

CHAPTER 3

Glycans and Lipids

Nucleic acids and proteins are the molecules whose synthesis is specified by the genetic code. There are two other broad classes of molecules of central importance to life—namely, **glycans** (carbohydrates) and **lipids**. These molecules, which are the focus of this chapter, are not directly coded for by DNA, but their presence is essential for the proper functioning of all cells.

Glycans are sugars and sugar polymers, in which the predominant chemical unit is hydroxylated carbon (HCOH). The building blocks of typical glycans have three to nine carbons, and are generally very soluble in water because of the presence of hydroxyl groups. Most of the carbons in glycans are chiral, leading to many stereochemical variations in the structures of sugars. The building blocks of glycans are linked to form large, complex, and diverse polymers that serve a wide variety of roles in cells, including energy storage, structural support, and cellular signaling. Glycans are among the most abundant organic compounds on Earth.

Lipids are diverse organic compounds that are soluble in nonpolar organic solvents but are insoluble in water. Examples of lipids include fats, waxes, and oils. The lipids that are most relevant for this chapter are **amphipathic** molecules—that is, they have both polar, hydrophilic parts and nonpolar, hydrophobic parts, a feature critical to their functions. A distinctive and critical property of such molecules is their ability to form cellular membranes. In contrast to the **biopolymers** (that is, the proteins, nucleic acids, and polysaccharides, all of which have many covalently linked residues), lipid structures are noncovalent assemblies of large numbers of molecules of moderate size (the molecular weight of a typical lipid molecule is roughly 1 kD). Membrane bilayers assembled from lipid molecules are critical for the formation of cells, and for forming compartments within cells. The development of membrane bilayers was necessarily an early feature in the evolution of life.

A. GLYCANS

3.1 Simple sugars are comprised primarily of hydroxylated carbons

The term “carbohydrate” comes from hydrated carbon. Addition of water molecules to carbons gives an empirical formula of $(\text{CH}_2\text{O})_n$, which reflects the composition of the core elements in the molecules of this class. Among compounds with this approximate formula, there are many isomers (that is, molecules with the same atoms but with differences in bonding and/or stereochemistry) that occur in biological systems. The molecules with small values of n are generally called **sugars** (one of which is sucrose, the common table sugar; see Figure 3.1). Many sugars taste sweet, but this term is used whether they taste sweet or not.

The simple sugars, and modified versions of them, also form polymers. Both homopolymers, made of just one kind of sugar building block, and heteropolymers, made from different kinds of building blocks, play important roles in

Glycans

Glycans are sugars or sugar derivatives, and their polymers. The simplest individual subunits of glycans have roughly the formula $(\text{HCOH})_n$ with n between 3 and 9, but most commonly 5, 6, or 7. Many glycans also have sugars with additional functionality relative to these simplest subunits. Glycans are also known as carbohydrates.

Lipids

Lipids are amphipathic molecules (also called amphiphilic, that is, part hydrophobic and part hydrophilic). Lipids are the primary components of biological membranes. Although membranes contain embedded proteins, most of the critical properties of membranes arise from the lipid components.

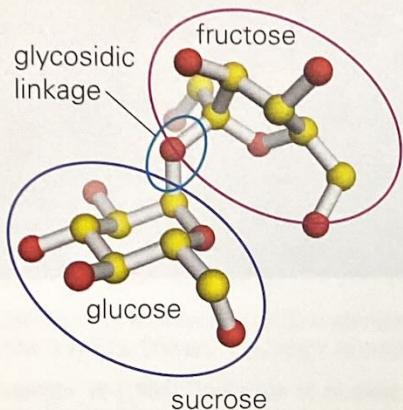


Figure 3.1 The structure of sucrose, common table sugar. The chemical formula of sucrose is $C_{12}H_{22}O_{11}$. Sucrose is made from two smaller building blocks, glucose and fructose, joined by a glycosidic linkage. The six-membered ring is glucose, and the five-membered ring is fructose. This depiction, like most molecular illustrations in this book, has yellow carbon atoms and red oxygen atoms. The hydrogen atoms are not shown.

living systems. Most of the common sugar building blocks contain five, six, or seven “hydrated carbon” units (CH_2O), which are best thought of as $H-C-O-H$ to reflect the typical bonding pattern.

Carbon always forms four bonds, so the hydrated carbons can form two additional bonds to other carbons, forming chains (for example, a six-carbon chain of CH_2O units is shown in Figure 3.2A). Note that there are unsatisfied bonds at the ends of the chain, so this would not actually be a stable molecule. One modification that satisfies the bonding is to shift one hydrogen from an OH to a terminal carbon, and add a second bond to the oxygen, making a carbonyl group, as shown in Figure 3.2B. When the oxygen double bond is at the end of the chain, the molecule is known as an **aldehyde**, whereas if the oxygen double bond is somewhere in the middle, the molecule is known as a **ketone**. Sugars are classified as **aldoses** or **ketoses**, respectively, depending on whether they contain aldehyde or ketone groups.

Two examples of common sugars are **glucose** (an aldohexose that is a major energy source in eukaryotic cells), and **fructose** (a ketohexose, also often called fruit sugar because of its common occurrence in fruit). The chemical structures of glucose and fructose, drawn using several different chemical conventions, are shown in Figure 3.3. There are many different ways to represent chemical structures, including their stereochemistry (consult an organic chemistry textbook to review these conventions). The different schematic representations shown in Figure 3.3 are those most commonly used for describing sugars and glycans. Some representations, such as the **Fischer projection**, are very convenient for comparing the stereochemistry of different sugar isomers, but are poor for seeing the three-dimensional relationship of atoms and the real geometry of the molecules. The representation indicating the three-dimensional structure provides a better sense of the conformational features, but may be less suited for seeing relationships between isomers.

Anomers

Anomers are forms of sugars that differ only in the equatorial or axial position of the hydroxyl at the position of ring closure.

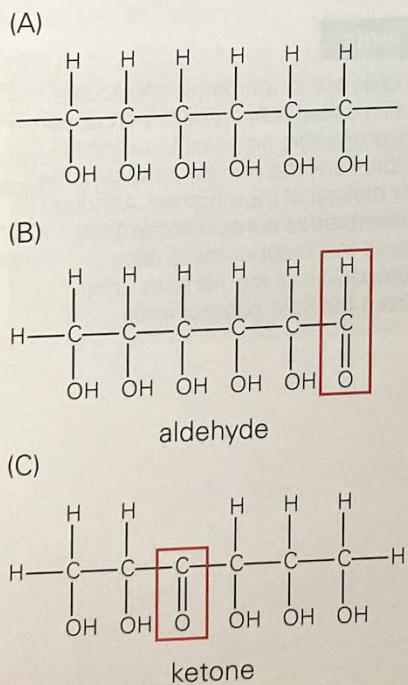


Figure 3.2 Sugar aldehydes and ketones. (A) A chain of bonded hydroxylated carbons leaves bonding unsatisfied at the ends. (B) In some sugars, a double bond to a terminal oxygen is formed, making an aldehyde. (C) In other sugars, the double-bonded oxygen is not at the end of the chain, making a ketone.

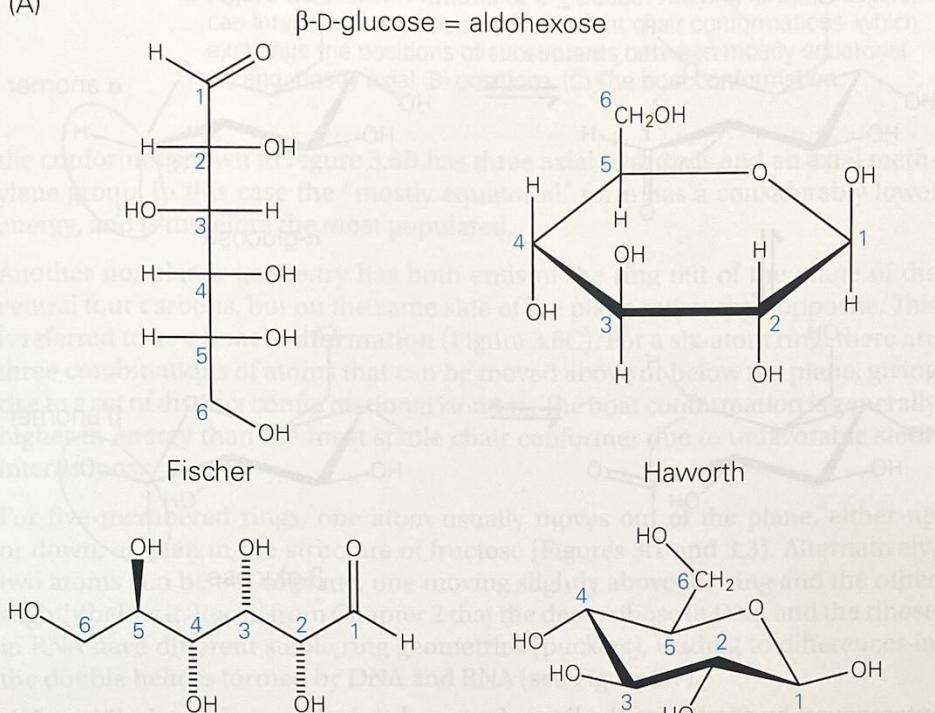
3.2 Many cyclic sugar molecules can exist in alternative anomeric forms

The presence of the carbonyl carbon allows for the facile cyclization of sugars, as shown in Figure 3.4. In this reaction, one of the hydroxyl groups of the sugar carries out an intramolecular nucleophilic addition to the carbonyl group, leading to ring structures. For the commonly occurring sugars in biological systems, the cyclic forms are favored (for glucose, about 99.98% is cyclic in water), but the open and ring structures are in equilibrium.

The closure to the ring form can occur through attack of a hydroxyl with the carbonyl in either of two orientations, leading to a different position of the hydroxyl derived from the oxygen of the carbonyl. The two isomers (that is, the two structures on the right hand side of Figure 3.4) are called **anomers**, and they differ only in the orientation of the hydroxyl group at the position of ring closure.

Alternative anomers are a general feature of the cyclic forms of sugars. The carbon at which this isomerism occurs is called the **anomeric carbon**. The anomers can

(A)



(B)

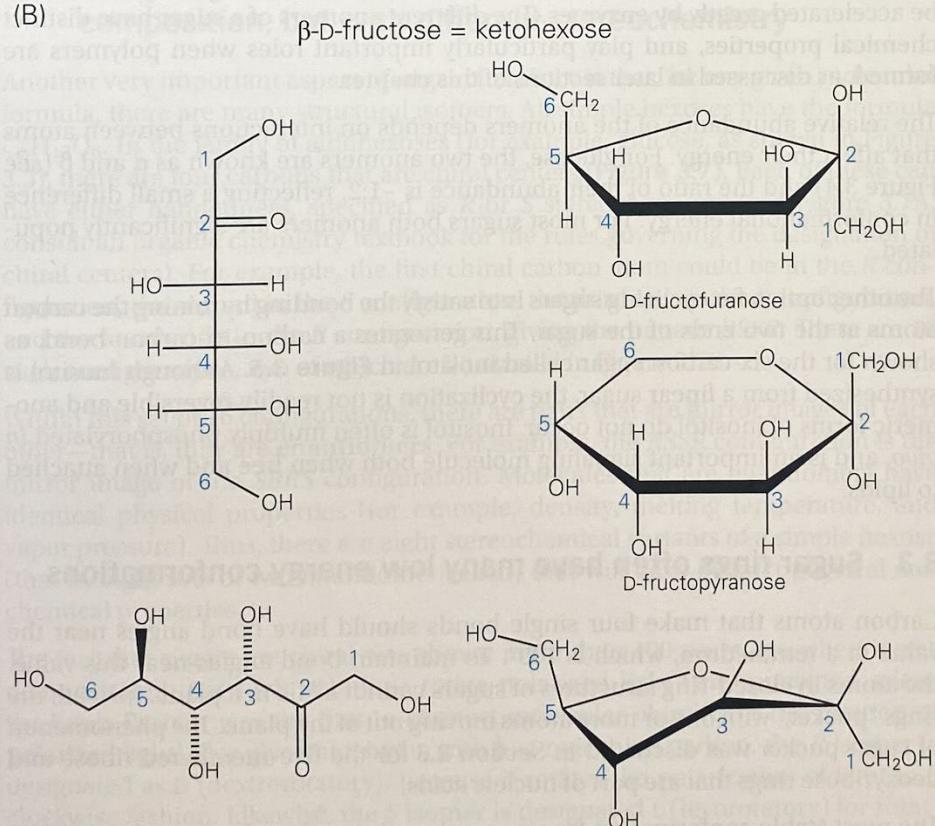
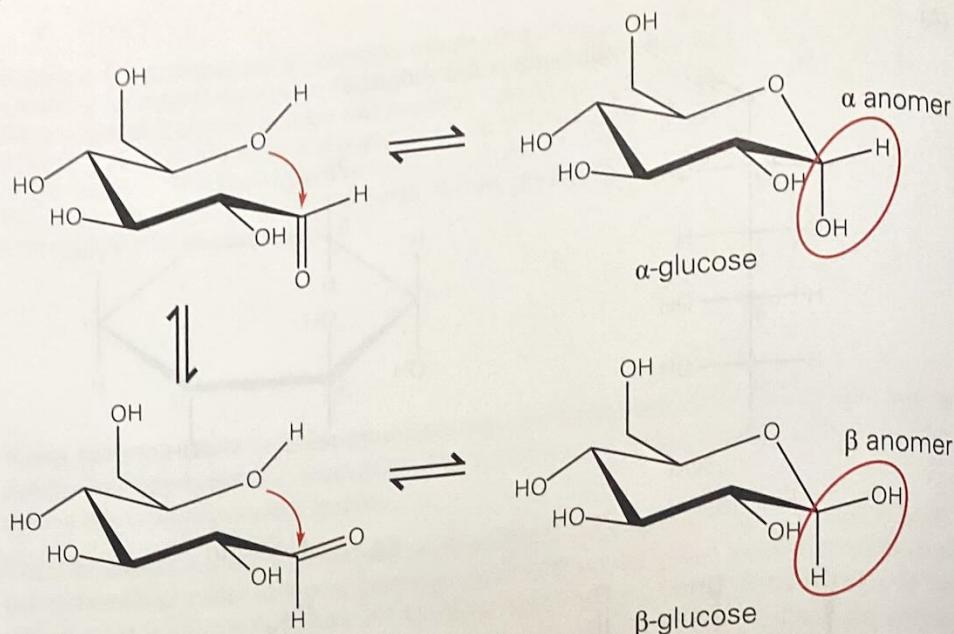


Figure 3.3 The structures of two sugars are shown. (A) Glucose (an aldohexose) and (B) fructose (a ketohexose that can cyclize in two different ways) are shown in several representations. Both exist as an equilibrium mixture of the open, linear chain and the closed, cyclic form (the cyclic forms being >99% of the mixture). In addition to molecular structures that specifically indicate stereochemistry, sugars are also commonly drawn in representations known as Fischer and Haworth projections. These are often easier for comparing the relative stereochemistries of different sugars. Carbons are numbered consecutively from the top carbon in the Fischer projection.

Figure 3.4 Linear and cyclized forms of sugars. The linear form of glucose is shown at the left in conformations that lead to the ring closure reaction. Depending on the orientation of the aldehyde relative to the attacking C5-hydroxyl, closure can lead to either the α or β anomer, shown at the right.



Axial and equatorial

Axial and equatorial are terms that define the directions of substituents of rings such as sugars, relative to the plane of the ring.

interconvert by opening to the linear form and reclosing, as shown in Figure 3.4. This process is relatively slow, taking minutes to hours in neutral water, but it can be accelerated greatly by enzymes. The different anomers of a sugar have distinct chemical properties, and play particularly important roles when polymers are formed, as discussed in later sections of this chapter.

The relative abundance of the anomers depends on interactions between atoms that affect their energy. For glucose, the two anomers are known as α and β (see Figure 3.4) and the ratio of their abundance is ~1:2, reflecting a small difference in conformational energy. For most sugars both anomers are significantly populated.

The other option for cyclizing sugars is to satisfy the bonding by joining the carbon atoms at the two ends of the sugar. This generates a carbon-to-carbon bond, as shown for the six-carbon sugar called inositol in Figure 3.5. Although inositol is synthesized from a linear sugar, the cyclization is not readily reversible and anomeric forms of inositol do not occur. Inositol is often multiply phosphorylated *in vivo*, and is an important signaling molecule both when free and when attached to lipids.

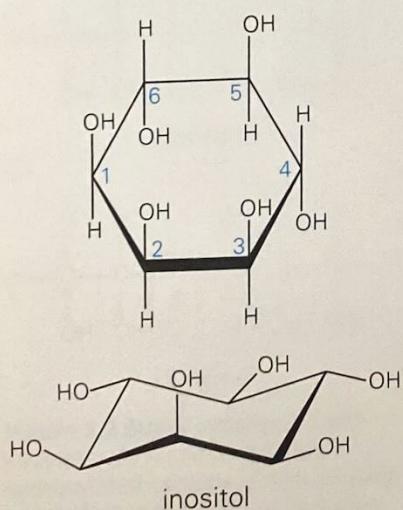


Figure 3.5 Inositol, a six-carbon sugar, is atypical in not having an oxygen in the ring. The stereochemistry of the natural isomer is shown.

3.3 Sugar rings often have many low energy conformations

Carbon atoms that make four single bonds should have bond angles near the value in a tetrahedron, which is 109° . To maintain bond angles near this value, the atoms in closed-ring structures of sugars cannot all lie in a plane. Instead, the rings “pucker,” with one or more atoms moving out of the plane. The phenomenon of sugar pucker was discussed in Section 2.6 for the five-membered ribose and deoxyribose rings that are part of nucleic acids.

The most stable conformation for six-membered rings has the atoms at opposite ends of the ring out of the plane, one atom above the plane and another below the plane of the other four, leading to a **chair conformation**. For a ring of six atoms, there are two distinct chair conformers, which are shown for α -glucose in Figure 3.6A and B. In a chair conformation, substituents, such as the hydroxyls, can either point out from the ring (termed **equatorial**) or point up or down (termed **axial**). There is less space for nonhydrogen atoms in the axial positions, so the conformational energy is lower when most of the hydroxyls are in equatorial positions. The chair conformer shown in Figure 3.6A has just one axial hydroxyl, while

Figure 3.6 Conformations of α -glucose. The ring forms of sugars can interconvert between two different chair conformations, which exchange the positions of substituents between mostly equatorial (A) and mostly axial (B) positions. (C) The boat conformation.

the conformer shown in Figure 3.6B has three axial hydroxyls and an axial methylene group. In this case the “mostly equatorial” form has a considerably lower energy, and is therefore the most populated.

Another nonplanar geometry has both ends of the ring out of the plane of the central four carbons, but on the same side of the plane rather than opposite. This is referred to as a **boat conformation** (Figure 3.6C). For a six-atom ring, there are three combinations of atoms that can be moved above or below the plane, giving rise to a set of distinct conformational isomers. The boat conformation is generally higher in energy than the most stable chair conformer due to unfavorable steric interactions.

For five-membered rings, one atom usually moves out of the plane, either up or down, as seen in the structure of fructose (Figures 3.1 and 3.3). Alternatively, two atoms can be out of plane, one moving slightly above the ring and the other slightly below it. Recall from Chapter 2 that the deoxyribose in DNA and the ribose in RNA have different stable ring geometries (puckers), leading to differences in the double helices formed by DNA and RNA (see Figure 2.7).

3.4 Many sugars are structural isomers of identical composition, but with different stereochemistry

Another very important aspect of sugar chemistry is that, for any given molecular formula, there are many structural isomers. All simple hexoses have the formula $C_6H_{12}O_6$. In the family of aldohexoses (for example, glucose, as shown in Figure 3.3), there are four carbons that are chiral centers (Figure 3.7). Each of these can have either handedness, specified as *R* or *S* stereochemistry (see Figure 3.7A; consult an organic chemistry textbook for the rules governing the designation of chiral centers). For example, the first chiral carbon atom could be in the *R* configuration, the second in the *S* configuration, the third one in the *S* configuration, and the fourth one in the *R* configuration, giving the pattern *RSSR*. There are 16 such configurations, two of which are illustrated in Figure 3.7B.

Within this set of 16 configurations, there are pairs that are mirror images of each other—that is, they are **enantiomers**. For example, the *RSSR* configuration is the mirror image of the *SRRS* configuration. Molecules that are enantiomers have identical physical properties (for example, density, melting temperature, and vapor pressure). Thus, there are eight stereochemical variants of a simple hexose (that is, eight sets of two enantiomers each) that will have distinct physical and chemical properties.

The fact that sugars are chiral was shown more than 100 years ago when it was demonstrated that their solutions rotate polarized light (the renowned scientist Louis Pasteur was the first to connect molecular chirality to the rotation of polarized light). For glyceraldehyde, with just one chiral carbon, the *R* isomer is designated as *D* (dextrorotatory), because it rotates the polarization of light in a clockwise fashion. Likewise, the *S* isomer is designated *L* (levorotatory) for rotating the polarization counterclockwise.

The stereochemical designation of sugars is based on the chirality of the carbon furthest from the aldehyde. If it matches the chirality of *D*-glyceraldehyde, then it is denoted a *D* sugar, whereas if it matches *L*-glyceraldehyde, then it is an *L* sugar. It turns out that only the *D* forms of the aldohexoses occur in nature. Of the eight possible *D* aldohexose sugars, three of them—glucose, mannose and galactose—occur quite commonly in biological systems; their structures are shown in Figure 3.8 in both linear (Fischer) and cyclic forms.

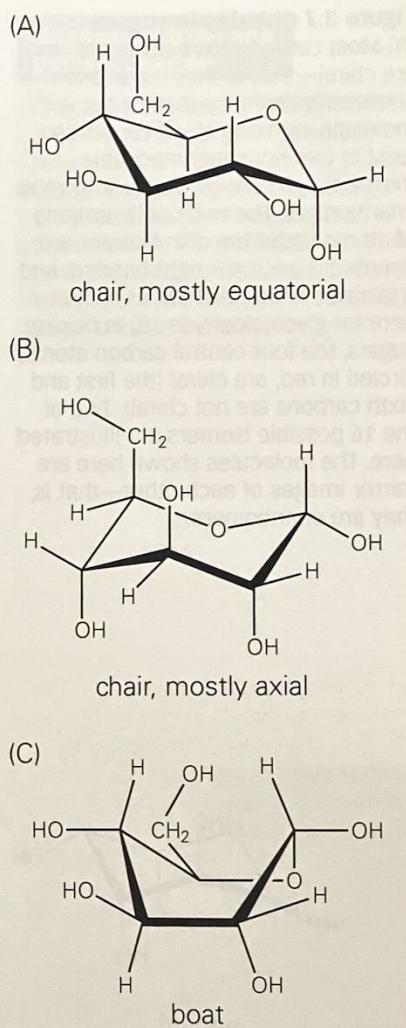


Figure 3.7 Chirality in sugars.

(A) Most carbon atoms in sugars are chiral—that is, they have four chemically distinct substituents. A molecule with one chiral center can exist in two nonsuperimposable mirror image configurations, known as enantiomers. The two configurations of atoms about the chiral center are denoted *R* (*rectus* = right handed) and *S* (*sinister* = left handed), as shown here for glyceraldehyde. (B) In hexose sugars, the four central carbon atoms, circled in red, are chiral (the first and sixth carbons are not chiral). Two of the 16 possible isomers are illustrated here. The molecules shown here are mirror images of each other—that is, they are enantiomers.

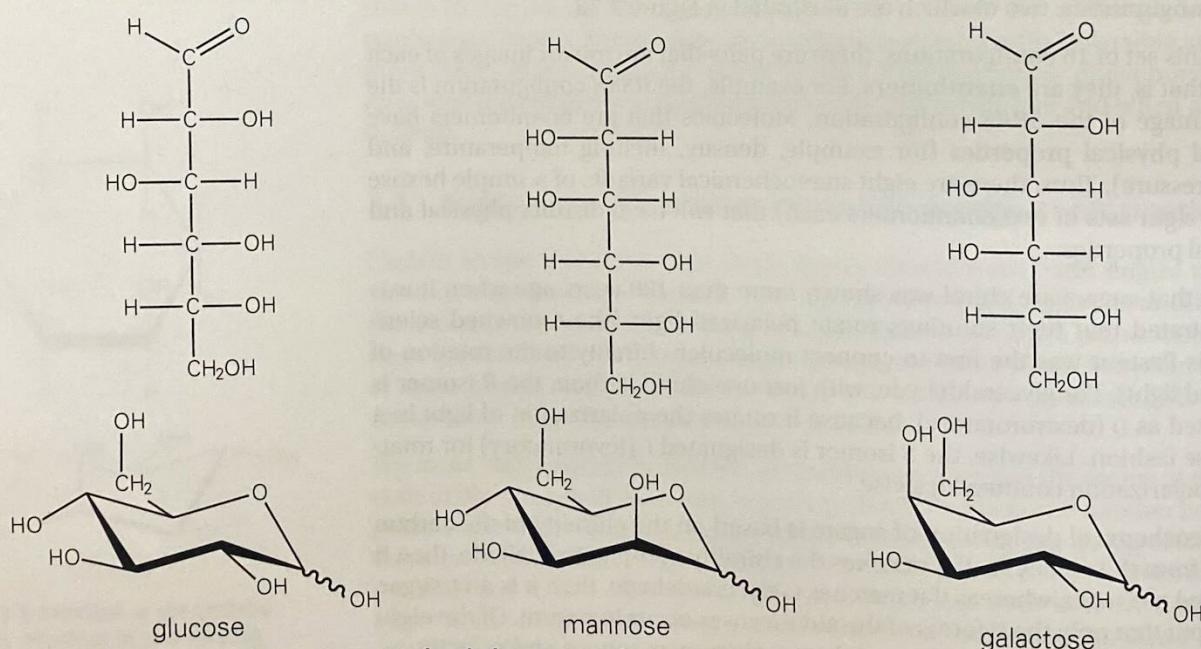
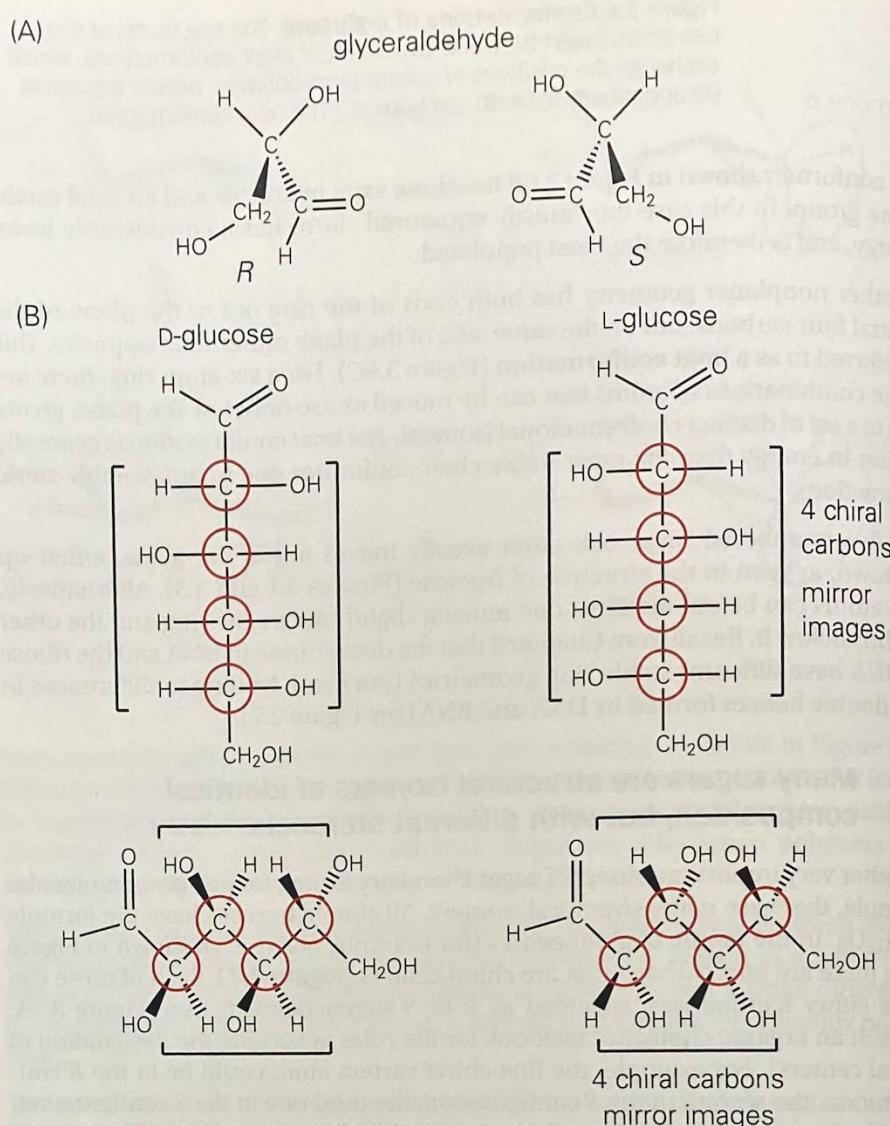


Figure 3.8 The structures of three common simple hexoses—glucose, mannose, and galactose—shown in both Fischer and stereochemical representations. Each sugar has α and β anomers. A “squiggly” line is used to indicate either possibility (that is, hydroxyl axial or equatorial) at the anomeric position.

3.5 Some sugars have other chemical functionalities in addition to alcohol groups

Oligosaccharides are polymeric glycans that appear on cell surfaces (the term **polysaccharide** is used if there are more than about 30 sugar units in the polymer). Oligosaccharides are built up of repeating units of the simple sugars mannose and galactose (see Figure 3.8), but also contain other sugars with additional chemical functionality. Several of the sugars that are present in oligosaccharides arise from relatively small modifications of the simple sugars. For example, replacing the commonly occurring $-\text{CH}_2\text{OH}$ group with a methyl group ($-\text{CH}_3$) gives **fucose** (which is also called 6-deoxy-L-galactose), and replacing one hydroxyl with an amino group gives **glucosamine** or **galactosamine**. These are usually acetylated to give *N*-acetylglucosamine, or *N*-acetylgalactosamine (Figure 3.9).

Some sugar residues have more extensive modification. Sialic acid, for example, has additional hydroxylated carbons, a carboxylic acid, and an acetylated amine. The structures of the most commonly occurring sugars in oligosaccharides are shown in Figure 3.9, and are listed in Table 3.1. Note that, as for amino acids and nucleotides, there are abbreviations for sugars that are a useful shorthand for specifying sugars and their combinations. It is important to remember that there are many additional modified sugars that occur in different particular contexts, some of which contain other distinct functional groups including carboxylates, phosphates, and/or sulfates.

oligosaccharide and polysaccharide

Oligosaccharides are glycans that are polymers of sugars. Specific small polymers are usually designated by the number of sugar units, such as monosaccharide, disaccharide, trisaccharide, etc. The term polysaccharide is usually used when there are more than about 30 sugar units in the polymer.

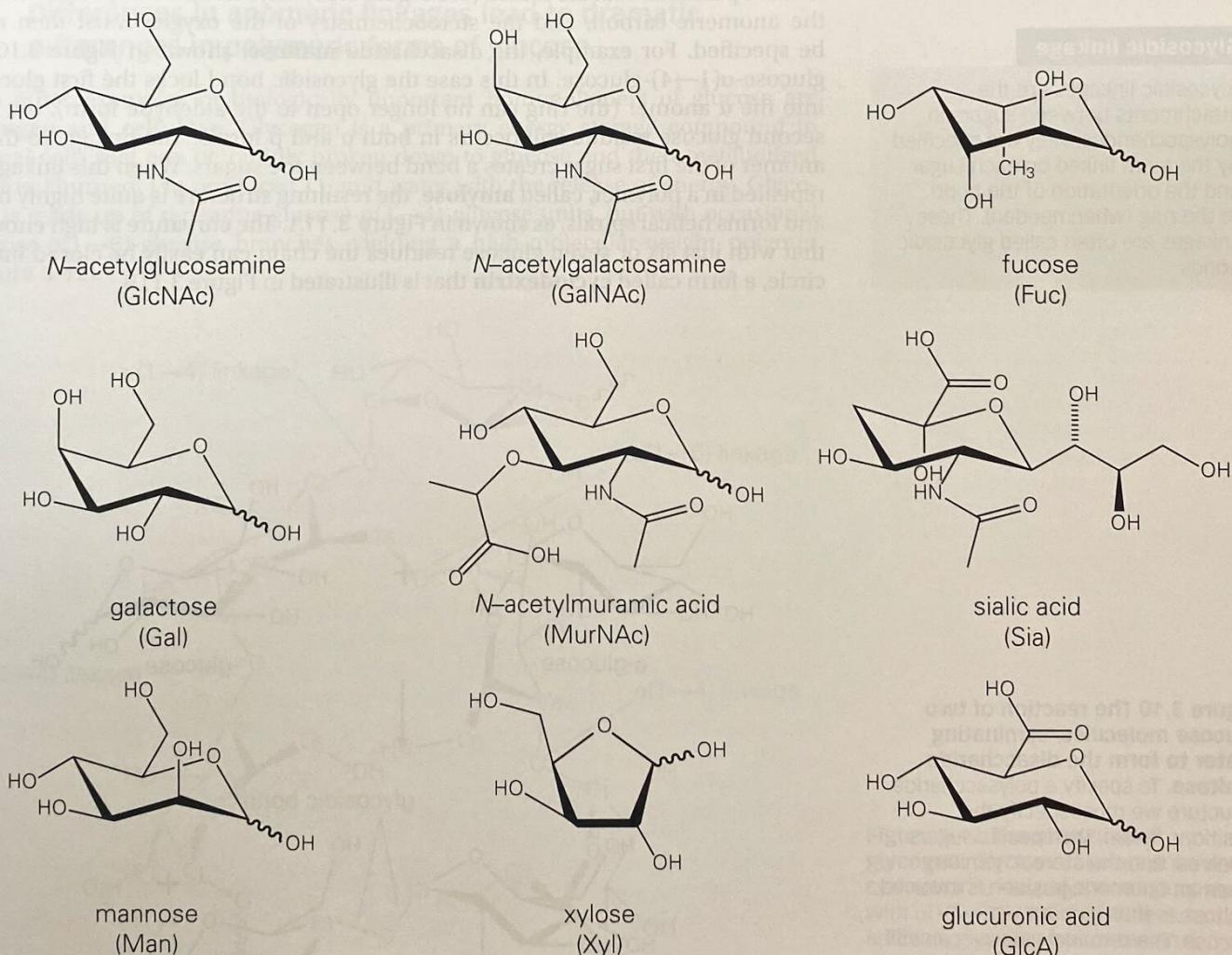


Figure 3.9 Nine sugars that occur commonly in oligosaccharides attached to proteins. The names of these sugars are shown together with their abbreviations. The squiggly line to oxygen indicates an anomeric position and that the oxygen can be either axial or equatorial at that position.

Table 3.1 Abbreviations used for the names of the sugars that commonly occur in oligosaccharides.

Glucose	Glc
Mannose	Man
Galactose	Gal
Fucose	Fuc
Xylose	Xyl
Sialic acid	Sia
N-Acetylglucosamine	GlcNAc
N-Acetylgalactosamine	GalNAc
N-Acetylmuramic acid	MurNAC
Glucuronic acid	GlcA

Glycosidic linkage

Glycosidic linkages are the attachments between sugars in polysaccharides. They are specified by the atom linked on each sugar and the orientation of the bond to the ring (when needed). These linkages are often called glycosidic bonds.

3.6 Glycans form polymeric structures that can have branched linkages

branched linkages

The bonds linking sugar residues into polymers are made by the elimination of a water molecule, resulting in the formation of ether bonds between the sugar residues, often referred to as the **glycosidic bonds** or **glycosidic linkages** (Figure 3.10). Although water can react with these bonds to break them, once formed they generally persist for long times in neutral solutions. As is common for most biological reactions, both the joining and the breaking of sugars is done by enzymes, which greatly accelerate these processes and allow the amounts of polymeric forms present to be regulated (see Chapter 16 for more about enzymes and regulation).

In proteins and nucleic acids, essentially all the bonds between residues involve identical chemical groups from linkage to linkage. For example, peptide bonds in genetically encoded proteins are made between backbone amines and backbone carboxylates, never through sidechains. For polysaccharides, however, any of the hydroxyls around a sugar ring may be used to link monomer subunits, giving oligosaccharides a variable covalent "skeleton" (in this case the term "backbone" no longer suffices). There may be more than one linkage to a particular sugar, leading to branching of the structure.

The linkages in oligosaccharides are specified by indicating the sugars involved and the positions that are linked (see Figure 3.10). Glycosidic bonds often involve the anomeric carbon, and the stereochemistry of the oxygen must then also be specified. For example, the disaccharide **maltose**, shown in Figure 3.10, is glucose- $\alpha(1\rightarrow 4)$ -glucose. In this case the glycosidic bond locks the first glucose into the α anomer (the ring can no longer open to the aldehyde form), but the second glucose residue still occurs in both α and β forms. The linkage to the α anomer of the first sugar creates a bend between the sugars. When this linkage is repeated in a polymer, called **amylose**, the resulting structure is quite highly bent and forms helical spirals, as shown in Figure 3.11A. The curvature is high enough that with just six or seven glucose residues the chain can easily be closed into a circle, a form called **cyclodextrin** that is illustrated in Figure 3.11B.

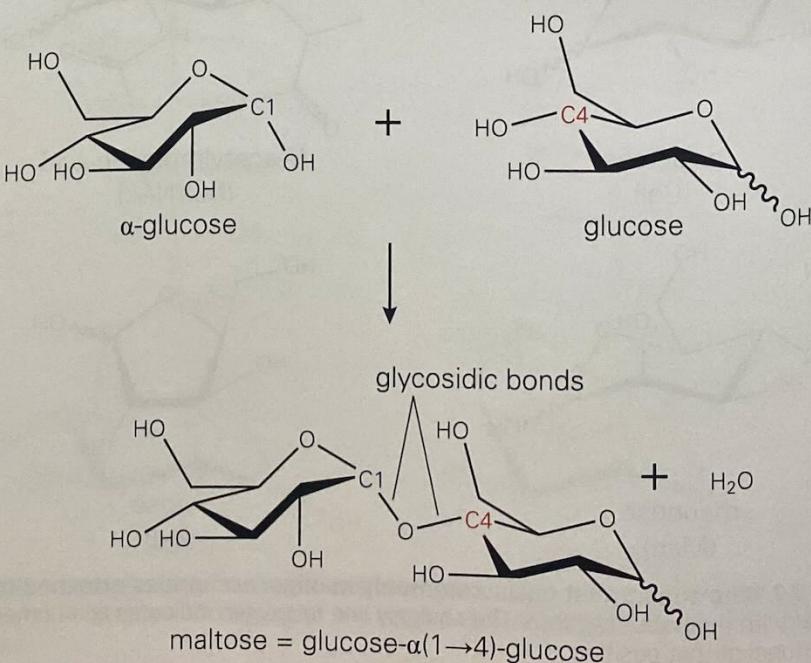
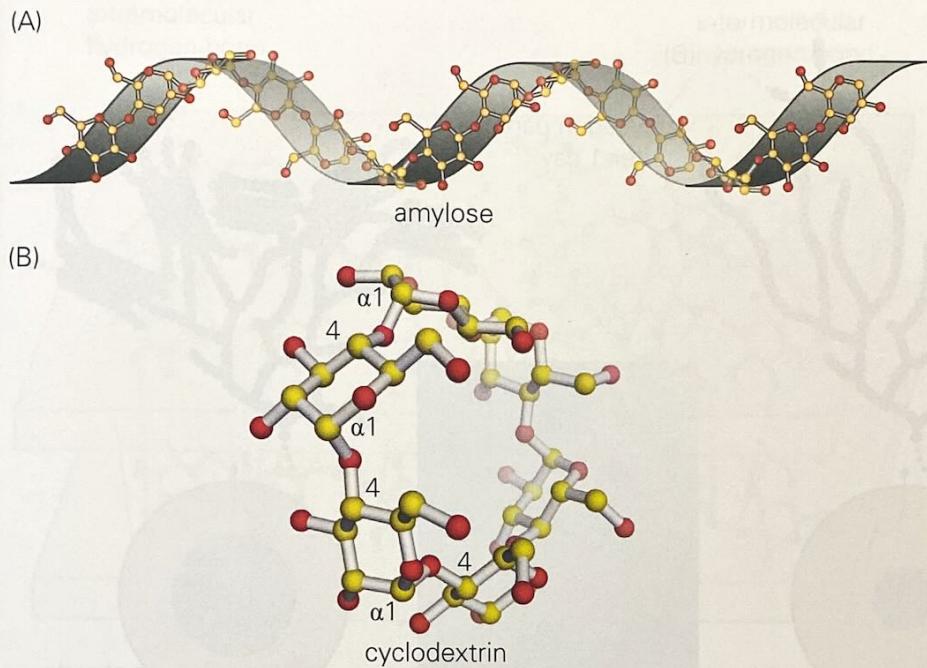


Figure 3.10 The reaction of two glucose molecules, eliminating water to form the disaccharide maltose. To specify a polysaccharide structure we must specify the positions linked, the specific sugars involved, and the stereochemistry when an anomeric position is involved. Maltose is thus glucose- $\alpha(1 \rightarrow 4)$ -glucose. The terminal glucose can still have both α and β anomers (giving α - and β -maltose), while linked anomeric positions are locked into the anomeric form present when they are linked.



3.7 Differences in anomeric linkages lead to dramatic differences in polymeric forms of glucose

Two other examples of biologically important homopolymers of glucose are **glycogen** and **cellulose**. Glycogen is a primary energy storage compound in animal cells that can be rapidly broken down to glucose and then metabolized (that is, “burned”) to produce CO_2 and water with the release of energy. Glycogen is made up of repeating glucose- $\alpha(1\rightarrow 4)$ -glucose units, but with occasional glucose- $\alpha(1\rightarrow 6)$ -glucose branches, yielding a high-molecular-weight polymer (Figure 3.12).

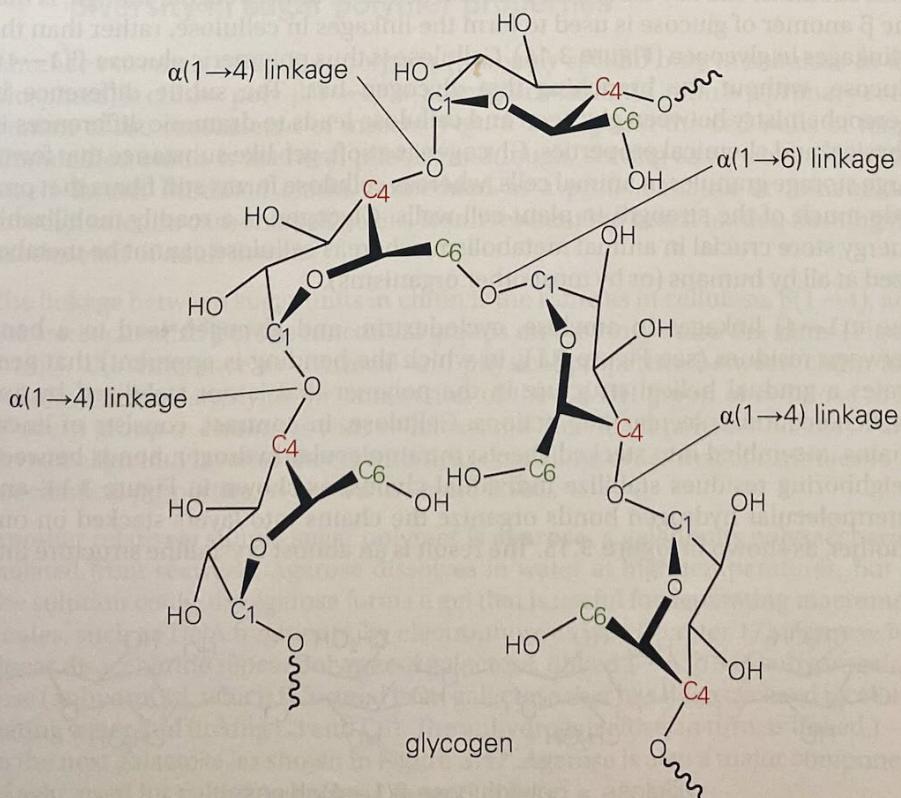


Figure 3.12 The structure of glycogen. Glycogen, like amylose and cyclodextrin, is a polymer of glucose with $\alpha(1\rightarrow 4)$ glycosidic linkages. In addition, glycogen has occasional $\alpha(1\rightarrow 6)$ linkages, as occurs in the central sugar in this structure. The C1, C4, and C6 carbons are colored blue, red, and green, respectively.

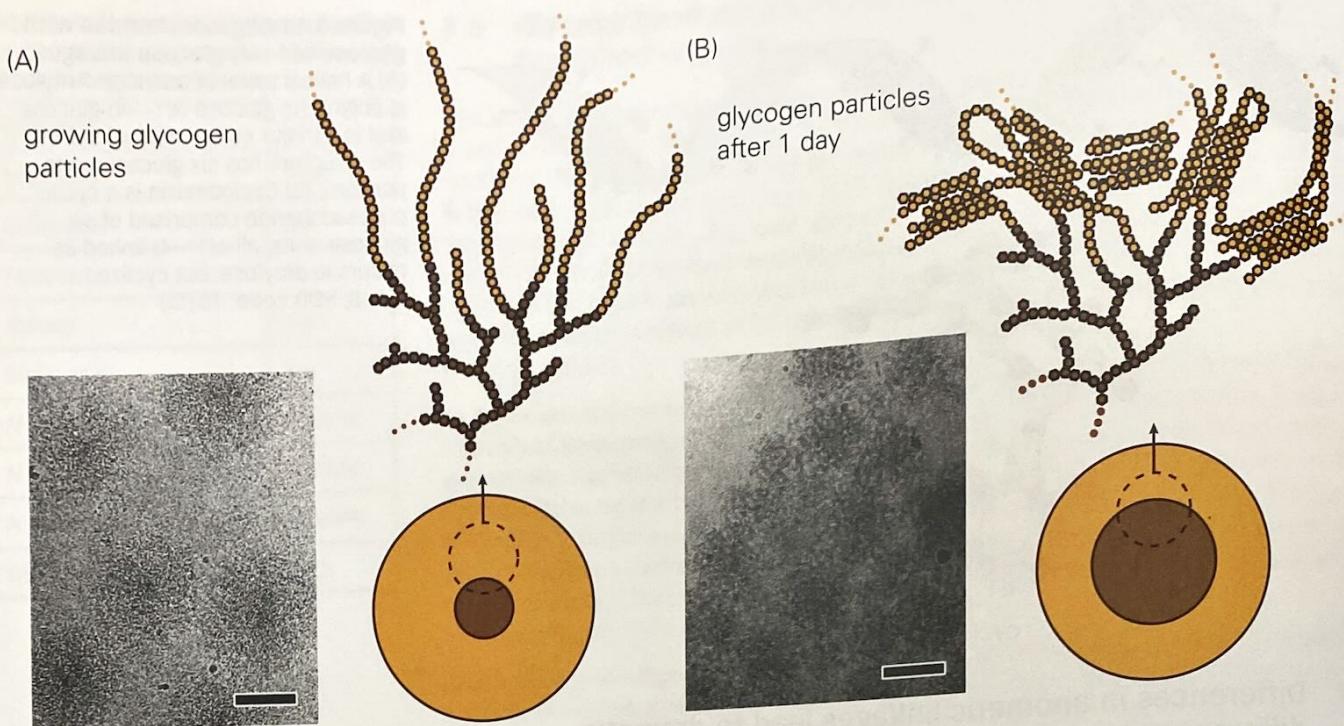


Figure 3.13 Glycogen particles.
These electron microscope images show glycogen particles growing due to the action of an enzyme.
(A) Glycogen particles are visible as dark grey circular objects in the electron micrograph on the left. The scale bar indicates 500 Å. Drawings of the sugar chains are shown, with each dot corresponding to one glucose unit. The inner circle of the schematic (brown) shows the initial branched glycogen core, while the outer region (orange) contains longer unbranched chains that have been added more recently. (B) The same glycogen particles, one day later. The chains are more highly branched and are denser.
(From A. Buleon, G. Veronese and J. Putaux, *Aust. J. Chem.* 60: 706–718, 2007. With permission from CSIRO.)

The primary linkages in glycogen are the same as those in amylose and, like amylose, glycogen also forms spiral structures. This spiral structure leaves space for additional glucose residues to be connected to the ones in the spiral, using the 1→6 linkages. These glucose residues are part of separate spirals, generating a highly branched structure. About one glucose residue in 12 has a branch, and glycogen polymers typically have about 50,000 glucose units. Glycogen chains are extended by the action of enzymes, as shown in Figure 3.13.

In plants, the cellulose, like glycogen, is a polymer made from 1→4 linked glucose subunits. The key difference between cellulose and glycogen, though, is that the β anomer of glucose is used to form the linkages in cellulose, rather than the α linkages in glycogen (Figure 3.14). Cellulose is thus polymeric glucose- $\beta(1\rightarrow 4)$ -glucose, without the branching that glycogen has. The subtle difference in stereochemistry between glycogen and cellulose leads to dramatic differences in physical and chemical properties. Glycogen is a soft, gel-like substance that forms large storage granules in animal cells, whereas cellulose forms stiff fibers that provide much of the strength in plant cell walls. Glycogen is a readily mobilizable energy store crucial in animal metabolism, whereas cellulose cannot be metabolized at all by humans (or by most other organisms).

The $\alpha(1\rightarrow 4)$ linkages in amylose, cyclodextrin, and glycogen lead to a bend between residues (see Figure 3.11, in which the bending is apparent) that generates a gradual helical structure in the polymer that is not stabilized by any particular intramolecular interactions. Cellulose, in contrast, consists of linear chains, assembled into stacked sheets; intramolecular hydrogen bonds between neighboring residues stabilize individual chains, as shown in Figure 3.14, and intermolecular hydrogen bonds organize the chains into layers stacked on one another, as shown in Figure 3.15. The result is an almost crystalline structure that

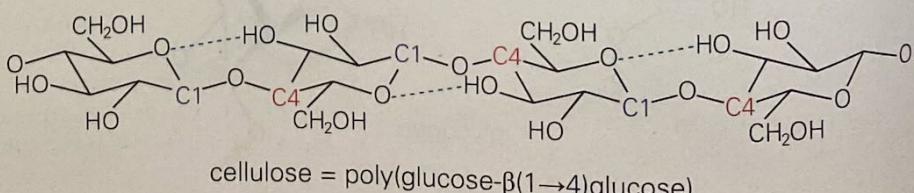


Figure 3.14 A chain of cellulose.
In regions of “crystalline” cellulose, multiple chains hydrogen bond to one another, making it mechanically stiff.

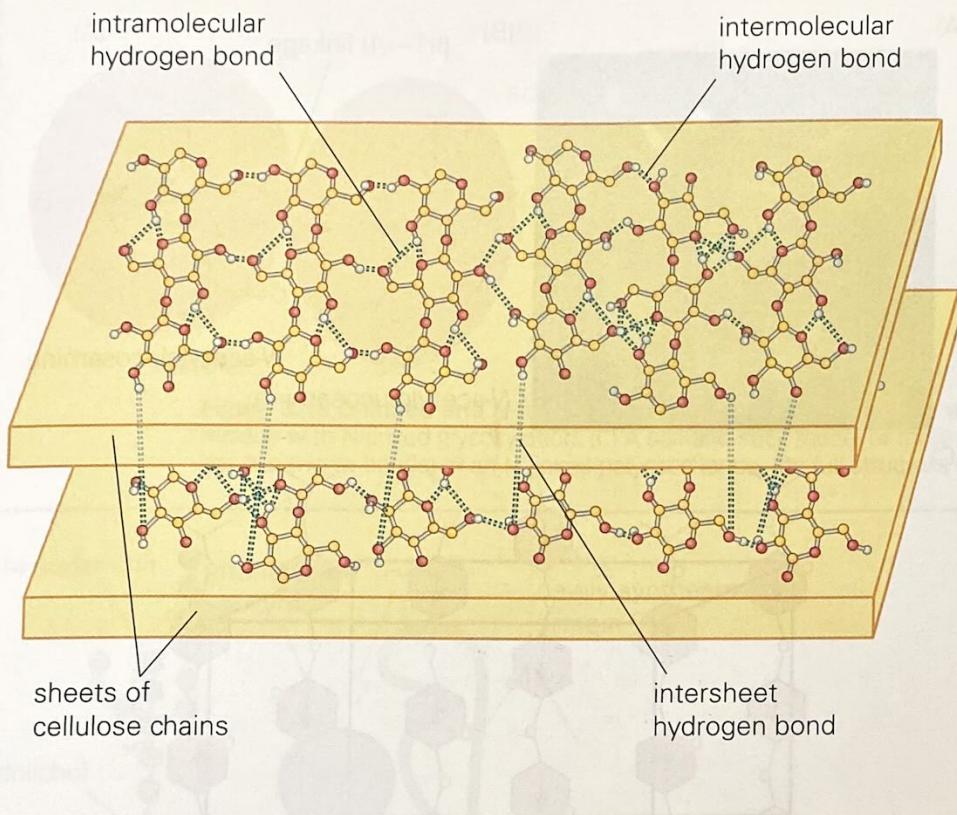


Figure 3.15 Layers in crystalline cellulose. Chains of cellulose molecules form sheets in which the chains run side by side and form hydrogen bonds with each other. These sheets are stacked upon one another (drawn artificially separated for clarity), and hydrogen bonds are also formed between molecules in different sheets.

is essentially insoluble in water. These interactions in cellulose are critical to its mechanical stability, and thus to its structural role in plants, where cellulose is responsible for much of the strength of the cell walls. Cellulose is the most abundant biopolymer in nature.

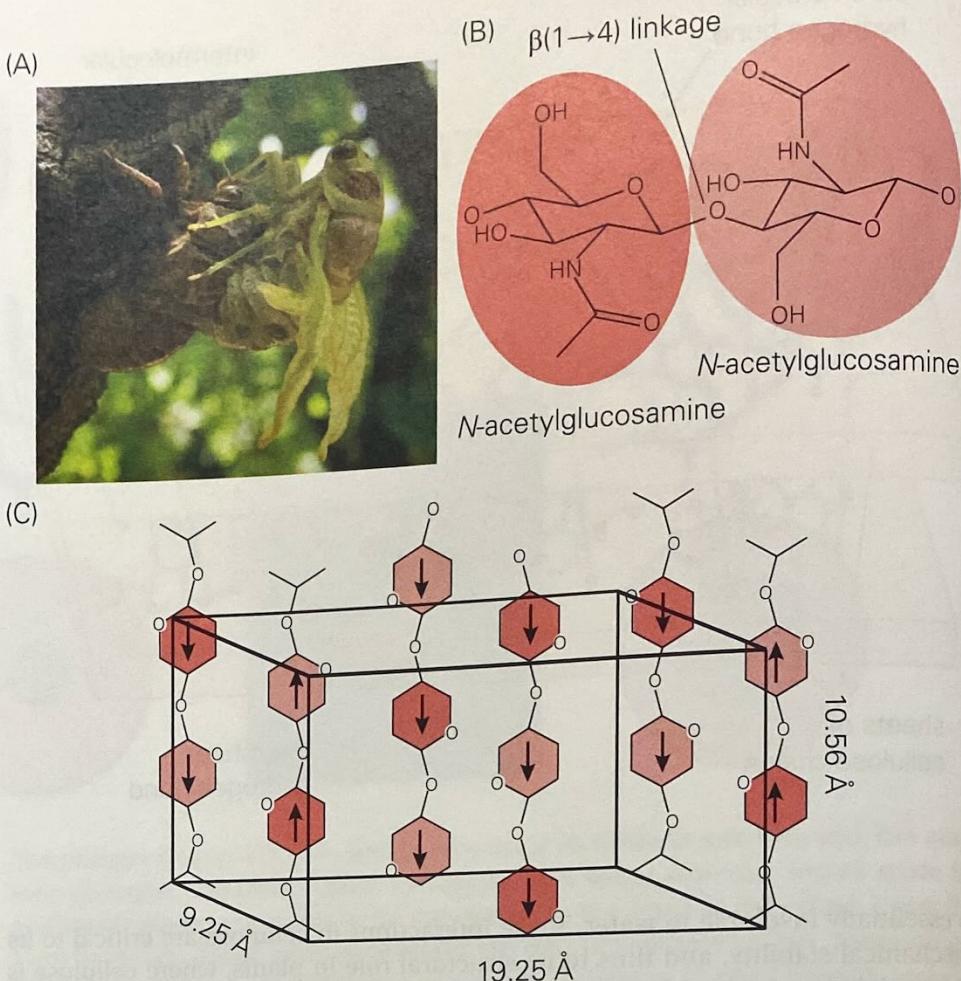
3.8 Acetylation or other chemical modification leads to diversity in sugar polymer properties

Another extremely abundant polymer (probably second only to cellulose in the biosphere) is **chitin**: poly- $\beta(1\rightarrow 4)$ -N-acetylglucosamine. Chitin is a primary component of the exoskeleton of insects (Figure 3.16A) and the cell walls of fungi. Unmodified chitin is leathery, pliable, and tough, serving as an outer “skin” for insect larvae. Modified chitin, embedded in a protein matrix or mineralized through calcium carbonate deposits, is still resilient but much harder, forming the exoskeleton of adult insects.

The linkage between sugar units in chitin is the same as in cellulose, $\beta(1\rightarrow 4)$, and the stereochemistry of the functional groups on the ring is also the same (Figure 3.16B). The difference in chemical and physical properties between chitin and cellulose is due largely to the conversion of the C2-OH group of cellulose to an N-acetyl group in chitin. As with cellulose, the individual polymer chains in chitin can align into layered sheets with near crystalline order that contributes to its mechanical properties, as shown in Figure 3.16C.

Another relatively simple sugar polymer is **agarose**, a gelatinous polysaccharide isolated from seaweed. Agarose dissolves in water at high temperatures, but as the solution cools, the agarose forms a gel that is useful for separating macromolecules, such as DNA fragments, by electrophoresis (see Chapter 17). Agarose is a linear disaccharide repeat polymer of galactose, linked 1 \rightarrow 4 to 3,6-anhydrogalactose (anhydroGal, which is formed from galactose that has been cyclized by eliminating water and linking C3 and C6). The anhydrogalactose, in turn, is linked 1 \rightarrow 3 to the next galactose, as shown in Figure 3.17. Agarose is also a major component of agar, used for culturing bacteria in petri dishes.

Figure 3.16 Chitin, a fundamental component of the exoskeletons of insects. (A) A chitin-rich exoskeleton is shed by a cicada. (B) Chitin consists of homopolymers of N-acetylglucosamine with $\beta(1 \rightarrow 4)$ linkages. (C) The basic organization of chains of polymers in chitin. The N-acetylglucosamine units are shown as hexagons, with the arrows pointing to the C1 atom. (A, courtesy of Jodelet/Lépinay, Wikipedia; C, from G.L. Clark and A.F. Smith, *J. Phys. Chem.* 40: 863–879, 1936. With permission from the American Chemical Society.)



3.9 Glycans may be attached to proteins or lipids

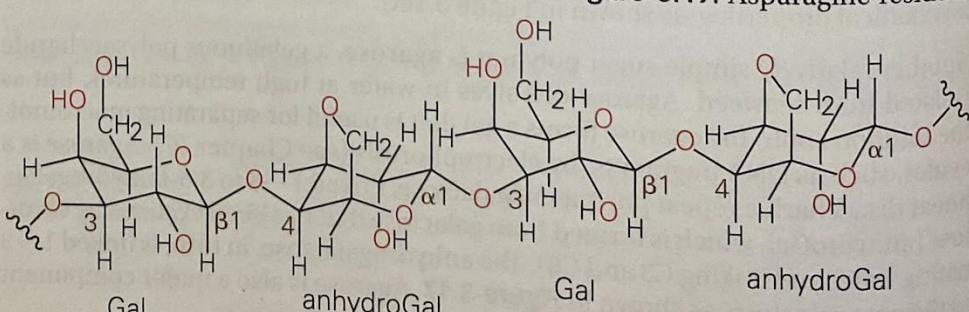
Many proteins in eukaryotic cells are “decorated” with covalently attached glycans. The process of attaching carbohydrate to a protein is called protein glycosylation, and it occurs in two principal forms. The glycan can be linked to the sidechain amide of an asparagine (Asn) residue, which is referred to as **N-linked glycosylation**. Alternatively, in **O-linked glycosylation**, the glycan is linked to the sidechain hydroxyl of a serine (Ser) or threonine (Thr) residue (Figure 3.18). Almost all proteins that are destined either to be secreted or to remain membrane bound are glycosylated, but only a small fraction of the Ser, Thr, and Asn residues in any particular protein are modified. Proteins destined to remain in the cytosol are rarely glycosylated, and then often with only a simple monosaccharide.

N-linked glycosylation accounts for about 90% of glycoprotein modifications, and is a co-translational process that starts in the endoplasmic reticulum with a nascent protein chain being synthesized and translocated across the rough endoplasmic reticulum membrane. As the protein emerges on the other side of the membrane, the enzyme oligosaccharyl transferase moves a 14-mer oligosaccharide *en bloc* from a lipid anchor (dolichol—see Figure 3.40) to the target Asn residue within the emerging protein, as shown in Figure 3.19. Asparagine residues

O-linked and N-linked glycosylation

O-linked and N-linked glycosylation refer to the attachment of polysaccharides through the sidechain hydroxyl of serine or threonine (O-linked), or the sidechain amide of asparagine (N-linked).

Figure 3.17 Agarose is composed of alternating galactose and anhydrogalactose units. Agarose is a polymer used to make gels that are useful in bioanalytical methods.



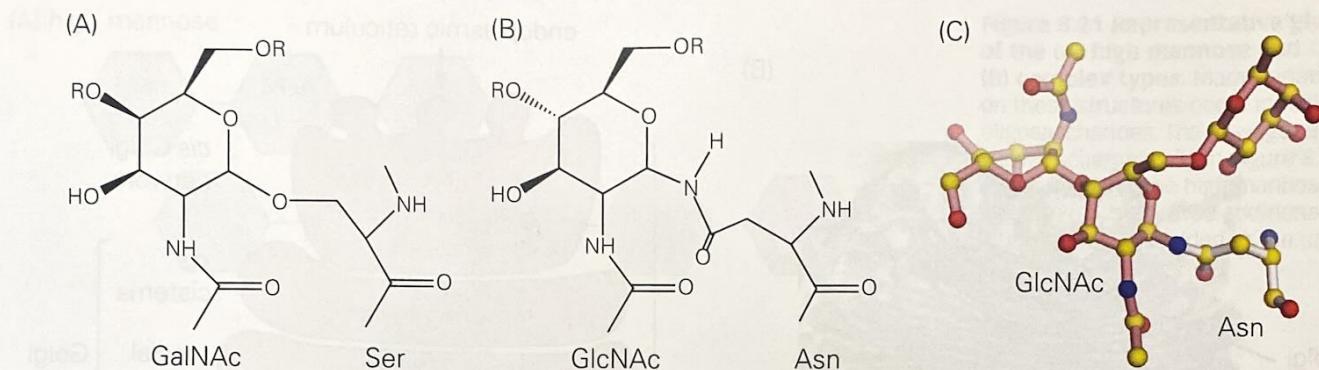


Figure 3.18 O-linked and N-linked glycosylation. (A) A Ser residue with O-linked glycosylation. (B) An Asn residue with N-linked glycosylation. (C) A ball-and-stick model of three sugar residues (colored bonds) closest to the Asn (white bonds) of an N-linked polysaccharide, the full structure of which is shown in Figure 3.22.

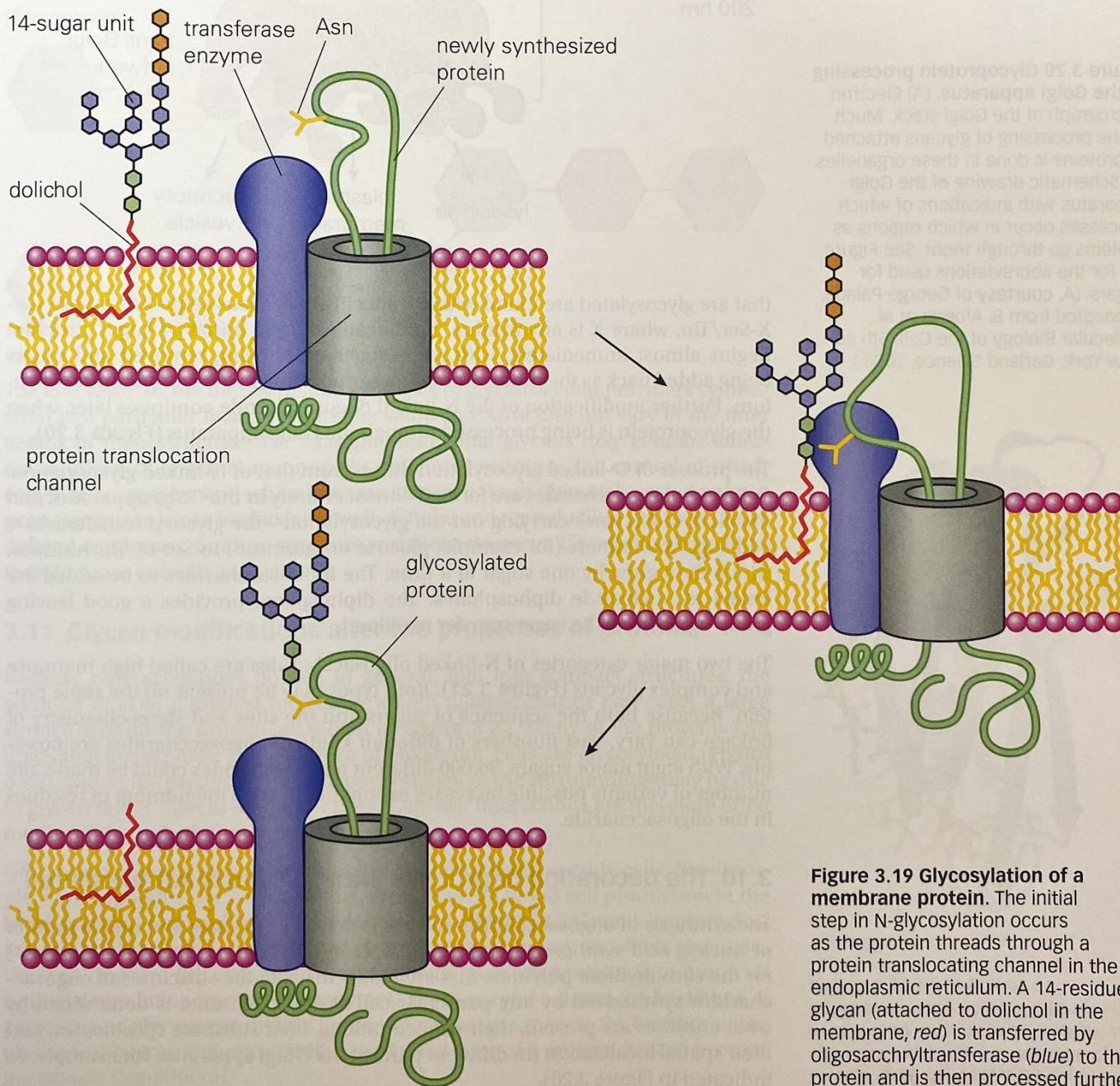


Figure 3.19 Glycosylation of a membrane protein. The initial step in N-glycosylation occurs as the protein threads through a protein translocating channel in the endoplasmic reticulum. A 14-residue glycan (attached to dolichol in the membrane, red) is transferred by oligosaccharyltransferase (blue) to the protein and is then processed further.

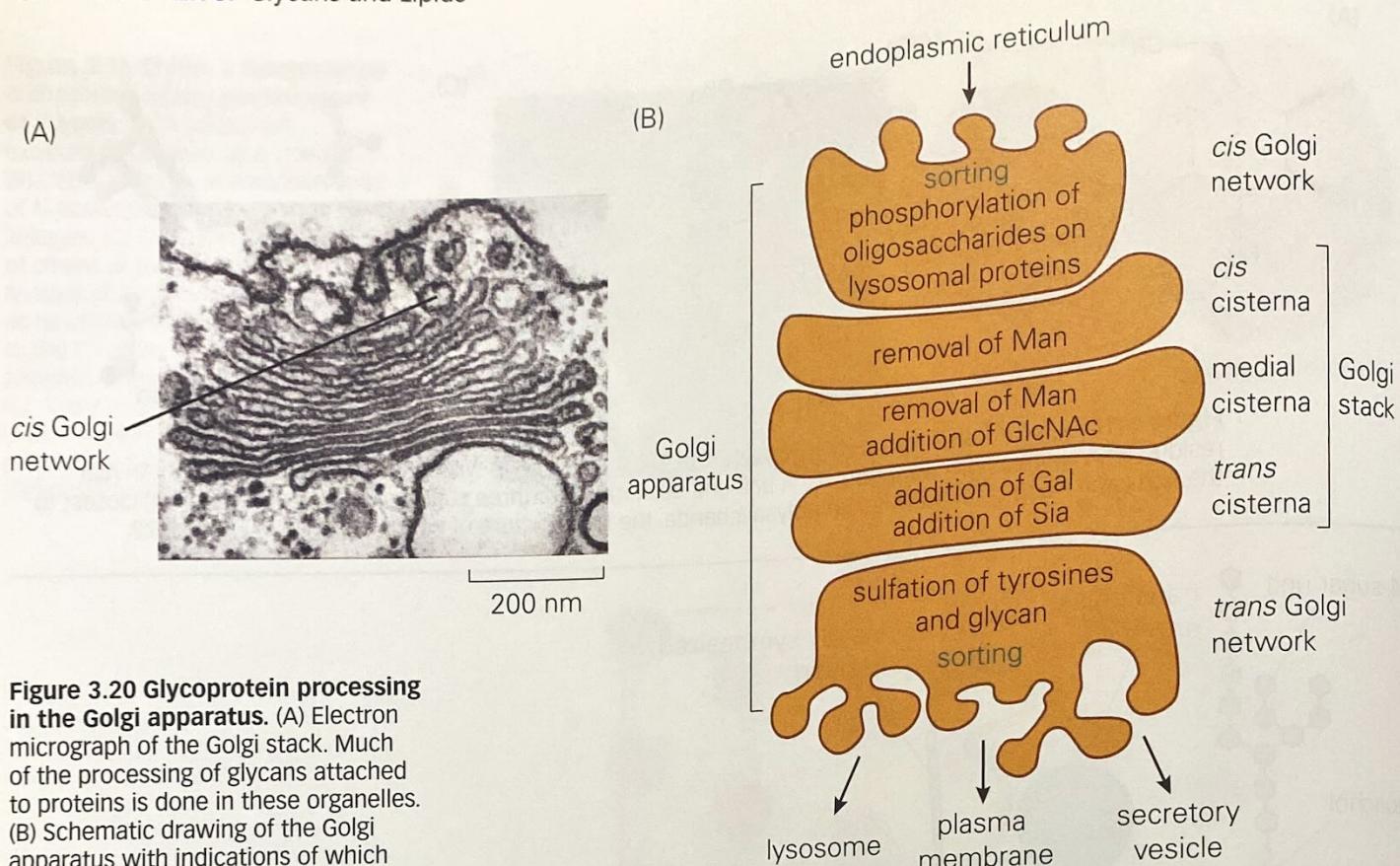


Figure 3.20 Glycoprotein processing in the Golgi apparatus. (A) Electron micrograph of the Golgi stack. Much of the processing of glycans attached to proteins is done in these organelles. (B) Schematic drawing of the Golgi apparatus with indications of which processes occur in which regions as proteins go through them. See Figure 3.9 for the abbreviations used for sugars. (A, courtesy of George Palade; B, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008.)

that are glycosylated are found within a specific consensus sequence motif: Asn-X-Ser/Thr, where X is any residue. Modification of the initial 14-sugar structure begins almost immediately, with some sugar units being removed and others being added back as the protein undergoes processing in the endoplasmic reticulum. Further modification of the N-linked oligosaccharide continues later, when the glycoprotein is being processed through the Golgi apparatus (**Figure 3.20**).

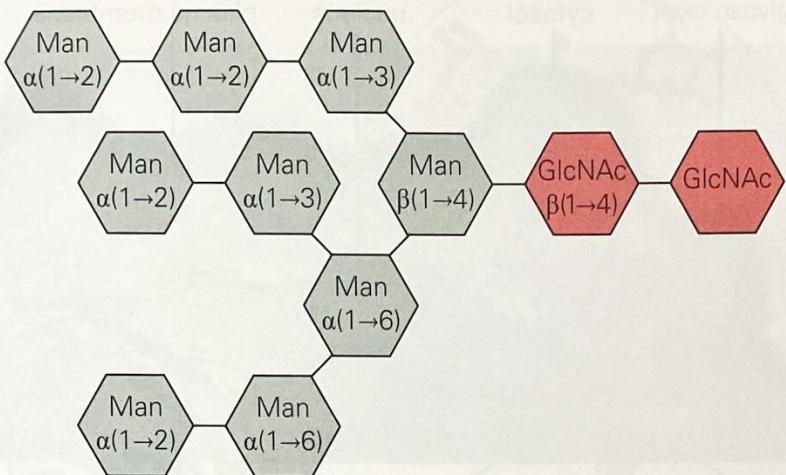
The process of O-linked glycosylation differs from that of N-linked glycosylation. O-linked oligosaccharides are formed almost entirely in the Golgi apparatus, and the class of enzymes carrying out the glycosylation—the glycosyl transferases—add sugar monomers (for example, glucose or mannose) to Ser or Thr residues, which then grow by one sugar at a time. The monosaccharides to be added are linked to nucleoside diphosphates. The diphosphate provides a good leaving group to facilitate the sugar transfer reaction.

The two major categories of N-linked oligosaccharides are called high mannose and complex glycans (**Figure 3.21**). Both types may be present on the same protein. Because both the sequence of sugars and the sites and stereochemistry of linkage can vary, vast numbers of different kinds of oligosaccharides are possible. With eight major sugars, 36,000 different tetrasaccharides could be made; the number of variants possible increases exponentially with the number of residues in the oligosaccharide.

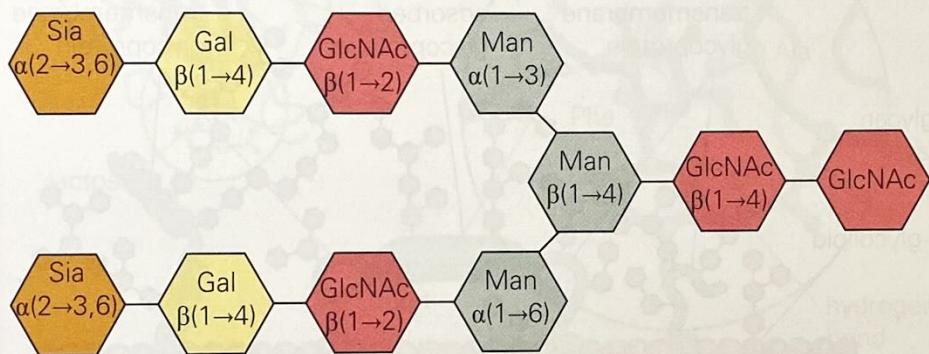
3.10 The decoration of proteins with glycans is not templated

The synthesis of oligosaccharides in cells is done by enzymes, but unlike protein or nucleic acid synthesis, there is no template (that is, no explicitly stored code) for the carbohydrate polymers that are made. Instead, the structures of oligosaccharides synthesized by any particular cell at any given time is determined by what enzymes are present, their concentrations, their substrate specificities, and their spatial localization (in different parts of the Golgi apparatus, for example, as indicated in Figure 3.20).

(A) high mannose



(B) complex



The end result of the nontemplated synthesis of glycans is that polymers synthesized in a particular location share the same core structure but vary in the finer details of their structure. For example, individual glycans may contain different numbers of terminal mannose or sialic acid residues. The control of which enzymes are present and where they are localized (both ultimately resulting from gene expression) does lead to fairly well-defined and reproducible products under defined conditions, but fine structure variations do occur. Glycosylated proteins, therefore, are quite heterogeneous.

3.11 Glycan modifications alter the properties of proteins

Because polysaccharides do not, in general, fold into compact structures, the addition of even a modest molecular weight oligosaccharide can cover a large surface area on the protein, as shown in **Figure 3.22**. Many of the proteins and lipids that are exposed at the surface of eukaryotic cells are glycosylated (**Figure 3.23**), and so the predominant surface presented to the surroundings may be due to glycan rather than to lipid or protein. Cell-cell interactions are therefore often mediated through glycan-protein interactions.

Glycans on proteins play important roles both inside and outside cells. For example, erythropoietin, the hormone that stimulates red blood cell production in the bone marrow, has multiple glycosylation sites (**Figure 3.24**), and its biological activity increases with the extent of glycosylation. One way in which glycosylation can improve the efficacy of a protein such as erythropoietin is to protect it against degradation by protease enzymes. These enzymes cleave the backbone of proteins, and glycans can make it harder for proteases to access the backbone. The presence of glycan can also stabilize the folded state of a protein, further protecting against degradation.

Figure 3.21 Representative glycans of the (A) high mannose and

(B) complex types. Many variations on these structures occur in natural oligosaccharides. The 14-sugar glycan shown schematically in Figure 3.19 corresponds to the high mannose sugar in (A), with three additional glucose residues added to the upper mannose branch.

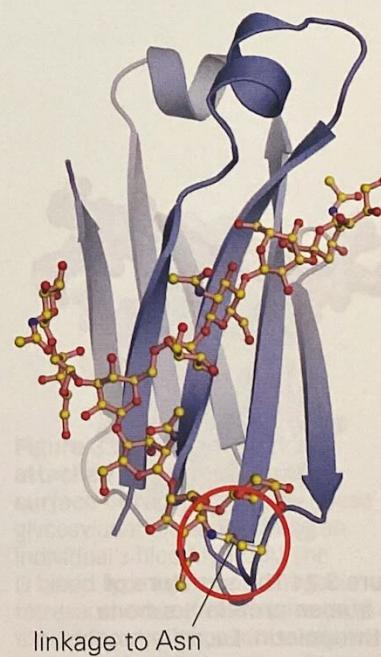
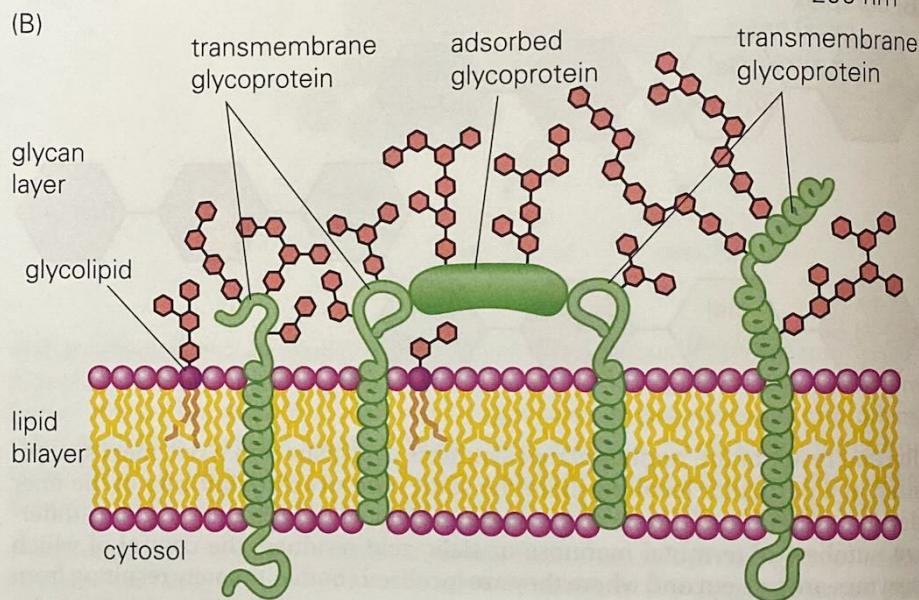
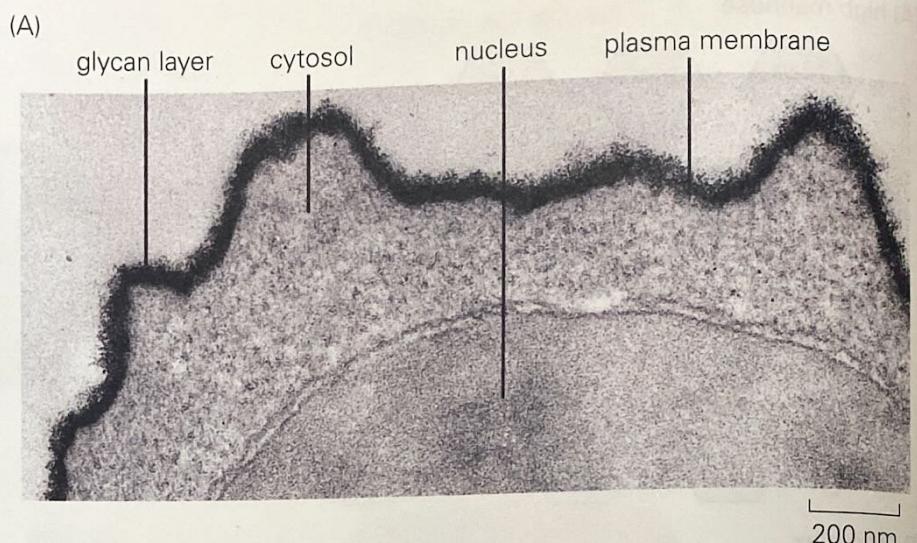


Figure 3.22 An N-linked oligosaccharide attached to an asparagine residue in a protein. The oligosaccharide covers one face of the protein. (PDB code: 1MCO.)

Figure 3.23 Cell-surface glycans.

(A) An electron micrograph of a fibroblast cell is shown with the outer glycan layer stained (dark layer). (B) Cell-surface glycolipids, transmembrane glycoproteins, and adsorbed glycoproteins all contribute to this thick glycan layer. (A, courtesy of Audrey M. Glauert and G.M.W. Cook; B, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. New York: Garland Science, 2008.)



Lectins

Lectins are proteins that recognize specific sugars, usually binding with moderate affinity and high specificity.

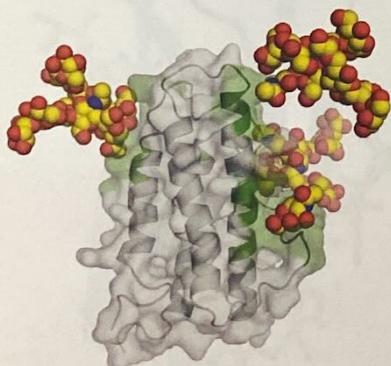


Figure 3.24 The structure of the human protein hormone erythropoietin. Each erythropoietin molecule has three N-linked complex carbohydrates and one O-linked GalNAc residue. Regions of the protein surface that interact with the sugars are colored green. (PDB code: 1BUY; glycans modeled by R.J. Woods.)

Glycosylation is also important for the trafficking of proteins inside cells. Within the endoplasmic reticulum, the attached sugars are indicators of the progress of protein folding, determining when a protein is allowed to progress from the endoplasmic reticulum to its next destination. For organelles in general (particularly those of secretory and endocytic pathways—namely, the endoplasmic reticulum, the Golgi apparatus, endosomes, lysosomes, and secretory vesicles), the oligosaccharides attached to proteins can act as “address labels,” determining where the proteins are transported to, including whether they are delivered to the cell surface (glycosylated proteins are enriched on cell surfaces). Glycans are also very important in the formation of cell walls, which is discussed in Section 3.23.

3.12 Protein–glycan interactions are important in cellular recognition

Cell surface glycans can interact with proteins on other cells, establishing stable connections between cells. A large group of proteins, known collectively as **lectins**, recognize specific sugars in glycans (Figure 3.25). In general, lectins bind only a small portion of a surface carbohydrate, but do so with high specificity. There are many glycans on each cell, and many copies of proteins that bind them, so these multivalent interactions lead to both tight and specific cell-cell associations.

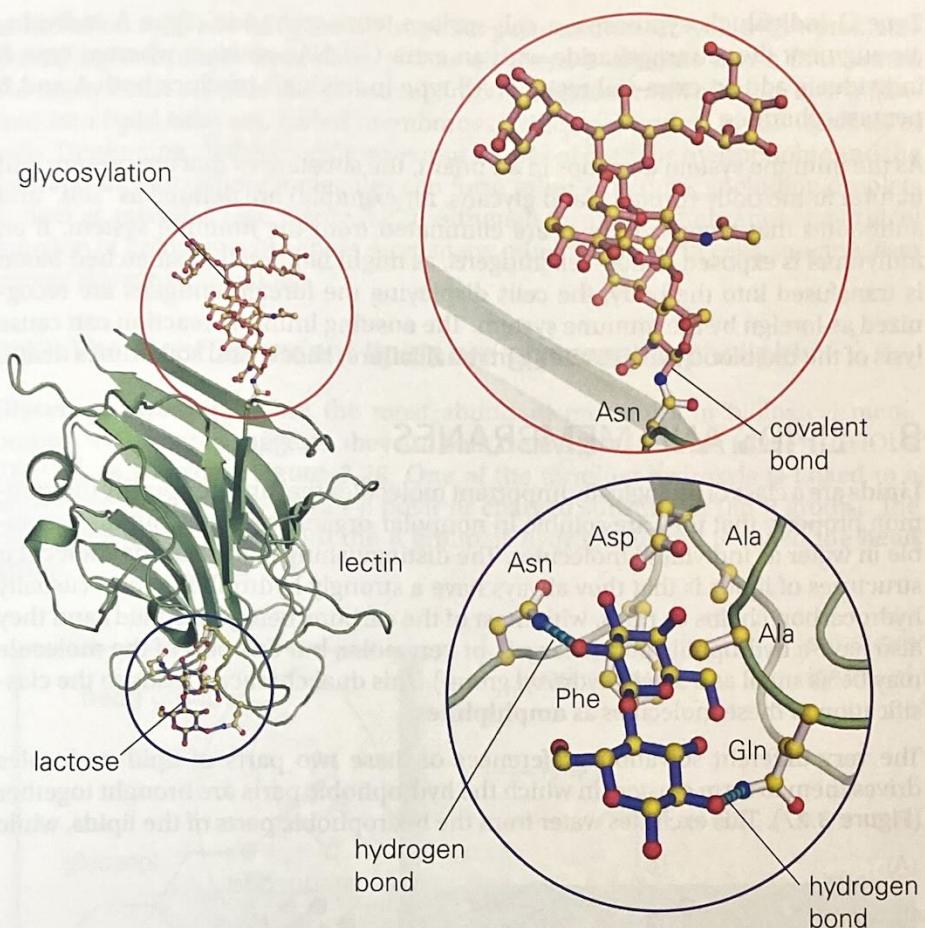


Figure 3.25 A lectin isolated from a legume. The lectin is shown in complex with lactose [the disaccharide β -galactose-(1 \rightarrow 4)- β -glucose, blue circles]. This lectin recognizes primarily the galactose end of the lactose, making both hydrogen bonds and hydrophobic contacts. The lectin itself is glycosylated (see the sugars at the top of the protein, red circles). (PDB code: 1LTE.)

Cell-surface glycoproteins can also be recognized by antibodies in the immune system. For example, glycans present on red blood cells define the “blood groups” A, B, AB, and O (Figure 3.26). Subtle variations in a polysaccharide on red cells arise from genetic differences in the enzymes that control the glycosylation.

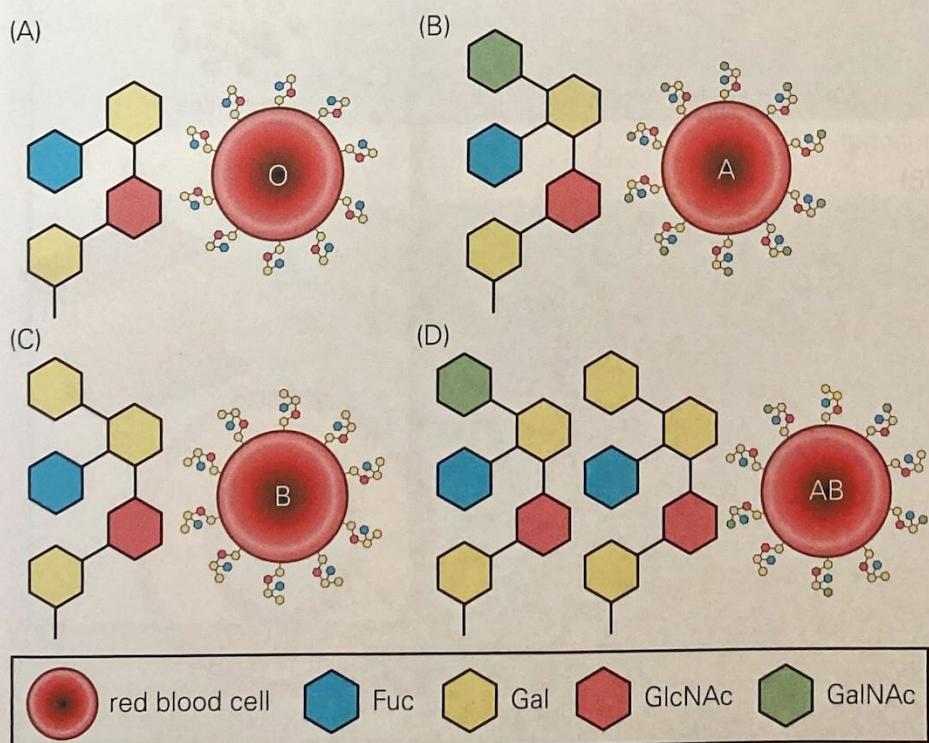


Figure 3.26 Glycans that are attached to proteins on the surface of red blood cells. These glycosylation patterns define an individual's blood type. (A) The O blood type has a non-antigenic tetrasaccharide that is common to the surface antigens of the other blood groups. A-type (B) and B-type (C) individuals produce oligosaccharides with an additional Gal or GalNAc residue, respectively. (D) AB-type individuals produce a mixture of the A- and B-type oligosaccharides.