

Unit 5. Introduction to Chemistry of Biomolecules

Contents

- 5.1. Carbohydrates
 - 5.1.1. Glucose
 - 5.1.2. The Structure and Configuration of Glucose
 - 5.1.3. Anomeric forms of Monosaccharides
 - 5.1.4. Glycosides
 - 5.1.5. Disaccharides
 - 5.1.6. Polysaccharides
- 5.2. Lipids
 - 5.2.1. Fatty Acids
 - 5.2.2. Fats & Oils
 - 5.2.3. Waxes
 - 5.2.4. Phospholipids
 - 5.2.5. Prostaglandins
 - 5.2.6. Terpenes
 - 5.2.7. Steroids
- 5.3. Proteins and Amino Acids
 - 5.3.1. α -Amino Acids
 - 5.3.2. Reactions of Amino Acids
 - 5.3.3. Synthesis of Amino Acids
 - 5.3.4. Peptides & Proteins
 - 5.3.5. The Primary Structure of Peptides
 - 5.3.6. Secondary & Tertiary Structure of Large Peptides and Proteins
- 5.4. Nucleic Acids
 - 5.4.1. Introduction to the chemistry of Nucleic Acids (Structure and Chemistry)
- 5.5. Summary
- 5.6. Exercises

Learning Objectives

After completing this unit the student should be able to:

- Classify various preparative methods of biological molecule such as carbohydrates, lipids, amino acids and proteins, and their important chemical properties.
- Draw and identify the structures of glucose, its anomers, and its epimers, as Fischer projections, Haworth and as chair conformations to indicate stereochemistry of saccharides.
- Draw the common types of glycosidic linkages, and these linkages in disaccharides and polysaccharides.
- Know the various chemical reactions of glucose.
- Understand the structure, function and chemistry of different kinds of lipids.

- Classify simple and complex lipids. Identify oil, waxes, triglycerides, phospholipids, steroids, prostaglandins and terpenes.
- Explain how unsaturation affects the properties of fats and oils. Compare the properties of saturated fats with those of polyunsaturated oils and partially hydrogenated vegetable oils.
- Draw the structure of amino acids and peptides. How amino acids are linked to form peptides and proteins.
- study physical and chemical properties of amino acids.
- Show how to synthesize amino acids from simpler compounds, and show how to combine amino acids in the proper sequence to synthesize a peptide.
- Recognize the structures of DNA and RNA, and draw the structures of the common ribonucleotides and deoxyribonucleotides. How nucleotides are linked to form nucleic acids.
- Understand the chemistry of nucleosides, nucleotides, and nucleic acids (RNA and DNA).
- Explain how the genetic messages encoded in DNA are transcribed into mRNA and then translated, and into protein.

Introduction to Chemistry of Biomolecules

Living systems are composed of millions of different types of molecules, each with their own distinct chemical composition, shape, and function. These molecules form the basis of life and are required for the growth and maintenance. Such molecules are called **biomolecules**. Biomolecules follow the same principles of structure and reactivity as the organic molecules we have discussed so far. Most biomolecules have more complicated structures than those of the organic compounds; however, there is great similarity between the organic reactions carried out in the laboratory and those performed by nature inside the living cell. All these biomolecules have a **turnover** - they are constantly being changed into some other biomolecules and also made from some other biomolecules. This breaking and making is through chemical reactions constantly occurring in living organisms. Together all these chemical reactions are called **metabolism**. Each of the metabolic reactions results in the transformation of biomolecules. A few examples for such metabolic transformations are: removal of CO_2 from amino acids to make an amine, removal of amino group in a nucleotide base; hydrolysis of a glycosidic bond in a disaccharide, etc. There are thousands of such examples carried out in living organism.

These biomolecules can be classified into four major groups: carbohydrates, lipids, proteins, and nucleic acids. Carbohydrates, proteins, and nucleic acids are all examples of *polymers* - large molecules composed of similar subunits (monomers) covalently bonded together in a long chain. In contrast, lipids, while they can aggregate together as in cell membranes, do not covalently bond, and therefore are not polymers.

5.1. Carbohydrates

Carbohydrates, the most abundant biomolecules in nature; more than half of all 'organic' carbon is found in carbohydrates. Carbohydrates are a direct link between solar energy and the chemical bond energy of living organisms. They are formed during photosynthesis, a biochemical process in which solar energy is captured by plants and convert energy-poor inorganic CO_2 and H_2O to energy-rich organic molecules (carbohydrates). They act as storehouses of chemical energy (glucose, starch, glycogen); are structural elements in plants (e.g., cellulose) and in animals (e.g., chitin); and are essential components of nucleic acids (D-ribose and 2-deoxy-D-ribose). Most carbohydrates contain carbon, hydrogen and oxygen in the ratio $(\text{CH}_2\text{O})_n$, hence the name "carbohydrate" for "hydrate of carbon". Later studies showed that they are not hydrates as they did not contain intact water molecule, but the name "carbohydrate" still persists.

Now carbohydrates are usually defined as polyhydroxy aldehydes and polyhydroxy ketones or compounds that hydrolyze to yield polyhydroxy aldehydes and polyhydroxy ketones. Generally they exist in their cyclic (hemiacetal or acetal) forms. Therefore, the chemistry of carbohydrates is essentially the chemistry of hydroxyl groups and carbonyl groups and of the acetal bonds formed between these two functional groups.

Classification:

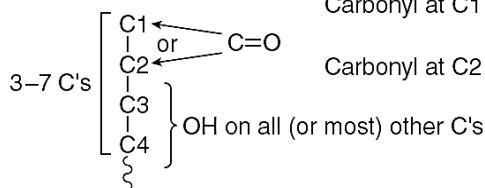
Carbohydrates are classified into two classes: *simple carbohydrates* and *complex carbohydrates*. Simple carbohydrates contain one sugar unit and they cannot be cleaved by hydrolysis to smaller carbohydrates. These are called **monosaccharides** (Latin *saccharum*, Greek *Sakcharon* for sugar). Complex carbohydrates possess two or more sugar units and they can easily be hydrolyzed to simple carbohydrates. Carbohydrates that undergo hydrolysis producing 2 molecules of monosaccharides are called **disaccharides**; those producing three molecules of monosaccharides are termed as **trisaccharides**; and so on. The term **oligosaccharide** (*oligos* is a Greek word, means “few”) is generally used for carbohydrates yielding 2-10 units of monosaccharides on hydrolysis. Carbohydrates yielding larger numbers of monosaccharide units are called **polysaccharides**.

5.1.1. Monosaccharides

Monosaccharides have the general formula $C_nH_{2n}O_n$ containing a chain of three to eight carbon atoms - one of the carbons being the carbonyl group of either an aldehyde or a ketone and the rest attached to hydroxyl groups. They are named according to

- (1) *the number of carbon atoms present in the molecule*. A monosaccharide containing three carbon atoms is called a *triose*; one containing four carbon atoms is called a *tetrose*; one containing five carbon atoms is a *pentose*; and one containing six carbon atoms is a *hexose*, and so on.
- (2) *whether they contain an aldehyde or keto group*. Polyhydroxyaldehydes are called **aldoses** (*ald-* is for *aldehyde* and *-ose* is the suffix for a sugar), and polyhydroxyketones are called **ketoses** (*ket-* for *ketone*, and *-ose* for sugar).

General structure of a monosaccharide



Carbonyl at C1 \longrightarrow aldehyde \longrightarrow

aldose

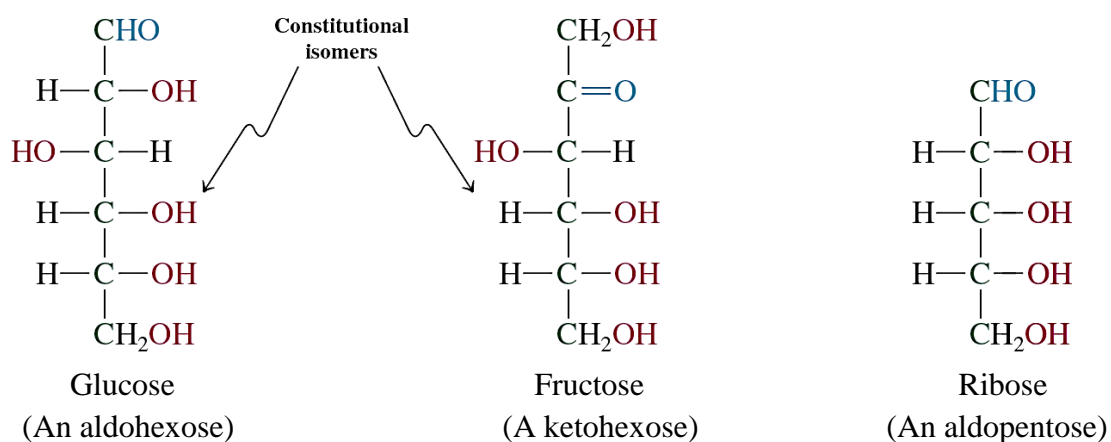
Carbonyl at C2 \longrightarrow ketone \longrightarrow

ketose

Both terminology are used in combination:

Number of Carbon atoms	Aldose	Ketose
Three	Aldotriose	Ketotriose
Four	Aldotetrose	Ketotetrose
Five	Aldopentose	Ketopentose
Six	Aldohexose	Ketohexose
Seven	Aldoheptose	Ketoheptose
Eight	Aldooctose	Ketooctose

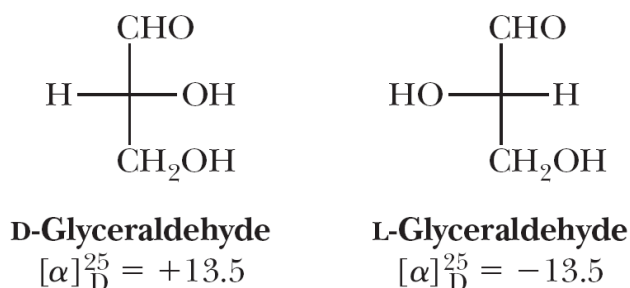
Glucose, also known as dextrose, blood sugar, or grape sugar (*glykys*, Greek, sweet), is a pentahydroxyhexanal and hence belongs to the class of aldohexoses. It occurs naturally in many fruits and plants and in concentrations ranging from 0.08 to 0.1% in human blood. A corresponding isomeric ketohexose is **fructose**, the sweetest natural sugar, which also is present in many fruits (*fructus*, Latin, fruit) and in honey. Another important natural sugar is the aldopentose **ribose**, a building block of the ribonucleic acids.



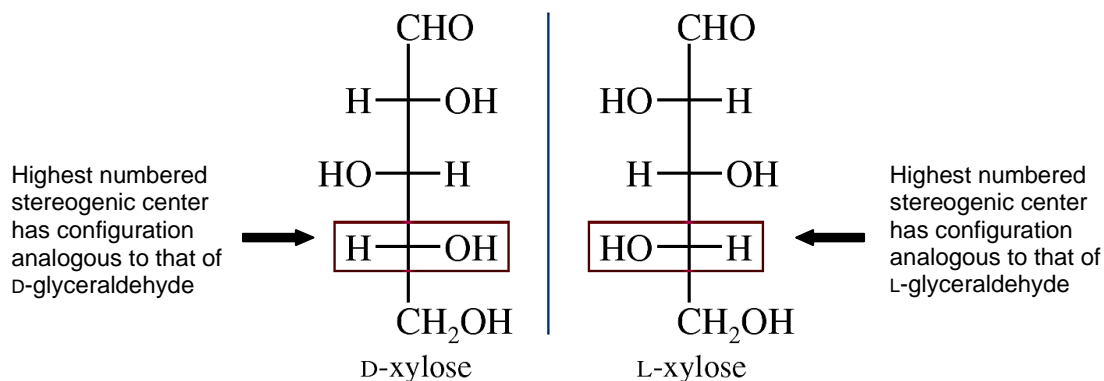
Monosaccharides are colorless, crystalline solids, although they often crystallize with difficulty. Because of hydrogen bonding between their polar $-\text{OH}$ groups and water, all monosaccharides are very soluble in water. They are only slightly soluble in ethanol and are insoluble in nonpolar solvents such as diethyl ether, chloroform, and benzene.

5.1.1.1. Relative configuration of monosaccharides (D- and L-notation)

The simplest chiral monosaccharide is glyceraldehyde, an aldotriose. It contains an asymmetric carbon and hence exists in two enantiomeric forms. The enantiomer having $-\text{OH}$ on the right side in Fischer projection is referred to as D-glyceraldehyde while one with the $-\text{OH}$ on left side as L-glyceraldehyde. D-Glyceraldehyde is (R)-(+)-glyceraldehyde and L-glyceraldehyde is (S)-(-)-glyceraldehyde. Almost all sugars found in nature are D-sugars. D- and L-configurations are mirror image of each other.

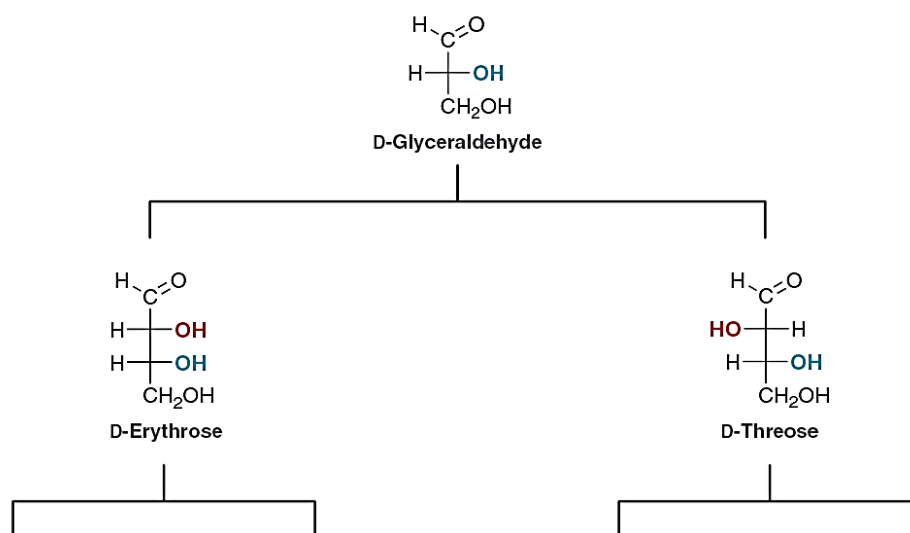


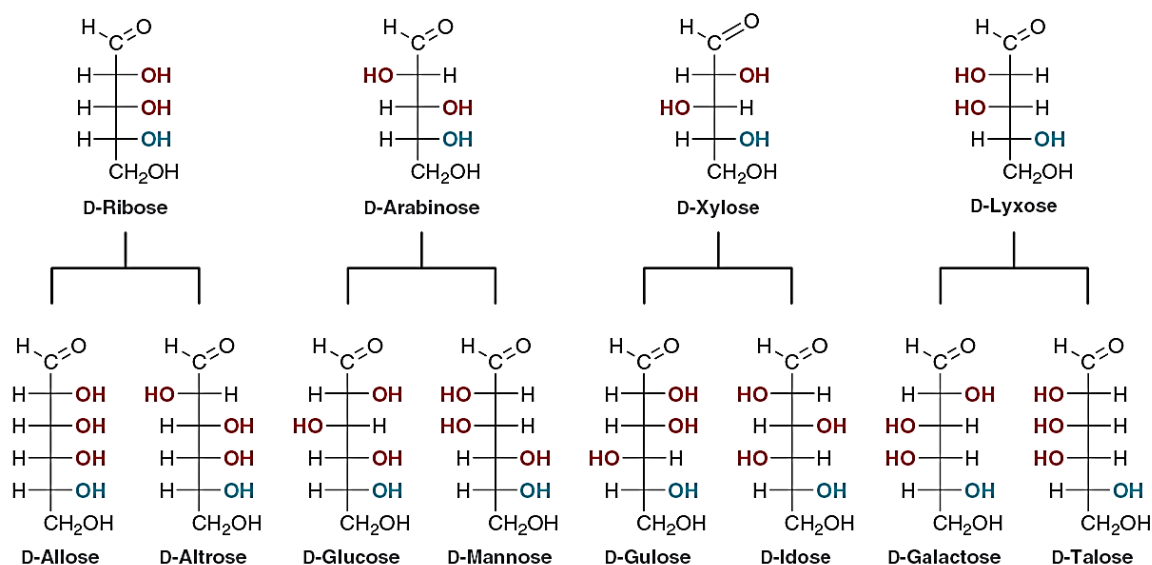
D- and L-glyceraldehyde serve as reference points for the assignment of relative configuration to all other aldoses and ketoses. For monosaccharides with two or more chiral centres, the designations D or L refer only to the configuration of the highest-numbered chiral centre that is, the chiral centre farthest from the aldehyde or ketone carbonyl group.



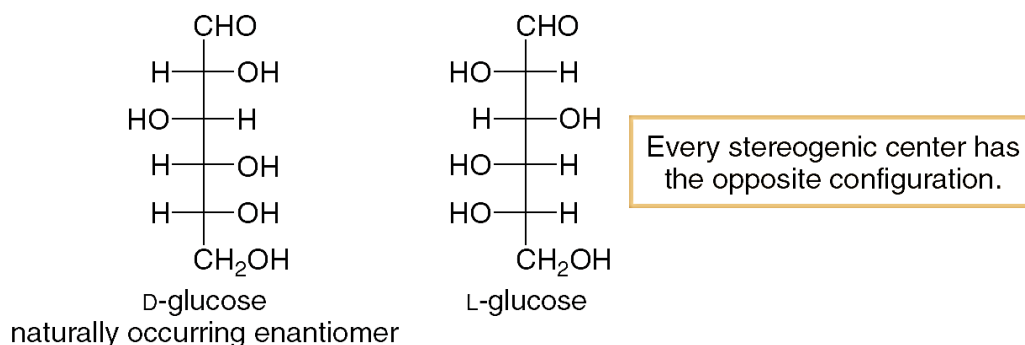
The D or L designation of a given monosaccharide does not specify the sign of the specific rotation of the compound. The sign of the rotation of plane polarized light is indicated by a + (plus) or a – (minus) sign preceding the name. For example, D-glucose, which is dextrorotatory, is designated D-(+)-glucose, and D-fructose, which is levorotatory, is designated D-(–)-fructose.

Aldotetroses have two asymmetric carbons and therefore four stereoisomers (2^n , where n is the number of asymmetric carbons). Two of the stereoisomers are D-sugars and two are L-sugars. Aldopentoses have three asymmetric carbons and therefore eight stereoisomers (four pairs of enantiomers), while aldohexoses have four asymmetric carbons and 16 stereoisomers (eight pairs of enantiomers). The family tree of D aldoses can be generated by starting with D-(+)-glyceraldehyde and adding another carbon at the top to generate two aldotetroses *erythrose* with the OH group of the new asymmetric carbon on the right, and *threose* with the new OH group on the left. Adding another carbon to these aldotetroses gives four aldopentoses, and adding a sixth carbon gives eight aldohexoses.

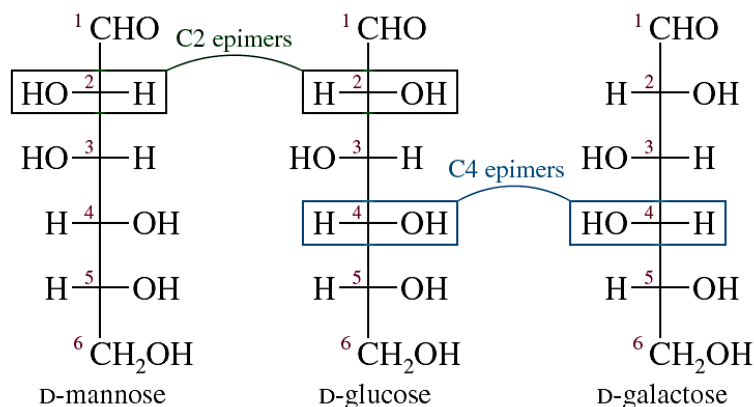




The L-aldoses are mirror images of these D-series of aldoses. For example, D-glucose and L-glucose are mirror images – every stereogenic center has the opposite configuration.

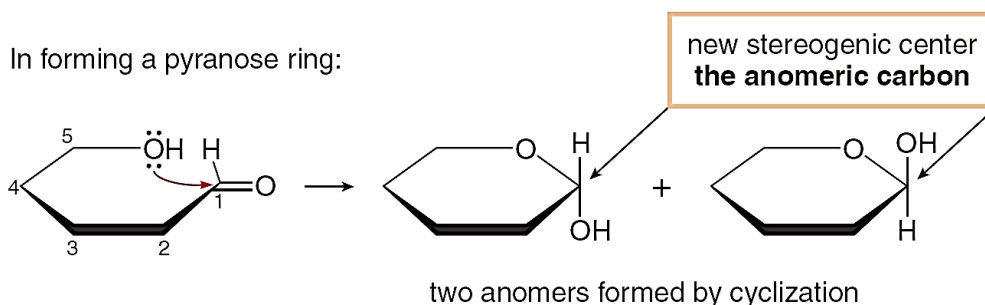


Stereoisomers that are mirror images are called enantiomers (e.g., D-glucose and L-glucose) while that are not mirror images are called diastereomers. The diastereomers that differ in the configuration around one stereogenic center only are called epimers. For example, D-glucose and D-mannose differ in configuration at C-2 only, so they are C2 epimers. Similarly, D-glucose and D-galactose are C4 epimers (differing in configuration at C-4 only).



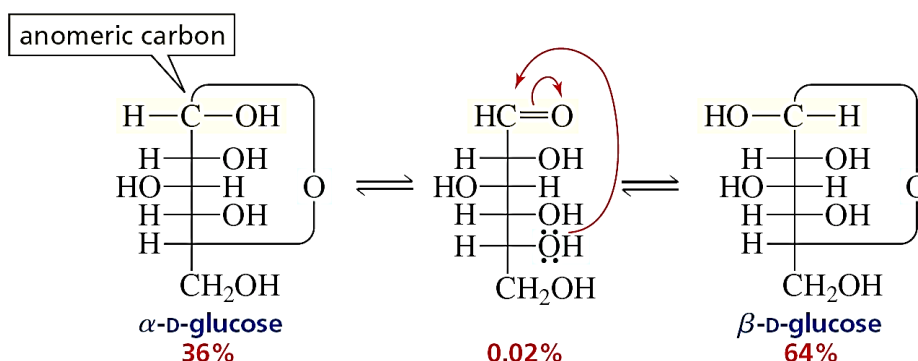
5.1.1.2. Cyclic structures of monosaccharides (Anomers)

We have already seen in unit-3 that carbonyl compounds react with alcohols and form hemiacetals. If these two groups ($>\text{C}=\text{O}$ and $-\text{OH}$) are present in the same molecule, they yield cyclic hemiacetals, preferably making five- or six-membered ring. Monosaccharides have hydroxyl and carbonyl groups in the same molecule. As a result, they too exist almost exclusively as five- and six-membered cyclic hemiacetals. Five-membered-ring sugars are called furanose (derived from furan) while six-membered-ring sugars are called pyranose (derived from pyran). Two different hemiacetals are formed because the carbonyl carbon of the open-chain sugar becomes a new asymmetric carbon in the hemiacetal. These are called anomers and the newly formed asymmetric carbon is called anomeric carbon.



Thus, anomers are stereoisomers of two cyclic sugars that differ in configuration only at the carbon that was the carbonyl carbon in the open-chain structure, i.e. the upper-most asymmetric carbon (*Ano* is Greek for “upper”). If the OH group bonded to the new asymmetric carbon is on the right in Fischer projection, the hemiacetal is α -anomer; if the OH group is on the left, the hemiacetal is β -anomer.

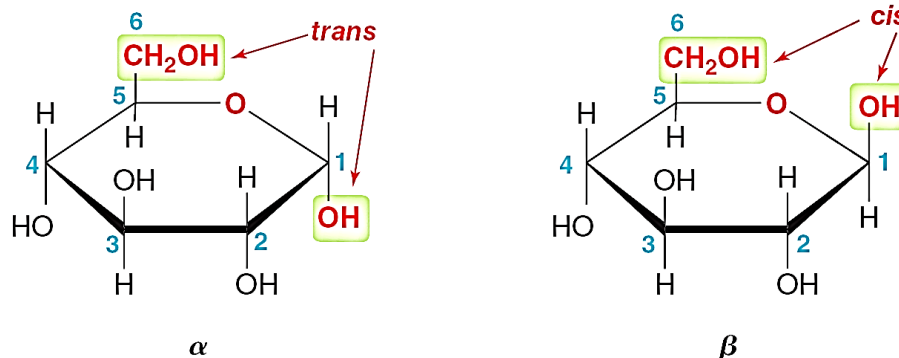
For example, glucose on cyclisation gives two compounds – α -D-glucose or α -D-glucopyranose (melting point 146°C and specific rotation $+112.2^\circ$), and β -D-glucose or β -D-glucopyranose (melting point 150°C and specific rotation $+18.7^\circ$). Thus, glucose exists in three different forms in equilibrium.



(i) Haworth Projections

Fischer projections are not the best way to show the structure of a cyclic sugar. A better way of depicting the hemiacetal forms of sugars is Haworth projection. The conventional way of writing the Haworth structure for a six-membered ring is to show the oxygen

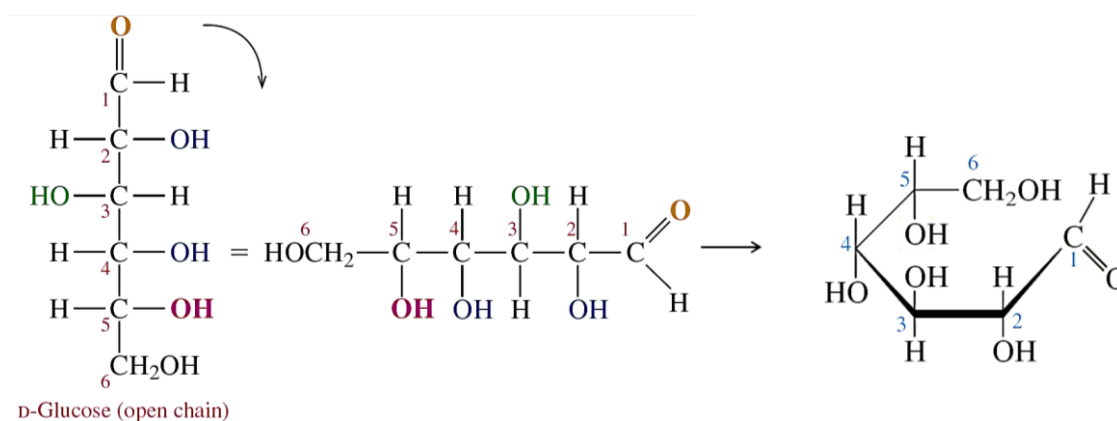
atom of the ring in the back right hand corner of hexagon, with the anomeric carbon (C-1) on the right-hand side and the primary -OH group drawn *up* from the back left-hand corner (C-5). All groups written on the *right* in the Fischer projection are written *down* in a Haworth projection, whereas groups written on the *left* are drawn *up* in a Haworth projection. The aldehydic carbon (C1) becomes a chirality center, and both configurations for that new chirality center are possible. In the **α -anomer**, the hydroxyl group at the anomeric position is *trans* to the CH_2OH group, while in the **β -anomer**, the hydroxyl group is *cis* to the CH_2OH group.



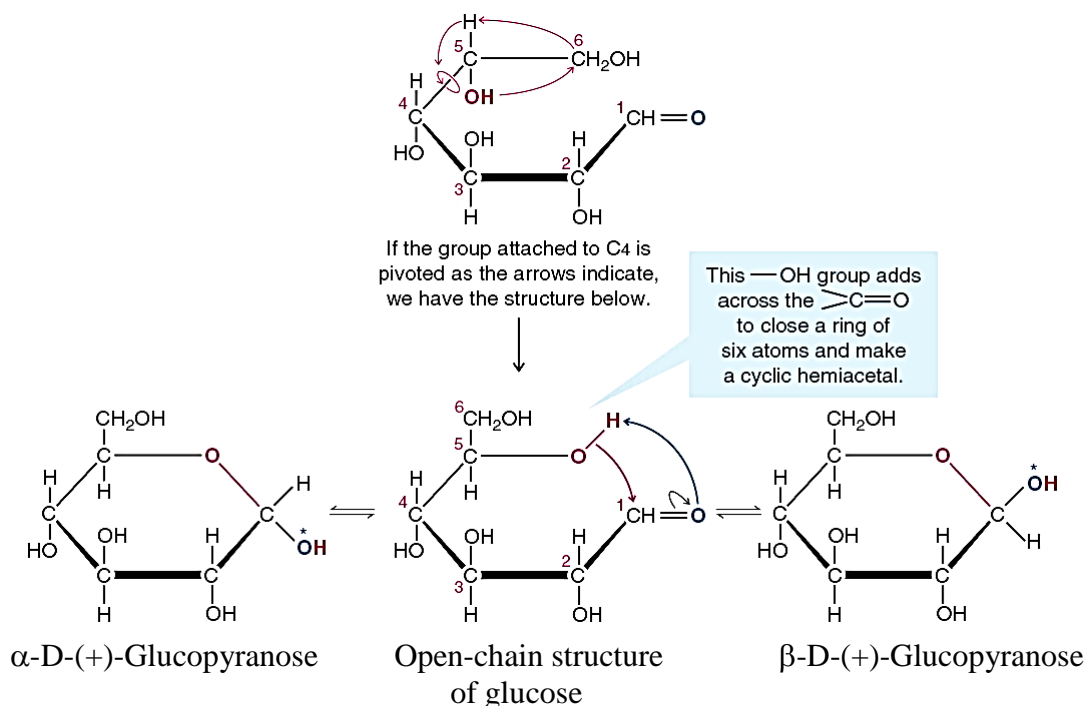
Drawing Haworth structure

Step 1: Turn the open-chain structure clockwise 90° . Now all the groups which were on right side will be *down*, and the groups on left side will be *up*.

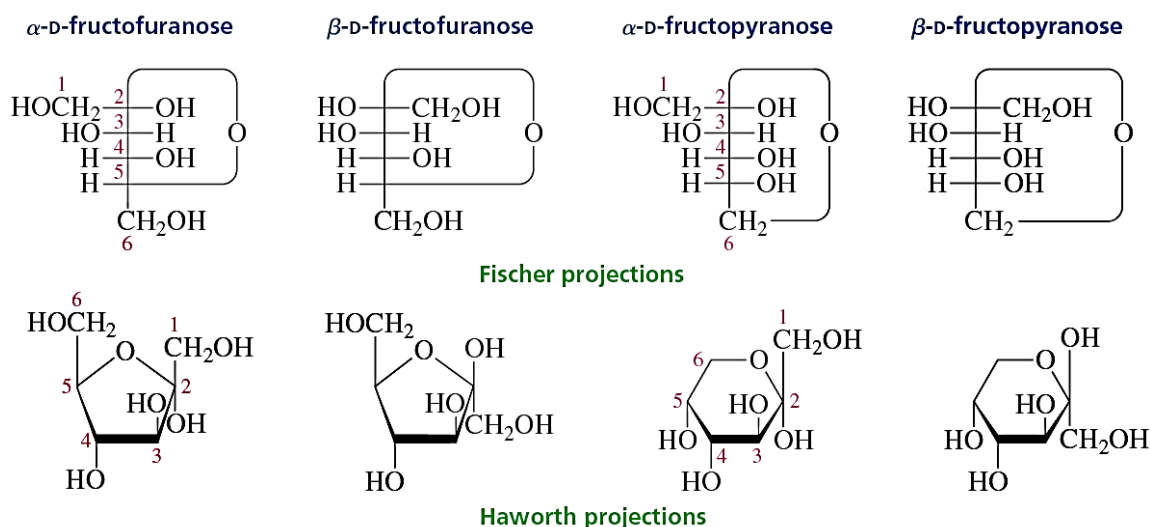
Step 2: Fold the carbon chain to make a hexagon. Folding of the $\text{-CH}_2\text{OH}$ end of the structure back side does not place the -OH group at C-5 in a proper orientation to make hemiacetal with aldehyde group.



Step 3: Rotate about C-4–C-5 bond in anticlockwise direction. Anticlockwise rotation about C-4–C-5 bond will change the position of all the three groups at C-5 so that the hydroxyl group will come closer to aldehyde group for ring formation. Finally, nucleophilic attack of -OH group on carbonyl carbon will result in the formation of hemiacetal. The new -OH group at C-1 is drawn *down* for α -anomer and *up* for β -anomer.

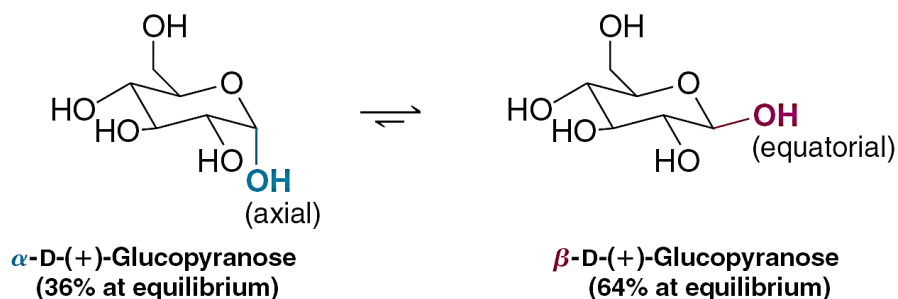


Ketoses also exist predominantly in cyclic forms. D-Fructose forms a five-membered-ring-hemiketal as a consequence of the C-5 OH group reacting with the ketone carbonyl group. If the new asymmetric carbon has the OH group on the right in a Fischer projection, the compound is α -D-fructose (or α -D-fructofuranose); if the OH group is on the left, the compound is β -D-fructose (or β -D-fructofuranose). In fructose the anomeric carbon is C-2, not C-1 as in aldoses. D-Fructose can also form a six-membered ring by using the C-6 OH group. Therefore, D-fructose exists as an equilibrium between an open-chain form, two pyranose forms (α and β), and two furanose forms (α and β). The pyranose form predominates in the monosaccharide, whereas the furanose form predominates when the sugar is part of a disaccharide.



(ii) Conformation structure

Haworth projections are useful to easily see the *cis* or *trans* configuration of the hydroxyl groups. A five-membered ring is so close to being planar that Haworth projections are adequate representations of furanoses. For pyranoses, however, the six-membered ring is more accurately represented as a chair conformation. X-ray analysis of glucose also confirmed that the actual conformation of the cyclic forms of glucose is the chair forms.

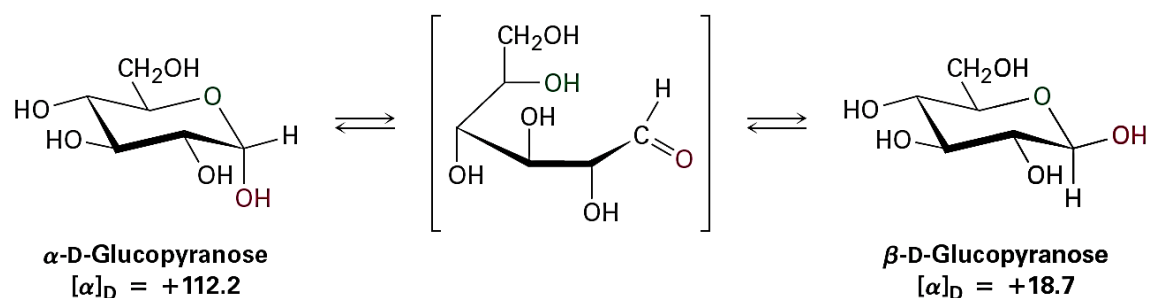


By convention the chair is drawn in the way that the back of the chair is on the left and the footrest is on the right. Then the oxygen atom is placed at the rear-right position, as shown in the figure above. The hydroxyl groups *up* in the Haworth projection remain *up* (whether *equatorial* or *axial*) and *down* will remain *down* in chair form. In D-sugars, the terminal $-\text{CH}_2\text{OH}$ group is on the top of the ring, whereas for L-sugars, the $-\text{CH}_2\text{OH}$ group is on the bottom of the ring. Notice that in a cyclic form of D-glucose all substituents ($-\text{OH}$ and $-\text{CH}_2\text{OH}$) larger than a hydrogen atom occupy equatorial positions making it more stable because large substituents are more stable in the equatorial position due to less steric strain in that position. This explains why D-glucose is the most common naturally occurring monosaccharide. Moreover, the $-\text{OH}$ group at C-1 (anomeric carbon) is in the equatorial position in β -anomer while it is in the axial position in α -anomer making β -anomer more stable than the α -anomer. Therefore, β -anomer predominates at equilibrium in an aqueous solution.

Mutarotation

Because anomers are diastereomers, they generally have different properties. For example, α -D-glucopyranose has a melting point of 146°C and a specific rotation of 112.2° while β -D-glucopyranose has a melting point of 150°C and a specific rotation of 18.7° . When glucose is crystallized from water at room temperature, pure crystalline α -D-glucopyranose is obtained. If glucose is crystallized from water by letting the water evaporate at a temperature above 98°C , crystals of pure β -D-glucopyranose are formed. When a sample of either pure anomer is dissolved in water, however, its optical rotation slowly changes until it reaches a constant value of $+52.6^\circ$. That is, the specific rotation of the α -anomer solution decreases from $+112.2^\circ$ to $+52.6^\circ$, and the specific rotation of the β -anomer solution increases from $+18.7^\circ$ to $+52.6^\circ$. This change (“mutation”) in the specific rotation is called **mutarotation**. Mutarotation occurs because the two anomers interconvert in solution to give an equilibrium mixture having 36% of α -anomer and 64% of β -anomer with traces of open-chain structure.

Mutarotation occurs by a reversible ring-opening of each anomer to the open-chain aldehyde, followed by re-closure. Although the equilibration is slow at neutral pH, it is catalyzed by both acid and base.



When D-fructose is dissolved in water, the following equilibrium concentrations are observed: 70% β -pyranose, 2% α -pyranose, 23% β -furanose, 5% α -furanose, and 0.7% open chain.

5.1.1.3. Chemical Reactions of Glucose

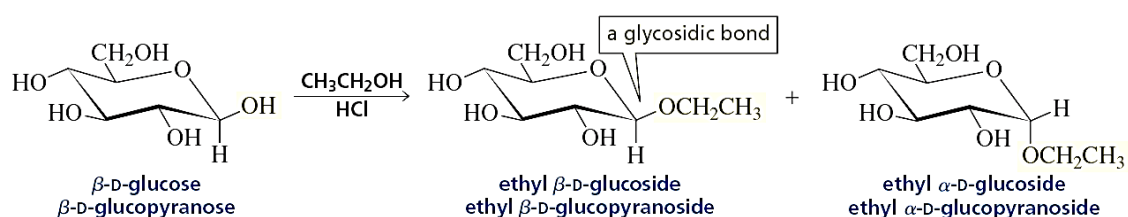
Because monosaccharides exist in solution in an equilibrium between acyclic and cyclic forms, they undergo three types of reactions:

- Reaction of the hemiacetal
- Reaction of the hydroxy groups
- Reaction of the carbonyl group

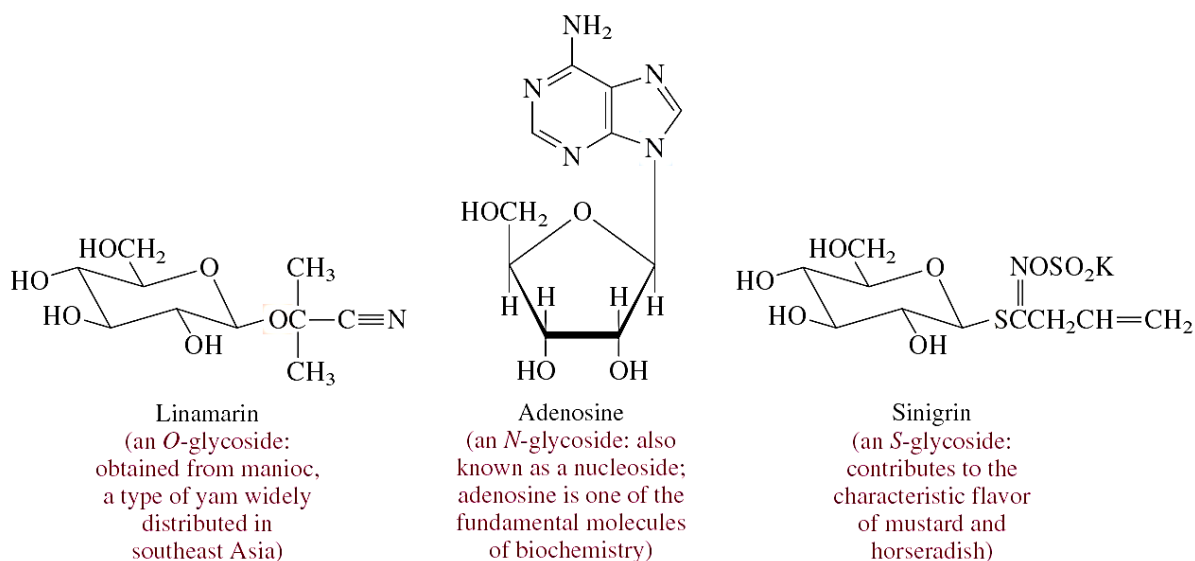
Glycosides

Glycosides are a large and very important class of carbohydrate derivatives characterized by the replacement of the anomeric hydroxyl group by some other substituent. We have already discussed earlier in unit-3 that the treatment of an aldehyde or ketone with one molecule of alcohol gives a hemiacetal and that treatment of the hemiacetal with another molecule of alcohol gives an acetal. Similarly, treatment of monosaccharides, all of which exist almost exclusively in a cyclic hemiacetal form, also gives acetals. Monosaccharides in the form of cyclic acetals are called **glycosides**, and the bond from the anomeric carbon to the $-OR$ group is called a **glycosidic bond**.

For example, treatment of α -D-glucose with CH_3OH and HCl forms two glycosides that are diastereomers at the acetal carbon. The α and β labels are assigned in the same way as anomers: with a D-sugar, an α -glycoside has the new OR group (OCH_2CH_3 group in this example) down, and a β -glycoside has the new OR group up. During the glycoside formation, only the anomeric hydroxyl is replaced.



Glycosides are named by listing the alkyl or aryl group bonded to oxygen followed by the name of the carbohydrate in which the ending *-e* is replaced by *-ide*. For example, the glycosides derived from β -D-glucopyranose are named β -D-glucopyranosides; those derived from β -D-ribofuranose are named β -D-ribofuranosides. They are termed *O*-glycosides, *N*-glycosides, *S*-glycosides, and so on, according to the atom in the substituent that replaces the anomeric hydroxyl group. For example, just as the anomeric carbon of a cyclic hemiacetal undergoes reaction with the -OH group of an alcohol to form a glycoside, it also undergoes reaction with the N-H group of an amine to form an *N*-glycoside.

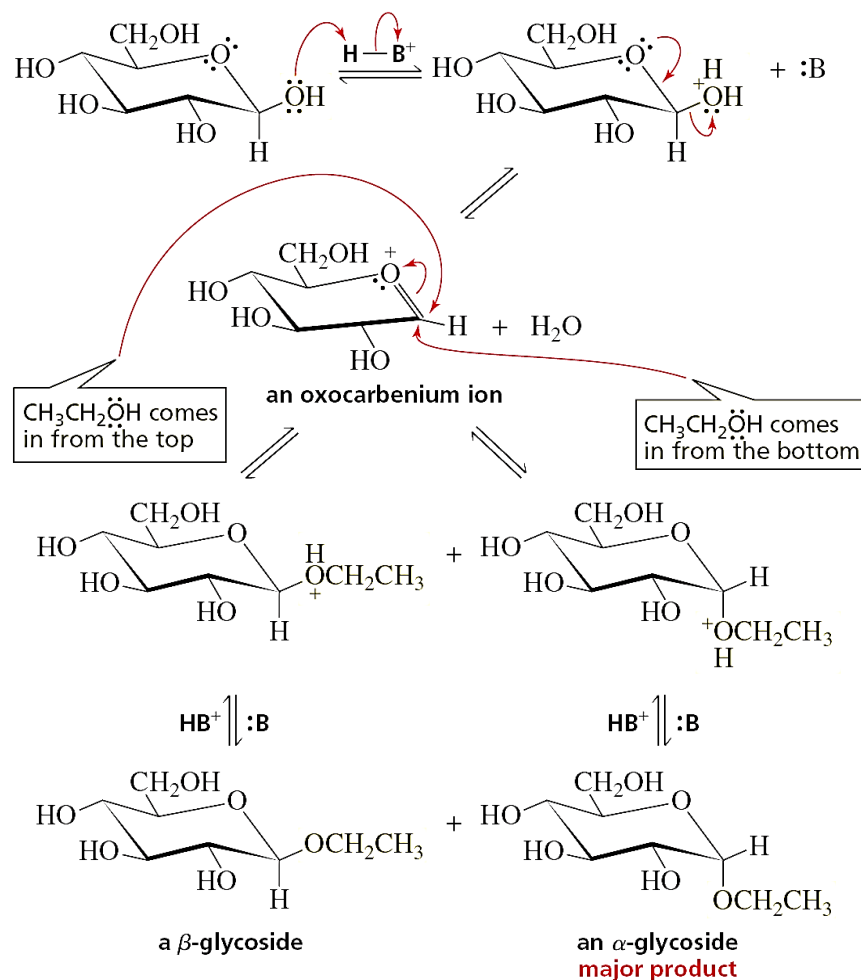


Aldoses and ketoses (hemiacetals) are in equilibrium with open chain structures, so they can mutarotate. However, mutarotation is not possible in a glycoside because an acetal is no longer in equilibrium with the open-chain carbonyl-containing compound. Therefore, aldoses and ketoses (hemiacetals) are reducing sugars while glycosides (acetals) are non-reducing sugars. Glycosides are stable in water and aqueous base, but like other acetals, they are hydrolyzed in aqueous acid back to an alcohol and a monosaccharide.

Mechanism:

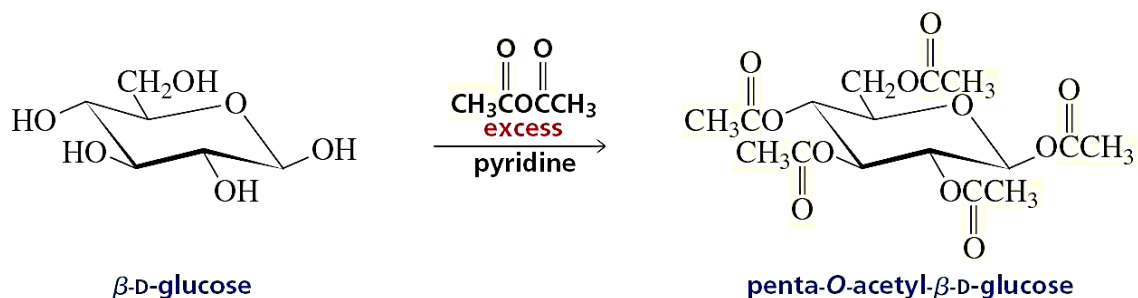
The mechanism of glycoside formation is analogous to the mechanism of acetal formation with carbonyl compounds discussed in unit 3. First, the anomeric hydroxyl group is protonated in acidic medium to generate a good leaving group, which is lost as water to yield a resonance-stabilized carbocation. This stabilization only occurs when the reaction takes place at the anomeric position because the positive charge is shared by a carbon and oxygen. The carbocation is then attacked by the alcohol, followed by loss of a proton to produce the glycoside.

As the carbocation is sp^2 hybridized, that part of the molecule is planar, and the nucleophile (alcohol) can attack either from the top of the plane to produce β -glycoside, or from the bottom of the plane to give α -glycoside.



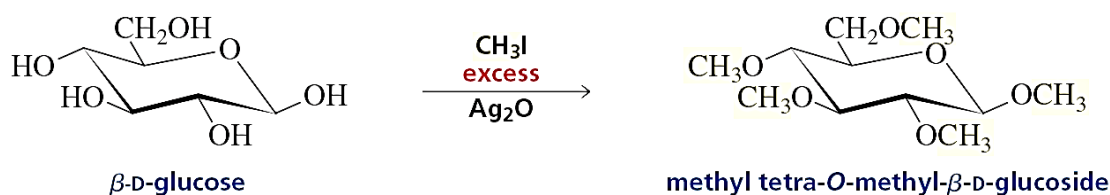
Acylation and alkylation (formation of ester and ether)

Monosaccharides are highly soluble in water (due to the presence of many hydroxyl groups) and are generally insoluble in most organic solvents. This property makes them difficult to purify by conventional methods. However, the ester and ether derivatives of monosaccharides are soluble in most organic solvents and are easily purified and crystallized. The $-\text{OH}$ groups of monosaccharides show the typical reactions of alcohols. Monosaccharides are converted into their ester derivatives when treated with an excess of acid chloride or acid anhydride in the presence of a weak base, such as pyridine or sodium acetate. Under these conditions, all five hydroxyl groups, including the anomeric hydroxyl, of β -D-glucopyranose are converted into ester groups.



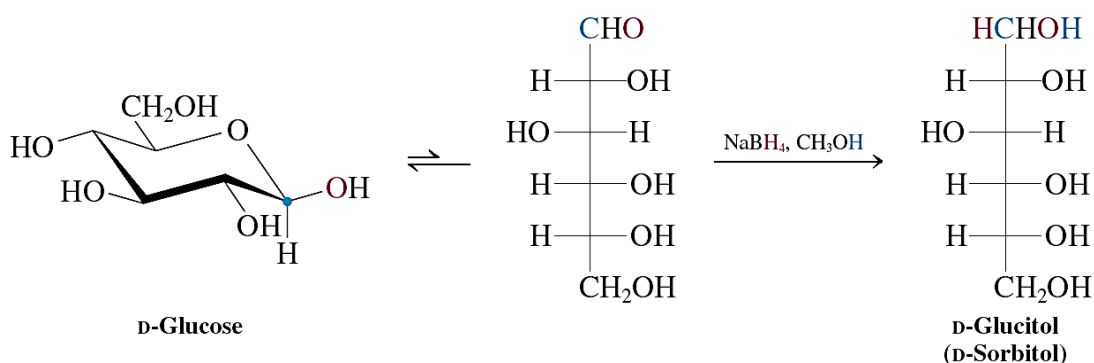
If the reaction is carried out at a low temperature (e.g., 0°C), the reaction occurs stereospecifically – the α -anomer gives the α -acetate and the β -anomer gives the β -acetate.

Similarly, monosaccharides react with methyl iodide/silver oxide to form ethers. The OH group is a relatively poor nucleophile, so silver oxide is used to increase the leaving tendency of the iodide ion in the S_N2 reaction. Under these conditions, all five hydroxyl groups of β -D-glucopyranose are converted into ether. Because no C–O bonds are broken, the configuration of all substituents groups in the starting material is **retained**, forming a single product.



Reduction (formation of alditol)

The carbonyl group of monosaccharides shows typical reactions of aliphatic carbonyl compounds. The reduction of carbonyl group of aldoses and ketoses with NaBH_4 or catalytic hydrogenation yields alditols; the carbonyl group is reduced to an alcohol. Reduction of an aldose forms one alditol, while reduction of a ketose forms two alditols because the reaction creates a new asymmetric carbon in the product, resulting in the formation of two epimers. Alditols are named by adding the suffix *-itol* to the root name of the sugar. For example, D-glucose is reduced to D-glucitol, more commonly known as D-sorbitol, when treated with NaBH_4 .

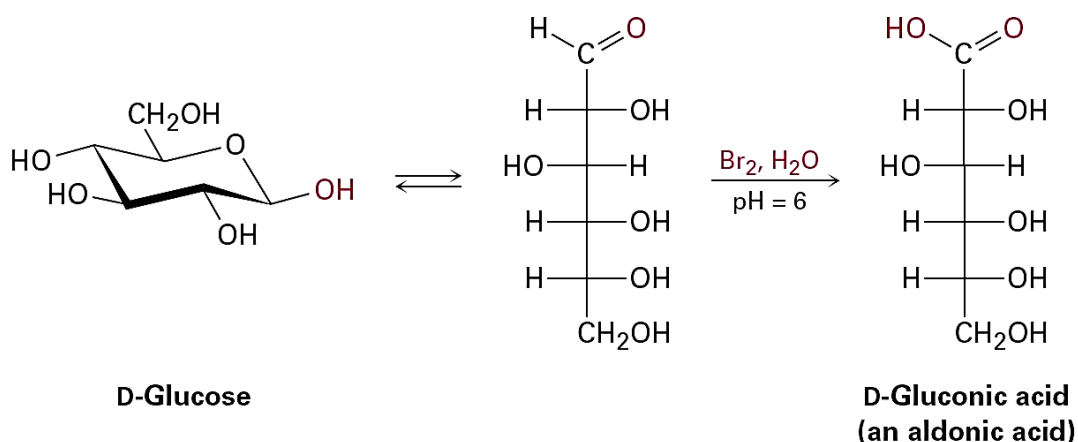


Although sugars exist mostly as cyclic hemiacetals in aqueous solution, they are in rapid equilibrium with their open-chain forms. As the small amount of open-chain form of the sugar is reduced, the equilibrium shifts from the unreactive cyclic hemiacetal to the product and consequently, the entire sugar is reduced to alditol.

Oxidation

(a) With bromine water.

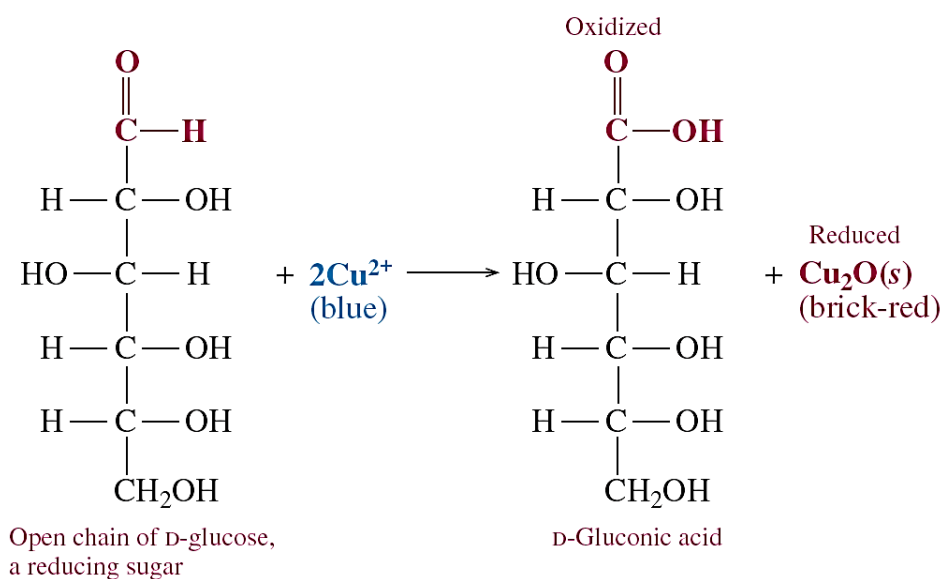
Bromine water oxidizes the aldehyde group of an aldose to a carboxylic acid, resulting in the formation of aldonic acid.



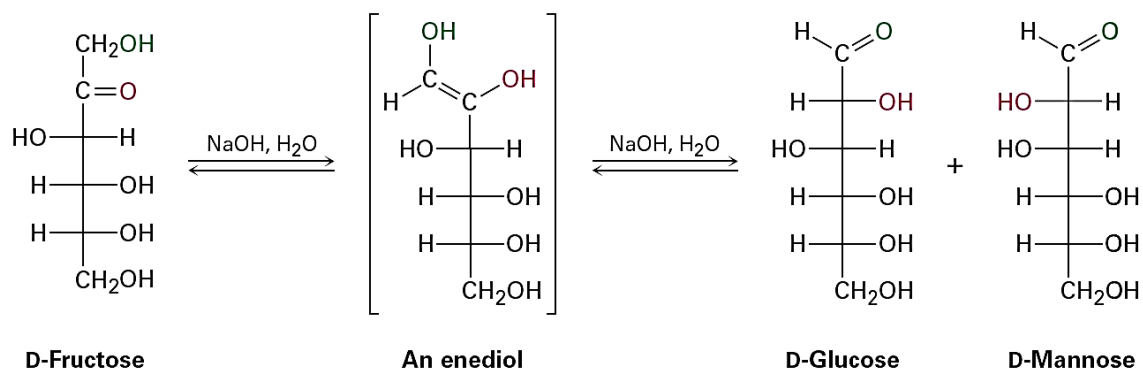
The reaction is used to distinguish aldoses from ketoses as bromine water oxidizes aldoses but not ketoses. To test aldoses, a small amount of bromine water is added to an unknown monosaccharide. Disappearance of the reddish-brown colour of bromine water confirms the sugar is aldose. If the colour persists, the sugar is a ketose.

(b) With Tollens reagent, Fehling's solution, and Benedict reagent.

Tollens reagent (Ag^+ in aqueous ammonia), Fehling's solution (Cu^{2+} in aqueous sodium tartrate), and Benedict reagents (Cu^{2+} in aqueous sodium citrate) are used to test reducing sugars. Tollens reagent oxidizes aldehydes to carboxylic acids using Ag_2O in NH_4OH , and forms a mirror of Ag as a by-product. **Benedict's** and **Fehling's reagents** use a blue Cu^{2+} salt as an oxidizing agent which is reduced to Cu_2O , a brick-red solid. All monosaccharides including aldoses and ketoses are oxidized with these reagents to aldonic acids. Aldoses contain aldehyde group which reduces these reagents and in turn, aldoses are oxidized to aldonic acid.



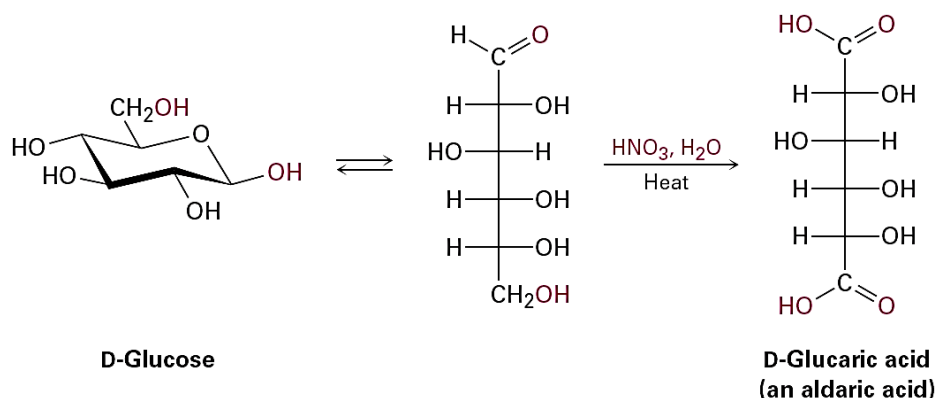
Ketoses also give positive tests with these reagents because under the basic conditions of the reaction, ketoses (α -hydroxy alcohols) are in equilibrium with aldoses by way of an enediol intermediate. The aldoses are then oxidized by these mild oxidizing agents.



- Sugars that give positive tests with these reagents are known as **reducing sugars**, and all carbohydrates that contain a *hemiacetal group* give positive tests because in aqueous solution the hemiacetal form of sugars exists in equilibrium with small concentration of open-chain aldehydes or α -hydroxy ketones. So as long as a sugar has an aldehyde, a ketone, a hemiacetal, or a hemiketal group, it is able to reduce an oxidizing agent and therefore is classified as **reducing sugar**. Without one of these groups, it is a nonreducing sugar. **All monosaccharides are reducing sugars.**
- Glycosides are acetals and they are not in equilibrium with the open-chain aldehyde (or ketone) in neutral or basic solution. Therefore, they cannot reduce Tollens reagent, Fehling's solution or Benedict reagent. Thus the carbohydrates that contain acetal group (called glycosides) are called *nonreducing sugars*.

(c) With concentrated nitric acid.

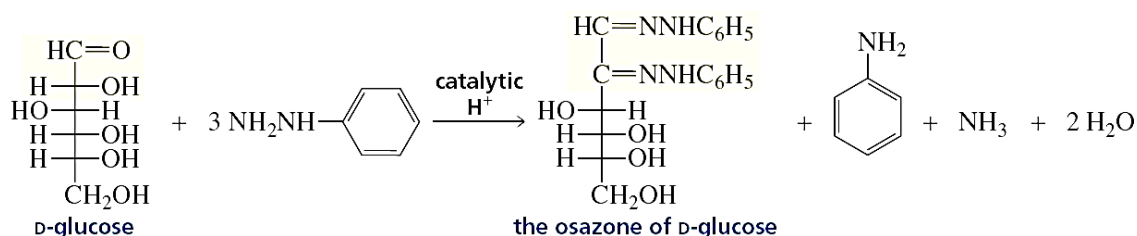
Nitric acid is strong oxidizing agent, which oxidizes not only the aldehyde group but also 1° alcohol of sugars to carboxylic acid to produce aldarcic acid. For example, D-glucose is oxidized to glucaric acid, if warm dilute nitric acid is used as oxidizing agent.



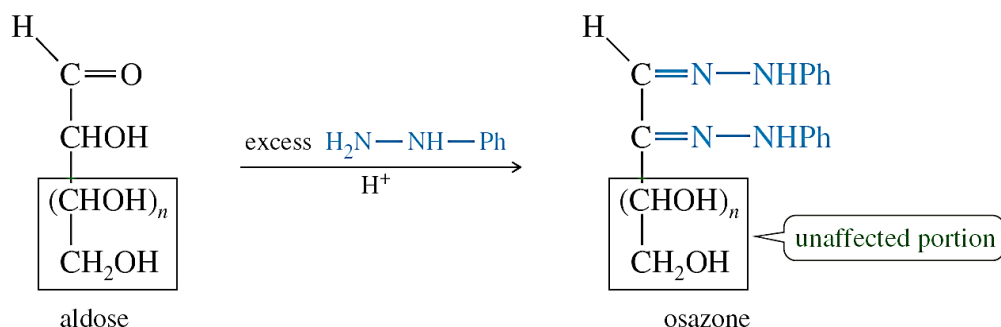
Osazone formation (reaction with phenylhydrazine)

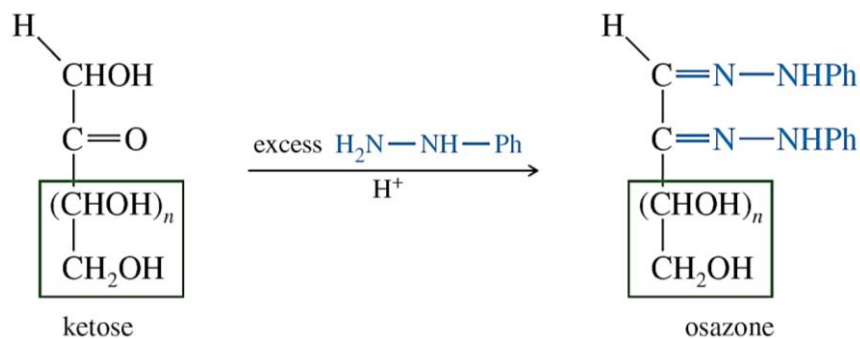
Carbohydrates make syrups that do not crystallize; therefore the purification and isolation of carbohydrates are difficult. Phenylhydrazine derivatives of the sugars, that are called osazone, are yellow crystalline solid that are insoluble in water and have sharp melting point. Thus the sugars can easily be isolated and purified by converting them to their osazones. Melting points of osazone derivatives provide valuable clues for the identification and comparison of sugars.

As discussed earlier in Unit 3 that aldehydes and ketones react with one equivalent of phenylhydrazine to yield phenylhydrazone. However, aldoses and ketoses react with three molecules of phenylhydrazine, resulting in the formation of osazones. One molecule of phenylhydrazine behaves as oxidizing agent and is reduced to aniline and ammonia, and other two molecules make imine with carbonyl groups.



In the formation of an osazone, both C1 and C2 are converted to phenylhydrazones. Therefore, a ketose gives the same osazone as its related aldose. The stereochemistry at C2 is lost in the phenylhydrazone, therefore, C2 epimers give the same osazone.





5.1.2. Disaccharides

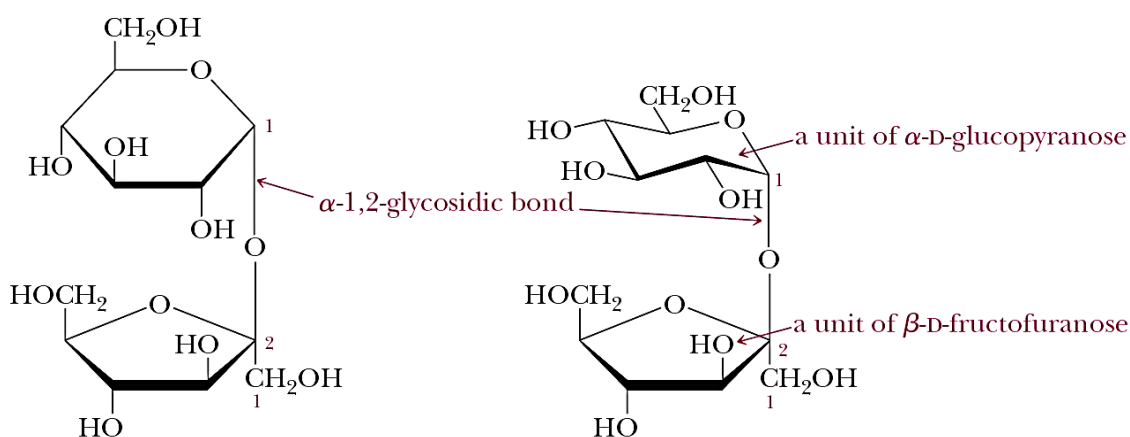
Disaccharides yield two same or different molecules of monosaccharides on hydrolysis. The general formula of disaccharides is $C_{12}H_{22}O_{11}$ that is sum of two monosaccharide units minus water molecule.

- The two monosaccharide units, either five-membered or six-membered, are joined through glycoside linkage between the anomeric carbon of one sugar and the $-OH$ group at any position on the other sugar.
- Glycosidic linkage may be α -oriented (drawn downwards) or β -oriented (drawn upwards).
- All disaccharides have **one acetal**, together with either a hemiacetal or another acetal.
- Disaccharides having hemiacetal group are reducing sugars while with acetal group are nonreducing sugars.
- The numbering starts with the anomeric carbon and goes on clockwise.
- A glycosidic linkage between C1 of the first sugar and the $-OH$ at C4 of the second sugar is particularly more common. Such a bond is called a 1→4 link.

The most common disaccharides are sucrose, maltose and lactose.

Sucrose.

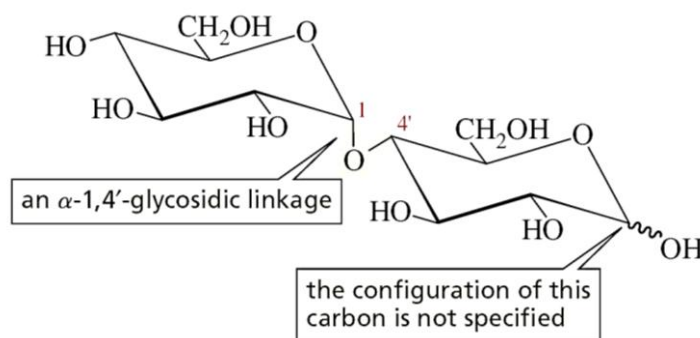
Sucrose is the most common disaccharide, also known as table sugar. It is present in varying concentrations in various fruits, flowers, seeds, and roots. Commercially it is produced from the sugar cane or beet roots. On hydrolysis sucrose gives a molecule of D-glucose and D-fructose each. Carbon 1 of α -D-glucopyranose (six-membered) is joined to carbon 2 of β -D-fructofuranose (five-membered) by an α -1,2-glycosidic bond to make sucrose molecule.



Because the anomeric carbons of both the glucopyranose and fructofuranose units are involved in formation of the glycosidic bond, sucrose is a non-reducing sugar and does not undergo mutarotation.

Maltose.

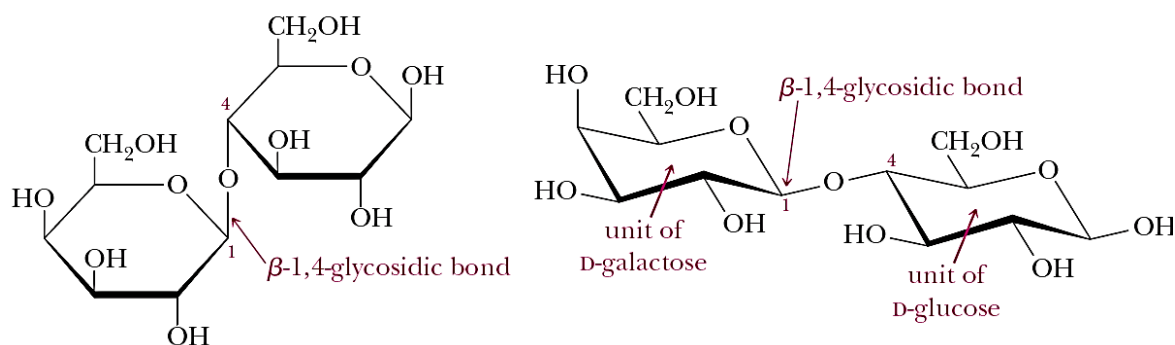
Maltose (or malt sugar) yields two units of glucose on hydrolysis. C-1 (anomeric carbon) of one unit of glucose is attached to C-4 of the other glucose unit by α -1,4-glycosidic linkage (called 1 \rightarrow 4 bond).



The -OH at anomeric carbon of right hand side unit may be α - (down or axial) or β - (up or equatorial) to give α -maltose and β -maltose, respectively. Maltose is a reducing sugar because the hemiacetal group on the right unit of D-glucopyranose is in equilibrium with the free aldehyde and can be oxidized to a carboxylic acid. Because it exists in both α - and β -forms, it undergoes mutarotation when dissolved in a solvent.

Lactose.

Lactose, milk sugar, is found in milk and milk products. It yields a molecule of D-glucose (as β -D-glucopyranose) and D-galactose (as β -D-galactopyranose) each. In lactose, C-1 of galactose joined to C-4 of D-glucose through β -1,4-glycosidic linkage.



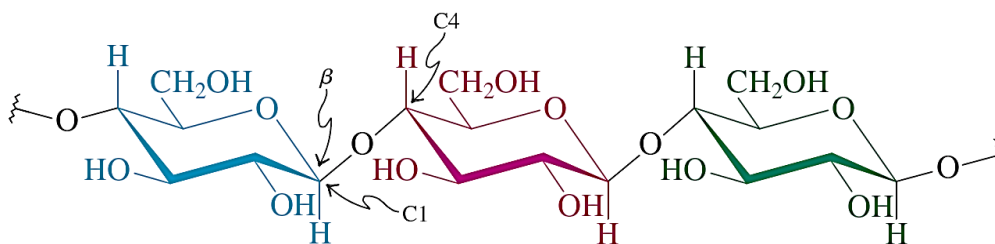
Lactose exists in two isomeric forms: the -OH at C-1 (anomeric carbon) of D-glucose may be up (β -isomer) or down (α -isomer). Because anomeric carbon of D-glucose is hemiacetal, it can form an open chain structure, making lactose a reducing sugar. Lactose undergoes mutarotation.

5.1.3. Polysaccharides

The polysaccharides, also known as glycans, are polymers of more than ten (even up to thousands) monosaccharides joined together through glycoside linkage. Four important polysaccharides—*amylose*, *amylopectin*, *cellulose*, and *glycogen*—are all polymers of D-glucose that differ only in the type of glycosidic bonds and the amount of branching in the molecule. All the anomeric carbons of polysaccharides are involved in acetal glycosidic bonds except the anomeric carbon at the end of the very long chain. Therefore, polysaccharides are not reducing sugars and they do not mutarotate. Three important polysaccharides are starch, glycogen, and cellulose.

Cellulose

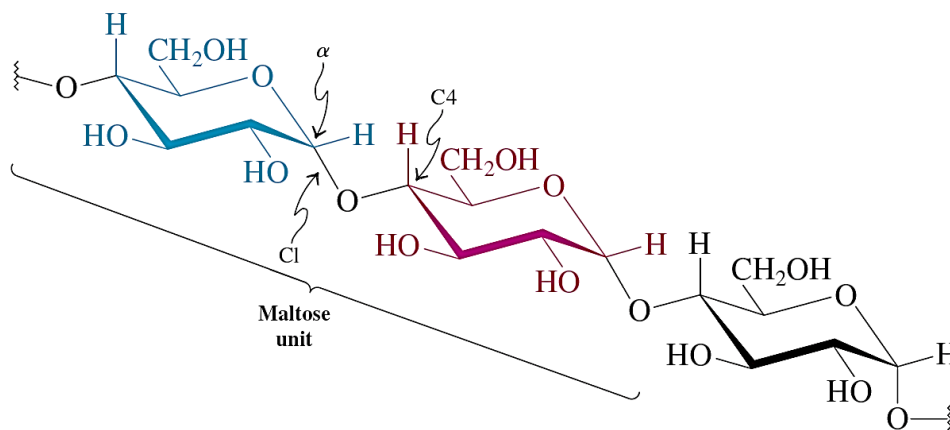
Cellulose is the most abundant and commonly occurring polysaccharide. It is found in the cell walls of nearly all plants, where it gives support and rigidity to wood and plant stems. About 50% of dry wood and about 90% of cotton fiber is cellulose. It is composed of about 3000 D-glucose units linked by β -1,4-glycosidic bonds making a long unbranched chain of molecular weight of about 500,000. It can contain as many as 12,000 units.



The β -glycosidic linkages of cellulose make cellulose chains essentially linear; they do not tend to coil into helical structures. The long linear chains of cellulose molecules stack in sheets, which are connected with one another by multiple hydrogen bonds. Therefore, only the few -OH groups are available on the surface to make hydrogen bonds with water, making this very stable, rigid, water insoluble material, which is ideal as cell-wall material for plants.

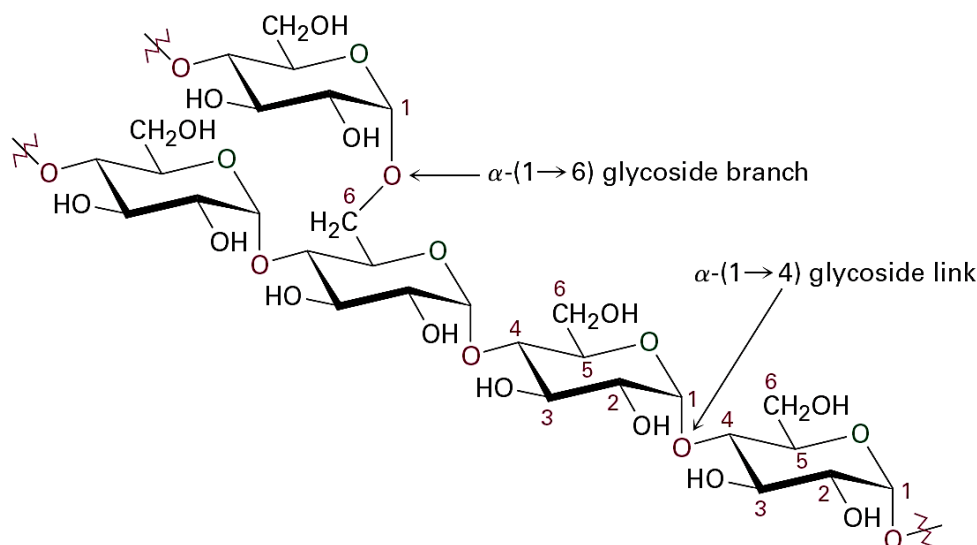
Starch

Starch, a storage form of glucose in plants, is found in all plant seeds and tubers, e.g., as insoluble granules in rice, wheat, potatoes, beans, and cereals. It is used for energy storage in plants, and is composed of two kinds of polysaccharides amylose and amylopectin. Although the starch from each plant is unique, most starches contain 20–25% amylose and 75–80% amylopectin. Amylose is insoluble in cold water, while amylopectin is soluble in cold water. Amylose is composed of linear chains of up to 4000 D-glucose units joined by α -1,4-glycosidic bonds (molecular weight, 150,000 – 600,000). Note that the disaccharide units in amylose are the same as those in maltose.



Also note that cellulose and amylose, both are composed of glucose units and both are linear molecules, however, cellulose has β -linkage, while amylose has α -linkage. The difference in the stereochemistry at the anomeric carbons leads to the strong tendency of amylose to form a helical polymer arrangement (not the straight chain shown in the structure above). The iodine molecules get inserted within helical coil to form a blue starch iodine complex (a test for starch with iodine solution).

Amylopectin is more complex in structure than amylose. Like amylose it is also joined by α -1,4-glycosidic bonds. However, unlike cellulose and amylose, which are linear polymers, amylopectin is branched polymers containing α -1,6-glycosidic linkages approximately every 24 to 30 glucose units.



Amylopectin can contain up to 10^6 glucose units, making it one of the largest molecules found in nature. Its molecular weight runs into the millions.

Glycogen

Glycogen, known as animal starch, is present in the liver and muscle tissues of animals as an energy-reserve carbohydrate. The excessive glucose in the body is stored in the form of glycogen, which gets converted to glucose, when required. Structurally, glycogen resembles amylopectin, and is a branched polymer of approximately 10^6 glucose units joined by α -1,4- and α -1,6-glycosidic bonds. It is more branched than amylopectin – branching occurs at every 8 to 12 glucose units. The high degree of branching in glycogen has important physiological effects. When the body needs energy, many individual glucose units can be simultaneously removed from the ends of many branches.

5.2. Lipids

Unlike other classes of organic compounds, lipids are defined by its solubility (a physical property) rather than by structure. The natural organic compounds, which are completely insoluble in water and soluble in nonpolar solvents, are classified as lipids. They are one of the most heterogeneous classes of naturally occurring organic compounds, which are major component of plasma membrane, and are involved in cellular communication. The word 'lipid' comes from the Greek *lipos*, which means 'fat'. Lipids include the fatty acids, fats, oils, waxes, phospholipids, prostaglandins, terpenes, steroids, and water insoluble vitamins. Lipids are classified into two main groups:

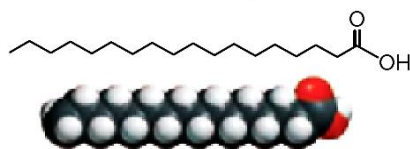
- (i) Those containing ester linkage and can be hydrolyzed, e.g., fats, waxes, etc.
- (ii) Those which do not have ester linkage and cannot be hydrolyzed, e.g., terpenes, steroids, prostaglandins, etc.

5.2.1. Fatty Acids

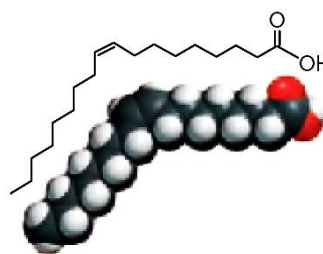
Fatty acids are the simplest type of lipids and are found as components in more complex lipids. They are long chain monocarboxylic acids containing a long, unbranched carbon chain attached to a carboxylic acid group at one end. Although the

carboxylic acid part is hydrophilic, the long hydrophobic carbon chain makes long-chain fatty acids insoluble in water.

- Most fatty acids in nature are unbranched and have even number of carbon atoms, most between 12 and 20.
- In unsaturated fatty acids, all the double bonds are *cis*- (or *Z*-) and are never conjugated.
- Saturated fatty acids have high melting point and fully extended carbon chain. Saturated fatty acids have higher melting points than unsaturated fatty acids because they pack together more tightly.
- In unsaturated fatty acids, the *cis* double bonds cause the carbon chain to bend or kink, giving the molecules an irregular shape, making them loosely bound, and thus the melting points of unsaturated fatty acids is lower than those of saturated fatty acids.



(a) Stearic acid



(b) Oleic acid

- As the number of double bonds increases, the melting point decreases. For example, an C_{18} fatty acid melts at 69°C if it is saturated, at 13°C if it has one double bond, at -5°C if it has two double bonds, and at -11°C if it has three double bonds.
- Triple bond rarely occurs in fatty acids.

The fatty acids without double bonds are called saturated fatty acids; with one double bond are called unsaturated fatty acids, while with more than one double bonds are called polyunsaturated fatty acids (PUFA). The number of carbons in a fatty acid and the number of carbon-carbon double bonds in its hydrocarbon chain are shown by two numbers separated by a colon. For example, linoleic acid is designated as 18:2 fatty acid; its 18-carbon chain contains two carbon-carbon double bonds. Some of the most abundant fatty acids in nature are:

Carbon Atoms/ Double Bonds*	Structure	Common Name	Melting Point ($^{\circ}\text{C}$)
Saturated Fatty Acids			
12:0	$\text{CH}_3(\text{CH}_2)_{10}\text{COOH}$	Lauric acid	44
14:0	$\text{CH}_3(\text{CH}_2)_{12}\text{COOH}$	Myristic acid	58
16:0	$\text{CH}_3(\text{CH}_2)_{14}\text{COOH}$	Palmitic acid	63
18:0	$\text{CH}_3(\text{CH}_2)_{16}\text{COOH}$	Stearic acid	70
20:0	$\text{CH}_3(\text{CH}_2)_{18}\text{COOH}$	Arachidic acid	77

Unsaturated Fatty Acids

16:1	$\text{CH}_3(\text{CH}_2)_5\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	Palmitoleic acid	1
18:1	$\text{CH}_3(\text{CH}_2)_7\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	Oleic acid	16
18:2	$\text{CH}_3(\text{CH}_2)_4(\text{CH}=\text{CHCH}_2)_2(\text{CH}_2)_6\text{COOH}$	Linoleic acid	-5
18:3	$\text{CH}_3\text{CH}_2(\text{CH}=\text{CHCH}_2)_3(\text{CH}_2)_6\text{COOH}$	Linolenic acid	-11
20:4	$\text{CH}_3(\text{CH}_2)_4(\text{CH}=\text{CHCH}_2)_4(\text{CH}_2)_2\text{COOH}$	Arachidonic acid	-49

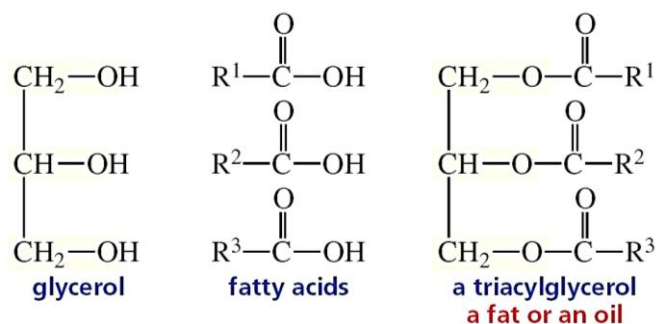
Essential Fatty acids (EFA)

The human body is capable of synthesizing most fatty acids from carbohydrates or other fatty acids. However, humans do not synthesize sufficient amounts of polyunsaturated fatty acids such as linoleic acid, linolenic acid, and arachidonic acid. Because they are obtained only from the diet, they are known as *essential fatty acids*. The most important essential fatty acid is **linoleic acid**, which serves as a basis for the production of other essential fatty acids. Vegetable oils are abundant in linoleic acid. Note that not all polyunsaturated fatty acids are essential fatty acids.

Fatty acids normally occur naturally as esters; fats, oils, phospholipids, and waxes all are different types of fatty acid esters.

5.2.2. Fats & Oils

Fats and oils, the most abundant naturally occurring lipids, are triesters of glycerol (1,2,3-propanetriol) with long-chain carboxylic acids and are also called **triglycerides** or **triacylglycerols**. Complete hydrolysis of triglycerides yield a molecule of glycerol and three molecules of long chain fatty acids. Simple triglycerides contain all the three fatty acid components same, while mixed triglycerides are made up of two or three different fatty acid components. Most fats and oils are mixed triacylglycerols that contain two or three different fatty acids.



The melting point of a triglyceride increases as the number of carbons in its hydrocarbon chains increases and as the number of carbon-carbon double bonds decreases. Triglycerides rich in saturated fatty acids, like palmitic and stearic acids, have high melting point, therefore, are solid or semisolid at room temperature. These are called fats. Triglycerides having high proportion of unsaturated or polyunsaturated fatty acids, like oleic and linoleic acid, have low melting point, therefore, are liquid at room temperature. They are called oils, for example olive oil, corn oil, etc.

The triglycerides of animal fats are richer in saturated fatty acids, whereas the triglycerides of plant oils (for example, corn, soybean, canola, olive, and palm oils) are richer in unsaturated fatty acids.

Fats are biological energy reserves, which are stored in the body's tissues until their metabolism leads ultimately to CO_2 and water. Oils have a similar function in plant seeds. Triglycerides, on metabolism, yield more than twice as many kilocalories per gram as do carbohydrates or proteins, because of the high proportion of carbon-hydrogen bonds per molecule.

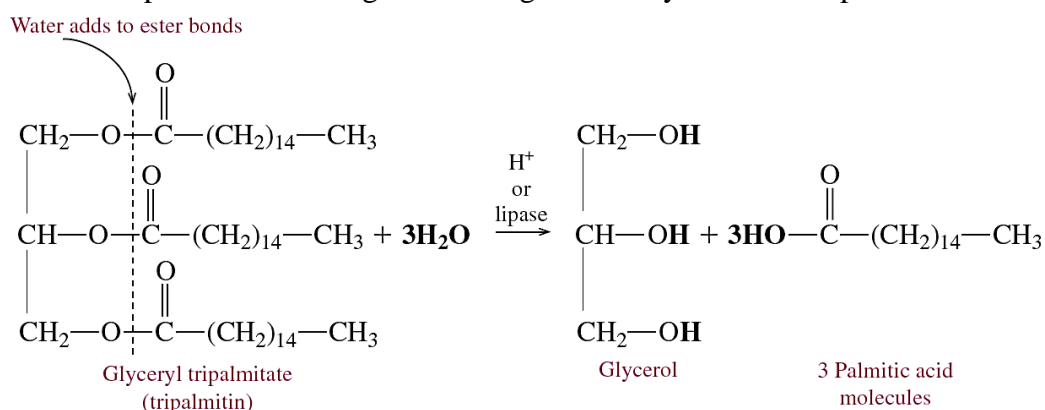
Chemical properties of triglycerides

i. Hydrogenation

Vegetable oils are mainly composed of unsaturated fatty acids, thus most of them are liquid. The catalytic hydrogenation of some or all double bonds of unsaturated fatty acids in vegetable oils under controlled conditions of temperature and pressure in the presence of finely divided nickel results in the formation of saturated oil (also called hydrogenated oil). The process is called hardening of oils, as the oil is converted to semisolid or solid fat. Complete hydrogenation gives a very brittle product, whereas the partial hydrogenation of a liquid vegetable oil changes it to a soft, semisolid fat. By controlling the amount of hydrogen, manufacturers produce various types of products such as soft margarines, solid stick margarines, and solid shortenings from soybean oil and safflower oil.

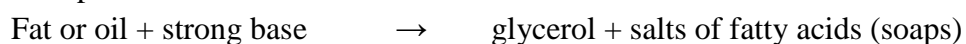
ii. Hydrolysis

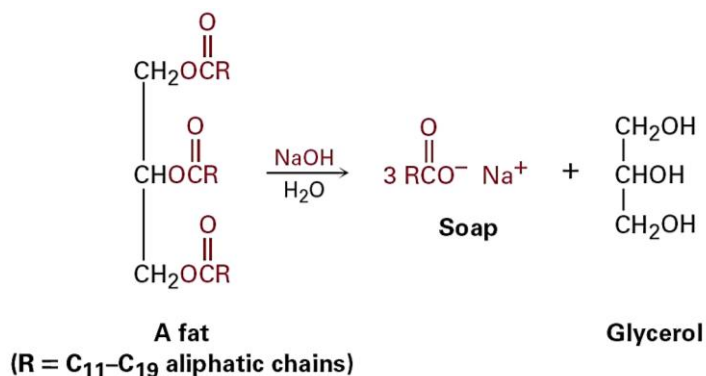
Triglycerides are hydrolyzed to a molecule of glycerol and three molecules of fatty acids in the presence of strong acids or digestive enzymes called lipases.



iii. Saponification

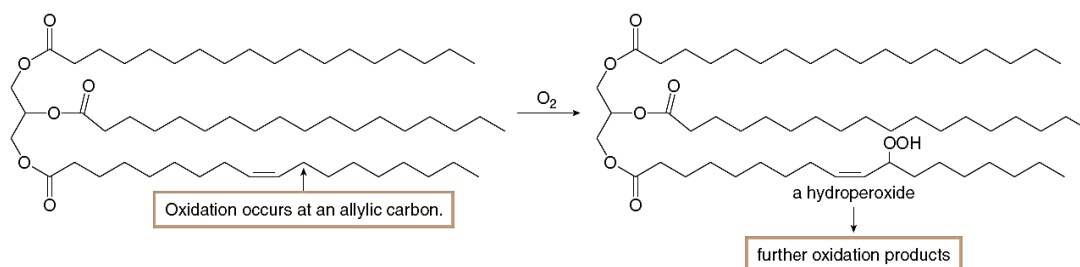
Triglycerides on heating with a strong base, like NaOH or KOH, hydrolyze to produce glycerol and salts of long chain fatty acids. The salts of long chain fatty acids are called **soaps**. NaOH produces solid soaps while KOH produces softer, liquid soap.





iv. Rancidity

Triglycerides containing highly unsaturated fatty acids, on exposure to air, develop unpleasant odour and flavor. The phenomenon is called rancidity. Triglycerides that contain unsaturated fatty acid residues are particularly susceptible to autooxidation at the allylic position to yield hydroperoxides as allylic C – H bonds are weaker than other C – H bonds. The hydroperoxide formed by this process is unstable and it undergoes further oxidation of the fatty acids to low molecular weight fatty acids and carbonyl compounds having foul smell. For example, the odour of rancid butter is largely due to the presence of butanoic acid formed on autooxidation of butter fat.



5.2.3. Waxes

Waxes are the simplest hydrolyzable lipids. They are esters (RCOOR') of long-chain carboxylic acids (RCOOH) with long-chain alcohols ($\text{R}'\text{OH}$). The carboxylic acid usually has an even number of carbons from 16 through 36, while the alcohol has an even number of carbons from 24 through 36. Because of the long hydrocarbon chain, they are highly hydrophobic. In nature, waxes are present as protective coating on leaves to reduce the water evaporation, and on feather of birds and animal fur to make them water repellent. The fatty acids and alcohols present in waxes are:

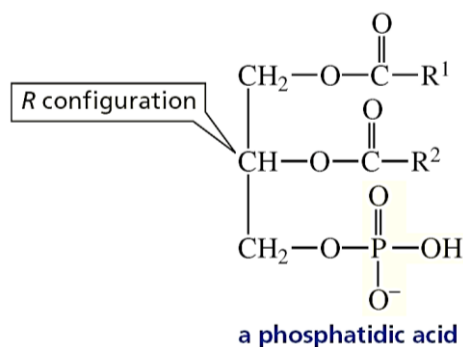
Fatty acid		Alcohols	
Palmitic acid	$\text{C}_{15}\text{H}_{31}\text{COOH}$	Cetyl alcohol	$\text{C}_{16}\text{H}_{33}\text{OH}$
Cerotic acid	$\text{C}_{25}\text{H}_{51}\text{COOH}$	Ceryl alcohol	$\text{C}_{26}\text{H}_{53}\text{OH}$
Myristic acid	$\text{C}_{13}\text{H}_{27}\text{COOH}$	Myricyl alcohol	$\text{C}_{30}\text{H}_{61}\text{OH}$

Some common example of waxes are:

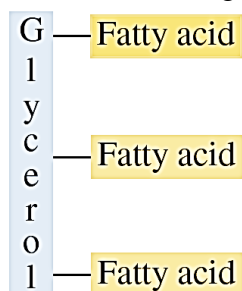
- (a) **Bee's wax**, the structural material of beehives, contains an ester (myricyl palmitate or triacontyl hexadecanoate, $C_{15}H_{31}COOC_{30}H_{61}$) of myricyl alcohol (1-triacontanol) and palmitic acid (hexadecanoic acid). It is used in shoe polishes, as water proof coatings, and for preparing candles.
- (b) **Carnauba wax**, also known as car wax, is an ester (myricyl cerotate or triacontyl pentadodecanoate, $C_{25}H_{51}COOC_{30}H_{61}$) of myricyl alcohol (1-triacontanol) and cerotic acid (pentadodecanoic acid). It is extracted from the leaves of a brazilian palm tree and is used for floor and automobile polishing.
- (c) **Spermaceti wax**, obtained from sperm whale, contains cetyl palmitate (hexadecyl hexadecanoate, $C_{15}H_{31}COOC_{16}H_{33}$). This is odourless and tasteless wax, white in colour and is widely used as emollient in a number of soaps and cosmetics.

5.2.4. Phospholipids

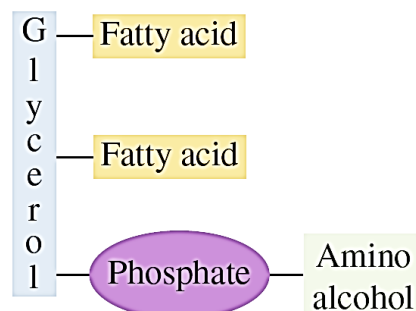
Phospholipids (or phosphoglycerides or glycerophospholipids), an important component of cell membrane, are second most abundant type of lipids. Like triglycerides, phosphoglycerides are also esters of glycerol with the main difference being that in phosphoglycerides only two of the hydroxy groups of glycerol are esterified with fatty acids. The third OH group is esterified with phosphoric acid instead of fatty acid, forming a phosphatidic acid. The fatty acids most common in phosphatidic acids are palmitic and stearic acids (both fully saturated) and oleic acid (one double bond in the hydrocarbon chain). The middle carbon of the glycerol backbone of all of these compounds is a stereogenic center, usually with the *R* configuration.



The phosphoric acid in phosphatidic acid is further esterified with low molecular-weight amino alcohols to give phospholipids.

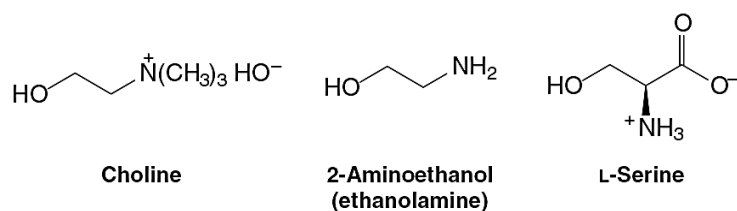


Triacylglycerol (triglyceride)

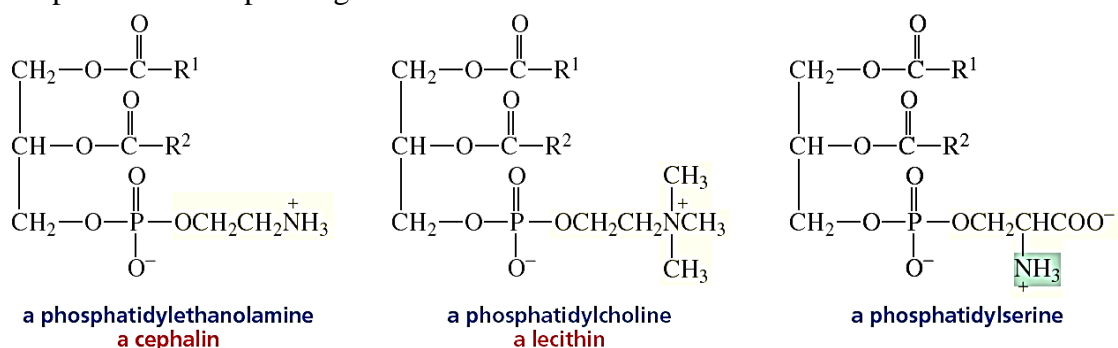


Glycerophospholipid

Phosphoglycerides derived from ethanolamine, choline and serine are particularly abundant in the cells of plants and animals.



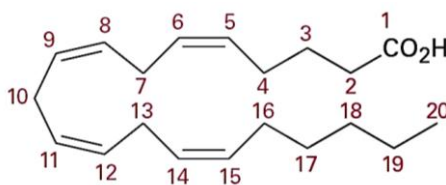
Phosphoglycerides that contain ethanolamine are called **cephalins**, while those that contain choline are called **lecithins**. **Lecithins** and **cephalins** are particularly abundant in brain and nerve tissues as well as in egg yolks, wheat germ, and yeast. Lecithin are added to foods such as mayonnaise as emulsifying agent to prevent the aqueous and fat component from separating.



Phospholipids are amphipathic in nature, that is, each has a hydrophilic head (the phosphate group plus amino alcohol), and a long hydrophobic tail (hydrocarbon chains of fatty acids).

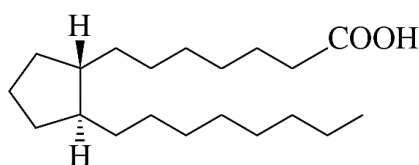
5.2.5. Prostaglandins

Prostaglandins are hormone-like substances produced in small amounts in most cells of the body and are responsible for regulating a variety of physiological responses, such as inflammation, blood pressure, blood clotting fever, pain, the induction of labor, and the sleep–wake cycle. The **prostaglandins** are a family of compounds derived from arachidonic acid, the polyunsaturated fatty acid.



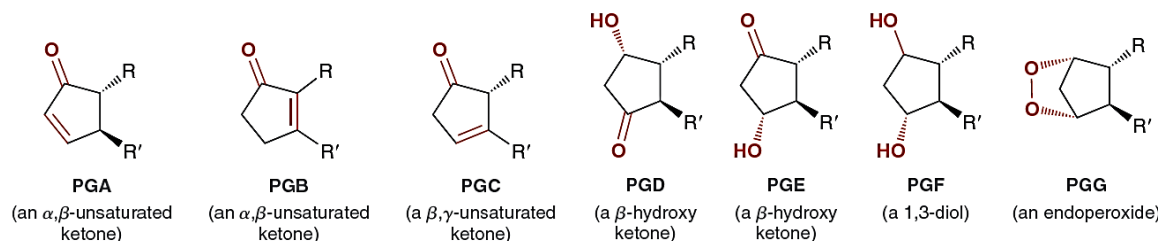
Arachidonic acid

All prostaglandins have the 20-carbon skeleton containing a five-membered ring with a seven-carbon carboxylic acid substituent and an eight-carbon hydrocarbon substituent. The two substituents are *trans* to each other.

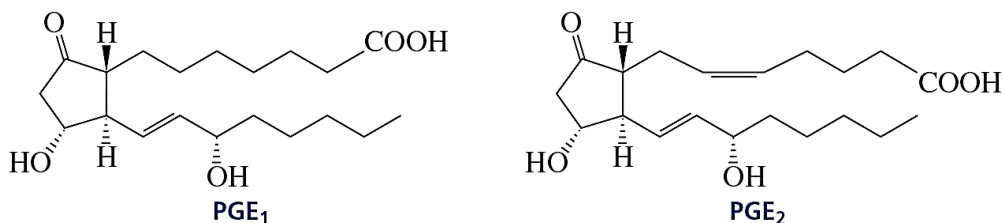


prostaglandin skeleton

All prostaglandins have hydroxyl group at C-15. The various kinds of prostaglandins differ by the substituents attached to five-carbon ring, the most common of which are:



In each case, the letters PG indicate that the compound is a prostaglandin and the third letter indicates the substitution pattern. PGAs, PGBs, and PGCs all exhibit a carbonyl group and a carbon-carbon π bond in the five-membered ring. These three substitution patterns differ from each other only in the location of the carbon-carbon π bond. PGDs and PGEs are β -hydroxyketones, PGFs are 1,3-diols, and PGGs are endoperoxides. The number of carbon-carbon π bonds in the side chains is indicated with a subscript after the letter for the substitution pattern. For example, PGE_1 and PGE_2 (dinoprostone) has the E substitution pattern and contain one and two double bonds in the side chains, respectively.

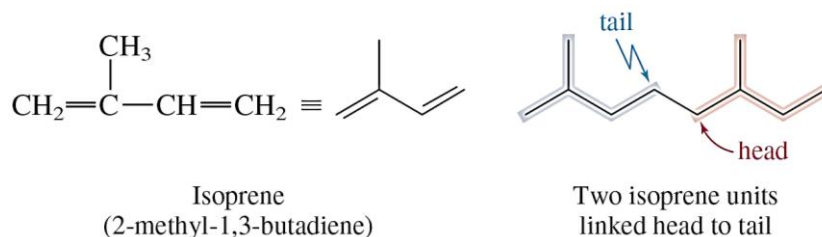


For the PGF substitution pattern, “ α ” or “ β ” is added to the name to indicate the configuration of the OH groups. A *cis* diol is designated as “ α ,” while a *trans* diol is designated as “ β ”.

5.2.6. Terpenes

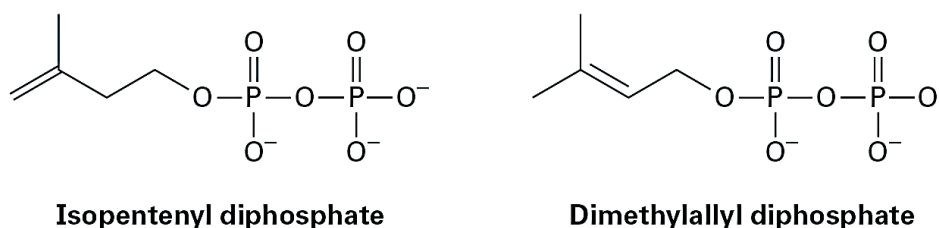
The fragrance of rose, jasmine, lilies, camphor, sandalwood, and flavor of orange, lemon, peppermint, cloves are because of certain volatile compounds called terpenes (or terpenoids). **Terpenes** are a diverse family of compounds with carbon skeletons composed of five-carbon units called isoprene (2-methyl-1,3-butadiene). Therefore, terpenes are also called **isoprenoids**. Consequently, the number of carbon atoms in terpenes is in multiple of 5. Carbon 1 (branched end) of an isoprene unit is called the

head; carbon 4 (unbranched end) is called the tail. Terpenes are formed by bonding the tail of one isoprene unit to the head of another.



That isoprene units are linked in a head-to-tail fashion to form terpenes is known as the **isoprene rule**.

Isoprene itself is not the true biological precursor of terpenoids. Nature instead uses two “isoprene equivalents”—isopentenyl diphosphate and its isomer dimethylallyl diphosphate.



Terpenes have various functional groups. They may be hydrocarbon, alcohols, esters, ketones, or aldehydes. Moreover, they may have acyclic or cyclic structures.

Classification

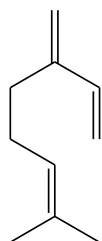
Terpenoids are classified according to the number of isoprene units they contain. A monoterpene contains two isoprene units (10 carbons), a sesquiterpene has three isoprene units (15 carbons), so on.

Name	Number of C-atoms	Number of isoprene units
Monoterpene	10	2
Sesquiterpene	15	3
Diterpene	20	4
Sesterterpene	25	5
Triterpene	30	6
Tetraterpene	40	8

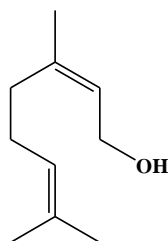
Monoterpenes

Monoterpenes have 10 carbon atom skeleton and are the simplest terpenes. These are dimers of isoprene. The monoterpenes are further subdivided as –

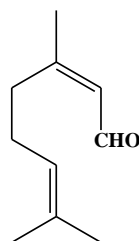
- (i) **Acyclic monoterpenes**, e.g., myrcene, geraniol, citral, linalool, etc.



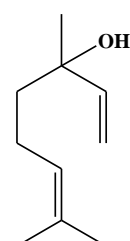
Myrcene
(bay oil)



Geraniol
(rose oil)

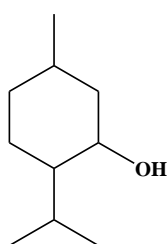


Citral
(oil of lemon grass)

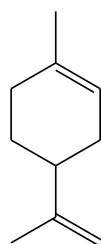


Linalool
(lavender oil)

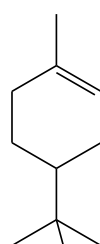
- (ii) **Monocyclic monoterpenes**, e.g., menthol, limonene, α -terpineol, piperitone, etc.



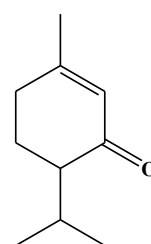
Menthol
(peppermint)



Limonene
(oil of orange)

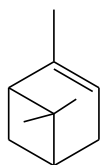


α -terpineol
(oil of cardamom)

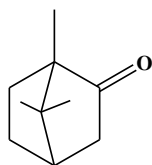


Piperitone
(eucalyptus)

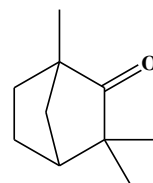
- (iii) **Bicyclic monoterpenes**, e.g., α -pinene, camphor, fenchone, etc.



α -Pinene
(turpentine oil)



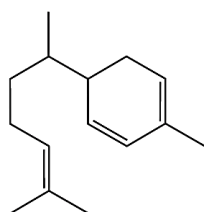
Camphor
(camphor tree)



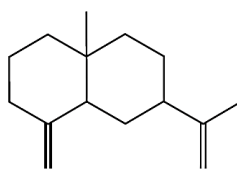
Fenchone
(oil of fennel)

Sesquiterpenes

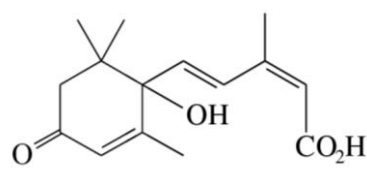
Sesquiterpenes have three isoprene units, thus 15 carbon atoms in their skeleton. They are also acyclic, monocyclic, bicyclic, or tricyclic. They have various types of structures. For example:



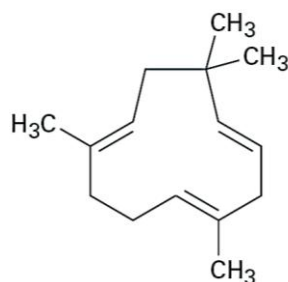
zingiberene
oil of ginger



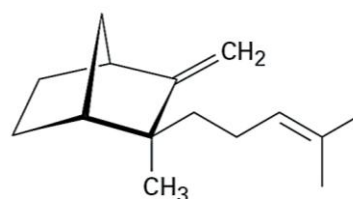
β -selinene
oil of celery



Absciscic acid
(a plant hormone)



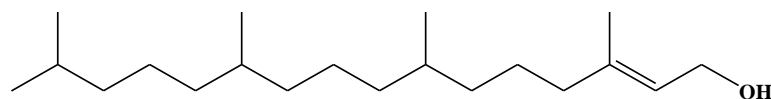
Humulene
(oil of hops)



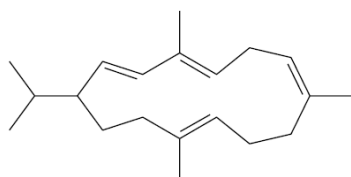
β -Santalene
(sandalwood oil)

Diterpenes

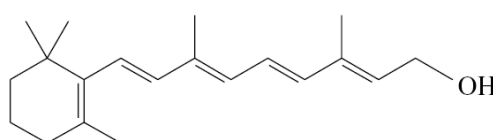
Diterpenes have four isoprene units and thus 20 carbon atoms. A well-known acyclic diterpene is phytol, an alcohol that occurs in the chlorophyll in green plants. Vitamin A is an example of monocyclic diterpene. Cembrene, a monocyclic diterpene, is found in pine tree.



Phytol



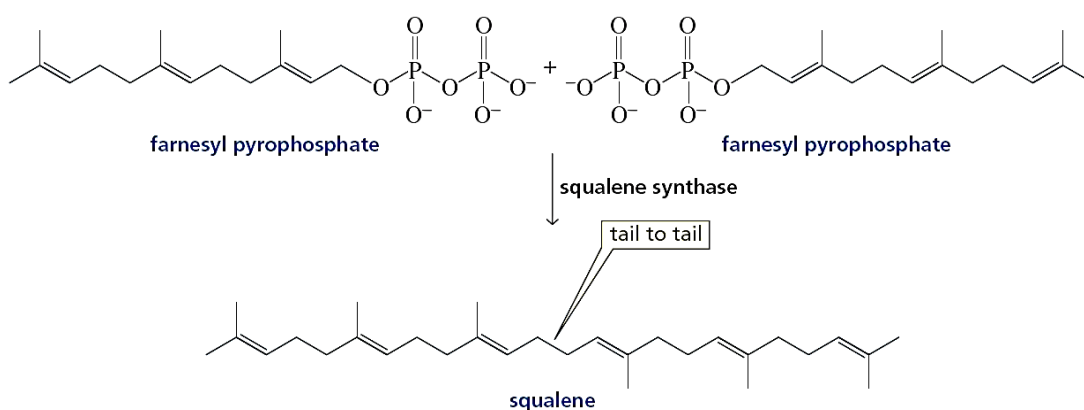
Cembrene
(pine)



Vitamin A
(present in mammalian tissue and fish oil;
important substance in the chemistry of vision)

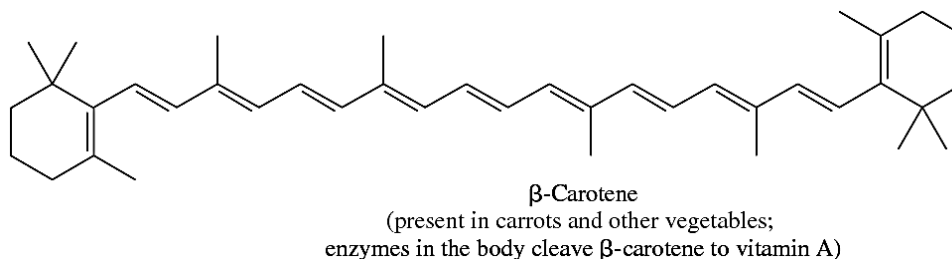
Triterpenes

Triterpenes are dimer of sesquiterpene skeleton (15 carbon skeleton). In triterpene, three isoprene units are joined in head-to-tail manner to make C_{15} unit. Then two C_{15} units are joined to each other in tail-to-tail manner. Squalene is triterpene acyclic hydrocarbon, which is found in liver oils of some of the shark species. They generally contain tetracyclic and pentacyclic skeletons.



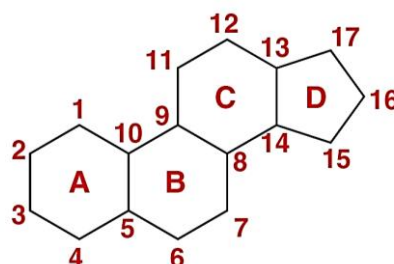
Tetraterpenes

Like triterpenes, tetraterpenes are dimer of C_{20} units. Four isoprene units are joined in head-to-tail manner to form C_{20} unit, and two C_{20} units are joined in tail-to-tail manner. Carotenoids, reddish yellow pigments found in plants and animals, are tetraterpenes.



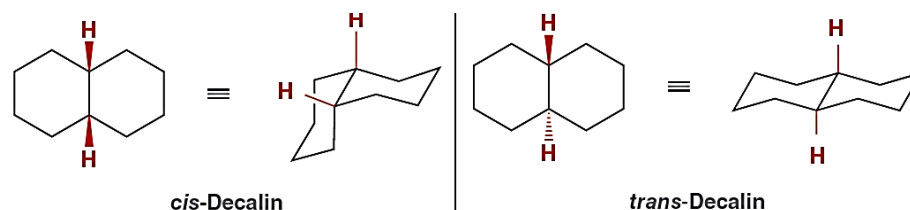
5.2.7. Steroids

Steroids, an important class of lipids are actually metabolic derivatives of terpenes. Steroids and their derivatives are among the most widely used therapeutic agents. They are used in birth control and hormone replacement therapy and in the treatment of inflammatory conditions and cancer. The structures of **steroids** are based on a tetracyclic ring system involving three six-membered rings and one five-membered ring, designated as A, B, C, and D.

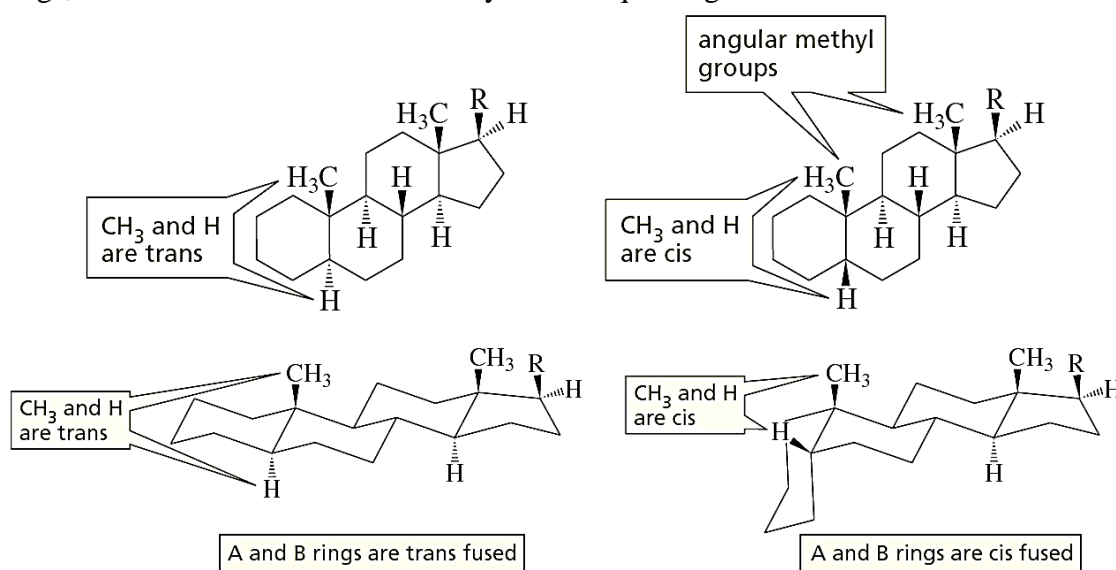


The tetracyclic skeleton of steroids

When six-membered rings are fused with a *cis* configuration, as in *cis*-decalin, the two rings are both free to exhibit ring flipping. Both groups at the ring-junction positions (the *angular* groups) are on the same side of the two rings. In contrast, in *trans*-decalin, the groups at the junctions are on opposite sides. It does not exhibit ring flipping and is a more rigid structure.

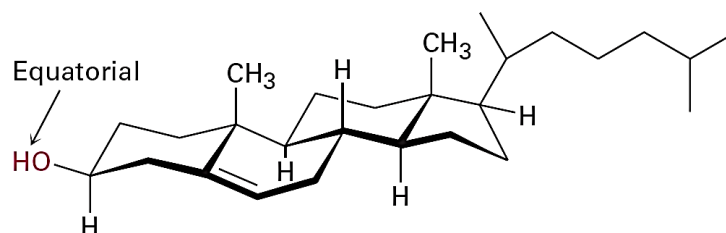


In steroids, the three 6-membered rings (A, B, and C) adopt chair conformations. The B, C, and D rings are all *trans* fused. The A and B rings may be *cis* or *trans* fused, but *trans* is more common in nature. Because of the *trans-trans-trans* fusion of all four rings, the structure of steroids is nearly flat and quite rigid.



Many steroids have methyl groups at the 10- and 13-positions. These are called **angular methyl groups**. They are denoted β , if are up; and are denoted α , if they are down. In A–B *trans* steroid the C19 angular methyl group is β , and the hydrogen atom at C5 is α , on opposite sides of the molecule. In A–B *cis* steroid, by contrast, both the C19 angular methyl group, and the C5 hydrogen atom are on the same side (β) of the molecule.

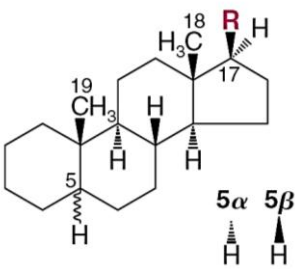
Substituent groups at different sites on the tetracyclic skeleton have axial or equatorial orientations that are fixed because of the rigid structure of the *trans*-fused rings. This fixed orientation influences chemical reactivity, largely due to the greater steric hindrance of axial groups versus their equatorial isomers. Thus an equatorial hydroxyl group is esterified more rapidly than its axial isomer.



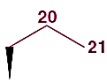
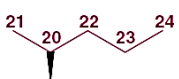
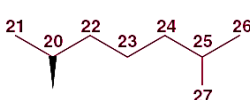


Cholesterol

In systematic nomenclature the nature of the R group at position 17 determines (primarily) the base name of an individual steroid.

Names of Steroid Hydrocarbons

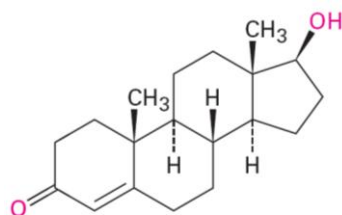


R	Name
	Androstane
 (with —H also replacing —CH ₃)	Estrane
	Pregnane
	Cholane
	Cholestane

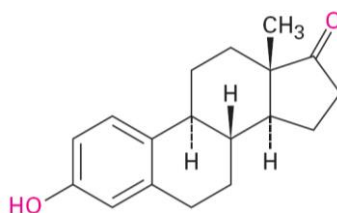
Classification

The steroids are classified into five major categories:

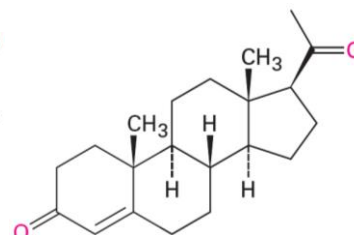
- (i) **Sterols:** These steroids are characterized by the presence of an alcoholic group and are found in plant oil, and animal fat, e.g., cholesterol.
- (ii) **Sex hormones:** Unlike other hormones, sex hormones are not secreted by ductless gland. But are produced in gonads. These are termed as steroidal hormones and are further classified into three major sub-categories:
 - (a) **Androgens:** They mediate the development of sexual characteristics in males e.g., testosterone
 - (b) **Estrogens:** They mediate the development of sexual characteristics in females e.g., estrone
 - (c) **Gestogens:** Corpus Luteum hormone e.g., Progesterone



Testosterone

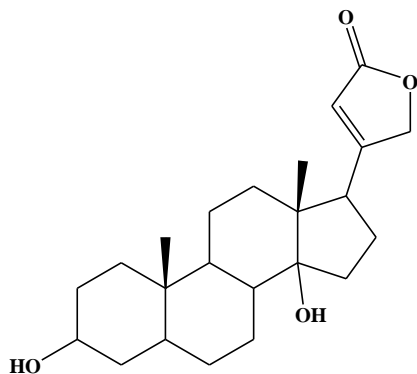


Estrone

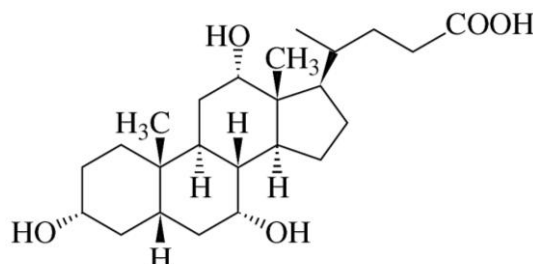


Progesterone

- (iii) **Cardiac Glycoside:** These steroids are extracted from plants and are used for cardiac therapy, e.g., digitoxigenin
- (iv) **Bile acids:** These are the organic compounds present in bile – a liquid synthesized by liver and stored in gall bladder, and secreted into the small intestine, where they act as emulsifying agents so that fats and oils can be digested by water-soluble digestive enzymes, e.g., cholic acid



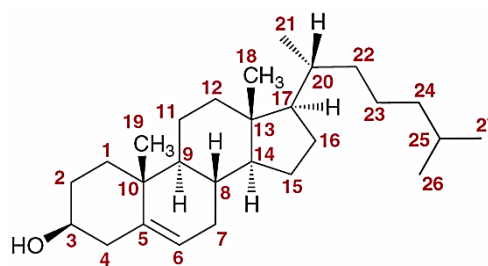
Digitoxigenin



cholic acid

- (v) **Sapogenin:** These are present in plant glycosides and their colloidal solution in water forms foam similar to soap. They are highly toxic.

The most abundant member of the steroid family is **cholesterol**, the precursor of all other steroids. It is an important component of cell membrane, and is biosynthesized from squalene, a triterpene. It has 8 asymmetric carbons, thus 256 stereoisomers (2^8) are possible, however only one exists in nature.

Cholesterol
(a steroid)

This substance is an integral part of human metabolism in two ways: (a) it is an essential component of biological membranes. (b) Sex hormones, cardiac glycosides, bile acids, and vitamin D all are synthesized from it. Thus cholesterol is the parent steroid.

5.3. Proteins and Amino Acids

The word “protein” is derived from the Greek word *proteios*, meaning “first” as proteins are the most important and fundamental structural components of the body. A protein is a polymer composed of a long chain of nitrogen containing monomers known as amino acids linked together by peptide bonds. The naturally occurring amino acids are mainly α -amino acids.

Proteins exhibit a virtually unlimited variety of sizes, configurations, and function due to (i) the large number of amino acids building a single protein molecule (thousands in many cases); (ii) the almost infinite number of combinations the different amino acids can form; (iii) the reactivity of side groups of the individual amino acids; (iv) their bonding with other

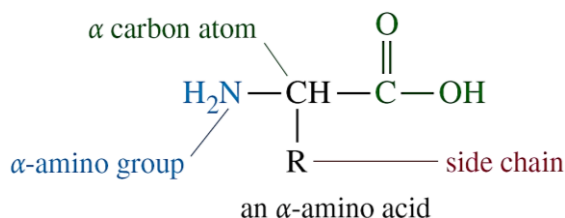
complex molecules such as lipids, carbohydrates, nucleic acids, and heme groups to form more complex structures.

Because of this versatility, proteins play an astounding variety of functions in living systems. Some of the functions of major classes of proteins are:

Class of Protein	Function	Example
Structural proteins	Provides structural components	<i>Collagen</i> in tendons and cartilage <i>Keratin</i> in hair, skin, and nails
Enzymes	Catalysed chemical reactions. Replicates and repairs DNA	<i>Sucrase</i> catalyse the hydrolysis of sucrose <i>DNA polymerase</i>
Transport proteins	Transport essential substances throughout the body	<i>Hemoglobin</i> transports oxygen <i>Lipoprotein</i> transport lipids
Contractile proteins	Cause contraction of muscles	<i>Actin, myosin</i>
Protective proteins	Complex with foreign proteins	<i>Antibodies</i>
Hormones	Regulates cell metabolism and the nervous system	<i>Insulin</i> regulates blood glucose level. <i>Growth hormone</i> regulates body growth
Toxins	Incapacitate prey	<i>Snake venoms</i>
Storage	Store nutrients	<i>Casein</i> stores protein in milk <i>Ferritin</i> stores iron in the spleen and liver

5.3.1. α -Amino Acids

α -Amino Acids are carboxylic acids containing amino group at α -carbon. Structurally, all amino acids have an amino ($-\text{NH}_2$) group, carboxylic ($-\text{COOH}$) group, a hydrogen, and a side chain ($-\text{R}$) – all attached to a carbon.



The R group (side chain) can be alkyl or aryl, and it can contain hydroxy, amino, mercapto, sulfide, carboxy, guanidino, or imidazolyl groups.

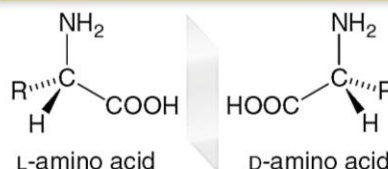
General features of α -amino acids

Though a large number of amino acids (more than 700) exist in nature, however, proteins in all living organisms mainly consist of twenty amino acids. The amino acids differ only in the side chain (R) attached to the α -carbon.

- The simplest amino acid is glycine which has $\text{R}=\text{H}$, and thus this is the only amino acid which is achiral.
- Except glycine, all amino acids have chiral center at α -carbon and are thus optically active. In the D,L convention, all naturally occurring amino acids are L-amino acids. D-amino acids are rarely found in nature According to the R,S

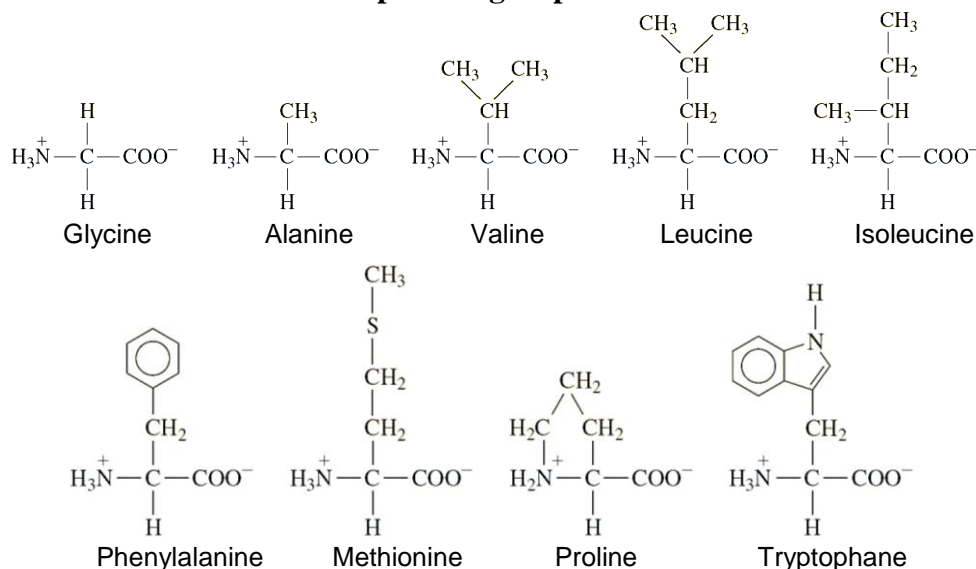
convention, all L-amino acids except cysteine have the S-configuration. Because of priority rules, the presence of the sulfhydryl group on the side chain of L-cysteine gives the chiral center the *R* configuration.

Two possible enantiomers when $R \neq H$

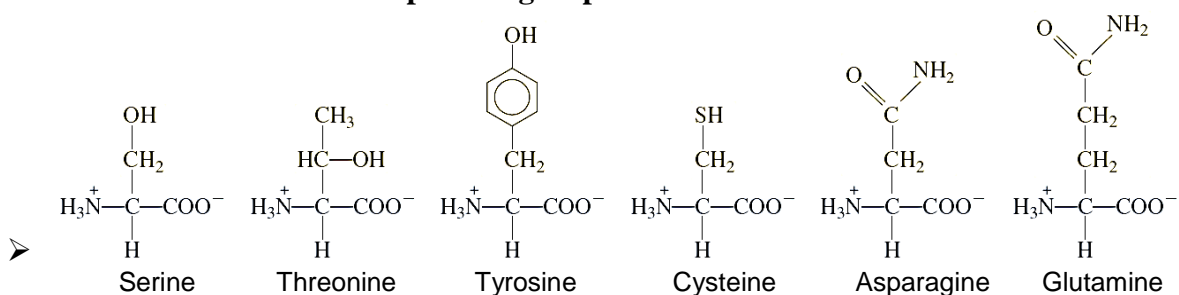


Only this isomer occurs in proteins.

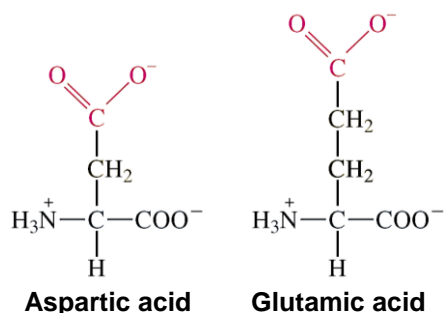
- The R group present in amino acids determines their properties in aqueous solution. The nonpolar amino acids have hydrogen, alkyl, or aromatic R groups, which make them *hydrophobic*. The polar amino acids have polar R groups that interact with water, which makes them *hydrophilic*.
- Amino acids with an additional COOH group in the side chain are called acidic amino acids. Those with an additional basic N atom in the side chain are called basic amino acids. All others are neutral amino acids.
- **Neutral amino acids with nonpolar R group**



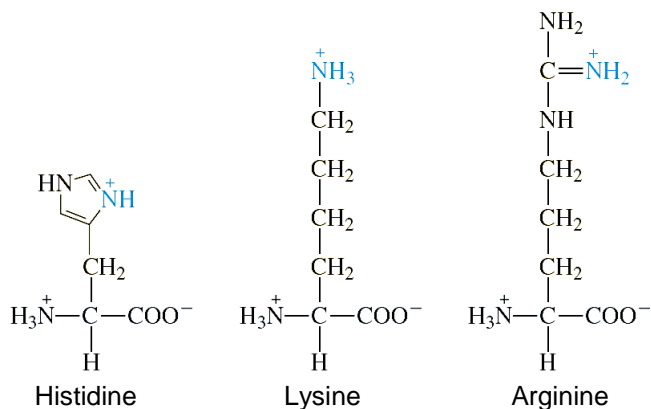
- **Neutral amino acids with polar R group.**



- **Acidic amino acids**

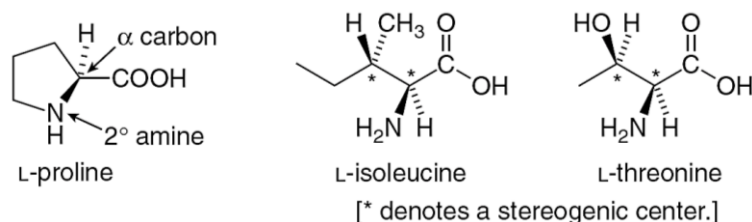


➤ **Basic amino acids**



Note that both cysteine (a thiol) and tyrosine (a phenol), although usually classified as neutral amino acids, nevertheless have weakly acidic side chains that can be deprotonated in a sufficiently basic solution.

- All amino acids have primary α -amino group except proline which has N-atom in five-membered ring, making it a secondary amine.
- **Isoleucine** and **threonine** contain an additional stereogenic center at the β -carbon, so there are four possible stereoisomers, only one is found in nature.



Classification based on nutritional importance

- (i) **Essential amino acids.** The amino acids that are not synthesized by body and are required to be taken externally through diet are termed as essential amino acids. For example, valine, leucine, isoleucine, phenylalanine, threonine, tryptophane, methionine, and lysine.
- (ii) **Semi-essential amino acids.** These amino acids are not synthesized in human body in sufficient amount during growth. However, these are essential in pregnant women, lactating mothers, and in growing children. Arginine and histidine fall under this category.

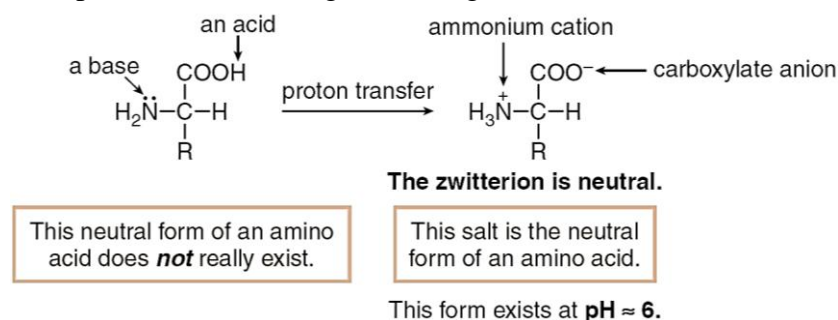
- (iii) **Non-essential amino acids.** These amino acids can be synthesized by human body from other material through a series of biochemical reactions. It is therefore, not necessary to be taken through diet. Ten amino acids have been classified under this category – glycine, alanine, proline, asparagine, glutamine, serine, tyrosine, cysteine, aspartic acid, and glutamic acid.

Note that humans can synthesize arginine, but it is needed for growth in greater amounts than can be synthesized. So arginine is an essential amino acid for children, but a nonessential amino acid for adults.

Physical properties of amino acids

- Amino acids are high melting, colourless crystalline substances
- They are highly soluble in water
- They have large dipole moment and high dielectric constant value.

Thus they resemble ionic compounds. The acidic ($-\text{COOH}$) and basic ($-\text{NH}_2$) groups of amino acid react with each other to form a dipolar ion or internal salt. The internal salt of amino acid is called zwitterion which has no net charge; it contains one positive and one negative charge.



This form exists at **pH ≈ 6.**

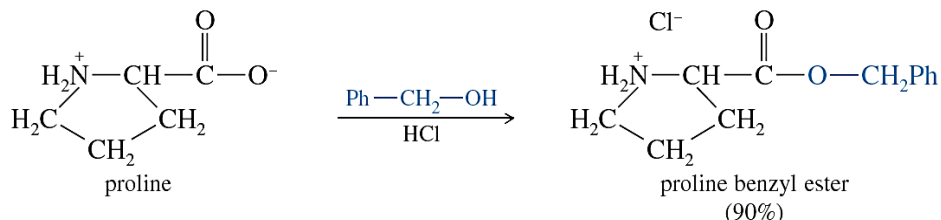
In zwitterion form, the molecule has no net charge and is said to be isoelectric. Zwitterion is dependent on the pH of the solution. The pH at which an amino acid exists in its zwitterion form is called the isoelectric point (pI) for that amino acid. Each amino acid has its characteristic isoelectric point depending on the nature of its side chain (R group).

5.3.2. Reactions of Amino Acids

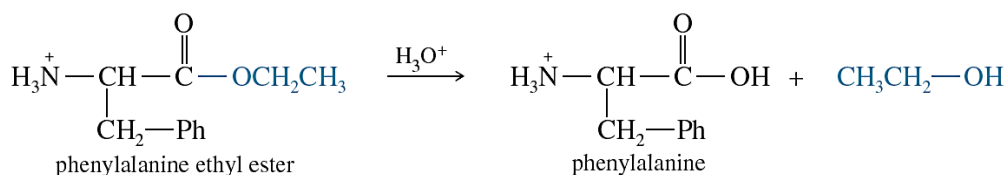
Amino acids undergo various typical reactions of both amines and carboxylic acids. However, in certain reactions these groups need to be protected, so that they do not interfere in the reactions of each other. Two most useful reactions are esterification of the carboxyl group, and acylation of the amino groups, which are used to protect these groups while the other group is being modified or coupled to another amino acid. Another reaction of amino group is with formaldehyde; the reaction is used for estimation of amino acids. Amino acids also undergo reactions that are specific to the α -amino acid structure. One of these unique amino acid reactions is the formation of colored product on treatment with ninhydrin. This reaction is used to detect amino acids.

Esterification of carboxyl group (reaction with alcohol)

The reaction of amino acids with large excess of alcohol in presence of acid catalyst (often gaseous HCl) results in the formation of acid ester. Under these acidic conditions, the amino group is present in its protonated ($-\text{NH}_3^+$) form, so it does not interfere with esterification.

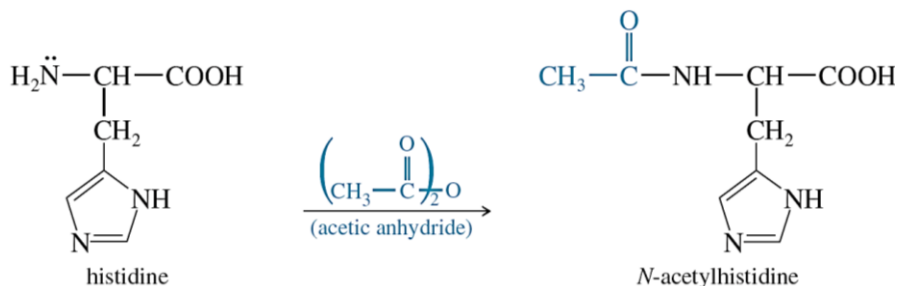


The reaction is used to protect carboxyl group from reacting in some undesired manner. Methyl, ethyl, and benzyl esters are the most common protecting groups. After the desired reaction of amino group is complete, the free amino acid can be regenerated by aqueous acid hydrolysis.

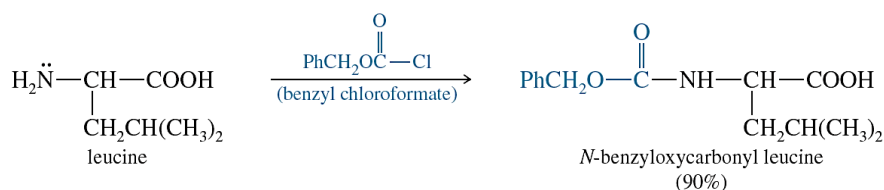


Acylation of amino group (formation of amides)

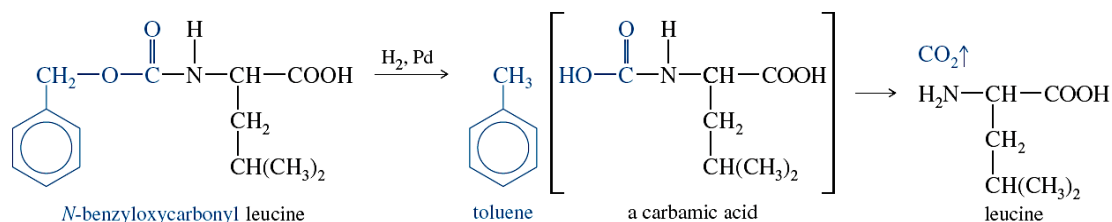
Amino group of amino acids react with acid chlorides or acid anhydrides to convert amino group to an amide.



Acylation of the amino group is often done to protect it from unwanted nucleophilic reactions. A wide variety of acid chlorides or anhydrides are used for acylation. Benzyl chloroformate acylates the amino group to give a benzyloxycarbonyl derivative, often used as a protecting group in peptide synthesis.

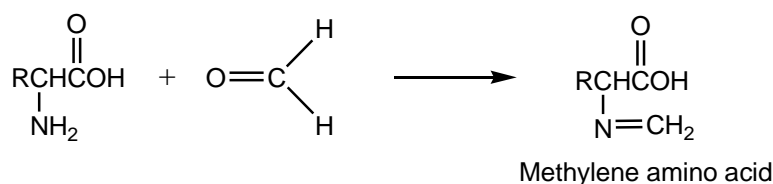


Catalytic hydrogenolysis of the *N*-benzyloxycarbonyl amino acid gives an unstable carbamic acid that quickly decarboxylates to give the deprotected amino acid.



Reaction with formaldehyde (formation of methylene amino acids)

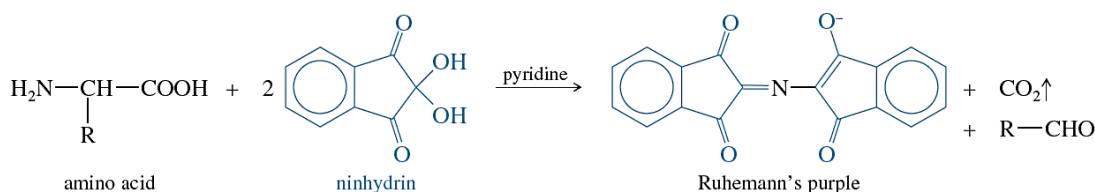
The amino group of amino acids reacts with formaldehyde with the elimination of water molecule, resulting in the formation of methylene amino acids.



The reaction is used for estimation of amino acid by direct titration with alkali. This method is known as *Sorensen's Formol Titration Method*. The direct titration of amino acid with alkali cannot be carried out because of the presence of free -NH_2 group. Reaction with formaldehyde converts amino group to methylene amino acid, which is a strong acid and can be titrated directly with an alkali.

Reaction with ninhydrin

Ninhydrin is a common reagent for visualizing spots or bands of amino acids that have been separated by chromatography or electrophoresis. Amino acids react with ninhydrin (hydrate of indane-1,2,3-trione) to give a deep violet, resonance-stabilized anion called *Ruhemann's purple*. Ninhydrin produces this same purple dye regardless of the structure of the original amino acid. The side chain of the amino acid is lost as an aldehyde.



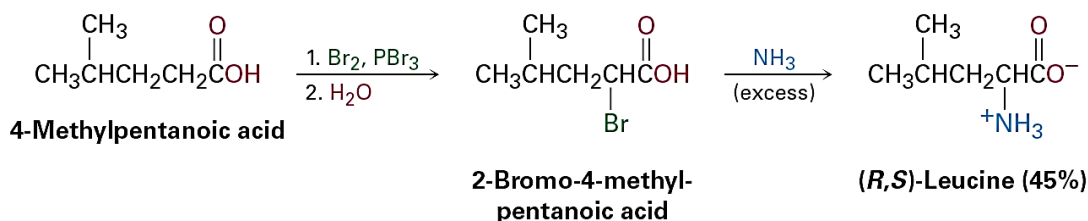
5.3.3. Synthesis of Amino Acids

The amino acids can be synthesized by a number of methods. However, a single method cannot be used for the preparation of all amino acids. The different methods are used extensively for preparation of different kinds of amino acids.

5.3.3.1. Amination of α -haloacids

This is one of the oldest methods to synthesize racemic α -amino acids. At first, the carboxylic acids are converted to α -bromo acids by Hall-Vallard-Zelinsky method using chlorine or bromine in presence of PBr_3 . This results in functionalization of α -position. The α -bromo acids on treatment with excess of

ammonia result in the substitution of bromine by -NH_2 via $\text{S}_{\text{N}}2$ reaction to yield α -amino acids. Glycine, alanine, valine, leucine, and aspartic acid can be prepared by this method.



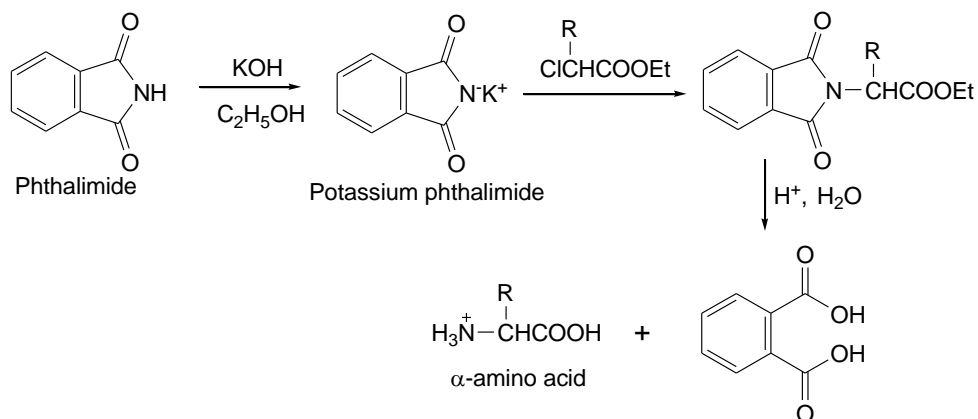
The reaction of ammonia with alkyl halide often results in polyalkylation. However, in this case, polyalkylation is not a problem because the alkyl halide is fairly large and steric hindrance prevents subsequent alkylations.

Unfortunately, this method gives relatively low yield of the product. Gabriel's procedure is better for the preparation of primary amines.

5.3.3.2. Gabriel phthalimide synthesis

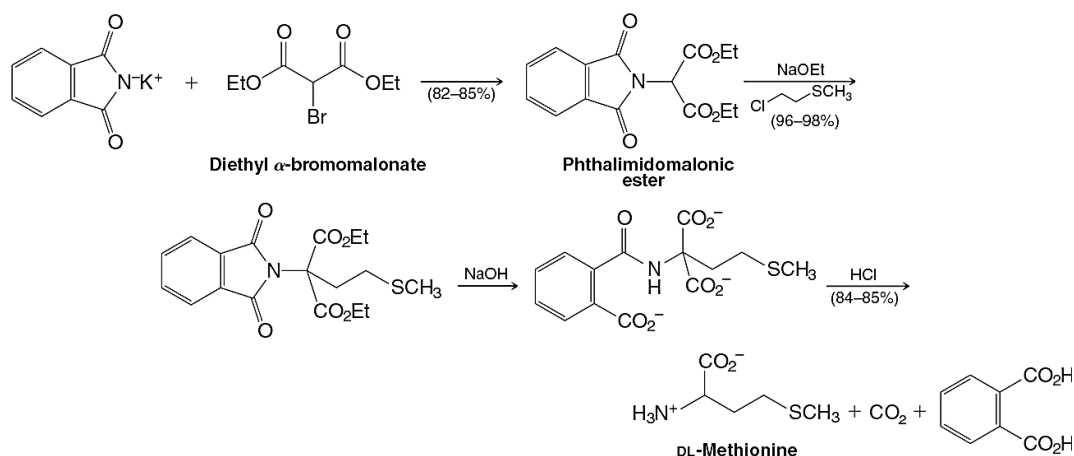
Phthalimide is converted to its reactive salt by reaction with ethanolic KOH. The potassium salt of phthalimide on reaction with α -halo acids (α -halo esters are preferred), followed by hydrolysis results in the formation of α -amino acids.

This method gives the product in high yield as compared to that obtained by the amination of α -halo acids. Glycine and leucine are best prepared by this method.



5.3.3.3. Phthalimido malonic ester synthesis

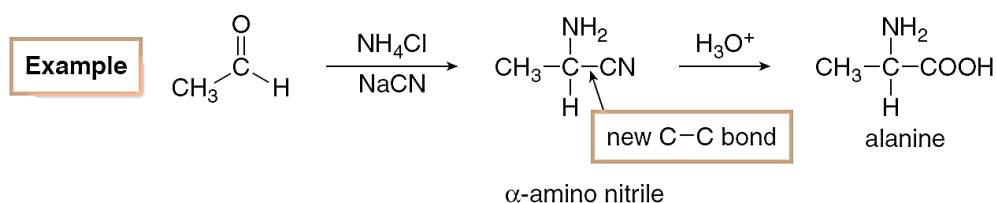
This method is modification of Gabriel phthalimide method and involves the reaction of potassium phthalimide with α -bromomalonate to produce phthalimidomalonate. The sodium salt of ester on reaction with an α -haloester followed by hydrolysis results in the formation of α -amino acid.



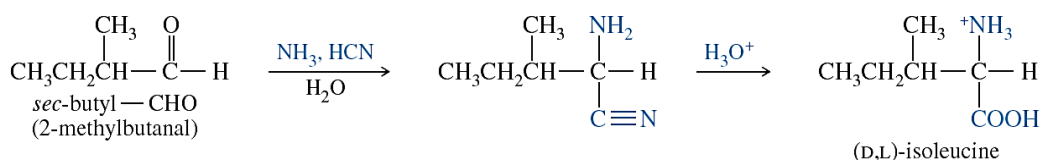
This method is used for the synthesis of methionine, and acidic amino acids and hydroxyl amino acids such as serine, glutamic acid and aspartic acid.

5.3.3.4. Strecker synthesis

Racemic α -amino acids can also be prepared from aldehydes via a two-step process called the **Strecker synthesis**. The first step involves conversion of the aldehyde into a α -amino nitrile by the reaction of an aldehyde with NaCN in presence of ammonia or ammonium chloride, and the second step involves hydrolysis of the cyano group to yield a carboxylic acid group. The reaction involves the conversion of aldehyde to α -amino acid having one more carbon atom than the parent aldehyde.



The **Strecker synthesis** can be used to synthesize a large number of amino acids from appropriate aldehydes. For example, as we have seen above that CH_3CHO undergoes Strecker synthesis to give alanine with CH_3 as side chain, similarly, as isoleucine has *sec*-butyl side chain, *sec*-butyl-CHO can be used to get isoleucine.



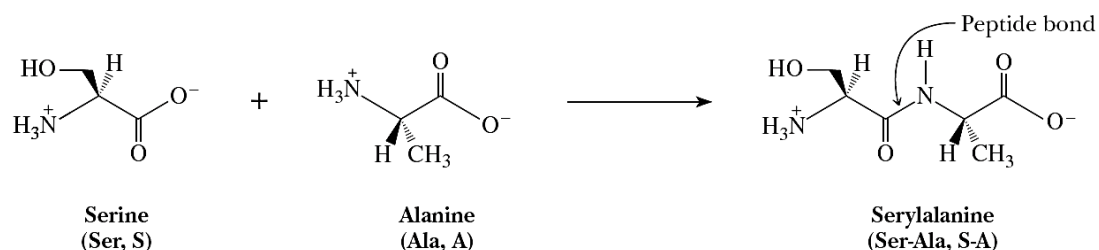
Resolution

All the above methods for synthesis of amino acids yield racemic mixture of amino acids. As the naturally occurring amino acids exist as a single enantiomer (L-isomer), two enantiomers obtained must be separated if they are to be used in biological applications. As two enantiomers have same physical and chemical

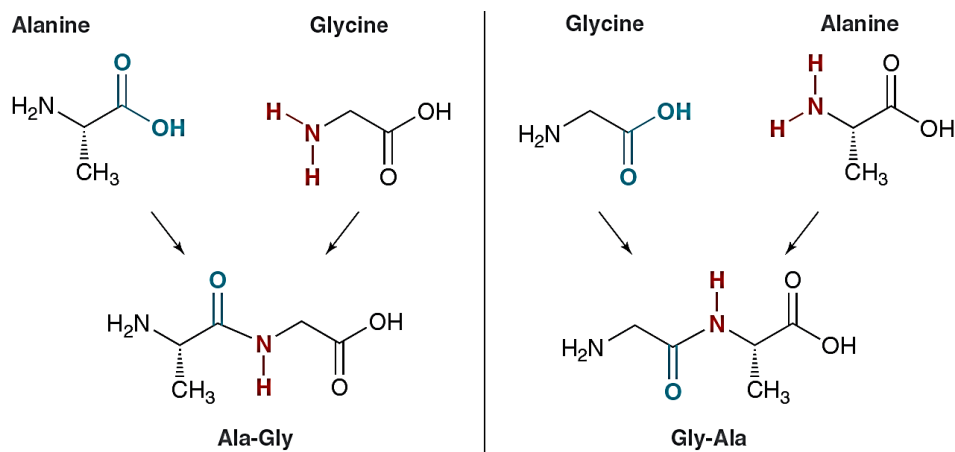
properties, they cannot be separated by distillation, chromatography, or even by chemical reactions. However, two diastereomers have different physical properties. The separation of a **racemic mixture into its component enantiomers is called resolution**. The strategy to separate two enantiomers is that they are converted first to diastereomers by the reaction with a chiral molecule followed by the separation of diastereomers by physical methods (fractional distillation or crystallisation), and then re-convert diastereomers to enantiomers. The resolution of racemic amino acids is typically done by protecting first the amino group as an amide and the resulting product is then treated with an optically active amine, such as the inexpensive alkaloid brucine. The two diastereomers thus formed are separated by fractional crystallisation. Thereafter, brucine is removed and amide is hydrolysed to get pure enantiomers.

5.3.4. Peptides & Proteins

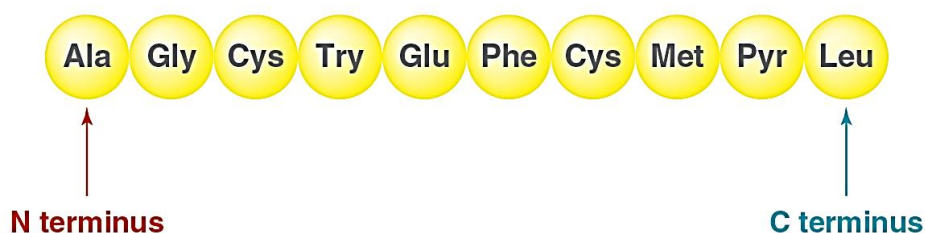
Peptides are long chains of amino acids joined together by amide bonds, known as peptide bonds, between the α -carboxyl group of one amino acid and the α -amino group of another. Amines and acids condense, with the loss of water, to form amides.



Peptide is the name given to a short polymer of amino acids. The long, repetitive sequence of $-N-CH-CO-$ atoms that makes up a continuous chain is called the peptide's (or protein's) backbone. When amino acids join together to form a peptide, the order in which they are connected is important. For example, consider a simple dipeptide made by joining alanine and glycine. The peptide bond can be formed between the $COOH$ group of alanine and the NH_2 group of glycine, or from the $COOH$ group of glycine and the NH_2 group of alanine.



By convention, peptides and proteins are written with the free amino group on the left and the free carboxyl group on the right. The left and right ends of the peptide are referred to as the **N terminus** (or amino terminus) and the **C terminus** (or carboxyl terminus), respectively. Individual amino acid components of peptides are often referred to as amino acid **residues**. The sequence of amino acid residues in a peptide can be abbreviated with one- or three letter abbreviations, starting with the N terminus. For example,



Depending on the number of amino acids present in a molecule, they are classified as **dipeptide** which contains two amino acid residues; **tripeptide** contains three amino acid residues; **oligopeptide** contains 2-10, and **polypeptide** contains many amino acid residues. Proteins are biological macromolecules consisting of one or more polypeptides chains that are made up of 40 to 4000 amino acid residues, thus ranging in molecular weight from about 5000 kiloDaltons (e.g. insulin) to 40 million kiloDaltons (e.g. tobacco mosaic virus protein).

- (i) **Fibrous proteins.** The molecules of fibrous proteins have long thread like structures that are joined to each other through hydrogen bonding. Fibrous proteins are insoluble in water and act as main supporting structural material of tissues. For example, keratin, collagen, myosin, fibroin, and so on.
- (ii) **Globular Proteins:** The molecules of globular proteins have folded structure that resembles a spherical shape. The intermolecular attraction between different molecules is weak. Globular proteins are water soluble and perform the function of maintenance and regulation of life processes. Most of the enzymes are globular proteins. For example, Haemoglobin, myoglobin, insulin, albumin, fibrinogen, and so on.

On the basis of solubility and physical properties, proteins are classified into two categories:

- (a) **Simple proteins.** Proteins which on complete hydrolysis result in the formation of amino acids only are called simple proteins. Example, albumin, globulin, keratin, elastin, collagen, etc.
- (b) **Conjugated proteins.** Apart from amino acids, these proteins contain a non-protein group known as *prosthetic group* such as a sugar, a nucleic acid, a lipid, or some other group. The protein part of conjugated proteins is termed as *apoprotein*. For example,

Class	Prosthetic Group	Examples
glycoproteins	carbohydrates	γ -globulin, interferon
nucleoproteins	nucleic acids	ribosomes, viruses
lipoproteins	fats, cholesterol	high-density lipoprotein
metalloproteins	a complexed metal	hemoglobin, cytochromes

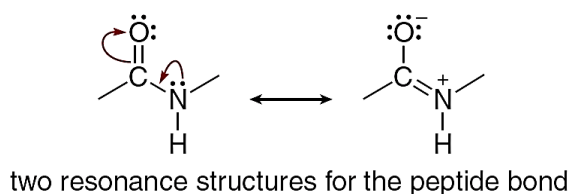
Structure of peptide and proteins

Proteins are fairly large organic compounds, and as such, the structure of a protein is more complex than the structures of simple organic compounds. In fact, proteins are generally described in terms of four levels of structure: primary, secondary, tertiary, and quaternary structure (if a protein has more than one polypeptide chain).

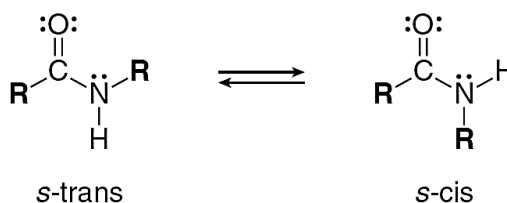
5.3.5. The Primary Structure of Peptides

The **primary structure** of a protein is the sequence of amino acid residues plus any disulfide links. In primary structure, the peptide bonds form the backbone and the side chains of amino acids project outside the peptide backbone.

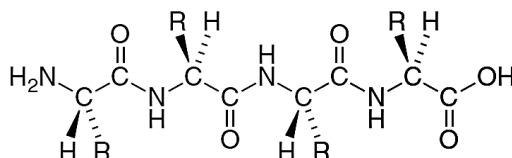
The carbonyl carbon of an amide (peptide bond) is **sp^2 hybridized** and has **trigonal planar** geometry. The non-bonded electron pair on the nitrogen atom is delocalized by the interaction with the carbonyl group.



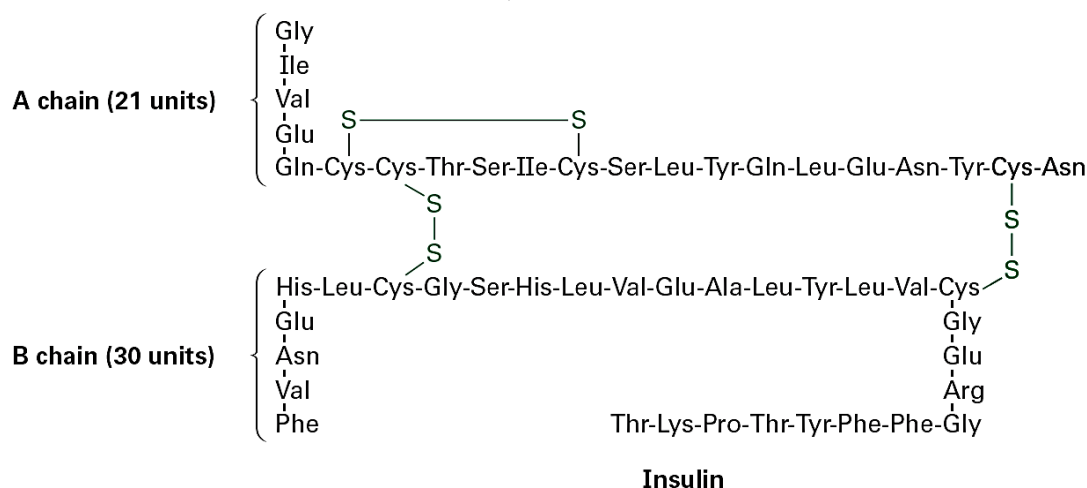
This resonance imparts a certain amount of double-bond character to the C–N bond and restricts rotation around it. As a result, there are two possible conformations – *s-trans* and *s-cis*.



The *s-trans* conformation of a peptide bond is typically more stable than the *s-cis*, because the *s-trans* has the two bulky R groups located farther from each other. The *s-trans* arrangement makes a long chain of amino acids in zigzag arrangement. A second consequence of resonance stabilization is that **all six atoms involved in the peptide bond lie in the same plane**. All bond angles are $\sim 120^\circ$ and the C–O and N–H bonds are oriented 180° from each other.



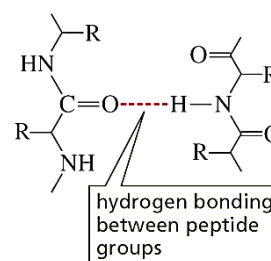
A disulfide bond between cysteine residues in different peptide chains links the separate chains together, while a disulfide bond between cysteine residues in the same chain forms a loop. For example, insulin is comprised of two polypeptide chains, called chains A and B, which are linked together by disulfide bridges.



5.3.6.

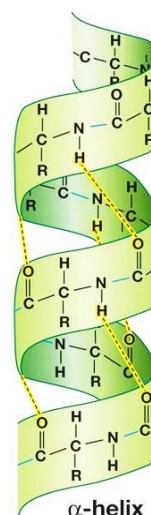
5.3.7. Secondary & Tertiary Structure of Large Peptides and Proteins

The three-dimensional conformations of localized regions of a protein are called its secondary structure. It results from the rigidity of peptide bond and from hydrogen bonds between the N–H proton of one peptide bond and C=O oxygen of another. Secondary structure usually takes the form of α -helix (spiral) or β -pleated sheet.



α -Helix. The most commonly found secondary structure of proteins is α -helix and almost all globular proteins contain many helical segments. An α helix is a right-handed or clockwise coil of the protein backbone, much like the coil of a spiral staircase. The important features of α -helix are:

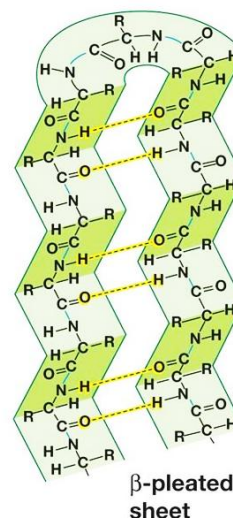
- Each turn of the helix contains 3.6 amino acid residues, with a distance between coils of 540 pm, or 5.4 Å.
- The helical structure is stabilized by intra-molecular hydrogen bonding that is formed between C=O group of one amino acid and N-H group that is four residues farther along on the chain, with N-H \cdots O distance of 2.8 Å. Thus, almost all C=O are hydrogen bonded to N-H proton.
- These hydrogen bonds are parallel to the axis of the helix, and the N-H and C-O bonds point along the axis of the helix. All C=O bonds point roughly upward, and all N-H bonds point in the opposite direction (downward).
- The R groups of the amino acids extend outward from the core of the helix.
- A proline residue lacks an N-H proton, therefore does not participate in hydrogen bonding. Thus, proline cannot be part of α -helix. The presence of proline is often associated with a bend in the peptide chain.



Example: The amino acids, which occur commonly in α -helix conformations, are alanine, leucine, and phenylalanine. The proteins like myosin in muscle and α -keratin in hair are composed almost entirely of α -helices.

β -Pleated sheet. The β -pleated sheet secondary structure forms when two or more peptide chains, called strands, line up side-by-side. The important features of β -pleated sheet structures are:

- The hydrogen bonding occurs between C=O oxygen and N-H proton of two or more adjacent extended polypeptide chains. The adjacent chains in β -pleated sheet structures are either parallel or anti-parallel, although the antiparallel arrangement is more common and energetically somewhat more favourable.
- The C-O and N-H bonds lie in the plane of the sheet.
- The **R groups are oriented alternate above and below the plane** of the sheet.



The β -pleated sheet arrangement most commonly occurs with amino acids with small R groups, like alanine and glycine. With larger R groups steric interactions prevent the chains from getting close together and so the sheet cannot be stabilized by hydrogen bonding.

Example: The β -keratin protein, which occurs in spider's web and reptilian claw, is a fully extended chain of β -pleated sheet. The silk fibroin protein (occurs in silk worms) is also rich in β -pleated sheet structure.

The **tertiary structure** of a protein refers to its three-dimensional shape. Proteins typically adopt conformations that maximize stability. In aqueous solution, a protein will adopt a conformation in which the polar groups are pointing toward the exterior of the protein and the nonpolar groups are located in the interior.

Because the tertiary structure of a globular protein is delicately held together by weak intramolecular attractions, a modest change in temperature or pH is often enough to disrupt that structure and cause the protein to become denatured. Loss of tertiary structure is typically accompanied by a loss of function, because the function of a protein is very much dependent on its three-dimensional shape. The denaturation process is irreversible in most cases.

5.4. Nucleic Acids

5.4.1. Introduction to the chemistry of Nucleic Acids (Structure and Chemistry)

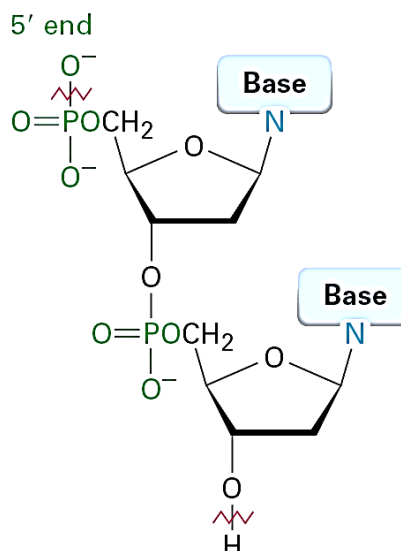
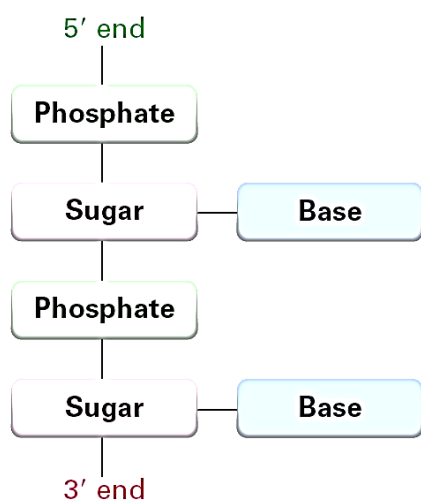
Nucleic acids are macromolecules present in the living organisms. They control the synthesis of protein and are required for the storage and expression of genetic information. There are two types of nucleic acids found in living organisms:

- (i) Deoxyribonucleic acid (DNA)
- (ii) Ribonucleic acid (RNA)

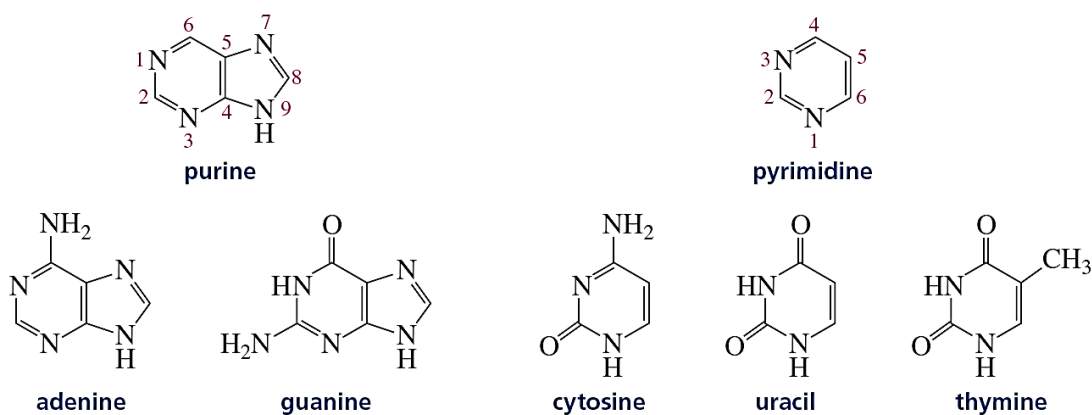
DNA molecules encode the “blueprints” for the manufacture of proteins. They are also involved in the continuance of genetic information from generation to generation. RNA molecules are involved in the manufacture of proteins through the processes of transcription and translation.

Structure of nucleic acids

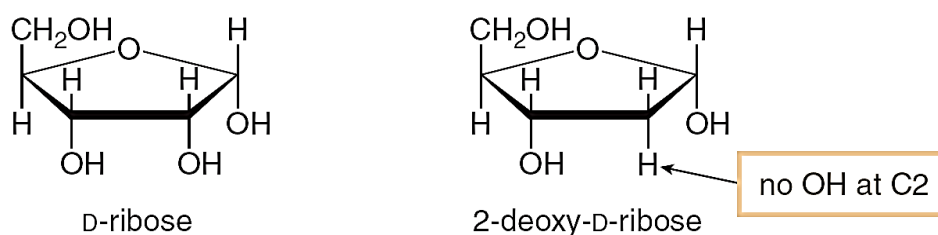
The nucleic acids are unbranched polymers of repeating monomer units known as **nucleotides**. A DNA molecule may contain several million nucleotides; smaller RNA molecules may contain up to several thousand. The complete hydrolysis of nucleic acids produces heterocyclic base, sugar residues and phosphoric acid. The basic structure of nucleic acids consists of a polymer chain (polyester) of phosphoric acid and sugar residue attached alternatively through **ester linkage**. The anomeric carbon of each sugar is bonded to a nitrogen of a heterocyclic base through **β -glycosidic linkage**. Thus, sugar residue and phosphoric acid are attached alternatively to make a long chain and heterocyclic base is attached to sugar residue as a side chain.



Nucleic acids contain four heterocyclic bases – two pyrimidine derivatives and two purine derivatives. The pyrimidine derivatives include cytosine, uracil, and thymine while purine derivatives are adenine and guanine. DNA contains adenine, guanine, cytosine and thymine. RNA also contains adenine, guanine and cytosine, but thymine is replaced by another pyrimidine base i.e. uracil. Thymine and uracil have the same structure; they differ only by a methyl group at C-5 position. Thymine is 5-methyluracil.



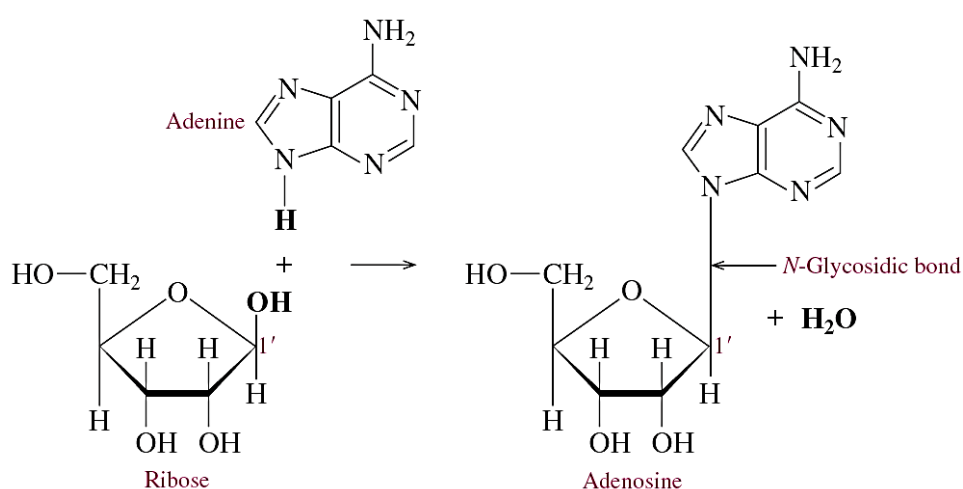
The sugar component in nucleic acids is a pentose sugar, which is D-ribose in RNA, and 2'-deoxy-D-ribose in DNA. The atoms in the pentose sugars are numbered with primes (1', 2', 3', 4' and 5') to differentiate them from the atoms in the bases. Thus the prefix 2'-deoxy in DNA indicates that oxygen is missing from the C-2' of ribose.



Nucleosides

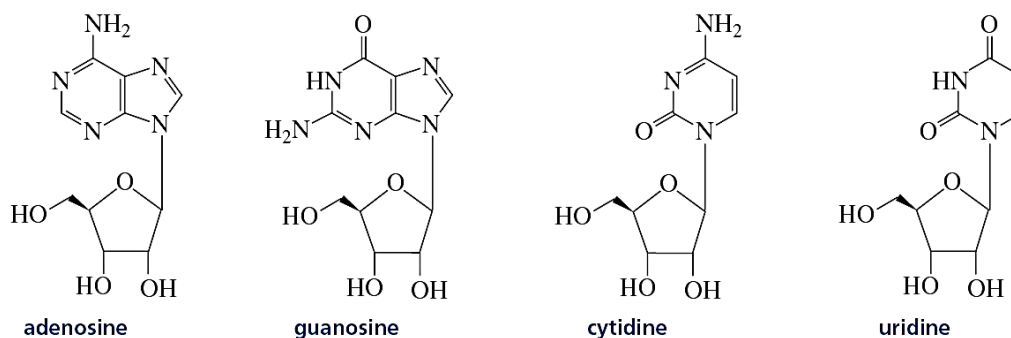
Nucleosides are *N*-glycosides of heterocyclic pyrimidine and purine bases. The anomeric carbon of ribose or 2'-deoxyribose sugar is bonded to the nitrogen atom of a heterocyclic pyrimidine or purine base by a β -*N*-glycosidic linkage. The sugar component of DNA is 2'-deoxy-D-ribose, whereas that of RNA is D-ribose. Thus, each nucleoside has two parts, a sugar and a base, joined together by a β -*N*-glycosidic linkage.

- In purine nucleoside, the C-1 of sugar is attached to N-9 of purine base.
- In pyrimidine nucleoside, the C-1 of sugar is attached to N-1 of pyrimidine base.

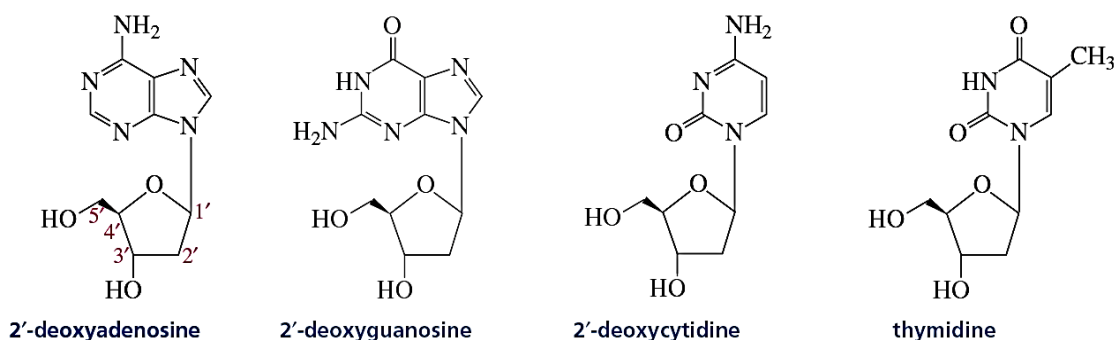


Nucleosides are named on the basis of the type of sugar and bases present. For example, adenosine (adenine + ribose), guanosine (guanine + ribose), cytidine (cytosine + ribose) and uridine (uracil + ribose) in RNA, and 2'-deoxyadenosine (adenine + 2'-deoxyribose), 2'-deoxyguanosine (guanine + 2'-deoxyribose), 2'-deoxycytidine (cytosine + 2'-deoxyribose), thymidine (thymine + 2'-deoxyribose) in DNA.

Four nucleosides that make up RNA:



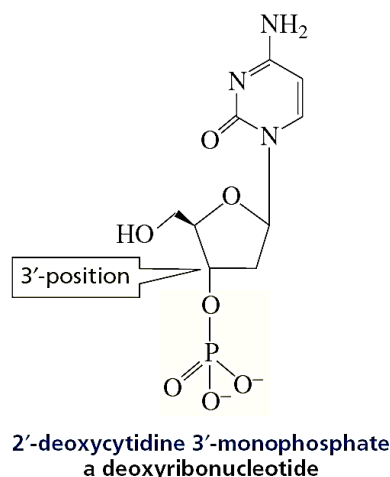
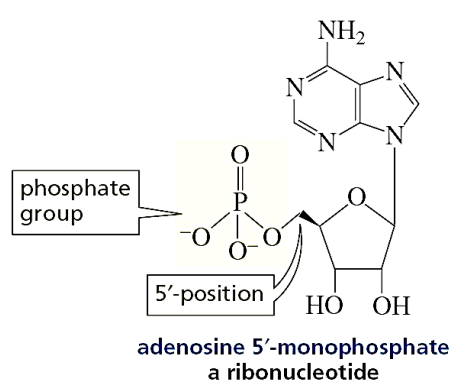
Four nucleosides that make up DNA:



Nucleotides

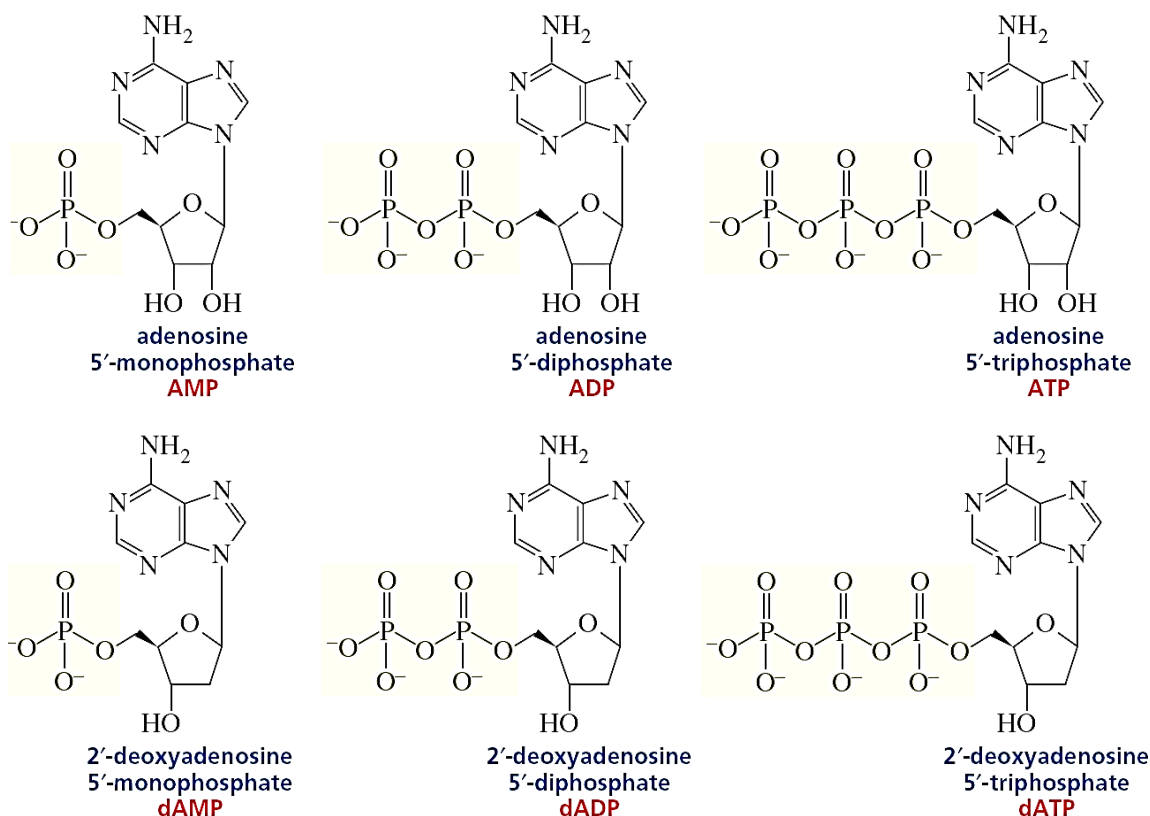
Nucleotides are phosphoric acid esters of nucleosides – a free hydroxyl group of the sugar of nucleoside is attached to a molecule of phosphoric acid by ester linkage.

- A deoxyribose sugar of nucleoside can attach to phosphoric acid through free –OH present at C-3' or C-5' to form ester linkage.
- A ribose sugar of nucleoside can attach to phosphoric acid through free –OH present at C-2', C-3' or C-5' to form ester linkage.



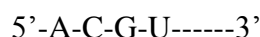
The nucleotides of RNA – where the sugar is D-ribose – are more precisely called ribonucleotides, whereas the nucleotides of DNA – where the sugar is 2'-deoxy-D-ribose – are called deoxyribonucleotides. A nucleotide is named by giving the name of the parent nucleoside followed by the word “monophosphate.” For example, adenosine 5'-monophosphate (AMP) and 2'-deoxycytidine 3'-monophosphate:

Nucleoside monophosphates can be further phosphorylated to form nucleoside diphosphates and nucleoside triphosphates.



Primary structure of nucleic acids

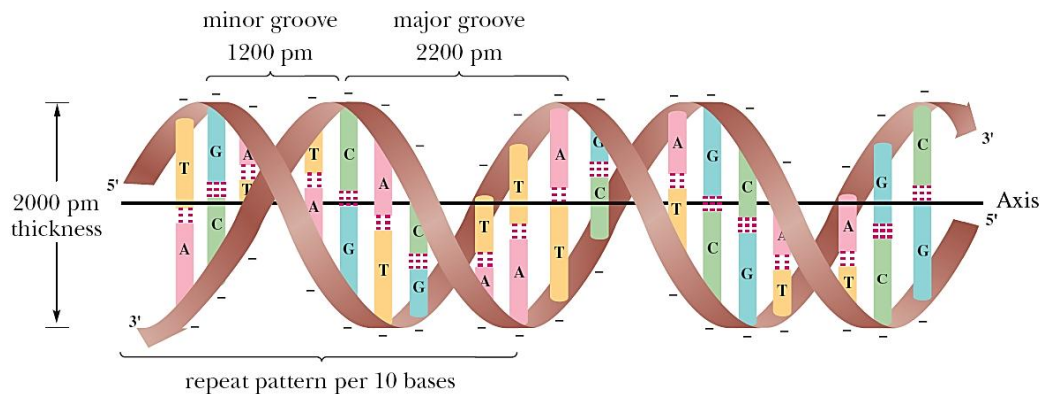
A dinucleotide contains two nucleotide subunits, an oligonucleotide contains three to ten subunits, and a polynucleotide contains more than ten nucleotides. Nucleic acids (DNA and RNA) are polynucleotides in which the 3'-OH group of the sugar in one nucleotide bonds to the phosphate group on the 5'-carbon atom in the sugar of the next nucleotide. This phosphate link between the sugars in adjacent nucleotides is referred to as a **phosphodiester bond**. As more nucleotides are added using phosphodiester bonds, a backbone forms that consists of alternating sugar and phosphate groups. The bases, which are attached to each sugar, extend out from the sugar-phosphate backbone. Thus, the nucleotide at one end of the strand has an unlinked 5'-phosphate group, and the nucleotide at the other end of the strand has an unlinked 3'-OH group. Each nucleic acid has its own unique sequence of bases, which is known as its **primary structure**. A nucleic acid sequence is read from the sugar with free 5'-phosphate to the sugar with the free 3'-hydroxyl group. The order of nucleotides is often written using only the letters of the bases. For example,



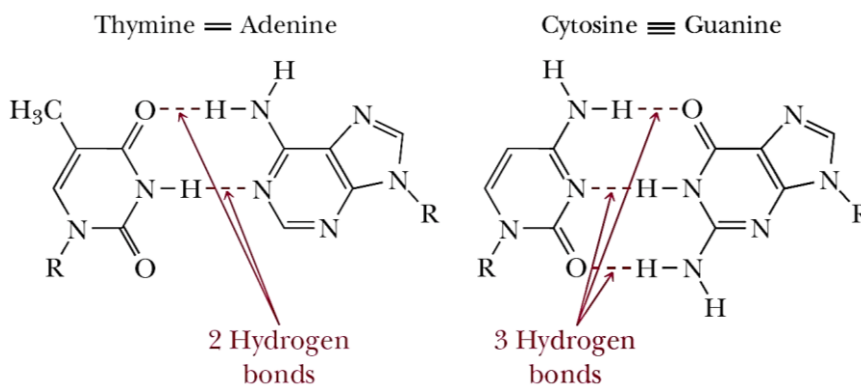
Secondary structure – the double helix

Watson and Crick postulated that a molecule of DNA (deoxyribonucleic acid) is a complementary double helical structure. It consists of two antiparallel polynucleotide strands coiled in a right-handed manner about the same axis to form a double helix. The

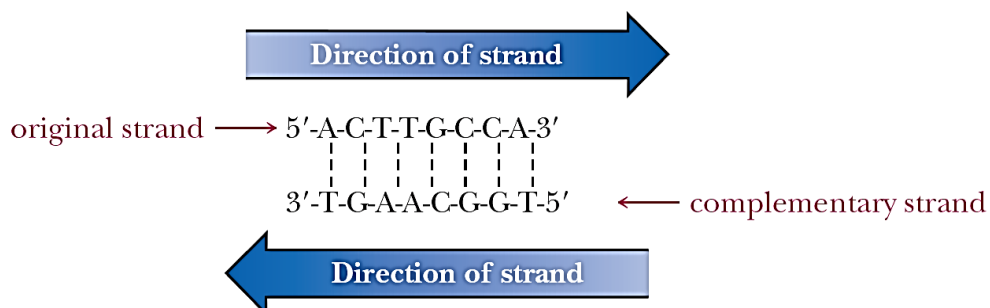
helices are held together by hydrogen bonds between the bases on the two strands. The bases in one strand are attached to complementary bases on the other strand through hydrogen bonding. The base pairs are found in the interior of the helix while charged and hydrophilic sugar-phosphate backbone is on the exterior.



Adenine (A) always pairs with thymine (T) through two hydrogen bonds while cytosine (C) pairs up with guanine (G) by three hydrogen bonds.



The base sequence is always written from the 5' end of the strand to the 3' end, that A pairs with T, and that G pairs with C. In double-stranded DNA, the two strands run in opposite (antiparallel) directions so that the 5' end of one strand is associated with the 3' end of the other strand.



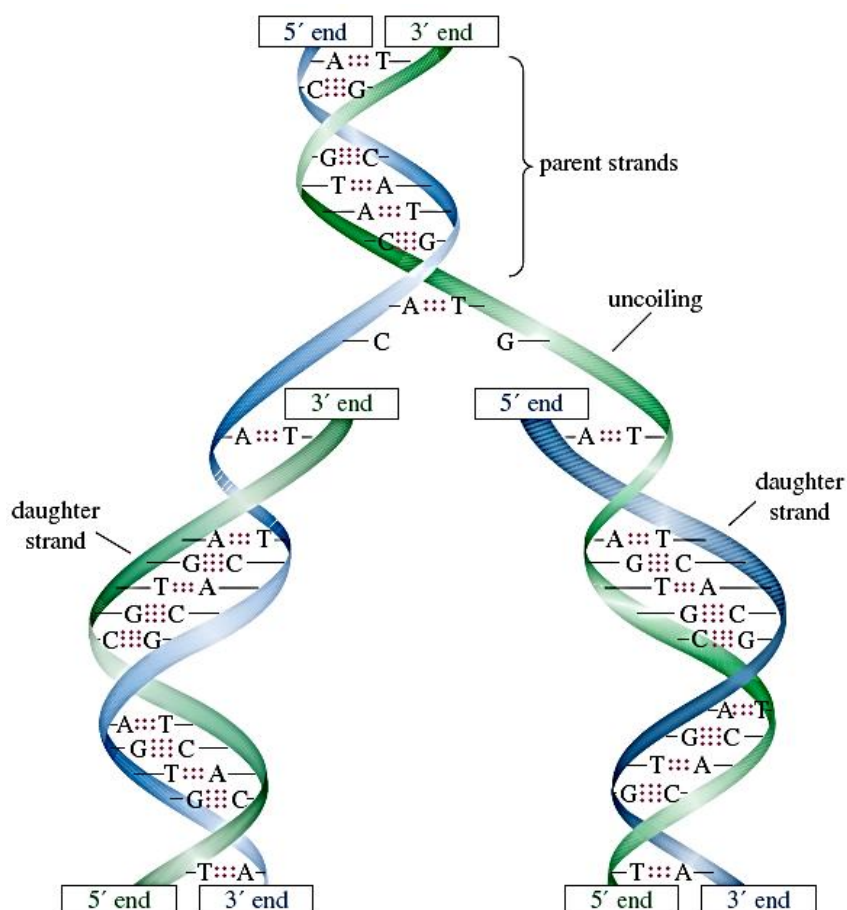
The structure of RNA (ribonucleic acid) is similar to that of DNA as it is also made up of a long, unbranched chains of nucleotides joined by phosphodiester groups between the 3'-OH of one ribose sugar and the 5'-OH of the next. There are, however, three major differences between the structure of DNA and RNA:

- RNA contains β -D-ribose sugar while DNA has β -2'-deoxy-D-ribose sugar component.
- Thymine in DNA is replaced by uracil in RNA.
- RNA exists as a single stranded helix rather than double-stranded DNA.

There are three types of RNA, namely messenger RNA (m-RNA), transfer RNA (t-RNA), and ribosomal RNA (r-RNA).

DNA replication

The DNA replication is the key to pass on the genetic information from one generation to the next. During cell division, the double helix of DNA unwinds into two strands – each acting as a template for the synthesis of new complementary DNA strands. In each new DNA molecule, one strand of the double helix is from the original DNA and one is a newly synthesized strand. This process produces two new DNAs called *daughter DNAs* that are identical to each other and exact copies of the original parent DNA. In the process of DNA replication, complementary base pairing ensures the correct placement of bases in the new DNA strands.



Thus, during cell division, exact copies of DNA are produced that transfer genetic information to the new cells.

Protein synthesis

DNA and RNA both are involved in protein synthesis. DNA stores all genetic information which is passed on to RNA. RNA, in turn, reads the information and decodes it to synthesize protein. Protein synthesis undergoes through two processes:

Transcription: Transcription means making copy. A copy of DNA fragment (known as gene) is formed in the form of mRNA (messenger RNA) in which deoxyribose sugar is replaced by ribose sugar and thymine is replaced by uracil. This process undergoes in nucleus. Thereafter, mRNA moves from nucleus to cytoplasm.

Translation: Translation means to change the language. mRNA has the genetic information in the form of sequence of nucleotides. This information is translated to the sequence of amino acids with the help of tRNA (transfer RNA) to make protein (a long chain of amino acids).

Summary

- Biomolecules are polymers of different monomers found in living organisms and perform different metabolic activities. These are mainly classified into four major groups – carbohydrates, lipids, proteins, and nucleic acids.
- **Carbohydrates** are polyhydroxylated aldehydes or ketones or compounds that produce polyhydroxylated aldehydes or ketones upon hydrolysis.
- They exist as monomer, dimer, oligomer, and polymers which are named as monosaccharides, disaccharides, oligosaccharides, and polysaccharides, respectively.
- Monosaccharides are classified according to the number of carbon atoms and the kind of carbonyl group they contain. Those contain aldehyde group are classified as **aldoses**; those have ketonic function are **ketoses**. Aldoses and ketoses are further classified according to the number of carbon atoms they contain – triose, tetrose, pentose, hexose, and so on.
- Monosaccharides are further classified as either D sugars or L sugars, depending on the stereochemistry of the chiral carbon farthest from the carbonyl group. Most natural carbohydrates belong to the D family; that is, the stereocenter farthest from the carbonyl group has the same configuration as that in D-(+)-glyceraldehyde].
- Stereoisomers that are mirror images – have opposite configuration at each carbon – are called enantiomers.
- Stereoisomers that are not mirror images are called diastereoisomers or diastereomers.
- Stereoisomers having opposite configuration only at one carbon are called epimers.
- The open-chain monosaccharides are in equilibrium with a cyclic hemiacetal structure. The cyclic hemiacetal is greatly favored and thus is found in large excess at equilibrium.
- Six-membered hemiacetal ring is called **pyranose ring**, and five-membered hemiacetal ring is called a **furanose ring**.
- Cyclization leads to the formation of a new chirality center called the anomeric center and the production of two diastereomeric hemiacetals called alpha (α) and beta (β) anomers.
- Cyclic monosaccharide structures are often drawn as **Haworth projections** in which the five-membered or six-membered cyclic hemiacetal is drawn as planar and perpendicular to the plane of the paper. The anomeric carbon is usually to the right with the hemiacetal oxygen atom in the back.
- In the α anomer, the hydroxyl group at the anomeric position is *trans* to the CH₂OH group, while in the β anomer, the hydroxyl group is *cis* to the CH₂OH group.
- A more accurate “chair” conformation can be drawn for six-membered ring hemiacetals, showing which groups are axial and which are equatorial.
- **Mutarotation** is the change in specific rotation that accompanies the interconversion of α - and β -anomers in aqueous solution. It is catalysed by either acid or base.
- Carbohydrates contain carbonyl and hydroxyl groups and thus, show the typical reactions of these functions. For example, the hydroxyl groups form ethers and esters. The carbonyl group of a monosaccharide can be reduced with NaBH₄ to form an

alditol, oxidized with aqueous Br_2 to form an aldonic acid, oxidized with HNO_3 to form an aldaric acid, or treated with an alcohol in the presence of acid to form a glycoside.

- Aldoses and ketoses react with three equivalents of phenylhydrazine, forming **osazones**.
- Sugars containing hemiacetal functions are called **reducing sugars**, because they readily reduce Tollens's and Fehling's solutions. Sugars in which the anomeric carbon has acetal function are nonreducing. Aldoses and ketoses are reducing sugars.
- Disaccharides are comprised of two monosaccharide units joined together via a glycosidic linkage. For example maltose, lactose, and sucrose. **Sucrose** is a disaccharide containing D-glucose joined to D-fructose by an α -1,2-glycosidic bond. **Lactose** is a disaccharide consisting of D-galactose joined to D-glucose by a β -1,4-glycosidic bond. **Maltose** is a disaccharide of two molecules of D-glucose joined by a β -1,4-glycosidic bond.
- Polysaccharides are polymers consisting of repeating monosaccharide units linked by glycoside bonds. For example starch, cellulose, and glycogen. Starch is made up of two types of polysaccharides – amylose and amylopectin. **Amylose** is a linear polymer of up to 4000 units of D-glucopyranose joined by α -1,4-glycosidic bonds. **Amylopectin** is a highly branched polymer of D-glucopyranose joined by α -1,4-glycosidic bonds and, at branch points, by α -1,6-glycosidic bonds
- Glycogen is similar to amylopectin, but it is more branched than amylopectin.
- **Cellulose**, the skeletal polysaccharide of plants, is a linear polymer of D-glucopyranose joined by β -1,4-glycosidic bonds.
- **Lipids** are a heterogeneous class of naturally occurring compounds grouped together on the basis of their solubility properties; they are insoluble in water and soluble in nonpolar solvents.
- Lipids are of two types – hydrolysable lipids, e.g. fats and waxes, and non-hydrolysable lipids, e.g. terpenes, prostaglandins.
- **Fatty acids** are carboxylic acids with long hydrocarbon chains. Nearly all fatty acids have an even number of carbon atoms, usually between 12 and 20.
- Double bonds in fatty acids have the *cis* configuration.
- Fatty acids with more than one double bond are called polyunsaturated fatty acids (PUFA). Double bonds in naturally occurring unsaturated fatty acids are not conjugated.
- Melting point of unsaturated fatty acids is lower than those of saturated fatty acids. The melting point decreases as the number of double bonds increases.
- Animal fats and vegetable oils are triesters of glycerol with long chain carboxylic acids, are also called triglycerides.
- Triglycerides with unsaturated fatty acid residues generally have lower melting points than triglycerides with saturated fatty acid residues. Triglycerides that are solids at room temperature are called fats, while those that are liquids at room temperature are

called oils. Animal fats are usually saturated, whereas vegetable oils usually have unsaturated fatty acid residues.

- When treated with aqueous base, triglycerides undergo hydrolysis, also called saponification.
- **Waxes** are esters formed from long-chain carboxylic acids and long-chain alcohols.
- Phospholipids are ester like derivatives of phosphoric acid. Phosphoglycerides are similar in structure to triglycerides except that one of the three fatty acid residues is replaced by a phosphoester group.
- The simplest kind of phosphoglyceride is a phosphoric monoester, called a phosphatidic acid.
- Phosphoglycerides that contain ethanolamine are called cephalins, while phosphoglycerides that contain choline are called lecithins.
- **Prostaglandins** are powerful regulators of biochemical processes and biosynthesized from arachidonic acid.
- Prostaglandins contain 20 carbon atoms and are characterized by a five-membered ring with two side chains.
- Terpenes are a class of naturally occurring compounds that contain carbon atoms in multiples of 5. They are made by joining together five-carbon isoprene units, usually in a head-to-tail fashion—the **isoprene rule**.
- Terpenoids are often isolated from the essential oils of plants, have an immense diversity of structure, and are produced biosynthetically from the five-carbon precursor isopentenyl diphosphate (IPP).
- Depending on the number of carbon atoms or number of isoprene units, terpenes are classified as monoterpenes containing 10 carbon atoms (or two isoprene units), sesquiterpenes with 15 carbon atoms (or three isoprene units), diterpene with 20 carbon atoms (or four isoprene units), and so on.
- Steroids are plant and animal lipids with a characteristic tetracyclic carbon skeleton. The steroid skeleton is rigid and relatively flat owing to a common *trans-trans-trans* orientation of ring fusions.
- Steroids are closely related to terpenoids and arise biosynthetically from the triterpenoid.
- The most abundant member of the steroid family in animals is **cholesterol**, the precursor of all other steroids.
- Proteins and peptides are polymers of amino acids.
- Twenty amino acids are commonly found in proteins. Each amino acid contains one amino group and one carboxylic acid group. Amino acids in which the two functional groups are separated by exactly one carbon atom are called alpha amino acids.
- The **amino acids** differ only in the substituent attached to the α -carbon.
- Except for glycine, which is achiral, all of the α -amino acids present in proteins are chiral and have the L configuration at the α -carbon according to D,L-convention.
- In R,S-convention, all amino acids have S-configuration, except cysteine, which has R-configuration.

- Isoleucine and threonine have two chiral centres.
- All amino acids have primary amino group, except proline, which contains secondary amino group.
- The most stable structure of a neutral amino acid is a **zwitterion**. The **isoelectric point** (pI) of an amino acid is the pH at which the amino acid has no net charge.
- Amino acids are said to be amphoteric, because they can function either as acids or as bases.
- Amino acids can be synthesized in racemic form by several methods, including amination of an α -bromo acid, Gabriel phthalimide synthesis, phthalimido malonic ester synthesis, and Strecker synthesis. Enantiomerically pure amino acids can be made by classical fractional crystallization of diastereomeric derivatives.
- Peptides are comprised of amino acid residues joined by peptide bonds.
- A dipeptide contains two amino acid residues, a tripeptide contains three, an oligopeptide contains three to 10, and a polypeptide contains many amino acid residues. Proteins have 40 to 4000 amino acid residues.
- Peptide chains have an amino group called the N terminus and a COOH group called the C terminus.
- Peptide bonds experience restricted rotation, giving rise to two possible conformations, called *s-trans* and *s-cis*. The *s-trans* conformation is generally more stable.
- Cysteine residues are uniquely capable of being joined to one another via disulfide bridges.
- The primary structure of a protein is the sequence of amino acid residues plus any disulfide bonds between two cysteine residues.
- The secondary structure of a protein refers to the three-dimensional conformations of localized regions of the protein. Two particularly stable arrangements are the α -helix and β -pleated sheet. The α -helix is stabilized by hydrogen bonds between C=O and N-H groups within a single polypeptide chain, while the β -pleated sheet is stabilized by hydrogen bonds between C=O and N-H groups of adjacent chains.
- The tertiary structure of a protein refers to its three-dimensional shape. The tertiary structure has a tremendous influence on the properties of the peptide and the biological role it plays.
- Under conditions of mild heating or change in pH, a protein can unfold, a process called denaturation.
- Proteins are classified as either globular or fibrous. Fibrous proteins such as α -keratin are tough, rigid, and water-insoluble; globular proteins such as myoglobin are water-soluble and roughly spherical in shape.
- DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) that act as chemical carriers of an organism's genetic information, are biological polymers of monomeric units called nucleotides.

- DNA contains four bases – adenine, cytosine, guanine, and thymine. In RNA thymine is replaced by uracil. Thus, RNA contains adenine, cytosine, guanine, and uracil. Adenine and guanine are purines; cytosine, thymine, and uracil are pyrimidines.
- Each of these four bases can couple to 2-deoxyribose (in DNA), or to ribose sugar (in RNA), giving rise to four possible deoxyribonucleosides (in DNA) or ribonucleosides (in RNA).
- Each of these four nucleosides can then be coupled to a phosphate group through either 3'-OH or 5'-OH, giving compounds called nucleotides, the phosphate esters of nucleosides. Thus, nucleotides are comprised of three parts: deoxyribose (or ribose), a nitrogen-containing base, and a phosphate group. When linked together, they serve as the building blocks of DNA (or RNA), a polynucleotide.

Nucleoside = base + Sugar

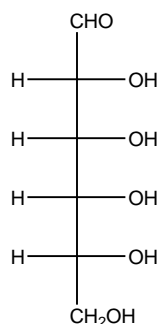
Nucleotide = base + Sugar + Phosphate

Nucleic acid (DNA or RNA) = polynucleotides

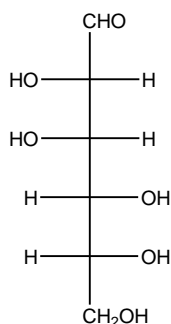
- The nucleotides are joined by phosphate links between the 5' phosphate of one nucleotide and the 3' hydroxyl on the sugar of another nucleotide.
- DNA exists as a double stranded pair of helices in which hydrogen bonds are responsible for complementary base pairing between adenine (A) and thymine (T), and between guanine (G) and cytosine (C).
- In the **transcription** stage of protein biosynthesis a molecule of **messenger RNA** (mRNA) having a nucleotide sequence complementary to that of DNA is assembled. Transcription is followed by **translation**, in which triplets of nucleotides of mRNA called **codons** are recognized by **transfer RNA** (tRNA) for a particular amino acid, and that amino acid is added to the growing peptide chain.

5.5. Exercises

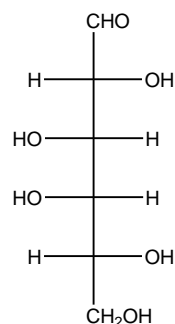
- 5.5.1. Write the Haworth structures and conformational structures of the following sugars:



I



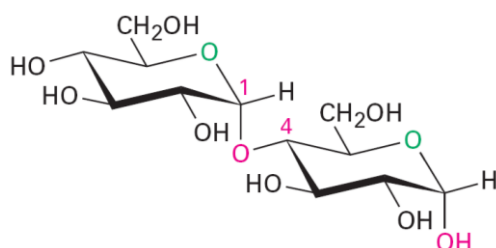
II



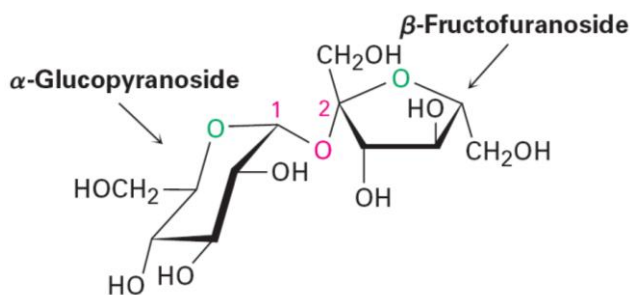
III

- 5.5.2. What is the meaning of the prefix *deoxy-* as it is used in carbohydrate chemistry?
 5.5.3. Find out the hemiacetal, hemiketal, acetal and ketal groups, if any, in the following compounds:

(a)

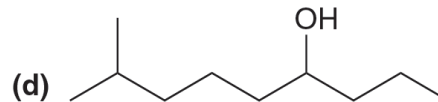
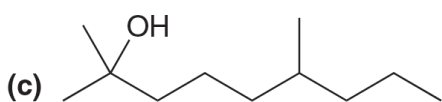
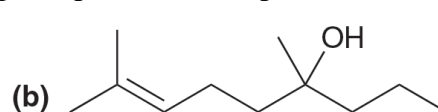
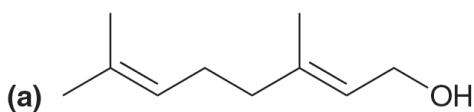


(b)

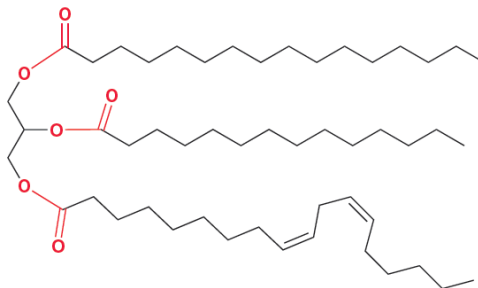


- 5.5.4. Identify each of the following fatty acids as saturated or unsaturated:
 (a) Palmitic acid (b) Myristic acid (c) Oleic acid
 (d) Lauric acid (e) Linoleic acid (f) Arachidonic acid

- 5.5.5. Determine whether each of the following compounds is a terpene:



- 5.5.6. Identify the products expected when the following triglyceride is hydrolyzed with aqueous base:



- 5.5.7. Explain primary, secondary and tertiary structures of protein.
- 5.5.8. Which bonds in the backbone of a peptide can rotate freely?
- 5.5.9. A mixture of seven amino acids (glycine, glutamate, leucine, lysine, alanine, isoleucine, and aspartate) is separated by TLC. Explain why only six spots show up when the chromatographic plate is sprayed with ninhydrin and heated.
- 5.5.10. If one of the strands of DNA has the following sequence of bases running in the 5'→3' direction,

5' G-G-A-C-A-A-T-C-T-G-C 3'

What is the sequence of bases in the complementary strand?