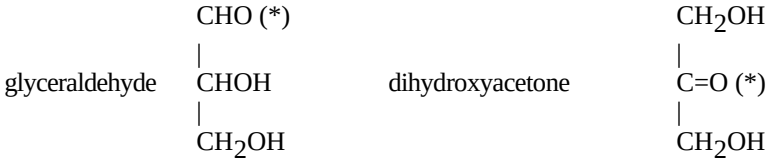
II. Chemistry of Sugars

Sugars-or saccharides- are the most abundant bio-molecule on the planet. They are important in a number of biological roles. Most obviously you will know them as a major component of your diet. Insoluble sugars also function as structural material in the cell walls of plants and bacteria and in the connective tissue and cell coats of animals. Sugar polymers serve to lubricate skeletal joints. Sugars can be attached to proteins (in glycoproteins) and lipids (in glycolipids) and these *glycosylated* compounds serve (i) as antigenic sites; (ii) provide signals that determine the cellular localization of proteins and (iii) function as signals that allow cells to recognize each other and adhere in the formation of tissues and organs.

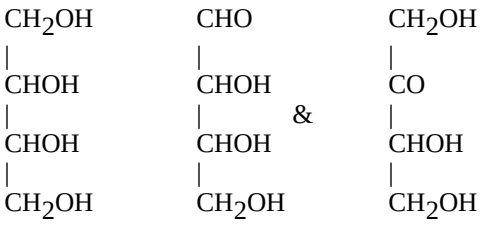
You will already have encountered two sugars as structural components of nucleic acids; these are ribose and deoxyribose,

A simple sugar is a derivative of a straight chain polyhydroxy-alcohol. Two classes of derivatives exist. In the first class, a terminal (primary) alcohol is oxidized to an aldehyde to yield a member of the ALDOSE series. In the second class, a penultimate (secondary) alcohol function is oxidized to a carbonyl to yield a member of the KETOSE family. We can consider glycerol to be the parent of all sugars.



ALDOSE SERIES KETOSE SERIES

Both series are extended by the consecutive addition of CHOH functions. This is most simply viewed as starting from the longer homologs of glycerol and each time oxidizing the terminal (aldoses) or penultimate (ketoses) alcohol function. For example:

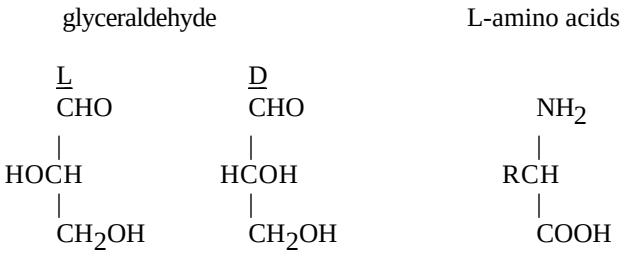


1,2,3,4 butanetetrol aldotetrose ketotetrose

Nomenclature: Prefix: aldo- (-CHO); keto- (-C=O) Suffix: -ose (=sugar). No. of C atoms = tri (3), tetr (4), pent (5), hex (6). E.g. aldohexose = 6 C sugar in CHO form. Sugars are numbered starting from the end that has been oxidized.

In glyceraldehyde, the middle carbon atom is chiral; i.e., it bears four different substituents, and consequently glyceraaldehyde has non-superimposable stereo isomers; it possesses enantiomeric (mirror-image) forms. The D- and L-forms of glyceraldehyde are identified as follows. Orient the glyceraldehyde as shown below with the eye of the observer

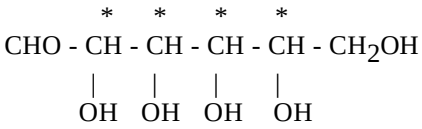
viewing from the primary alcohol towards the aldehyde.



If the secondary OH is on the right we have D-glyceraldehyde, and vice versa. (This is the same rule as for amino acids oriented as shown.) For longer chains the D/L distinction is based on the orientation of the secondary OH furthest from the C=O, C5 in hexose. Note that D/L in this context refers only to the configuration about this carbon atom. It does not specify the optical activity of the sugar; the latter is denoted by d (or +, dextrorotatory) or l (-, levorotatory). The D/L system has now been officially replaced by the Cahn-Ingold-Prelog or R, S system for designating configuration at sp3 carbon centers (this is reviewed in Additional Material). D-sugars = R-sugars.

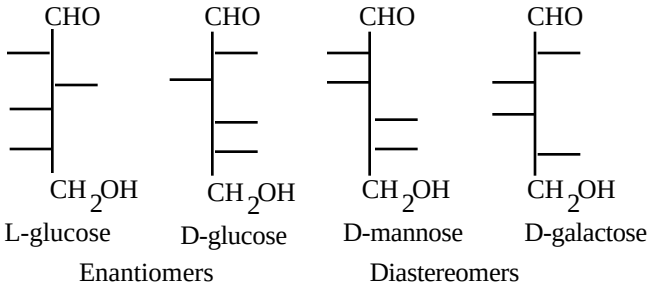
Stereochemical Isomers

Every secondary alcohol function in a sugar is a chiral center. For example, in the aldohexose



the "*" identify the chiral carbon centers. N chiral centers yield 2^N isomers; there are consequently 16 isomers of aldohexose. These are divided into 2 families, D/L, depending on the configuration at C5 (the last starred carbon on the right). The D-isomers are the important ones. We will deal mainly with one of them, D-GLUCOSE, which is the most abundant monosaccharide from the aldose family, but two others, D-GALACTOSE (a component of milk sugar) and D-MANNOSE, are important biochemically.

These sugars are shown below using a shorthand form of the Fischer projection formula (see appendix); the sideways-pointing lines identify the orientation of the OH substituent of the secondary alcohols. Note that in the L-forms the orientations of all four secondary OH's are reversed relative to the D form. Not all of these isomers are mirror images of one another. Those that are mirror-images are called ENANTIOMERS; the other stereo-isomers are called DIASTEREOMERS. D- and L-glucose are enantiomers. Glucose, galactose and mannose are diastereomers. D-Glucose and D-mannose, which differ by the orientation of a -OH at a single chiral center, are called EPIMERS (epimers are thus a subset of diastereomers). Mannose and galactose are epimers of glucose but not of each other.



Other sugars you will (or already have) encounter are THREOSE (4C), RIBOSE and DEOXYRIBOSE (5C), FRUCTOSE (6C ketose) and SEDOHEPTULOSE (7C ketose) . As ketoses have one less chiral center than hexoses they have half as many structural isomers for the same chain length; otherwise fructose resembles glucose with the C=O

moved down 1 position.

Figs 10-1 and 10-2 of your text book summarizes the "family trees" for Aldoses and Ketoses.

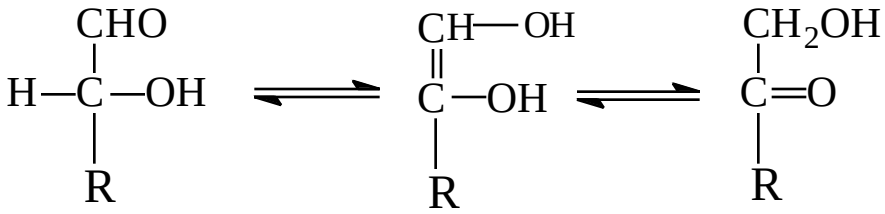
Chemical Reactions Important for Structural Analysis

(1) Ether formation: Base-induced reaction of all -OH's with alkylating agents e.g., methyl iodide plus dimethyl sulfoxide (anion, scavenges proton) plus sodium hydride (catalyst; very strong base) to produce methyl ether ($-OCH_3$; -OH's react as strongly nucleophilic $-O^-$). The ether link is very stable. The oxygen at C1 also reacts (it is a hemi acetal which will be clarified soon) but the $-OCH_3$ at C1 is atypical and hydrolyzes to -OH in dilute acid. Because of this the product from reaction of glucose is 2,3,4,5,6 pentamethyl glucose.

(3) Reduction (with borohydride) converts CHO to CH_2OH (aldoses and ketoses give an alditol) e.g. glucose to glucitol (a.k.a. sorbitol); fructose gives glucitol *and* mannitol (why?).

(4) Oxidation. Mild oxidation is often used to test for the free C=O group because with weak oxidizing reagents such as alkaline Ag^+ (Tollen's Reagent) or alkaline Cu^{2+} (Fehling's Reagent) the products are the metallic Ag or the red insoluble Cu_2O both of which can be easily visualized. Sugars that give a positive response in these tests are called REDUCING sugars. If the -C=O group is tied up (see below) so that it cannot isomerize to a free -C=O then we have a NON-REDUCING sugar (e.g. sucrose). Note that fructose is a reducing sugar. Even though the C=O group of fructose is at C_2, at high pH the C=O rearranges to C_1; this is the Lobry de Bruyn-van Eckenstein re-arrangement. It's catalyzed by a base removing the α-proton to form the enediol. This can then rearrange to yield fructose or re-capture the proton to yield glucose + mannose (because the return of the proton can be to either face).

Interconversion of Glucose and Fructose via the enediol

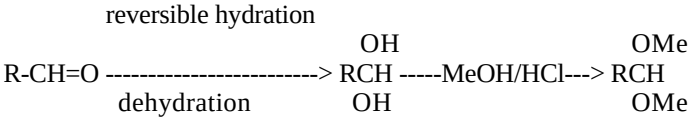


This re-arrangement is found twice in the glycolytic pathway (where it is enzyme catalyzed).

Some nomenclature:
a) CHO to COOH (aldose to aldonic acid) e.g., gluconic acid
b) CH_2OH to COOH (aldose to uronic acid) e.g., glucuronic acid.
c) CH_2OH **and** CHO to COOH (aldose to aldaric acid) e.g. glucaric acid.

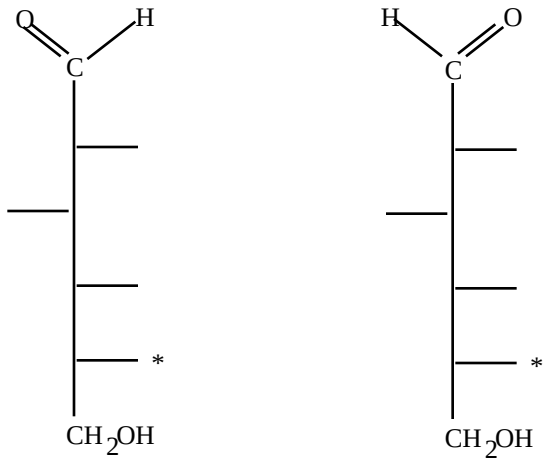
Sugars are Rings in Solution

Up to this point we have been representing the sugars as "linear" molecules. In reality the linear form is normally a minor species (often less than 0.1%) The initial evidence for this came from the following anomalous chemical behavior. The typical reaction of carbonyl-containing compounds with weak methylating reagents, e.g. MeOH/HCl, yields a dimethyl derivative called an acetal.

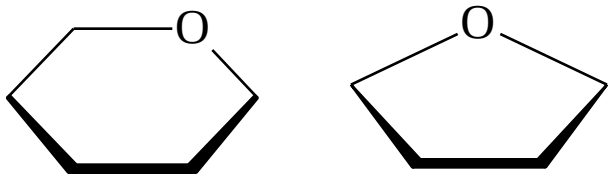


aldehyde hydrate acetal

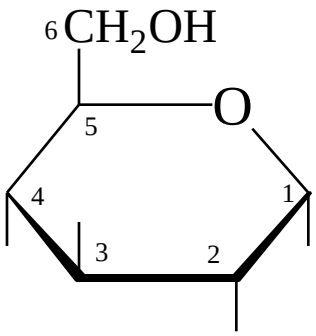
But with glucose, we obtain two different monomethyl products. This result is evidence that the C=O is abnormal. In fact, we know that it is "tied-up" as a *cyclic* hemiacetal. To visualize this we recognize that the aldehyde group is planar (sp^2 hybridization) and we can therefore think of two extreme orientations shown below:



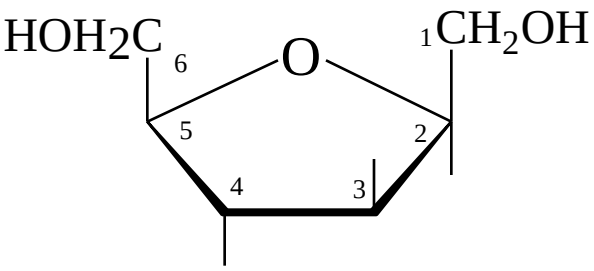
Ring closure occurs by attack of a secondary -OH (presumably as the nucleophilic anion) on the carbon of the electron deficient C=O . This attack can occur on either face of the planar CHO with the result that the -OH group that is created at C_1 can be oriented in either of two directions (if attack is on the left structure the -OH created will point to the left and vice versa). The two forms of glucose that are formed are called **anomers** and the C bearing the C=O is the anomeric carbon. When the newly created -OH has the same orientation as the -OH that did the attacking (the two -OH's are cis) we have the α-anomer, otherwise it is the β-anomer. Usually the -OH that does the attacking is located on C_5 (indicated with the *) and a 6-membered ring is formed. This is called a pyranoside by analogy to tetrahydopyran (below, left). Note that by closing the ring we have created another chiral center so that the ring forms of aldohexoses have twice as many isomers as the straight chain forms.



Less common is attack by the C4 -OH which leads to the sterically strained, 5-membered furanoside (by analogy to tetrahydrofuran). In ketohexoses the C=O is at C_2 and so attack by C_5 still yields a 5-membered ring (as you will have seen in ribose). The cyclic sugars are normally represented using Haworth structures; in this representation the lines denote -OH and there is an implicit -H on the other end. The procedure for converting a Fisher to a Haworth structure is given in Additional Reading at the end of the chapter.



Glucose



α-Fructose

An important result of the formation of the cyclic hemiacetal is that the hydroxyl group "created" on the carbon atom, C_1, bearing the original aldehyde function can be oriented in either of two ways, i.e., a new chiral center is created. Thus the α and β forms just described are geometric isomers of each other. α and β glucose can be crystallized separately and can be distinguished by a measurement of optical activity. For glucose, both α and β forms are dextrorotatory. However, dissolving the pure (α or β) solid in water leads to the production of the mixture,

α-glucopyranoside ⇔ open form ⇔ β-glucopyranoside

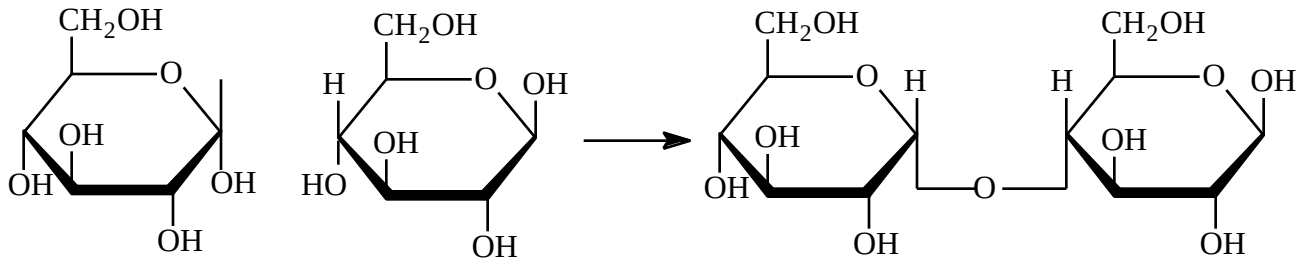
40% <.1% 60%

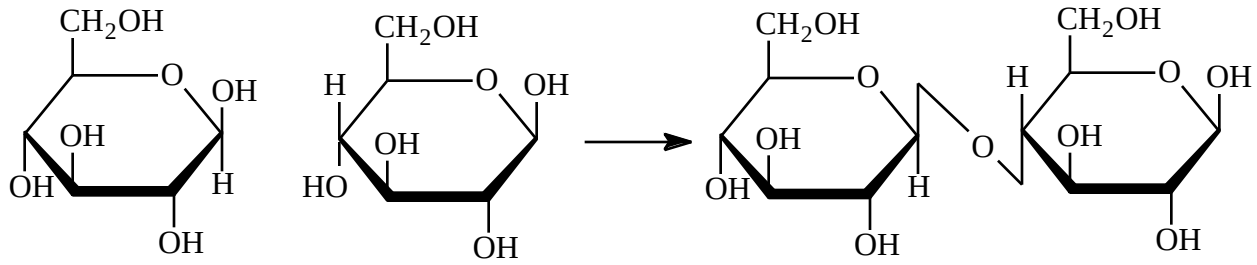
a phenomenon called **mutarotation** because the optical activity changes from that of the pure form observed immediately on dissolution to that of the equilibrium mixture some five hours later. Pure α = $+113^o$, pure β = $+19^o$, equilibrium solution = $+52^o$.

The ring forms of the sugars are the ones that predominate in solution. For example, the proportion of the straight-chain form of glucose is quite small, about 0.1% of the total molecules. Glucose slowly inter converts between the three structures (half time ~ 1 hour). Note however that when a sugar is bound to the active site of an enzyme it is frequently converted to the straight-chain form.

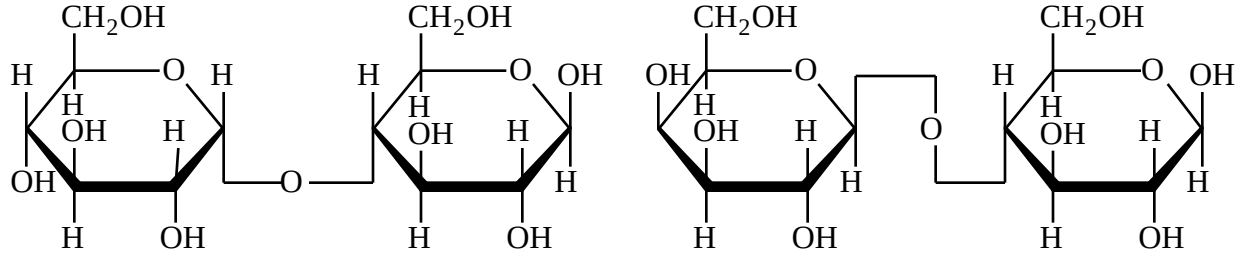
Di-, Oligo- and Poly-saccharides

An extremely important biochemical reaction is the condensation of two (or more) monosaccharides by the elimination of water from an OH group present on each of the two sugars. Most commonly the reaction occurs between the OH present on C_1 of one monosaccharide and that present on C_4 of the second to form a 1 4 GLYCOSIDIC linkage. Because the reaction involves C_1, which can exist in either α- or β- forms, we can obtain either an α (1 4) or a β (1 4) **glycoside.**





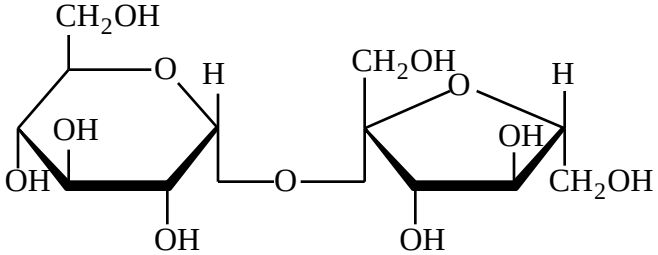
Important Disaccharides: The dimer of glucose with the α **(1⇒4)** glycosidic link is the disaccharide MALTOSE. It is an intermediate in the hydrolysis of STARCH; its full name is α-D(glucopyranosyl) (1 4)-?−D-glucopyranose (top right, also next left). The ? (my notation) reminds us that the "right-hand" sugar will exist in both α and β anomeric forms. The picture shown is the β-anomer. The dimer of glucose with the β **(1⇒4)** glycosidic link is called cellobiose; it is an intermediate in the degradation of cellulose (grass and tree trunks).



MALTOSE LACTOSE

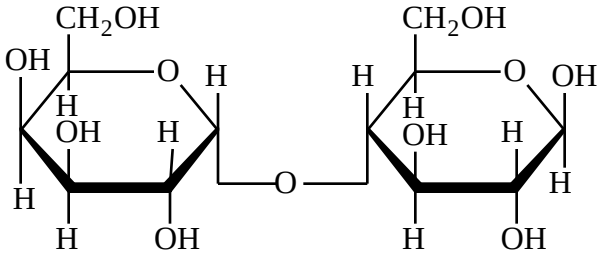
.The disaccharide LACTOSE present in milk is galactose β(1 ⇒ 4) glucose with the full systematic name of β-D (galactopyranosyl)- (1 ⇒ 4)-?-D glucopyranose.

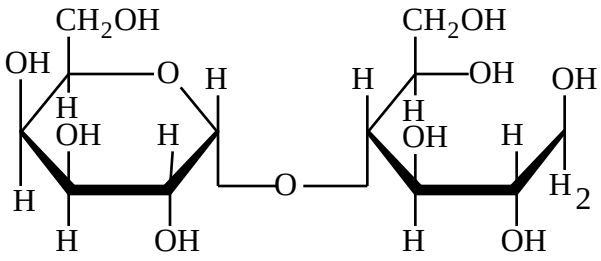
Glycoside links to other carbon atoms are fairly common, notably 1 ⇒ 2 and 1 ⇒ 6**.** SUCROSE (domestic sugar) is glucose-α (1 ⇒ 2) -β-fructose.

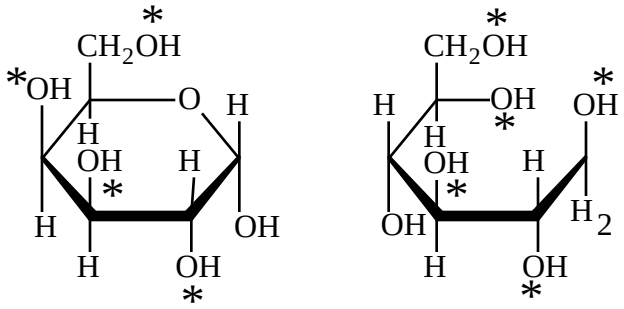


Know fructose, maltose and lactose. Which are reducing? Why?

Structure determination of disaccharides.







(1) Establish if the disaccharide is reducing.

(2) Hydrolyze and identify component monosaccharides. They are?

(3) Identification of C atoms present in glycosidic link. Extensive methylation of disaccharide (care: is the disaccharide a reducing sugar? If so pre-treat with borohydride to convert nascent C=O to C-OH) followed by hydrolysis of glycosidic bond (mild acid) and subsequent identification of the methylated monosaccharides (gas liquid chromatography + mass spectrometrys).

They are?

(4) Nature of glycosidic link. Which enzyme hydrolyzes the disaccharide?

yeast maltase ⇒ α (this one)

almond emulsin ⇒ β (not this one)

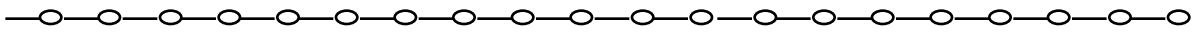
Polysaccharides

By the consecutive application of the glycosidic condensation reaction we obtain polysaccharides. In contrast to proteins and nucleic acids which are linear polymers with variable monomeric units, the multiple -OH groups present on the sugar ring allow the possibility of both linear and branched polymers most commonly of the same monomeric unit, namely glucose. When only a single monosaccharide is present the polysaccharide is called simple: (e.g., starch and cellulose). Complex polysaccharides contain several sugars, including substituted sugars (amino-sugars). The blood group (ABO) antigens are important examples of the latter class.

Linear polysaccharides.

The best known contain (1 4) glycosidic links. These can have either the α or the β anomeric configuration but not both.

Amylose is an edible, **linear** polymer composed of between 5,000-500,000 (n) glucose residues connected by α (1 4) links. (it is maltose perpetuated.)



Hydrolysis of amylose gives **n** glucose residues; methylation followed by hydrolysis gives (**n**-1) 2,3,6 trimethyl glucose plus one 2,3,4,6 tetramethyl glucose (from left-most residue). Amylose is broken down in the body by α-amylase (in saliva), an endo-glycosidase which hydrolyzes internal α(1 4) glycosidic bonds at random locations to produce smaller fragments; note that the α in α-amylase does not refer to the glycosidic link but is used to distinguish it from another enzyme, β-amylase, found in malt. Because the α glycosidic link imparts a constant twist to the molecule the polymer is helical (left-handed): see V&V Fig. 10-17).

Cellulose (Plant Structural Material) is a **linear** β(1 4) polymer of glucose (containing 300-25,000 residues). The human polysaccharide degrading enzymes present in the saliva and pancreatic juice hydrolyze α (1 4) but not β(1 4) links; consequently amylose, but not cellulose, is a source of food. Cellulose is the most abundant organic compound on this planet; about 50% of organic mass is cellulose. Cotton is pure cellulose. The β glycosidic link forces each glucose to be flipped 180^{o} with respect to the preceding but the overall chain is essentially linear. Chains then assemble first into 2D sheets which can stack into three dimensional bundles that give cotton its strength. The interactions between chains and between sheets are stabilized by hydrogen bonds (See V&V Fig 10-15).

Branched polysaccharides.

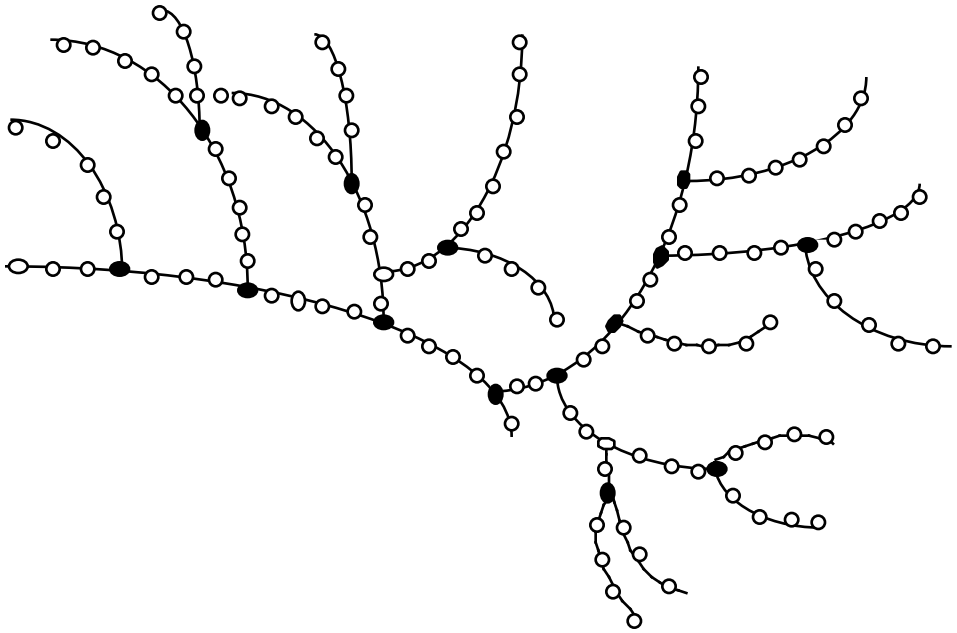
Glycogen (Animal Starch) is an edible, **highly branched** polymer (MW ~1 million, ~5000 glucose units) composed of linear segments of α (1 4) glucose (o) which branch at about every tenth glucose residue by virtue of α (1 6) links (•) (see next page). Complete degradation in the intestine depends on the combined action of α(1 4) amylase and an α(1 6) amylase Within the liver cell degradation is accomplished by the cooperation of glycogen phosphorylase (which cleaves the 1 4 bonds) plus a debranching enzyme which clips out the branch points in the "limit dextrin" produced by the enzyme phosphorylase (see lecture on metabolism of carbohydrates, to follow).

Amylopectin resembles glycogen except that the branches are further apart, about every 25 residues on average.

Plant starch, an important constituent of our diet, is about 25% amylose and 75% amylopectin.

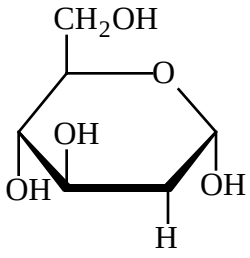
Dextran is a linear polysaccharide with α(1 6) glycosidic links. Dextrans that have been *chemically* cross-linked (e.g. with epichlorhydrin) yield Sephadex and Biogel, the column materials used in gel filtration.

Polysaccharides rather than monosaccharides are used as storage forms because osmotic pressure is proportional to the number of molecules and a polymer of **n** monosaccharides has 1/n'th the osmotic pressure of its n component sugars.

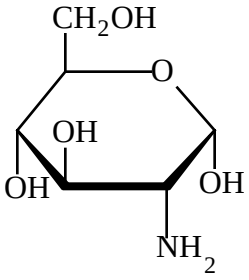


Modified Sugars:

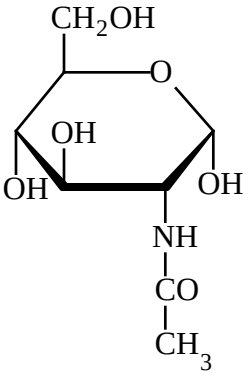
Modified sugars are important constituents of complex carbohydrates, glycoproteins and glycolipids. They are commonly of two kinds, deoxy and amino. In deoxy sugars an -OH is replaced by -H or a $-CH_2OH$ by CH_3. The former typically occurs at C_2 and the latter at C_6. You have already encountered 2-deoxyribose. In amino sugars the –OH is replaced by $-NH_2$, commonly at C_2. The amino group may subsequently be acetylated to give the N-acetyl derivative.



deoxy-sugar



amino-sugar



N-acetyl amino sugar

Additional Material

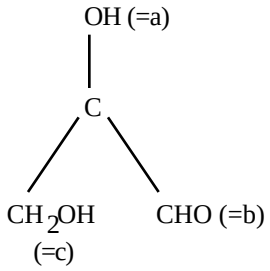
A review. The Cahn-Ingold-Prelog or R,S System for Designating Configuration at sp3 carbon centers (See Loudon, Organic Chemistry, p. 186).

1. Assign priorities to the four atoms attached to the chiral carbon with the highest priority (a) being assigned to the atom with the largest atomic number and the lowest priority (d) to the atom with the smallest atomic number. For glyceraldehyde, a = OH, d = H. However, the other two substituents are both carbons. Which is b and which is c?

2. To resolve this we now compare the substituents on the two "deadlocked" carbon atoms, hoping to find that one of these C atoms has a substituent with a higher priority than the other. If this is the case, the C atom bearing the higher priority substituent is assigned to b; the other is then c, e.g., for the two underlined C atoms, -$\underline{C}H_2CH_3$ has a higher priority than -$\underline{C}H_3$ because in the former one of the substituents on the C is another C, whereas in the latter all the substituents are H.

3. Finally, when a double bound is present on one of the carbon atoms (e.g., C=0) the oxygen substituent on this carbon is duplicated. So C=0 counts as CO2 By this rule -CHO (i.e. CHO2) has a higher priority than CH_2OH. Hence CHO = b and CH_2OH = c.

Orient the chiral carbon so that you are looking from it towards substituent d; i.e., along the CH bond. Observe the relative orientation of groups a, b, and c



Trace the path a --> b --> c. If it is clockwise, the chiral carbon is assigned the R designation (R = rectus); if it is counter-clockwise, it is assigned the S designation (S = sinister). D-glyceraldehyde = R-glyceraldehyde. N.B.. D/L applies only to C5 (glucose); R/S applies to each chiral center independently.

Establishing the size of the ring.

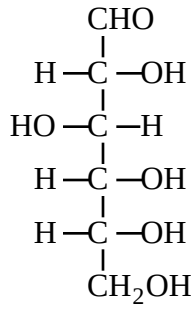
1) Completely methylate all -OH groups (e.g. methyl iodide + DMSO + NaH)
2) Hydrolyze terminal -OCH_3 (i.e., at C_1) (not an ether; hemi-acetal hydrolyzes in dilute acid)
3) Oxidize carbons not protected by methylation
   - Pyranosides -----> trimethoxyglutarate
   - Furanosides ----- > dimethoxytartrate

(The reduction in chain length is due to the breaking of the bond between C_5 and C_6 (in pyranosides); as C_5 is oxidized to COOH it would become the unthinkable 5-valent carbon unless some other bond broke).

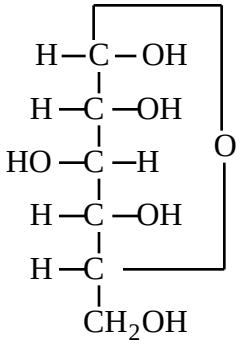
Representations of Sugar Structure

There are several ways to draw the structure of glucose. The simplest is the method introduced by Emil Fischer which represents the sugar as a straight chain of carbon atoms with the lowest numbered at the top and the OH's of the secondary alcohol functions shown to the right or left. The horizontal lines are to be visualized as projecting out of the page; the vertical lines project into the page so that the carbon backbone has the overall profile of a banana, the top and bottom of which lies behind the plane of this page. The Fischer representation can be moved around in the plane of the paper. However it should not be lifted out of the plane or turned over. The assignment of D or L depends on the orientation of the OH in the penultimate carbon (C_5 in hexoses) and the orientation of the other OH's are relative to this one. However, sugars exist mainly in the ring form. In the Fischer representation the open chain form is converted to the ring form by drawing a "box" connecting the C atom bearing the keto function (C_1 or C_2) to C_5. If the keto function is at C_1 the C_1 ---> C_5 link yields the pyranose ring; the C_2 ---> C_5 link gives the furanose ring. The box is drawn on the side of the main chain as specified by the orientation of the OH at C_5 (or C_4). C_1 is now chiral; if the OH created at C_1 is cis to the OH at C_5 we have the α anomeric form, and vice versa.

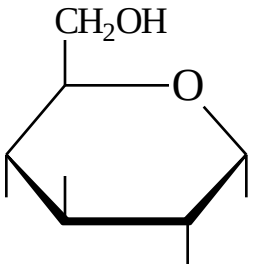
A popular alternative is the Haworth representation which attempts to convey more three-dimensional information. The sugar is drawn as a hexagon (or pentagon), and you must imagine that the hexagon is projecting out of (and behind) the plane of the page. This is sometimes facilitated by thickening the lower 3 edges of the hexagon. The OH's are then explicitly shown and lie either above or below the plane of the hexagon - as indicated. In the drawing the anomeric carbon and the orientation for α are indicated. However in this Haworth representation defining the sense of the anomeric carbon is not easy; the Fischer representation is better for this purpose. Three representations of glucose are:



Original Fischer



Fischer Ring Form



Haworth

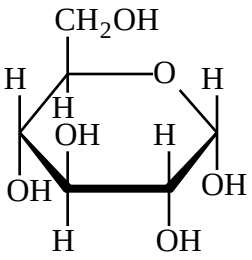
The correct configuration of the groups on the ring can be determined using three rules.

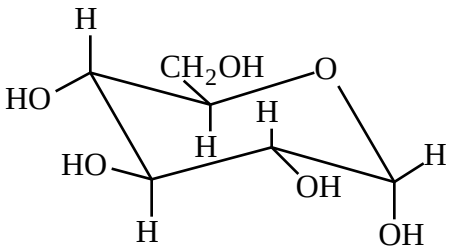
[1] If the ring closes on a hydroxyl which is on the right in the Fischer projection, the hydroxymethyl group (tail) points up; if it closes on a hydroxyl which was on the left in the Fischer projection, the tail points down.

[2] The ring hydroxyls point down if they are on the right in the Fischer projection, and up if they are on the left in the Fischer projection. (Note that these are also the positions they are in after the Fischer projection has been rotated 90o clockwise to lie on its side.)

[3] The hydroxyl on the anomeric carbon points down in the D series if it is α and up if β. In the L series, α is up and β is down.

Finally, the most realistic representation of structure would show that the ring form of glucose is not planar. The predominant forms in solution are both "chair" forms; however, a very small percentage of the sugar molecules are present in the boat form. Glucose in the Haworth and chair forms follows:


CH2OH
O
H
H
H
OH
H
OH
OH
H
OH


H
HO
CH2OH
O
H
H
HO
H
OH
H
H
OH