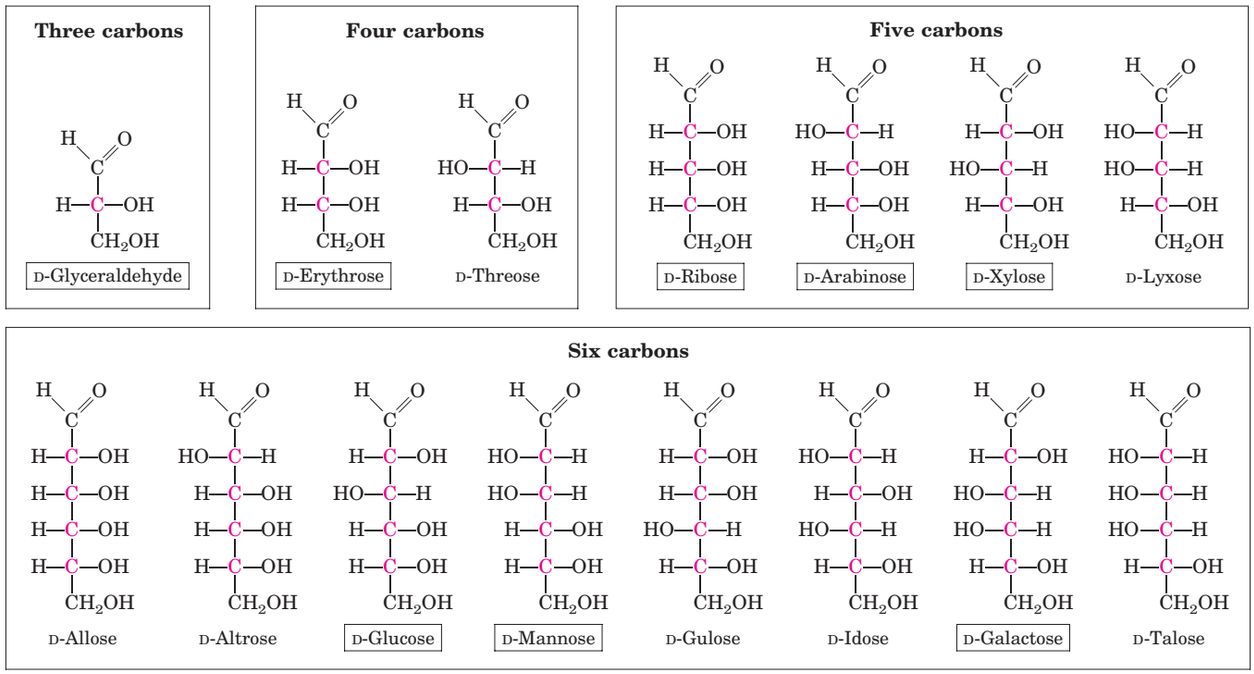


FIGURE 7–3 **Aldoses and ketoses.** The series of **(a)** D-aldoses and **(b)** D-ketoses having from three to six carbon atoms, shown as projection formulas. The carbon atoms in red are chiral centers. In all these D isomers, the chiral carbon *most distant from the carbonyl carbon* has the same configuration as the chiral carbon in D-glyceraldehyde. The sugars named in boxes are the most common in nature; you will encounter these again in this and later chapters.

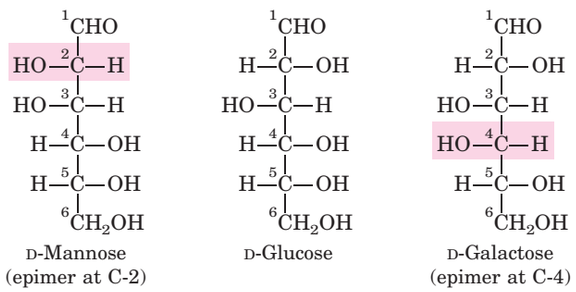


FIGURE 7–4 **Epimers.** D-Glucose and two of its epimers are shown as projection formulas. Each epimer differs from D-glucose in the configuration at one chiral center (shaded red).

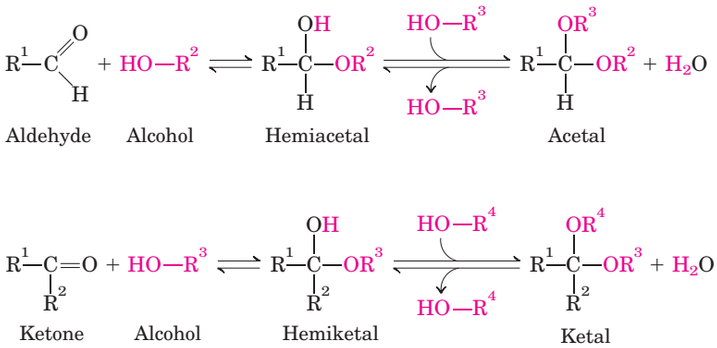


FIGURE 7–5 Formation of hemiacetals and hemiketals. An aldehyde or ketone can react with an alcohol in a 1:1 ratio to yield a hemiacetal or hemiketal, respectively, creating a new chiral center at the carbonyl carbon. Substitution of a second alcohol molecule produces an acetal or ketal. When the second alcohol is part of another sugar molecule, the bond produced is a glycosidic bond (p. 245).

group along the chain. The formation of these ring structures is the result of a general reaction between alcohols and aldehydes or ketones to form derivatives called **hemiacetals** or **hemiketals** (Fig. 7–5), which contain an additional asymmetric carbon atom and thus can exist in two stereoisomeric forms. For example, D-glucose exists in solution as an intramolecular hemiacetal in which the free hydroxyl group at C-5 has reacted with the aldehydic C-1, rendering the latter carbon asymmetric and producing two stereoisomers, designated α and β (Fig. 7–6). These six-membered ring compounds are called **pyranoses** because they resemble the six-membered ring compound pyran (Fig. 7–7). The systematic names for the two ring forms of D-glucose are α-D-glucopyranose and β-D-glucopyranose.

Aldohexoses also exist in cyclic forms having five-membered rings, which, because they resemble the five-membered ring compound furan, are called **furanoses.** However, the six-membered aldopyranose ring is much more stable than the aldofuranose ring and predominates in aldohexose solutions. Only aldoses having five or more carbon atoms can form pyranose rings.

Isomeric forms of monosaccharides that differ only in their configuration about the hemiacetal or hemiketal carbon atom are called **anomers.** The hemiacetal (or carbonyl) carbon atom is called the **anomeric carbon.** The α and β anomers of D-glucose interconvert in aqueous solution by a process called **mutarotation.** Thus, a solution of α-D-glucose and a solution of β-D-glucose eventually form identical equilibrium mixtures having identical optical properties. This mixture consists of about one-third α-D-glucose, two-thirds β-D-glucose, and very small amounts of the linear and five-membered ring (glucofuranose) forms.

Ketohexoses also occur in α and β anomeric forms. In these compounds the hydroxyl group at C-5 (or C-6) reacts with the keto group at C-2, forming a furanose (or pyranose) ring containing a hemiketal linkage (Fig. 7–5). D-Fructose readily forms the furanose ring (Fig. 7–7); the more common anomer of this sugar in combined forms or in derivatives is β-D-fructofuranose.

Haworth perspective formulas like those in Figure 7–7 are commonly used to show the stereochemistry of ring forms of monosaccharides. However, the six-membered pyranose ring is not planar, as Haworth perspectives suggest, but tends to assume either of two "chair" conformations (Fig. 7–8). Recall from Chapter 1 (p. 19) that two *conformations* of a molecule are interconvertible without the breakage of covalent bonds,

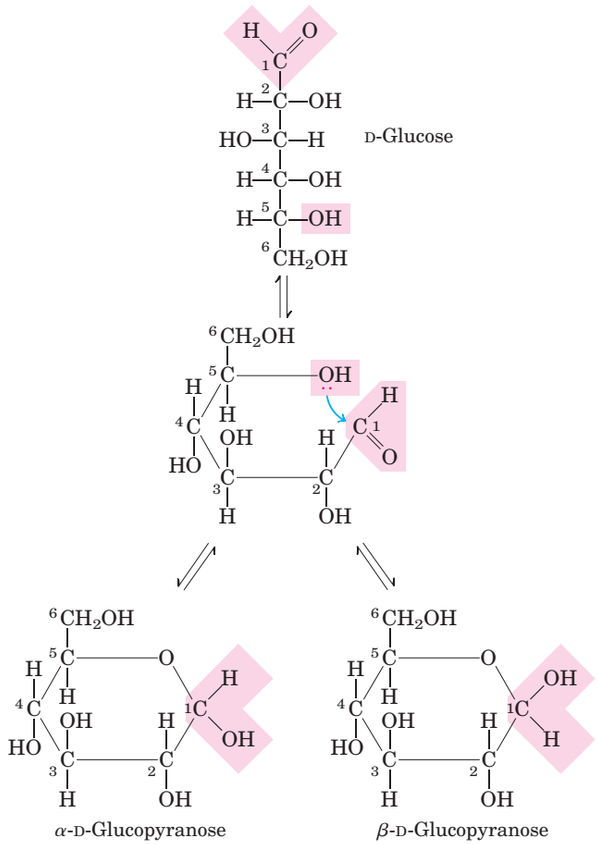


FIGURE 7–6 Formation of the two cyclic forms of D-glucose. Reaction between the aldehyde group at C-1 and the hydroxyl group at C-5 forms a hemiacetal linkage, producing either of two stereoisomers, the α and β anomers, which differ only in the stereochemistry around the hemiacetal carbon. The interconversion of α and β anomers is called mutarotation.

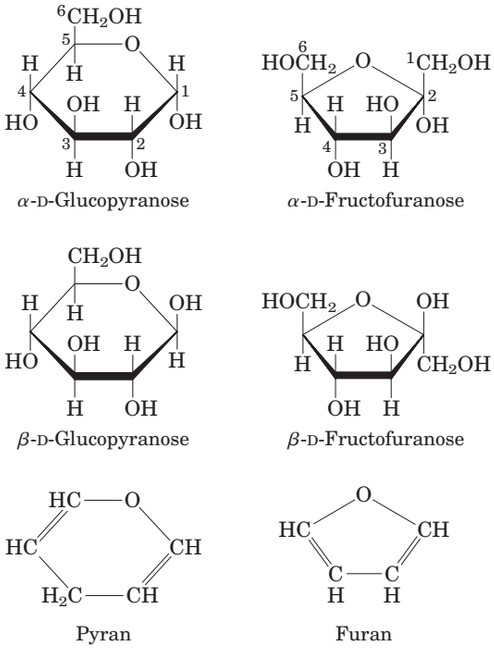


FIGURE 7–7 Pyranoses and furanoses. The pyranose forms of D-glucose and the furanose forms of D-fructose are shown here as Haworth perspective formulas. The edges of the ring nearest the reader are represented by bold lines. Hydroxyl groups below the plane of the ring in these Haworth perspectives would appear at the right side of a Fischer projection (compare with Fig. 7–6). Pyran and furan are shown for comparison.

whereas two *configurations* can be interconverted only by breaking a covalent bond—for example, in the case of α and β configurations, the bond involving the ring oxygen atom. The specific three-dimensional conformations of the monosaccharide units are important in determining the biological properties and functions of some polysaccharides, as we shall see.

Organisms Contain a Variety of Hexose Derivatives

In addition to simple hexoses such as glucose, galactose, and mannose, there are a number of sugar derivatives in which a hydroxyl group in the parent compound is replaced with another substituent, or a carbon atom is oxidized to a carboxyl group (Fig. 7–9). In glucosamine, galactosamine, and mannosamine, the hydroxyl at C-2 of the parent compound is replaced with an amino group. The amino group is nearly always condensed with acetic acid, as in *N*-acetylglucosamine. This glucosamine derivative is part of many structural polymers, including those of the bacterial cell wall. Bacterial cell walls also contain a derivative of glucosamine, *N*-acetylmuramic acid, in which lactic acid (a three-carbon carboxylic acid) is ether-linked to the oxygen at C-3 of *N*-acetylglucosamine. The substitution of a hydrogen for the hydroxyl group at C-6 of L-galactose or L-mannose produces L-fucose or L-rhamnose, respectively; these deoxy sugars are found in plant polysaccharides and in the complex oligosaccharide components of glycoproteins and glycolipids.

Oxidation of the carbonyl (aldehyde) carbon of glucose to the carboxyl level produces gluconic acid; other aldoses yield other **aldonic acids.** Oxidation of the carbon at the other end of the carbon chain—C-6 of glucose, galactose, or mannose—forms the corresponding **uronic acid:** glucuronic, galacturonic, or mannuronic acid. Both aldonic and uronic acids form stable intramolecular esters called lactones (Fig. 7–9, lower left). In addition to these acidic hexose derivatives, one nine-carbon acidic sugar deserves mention: *N*-acetylneuraminic acid (a sialic acid, but often referred to simply as "sialic acid"), a derivative of *N*-acetylmannosamine, is a component of many glycoproteins and glycolipids in animals. The carboxylic acid groups of the acidic sugar derivatives are ionized at pH 7, and the compounds are therefore correctly named as the carboxylates—glucuronate, galacturonate, and so forth.

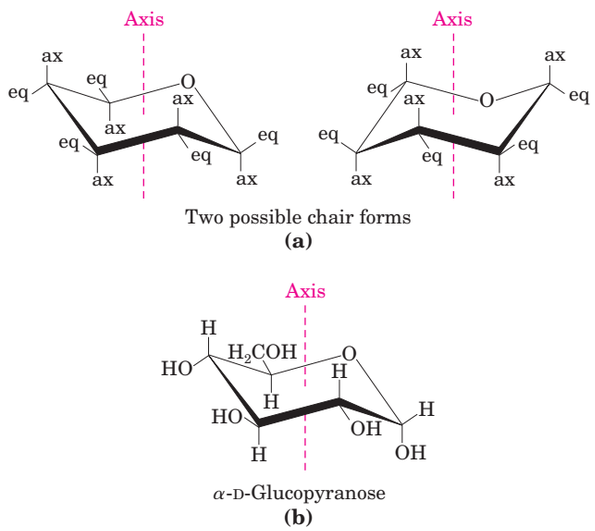


FIGURE 7–8 Conformational formulas of pyranoses. (a) Two chair forms of the pyranose ring. Substituents on the ring carbons may be either axial (ax), projecting parallel to the vertical axis through the ring, or equatorial (eq), projecting roughly perpendicular to this axis. Two *conformers* such are these are not readily interconvertible without breaking the ring. However, when the molecule is "stretched" (by atomic force microscopy), an input of about 46 kJ of energy per mole of sugar can force the interconversion of chair forms. Generally, substituents in the equatorial positions are less sterically hindered by neighboring substituents, and conformers with bulky substituents in equatorial positions are favored. Another conformation, the "boat" (not shown), is seen only in derivatives with very bulky substituents. **(b)** A chair conformation of α-D-glucopyranose.

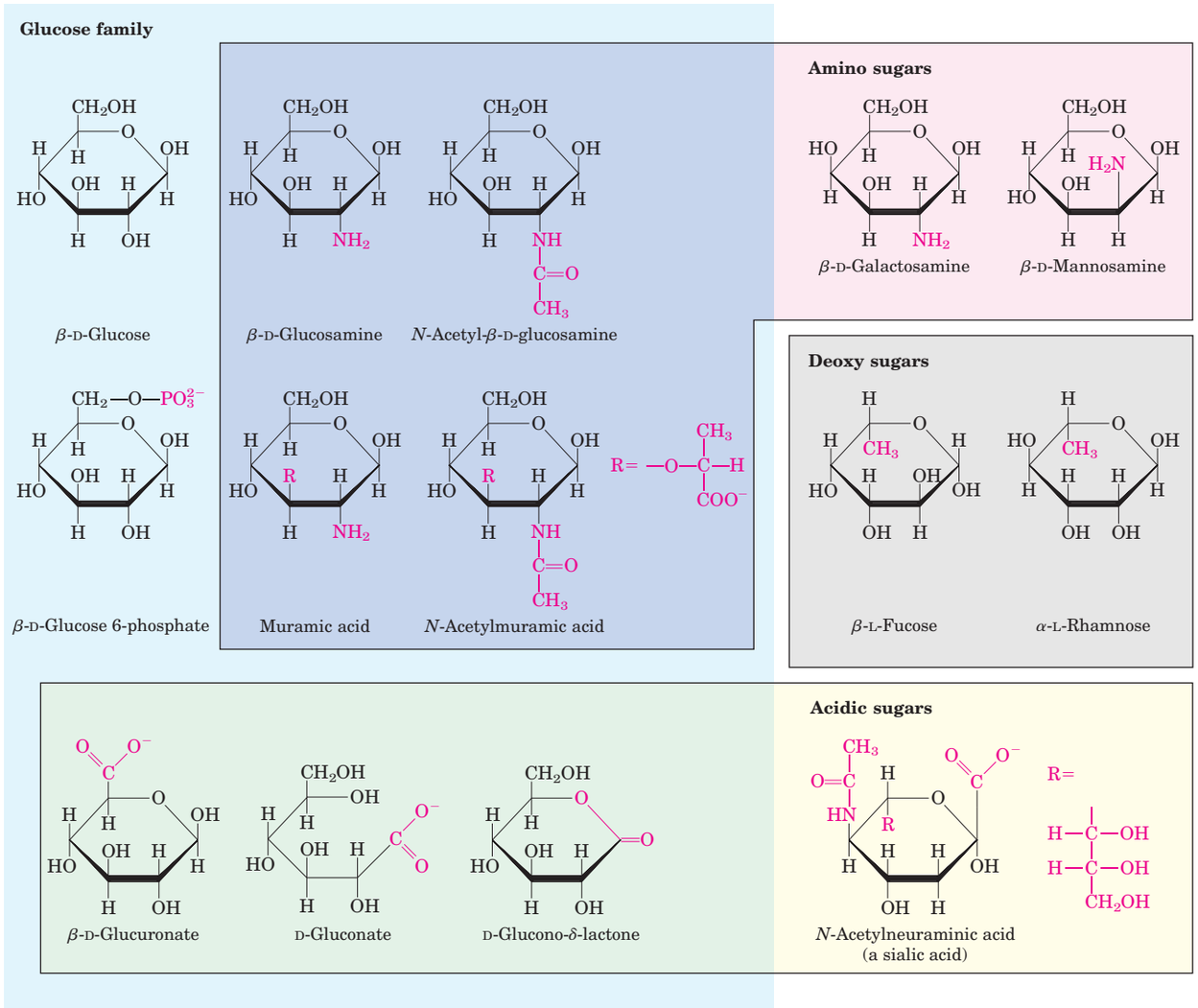


FIGURE 7–9 Some hexose derivatives important in biology. In amino sugars, an —NH_2 group replaces one of the —OH groups in the parent hexose. Substitution of —H for —OH produces a deoxy sugar; note that the deoxy sugars shown here occur in nature as the L isomers. The acidic sugars contain a carboxylate group, which confers a negative charge at neutral pH. D-Glucono-δ-lactone results from formation of an ester linkage between the C-1 carboxylate group and the C-5 (also known as the δ carbon) hydroxyl group of D-gluconate.

In the synthesis and metabolism of carbohydrates, the intermediates are very often not the sugars themselves but their phosphorylated derivatives. Condensation of phosphoric acid with one of the hydroxyl groups of a sugar forms a phosphate ester, as in glucose 6-phosphate (Fig. 7–9). Sugar phosphates are relatively stable at neutral pH and bear a negative charge. One effect of sugar phosphorylation within cells is to trap the sugar inside the cell; most cells do not have plasma membrane transporters for phosphorylated sugars. Phosphorylation also activates sugars for subsequent chemical transformation. Several important phosphorylated derivatives of sugars are components of nucleotides (discussed in the next chapter).

Monosaccharides Are Reducing Agents

Monosaccharides can be oxidized by relatively mild oxidizing agents such as ferric (Fe^{3+}) or cupric (Cu^{2+}) ion (Fig. 7–10a). The carbonyl carbon is oxidized to a carboxyl group. Glucose and other sugars capable of reducing ferric or cupric ion are called **reducing sugars.** This property is the basis of Fehling's reaction, a qualitative test for the presence of reducing sugar. By measuring the amount of oxidizing agent reduced by a solution of a sugar, it is also possible to estimate the concentration of that sugar. For many years this test was used to detect and measure elevated glucose levels in blood and urine in the diagnosis of dia-

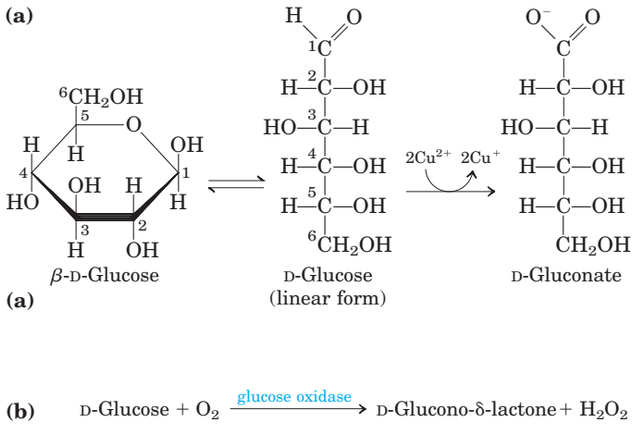


FIGURE 7–10 **Sugars as reducing agents. (a)** Oxidation of the anomeric carbon of glucose and other sugars is the basis for Fehling's reaction. The cuprous ion (Cu^{+}) produced under alkaline conditions forms a red cuprous oxide precipitate. In the hemiacetal (ring) form, C-1 of glucose cannot be oxidized by Cu^{2+}. However, the open-chain form is in equilibrium with the ring form, and eventually the oxidation reaction goes to completion. The reaction with Cu^{2+} is not as simple as the equation here implies; in addition to D-gluconate, a number of shorter-chain acids are produced by the fragmentation of glucose. **(b)** Blood glucose concentration is commonly determined by measuring the amount of H_2O_2 produced in the reaction catalyzed by glucose oxidase. In the reaction mixture, a second enzyme, peroxidase, catalyzes reaction of the H_2O_2 with a colorless compound to produce a colored compound, the amount of which is then measured spectrophotometrically.

betes mellitus. Now, more sensitive methods for measuring blood glucose employ an enzyme, glucose oxidase (Fig. 7–10b). ■

Disaccharides Contain a Glycosidic Bond

Disaccharides (such as maltose, lactose, and sucrose) consist of two monosaccharides joined covalently by an ***O*-glycosidic bond,** which is formed when a hydroxyl group of one sugar reacts with the anomeric carbon of the other (Fig. 7–11). This reaction represents the formation of an acetal from a hemiacetal (such as glucopyranose) and an alcohol (a hydroxyl group of the second sugar molecule) (Fig. 7–5). Glycosidic bonds are readily hydrolyzed by acid but resist cleavage by base. Thus disaccharides can be hydrolyzed to yield their free monosaccharide components by boiling with dilute acid. *N*-glycosyl bonds join the anomeric carbon of a sugar to a nitrogen atom in glycoproteins (see Fig. 7–31) and nucleotides (see Fig. 8–1).

The oxidation of a sugar's anomeric carbon by cupric or ferric ion (the reaction that defines a reducing sugar) occurs only with the linear form, which exists in equilibrium with the cyclic form(s). When the anomeric carbon is involved in a glycosidic bond, that sugar residue cannot take the linear form and therefore becomes a nonreducing sugar. In describing disaccharides or polysaccharides, the end of a chain with a free anomeric carbon (one not involved in a glycosidic bond) is commonly called the **reducing end.**

The disaccharide maltose (Fig. 7–11) contains two D-glucose residues joined by a glycosidic linkage between C-1 (the anomeric carbon) of one glucose residue and C-4 of the other. Because the disaccharide retains a free anomeric carbon (C-1 of the glucose residue on the right in Fig. 7–11), maltose is a reducing sugar. The configuration of the anomeric carbon atom in the glycosidic linkage is α. The glucose residue with the free anomeric carbon is capable of existing in α- and β-pyranose forms.

To name reducing disaccharides such as maltose unambiguously, and especially to name more complex oligosaccharides, several rules are followed. By convention, the name describes the compound with its nonreducing end to the left, and we can "build up" the name in the following order. (1) Give the configuration (α or β) at the anomeric carbon joining the first monosaccharide unit (on the left) to the second. (2) Name the

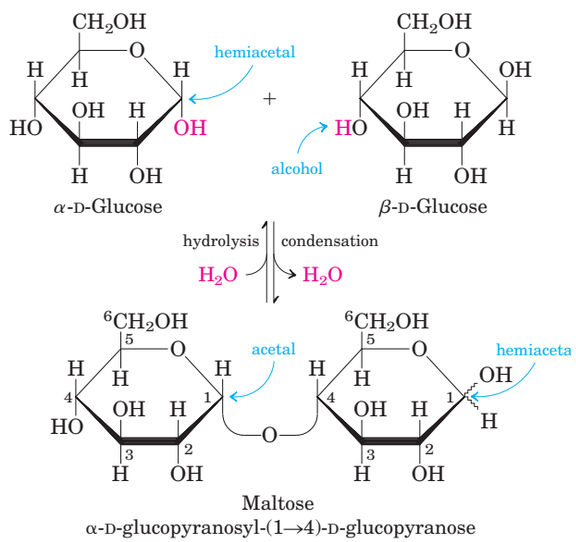


FIGURE 7–11 **Formation of maltose.** A disaccharide is formed from two monosaccharides (here, two molecules of D-glucose) when an —OH (alcohol) of one glucose molecule (right) condenses with the intramolecular hemiacetal of the other glucose molecule (left), with elimination of H_2O and formation of an *O*-glycosidic bond. The reversal of this reaction is hydrolysis—attack by H_2O on the glycosidic bond. The maltose molecule retains a reducing hemiacetal at the C-1 not involved in the glycosidic bond. Because mutarotation interconverts the α and β forms of the hemiacetal, the bonds at this position are sometimes depicted with wavy lines, as shown here, to indicate that the structure may be either α or β.

nonreducing residue; to distinguish five- and six-membered ring structures, insert "furano" or "pyrano" into the name. (3) Indicate in parentheses the two carbon atoms joined by the glycosidic bond, with an arrow connecting the two numbers; for example, (1→4) shows that C-1 of the first-named sugar residue is joined to C-4 of the second. (4) Name the second residue. If there is a third residue, describe the second glycosidic bond by the same conventions. (To shorten the description of complex polysaccharides, three-letter abbreviations for the monosaccharides are often used, as given in Table 7–1.) Following this convention for naming oligosaccharides, maltose is α-D-glucopyranosyl-(1→4)-D-glucopyranose. Because most sugars encountered in this book are the D enantiomers and the pyranose form of hexoses predominates, we generally use a shortened version of the formal name of such compounds, giving the configuration of the anomeric carbon and naming the carbons joined by the glycosidic bond. In this abbreviated nomenclature, maltose is Glc(α1→4)Glc.

The disaccharide lactose (Fig. 7–12), which yields D-galactose and D-glucose on hydrolysis, occurs naturally only in milk. The anomeric carbon of the glucose residue is available for oxidation, and thus lactose is a reducing disaccharide. Its abbreviated name is Gal(β1→4)Glc. Sucrose (table sugar) is a disaccharide of glucose and fructose. It is formed by plants but not by animals. In contrast to maltose and lactose, sucrose contains no free anomeric carbon atom; the anomeric carbons of both monosaccharide units are involved in the glycosidic bond (Fig. 7–12). Sucrose is therefore a nonreducing sugar. Nonreducing disaccharides are named as glycosides; in this case, the positions joined are the anomeric carbons. In the abbreviated nomenclature, a double-headed arrow connects the symbols specifying the anomeric carbons and their configurations. For example, the abbreviated name of sucrose is either Glc(α1↔2β)Fru or Fru(β2↔1α)Glc. Sucrose is a major intermediate product of photosynthesis; in many plants it is the principal form in which sugar is transported from the leaves to other parts of the plant body. Trehalose, Glc(α1↔1α)Glc (Fig. 7–12)—a disaccharide of D-glucose that, like sucrose, is a nonreducing sugar—is a major constituent of the circulating fluid (hemolymph) of insects, serving as an energy-storage compound.

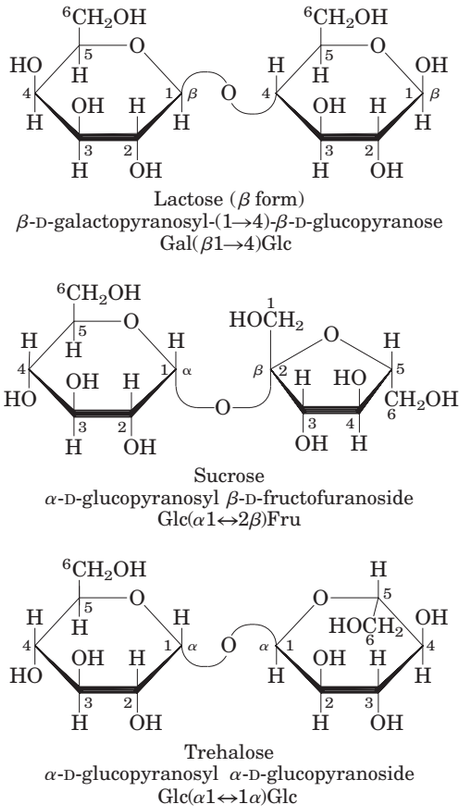


FIGURE 7–12 **Some common disaccharides.** Like maltose in Figure 7–11, these are shown as Haworth perspectives. The common name, full systematic name, and abbreviation are given for each disaccharide.

TABLE 7–1 **Abbreviations for Common Monosaccharides and Some of Their Derivatives**

Abequose	Abe	Glucuronic acid	GlcA
Arabinose	Ara	Galactosamine	GalN
Fructose	Fru	Glucosamine	GlcN
Fucose	Fuc	*N*-Acetylgalactosamine	GalNAc
Galactose	Gal	*N*-Acetylglucosamine	GlcNAc
Glucose	Glc	Iduronic acid	IdoA
Mannose	Man	Muramic acid	Mur
Rhamnose	Rha	*N*-Acetylmuramic acid	Mur2Ac
Ribose	Rib	*N*-Acetylneuraminic acid (a sialic acid)	Neu5Ac
Xylose	Xyl		

SUMMARY 7.1 Monosaccharides and Disaccharides

- Sugars (also called saccharides) are compounds containing an aldehyde or ketone group and two or more hydroxyl groups.
- Monosaccharides generally contain several chiral carbons and therefore exist in a variety of stereochemical forms, which may be represented on paper as Fischer projections. Epimers are sugars that differ in configuration at only one carbon atom.
- Monosaccharides commonly form internal hemiacetals or hemiketals, in which the aldehyde or ketone group joins with a hydroxyl group of the same molecule, creating a cyclic structure; this can be represented as a Haworth perspective formula. The carbon atom originally found in the aldehyde or ketone group (the anomeric carbon) can assume either of two configurations, α and β, which are interconvertible by mutarotation. In the linear form, which is in equilibrium with the cyclized forms, the anomeric carbon is easily oxidized.
- A hydroxyl group of one monosaccharide can add to the anomeric carbon of a second monosaccharide to form an acetal. In this disaccharide, the glycosidic bond protects the anomeric carbon from oxidation.
- Oligosaccharides are short polymers of several monosaccharides joined by glycosidic bonds. At one end of the chain, the reducing end, is a monosaccharide unit whose anomeric carbon is not involved in a glycosidic bond.
- The common nomenclature for di- or oligosaccharides specifies the order of monosaccharide units, the configuration at each anomeric carbon, and the carbon atoms involved in the glycosidic linkage(s).

7.2 Polysaccharides

Most carbohydrates found in nature occur as polysaccharides, polymers of medium to high molecular weight. Polysaccharides, also called **glycans,** differ from each other in the identity of their recurring monosaccharide units, in the length of their chains, in the types of bonds linking the units, and in the degree of branching. **Homopolysaccharides** contain only a single type of monomer; **heteropolysaccharides** contain two or more different kinds (Fig. 7–13). Some homopolysaccharides serve as storage forms of monosaccharides that are used as fuels; starch and glycogen are homopolysaccharides of this type. Other homopolysaccharides (cellulose and chitin, for example) serve as structural elements in plant cell walls and animal exoskeletons. Heteropolysaccharides provide extracellular support for organisms of all kingdoms. For example, the rigid layer of the bacterial cell envelope (the peptidoglycan) is composed in part of a heteropolysaccharide built from two alternating monosaccharide units. In animal tissues, the extracellular space is occupied by several types of heteropolysaccharides, which form a matrix that holds individual cells together and provides protection, shape, and support to cells, tissues, and organs.

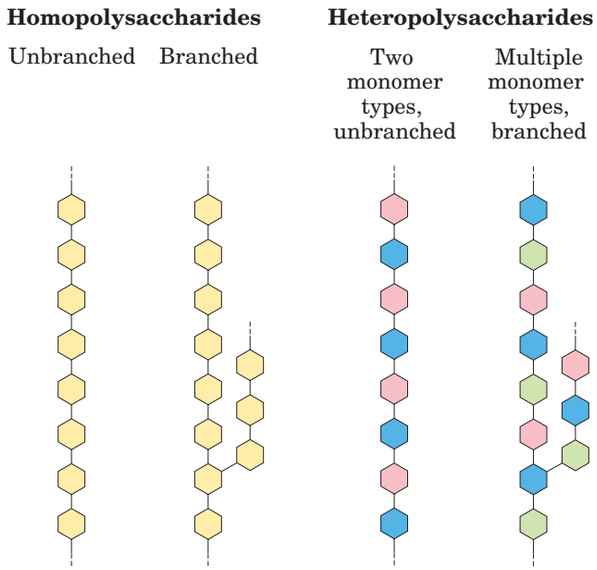


FIGURE 7–13 Homo- and heteropolysaccharides. Polysaccharides may be composed of one, two, or several different monosaccharides, in straight or branched chains of varying length.

Unlike proteins, polysaccharides generally do not have definite molecular weights. This difference is a consequence of the mechanisms of assembly of the two types of polymers. As we shall see in Chapter 27, proteins are synthesized on a template (messenger RNA) of defined sequence and length, by enzymes that follow the template exactly. For polysaccharide synthesis there is no template; rather, the program for polysaccharide synthesis is intrinsic to the enzymes that catalyze the polymerization of the monomeric units, and there is no specific stopping point in the synthetic process.

Some Homopolysaccharides Are Stored Forms of Fuel

The most important storage polysaccharides are starch in plant cells and glycogen in animal cells. Both polysaccharides occur intracellularly as large clusters or granules (Fig. 7–14). Starch and glycogen molecules are heavily hydrated, because they have many exposed hydroxyl groups available to hydrogen-bond with water. Most plant cells have the ability to form starch, but it is

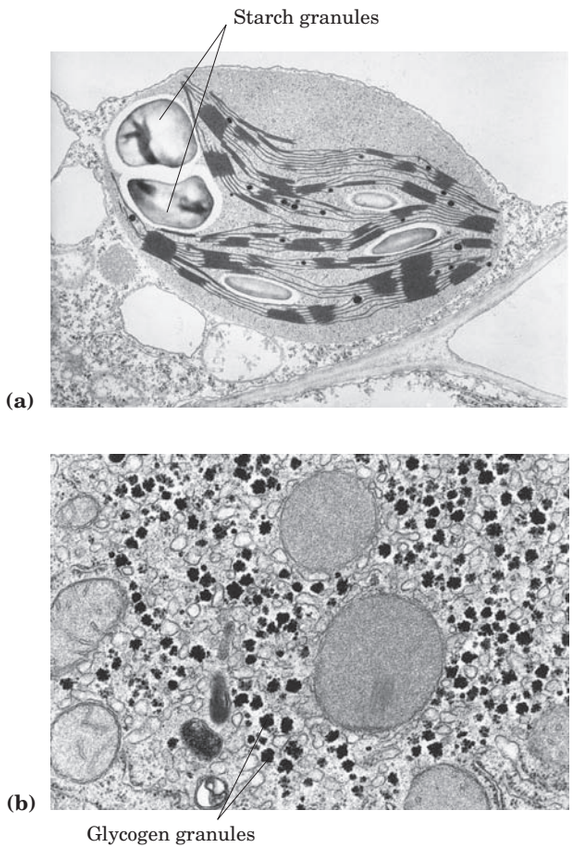


FIGURE 7–14 Electron micrographs of starch and glycogen granules. **(a)** Large starch granules in a single chloroplast. Starch is made in the chloroplast from D-glucose formed photosynthetically. **(b)** Glycogen granules in a hepatocyte. These granules form in the cytosol and are much smaller (~0.1 μm) than starch granules (~1.0 μm).

especially abundant in tubers, such as potatoes, and in seeds.

Starch contains two types of glucose polymer, amylose and amylopectin (Fig. 7–15). The former consists of long, unbranched chains of D-glucose residues connected by (α1→4) linkages. Such chains vary in molecular weight from a few thousand to more than a million. Amylopectin also has a high molecular weight (up to 100 million) but unlike amylose is highly branched. The glycosidic linkages joining successive glucose residues in amylopectin chains are (α1→4); the branch points (occurring every 24 to 30 residues) are (α1→6) linkages.

Glycogen is the main storage polysaccharide of animal cells. Like amylopectin, glycogen is a polymer of (α1→4)-linked subunits of glucose, with (α1→6)-linked branches, but glycogen is more extensively branched (on average, every 8 to 12 residues) and more compact than starch. Glycogen is especially abundant in the liver, where it may constitute as much as 7% of the wet weight; it is also present in skeletal muscle. In hepatocytes glycogen is found in large granules (Fig. 7–14b), which are themselves clusters of smaller granules composed of single, highly branched glycogen molecules with an average molecular weight of several million. Such glycogen granules also contain, in tightly bound form, the enzymes responsible for the synthesis and degradation of glycogen.

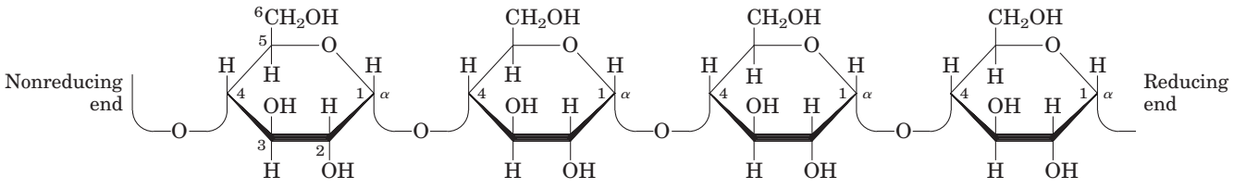
Because each branch in glycogen ends with a nonreducing sugar unit, a glycogen molecule has as many nonreducing ends as it has branches, but only one reducing end. When glycogen is used as an energy source, glucose units are removed one at a time from the nonreducing ends. Degradative enzymes that act only at nonreducing ends can work simultaneously on the many branches, speeding the conversion of the polymer to monosaccharides.

Why not store glucose in its monomeric form? It has been calculated that hepatocytes store glycogen equivalent to a glucose concentration of 0.4 M. The actual concentration of glycogen, which is insoluble and contributes little to the osmolarity of the cytosol, is about 0.01 μM. If the cytosol contained 0.4 M glucose, the osmolarity would be threateningly elevated, leading to osmotic entry of water that might rupture the cell (see Fig. 2–13). Furthermore, with an intracellular glucose concentration of 0.4 M and an external concentration of about 5 mM (the concentration in the blood of a mammal), the free-energy change for glucose uptake into cells against this very high concentration gradient would be prohibitively large.

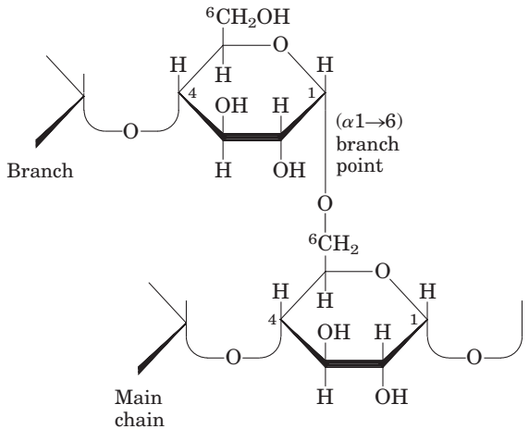
Dextrans are bacterial and yeast polysaccharides made up of (α1→6)-linked poly-D-glucose; all have (α1→3) branches, and some also have (α1→2) or (α1→4) branches. Dental plaque, formed by bacteria growing on the surface of teeth, is rich in dextrans. Synthetic dextrans are used in several commercial products (for example, Sephadex) that serve in the fractionation of proteins by size-exclusion chromatography (see Fig. 3–18b). The dextrans in these products are chemically cross-linked to form insoluble materials of various porosities, admitting macromolecules of various sizes.

Some Homopolysaccharides Serve Structural Roles

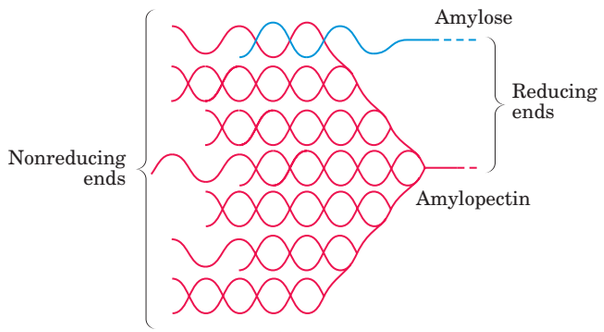
Cellulose, a fibrous, tough, water-insoluble substance, is found in the cell walls of plants, particularly in stalks, stems, trunks, and all the woody portions of the plant body. Cellulose constitutes much of the mass of wood, and cotton is almost pure cellulose. Like amylose and the main chains of amylopectin and glycogen, the cellulose molecule is a linear, unbranched homopolysaccharide, consisting of 10,000 to 15,000 D-glucose units. But there is a very important difference: in cellulose the glucose residues have the β configuration (Fig. 7–16),



(a) amylose



(b)

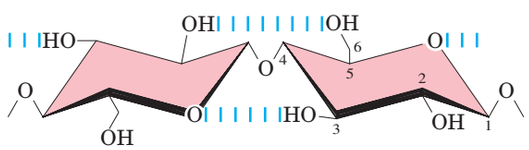


(c)

FIGURE 7–15 Amylose and amylopectin, the polysaccharides of starch. (a) A short segment of amylose, a linear polymer of D-glucose residues in (α1→4) linkage. A single chain can contain several thousand glucose residues. Amylopectin has stretches of similarly linked residues between branch points. **(b)** An (α1→6) branch point of amylopectin. **(c)** A cluster of amylose and amylopectin like that believed to occur in starch granules. Strands of amylopectin (red) form double-helical structures with each other or with amylose strands (blue). Glucose residues at the nonreducing ends of the outer branches are removed enzymatically during the mobilization of starch for energy production. Glycogen has a similar structure but is more highly branched and more compact.

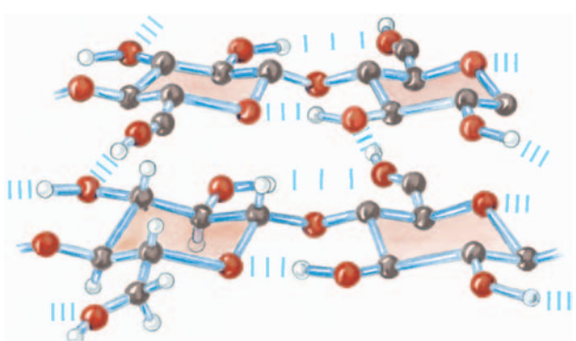
whereas in amylose, amylopectin, and glycogen the glucose is in the α configuration. The glucose residues in cellulose are linked by (β1→4) glycosidic bonds, in contrast to the (α1→4) bonds of amylose, starch, and glycogen. This difference gives cellulose and amylose very different structures and physical properties.

Glycogen and starch ingested in the diet are hydrolyzed by α-amylases, enzymes in saliva and intestinal secretions that break (α1→4) glycosidic bonds between glucose units. Most animals cannot use cellulose as a fuel source, because they lack an enzyme to hydrolyze the (β1→4) linkages. Termites readily digest cellulose



(β1→4)-linked D-glucose units

(a)



(b)

FIGURE 7–16 The structure of cellulose. (a) Two units of a cellulose chain; the D-glucose residues are in (β1→4) linkage. The rigid chair structures can rotate relative to one another. **(b)** Scale drawing of segments of two parallel cellulose chains, showing the conformation of the D-glucose residues and the hydrogen-bond cross-links. In the hexose unit at the lower left, all hydrogen atoms are shown; in the other three hexose units, the hydrogens attached to carbon have been omitted for clarity as they do not participate in hydrogen bonding.

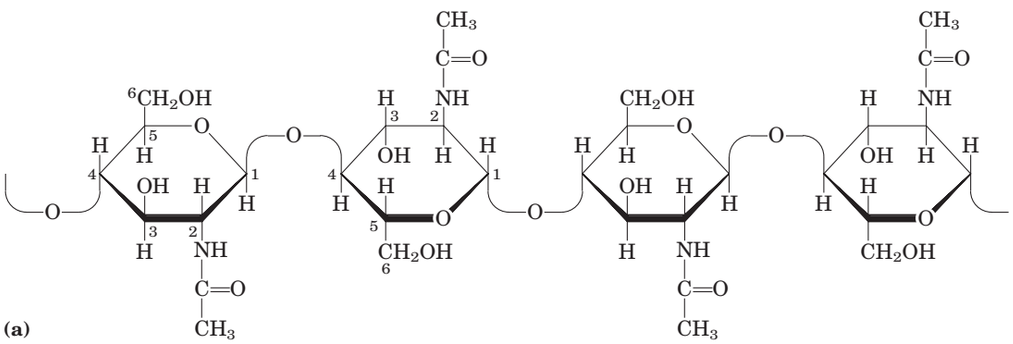
FIGURE 7–17 Cellulose breakdown by wood fungi. A wood fungus growing on an oak log. All wood fungi have the enzyme cellulase, which breaks the (β1→4) glycosidic bonds in cellulose, such that wood is a source of metabolizable sugar (glucose) for the fungus. The only vertebrates able to use cellulose as food are cattle and other ruminants (sheep, goats, camels, giraffes). The extra stomach compartment (rumen) of a ruminant teems with bacteria and protists that secrete cellulase.

(and therefore wood), but only because their intestinal tract harbors a symbiotic microorganism, *Trichonympha*, that secretes cellulase, which hydrolyzes the (β1→4) linkages. Wood-rot fungi and bacteria also produce cellulase (Fig. 7–17).

Chitin is a linear homopolysaccharide composed of *N*-acetylglucosamine residues in β linkage (Fig. 7–18). The only chemical difference from cellulose is the replacement of the hydroxyl group at C-2 with an acetylated amino group. Chitin forms extended fibers similar to those of cellulose, and like cellulose cannot be digested by vertebrates. Chitin is the principal component of the hard exoskeletons of nearly a million species of arthropods—insects, lobsters, and crabs, for example—and is probably the second most abundant polysaccharide, next to cellulose, in nature.

Steric Factors and Hydrogen Bonding Influence Homopolysaccharide Folding

The folding of polysaccharides in three dimensions follows the same principles as those governing polypeptide structure: subunits with a more-or-less rigid structure dictated by covalent bonds form three-dimensional macromolecular structures that are stabilized by weak interactions within or between molecules: hydrogen-bond, hydrophobic, and van der Waals interactions, and, for polymers with charged subunits, electrostatic interactions. Because polysaccharides have so many hydroxyl groups, hydrogen bonding has an especially important influence on their structure. Glycogen, starch, and cellulose are composed of pyranoside subunits (having six-membered rings), as are the oligosaccharides of glycoproteins and glycolipids to be discussed later. Such molecules can be represented as a series of rigid pyranose rings connected by an oxygen atom bridging two carbon atoms (the glycosidic bond). There is, in princi-



(a)

(b)

FIGURE 7–18 Chitin. (a) A short segment of chitin, a homopolymer of *N*-acetyl-D-glucosamine units in (β1→4) linkage. **(b)** A spotted June beetle (*Pellidnota punetatia*), showing its surface armor (exoskeleton) of chitin.

ple, free rotation about both C—O bonds linking the residues (Fig. 7–16a), but as in polypeptides (see Figs 4–2, 4–9), rotation about each bond is limited by steric hindrance by substituents. The three-dimensional structures of these molecules can be described in terms of the dihedral angles, ϕ and ψ, made with the glycosidic bond (Fig. 7–19), analogous to angles ϕ and ψ made by the peptide bond (see Fig. 4–2). Because of the bulkiness of the pyranose ring and its substituents, their size and shape place constraints on the angles ϕ and ψ; certain conformations are much more stable than others, as can be shown on a map of energy as a function of ϕ and ψ (Fig. 7–20).

The most stable three-dimensional structure for starch and glycogen is a tightly coiled helix (Fig. 7–21), stabilized by interchain hydrogen bonds. In amylose (with no branches) this structure is regular enough to allow crystallization and thus determination of the structure by x-ray diffraction. Each residue along the amylose chain forms a 60° angle with the preceding residue, so the helical structure has six residues per turn. For amylose, the core of the helix is of precisely the right dimensions to accommodate iodine in the form I^{3-} or I^{5-}(iodide ions), and this interaction with iodine is a common qualitative test for amylose.

For cellulose, the most stable conformation is that in which each chair is turned 180° relative to its neighbors, yielding a straight, extended chain. All —OH groups are available for hydrogen bonding with neighboring chains. With several chains lying side by side, a stabilizing network of interchain and intrachain hydrogen bonds produces straight, stable supramolecular

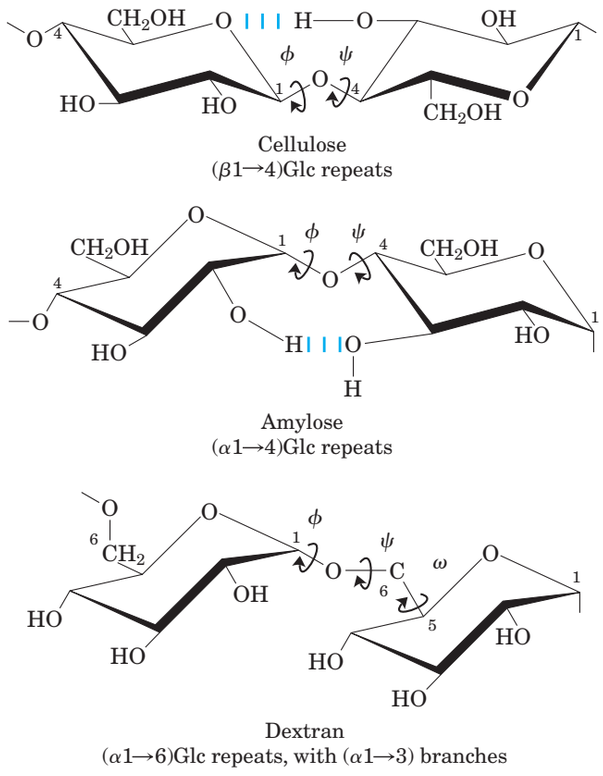


FIGURE 7–19 Conformation at the glycosidic bonds of cellulose, amylose, and dextran. The polymers are depicted as rigid pyranose rings joined by glycosidic bonds, with free rotation about these bonds. Note that in dextran there is also free rotation about the bond between C-5 and C-6 (torsion angle ω (omega)).

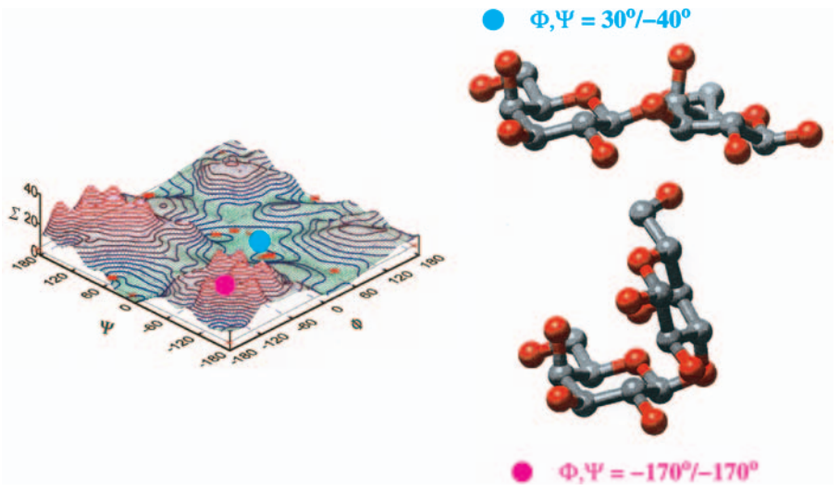


FIGURE 7–20 A map of favored conformations for oligosaccharides and polysaccharides. The torsion angles ψ and ϕ (see Fig. 7–19), which define the spatial relationship between adjacent rings, can in principle have any value from 0° to 360°. In fact, some of the torsion angles would give conformations that are sterically hindered, whereas others give conformations that maximize hydrogen bonding. When the relative energy is plotted for each value of ϕ and ψ, with isoenergy ("same energy") contours drawn at intervals of 1 kcal/mol above the minimum energy state, the result is a map of preferred conformations. This is analogous to the Ramachandran plot for peptides (see Figs 4–3, 4–9). The known conformations of the three polysaccharides shown in Figure 7–19 have been determined by x-ray crystallography, and all fall within the lowest-energy regions of the map.

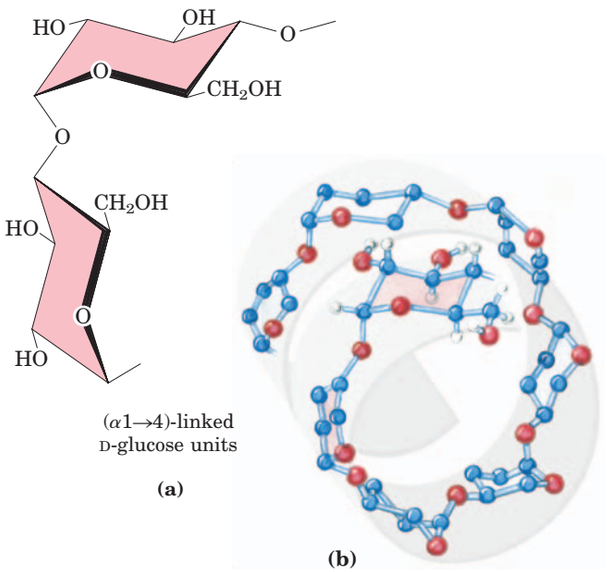


FIGURE 7–21 The structure of starch (amylose). (a) In the most stable conformation, with adjacent rigid chairs, the polysaccharide chain is curved, rather than linear as in cellulose (see Fig. 7–16). **(b)** Scale drawing of a segment of amylose. The conformation of (α1→4) linkages in amylose, amylopectin, and glycogen causes these polymers to assume tightly coiled helical structures. These compact structures produce the dense granules of stored starch or glycogen seen in many cells (see Fig. 7–14).

fibers of great tensile strength (Fig. 7–16b). This property of cellulose has made it a useful substance to civilizations for millennia. Many manufactured products, including papyrus, paper, cardboard, rayon, insulating tiles, and a variety of other useful materials, are derived from cellulose. The water content of these materials is low because extensive interchain hydrogen bonding between cellulose molecules satisfies their capacity for hydrogen-bond formation.

Bacterial and Algal Cell Walls Contain Structural Heteropolysaccharides

The rigid component of bacterial cell walls is a heteropolymer of alternating (β1→4)-linked *N*-acetylglucosamine and *N*-acetylmuramic acid residues (Fig. 7–22). The linear polymers lie side by side in the cell wall, cross-linked by short peptides, the exact structure of which depends on the bacterial species. The peptide cross-links weld the polysaccharide chains into a strong sheath that envelops the entire cell and prevents cellular swelling and lysis due to the osmotic entry of water. The enzyme lysozyme kills bacteria by hydrolyzing the (β1→4) glycosidic bond between *N*-acetylglucosamine and *N*-acetylmuramic acid (see Fig. 6–24). Lysozyme is notably present in tears, presumably as a defense against bacterial infections of the eye. It is also produced by certain bacterial viruses to ensure their release from the host bacterial cell, an essential step of the viral infection cycle. Penicillin and related antibiotics kill bacteria by preventing synthesis of the cross-links, leaving the cell wall too weak to resist osmotic lysis (see Box 20–1).

Certain marine red algae, including some of the seaweeds, have cell walls that contain **agar,** a mixture of sulfated heteropolysaccharides made up of D-galactose and an L-galactose derivative ether-linked between C-3 and C-6 (Fig. 7–23). The two major components of agar are the unbranched polymer **agarose** (M_r ~120,000) and a branched component, agaropectin. The remarkable gel-forming property of agarose makes it useful in the biochemistry laboratory. When a suspension of agarose in water is heated and cooled, the agarose forms a double helix: two molecules in parallel orientation twist together with a helix repeat of three residues; water molecules are trapped in the central cavity. These struc-

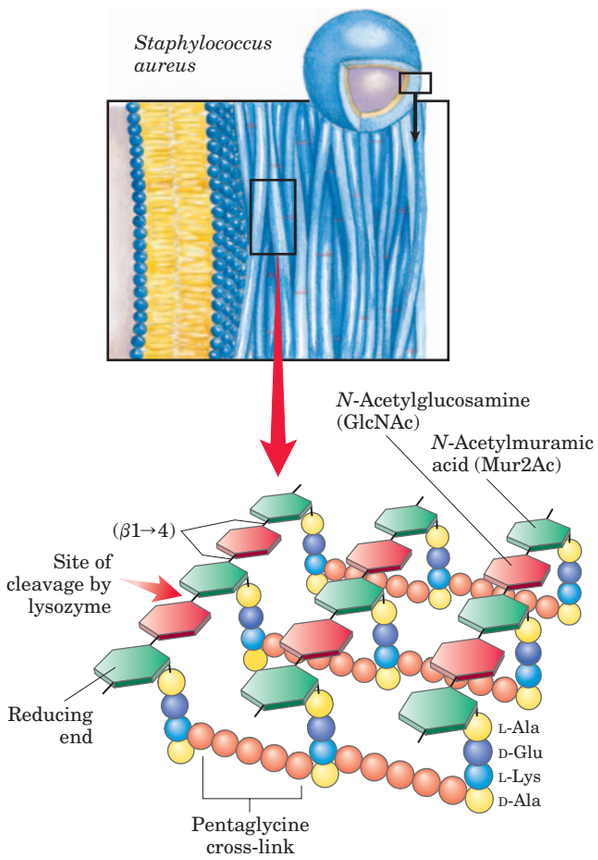
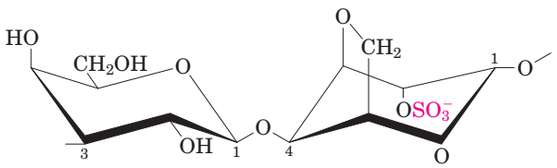


FIGURE 7–22 Peptidoglycan. Shown here is the peptidoglycan of the cell wall of *Staphylococcus aureus,* a gram-positive bacterium. Peptides (strings of colored spheres) covalently link *N*-acetylmuramic acid residues in neighboring polysaccharide chains. Note the mixture of L and D amino acids in the peptides. Gram-positive bacteria have a pentaglycine chain in the cross-link. Gram-negative bacteria, such as *E. coli,* lack the pentaglycine; instead, the terminal D-Ala residue of one tetrapeptide is attached directly to a neighboring tetrapeptide through either L-Lys or a lysine-like amino acid, diaminopimelic acid.



Agarose
3)D-Gal(β1→4)3,6-anhydro-L-Gal2S(α1 repeats

FIGURE 7–23 The structure of agarose. The repeating unit consists of D-galactose (β1→4)-linked to 3,6-anhydro-L-galactose (in which an ether ring connects C-3 and C-6). These units are joined by (α1→3) glycosidic links to form a polymer 600 to 700 residues long. A small fraction of the 3,6-anhydrogalactose residues have a sulfate ester at C-2 (as shown here).

tures in turn associate with each other to form a gel—a three-dimensional matrix that traps large amounts of water. Agarose gels are used as inert supports for the electrophoretic separation of nucleic acids, an essential part of the DNA sequencing process (p. 8-24). Agar is also used to form a surface for the growth of bacterial colonies. Another commercial use of agar is for the capsules in which some vitamins and drugs are packaged; the dried agar material dissolves readily in the stomach and is metabolically inert.

Glycosaminoglycans Are Heteropolysaccharides of the Extracellular Matrix

The extracellular space in the tissues of multicellular animals is filled with a gel-like material, the **extracellular matrix,** also called ground substance, which holds the cells together and provides a porous pathway for the diffusion of nutrients and oxygen to individual cells. The extracellular matrix is composed of an interlocking meshwork of heteropolysaccharides and fibrous proteins such as collagen, elastin, fibronectin, and laminin. These heteropolysaccharides, the **glycosaminoglycans,** are a family of linear polymers composed of repeating disaccharide units (Fig. 7–24). One of the two monosaccharides is always either *N*-acetylglucosamine or *N*-acetylgalactosamine; the other is in most cases a uronic acid, usually D-glucuronic or L-iduronic acid. In some glycosaminoglycans, one or more of the hydroxyls of the amino sugar are esterified with sulfate. The combination of sulfate groups and the carboxylate groups of the uronic acid residues gives glycosaminoglycans a very high density of negative charge. To minimize the repulsive forces among neighboring charged groups, these molecules assume an extended conformation in solution. The specific patterns of sulfated and nonsulfated sugar residues in glycosaminoglycans provide for specific recognition by a

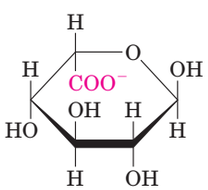
FIGURE 7–24 Repeating units of some common glycosaminoglycans of extracellular matrix. The molecules are copolymers of alternating uronic acid and amino sugar residues, with sulfate esters in any of several positions. The ionized carboxylate and sulfate groups (red) give these polymers their characteristic high negative charge. Heparin contains primarily iduronic acid (IdoA) and a smaller proportion of glucuronic acid (GlcA), and is generally highly sulfated and heterogeneous in length. Heparan sulfate (not shown) is similar to heparin but has a higher proportion of GlcA and fewer sulfate groups, arranged in a less regular pattern.

Glycosaminoglycan	Number of disaccharides per chain	Repeating disaccharide
Hyaluronate	~50,000	GlcA, GlcNAc
Chondroitin 4-sulfate	20–60	GlcA, GalNAc4S
Keratan sulfate	~25	Gal, GlcNAc6S
Heparin	15–90	GlcA2S or IdoA2S, GlcNS3S6S

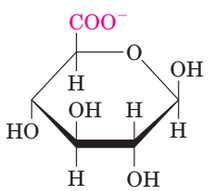
variety of protein ligands that bind electrostatically to these molecules. Glycosaminoglycans are attached to extracellular proteins to form proteoglycans (Section 7.3).

The glycosaminoglycan **hyaluronic acid** (hyaluronate at physiological pH) contains alternating residues of D-glucuronic acid and *N*-acetylglucosamine (Fig. 7–24). With up to 50,000 repeats of the basic disaccharide unit, hyaluronates have molecular weights greater than 1 million; they form clear, highly viscous solutions that serve as lubricants in the synovial fluid of joints and give the vitreous humor of the vertebrate eye its jelly-like consistency (the Greek *hyalos* means "glass"; hyaluronates can have a glassy or translucent appearance). Hyaluronate is also an essential component of the extracellular matrix of cartilage and tendons, to which it contributes tensile strength and elasticity as a result of its strong interactions with other components of the matrix. Hyaluronidase, an enzyme secreted by some pathogenic bacteria, can hydrolyze the glycosidic linkages of hyaluronate, rendering tissues more susceptible to bacterial invasion. In many organisms, a similar enzyme in sperm hydrolyzes an outer glycosaminoglycan coat around the ovum, allowing sperm penetration.

Other glycosaminoglycans differ from hyaluronate in two respects: they are generally much shorter polymers and they are covalently linked to specific proteins (proteoglycans). Chondroitin sulfate (Greek *chondros,* "cartilage") contributes to the tensile strength of cartilage, tendons, ligaments, and the walls of the aorta. Dermatan sulfate (Greek *derma,* "skin") contributes to the pliability of skin and is also present in blood vessels and heart valves. In this polymer, many of the glucuronate (GlcA) residues present in chondroitin sulfate are replaced by their epimer, iduronate (IdoA).



β-L-Iduronate (IdoA)



β-D-Glucuronate (GlcA)

Keratan sulfates (Greek *keras,* "horn") have no uronic acid and their sulfate content is variable. They are present in cornea, cartilage, bone, and a variety of horny structures formed of dead cells: horn, hair, hoofs, nails, and claws. Heparin (Greek *hēpar,* "liver") is a natural anticoagulant made in mast cells (a type of leukocyte) and released into the blood, where it inhibits blood coagulation by binding to the protein antithrombin. Heparin binding causes antithrombin to bind to and inhibit thrombin, a protease essential to blood clotting. The interaction is strongly electrostatic; heparin has the highest negative charge density of any known biological macromolecule (Fig. 7–25). Purified heparin is routinely added to blood samples obtained for clinical analysis, and to blood donated for transfusion, to prevent clotting.

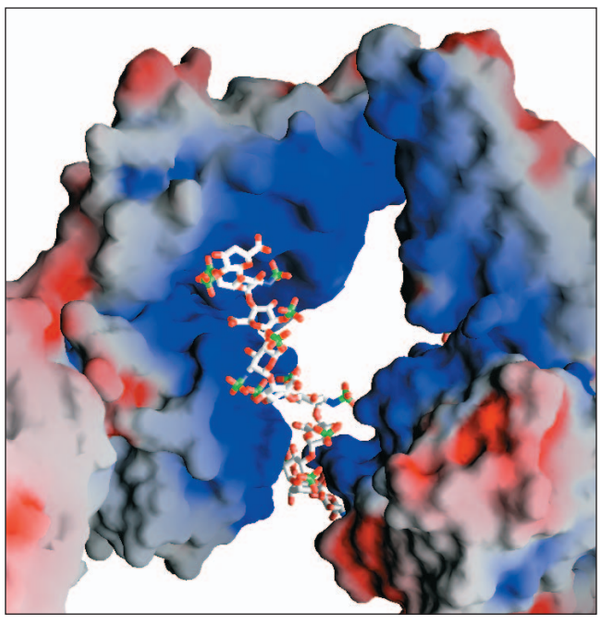


FIGURE 7–25 Interaction between a glycosaminoglycan and its binding protein. Fibroblast growth factor (FGF1), its cell surface receptor (FGFR), and a short segment of a glycosaminoglycan (heparin) were co-crystallized to yield the structure shown here (PDB ID 1E0O). The proteins are represented as surface contour images, with color to represent surface electrostatic potential: red, predominantly negative charge; blue, predominantly positive charge. Heparin is shown in a ball-and-stick representation, with the negative charges ($—SO_3^-$ and $—COO^-$) attracted to the positive (blue) surface of the FGF protein. Heparin was used in this experiment, but, in vivo, the glycosaminoglycan that binds FGF is heparan sulfate on the cell surface.

Table 7–2 summarizes the composition, properties, roles, and occurrence of the polysaccharides described in Section 7.2.

SUMMARY 7.2 Polysaccharides

- Polysaccharides (glycans) serve as stored fuel and as structural components of cell walls and extracellular matrix.
- The homopolysaccharides starch and glycogen are stored fuels in plant, animal, and bacterial cells. They consist of D-glucose with linkages, and all three contain some branches.
- The homopolysaccharides cellulose, chitin, and dextran serve structural roles. Cellulose, composed of (β1→4)-linked D-glucose residues, lends strength and rigidity to plant cell walls. Chitin, a polymer of (β1→4)-linked *N*-acetylglucosamine, strengthens the

TABLE 7–2 Structures and Roles of Some Polysaccharides

Polymer	*Type**	*Repeating unit*†	*Size (number of monosaccharide units)*	*Roles/significance*
Starch				Energy storage: in plants
Amylose	Homo-	(α1→4)Glc, linear	50–5,000	
Amylopectin	Homo-	(α1→4)Glc, with (α1→6)Glc branches every 24–30 residues	Up to 10^6	
Glycogen	Homo-	(α1→4)Glc, with (α1→6)Glc branches every 8–12 residues	Up to 50,000	Energy storage: in bacteria and animal cells
Cellulose	Homo-	(β1→4)Glc	Up to 15,000	Structural: in plants, gives rigidity and strength to cell walls
Chitin	Homo-	(β1→4)GlcNAc	Very large	Structural: in insects, spiders, crustaceans, gives rigidity and strength to exoskeletons
Dextran	Homo-	(α1→6)Glc, with (α1→3) branches	Wide range	Structural: in bacteria, extracellular adhesive
Peptidoglycan	Hetero-; peptides attached	4)Mur2Ac(β1→4)GlcNAc(β1	Very large	Structural: in bacteria, gives rigidity and strength to cell envelope
Agarose	Hetero-	3)D-Gal(β1→4)3,6-anhydro-L-Gal(α1	1,000	Structural: in algae, cell wall material
Hyaluronate (a glycosaminoglycan)	Hetero-; acidic	4)GlcA(β1→3)GlcNAc(β1	Up to 100,000	Structural: in vertebrates, extracellular matrix of skin and connective tissue; viscosity and lubrication in joints

*Each polymer is classified as a homopolysaccharide (homo-) or heteropolysaccharide (hetero-).

†The abbreviated names for the peptidoglycan, agarose, and hyaluronate repeating units indicate that the polymer contains repeats of this disaccharide unit. For example, in peptidoglycan, the GlcNAc of one disaccharide unit is (β1→4)-linked to the first residue of the next disaccharide unit.

exoskeletons of arthropods. Dextran forms an adhesive coat around certain bacteria.

- Homopolysaccharides fold in three dimensions. The chair form of the pyranose ring is essentially rigid, so the conformation of the polymers is determined by rotation about the bonds to the oxygen on the anomeric carbon. Starch and glycogen form helical structures with intrachain hydrogen bonding; cellulose and chitin form long, straight strands that interact with neighboring strands.
- Bacterial and algal cell walls are strengthened by heteropolysaccharides—peptidoglycan in bacteria, agar in red algae. The repeating disaccharide in peptidoglycan is GlcNAc(β1→4)Mur2Ac; in agarose, it is D-Gal(β1→4)3,6-anhydro-L-Gal.
- Glycosaminoglycans are extracellular heteropolysaccharides in which one of the two monosaccharide units is a uronic acid and the other an *N*-acetylated amino sugar. Sulfate esters on some of the hydroxyl groups give these polymers a high density of negative charge, forcing them to assume extended conformations. These polymers (hyaluronate, chondroitin sulfate, dermatan sulfate, keratan sulfate, and heparin) provide viscosity, adhesiveness, and tensile strength to the extracellular matrix.

7.3 Glycoconjugates: Proteoglycans, Glycoproteins, and Glycolipids

In addition to their important roles as stored fuels (starch, glycogen, dextran) and as structural materials (cellulose, chitin, peptidoglycans), polysaccharides and oligosaccharides are information carriers: they serve as destination labels for some proteins and as mediators of specific cell-cell interactions and interactions between cells and the extracellular matrix. Specific carbohydrate-containing molecules act in cell-cell recognition and

adhesion, cell migration during development, blood clotting, the immune response, and wound healing, to name but a few of their many roles. In most of these cases, the informational carbohydrate is covalently joined to a protein or a lipid to form a **glycoconjugate,** which is the biologically active molecule.

Proteoglycans are macromolecules of the cell surface or extracellular matrix in which one or more glycosaminoglycan chains are joined covalently to a membrane protein or a secreted protein. The glycosaminoglycan moiety commonly forms the greater fraction (by mass) of the proteoglycan molecule, dominates the structure, and is often the main site of biological activity. In many cases the biological activity is the provision of multiple binding sites, rich in opportunities for hydrogen bonding and electrostatic interactions with other proteins of the cell surface or the extracellular matrix. Proteoglycans are major components of connective tissue such as cartilage, in which their many noncovalent interactions with other proteoglycans, proteins, and glycosaminoglycans provide strength and resilience.

Glycoproteins have one or several oligosaccharides of varying complexity joined covalently to a protein. They are found on the outer face of the plasma membrane, in the extracellular matrix, and in the blood. Inside cells they are found in specific organelles such as Golgi complexes, secretory granules, and lysosomes. The oligosaccharide portions of glycoproteins are less monotonous than the glycosaminoglycan chains of proteoglycans; they are rich in information, forming highly specific sites for recognition and high-affinity binding by other proteins.

Glycolipids are membrane lipids in which the hydrophilic head groups are oligosaccharides, which, as in glycoproteins, act as specific sites for recognition by carbohydrate-binding proteins.

Proteoglycans Are Glycosaminoglycan-Containing Macromolecules of the Cell Surface and Extracellular Matrix

Mammalian cells can produce at least 30 types of molecules that are members of the proteoglycan superfamily. These molecules act as tissue organizers, influence the development of specialized tissues, mediate the activities of various growth factors, and regulate the extracellular assembly of collagen fibrils. The basic proteoglycan unit consists of a "core protein" with covalently attached glycosaminoglycan(s). For example, the sheetlike extracellular matrix (basal lamina) that separates organized groups of cells contains a family of core proteins (M_r 20,000 to 40,000), each with several covalently attached heparan sulfate chains. (Heparan sulfate is structurally similar to heparin but has a lower density of sulfate esters.) The point of attachment is commonly a

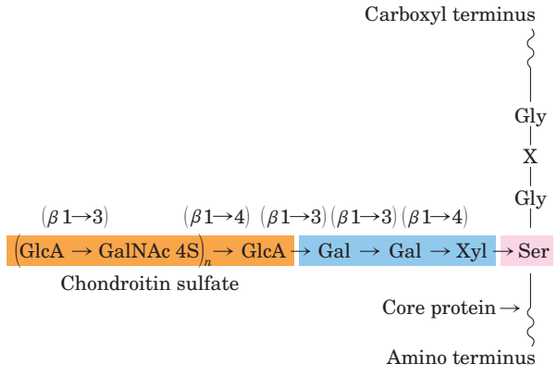


FIGURE 7–26 Proteoglycan structure, showing the trisaccharide bridge. A typical trisaccharide linker (blue) connects a glycosaminoglycan—in this case chondroitin sulfate (orange)—to a Ser residue (red) in the core protein. The xylose residue at the reducing end of the linker is joined by its anomeric carbon to the hydroxyl of the Ser residue.

Ser residue, to which the glycosaminoglycan is joined through a trisaccharide bridge (Fig. 7–26). The Ser residue is generally in the sequence –Ser–Gly–X–Gly– (where X is any amino acid residue), although not every protein with this sequence has an attached glycosaminoglycan. Many proteoglycans are secreted into the extracellular matrix, but some are integral membrane proteins (see Fig. 11–7). For example, syndecan core protein (M_r 56,000) has a single transmembrane domain and an extracellular domain bearing three chains of heparan sulfate and two of chondroitin sulfate, each attached to a Ser residue (Fig. 7–27a). There are at least four members of the syndecan family in mammals. Another family of core proteins is the glypicans, with six members. These proteins are attached to the membrane by a lipid anchor, a derivative of the membrane lipid phosphatidylinositol (Chapter 11).

The heparan sulfate moieties in proteoglycans bind a variety of extracellular ligands and thereby modulate the ligands' interaction with specific receptors of the cell surface. Detailed examination of the glycan moiety of proteoglycans has revealed a sequence heterogeneity that is not random; some domains (typically 3 to 8 disaccharide units long) differ from neighboring domains in sequence and in ability to bind to specific proteins. Heparan sulfate, for example, is initially synthesized as a long polymer (50 to 200 disaccharide units) of alternating *N*-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) residues. This simple chain is acted on by a series of enzymes that introduce alterations in specific regions. First, an *N*-deacetylase: *N*-sulfotransferase replaces some acetyl groups of GlcNAc residues with sulfates, creating clusters of *N*-sulfated glucosamine (GlcN) residues. These clusters then attract enzymes that carry out further modifications: an epimerase con-

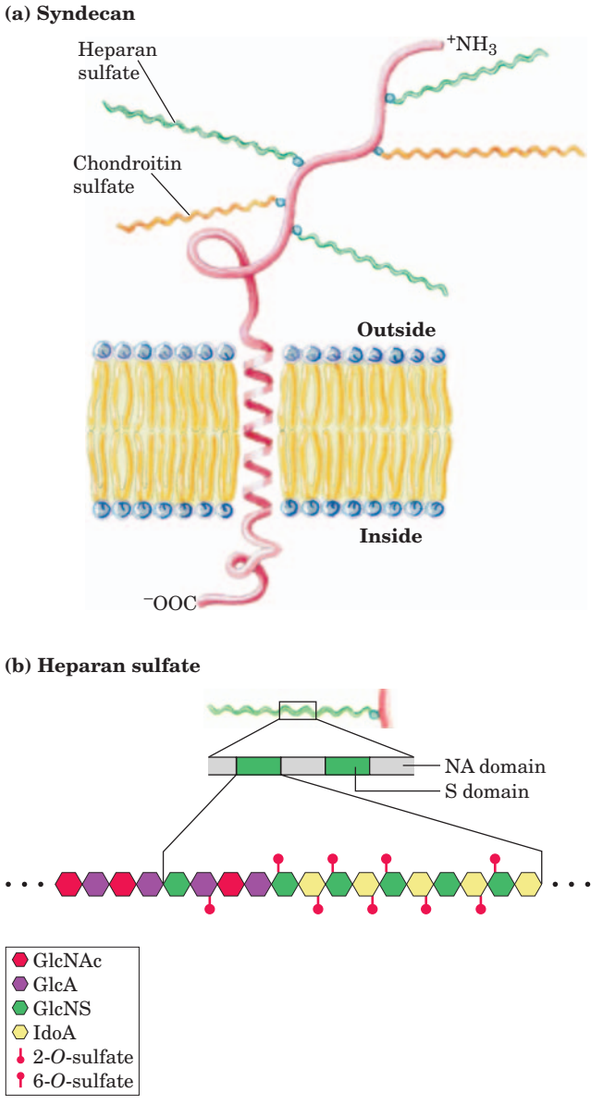


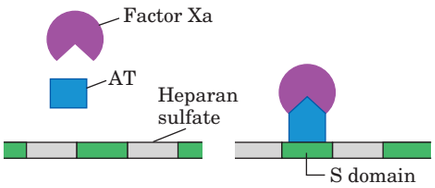
FIGURE 7–27 Proteoglycan structure of an integral membrane protein. (a) Schematic diagram of syndecan, a core protein of the plasma membrane. The amino-terminal domain on the extracellular surface of the membrane is covalently attached (by trisaccharide linkers such as those in Fig. 7–26) to three heparan sulfate chains and two chondroitin sulfate chains. Some core proteins (syndecans, as here) are anchored by a single transmembrane helix; others (glypicans), by a covalently attached membrane glycolipid. In a third class of core proteins, the protein is released into the extracellular space, where it forms part of the basement membrane. **(b)** Along a heparan sulfate chain, regions rich in sulfated sugars, the S domains (green), alternate with regions with chiefly unmodified residues of GlcNAc and GlcA, the NA domains (gray). One of the S domains is shown in more detail, revealing a high density of modified residues: GlcA, with a sulfate ester at C-6; and IdoA, with a sulfate ester at C-2. The exact pattern of sulfation in the S domain differs among proteoglycans. Given all the possible modifications of the GlcNAc–IdoA dimer, at least 32 different disaccharide units are possible.

verts GlcA to IdoA; sulfotransferases then create sulfate esters at the C-2 hydroxyl of IdoA and the C-6 hydroxyl of *N*-sulfated GlcN, but only in regions that already have *N*-sulfated GlcN residues. The result is a polymer in which highly sulfated domains (S domains) alternate with domains having unmodified GlcNAc and GlcA residues (*N*-acetylated, or NA, domains) (Fig. 7–27b). The exact pattern of sulfation in the S domain differs in different proteoglycans; given the number of possible modifications of the GlcNAc–IdoA dimer, at least 32 different disaccharide units are possible. Furthermore, the same core protein can display different heparan sulfate structures when synthesized in different cell types.

The S domains bind specifically to extracellular proteins and signaling molecules to alter their activities. The change in activity may result from a conformational change in the protein that is induced by the binding (Fig. 7–28a), or it may be due to the ability of adjacent domains of heparan sulfate to bind to two different proteins, bringing them into close proximity and enhancing protein-protein interactions (Fig. 7–28b). A third general mechanism of action is the binding of extracellular signal molecules (growth factors, for example) to heparan sulfate, which increases their local concentrations and enhances their interaction with growth factor receptors in the cell surface; in this case, the heparan sulfate acts as a coreceptor (Fig. 7–28c). For example, fibroblast growth factor (FGF), an extracellular protein signal that stimulates cell division, first binds to heparan sulfate moieties of syndecan molecules in the target cell's plasma membrane. Syndecan presents FGF to the FGF plasma membrane receptor, and only then can FGF interact productively with its receptor to trigger cell division. Finally, the S domains interact—electrostatically and otherwise—with a variety of soluble molecules outside the cell, maintaining high local concentrations at the cell surface (Fig. 7–28d). The importance of correctly synthesizing sulfated domains in heparan sulfate is demonstrated in "knockout" mice that lack the enzyme that places sulfates at the C-2 hydroxyl of IdoA. Such animals are born without kidneys and with very severe abnormalities in development of the skeleton and eyes.

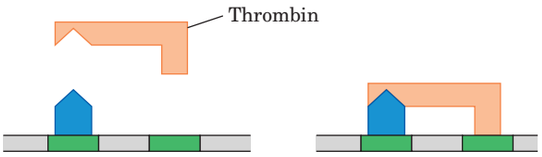
Some proteoglycans can form **proteoglycan aggregates,** enormous supramolecular assemblies of many core proteins all bound to a single molecule of hyaluronate. Aggrecan core protein (M_r ~250,000) has multiple chains of chondroitin sulfate and keratan sulfate, joined to Ser residues in the core protein through trisaccharide linkers, to give an aggrecan monomer of M_r ~2×10^6. When a hundred or more of these "decorated" core proteins bind a single, extended molecule of hyaluronate (Fig. 7–29), the resulting proteoglycan aggregate ($M_r > 2 \times 10^8$) and its associated water of hydration occupy a volume about equal to that of a bacterial cell! Aggrecan interacts strongly with collagen in the

(a) Conformational activation



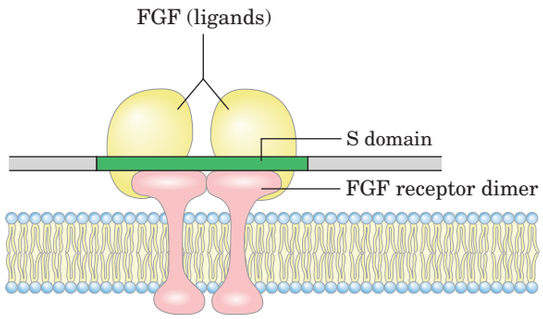
A conformational change induced in the protein antithrombin (AT) on binding a specific pentasaccharide S domain allows its interaction with Factor Xa, a blood clotting factor, preventing clotting.

(b) Enhanced protein-protein interaction



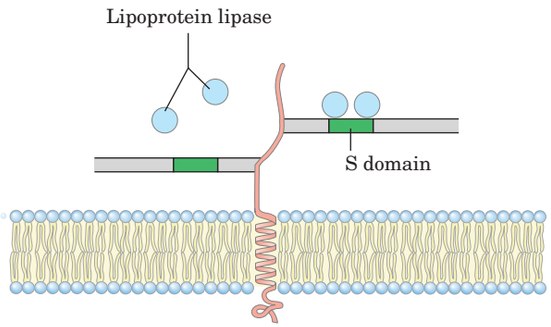
Binding of AT and thrombin to two adjacent S domains brings the two proteins into close proximity, favoring their interaction, which inhibits blood clotting.

(c) Coreceptor for extracellular ligands



S domains interact with both the fibroblast growth factor (FGF) and its receptor, bringing the oligomeric complex together and increasing the effectiveness of a low concentration of FGF.

(d) Cell surface localization/concentration



The high density of negative charges in heparan sulfate brings positively charged molecules of lipoprotein lipase into the vicinity and holds them by electrostatic interactions as well as by sequence-specific interactions with S domains. Such interactions are also central in the first step in the entry of certain viruses (such as herpes simplex viruses HSV-1 and HSV-2) into cells.

FIGURE 7–28 **Four types of protein interactions with S domains of heparan sulfate.**

extracellular matrix of cartilage, contributing to the development and tensile strength of this connective tissue.

Interwoven with these enormous extracellular proteoglycans are fibrous matrix proteins such as collagen, elastin, and fibronectin, forming a cross-linked meshwork that gives the whole extracellular matrix strength and resilience. Some of these proteins are multiadhesive, a single protein having binding sites for several different matrix molecules. Fibronectin, for example, has separate domains that bind fibrin, heparan sulfate, collagen, and a family of plasma membrane proteins called integrins that mediate signaling between the cell interior and the extracellular matrix (see Fig. 11–24). Integrins, in turn, have binding sites for a number of other extracellular macromolecules. The overall picture of cell-matrix interactions that emerges (Fig. 7–30) shows an array of interactions between cellular and extracellular molecules. These interactions serve not merely to anchor cells to the extracellular matrix but also to provide paths that direct the migration of cells in developing tissue and, through integrins, to convey information in both directions across the plasma membrane.

Glycoproteins Have Covalently Attached Oligosaccharides

Glycoproteins are carbohydrate-protein conjugates in which the carbohydrate moieties are smaller and more structurally diverse than the glycosaminoglycans of proteoglycans. The carbohydrate is attached at its anomeric carbon through a glycosidic link to the —OH of a Ser or Thr residue (*O*-linked), or through an *N*-glycosyl link to the amide nitrogen of an Asn residue (*N*-linked) (Fig. 7–31). Some glycoproteins have a single oligosaccharide chain, but many have more than one; the carbohydrate may constitute from 1% to 70% or more of the glyco-

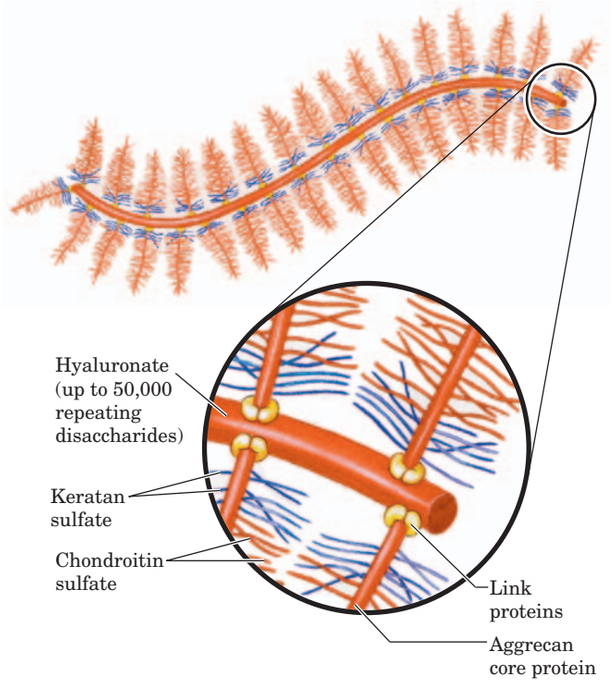


FIGURE 7–29 Proteoglycan aggregate of the extracellular matrix. One very long molecule of hyaluronate is associated noncovalently with about 100 molecules of the core protein aggrecan. Each aggrecan molecule contains many covalently bound chondroitin sulfate and keratan sulfate chains. Link proteins situated at the junction between each core protein and the hyaluronate backbone mediate the core protein–hyaluronate interaction.

protein by mass. The structures of a large number of *O*- and *N*-linked oligosaccharides from a variety of glycoproteins are known; Figure 7–31 shows a few typical examples.

As we shall see in Chapter 11, the external surface of the plasma membrane has many membrane glycoproteins with arrays of covalently attached oligosaccharides of varying complexity. One of the best-characterized membrane glycoproteins is glycophorin A of the erythrocyte membrane (see Fig. 11–8). It contains 60% carbohydrate by mass, in the form of 16 oligosaccharide chains (totaling 60 to 70 monosaccharide residues) covalently attached to amino acid residues near the amino terminus of the polypeptide chain. Fifteen of the oligosaccharide chains are *O*-linked to Ser or Thr residues, and one is *N*-linked to an Asn residue.

Many of the proteins secreted by eukaryotic cells are glycoproteins, including most of the proteins of blood. For example, immunoglobulins (antibodies) and certain hormones, such as follicle-stimulating hormone, luteinizing hormone, and thyroid-stimulating hormone, are glycoproteins. Many milk proteins, including lactalbumin, and some of the proteins secreted by the pancreas (such as ribonuclease) are glycosylated, as are most of the proteins contained in lysosomes.

A number of cases are known in which the same protein produced in two types of tissues has different glycosylation patterns. For example, the human protein interferon IFN-β1 has one set of oligosaccharide chains when produced in ovarian cells and a different set when produced in breast epithelial cells. The biological significance of these **tissue glycoforms** is not understood, but in some way the oligosaccharide chains represent a tissue-specific marker.

The biological advantages of adding oligosaccharides to proteins are not fully understood. The very hydrophilic clusters of carbohydrate alter the polarity and solubility of the proteins with which they are conjugated. Oligosaccharide chains that are attached to newly synthesized proteins in the endoplasmic reticulum and elaborated in the Golgi complex may also influence the sequence of polypeptide-folding events that determine the tertiary structure of the protein (see Fig. 27–34). Steric interactions between peptide and oligosaccharide may preclude one folding route and favor another. When numerous negatively charged oligosaccharide chains are clustered in a single region of a protein, the charge repulsion among them favors the formation of an extended, rodlike structure in that region. The bulkiness and negative charge of oligosaccharide chains also

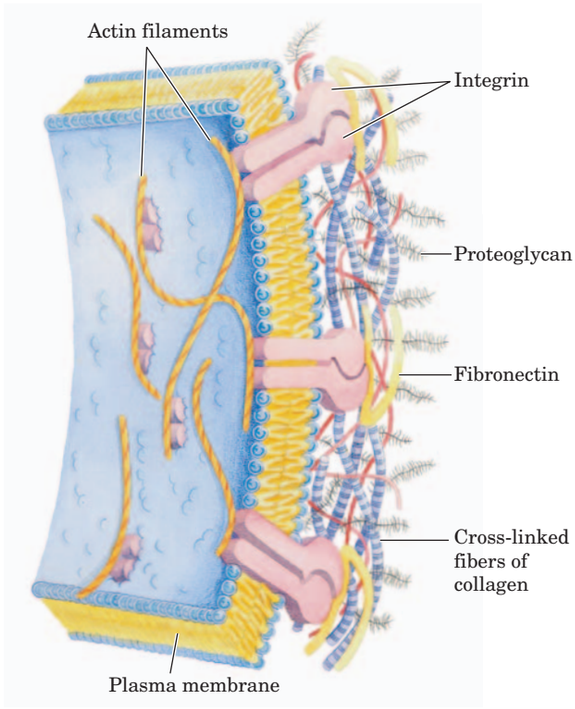


FIGURE 7–30 Interactions between cells and the extracellular matrix. The association between cells and the proteoglycan of the extracellular matrix is mediated by a membrane protein (integrin) and by an extracellular protein (fibronectin in this example) with binding sites for both integrin and the proteoglycan. Note the close association of collagen fibers with the fibronectin and proteoglycan.

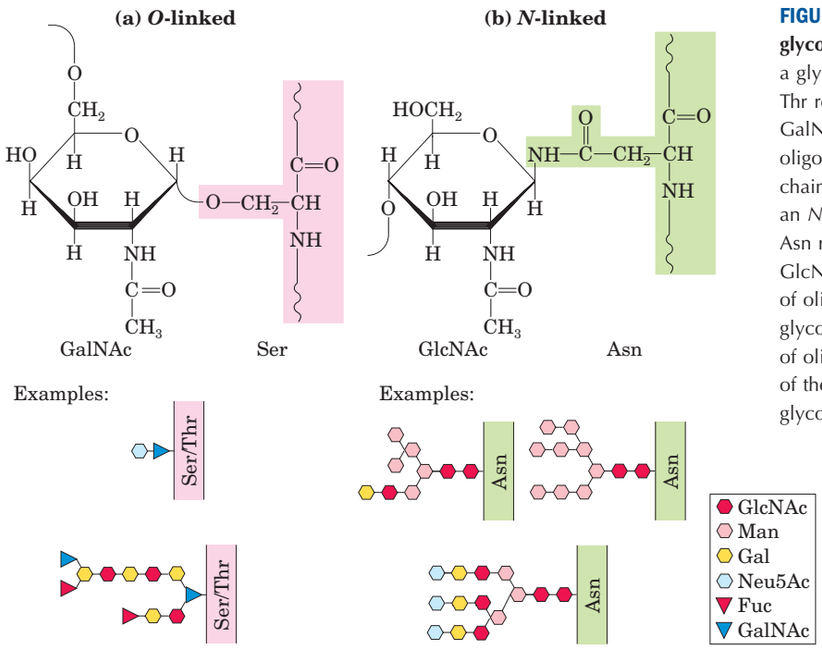


FIGURE 7–31 Oligosaccharide linkages in glycoproteins. (a) *O*-linked oligosaccharides have a glycosidic bond to the hydroxyl group of Ser or Thr residues (shaded pink), illustrated here with GalNAc as the sugar at the reducing end of the oligosaccharide. One simple chain and one complex chain are shown. **(b)** *N*-linked oligosaccharides have an *N*-glycosyl bond to the amide nitrogen of an Asn residue (shaded green), illustrated here with GlcNAc as the terminal sugar. Three common types of oligosaccharide chains that are *N*-linked in glycoproteins are shown. A complete description of oligosaccharide structure requires specification of the position and stereochemistry (α or β) of each glycosidic linkage.

protect some proteins from attack by proteolytic enzymes. Beyond these global physical effects on protein structure, there are also more specific biological effects of oligosaccharide chains in glycoproteins (Section 7.4).

Glycolipids and Lipopolysaccharides Are Membrane Components

Glycoproteins are not the only cellular components that bear complex oligosaccharide chains; some lipids, too, have covalently bound oligosaccharides. **Gangliosides** are membrane lipids of eukaryotic cells in which the polar head group, the part of the lipid that forms the outer surface of the membrane, is a complex oligosaccharide containing sialic acid (Fig. 7–9) and other monosaccharide residues. Some of the oligosaccharide moieties of gangliosides, such as those that determine human blood groups (see Fig. 10–14), are identical with those found in certain glycoproteins, which therefore also contribute to blood group type determination. Like the oligosaccharide moieties of glycoproteins, those of membrane lipids are generally, perhaps always, found on the outer face of the plasma membrane.

Lipopolysaccharides are the dominant surface feature of the outer membrane of gram-negative bacteria such as *Escherichia coli* and *Salmonella typhimurium*. These molecules are prime targets of the antibodies produced by the vertebrate immune system in response to bacterial infection and are therefore important determinants of the serotype of bacterial strains (serotypes are strains that are distinguished on the basis of antigenic properties). The lipopolysaccharides of *S. typhimurium* contain six fatty acids bound to two glucosamine residues, one of which is the point of attachment for a complex oligosaccharide (Fig. 7–32). *E. coli* has similar but unique lipopolysaccharides. The lipopolysaccharides of some bacteria are toxic to humans and other animals; for example, they are responsible for the dangerously lowered blood pressure that occurs in toxic shock syndrome resulting from gram-negative bacterial infections. ■

SUMMARY 7.3 Glycoconjugates: Proteoglycans, Glycoproteins, and Glycolipids

- Proteoglycans are glycoconjugates in which a core protein is attached covalently to one or more large glycans, such as heparan sulfate, chondroitin sulfate, or keratan sulfate. The glycan is the greater portion (by mass) of the molecule. Bound to the outside of the plasma membrane by a transmembrane peptide or a covalently attached lipid, proteoglycans provide points of adhesion, recognition, and information transfer between cells, or between the cell and the extracellular matrix.
- Glycoproteins contain covalently linked oligosaccharides that are smaller but more structurally complex, and therefore more information-rich, than glycosaminoglycans. Many cell surface or extracellular proteins are glycoproteins, as are most secreted proteins. The covalently attached oligosaccharides influence the folding and stability of the proteins, provide critical information about the

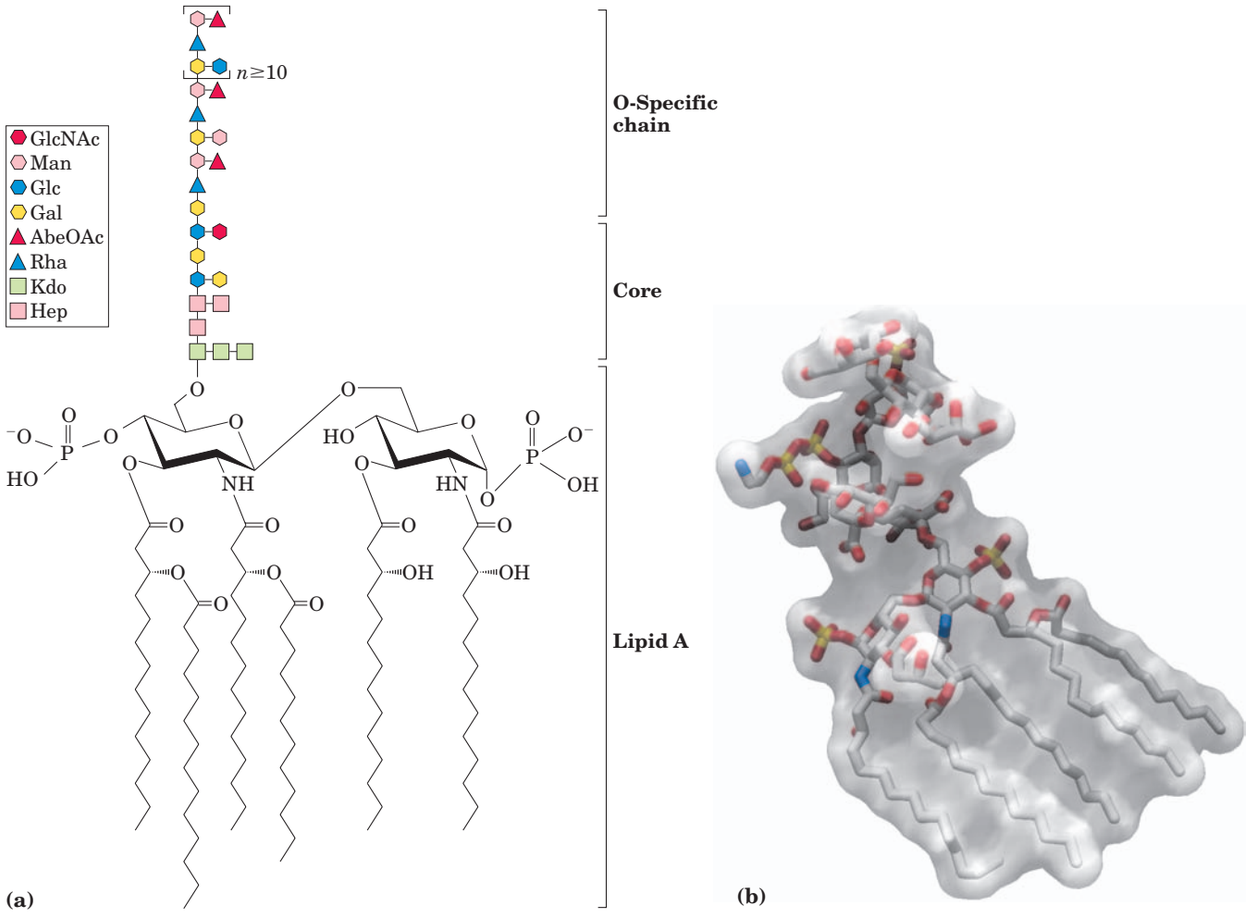


FIGURE 7–32 **Bacterial lipopolysaccharides. (a)** Schematic diagram of the lipopolysaccharide of the outer membrane of *Salmonella typhimurium*. Kdo is 3-deoxy-D-manno-octulosonic acid, previously called *keto*deoxyoctonic acid; Hep is L-glycero-D-mannoheptose; AbeOAc is abequose (a 3,6-dideoxyhexose) acetylated on one of its hydroxyls. There are six fatty acids in the lipid A portion of the molecule. Different bacterial species have subtly different lipopolysaccharide structures, but they have in common a lipid region (lipid A), a core oligosaccharide, and an "O-specific" chain, which is the principal determinant of the serotype (immunological reactivity) of the bacterium. The outer membranes of the gram-negative bacteria *S. typhimurium* and *E. coli* contain so many lipopolysaccharide molecules that the cell surface is virtually covered with O-specific chains. **(b)** The stick structure of the lipopolysaccharide of *E. coli* is visible through a transparent surface contour model of the molecule. The position of the sixth fatty acyl chain was not defined in the crystallographic study, so it is not shown.

targeting of newly synthesized proteins, and allow for specific recognition by other proteins.

- Glycolipids and lipopolysaccharides are components of the plasma membrane with covalently attached oligosaccharide chains exposed on the cell's outer surface.

7.4 Carbohydrates as Informational Molecules: The Sugar Code

Glycobiology, the study of the structure and function of glycoconjugates, is one of the most active and exciting areas of biochemistry and cell biology. As is becoming increasingly clear, cells use specific oligosaccharides to encode important information about intracellular targeting of proteins, cell-cell interaction, tissue development, and extracellular signals. Our discussion uses just a few examples to illustrate the diversity of structure and the range of biological activity of the glycoconjugates. In Chapter 20 we discuss the biosynthesis of polysaccharides, including the peptidoglycans; and in Chapter 27, the assembly of oligosaccharide chains on glycoproteins.

Improved methods for the analysis of oligosaccharide and polysaccharide structure have revealed remarkable complexity and diversity in the oligosaccharides of glycoproteins and glycolipids. Consider the oligosaccharide chains in Figure 7–31, typical of those

found in many glycoproteins. The most complex of those shown contains 14 monosaccharide residues of four different kinds, variously linked as (1→2), (1→3), (1→4), (1→6), (2→3), and (2→6), some with the α and some with the β configuration. Branched structures, not found in nucleic acids or proteins, are common in oligosaccharides. With the reasonable assumption that 20 different monosaccharide subunits are available for construction of oligosaccharides, we can calculate that 1.44×10^{15} different hexameric oligosaccharides are possible; this compares with 6.4×10^{7} (20^6) different hexapeptides possible with the 20 common amino acids, and 4,096 (4^6) different hexanucleotides with the four nucleotide subunits. If we also allow for variations in oligosaccharides resulting from sulfation of one or more residues, the number of possible oligosaccharides increases by two orders of magnitude. Oligosaccharides are enormously rich in structural information, not merely rivaling but far surpassing nucleic acids in the density of information contained in a molecule of modest size. Each of the oligosaccharides represented in Figure 7–31 presents a unique, three-dimensional face—a word in the sugar code—readable by the proteins that interact with it.

Lectins Are Proteins That Read the Sugar Code and Mediate Many Biological Processes

Lectins, found in all organisms, are proteins that bind carbohydrates with high affinity and specificity (Table 7–3). Lectins serve in a wide variety of cell-cell recognition, signaling, and adhesion processes and in intracellular targeting of newly synthesized proteins. In the laboratory, purified lectins are useful reagents for detecting and separating glycoproteins with different oligosaccharide moieties. Here we discuss just a few examples of the roles of lectins in cells.

Some peptide hormones that circulate in the blood have oligosaccharide moieties that strongly influence their circulatory half-life. Luteinizing hormone and thyrotropin (polypeptide hormones produced in the adrenal cortex) have *N*-linked oligosaccharides that end with the disaccharide GalNAc4S(β1→4)GlcNAc, which is recognized by a lectin (receptor) of hepatocytes. (GalNAc4S is *N*-acetylgalactosamine sulfated on the —OH group of C-4.) Receptor-hormone interaction mediates the uptake and destruction of luteinizing hormone and thyrotropin, reducing their concentration in the blood. Thus the blood levels of these hormones undergo a periodic rise (due to secretion by the adrenal cortex) and fall (due to destruction by hepatocytes).

The importance of the oligosaccharide moiety of these hormones is apparent from studies of individuals with a defective enzyme in the pathway that produces this oligosaccharide. Females with this congenital defect often fail to undergo the sexual changes of puberty (although males with the same defect develop normally). ■

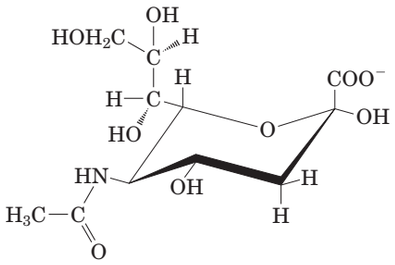
The residues of Neu5Ac (a sialic acid) situated at the ends of the oligosaccharide chains of many plasma glycoproteins (Fig. 7–31) protect those proteins from uptake and degradation in the liver. For example, ceruloplasmin, a copper-containing serum glycoprotein, has several oligosaccharide chains ending in Neu5Ac. Re-

TABLE 7–3 Some Lectins and the Oligosaccharide Ligands They Bind

Lectin source and lectin	*Abbreviation*	*Ligand(s)*
Plant		
Concanavalin A	ConA	Manα1—OCH_3
Griffonia simplicifolia lectin 4	GS4	Lewis b (Le^b) tetrasaccharide
Wheat germ agglutinin	WGA	Neu5Ac(α2→3)Gal(β1→4)Glc GlcNAc(β1→4)GlcNAc
Ricin		Gal(β1→4)Glc
Animal		
Galectin-1		Gal(β1→4)Glc
Mannose-binding protein A	MBP-A	High-mannose octasaccharide
Viral		
Influenza virus hemagglutinin	HA	Neu5Ac(α2→6)Gal(β1→4)Glc
Polyoma virus protein 1	VP1	Neu5Ac(α2→3)Gal(β1→4)Glc
Bacterial		
Enterotoxin	LT	Gal
Cholera toxin	CT	GM1 pentasaccharide

Source: Weiss, W.I. & Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65,** 441–473.

moval of the sialic acid residues by the enzyme sialidase (also called neuraminidase) is one way in which the body marks "old" proteins for destruction and replacement. The plasma membrane of hepatocytes has lectin molecules (asialoglycoprotein receptors; "asialo-" indicating "without sialic acid") that specifically bind oligosaccharide chains with galactose residues no longer "protected" by a terminal Neu5Ac residue. Receptor-ceruloplasmin interaction triggers endocytosis and destruction of the ceruloplasmin.



N-Acetylneuraminic acid (Neu5Ac)
(a sialic acid)

A similar mechanism is apparently responsible for removing old erythrocytes from the mammalian bloodstream. Newly synthesized erythrocytes have several membrane glycoproteins with oligosaccharide chains that end in Neu5Ac. When the sialic acid residues are removed by withdrawing a sample of blood, treating it with sialidase in vitro, and reintroducing it into the circulation, the treated erythrocytes disappear from the bloodstream within a few hours; those with intact oligosaccharides (erythrocytes withdrawn and reintroduced without sialidase treatment) continue to circulate for days.

Several animal viruses, including the influenza virus, attach to their host cells through interactions with oligosaccharides displayed on the host cell surface. The lectin of the influenza virus, the HA protein, is essential for viral entry and infection (see Fig. 11–25). After initial binding of the virus to a sialic acid–containing oligosaccharide on the host surface, a viral sialidase removes the terminal sialic acid residue, triggering the entry of the virus into the cell. Inhibitors of this enzyme are used clinically in the treatment of influenza. Lectins on the surface of the herpes simplex viruses HS-1 and HS-2 (the causative agents of oral and genital herpes, respectively) bind specifically to heparan sulfate on the cell surface as a first step in their infection cycle; infection requires precisely the right pattern of sulfation on this polymer.

Selectins are a family of plasma membrane lectins that mediate cell-cell recognition and adhesion in a wide range of cellular processes. One such process is the movement of immune cells (T lymphocytes) through the capillary wall, from blood to tissues, at sites of infection or inflammation (Fig. 7–33). At an infection site, P-selectin on the surface of capillary endothelial cells interacts with a specific oligosaccharide of the glycoproteins of circu-

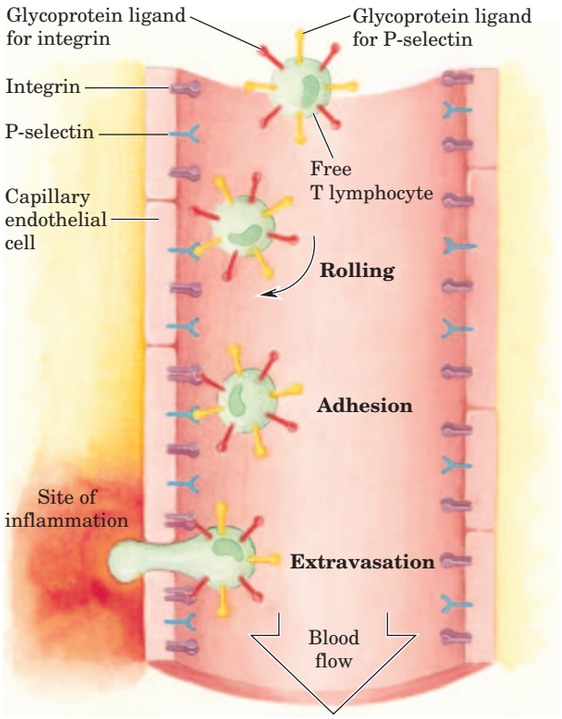


FIGURE 7–33 Role of lectin-ligand interactions in lymphocyte movement to the site of an infection or injury. A T lymphocyte circulating through a capillary is slowed by transient interactions between P-selectin molecules in the plasma membrane of the capillary endothelial cells and glycoprotein ligands for P-selectin on the T-cell surface. As it interacts with successive P-selectin molecules, the T cell rolls along the capillary surface. Near a site of inflammation, stronger interactions between integrin in the capillary surface and its ligand in the T-cell surface lead to tight adhesion. The T cell stops rolling and, under the influence of signals sent out from the site of inflammation, begins extravasation—escape through the capillary wall—as it moves toward the site of inflammation.

lating T cells. This interaction slows the T cells as they adhere to and roll along the endothelial lining of the capillaries. A second interaction, between integrin molecules (see p. XXX) in the T-cell plasma membrane and an adhesion protein on the endothelial cell surface, now stops the T cell and allows it to move through the capillary wall into the infected tissues to initiate the immune attack. Two other selectins participate in this "lymphocyte homing": E-selectin on the endothelial cell and L-selectin on the T cell bind their cognate oligosaccharides on the T cell and endothelial cell, respectively.

Some microbial pathogens have lectins that mediate bacterial adhesion to host cells or toxin entry into cells. The bacterium believed responsible for most gastric ulcers, *Helicobacter pylori,* adheres to the inner surface of the stomach by interactions between bacterial membrane lectins and specific oligosaccharides of membrane glycoproteins of the gastric epithelial cells

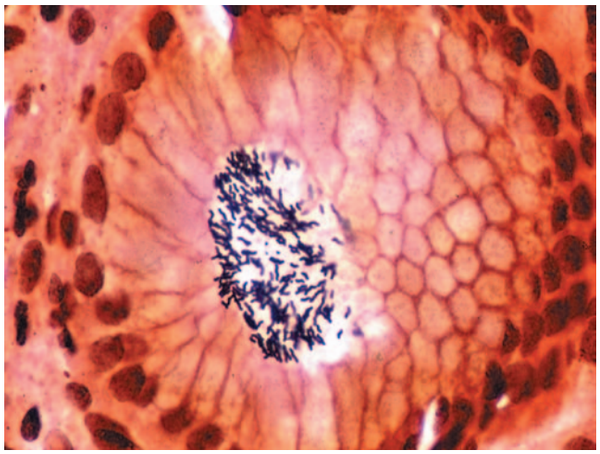


FIGURE 7–34 An ulcer in the making. *Helicobacter pylori* cells adhering to the gastric surface. This bacterium causes ulcers by interactions between a bacterial surface lectin and the Le^b oligosaccharide (a blood group antigen) of the gastric epithelium.

(Fig. 7–34). Among the binding sites recognized by *H. pylori* is the oligosaccharide Le^b when it is part of the type O blood group determinant. This observation helps to explain the severalfold greater incidence of gastric ulcers in people of blood type O than in those of type A or B. Chemically synthesized analogs of the Le^b oligosaccharide may prove useful in treating this type of ulcer. Administered orally, they could prevent bacterial adhesion (and thus infection) by competing with the gastric glycoproteins for binding to the bacterial lectin.

The cholera toxin molecule (produced by *Vibrio cholerae*) triggers diarrhea after entering intestinal cells responsible for water absorption from the intestine. The toxin attaches to its target cell through the oligosaccharide of ganglioside GM1, a membrane phospholipid (for the structure of GM1 see Box 10–2, Fig. 1), on the surface of intestinal epithelial cells. Similarly, the pertussis toxin produced by *Bordetella pertussis,* the bacterium that causes whooping cough, enters target cells only after interacting with an oligosaccharide (or perhaps several oligosaccharides) with a terminal sialic acid residue. Understanding the details of the oligosaccharide-binding sites of these toxins (lectins) may allow the development of genetically engineered toxin analogs for use in vaccines. Toxin analogs engineered to lack the carbohydrate binding site would be harmless because they could not bind to and enter cells, but they might elicit an immune response that would protect the recipient if later exposed to the natural toxin. It is also possible to imagine drugs that would act by mimicking the oligosaccharides of the cell surface, binding to the lectins of bacteria or toxins and preventing their productive binding to cell surfaces. ■

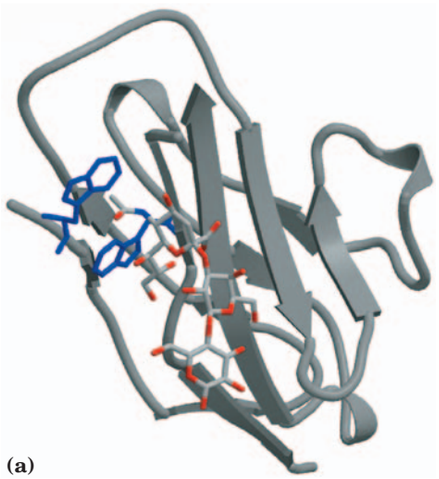
Lectins also act intracellularly. An oligosaccharide containing mannose 6-phosphate marks newly synthesized proteins in the Golgi complex for transfer to the lysosome (see Fig. 27–36). A common structural feature on the surface of these glycoproteins, the signal patch, causes them to be recognized by an enzyme that phosphorylates a mannose residue at the terminus of an oligosaccharide chain. This mannose phosphate residue is recognized by the cation-dependent mannose 6-phosphate receptor, a membrane-associated lectin with its mannose phosphate binding site on the lumenal side of the Golgi complex. When a section of the Golgi complex containing this receptor buds off to form a transport vesicle, proteins containing mannose phosphate residues are dragged into the forming bud by interaction of their mannose phosphates with the receptor; the vesicle then moves to and fuses with a lysosome, depositing its cargo therein. Many, perhaps all, of the degradative enzymes (hydrolases) of the lysosome are targeted and delivered by this mechanism.

Lectin-Carbohydrate Interactions Are Very Strong and Highly Specific

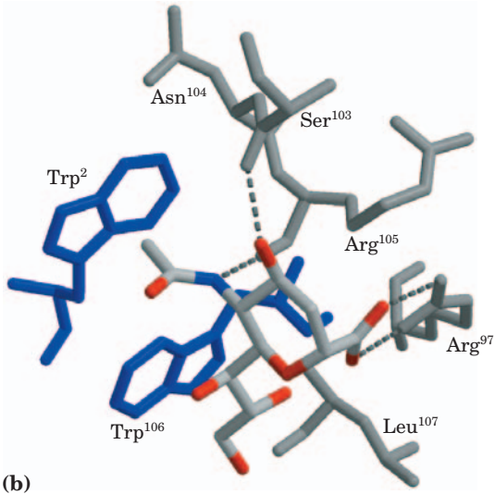
In all the functions of lectins described above, and in many more known to involve lectin-oligosaccharide interactions, it is essential that the oligosaccharide have a unique structure, so that recognition by the lectin is highly specific. The high density of information in oligosaccharides provides a sugar code with an essentially unlimited number of unique "words" small enough to be read by a single protein. In their carbohydrate-binding sites, lectins have a subtle molecular complementarity that allows interaction only with their correct carbohydrate cognates. The result is extraordinarily high specificity in these interactions.

X-ray crystallographic studies of the structures of several lectin-carbohydrate complexes have provided rich details of the lectin-sugar interaction. Sialoadhesin (also called siglec-1) is a membrane-bound lectin on the surface of mouse macrophages that recognizes certain sialic acid–containing oligosaccharides. This protein has a β sandwich domain (see this motif in the CD8 protein in Fig. 4–22) that contains the sialic acid binding site (Fig. 7–35a). Each of the ring substituents unique to Neu5Ac is involved in the interaction between sugar and lectin; the acetyl group at C-5 undergoes both hydrogen-bond and van der Waals interactions with the protein; the carboxyl group makes a salt bridge with Arg^{97}; and the hydroxyls of the glycerol moiety hydrogen-bond with the protein (Fig. 7–35b).

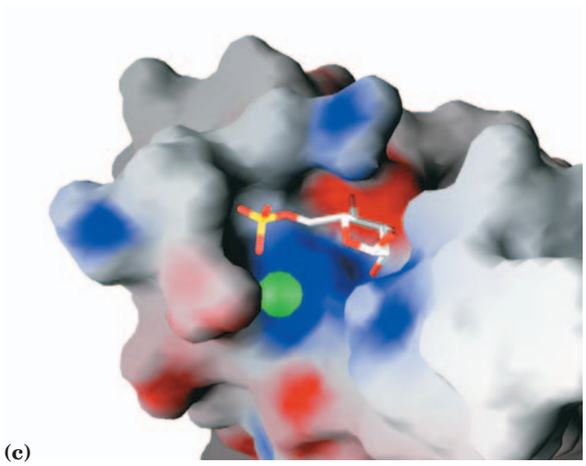
The structure of the mannose 6-phosphate receptor/lectin has also been resolved crystallographically, revealing details of its interaction with mannose 6-phosphate that explain the specificity of the binding and the necessity for a divalent cation in the lectin-sugar interaction (Fig. 7–35c). Arg^{111} of the receptor is hydrogen-bonded to the C-2 hydroxyl of mannose and coordinated with Mn^{2+}. His^{105} is hydrogen-bonded to



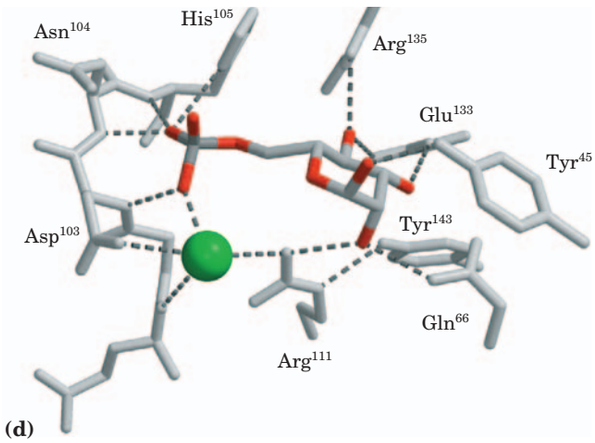
(a)



(b)



(c)



(d)

FIGURE 7–35 Details of lectin-carbohydrate interaction. (a) X-ray crystallographic studies of a sialic acid–specific lectin (derived from PDB ID 1QFO) show how a protein can recognize and bind to a sialic acid (Neu5Ac) residue. Sialoadhesin (also called siglec-1), a membrane-bound lectin of the surface of mouse macrophages, has a β sandwich domain (gray) that contains the Neu5Ac binding site (dark blue). Neu5Ac is shown as a stick structure. **(b)** Each ring substituent unique to Neu5Ac is involved in the interaction between sugar and lectin: the acetyl group at C-5 has both hydrogen-bond and van der Waals interactions with the protein; the carboxyl group makes a salt bridge with Arg^{97}; and the hydroxyls of the glycerol moiety hydrogen-bond with the protein. **(c)** Structure of the bovine mannose 6-phosphate receptor complexed with mannose 6-phosphate (PDB ID 1M6P). The protein is represented here as a surface contour image, with color to indicate the surface electrostatic potential: red, predominantly negative charge; blue, predominantly positive charge. Mannose 6-phosphate is shown as a stick structure; a manganese ion is shown in green. **(d)** In this complex, mannose 6-phosphate is hydrogen-bonded to Arg^{111} and coordinated with the manganese ion (green). The His^{105} hydrogen-bonded to a phosphate oxygen of mannose 6-phosphate may be the residue that, when protonated at low pH, causes the receptor to release mannose 6-phosphate into the lysosome.

one of the oxygen atoms of the phosphate (Fig. 7–35d). When the protein tagged with mannose 6-phosphate reaches the lysosome (which has a lower internal pH than the Golgi complex), the receptor apparently loses its affinity for mannose 6-phosphate. Protonation of His^{105} may be responsible for this change in binding.

In addition to these very specific interactions, there are more general interactions that contribute to the binding of many carbohydrates to their lectins. For example, many sugars have a more polar and a less polar side (Fig. 7–36); the more polar side hydrogen-bonds with the lectin, while the less polar undergoes hydrophobic interactions with nonpolar amino acid residues. The sum of all these interactions produces high-affinity binding (K_d often 10^{-8} M or less) and high specificity of lectins for their carbohydrates. This represents a kind of information transfer that is clearly central in many processes within and between cells. Figure 7–37 summarizes some of the biological interactions mediated by the sugar code.

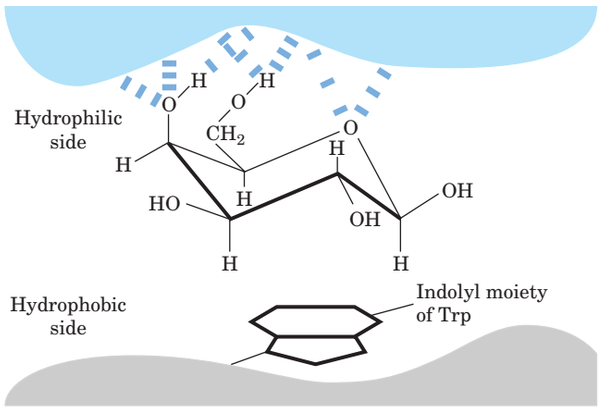


FIGURE 7–36 Hydrophobic interactions of sugar residues. Sugar units such as galactose have a more polar side (the top of the chair, with the ring oxygen and several hydroxyls), available to hydrogen-bond with the lectin, and a less polar side that can have hydrophobic interactions with nonpolar side chains in the protein, such as the indole ring of tryptophan.

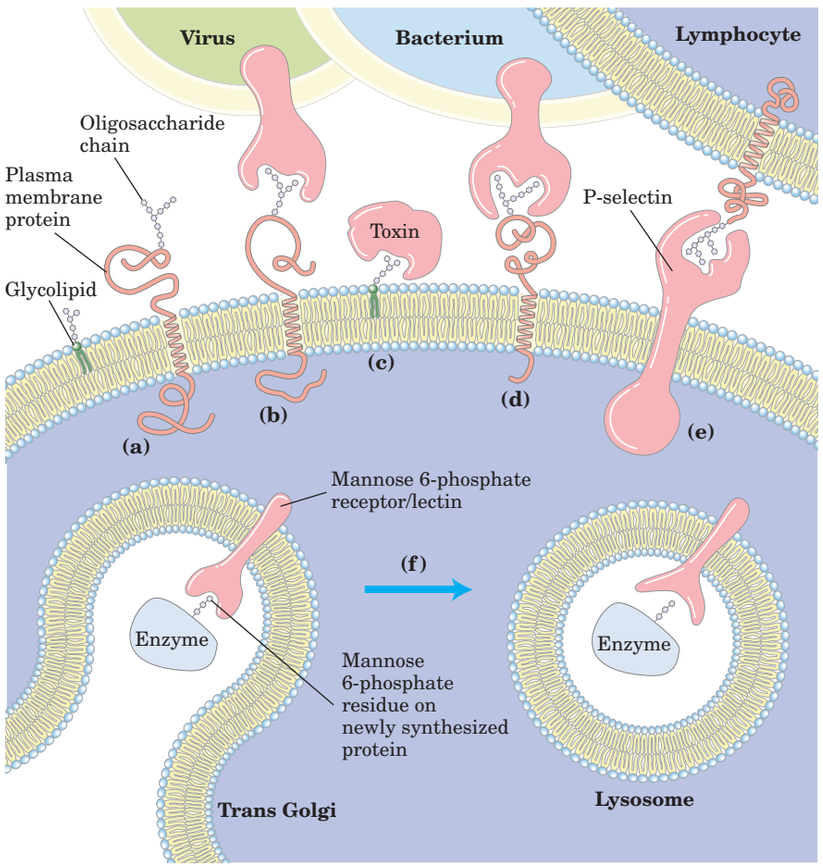


FIGURE 7–37 Roles of oligosaccharides in recognition and adhesion at the cell surface. **(a)** Oligosaccharides with unique structures (represented as strings of hexagons), components of a variety of glycoproteins or glycolipids on the outer surface of plasma membranes, interact with high specificity and affinity with lectins in the extracellular milieu. **(b)** Viruses that infect animal cells, such as the influenza virus, bind to cell surface glycoproteins as the first step in infection. **(c)** Bacterial toxins, such as the cholera and pertussis toxins, bind to a surface glycolipid before entering a cell. **(d)** Some bacteria, such as *H. pylori*, adhere to and then colonize or infect animal cells. **(e)** Selectins (lectins) in the plasma membrane of certain cells mediate cell-cell interactions, such as those of T lymphocytes with the endothelial cells of the capillary wall at an infection site. **(f)** The mannose 6-phosphate receptor/lectin of the trans Golgi complex binds to the oligosaccharide of lysosomal enzymes, targeting them for transfer into the lysosome.

SUMMARY 7.4 Carbohydrates as Informational Molecules: The Sugar Code

- Monosaccharides can be assembled into an almost limitless variety of oligosaccharides, which differ in the stereochemistry and position of glycosidic bonds, the type and orientation of substituent groups, and the number and type of branches. Oligosaccharides are far more information-dense than nucleic acids or proteins.
- Lectins, proteins with highly specific carbohydrate-binding domains, are commonly found on the outer surface of cells, where they initiate interaction with other cells. In vertebrates, oligosaccharide tags "read" by lectins govern the rate of degradation of certain peptide hormones, circulating proteins, and blood cells.
- The adhesion of bacterial and viral pathogens to their animal-cell targets occurs through binding of lectins in the pathogens to

oligosaccharides in the target cell surface. Lectins are also present inside cells, where they mediate intracellular protein targeting.

- X-ray crystallography of lectin-sugar complexes shows the detailed complementarity between the two molecules, which accounts for the strength and specificity of their interactions with carbohydrates.
- Selectins are plasma membrane lectins that bind carbohydrate chains in the extracellular matrix or on the surfaces of other cells, thereby mediating the flow of information between cell and matrix or between cells.

7.5 Working with Carbohydrates

The growing appreciation of the importance of oligosaccharide structure in biological recognition has been the driving force behind the development of methods for analyzing the structure and stereochemistry of complex oligosaccharides. Oligosaccharide analysis is complicated by the fact that, unlike nucleic acids and proteins, oligosaccharides can be branched and are joined by a variety of linkages. Oligosaccharides are generally removed from their protein or lipid conjugates before analysis, then subjected to stepwise degradation with specific reagents that reveal bond position or stereochemistry. Mass spectrometry and NMR spectroscopy have also become invaluable in deciphering oligosaccharide structure.

The oligosaccharide moieties of glycoproteins or glycolipids can be released by purified enzymes—glycosidases that specifically cleave *O*- or *N*-linked oligosaccharides or lipases that remove lipid head groups. Mixtures of carbohydrates are resolved into their individual components (Fig. 7–38) by some of the same techniques useful in protein and amino acid separation: fractional precipitation by solvents, and ion-exchange and size-exclusion chromatography (see Fig. 3–18). Highly purified lectins, attached covalently to an insoluble support, are commonly used in affinity chromatography of carbohydrates (see Fig. 3–18c). Hydro-lysis of oligosaccharides and polysaccharides in strong acid yields a mixture of monosaccharides, which, after conversion to suitable volatile derivatives, may be separated, identified, and quantified by gas-liquid chromatography (p. XXX) to yield the overall composition of the polymer.

For simple, linear polymers such as amylose, the positions of the glycosidic bonds are determined by treating the intact polysaccharide with methyl iodide in a strongly basic medium to convert all free hydroxyls to acid-stable methyl ethers, then hydrolyzing the methylated polysaccharide in acid. The only free hydroxyls present in the monosaccharide derivatives so produced are those that were involved in glycosidic bonds. To determine the sequence of monosaccharide residues, including branches if they are present, exoglycosidases of known specificity are used to remove residues one at a time from the nonreducing end(s). The specificity of these exoglycosidases often allows deduction of the position and stereochemistry of the linkages. Polysaccharides and large oligosaccharides can be treated chemically or with endoglycosidases to split specific internal glycosidic bonds, producing several smaller, more easily analyzable oligosaccharides.

Oligosaccharide analysis relies increasingly on mass spectrometry and high-resolution NMR spectroscopy (see Box 4–4). Matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) and tandem mass spectrometry (MS/MS) (described in Box 3–2), are readily applicable to polar compounds like oligosaccharides. MALDI MS is a very sensitive method for determining the mass of the molecular ion (the entire oligosaccharide chain). Tandem MS reveals the mass of the molecular ion and many of its fragments, which are usually the result of breakage of the glycosidic bonds. A comparison of the masses of each fragment therefore gives information about the sequence of monosaccharide units. NMR analysis alone, especially for oligosaccharides of moderate size, can yield much information about sequence, linkage position, and anomeric carbon configuration. Automated procedures and commercial instruments are used for the routine determination of oligosaccharide structure, but the sequencing of branched oligosaccharides joined by more than one type of bond remains a far more formidable task than determining the linear sequences of proteins and nucleic acids, with monomers joined by a single bond type.

SUMMARY 7.5 Working with Carbohydrates

- Establishing the complete structure of oligosaccharides and polysaccharides requires determination of branching positions, the sequence in each branch, the configuration of each monosaccharide unit, and the positions of the glycosidic links—a more complex problem than protein and nucleic acid analysis.
- The structures of oligosaccharides and polysaccharides are usually determined by a combination of methods: specific enzymatic hydrolysis to determine stereochemistry and produce smaller fragments for further analysis; methylation analysis to locate glycosidic bonds; and stepwise degradation to determine sequence and configuration of anomeric carbons.
- Mass spectrometry and high-resolution NMR spectroscopy, applicable to small samples of carbohydrate, yield essential information about sequence, configuration at anomeric and other carbons, and positions of glycosidic bonds.

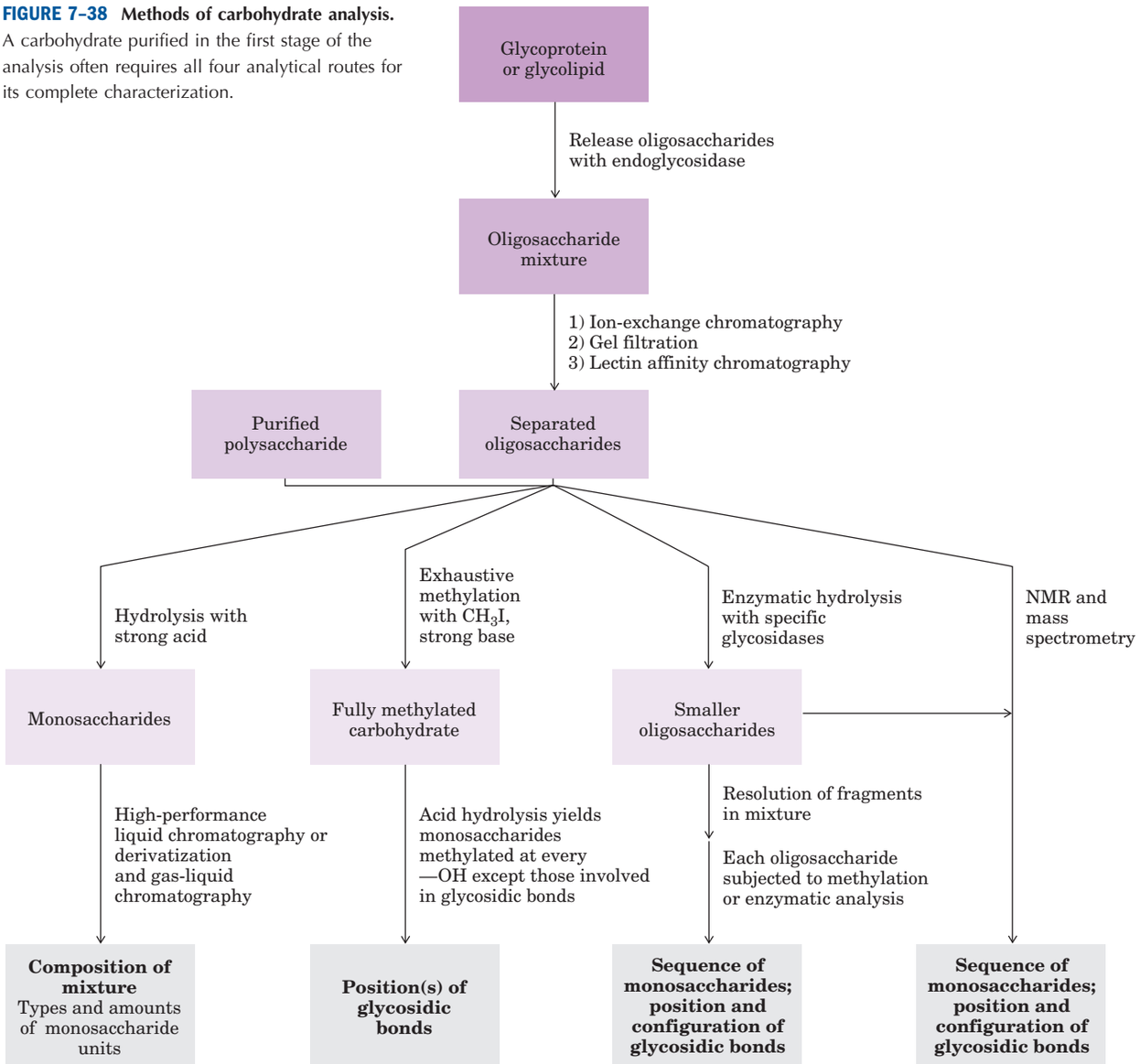


FIGURE 7–38 Methods of carbohydrate analysis. A carbohydrate purified in the first stage of the analysis often requires all four analytical routes for its complete characterization.

Key Terms

Terms in bold are defined in the glossary.

glycoconjugate 238
monosaccharide 238
oligosaccharide 238
disaccharide 238
polysaccharide 238
aldose 239
ketose 239
Fischer projection formulas 240
epimers 240
hemiacetal 242
hemiketal 242
pyranose 242
furanose 242
anomers 242
anomeric carbon 242
mutarotation 242
Haworth perspective formulas 242
reducing sugar 244
glycosidic bonds 245
reducing end 245
glycan 247
starch 248
glycogen 248
extracellular matrix 253
glycosaminoglycan 253
hyaluronic acid 254
proteoglycan 256
glycoprotein 256
glycolipid 256
lectin 262
selectins 263

Further Reading

General Background on Carbohydrate Chemistry

Aspinall, G.O. (ed.) (1982, 1983, 1985) *The Polysaccharides,* Vols 1-3, Academic Press, Inc., New York.

Collins, P.M. & Ferrier, R.J. (1995) *Monosaccharides: Their Chemistry and Their Roles in Natural Products,* John Wiley & Sons, Chichester, England.

A comprehensive text at the graduate level.

Fukuda, M. & Hindsgaul, O. (1994) *Molecular Glycobiology,* IRL Press at Oxford University Press, Inc., New York.

Thorough, advanced treatment of the chemistry and biology of cell surface carbohydrates. Good chapters on lectins, carbohydrate recognition in cell-cell interactions, and chemical synthesis of oligosaccharides.

Lehmann, J. (Haines, A.H., trans.) (1998) *Carbohydrates: Structure and Biology,* G. Thieme Verlag, New York.

The fundamentals of carbohydrate chemistry and biology, presented at a level suitable for advanced undergraduates and graduate students.

Morrison, R.T. & Boyd, R.N. (1992) *Organic Chemistry,* 6th edn, Benjamin Cummings, San Francisco.

Chapters 34 and 35 cover the structure, stereochemistry, nomenclature, and chemical reactions of carbohydrates.

Pigman, W. & Horton, D. (eds) (1970, 1972, 1980) *The Carbohydrates: Chemistry and Biochemistry,* Vols IA, IB, IIA, and IIB, Academic Press, Inc., New York.

Comprehensive treatise on carbohydrate chemistry.

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., & Marth, J. (1999) *Essentials of Glycobiology,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Structure, biosynthesis, metabolism, and function of glycosaminoglycans, proteoglycans, glycoproteins, and glycolipids, all presented at an intermediate level and very well illustrated.

Glycosaminoglycans and Proteoglycans

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

Esko, J.D. & Selleck, S.B. (2002) Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71,** 435–471.

Iozzo, R.V. (1998) Matrix proteoglycans: from molecular design to cellular function. *Annu. Rev. Biochem.* **67,** 609–652.

A review focusing on recent genetic and molecular biological studies of the matrix proteoglycans. The structure-function relationships of some paradigmatic proteoglycans are discussed in depth, and novel aspects of their biology are examined.

Jackson, R.L., Busch, S.J., & Cardin, A.D. (1991) Glycosaminoglycans: molecular properties, protein interactions, and role in physiological processes. *Physiol. Rev.* **71,** 481–539.

An advanced review of the chemistry and biology of glycosaminoglycans.

Roseman, S. (2001) Reflections on glycobiology. *J. Biol. Chem.* **276,** 41,527–41,542.

A masterful review of the history of carbohydrate and glycosaminoglycan studies, by one of the major contributors to this field.

Turnbull, J., Powell, A., & Guimond, S. (2001) Heparan sulfate: decoding a dynamic multifunctional cell regulator. *Trends Cell Biol.* **11,** 75–82.

Review of the chemistry and biology of high-sulfate domains in heparan sulfate.

Glycoproteins

Gahmberg, C.G. & Tolvanen, M. (1996) Why mammalian cell surface proteins are glycoproteins. *Trends Biochem. Sci.* **21,** 308–311.

Opdenakker, G., Rudd, P., Ponting, C., & Dwek, R. (1993) Concepts and principles of glycobiology. *FASEB J.* **7,** 1330–1337.

This review considers the genesis of glycoforms, functional roles for glycosylation, and structure-function relationships for several glycoproteins.

Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3,** 97–130.

Glycobiology and the Sugar Code

Angata, T. & Brinkman-Van der Linden, E. (2002) I-type lectins. *Biochim. Biophys. Acta* **1572,** 294–316.

Aplin, A.E., Howe, A., Alahari, S.K., & Juliano, R.L. (1998) Signal transduction and signal modulation by cell adhesion receptors: the role of integrins, cadherins, immunoglobulin-cell adhesion molecules, and selectins. *Pharmacol. Rev.* **50,** 197–263.

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

Extensive review of the biological roles of heparan sulfate.

Bertozzi, C.R. & Kiessling, L.L. (2001) Chemical glycobiology. *Science* **291,** 2357–2363.

A review of applications of chemical synthesis of carbohydrates to an understanding of the biological roles of oligosaccharides.

Borén, T., Normark, S., & Falk, P. (1994) *Helicobacter pylori:* molecular basis for host recognition and bacterial adherence. *Trends Microbiol.* **2,** 221–228.

A look at the role of the oligosaccharides that determine blood type in the adhesion of *H. pylori* to the stomach lining, producing ulcers.

Cooper, D.N. (2002) Galectinomics: finding themes in complexity. *Biochim. Biophys. Acta* **1572,** 209–231.

A review of the genomic evidence for the conservation of the galectins, a family of lectins.

Cornejo, C.J., Winn, R.K., & Harlan, J.M. (1997) Anti-adhesion therapy. *Adv. Pharmacol.* **39,** 99–142.

Analogs of recognition oligosaccharides are used to block adhesion of a pathogen to its host-cell target.

Dahms, N.M. & Hancock, M.K. (2002) P-type lectins. *Biochim. Biophys. Acta* **1572,** 317–340.

Gabius, H.-J. (2000) Biological information transfer beyond the genetic code: the sugar code. *Naturwissenschaften* **87,** 108–121.

Description of the basis for the high information density in oligosaccharides, with examples of the importance of the sugar code.

Gabius, H.-J., Andre, S., Kaltner, H., & Siebert, H.C. (2002) The sugar code: functional lectinomics. *Biochim. Biophys. Acta* **1572,** 165–177.
This review examines the reasons for the relatively late appreciation of the informational roles of oligosaccharides and polysaccharides.

Ghosh, P., Dahms, N.M., & Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twists in the tale. *Nat. Rev. Mol. Cell Biol.* **4,** 202–212.

Helenius, A. & Aebi, M. (2001) Intracellular functions of *N*-linked glycans. *Science* **291,** 2364–2369.
Review of the synthesis of *N*-linked oligosaccharides and their targeting functions.

Hooper, L.A., Manzella, S.M., & Baenziger, J.U. (1996) From legumes to leukocytes: biological roles for sulfated carbohydrates. *FASEB J.* **10,** 1137–1146.
Evidence for roles of sulfated oligosaccharides in peptide hormone half-life, symbiont interactions in nitrogen-fixing legumes, and lymphocyte homing.

Horwitz, A.F. (1997) Integrins and health. *Sci. Am.* **276** (May), 68–75.
Article on the role of integrins in cell-cell adhesion, and possible roles in arthritis, heart disease, stroke, osteoporosis, and the spread of cancer.

Iozzo, R.V. (2001) Heparan sulfate proteoglycans: intricate molecules with intriguing functions. *J. Clin. Invest.* **108,** 165–167.
Introduction to a series of papers on heparan sulfates published in this issue; all are rewarding reading.

Kilpatrick, D.C. (2002) Animal lectins: a historical introduction and overview. *Biochim. Biophys. Acta* **1572,** 187–197.
Introduction to a series of excellent reviews on lectins and their biological roles, all published in this issue.

Loris, R. (2002) Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* **1572,** 198–208.

McEver, R.P., Moore, K.L., & Cummings, R.D. (1995) Leukocyte trafficking mediated by selectin-carbohydrate interactions. *J. Biol. Chem.* **270,** 11,025–11,028.
This short review focuses on the interaction of selectins with their carbohydrate ligands.

Reuter, G. & Gabius, H.-J. (1999) Eukaryotic glycosylation: whim of nature or multipurpose tool? *Cell. Mol. Life Sci.* **55,** 368–422.
Excellent review of the chemical diversity of oligosaccharides and polysaccharides and of biological processes dependent upon protein-carbohydrate recognition.

Selleck, S. (2000) Proteoglycans and pattern formation: sugar biochemistry meets developmental genetics. *Trends Genet.* **16,** 206–212.
A short, intermediate-level review of the genetic evidence for proteoglycans as determinants of development.

Weigel, P.H. & Yik, J.H. (2002) Glycans as endocytosis signals: the cases of the asialoglycoprotein and hyaluronan/chondroitin sulfate receptors. *Biochim. Biophys. Acta* **1572,** 341–363.

Weiss, W.I. & Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65,** 441–473.
Good treatment of the chemical basis of carbohydrate-protein interactions.

Working with Carbohydrates

Chaplin, M.F. & Kennedy, J.F. (eds) (1994) *Carbohydrate Analysis: A Practical Approach,* 2nd edn, IRL Press, Oxford.
Very useful manual for analysis of all types of sugar-containing molecules—monosaccharides, polysaccharides and glycosaminoglycans, glycoproteins, proteoglycans, and glycolipids.

Dell, A. & Morris, H.R. (2001) Glycoprotein structure determination by mass spectroscopy. *Science* **291,** 2351–2356.
Short review of the uses of MALDI MS and tandem MS in oligosaccharide structure determination.

Dwek, R.A., Edge, C.J., Harvey, D.J., & Wormald, M.R. (1993) Analysis of glycoprotein-associated oligosaccharides. *Annu. Rev. Biochem.* **62,** 65–100.
Excellent survey of the uses of NMR, mass spectrometry, and enzymatic reagents to determine oligosaccharide structure.

Fukuda, M. & Kobata, A. (1993) *Glycobiology: A Practical Approach,* IRL Press, Oxford.
A how-to manual for the isolation and characterization of the oligosaccharide moieties of glycoproteins, using the whole range of modern techniques. Available as part of the IRL Press *Practical Approach Series* on CD-ROM, from Oxford University Press (www.oup-usa.org/acadsci/pasbooks.html).

Jay, A. (1996) The methylation reaction in carbohydrate analysis. *J. Carbohydr. Chem.* **15,** 897–923.
Thorough description of methylation analysis of carbohydrates.

Lennarz, W.J. & Hart, G.W. (eds) (1994) *Guide to Techniques in Glycobiology,* Methods in Enzymology, Vol. 230, Academic Press, Inc., New York.
Practical guide to working with oligosaccharides.

McCleary, B.V. & Matheson, N.K. (1986) Enzymic analysis of polysaccharide structure. *Adv. Carbohydr. Chem. Biochem.* **44,** 147–276.
On the use of purified enzymes in analysis of structure and stereochemistry.

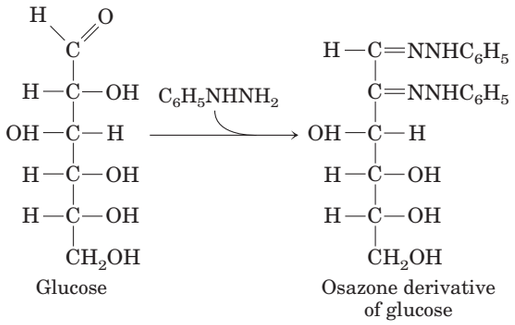
Rudd, P.M., Guile, G.R., Kuester, B., Harvey, D.J., Opdenakker, G., & Dwek, R.A. (1997) Oligosaccharide sequencing technology. *Nature* **388,** 205–207.

Problems

1. Determination of an Empirical Formula An unknown substance containing only C, H, and O was isolated from goose liver. A 0.423 g sample produced 0.620 g of CO_2 and 0.254 g of H_2O after complete combustion in excess oxygen. Is the empirical formula of this substance consistent with its being a carbohydrate? Explain.

2. Sugar Alcohols In the monosaccharide derivatives known as sugar alcohols, the carbonyl oxygen is reduced to a hydroxyl group. For example, D-glyceraldehyde can be reduced to glycerol. However, this sugar alcohol is no longer designated D or L. Why?

3. Melting Points of Monosaccharide Osazone Derivatives Many carbohydrates react with phenylhydrazine ($C_6H_5NHNH_2$) to form bright yellow crystalline derivatives known as osazones:



The melting temperatures of these derivatives are easily determined and are characteristic for each osazone. This information was used to help identify monosaccharides before the development of HPLC or gas-liquid chromatography. Listed below are the melting points (MPs) of some aldose-osazone derivatives:

Monosaccharide	*MP of anhydrous monosaccharide (°C)*	*MP of osazone derivative (°C)*
Glucose	146	205
Mannose	132	205
Galactose	165–168	201
Talose	128–130	201

As the table shows, certain pairs of derivatives have the same melting points, although the underivatized monosaccharides do not. Why do glucose and mannose, and galactose and talose, form osazone derivatives with the same melting points?

4. Interconversion of D-Glucose Forms A solution of one stereoisomer of a given monosaccharide rotates plane-polarized light to the left (counterclockwise) and is called the levorotatory isomer, designated (−); the other stereoisomer rotates plane-polarized light to the same extent but to the right (clockwise) and is called the dextrorotatory isomer, designated (+). An equimolar mixture of the (+) and (−) forms does not rotate plane-polarized light.

The optical activity of a stereoisomer is expressed quantitatively by its *optical rotation,* the number of degrees by which plane-polarized light is rotated on passage through a given path length of a solution of the compound at a given concentration. The *specific rotation* $[\alpha]_D^{25°C}$ of an optically active compound is defined thus:

$$[\alpha]_D^{25°C} = \frac{\text{observed optical rotation (°)}}{\text{optical path length (dm)} \times \text{concentration (g/mL)}}$$

The temperature and the wavelength of the light employed (usually the D line of sodium, 589 nm) must be specified in the definition.

A freshly prepared solution of α-D-glucose shows a specific rotation of +112°. Over time, the rotation of the solution gradually decreases and reaches an equilibrium value corresponding to $[\alpha]_D^{25°C} = +52.5°$. In contrast, a freshly prepared solution of β-D-glucose has a specific rotation of +19°. The rotation of this solution increases over time to the same equilibrium value as that shown by the α anomer.

(a) Draw the Haworth perspective formulas of the α and β forms of D-glucose. What feature distinguishes the two forms?

(b) Why does the specific rotation of a freshly prepared solution of the α form gradually decrease with time? Why do solutions of the α and β forms reach the same specific rotation at equilibrium?

(c) Calculate the percentage of each of the two forms of D-glucose present at equilibrium.

5. A Taste of Honey The fructose in honey is mainly in the β-D-pyranose form. This is one of the sweetest carbohydrates known, about twice as sweet as glucose. The β-D-furanose form of fructose is much less sweet. The sweetness of honey gradually decreases at a high temperature. Also, high-fructose corn syrup (a commercial product in which much of the glucose in corn syrup is converted to fructose) is used for sweetening *cold* but not *hot* drinks. What chemical property of fructose could account for both these observations?

6. Glucose Oxidase in Determination of Blood Glucose The enzyme glucose oxidase isolated from the mold *Penicillium notatum* catalyzes the oxidation of β-D-glucose to D-glucono-δ-lactone. This enzyme is highly specific for the β anomer of glucose and does not affect the α anomer. In spite of this specificity, the reaction catalyzed by glucose oxidase is commonly used in a clinical assay for total blood glucose—that is, for solutions consisting of a mixture of β- and α-D-glucose. How is this possible? Aside from allowing the detection of smaller quantities of glucose, what advantage does glucose oxidase offer over Fehling's reagent for the determination of blood glucose?

7. Invertase "Inverts" Sucrose The hydrolysis of sucrose (specific rotation +66.5°) yields an equimolar mixture of D-glucose (specific rotation +52.5°) and D-fructose (specific rotation −92°). (See Problem 4 for details of specific rotation.)

(a) Suggest a convenient way to determine the rate of hydrolysis of sucrose by an enzyme preparation extracted from the lining of the small intestine.

(b) Explain why an equimolar mixture of D-glucose and D-fructose formed by hydrolysis of sucrose is called invert sugar in the food industry.

(c) The enzyme invertase (now commonly called sucrase) is allowed to act on a 10% (0.1 g/mL) solution of sucrose until hydrolysis is complete. What will be the observed optical rotation of the solution in a 10 cm cell? (Ignore a possible small contribution from the enzyme.)

8. Manufacture of Liquid-Filled Chocolates The manufacture of chocolates containing a liquid center is an interesting application of enzyme engineering. The flavored liquid center consists largely of an aqueous solution of sugars rich in fructose to provide sweetness. The technical dilemma is the following: the chocolate coating must be prepared by pouring hot melted chocolate over a solid (or almost solid) core, yet the final product must have a liquid, fructose-rich center. Suggest a way to solve this problem. (Hint: Sucrose is much less soluble than a mixture of glucose and fructose.)

9. Anomers of Sucrose? Although lactose exists in two anomeric forms, no anomeric forms of sucrose have been reported. Why?

10. Physical Properties of Cellulose and Glycogen The almost pure cellulose obtained from the seed threads of *Gossypium* (cotton) is tough, fibrous, and completely insoluble in water. In contrast, glycogen obtained from muscle or liver disperses readily in hot water to make a turbid solution. Although they have markedly different physical properties, both substances are composed of (1→4)-linked D-glucose polymers of comparable molecular weight. What structural features of these two polysaccharides underlie their different physical properties? Explain the biological advantages of their respective properties.

11. Growth Rate of Bamboo The stems of bamboo, a tropical grass, can grow at the phenomenal rate of 0.3 m/day under optimal conditions. Given that the stems are composed almost entirely of cellulose fibers oriented in the direction of growth, calculate the number of sugar residues per second that must be added enzymatically to growing cellulose chains to account for the growth rate. Each D-glucose unit contributes ~0.5 nm to the length of a cellulose molecule.

12. Glycogen as Energy Storage: How Long Can a Game Bird Fly? Since ancient times it has been observed that certain game birds, such as grouse, quail, and pheasants, are easily fatigued. The Greek historian Xenophon wrote, "The bustards . . . can be caught if one is quick in starting them up, for they will fly only a short distance, like partridges, and soon tire; and their flesh is delicious." The flight muscles of game birds rely almost entirely on the use of glucose 1-phosphate for energy, in the form of ATP (Chapter 14). In game birds, glucose 1-phosphate is formed by the breakdown of stored muscle glycogen, catalyzed by the enzyme glycogen phosphorylase. The rate of ATP production is limited by the rate at which glycogen can be broken down. During a "panic flight," the game bird's rate of glycogen breakdown is quite high, approximately 120 μmol/min of glucose 1-phosphate produced per gram of fresh tissue. Given that the flight muscles usually contain about 0.35% glycogen by weight, calculate how long a game bird can fly. (Assume the average molecular weight of a glucose residue in glycogen is 160 g/mol.)

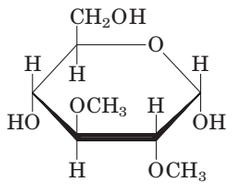
13. Volume of Chondroitin Sulfate in Solution One critical function of chondroitin sulfate is to act as a lubricant in skeletal joints by creating a gel-like medium that is resilient to friction and shock. This function appears to be related to a distinctive property of chondroitin sulfate: the volume occupied by the molecule is much greater in solution than in the dehydrated solid. Why is the volume occupied by the molecule so much larger in solution?

14. Heparin Interactions Heparin, a highly negatively charged glycosaminoglycan, is used clinically as an anticoagulant. It acts by binding several plasma proteins, including antithrombin III, an inhibitor of blood clotting. The 1:1 binding of heparin to antithrombin III appears to cause a conformational change in the protein that greatly increases its ability to inhibit clotting. What amino acid residues of antithrombin III are likely to interact with heparin?

15. Information Content of Oligosaccharides The carbohydrate portion of some glycoproteins may serve as a cellular recognition site. In order to perform this function, the oligosaccharide moiety of glycoproteins must have the potential to exist in a large variety of forms. Which can produce a greater variety of structures: oligopeptides composed of five different amino acid residues or oligosaccharides composed of five different monosaccharide residues? Explain.

16. Determination of the Extent of Branching in Amylopectin The extent of branching (number of (α1→6) glycosidic bonds) in amylopectin can be determined by the following procedure. A sample of amylopectin is exhaustively methylated—treated with a methylating agent (methyl iodide) that replaces all the hydrogens of the sugar hydroxyls with methyl groups, converting —OH to —OCH_3. All the glycosidic bonds in the treated sample are then hydrolyzed in aqueous acid. The amount of 2,3-di-*O*-methylglucose in the hydrolyzed sample is determined.

(a) Explain the basis of this procedure for determining the number of (α1→6) branch points in amylopectin. What happens to the unbranched glucose residues in amylopectin during the methylation and hydrolysis procedure?



2,3-Di-*O*-methylglucose

(b) A 258 mg sample of amylopectin treated as described above yielded 12.4 mg of 2,3-di-*O*-methylglucose. Determine what percentage of the glucose residues in amylopectin contain an (α1→6) branch. (Assume that the average molecular weight of a glucose residue in amylopectin is 162 g/mol.)

17. Structural Analysis of a Polysaccharide A polysaccharide of unknown structure was isolated, subjected to exhaustive methylation, and hydrolyzed. Analysis of the products revealed three methylated sugars in the ratio 20:1:1. The sugars were 2,3,4-tri-*O*-methyl-D-glucose; 2,4-di-*O*-methyl-D-glucose; and 2,3,4,6-tetra-*O*-methyl-D-glucose. What is the structure of the polysaccharide?