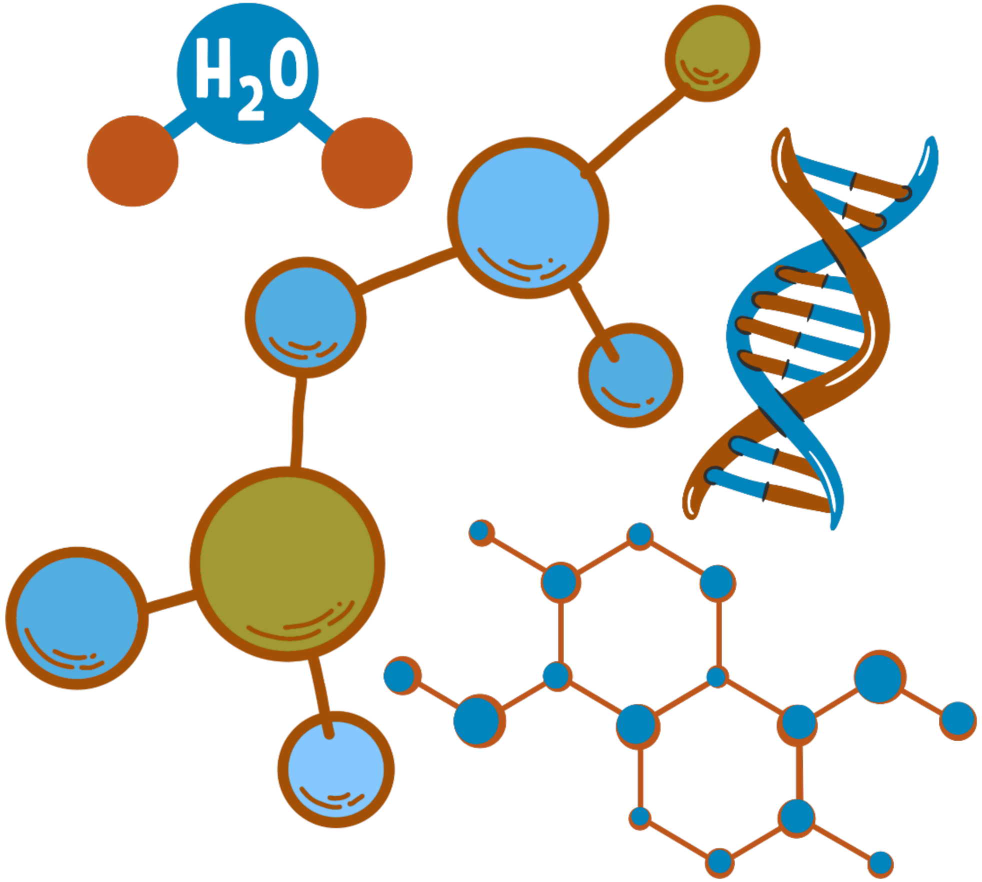
Introduction to Organic Chemistry
CHEM 109

جامعة الملك سعود
King Saud University
COLLEGE OF SCIENCES
CHEMISTRY DEPARTMENT

Chapter 6:
Biomolecules and Their Biological Relevance



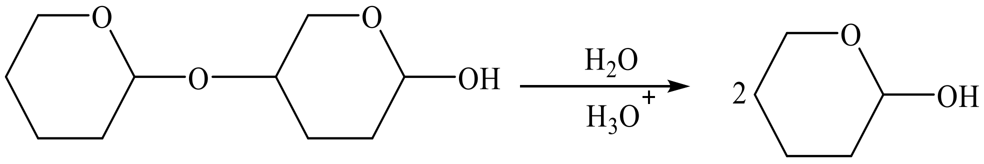
Carbohydrates

6.1.1 Carbohydrates

- **Carbohydrates** are polyhydroxy aldehydes or ketones, or compounds that can be hydrolyzed to produce such molecules.
- Carbohydrates typically follow the general chemical formula $Cx(H_2O)y$.

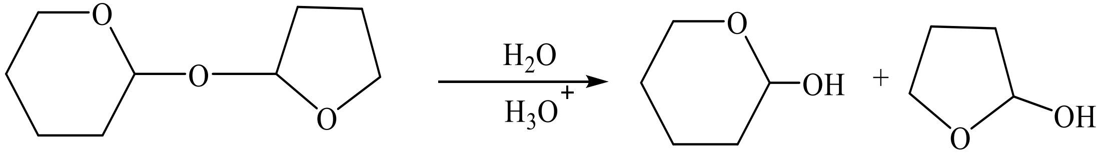
6.1.1.1 Classification of Carbohydrates

- **Monosaccharides** are the simplest carbohydrates and cannot be further hydrolyzed into smaller carbohydrate units.
- **Disaccharides** can be broken down by hydrolysis into two monosaccharide units.
- **Oligosaccharides** produce between two and ten monosaccharides upon hydrolysis.
- **Polysaccharides** yield more than ten monosaccharides when hydrolyzed.



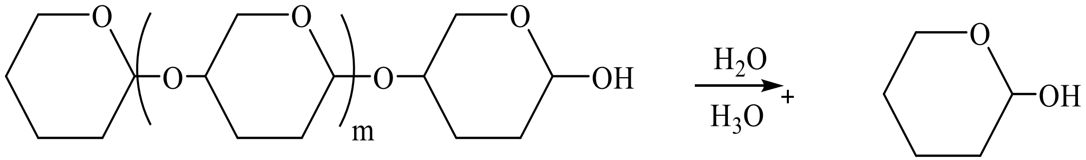
1 mol of maltose
A disacchride

2 mol of glucose
A monosacchride



1 mol of sucrose
A disacchride

1 mol of glucose + 1 mol of fructose
Monosacchrides



1 mol of starch or
1 mole of cellulose
polysacchride

many moles of glucose
Monosacchride

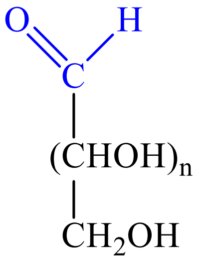
6.1.2 Monosaccharides

6.1.2.1 Classification of Monosaccharides

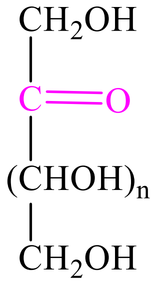
- Monosaccharides are classified based on two main features:
  1. The number of carbon atoms they contain.
  2. The type of carbonyl group present — either an aldehyde or a ketone group.

6.1.2.2 D and L Designations of Monosaccharides

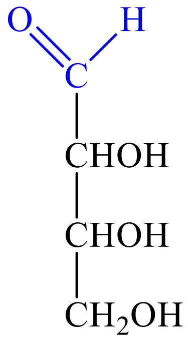
- **Chirality** is the property of a molecule that makes it non-superimposable on its mirror image, usually due to the presence of a chiral (asymmetric) carbon atom bonded to four different groups.
- The simplest carbohydrates are glyceraldehyde, which is chiral and assigned as L or D-configuration, and dihydroxyacetone, which is achiral.



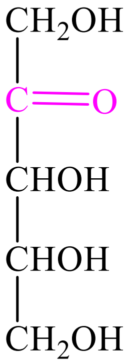
An aldose



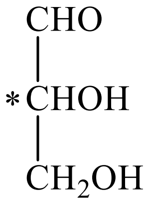
A Ketose



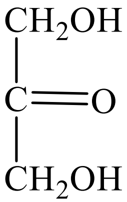
An aldotetrose
C4



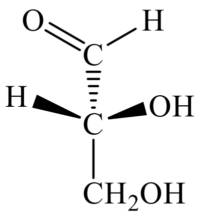
A Ketopentose
C5



Glyceraldehyde
(an aldotriose)

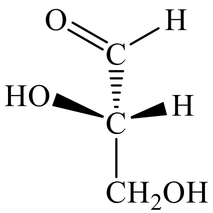


Dihydroxyacetone
(a ketotriose)



(+)-Glyceraldehyde

and

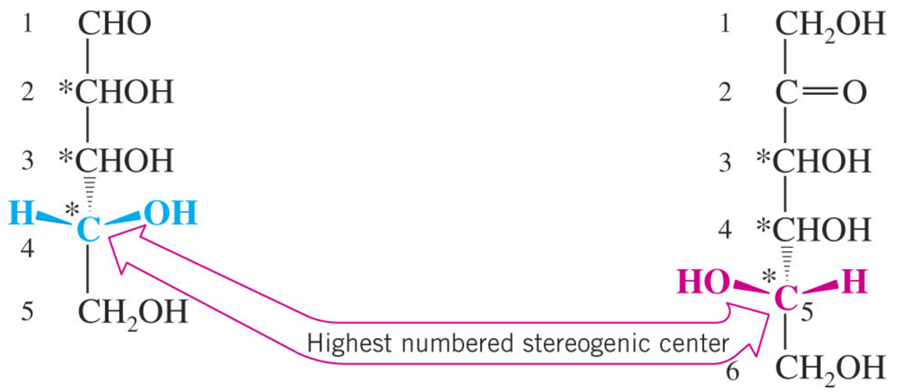


(-)-Glyceraldehyde

6.1.2 Monosaccharides

6.1.2.2 D and L Designations of Monosaccharides

- A monosaccharide is classified as a D-sugar if its highest-numbered chiral center has the same configuration as D-(+)-glyceraldehyde.
- A monosaccharide is identified as an L-sugar if its highest-numbered chiral center shares the same configuration as L-(–)-glyceraldehyde.

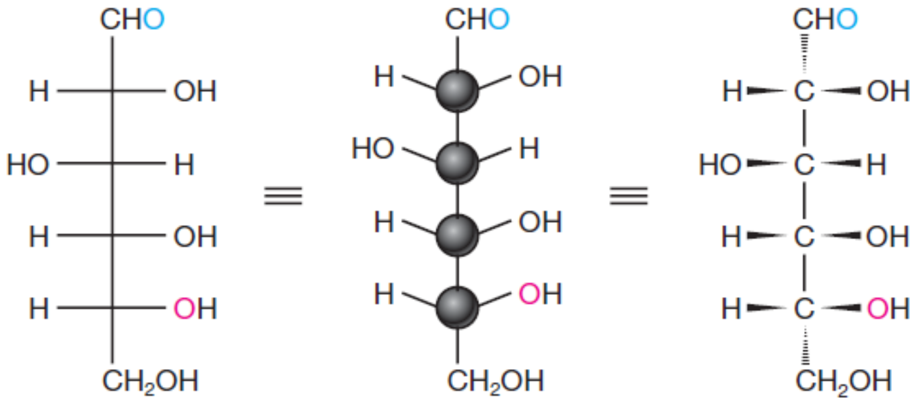


A D-aldopentose

An L-ketohexose

6.1.2.3 Structural Formulas for Monosaccharides

- **Fischer projection**
  - Represents sugar stereochemistry in the open-chain form.
  - Horizontal lines → bonds projecting out of the plane (toward the viewer).
  - Vertical lines → bonds behind the plane (away from the viewer).
  - The –CHO group is placed at the top, and the –CH_2OH group at the bottom.



Fischer projection formula

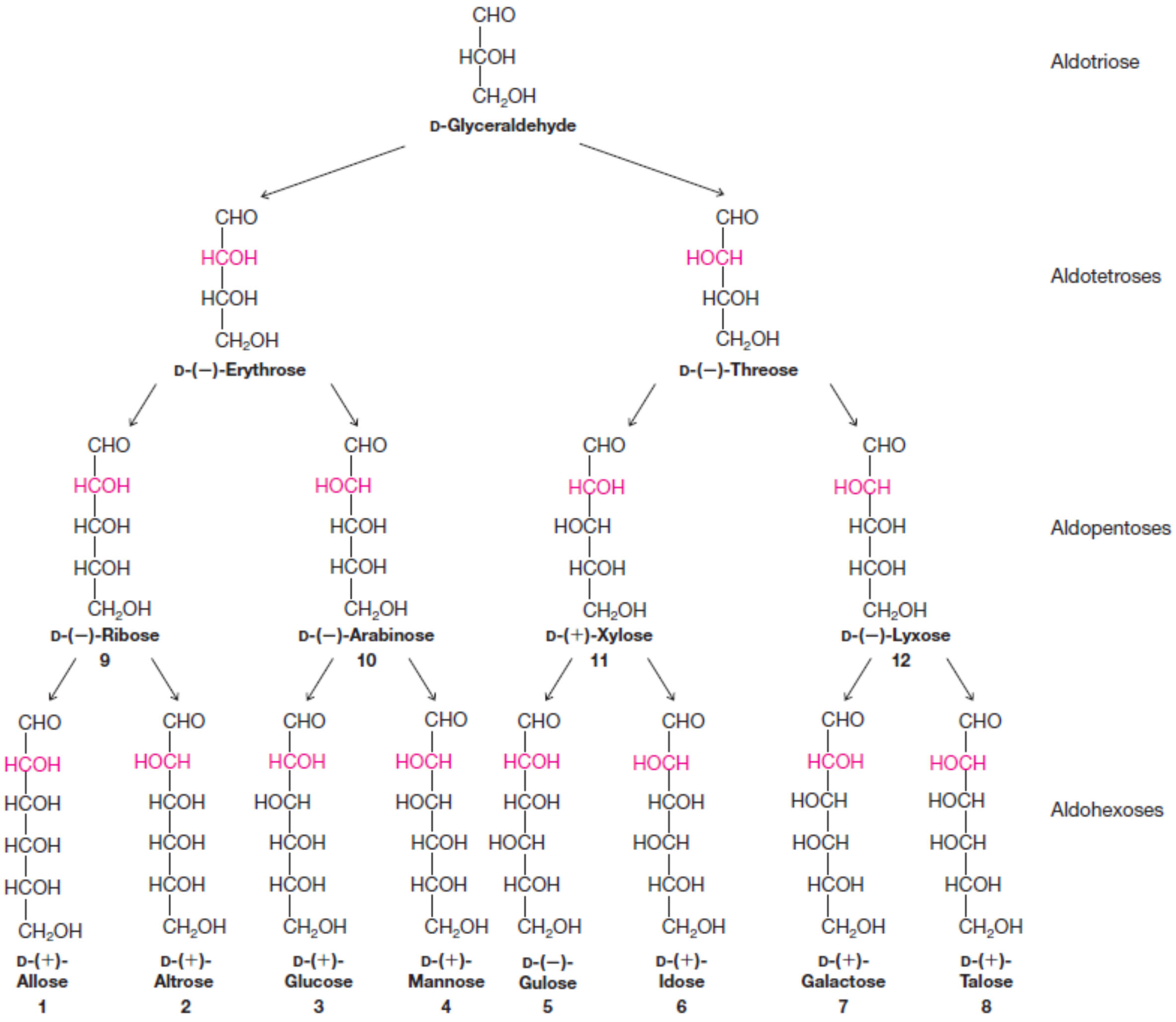
Circle-and-line formula

Wedge–line–dashed wedge formula

6.1.2 Monosaccharides

6.1.2.3 Structural Formulas for Monosaccharides

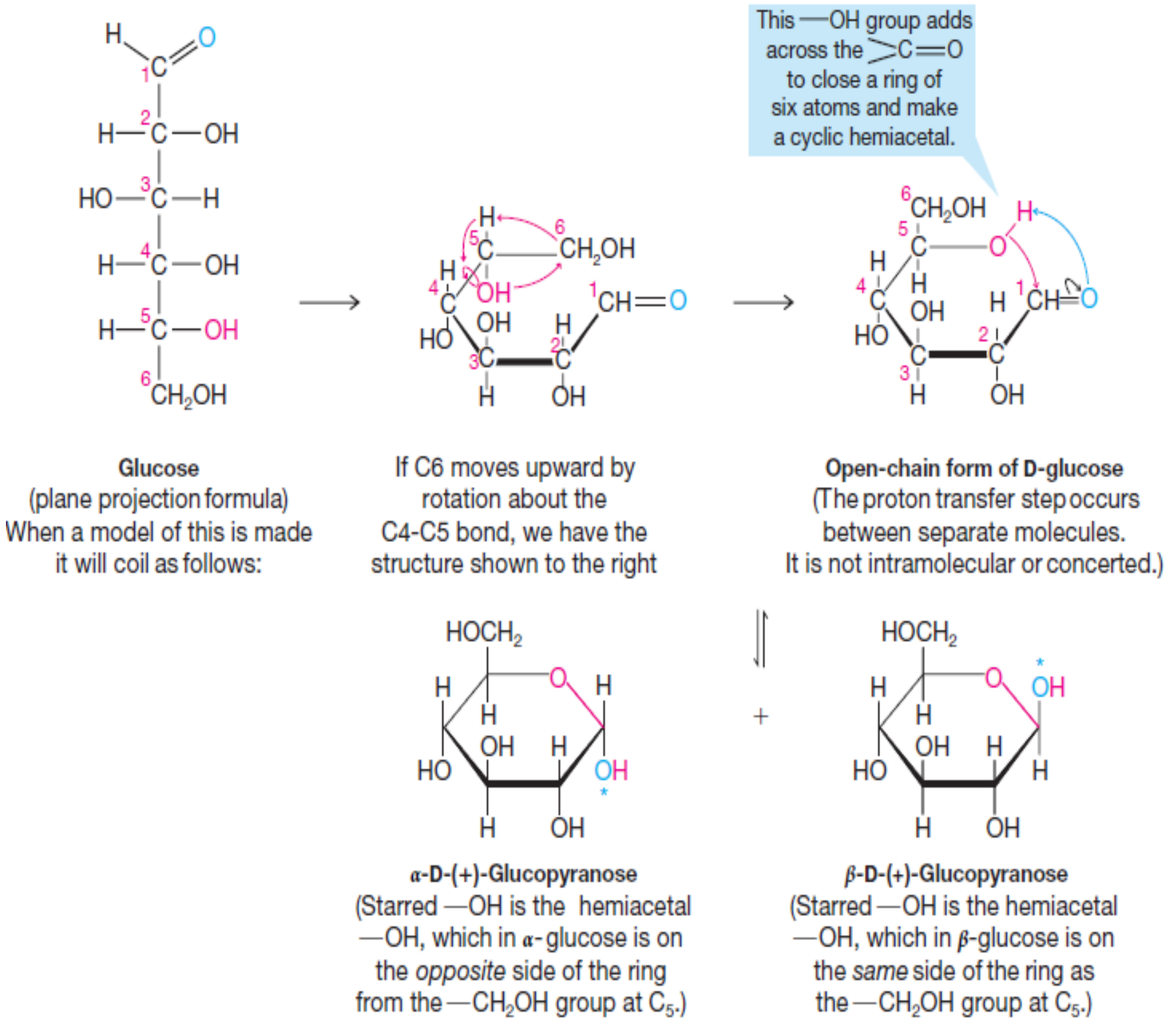
- The Ruff degradation and the Kiliani–Fischer synthesis allow us to place all of the aldoses into families or "family trees" based on their relation to D- or L-glyceraldehyde.
- Most, but not all, of the naturally occurring aldoses belong to the d family, with D-(+)-glucose being by far the most common.



6.1.2 Monosaccharides

6.1.2.3 Structural Formulas for Monosaccharides

- **Haworth formulas**
  - Represent sugar stereochemistry in the cyclic form, which exists in equilibrium with the open-chain form.
  - The cyclic (hemiacetal) forms are produced by an intramolecular reaction between the –OH group at C5 and the aldehyde group at C1.
  - Cyclization creates a new chiral center at C1, explaining the existence of two cyclic forms.
  - These two forms are diastereomers, differing only in the configuration at C1 (the anomeric carbon).
    - α-anomer → the hemiacetal –OH is *trans* to the –CH$_2$OH group.
    - β-anomer → the hemiacetal –OH is *cis* to the –CH$_2$OH group.



This —OH group adds across the >C=O to close a ring of six atoms and make a cyclic hemiacetal.

Glucose
(plane projection formula)
When a model of this is made it will coil as follows:

If C6 moves upward by rotation about the C4-C5 bond, we have the structure shown to the right

Open-chain form of D-glucose
(The proton transfer step occurs between separate molecules. It is not intramolecular or concerted.)

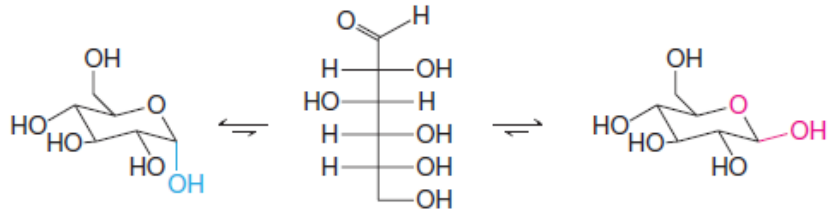
α-D-(+)-Glucopyranose
(Starred —OH is the hemiacetal —OH, which in α-glucose is on the *opposite* side of the ring from the —CH$_2$OH group at C$_5$.)

β-D-(+)-Glucopyranose
(Starred —OH is the hemiacetal —OH, which in β-glucose is on the *same* side of the ring as the —CH$_2$OH group at C$_5$.)

6.1.2 Monosaccharides

6.1.2.3 Structural Formulas for Monosaccharides

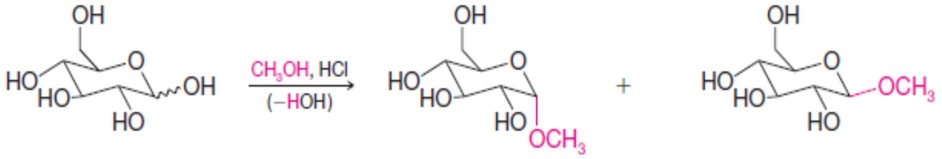
- **Mutarotation:** a change in specific rotation toward an equilibrium value.
  - It results from the equilibrium established between the open-chain form of D-(+)-glucose and its cyclic α- and β-hemiacetal forms.
  - In solutions: The concentration of open-chain is negligible, α-anomer 36%, and β-anomer 64%, reflecting a greater stability for β-D-(+)-glucopyranose due to the equatorial OH group on C1.
  - Carbohydrate acetals are generally called glycosides.
  - Reaction of D-(+)-glucose with methanol in presence of anhydrous (gaseous) hydrogen chloride yields anomeric methyl acetal.
  - An acetal of glucose is called a glucoside, acetals of mannose are mannosides, acetals of fructose are fructosides, and so on.)



α-D-(+)-Glucopyranose (mp 146 °C; $[\alpha]_D^{25} = +112$)

Open-chain form of D-(+)-glucose

β-D-(+)-Glucopyranose (mp 150 °C; $[\alpha]_D^{25} = +18.7$)



D-(+)-Glucose

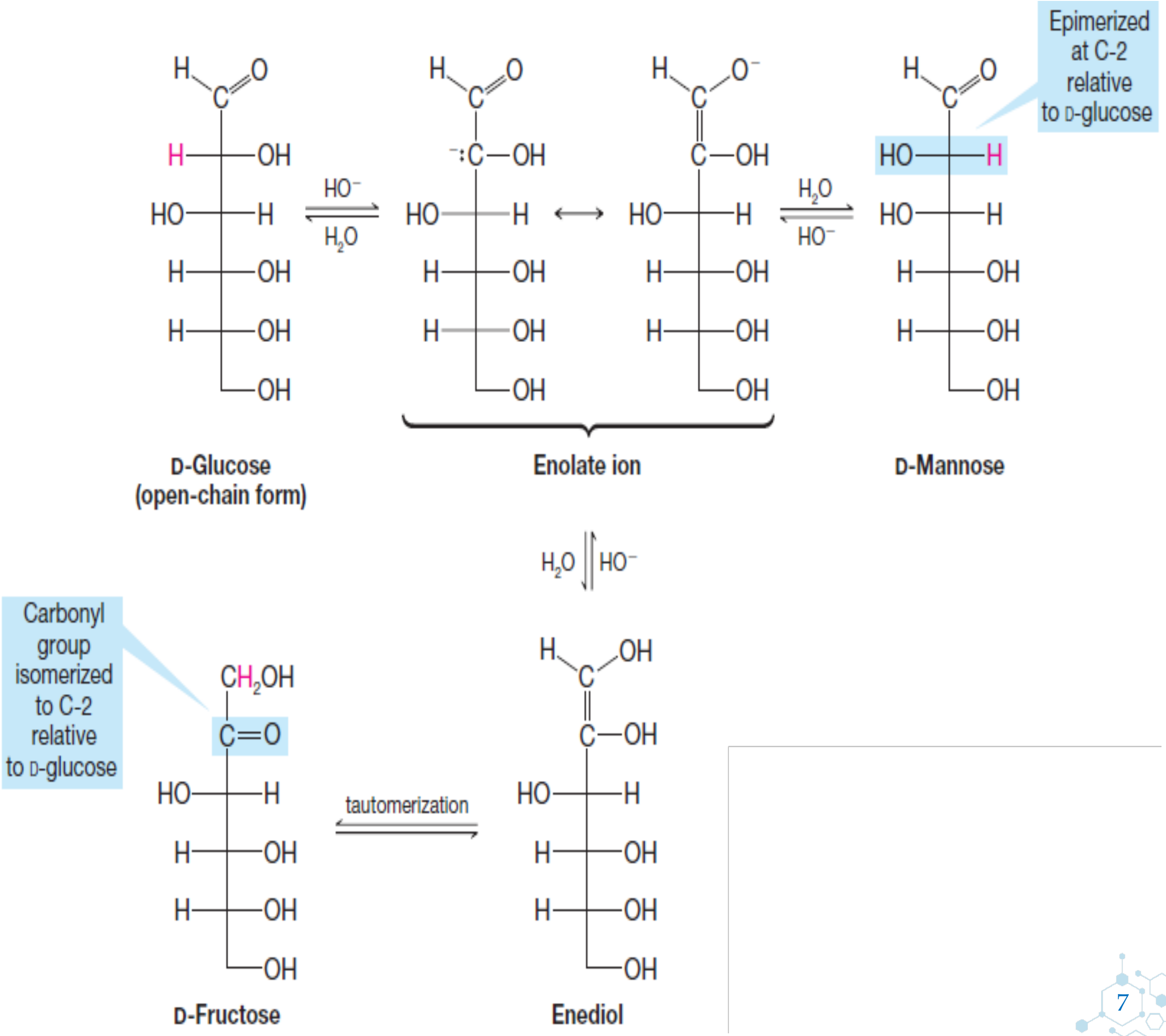
Methyl α-D-glucopyranoside (mp 165 °C; $[\alpha]_D^{25} = +158$)

Methyl β-D-glucopyranoside (mp 107 °C; $[\alpha]_D^{25} = -33$)

6.1.2 Monosaccharides

6.1.2.4 Isomerization

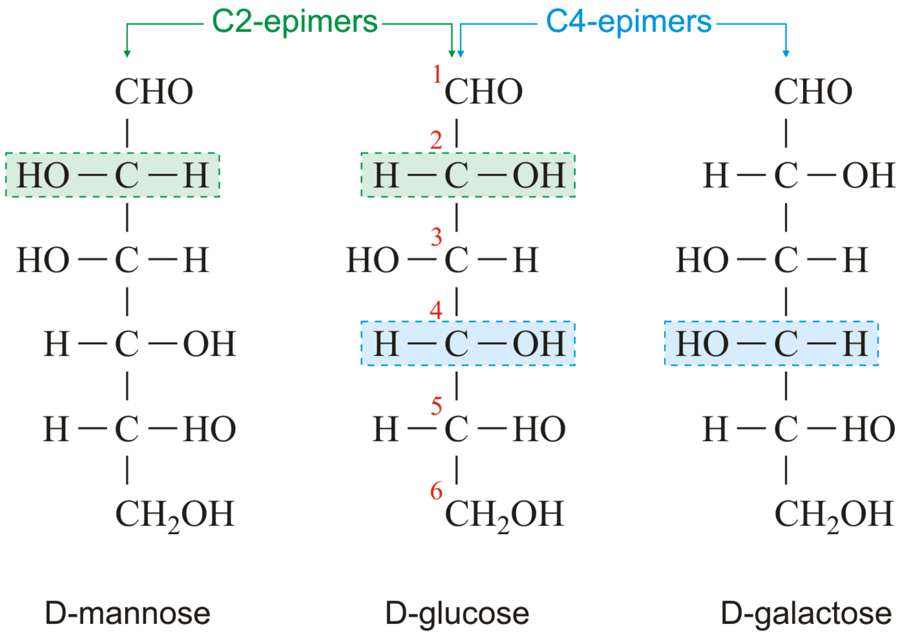
- Dissolving monosaccharides in aqueous base causes them to undergo a series of enolizations and keto–enol tautomerizations that lead to isomerizations.
- For example, if a solution of D-glucose containing calcium hydroxide is allowed to stand for several days, a number of products can be isolated, including D-fructose and D-mannose.
- Converting a monosaccharide to a methyl glycoside protects the anomeric carbon by preventing isomerization and preserving stereochemistry under basic conditions.



6.1.2 Monosaccharides

6.1.2.4 Isomerization

- **Epimers:** A special name is given to diastereomers that differ in configuration at only one stereogenic center as shown.



D-glucose and D-mannose are epimers (at C-2).

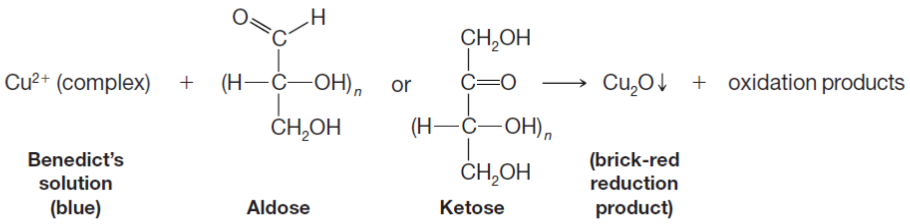
D-glucose and D-galactose are epimers (at C-4).

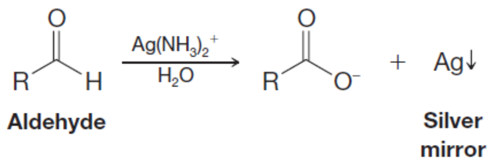
6.1.2 Monosaccharides

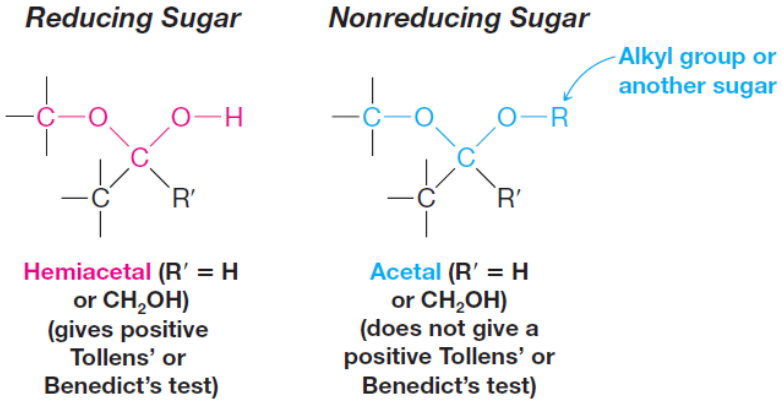
6.1.2.5 Oxidation Reactions of Monosaccharides

a. Benedict's or Tollens' Reagents: Reducing Sugars

- **Benedict's reagent:** an alkaline solution containing a cupric citrate complex ion.
- **Tollens' solution:** $[\overset{+}{Ag}(NH_3)_2\bar{O}H]$
    - They both oxidize and give positive tests with aldoses and ketoses even those that exist primarily as cyclic hemiacetals.
    - In aqueous solution, the hemiacetal form of sugars is in equilibrium with small amounts of open-chain aldehyde or α-hydroxy ketone forms, which undergo oxidation until the reactants are consumed.
    - Carbohydrates that contain only acetal groups do not give positive tests with Benedict's or Tollens' solutions, and they are called *nonreducing sugars*.
    - Acetals do not exist in equilibrium with aldehydes or α-hydroxy ketones in the basic aqueous media of the test reagents.







6.1.2 Monosaccharides

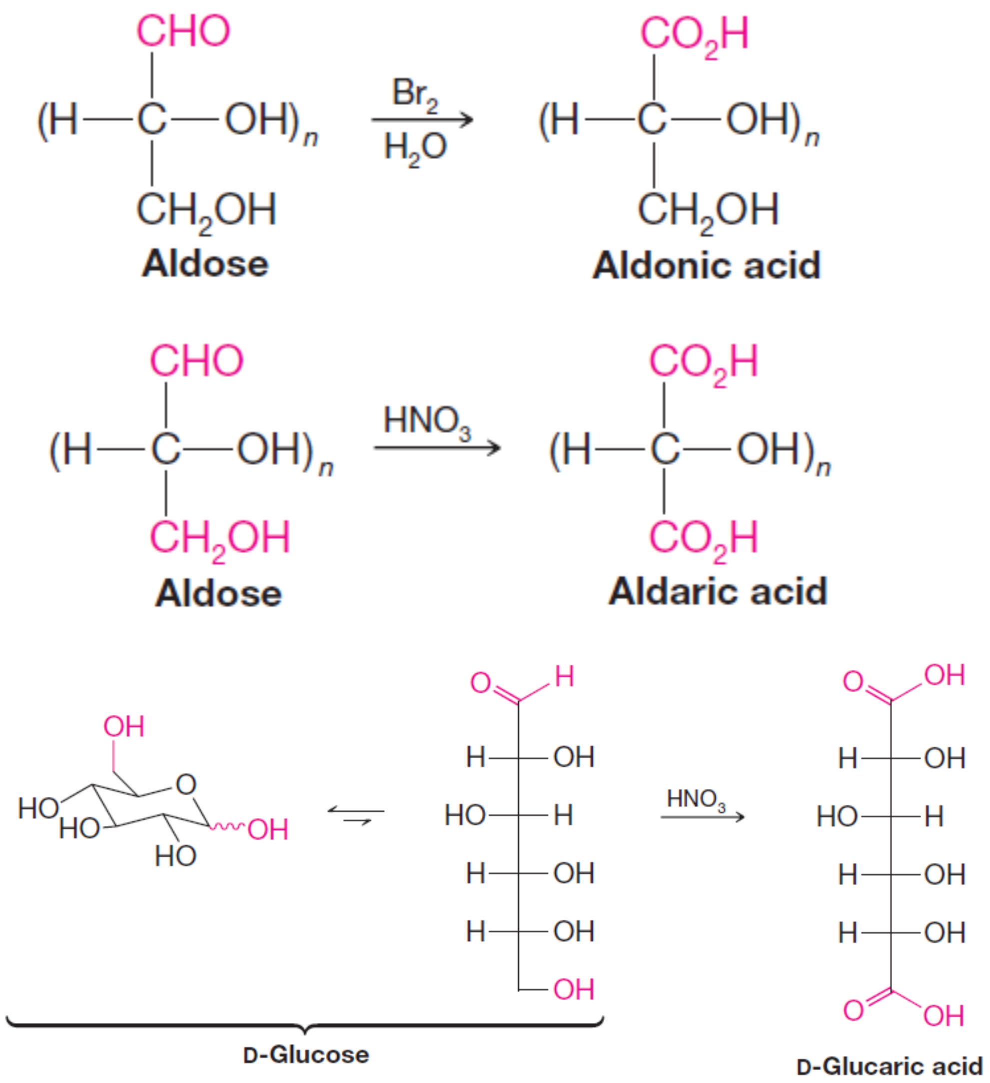
6.1.2.5 Oxidation Reactions of Monosaccharides

b. Bromine Water: The Synthesis of Aldonic Acids

- Bromine water is a general reagent that selectively oxidizes the $-CHO$ group to a $-CO_2H$ group, thus converting an aldose to an **aldonic acid**.

c. Nitric Acid Oxidation: Aldaric Acids

- Dilute nitric acid is a stronger oxidizing agent than bromine water , it oxidizes both the $-CHO$ group and the terminal $-CH_2OH$ group of an aldose to $-CO_2H$ groups, forming dicarboxylic acids are known as **aldaric acids**.
- Bromine water is a general reagent that selectively oxidizes the $-CHO$ group to a $-CO_2H$ group, thus converting an aldose to an **aldonic acid**.

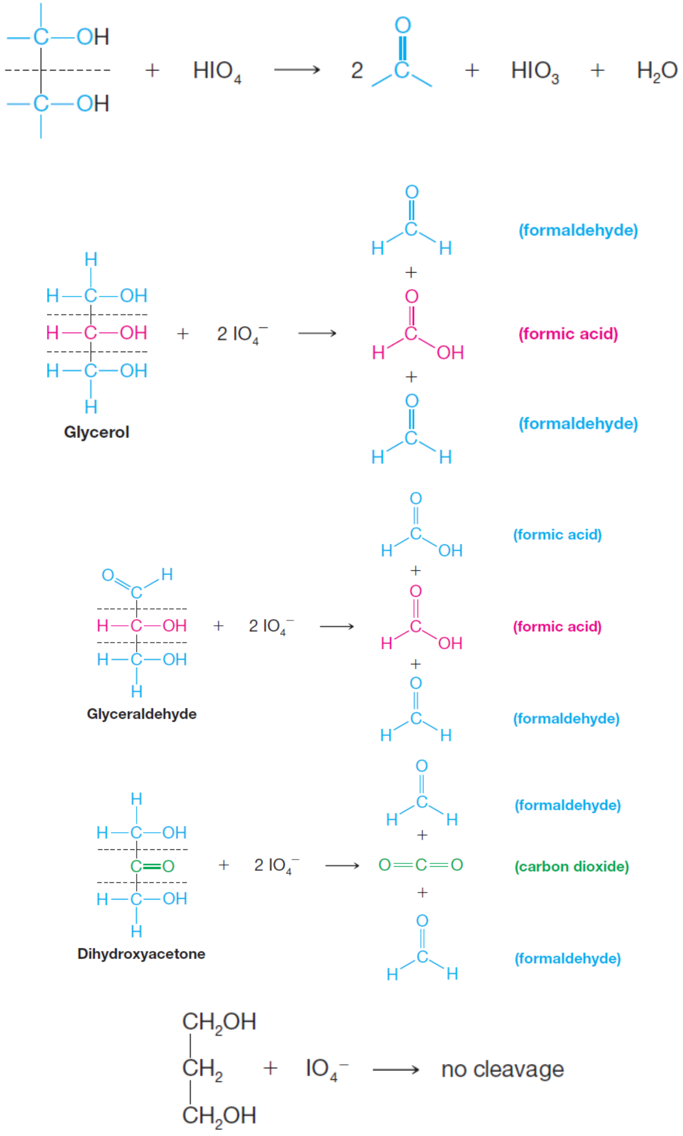


6.1.2 Monosaccharides

6.1.2.5 Oxidation Reactions of Monosaccharides

d. Periodate Oxidations: Oxidative Cleavage of Polyhydroxy Compounds

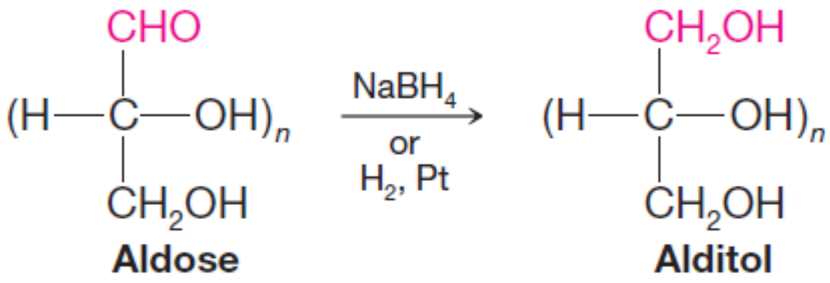
- Compounds with adjacent hydroxyl groups (vicinal diols) undergo oxidative cleavage by aqueous periodic acid (HIO_4), breaking the C–C bond to yield carbonyl compounds (aldehydes, ketones, or carboxylic acids).
- In chains with three or more contiguous –CHOH groups, the internal ones produce formic acid.
- Cleavage also occurs when an –OH group is adjacent to the carbonyl of an aldehyde or ketone (but not acids or esters).

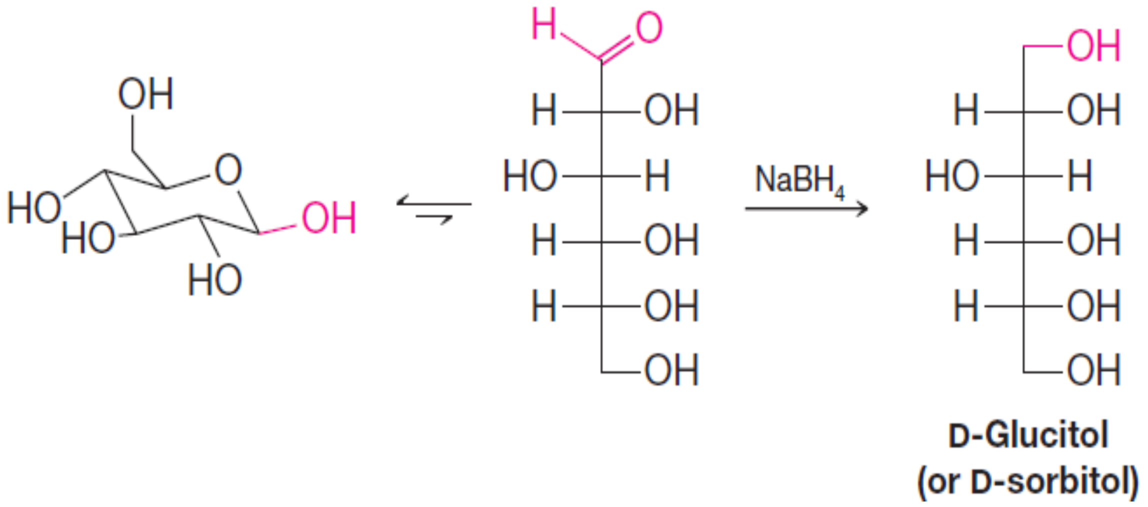


6.1.2 Monosaccharides

6.1.2.6 Reduction of Monosaccharides: Alditols

- Aldoses (and ketoses) can be reduced with sodium borohydride to give compounds called **alditols**.
- Reduction of D-glucose, for example, yields D-glucitol





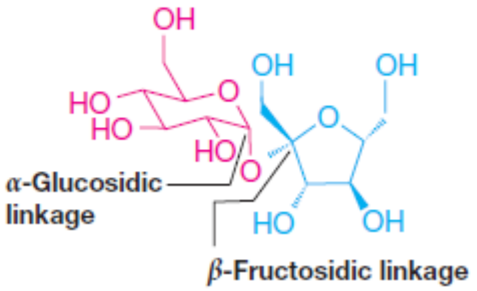
6.1.3 Disaccharides

6.1.3.1 Sucrose

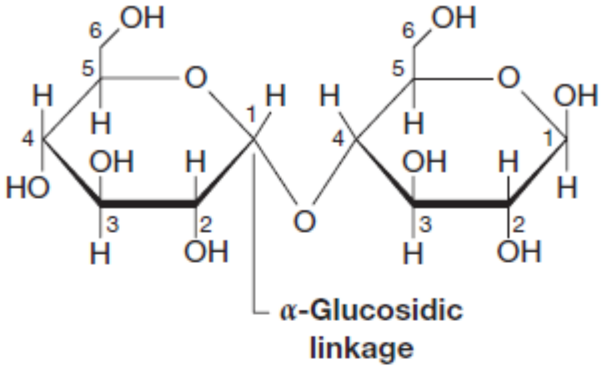
- Sucrose is the most widely occurring disaccharide, obtained commercially from sugarcane or sugar beets.
- It is a nonreducing sugar and has the molecular formula $C_{12}H_{22}O_{11}$.
- Acid-catalyzed hydrolysis of 1 mole of sucrose yields 1 mole of D-glucose and 1 mole of D-fructose.

6.1.3.2 Maltose

- Partial hydrolysis of starch gives Maltose.
- When 1 mol of maltose is subjected to acid-catalyzed hydrolysis, it yields 2 mol of D-(+)-glucose.
- Maltose is a reducing sugar; it gives positive tests with Benedict’s, and Tollens’ solutions.



(+)-Sucrose
α-D-Glucopyranosyl-(1→2)-β-D-fructofuranoside



(+)-maltose,
α-D-Glucopyranosyl-(1→4)- β- D-glucopyranose

6.1.3 Disaccharides

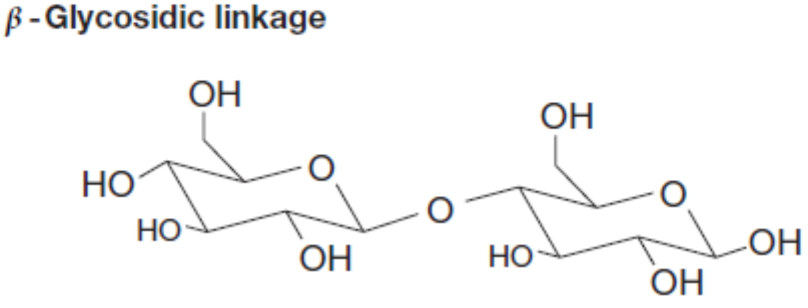
6.1.3.3 Cellobiose

- Partial hydrolysis of cellulose gives cellobiose.
- Cellobiose resembles maltose in every respect except one: the configuration of its 1→4 glycosidic linkage.
- Cellobiose is a reducing sugar that, on acid-catalyzed hydrolysis, yields two molar equivalents of D-glucose

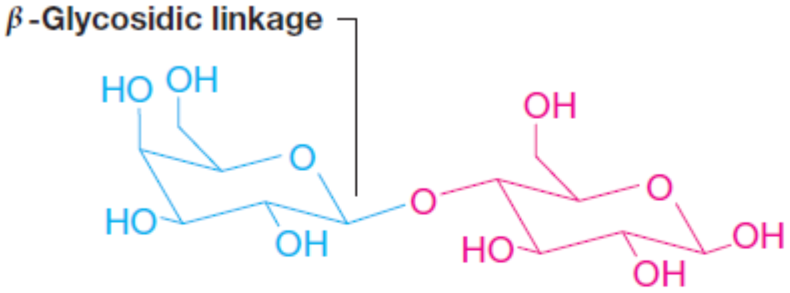
6.1.3.3 Lactose

- Partial hydrolysis of cellulose gives cellobiose.
- Lactose is a disaccharide present in the milk of humans, cows, and almost all other mammals.
- Lactose is a reducing sugar that hydrolyzes to yield d-glucose and D-galactose; the 1→4 glycosidic linkage is β.

c. Cellobiose



Cellobiose,
β-D-Glucopyranosyl-(1→4)- β-D-glucopyranose



Lactose,
β-D-Galactopyranosyl-(1→4)-β-D-glucopyranose

6.1.4 Polysaccharides

6.1.3.3 Cellobiose

- Polysaccharides are polymers consisting of monosaccharides joined together by glycosidic linkages.
- Homopolysaccharides are polymers made of a single type of monosaccharide.
- Heteropolysaccharides are polymers made of more than one type of monosaccharide.

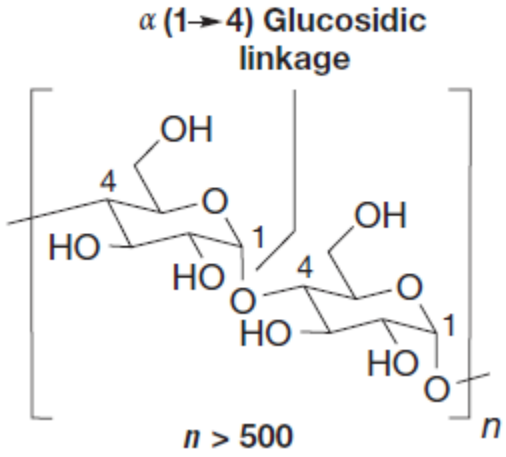
6.1.4.1 Starch

- Starch occurs as microscopic granules in roots, tubers, and seeds (e.g., corn, potatoes, wheat, rice).
- Heating starch with water swells the granules, forming a colloidal suspension.
- Starch is a nonreducing polysaccharide composed of amylose (10–20%) and amylopectin (80–90%).

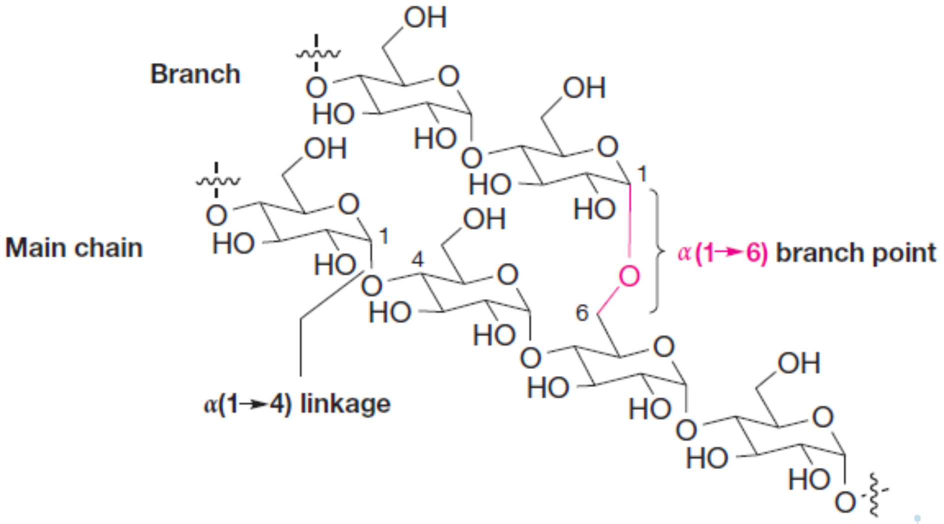
6.1.4 Polysaccharides

6.1.4.1 Starch

- **Amylose** typically consists of more than 1000 D-glucopyranoside units connected in α linkages between C1 of one unit and C4 of the next.
- **Amylopectin** has a structure similar to amylose, with $\alpha(1\rightarrow4)$ glycosidic links in the linear chains, but its structure is branched by $\alpha(1\rightarrow6)$ glycosidic bonds.
- The branching occurs between C6 of one glucose unit and C1 of another.
- Branch points appear approximately every 20–25 glucose units.
- Amylopectin has a molecular weight of 1–6 million, consisting of hundreds of interconnected chains, each containing about 20–25 glucose units.



Partial structure of Amylose



Partial structure of Amylopectin

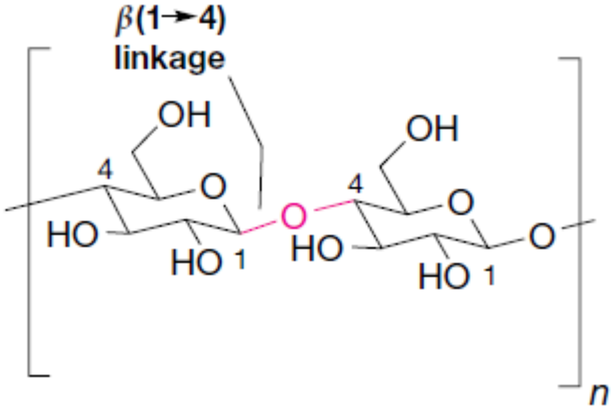
6.1.4 Polysaccharides

6.1.4.2 Glycogen

- Glycogen has a structure very much like that of amylopectin; however, in glycogen the chains are much more highly branched.
- Glycogen has a very high molecular weight, and samples isolated under conditions minimizing hydrolysis show molecular weights as high as 100 million.

6.1.4.2 Cellulose

- Cellulose consists of D-glucopyranose units linked by β(1→4) glycosidic bonds forming very long, unbranched chains.
- The β-linkages make cellulose chains linear, unlike the helical α(1→4) chains in starch and glycogen.
- Hydroxyl (–OH) groups on adjacent chains form extensive hydrogen bonds, producing a rigid, insoluble, fibrous structure.
- The precise stereochemistry of D-glucose allows the formation of strong linear fibers ideal for plant cell walls.

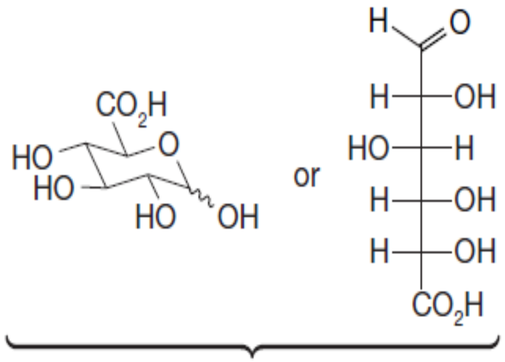


Partial structure of Cellulose

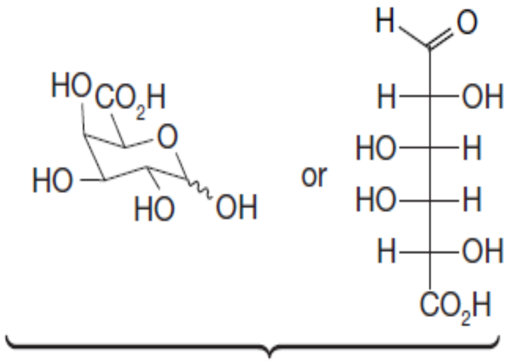
6.1.5 Importance of Carbohydrate

6.1.5.1 Biological Important Carbohydrate

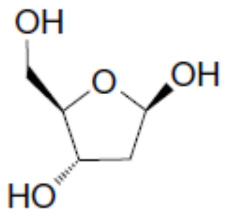
- Monosaccharide derivatives in which the –CH_2OH group at C6 is oxidized to a carboxyl group are called uronic acids.
- Their names are derived from the parent monosaccharide; for example, oxidation of C6 in glucose forms glucuronic acid, while oxidation of C6 in galactose yields galacturonic acid.
- Monosaccharides in which an **–OH** group is replaced by **–H** are called **deoxy sugars**; the most important is deoxyribose (found in DNA), while L-rhamnose and L-fucose are common in polysaccharides.



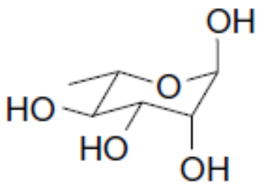
D-Glucuronic acid



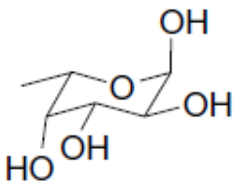
D-Galacturonic acid



β-2-Deoxy-D-ribose



α-L-Rhamnose
(6-deoxy-L-mannose)

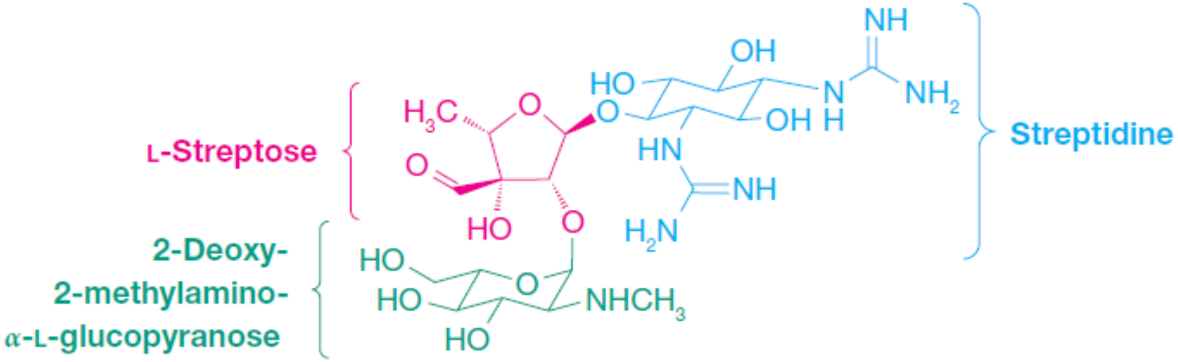


α-L-Fucose
(6-deoxy-L-galactose)

6.1.5 Importance of Carbohydrate

6.1.5.2 Carbohydrate Antibiotics

- One of the important discoveries in carbohydrate chemistry was the isolation (in 1944) of the carbohydrate antibiotic streptomycin.
- Streptomycin disrupts bacterial protein synthesis.
- Composed of three subunits:
  - An amino sugar based on L-glucose.
  - A branched-chain monosaccharide (streptose).
  - An amino cyclitol (streptidine), a cyclohexane derivative.



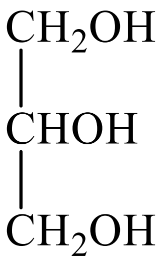
Lipids

6.2.1 Lipids

- **Lipids** are compounds of biological origin that dissolve in nonpolar solvents, such as chloroform and diethyl ether.
- The name lipid comes from the Greek word *lipos*, for fat.

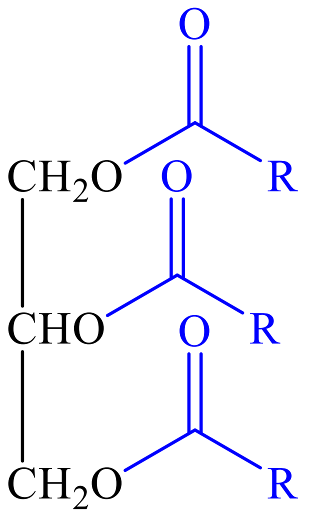
6.2.1.1 Fatty Acids and Triacylglycerols

- **Triacylglycerols** (**Triglycerides**) are the oils of plants and the fats of animal origin.
- **Triacylglycerols** are carboxylic acids of biological origin that are found as esters of glycerol.
- The groups R, R′, and R″ are usually long chain alkyl groups.
- R, R′, and R″ may also contain one or more carbon–carbon double bonds.
- R, R′, and R″ may all be different.
- **Triacylglycerols** include such common substances as peanut oil, soybean oil, corn oil, sunflower oil, butter, lard, and tallow.



$$\begin{array}{l} CH_2OH \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

Glycerol

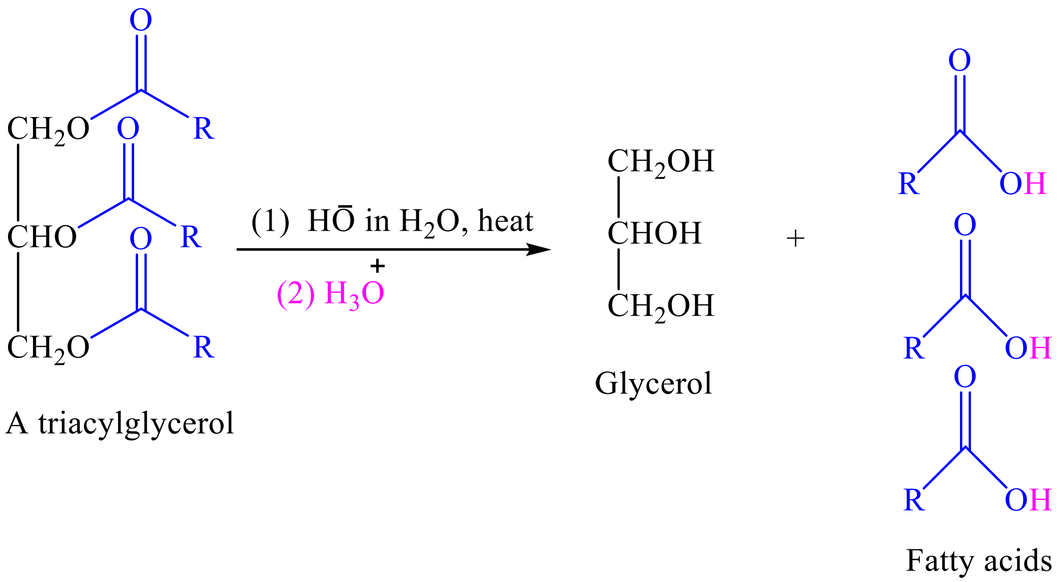


A triacylglycerol

6.2.1 Lipids

6.2.1.1 Fatty Acids and Triacylglycerols

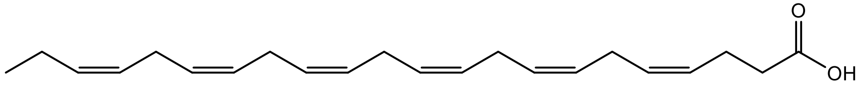
- **Triacylglycerols** are known as oils and fats.
- **Fats** are solid at room temperature due to saturated fatty acids.
- **Oils** are liquid because of higher unsaturation.
- Triacylglycerols can be
  - **Simple triacylglycerols**; all three acyl groups are the same.
  - **Mixed triacylglycerols**; the acyl groups are different.
- Hydrolysis of fats or oils yields fatty acids that are typically unbranched and contain an even number of carbon atoms.



6.2.1 Lipids

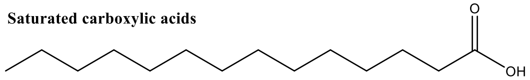
6.2.1.1 Fatty Acids and Triacylglycerols

- **Unsaturated fatty acids,**
  - Unsaturated fatty acids contain cis double bonds
  - Many of them have two or three double bonds
  - Such fats are called polyunsaturated fats or oils
  - The first double bond usually occurs between C9 and C10
- **Saturated fatty acids** have relatively high melting points.
- **Omega-3 fatty acids** are those where the third to last carbon in the chain is part of a carbon–carbon double bond.

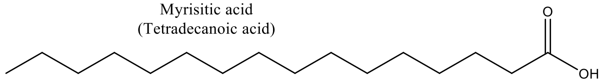


DHA, an omega-3-fatty acid
[(4Z,7Z,10Z,13Z,16Z,19Z)-4,7,10,13,16,19-docosahexaenoic acid]

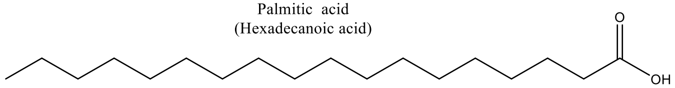
Saturated carboxylic acids



Myrisitic acid
(Tetradecanoic acid)

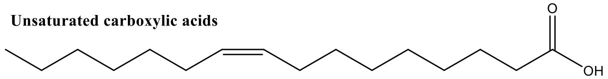


Palmitic acid
(Hexadecanoic acid)

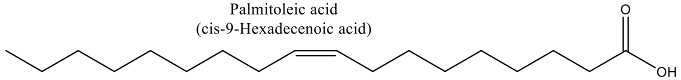


Stearic acid
(Octadecanoic acid)

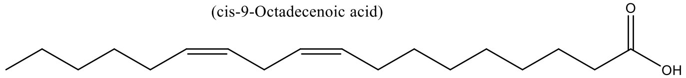
Unsaturated carboxylic acids



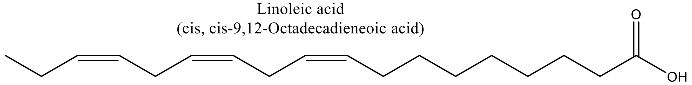
Palmitoleic acid
(cis-9-Hexadecenoic acid)



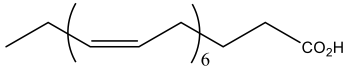
Oleic acid
(cis-9-Octadecenoic acid)



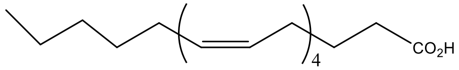
Linoleic acid
(cis, cis-9,12-Octadecadieneoic acid)



Linolenic acid
(cis, cis, cis-9,12,15-Octadecatrieneoic acid)



DHA, an omega-3-fatty acid
[(4Z,7Z,10Z,13Z,16Z,19Z)-4,7,10,13,16,19-docosahexaenoic acid]

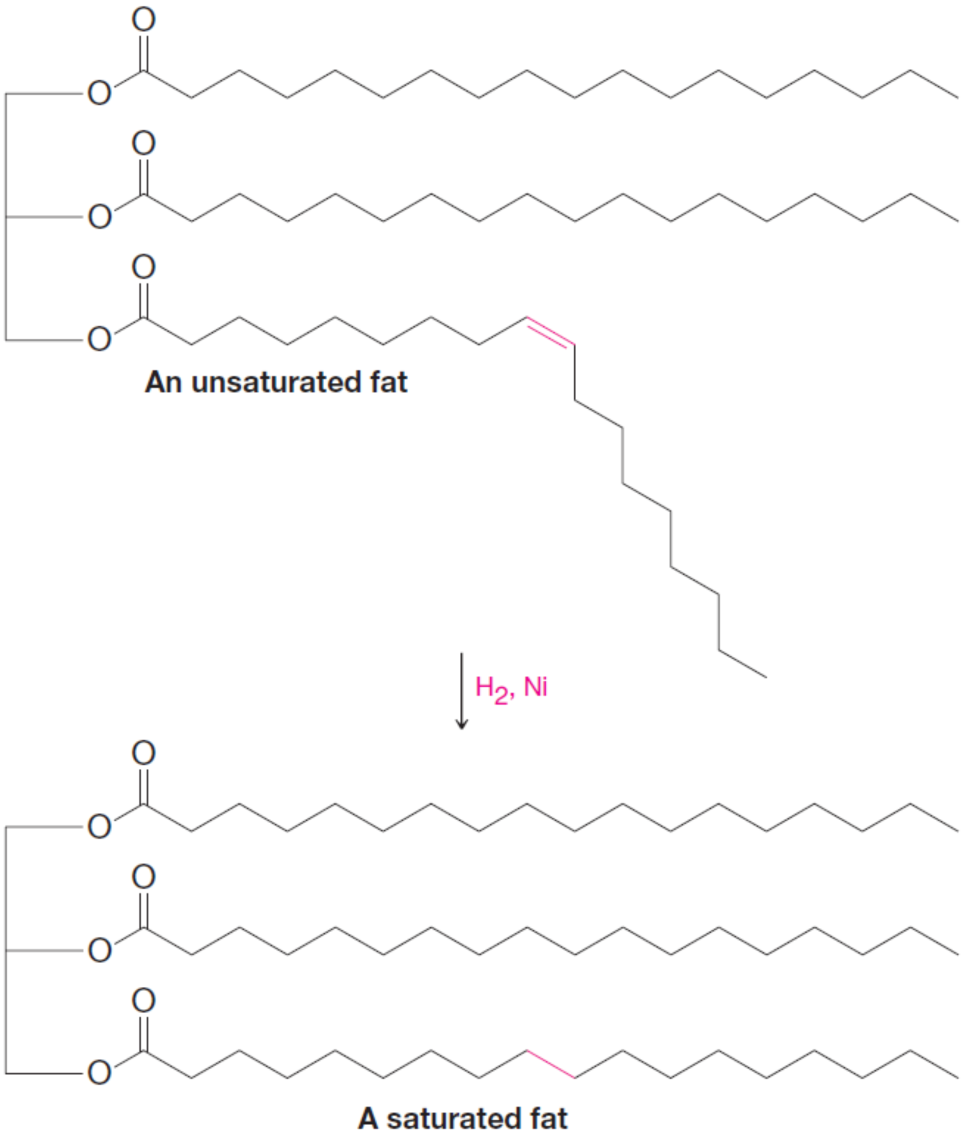


Arachidonic acid, an omega-6fatty acid
[5Z, 8Z,11Z,14Z)-5,8,11,14-eicosatetraenoic acid]

6.2.1 Lipids

6.2.1.2 Hydrogenation of Triacylglycerols

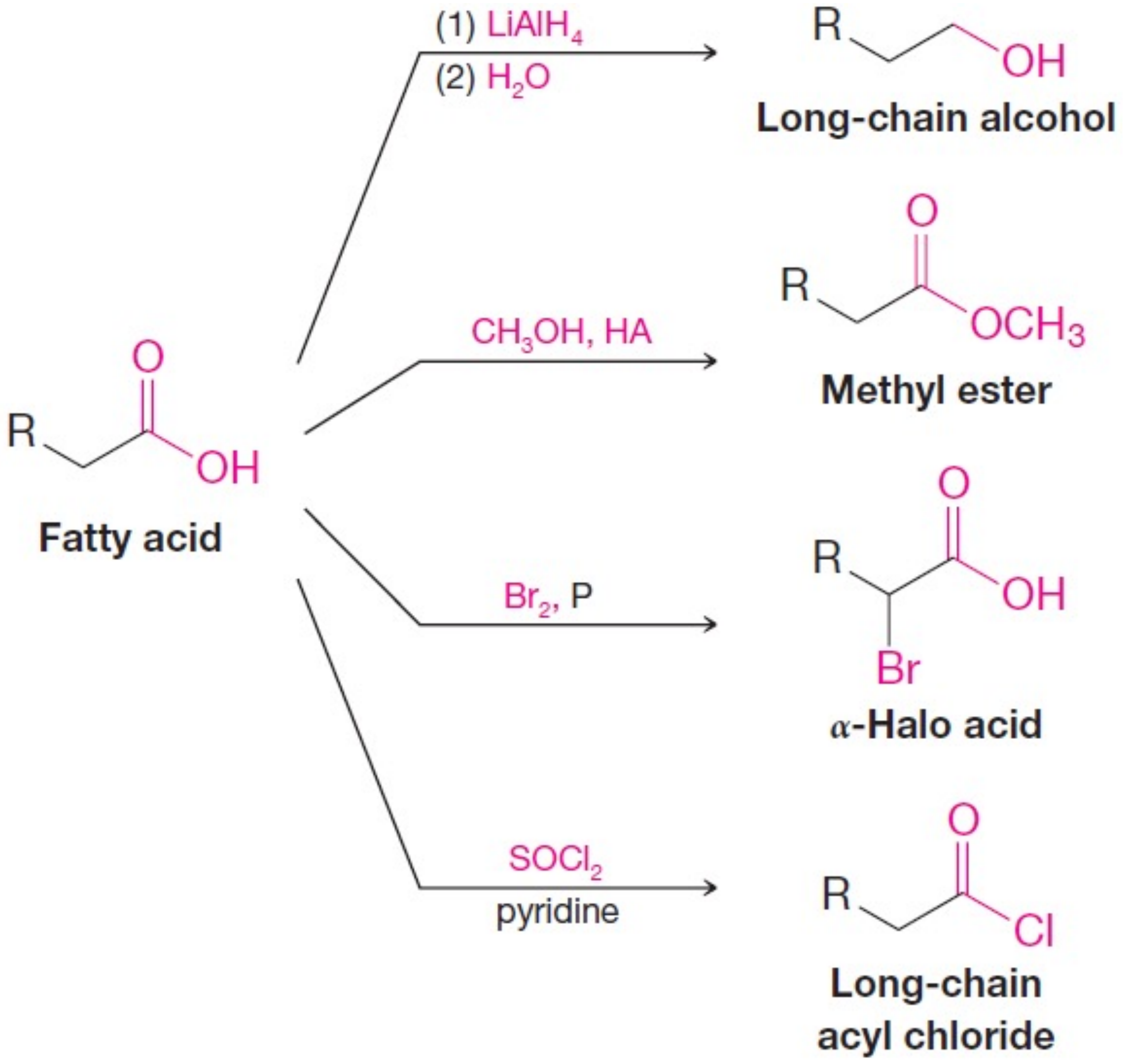
- **Fats:** Triacylglycerols high in saturated fatty acids → solid at room temperature, high melting points.
- **Oils:** Triacylglycerols high in unsaturated/polyunsaturated fatty acids → liquid at room temperature, lower melting points.
- Catalytic hydrogenation converts unsaturated triacylglycerols into more saturated forms.
- Commercial solid fats are produced by partial hydrogenation of vegetable oils, while complete hydrogenation is avoided because fully saturated triacylglycerols are hard and brittle.



6.2.1 Lipids

6.2.1.3 Reactions of the Carboxyl Group of Fatty Acids

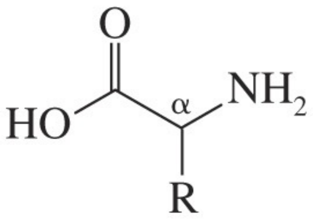
- Fatty acids undergo reactions typical of carboxylic acids.
  - They react with $LiAlH_4$ to form alcohols,
  - They react with alcohols and mineral acid to form esters,
  - They react with bromine and phosphorus to form α-halo acids,
  - They react with thionyl chloride to form acyl chlorides



Amino Acids and Proteins

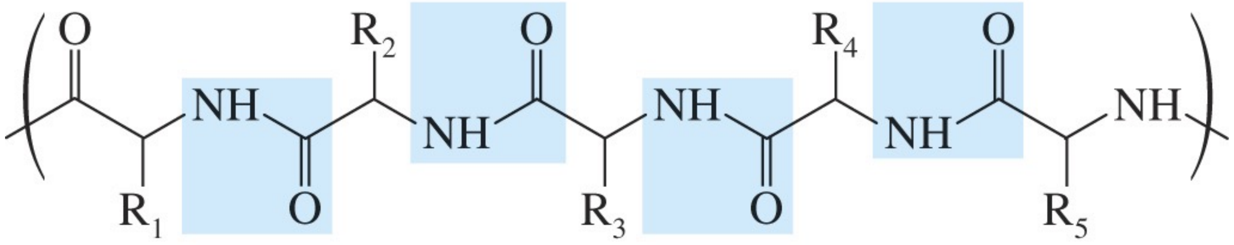
6.3.1 Introduction

- **Proteins** have many diverse functions; they are major components of the following biomolecules.
- **Enzymes and hormones** which catalyze and regulate biological reactions.
- **Muscles and tendons** which provide the body with means for movement.
- **Hemoglobin** which carries oxygen to all parts of the body.
- Antibodies they are integral parts of the immune system.
- **The 20 α-amino acids** are the essential building blocks of proteins and peptides, forming polyamides through peptide bonds.



An α-amino acid

R is a side chain at the α carbon that determines the identity of the amino acid (Table 24.1).



A portion of a protein molecule

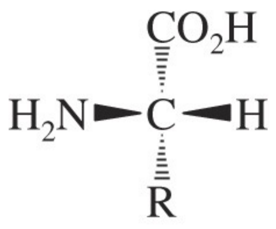
Amide (peptide) linkages are shaded.

R_1–R_5 may be any of the possible side chains.

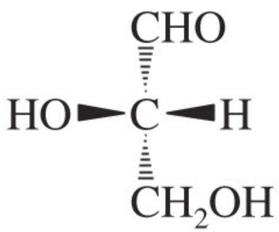
6.3.2 Amino acids

6.3.2.1 Structure and Names

- **All natural occurring amino acids** are chiral (except glycine which is achiral) and have the L-configuration (as related to glyceraldehyde) at the α-carbon.
- **Nomenclature and Classification of amino acids:**
- The 20 amino acids are known by their common names (IUPAC nomenclature could be applied) and abbreviated by **three letter** based on its common name, which is used when writing the formulas of peptides, and a **one-letter** abbreviation used to describe the amino acid sequence in a protein (see the following Table). **For example; Glycine = Gly or (G); Alanine = Ala or (A); Valine = Val or (V), etc..**



An L-α-amino acid [usually an (*S*)-α-amino acid]



L-Glyceraldehyde [(*S*)-glyceraldehyde]

Fischer projections for an L-α-amino acid and L-glyceraldehyde

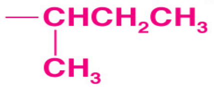
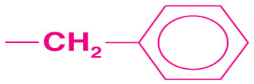
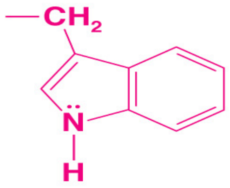
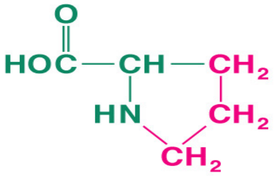
6.3.2 Amino acids

6.3.2.2 Classification of amino acids

- Amino acids are classified according to the structure of R group:
  1. The acidity of R group: neutral or acidic or basic as shown in the Table.
     - Neutral R Group: Are those which containing one amino group and one carboxylic group. The majority of amino acids contain neutral R group.
     - Acidic R group: Are those amino acids that contain more than one carboxylic group such as
  2. Polarity of the R group: Amino acid could be classified into polar (hydrophilic) or non-polar (hydrophobic).
- Essential and non-essential amino acid:
  1. Essential amino acid are those amino acids which cannot be synthesized by adult humans and therefore must be included in the diet in the form of proteins. There are eight amino acids Valine. Leucine, Isoleucine, Threonine, Methionine, Phenylalanine, Tryptophan, and Lysine.
  2. Non-essential amino acids: These can be synthesized in the body from other foods and include Glycine, Alanine, Serine, Cysteine, Proline, Tyrosine, Aspartic acid, Glutamic acid, Asparagine, Glutamine, Arginine, and Histidine.

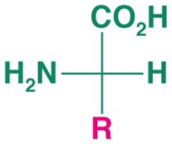
6.3.2 Amino acids

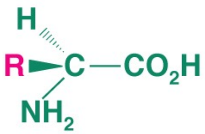
6.3.2.2 Classification of amino acids

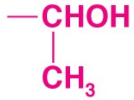
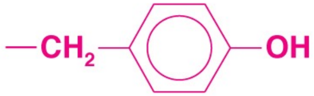
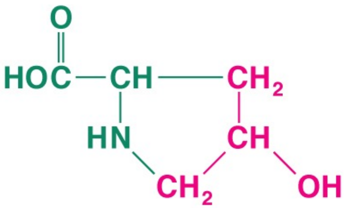
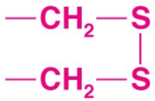
Structure of R	Name	Abbreviations[a]	pK_{a_1} α-CO_2H	pK_{a_2} α-NH_3^+	pK_{a_3} R group	pI
Neutral Amino Acids						
$-H$	Glycine	G or Gly	2.3	9.6		6.0
$-CH_3$	Alanine	A or Ala	2.3	9.7		6.0
$-CH(CH_3)_2$	Valine[b]	V or Val	2.3	9.6		6.0
$-CH_2CH(CH_3)_2$	Leucine[b]	L or Leu	2.4	9.6		6.0
 $-\underset{\mid \atop CH_3}{CH}CH_2CH_3$	Isoleucine[b]	I or Ile	2.4	9.7		6.1
	Phenylalanine[b]	F or Phe	1.8	9.1		5.5
$-CH_2CONH_2$	Asparagine	N or Asn	2.0	8.8		5.4
$-CH_2CH_2CONH_2$	Glutamine	Q or Gln	2.2	9.1		5.7
	Tryptophan[b]	W or Trp	2.4	9.4		5.9
 (complete structure)	Proline	P or Pro	2.0	10.6		6.3

6.3.2 Amino acids

6.3.2.2 Classification of amino acids

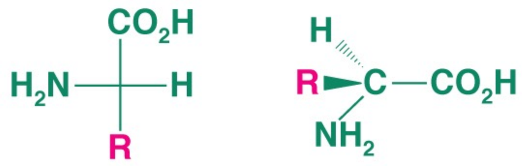


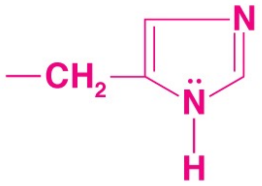


Structure of **R**	Name	Abbreviations^a	pK_{a_1} α-CO_2H	pK_{a_2} α-NH_3^+	pK_{a_3} **R** group	pI
Neutral Amino Acids						
$—CH_2OH$	Serine	S or Ser	2.2	9.2		5.7
 $—CH(OH)CH_3$	Threonine^b	T or Thr	2.6	10.4		6.5
	Tyrosine	Y or Tyr	2.2	9.1	10.1	5.7
 (complete structure)	Hydroxyproline	Hyp	1.9	9.7		6.3
$—CH_2SH$	Cysteine	C or Cys	1.7	10.8	8.3	5.0
 $—CH_2—S—S—CH_2—$	Cystine	Cys-Cys	1.6 2.3	7.9 9.9		5.1
$—CH_2CH_2SCH_3$	Methionine^b	M or Met	2.3	9.2		5.8

6.3.2 Amino acids

6.3.2.2 Classification of amino acids



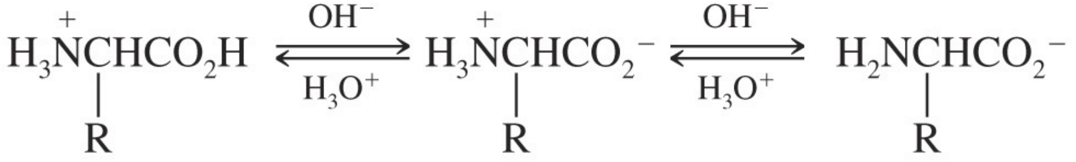
Structure of R	Name	Abbreviations[a]	pK_{a_1} α-CO_2H	pK_{a_2} α-NH_3^+	pK_{a_3} R group	pI
R Contains an Acidic (Carboxyl) Group						
$-CH_2CO_2H$	Aspartic acid	D or Asp	2.1	9.8	3.9	3.0
$-CH_2CH_2CO_2H$	Glutamic acid	E or Glu	2.2	9.7	4.3	3.2
R Contains a Basic Group						
$-CH_2CH_2CH_2CH_2NH_2$	Lysine[b]	K or Lys	2.2	9.0	10.5[c]	9.8
 $-CH_2CH_2CH_2NH-\overset{\overset{NH}{\Vert}}{C}-NH_2$	Arginine	R or Arg	2.2	9.0	12.5[c]	10.8
	Histidine	H or His	1.8	9.2	6.0[c]	7.6

[a] Single-letter abbreviations are now the most commonly used form in current biochemical literature.
[b] An essential amino acid.
[c] pK_a is of protonated amine of R group.

6.3.2 Amino acids

6.3.2.3 Amino acids as dipolar ions

- In the dry solid state amino acids exist as dipolar ions (zwittrions or inner salts).
- In aqueous solutions an equilibrium exists between the dipolar ion, the cationic and the anionic forms of the amino acid.
- The predominant form depends on the pH of the solution.
- At low pH, the amino acid exists primarily in the cationic form.
- At high pH, the amino acid exists primarily in the anionic form.
- At some intermediate pH, called the P*I* (*isoelectric point*), the concentration of the dipolar ion is at maximum and the concentrations of anionic and cationic forms are equal
- Each individual amino acid has a characteristic P*I*.



$$\underset{\displaystyle R}{\overset{+}{H_3N}CHCO_2H} \underset{H_3O^+}{\overset{OH^-}{\rightleftharpoons}} \underset{\displaystyle R}{\overset{+}{H_3N}CHCO_2^-} \underset{H_3O^+}{\overset{OH^-}{\rightleftharpoons}} \underset{\displaystyle R}{H_2NCHCO_2^-}$$

Cationic form (predominant in strongly acidic solutions, e.g., at pH 0)

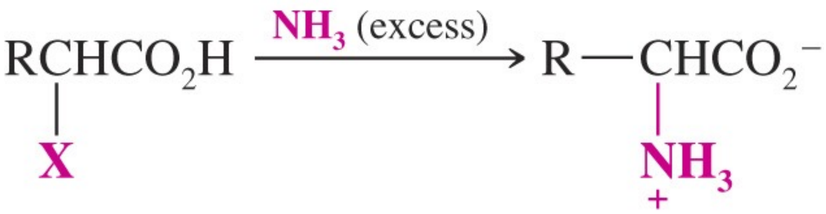
Dipolar ion

Anionic form (predominant in strongly basic solutions, e.g., at pH 14)

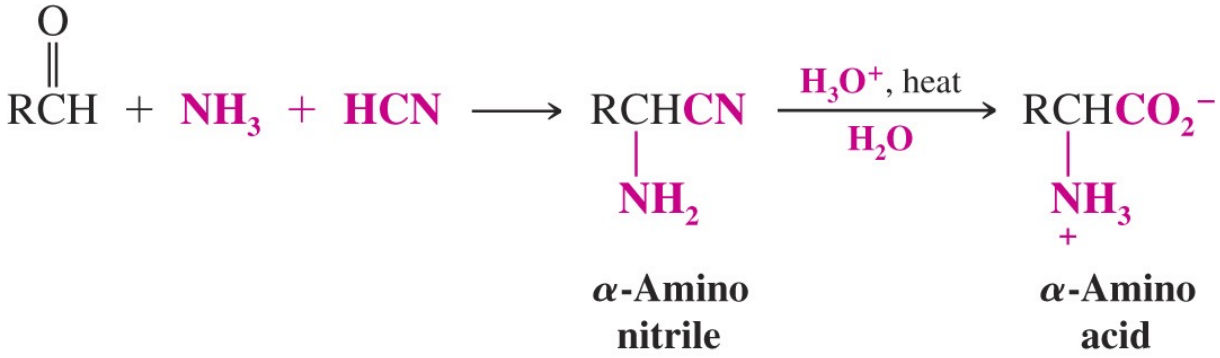
6.3.2 Amino acids

6.3.2.4 Synthesis of Amino acids

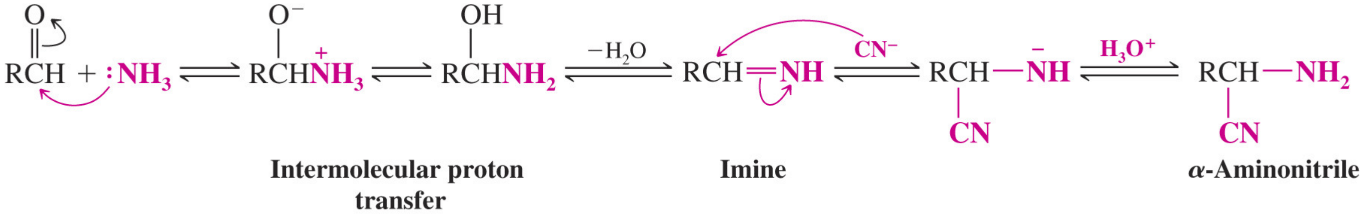
- There are many methods to prepare amino acids, two methods will be mentioned here.
  1. Direct Ammonolysis of α-halo-carboxylic acid with this method, yields tend to be poor.
  2. The Strecker Synthesis (Reductive amination) treatment of an aldehyde with ammonia and hydrogen cyanide yields an α-aminonitrile which is hydrolyzed to the α-amino acid. The mechanism proceeds via an intermediate imine



Direct ammonolysis of α-halocarboxylic acids

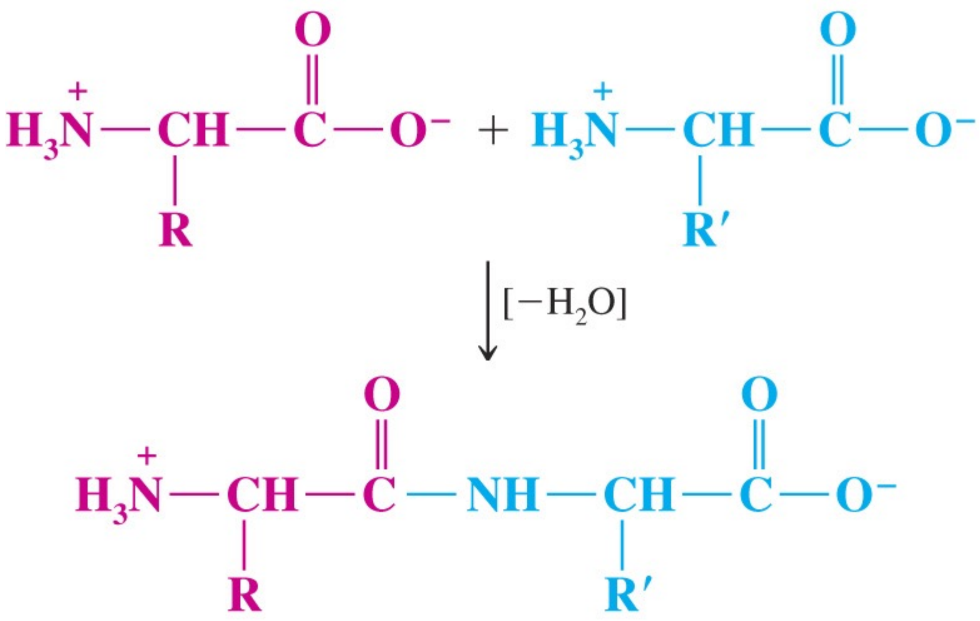


The Strecker synthesis



6.3.3 Polypeptides and Proteins

- Enzymes polymerize amino acids by forming amide linkages.
- The polymer is called a peptide and the amide linkages are called peptide bonds or peptide linkages.
- Each amino acid in the peptide is called an amino acid residue.
- Proteins can contain one or more peptide chains and others associated molecules or metal ions.

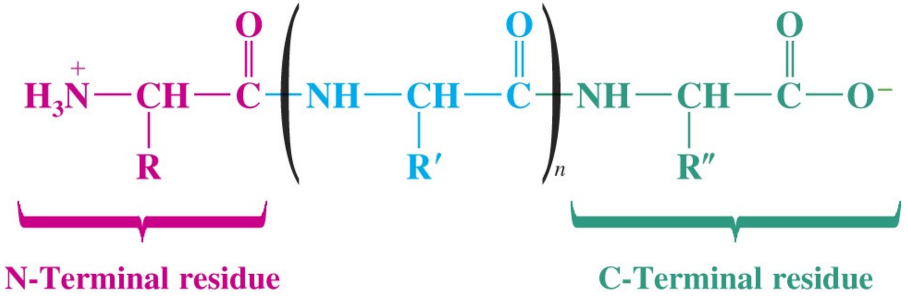


$$H_3\overset{+}{N}-\underset{\displaystyle R}{CH}-\overset{\displaystyle O}{\overset{\|}{C}}-O^- + H_3\overset{+}{N}-\underset{\displaystyle R'}{CH}-\overset{\displaystyle O}{\overset{\|}{C}}-O^- \xrightarrow{[-H_2O]} H_3\overset{+}{N}-\underset{\displaystyle R}{CH}-\overset{\displaystyle O}{\overset{\|}{C}}-NH-\underset{\displaystyle R'}{CH}-\overset{\displaystyle O}{\overset{\|}{C}}-O^-$$

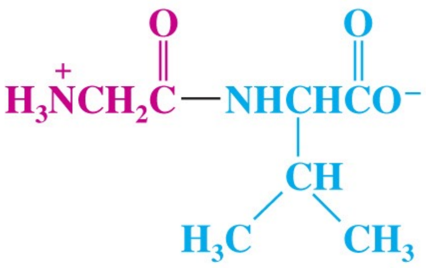
A dipeptide

6.3.3 Polypeptides and Proteins

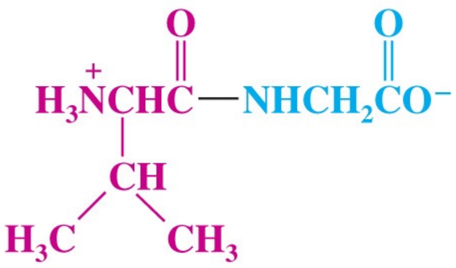
- Polypeptides are linear polymers. One end of a polypeptide chain in an amino acid residue has a free amino group -$\overset{+}{N}H_3$, the other terminate in amino acid residue with free carboxylic group –CO_2^-. These two groups are called the N-terminal and the C-terminal residues, respectively.
- Polypeptides are usually written with the N-terminal residue to the left and the C-terminal residue to the right.



$$H_3\overset{+}{N}-\underset{R}{CH}-\overset{O}{\overset{\|}{C}}\left(NH-\underset{R'}{CH}-\overset{O}{\overset{\|}{C}}\right)_n NH-\underset{R''}{CH}-\overset{O}{\overset{\|}{C}}-O^-$$



Glycylvaline
(Gly · Val)

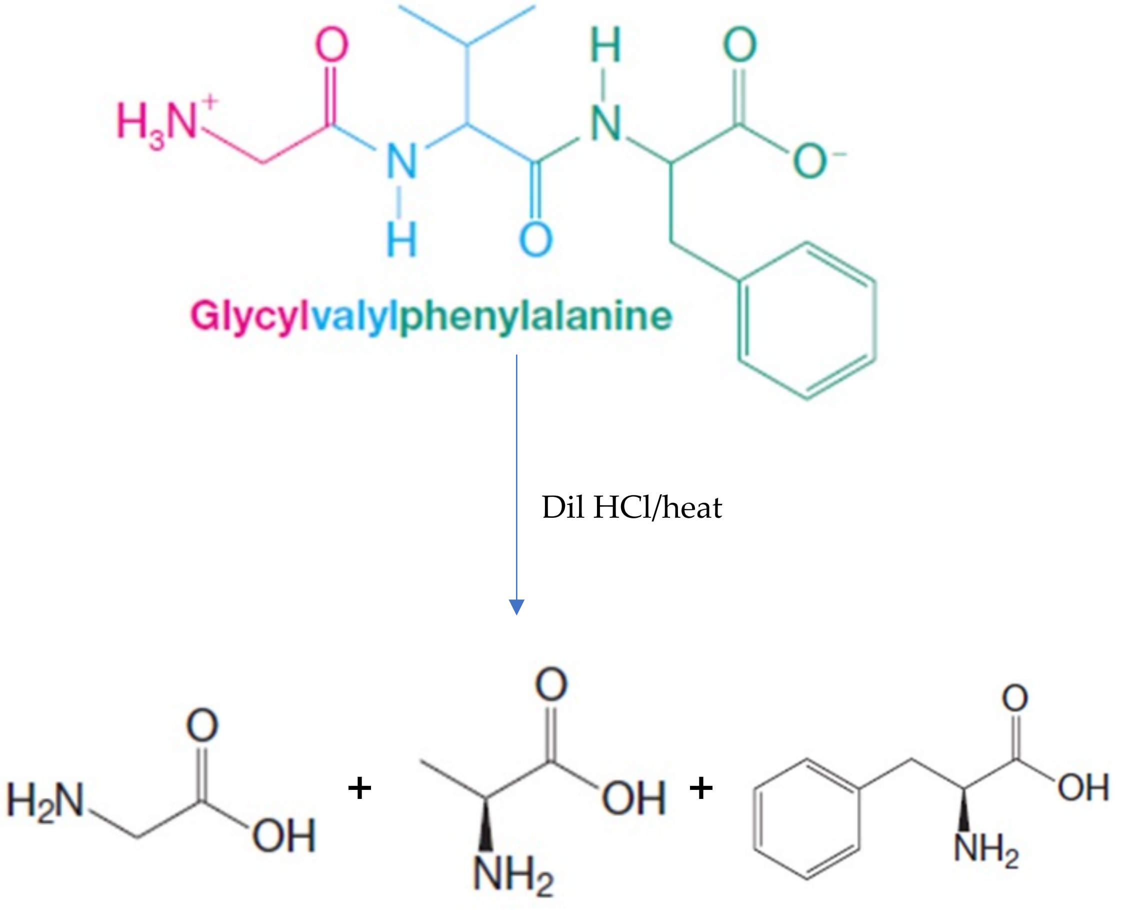


Valylglycine
(Val · Gly)

6.3.3 Polypeptides and Proteins

6.3.3.1 Hydrolysis of Polypeptides

- Hydrolysis of all amide linkages of a protein or polypeptide usually takes place by refluxing with 6M HCl for 24 hours. This hydrolysis results in the liberation of amino acids. The individual amino acids can be separated from each other using chromatographic methods.
- For example, the complete hydrolysis of the tripeptide Glycylvalylpheneylalanine (Gly-Val-Phe) will give the three amino acids glycine, valine, and phenylalanine.

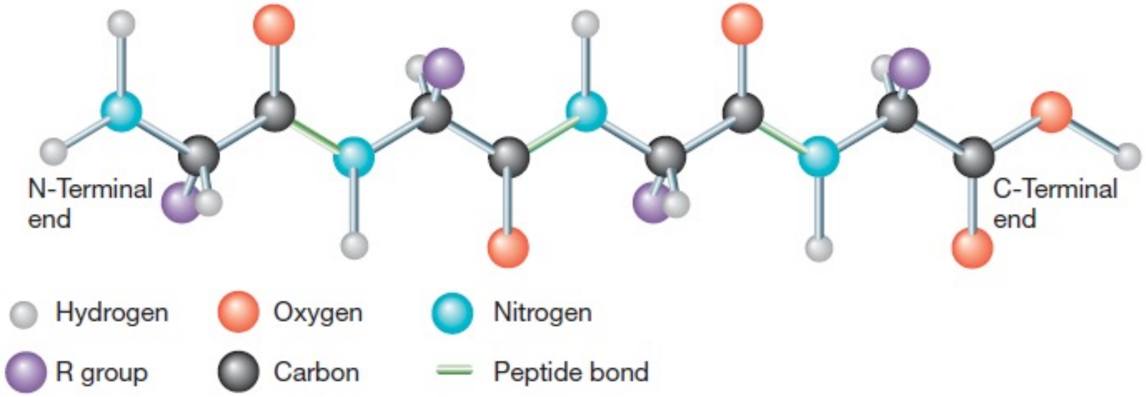


Glycine (Gly) **Valine (Val)** **Phenylalanine (Phe)**

6.3.3 Polypeptides and Proteins

6.3.3.2 Primary Structure of Polypeptides and Proteins

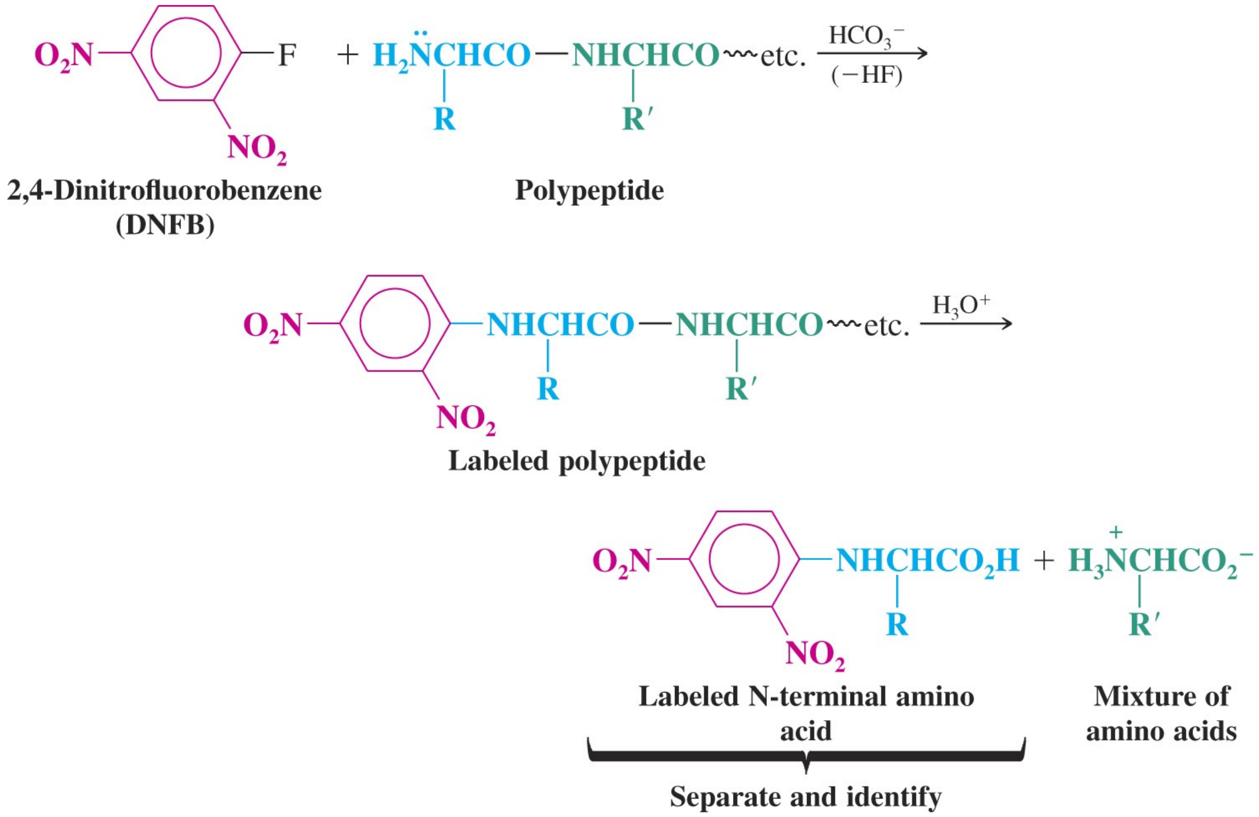
- Primary structure = specific sequence of amino acid residues in a polypeptide or protein
- For a peptide with n different amino acids → maximum possible sequences = n! (n factorial).
- Example: Tetrapeptide (4 amino acids) → up to 24 different sequences (4! = 4 × 3 × 2 × 1 = 24)
- Determining the exact sequence is very important.
- First step: Identify the N-terminal and C-terminal residues.



6.3.3 Polypeptides and Proteins

6.3.3.2 Primary Structure of Polypeptides and Proteins

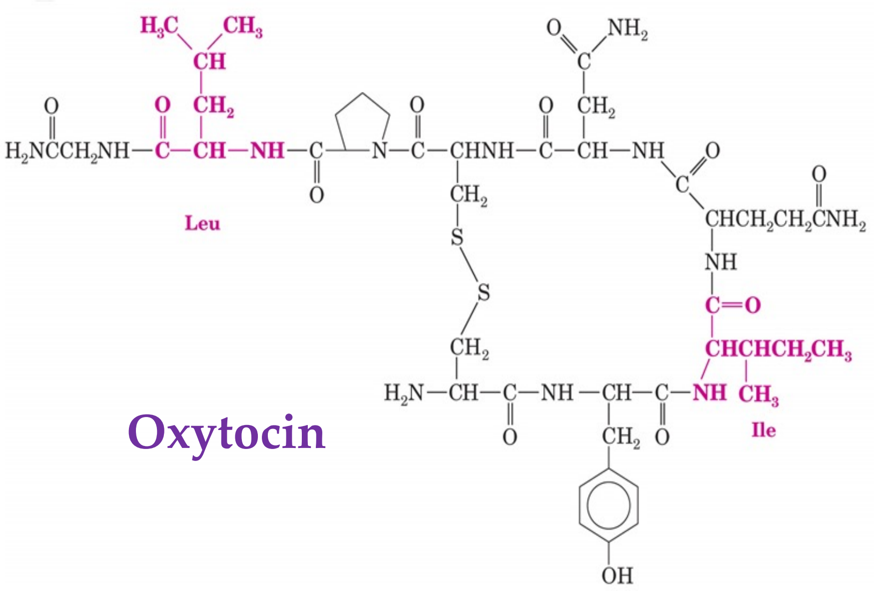
- The *N*-terminal end of the polypeptide is labeled with 2,4-dinitrofluorobenzene and the polypeptide is hydrolyzed. Then, the labeled N-terminal amino acid is separated from the mixture and identified.
- The C-terminal end Enzymes called carboxypeptidase hydrolyze C-terminal amino acids selectively. The enzyme continues to release each newly exposed C-terminal amino acid as the peptide is hydrolyzed.



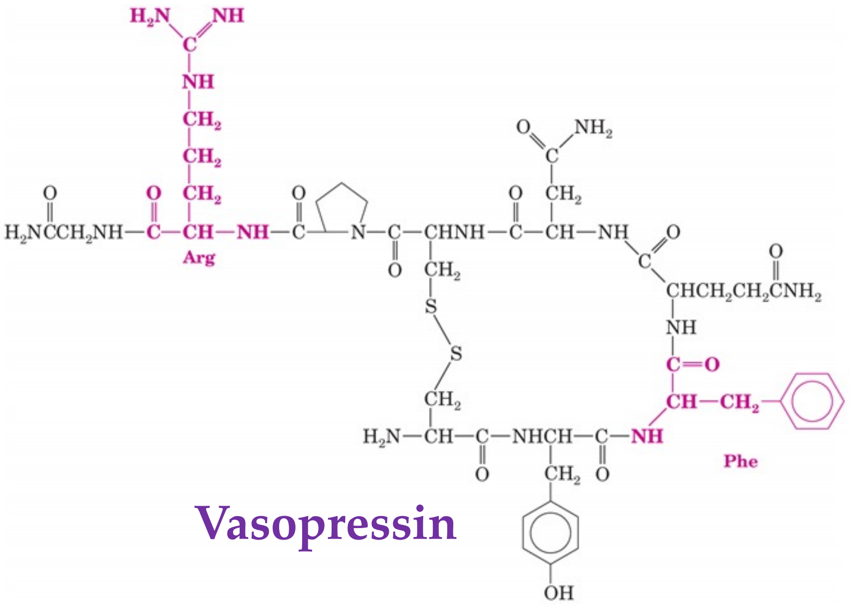
6.3.3 Polypeptides and Proteins

6.3.3.3 Examples of Polypeptide and Protein Primary Structure

- Polypeptides and proteins play crucial biological and physiological roles in living organisms.
- Key examples:
  - Oxytocin and vasopressin — two small, strikingly similar nonapeptides (9 amino acids) that differ in only two residues (highlighted in red).
    - Oxytocin stimulates uterine contractions during childbirth.
    - Vasopressin causes contraction of peripheral blood vessels, resulting in increased blood pressure.



Oxytocin



Vasopressin

6.3.3 Polypeptides and Proteins

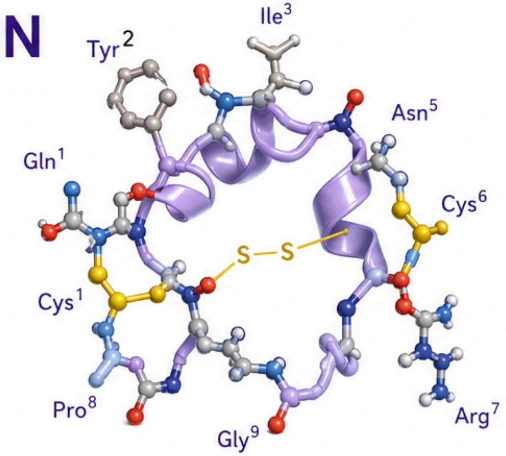
6.3.3.3 Examples of Polypeptide and Protein Primary Structure

OXYTOCIN

(Human Oxytocin)

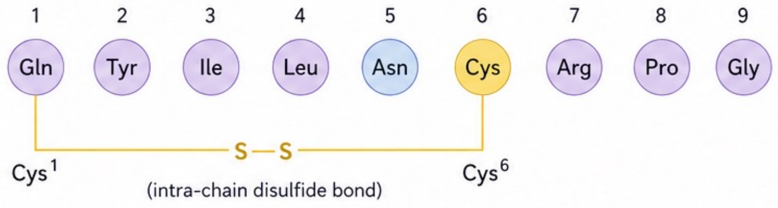
- Nonapeptide hormone
- 9 amino acids
- Cyclic structure
- Contains one disulfide bond (intra-chain)

PDB ID: 7DFG



PRIMARY STRUCTURE (Amino Acid Sequence)

Human Oxytocin (9 amino acids)



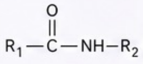
Sequence: Gln–Tyr–Ile–Leu–Asn–Cys–Arg–Pro–Gly

(One-letter code: Q–Y–I–L–N–C–R–P–G)

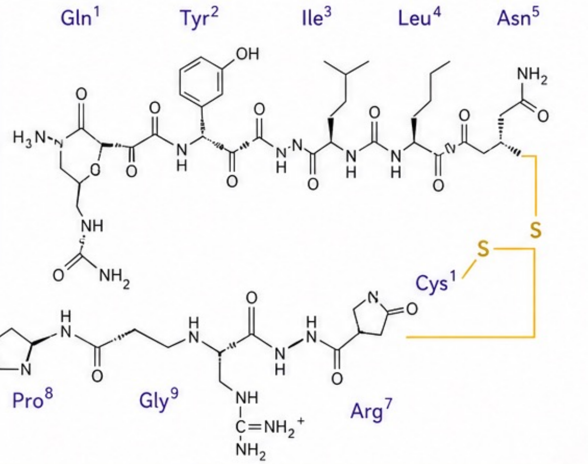
Molecular formula: $C_{43}H_{66}N_{12}O_{12}S_2$ | Molecular weight: 1007.19 g/mol

CHEMICAL STRUCTURE (Peptide Bonds)

Peptide bond
(amide bond)



The peptide bond links amino acids in a chain.

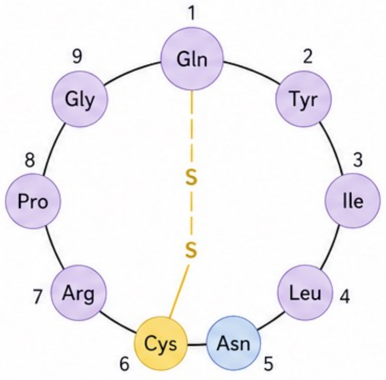


Source: UniProtKB – P01178 | PubChem CID: 439302 | DrugBank ID: DB00204

KEY FEATURES

- Oxytocin is a cyclic nonapeptide hormone.
- Contains one intra-chain disulfide bond between Cys^1 and Cys^6.
- The cyclic structure is essential for its biological activity.
- Synthesized in the hypothalamus and released by the posterior pituitary.
- Major functions:
  - Stimulates uterine contractions during labor
  - Causes milk ejection (let-down reflex) during breastfeeding
  - Involved in social bonding and behavior

DISULFIDE BOND MAP



One intra-chain disulfide bond
(Cys^1 – Cys^6)

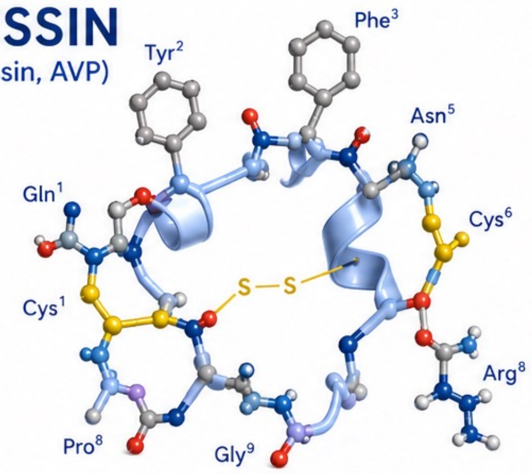
6.3.3 Polypeptides and Proteins

6.3.3.3 Examples of Polypeptide and Protein Primary Structure

VASOPRESSIN
(Arginine Vasopressin, AVP)

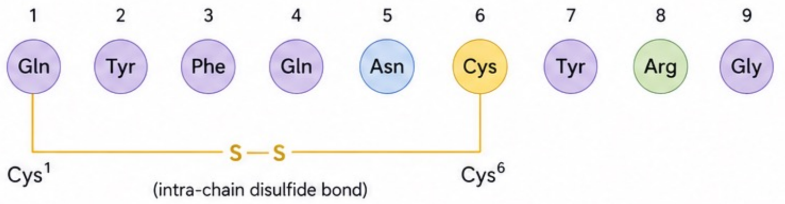
- Nonapeptide hormone
- 9 amino acids
- Cyclic structure
- Contains one disulfide bond (intra-chain)
- Synthesized in the hypothalamus and released by the posterior pituitary

PDB ID: 1AVP



PRIMARY STRUCTURE (Amino Acid Sequence)

Human Vasopressin (9 amino acids)

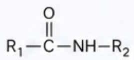


Sequence: Gln–Tyr–Phe–Gln–Asn–Cys–Tyr–Arg–Gly
(One-letter code: Q–Y–F–Q–N–C–Y–R–G)

Molecular formula: $C_{46}H_{65}N_{15}O_{12}S_2$ | Molecular weight: 1084.23 g/mol

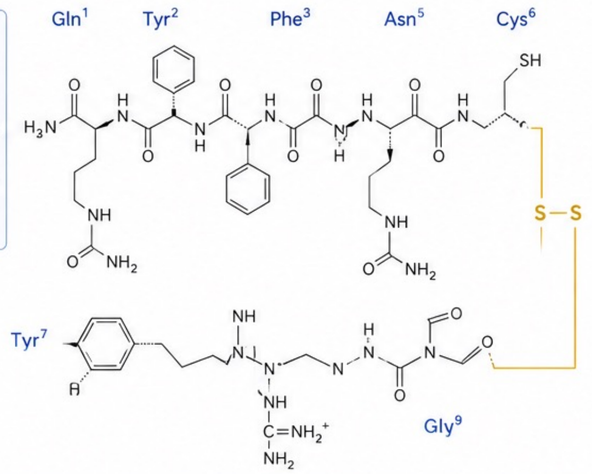
CHEMICAL STRUCTURE (Peptide Bonds)

Peptide bond
(amide bond)



$R_1-\overset{\overset{\displaystyle O}{\|}}{C}-NH-R_2$

The peptide bond links amino acids in a chain.



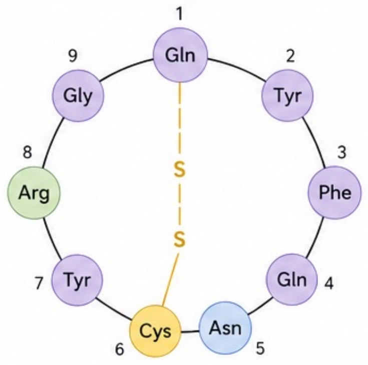
Sources: UniProt (P01185) | PubChem CID: 439302 | DrugBank ID: DB00161

KEY FEATURES

- Vasopressin is a cyclic nonapeptide hormone.
- Contains one intra-chain disulfide bond between Cys^1 and Cys^6.
- The cyclic structure is essential for its stability and biological activity.
- Synthesized in the hypothalamus (supraoptic and paraventricular nuclei) and released by the posterior pituitary.
- Major functions:
  - Antidiuretic effect: increases water reabsorption in the kidneys (V_2 receptors)
  - Vasoconstriction: increases blood pressure (V_1 receptors)
- More potent vasoconstrictor than oxytocin.

● Aromatic amino acid
● Polar uncharged amino acid
● Basic amino acid
● Cysteine (disulfide bond)
● Other amino acid

DISULFIDE BOND MAP

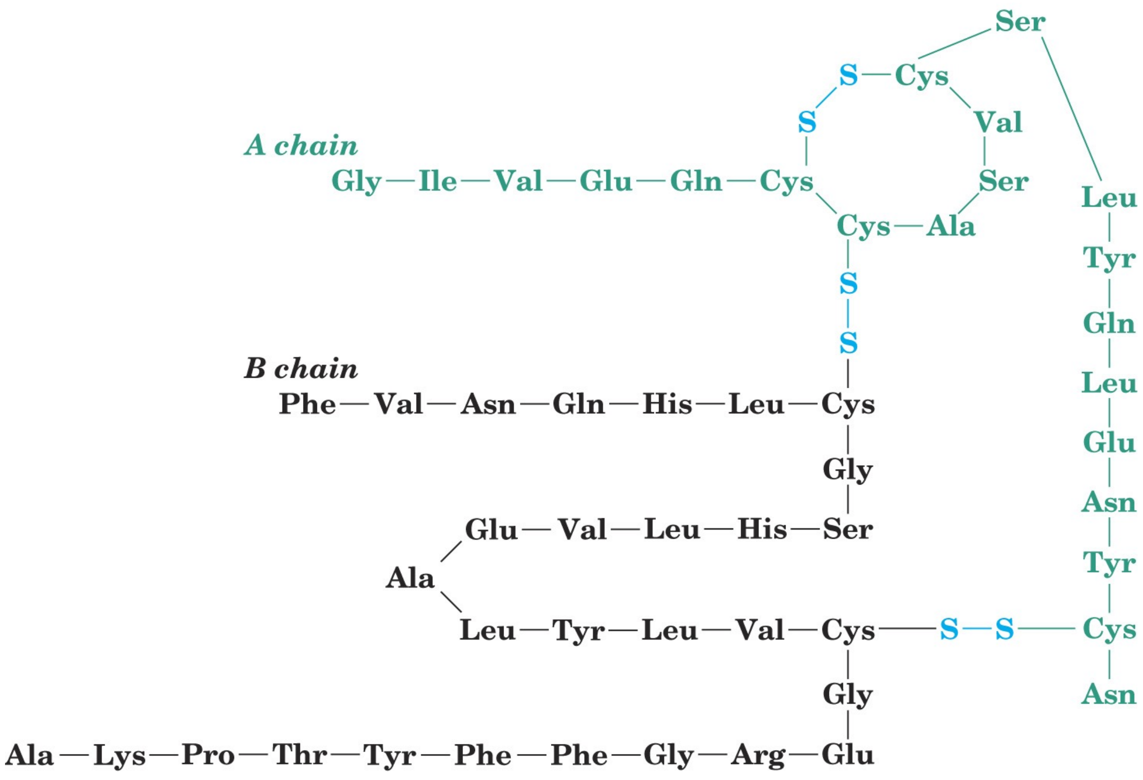


One intra-chain disulfide bond
(Cys^1 – Cys^6)

6.3.3 Polypeptides and Proteins

6.3.3.3 Examples of Polypeptide and Protein Primary Structure

- Insulin – Key Facts
- Hormone regulating glucose metabolism
- Bovine insulin: 51 amino acids
  - A-chain: 21 aa
  - B-chain: 30 aa
- Chains connected by disulfide linkages
- Insulin deficiency → major cause of diabetes mellitus
- Bovine insulin differs from human in only 3 amino acids



Insulin

6.3.3 Polypeptides and Proteins

6.3.3.3 Examples of Polypeptide and Protein Primary Structure

INSULIN

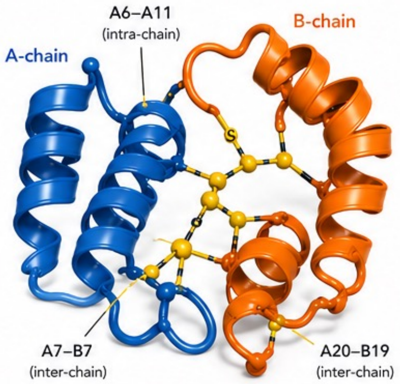
(Human Insulin)

- Peptide hormone
- 51 amino acids
- Two polypeptide chains
  - – A-chain (21 aa)
  - – B-chain (30 aa)
- Linked by three disulfide bonds
- Essential for glucose homeostasis

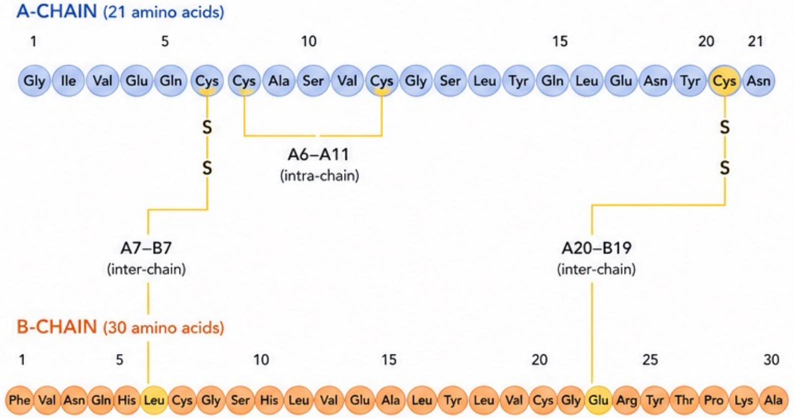
● A-chain
● B-chain
— Disulfide bond (S–S)

PDB ID: 1TRZ

3D STRUCTURE



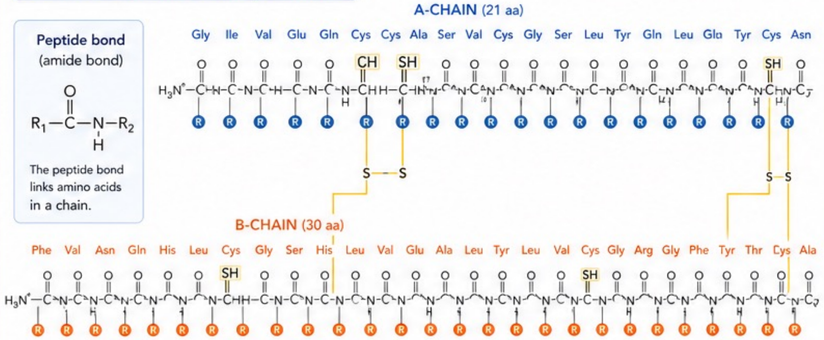
PRIMARY STRUCTURE (Amino Acid Sequence)



CHEMICAL STRUCTURE (Peptide Bonds)

Peptide bond
(amide bond)

The peptide bond links amino acids in a chain.



Molecular formula: $C_{257}H_{383}N_{65}O_{77}S_6$ **Molecular weight:** 5808.6 g/mol (monomer)

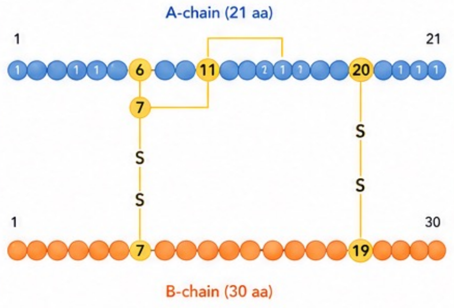
- A-chain: 21 amino acids
- B-chain: 30 amino acids
- Three disulfide bonds: A6–A11 (intra-chain), A7–B7 (inter-chain), A20–B19 (inter-chain)

KEY FEATURES

- Insulin consists of two polypeptide chains:
  - – A-chain (21 amino acids)
  - – B-chain (30 amino acids)
- The chains are linked by three disulfide bonds.
- Disulfide bonds are crucial for maintaining the 3D structure and biological activity.
- Synthesized in the β-cells of the pancreas as preproinsulin, processed to proinsulin, then to mature insulin.
- Major function: promotes cellular uptake of glucose and regulates carbohydrate, fat and protein metabolism.

○ Nonpolar amino acid
● Polar uncharged amino acid
● Basic amino acid
● Acidic amino acid
● Cysteine (forms disulfide bond)

DISULFIDE BOND MAP



Disulfide bonds (S–S):
- A6–A11 (intra-chain)
- A7–B7 (inter-chain)
- A20–B19 (inter-chain)

Nucleic Acids

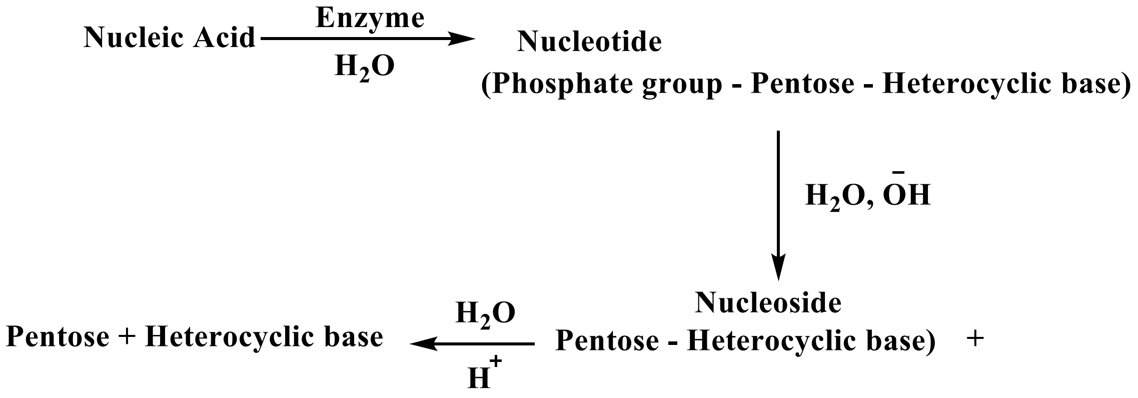
6.4.1 Nucleic Acids

6.3.3.2 Primary Structure of Polypeptides and Proteins

- Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are molecules that carry genetic information in cells.
- DNA is the molecular archive of instructions for protein synthesis.
- RNA molecules transcribe and translate the information from DNA for the mechanics of protein synthesis.

6.4.2 Nucleotides and Nucleosides

- Nucleic acids are linear polymers made up of nucleotide units.
- Hydrolysis of nucleic acids releases nucleotides, the basic building blocks of these molecules.



Nucleic Acid $\xrightarrow[H_2O]{\text{Enzyme}}$ **Nucleotide**
(Phosphate group - Pentose - Heterocyclic base)

$\downarrow$ $H_2O, \bar{O}H$

Nucleoside
Pentose - Heterocyclic base) +

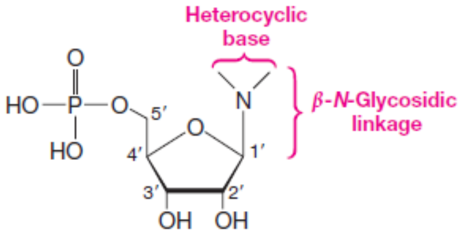
Pentose + Heterocyclic base $\xleftarrow[H^+]{H_2O}$ **Pentose - Heterocyclic base)**

6.4.2 Nucleotides and Nucleosides

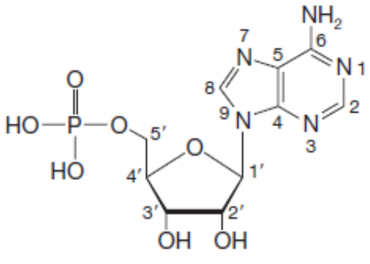
6.4.2.1 The General Structure of Nucleic acids

A. Nucleotides

- The Nucleotide Unit is composed of:
  1. A heterocyclic base from either the purine or pyrimidine family.
     - **The heterocyclic base** of a nucleotide is attached through an N-glycosidic linkage to C1′ of the ribose or deoxyribose unit, and this linkage is always β.
  2. A five-carbon monosaccharide that is either d-ribose or deoxy-d-ribose.
  3. A phosphate ion.
     - **The phosphate group** of a nucleotide is present as a phosphate ester and may be attached at C5′ or C3′.



A nucleotide



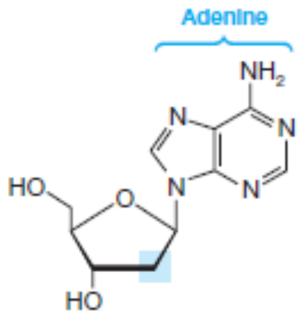
Adenylic acid

6.4.2 Nucleotides and Nucleosides

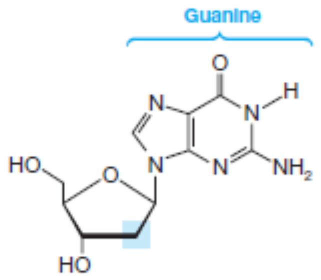
6.4.2.1 The General Structure of Nucleic acids

B. Nucleosides

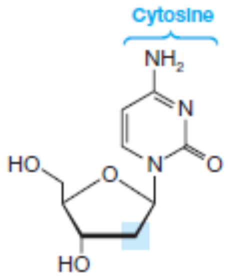
- Removal of the phosphate group of a nucleotide converts it to a compound known as **a nucleoside**.
- DNA-derived nucleosides consist of 2-deoxy-D-ribose and one of four bases: adenine, guanine, cytosine, or thymine.
- RNA nucleosides contain D-ribose and uracil instead of thymine, which is the main structural difference from DNA.
- The nucleosides obtained from DNA contain 2-deoxy-D-ribose as the pentose sugar and one of four heterocyclic nitrogenous bases: adenine, guanine, cytosine, or thymine.



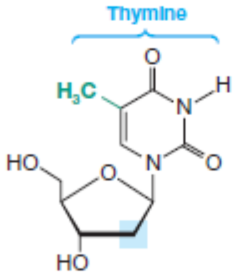
2′-Deoxyadenosine



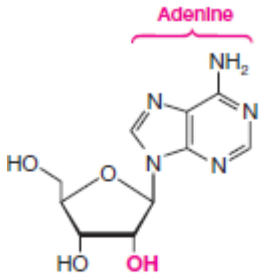
2′-Deoxyguanosine



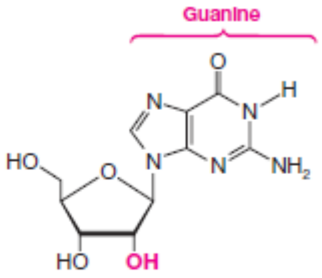
2′-Deoxycytidine



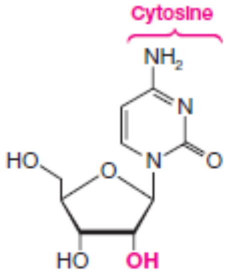
2′-Deoxythymidine



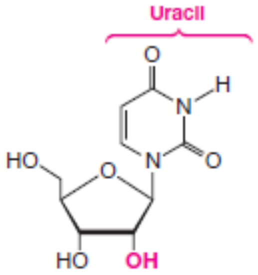
Adenosine



Guanosine



Cytidine



Uridine

6.4.2 Nucleotides and Nucleosides

6.4.2.1 The General Structure of Nucleic acids

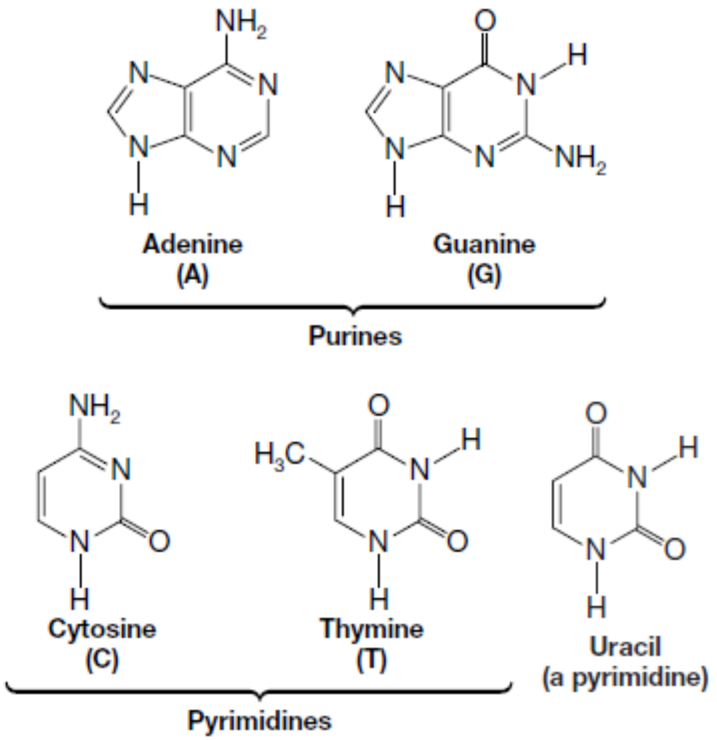
A. Nucleotides

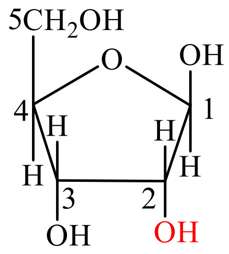
A-1: Heterocyclic Bases

- The nitrogenous base—either a purine or a pyrimidine—is linked to the 1′-carbon of the pentose sugar via replacement of the hydroxyl group.
- Uracil replaces thymine in an RNA nucleoside (or nucleotide).

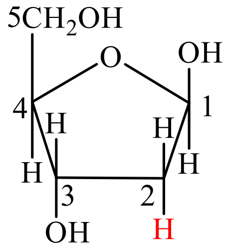
A-2: A five-carbon monosaccharide

- The five-carbon monosaccharide in nucleic acids is either ribose or 2-deoxyribose, which are found in RNA and DNA, respectively.
- The key structural difference is at the C2′ carbon: ribose contains a hydroxyl (–OH) group, whereas in deoxyribose this group is replaced by hydrogen (–H).





D-ribose



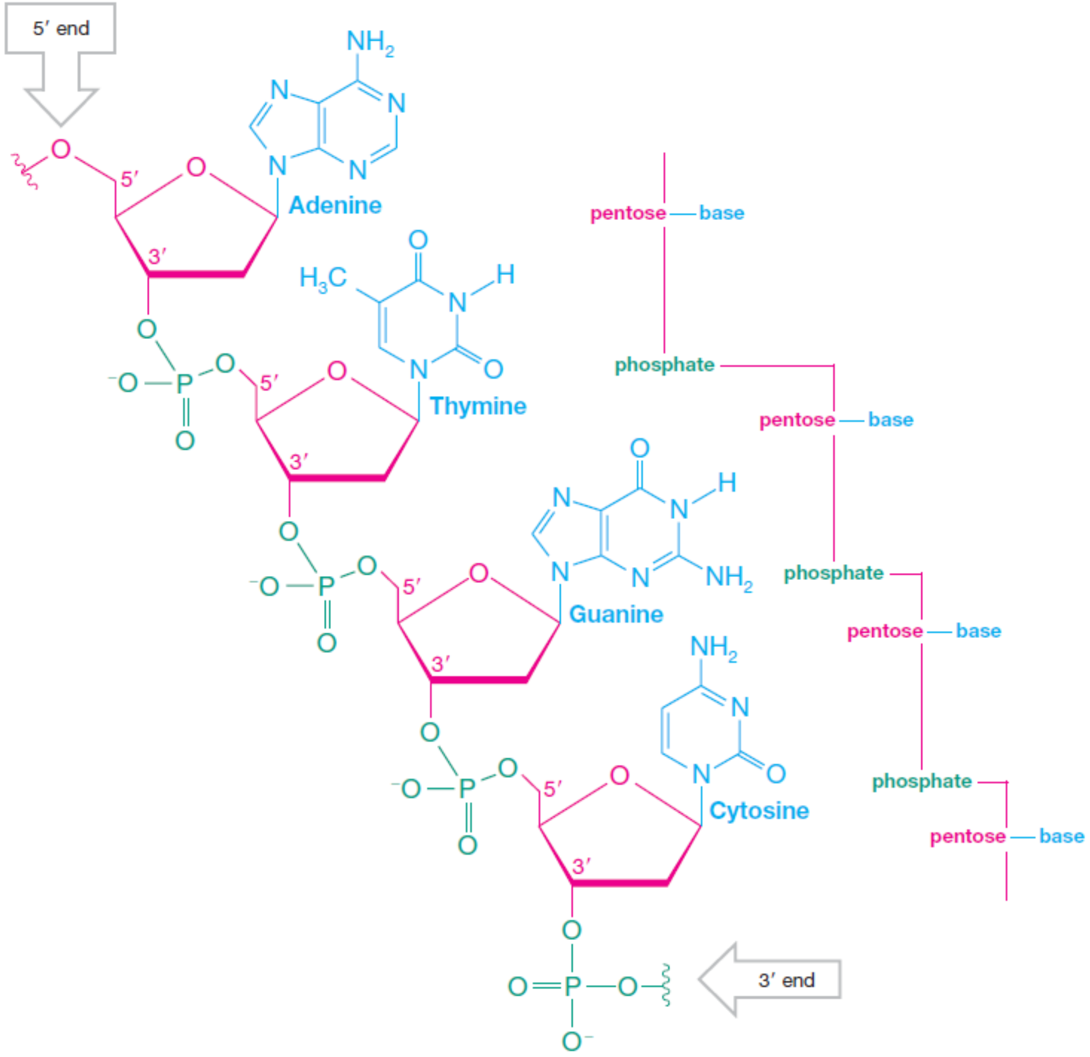
2-deoxy-D-ribose

6.4.2 Nucleotides and Nucleosides

6.4.2.1 The General Structure of Nucleic acids

A-3: Phosphate group

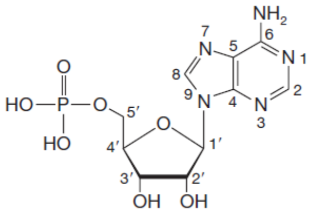
- Phosphate groups link nucleotides through 3′–5′ phosphodiester bonds, forming the sugar–phosphate backbone of nucleic acids.
- The phosphate is attached to the 5′-carbon of one sugar and the 3′-hydroxyl of the next.



6.4.2 Nucleotides and Nucleosides

6.4.2.2 Nucleotide Nomenclature

- Nucleotides are named in several ways. For example Adenylic acid which derived from Adenine base can be written as adenosine monophosphate, and the position of phosphate group can be mentioned adenosine 5′-monophosphate or 5′-adenylic acid

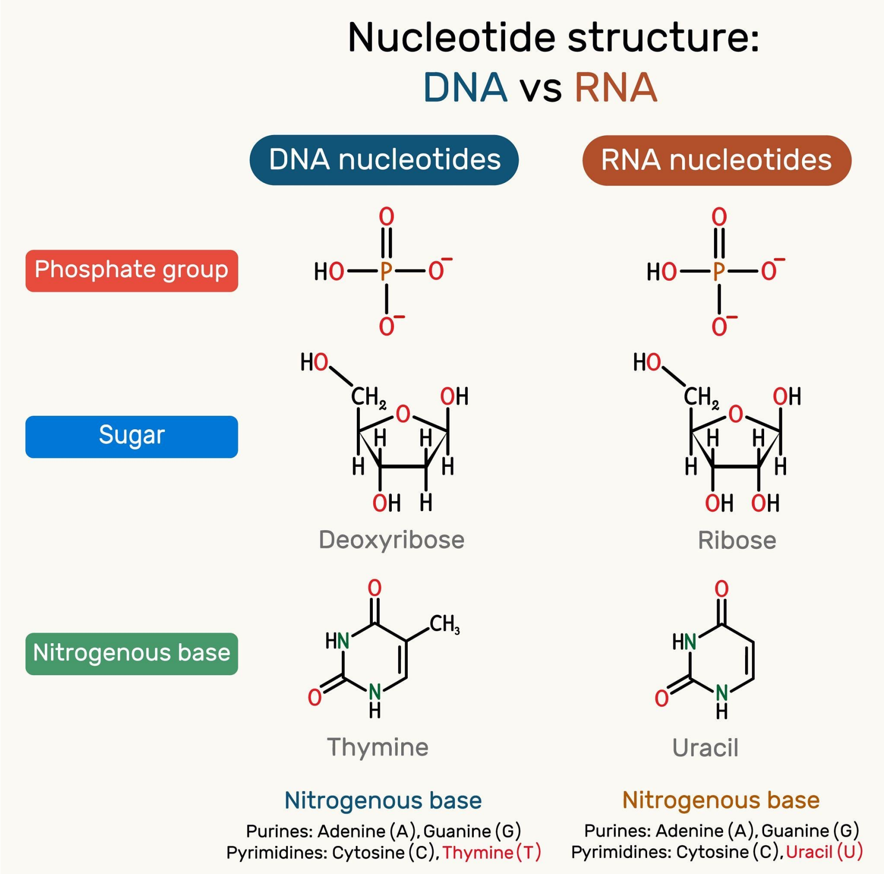


Adenylic acid

Nitrogenous base	Nucleoside	Nucleotide (common name)	Systematic name
Adenine	Adenosine	Adenylic acid	Adenosine 5′-monophosphate (AMP)
Guanine	Guanosine	Guanylic acid	Guanosine 5′-monophosphate (GMP)
Cytosine	Cytidine	Cytidylic acid	Cytidine 5′-monophosphate (CMP)
Uracil	Uridine	Uridylic acid	Uridine 5′-monophosphate (UMP)
Thymine	Thymidine	Thymidylic acid	Thymidine 5′-monophosphate (TMP)

6.4.2.3 The structural difference between DNA and RNA

	DNA	RNA
Sugar	2-deoxy-D-ribose	D-ribose
A heterocyclic base	Adenine (A), Thymine (T), Guanine (G), Cytosine (C)	Adenine (A), Uracil (U), Guanine (G), Cytosine (C)
Number of strands	double-stranded (double helix)	single-stranded



6.4.2 Nucleotides and Nucleosides

6.4.2.3 DNA and RNA – Distinctions Summary