## Carbohydrate Chemistry and Nomenclature

The term "carbohydrate" is derived from the $C_n(H_2O)_n$ general chemical formula exhibited by most of these molecules. Carbohydrates are the most common organic molecules, and are far more common than any other type of biological compound.

Carbohydrates have a variety of roles. Carbohydrates are important components of extracellular molecules, including secreted soluble proteins, cell-surface proteins, and various types of extracellular matrix molecules. In the form of cellulose, carbohydrates act as one of the main structural components in plants. Carbohydrates are also used in bacterial cell walls, and form the major components of arthropod exoskeletons.

Carbohydrates are used as energy sources in essentially all organisms. Glucose polymers, such as starch in plants and glycogen in animals and bacteria, are critical energy storage molecules. In addition, glucose is a source of biosynthetic intermediates for nearly all anabolic pathways.

## Nomenclature and structure

Most carbohydrate nomenclature is based on historical trivial names. However a few general rules are commonly used.

The word **saccharide** (from the Sanskrit word *śarkarā* via the Greek word *sakcharon (σακχαρον)*, meaning "sugar") is also used to refer to carbohydrates. It is primarily used to refer to the size of the molecule. "Monosaccharide" refers to the individual carbohydrates; "oligosaccharide" and "polysaccharide" refer to carbohydrate polymers of varying sizes.

The suffix "**-ose**" designates a carbohydrate. The number of carbon atoms is designated by a prefix for some forms of the compound names (*e.g.*, triose and hexose).

Some carbohydrates contain ketones; these are termed **ketoses**, and are sometimes (but not always) identified by the "**-ulose**" suffix. Most other carbohydrates are termed **aldoses**, because they contain aldehydes.

The simplest carbohydrate, dihydroxyacetone, is shown below. Dihydroxyacetone is a ketotriose, with the molecular formula $C_3(H_2O)_3$.

$$
\begin{array}{l}
HO-CH_2 \\
\quad\ \ | \\
\quad\ \ C=O \qquad \text{Dihydroxyacetone} \\
\quad\ \ | \\
HO-CH_2
\end{array}
$$
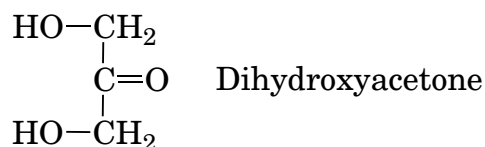
With the exception of dihydroxyacetone (which is a symmetrical molecule), all carbohydrates are **chiral** molecules. In general, aldoses have $(n-2)$ chiral centers, while ketoses have $(n-3)$ chiral centers. Glyceraldehyde, the simplest aldotriose, has the same $C_3(H_2O)_3$ formula as dihydroxyacetone, although it is both

structurally and chemically different. The possible structures for glyceraldehyde are shown below.

$$
\begin{array}{cc}
\text{D-Glyceraldehyde} & 
\begin{array}{c}
\text{O} \\
\parallel \\
\text{H}-\text{C} \\
\textbf{H} \blacktriangleright \bar{\text{C}} \blacktriangleleft \textbf{OH} \\
\text{HO}-\text{CH}_2
\end{array}
\qquad
\begin{array}{c}
\text{O} \\
\parallel \\
\text{H}-\text{C} \\
\textbf{HO} \blacktriangleright \bar{\text{C}} \blacktriangleleft \textbf{H} \\
\text{HO}-\text{CH}_2
\end{array}
& \text{L-Glyceraldehyde}
\end{array}
$$

Note that glyceraldehyde has two possible configurations at the second carbon. In the modified **Fischer projection** above, the hydroxyl at the carbon points either to the right to yield the D configuration, or to the left to yield the L configuration. Most, although not all, biological carbohydrates have D stereochemistry.

For carbohydrates, the chiral carbon *furthest* from the carbonyl is used to determine the D or L nomenclature.[19] The structure below is a D-aldohexose (recall that "hex-" means "six"), because it has the D configuration at the fifth carbon.
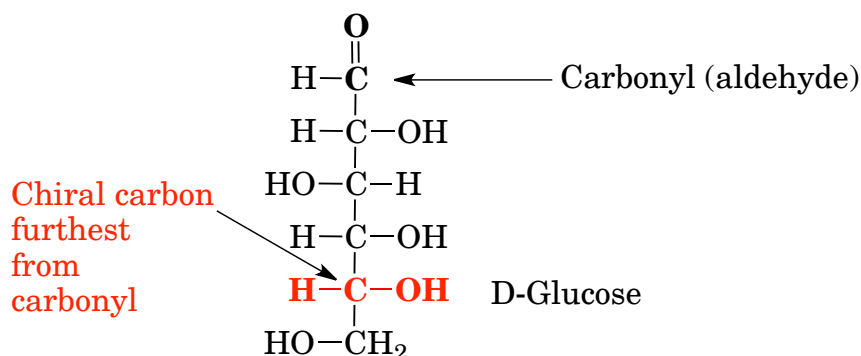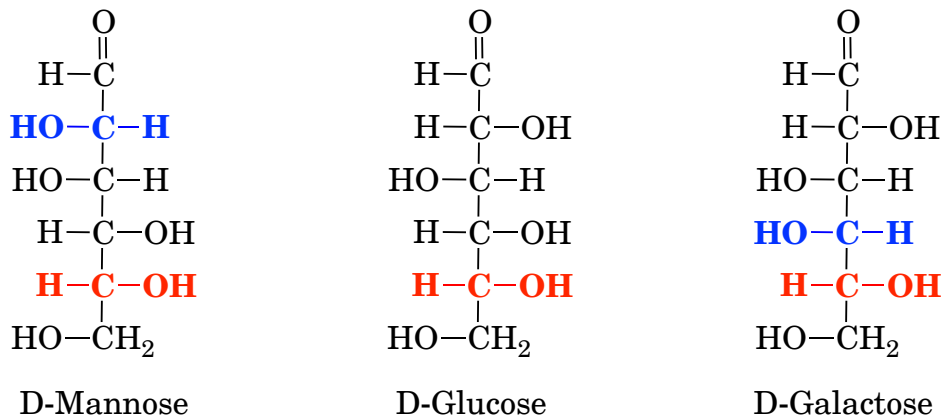


The structure above has a total of four stereocenters. The configuration at the other three chiral carbons defines the specific identity of the compound; in this case, the compound is D-glucose (sometimes called "dextrose", especially in older literature).[20]

---

[19] Carbohydrate nomenclature is, however, somewhat more complex than the simple statement that the chiral carbon furthest from the carbonyl determines D or L. For example, the enantiomer of D-glucose is defined as L-glucose, the carbohydrate that has all 4 stereocenters inverted; it is not the compound that merely has inversion of the 5-carbon stereocenter.
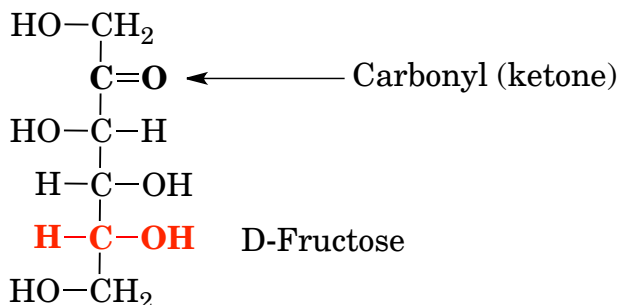
[20] **Side note: optical activity**
Most chiral molecules rotate plane-polarized light. In the older literature, a chiral molecule would have a prefix of a lower case "d" or "l", indicating the direction that the molecule rotated the polarized light. This should not be confused with "D" and "L", which refer only to absolute configuration, and not to the rotation of polarized light. To avoid this possible confusion, current literature uses "+" and "–" to give the direction of polarized light rotation by the compound. Some compounds with "D" configuration, such as D-glyceraldehyde and D-glucose, rotate light in the (+) direction, and others, such as D-fructose, rotate light in the (–) direction (in older literature, fructose was called "levulose" to distinguish it from "dextrose"). In addition, because the different anomeric forms of carbohydrates (see below) are different chiral compounds, different anomers frequently have different optical activities.

It is obviously possible to have other configurations at these other chiral carbons. Mannose and galactose (shown below) each differ from glucose by the configuration at a single carbon. These carbohydrates are **epimers** of glucose (mannose is the C-2 epimer and galactose is the C-4 epimer of glucose). Mannose and galactose differ at two positions and are not epimers of one another.

$$
\begin{array}{ccc}
\overset{\displaystyle O}{\underset{\displaystyle H-C}{\|}} & \overset{\displaystyle O}{\underset{\displaystyle H-C}{\|}} & \overset{\displaystyle O}{\underset{\displaystyle H-C}{\|}} \\
\mathbf{HO-C-H} & H-C-OH & H-C-OH \\
HO-C-H & HO-C-H & HO-C-H \\
H-C-OH & H-C-OH & \mathbf{HO-C-H} \\
\mathbf{H-C-OH} & \mathbf{H-C-OH} & \mathbf{H-C-OH} \\
HO-CH_2 & HO-CH_2 & HO-CH_2 \\
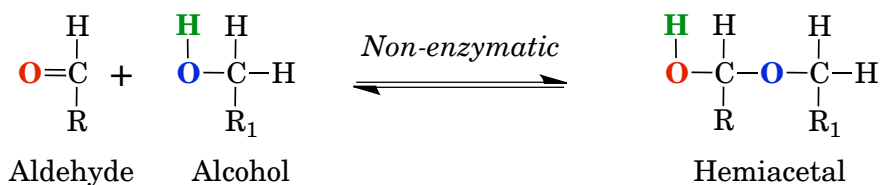\text{D-Mannose} & \text{D-Glucose} & \text{D-Galactose}
\end{array}
$$

Ketoses use similar nomenclature. The structure below is the most important biological ketohexose, and one of the most important ketoses: D-fructose.

$$
\begin{array}{l}
HO-CH_2 \\
\mathbf{C=O} \longleftarrow \text{Carbonyl (ketone)} \\
HO-C-H \\
H-C-OH \\
\mathbf{H-C-OH} \quad \text{D-Fructose} \\
HO-CH_2
\end{array}
$$

Note that fructose has a similar configuration to glucose at each of its chiral carbons, but has a C-2 ketone, instead of the C-1 aldehyde. Glucose and fructose (and mannose and galactose) are isomers, because they all have the same chemical formula.

The linear structures discussed above do exist, but are minor forms in aqueous solution. Most biological carbohydrates exist as cyclic structures. This is especially true for the pentoses and hexoses, because these compounds can form stable 5-atom and 6-atom ring structures. The ring structures are possible because of the reaction shown below; an alcohol and a carbonyl carbon can reversibly form a covalent bond. If the carbonyl is an aldehyde, the product is a hemiacetal (as shown); if the carbonyl is a ketone, the product is a hemiketal.

$$
\underset{\text{Aldehyde}}{\overset{\displaystyle H}{\underset{\displaystyle R}{O=C}}} + \underset{\text{Alcohol}}{\overset{\displaystyle H \quad H}{\underset{\displaystyle R_1}{O-C-H}}} \xrightarrow[\quad]{\textit{Non-enzymatic}} \underset{\text{Hemiacetal}}{\overset{\displaystyle H \quad H \qquad H}{\underset{\displaystyle R \quad R_1}{O-C-O-C-H}}}
$$

67

Glucose, like most pentoses and hexoses, can form several possible **anomers**, depending on the side of the planar carbonyl that experiences the attack by the alcohol, and depending on which alcohol attacks the carbonyl.



α-D-Glucose                D-Glucose                β-D-Glucose
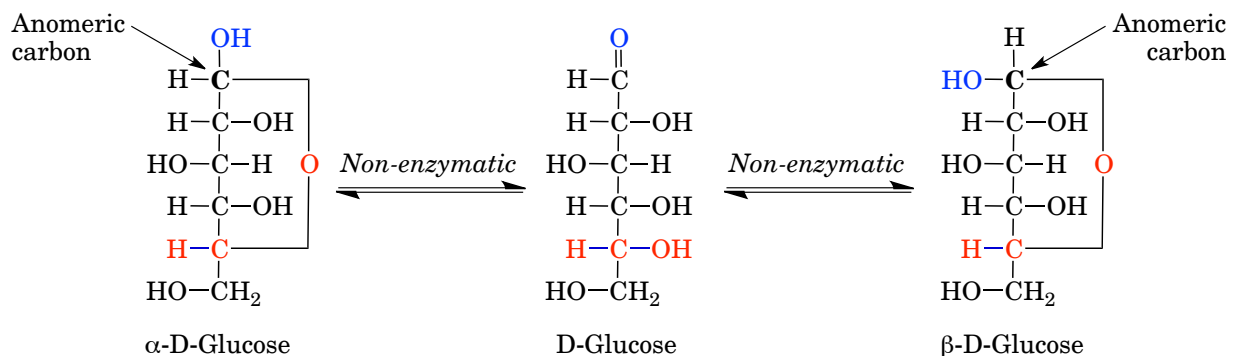
In general, for carbon compounds, five- and six-membered rings are the most stable forms. This means that, although four-membered, five-membered, six-membered, and seven-membered rings are possible, most carbohydrates tend to predominantly form five- and six-membered rings. However, because the ring formation is reversible, a solution of one of these a compound will exist as a distribution of different forms. For glucose in aqueous solution, roughly 66% is present as the six-membered ring β-D-glucose, 33% as the six-membered ring α-D-glucose, about 1% as the linear free aldehyde, with some traces of five-membered ring forms of the molecule.

The representations shown below for glucose are somewhat closer to the three-dimensional structures than those shown above. Recall that the groups attached to the ring carbons can be in the axial position or the equatorial position. The equatorial position is less sterically hindered, and is therefore preferred for bulkier groups. Note that the β-D-glucose structure has all of its bulky ring-substituents in the equatorial position.



β-D-Glucose
(chair representation)

α-D-Glucose
(chair representation)
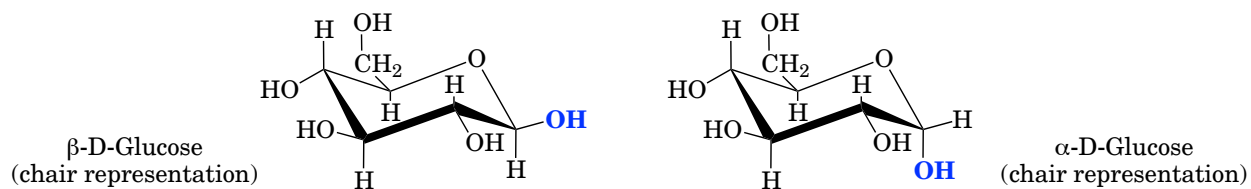
The chair-form structures are somewhat difficult for many students to draw correctly. As a result, the **Haworth projections** are frequently used to depict carbohydrate ring structures. The drawing below shows the Haworth projections for the two anomers of glucose.



β-D-Glucose
(Haworth representation)

α-D-Glucose
(Haworth representation)
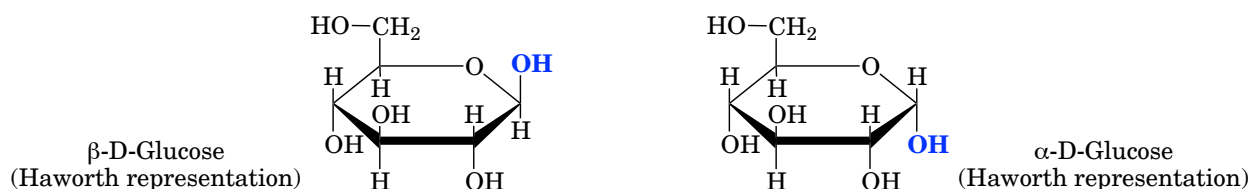
The nomenclature used for groups that protrude above and below rings defines groups above the ring as β and groups below the ring as α. This clearly only works if a standard ring orientation is used for the drawings. For carbohydrates, the anomeric carbon (the carbon that forms a carbonyl in the linear structure) is drawn on the right, and the ring oxygen further from the viewer than the rest of the molecule, and **the relative position of the hydroxyl formed from the carbonyl oxygen determines the α or β designation**. Note the position of the anomeric hydroxyl in the two forms of fructose shown below.



β-D-Fructose          α-D-Fructose
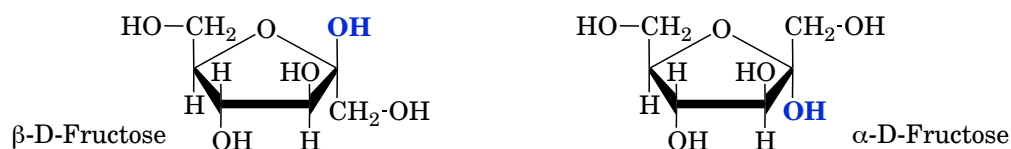
Comparison of the structural figures of glucose and fructose shown above reveals that some carbohydrates will form five- or six-membered rings. Six-membered carbohydrate rings (which are usually, but not always, more stable) are designated "pyranose" based on their similarity to pyran; five-membered rings are designated "furanose". The structures of fructose above can therefore be considered to be β- and α-D-fructofuranose.



Pyran          Furan

The same carbohydrate can form different ring structures, depending on the hydroxyl that performs the attack on the carbonyl. For glucose, the pyranose structure is the most stable and the most commonly used in enzymatic reactions. For fructose, although both furanose and pyranose structures exist in aqueous solution, the furanose structure is more common at 37°C and is the form used as the substrate for most of the relevant enzymes.[21]

### Reducing and non-reducing sugars
The cyclic structures of the carbohydrates do not contain free carbonyls. However, the anomeric carbon of these compounds is capable of dissociating from the hemiacetal or hemiketal structure to release the free carbonyl. Because free carbonyls are reactive compounds, these compounds are capable of forming non-enzymatic covalent bonds to free amino groups within proteins. Formation of these non-enzymatic adducts may result in deleterious alterations in protein structure.

---

[21] Fructose, in the six-membered ring form (fructopyranose) is about 70% sweeter than sucrose, while the five-membered ring form fructofuranose is comparable in sweetness to sucrose. The equilibrium between fructopyranose and fructofuranose is temperature dependent, with the fructopyranose being more the stable form at lower temperatures, while fructofuranose is the more stable form at room temperature. As a result, beverages sweetened with fructose are sweeter if consumed at lower temperatures (*i.e.* near 4°C). Because Europeans are less likely to use ice in their beverages, sodas sold in Europe frequently have the statement "best if served cold" printed on the packaging.

D-Glucose → Schiff base adduct (Non-enzymatic reactions)
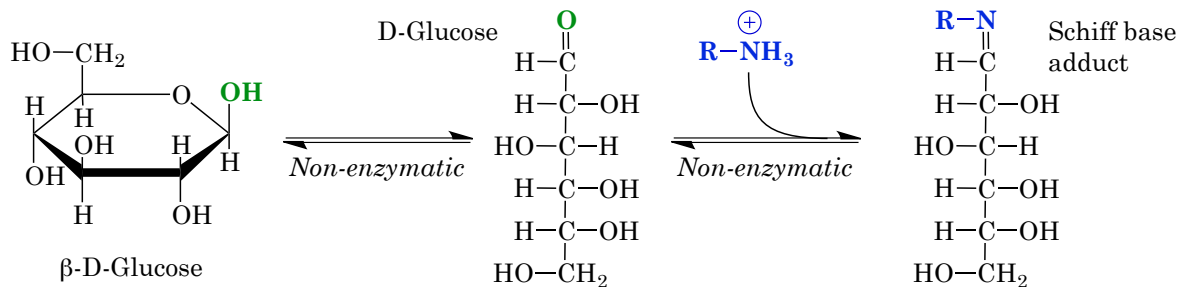
β-D-Glucose

While free glucose in circulation is observed to covalently modify proteins and other molecules (a process that is involved in some of the damage caused by elevated glucose concentrations in diabetes mellitus), some biological systems have evolved to prevent the release of a free carbonyl from carbohydrates during storage. In order for the free carbonyl to be formed, the anomeric carbon must have a free hydroxyl group. A covalent bond between the oxygen and any non-hydrogen atom prevents the formation of the free carbonyl. As an example, glucose-1-phosphate cannot form a free carbonyl because of the covalent bond to the phosphate. This has two consequences: 1) it prevents non-enzymatic reactions, and 2) it fixes the anomeric carbon in a single configuration (such as the α-configuration shown below).



α-D-Glucose-1-phosphate
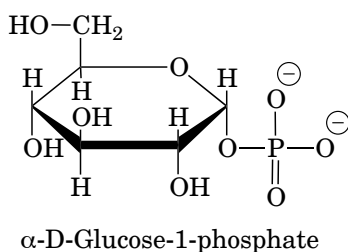
Another example of a carbohydrate locked in one configuration is provided by adenosine and the other nucleosides. All of these compounds are based on the aldopentose ribose. The attachment of the purine or pyrimidine base to the anomeric carbon, a process that occurs in a series of reactions beginning with α-D-ribose-5-phosphate, results in an anomeric carbon incapable of dissociating into the linear form.



α-D-Ribose



Adenosine Triphosphate (ATP)
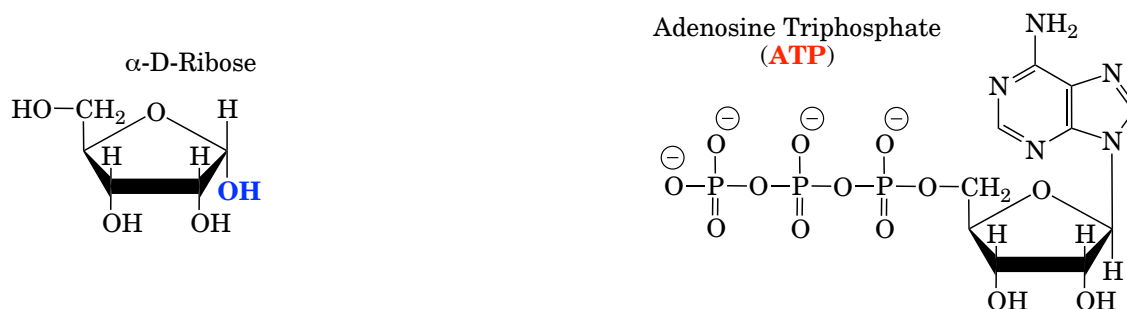
Until relatively recently, clinical tests for carbohydrates utilized chemical reactions that involved the free carbonyl. Most of these tests allow the carbohydrate to react with another compound; if a free carbonyl is present, the result is reduction of the

other compound along with oxidation of the carbohydrate.

Because these tests for free carbonyls resulted in reduction of the test reagent, carbohydrates that contain unmodified anomeric hydroxyl groups are referred to as **reducing sugars**. Carbohydrates that contain modified anomeric hydroxyls (such as glucose-1-phosphate), are therefore termed **non-reducing sugars**.

An example of this type of reaction is the Fehling's test, in which blue $Cu^{2+}$ ions are reduced to $Cu^+$, followed by precipitation of the orange-yellow or orange-red copper oxide. The reaction for glucose is shown below. Glucose is oxidized in this reaction to a carboxylic acid; some keto-sugars (*e.g.*, fructose) are also oxidized in this type of reaction, probably after isomerization to an aldose.



β-D-Glucose
(pyranose)

D-Glucose
(linear form)

D-Gluconate
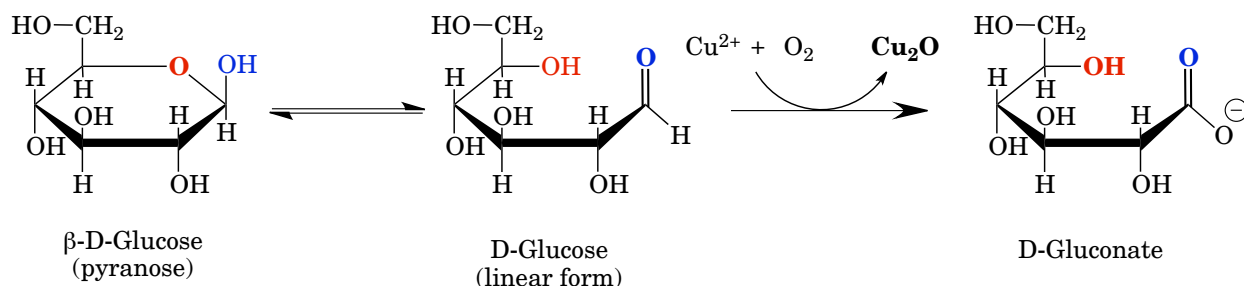
## Side note: measuring glucose concentration in biological fluids

The prevalence of diabetes mellitus has led to a variety of methods for measuring glucose concentration in both urine and blood. One current method uses a bacterial enzyme, glucose oxidase, which performs the reaction shown below. This reaction generates hydrogen peroxide, which can be readily measured. The glucose oxidase reaction is specific for glucose. Although the reducing sugar tests were not chemically specific for glucose, they were also useful for diagnostic purposes in diabetes mellitus, because both in normal individuals and in diabetic patients, glucose concentrations are far higher than those of any other reducing sugar. The enzymatic tests currently used are preferred because they are considerably more sensitive than the reducing sugar tests, and may be performed using very small amounts of blood.
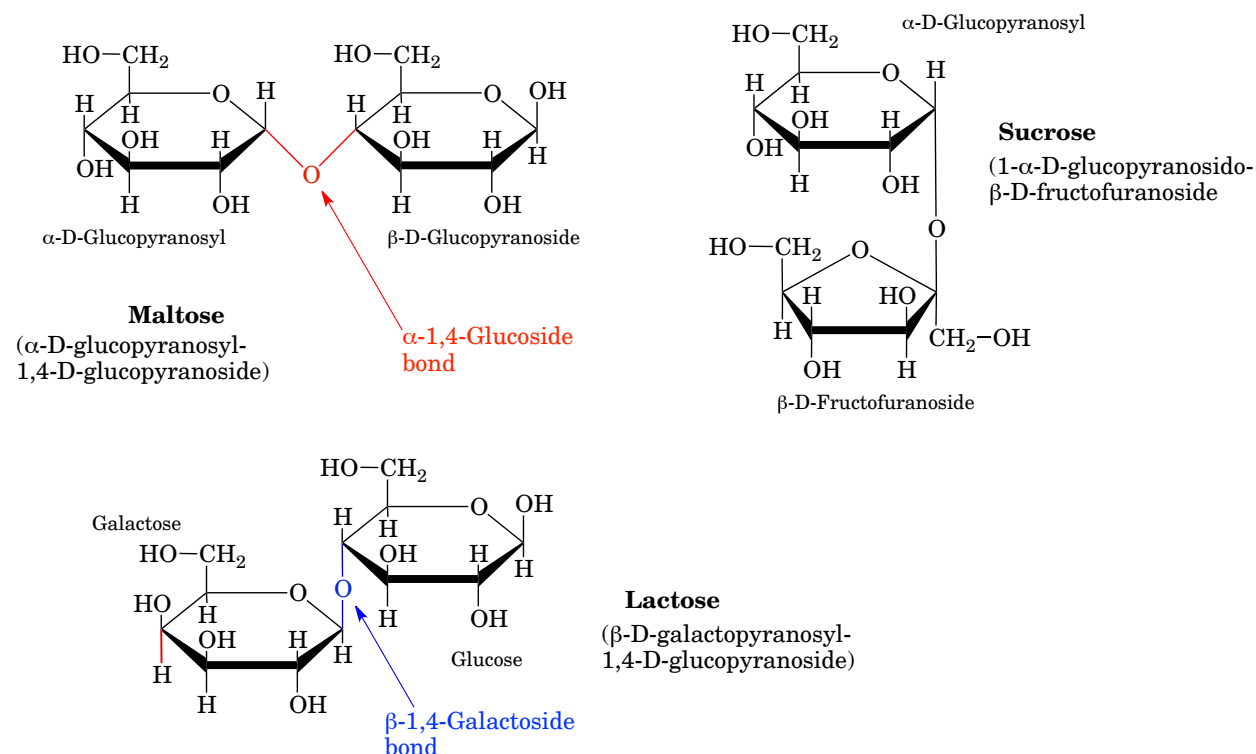


β-D-Glucose

*Glucose oxidase*

Gluconolactone
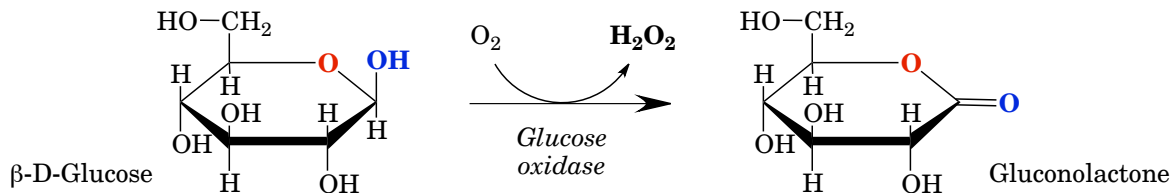
## Complex carbohydrates

In addition to monosaccharides, biological systems use a variety of more complex carbohydrates, in which glycoside bonds link two or more monosaccharides. The nomenclature of these compounds specifies the carbohydrates involved, the carbons connected by the bond, and, when the anomeric hydroxyl is modified, the

configuration at the anomeric carbon.

Examples of common disaccharides that illustrate this nomenclature are shown below.

HO—CH₂

H

H

OH

OH

H

O

H

OH

α-D-Glucopyranosyl

**Maltose**
(α-D-glucopyranosyl-
1,4-D-glucopyranoside)

α-1,4-Glucoside bond

HO—CH₂

H

H

OH

H

O

OH

H

H

OH

β-D-Glucopyranoside

α-D-Glucopyranosyl

HO—CH₂

H

H

OH

OH

O

H

H

OH

**Sucrose**
(1-α-D-glucopyranosido-
β-D-fructofuranoside

HO—CH₂

H

O

HO

H

CH₂—OH

OH

H

β-D-Fructofuranoside

O

Galactose

HO—CH₂

HO

H

OH

H

O

H

H

OH

H

HO—CH₂

H

H

OH

O

OH

H

OH

O

OH

H

Glucose

**Lactose**
(β-D-galactopyranosyl-
1,4-D-glucopyranoside)

β-1,4-Galactoside bond

Maltose contains a link between the 1-position of one monosaccharide and the 4-position of the other. This bond is a **"α-1,4 glucoside bond"**, because the anomeric carbon is in the α-position. The "glucoside" refers to the fact that the compound accepting the bond is glucose. In nearly all cases, the carbohydrates will be in the ring form; in most cases, the terms "glucoside" and "glucopyranoside" are therefore synonymous.

Lactose contains a β-anomeric carbon in the glycoside bond. In lactose, the monosaccharide is galactose, and therefore the bond is a "**β-1,4-galactoside**" bond. The specific type of bond present can be very important, because the ability of enzymes to use these molecules as substrates can depend on the configuration. Lactose is a prime example of this; during childhood, many individuals cease expressing the enzyme β-galactosidase (also known as lactase) that is capable of cleaving β-galactoside bonds, and can therefore no longer metabolize lactose.
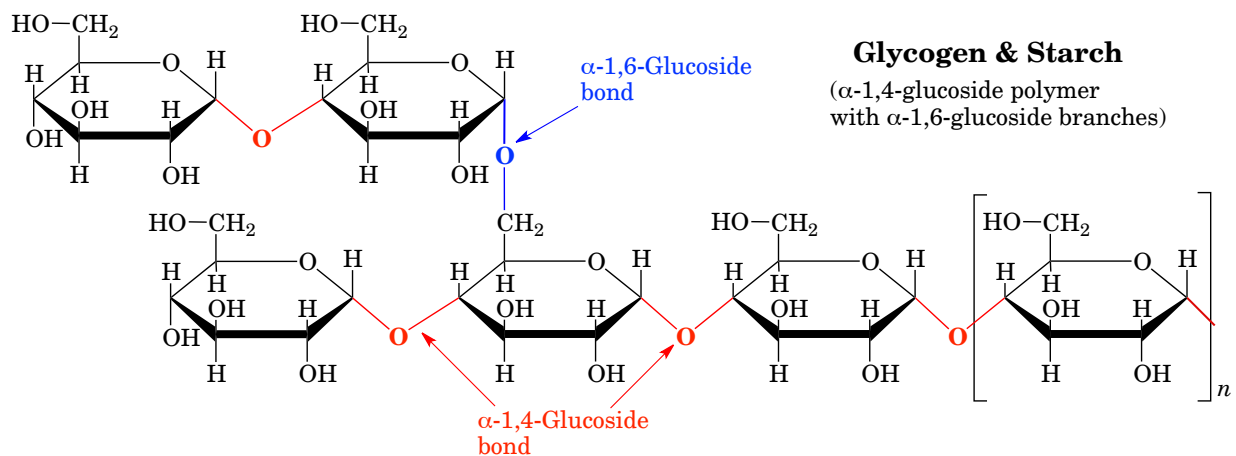
Of the disaccharide structures above, sucrose is the only one in which the configuration of the anomeric carbon is given for both monosaccharide units. This is because only sucrose has both anomeric hydroxyls involved in covalent bonds, while the others contain one free anomeric hydroxyl.

## Polysaccharides

Osmotic pressure is proportional to the number of molecules present on different sides of a barrier. Storing large numbers of glucose molecules within a cell would: 1) result in the influx of large amounts of water; 2) would create a large, unfavorable concentration gradient that would force cells to expend energy to maintain their supply of glucose; and 3) would risk cellular damage if large amounts of reducing sugars were present within the cell. In contrast, large polymers of saccharide units do not result in large osmotic pressures, and safely confine the reducing power of the free anomeric carbon within glycoside bonds.

Although a variety of carbohydrate polymers are used for biological purposes, we will look at only a few major types. **Starch** and **glycogen** are used as **glucose storage polymers**. Starch is used in plants, while glycogen is used in animals and bacteria. Both starch and glycogen consist of α-1,4-glycoside bonds, with some α-1,6-glycoside bonds as branch points. The main difference between starch and glycogen is the number of branch points, with glycogen being more heavily branched.
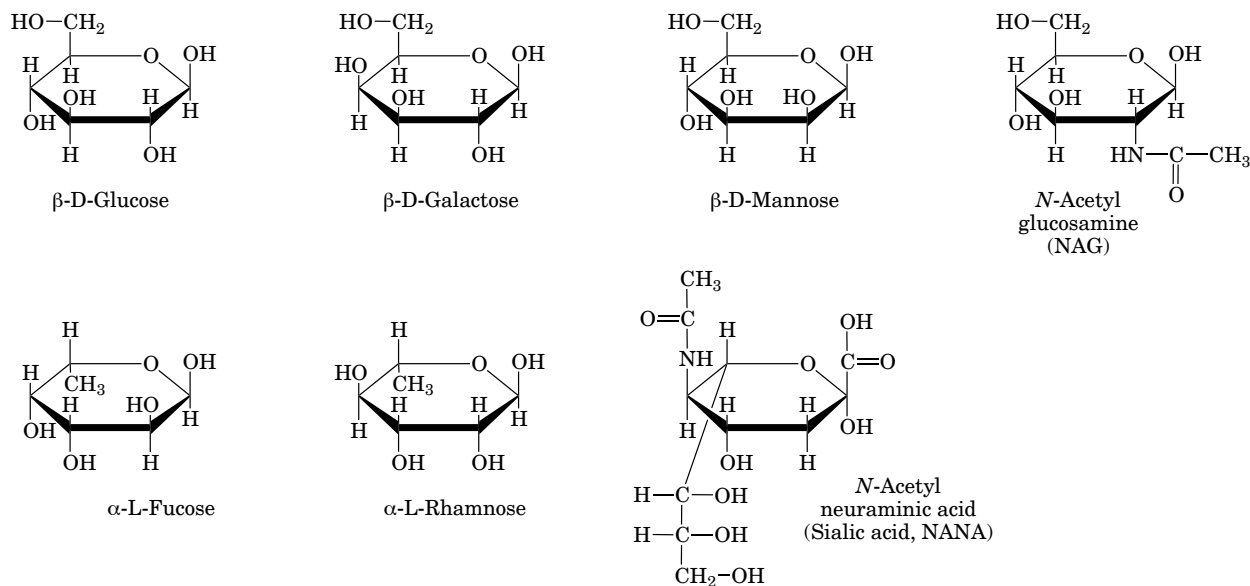


**Glycogen & Starch**
(α-1,4-glucoside polymer with α-1,6-glucoside branches)

α-1,6-Glucoside bond

α-1,4-Glucoside bond

In contrast to starch and glycogen, which act as energy storage molecules, the linear β-1,4 polymer **cellulose** is used for **structural purposes**. Cellulose, because of the ability of the its hydroxyl groups to form large numbers of stable hydrogen bonds, makes an extremely stable, rigid, and strong molecule, and is a major component of the structure of plants and trees. In addition, the β-1,4-glycoside bond is resistant to cleavage, and in fact is impervious to digestion by animals. Even cows, termites, and other animals that rely exclusively on plants for sustenance must use bacteria to perform the actual degradation of the cellulose molecule.



**Cellulose**
(β-1,4-glucoside polymer)

β-1,4-Glucoside bond
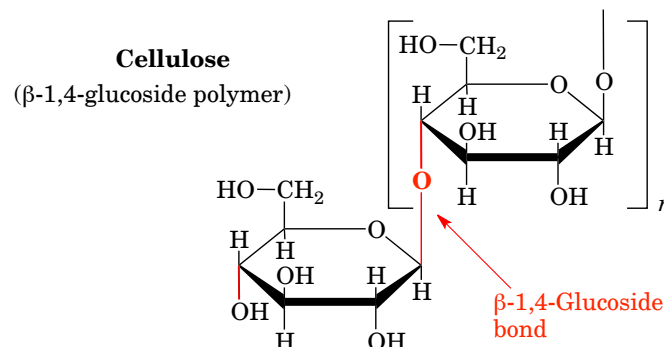
## Glycoproteins

Many secreted and cell-surface proteins contain carbohydrates. The carbohydrate groups are added to the protein in the endoplasmic reticulum and Golgi apparatus during and following synthesis of the polypeptide chain. The saccharide units include those discussed above, but also include a variety of other compounds. Structures of some of the monosaccharide units more commonly used in glycoproteins are shown below.



β-D-Glucose    β-D-Galactose    β-D-Mannose    *N*-Acetyl glucosamine (NAG)

α-L-Fucose    α-L-Rhamnose    *N*-Acetyl neuraminic acid (Sialic acid, NANA)
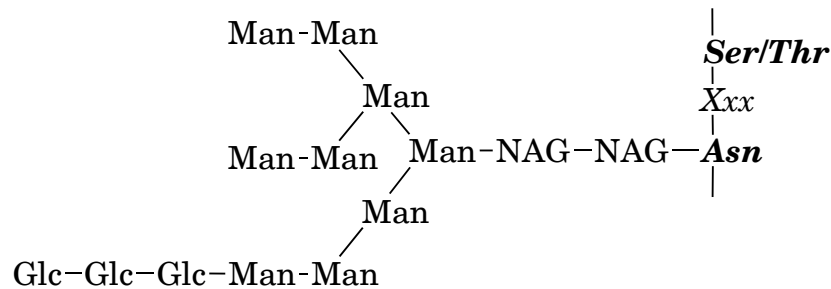
The carbohydrate groups are either **N-linked** (attached to the amide group of asparagine residues) or **O-linked** (attached to serine or threonine side-chains). In general, the *N*-linked carbohydrate groups are larger, more complex chains of carbohydrates. The carbohydrate groups can form a significant part of the protein, with some glycoproteins having 50% or more of their molecular weight comprised of carbohydrate.

Glycosylation sites are probably specified by the three-dimensional structure of the protein. This is especially true for *O*-linked attachments, which do not seem to have any sequence signals. While *N*-linked groups are attached to the asparagine side-chain of Asn-X-Ser/Thr sequences, it is likely that *N*-linked carbohydrate attachment is also in part dependent upon structure, because not all of the Asn-X-Ser/Thr sites in glycoproteins are glycosylated.

Carbohydrate groups that will be attached to *N*-linked sites are first synthesized on **dolichol pyrophosphate**, and then transferred co-translationally to the protein.



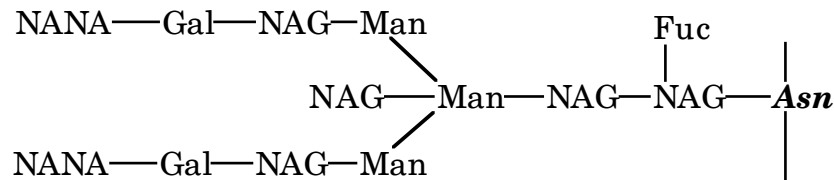Dolichol pyrophosphate ($n = 14 - 24$)
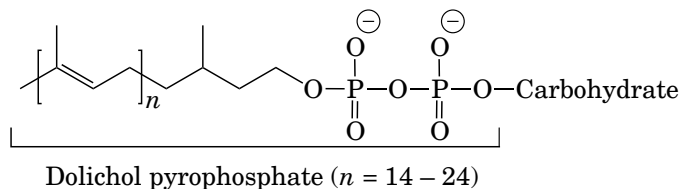
In the drawing below, the basic structure of the added polysaccharide is shown. The drawing uses the common abbreviations for the different monosaccharides, with Man = mannose, NAG = *N*-acetyl glucosamine, and Glc = glucose.

```
Man-Man                                    |
      \                                 Ser/Thr
       Man                                 |
      /   \                               Xxx
Man-Man   Man-NAG−NAG—Asn                  |
      \   /                                |
       Man
      /
Glc−Glc−Glc−Man-Man
```

Following addition to the protein, the carbohydrate groups are modified by removal or addition of monosaccharide units. The carbohydrate groups are frequently heterogeneous, with different protein molecules, and different sites within the same protein containing at least slightly different modifications. Depending on the protein, the linkages may also vary; 1,4, and 1,6 linkages are the most common, but several other linkages have also been observed.

An example of a "mature" carbohydrate unit is shown below. Note that during the maturation process, several of the mannose residues have been removed, and a few new saccharide units have been added (Fuc = fucose, Gal = galactose, and NANA = *N*-acetylneuraminic acid).

```
NANA——Gal——NAG—Man              Fuc
                   \              |       |
              NAG——Man——NAG——NAG——Asn
                   /              |
NANA——Gal——NAG—Man               |
```

The carbohydrate portion of proteins is thought to assist in maintaining the proper conformation. The carbohydrate can also act as a signal for processing. This is especially the case for asialoglycoproteins; after the enzyme neuraminidase removes terminal *N*-acetyl neuraminic acid residue from a protein, the asialoglycoprotein receptor removes the protein from circulation.

**Summary**

Carbohydrates are partially oxidized organic compounds widely used in biological systems for structural roles, for metabolism to generate energy, and for metabolism to generate biosynthetic intermediates.

Most carbohydrates are chiral compounds. The most commonly used carbohydrates have the D configuration, although some carbohydrates used for post-translational modification and some other non-metabolic purposes have the L configuration.

The vast majority of carbohydrates contain either a ketone or an aldehyde. These groups are reactive; in carbohydrates of five or more carbons, the carbonyl carbon tends to undergo a non-enzymatic attack by one of the hydroxyl groups, resulting in a hemiacetal or hemiketal ring structure.

Unmodified carbohydrates are present in aqueous solution as a mixture of the ring and linear forms. The free carbonyl of the linear form may react with other biological compounds to produce potentially deleterious covalent adducts.

Carbohydrates may form chains involving ether linkages between monosaccharide units. At least one of the carbons involved in these linkages is the anomeric carbon. Oligosaccharide units are used in glycoproteins. Polysaccharides act as structural molecules (such as cellulose), and as energy storage molecules (sucrose, starch, and glycogen).