Taras Shevchenko National University of Kyiv
The Institute of Biology and Medicine

T.B. Synelnyk, O.S. Kostiuk, L.I. Ostapchenko

BIOORGANIC CHEMISTRY

Kyiv-2021

Authors:
T.B. Synelnyk, O.S. Kostiuk, L.I. Ostapchenko

Reviewers:
Dr. Sc, Professor Zholos O. V.
Dr. Sc, Dvorshchenko K. O.

Approved by Academic Council of Educational Scientific Center "Institute of Biology and Medicine" (protocol № 4 from November 26, 2020)

Bioorganic chemistry: textbook / T.B. Synelnyk, O.S. Kostiuk, L.I. Ostapchenko. – K.: Online edition, 2021. – 367 p.

The textbook provides a concise yet comprehensive overview of key topics essential for students in medical and biological specializations and is focused on the structure, properties, and biological roles of key organic compounds – carboxylic acids, heterofunctional compounds, fatty acids, lipids, amino acids, proteins, carbohydrates, heterocyclic compounds, and nucleic acids – with direct relevance to human physiology and biochemistry and medical science. Emphasizing clinical significance and biochemical mechanisms, this book provides a solid foundation for understanding the molecular basis of health and disease, essential for future medical practice and biomedical research.

The textbook is designed for medical students and students of related biological specializations studying in English.

CONTENT

THE LIST OF ABBREVIATIONS

ABC – ATP-binding cassette transporters
AchR – nicotinic acetylcholine receptor
ADH – antidiuretic hormone (the same as vasopressin)
AMP – adenosine-5'-monophosphate
Apo – apoprotein
AQPs – aquaporines
ATP – adenosine 5'-triphosphate
cADP-ribose – cyclic ADP-ribose
cAMP – cyclic adenosine 3',5'-monophosphate
CFTR – cystic fibrosis transmembrane conductance regulator)
CoA – coenzyme A (coenzyme of acetylation)
CS A – chondroitin-4-sulphate
CS C – chondroitin-6-sulphate
CTP – cytidine 5'-triphosphate
dATP – deoxyadenosine 5'-triphosphate
dCTP – deoxycytidine 5'-triphosphate
DG – diacylglycerols
dGTP – deoxyguanosine 5'-triphosphate
DHPS – dihydropteroate synthase
DNA – deoxyribonucleinic acid
DS – dermatan sulphate (also known as CS B)
dsDNA – double strand DNA
dTTP – thymidine/deoxythymidine 5'-triphosphate
ECM – extracellular matrix
ELG – extracellular ligand-gated () Cl^--channels
FAD – flavine adenindinucleotide
FMN – flavine mononucleotide
GABA – γ-aminobutyric acid
GAGs – glycosaminoglycans
GLUTs – glucose transporters
GTP – guanosine 5'-triphosphate
HA – hyaluronic acid (also called hyaluronan)
Hb – hemoglobin
HbA – adult Hb

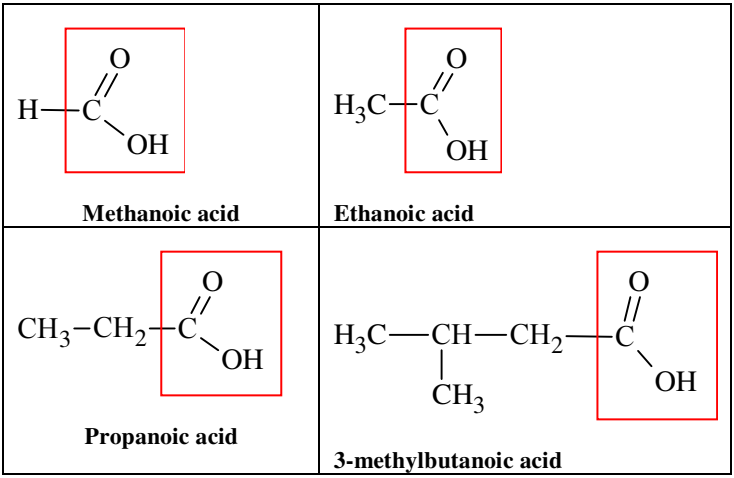
HbF – fetal Hb
HbS – hemoglobin in sickle-cell anemia
HDLs – high density lipoproteins
hnRNA – heterogeneous nuclear RNA
HS – heparan sulphate
HVA – high voltage activated Ca^{2+}-channels
Ig – immunoglobulin
IUPAC – International Union of Pure and Applied Chemistry
IVA – intermediate-voltage-activated Ca^{2+}-channels
KS – keratan sulphate
LDLs – low density lipoproteins
LSD – lysergic acid
LT – leukotrienes
LVA – low voltage activated Ca^{2+}-channels
MDRs – multidrug-resistance glycoproteins
MG – monoacylglycerols
mRNA – messenger RNA
MRPs – multidrug-resistance protein
NMDA-R – receptor for glutamate
NSAIDs – nonsteroidal anti-inflammatory drugs
PABA – *p*-aminobenzoic acid
PGIs – prostacyclins
P-gp – P-glycoproteins
PGs – Proteoglycans
PMCA – plasma membrane Ca^{2+}-ATPase
RBCs – red blood cells
RFOs – raffinose family of oligosaccharides
RNA – ribonucleinic acid
rRNA – ribosomal RNA
s/u – subunit
scRNA – small cytoplasmic RNA
SDS – sodium dodecylsulfate
SERCA – sarco/endoplasmic reticulum Ca^{2+}-ATPase
snoRNA – small nucleolar RNA
snRNA – small nuclear RNA
ssDNAs – single-stranded DNAs
T3 – triiodothyronine

T4 – thyroxine
TEA – tetraethyl ammonium
TG – triacylglycerols
T_m – melting temperature of DNA
tRNA – transfer RNA
TXs – thromboxanes
UTP – uridine 5'-triphosphate
UV – ultraviolet
VLDL – very low density lipoproteins

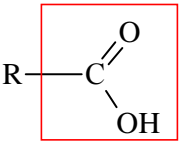
PART 1. CARBOXYLIC ACIDS AND HETEROFUNCTIONAL COMPOUNDS

1.1. CARBOXYLIC ACIDS

Carboxylic acids are the organic compounds whose characteristic functional group is the **carboxyl group −COOH**. In this group a carbon (C) atom is bonded to an oxygen (O) atom by a double bond and to a hydroxyl group (−OH) by a single bond. The carboxyl (COOH) group is so-named because of the ***carb***onyl group (C=O) and hydr***oxyl*** group (OH):



Therefore the common formula of carboxylic acid is:



Carboxylic acids and their derivates are widespread in nature and are important for studying of biochemistry, pharmacology and other medical disciplines. They get into organism with food and further are

converted into other substances. *Fatty acids* are components of lipids, *lactic acid* (found in sour-milk products) and *citric acid* (found in citrus fruits), as well as many *keto acids* are important metabolic products that exist in most living cells. Proteins are made up of *amino acids*, which also contain carboxyl groups.

Some carboxylic acids and their salts are medical products. For example, *isovaleric acid* is a part of validol, *sodium benzoate* is important disinfectant, and *aspirin* - the ester of *salicylic acid* - is prepared from *acetic acid.*

1.1.1. Nomenclature of carboxylic acids

Systematic names of carboxylic acids are derived from the name of the longest carbon chain that contains the carboxyl group via dropping the final -e from the name of the parent alkane and adding the suffix -oic followed by the word "acid" (Fig. 1.1)**.** The chain is numbered beginning with the carbon of the carboxyl group.

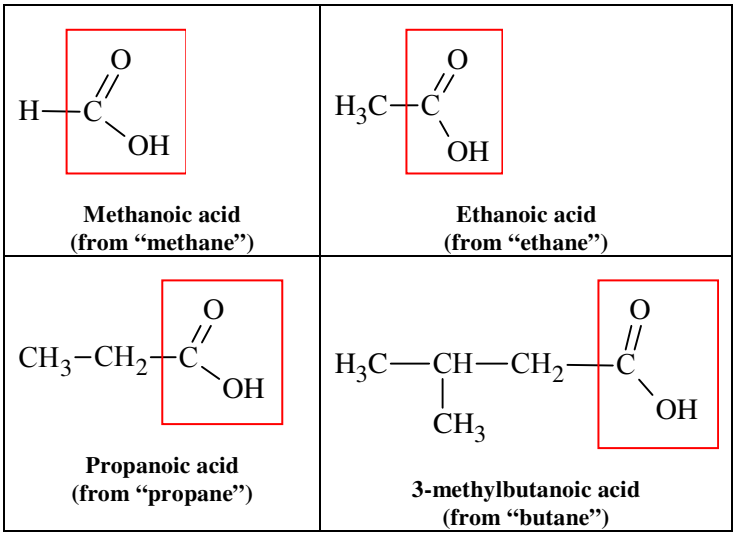


Fig 1.1. The systematic names of some carboxylic acids

For example, the compound CH_3CH_2COOH has three carbon atoms and is called *propanoic acid*, from *propane*, the name for a three-carbon chain, with *-oic acid*.

If the carboxylic acid contains a carbon-carbon double bond, the ending is changed from *-anoic acid* to *-enoic acid* to indicate the presence of the double bond, and a number is used to show the location of the double bond.

Most simple carboxylic acids were originally isolated from biological sources; because their structural formulas were often unknown at the time of isolation they were given names – **trivial, or common names** – that were generally derived from the names of the sources (Fig. 1.2). For example, $CH_3CH_2CH_2COOH$, *butyric acid*, first obtained from butter.

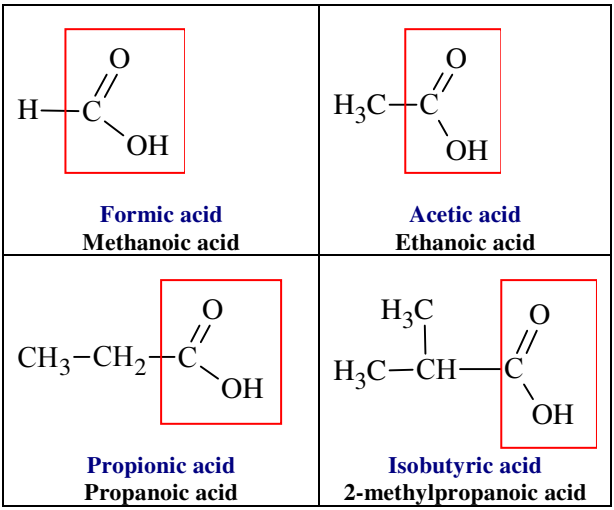
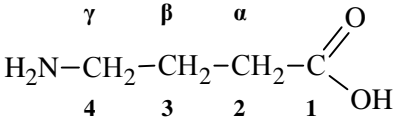


Fig 1.2. The trivial (common; blue color) and systematic names of some carboxylic acids

When common names are used, substituents on the hydrocarbon chain are designated by Greek letters rather than by numbers, and counting begins not with the carboxyl carbon but with

the adjacent carbon. For example, the common name of the *4-aminobutanoic acid* (the IUPAC (International Union of Pure and Applied Chemistry) name) is *γ-aminobutyric acid* (abbreviated *GABA*):



4-aminobutanoic acid
(γ-aminobutyric acid, GABA)

This compound is an inhibitory neurotransmitter in the central nervous system of humans.

Salts of carboxylic acids are named in the same manner as are the salts of inorganic compounds: the cation is named first and then the anion, as in sodium chloride. For carboxylic acids, the name of the anion is derived by changing the ending -oic acid of the IUPAC name or -ic acid of the common name to -ate. For example, CH_3COONa is called *sodium acetate*, $HCOONH_4$ - *ammonium formate,* $CH_3CH_2CH_2COOK$ - *potassium butanoate (potassium butyrate).*

1.1.2. Classification of carboxylic acids

Carboxylic acids can be classified depending on the nature of hydrocarbon radical:

- **saturated acids** are the acids, which have only simple bonds in molecule: *formic acid, butanoic acid* (Tab. 1.1)**;**
- **unsaturated acids** are the acids, which have both as simple bonds and double bonds in molecule: *oleic acid, linoleic acid, linolenic acid, arachidonic acid* (Tab. 1.2)**;**

- **aromatic acids** are the acids, which contain aromatic ring: *benzoic acid.*

According to another classification, which is based on the number of carboxyl groups, carboxylic acids can be divided into:

- **monocarboxylic acids**, which have one carboxylic group in molecule: *acetic acid, formic acid, butanoic acid;*
- **dicarboxylic acids**, which have two carboxylic group in molecule: *oxalic acid, malonic acid* (Tab. 1.3).

1.1.3. Properties of carboxylic acids

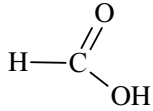
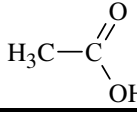
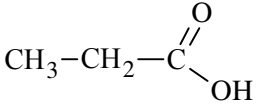
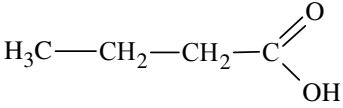
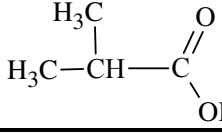
Acidity. Carboxylic acids donate H^+ (hydrogen ion, also called a proton), to another compound, termed a *base*:

$$CH_3COOH + NaOH \leftrightarrow CH_3COONa + H_2O$$

Carboxylic acids are generally more acidic than other organic compounds with hydroxyl groups but are generally weaker than the familiar mineral acids (e.g., hydrochloric (HCl), sulfuric (H_2SO_4) or nitric (HNO_3) acids).

Solubility. The solubility of carboxylic acids in water is similar to that of alcohols, aldehydes and ketones: acids with fewer than about five carbons dissolve in water, whereas those with a higher molecular weight are insoluble throw the larger hydrocarbon chain, which is hydrophobic. The sodium, ammonium, and potassium salts of carboxylic acids, however, are generally quite soluble in water – so, almost any carboxylic acid can be made to dissolve in water by converting it to salt, which is easily done by adding a strong base — most commonly sodium hydroxide (NaOH) or potassium hydroxide (KOH).

Table 1.1. The names of some saturated monocarboxylic acids

Structural formula	Name of nomenclature		
	Trivial	**Substitute**	**Rational**
	Formic acid	Methanoic acid	-
	Acetic acid	Etanoic acid	Acetic acid
	Propionic acid	Propanoic acid	Methylacetic acid
	Butyric acid	Butanoic acid	Ethylacetic acid
	Isobutyric acid	2-methylpropanoic acid	Dimethylacetic acid

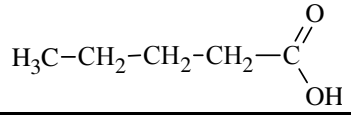
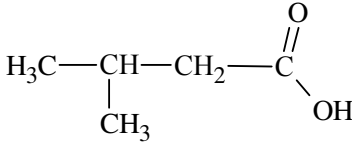
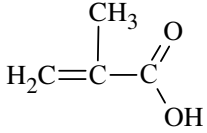
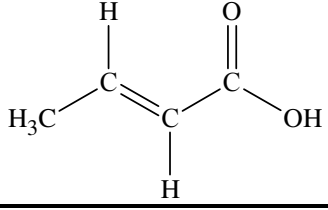
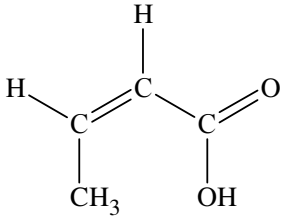
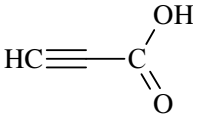
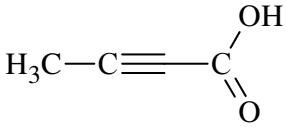
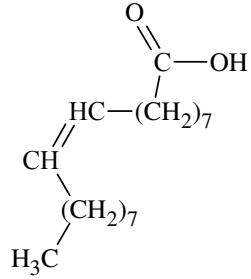
	Valeric acid	Pentanoic acid	Propylacetic acid
	Iso valeric acid	3-methylbutanoic acid	Methylethylacetic acid
CH_3-$(CH_2)_4$-COOH	Capronic acid	Hexanoic acid	n-butylacetic acid
CH_3-$(CH_2)_{10}$-COOH	Lauric acid	Dodecanoic acid	-
CH_3-$(CH_2)_{12}$-COOH	Myristic acid	Tetradecanoic acid	-
CH_3-$(CH_2)_{14}$-COOH	Palmitic acid	Hexadecanoic acid	-
CH_3-$(CH_2)_{16}$-COOH	Stearic acid	Octadecanoic acid	-

Table 1.2. The names of some unsaturated monocarboxylic acids

Structural formula	Name of nomenclature	
	Trivial	**Substitute**
CH_2=CH-COOH	Acrylic acid	Propenoic acid
	Methacrylic acid	2-methylpropenoic acid
CH_2=CH-CH_2-COOH	Vinyl acetic acid	3-butenoic acid
	Crotonic acid	*trans*-2-butenoic acid

	Iso crotonic acid	*cis*-2-butenoic acid
	Propiolic acid	Propionoic acid
	Tetrolic acid	2-butynoic acid
	Oleic acid	*cis*-9-octadecenoic acid

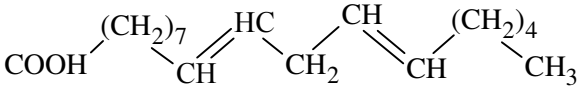
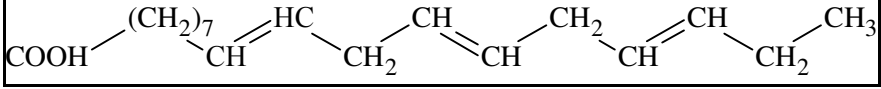
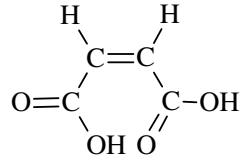
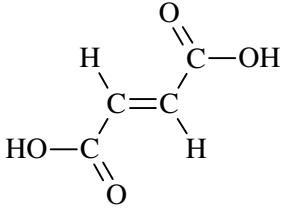
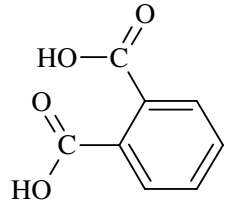
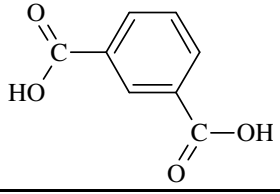
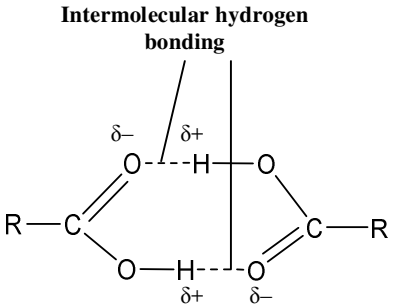
	Linoleic acid	*cis*-9-*cis*-12-octadecadienoic acid
	linolenic acid	*cis*-9-*cis*-15-octadecatrienoic acid

Table 1.3. The names of some dicarboxylic acids

Structural formula	**Name of nomenclature**	
	trivial	**substitute**
HOOC-COOH	oxalic acid	ethandioic acid
HOOC-CH_2-COOH	malonic acid	propandioic acid
HOOC-CH_2-CH_2-COOH	succinic acid	butandioic acid
HOOC-CH_2-CH_2-CH_2-COOH	glutaric acid	pentandioic acid
HOOC-CH_2-CH_2-CH_2-CH_2-COOH	adypinic acid	hexandioic acid
HOOC-$(CH_2)_5$-COOH	pimelic acid	heptadioic acid
HOOC-$(CH_2)_6$-COOH	cork acid	octandioic acid

	maleic acid	*cis*-butendioic acid
	fumaric acid	*trans*-butendioic acid
	phthalic acid	1,2-benzoldicarbonic acid
	iso phthalic acid	1,3-benzoldicarbonic acid

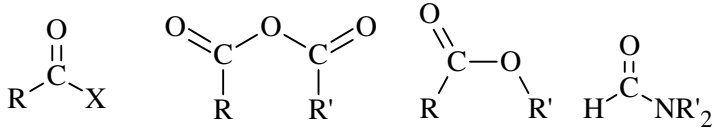
Boiling point. Carboxylic acids have much higher boiling points than hydrocarbons, alcohols, ethers, aldehydes or ketones with similar molecular weight. So, formic acid boils at 101 °C whereas ethanol (ethyl alcohol), C_2H_5OH – at 78.5 °C, although the two have nearly identical molecular weights. This difference is due to two molecules of a carboxylic acid form two hydrogen bonds with each other (two alcohol molecules can form only one). Thus, carboxylic acids exist as dimers (pairs of molecules):



Therefore, if the dimer is broken upon boiling, more energy is required to break these two hydrogen bonds.

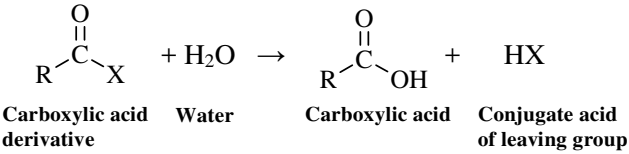
1.1.4. Chemical properties of carboxylic acids. Carboxylic acid derivatives

Compounds in which the −OH of the carboxyl group is replaced by certain other groups are called **carboxylic acid derivatives**, the most important of which are:



Acyl halide (X = F, Cl, Br, I) | **Acyl anhydride** | **Ester** | **Amide**

All of them may be converted to carboxylic acids by hydrolysis:



The scheme of some benzoic acid derivatives formation is given on Fig. 1.3.

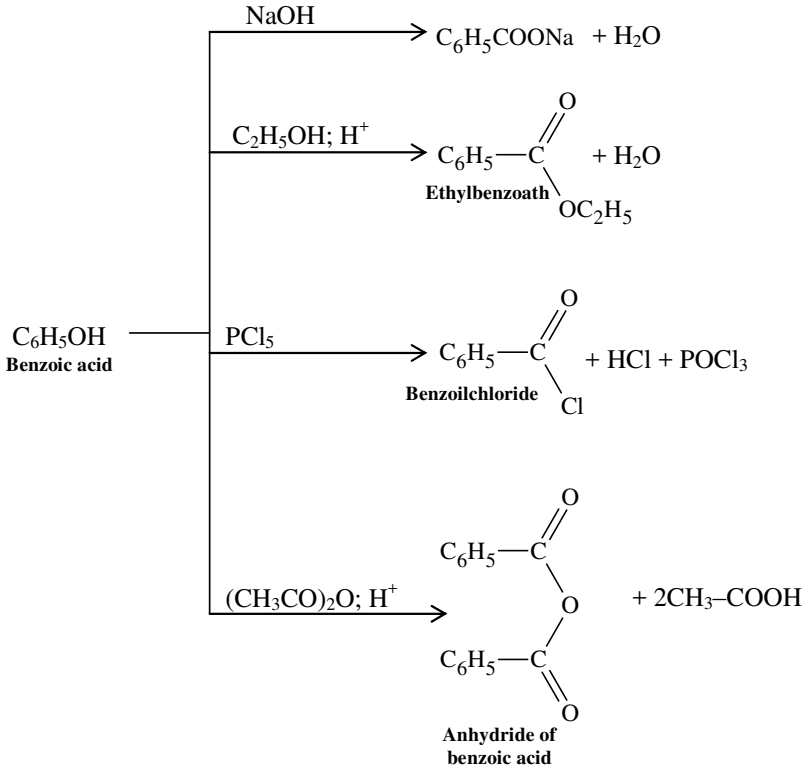
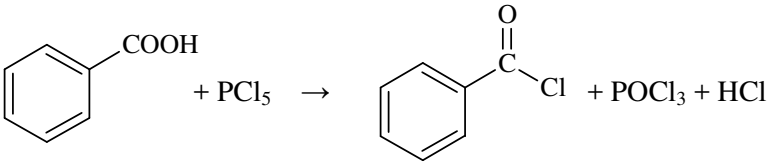
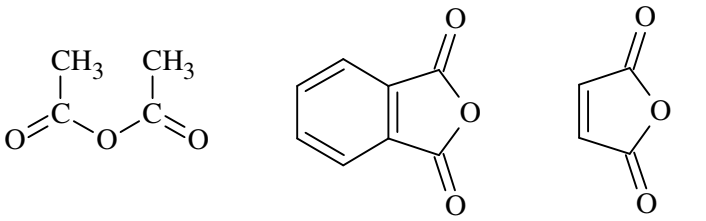


Fig. 1.3. The scheme of some benzoic acid derivatives (salts, esters, halides, anhydrides) formation

1. Reaction with halo-compounds passes through such a scheme:



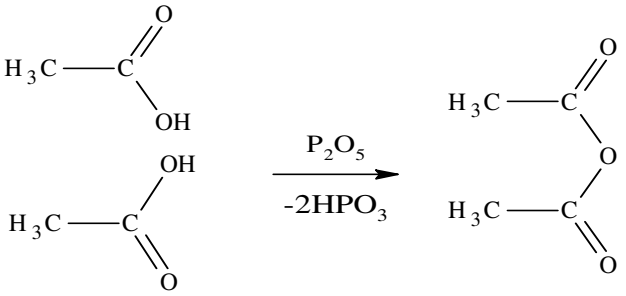
2. The schemes of **carboxylic acid anhydrides formation** are given below. Acid anhydrides (Fig. 1.4) are the most reactive carboxylic acid derivatives.



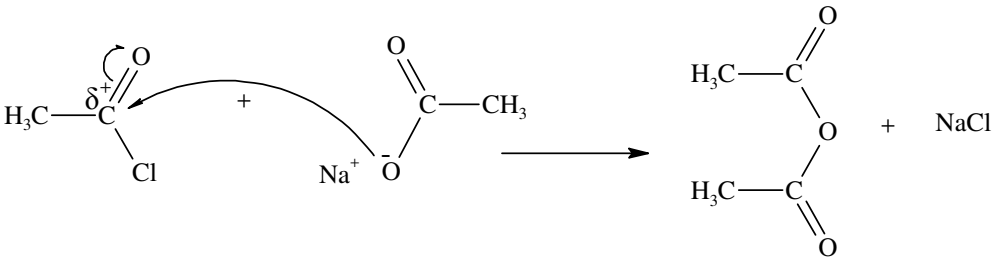
Acetic anhydride **Phthalic anhydride** **Maleic anhydride**

Fig. 1.4. Carboxylic acids anhydrides examples

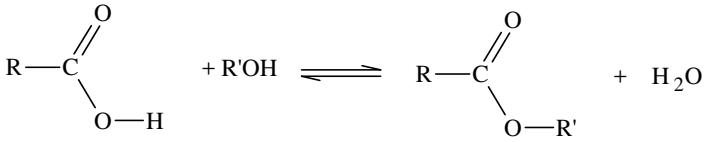
Formation of anhydrides can be a consequence of the reaction of carboxylic acids with phosphorus oxide (V):

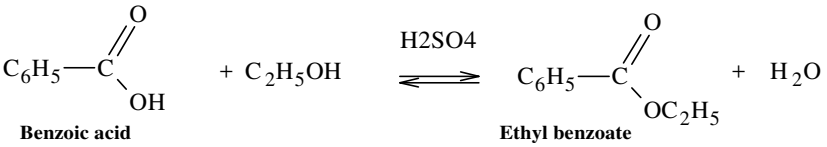


or of the reaction of halogen anhydrides with carboxylic acid salts (without water):



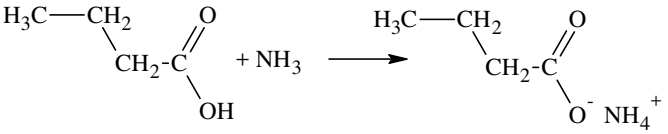
3. Esterification is the mainly reaction between alcohols and carboxylic acids to make esters:



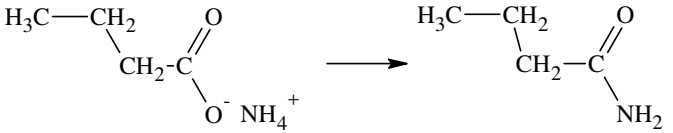


4. The amides formation. The most common reaction of this type is the reaction of carboxylic acids with ammonia or amines to give amides. For example, ammonia reacts with butyric acid with butyramide forming. This reaction involves two stages:

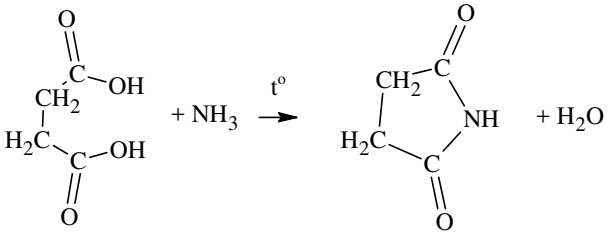
1) at room temperature butyric acid reacts with the weak base ammonia to give the salt ammonium butyrate:



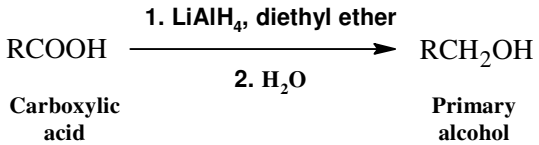
2) ammonium butyrate is perfectly stable at normal temperatures. However, pyrolysis of this salt results in the elimination of water and formation of the amide:



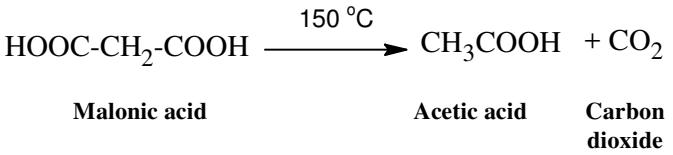
Another example of this type reactions is the succynimide formation when ammonia reacts with succinic acid:



5. Reduction reaction. Carboxylic acids are reduced to primary alcohols by the powerful reducing agent lithium aluminum hydride.



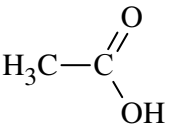
6. Decarboxylation of carboxylic acids is the loss of a molecule of carbon dioxide from a carboxylic acid:



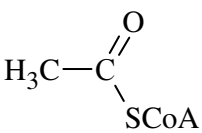
Unsaturated acids exhibit chemical properties expected of compounds that contain both a COOH group and one or more carbon-carbon double bonds. **Like all carboxylic acids** they are *acidic*; they *can be reduced to alcohols*; they *can be converted to acid derivatives*. **Like other compounds containing double bonds** they *can undergo the normal double-bond addition reactions*; they *can undergo the oxidation-reduction reactions*.

1.1.5. The most important representatives of saturated carboxylic acids

Acetic acid (CH_3COOH) is important in the metabolic processes of humans and all animals and plants. In these processes the *acetyl group (CH_3CO-)* of the acetic acid molecule is attached to a large biochemical molecule called *coenzyme A (HS−CoA) with acetyl coenzyme A (acetyl−S−CoA)* formation.



Acetic acid



Acetyl-SCoA

In the processes of metabolism the carbon atoms of carbohydrates, fats, and, to some degree, proteins are converted to acetyl groups that are bonded to coenzyme A forming acetyl coenzyme A. The acetyl group of acetyl coenzyme A is then converted via the tricarboxylic acid cycle (or Krebs's cycle) and oxidative phosphorylation to energy (in the form of adenosine triphosphate, or ATP) and carbon dioxide (CO_2), which is exhaled. Some of the acetyl groups aren't converted to energy, but are used to fatty acids, terpenes, steroids and other needed molecules synthesis (Fig. 1.5).

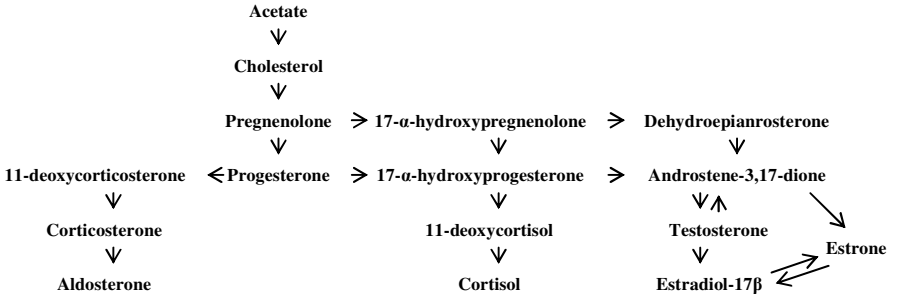


Fig. 1.5. The scheme of steroid hormones formation from acetic acid salts

The fatty acids (acids of fats) from 4 to 10 carbon atoms are mostly found in milk fats. Therefore, *butanoic (butyric) acid*, $CH_3CH_2CH_2COOH$, is an important component of cow's milk, and goat's milk is rich in fats containing the 6-, 8-, and 10-carbon acids - *hexanoic (caproic), octanoic (caprylic),* and *decanoic (capric)* acids, respectively. The higher saturated fatty acids, from C12 to C18 (*lauric, myristic, palmitic,* and *stearic*), are present in the fats and oils of many animals and plants. Saturated fatty acids higher than C18 are much less common in fats but do occur in some waxes.

Waxes obtained from animal and plant sources typically consist of esters derived from long-chain acids and long-chain alcohols (Fig. 1.6).

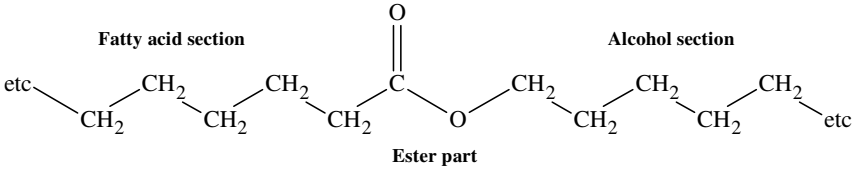
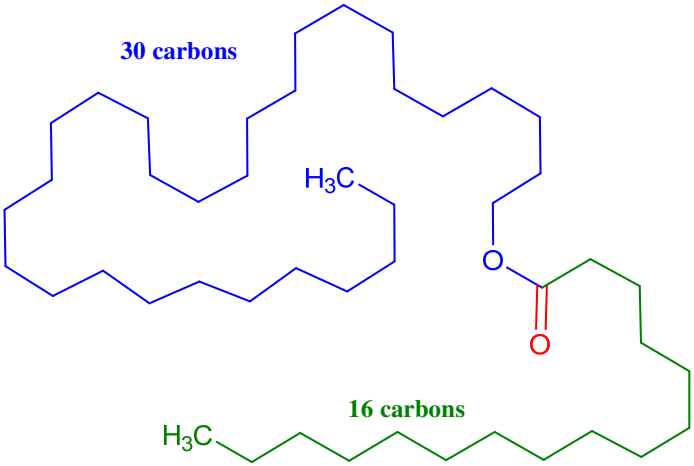


Fig. 1.6. The scheme of waxes structure

For example, beeswax contains, among many other components, the ester made from *cerotic acid (C26)* and *triacontanol* (the

unbranched-chain alcohol containing 30 carbons). Triacontanyl palmitate is another component of beeswax:



Tricontanyl palmitate

$CH_3(CH_2)_{29}O–CO–(CH_2)_{14}CH_3$

1.1.6. The most important representatives of unsaturated carboxylic acids

Unsaturated fatty acids have some peculiarities. If they have two or more carbon-carbon double bonds, each double bond is separated from the next by a −CH2− group (methylene group) – so they are called *the isolated double bonds***.** In addition, double bonds have *cis-configuration*. Unsaturated fatty acids with above 12 carbon atoms (*long-chain fatty acids*) are liquids at room temperature, in contrast to the saturated fatty acids, which are solids.

The fats are esters of these long-chain fatty acids and the alcohol glycerol. **Solid fats**, obtained mostly from animal sources, have a

high percentage of saturated fatty acids. **Liquid fats** (called *oils*), obtained mainly in plant, have a high percentage of unsaturated fatty acids.

Essential fatty acids – ***linoleic*** and ***linolenic acids*** – are necessary for the human body, but the body can't synthesize them, and they must be obtained in the diet.

One of the unsaturated fatty acids - ***arachidonic acid -*** is used by the human body to syntheze of two kinds of essential substances, *prostaglandins* (for example, PGE1) and *leukotrienes* (for example, LTB4) (Fig. 1.7).

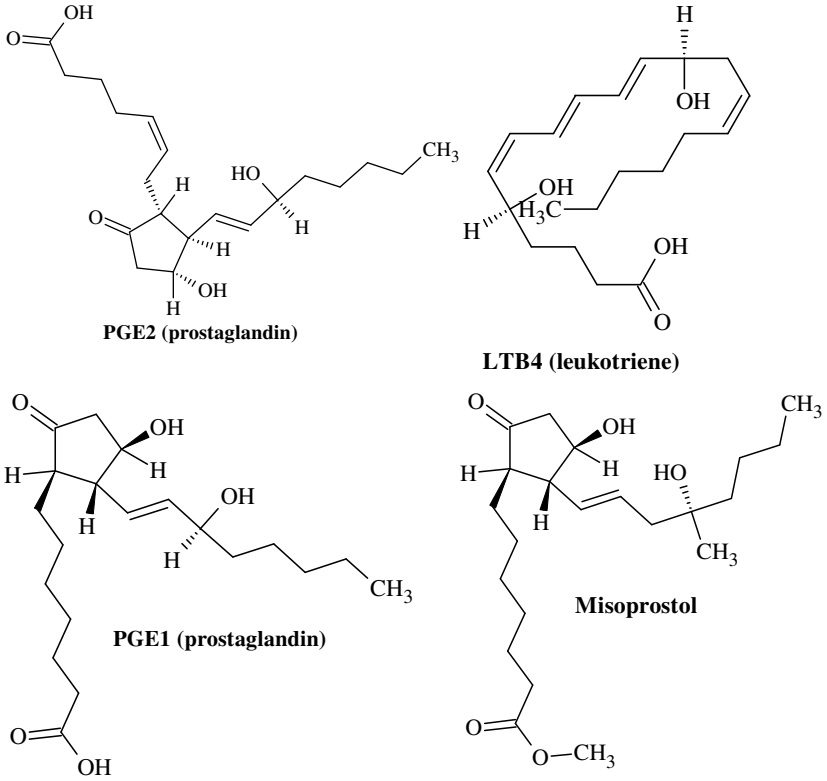
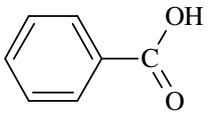


Fig. 1.7. The structures of some prostaglandins, leukotrienes and PGE1 analog misoprostol

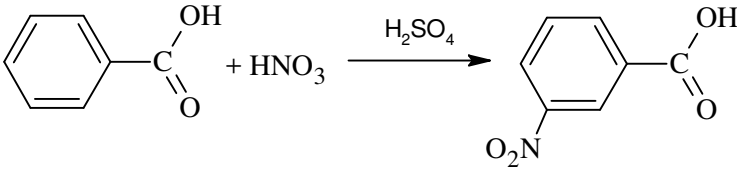
Prostaglandins and leukotrienes (which together with tromboxanes are called *eicosanoids*) are made in small amounts and act as hormone mediators. Some prostaglandins raise blood pressure, whereas others lower it. The PGE1 analog, *misoprostol*, is currently used to prevent ulceration associated with the use of nonsteroidal anti-inflammatory drugs (NSAIDs).

1.1.7. Aromatic carboxylic acids

These compounds contain a COOH group bonded to an aromatic ring. The simplest aromatic acid is *benzoic acid.*



Aromatic carboxylic acids show the acidity and other reactions expected of carboxylic acids (as an acid, benzoic acid is slightly stronger than acetic acid). They also similar to other aromatic compounds can undergo electrophilic substitution reactions (for example, the nitration reaction):



Benzoic acid occurs in various plants, both in free acid form and in ester form. It also presents in the urine of certain animals, especially horses, as an amide of glycine called **hippuric acid**, $C_6H_5CONHCH_2COOH$. It is a solid at room temperature (melting point 122 °C); its sodium salt, *sodium benzoate*, is used as a preservative in many foods.

Some other important aromatic acids are given on Fig. 1.8.

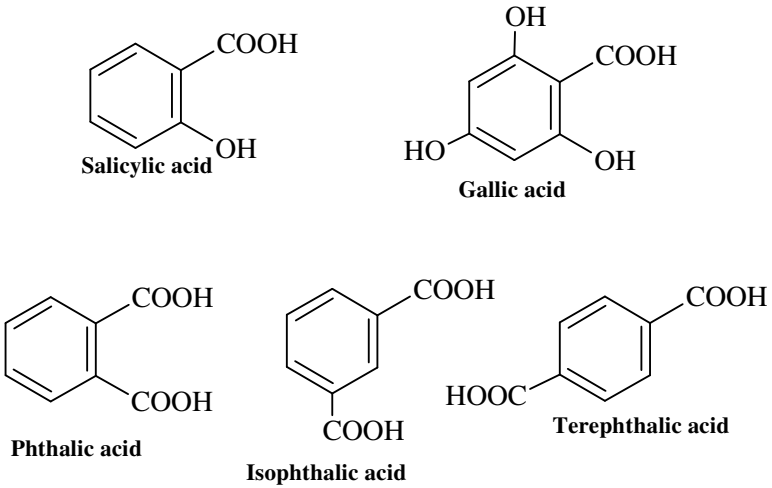


Fig. 1.8. The structures of some aromatic acids

Salicylic acid and **gallic acid** are both aromatic carboxylic acids and phenols (*phenoloacids*).

1.1.8. Polycarboxylic acids

Polycarboxylic acids are unbranched-chain dicarboxylic acids contain two COOH groups, so they can form more than one kinds of salts. For example, for *oxalic acid*, HOOC−COOH, its half-neutralization with NaOH (the acid and base are in a 1:1 molar ratio) results in HOOC−COONa, called *sodium acid oxalate*, formation. Because one COOH group is still present in this compound, it has the properties of both a salt and an acid. Full neutralization of oxalic acid (its treatment with NaOH in a 1:2 acid-to-base molar ratio) forms NaOOC−COONa, *sodium oxalate*. Moreover, the half-neutralization can be done with one base and the rest – with another – so a mixed salt, KOOC−COONa — *sodium potassium oxalate* – is produced.

All dicarboxylic acids can be neutralized or half-neutralized in a similar manner.

Oxalic acid, in the form of its monopotassium salt, is found in many vegetables and fruits. Of much greater importance than other dicarboxylic acid - **malonic acid** - is its diethyl ester called *diethyl malonate* $CH_2(COOCH_2CH_3)_2$. When heated with urea and sodium ethoxide, this compound forms **barbituric acid**. Derivatives of barbituric acid are called *barbiturates* – for example, *pentobarbital* and *thiopental* (Fig. 1.9).

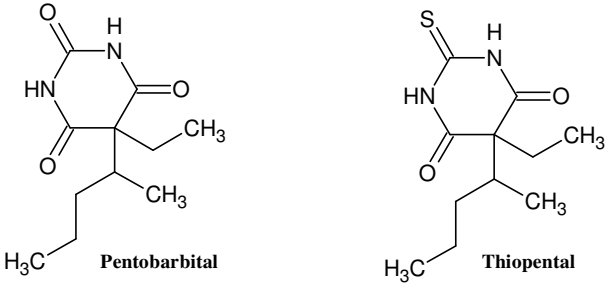


Fig. 1.9. The structures of barbiturates (pentobarbital and thiopental)

Barbiturates have two principal effects: in small doses, they are *sedatives* (*tranquilizers*) whereas in larger doses they have hypnotic effects (induce sleep).

Succinic acid is an important component of the Krebs's cycle, a part of the process by which animals convert food to energy.

1.2. HETEROFUNCTIONAL COMPOUNDS

Functional group is a group of atoms that occur within organic molecule and is involved in characteristic chemical reactivity of this molecule. The functional groups which the most often occur in the heterofunctional compounds are given on Fig. 1.10. Therefore, functional groups confer specific chemical properties of molecules.

- hydroxyl (−OH),

- amino (−NH_2),

- oxo (C=O or −C=O)

- carboxyl group (−C=O)

Fig. 1.10. The main functional groups in organic compounds

There are mono-, poly- and heterofunctional organic compounds. *Monofunctional compounds* contain only one functional group, *polyfunctional* ones – several similar functional groups, whereas *heterofunctional* compounds include several different functional group (Fig. 1.11).

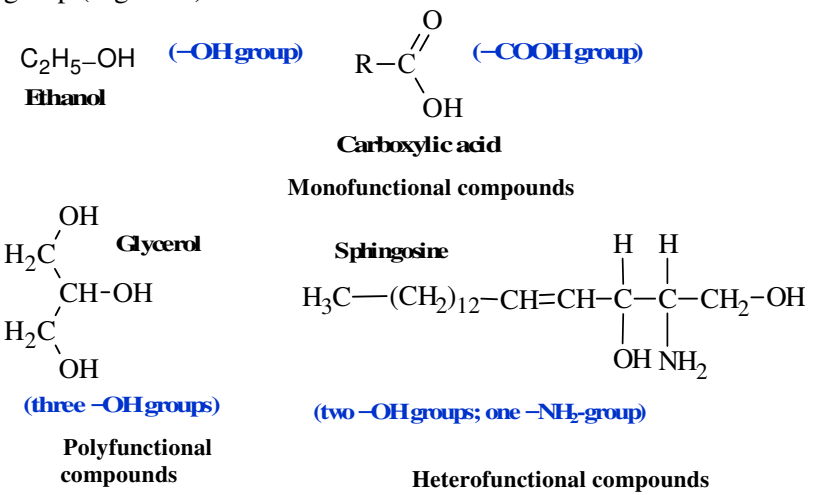


Fig. 1.10. The examples of mono-, poly- and heterofunctional organic compounds

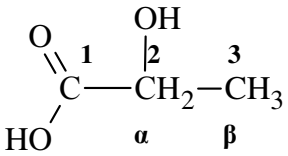
A combination of different functional groups results in the formation of mixed classes of organic compounds such as:

- hydroxyacids,
- phenoloacids
- aminoacids,
- oxoacids.

The chemical behavior of heterofunctional compounds can be represented as a sum of properties of separate monofunctional classes. **Oxo acids** (for example, *pyruvic acid*) can be esterified via their −COOH group and transformed into derivatives on their carbonyl group. **Hydroxy acids** (*salicylic* and *lactic acids*) via −COOH group form esters in the reaction with alcohols, as well as their hydroxyl group can be acylated or alkylated. As *salicylic acid* also belong to **phenoloacids** and has aromatic ring, so it also has properties of aromatic compounds.

1.2.1. Hydroxyacids

Hydroxyacids are the derivatives of carboxyl acids that contain one or more hydroxyl group. Therefore the hydroxyacids have both carboxyl (COOH) and hydroxyl (OH) groups:



2-hydroxypropanoic acid
α-hydroxypropanoic acid

The quantity of carboxyl groups defines **basicity of acids –** they can be *monobasic*, *dibasic, tribasic* ones, etc. Also hydroxyacids are divided into *aliphatic* and *aromatic* (or *phenoloacids*).

The **rational** and **IUPAC nomenclatures** are used for making the names of hydroxyacids. However **trivial names** of hydroxy-(oxo-) acids are more often used (Tab. 1.4).

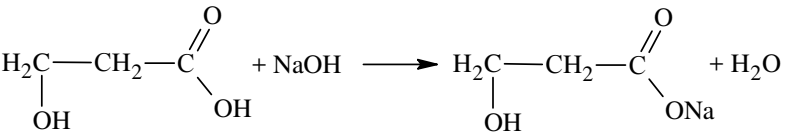
Hydroxycarboxylic acids are colorless liquids or crystalline substance, soluble in water. In the molecule of these compounds either –OH group or –COOH group can react**.**

Table 1.4.
Nomenclature of some hydroxyacids

Trivial	Rational	IUPAC
Lactic acid	α-oxypropionic acid	2-oxypropanoic acid
Malic acid	α-oxysuccinic acid	2-oxybutanedioic acid
Tartaric acid	α, β-dioxysuccinic acid	2,3-dioxybutanedioic acid

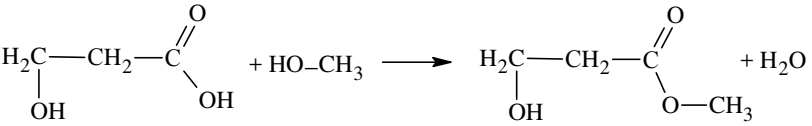
In the reactions **on carboxyl group** can be formed:

a) salts (with bases):



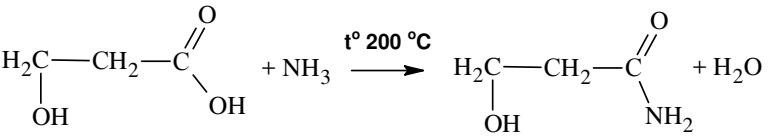
Sodium β-hydroxypropanoic acid

b) *esters (with alcohols):*



Methyl-β-hydroxypropanoate

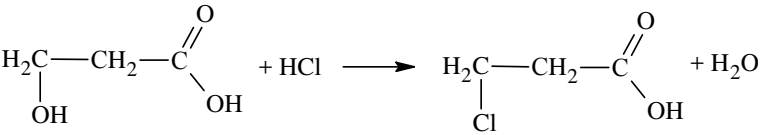
c) Amides (with ammonium, t°):



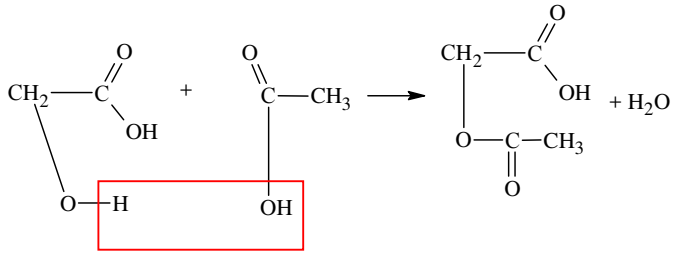
Amide of β-hydroxypropanoic acid

In the reactions **on –OH group** can be formed:

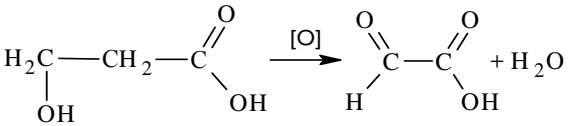
a) *halogens derivates* (*with hydrohalogens* (HCl, HBr, HI, HF)):



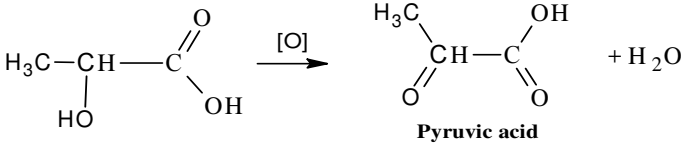
b) *esters* (as a result of the *interaction of the hydroxyl group of the hydroxyacid with carboxylic acid*):



c) *aldo- or ketoacids* (*as a result of oxidize*)

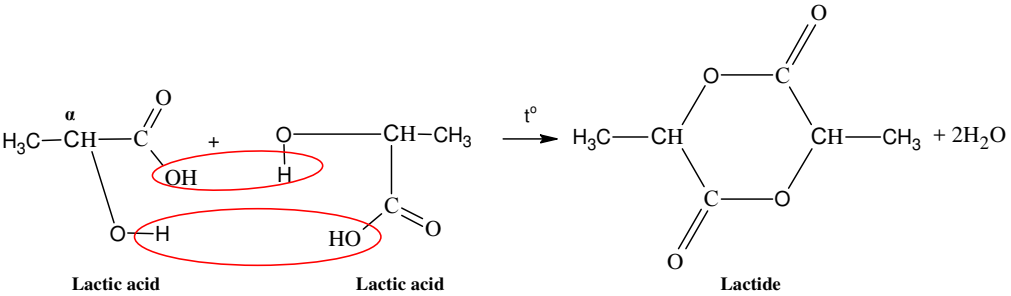


Glyoxalic acid



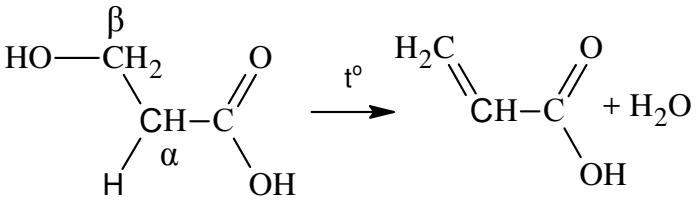
Pyruvic acid

Heating of α-hydroxyacids is accompanied by the release of water and the formation of *cyclic esters – lactides:*

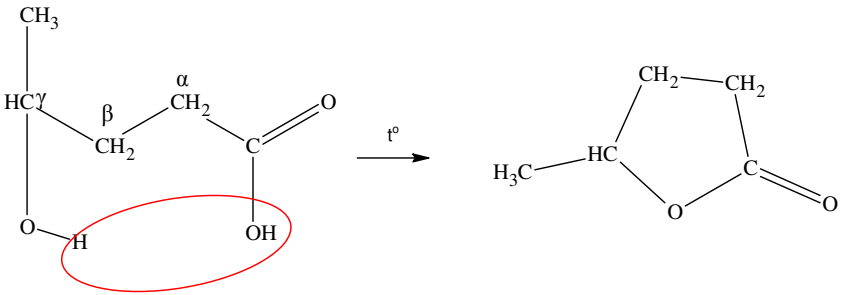


Lactic acid Lactic acid Lactide

When heated, β-hydroxyacids form *unsaturated acids* by the intramolecular release of the water molecule:



Elimination of water from one molecule of an acid and formation of *cyclic esters – lactones* is observed at the heating of γ-hydroxyacids:



Many hydroxyacids have optical activity, i.e. ability to rotate a plane polarized light. The concept of mirror images is the key to understanding the optical activity. All objects, including all molecules, have mirror images. The **mirror image** of an object is the object's reflection in a mirror - for example, human hands (Fig. 1.12).

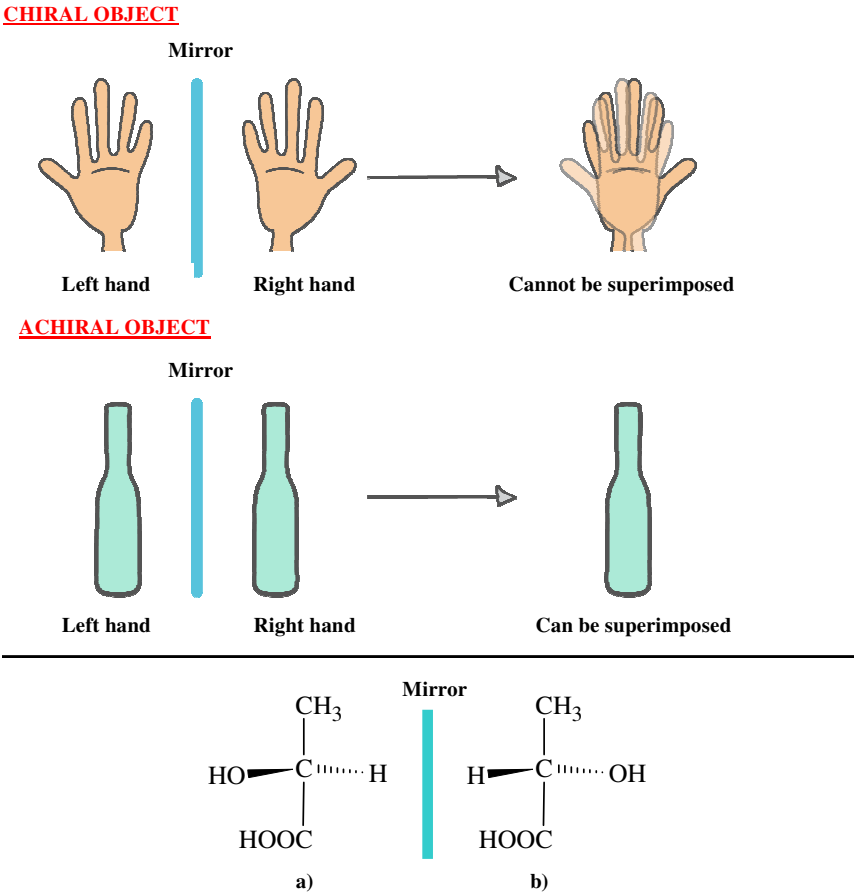


Fig. 1.12. The principle of mirror images

An optically active compound is a compound that rotates the plane of polarized light. Optical activity of a substance occurs if the molecule cannot be combined with its mirror image. Optical active molecules always have a chiral centre in their structure.

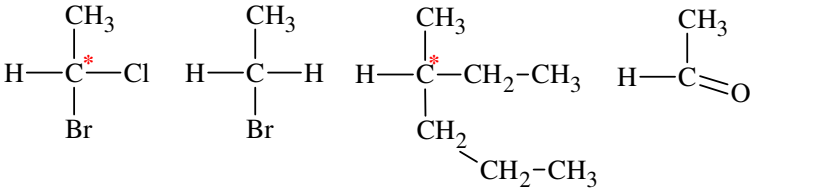
Chiral center or an **asymmetric carbon atom** is a carbon atom in molecule, which is connected with four different groups. **To find chiral carbon, you need:**

S t e p 1 . To number carbon atoms in molecule.

S t e p 2 . Research all carbon atoms in molecule to find carbon atoms, which are connected with four different groups.

The next carbon atoms cannot be chiral centres (Fig. 1.13)**:**

- Carbons in $-CH_3$ or $-CH_2$ groups
- Carbon, connected with other atom via double or triple bond (for example, with oxygen in ketone group).



CHIRAL (has 4 different atoms bonded to the carbon

ACHIRAL (does not have 4 different atoms or groups bonded to the carbon (has 2 hydrogen atoms)

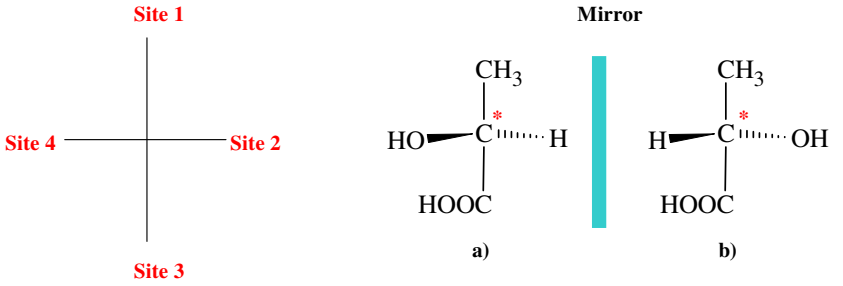
CHIRAL (has 4 different atoms bonded to the carbon

ACHIRAL (only has 3 atoms bonded to the carbon (has double bound with oxygen)

Fig. 1.13. The examples of chiral centres identification

Any molecule that has chiral center has optical activity. **Enantiomers** are stereoisomers whose molecules are mirror images of each other. They also are called **mirror** or **optical isomers.** In general, **a compound that has *n* chiral centers may exist in a maximum of 2^n stereoisomeric forms**. For example, when three chiral centers are present, at most eight stereoisomers ($2^3 = 8$) are possible (four pairs of enantiomers).

The lactic acid molecule has one chiral center, because of there are four different groups connected with this carbon atom:



- Site 1 (−CH_3)
- Site 2 (−H)
- Site 3 (−COOH)
- Site 4 (−OH)

Therefore, the lactic acid molecule has two (2^1) stereoisomers, called enantiomers.

Enantiomers have the same boiling point, melting point, density, formula, mass, connections. They differ by two characteristics: *their configuration* (they differ one from another, as an object and its image in a mirror) and *direction of rotation of plane polarized light* (right or left).

Optical isomers can have **D-** or **L-configurations.** Identification of configurations of optical isomers via D-, L-system is especially powerful in the chemistry of carbohydrates and amino acids. **Glyceraldehyde is used as the standard** for identification of optical isomers configuration**. D–isomer** was called a substance, which *had −OH group to the right from asymmetric carbon atom* (Fig. 1.14). Substances which formulas can be formed of D-glyceric aldehyde by the adjustment of a hydrocarbon chain from the side of aldehyde group belong to a D-series, and from L-glyceric aldehyde – to L-series.

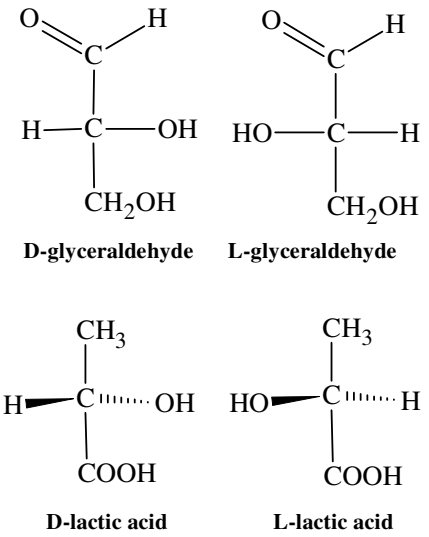


Fig. 1.14. D- and L-configurations of glyceraldehyde and lactic acid

Optical activity is a unique characteristic of enantiomers and can be measured by means of a *polarimeter* (Fig. 1.15)**.** Unpolarized monochromatic light from a suitable source pass through a polarizer (for example, pair of crossed Nicol prisms).

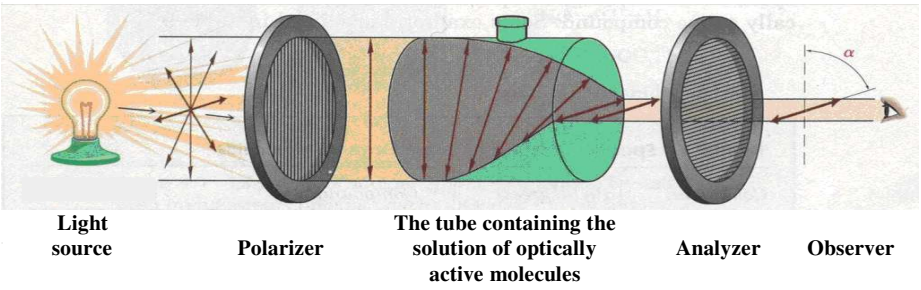


Fig. 1.15. The principle of a polarimeter working

The light leaving the polarizer is plane polarized and the angle of the Nichol prisms has been adjusted to produce vertically polarized

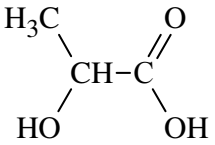
light. This light is then passed through a tube containing the solution of optically active molecules, and the plane of the polarized light will be rotated.

The light from the tube with solution then pass through analyzer (another pair of crossed Nichol prisms). The analyzer prisms are then rotated so that the transmitted light is again vertically polarized. The angle through which the analyzer has been turned is called the angle of polarization. *Clockwise rotation* of polarized light is designated as **(+)**; *anticlockwise rotation* as **(-).** Therefore, notation (+) means rotation to the right (clockwise), and (-) means rotation to the left (anticlockwise). Not in all cases substances of D-series rotate a surface of plane polarized light to the right, and substances of L-series – to the left. For example, D-enantiomer of glucose is (+)-glucose.

In general, **letters D** or **L** are used before the name of substance designation of a *configuration.* Also before name of the substance must be written signs **(+)** or **(-)** corresponding to *the rotation to the right or to the left.*

An equimolar mixture of two enantiomers is called a **racemic mixture**, or a **racemate.** Since a racemic mixture contains equal numbers of dextrorotating and levorotating molecules, the net optical rotation is zero - such mixture is optically inactive. A racemic mixture is often specified by prefixing the name of the compound with the symbol (±);

Hydroxyacids are widespread in the nature; also they are formed in body. **Lactic acid** (hydroxyacid with one OH-group and one COOH-group) firstly was extracted from the milk.

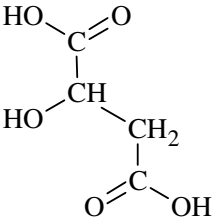


Lactic acid

Fig. 1.12. The lactic acid structure

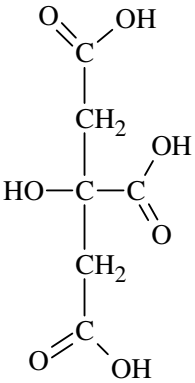
It is produced (in the form of its salt *lactate*) in skeletal muscles in process of the anaerobic breakdown of glucose (glycolysis); in hard qualities it is formed in muscles during hard and prolonged work, when the body's supply of oxygen is reduced. Salts of milk acid are used in medicine.

Malic and **citric acids** (hydroxyacids with one OH-group and two COOH-groups and one OH-group and three COOH-groups, respectively) are formed in a cycle of tricarboxylic acids (citric acid cycle, Krebs' cycle). Malic acid is found in many fruits, including apples:



Malic acid

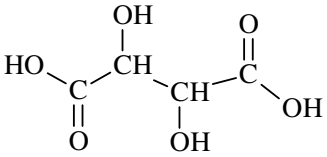
Citric acid is present in orange, lemon and other citric fruits:



Citric acid

These compounds are used in medicine for synthesis of some medical preparations.

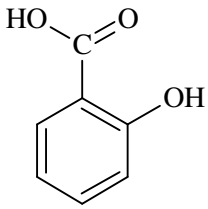
Tartaric acid occurs in grapes and is used in medicine for synthesis of some medical preparations:



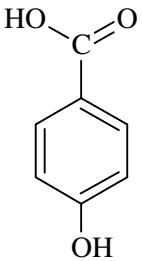
Tartaric acid

1.2.2. Phenolacids as an example of aromatic hydroxyacids

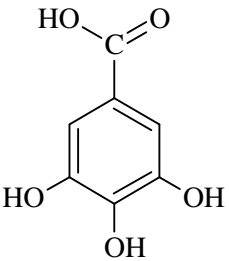
Phenolacids are the derivatives of aromatic carboxyl acids that contain –OH group (one or more) (Fig. 1.16).



Salicylic acid, 2-hydroxybenzoic acid



4-hydroxybenzoic acid



Gallic acid, 3,4,5-trihydroxybenzoic acid

Fig. 1.16. The structures of phenolacids representatives

Chemical properties of phenolacids are due to the presence in their structure *carboxyl group, phenolic hydroxyl* as well as *the aromatic nucleus*.

For example, salicylic acid is both a carboxylic acid and a phenol, so it can be esterified in two ways: in *methyl salicylate* (when COOH group of salicylic acid is esterified with methanol (CH_3OH) at

presence H_2SO_4), and in *acetylsalicylic acid (aspirin)* (in this case the phenolic −OH group of salicylic acid is esterified with acetic acid in reaction of salicylic acid with acetic acid anhydride or acetic acid chloride) (Fig. 1.17). Aspirin is an analgesic and has anti-inflammatory effect.

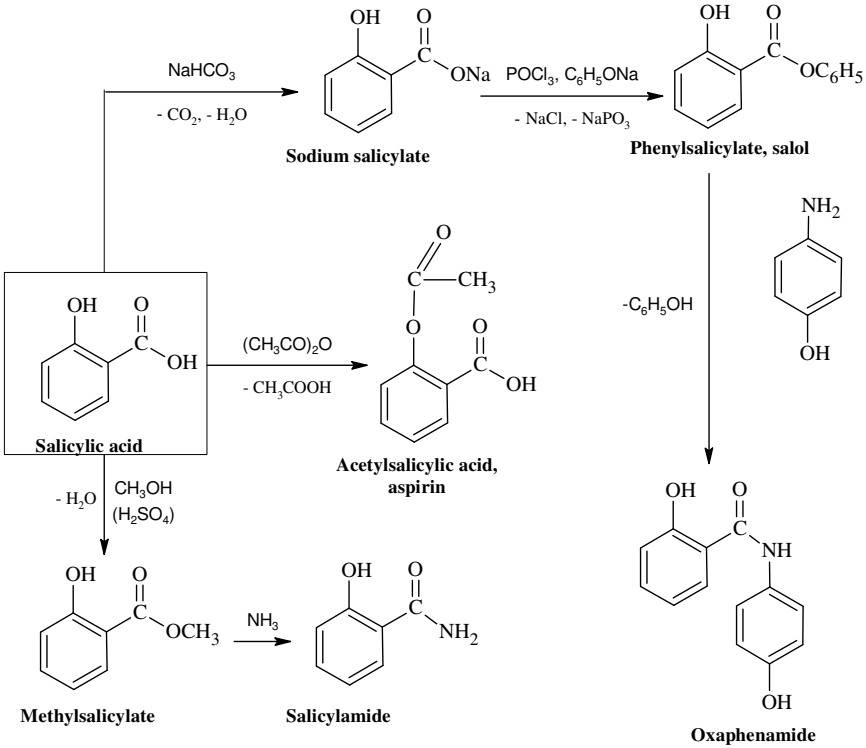
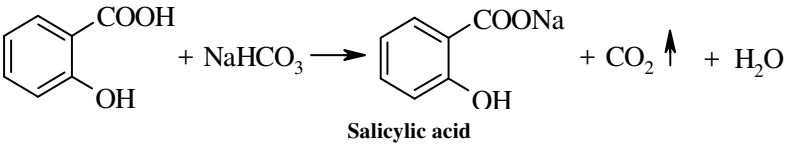


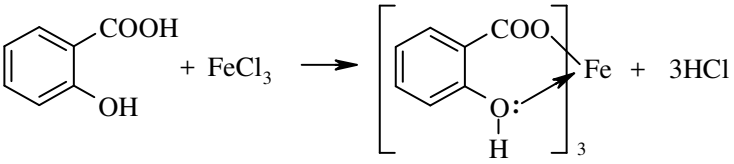
Fig. 1.17. The scheme of the some salicylic acid derivatives formation

Reactions **on carboxyl group** include**:**

a) With salts:

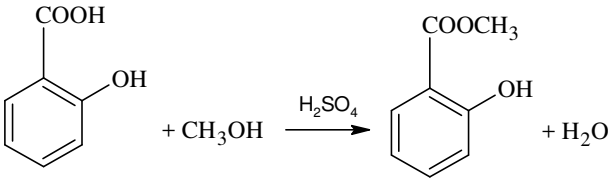


Salicylic acid

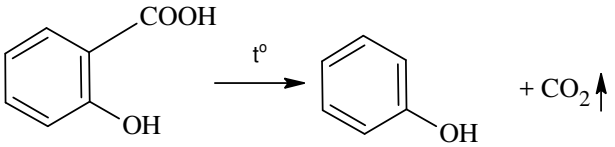


Complex helatic salt of salicylic acid (violet colour)

b) With alcohols:



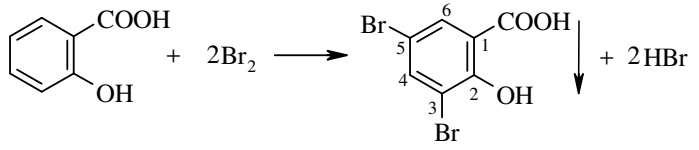
c) Decarboxylation:



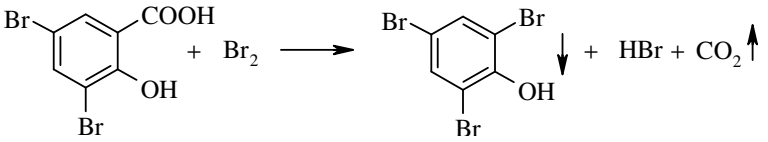
Salicylic acid **Phenol**

One of the main **reactions on phenolic hydroxyl** is *acetylation of the phenolic hydroxyl group* that was given above (Fig. 1.17) – it is used to prepare aspirin.

The example of **reactions on the aromatic nucleus** is the halogenic derivatives formation**:**



3,5-dibromsalicylic acid (white precipitate)



Yellow precipitate

Salicylic acid is widely widespread in the nature and together with its derivatives enters into the composition of many drugs. Derivatives of salicylic acid have anesthetizing, febrifugal and anti-inflammatory effects. Anti-inflammatory activity is caused by the presence of −OH group: all derivatives of phenol which contain this group posse a bactericidal action.

Methylsalicylate is a liquid without color with a fragrant smell. It shows irritating, anesthetizing and anti-inflammatory action, it is used for massages and compresses (with vegetable oil at rheumatoid arthritis). **Phenylsalicylate** (**salol**) is an ester of a salicylic acid and phenol. **Aspirin** is a white crystalline powder with a slight smell and faintly acid taste; it is slightly soluble in water possessing antipyretic and analgesic properties. It is used in the form of powders ant tablets. Aspirin is often combined with others analgesics (ascophene, citramonum, etc.).

Gallic acid is found in tea, as well as in other plants, and it also occurs as part of a larger molecule, called tannin, which is present in galls (such as the swellings of the tissue of trees caused by the attack of wasps).

Three of the other most important aromatic dicarboxylic acids are called **phthalic**, **isophthalic**, and **terephthalic** acids, for the *ortho, meta,* and *para* isomers, respectively.

1.2.3. Oxoacids

Oxoacids include **aldehydo-** and **ketoacids**. These compounds have in their structures both the carboxyl group and aldehyde functional group or ketone functional group. The structures of some ketoacids are given on Fig. 1.18. *Pyruvic acid* and *acetoacetic acid* are the simplest and the most important of the α-keto and β-keto acids, respectively.

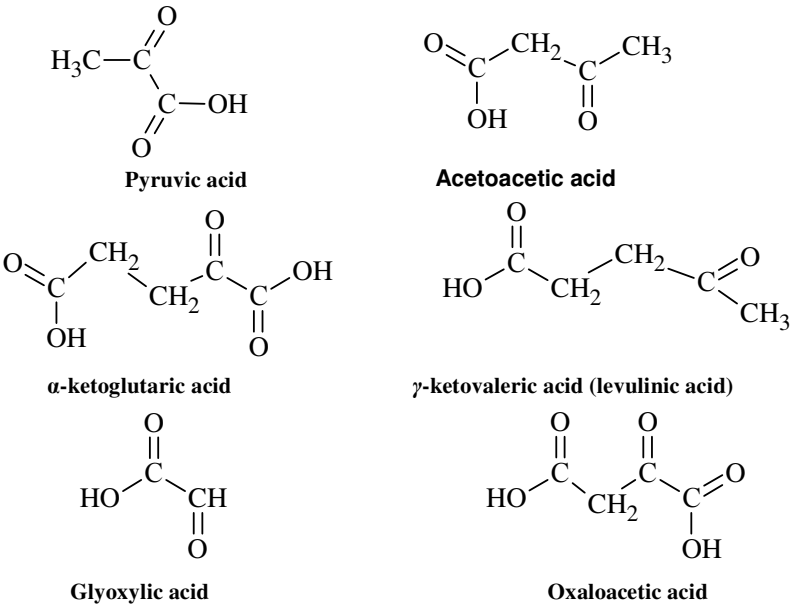
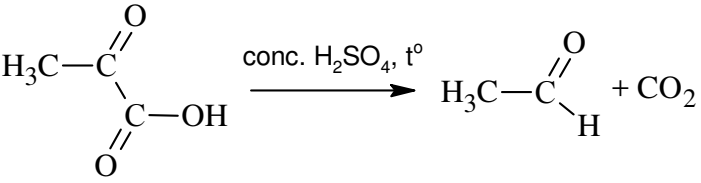


Fig. 1.18. The structures of some ketoacids

Aldo- and ketoacids have the properties of carboxylic acids and aldehydes or ketones. They form salts, esters, etc. on carboxyl group; on carbonyl group they enter into the addition reactions, forming oxide, cyanhydride, etc.

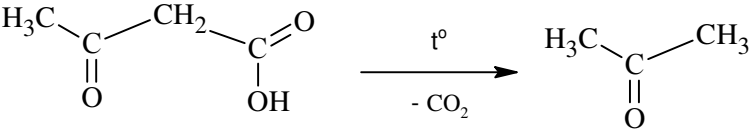
Reactions on carboxyl group include**:**

a) Decarboxylation of α-oxoacids:



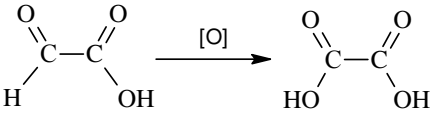
Pyruvic acid **Acetaldehyde**

b) Decarboxylation of β-oxoacids:

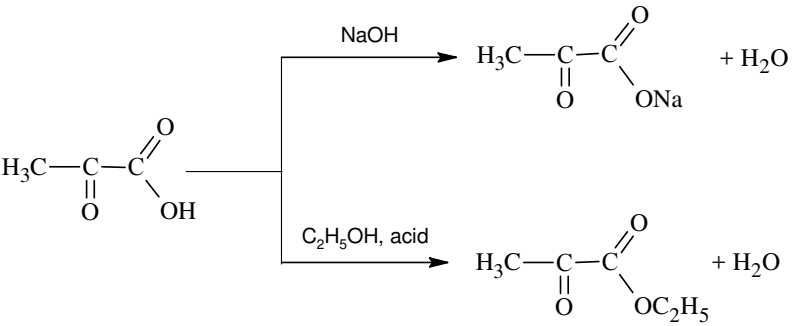


Acetoacetic acid **Acetone**

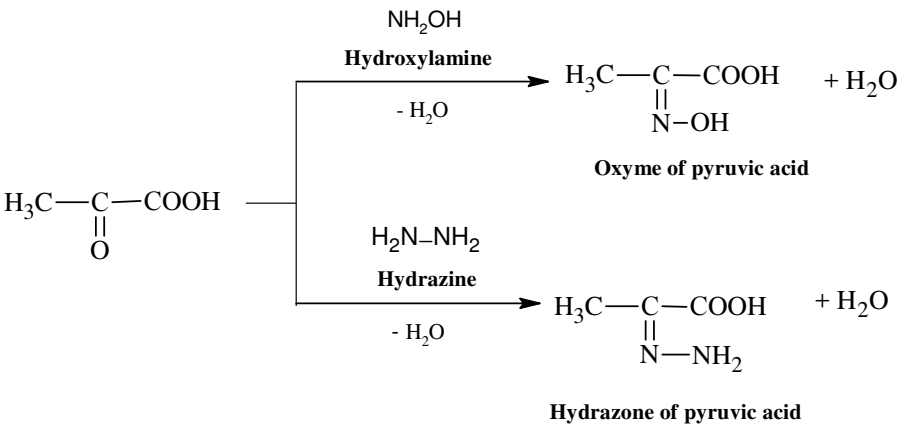
c) Aldoacids are easily oxidized with formation of the dibasic acids:



d) Forming the salts with bases and esters with alcohols:



The examples of **reactions on carbonyl group** are *the reactions with compounds like NH_2–X* (hydroxilamine NH_2–OH; hydrazine NH_2–NH_2) *with oxymes and hydrazones formation*:



Pyruvic acid (in the form of its salt pyruvate) is involved in the normal metabolism of carbohydrates as the final product of glycolysis – the process that consists of 10 (aerobic) or 11 (anaerobic) reactions starting from glucose. After formation it is converted (by loss of carbon dioxide) to acetyl coenzyme A, which enters the tricarboxylic acid cycle. Pyruvate is also used by the body to synthesize alanine, an amino acid required for the synthesis of proteins.

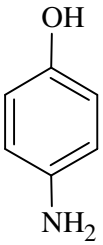
The acetoacetic acid is unstable and loses carbon dioxide to give *acetone*:

$$CH_3COCH_2COOH \rightarrow CH_3COCH_3 + CO_2$$

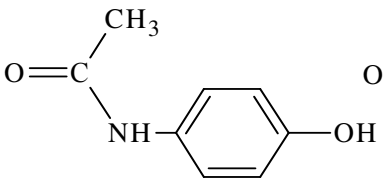
In severe diabetes the body converts acetyl coenzyme A to acetoacetic acid and acetone — excess quantities of which occur in the blood and are secreted in the urine. These two compounds, along with *β-hydroxybutyric acid* (in which the acids are in the form of their salts), are collectively called **ketone bodies**, although the third of these is not a ketone; they are used to diagnose diabetes.

1.2.4. Heterofunctional compounds of benzene series

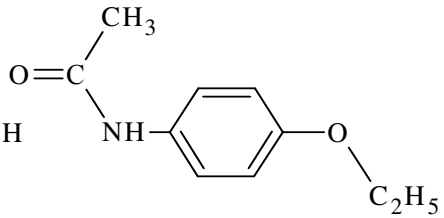
A lot of heterofunctional compounds of benzene series belong to the drugs and physiologically active substances: non-narcotic analgesics, local anesthetics, antiseptic and anti-parasitic medications, drugs for tuberculosis treatment. One of the representatives of this group of compounds is ***p*-aminophenol** which contains hydroxyl group and amino group in *para* position and therefore exhibits properties of phenols and amines:



The main derivatives of *p*-aminophenol are the **paracetamol** and **phenacetin**:



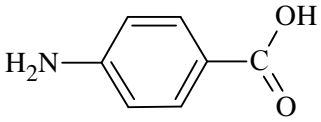
Paracetamol



Phenacetin

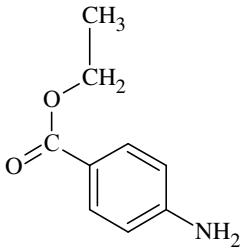
They are non-narcotic analgesics which have anesthetizing and antipyretic actions. Paracetamol is a product of acylation of *p*-aminophenol with acyl halides of acetic acid, whereas phenacetin is a derivative of *p*-aminophenol, in which amino group is acetylated and hydroxyl group is esterified with ethanol.

***p*-aminobenzoic acid (PABA)** is another substance that belong to heterofunctional compounds of benzene series:

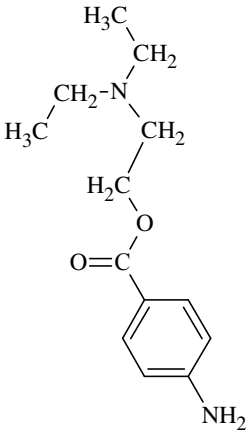


***p*-aminobenzoic acid (PABA)**

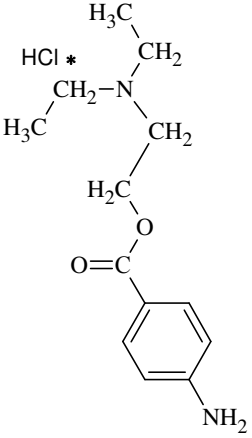
It occurs in the yeast, liver, eggs, milk, etc. This compound is the factor of growth of microorganisms and takes part in many biochemical processes. Deficiency of *p*-aminobenzoic acid in food leads to graying and loss of hair. Derivatives of *p*-aminobenzoic acid (*benzocaine (anesthesin) and novocaine (procaine)*) are local anesthetics (Fig. 1.19).



Benzocaine (anesthesin)



Novocaine (procaine)



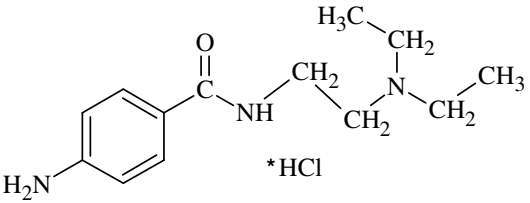
Novocaine hydrochloride

Fig. 1.19. Local anesthetics - derivatives of *p*-aminobenzoic acid

Anesthesin – ethyl ether of PABA – does not form stable water-soluble salts with acids, and therefore it is used in the form of powders, ointments, pastes at burns, ulcers on skin. One of the general local anesthetics is **novocaine** (also called **procaine**) - 2-

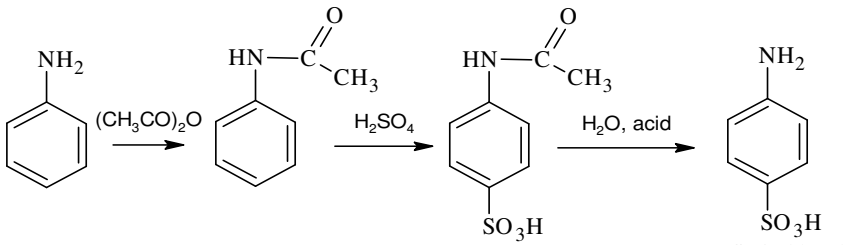
diethylaminoethyl ether of PABA. Novocaine is used as a salt - **novocaine hydrochloride** – this modification enhances its solubility in water. Novocaine is hydrolyzed quickly in the organism with formation of *aminobenzoic acid*. Novocaine anaesthesia lasts near 30 minutes.

Introduction of amide group to the molecule of novocaine changes its physiological action. **Procainamide hydrochloride** is an amide of *p*-aminobenzoic acid, it's an antidysrhythmic drug and used for treatment for heart rhythm disorders:



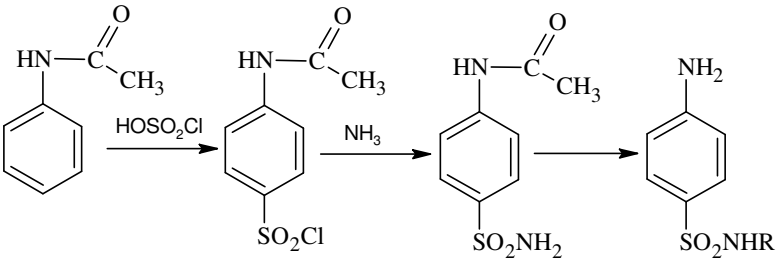
Procainamide hydrochloride

Sulfanilic acid and **sulfanilamides** also belong to heterofunctional compounds of benzene series. Sulfonic acid of *p*-aniline (*sulfanilic acid*) is obtained as a result of aniline sulfonation. In this case amino group should be protected preliminarily, for example, using acylation:

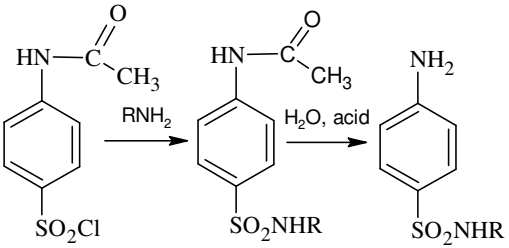


Acetanilide **Sulfanilic acid**

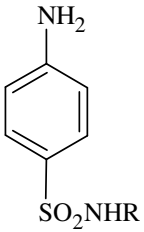
If sulfonation is performed without sulfuric acid but using chlorosulfonic acid, after the interaction with ammonia, amide of sulfanilic acid - *sulfanilamide*, or **streptocide,** that has bactericidal action – will be formed:



If amines are used instead of ammonia substituted, **sulfanilamides (sulfa drugs)** can be obtained:



In this case, antibacterial action and toxicity of generated sulfanilamides depend on the nature of radical (R), connected with amide nitrogen:



There are more than 5000 synthesized derivatives of sulfanilamide, however near 30 preparations have found practical

use; almost all of them contain heterocyclic compounds in radical. Sulfa drugs are broad-spectrum drugs - they are effective against many types of bacteria.

The mechanism of sulfanilamides action is as follows. PABA is needed in enzymatic reactions that produce tetrahydrofolic acid, which, in turn, acts as a coenzyme in the synthesis of purines and pyrimidines. Mammals don't synthesize their own folic acid so are unaffected by PABA inhibitors, which selectively kill bacteria. They, in contrast to bacteria, acquire folate through the diet.

Since tetrahydrofolate is important for pyrimidines and so for nucleic acid synthesis, drugs that inhibit its synthesis are bacteriostatic because the bacteria can't reproduce if they don't have enough tetrahydrofolic acid to make new DNA, RNA and proteins.

Sulfa drugs inhibit the enzyme that catalyzes the first step of this pathway - *dihydropteroate synthase* (DHPS) (Fig. 1.20). The real substrate of this enzyme is the PABA. Acting as a substrate analogue, Sulfa drugs are **competitive inhibitors** of this enzyme (Fig. 1.21).

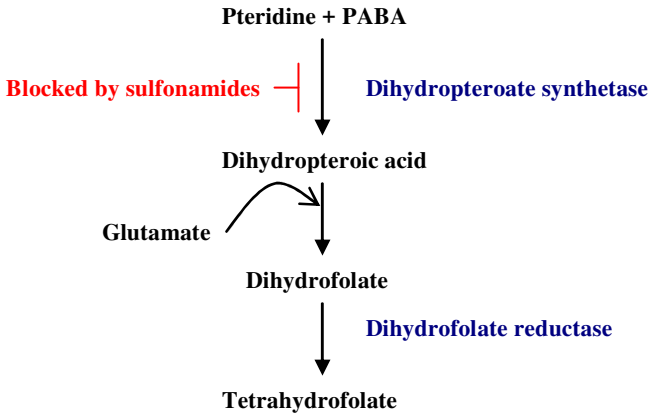


Fig. 1.20. Scheme of tetrahydrofolate biosynthesis and Sulfa drug action

Sulfonamides mimic PABA which is the normal substrate for dihydropteroate synthetase. This means that sulfonamide will bind in

the same manner as PABA. This binding is reversible, because of that sulfonamides have bacteriostatic effect but not bactericidal.

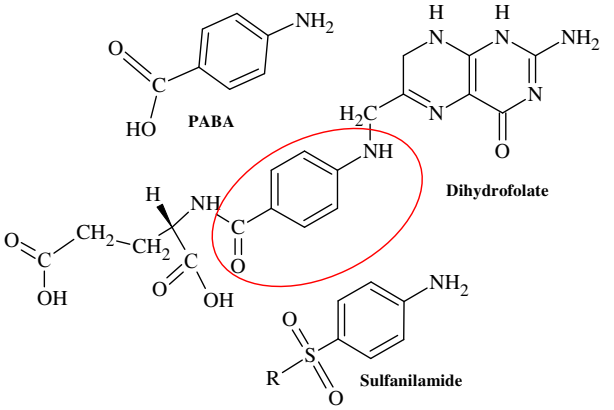


Fig. 1.21. Structural similarity of PABA, Sulfa drugs and dihydrofolate - intermediate of tetrahydrofolic acid biosynthesis

Because sulfonamides are competitive inhibitors for the enzyme, the bacteria can increase the production of PABA to compete with sulfonamide at the active site and become resistant to sulfa drugs. In such case, the dose of sulfonamide agents should be increased to overcome this resistant mechanism. But this high dose is accompanied with an increase in side effects especially the crystalluria.

1.3. Test questions

1. Choose from the names of the carboxylic acids the name that is trivial:

A. propanoic acid
B. 4-aminobutanoic acid
C. 2-methylbutanoic acid
D. formic acid
E. ethanoic acid

2. Which of these carboxylic acids is aromatic?

A. formic acid
B. butanoic acid
C. oleic acid
D. arachidonic acid
E. benzoic acid

3. Which of these carboxylic acids is unsaturated?
A. formic acid
B. butanoic acid
C. oleic acid
D. valeric acid
E. benzoic acid

4. Which of these carboxylic acids is dicarboxylic?
A. acetic acid
B. formic acid
C. butanoic acid
D. oxalic acid
E. oleic acid

5. Which carboxylic acid attaches to a large biochemical molecule called coenzyme A (HS−CoA) and in this form enters to the tricarboxylic acid cycle (or Krebs's cycle)?
A. acetic acid
B. formic acid
C. butanoic acid
D. oxalic acid
E. oleic acid

6. Which of these fatty acids is essential?
A. oleic acid
B. stearic
C. palmitic
D. linolenic
E. myristic

7. Which of these fatty acids is a source to prostaglandins and leukotrienes synthesis?

A. oleic acid
B. stearic
C. palmitic
D. arachidonic
E. myristic

8. Hippuric acid is the derivate of:

A. acetic acid
B. butanoic acid
C. oleic acid
D. arachidonic acid
E. benzoic acid

9. Barbituric acid and barbiturates are the derivates of:

A. hippuric acid
B. oxalic acid
C. malonic acid
D. arachidonic acid
E. benzoic acid

10. Which of these compouns belong to heterofunctional compounds?

A. hippuric acid
B. oxalic acid
C. malonic acid
D. lactic acid
E. benzoic acid

PART 2. FATTY ACIDS AND LIPIDS

The lipids are biological substances with various structures that are insoluble in water and soluble in non-polar organic solvents such as chloroform, benzene and ether. They are hydrophobic in nature because of the high amount of hydrocarbons in their structure.

Lipids are the structural components of cell membranes and therefore they take part in the regulation of cell membranes permeability, are involved in the creation of intercellular contacts and participate in the nerve impulse transmittion. They are also energy storage and can protect the internal organs. Moreover, since fat is a bad conductor of heat, it provides excellent insulation. Besides lipids are important precursors of a number of biologically active substances.

2.1. Lipids classification

Lipids can be divided into simple and complex lipids as well as lipids derivatives.

Simple lipids are classified into:

- **neutral fats**, or **acylglycerols** – they are represented by fats and oils that are esters of fatty acids and glycerol;
- **waxes** – they are esters of fatty acids and higher molecular weight monohydric alcohols;
- **steroids,** that contain the cyclopentaneperhydrofenanthrene nucleus (also called steroid nucleus).

Complex lipids have additional components:

- **phospholipids** contain *phosphoric acid residue*:
  - glycerophospholipids;
  - sphingophospholipides;
- **glycolipids** contain *hexoses, sulfated hexoses, hexosamines, sialic acids:*
  - glyceroglycolipids (not in mammalians but in plants);

- glycosphingolipids (cerebrosides, globosides, gangliosides, sulphatides);

- **lipoproteins** that have a protein part.

Lipid Derivatives are the next:

- **fatty acids;**
- derivatives of polyunsaturated arachidonic acid – **eicosanoisds** (*prostaglandins, thromboxanes* and *leukotrienes*);
- **derivatives of cholesterol** - *steroid hormones, bile acids, vitamin D3*;
- **izoprenoids** (*fat soluble vitamins - vitamin A, vitamin K, vitamin E, terpenes*).

Lipids are also divided into saponifiable and non-saponifiable. *Saponification* is the base hydrolysis of ester bonds.

Saponifiable lipids are lipids that have an ester functional group and therefore can be hydrolyzed under basic conditions. The hydrolyzed products always include fatty acids (or salts of fatty acids), and in some cases glycerol and other molecular components contained in the lipid. The examples of saponifiable lipids are:

- neutral fats (acylglycerols),
- phospholipids,
- glycolipids,
- waxes.

Non-saponifiable lipids are those lipids that cannot be hydrolyzed to fatty acids, for example:

- steroids
- terpenes (have several five-carbon building blocks called *isoprene units* - for example, fat-soluble vitamins A, E, K) (Fig. 2.1)

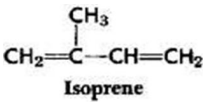


Fig. 2.1. Structure of isoprene - five-carbon building blocks of terpenes and some other compounds

2.2. Fatty acids as main lipids components

Fatty acids are derived from saponifiable lipids. They usually have non-branching chain and even number of carbon atoms; the length of the aliphatic chain is usually from between 10 and 20 carbons. They can be saturated (alkanes) and unsaturated (alkenes). In most unsaturated fatty acids, the *cis* isomer predominates; the *trans*-isomer is rare.

There are three groups of fatty acids (Fig. 1.2):

- **saturated**: butyric (C4); capronic (C6); lauric (C12); palmitic (C16); stearic (C18) and others;
- **monounsaturated** (or **monoenic**): crotonic (C4), palmitooleic (C16), oleic (C18), and others;
- **polyunsaturated** (or **polyenes**): linoleic (C18, two double bonds), linolenic (C18, three double bonds), arachidonic (C20, four double bonds).

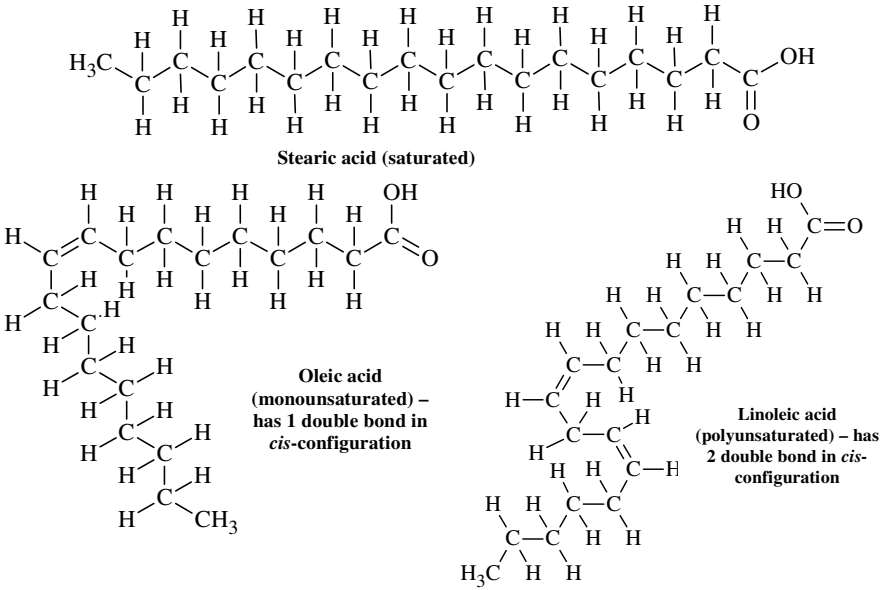


Fig. 2.2. Saturated, monounsaturated and polyunsaturated fatty acids

2.2.1. Nomenclature of fatty acids

Fatty acids are either called by their *systematic names* according to the IUPAC nomenclature or by their *trivial (common) names*. Actually trivial names are used much more frequently than systematic names. Trivial names are non-systematic historical names that seldom convey structural information and often reflecting a common or early source of the acid. They are the most frequent naming system used in literature. For example (first – systematic names; written in italics in parentheses – trivial (common) names):

- hexadecanoic acid (*palmitic acid*),
- octadecenoic acid (*oleic acid*),
- octadecatrienoic acid (*linolenic acid*)

Another way of naming fatty acids would be a *shorthand nomenclature* (it will be described later).

Numbering of the carbon atoms in fatty acids. The position of the carbon atoms in a fatty acid can be indicated (Fig. 2.3):

- **from the –COOH** (or carboxyl) **end**,
- **from the –CH_3** (or methyl) **end**.

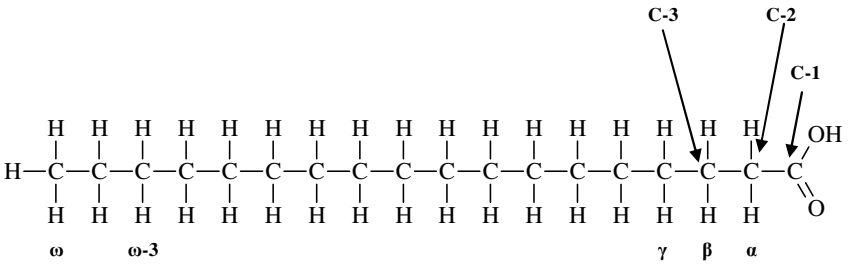


Fig. 2.3. The principle of numbering of the carbon atoms in fatty acids

In systematic and common names the first way is used. When the fatty acid is indicated from the -COOH end, then the C-1, C-2, C-3, etc. notation is used. In addition in common names Greek letters or symbols (α, β, γ, δ, ε, etc.) are used to number Carbon atoms in fatty acid molecules (Fig. 2.3, Tab. 2.1). **Carboxyl Carbon** in a fatty acid molecule is always considered as the First Carbon (C-1), it has no

corresponding Greek letter or symbol in Common name. **Methyl carbon** is always named **ω-Carbon**, or **n-carbon.**

Table 2.1.

The principle of numeration of carbon atoms in fatty acids molecules

Systematic name	**Common name**
The first Carbon atom **(C-1, Carboxyl carbon)** in Fatty Acid molecule	**-**
Second Carbon atom **(C-2)** in Fatty Acid molecule	α-Carbon atom
Third Carbon atom (C-3) in Fatty Acid molecule	β-Carbon atom
…..	……
Last or Terminal Carbon atom in a fatty acid molecule	**ω-Carbon** or the **n-carbon atom**

Systematic names of fatty acids are formed:

- from **alkane** name with the same number of carbon atoms by replacing the ending **-e** with suffix **-oic acid** *(for saturated fatty acids);*
- from **alkene** name with the same number of carbon atoms by replacing the ending **-e** with suffix **-enoic acid** *(for monounsaturated fatty acids).*

For example, systematic and trivial names of some saturated fatty acids are the next:

- 16:0 – hexadecan**oic acid** (*palmitic acid* **-** from latin palm tree) - from Hexadecan**e**;
- 18:0 – octadecan**oic acid** (*stearic acid*) - from octadecan**e;**
- 20:1 - eicosan**oic acid** (*arachidic acid* **-** from Greek arachne, spider) **–** from eicosan**e.**

Systematic and trivial names of monounsaturated fatty acids:

- 16:1 – hexadecen**enoic acid** (*palmitooleic acid*) – from hexadecane;
- 18:1 – octadec**enoic acid** (*oleic acid*) – from octadecen**e**.

Examples of systematic and trivial names of polyunsaturated fatty acids:

- C18:2 – octadecadi**enoic acid** (*linoleic acid*) – from octadecadien**e**;
- C18:3 – octadecatr**ienoic acid** (*linolenic acid*) – from octadecatrien**e**;
- C20:4 – eicosatetra**enoic acid** (*arachidonic acid*) – from eicosatetraen**e**.

In **shorthand nomenclature** fatty acids are named by their number of carbon atoms, and their number of double bonds after a colon, e.g. 16:0 (Fig. 2.4).

– **16:0 (palmitic)** – **18: 1 (oleic)** – **20: 3 (linolenic)**	**The first number refers to the total number of carbon atoms, and the second one – to the number of double bonds**

Fig. 2.4. The principle of writing the name of a fatty acid via shorthand nomenclature

For unsaturated fatty acids the shorthand nomenclature can also contain the positioning of the double bonds – it can be pointed either **from COOH-end** of fatty acid molecule (*Δx (or delta-x) nomenclature*), or **from CH_3-end** (*ω−x, omega-x* or *n−x nomenclature*) (Fig. 2.3)

In the first pathway **- Δ*x*-nomenclature** (the "Δ" is the Greek letter "delta", which translates into "D" (for Double bond) in the Roman alphabet) - **the location of the double bond is counting from the -COOH end** (**Carboxyl Carbon** is also taken into account!!!) and is denoted by the symbol Δ with the number (Fig. 2.5).

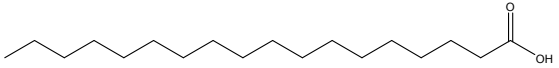
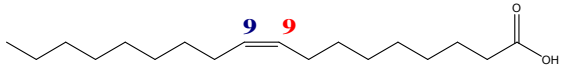
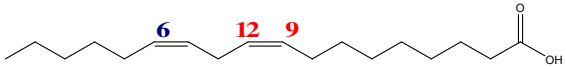
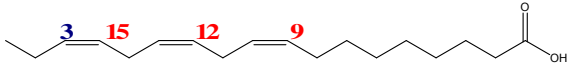
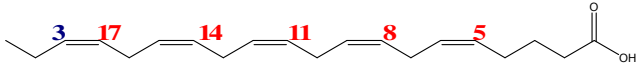
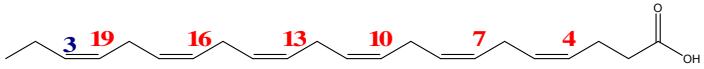
Trivial name and ω–*x*- nomendature	Methyl end — Carboxyl end	Saturation	Δ*x*- nomendature
Stearic acid, 18:0		Saturated	18:0
Oleic acid, 18:1, ω-9		Monoenic	18:1 Δ^{9}
Linoleic acid, 18:2, ω-6		Polyenic	18:2 $\Delta^{9,12}$
Linolenic acid, 18:3, ω-3		Polyenic	18:3 $\Delta^{9,12,15}$
Eicosapentaenoic acid, 20:5, ω-3		Polyenic	20:5 $\Delta^{5,8,11,14,17}$
Docosahexaenoic acid, 22:6, ω-3		Polyenic	22:6 $\Delta^{4,7,10,13,16,19}$

Fig. 2.5. The principle of fatty acid naming via ω−x- and Δx-shorthand nomenclature

For example (first - Δ*x*-nomenclature; second – systematic name; written in italics in parentheses – trivial name):

- $C_{18:2,}\Delta^{9,12}$ – octadecadienoic acid (*linoleic acid*) - this means that the first double bond is contained between 9 and 10 carbon atoms (the COOH group is also taken into account), and the second one is between 12 and 13;
- $C_{18:3,}\Delta^{9,12,15}$ – octadecatrienoic acid (*linolenic acid*);
- $C_{20:4,}\Delta^{5,8,11,14}$ – eicosatetraenoic acid (*arachidonic acid*).

Another pathway **- ω−*x* nomenclature -** is used to indicate the position of a double bond by counting **from the $-CH_3$-end (ω-Carbon)** ((Fig. 2.5). Omega (ω) is the last letter in the Greek alphabet, and is therefore used to indicate the "last" carbon atom in the fatty acid chain. For example (first - ω−*x*-nomenclature; second written in italics – trivial name):

- 18:3ω3 – *linolenic acid* - this means that the first double bond is contained between 3 and 4 carbon atoms, if we count from the $-CH_3$ group (this group is also taken into account);
- 18:1ω9 - *oleic acid* – it contains a double bond between C9 and C10 counting from the ω-carbon atom (i.e., from the last Carbon atom in the fatty acid molecule);
- 18:2ω6 - *linoleic acid* – it contains the first double bond between C6 and C7 counting from the ω-carbon atom.

In Omega numbering system the position of the second double bond is not indicated; it is because of double bonds in polyunsaturated fatty acids ARE NOT in the conjugated form such as:

– CH = CH – CH = CH –

Incorrect

They are separated by two single bonds, not by just one single bond:

– CH = CH – CH_2 – CH = CH – CH_2 – CH = CH –

Correct

Thus, in linoleic acid (C18:2ω6) the second double bond will be between C9 and C10 from the ω-carbon.

According to ω−*x*–nomenclature, unsaturated fatty acids are divided into groups, the most important of which are next:

- **ω9 series –** there are fatty acids in which the first double bond is between C9 and C10 counting from the ω-carbon (for ex., *oleic acid*);
- **ω6 series -** these fatty acids have the first double bond between C6 and C7 counting from the ω-carbon (for ex., *linoleic acid* and *arachidonic acid*);
- **ω3 series -** fatty acids in which the first double bond is between C3 and C4 counting from the ω-carbon (example **-** *linolenic acid*).

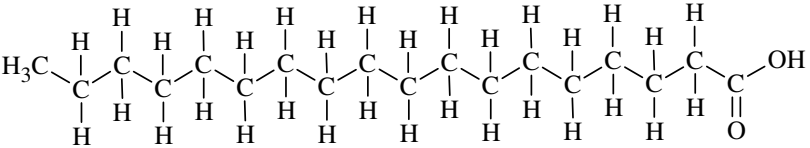
The unsaturated fatty acids that cannot be biosynthesized in tissues of some animals including humans (because of luck the desaturases – enzymes needed for some double bonds formation) are called **essential fatty acids**. Thus they must be obtained in the diet. The most of the essential fatty acids are members of the ω6 and ω3 series – for example, linoleic and linolenic acids. Some animals including humans can synthesize arachidonic acid from linoleic acid obtained in the diet.

2.2.2. Some properties of fatty acids

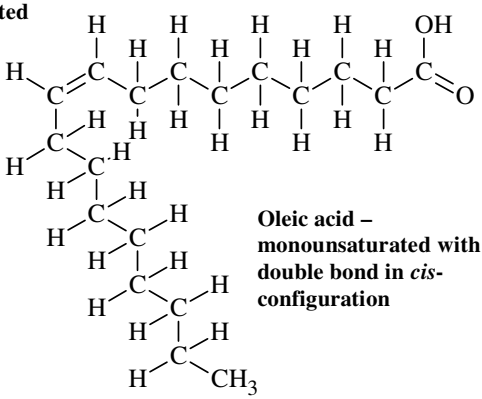
Fatty acids are **amphipathic**, because of the hydrophobic tail and hydrophilic (–COOH) head. The longer the hydrocarbon chain of the fatty acid, the higher its melting point, and the greater the number of double bonds in the fatty acid - the lower its melting point. Therefore unsaturated fatty acids have substantially lower melting points than saturated ones, and the greater the degree of unsaturation, the lower the melting point (Fig. 2.6). This is due to hydrocarbon chains of saturated fatty acids can lie parallel with strong dispersion forces between their chains (Fig. 2.7); they pack into well-ordered, compact crystalline forms and melt above room temperature. In contrast, because of the *cis*-configuration of the double bonds in unsaturated fatty acids, their hydrocarbon chains have a less ordered structure and dispersion forces between them are weaker; these triglycerides have melting points below room temperature

	Carbon Atoms: Double Bonds	Common Name	Melting Point (°C)
	Saturated Fatty Acids		
Higher mp	**12:0**	**Lauric acid**	**44**
	14:0	**Myristic acid**	**58**
	16:0	**Palmitic acid**	**63**
	18:0	**Stearic acid**	**70**
	20:0	**Arachidic acid**	**77**
	Unsaturated Fatty Acids		
Lower mp	**16:1**	**Palmitoleic acid**	**1**
	18:1	**Oleic acid**	**16**
	18:2	**Linoleic acid**	**-5**
	18:3	**Linolenic acid**	**-11**
	20:4	**Arachidonic acid**	**-49**

Fig. 2.6. Some fatty acids and their melting points (MP)



Stearic acid – saturated



Oleic acid – monounsaturated with one double bond in *cis*-configuration

Fig. 2.7. The peculiarities of saturated and unsaturated fatty acids hydrocarbon chains structure

Salts of fatty acids are called **soaps** and form micelles in aqueous (water is solvent) solutions. These spherical structures allow the charged head-groups to be in contact with water while at the same time removing the non-polar tails from contact with water. More details about these formations will be given later.

2.3. The neutral fats

Neutral fats are represenrated by glycerides, more correctly known as **acylglycerols.** These compounds are esters formed from glycerol and fatty acids. Glycerol, which is also called glycerin, is alcohol that has three hydroxyl functional groups, which can be esterified with one, two, or three fatty acids to form monoacylglycerols (MG), diacylglycerols (DG) or triacylglycerols (TG), respectively. Schematic representation of these compounds is given on Fig. 2.8.

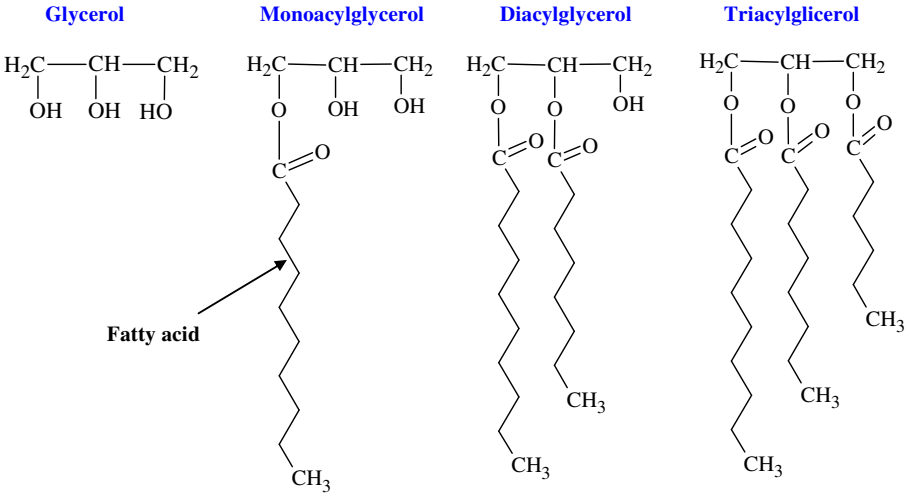


Fig. 2.8. The structures of glycerol, monoacylglycerol, diacylglycerol and triacylglycerol

The fatty acids can react with the hydroxyl groups to form esters. For triglycerides, all three hydroxyls of the glycerol have a fatty acid residue attached to it (Fig. 2.9).

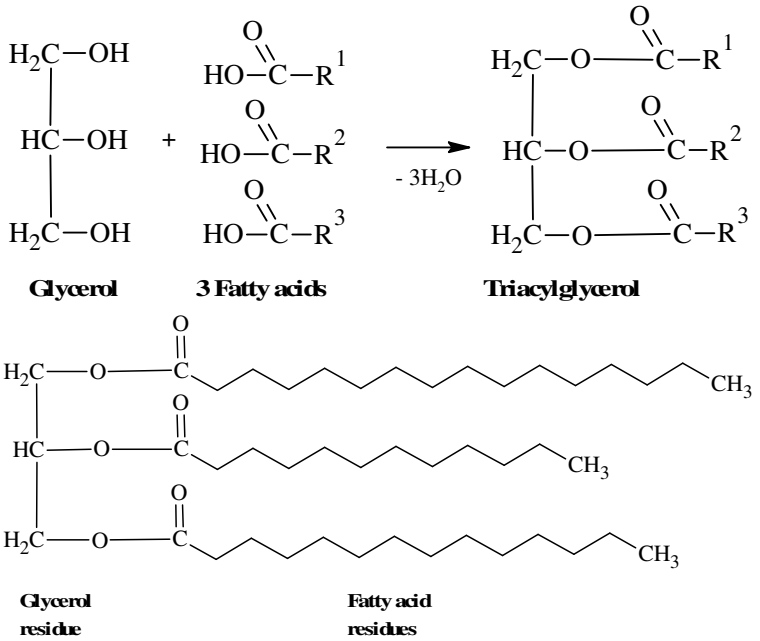


Fig. 2.9. Structure of triacylglycerol and scheme of its formation

Neutral fats are neutral because they are uncharged and do not contain acidic or basic groups. They are nonpolar and hydrophobic.

2.3.1. The triacylglycerols classification, nomenclature, localization and functions.

The triacylglycerols can be divided into simple and mixed ones.

The simple triacylglycerols contain identical fatty acid residues. For example, *triolein* (or *tri-oleoyl-glycerol*) contains three molecules of oleic acid residues esterified to a molecule of glycerol,

and *tristearin* (or *tri-stearoyl-glycerol*) contains three stearic acid residues esterified to a molecule of glycerol.

Mixed triacylglycerols contain two or three different types of fatty acid residues esterified to a molecule of glycerol. They are named according to the placement of the fatty acid residues on the glycerol molecule, for example *1-palmitoleoyl-2-linoleoyl-3-stearoyl-glycerol; 1,3-dipalmitoleoylstearoyl-glycerol.*

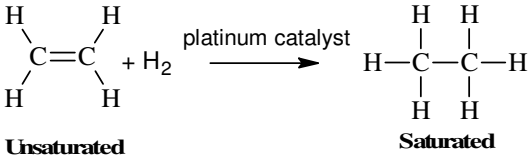
In the blood triacylglycerols exist as a part of lipoproteins. In the body they are mainly localized in the fat cells – *adipocytes* – in the *adipose tissue*, where they serve as a form of stored energy. Under condition called **obesity** triacylglycerols can also be found as the structural components of other cells – for example, hepatocytes or tubular cells of kidney**.**

2.3.2. The physical and chemical properties of triacylglycerols

Physical properties of triacylglycerols depend on the fatty acid components. Their melting point increases as the number of carbons in its hydrocarbon chains increases and as the number of double bonds decreases. Triglycerides rich in unsaturated fatty acids are generally liquid at room and form **oils**; triglycerides from plants tend to have a higher proportion of unsaturated fatty acids. Triglycerides rich in saturated fatty acids are generally solids at room temperature and form **fats**. Triglycerides from animals tend to have a higher proportion of saturated fatty acids.

Reactions that involve triglycerides include hydrogenation, oxidation and base-catalyzed hydrolysis (*saponification*).

Hydrogenation reaction is typical for triacylglycerols rich on unsaturated fatty acids that can be hydrogenated to produce saturated fats. In hydrogenation, double bonds in unsaturated fatty acids react with H_2 in the presence of a Ni or Pt catalyst:



If triacylglycerol in its composition contains several residues of unsaturated fatty acids, the hydrogenation may be complete or partial. In *complete hydrogenation* all unsaturated fatty acid residues of this triacylglycerol are subject to reduction, whereas in *partial hydrogenation* only some unsaturated fatty acid-components of triacylglycerol are converted into corresponding saturated fatty acids (Fig. 2.10)

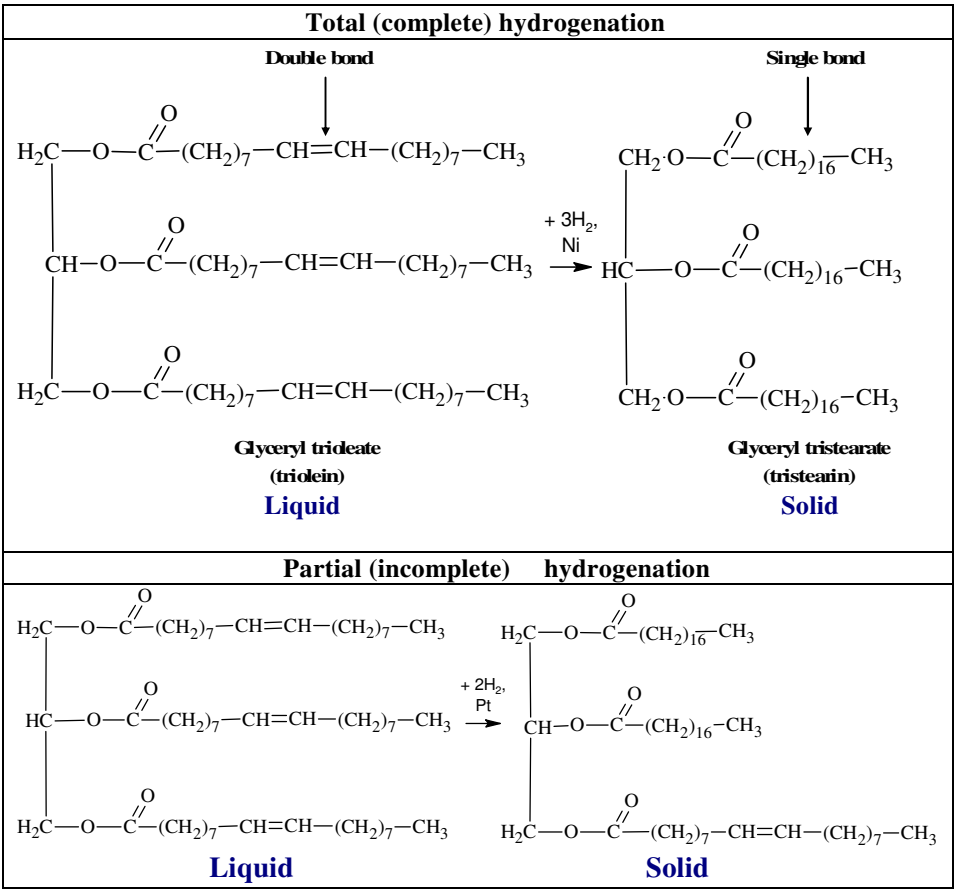


Fig. 2.10. The schemes of complete and partial hydrogenation

In fats unsaturated fatty acids normally found as a *cis*-isomers (Fig. 2.11). Their *trans*-isomers called **trans-fatty acids** occur in **trans fats**.

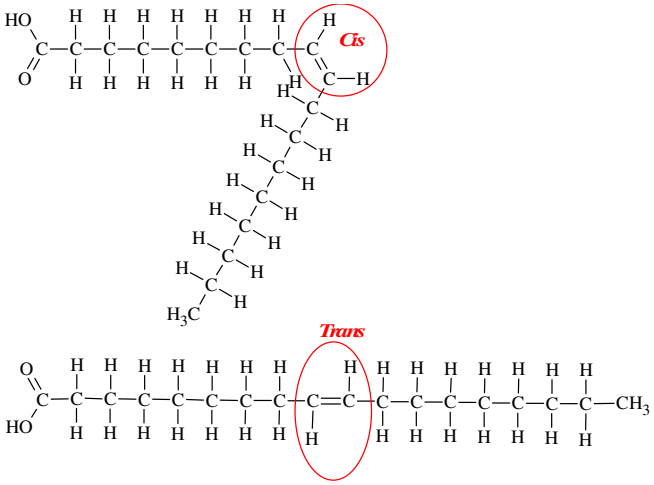


Fig. 2.11. Cis- and trans-fatty acids

Trans-fatty acids are formed from unsaturated fatty acids as a side product of their partial hydrogenation (Fig. 2.12).

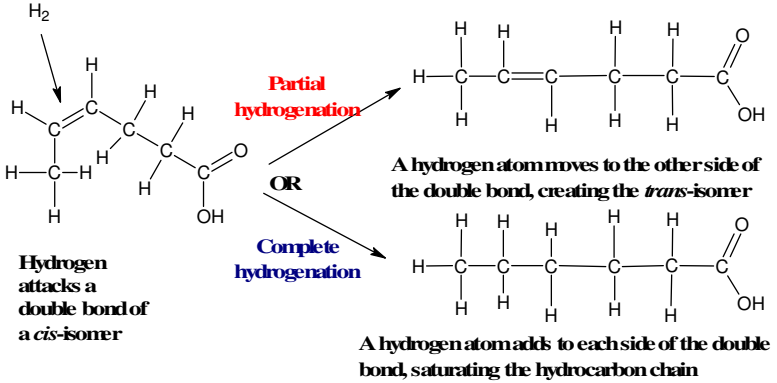


Fig. 2.12. The scheme of trans-fatty acids forming at partial hydrogenation

Trans fat are found in margarine, deep-fried foods like fried chicken and French-fried potatoes, snack chips, imitation cheese, ect (Fig. 2.13).

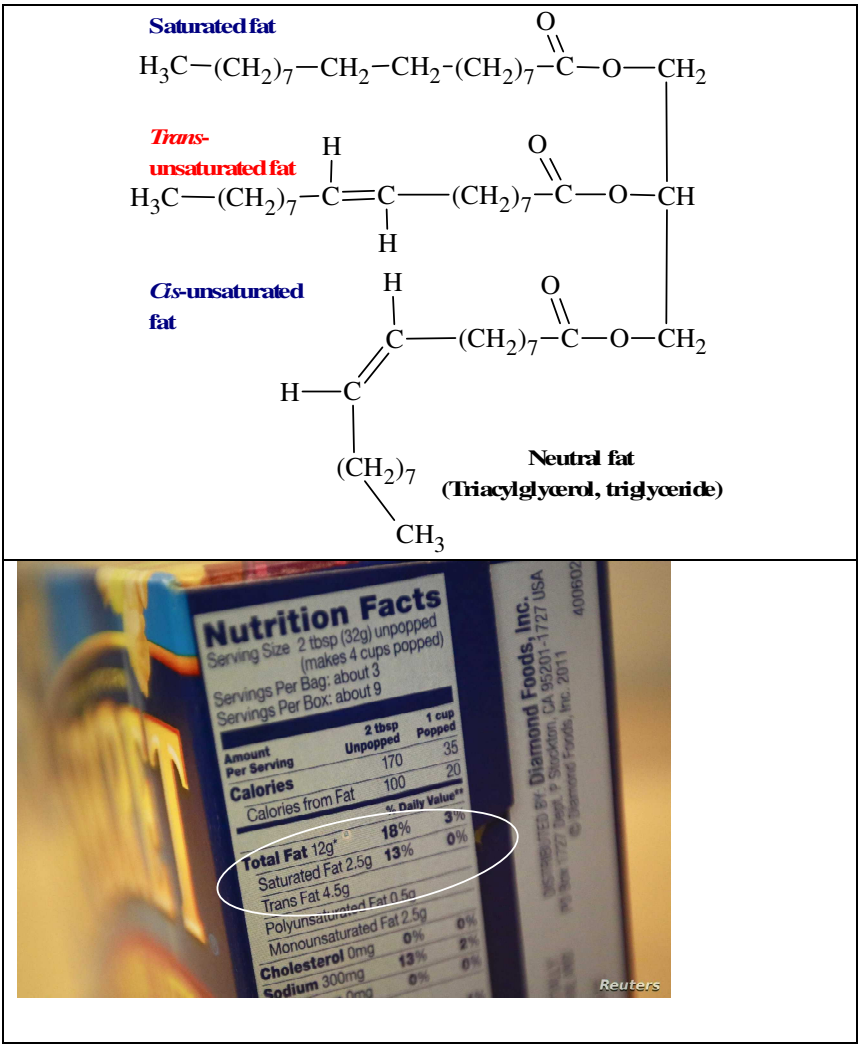


Fig. 2.13. The trans fat structure and example of trans fat courses

Trans fat tends to raise cholesterol levels in the blood, although not as much as saturated fat.

Rancidity of fat is the physico-chemical changes in the natural properties of fat leading to the development of unpleasant odor or taste or abnormal color because of atmospheric oxygen, light, bacterial or fungal enzymes or heat action. There are two main causes of fat rancidity (Fig. 2.14):

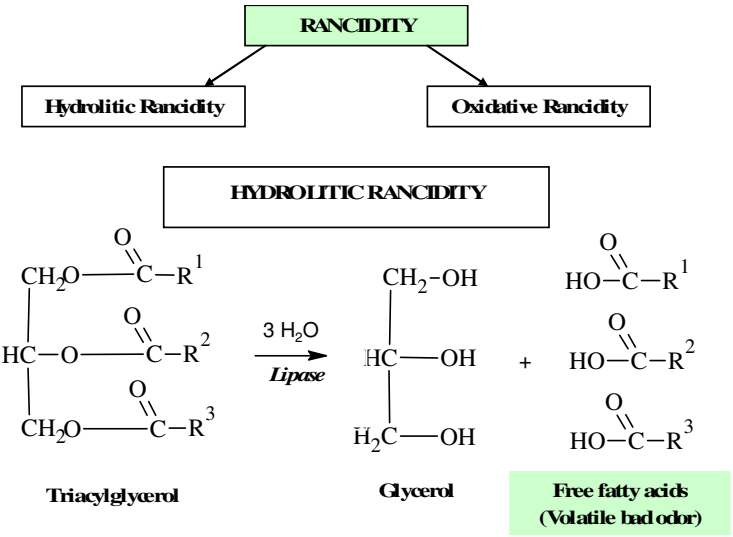


Fig. 2.14. The rancidity pathways

- *fat oxidation* (peroxides, aldehydes, ketones and dicarboxylic fatty acids are formed that smell bad) (Fig. 2.15);
- *fat hydrolysis* (with free fatty acids and glycerol formation, and free fatty acids have unpleasant odor).

Saturated fats are more resistant to rancidity then unsaturated. The products of rancidity are toxic and can cause food poisoning and cancer. Moreover, rancidity destroys fat-soluble vitamins and polyunsaturated essentially fat acids.

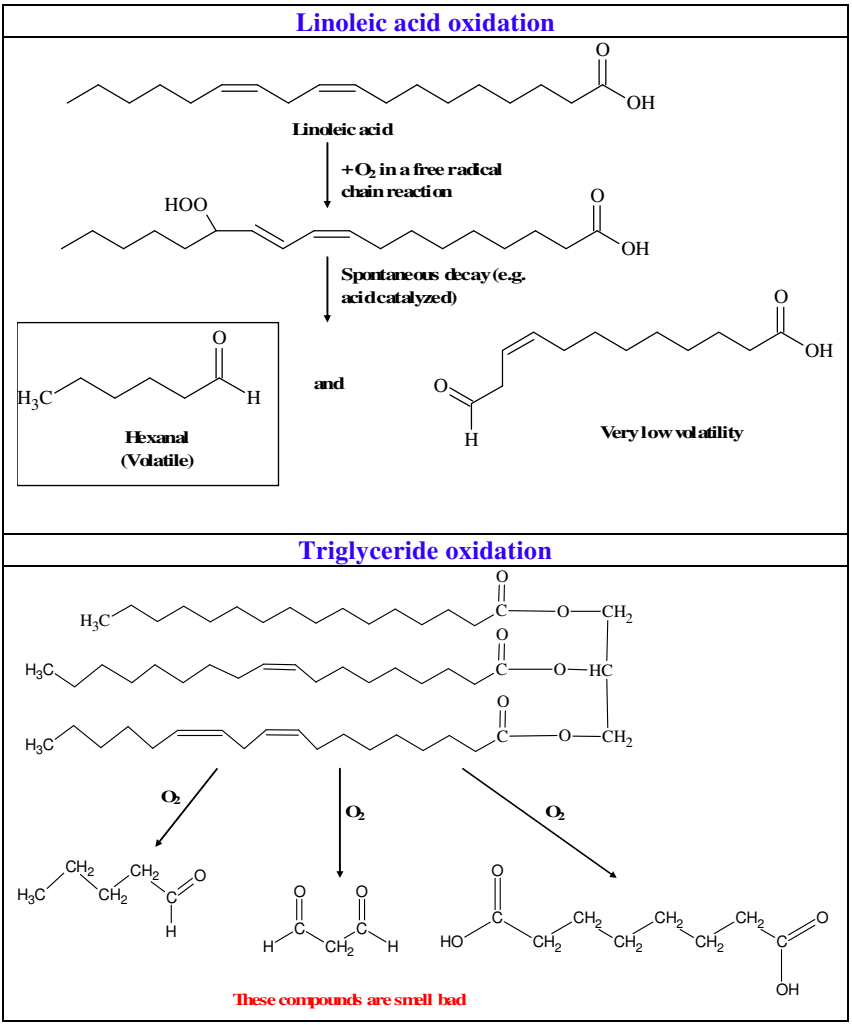


Fig. 2.15. The scheme of rancidity of fat caused by oxidation

Saponification (base-catalyzed hydrolysis) of fat. Saponification is the base-catalyzed hydrolysis of the ester bonds in a triglyceride (Fig. 2.16).

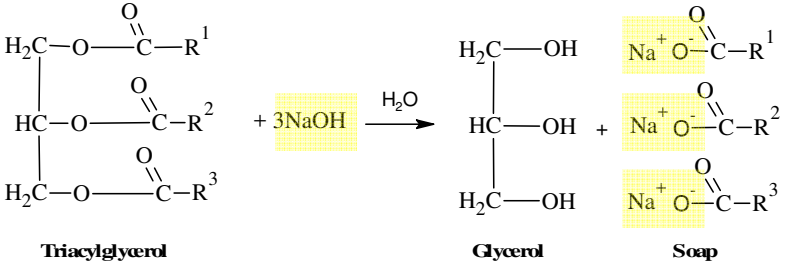


Fig. 2.16. The scheme of fat saponification

The products of this hydrolysis are carboxylic fatty acids and an alcohol glycerol. Because the reaction is base-catalyzed, the fatty acids then react with base forming salts of fatty acids – **soaps.** Soaps form **micelle** in the water solutes – in these structures hydrophobic chains of fatty acids are directed inside of micelle, whereas carboxylic group together with ions (Na^+ or K^+) are directed to water (Fig. 2.17).

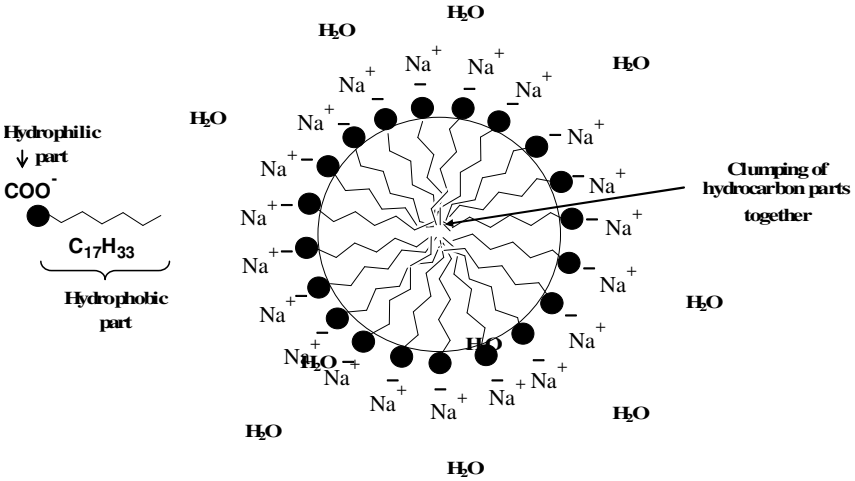


Fig. 2.17. The scheme of soap micelle structure

2.4. Waxes

Waxes are esters of long chain fatty acids and long chain alcohol (Fig. 2.18).

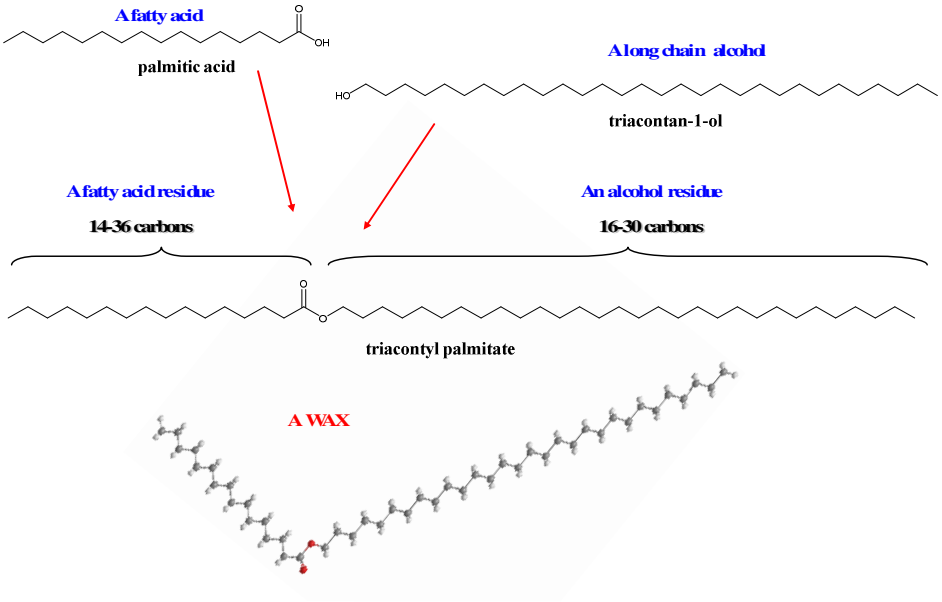


Fig. 2.18. The wax structure

Waxes are very hydrophobic and are used by plants and animals for protective, water-proof coatings. Some examples of waxes are given on Fig. 2.19 and on Tab. 2.2.

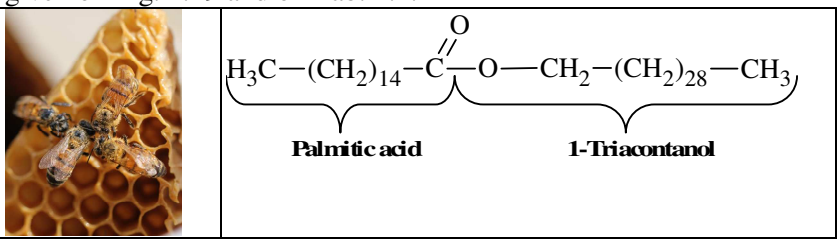


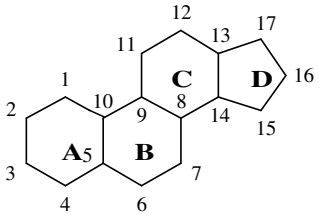
Fig. 2.19. The beeswax structure

Table 2.2.
Composition of some waxes

Name	Formula	Source	Use
Beewax	$CH_3(CH_2)_{14}CO_2CH_2(CH_2)_{28}CH_3$	Honeycomb	Candles
Carnauba wax	$CH_3(CH_2)_{24}CO_2CH_2(CH_2)_{28}CH_3$	Palm	Furniture wax
Insect wax	$CH_3(CH_2)_{24}CO_2CH_2(CH_2)_{48}CH_3$	Insect	Polish

2.5. Steroids

Steroids are a type of lipid that is not derived from a fatty acid. The structure of steroids is based on the tetracyclic (four-ring) system – **cyclopentanoperhydrophenanthrene (steroid) nucleus**:



Cyclopentanoperhydrophenanthrene (steroid) nucleus

Three of the rings are six-membered, while the fourth is five-membered.

Cholesterol is the steroid that used as the starting point for the synthesis of other steroids (Fig. 2.20). Besides being used to synthesize the other steroids, cholesterol is dissolved in membranes to keep them fluid.

The most of cholesterol is bonded through ester links to fatty acids (*esterified cholesterol*), but some is found in a free form (*free cholesterol*).

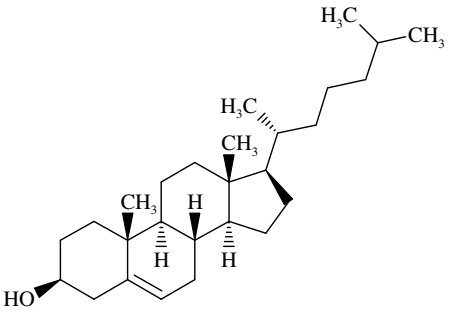


Fig. 2.20. Cholesterol structure

Bile acids are one of the main cholesterol derivatives (Fig. 2.21). They are synthesized in the liver, stored in the gallbladder, and secreted into the intestine where their function is emulsify dietary fats and aid in their absorption and digestion.

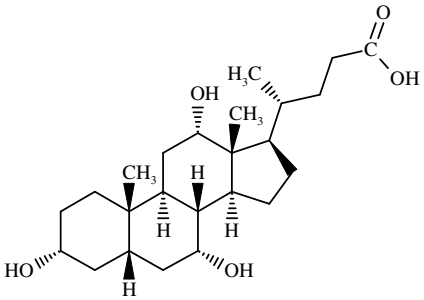


Fig. 2.21. Structure of trihydroxy bile acid – cholic acid

Male and female sex hormones - androgens and estrogens, respectively – are also cholesterol derivatives (Fig. 2.22) as well as corticosteroids (Fig. 2.23).

Corticosteroids are a class of steroid hormones – derivatives of cholesterol that are produced in the adrenal cortex of vertebrates and include both glucocorticoids and mineralocorticoids.

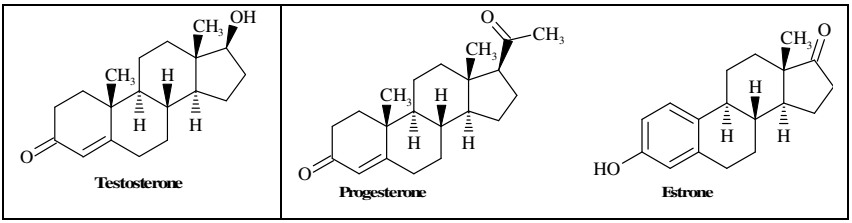


Fig. 2.22. Structure of the main sex hormones

Glucocorticoids (for example, cortisol – Fig. 2.23) are involved in glucose metabolism, whereas **mineralocorticoids** regulate blood pressure and volume by stimulating the kidneys to absorb Na^+.

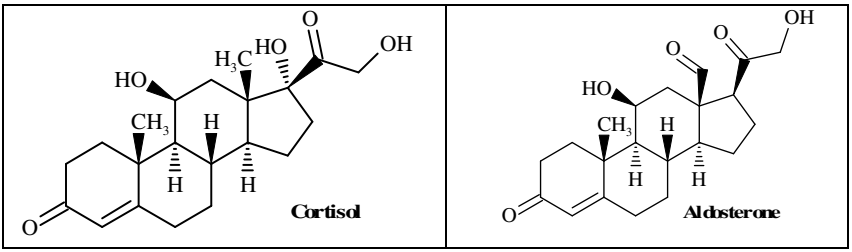


Fig. 2.23. Structures of the corticosteroids

Vitamin D (cholecalciferol) in its active form – *1,25-dioxycholecalciferol* - play a role in the regulation of calcium and phosphorus metabolism (Fig. 2.24).

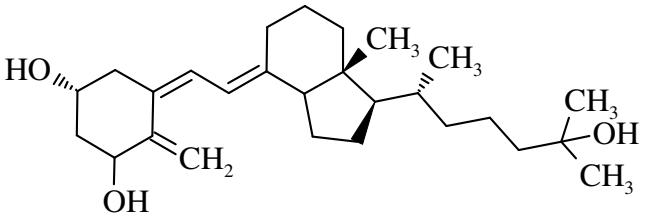


Fig. 2.24. Structures of Vitamin D3

Cardiac glycosides increase the force of cardiac muscle contraction via acting on the cellular Na^+/K^+ ATPase pump. The general structure of a cardiac glycoside consists of a steroid ring attached to a sugar and an R group (Fig. 2.25).

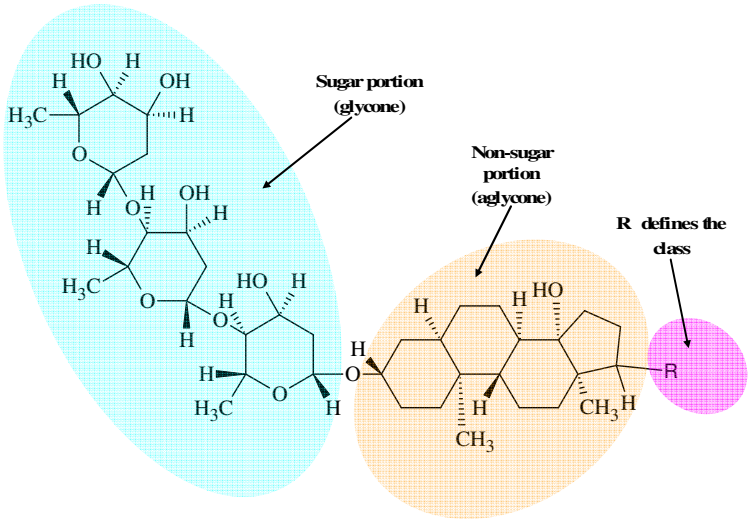


Fig. 2.25. The general structures of cardiac glycoside

Examples of cardic glycosides are *convallotoxin* (from *Convallaria majalis*), *ouabain* and *strophanthin* (from *Strophanthus kombe*), *digoxin* and *digitoxin* (from *Digitalis purpurea*) and *oleandrin* (*Nerium oleander*) (Fig. 2.26).

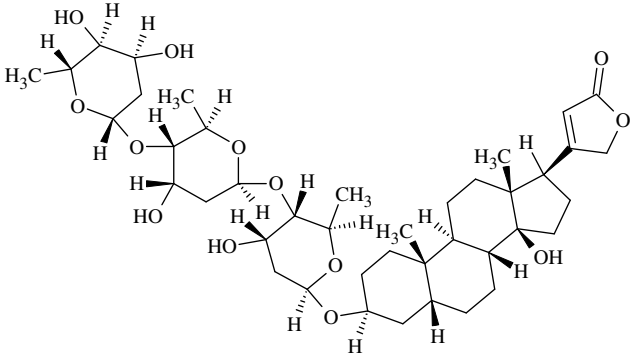


Fig. 2.26. The structure of digitoxin from *Digitalis purpurea*

They are used as treatments for congestive heart failure and cardiac arrhythmias; however, their relative toxicity prevents them from being widely used.

2.6. Complex lipids

Complex lipids are the major constituents of cell membranes but are also found in circulating fluids (Fig. 2.27). They can be classified into two groups:

- lipids with a *phosphate residue* called **phospholipids;**
- lipids containing a *carbohydrate residue* (one or more) called **glycolipids.**

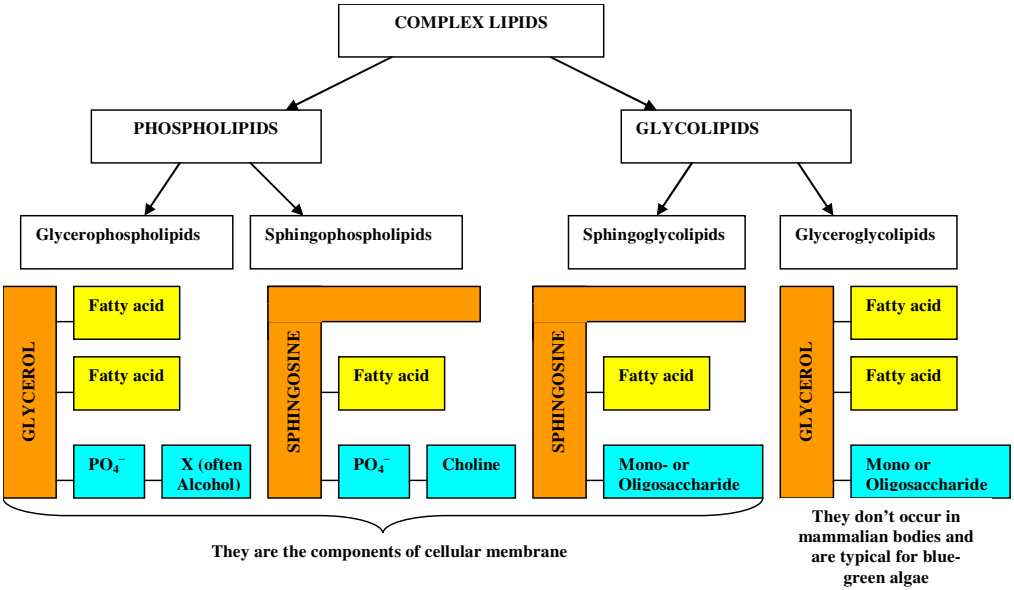


Fig. 2.27. Complex lipids classification

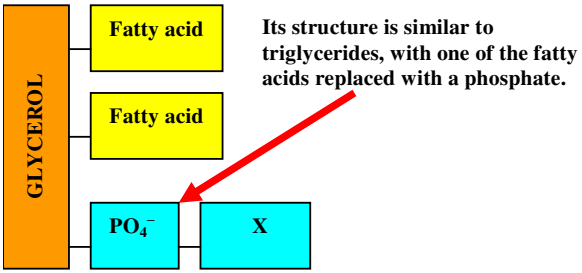
Both groups of lipids can be divided into two subgroups dependent on alcohol that is contained in their structures – **glycerol** (such lipids are called *glycerophospholipids* or *glyceroglycolipids*, respectively) or **sphingosine** (*sphingophospholipids* or *sphingoglycolipids*).

There are also **lipoproteins** – structures that includes *different types of lipids* plus protein part called *apoproteins*.

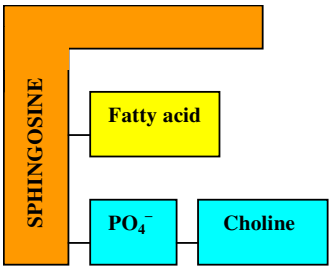
2.6.1. Phospholipids

These lipids contain phosphoric acid residue and are divided into:

- **glycerophospholipids** (or *phosphoglycerides*; contain alcohol **glycerol**):



- **sphingophospholipides** (e.g., *sphingomyelin*) – contain unsaturated amino alcohol **sphingosine**, that has 18 carbon atoms:



Phospholipids are found almost exclusively in plant and animal membranes, which typically consist of 40% - 50% phospholipids and 50% - 60% proteins.

2.6.1.1. Glycerophospholipids

Glycerophospholipids may be regarded as *derivatives of phosphatidic acid* (glycerol esterified with two molecules of fatty acid and one of phosphoric acid) (Fig. 2.28).

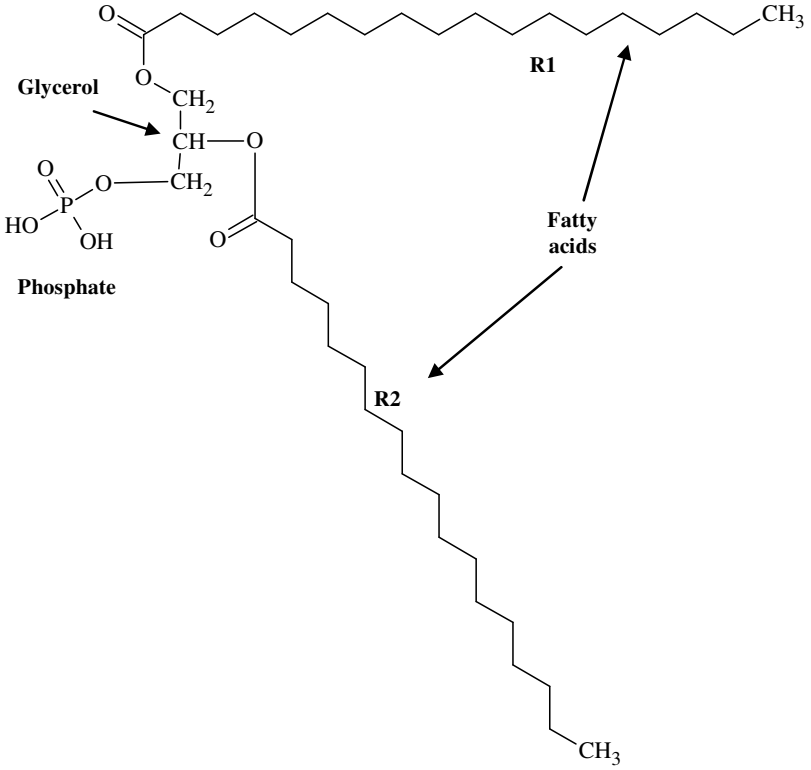


Fig. 2.28. Distearoyl phosphatidic acid structure

Any glycerophospholipid is made up of (Fig. 2.29):

- *3-phosphorylated glycerol* (glycerol-3-phosphate);
- *two fatty acids residues (R1 and R2)* linked to OH-groups near C-1 and C-2 by ester bonds;
- *a nitrogenous compound or cyclic alcohol inositol ("X")* linked to phosphate group of glycerol-3-phosphate by a third ester bond.

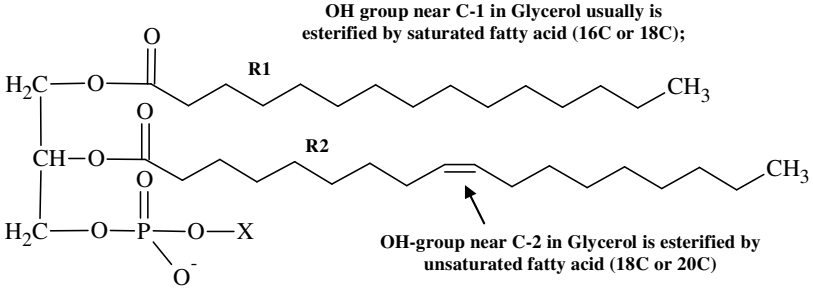


Fig. 2.29. Common structure of glycerophospholipids

Three most abundant fatty acids in phosphatidic acids and glycerophospholipids are palmitic acid (16:0) and stearic acid (18:0) (in R1 position), as well as oleic acid (18:1) (in R2 position). Besides polyunsaturated fatty acids (such as arachidonic, linolic and linolenic ones) residues can be found in the second position.

There are 6 classes of glycerophospholipides (each of them has some compounds differ by fatty acid residues) (Fig. 2.30):

- *phosphatidylcholines* (its another name is Lecithin) ("X" - choline);
- *phosphatidylethanolamines* ("X" - ethanolamine);
- *phosphatidylserine* ("X" - serine);
- *phosphatidylinositol* ("X" - inositol);
- *plasmalogenes;*
- *cardiolipins.*

Phosphatidylcholines (*lecithins*) are the most abundant phospholipids of the cell membrane. They are present a large proportion of the body's store of choline. **Phosphatidyl ethanolamines** (*cephalins*) and **phosphatidyl serines** are functional similar to lecithin. Brain and nervous tissue are rich in cephalin, whereas phosphatidyl serine is found in most tissues. Such representative of **phosphatidylinositols** as *phosphatidylinositol 4,5-bisphosphate* is an important constituent of cell membrane phospholipids; upon stimulation by a suitable hormone agonist, it is cleaved into *diacylglycerol* and *inositol-1,4,5-trisphosphate***,** both of which act as internal signals or second messengers.

Cardiolipins and plasmalogens are structurally different from other Glycerophospholipids (Fig. 2.31). **Cardiolipins** are found in inner mitochondrial membrane where is involved in respiratory chain functioning, whereas **plasmalogens** – brain and muscle, where its content is about 10% of the phospholipids. In plasmalogen molecule the OH-group near 1-C of glycerol is bond with a *cis-α, β-unsaturated alcohol* instead of a saturated fatty acid as in other glycerophospholipids. Polar Head group "X" of plasmalogen can either be Choline, Ethanolamine, Choline or Serine.

Lysophospholipids are the glycerophospholipids that contain only one fatty acid residue in their molecule. For example, lecithin (phosphatidylcholine) contains 2 fatty acyl residues in its molecule, and *lysolecithin* contains only one fatty acyl residue.

Phosphoglycerides are *amphipathic* because they contain:

- two non-polar aliphatic hydrophobic chains ("*tails*");
- polar hydrophilic phosphoryl-X group ("*head*").

Therefore, when these molecules are mixed with water, they spontaneously form **micelles** (spherical structures that are formed by one lipid monolayer – Fig. 2.17) or **lipid bilayers** (planar or spherical – the latter are called **liposomes**); in this structures hydrophobic groups of glycerophospholipids are contacted each with other, and hydrophilic groups - with water (it is the basis of membrane structure) (Fig. 2.32).

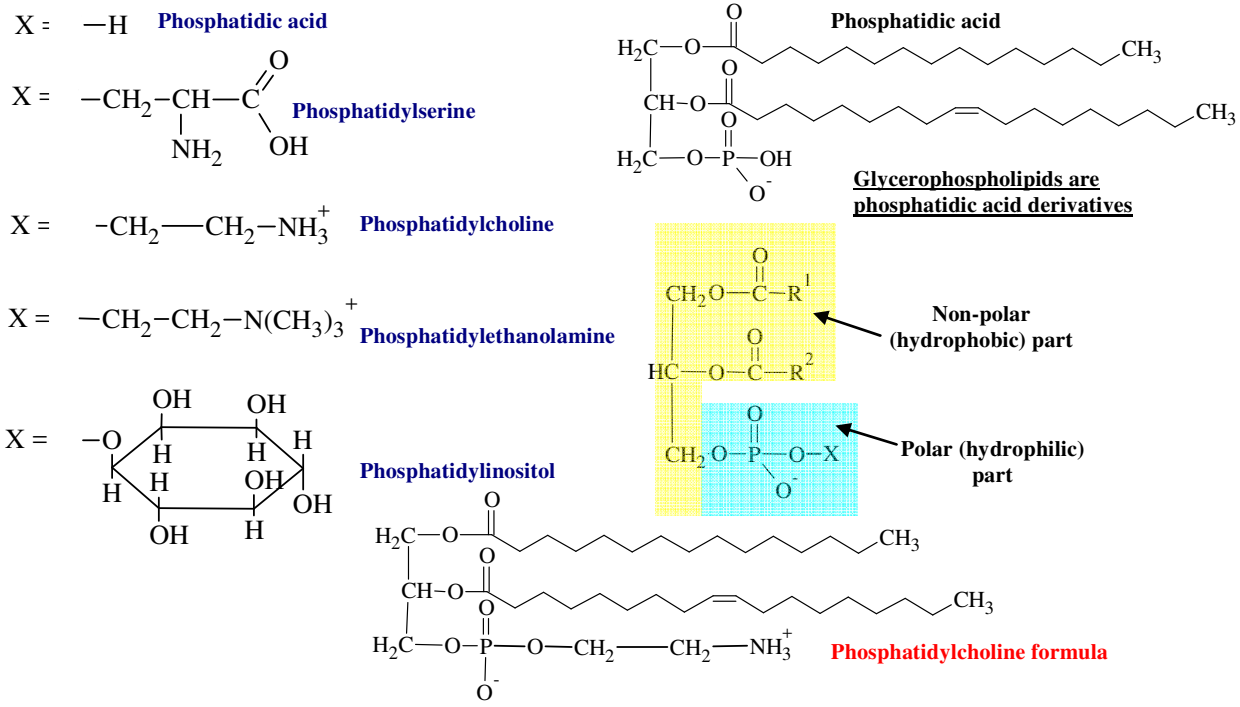


Fig. 2.30. Peculiarity of structure of different classes of glycerophospholipids

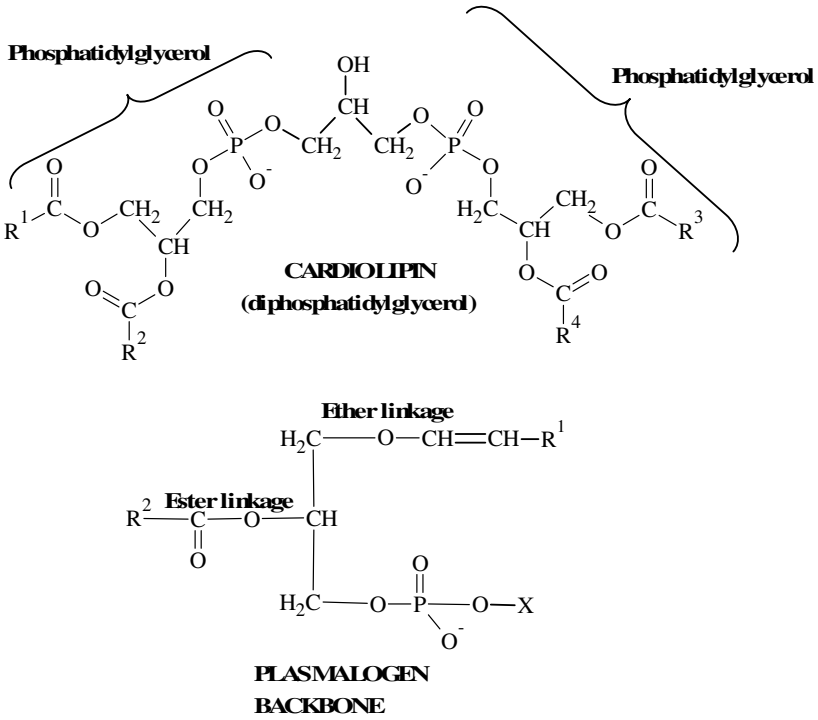


Fig. 2.31. Structures of cardiolipin and plasmalogen

On the surface of the water / air separation, phospholipids form a **monolayer** in which the hydrophilic heads are directed to the aqueous phase and the hydrophobic fatty acid tails - to the air (Fig. 2.32).

In general, the main function of glycerophospholipids is their role in biological membranes structure and functions: fluid-mosaic model of biological membrane showing the lipid bilayer and membrane proteins on the inner and outer surfaces of the membrane (Fig. 2.33).

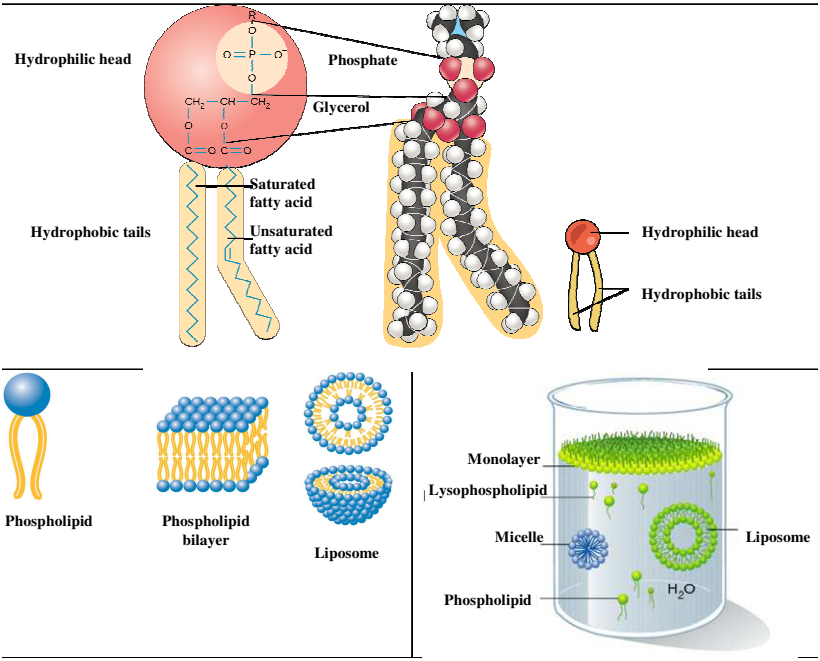


Fig. 2.32. Amphipatic properties of glycerophospholipids; monolayer, micelle, lipid bilayer and liposome

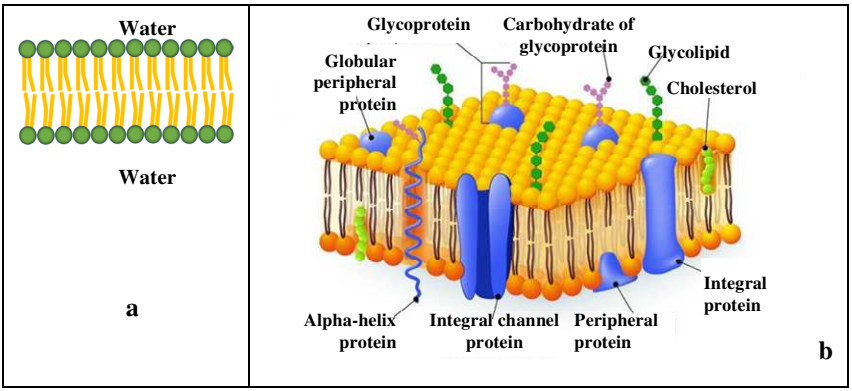


Fig. 2.33. Phospholipid bilayer as a main part of biological membranes (a) and fluid-mosaic model of biological membrane (b)

2.6.1.2. Sphingophospholipids

The sphingophospholipids function is similar to the glycerophospholipids, but structurally they are different: there is no glycerol core, and glycerol and one of the fatty acids found in glycerophospholipids is replaced with a molecule called **sphingosine** (Fig. 2.34).

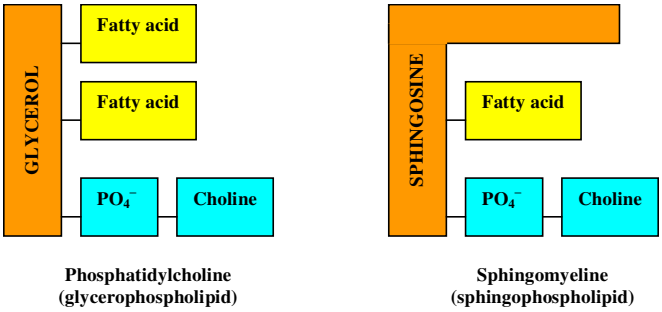


Fig. 2.34. Comparative structures of glycerophospholipids and sphingophospholipids

Compound formed when a fatty acid molecule is linked to -NH_2 group in sphingosine via amide bond is called **ceramide** (Fig. 3.35).

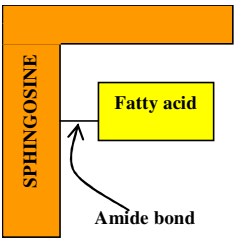


Fig. 2.35. Ceramide structure

Ceramides form the core structure of naturally occurring sphingolipids – both sphingophospholipids and sphingoglycolipids that are founded in cell membranes.

The main representative of sphingophospholipids is **sphingomyeline**. It contains (Fig. 2.36) *sphingosine residue, one fatty acid residue* linked to -NH_2 group in sphingosine via amide bond as well as *phosphate* linked with *choline* (*phosphocholine*) via ester linkage. Sphingosine and fatty acid residue together forms ceramide.

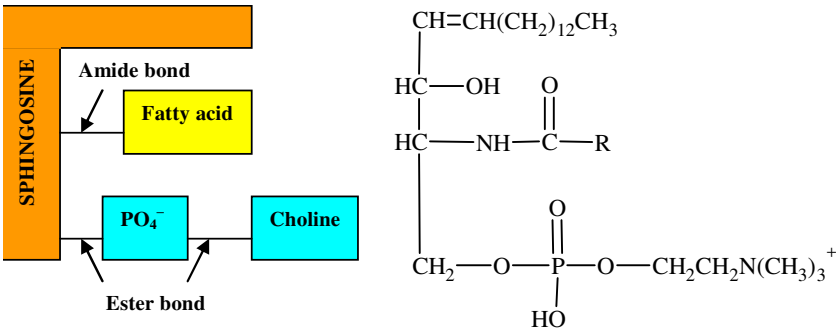


Fig. 2.36. Structure of sphingophospholipid sphingomyeline

Sphingomyelins are found in all cell membranes and are important structural components of the myelin sheath, the protective and insulating coating that surrounds nerves.

2.6.2. Glycolipids

These lipids have no phosphate but contain a carbohydrate residue (one or more) and are divided into:

- **glyceroglycolipids** (contain alcohol **glycerol**). Glyceroglycolipids don't widely occur in the animal tissues, but is typical fore some bacteria and plants). Sphingoglycolipids occur particularly in the outer leaflet of the plasma membrane, where they contribute to cell surface carbohydrates. They can act as cell surface receptors for various hormones and growth factors, play important role in cellular interactions, growth and development. For example, ganglioside GM1 acts as a receptor for cholera toxin in

human intestine. They are also an important source of blood group antigens and various embryonic antigens.

- **sphingoglycolipids** (in which one fatty acid residue linked to -NH_2 group in **sphingosine** via amide bond; a carbohydrate part links to sphingosine by glycosidic bond) (Fig. 2.37). The comparative structures of sphingophospholipids and sphingoglycolipids are given on Fig. 2.38.

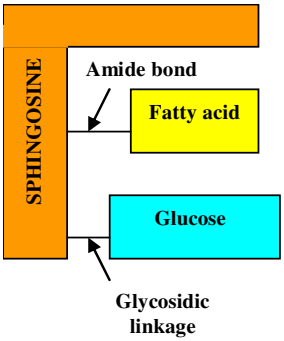


Fig. 2.37. Structure of the simplest sphingoglycolipid glucocerebroside

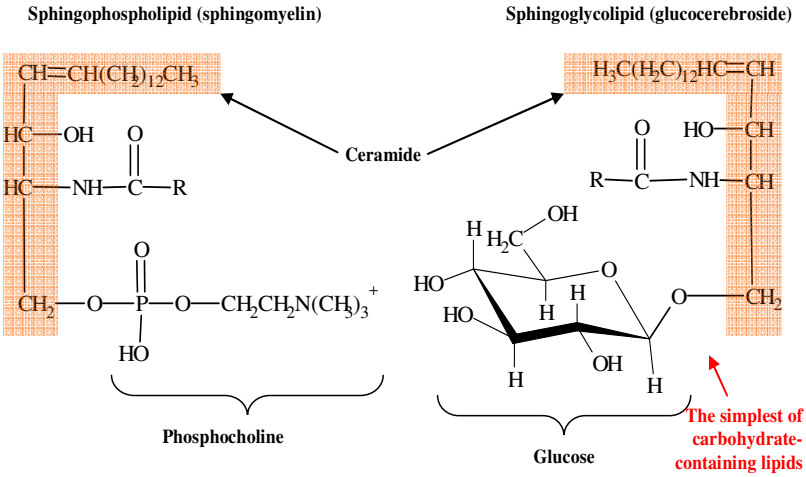


Fig. 2.38. The comparative structures of sphingophospholipids and sphingoglycolipids

Sphingoglycolipids are divided into next classes (Fig, 2.39):

- *cerebrosides;*
- *sulfatides;*
- *globosides;*
- *gangliosides.*

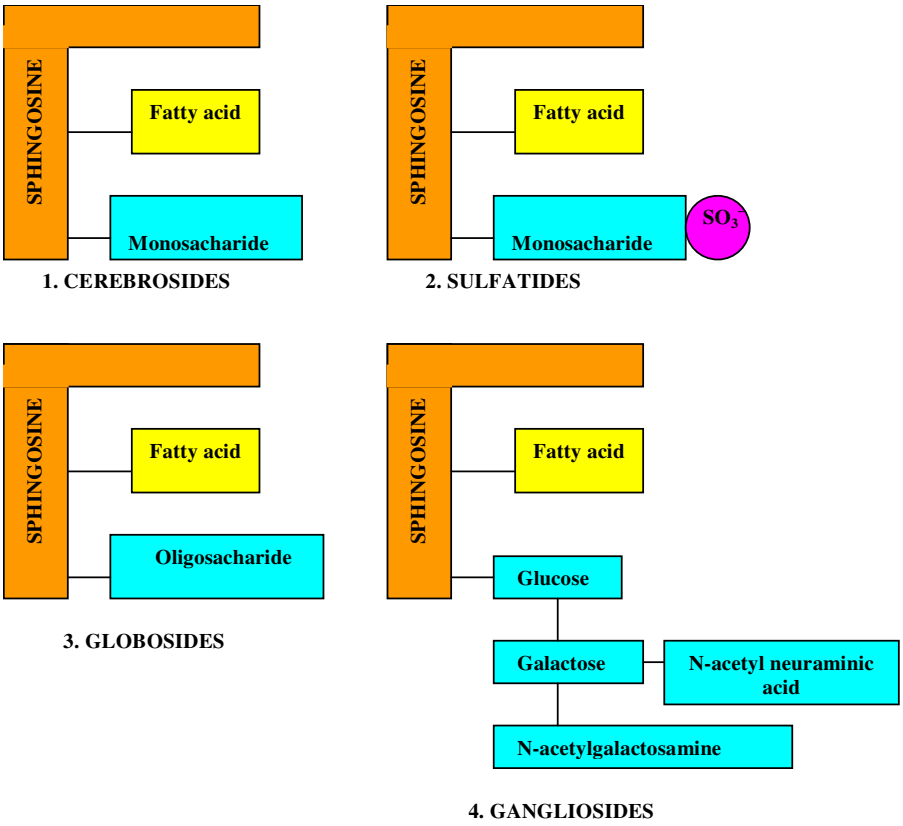


Fig. 2.39. The comparative structures of different classes of sphingoglycolipids

Cerebrosides have one residue of glucose or galactose (monosaccharide) as carbohydrate part (*glucocerebrosides* or

galactocerebrosides that also may be called as *glucosylceramides* and *galactosylceramides*) (Fig. 2.40).

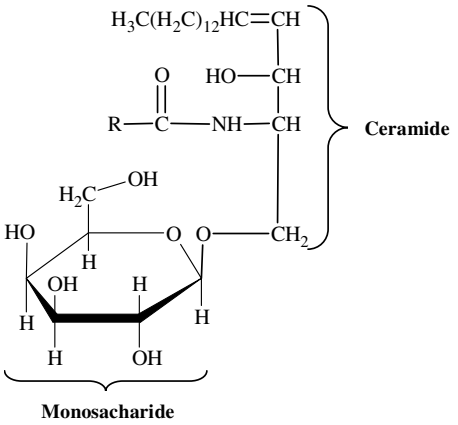


Fig. 2.40. Cerebroside structure

These compounds are found in the brain and nervous tissue with high concentration in myelin sheath.

Sulfatides are cerebrosides that contain sulfated galactosyl residues (Fig. 2.41).

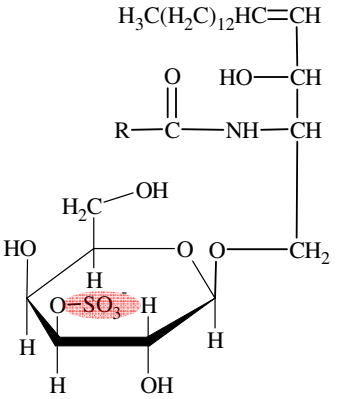


Fig. 2.41. Sulfatides structure

They are negatively charged at physiological pH and found predominantly in nerve tissue and kidney. Their failure of degradation causes them to accumulate in nervous tissues.

Globosides have two or more glucose or galactose residues (oligosaccharide) in their structure (Fig. 2.42).

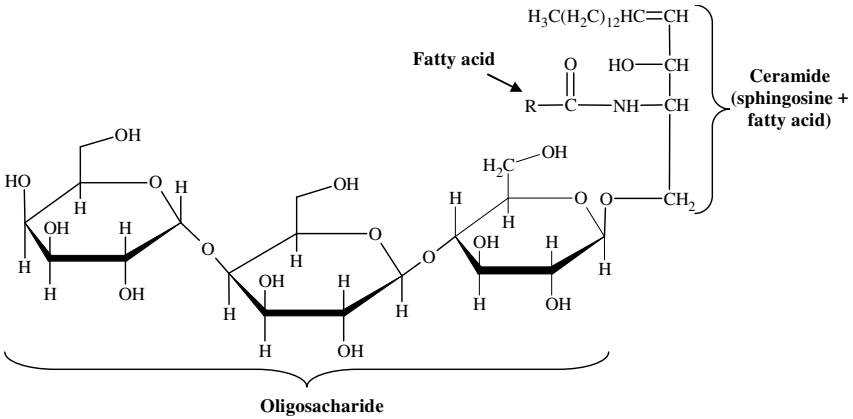


Fig. 2.42. Globoside structure

They are produced by attaching additional monosaccharides to glucocerebroside.

Gangliosides are negatively charged at physiological pH. The negative charge is imparted by *N-acetyl neuraminic acid* (*sialic acid*) (Fig. 2.43).

Brain gangliosides may contain up to four sialic acid residues and based on that they are GM, GD, GT and GQ, containing 1,2,3 or 4 sialic acid residues.

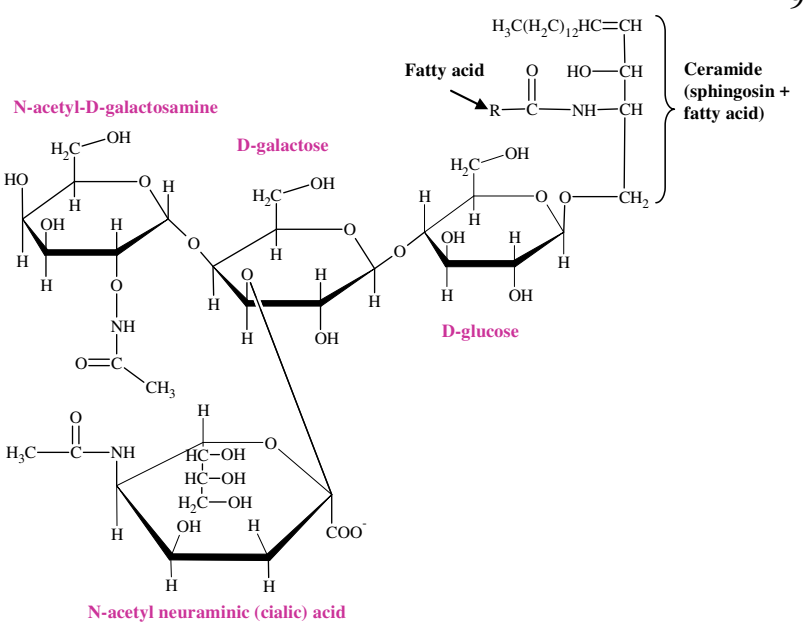


Fig. 2.43. Ganglioside GM_2 structure

2.6.3. Lipoproteins

Lipoproteins are the molecular complexes found in blood plasma. They are used to transport lipids (triglycerides, phospholipids and cholesterol) in the blood (Fig. 2.44).

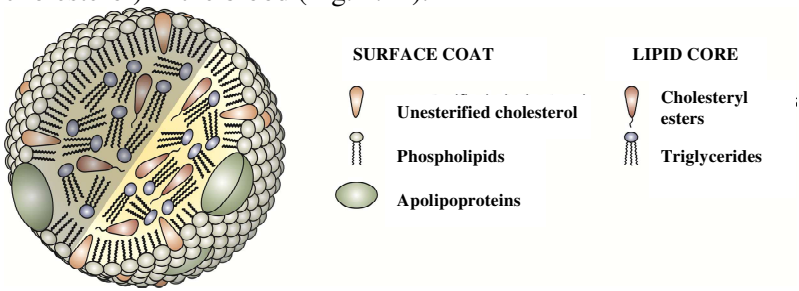


Fig. 2.44. The general lipoproteins structure

All of them contain:

- *neutral lipid core with cholesterol esters and triglycerides;*
- *a layer of phospholipids, free cholesterol and proteins* (called *apoproteins*) *around core.*

The next lipoprotein classes can be found in the blood (Fig. 2.45):

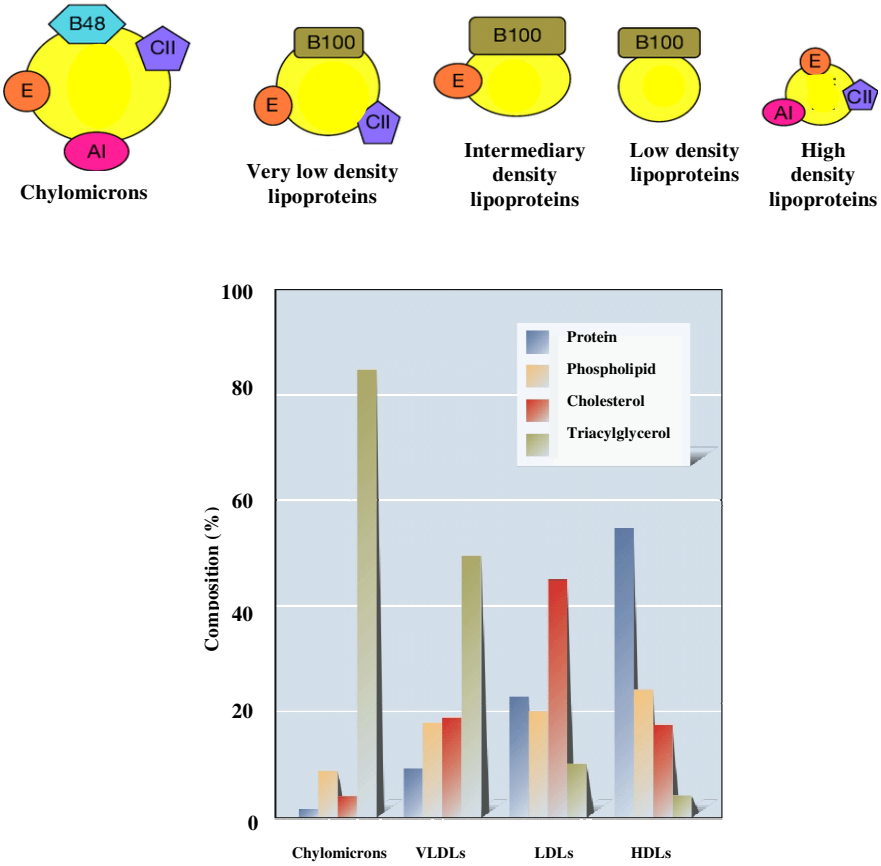


Fig. 2.45. Comparative sizes, types of apoproteins and lipid composition of the main lipoproteins classes

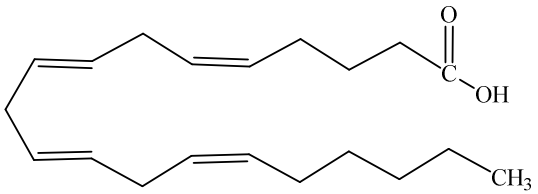
- **chylomicrons** – very large complexes that transport primarily *triacylglycerols* from the digestive track to adipose tissue;
- **VLDL** (very low density lipoproteins) are made in liver; they transport mainly *cholesterol* and *triacylglycerols* to the tissues;
- **LDLs** (low density lipoproteins) which transport mainly *cholesterol* to the tissues;
- **HDLs** (high density lipoproteins) - they are formed in liver; HDLs scavenge excess cholesterol esters and transport cholesterol and phospholipids back to the liver. They are also storage of apoproteins.

Comparative sizes, types of apoproteins and lipid composition of the main lipoproteins classes are given on Fig. 2.45

Cholesterol of HDLs is a "*good cholesterol*", and cholesterol of VLDLs and LDLs is a "*bad cholesterol*" – its high blood levels may cause atherosclerosis development.

2.7. Eicosanoids

Eicosanoids are derived from *arachidonic acid*:



Arachidonic acid
(*all-cis*-5,8,11,14-Eicosatetraenoic acid)

Their examples are (Fig. 2.46):

- *leukotrienes (LTs);*
- *prostaglandins (PGs);*
- *prostacyclins (PGIs);*
- *thromboxanes (TXs).*

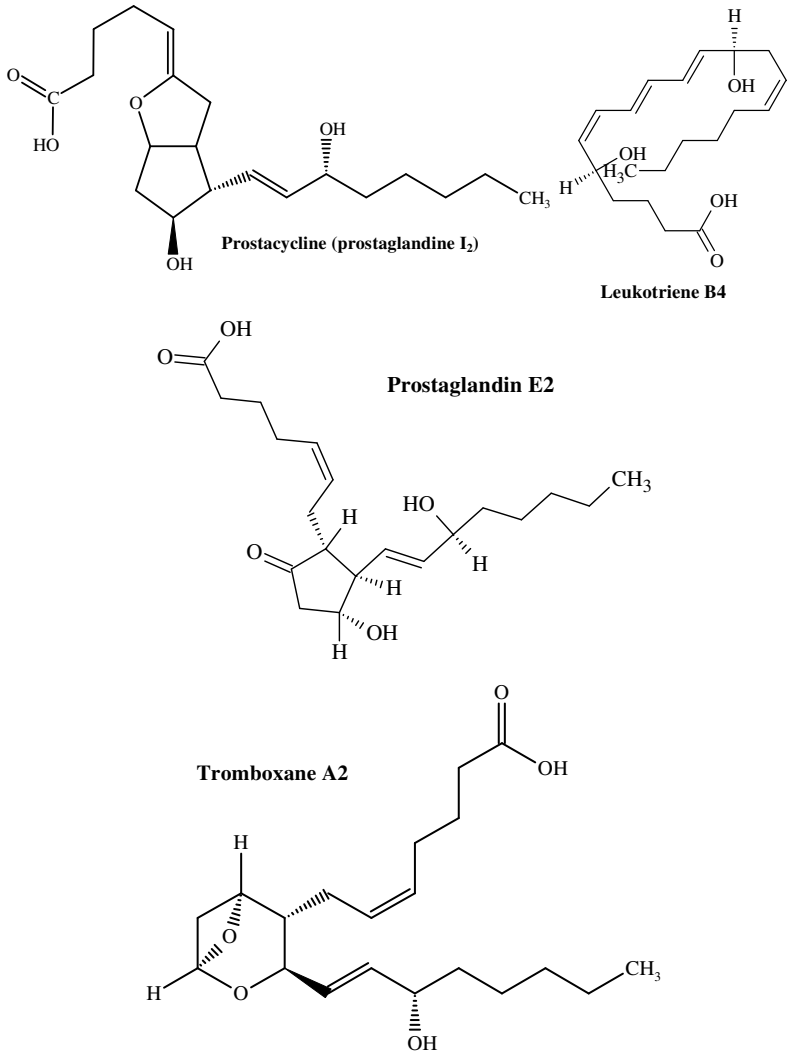


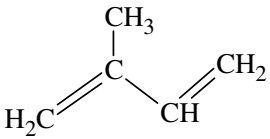
Fig. 2.46. Examples of eicosanoids

These compounds are called *messengers of pain and inflammation* because of different eicosanoids take part in the

maintenance or inhibition of inflammation, allergy, fever and other immune responses as well as serve as the contributes of the perception of pain. They also can control cell growth, blood pressure and modulate the regional flow of blood to tissues. For example, **prostaglandins** are powerful locally acting vasodilators and inhibit the aggregation of blood platelets. Through their role in vasodilation, prostaglandins are also involved in inflammation – therefore, the medication that inhibits production of these compounds can reduce inflammation, fever, and pain. Conversely, **thromboxanes** - eicosanoids, produced by platelets - are vasoconstrictors and facilitate platelet aggregation. Their name comes from their role in clot formation (thrombosis). **Leukotrienes** also have a powerful effect in bronchoconstriction and contribute to the pathophysiology of asthma – so drugs that block leukotrienes action are used clinically as maintenance treatment for allergen-induced asthma and rhinitis.

2.8. Isoprenoids

Isoprenoids contain a repeating five-carbon unit know as *isoprene*:



Isoprene

They are synthesized from isopentenyl pyrophosphate. One of the lagest groups of isoprenoids are **terpenes** – an enormous group of molecules that are found largely in the essential oils of plants. Terpenes are divided into:

- **monoterpenes**: 2 isoprene units - *geraniol* (in germaniums);
- **sesquiterpenes**: 3 isoprene units - *farnesene* (part of citronells oil);

- **diterpenes**: 4 isoprene units - *phytol* (a plant alcohol);
- **tetraterpenes**: 8 isoprene units - *carotenoids* (orange pigment in plants).

Their structures are given on Fig. 2.4.

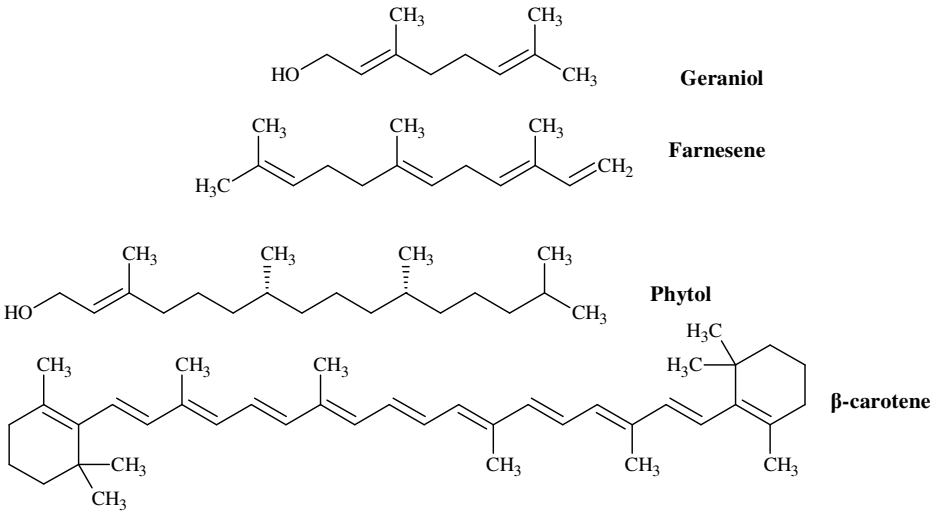
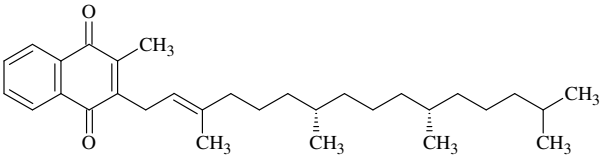


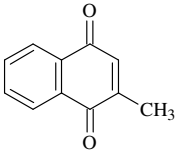
Fig. 2.47. Examples of terpenes

There are also **mixed terpenoids –** biomolecules which have isoprenoid (isoprenyl) components. Their examples include some fat soluble vitamins (*vitamin E*, *vitamin K*, *vitamin A*) and *ubiquinone*. Moreover, some proteins (for example, G proteins which is an important component of cell transduction) can be *isoprenylated* (attached to isopreniod groups) – it causes .their attachment to plasmatic membrane.

Vitamin K - the name of this vitamin comes from *Germ. koagulation*, signifying its important role in the blood-clotting process:

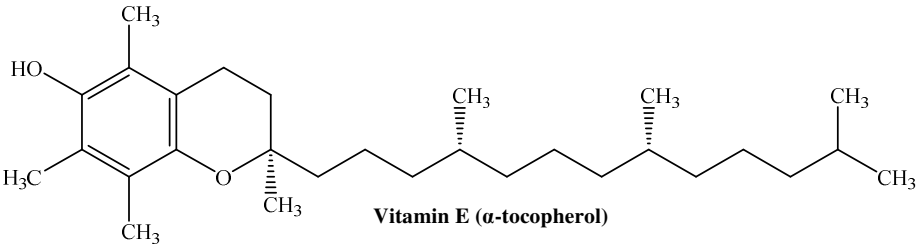


Vitamin K_1



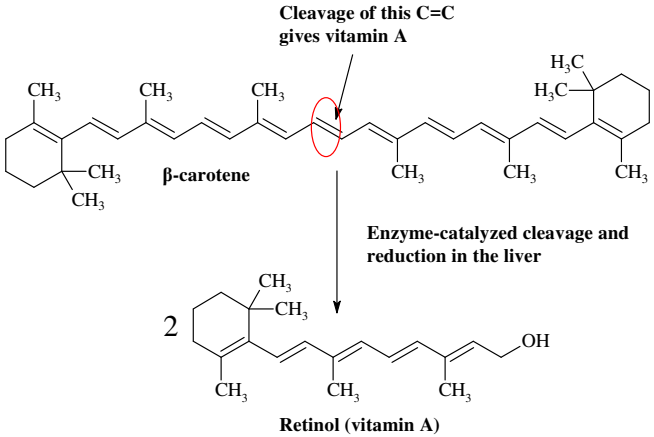
Menadione (a synthetic vitamin K analog)

Vitamin E (α-tocopherol) functions as an antioxidant:



Vitamin E (α-tocopherol)

Vitamin A occurs only in the animal world. It is found in the plant world in the form of a provitamin in a group of pigments called carotenes:



Retinol (vitamin A)

2.9. Test questions

1. Which of these lipids built from fatty acids and glycerol?

A. waxes
B. neutral fats
C. glycerophospholipids
D. glycerosphingolipids
E. steroids

2. Which of these lipids belong to complex lipids?

A. waxes
B. neutral fats
C. glycerophospholipids
D. derivatives of cholesterol
E. steroids

3. Bile acids and vitamin D3 are derivatives of:

A. Arachidonic acid
B. Cholesterol
C. Waxes
D. Neutral fat
E. Sphingolipids

4. Vitamin A, vitamin K, vitamin E as well as terpenes belong to:

A. derivatives of cholesterol
B. fatty acids derivatives
C. eicosanoisds
D. izoprenoids
E. waxes

5. The examples of non-saponifiable lipid are:

A. neutral fats,
B. phospholipids,
C. steroids
D. glycolipids
E. waxes

6. Which of these fatty acids has such shorthand nomenclature -$C_{20:4}$,$\Delta^{5,8,11,14}$:

A. palmitic
B. oleic
C. linolenic
D. linoleic acid
E. arachidonic acid

7. Triolein and 1-palmitoleoyl-2-linoleoyl-3-stearoyl-glycerol are:

A. neutral fats,
B. phospholipids,
C. steroids
D. glycolipids
E. waxes

8. Adipocytes are the cells of the adipose tissue. Their function is to deposite:

A. waxes
B. neutral fats
C. glycerophospholipids
D. derivatives of cholesterol
E. steroids

9. The physico-chemical changes in the natural properties of fat leading to the development of unpleasant odor or taste or abnormal color because of atmospheric oxygen, light, bacterial or fungal enzymes or heat action are called:

A. Saponification
B. Partial (incomplete) hydrogenation
C. Complete hydrogenation
D. Rancidity
E. Esterification

10. Phosphatidylethanolamines and phosphatidylcholines belong to:

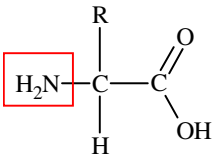
A. waxes
B. neutral fats
C. glycerophospholipids
D. derivatives of cholesterol
E. sphingophospholipids

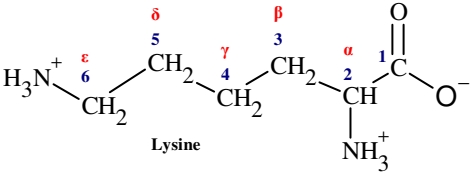
PART 3. AMINO ACIDS AND PROTEINS

Proteins are the biopolymers, the monomers of which are α-amino acids linked by peptide bonds. These molecules can have the next functions in the body:

- **structural** (*collagen*, *elastin* - proteins of connective tissue; *keratin* - nails and skin; *glycoproteins* - shells and cell walls);
- **catalytic** (*enzymes*);
- **regulatory** (*protein hormones*, *inhibitors and activators of enzymes*);
- **receptors**;
- **transport** (*hemoglobin*; *serum albumin*; *transferrin*);
- **protective** - *immunoglobulins* of the blood, *plasma fibrinogen*, *interferon*;
- **contractile** (*actin*, *myosin*);
- **nutritious** (*casein milk*, *ovalbumin*, *wheat gliadin*);
- **energy** (1 g of protein, when degraded to the end products, yields 17,6 kJ of energy).

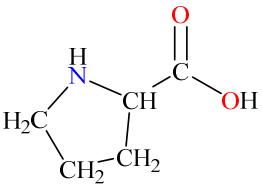
α-amino acids are derivatives of carboxylic acids, in which one hydrogen atom near the second (alpha) carbon atom is replaced by an amino group:





Lysine

In the amino acid *proline*, the amino group is included in the ring as α-imino group, therefore proline is more likely an α-imino acid than an α-amino acid:



20 amino acids are the part of the proteins. Others are called *non-proteinogenic* or *non-standard amino acids*. **Non-standard amino acids** that are found in proteins are formed by post-translational modification, which is modification after translation during protein synthesis. The examples of such amino acids are:

- *hydroxyproline* (is formed from proline);
- *hydroxylysine* (is formed from lysine);
- *γ-carboxyglutaminic acid* (is formed from glutaminic acid);
- *selenocysteine* (is formed from cysteine).

These modifications are often essential for the function or regulation of a protein. For example, the carboxylation of glutamate allows for better binding of calcium cations, and the hydroxylation of proline is critical for maintaining connective tissues.

Other amino acids - **non-proteinogenic** ones - occur in the body, but are not found in proteins. They are:

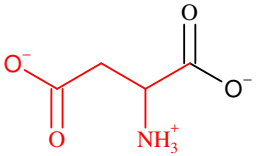
- *citrulline,*
- *ornithine,*
- *carnitine,*
- *homocysteine,*
- *taurine,*
- *GABA,*
- *iodinated amino acids* (thyroid hormones T3 and T4).

These amino acids often occur as intermediates in the metabolic pathways for standard amino acids — for example, ornithine and citrulline occur in the urea cycle, part of amino acid catabolism.

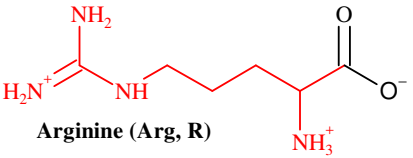
α-amino acids differ one from other by their radical groups (R). These radicals may contain hydrophilic or hydrophobic functional groups, can be positively or negatively charged and can have aliphatic or cyclic structures.

3.1. Nomenclature and classification of α-amino acids

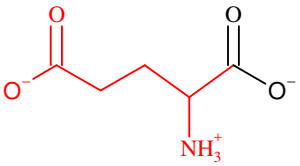
Amino acids can be named by three letters (Ala, Trp) or by one letter (A, W) (Fig. 3.1):



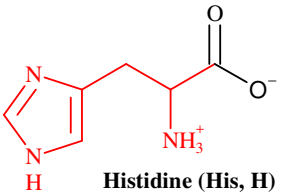
Aspartic acid (Asp, D)



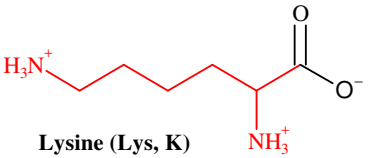
Arginine (Arg, R)



Glutamic acid (Glu, E)



Histidine (His, H)



Lysine (Lys, K)

Amino acids	1-Letter abbrevi-ation	3-Letter abbrevi-ation
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His

Amino acids	1-Letter abbrevi-ation	3-Letter abbrevi-ation
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

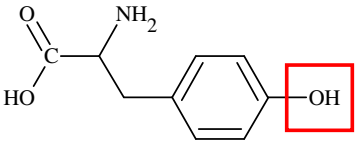
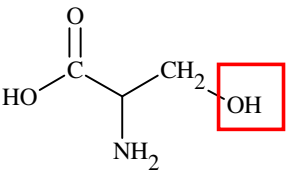
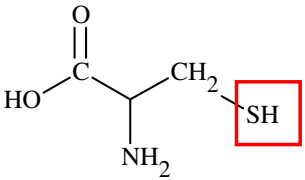
Fig. 3.1. The examples of amino acids names

There are several classifications of amino acids:

1. **Based on polarity,** amino acids are classified into four groups as follows:

- *non-polar (hydrophobic) amino acids* - Ala, Val, Lei, Ile, Met, Pro, Phe, Trp, Gly;
- *polar (hydrophilic) amino acids with no charge* - Ser, Thr, Cys, Tyr, Asn, Gln;
- *polar (hydrophilic) amino acids with positive charge – basic amino acids* Lys, Arg, His, which are *positively charged* at the physiological pH;
- *polar (hydrophilic) amino acids with negative charge – acidic amino acids* Asp, Glu, which are *negatively charged* at the physiological pH.

Polar and non-polar amino acids differ by occurrence of hydrophilic or hydrophobic functional groups in their radicals (R) (Fig. 3.2).

<table>
<tr><td colspan="3">Polar functional groups in the amino acids radicals (R): amino acids ALWAYS belong to POLAR when they are present in the radical</td></tr>
<tr><td>–OH</td><td>Tyrosine
  </td><td>Serine
  </td></tr>
<tr><td>–SH</td><td colspan="2"> 
Cysteine</td></tr>
</table>

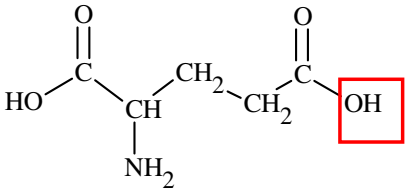
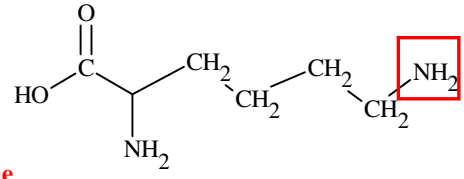
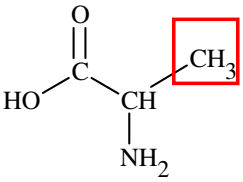
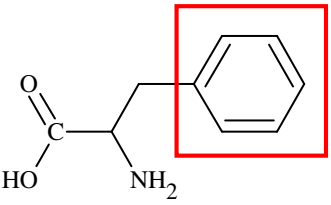
–COOH (–COO^-)	 Glutamic acid
–NH_2 (–NH_3^+)	 Lysine
Non-polar functional groups in the amino acid radical (R): when they are present in the radical and in the absence of the polar substituents, the amino acids belong to the NON POLAR ones	
–CH_3 and fatty acid chains	 Alanine
Cyclic structures	 Phenylalanine

Fig. 3.2. Polar and non-polar functional groups in the amino acids radicals (R)

Examples of polar and non polar amino acids are given on Fig. 3.3 and Fig. 3.4.

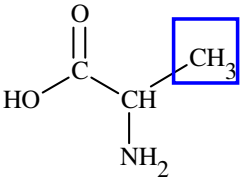
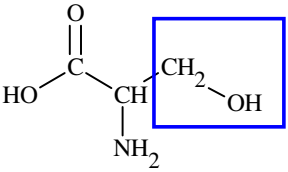
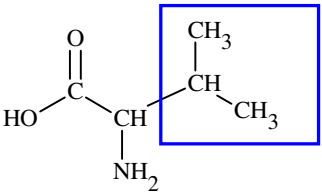
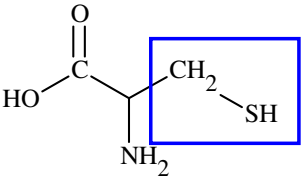
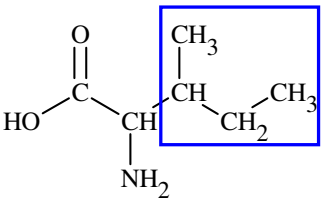
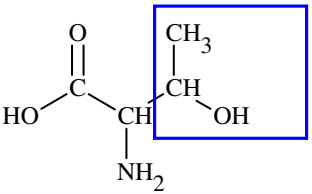
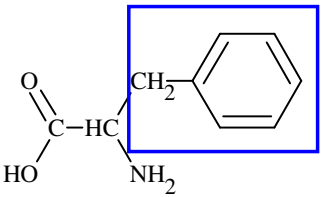
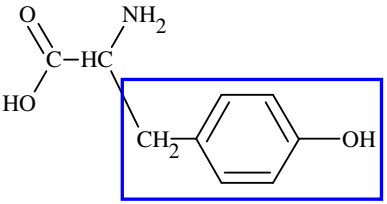
Non polar amino acids	Polar amino acids
Alanine 	**Serine** 
Valine 	**Cysteine** 
Isoleucine 	**Threonine** 
Phenylalanine 	**Tyrosine** 

Fig. 3.3. Examples of polar and non-polar amino acids (a quadrilateral is a radical of amino acids)

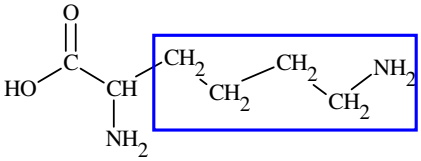
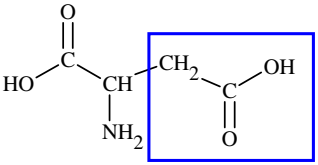
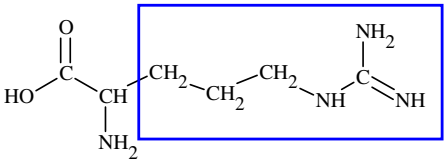
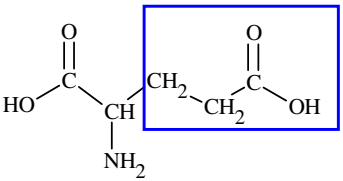
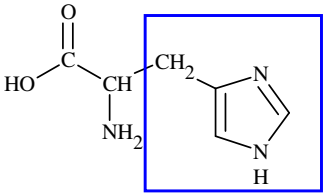
Positively charged amino acids	Negatively charged amino acids
Lysine 	**Aspartic acid** 
Arginine 	**Glutamic acid** 
Histidine 	

Fig. 3.4. Examples of polar charged amino acids (a quadrilateral is a radical of amino acid; –COOH and $–NH_2$ groups in radicals of these amino acids exist as $–COO^-$ and $–N^+H_3$ groups, respectively)

2. According to number of amino- and carboxylic groups (Fig. 3.5) amino acids are divided into:

- *monoamino monocarboxylic acids* (ex. Ser, Gli);
- *monoamino dicarboxylic acids (acidic amino acids* Glu, Asp) and their amides (Gln, Asn)
- *diamino monocarboxylic acids (basic amino acids* Lys, Arg);

• *diamino dicarboxylic acids* (*cystine* – the amino acid derived from Cys).

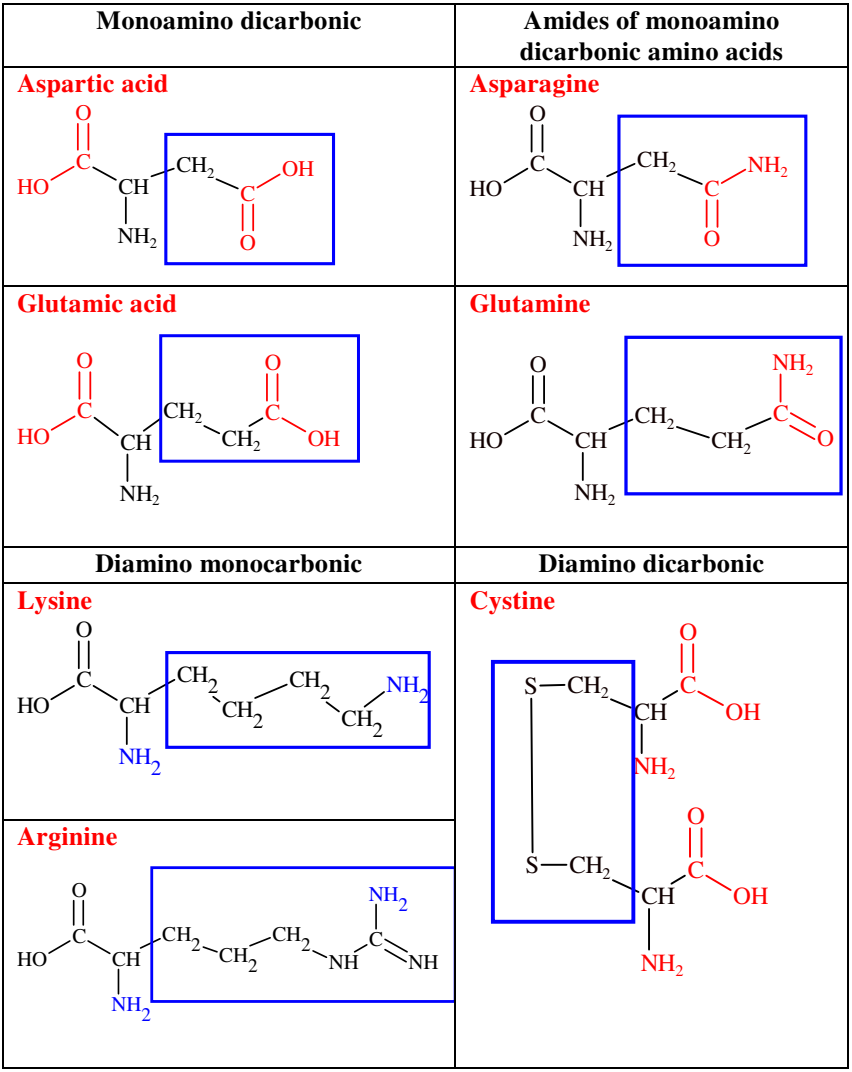


Fig. 3.5. Examples of monoamino dicarbonic, diamino monocarbonic, diamino dicarbonic amino acids and amides of monoamino dicarbonic amino acids

3. **Based on the presence of cyclic structures:**

- *cyclic* (*aromatic* – Phe, Tyr; *heterocyclic* – Trp, His, Pro);
- *acyclic.*

4. **Nutritional classification:**

- *essential amino acids:* 8 amino acids that can't be formed in the body and so, it is essential to be taken in diet. Their deficiency affects growth, health and protein synthesis - **незамінні**: (Val, Leu, Ile, Thr, Lys, Met, Phe, Trp);
- *semiessential amino acids (conditional amino acids):* these are formed in the body but not in sufficient amount for body requirements especially in children) – His, Arg;
- *non essential amino acids.*

There is such short formula for remembering of essential and semiessential amino acids:

Villa HM = Ten Thousands Pound,

Where V – valine, i – isoleucine, l – lysine, l – leucine, A – arginine*, H – histidine*, M – methionine, T – tryptophan, Th – threonine, P – phenylalanine (* - arginine and histidine are semiessential).

5. Metabolic classification – according to metabolic or degradation products of amino acids they may be:

- *ketogenic amino acids* which give ketone bodies. *Lysine* and *leucine* are the only pure ketogenic amino acids.
- *mixed ketogenic and glucogenic amino acids* which give both ketonbodies and glucose. These are: *isoleucine, phenyl alanine, tyrosine and tryptophan.*
- *glucogenic amino acids* which give glucose. They include *the rest of amino acids*. These amino

acids in their catabolism yield products that enter in glycogen and glucose formation.

3.2. Amphoteric and stereochemical properties of amino acids. Chiral carbon atom

Amino acids have both basic and acidic groups and so can act as a base or acid (have an **amphoteric properties).**

For all amino acids *except for glycine*, **α-carbon atom (C*) is chiral because of it bonds to four different groups** (Fig. 3.6)**:**

- a carboxyl group (1),
- an amino group (2),
- a R-group (different radicals) (3),
- a hydrogen atom (4).

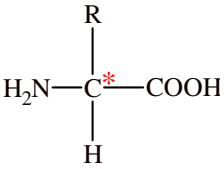


Fig. 3.6. Existence of chiral carbon atom in amino acids molecules

The presence of a chiral center determines **the stereochemical properties of amino acids:**

- *their absolute configuration;*
- *optical activity.*

According to **absolute configuration** amino acids are divided into *amino acids of D- and L-series* (by analogy with the absolute configuration of D- and L-glyceraldehyde) (Fig. 1.7). These stereoisomers called **enanthiomers** are identical by chemical and physical properties, but they rotate the plane of polarized light to the right (+) or to the left (-). The natural proteins include only L-amino acids, and D-isomers have been found only in small peptides of bacteria cell walls or in some peptide antibiotics. While L-amino

acids are included in proteins during translation on the ribosome, D-amino acids found in some proteins are produced by post translational modifications.

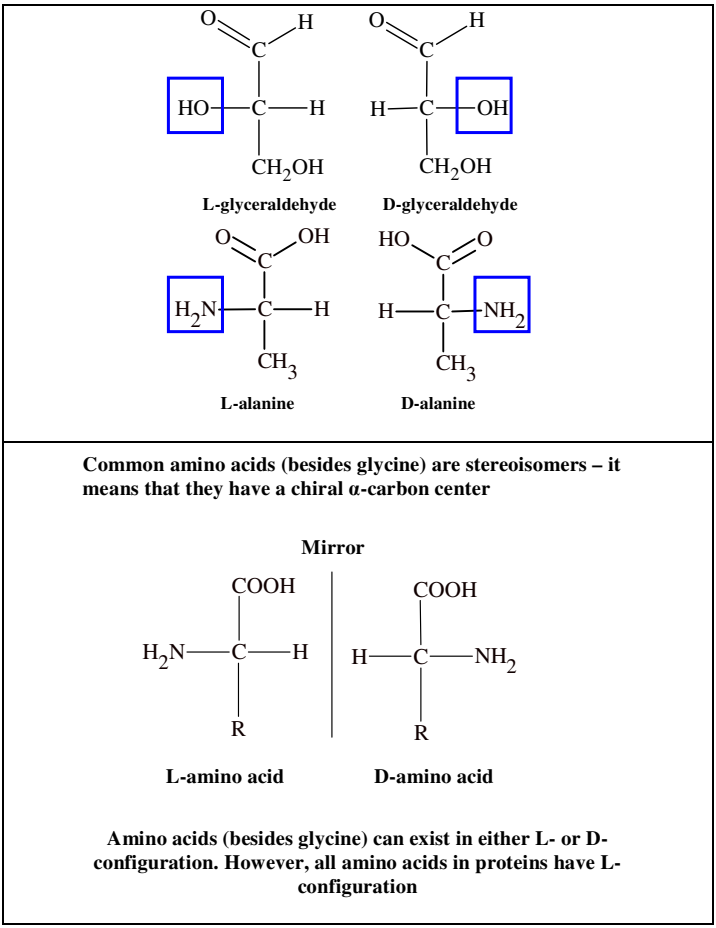
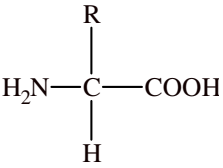
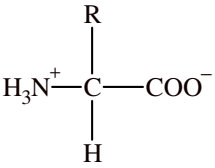


Fig. 3.7. Absolute configuration of amino acids: amino acids of D- and L-series. Carbons are lined up vertically, with the chiral atom in the center and carboxyl group from above. When α-amino group is on the left site of the vertical line, amino acid is in L-form, and when α-amino group is on the right site of the vertical line, amino acid is in D-form.

3.3. Isoelectric point of amino acid. Titration curves.

Neutral amino acids (monobasic, monocarboxylic) exist in aqueous solution as **Zwitter ion** – i.e. contain both positive and negative charge. Zwitter ion is electrically neutral and can't migrate into electric field:

	
The common formula of amino acids	**Amino acid in Zwitter ion form**

In solutions, amino acids exist in the form of anions, cations, Zwitter ions and their mixtures. The ratio of these forms depends on the pH of the solution (Fig. 3.8). *In acidic medium* the most of amino acids are positively charged, so they behave as a base (proton acceptor). *In alkaline medium* the most of amino acids are negatively charged, so they behave as an acid (proton donor). The characteristic pH at which the electric charge of amino acid is zero is called the **isoelectric point** of this amino acid, or "**pI**".

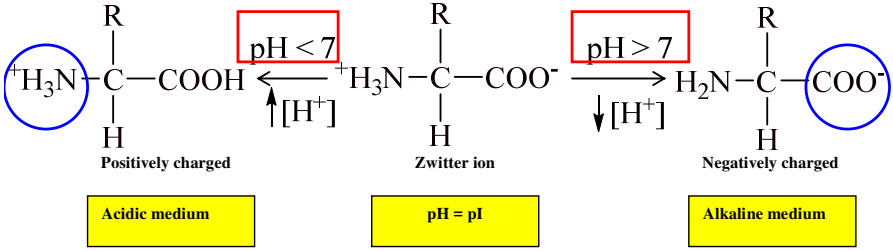


Fig. 3.8. The amino acid electric charge in solutes.

The isoelectric points of different amino acids are given on Fig. 3.9.

Neutral amino acids - pI = 5-6 (amino acids with non polar and polar non charged side chains)	
Amino acid	pI
Alanine	6.01
Asparagine	5.41
Cysteine	5.07
Glutamine	5.65
Glycine	5.97
Isoleucine	6.02
Leucine	6.02
Methionine	5.74
Phenylalanine	5.48
Proline	6.48
Serine	5.68
Threonine	5.87
Tyrosine	5.66
Tryptophan	5.89
Valine	5.97
Acidic amino acids = pI ~3.	
Aspartic acid	2.77
Glutamic acid	3.22
Basic amino acids - pI ~9.	
Arginine	10.76
Histidine	7.59
Lysine	9.74

Fig. 3.8. The isoelectric points of different amino acids

Titration curve of amino acid (the dependence of pH of amino acid solution from volume of titrant (NaOH) that has been added) is used to determine:

- *amino acid isoelectric point;*
- *amino acid buffer properties.*

3.3.1. The typical titration curve of amino acids with uncharged radical

The typical titration curve of amino acids with uncharged radical (for example, alanine) is given on Fig. 3.9.

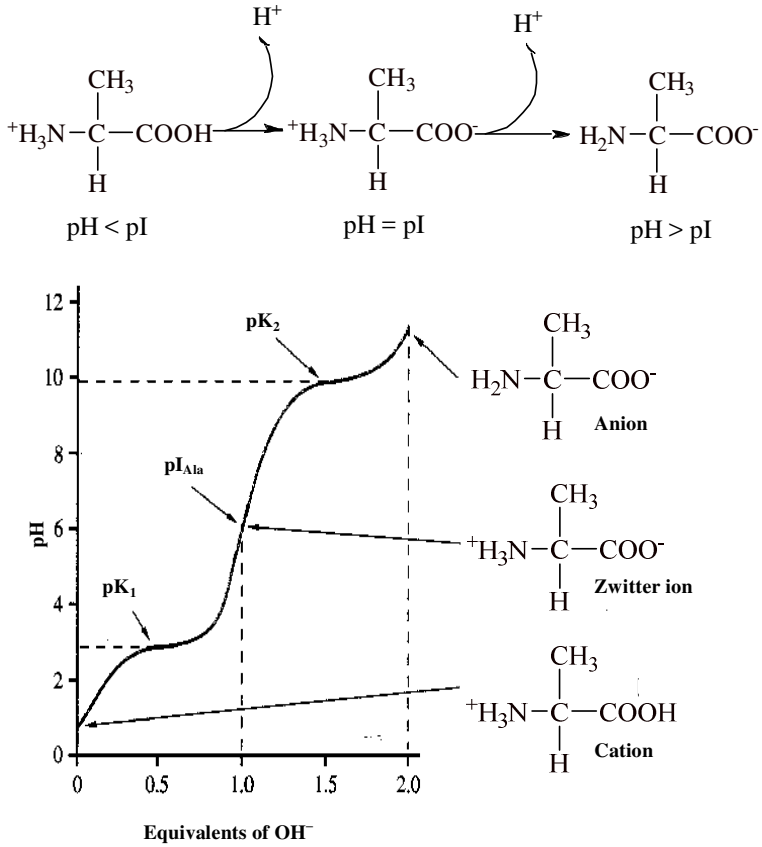


Fig. 3.9. The titration curve of alanine

The next formula is used to determine pI of amino acids with uncharged radical:

$$pI = (pK_1 + pK_2)/2 = 6{,}02 \text{ (}\textit{for alanine}\text{)},$$

where pK_1 is the dissociation constant of the COOH-group, and pK_2 is the dissociation constant of the amino group.

At pH<pI, alanine is charged positively, moves to the cathode, and at pH>pI it charged negatively and moves to the anode. At pH=pK_1 ½ of alanin molecules are in protonated form and ½ - in the Zwitter-ion form; at pH=pK_2 ½ of alanin molecules are in Zwitter-ion form and ½ - in the anion form.

3.3.2. The titration curves of amino acids with charged radical

The titration curve of histidine, glutaminic acid and lysine is given on Fig. 3.10 – 3.12.

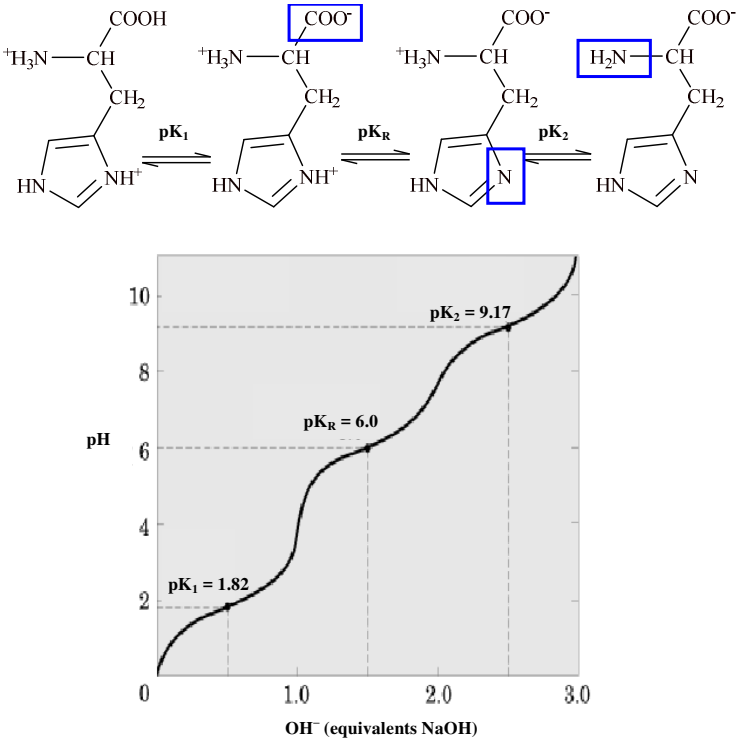


Fig. 3.10. The titration curve of histidine

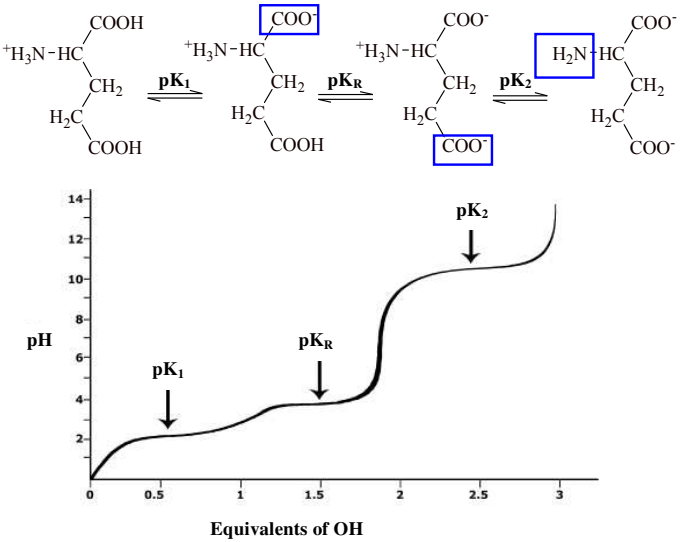


Fig. 3.11. The titration curve of gluramic acid

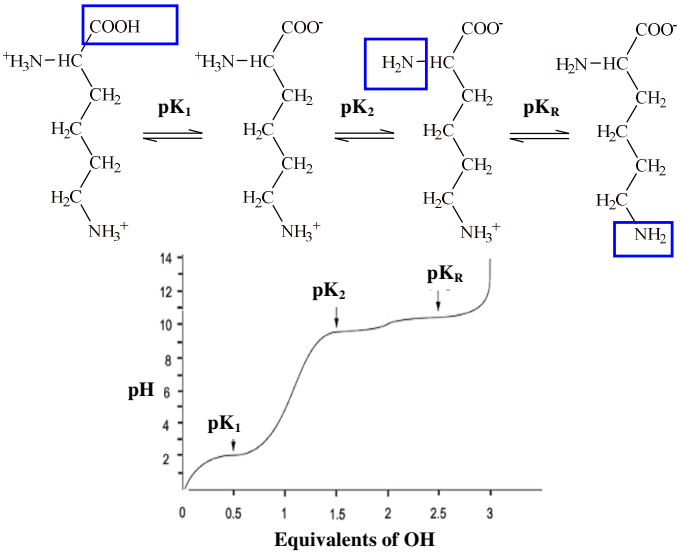


Fig. 3.12. The titration curve of lysine

The next formulas are used to determine pI of acidic amino acids:

$$\mathbf{pI = ½ (pK_1 + pK_R)}$$

and pI of **basic amino acids:**

$$\mathbf{pI = ½ (pK_R + pK_2),}$$

where pK_1 is the dissociation constant of the COOH-group; pK_2 – dissociation constant of amino group; pK_R – dissociation constant of radical R).

Therefore:

- **pI for glutamic acid:**

$$pI = (pK_1 + pK_R)/2 = (2,19 + 4,25)/2 = 3,22$$

- **pI for histidine:**

$$pI = (pK_R + pK_2)/2 = (6 + 9,17)/2 = 15,17/2 = 7,6$$

- **pI for lysine:**

$$pI = (pK_R + pK_2)/2 = 9,74$$

3.4. Chemical properties of amino acids and some methods for amino acids determination and separation

Chemical properties of amino acids are explained by the presence of –COOH and $–NH_2$ groups in the amino acid, as well as the presence of different functional groups in radical R.

- **Reactions on –COOH-group:**
  - they can react with bases that results in salts formation;
  - with alcohols → esters formation;
  - with amines → amides generation;

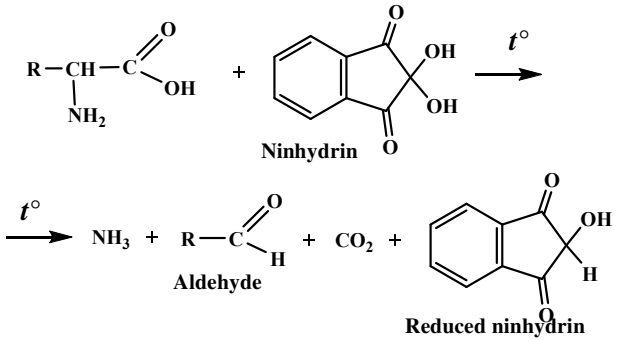
- decarboxylation with amines formation.

- **Reactions on –NH_2-group:**
  - they can react with acid that results in salt formation;
  - desamination with keto acids formation;
  - ninhydrinic reaction.

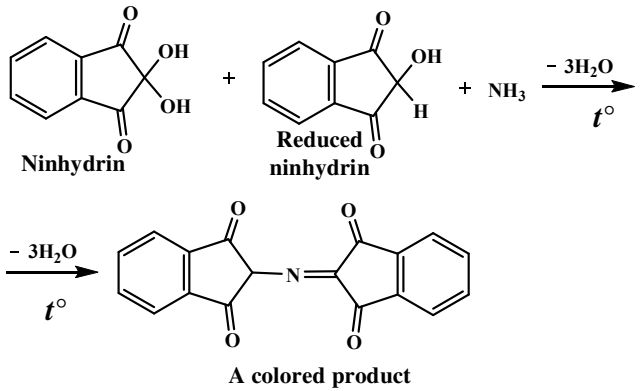
- **Reactions on the functional groups in radical (R) –** *on cyclic structures, SH-groups* etc.).

One of the reactions that are widely used *for qualitative detection of amino acids in solutions* is **ninhydrinic reaction.** The principle of reaction is that amines (including α-amino acids) react with ninhydrin to give a coloured product:

1)



2)



The α-amino acids typically give a blue-purple product, whereas proline, a secondary amine, gives a yellow-orange product.

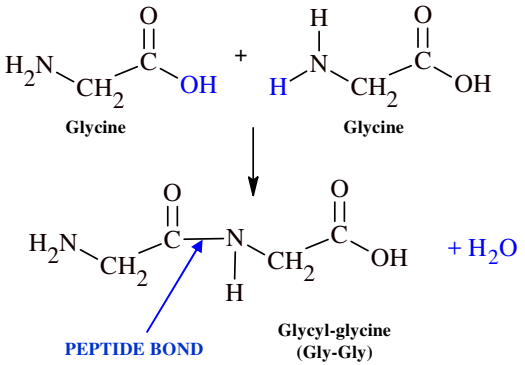
Sanger's and **Edman reactions** (with *dinitrofluorobenzene* and *phenyl isothiocyanate*, respectively) are used as *qualitative reactions on the N-terminal amino acid in the polypeptides* to determine of the polypeptide amino acid sequence.

Chromatography is a method used for *amino acids separation.* There are next variants of chromatography methods:

- **for the purpose:**
  - *analytical,*
  - *preparative,*
  - *to physical and chemical research;*
- **by the type of stationary phase:**
  - *columns chromatography*
  - *a thin layer chromatography;*
- **by the type of mobile phase:**
  - *a gas chromatography,*
  - *a liquid chromatography;*
- **by the type of sorbents that are used as a stationary phase, and by the nature of the interactions that cause the redistribution of components** between moving and stationary phases:
  - *adsorption chromatography*
  - *partitial chromatography,*
  - *precipitation chromatography*
  - *ion exchange chromatography*
  - *affinity chromatography,*
  - *size-exclusion, or gel filtration chromatography.*

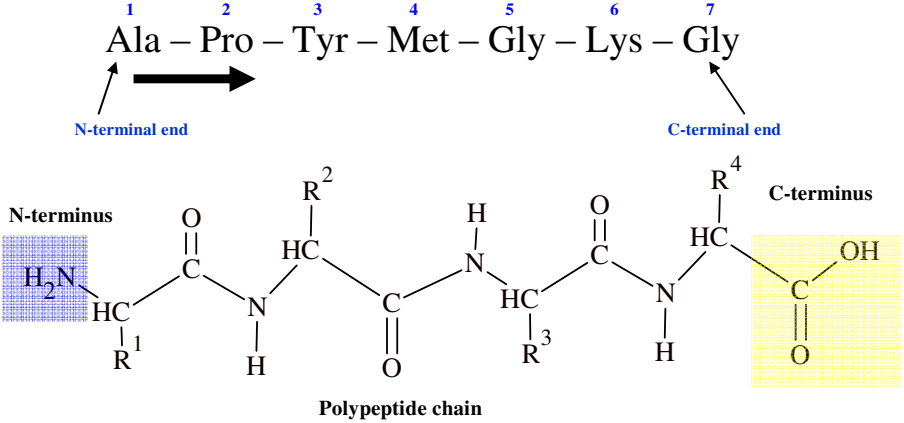
3.5. Peptide bond formation. Peptides.

Peptide bond is a covalent linkage between α-carboxyl group of one amino acid and α-amino group of another amino acid:



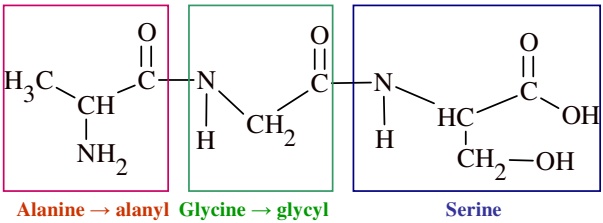
If peptide is formed by 2 amino acids – this is a *dipeptide*; 3 – *tripeptide*, 4 – *tetrapeptide*, etc. If peptide has 40-50 amino acids - this is a *polypeptide*.

The one end of the peptide chain has *a free amino group* (therefore it is **N-terminal end**), and the other is *a free carboxyl group* (**C-terminal end**):



Peptides are named from the N-terminus**.** The ending **"-ine"** in the name of each amino acid except of the last one is changed to **"-**

yl"; the name of the last amino acid on the C-terminal end does not change:



3.6. Proteins. The levels of protein molecules organization

Proteins are biopolymers, the monomers of which are amino acids, linked together by peptide bonds. Schematic representation of the levels of protein molecules organization is given on Fig. 3.12.

3.6.1. Primary protein structure

The **primary structure of a protein is its unique sequence of amino acids**. It is represented by polypeptide chain, in which amino acids are linked together by **peptide bonds.**

There are next peculiarities of peptide bond:

- all atoms of peptide bond are located in one plane;
- the length of the peptide bond – 0,132 nm - has an intermediate position between the ordinary single C–N link (0,147 nm) and double bond (e.g. C=O – 0,123 nm); therefore the peptide bond has a partial double-bond character and is shorter than a single bond (Fig. 3.14); because of this, the peptide bond is rigid, and the area located near it (CHR) is mobile and can rotate around the bond;

- the carbonyl oxygen and the amide hydrogen are in *trans*-positions;
- atoms H and O can form hydrogen bonds with other functional groups.

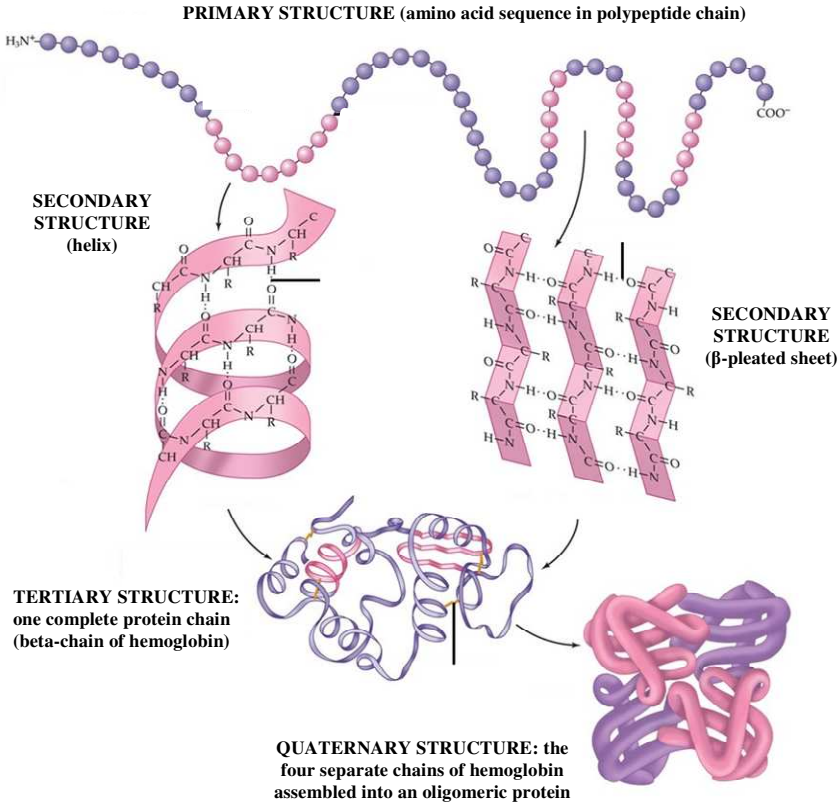


Fig. 3.13. Schematic representation of the levels of protein molecules organization

There are two dihedral angles that characterize rotation around bonds (Fig. 3.14)**:**

- the angle of rotation about the bond between the nitrogen and the α-carbon atoms is called **phi (φ)**;
- the angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms is called **psi (ψ)**.

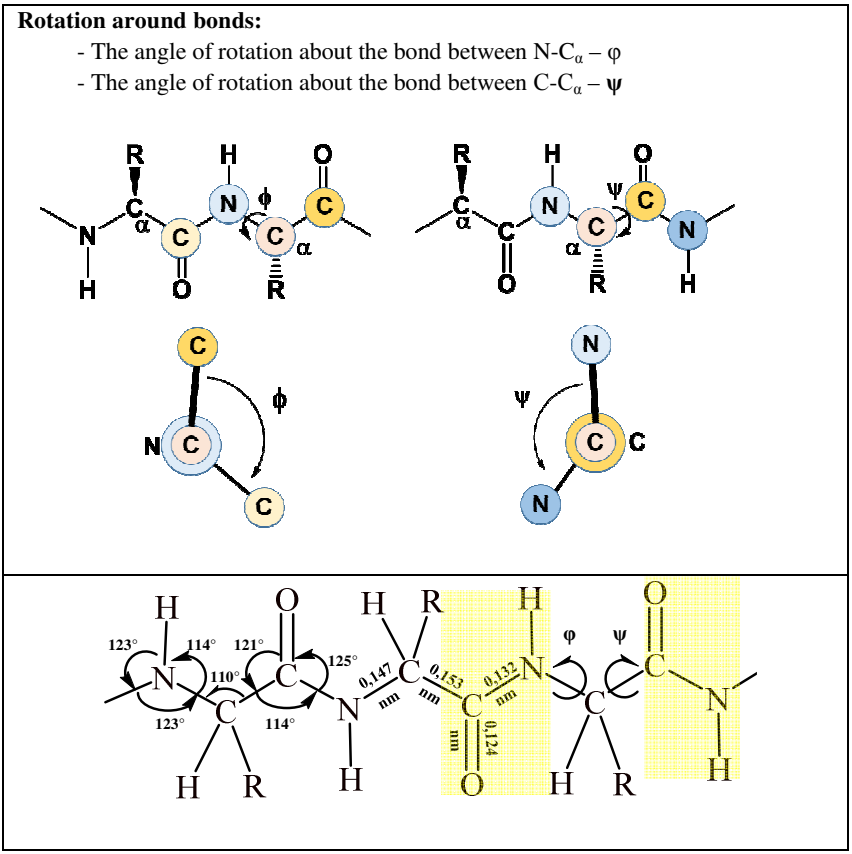


Fig. 3.14. The peculiarities of peptide bond

The precise primary structure of a protein is determined by inherited genetic information. Even a slight change in primary structure can affect a protein's conformation and ability to function. In individuals with sickle cell disease, abnormal oxygen-carrying protein hemoglobin is formed because of a single amino acid substitution. These abnormal hemoglobins crystallize, deforming the red blood cells and leading to clogs in tiny blood vessels.

3.6.2. The secondary structure of proteins: types

The folding of the polypeptide chain into specific coiled structure held together by H bonds is called **secondary structure of protein.** The main types of secondary structure are the next:

- *alpha-helix;*
- *beta-pleated sheet (β-structure);*
- *collagen helix.*

The type of secondary structure is determined by the primary structure of the protein (amino acid sequence). α-helixes and β-structures are detected by spectropolarimetry and UV spectrophotometry.

Alpha-helix is characterized by the next parameters:

- **right-handed turn**
- **diameter** – 1,05 nm;
- **pitch** (the vertical distance between consecutive turns of the helix) is 0,54 nm (1 turn);
- **amino acids per turn** – 3.6;
- **1 amino acid residue** is 0,15 nm;
- **identity period** - 5 turns (or 18 amino acids residues);
- **φ =** -57°, **ψ =** -47°

This type of secondary structure is stabilized by **hydrogen bonds** (the NH group of an amino acid forms this bond with the C=O group of the amino acid *four* residues earlier) and sometimes – by **disulfide bonds** formed between cysteine residues (Fig. 3.15)

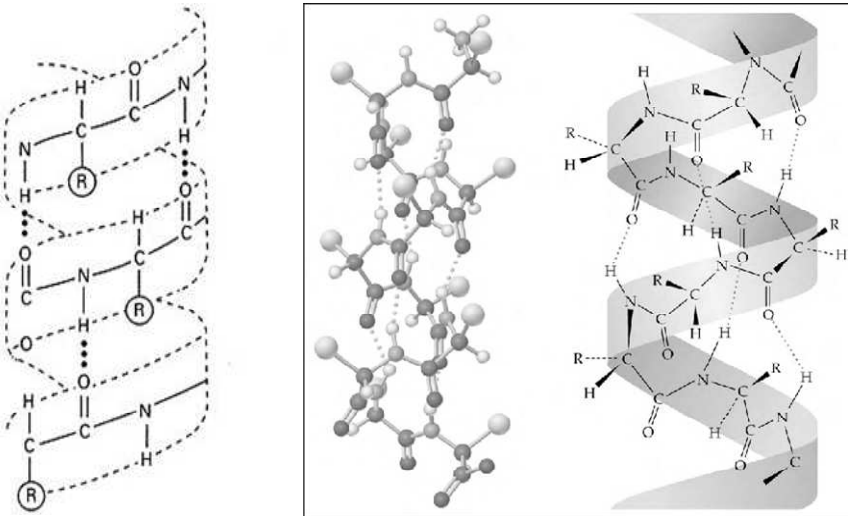
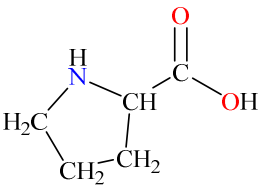


Fig. 3.15. The hydrogen bonds in α-helix stabilization

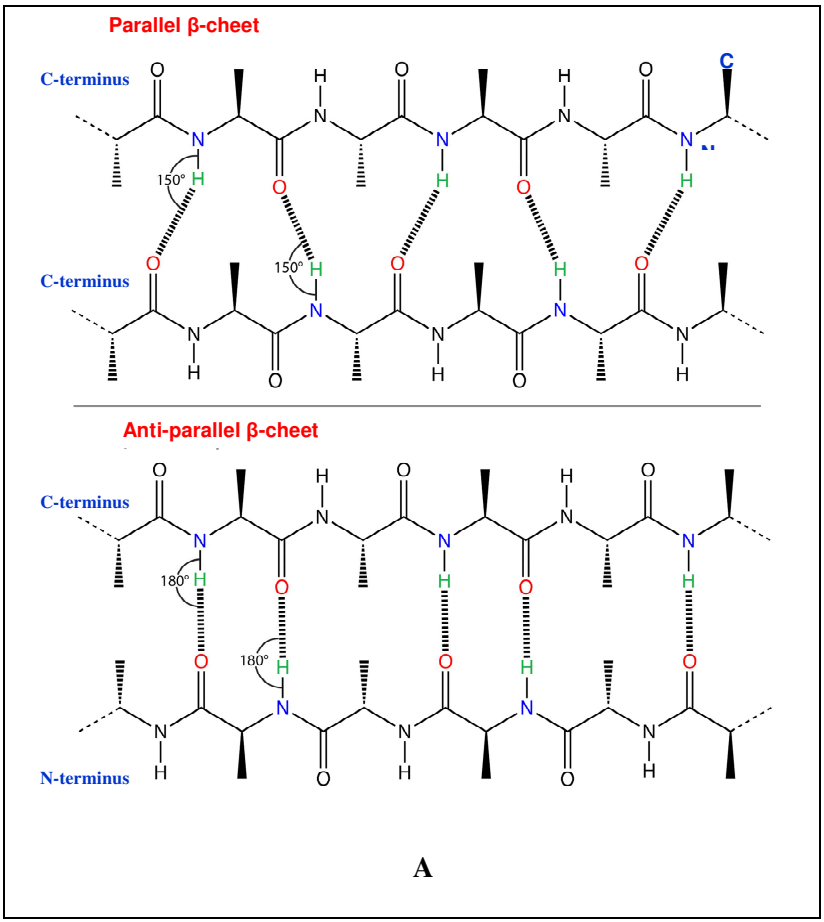
There are only right-handed helixes in natural proteins since they contain only L-amino acids.

Different amino-acid sequences have different propensities for forming α-helical structure. **Methionine, alanine, leucine, glutamate,** and **lysine** (**"MALEK"** in the amino-acid 1-letter codes) all have *especially high helix-forming propensities*. There are also amino acids that *prevent the formation of alpha spirals*. For example, **proline** either breaks or kinks a helix, both because it cannot donate an amide hydrogen bond (having no amide hydrogen), and also because the nitrogen atom is contained in the ring, which makes it impossible to rotate around N – C bond:



However, proline is often seen as the first residue of a helix. **Many residues of the same charged amino acids located near** (at pH = 7), and also **Asp, Ser, Tre, Leu residues localized near each other** - due to the large size and form of radicals (dimentional discrepancy) – prevent the formation of alpha helixes too.

Beta-pleated sheet arises due to the formation of **hydrogen bonds** between parallel or antiparallel segments of one polypeptide chain called *beta strands* (depending on whether the strand directions (N-terminus to C-terminus) are the same or opposite) (Fig. 3.16).



A

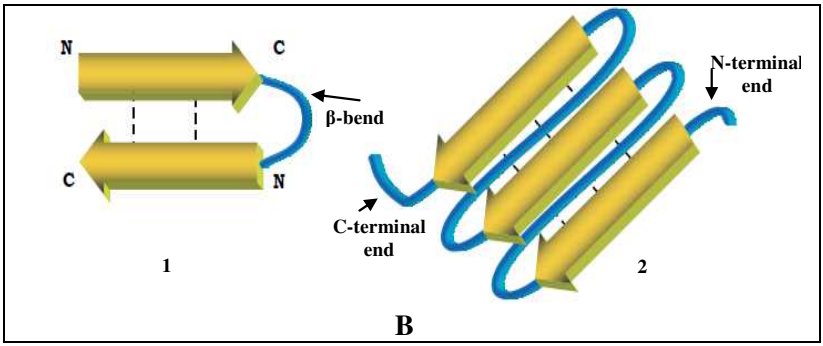


Fig. 3.16. A - β-sheet atructure; B - antiparallel (1) and parallel (2) β-sheets. β-structures are indicated by wide arrows.

In addition, it can be stabilized by **disulfide bonds.** The anti-parallel ß-sheet is more stable due to the more well-aligned hydrogen bonds.

This structure is dominated by **amino acids with small radicals (Gly, Ala)**. Such structures form mainly fibrillar proteins (β-keratin, fibroin), superoxide dismutase and carboxypeptidase.

β-sheet is characterized by the next parameters:

- **the distance between the segments of the chain** – 0,95 nm;
- **periods of identity along the chain**= 0,70 nm (*parallel β-strand*) and 0,65 nm (*anti-parallel β-strand*);
- most layers contain no more than **6 β-strands with 6 amino acid residues in the length of each**;
- **dimensions of this layer:**
  - **width** = 2,5 nm;
  - **length** = 2,0 nm;
- **dihedral angles for β-sheet with anti-parallel β-strands:** $\varphi = -139°$; $\psi = +135°$;
- **dihedral angles for β-sheet with parallel β-strands:** $\varphi = -119°$; $\psi = +113°$.

In the same protein, both α-helical regions and β-structures, as well as areas with a disordered (chaotic) conformation may occur. For example, *chymotrypsin* consists of α-helix (14% of polypeptide chain); β-sheet (45%) and chaotic conformation (61%) (Fig. 3.17); *myoglobin* and *tropomyosin* have 80% and 100% α-helix, respectively, whereas all *fibroin* is in β-sheet structure.

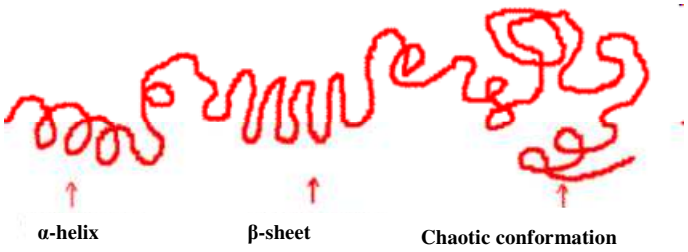


Fig. 3.17. Percentage contain of different types secondary structures in chymotrypsin molecule

Collagen helix is also called **α-chain (NOT α-HELIX!!!)** and has the next particularities and parameters:

- **left-handed helical structure;**
- **amino acids per turn** – 3,3 amino acid residues;
- **is rich in proline** (10%);
- **each third amino acid is glycine** (Fig. 3.18);
- has **hydroxyproline** (10%) and **hydroxylysine** (1%).

These non-standard amino acids are formed after collagen is synthesized (*by post-translational modifications*) *from proline and lysine* by *prolyl- and lysyl-hydroxylase*, respectively. This reaction is vitamin C-dependent.

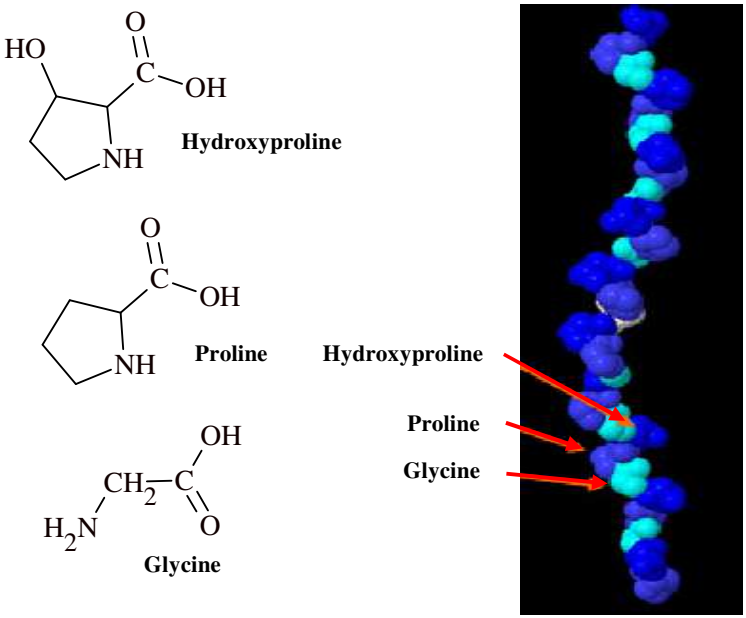


Fig. 3.18. Particularities and parameters of collagen helix

3.6.3. Super secondary structure

Structures that are the result of the interaction of several secondary structures are the examples of the **super secondary structure** of proteins. It is intermediate between secondary and tertiary structures of protein.

Collagen triple helix is an example of super secondary structure - three parallel left-handed collagen α-chains are twisted around each other to form right-handed triple-helical structure **– tropocollagen** (Fig. 3.19). These three α-chains are hydrogen bonded to each other. The hydrogen bond forms between peptide NH groups of glycine residues and CO groups of residues on the other chains.

The OH-group of hydroxyproline also participates in hydrogen bonding.

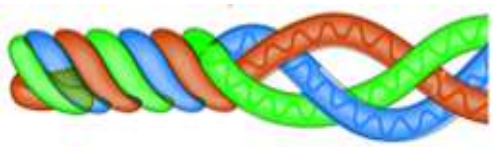


Fig. 3.19. Right-handed triple-helical structure of tropocollagen

Collagen triple helices form a rod-like fibril or sheet aggregate that is somewhat flexible, not extensible and can be very strong. Cross-linking increases its strength.

Other examples of super secondary structure of proteins are the interactions of two α-helix (**tropomyosin**) or three α-helix (**fibrinogen**) as well as the combinations of β-sheets alone or β-sheets and α-helixes (they are called **motifs**) (Fig. 3.20). Some of them have specific names:

- **"α-helix - turn - α-helix"** - two α-helix (one shorter, another longer), linked by the turning of the polypeptide chain (in many DNA-binding proteins);
- **β-β corner** (or **β corner**) - consists of two anti-parallel beta strands;
- **EF-hand** – two helices connected by a loop that contains residues to coordinate calcium ion;
- **"zinc finger“** - is a protein fragment containing about 20 amino acid residues, in which the zinc atom is usually linked with two cysteine residues and two residues of histidine. It is found in many DNA-binding proteins.

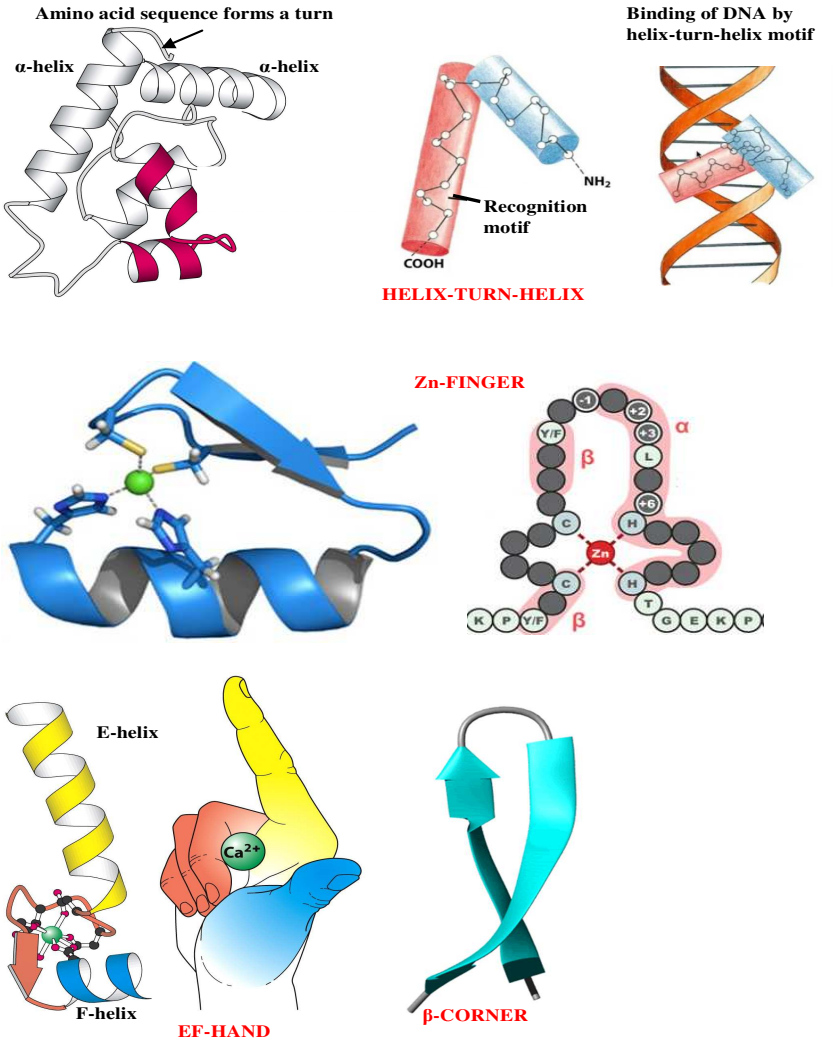


Fig. 3.20. Some examples of super secondary structures (motifs)

3.6.4. Tertiary Structure of proteins

The tertiary structure defines the specific overall 3-D shape of the one polypepride chain of protein (Fig. 3.21).

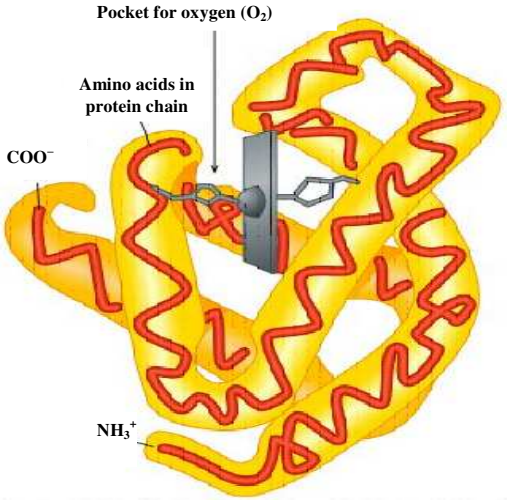


Fig. 3.21. The tertiary structure of myoglobin

The primary structure of a polypeptide chain determines its tertiary structure. The tertiary structure is based on various types of interactions between the side-chains of amino acids in the peptide chain (Fig. 3.22):

- **electrostatic, or salt bridge** (*a*) - between the oppositely charged radicals Asp, Glu, Arg, Liz;
- **hydrogen bond** (*b*) - (between the side radicals of the amino acids of different regions of the chain);
- **hydrophobic**, or **van der Waals (nonpolar) interaction** (*d*) - between hydrophobic radicals Leu, Ile, Phe, Trp);

- **disulfide (-S-S-) bond** (*c*) - covalent bond between the cysteine residues

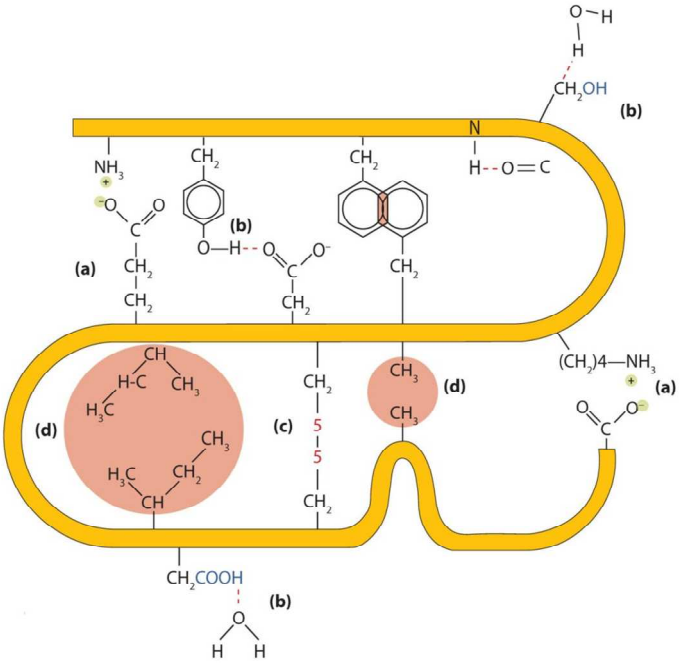


Fig. 3.22. Various types of interactions between the side-chains of amino acids in the peptide chain in stabilization of tertiary structure of proteins

The function of a protein depends on its tertiary structure. If the latter is disrupted, the protein loses its activity.

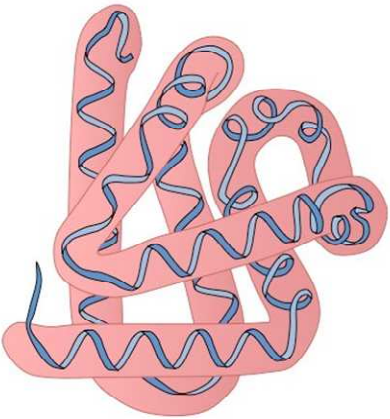
There are 2 types of tertiary structure:

- *globular (elipsoid);*
- *fibrous (rod-shaped* or *filiform):*

The type of tertiary structure is determined by the sequence of amino acids. *Globular proteins are mostly water soluble* – their hydrophobic side chains are buried in the interior, whereas hydrophilic groups are generally found on the surface of the molecule. *Fibrous proteins are insoluble in water* - they contain a

high proportion of nonpolar amino acid residues both in their interiors and on their surfaces.

Fibrous protein



Globular protein

Fibrous proteins consist of long fibers and are mainly structural proteins. They have *an axial ratio of more than 10* (***axial ratio*** = Length/*Width* of the protein molecule). They are fairly stable proteins. Their examples are:

- *keratin proteins* in hairs, wool, skin, and most cells. In its native state, it is present in the form of coiled polypeptide chains called *∝-keratin*. It can be stretched by denaturation forming *β-keratin*:

α-keratins are fibrous proteins that make hair, fur, nails and skin

β-keratins are fibrous proteins found in feathers and scales that are made up mostly of β-pleated sheets

- *myosin* is the major protein of muscles. During muscle relaxation it is called *∝-myosin* but during muscle contraction, it undergoes a change in its structure and it becomes *β-myosin*.

Globular proteins fold up into compact, spherical shapes; *their axial ratio is less than 10.* Their functions include biosynthesis, transport and metabolism. For example, *myoglobin* is a globular protein that stores oxygen (O_2) in the muscles. Its molecule has a single peptide chain that is mostly α-helix. O_2-binding pocket is formed by a heme group and specific amino acid side-chains that are brought into position by the tertiary structure (Fig. 3.21). Other examples of globular proteins are *albumins* and *globulins* of blood.

3.6.5. Domain structure of proteins

Domains are formed *within a single polypeptide chain.* Most domains consist of 40 to 200 amino acids, average diameter of ~25 Å. The core of a domain is built from combinations of super secondary structural elements (motifs). Folding of the peptide chain within a domain usually occurs independently of folding in other domains. Separate domains are functionally autonomous formations in the protein molecule; for example, they can catalyze the various stages of the complex catalytic process (for example, such enzymes with domain structure as *glyceraldehyde-3-phosphate dehydrogenase* and *phosphoglycerate kinase*). Individual domains often have specific function, i.e. binding of the dinucleotide NAD^+ by nucleotide-binding site in glyceraldehyde-3-phosphate dehydrogenase structure (Fig. 3.23).

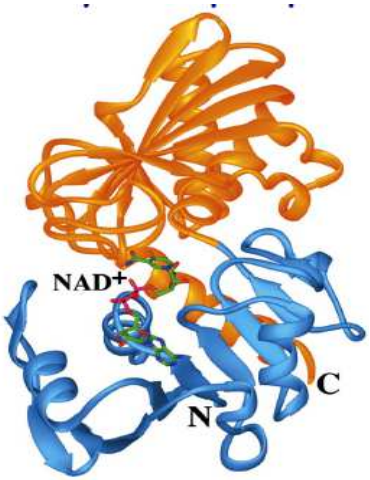


Fig. 3.23. Domain structure of glyceraldehyde-3-phosphate dehydrogenase – the enzyme that has 2 globular domains and dinucleotide binding site in N-term domain

Domains are bound together by a very limited number of peptide bonds, which are relatively easily torn under the action of proteolytic enzymes

3.6.6. Quaternary Structure of proteins

Many proteins are made up of multiple polypeptide chains, often referred to as **protein subunits** (s/u). These proteins are **oligomers**, and their subunits are called **protomers**. These subunits may be the same (as in a *homodimer*) or different (as in a *heterodimer*). Examples of proteins that have more then one subunits are:

- **hemoglobin A** - it is a heterotetramer (4 s/u – $2\alpha 2\beta$) (Fig. 3.24);
- **RNA polymerase** (E. coli) – $\alpha\alpha\beta\beta\sigma$ (5 s/u);
- **tobacco mosaic virus protein** - 2130 s/u (M.w. = 40 mln Da).

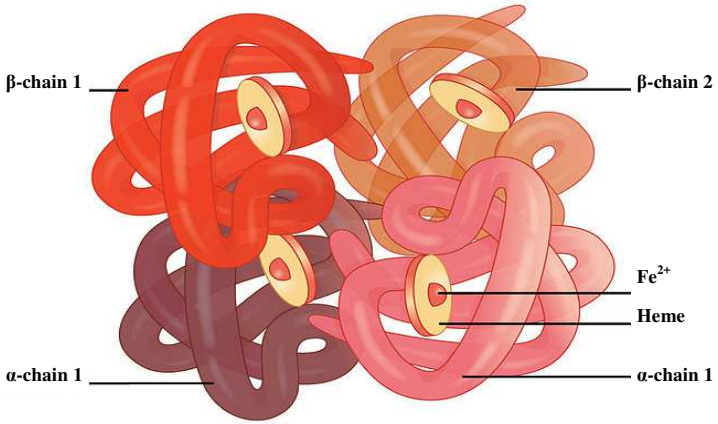


Fig. 3.24. The quaternary structure of hemoglobin A

The quaternary structure refers to how these protein subunits interact with each other and arrange themselves to form a larger aggregate protein complex. The quaternary structure of the protein is stabilized by:

- *hydrogen bonds,*
- *disulfide-bridges,*
- *salt bridges,*
- *hydrophobic interactions.*

In hetero-oligomeric proteins, subunits can perform various biochemical functions. All factors breaking hydrogen bonds, remove the charge of the protein molecule or reduce the disulfide bonds, destroy the quaternary structure (eg, change in pH, temperature, ionic strength) and violate the specific biological activity of the protein.

Phenomenon of cooperative interaction (or **concept of cooperativity**) is closely related with the quaternary structure of proteins. This phenomenon is that modification of one of the polypeptide chains (1 subunit) causes a change in the tertiary

structure of the remaining subunits, that results in changes of the remaining subunits properties.

For example, *myoglobin* - the small protein that contains 150 amino acids - has 1 subunit and is characterized by a very high affinity for oxygen. Therefore O_2 very quickly binds to the Fe^{2+} of myoglobin heme (Fig. 3.25).

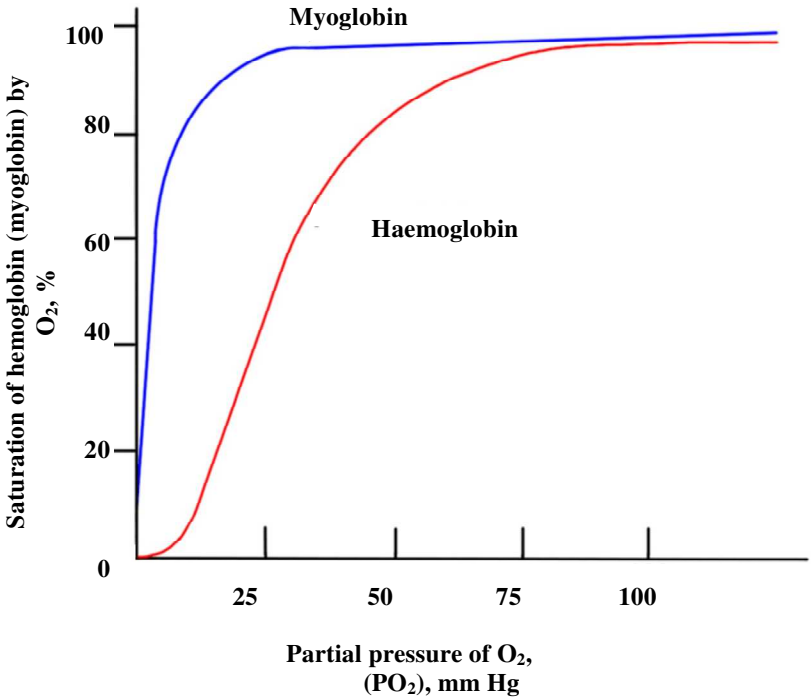


Fig. 3.25. The curve of saturation of hemoglobin (myoglobin) by O_2: a - hemoglobin (4 s/u, each one can attach 1 molecule of oxygen) - low affinity for oxygen; b - Myoglobin (1s/u) - high affinity for oxygen

In contrast, *hemoglobin* consists of four subunits that are similar to myoglobin. The affinity of hemoglobin for O_2 is lower than that of

myoglobin - so the binding of the first O_2 molecule to the first subunit of hemoglobin occurs very slowly and causes the changes in the structures of the other subunits of this protein. As a result, the second and third subunits get much higher affinity for O_2 than the first. Therefore a sigmoid curve is produced because of the effect of one substrate binding to one active site increasing the activity at the other active sites.

The phenomenon of cooperative interaction also explains the mechanisms of alosteric regulation of the enzymatic activity.

3.7. Methods of extraction and purification of proteins

3.7.1. Methods of extraction of proteins from cells or tissues in the dissolved state.

There are next methods to extract proteins from biological material:

- **homogenization of biological material** - a disruption of soft tissues such as liver, heart, brain, and muscle or other sample. Homogenization methods can be divided into five major cathegoties:
  - *mechanical,*
  - *ultrasonic,*
  - *freeze/thaw,*
  - *osmotic lysis,*
  - *detergent lysis;*
- **destruction of membranes** (for membrane proteins extraction; detergents are used).

After homogenization **crude extract** is formed that contains the protein or enzyme of our interest plus a mixture of other compounds.

Dialysis is commonly used for removing the salt and other low molecular weight compounds from the proteins. Dialysis is a process that separates molecules *according to size* through the use of *semi-permeable membranes* containing pores of less than macromolecular

dimensions. Pores in the membrane allow solvents, salts and small metabolites to diffuse across but block larger molecules.

The cell (tissue) extract is simply put in a bag made of a semi-permeable membrane that is permeable to small molecules (e.g. salts) and not permeable to proteins; this bag put into distilled water (Fig. 3.26). Low molecular compounds can cross the bag membrane and go into water whereas the proteins with high M. can't. So the proteins stay in bag whereas salts, protein fragments and other molecules smaller then the pores size pass through.

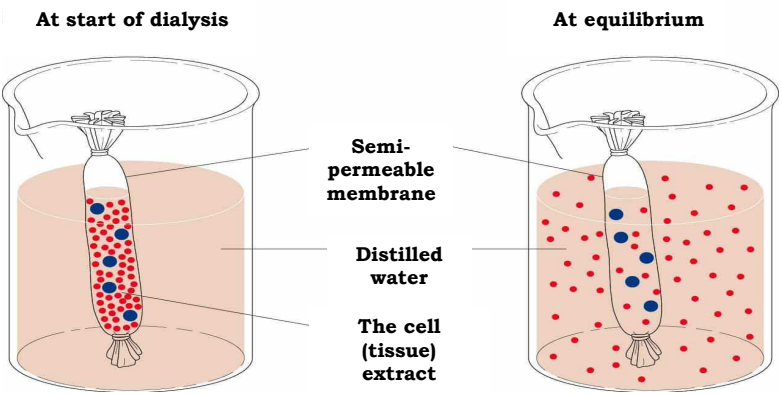


Fig. 3.26. The scheme of dialysis method

3.7.2. Methods of separating a mixture of proteins

There are next methods for protein mixture separation:

- *fractionation,*
- *ultracentrifugation,*
- *chromatography,*
- *eletrophoresis,*
- *isoelectric focusing.*

Fractionation is a fractional precipitation of proteins by solutions of neutral salts (ammonium sulfate $(NH_4)_2SO_4$, sodium chloride) with a gradual increase in concentration. For example, **salting out** is based on the addition of ammonium sulfate ($(NH_4)_2SO_4$) for differential precipitation of proteins. The mechanism of salting out is the interaction of anions (SO_4^{2-}) and cations of salt (Na^+, NH_4^+) with positively and negatively charged groups in radicals (R) of amino acids, resulting in a loss of charge and mutual repulsion of molecules. At the same time, the hydrate shell around the protein is sharply reduced (dehydration). The consequence is the adhesion of molecules and deposition.

Different proteins differ in amino acid composition, size and charge - it is possible to pick up such concentrations of salt, which precipitate less stable proteins, while others will still be soluble:

- *the proteins with a lower charge are precipitate first* (less salt is required to neutralize their charge and remove the hydrated shell). Therefore, globulins of blood plasma precipitate in 50% -saturation of the solution with neutral salts; albumins (more polar molecules) - at 100% saturation with salt;
- *the hydrated shell of protein is smaller, the less salt is needed to precipite it*. Globulins are large and heavy molecules with a small water shell, albumins - smaller and surrounded by a large one.

This process is reversed - after removal of salt (dialysis, dilution), the protein regains its natural properties.

Ultracentrifugation - particulate substances will reach the bottom of the tube (i.e. they form a *pellet*), and the remaining liquid is called the *supernate* or *supernatant liquid* (Fig. 3.27).

This method causes the separation of mixture (that may contain cells, organelles, proteins, nucleic acids, lipids, etc.) *on the basis of differences in its components weight.* For example, ultracentrifugation can separate:

- red cells from plasma of blood,
- nuclei from mitochondria in cell homogenates,
- one protein from another in complex mixtures

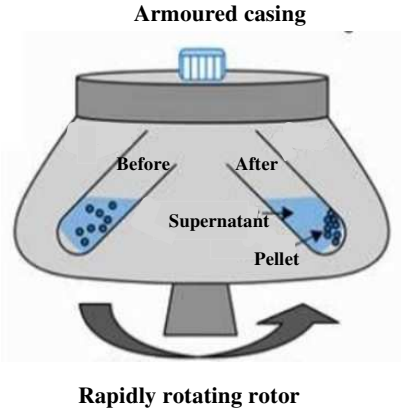


Fig. 3.27. The scheme of ultracentrifugation method

Chromatography – these methods were described in 3.5 subsection.

Eletrophoresis – the separation of *charged proteins* in an electric field. It is the movement of charged particles in an electric field towards the oppositely charged electrode (Fig. 3.28).

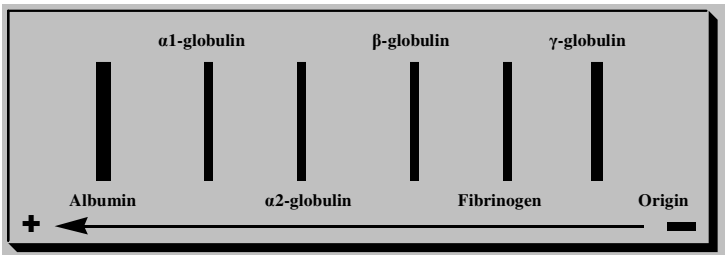


Fig. 3.28. The scheme of electrophoretic method of protein separation

Depending on size and shape of proteins electrophoresis can be carried out:

- *on paper,*
- *in polyacrylamide gel (PAAG),*
- *in starch gel.*

Isoelectric focusing is *an eletrophoresis in pH gradient***.** Gel contains a mixture of low molecular weight organic acids and bases

(ampholytes) with different pI value. After application of electric field each protein migrates until it reaches the pH corresponding to its pI (Fig. 3.29).

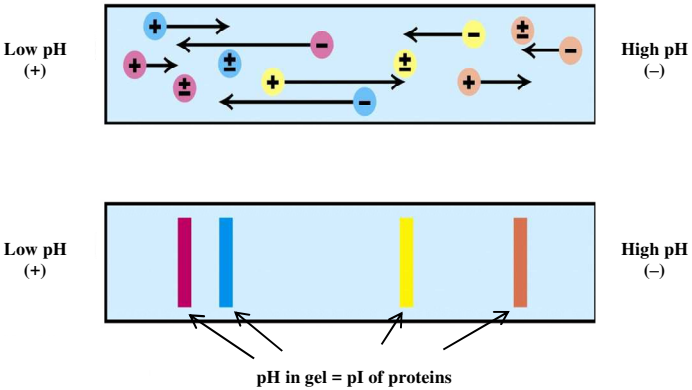


Fig. 3.29. The scheme of isoelectric focusing method

3.8. Physical and chemical properties of proteins

Molecular weight of proteins is 6000 - 1 million Da or more; it depends on the number of amino acids residues and the number of protein subunits. **The protein form** may be globular or fibrillar.

Total charge is conditioned by the presence of functional groups in the protein molecules that are capable of ionization at a certain pH of the medium. The total charge of the protein molecule depends on the ratio of the ionized anionic radicals of Glu and Asp and the cationic radicals of Liz, Arg and Gis. *In the state of the isoelectric point, it is equal to 0* - with the protein falling into the precipitate. In the neutral pH, most of the natural proteins are charged, as well as the cytoplasm, negatively. *At pH<pI the protein molecules become positive, at pH>pI - negative.*

Amphotericity – is explained by amphoteric properties of amino acids: proteins can interact with acids and alkalis, neutralizing them.

The methods of proteins precipitation from aqueous solutions by the action of trichloroacetic acid are based on this property. **Buffer properties** are associated with amphotericity (for example, albumins form protein buffer system of blood).

Hydrophilicity is due to the presence in protein composition of a large number of functional groups, capable for ionization (-COON, -OH, -SH, $-NH_2$). Dipoles of water are oriented around them and form a hydrated shell. About ¾ of proteins hydrophilicity is determined by the presence of peptide bonds: each of them binds 4 molecules of water, -COOH group - 3, -OH group - 2. 100 g of protein binds 50 g of water. Hydrophilicity is carefully connected with **solubility** that is due to the presence of amino acid residues with polar radicals; it depends on the size, shape, charge of the molecule, the properties of the solvents, pH of the medium, temperature, ionic strength of the solution.

Precipitation of proteins is their transition from solution to precipitate. Precipitation can be:

- **inverse** - when the solvent is added to the precipitate, the precipitate is dissolved. This type of precipitation is carried out with high concentrations of electrolyte salts (salting out – the more information about this method is in 3.8 subsection), alcohol and acetone (at low temperatures and short-lived) and used to isolate native proteins. *The mechanism of precipitation - "removal" of the hydrated shell and charge from the protein molecule.* It is easier to precipitate proteins at pH = pH.
- **non-inverse**, or **denaturation**, which *doesn't lead to the destruction of the primary structure of the protein, but is an irreversible process.*

The next factors can be used as denaturators:

- **heavy metal cations** - they destroy disulfide bonds;
- **detergents** – they break hydrophobic interactions and hydrogen bonds;
- **organic acids** (trichloroacetic, sulfosalicylic) – they break hydrogen and electrostatic bonds as well as hydrophobic interactions;
- **reducing agents** – they destroy disulfide bonds;

- **concentrated mineral acids and bases** – they break hydrogen and electrostatic bonds);
- **heating** – it destroys hydrogen bonds and hydrophobic interactions) (Fig. 3.30);
- **UV and ionizing radiation**.

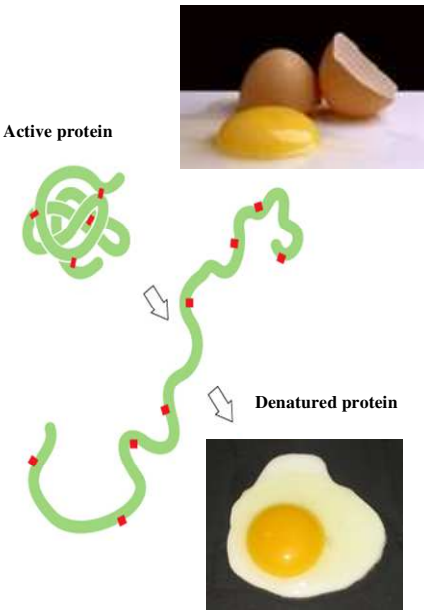


Fig. 3.30. The scheme of heating-caused protein denaturation

Denatured proteins are better exposed to proteolytic enzymes; they lose biological activity. Artificially synthesized and natural peptides are not able to denaturation. The proteins more easily denature in their isoelectric points.

Renaturation is possible at the early stages of denaturation under conditions of removal of denaturing agents.

The size of the protein molecules (1-100 nm) corresponds to the size of the particles of colloidal systems – therefore, they have **colloidal properties,** and the osmotic properties and electrokinetic phenomena are typical for proteins. Due to **osmotic properties**, the

protein molecules in the solution slowly diffuse and do not pass through the semi-permeable membrane. *Osmosis* is an one-sided diffusion of solvent molecules through a semi-permeable membrane into a more concentrated solution. *Osmotic pressure* - the pressure required to prevent osmosis. *Oncotic pressure* is a part of osmotic pressure, created by proteins (0,5% of all osmotic blood pressure, or 3,5 – 3,9 kPa). With decreasing protein content, oncotic pressure drops, and edema develops. **Electrokinetic phenomena** (electrophoresis, electroosmosis) – are due to the existence of electric charge on the surface of particles. *Electrophoresis* is the movement of particles in the constant electric current field to one of the electrodes; the velocity of motion depends on the charge of the molecules, the viscosity of the medium, the size of the molecules, and their molecular weight. Electrophoresis applies in medical diagnostics and to determine the purity of biochemical drugs. **The Tindal effect** - the ability to disperse light rays and to opalescence – is also typical for protein solution.

Antigenic properties of proteins. Proteins are antigens. When a protein from a tissue of animal of one species get into the blood of animal of another species, certain cells of the latter synthesize antibodies that enter the bloodstream, causing *immunological reactions*. Immunological reactions are characterized by weak non-covalent interactions between the *antibody* and a specific region of the foreign protein, the so-called *antigenic determinant*. If *precipitate* is formed - this is non inverse reaction of a precipitation. In the absence of a precipitate, the immunological reaction remains inversed.

3.9. Protein classification

Protein classification can be based on:

- **functions** (enzymes, transport proteins, structure proteins, defense proteins, contractile proteins, proteins-hormones, proteins-regulators of genome activity (proteins-transcription factors), toxic proteins, receptor proteins; proteins-inhibitors of enzymes;

proteins of the viral membrane; proteins with other functions) (tab. 3.1);

- **the form of the molecules** (*globular*, *fibrillar* – tab. 3.2);
- **the physical and chemical properties:**
  - **according to the electrochemical properties** - *acidic*, or *polyanionic* (for example, blood proteins); *basic*, or *polycationic* (histones), *neutral*;
  - **according to polarity** - *polar* or *hydrophilic*; *nonpolar*, or *hydrophobic*; *amphiphilic*);
- **the molecular weight** (*low molecular weight*, *high molecular weight*, etc.).
- **their localization in the cell** (cytosolic, membranous, mitochondrial, nuclear, lisosomal, ect);
- **their localization in an organism** (proteins of blood, liver, heart, etc.);
- **the possibility to adaptively regulate the amount of proteins**: proteins synthesized at constant speed (*constitutive*), and proteins whose synthesis may be enhanced by the action of environmental factors (*inducible*);
- **their chemical structure** (*simple, conjugate*). It is generally accepted but it doesn't take into account the presence of posttranslational modifications in some proteins (fatty acids residues, phosphoric acid, carbohydrates, methyl groups, etc.) (Tab. 3.3)

Table 3.1.
Protein classification based on their function

Type	Examples	Functions
Contractile proteins	Actin	Thin filaments in myofibril participated in muscle contraction
	Myosin	Thick filaments in myofibril participated in muscle contraction
	Dynein	Cilia and flagella protein, unvolved in protozoa movement
Enzymes	Hexokinase	Cytosolic protein involved in glucose phosphorylation
	Lactate dehydrogenase	Cytosolic protein involved in lactate dehydrogenation
	Cytochrome c	Mitochondrial protein involved in transferring of electrons
	DNA polymerase	Nuclear protein involved in replication and reparation of DNA
Hormones	Insulin	Regulates glucose metabolism
	Adrenocortico-tropic hormone	Regulates corticosteroids synthesis
	Growth hormone	Stimulates the growth of bone
Receptors	Ion channel receptors	Proteins of cell membrane or membranes of organelles participated in signal transduction
	G-protein linked receptors	
	Tyrosine kinase receptors	
Other regulatory proteins	Lac repressor	Genetic switch that turns off bacterial genes involved in lactose catabolism
	Transcriptional factors	Proteins that regulate the genes expression
Toxins	Clostridium botulinum toxin	Causes bacterial food poisoning
	Diphtheria toxin	Bacterial toxin
	Snake venom	Enzymes that hydrolyzed glycerophospholipids
	Ricin	Toxic protein from castor bean
	Gossypin	Toxic protein from cottonseed

Storage proteins	Ovalbumin	Egg-white protein
	Casein	A milk protein
	Ferritin	Iron storage in spleen
	Gliadin	Seed protein of wheat
	Zein	Seed protein of corn
Defensive proteins	Antibodies	Form complexes with foreign proteins
	Interferons	Proteins, produced by higher animals that interfere with viral replication
	Fibrinogen	Precursor of fibrin in blood clotting
	Thrombin	Component of blood clotting
Transport proteins	Hemoglobin	Transports O_2 in blood of vertebrates
	Hemocyanin	Transports O_2 in blood of some invertebrates
	Myoglobin	Transports O_2 in muscles
	Serum albumin	Transports fatty acids, bilirubine etc. in blood
	Ceruloplasmin	Transports coopers in blood
	Apoproteins	Protein components of lipoproteins that are participate in triacylglycerol, cholesterol transport
	Glucose permeases (GLUTs) ViraA	Proteins that take part in facilitate diffusion of glucose through plasmatic membrane (for example, in basolateral membrane of enterocyte)
Structural proteins	Viral coat proteins	Sheath around nucleic acid
	Some glycoproteins	Cell coats and walls
	Keratins	Fibrous proteins of hair, skin, nails, etc.
	Sclerotin	Exoskeletons of insects
	Collagens	Fibrous proteins of connective tissue and bones component, that

		serve as a scaffolding for support of tissues and organs
	Elastins	Fibrous proteins of connective tissue have rubber-like properties that allow them to stretch to several times their normal lenght

Table 3.2.
Examples of globular and fibrillar proteins

Group	Protein	Functions
Globular	hemoglobin	Transport proteins (transport of O_2)
	myoglobin	
	cytochrome c	Transport protein (electron transferring)
	lysozyme	Enzyme (bacterial wall hydrolysis)
	chymotrypsin	Enzyme (hydrolysis of proteins in gastrointestinal tract)
	ribonuclease	Enzyme (hydrolysis of RNA)
	lactate dehydrogenase	Enzyme (metabolism of carbohydrates)
	immunoglobulins	Defense proteins (antibodies)
	G-actin	Muscle protein participated in muscle contraction
Fibrillar	collagens	Structural proteins
	elastin	
	keratins	
	fibrin	
	myosin	Muscle protein participated in muscle contraction

Table 3.3.
Protein classification based on their chemical structure

Simple proteins	Conjugated proteins
Albumins	Phosphoproteins
Globulins	Glycoproteins
Protamines	Nucleoproteins
Histones	Lipoproteins
Glutalins	Chromoproteins
Gliadins (or prolamins)	Metalloproteins
Scleroproteins	

3.9.1. Simple proteins

Simple proteins are proteins which on hydrolysis produce amino acids only. Simple proteins are subdivided according to their physical properties, solubility, molecular weight and amino acid composition into subclasses (Tab. 3.3).

Albumins exist in both plant and animal kingdoms. They are present in serum (serum albumin), eggs (ovalbumin), milk (lactalbumin). They have a small M.w. (68 KDa), contain little glycine, have an acid properties and negative charge, are soluble in water and salt solution. They are precipitated by full (100%) saturation with ammonium sulfate and are coagulated by heat. Albumins can functioning as the transporting proteins for elements, vitamins, hormones, fatty acids, bilirubine, some medicals, other substances. They also keep blood osmotic pressure (*oncotic blood pressure* is created by albumins - it is about 80% of osmotic blood pressure).

Globulins are found in animals. They are bigger then albumins, aren't soluble in water, but are soluble in salt solution. They are more easily precipitated than albumins and this can be done by half (50%) saturation with ammonium sulfate. Thus half-saturation with ammonium sulfate can be used to separate globulins from albumin; this process is called salting out. Globulins are weakly acidic or neutral proteins that have more glycine. Some of globulins take part

in the transport of the other blood compounds, have defense function or are a proteases inhibitors; other are antibodies.

There are next globulins classes:

- **α1-globulins** of blood - some of them are a *protein parts of high-density lipoproteins*; other examples include *antitrypsin*, *antichymotrypsin*, *orozomukoid*;
- **α2-globulins** of blood - for example, *ceruloplasmin*, *α2-macroglobulin, haptoglobin, C-reactive protein.*
- **β-globulins** - *transferrin, prothrombin, a protein part of low density lipoproteins, complement components*.
- **γ-globulins** - they are *antibodies* (Ig G, A, M. D, E).

The index of total blood protein in normal condition is 60-85 g/l. The amount of albumin is 50-70% of blood proteins (40-50 g/l), α1-globulins - 3-6%, α2-globulins - 9-15%, β-globulins - 8-18% and γ-globulins - 8-17%. **Protein ratio** - the ratio of albumin to globulin content - normaly is about 1,2-2.

Protamines and histones are the main proteins of cell nucleus that are involved in the formation of the structure and in the functioning of the nuclear nucleoprotein complex. Both are polycationic proteins with a small molecular weight and a similar function.

Protamines are water-soluble and ammonia soluble proteins, non-coagulable. Their molecule is small (M. w. 4 – 12 κDa) and contains not more than 20 amino acid residues. They are strongly basic, due to the presence of large amount of basic amino acids specially arginine (80-90%). These proteins lack in both tyrosine and tryptophan. Protamines are present in sperms and ova, in plants - in pollen grains. Examples of protamines from fish are: *salmine* from salmon; *clupeine* from herring sperm (*Clupea*); *iridine* from rainbow trout; *thinnine* from tunafish (*Thunnus*); *stelline* from starry sturgeon (*Acipenser stellatus*); *scylliorhinine* from dogfish (*Scylliorhinus*)

Histones are basic proteins too. The most important amino acids entering in their structure (20-35%) are arginine, histidine and less

lysine, but the amount of their positive charge and basic properties is less than in protamines. They are usually present in a form of *nucleoproteins* (histones associated with DNA) as they participate in the structural organization of chromatin and can bind to DNA (because of DNA is charged negatively, and histones - positively).

There are next histone types that differ in the Arg and Lyz content:

- H1,
- H2a,
- H2b,
- H3,
- H4.

Histones are characterized by being soluble in water, dilute acids, and insoluble in dilute ammonia – the latter differs them from protamines.

Gliadins and glutelins are vegetable proteins, mainly from cereal seeds, form the bulk of gluten. *Gliadins* (also called *prolamins* due to the presence of a high percentage of the amino acids **prol**ine (10-14%) and glut**amine**). They are a group of plant storage proteins found in the seeds of cereal grains: wheat (*gliadin*), barley (*hordein*), rye (*secalin*), corn (*zein*), sorghum (*kafirin*). They are never present in animals.

They are insoluble in water, saline solutions, acids and alkalis, and generally soluble only in strong (70-80%) ethanol.

Glutelins are plant proteins too. They are soluble in weak alkalis and acids but not in water, alcohol or diluted salt solutions. They are also very rich in glutamic acid so they have acidic properties. Glutelins have very large molecular weight and are heat coagulable. Examples of this protein are *oryzenin* of rice and *glutelin* of wheat.

Scleroproteins are characterized by their extreme insolubility in water, dilute acids and the most common reagents. They are strong fibrous structural proteins that are never present in plants; they rich in sulfur containing amino acids and hence disulfide bonds. The main important examples of scleroproteins are elastins, collagens and keratins.

Elastins are present in particular in tendons and ligaments. They are rich in alanine, leucine, valine and proline but deficient in sulfur containing amino acids (cysteine and methionine), lysine and histidine. *Collagens* are found in connective tissues, tendons and bones. They are insoluble in water, dilute acids and alkalis. From this point of view they are similar to elastins. *Gelatin* is a collagen-derived protein obtained by the partial hydrolysis of collagen. It is the product of prolonged boiling of collagen in water. Collagen is rich in glycine, proline and hydroxyproline but low in sulfur containing amino acids and tryptophan. *Keratins* are highly insoluble compounds. The sulfur content of keratin is high. It is present in the form of cystine, which is responsible for the stability and insolubility of keratins.

3.9.2. Conjugated proteins

Conjugated proteins are those which on hydrolysis give amino acids and non-protein group called **prosthetic group**. In these proteins, the prosthetic group is strongly bound to the protein part called **apoprotein**. The types of conjugated proteins and their prosthetic groups are given on Tab. 3.4.

Phosphoproteins are the proteins conjugated with phosphate. Phosphate is attached to OH group of serine, tyrosine or threonine present in protein. They are found in milk (*casein*), in egg yolk (*vitellin*, also called *phosvitin*, and *vitellenin*), in fish caviar (*ichtulin*). These proteins are nutrients during the growth of embryos and young organisms. A number of key enzymes that regulate the processes of intracellular metabolism as well as receptors and other proteins exist both in phosphorylated and unphosphorylated form, but they are not considered as phosphoproteins. Phosphorylation reactions are catalyzed by *protein kinases*, and dephosphorylation reactions – by *protein phosphatases*.

Glycoproteins are proteins conjugated with carbohydrates in varying amounts attached as short or long chains. They are divided into *true glycoproteins* (the prosthetic group is represented by oligosaccharides containing 15-20 monosaccharide residues) and

proteoglycans (the prosthetic group is represented by heteropolysaccharides *glycosaminoglycans* also called *mucopolysaccharides*, mainly contained in the connective tissue).

Table 3.4.
The types of conjugated proteins and their prosthetic groups

Conjugated proteins types	Prosthetic group	Examples
Phosphoproteins	phosphoric acid residue	Casein of milk
Glycoproteins	the residue of the oligosaccharide or heteropolysaccharide (i.e. carbohydrate residue)	Some hormones, receptors, etc.; proteoglycans of connective tissue
Lipoproteins	lipids	β-lipoprotein of blood (or low density lipoprotein)
Nucleoproteins	nucleic acids	Chromatin (DNA + histones), ribosomes (rRNA + ribosomal proteins)
Metalloproteins	metals	Ferritin (Fe), ceruloplasmin (Cu), xanthine oxidase (Mo), alcohol dehydrogenase (Zn), superoxide dismutase (Cu)
Chromoproteins	colored compound (retinol, isoaloxazine, heme)	Rhodopsin (contains retinol), flavoproteins such as succinate dehydrogenase (contain isoaloxazine), hemoglobine and myoglobine (contain heme)

Glycoproteins can have next functions in the body:

- receptor (receptors for hormones, neurotransmitters);

- transport (transcortin);
- catalytic (enterokinase, alpha-amylase, peroxidase, glucose oxidase);
- regulatory (gonadotrophin, corticotropin, thyroglobulin);
- protective (some factors of blood clotting, mucinous);
- structural and mechanical (they are the components of the intercellular substance of the connective tissue, skin, synovial fluid of the joint capsule, tendons, heart valves, vitreous body, cornea of the eye).

A number of proteins (eg, hemoglobin) may occur both in the glycosylated and in non-glycosylated forms so they are not considered as glycoproteins.

Lipoproteins are complexes of proteins (*apoproteins*) conjugated with lipids converting them into water soluble substances. Their prosthetic group is represented by lipids such as (Fig. 3.31):

- cholesterol (free and esterified);
- phospholipids (mainly phosphatidylcholine);
- triacylglycerols.

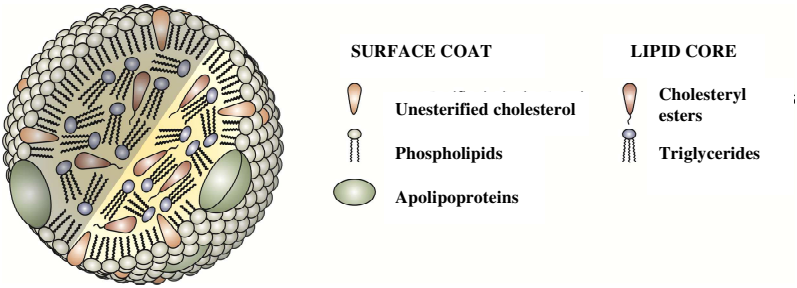


Fig. 3.31. Schematic structure of lipoproteins

They are present in blood, brain and egg. They are soluble in water, insoluble in organic solvents. A lot of them occur in the nerve tissue, in the biological membranes, and also in the blood plasma, lymph, milk, egg yolk.

There are next classes of blood lipoproteins:

- *chylomicrons* – are rich in triacylglycerols (TGs); their function is to deliver TG's to body cells to be used as fuel. They contain Apo B-48, Apo C-II, Apo E as an apoproteins.
- *very low density lipoproteins (VLDL, pre-β-lipoproteins)* - rich in TGs and cholesterol; their function is to deliver these compounds to body cells. Contain Apo B-100, Apo C-II, Apo E as an apoproteins.
- *low density lipoproteins (LDL, β-lipoproteins)* - rich in cholesterol; their function is to deliver cholesterol to all body cells. Contain Apo B-100 as an apoprotein.
- *high density lipoproteins (HDL, α-lipoproteins)* - rich in apoproteins (Apo A, Apo C, Apo E) and phospholipids, their functions are to deliver phospholipids to all body cells and pick up cholesterol from body cells and take it back to the liver ("reverse cholesterol transport"). They also are sourse of apoproteins for other lipoproteins types.

Nucleoproteins are the proteins (protamines or histones) conjugated with nucleic acids (DNA or RNA). For example, *chromatin* is histones conjugated with DNA (*deoxyribonucleoproteins*). A *nucleosome* is the basic unit of DNA packaging in chromatin (Fig. 3.32). It consists of a segment of DNA (about 146 base pears) wound around the histone octamer (two molecules of each histones H2A, H2B, H3, and H4). Adjacent nucleosomes are joined by a stretch of free DNA termed *linker DNA* (which varies from 10 - 80 bp in length depending on species and tissue type). The H1 protein binds to the linker DNA region between nucleosomes.

Nucleosomes form the *"beads on the string" structures*, which are folded into 30 nm fiber due to a binding of H1 histone followed by further folding into higher-order *loop structures* (non-histone nuclear proteins are involved in this process), that, in turn, form *the secondary loops* and *metaphase chromosomes* (Fig. 3.33).

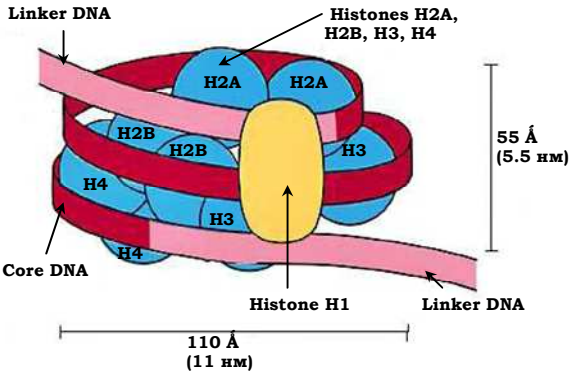


Fig. 3.32. The nucleosome structure

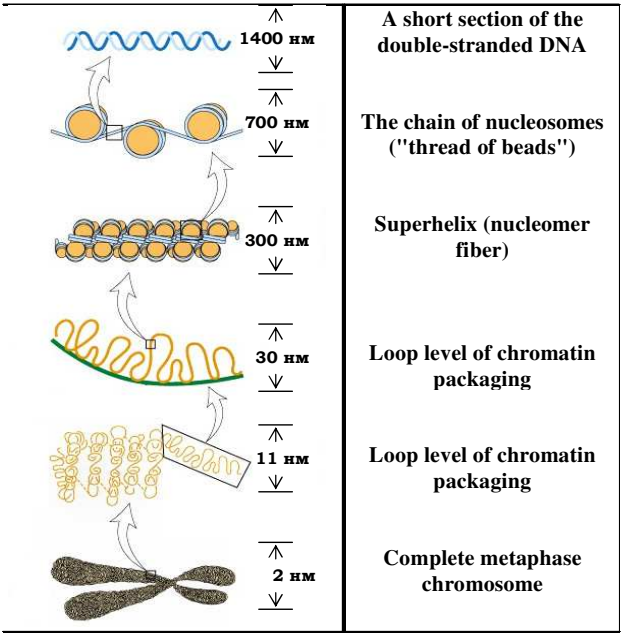


Fig. 3.33. The levels of chromatin structural organization

The examples of *ribonucleoproteins* are *ribosomes* - they are proteins conjugated with RNA.

Metalloproteins are proteins conjugated with metals with no heme group such as iron (e.g., *ferritin* is intracellular iron-binding protein and *transferrin* is an iron-binding transport protein in the blood) and copper (e.g., *ceruloplasmin* – a protein responsible for the oxidation of Fe^{2+} ions to Fe^{3+} ions which present in blood).

Chromoproteins are the proteins conjugated with colored pigment. Depending on the type of colored prosthetic group are divided into:

- *flavoproteins* – contain a prosthetic group in the form of **iso-aloxazine** (a large number of redox enzymes, which co-enzymes have flavine mononucleotide (FMN) or flavine adenindinucleotide (FAD));
- *retinolproteins* - prosthetic group is in the form of **retinol** (rhodopsin of the eye retina)
- *hemoproteins* contain a prosthetic group in the form of **porphyrin** (myoglobin, hemoglobin and enzymes catalase, peroxidase, cytochrome oxidase)

For example, one of the main body hemoproteins - hemoglobin - has following structure. Its non-protein part is represented by *heme-chelate complex of protoporphyrin IX with an iron atom* (Fig. 3.34).

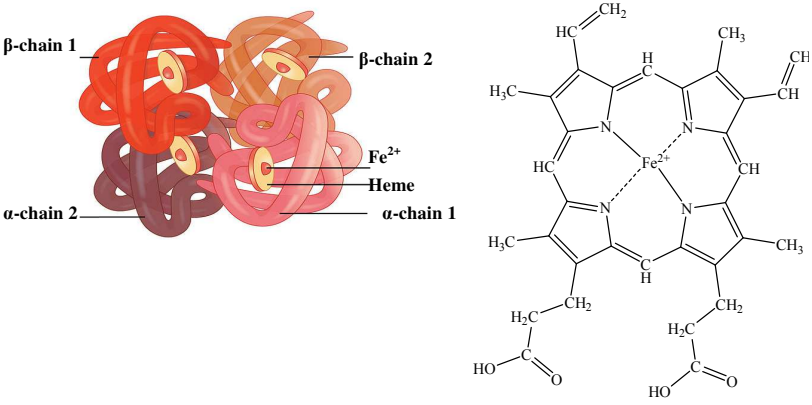


Fig. 3.34. The structures of hemoglobin (left) and heme (right)

The iron atom is co-ordinated with 4 nitrogen atoms of the pyrrole rings, and the last 2 coordinate bonds of the iron are directed perpendicular to the heme plane. One of them joins the globin, and the other - such *ligands* as oxygen, CO, cyanides.

Adult hemoglobin molecule has 4 subunits (2α, 2β) of protein called *globin*, and 4 heme groups, each of which is wrapped around by the one polypeptide chain.

There are next types of hemoglobin depending on the ligands (Fig. 3.35):

- **oxyhemoglobin** (with **O_2**);
- **carboxyhemoglobin** (with **CO**, normally not present).

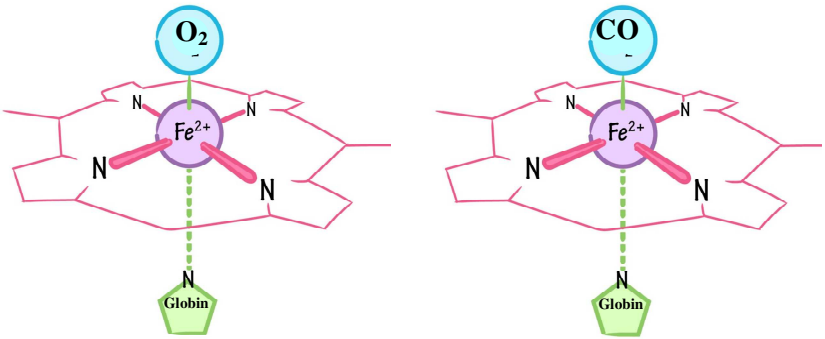


Fig. 3.35. The structures of heme in oxyhemoglobin and carboxyhemoglobin

There are also other types of hemoglobin:

- **carbhemoglobin** (has CO_2 attached not to the heme iron but to the terminal amino groups of the protein part);
- **methemoglobin** (contains Fe^{3+} instead of Fe^{2+}, normally not more than 1%);

- **Embryonic hemoglobins** (αααε; ζζεε; εεεε);
- **Fetal hemoglobin** (ααγγ; in newborns; by the end of the 1st year of life it is completely replaced by adult hemoglobin (hemoglobin A);
- **minor hemoglobins of an adult** (A2, A3, A1a, A1б, A1c);
- **glycosylated hemoglobin** (a stable glucose complex with a minor A1 fraction of hemoglobin, normally 5-7% of total hemoglobin);
- **genetically modified types of hemoglobin** (e.g., *hemoglobin S* in sickle cell anemia) that occur in *hemoglobinopathies;*
- **"old" hemoglobin** (its complex with glutathione).

3.10. Test questions

1. Amino acids hydroxyproline and hydroxylysine are typical for:

A. antiparallel β-sheet
B. parallel β-sheet
C. α-helix
D. chaotic conformation
E. Collagen helix

2. Which amino acid prevents the α-helix formation?

A. Methionine
B. Proline
C. Alanine
D. Leucine
E. Glutamate

3. Which of the following amino acids belong to the polar uncharged?

A. methionine
B. cysteine;
C. glutamic acid;

D. histidine;
E. phenylalanine

4. Which of the following amino acids is the amide of monoamino dicarbonic amino acid?

A. cystine;
B. glutamic acid;
C. glutamine;
D. arginine;
E. tryptophan

5. Select from this list the cyclic amino acid:

A. tyrosine;
B. valine;
C. cysteine;
D. methionine;
E. aspartic acid;

6. For which of the following amino acids, the value of the isoelectric point is determined by the formula pI = ½ (pK_1 + pK_R)?

A. glycine;
B. aspartic acid;
C. alanine;
D. tryptophan;
E. histidine

7. What type of secondary structure is characteristic of fibrillar proteins β-keratin and fibroids?

A. β-sheet
B. α-sheet
C. α-helix
D. chaotic conformation
E. collagen helix

8. Dialysis is a method of:

A. chromatographic separation of the mixture of proteins;

B. electrophoretic separation of the mixture of proteins;
C. enzymatic destruction of proteins;
D. separation of proteins from low molecular weight impurities and solvent;
E. determination of the isolectric point of the protein using electrophoresis in the pH gradient.

9. The main proteins of the nucleus of the cell, involved in the formation of the structure and in the functioning of the nuclear nucleoprotein complex and rich in alkaline amino acids arginine, histidine, lysine, are representatives of:
A. scleroproteins;
B. albumin;
C. globulins;
D. histones;
E. glutenins

10. Chromatin is an example of:
A. deoxyribonucleoproteins;
B. lipoproteins;
C. ribonucleoproteins;
D. chromoproteins;
E. metal proteins.

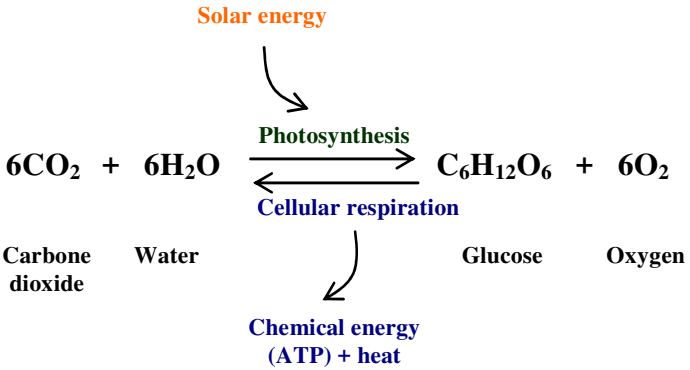
PART 4. CARBOHYDRATES AS BIOORGANIC MOLECULES

Carbohydrates are polyhydroxy aldehydes or ketones (the simplest carbohydrates, or monosaccharides), or substances that yield such compounds on hydrolysis (large carbohydrate molecules called polysaccharides):

$$\underset{\text{(polysaccharide)}}{\text{cellulose}} \xrightarrow{\textit{hydrolysis}} \underset{\text{(monosaccharide)}}{\text{D-glucose}}$$

The general empirical structure for carbohydrates is $\mathbf{(CH_2O)_n}$ suggesting that they are carbon "hydrates". For example, the empirical formula of glucose is $C_6H_{12}O_6$, which can also be written $(CH_2O)_6$ or $C_6(H_2O)_6$. But although many common carbohydrates conform to the empirical formula $(CH_2O)_n$, others do not; some carbohydrates also contain nitrogen, phosphorus, or sulfur.

Carbohydrates are the most abundant biomolecules in nature. They are not genetically encoded and are produced in a form of glucose by *photosynthesis* in plants – this process is a direct link between solar energy and the chemical bond energy in living organisms:



Carbohydrates (mainly *starch* in plants and *glycogen* in animals) are the source of rapid energy production. Some of carbohydrates are the structural elements in the cell walls of bacteria (*peptidoglycan* or

murein), plants (*cellulose*) and animals (*chitin*). *Glycosaminoglycans* of extracellular matrix also play a structural role, and some gel-like carbohydrates play the role of "lubricant" for joints. Carbohydrates may be linked to proteins and lipids with glycoproteins and glycolipids forming. Such *conjugated carbohydrates* are important in cell-cell communication and in interactions between cells and other elements in the cellular environment (e.g., other cells, hormones, and viruses). For example, they are involved in the recognition and adhesion of cells, promote the realization of antigenic function

Carbohydrates may be also the components of several metabolic pathways and some of them are the substrates for the fats, non-essential amino acids and steroids synthesis. *Glucuronic acid*, which is a product of glucose oxidation, is used to neutralize the products of proteins decay in the intestine, as well as in the metabolism of bile pigments.

Disorders of the carbohydrates metabolism are characteristic of a number of diseases (diabetes mellitus, liver and nervous system damage), or cause a number of pathological conditions (intolerance to lactose, fructose, galactosemia, glycogenosis and aglycogenosis). Carbohydrates and their derivatives are used in medical practice: glucose 40%; cardiac glycosides (e.g., digitalis), blood substitute dextran.

There are three major classes of carbohydrates:

- monosaccharides,
- oligosaccharides,
- polysaccharides.

Monosaccharides consist of a single polyhydroxy aldehyde (aldoses) or ketone unit (ketoses). The most abundant monosaccharide in nature is the sixcarbon D-glucose.

Oligosaccharides consist of short chains of monosaccharide units joined together by *glycosidic linkages*. The most abundant are the *disaccharides* with two monosaccharide units – for example, sucrose (consists of the six-carbon sugars D-glucose and D-fructose joined covalently). All common names of monosaccharides and disaccharides have ending with the suffix **"-ose"**. Most oligosaccharides having three or more units do not occur as free

entities but are joined to lipids or proteins in structures called *glycoconjugates*.

Polysaccharides consist of long chains that have hundreds or thousands of monosaccharide units. Polysaccharides that have the same monosaccharide units are called *homopolysaccharides*; when they consist of the different monosaccharide units – there are *heteropolysaccharides*. Some polysaccharides, such as cellulose, occur in *linear* chains, whereas others, such as glycogen, have *branched* chains. The most abundant homopolysaccharides, starch and cellulose made by plants, consist of recurring units of D-glucose, but they differ in the type of glycosidic linkage.

4.1. Monosaccharides (simple sugar)

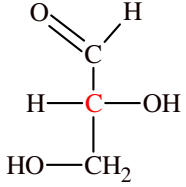
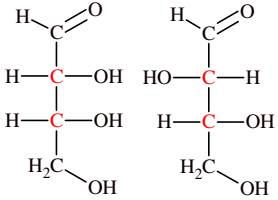
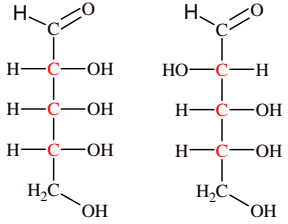
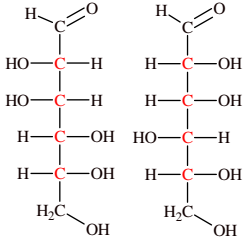
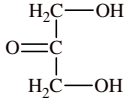
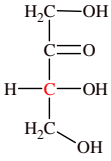
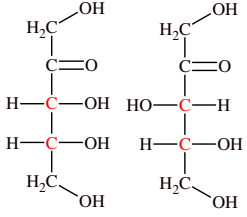
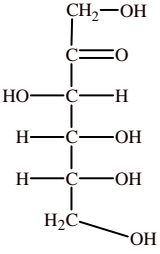
Monosaccharides are the carbohydrates that cannot be hydrolyzed to simpler carbohydrates. They have the general formula $\mathbf{C_nH_{2n}O_n}$, where n > 3. They are colorless, crystalline solids, soluble in water but insoluble in nonpolar solvents. One of the carbon atoms of these substances is double-bonded to an oxygen atom to form a carbonyl group; each of the other carbon atoms has a hydroxyl group.

4.1.1. Monosaccharides classification

They may be subcategorized as *aldoses* or *ketoses*, if the molecule contains an aldehyde or ketone functional groups respectively. **Aldoses** have a carbonyl group in the form of an aldehyde on the end of the carbon chain and **ketoses** have a ketone group somewhere along the sugar backbone. The simplest ketose is *dihydroxyacetone* while the simplest aldose is *glyceraldehyde*, which can be found as either the L- or D-enantiomer.

Monosaccharides may be also classified based on the number of carbon atoms as *trioses* (n=3), *tetroses* (n=4), *pentoses* (n=5), *hexoses* (n=6), *heptose* (n=7), etc (Tab. 4.1).

Table 4.1. Monosaccharides classification (with examples)

	Trioses	Tetroses	Pentoses	Hexoses
Aldoses	Three carbons  D-glyceraldehyde	Four carbons  D-Erythrose D-Threose	Five carbons  D-Ribose D-Arabinose	Six carbons  D-Mannose D-Gulose
	Aldotriose	**Aldotetroses**	**Aldopentoses**	**Aldohexoses**
Ketoses	Three carbons  Dihydroxyacetone	Four carbons  D-Erythrulose	Five carbons  D-Ribulose D-Xylulose	Six carbons  D-Fructose
	Ketotriose	**Ketotetrose**	**Ketopentoses**	**Ketohexoses**

So dihydroxyacetone and glyceraldehyde are trioses, the example of tetrose is erythrulose, penthoses – ribulose, ribose and xylulose, and hexoses – glucose, mannose, fructose and galactose. Glucose is known as a *aldohexose* (a six carbon aldose) and ribose is considered an *aldopentose* (a five carbon aldose).

4.1.2. Stereochemical properties of monosaccharides.

All monosaccharides (with the exception of ketotriose dioxyacetone) have one or more chiral centers, also called asymmetric atoms - carbon atoms in molecule, which are connected with four different groups.

For aldoses the number of chiral centers (n) equals:

$$n = (\text{the number of carbon atoms, C}) - 2$$

For example aldotriose called glyceraldehyde (C=3) has:

$$n = 3 - 2 = 1 \text{ asymmetric center.}$$

For ketoses:

$$n = (\text{the number of carbon atoms, C}) - 3$$

For example, ketotriose called dioxyacetone (C=3) has:

$$n = (3 - 3) = 0 \text{ asymmetric center}$$

As compounds, that have chiral center**,** carbohydrates hae optical activity (their solutions can rotate the plane polarized light right or left) and may exist in a form of stereoisomers called *enantiomers* (mirror or optical isomers), which have the same formula, mass, density, boiling and melting point, connections, but are different by their configuration (they differ one from another, as an object and its image in a mirror) and direction of rotation of plane polarized light (right or left). In general, a compound that has n chiral centers may exist in 2^n stereoisomeric forms:

$$N = 2^n,$$

where N is the number of stereoisomers, n is the number of asymmetric atoms. For example, when 4 chiral centers are present, at most 32 stereoisomers ($N = 2^4 = 32$) are possible (16 pairs of enantiomers). Glyceraldehyde molecule has 1 chiral centers – and therefore can exist in the form of two enantiomers – L- and D-glycedraldehyde ($N = 2^1 = 2$).

Identification of **configurations of optical isomers** via D-, L-system is especially powerfull in the chemistry of carbohydrates. Fischer projection formulas is uzed for it. *Fischer projection formula* is a two dimensional representation uzed for showing the configuration of carbohydrates. It represents a three-dimensional sugar structure (its *perspective formula*) on paper (Fig. 4.1, Fig. 4.2).

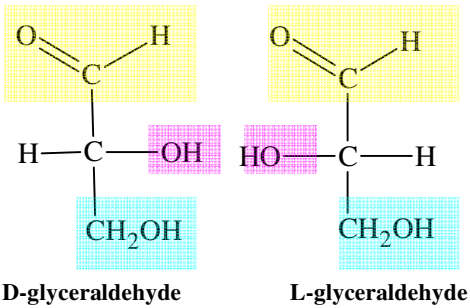


Fig. 4.1. Fischer projection formulas of D- and L-glyceraldehydes

In this way horizontal lines represent bonds projecting forward, and vertical lines represent bonds projecting to the rear. The more highly oxidized (carbonyl) carbon is shown at the top.

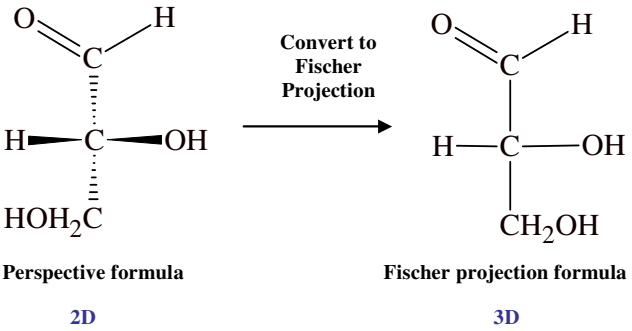
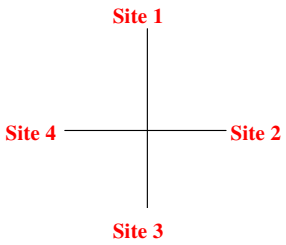
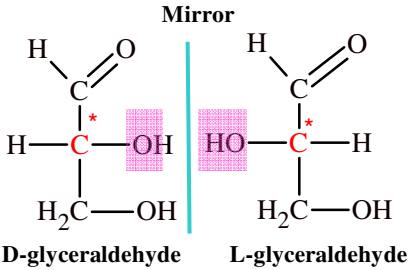


Fig. 4.2. The principle of conversion of perspective 3D formula of D-glyceraldehyde to Fischer projection formula (2D).

The simplest aldose, glyceraldehyde, contains one chiral center (the middle carbon atom - because of there are four different groups connected with this carbon atom) and therefore has two different optical isomers called enantiomers:

- **Site 1** (−COH)
- **Site 2** (−OH)
- **Site 3** (−H)
- **Site 4** (−CH_2OH)





By convention, one of these two forms is designated the D isomer of glyceraldehyde; the other is the L isomer (Fig. 4.1). Therefore **glyceraldehyde is used as the standard –** its D–isomer has −OH group to the right from asymmetric carbon atom.

In *monosaccharides that have more than one asymmetric center*, D- and L-configuration defines **by the position of the OH group at the highest numbered chiral carbon**. This carbon is called *penultimate* because it is "next to last" in the carbon sequence. In hexoses molecules the penultimate carbon is C5 because of this asymmetric carbon atom maximally distant from the aldehyde or ketone group (Fig. 4.3).

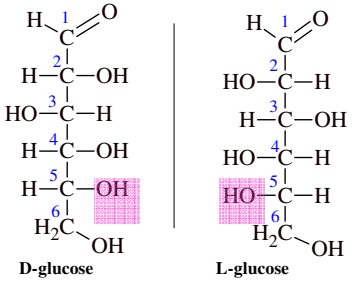


Fig. 4.3. The penultimate carbon and D- and L- configuration definition in glucose molecule

Optical activity – the ability to rotate the plane polarized light right or left – is a unique characteristic of enantiomers and can be measured by means of a polarimeter**.** The notation (+) means rotation to the right (clockwise), and the notation (-) means rotation to the left (anticlockwise). Not in all cases monosaccharides of D-series rotate a surface of plane polarized light to the right, and of L-series – to the left:

Type of sugar	Angle of optic rotation
D-glucose	+52.7
D-fructose	-92.4

An equimolar mixture of two enantiomers is called a racemic mixture, or a racemate**.** Since a racemic mixture contains equal numbers of dextrorotating and levorotating molecules, the net optical rotation is zero - such mixture is optically inactive. A racemic mixture is often specified by prefixing the name of the compound with the symbol (±).

4.1.3. Open chain (acyclic) and closed chain (cyclic) forms of monosaccharides. Haworth projections

Monosaccharides molecules can exist either as *an open chain (acyclic) form* or *a closed chain (cyclic, or ring) form.*

The open chain form of monosaccharides is illustrated with Fischer projections which show the skeleton of the acyclic monosaccharide (Fig. 4.1, Fig. 4.2). *A Haworth projection* is used to represent the cyclic form of monosaccharides (Fig. 4.4).

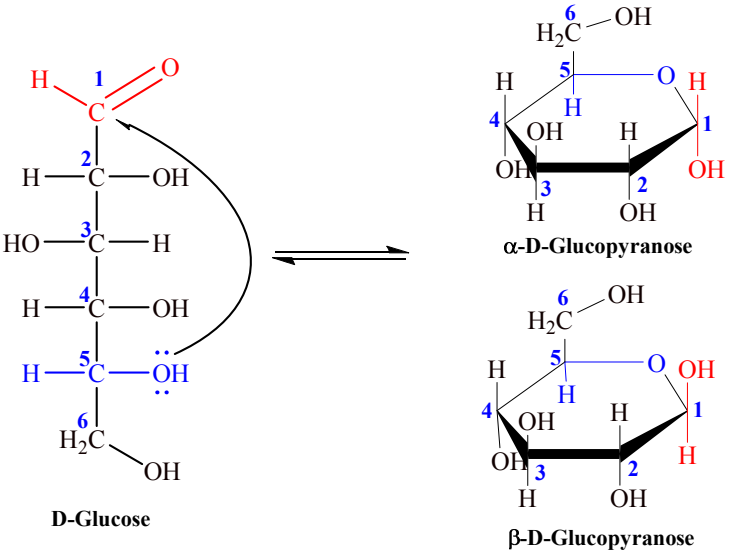


Fig. 4.4. The principle of conversion of Fischer projection formula of D-glucose into Haworth projection.

Therefore, Fischer projection is a way of representing an acyclic (open chain, a linear form) carbohydrate structure. The top of the Fischer projection of glucose contains an aldehyde group and that there are six carbons in the polyhydroxy chain. Haworth Projection is a way of representing a cyclic (closed chain) carbohydrate. Substituents can either point up or down on this ring. *All the groups localized to the right of linear form are pointing down and all the groups to the left are pointing up in the cyclic projection.* The ring structure is more energetically stable and is more common in the case of glucose, fructose and ribose.

Monosaccharides with five or more carbon atoms in the backbone have the ability to form cyclic structures, typically five or six member heterocyclic rings (Fig. 4.5).

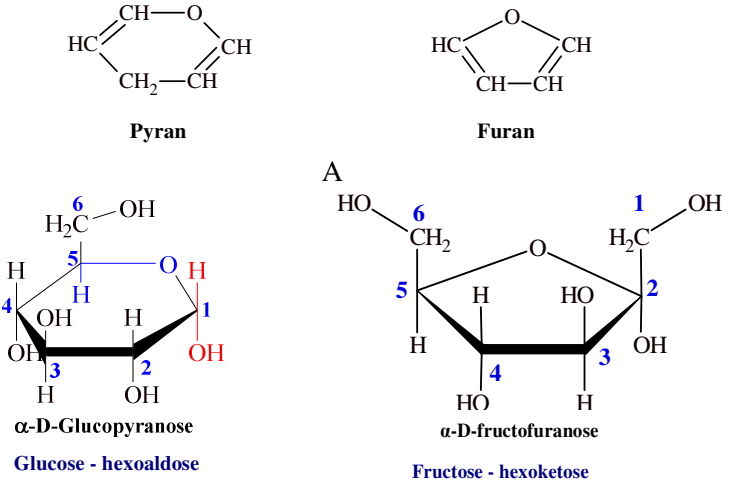


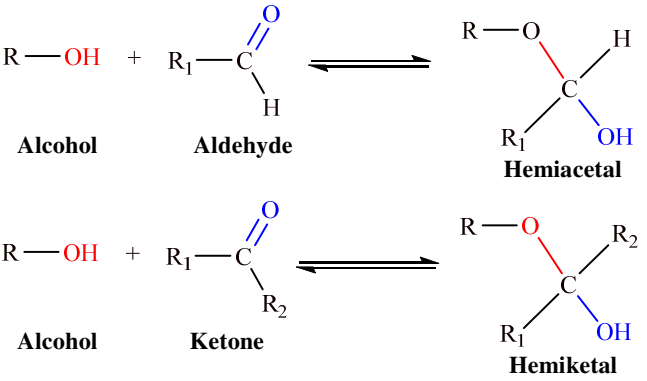
Fig. 4.5. The structures of pyran and furan (A) and α-D-glucopyranose and α-D-fructofuranose (B)

There are two possible ring formations for these sugars, known as *pyranose* and *furanose* formations. **Six-membered cyclic forms of sugars** are called **pyranoses** (for example, *glucopyranose*) because

they resemble the 6-membered ring compound *pyran.* Pyranose rings are more typical for aldoses.

Hexoses also exist in **cyclic forms having five-membered rings**, which, because they resemble the 5-membered ring compound *furan*, are called **furanoses** (for example, *fructofuranose*) (Fig. 4.5). Furanose rings are more typical for ketoses.

The mechanism of aldoses and ketoses cyclic form formation is as follows. In general, alcohols react with carbonyl group of aldehyde or ketone to form hemiacetals and hemiketals:



Carbohydrates have both alcohol and carbonyl groups, so they can form intramolecular hemiacetales (for aldoses) and hemiketales (for ketoses) in a process called **intramolecular cyclization.**

In the case of hemiacetal form of aldose formation the carbonyl group in aldoses reacts with an alcohol group within the same sugar molecule. This intramolecular reaction produces an **intramolecular hemiacetal.** Usually in hexoaldoses *the C-5 hydroxyl group acts as the alcohol (nucleophile) that attacks the C-1 aldehyde carbonyl carbon (electrophile) to form the 6 membered ring, pyranose* (Fig. 4.6). **In the case of hemiketal form of ketose formation** the carbonyl group in ketoses reacts with an alcohol group within the same sugar molecule. This intramolecular reaction produces an **intramolecular hemiketal** (Fig. 4.7).

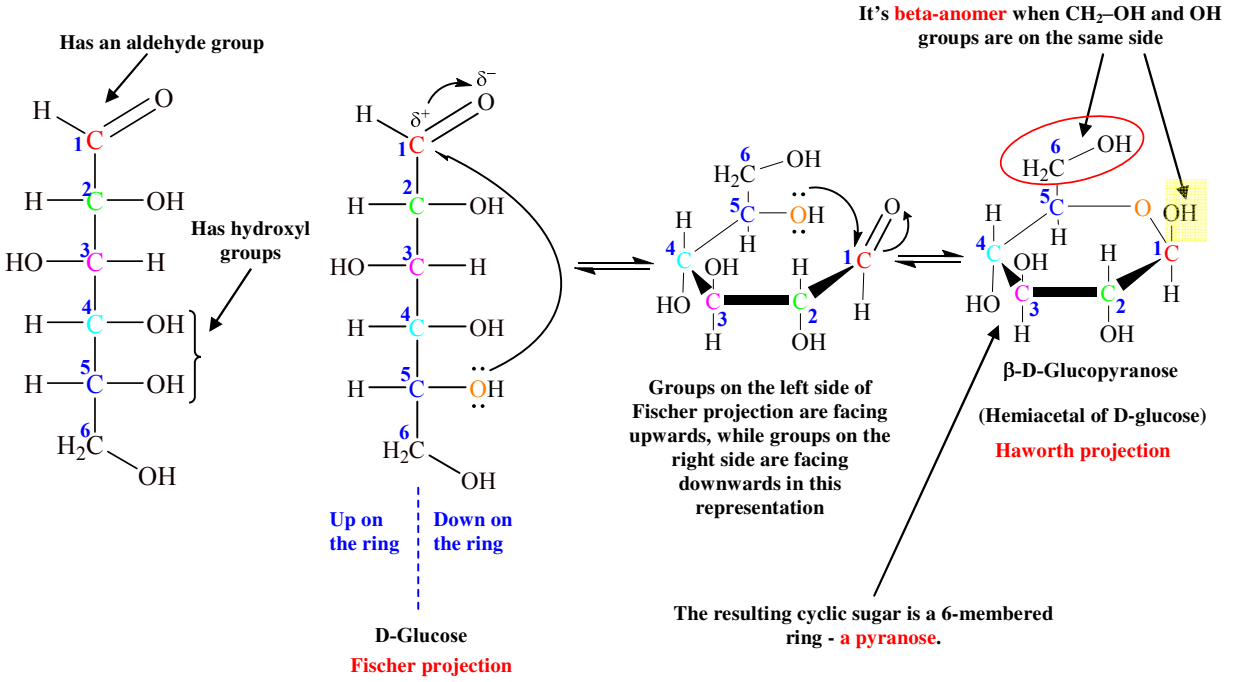


Fig. 4.6. The scheme of intramolecular cyclization of D-glucose with hemiacetal form (β-D-glucopyranose) formation

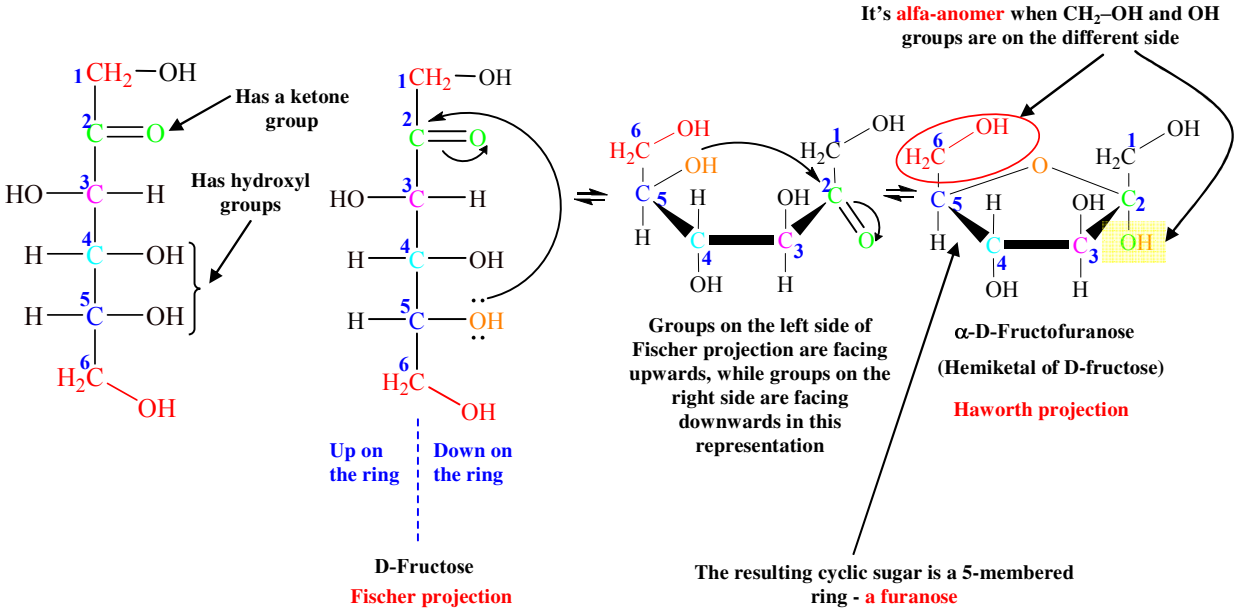


Fig. 4.7. The scheme of intramolecular cyclization of D-fructose with hemiacetal form (α-D-fructofuranose) formation

To form a intramolecular hemiketal, *the C-5 hydroxyl group of a ketohexose attacks the C-2 keto group to form the 5 membered ring, furanose* (Fig. 4.7). Comparison of mechanisms of hemiacetal and hemiketal forms of hexoses formation is given on Fig. 4.8. Glucofuranose and fructopyranose forms are also possible, but they are less stable – their structures together with the scheme of their formation mechanism are given on Fig. 4.9.

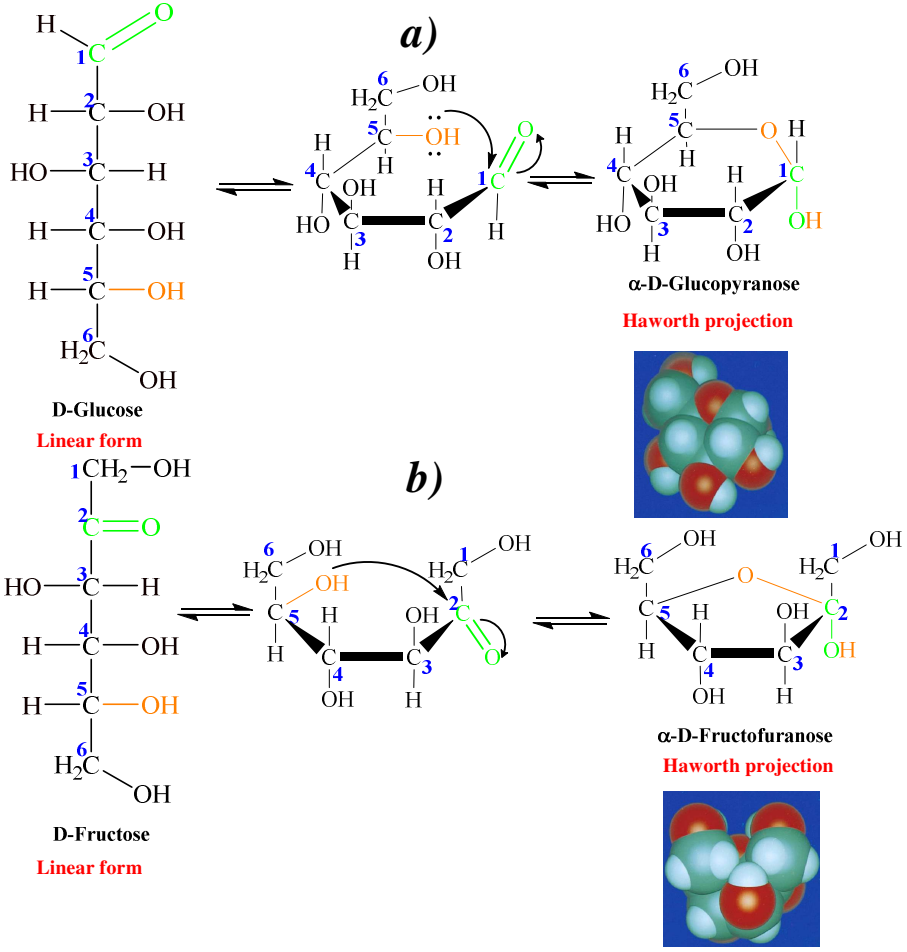


Fig. 4.8. Comparison of mechanisms of hemiacetal (a) and hemiketal (b) forms of hexoses formation

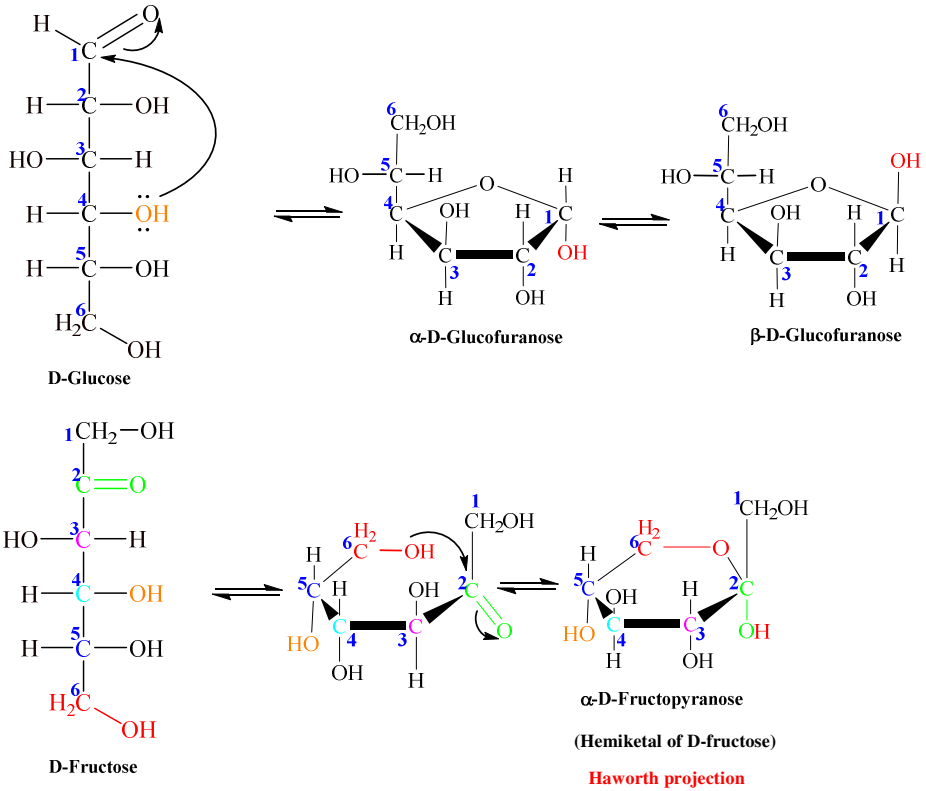


Fig. 4.9. The structures of glucofuranose and fructopyranose forms and the scheme of their formation mechanism

4.1.4. Anomeric carbon. α and β anomers

Cyclization process creates a new stereogenic center called **the anomeric carbon** (other names – **hemiacetal carbon** for aldoses and **hemiketal carbon** for ketoses). *In hemiacetal forms this center is C-1, and in hemiketal forms – C-2.* The functional group linked

with anomeric carbon both in hemiacetals and hemiketals is called the **hemiacetal group** (Fig. 4.10)**.**

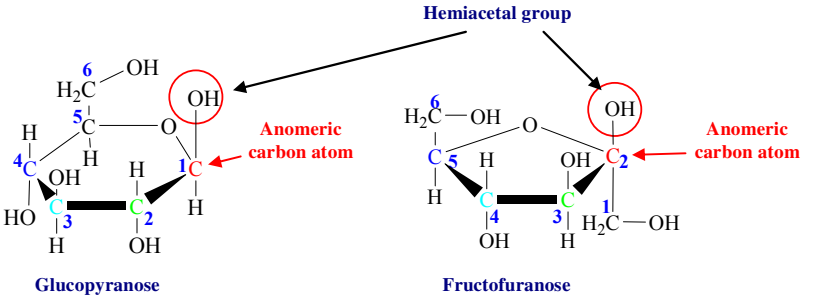


Fig. 4.10. Localization of hemiacetal group and anomeric carbon in hemiacetal and hemiketal forms of carbohydrates (glucopyranose and fructofuranose, respectively)

In cyclic monosaccharide molecules hemiacetal group is –OH (in glucose at C1 and in fructose at C2).

Since cyclic forms of monosaccharides contain an additional asymmetric carbon atom (anomeric carbon) they can exist in two additional stereoisomeric forms. Such isomeric forms that differ from each other only in their configuration about the anomeric carbon atom (α-D-glucopyranose and β-D-glucopyranose; α-D-fructofuranose and β-D-fructofuranose) are called **anomers**. They can either be *α- or β-* depending on the relative position of the -CH_2OH group and the -OH group on the anomeric carbon (Fig. 4.11).

4.1.5. Chair and boat 3D conformations of pyranoses and envelope 3D conformation of furanoses

The six-membered pyranose ring (for example, in D-glucopyranose molecule) is not actually planar, as Haworth perspectives suggest, but tends to assume either the "boat" or the "chair" conformation.

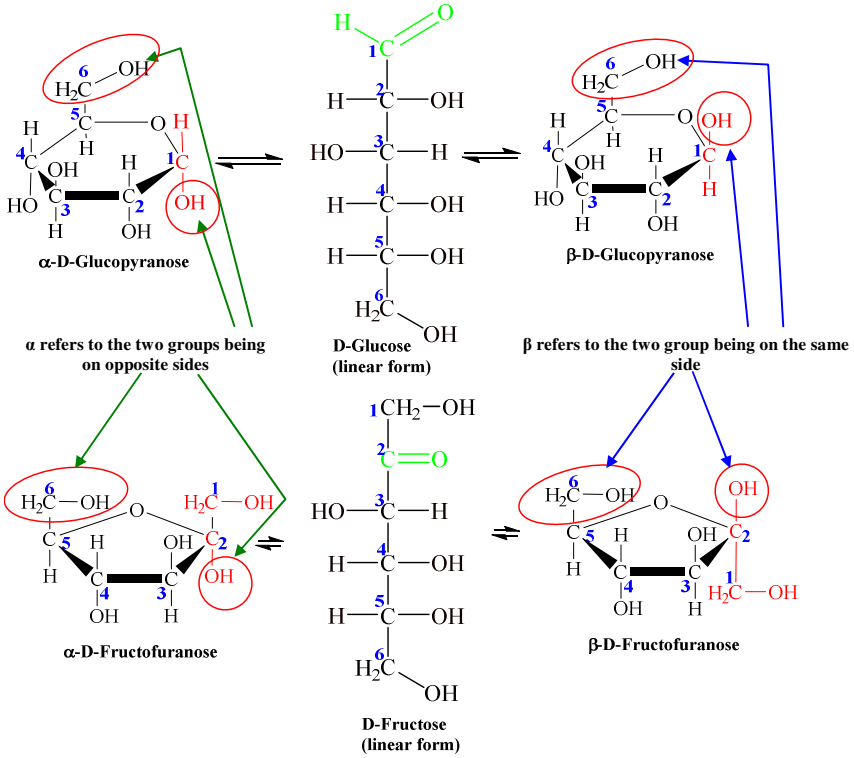


Fig. 4.11. Definition of α- and β- anomers of glucopyranose and fructofuranose

In the chair conformations the carbons 2,3,5,6 lies in the same plane, C1 lies above and C4 lies below (Fig. 4.12). The carbon atoms are antisocial they keep apart from each other thus increasing the space and reducing steric hindrance (electron repulsion) from the groups.

In the boat conformation carbons C2, C3, C5 and C6 are coplanar, while C1 and C4 (the 'bow' and the 'stern') are displaced

away from that plane in the same direction. These two groups facing up are brought close enough to each other to cause steric strain. It makes the structure unstable and energetically unfavorable. Thus, *chair conformation is more stable, energetically feasible and preferred.*

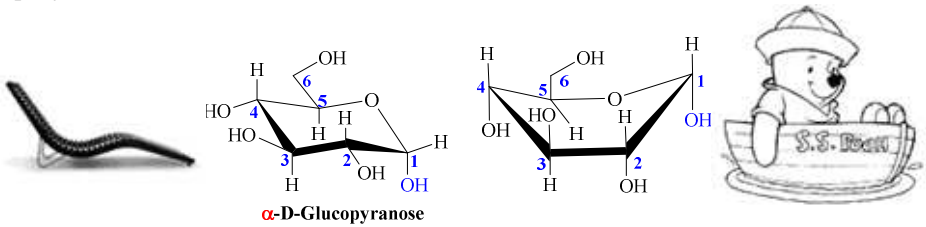


Fig. 4.12. Definition of chair and boat 3D conformations of pyranoses

The lowest energy conformation of *furanose ring* (for example, in ribose, deoxyribose or fructofuranose molecule) is known as the **'envelope' conformation.** It is characterized by four of the ring atoms in the same plane and one out of plane. The out-of-plane carbon is said to be in the ***endo*** position ('*endo*' means 'inside') (Fig. 4.13).

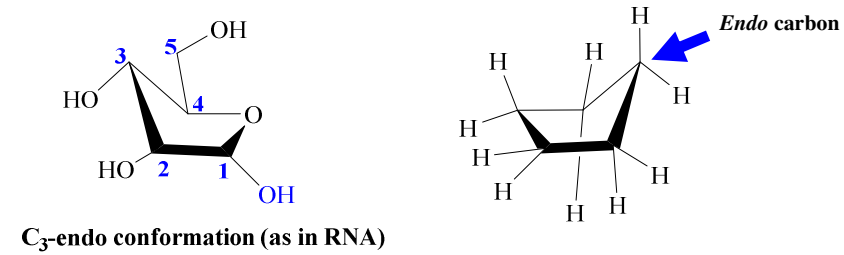


Fig. 4.13. Definition of envelope conformations of furanoses

4.1.6. Mutarotation as an interconverce of α and β forms of D-glucose

Mutarotation is the change in the optical rotation because of the change in the equilibrium between two anomers. It was discovered by French chemist Dubrunfaut in 1846, when he noticed that the specific rotation of aqueous sugar solution changes with time.

Cyclic sugars show mutarotation because of α and β anomers interconvert in aqueous solution. The optical rotation of the solution depends on the optical rotation of each anomer and their ratio in the solution (Tab. 4.2).

Table 4.2.

The optical rotation of monosaccharide solution before and after mutarotation and percent of monosaccharides anomers that presents at equilibrium

Monosaccharide	**[α] (optical rotation, °)**	**[α] after mutarotation**	**% present at equilibrium**
α-D-Glucose	+112,2	+52,7	36
β-D-Glucose	+18,7	+52,7	64
α-D-Galactose	+150,7	+80,2	28
β-D-Galactose	+52,8	+80,2	72

For example, if a solution of β-D-glucopyranose is dissolved in water, its specific optical rotation will be +18,7°. Over time, some of the β-D-glucopyranose will undergo mutarotation to become α-D-glucopyranose, which has an optical rotation of +112,2°. Thus the rotation of the solution will increase from +18.7 to an equilibrium value of +52.7° as some of the β form is converted to α form. The equilibrium mixture is actually about 64% of β-D-glucopyranose and about 36% of α-D-glucopyranose (as the β form is more stable), though there are also some other forms including furanoses and open chained form (Fig. 4.14).

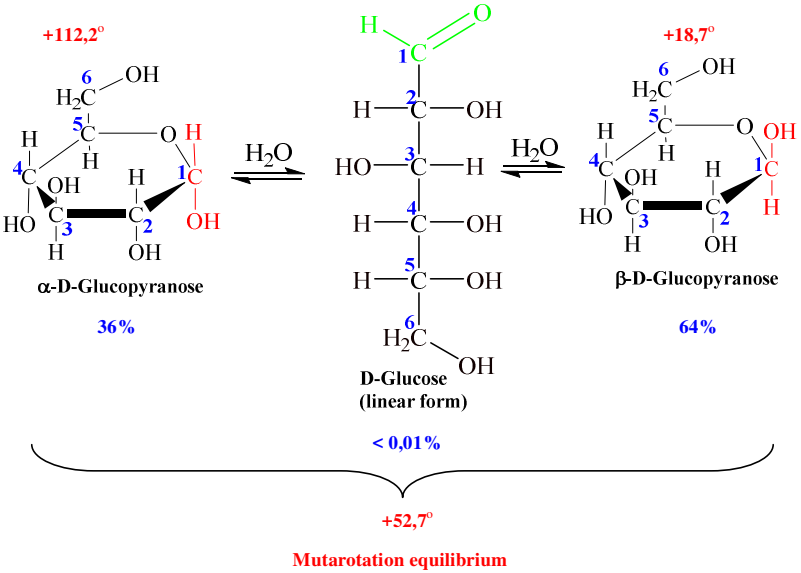
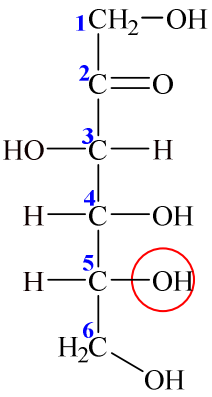


Fig. 4.14. The scheme of glucopyranose mutarotation

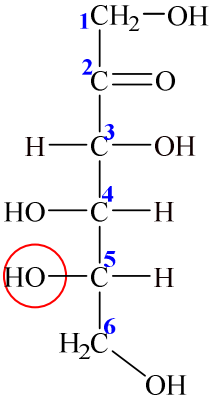
4.1.7. Important monosaccharides

The most important monosaccharides as far as the human body is concerned are glucose, galactose and fructose. They have the same molecular formula but own different structural formulas - they are isomers.

D-fructose belong to ketoses and is often referred as fruit sugar and is found in some vegetables and honey. It is twice as sweet as sucrose (per gram basis) and is used as sweeting agent in processed food products. It is present in large amounts in male reproductive tract and is synthesised in the seminal vesicles. **Fructosemia**, or **fructose intolerance**, is an inherited disease due to a deficiency of the enzyme *fructose 1-phosphate aldolase*. An infant suffering from this disease experiences hypoglycemia, vomiting, and severe malnutrition.



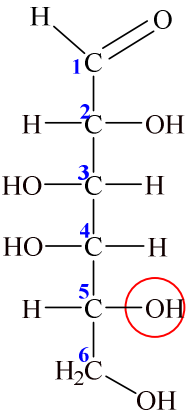
D-Fructose



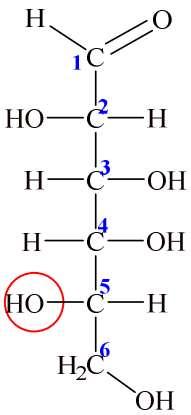
L-Fructose

Such condition is treated by placing the infant on a low-fructose diet

Galactose is necessary to synthesize a variety of biomolecules (*lactose* in mammalary glands, *glycolipids*, certain *phospholipids*, *proteoglycans*, and *glycoproteins*). Galactose and glucose are epimers at carbon 4 and interconversion is catalysed by enzyme *epimerase*.



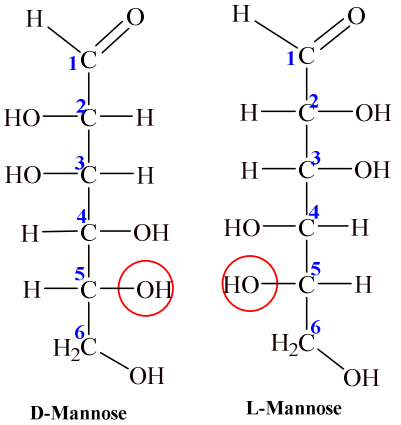
D-Galactose



L-Galactose

The main medical problem in galactose metabolism is genetic disorder **galactosemia** where enzyme to metabolize galactose is missing; accumulation of galactose in the body can cause liver damage, cataracts, and severe mental retardation.

Mannose is a C-2 epimer of glucose. It is important in human metabolism, especially in the glycosylation of certain proteins.



Several **congenital disorders of glycosylation** are associated with mutations in enzymes involved in mannose metabolism

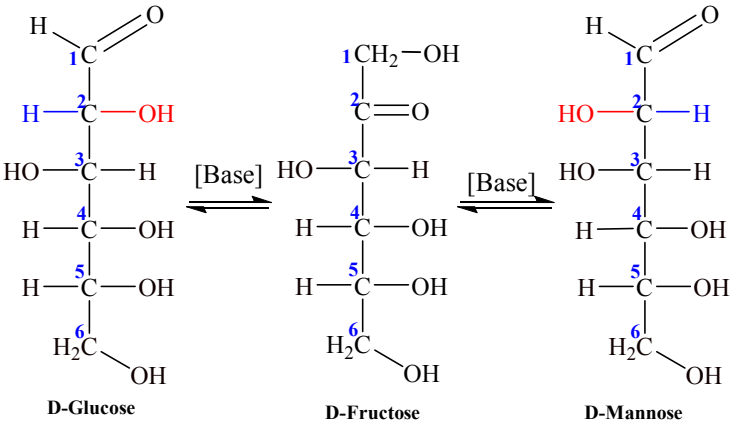
4.1.8. Physical and chemical properties of monosaccharides

Monosaccharides such as glucose and fructose are crystalline solids at room temperature. At the same time, due to existence of several OH groups that readily engage in hydrogen bonds forming they are polar compounds that quite soluble in water and insoluble in nonpolar solvents. They have high melting points (because of OH groups too). After dissolving in water, all monosaccharides give colorless solutes with sweet taste.

Monosaccharides contain *some chiral centers* as well as both *carbonyl group* (aldehyde or keto-group) and *alcohol groups***.** Therefore they can serve as substrates for:

1. izomerization reaction (with epimers forming);
2. oxidation / reduction reactions;
3. formation of glycosides.

Izomerization reaction (with epimers forming). In the presence of base D-glucose may be converted into D-fructose and D-mannose, and D-fructose may be converted into D-glucose and D-mannose:



Galactose can belong interconversion to glucose.

Oxidation of monosaccharides. The monosaccharides can be oxidized at C1, C6 as well as at both sites.

If the aldehyde end of the molecule (at C1) is oxidized (by relatively mild oxidizing agents such as Ag^+ or Cu^{2+} ions – for example, in *Tollens', Benedict's, or Fehling's solutions*), **aldonic acids** are formed. So, when the aldehyde end of *glucose* is oxidized, the product is called *gluconic acid* (Fig. 4.15).

If the alcohol at the end opposite the aldehyde (at C-6) is oxidized, hexoses form **uronic acid**. Oxidation at C-6 of *glucose*, *galactose*, or *mannose* forms *glucuronic*, *galacturonic*, or *mannuronic acids*, respectively. D-glucoronic acid and *L-iduronic acid* both are derived from oxidation of C6 of D-glucose – they are optical isomers (epimers). They are important acids in animals – in liver cells glucuronic acid combines with steroids, certain drugs, and bilirubin

to improve water solubility therby helping the removal of waste products from the body; these acids are abundant in the connective tissue carbohydrate components.

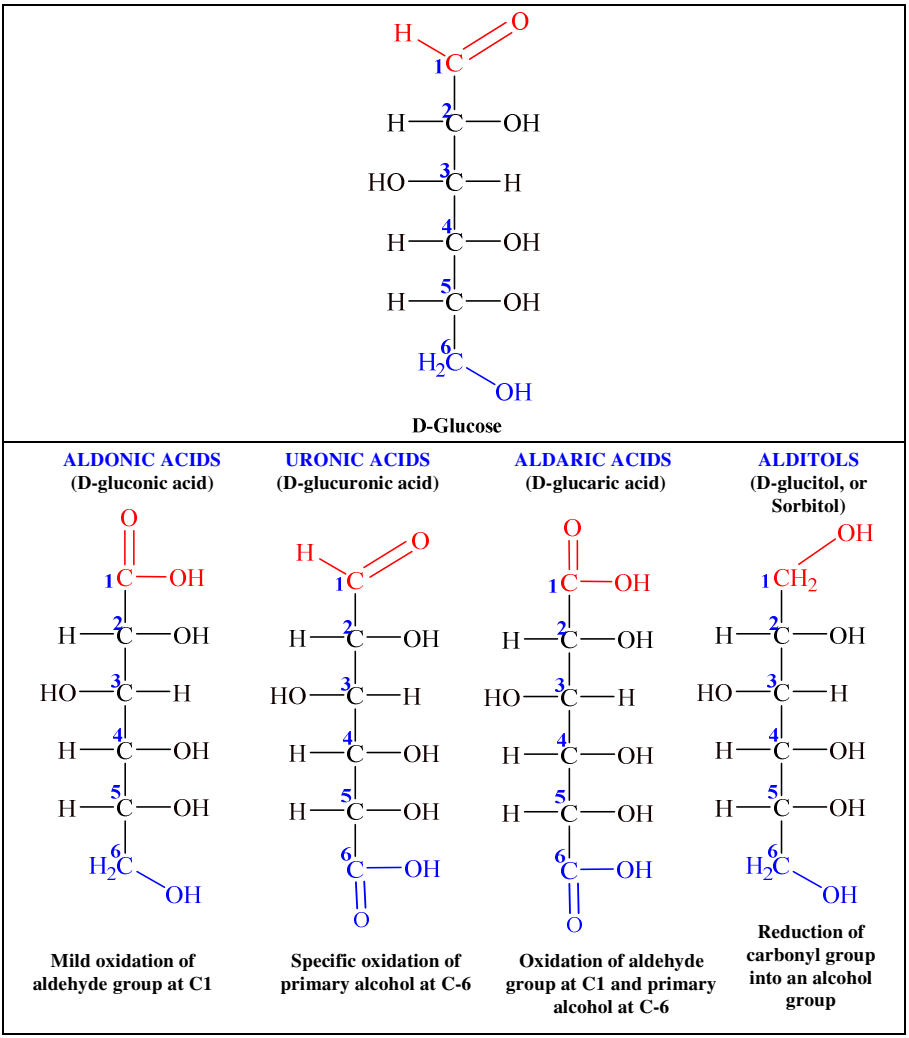


Fig. 4.15. Sites for oxidation and reduction of glucose molecule and the structures of the corresponding glucose derivatives

If both ends of an aldose chain are oxidized to carboxylic acids (for example, by such **strong oxidizing agent** as nitric acid the product is called an **aldaric acid.**

Reducing Sugars. *Reducing sugar* is any sugar that is capable of acting as a reducing agent because it *has a free aldehyde group.* In this way the aldehydes (including aldoses) can be oxidized (to acid) via a redox reaction in which another compound is reduced. A sugar is classified as a reducing sugar only if it has an open-chain form with an aldehyde group or a free hemiacetal group - such cyclic hemiacetal forms of aldoses can open to reveal an aldehyde (Fig. 4.16).

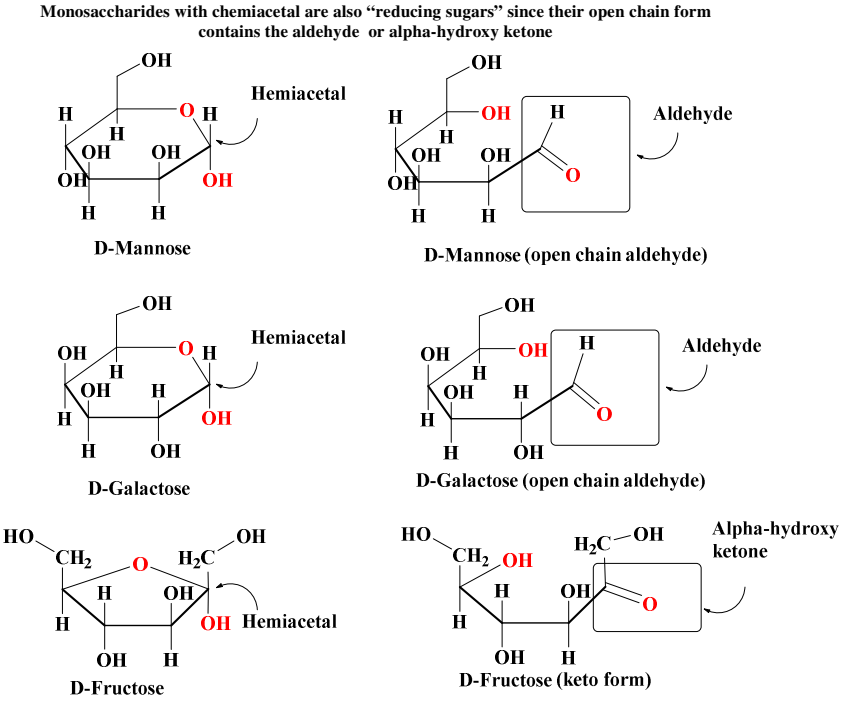
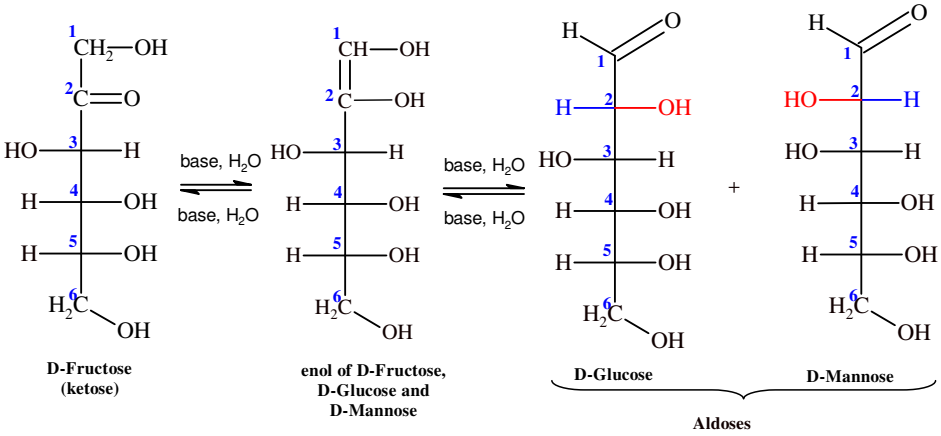


Fig. 4.16. Aldoses and ketoses as reducing sugars

Ketoses haven't a free aldehyde group as they have a ketone group, but they are also redicing sugars. It is because sugars with ketone groups can isomerize to aldoses under basic conditions**:**



Therefore all monosaccharides are reducing sugars, because all monosaccharides have an aldehyde group (if they are aldoses) or can tautomerize in solution to form an aldehyde group (if they are ketoses). Some disaccharides and oligosaccharides also are reducing.

Visual tests for aldehydes. Three visual tests for aldehydes (including aldoses) that you might encounter in an introductory bioorganic chemistry lab are the following (Fig. 4.17):

- **Test with Fehling's solution**, where an aldehyde changes the color of a blue Cu (II) solution to red Cu (I) *[as Cu_2O]*.
- **Test with Benedict's solution -** a slightly modified version of Fehling's solution
- **Tollens' test**, where aldehyde oxidation results in a beautiful "mirror" of silver metal to precipitate on the reaction vessel.

These reactions are the basis for some tests on sugar in the urine and in the blood. Importantly, ketones don't react under any of these conditions. The above tests were also a useful way of distinguishing

aldehydes from ketones. To visualize the presence of ketones with these tests it is needed to add the base in studied solutions.

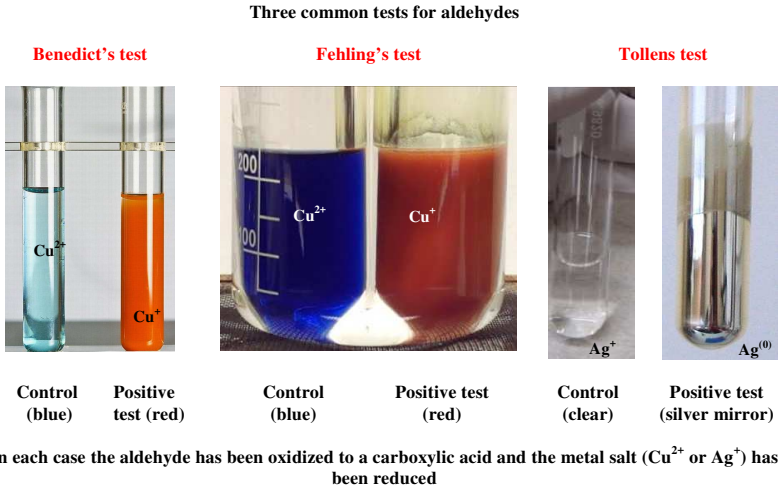
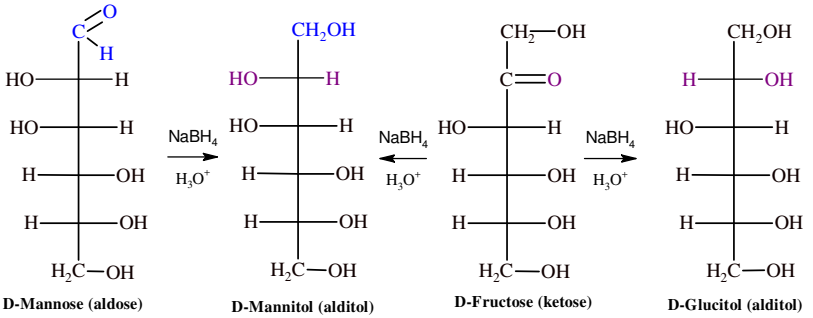


Fig. 4.17. Visual tests for qualitative definition of aldoses

Reduction of monosaccharides. Monosaccharides can be reduced via turning the carbonyl group into an alcohol group. The products that are formed in these reactions generally are called **alditols**.

Reduction of aldose forms one alditol whereas reduction of ketose forms two alditols:



The separate products are named by replacing the ***-ose*** ending with ***–itol*** – therefore:

- *D-Glucose* reduces to *D-Glucitol*, also called *D-Sorbitol*.
- Reduction of *galactose* yields *dulcitol*.
- *D-Mannose* reduces to *D-Mannitol*
- Reduction of *fructose* yields a mixture of *mannitol* and *sorbitol*.

Alditols are characteristically sweet tasting, and are widely used as sweetening agents (Fig. 4.18).

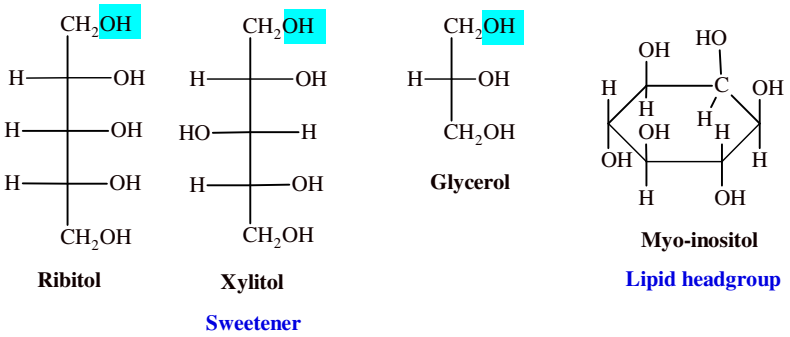


Fig. 4.18. The examples of alditols

Sorbitol accumulation in the eyes is a major factor in the formation of cataracts due to diabetes. Mannitol is used in the treatment of malignant brain tumors.

Glycoside bond and formation of Glycosides (Acetals). A covalent **glycoside bond** is formed between the hemiacetal group (OH-group linked with anomeric carbon) of a carbohydrate and a hydroxyl group of another organic compound by a condensation reaction. The products of these reactions are **the acetals** (also called **glycosides**) and one water molecule. Acetals are stable in water and bases but they are hydrolyzed in acids.

There are two configurations of glycosidic bonds, α- and β-glycosidic bonds. It depends on a configuration of the sugar with anomeric carbon that involves in glycoside bond forming (Fig. 4.19).

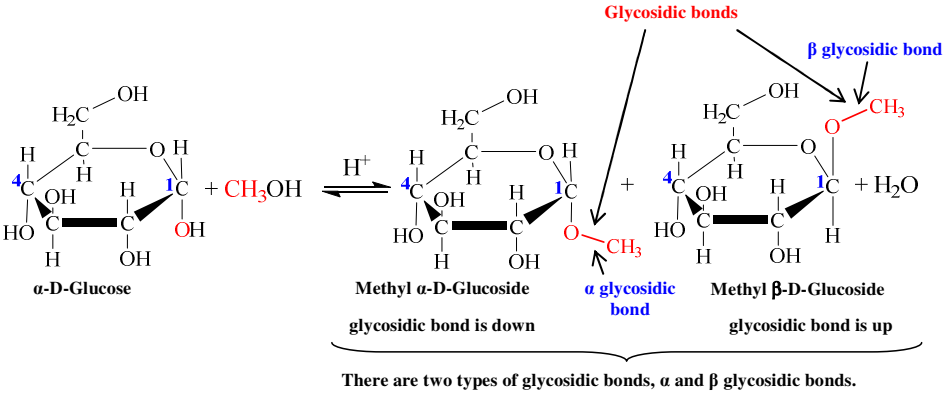


Fig. 4.19. The structures of α- and β-glycosidic bonds

Alpha linkage has the oxygen of hemiacetal group below the ring and the beta has it above the ring (Fig. 4.19). There are enzymes called **glycoside hydrolases** (or **glycosidases**) that can break glycoside bonds, but they are only able to break either the alpha or beta conformation of these bonds, not both.

There are also *O*- and *N-types of glycoside bonds*.

O-glycosidic bond is formed when the hemiacetal group (OH-group linked with anomeric carbon) of one monosaccharide reacts with any OH-group of the other monosaccharide (Fig. 4.20).

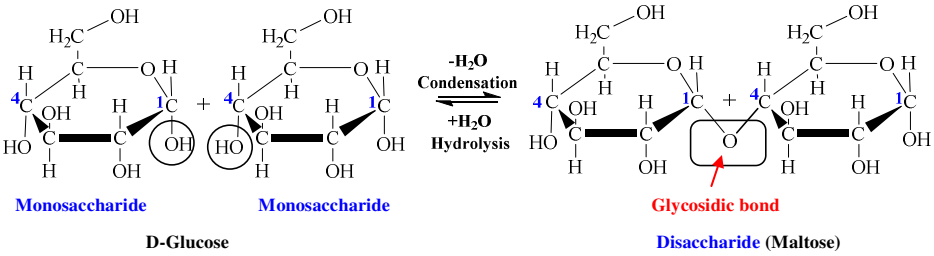


Fig. 4.20. The principle of O-glycosidic bond formation in disaccharide maltose molecule

O-glycosidic bond links monosaccharides in *di-*, *oligo-* and *polysaccharides*. For example, disaccharides such as maltose, lactose, and sucrose consist of two monosaccharides joined by an O-glycosidic bond. Glycosidic bonds are readily hydrolyzed by acid (but resist cleavage by base). Thus disaccharides can be hydrolyzed to yield their free monosaccharide components by boiling with dilute acid.

O-glycosidic bond also creates a covalent linkage between a OH-group of serine or threonine side chain and sugar (under posttranslational protein modification by *glycosylation*).

N-glycosidic bond joins the anomeric carbon of a monosaccharide to:

- a *nitrogen atom of nitrogenous base* (purines, pyrimidines) in *nucleosides* and *nucleotides* (Fig. 4.21)
- *a nitrogen atom of -NH_2-group of asparagine side chain* (under posttranslational protein modification by *glycosylation*).

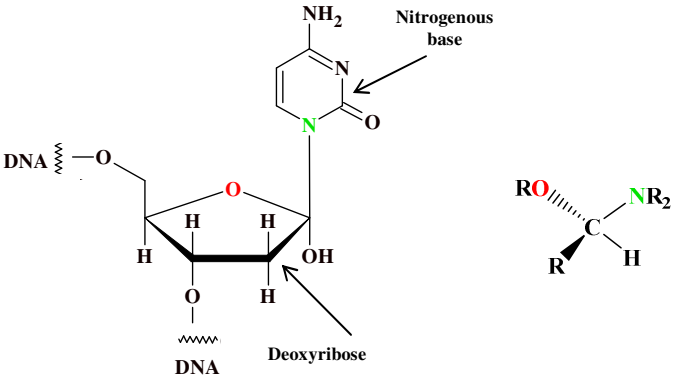
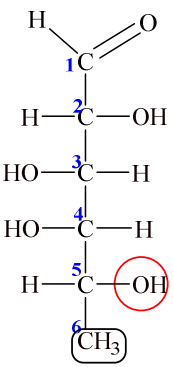


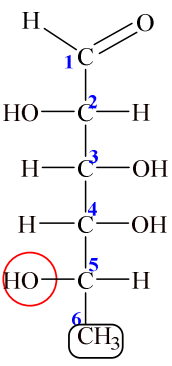
Fig. 4.21. The principle of N-glycosidic bond formation in nucleotide molecule

4.1.9. Natural sugar derivatives

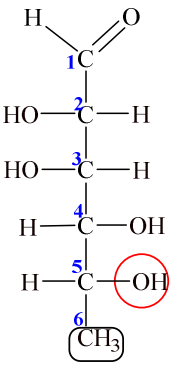
Deoxy sugars. They are formed by replacement of OH by H. The change of hydroxyl group at C-6 of galactose or mannose on hydrogen produces *fucose* or *rhamnose*, respectively:



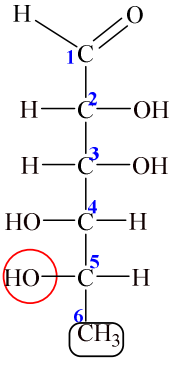
D-Fucose



L-Fucose



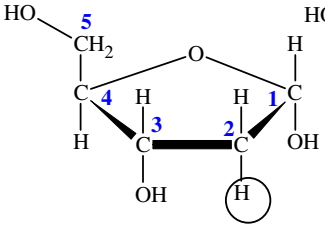
D-Ramnose



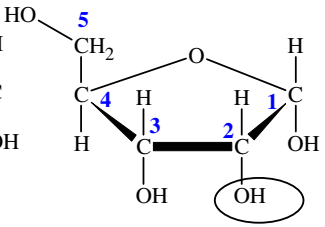
L-Ramnose

These deoxy sugars are found in the complex oligosaccharide components of glycoproteins and glycolipids.

Other example of deoxy sugars is *D-2-deoxyribose* derived from the sugar ribose by loss of an oxygen atom:



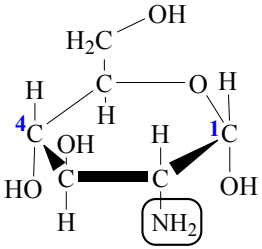
Deoxyribose – it is the pentose sugar component of DNA



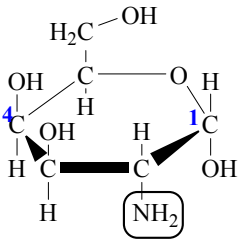
Ribose – it is the pentose sugar component of RNA

As a component of DNA, 2-deoxyribose has an important role in biology. The absence of the 2′ hydroxyl group in deoxy ribose is apparently responsible for the increased mechanical flexibility of DNA compared to RNA, which allows it to assume the double-helix conformation, and also (in the eukaryotes) to be compactly coiled within the small cell nucleus.

Amino sugars. In *glucosamine*, *galactosamine*, and *mannosamine*, the hydroxyl at C-2 of the parent compound is replaced with an amino group:



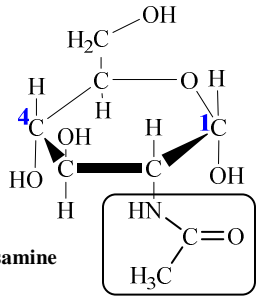
α-D-glucosamine (2-amino-2-deoxy-α-D-glucopyranose)



α-D-galactosamine (2-amino-2-deoxy-α-D-galactopyranose)

Glucosamine helps keep the cartilage in joints healthy. But natural glucosamine levels drop as people age, and, as a supplement, glucosamine is most often used to try to ease the joint pain caused by arthritis.

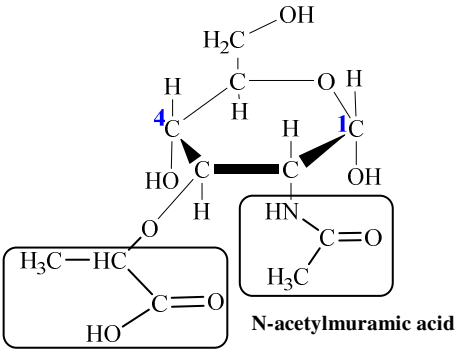
The amino group may be condensed with acetic acid, as in *N-acetylglucosamine*:



N-acetylglucosamine

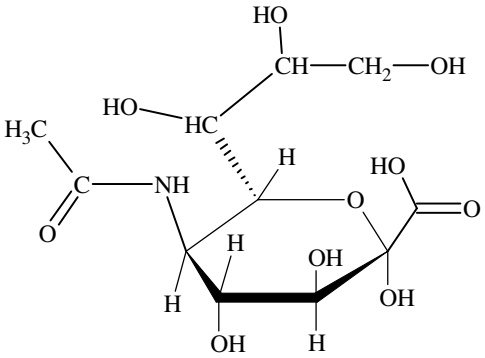
This glucosamine derivative is a part of many structural polymers, including those of the bacterial cell wall.

Another derivative of glucosamine is *N-acetylmuramic acid* - a part of the peptidoglycan polymer of Gram-positive bacterial cell walls:



N-acetylmuramic acid

N-acetylneuraminic acid (sialic acid), a nine-carbon derivative of N-acetylmannosamine, is a component of many glycoproteins and glycolipids in higher animals:



N-acetylneuraminic acid

Sugar phosphates. These compounds are relatively stable at neutral pH, and bear a negative charge. One effect of sugar phosphorylation within cells is to prevent the diffusion of the sugar

out of the cell; highly charged molecules do not, in general, cross biological membranes without specific transport systems. Phosphorylation also activates sugars for subsequent chemical transformation.

4.2. Disaccharides

A **disaccharide** is formed when a OH-group of one monosaccharide reacts with the hemiacetal group (OH-group linked with anomeric carbon) of another monosaccharide to form a glycosidic bond. This bond can be broken by hydrolysis to form two monosaccharides.

There are next examples of disaccharides:

- *Lactose* (sugar in milk) – consists of β-galactose and α-glucose residues
- *Maltose* (malt sugar) – has two molecules of α-glucose as the components
- *Cellobiose* – has two molecules of β-glucose
- *Sucrose* (table sugar) - consists of α-glucose and β-fructose
- *Trehalose* – its ccmponents are represented by two molecules of α-glucose

Each disaccharide has a specific glycosidic linkage (depending on which hydroxyl of one monosaccharide reacts with OH-group near anomer carbon of another). There are three naturally occurring glycosidic linkages in disaccharides molecules. They are:

- **1-4' link:** the anomeric carbon is bonded to oxygen on C4 of second sugar (for example, in *maltose*) (Fig. 4.22);
- **1-1' link:** the anomeric carbons of two sugars are bonded through an oxygen (for example, in *trehalose*);
- **1-2' link:** the anomeric carbons of two sugars are bonded through an oxygen ((or example, in *sucrose*).

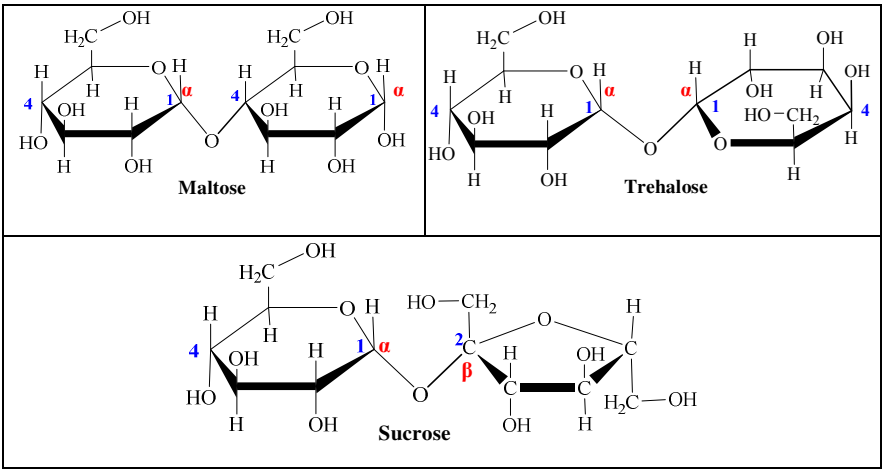


Fig. 4.22. The types of glycosidic bonds in some disaccharides molecules

Alfa or *beta* configurations of glycoside bonds *depend on a configuration of the sugar with* anomeric carbon that involves in glycoside bond forming.

Digestion of disaccharides is needed in enzymes. Defficiency of any of these enzymes causes unpleasant symptoms: much of the disaccharides moves to the colon, where *bacterial fermentation* generates large quantities of CO_2, H_2 and irritating organic acids. These products cause painful digestive upset. Most common defficiency is an ancestoral disorder, *lactose intolerance* caused by reduced synthesis of lactase.

4.2.1. Reducing and non reducing disaccharides

Disaccharide is **the reducing sugar** *if one of its monosaccharide monomers has a free hemiacetal group* (OH-group linked with anomeric carbon – for example, in glucose or galactose molecules

this group is linked with C1 and in fructose molecule – with C2) (Fig. 4.23).

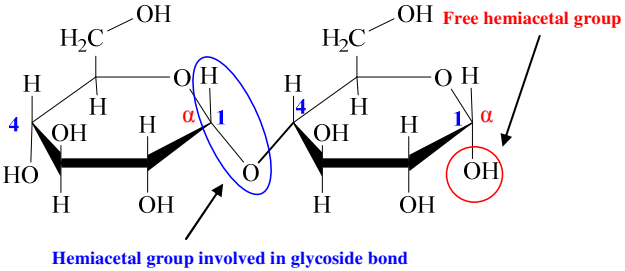


Fig. 4.23. The localization of free hemiacetal group in reducing disaccharide

Such cyclic structures can open to give a free aldehyde group to oxidize and can undergo mutarotation.

If the glycosidic bond involves OH-groups of both anomeric carbons, none of disaccharides components can be changed to an aldehyde or ketone and none of them can undergo mutarotation – *such disaccharides are* **non-reducing sugars** (Fig. 4.24).

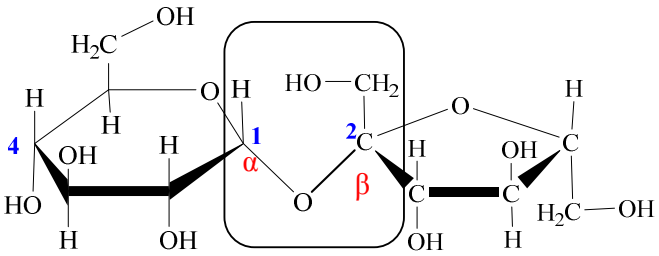


Fig. 4.24. The absence of free hemiacetal groups in non-reducing disaccharide

In reducing di- and oligosaccharides as well as in polysaccharides, the end of a chain that has a free hemiacetal group (OH-group linked with anomeric carbon) which is not involved in a glycosidic bond is commonly called **the reducing end of the chain** (Fig. 4.25)**.** The reducing end is capable of converting corresponding monosaccharide unit to the open-chain form. The reducing disaccharide will have only one reducing end, as disaccharide forms by glycosidic bond, which involve at least one hemiacetal group.

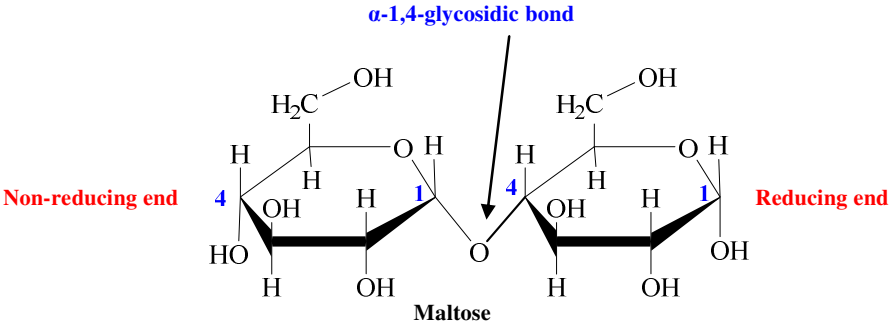


Fig. 4.25. Reducing and non-redusing ends in reducing disaccharide molecule

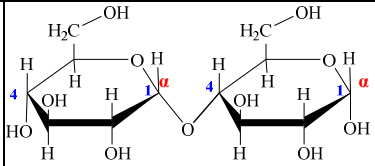
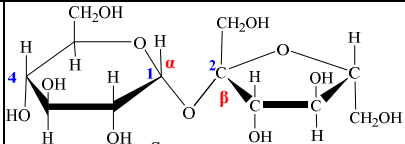
A non-reducing disaccharide has both anomeric carbons tied up in the glycosidic bond and therefore has no reducing ends.

4.2.2. Nomenclature of Disaccharides

The main rules for naming disaccharides are given on Tab. 4.3.

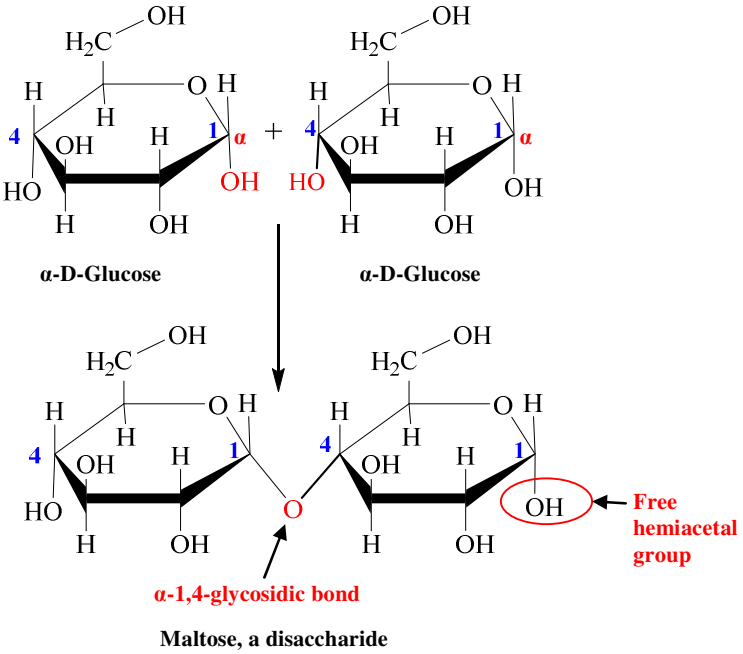
Table 4.3.

The Rules for Naming Disaccharides

№	RULES	Example - *maltose*	Example - *sucrose*
1	Naming of disaccharides proceeds from left to right (from the non-reducing end)	 **Maltose**	 **Sucrose**
2	Name of first monosaccharide is changed on …pyranosyl (6-ring) or …furanosyl (5-ring)	**α-D-glucopyranosa is changed to α-D-glucopyranosyl**	**α-D-glucopyranosa is changed to α-D-glucopyranosyl**
3	The numbers of carbons, OH-groups of which are involved in glycoside bond forming, are pointed (1→4)	**(1→4)**	**(1→2)**
4	Name of second monosaccharide: <u>- is not changed</u> (…pyranose or …furanose) – if it has a free OH-group near anomeric carbon (in reducing disaccharides) <u>- is changed</u> on …pyranoside or …furanoside if both OH-groups linked with anomer carbones are involved in glycoside bond forming (in non-reducing disaccharides)	**α-D-glucopyranose**	**β-D-fructofuranoside**
	RESULT	***α-D-glucopyranosyl-(1→4)-α-D-glucopyranose***	***α-D-glucopyranosyl-(1→2)-β-D-fructofuranoside***

4.2.3. The main representatives of disaccharides

Maltose (malt sugar or corn sugar; α-D-Glucopyranosyl-(1→4)-α-D-glucose) consists of two glucose molecules linked by an α-1,4-glycosidic bond:

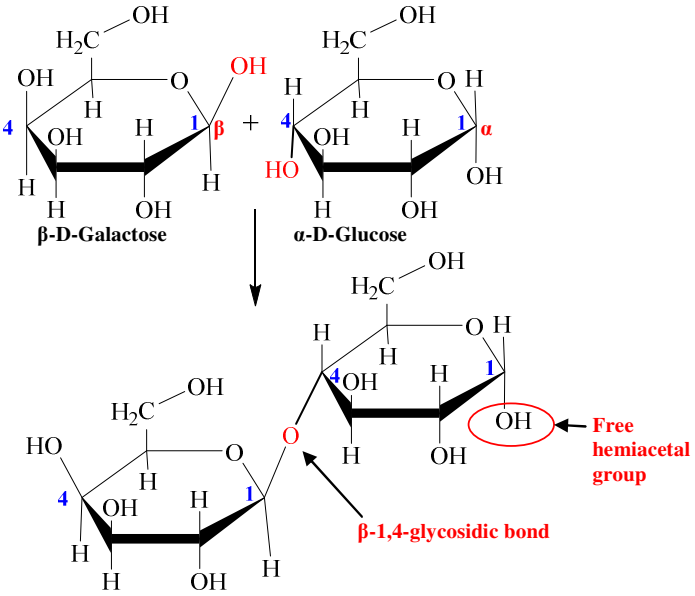


Maltose, a disaccharide

It comes *from partial hydrolysis of starch* by the enzyme *amylase*, which is in saliva and also in grains (like barley). *Maltase* is the enzyme that hydrolyzes maltose.

Since one of the glucose monomers has a free hemiacetal group (at C1) and can open to give a free aldehyde group to oxidize - maltose can undergo mutarotation and *is a reducing sugar***.**

Lactose (milk sugar; β-D-galactopyranosyl-(1→4)-α-D-glucose) consists of one glucose molecule and one galactose molecule linked by a β-1,4 glycosidic bond:

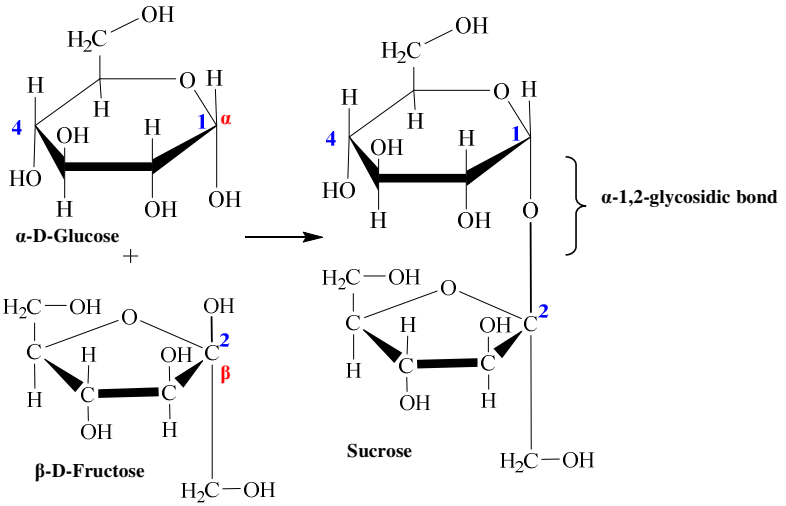


Lactase is the enzyme in humans that hydrolyzes lactose back to its two monosaccharides. Some people don't produce enough lactase, the enzyme that hydrolyzes lactose, and so can't digest lactose. Many adults become *lactose intolerant*, and develop abdominal cramps, nausea and diarrhea. Lactase can be added to milk products (or taken as a supplement) to combat this problem

Since the glucose component has free hemiacetal group (at C1) and can open to give a free aldehyde group to oxidize – lactose can undergo mutarotation and *is a reducing sugar*.

Sucrose (table sugar; α-D-glucopyranosyl-(1→2)-β-D-fructofuranoside) consists of one glucose molecule and one fructose molecule linked by an α-1,2-glycosidic bond. Sucrose is formed by plants but not by higher animals. Sucrose is a major intermediate product of photosynthesis; in many plants it is the principal form in which sugar is transported from the leaves to other portions of plants via their vascular systems. It is the most abundant disaccharide and is commercially produced from sugar cane and sugar beets.

Since the glycosidic bond in sucrose involves both OH-groups linked with anomeric carbons (C1 in glucose and C2 in fructose), neither glucose nor fructose can be changed to a aldehyde or ketone as well as neither monosaccharide can undergo mutorotation, and therefore *sucrose is not a reducing sugar:*



Sucrose is hydrolyzed by the enzyme *sucrase*, which is secreted in the small intestine. The glucose and fructose can then be absorbed into the bloodstream (disaccharides are too large to be absorbed).

Some molecules bind to a receptor on a taste bud cell of the tongue and give sweet taste. When this molecule binds, a nerve impulse passes from the taste bud to the brain, where the molecule is interpreted as being sweet. The degree of sweetness differs for different sugars. Sucrose is very sweet, but contains many calories. Sucrose is used as the prototypical example of a sweet substance. Sucrose in solution has a sweetness perception rating of 1, and other substances are rated relative to this (Tab. 4.4). For example, another sugar, fructose, is somewhat sweeter, being rated at 1.7 times the sweetness of sucrose.

Table 4.4.
Relative sweetness of some carbohydrates and artificial sweeteners

Compound	Relative sweetness
Sorbitol	0,60
Glucose	0,75
Sucrose	1,00
Fructose	1,75
Aspartame	150
Saccharin	350
Sucralose	600

To reduce caloric intake, many **artificial sweeteners** have been developed such as aspartame, saccharin and sucralose.

Aspartame (Fig. 4.26) was discovered in 1965 by a group of scientists working for G.D. Searle, that later become a Monsanto company, to develop a new anti-ulcer drug based on a tetrapeptide. It is 150 times sweeter than sucrose. It undergoes slow hydrolysis in solution and it also decomposes with heat. So, we cannot use for soft drinks and cooking purposes. It is hydrolyzed into phenylalanine, which cannot be processed by those individuals with the condition *phenylketonuria.*

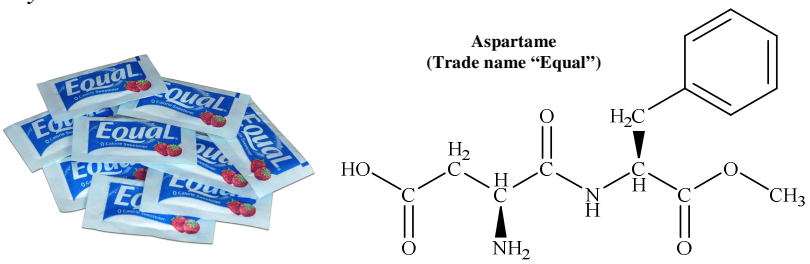


Fig. 4.26. Aspartame structure and examples of aspartame-contained products

Saccharine (Fig. 4.27) is the first synthetic sweetener, was discovered by Ira Remsen and his student Constantine Fahlberg at Johns Hopkins University in 1878. It was used extensively during

World War I. There were concerns in the 1970s that saccharin causes cancer. It is 350 times sweeter than glucose.

Saccharin (Trade name "Sweet'n Low")

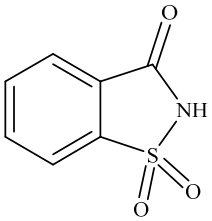
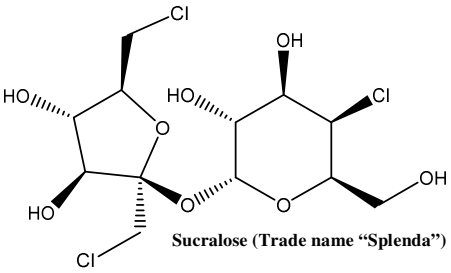


Fig. 4.27. Saccharine structure and examples of saccharine-contained products

Sucralose has a very similar structure to sucrose (Fig. 4.28) as it is a trichloro derivative of latter. Sucralose is 600 times sweeter than sucrose. It tastes like sugar and used for cooking and baking.

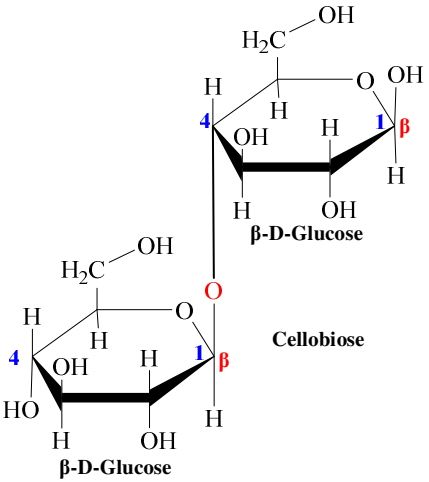


Sucralose (Trade name "Splenda")

Fig. 4.28. Sucralose structure and examples of sucralose-contained products

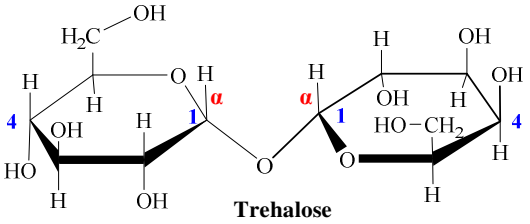
Cellobiose (β-D-Glucopyranosyl-(1→4)-β-D-glucose) is also made from two glucose molecules. Unlike maltose it is bonded by a

β-1,4-glycosidic linkage, making it a non-digestable sugar due to the equatorial linkage:



Cellobiose **is a reducing sugar** because it has free hemiacetal group (OH-group linked with anomeric carbon at C1) that can open to give a free aldehyde group to oxidize.

Trehalose (α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside) is a disaccharide formed by an α,α-1,1-glucoside bond between two α-glucose units:



It can be synthesised by bacteria, fungi (for example, it is prevalent in some mushrooms), plants, and invertebrate animals. Trehalose is a major constituent of the circulating fluid (hemolymph) of insects, in which it serves as an energy storage compound. **It is a non-reducing sugar**, because the glycosidic bond in its molecule

involves OH-groups of both anomeric carbons (C1) in glucose monomers. it is rapidly broken down into glucose by the enzyme *trehalase*, which is present in the brush border of the intestinal mucosa. Trehalase deficiency is unusual in humans.

4.3. Oligosaccharides

Oligosaccharides are small polymers (typically consist of three to ten monosaccharide units).

4.3.1. Mammalian oligosaccharides

Mammalian oligosaccharides often found attached to polypeptides or lipids in glycoproteins and some glycolipids, respectively. They are attached to membrane and secretory proteins found in endoplasmic reticulum and Golgi complex of various cells.

There are two classes of mammalian oligosaccharides - N-linked and O-linked ones (Fig. 4.29).

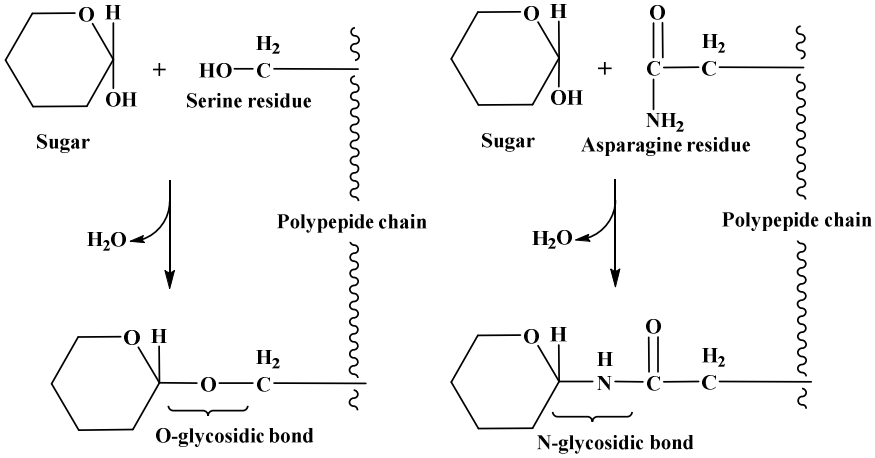


Fig. 4.29. The principle of N- and O-linked oligosaccharides formation in glycoproteins

N-linked oligosaccharides are always pentasaccharides attached to asparagine via a beta linkage to the amine nitrogen of the side chain. **O-linked oligosaccharides** are generally attached to threonine or serine on the alcohol group of the side chain.

Oligosaccharides can have many functions including cell recognition and cell binding. For example, glycolipids have an important role in the immune response.

4.3.2. Plant oligosaccharides

Raffinose is a trisaccharide composed of galactose, glucose, and fructose (Fig. 4.30).

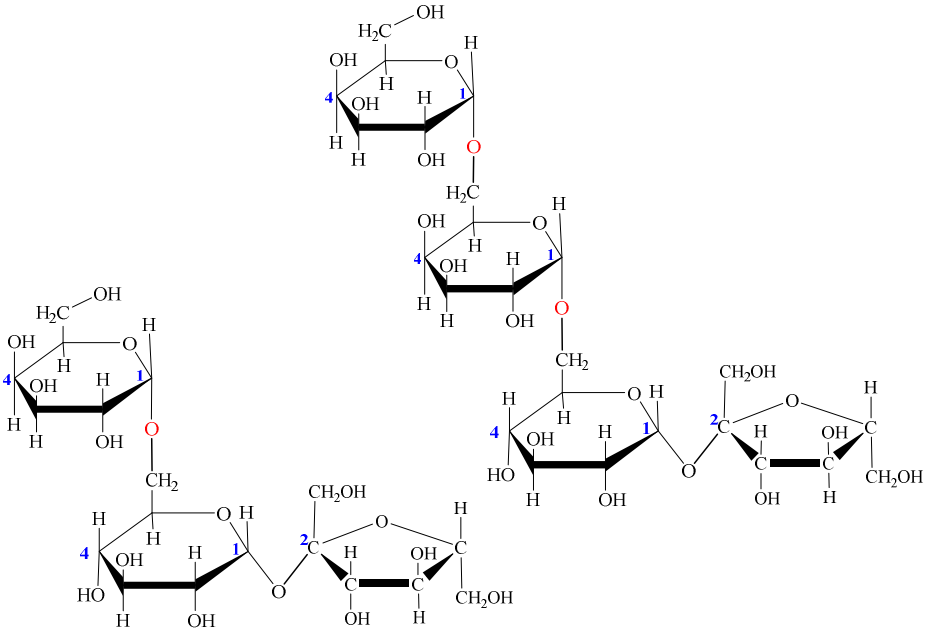


Fig. 4.29. The structures of plant oligosaccharides raffinose (left) and stachyose (right)

It can be found in beans, cabbage, brussels sprouts, broccoli, asparagus, other vegetables, and whole grains. Raffinose can be hydrolyzed to D-galactose and sucrose by the enzyme *α-galactosidase*. This enzyme also metabolizes such plant oligosaccharides as tetrasaccharide **stachyose** and pentasaccharide **verbascose** (Fig. 4.29). These compounds together with raffinoze belong to **the raffinose family of oligosaccharides (RFOs)** and are alpha-galactosyl derivatives of sucrose (Fig. 4.30). RFOs are almost ubiquitous in the plant kingdom, being found in a large variety of seeds from many different families. α-galactosidase that breaks down RFOs isn't found in the human digestive tract, so these oligosaccharides pass undigested through the stomach and upper intestine.

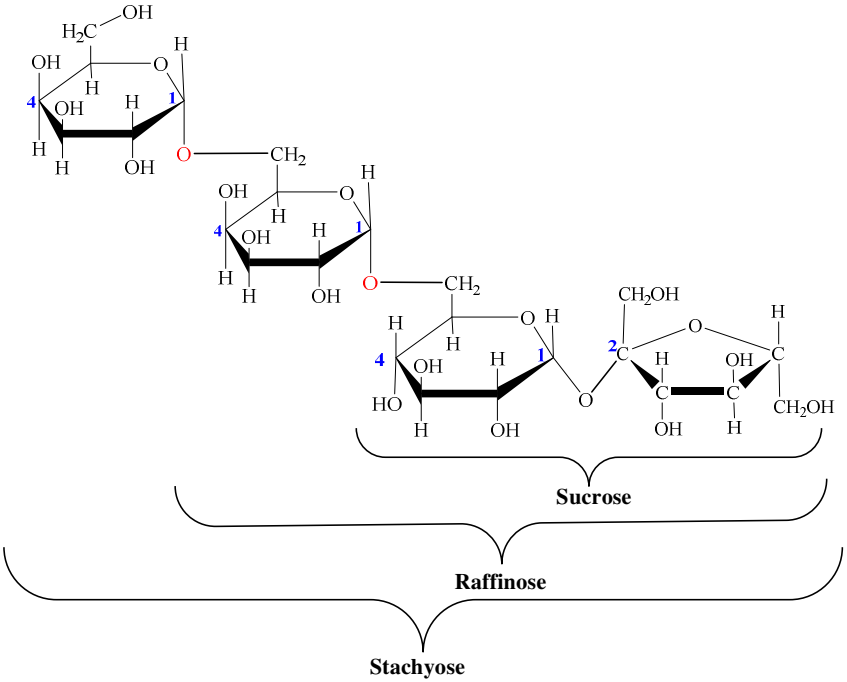


Fig. 4.30. The representatives of plant oligosaccharides from RFOs that are derivatives of sucrose

In the lower intestine, they *are fermented by gas-producing bacteria* that possess the α-galactosidase and make CO_2, methane or H_2 leading to the flatulence commonly associated with eating beans and other vegetables.

4.4. Polysaccharides (Glycans)

Polysaccharides are the polymers of multiple units of monosaccharides linked together by glycosidic linkages. Their classifications can be based on (Fig. 4.31):

- **their functions** (*storage* or *structural*);
- **main monosaccharide components** (*homopolysaccharides* and *heteropolysaccharides*);
- **the amount of branching in the polymer** (*branched* and *unbranched*);
- **the type of links between monosaccharides** (polysaccharides are formed by glycosidic bonds on different carbons of the monosaccharides).

Homopolysaccharides are composed of one type of monosaccharids whereas heteropolysaccharides can be composed of two or more types of monosaccharides. In general polysaccharides may be branched or unbranched; heteropolysaccharides more often are the structural compouns (for example, glycoaminoglycans (GAGs) and murein).

Unlike proteins, polysaccharides generally do not have definite molecular weights. This difference is a consequence of the mechanisms of assembly of the two types of polymers. Proteins are synthesized on a template (messenger RNA) of defined sequence and length, by enzymes that copy the template exactly. For polysaccharide synthesis, there is no template; rather, the program for polysaccharide synthesis is intrinsic to the enzymes that catalyze the polymerization of monomeric units. For each type of monosaccharide to be added to the growing polymer there is a separate enzyme, and each enzyme acts only when the enzyme that inserts the preceding subunit has acted. The alternating action of several enzymes produces a polymer with a precisely repeating

sequence, but the exact length varies from molecule to molecule, within a general size class. The mechanisms that set the upper size limits are unknown.

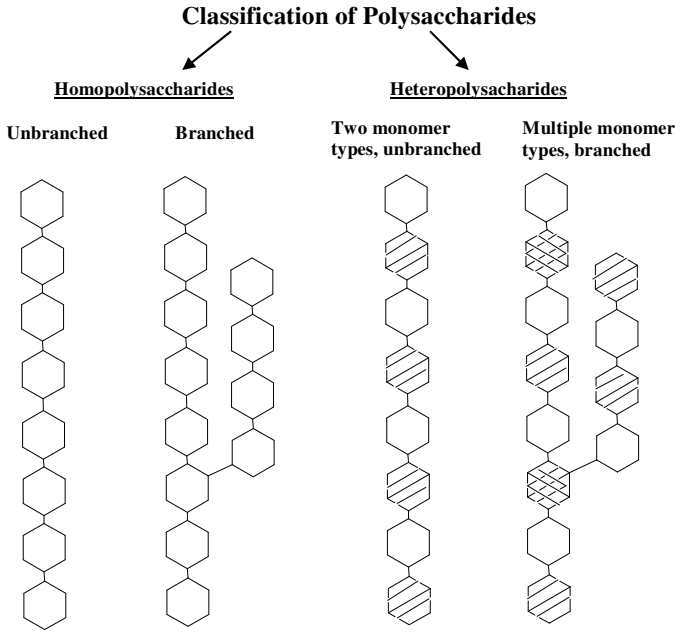


Fig. 4.31. The main aspects of polysaccharides classification

4.4.1. Homopolysaccharides

Homopolysaccharides are composed of one type of monosaccharids. They can be branched or unbranched and can functioning as storage of energy (starch, glycogen, and dextrans) or have structural roles (cellulose, chitin).

4.4.1.1. Starch and glycogen as storage homopolysaccharides.

Their function is to store the energy.

Glycogen and starch ingested in the diet are hydrolyzed by *α-amylases* - enzymes of saliva and intestinal juice, that break *α-(1→4)-glycosidic bonds* between glucose units. Therefore, digestion of these homopolysaccharides occurs in the mouth and in the intestine. They are hydrolyzed to *dextrins* (the mixture of less long polymers of α-D-glucose units) and disaccharide *maltose* (**partial hydrolysis**) and to monosaccharide *α-D-glucose* (**complete hydrolysis**). Therefore, *α-D-glucose is their monomeric unit* and *such units in these polymers are linked by α-(1,4)- and α-(1,6)-glycosidic bonds*.

The (α→4) linkages of starch and glycogen cause these polymers to form *a tightly coiled helical structure (the three-dimensional structure).* This compact structure produces the dense granules of stored starch or glycogen seen in many cells. Starch and glycogen molecules are heavily hydrated because they have many exposed hydroxyl groups available to hydrogen bond with water.

Starch is a homopolysaccharide, a polymer of glucose with formulae $(C_6H_{10}O_5)_n$ that is constructed by α-D-glucose units bonded by α-glycosidic bonds. Starch serves as a *major source of energy storage in plants* – it is found chiefly in the seeds, fruits, tubers, roots, and stem pith of plants, notably in corn, potatoes, wheat, and rice, and varying widely in appearance according to source but commonly prepared as a white amorphous tasteless powder. It is also a main nutrition substance to most animals, including humans.

Starch can be separated into two fractions – amylose and amylopectin. **Amylose** is composed of *continuous, unbranched (linear) chains* of up to 4000 D-glucose units joined by *α-(1,4)-glycoside bonds*. Such chains vary in molecular weight from a few thousand to 500,000 (Fig. 4.32).

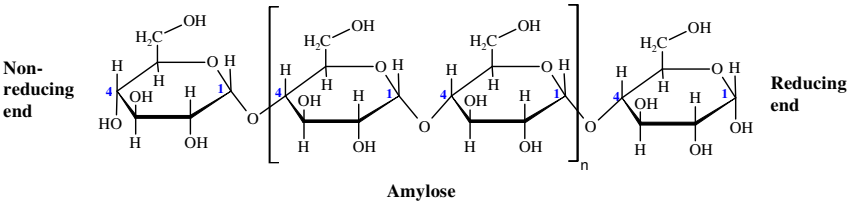


Fig. 4.32. Amylose structure

Amylopectin also has a high molecular weight (up to 1 million). It is *a highly branched polymer* of D-glucose ***containing both α-(1,4) and α-(1,6) glycosidic linkages*** (Fig. 4.33).

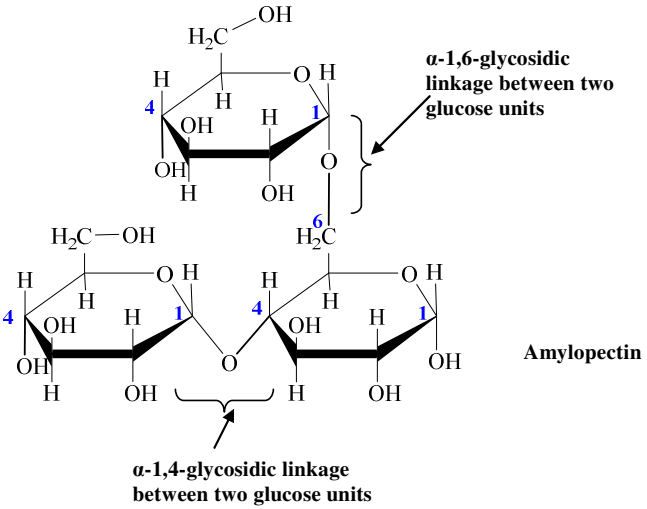


Fig. 4.33. Amylopectin structure

Its linear chains consist of 24-30 units of D-glucose joined by α-1,4-glycoside bonds; branches created by α-1,6-glycoside bonds. These branch points prevent helix formation.

Glycogen, also known as "animal starch", is the storage form of glucose and energy in animals and humans which is analogous to the starch in plants. Glycogen is synthesized and stored mainly in the liver and the muscles. The total amount of glycogen in the body of a well-nourished adult is about 350 g (about 3/4 of a pound) divided almost equally between liver and muscle. In hepatocytes glycogen is found in large granules, which are themselves clusters of smaller granules composed of single, highly branched glycogen molecules with an average molecular weight of several million (Fig. 4.34).

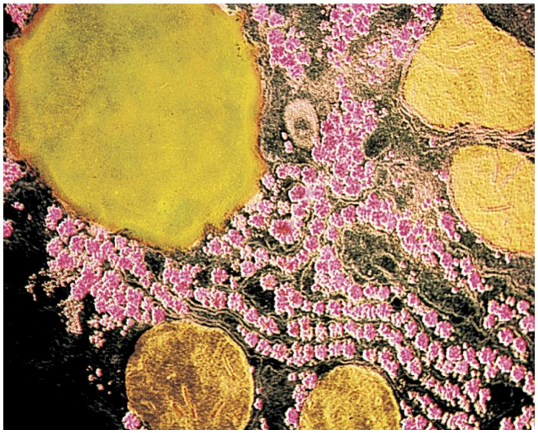


Fig. 4.34. Glycogen (pink granules) in liver cells

Such glycogen granules also contain, in tightly bound form, the enzymes responsible for the synthesis and degradation of glycogen.

Glycogen is constructed by α-D-glucose units bonded by α-glycosidic bonds. Structurally, *glycogen is very similar to amylopectin*; however, it has even more branching and more glucose residues (1,700-600,000 units) are present than in amylopectin and so is more compact than starch. In the linear regions of glycogen D-glucose units are joined by α-(1,4)-glycosidic linkages; the branches are formed by α-(1,6) glycosidic linkages. In glycogen, the branches

occur at intervals of 8-10 glucose units, while in amylopectin the branches are separated by 24 - 30 glucose units (Fig. 4.35).

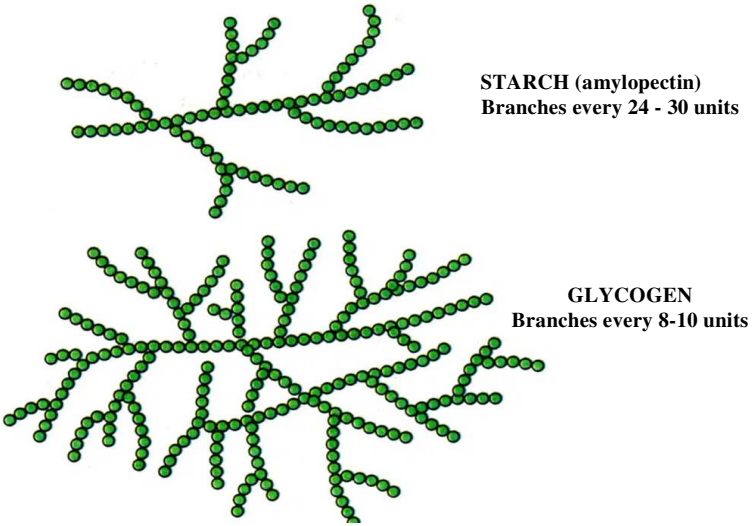
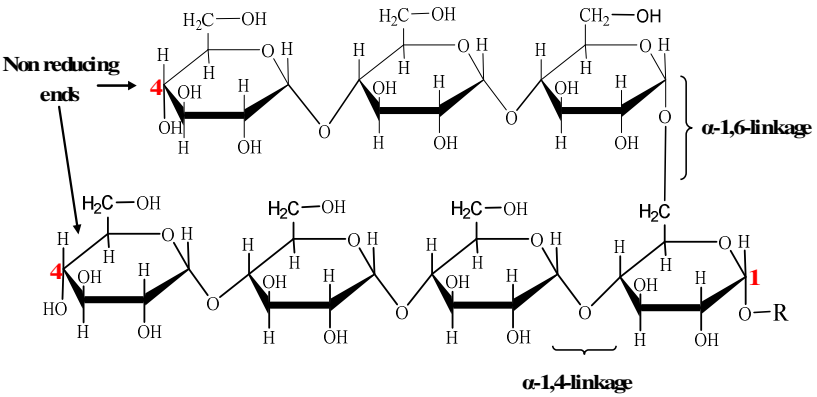
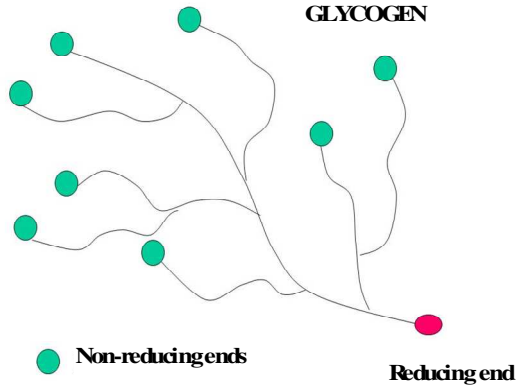


Fig. 4.35. The comparative structures of amylopectin and glycogen

Branched structure allows several sites for simultaneous synthesis and degradation (and therefore branching speeds up degradation) and makes glycogen compact and coiled is so an efficient way to store glucose.

Glycogen and starch have *as many non-reducing ends* (glucose residue with free –OH on C4) *as they have branches*, but *only one reducing end* (glucose residue with free –OH on C1 (anomeric carbon)). No matter how large the glycogen molecule is or how many branches it has, each branch ends with a non-reducing sugar residue (one without a free anomeric carbon). When glycogen and starch are broken down to be used as an energy source, glucose units are removed from the nonreducing ends by enzymes:



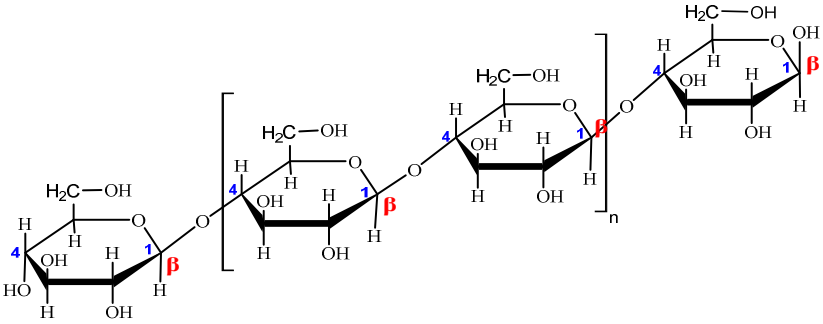


4.4.1.2. Cellulose

Cellulose - primary component of plant cells that has a structural roles. This fibrous, tough, water-insoluble substance is found in plants as microfibrils (2-20 nm diameter and 100 - 40 000 nm long, corresponding to approximately 2800 D-glucose units per molecule) that form the structurally strong framework in the cell walls.

Cellulose is hydrolyzed (not in human organism) to disaccharide *cellobiose* (**partial hydrolysis**) and *β-D-glucose* (**complete hydrolysis**), so β-D-glucose is its monomeric unit.

Cellulose is a *linear, unbranched polymer of β-D-glucose units linked by β-(1,4)-glycoside bonds* in a linear fashion:



The three-dimensional structures of starch, glycogen and cellulose differ each from other (Fig. 4.36). The linkage by α-1,4-glycosidic bonds in amylose fraction of starch and also in glycogen give them a very different structure than cellulose (linked by β-1,4-glycosidic bonds). *The α-1,4-glycosidic linkage results in a hollow helical structure* that is more suitable for energy storage. *The β-1-4-glycosidic bonds allows cellulose to form linear chains* that are stabilized by hydrogen-bonding with adjacent chains to form tensile fibers.

Intermolecular and intramolecular hydrogen bonds allow the more hydrophobic part of cellulose to stack, leading to complete insolubility of cellulose in normal aqueous solution. Although an individual hydrogen bond is relatively weak, many such bonds acting together can impart great stability to certain conformations of large molecules.

Most animals cannot digest cellulose as a food, and in the diets of humans this part of our vegetable intake functions as roughage and is eliminated largely unchanged. Termites readily digest cellulose (and therefore wood), but only because their intestinal tract harbors a symbiotic microorganism, *Trichonympha*, which secretes beta-glycosidase enzyme *cellulase* that hydrolyzes β-(1,4) linkages between glucose units. Wood-rot fungi and bacteria also produce cellulase.

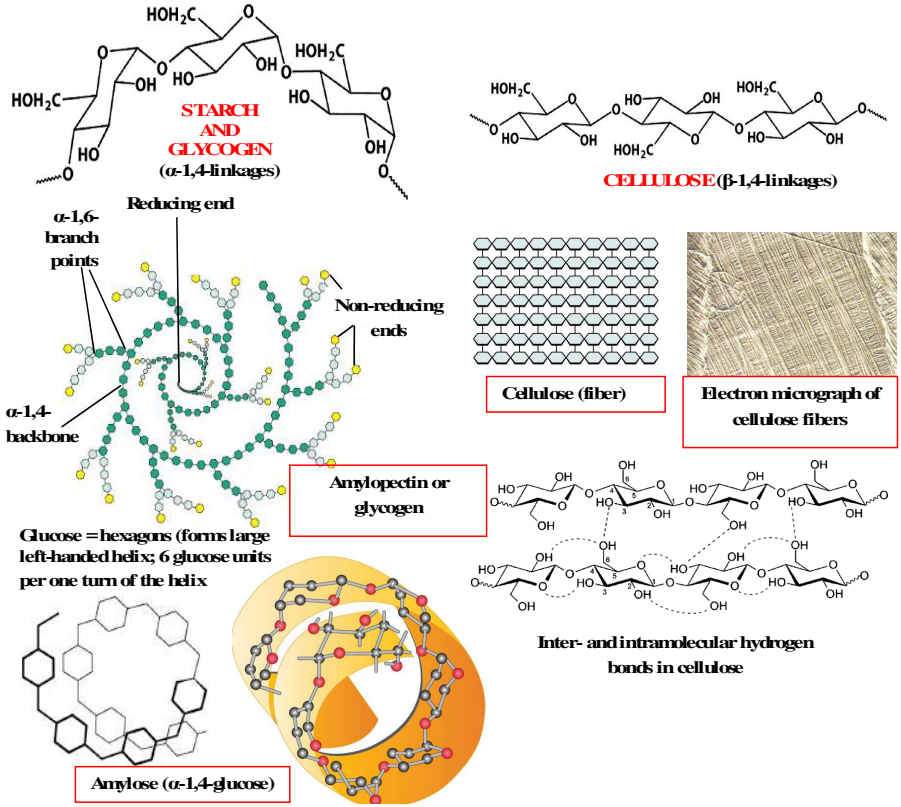


Fig. 4.36. The differences between the three-dimensional structures of starch, glycogen and cellulose

The only vertebrates able to use cellulose as food are cattle and other ruminant animals (sheep, goats, camels, giraffes). The extra stomachs (rumens) of these animals teem with bacteria that secrete cellulase.

Although cellulose in the human body is not digested, it is a necessary component of a healthy diet. Plant (or dietary) fibers

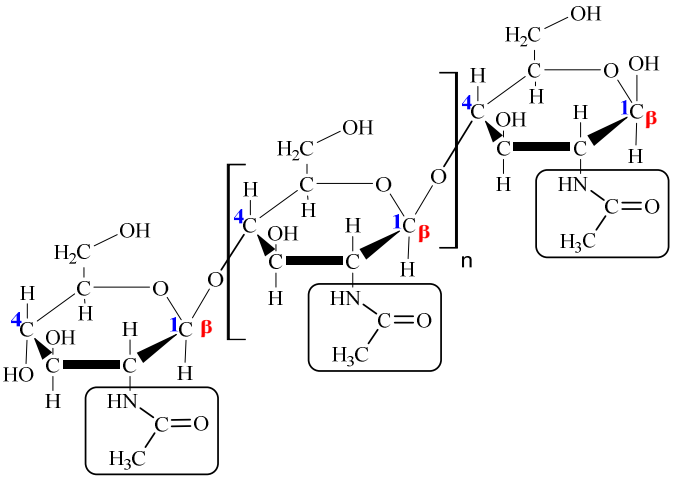
formed by cellulose adsorb a large amount of water and increase the volume of the food lobes, which stimulates intestinal peristalsis. Unperfected cellulose contributes to the formation of feces. Dietary fibers also absorb a number of toxic compounds, including carcinogens, which facilitates their excretion from the body.

In humans, dietary fibers contribute to the prevention of disorders of the colon, prevent the emergence of a number of its diseases, including cancer.

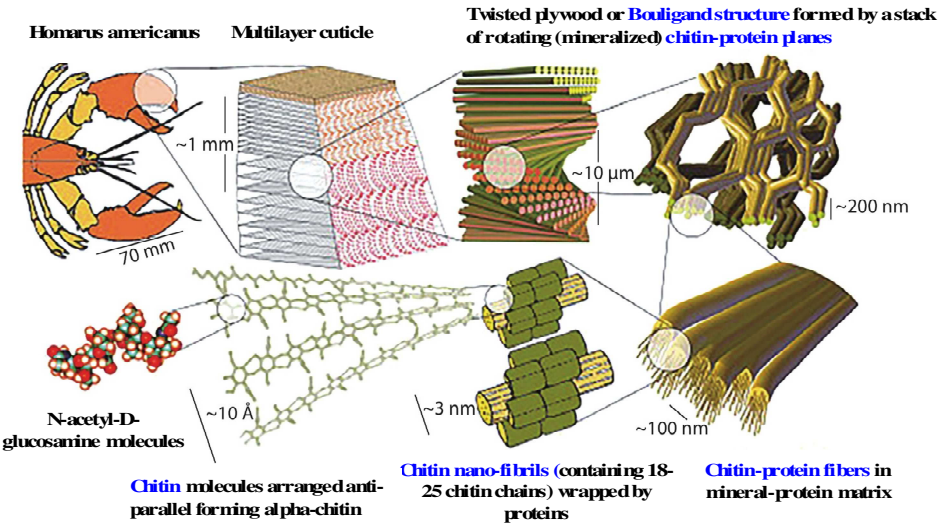
In addition to cellulose, other plant polysaccharides also are got from food to the digestive canal of the human. These polysaccharides include *heteropolysaccharides hemicelluloses and pectins* that along with cellulose form dietary fibers.

4.4.1.3. Chitin

Chitin is a homopolysaccharide made from repeating units of a derivative of glucose, *N-acetyl-D-glucosamine*. These units are connected linearly through *β-(1,4)-glycoside bonds*:

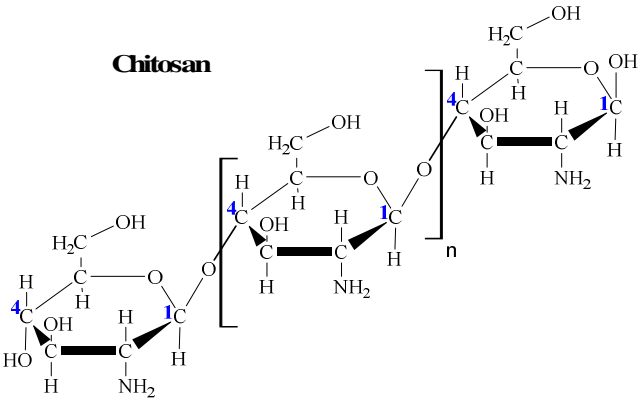


Its structure and linkages are similar to that of cellulose, except that the hydroxyl group on the 2' carbon of glucose is replaced by an acetylamine group. Therefore chitin forms fiber similar to those of cellulose, and *like cellulose is indigestible by vertebrate animals*:



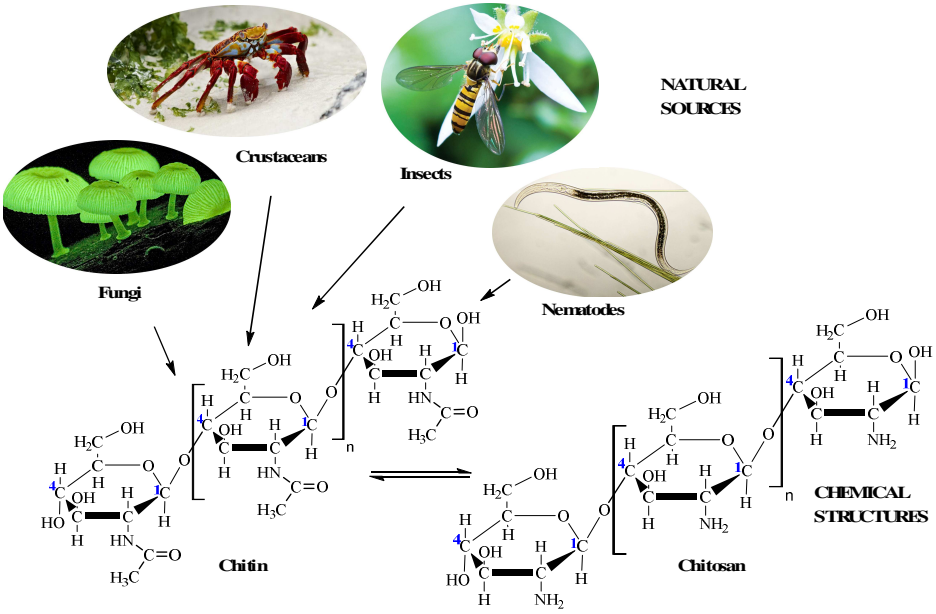
Chitin is a very important structural component making up the cell walls of mushrooms, the exoskeletons of lobsters, crabs, insects, shells and spiders and is probably the second most abundant polysaccharide, next to cellulose, in nature. Chitin is a very firm material and it helps to protect an insect against harm and pressure.

The main derivative of chitin (*deacetylated chitin*) is called **chitosan:**



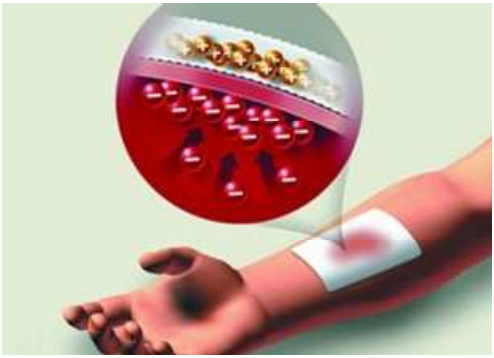
It is a linear polysaccharide composed of *β-(1→4)-linked D-glucosamine* (deacetylated unit of chitin). The molecular weight of commercially produced chitosan is between 3800 and 20,000 Da.

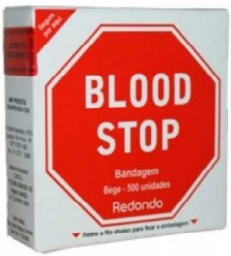
A common method for the chitosan synthesis is the deacetylation of chitin (which is the structural element of the exoskeleton of crustaceansand and cell walls of mushrooms) using sodium hydroxide in excess as a reagent and water as a solvent:



Chitosan can be used both internally and externally to treat wounds, obesity, high cholesterol problems.

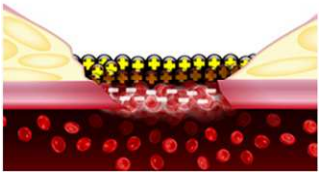
Wounds and hemostatic agents based on chitosan. Chitosan's properties allow it to rapidly clot blood, and have recently gained approval in the United States and Europe for use in bandages and other hemostatic agents:





Chitosan hemostatic products have been sold to the U.S. Army and are currently used by the UK military. Chitosan is hypoallergenic and has natural antibacterial properties, which further support its use in field bandages. Chitosan's hemostatic properties also allow it to reduce pain by blocking nerve endings.

Celox™ is a *granular chitosan* available in powder, gauze, and nasal tampon forms as a hemostatic agent. It is made of a proprietary composition which contains chitosan. Chitosan hemostatic agents are often chitosan salts made from mixing chitosan with an organic acid (such as succinic or lactic acid). The hemostatic agent works by an interaction between the cell membrane of erythrocytes and platelets (negative charge) and the protonated chitosan (positive charge) leading to involvement of platelets and rapid thrombus formation (Fig. 4.37).



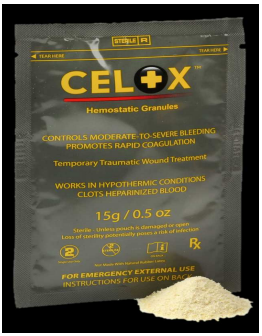


Fig. 4.37. The mechanism of hemostatic agents based on chitosan action

Therefore these hemostatic agents work independently of the body's normal clotting mechanism.

Chitosan and Weight Loss. Chitosan is marketed in a tablet form as a "fat binder".

It can't be digested by human organism, so when a person consumes chitosan, it acts in a similar way to the plant fiber:

- creates a feeling of fullness in stomach, which helps a person eat less;
- blocks the absorption of fats and carbohydrates in intestine;
- reabsorbs both some of the fat as cholesterol from the food.

In these ways chitosan helps people lose weight.

4.4.1.4. Inulin

Inulin is linear (no branching) plant homopolysaccharide formed by *β-(2,1)-linked fructofuranose units* (about 30–35 fructose units per molecule). On hydrolysis, however, inulin also yields a small amount of glucose besides fructose molecules. It is now thought that *there are 2 glucose units in the inulin molecule*, one somewhere in the centre and the other at the reducing end of the chain. It is a nonreducing sugar (Fig. 4.38).

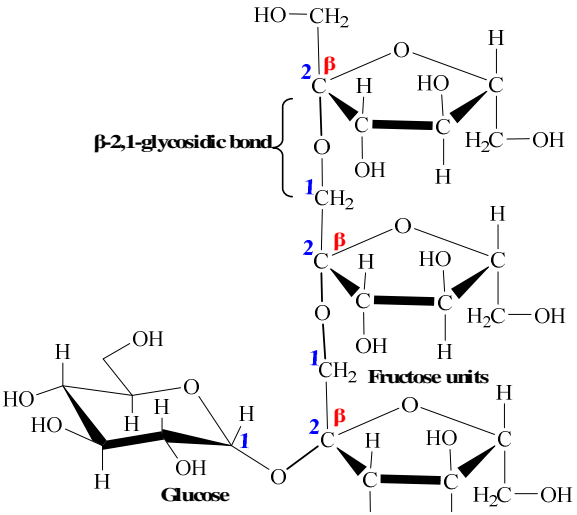


Fig. 4.38. The structure of inulin

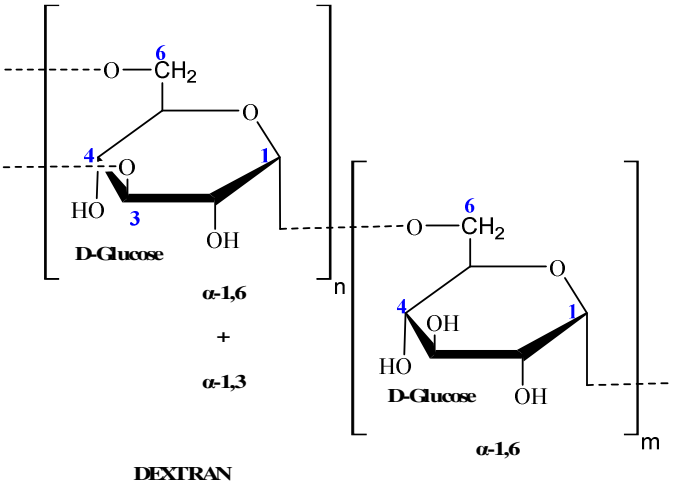
Inulin is generally abundant as a storing energy form instead of starch in roots and rhizomes in more than 36,000 species of plants such as wheat, onion, garlic, dandelions, bananas, asparagus, dahlias, chicory and artichokes – these plants are the source of inulin:

Jerusalem artichokes

Inulin is not digested by the endogenous enzymes of the human digestive system therefore it is considered as dietary fiber. It is used in medicine as diagnostic agent for the evaluation of glomerular filtration rate (GFR) in renal function tests.

4.4.1.5. Dextrans

Dextrans are bacterial and yeast homopolysaccharides made up of linear chain of *α-D-glucose residues* linked by *α-(1,6)-bonds*. They all have *branches formed by α-(1,3)-, α-(1,2) or α-(1,4)-bonds*:

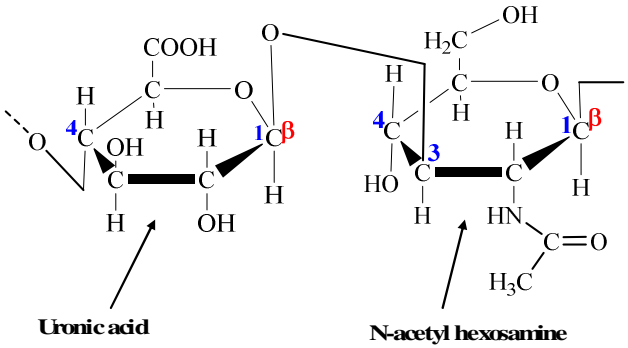


Dental plaque, formed by bacteria growing on the surface of teeth, is rich in dextrans.

Dextrains are used in clinical medicine as plasma and blood substitutes (pharmaceuticals *Polyglyukin* and *Reopoliglyukin*). Artificial dextrans are utilized in many commercial products such as *sephadex* which serve in biochemical practice as gels for the fractionation of proteins by size-exclusion chromatography.

4.4.2. Heteropolysaccharides

The main **plant heteropolysaccharides** are *pectins, hemicelluloses, natural gums* and *agar*. They are the components of dietary fibers. **Bacterial heteropolysaccharides** are components of bacterial cell wall. **Animal heteropolysaccharides** can be represented by *glycosaminoglycans*, or *mucopolysaccharides* - unbranched polymers, constructed of repeatable disaccharide units. The disaccharide unit consists of two components, one of which is a representative of *uronic acids* (glucuronic or iduronic), and the other is *N-acetylated derivative of hexosamines* (glucosamine, galactosamine):



4.4.2.1. Pectins

Pectins, also known as pectic polysaccharides, are the structural heteropolysaccharides contained in the cell walls of plants. Pears, apples, guavas, quince, plums, gooseberries, and oranges and other citrus fruits contain large amounts of pectins. These substances enter the human body with plant foods. Pectins are highly water-soluble and are almost completely metabolized by colonic bacteria.

Pectins are complex branched heteropolysaccharides primarily containing a *linear backbone from α-(1-4) polygalacturonic acid. residues*, some of which can be methylated (Fig. 4.39).

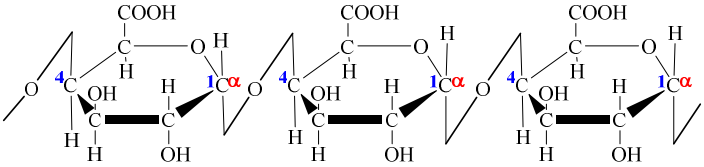


Fig. 4.39. The structure of α-(1-4) polygalacturonic acid linear backbone of pectines

Some of pectins are also characterized by the presence of other saccharide residues - such *D-xylose*, *α-L-rhamnose*, *D-galactose*, *L-arabinose* branching from a backbone of D-galacturonic acid residues.

In medicine, pectin increases viscosity and volume of stool so that it is used against constipation and diarrhea. The typical property of pectins is the ability of their solutions to gelate. Due to their gelling behaviour, these soluble polysaccharides may decrease the rate of gastric emptying and influence small intestinal transit time. This explains their hypoglycemic properties. Pectins are also used to make gels and are the basis of a number of drugs as well as to heavy metal removal from biological systems.

4.4.2.2. Natural gums

Natural gums are the types of plant fibres that are not cell wall components but are formed in specialized secretory plant cells:

Certain tropical trees is a spontaneous formation of exudates gums, which are exuded as viscous fluids at sites of injury and after dehydration give hard, clear nodules rich in polysaccharides.

Natural gums are highly branched plant polysaccharides that contain monosaccharide residues (*D-galactose*, *D-glucose, L-arabinose, L-rhamnose*, etc.) and *uronic acids*.

The one of the main representatives of natural gums is **gum arabic** which includes the residues of *L-arabinose, D-galactose, D-glucuronic acid and methylpentose*. Natural gums are capable to cause a large increase in a solution's viscosity, even at small concentrations; they also can form gels, bind water and other organic materials. They are used in pharmaceutical practice in the manufacture of medical emulsions and are also the component of dietary fibres.

4.4.2.3. Hemicelluloses

Hemicelluloses are the world's second most abundant carbohydrate (after cellulose). Unrefined cereals, some fruits and vegetables and whole wheat are rich sources of hemicelluloses. They function as storage and supporting substances in plants.

They are characterized as a group of cell wall branched, amorphous heteropolysaccharides with low molecular weight that are neither cellulose (polymer of β-(1,4)-D-glucose) nor pectins (polymers of galacturonic acid) and *have a β-(1→4)-linked backbones of xylose, mannose or glucose, linked in equatorial configuration*. Other monomer units founded in hemicelluloses are *galactose, rhamnose*, and *arabinose units*. Hemicelluloses contain *most of the D-pentose sugars*, and occasionally small amounts of L-sugars as well. Xylose is in most cases the sugar monomer present in the largest amount, although in softwoods mannose can be the most abundant sugar. Not only regular sugars can be found in hemicellulose, but also their acidified form, for instance *glucuronic acid* and *galacturonic acid* can be present.

There are next differences between cellulose and hemicelluloses. While cellulose is crystalline, strong and resistant to hydrolysis, hemicelluloses have a random, amorphous structure with little strength. They are easily hydrolyzed by dilute acid or base as well as

the hemicellulase enzymes. Cellulose contains only β-D-glucose, whereas hemicelluloses are mixed polymers contain many different sugar monomers (glucose, xylose, mannose, galactose, rhamnose and arabinose). Apart from cellulose that is a long unbranched polymer, hemicelluloses are either heavily branched or have short side-chains. While cellulose has a very high degree of polymerization (7,000–15,000 glucose molecules per polymer) and so has a high molecular weight, hemicelluloses consist of 500–3,000 sugar units and are low molecular weight polymers

Therefore hemicelluloses together with cellulose and pectins are localized in the plant cell wall and are the main carbohydrate components of dietary fibers. Hemicelluloses are embedded in the cell walls of plants, sometimes in chains that form a “ground” - they bind with pectin to cellulose to form a network of cross-linked fibres (Fig. 4.40).

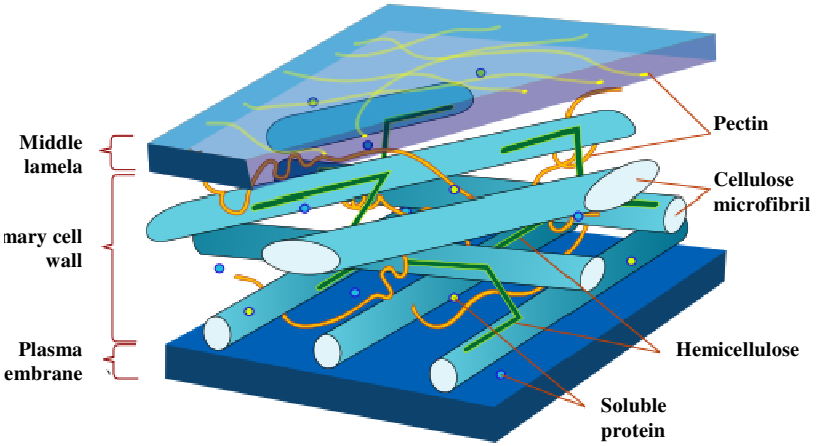


Fig. 4.40. The schematic structure of plant cell wall and the role of cellulose, pectins and hemicelluloses in its stabilization

Dietary fibers consist of non-digestible plant carbohydrates. They are resistant to digestion (are not hydrolyzed by the endogenous enzymes) and absorption in the human small intestine,

but some of them can be partially or common hydrolyzed by colonic bacteria.

Dietary fibers absorb water and can significantly increase stool weight and regularity as well as may attenuate the absorption. Because of this these structures promote beneficial physiologic effects including losing weight, reducing risk of colon cancer, blood cholesterol and glucose attenuation.

The main hemicelluloses representatives include *xylans*, *mannans* and *xyloglucans*.

Xylans are the most abundant and widely distributed carbohydrates after cellulose. Xylans have *a linear β-(1→4)-D-xylopyranose backbone* (consists of an aldopentose D-xylose in β-1→4 linkage) and *the side chains of 4–O–methylglucuronic acid and/or arabinose*. *Glucuronoxylans* are the most common hemicelluloses in the cell walls of hardwoods; its backbone consists exclusively of xylose (Fig. 4.41).

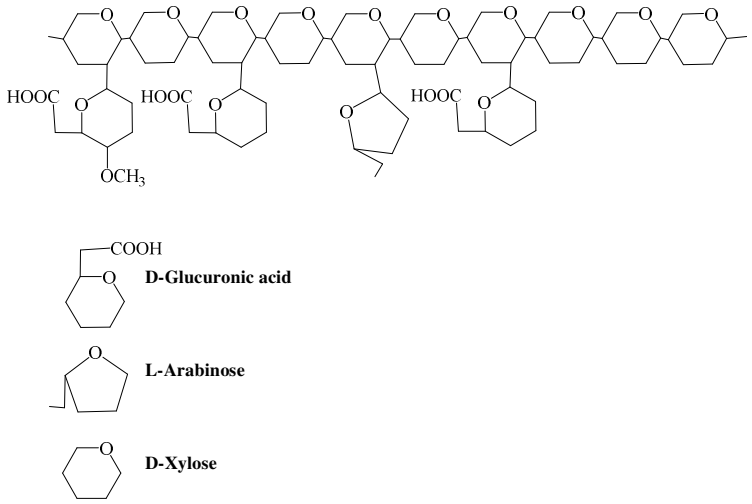


Fig. 4.41. The schematic structure of glucuronoxylanes

Mannans are the heteropolymers possessing a *β-(1→4)-D-mannopyranose backbone*. *Galactoglucomannans* are the most

common hemicelluloses in the cell walls of softwoods; mannose : glucose : galactose ratio in softwoods is 3.5-4.5 : 1 : 0.5-1.1 (Fig. 4.42).

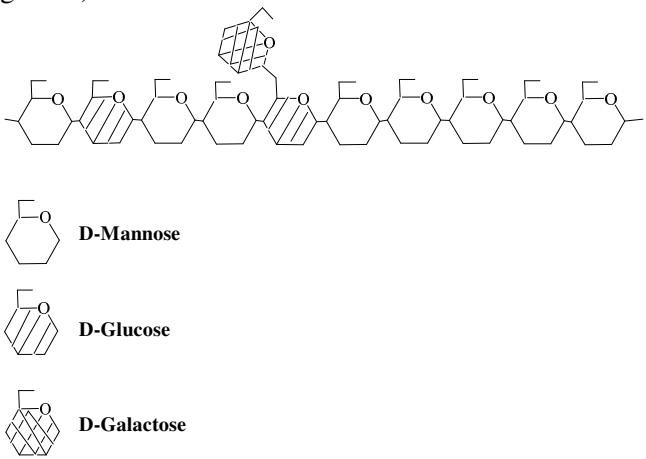


Fig. 4.41. The schematic structure of galactoglucomannans

Galactomannans are abundant in cell walls of storage tissues, notably those from the endosperm of leguminous seeds (guar, locust bean, tara gum, etc.). The amount of galactose residues influences solubility, viscosity and interactions with other polysaccharides (Fig. 4.42).

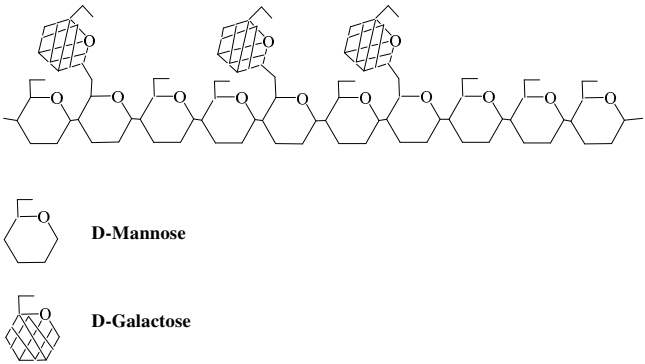
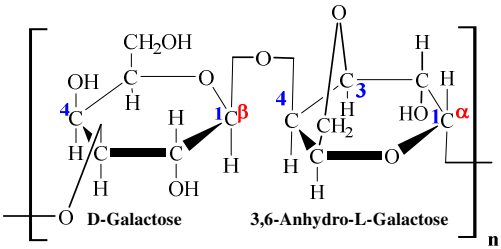


Fig. 4.42. The schematic structure of galactomannans

4.4.2.4. Agar

Agar, currently known as *"agar-agar"* (from Malaysian name of seaweeds), occurs as structural carbohydrate in the cell walls of agarophytes algae belonging to *Gelidium*, *Pterocladia*, *Petrocladiella*, *Gracilaria* and *Hypnea* species. Agar is a complex mixture of polysaccharides composed of two major fractions - *agarose*, a neutral polymer, and *agaropectin*, a charged, sulfated polymer.

Agarose - *the gelling fraction* - is a neutral linear molecule essentially free of sulfates, consisting of chains of repeating alternate units of *β-1,3-linked D-galactose* and *α-1,4-linked 3,6-anhydro-L-galactose:*



Agaropectin - *the non-gelling fraction* - is a sulfated polysaccharide (3% to 10% sulfate), composed of *agarose and varying percentages of ester sulfate, D-glucuronic acid, and small amounts of pyruvic acid.*

The proportion of these two polymers varies according to the species of seaweed, but agarose normally represents at least two-thirds of the natural agar-agar.

Agar is used as an ingredient in desserts throughout Asia, and also as a solid substrate to contain culture media for microbiological work:



Agar can be used as an appetite suppressant, vegetarian substitute for gelatin, a thickener for soups, in fruit preserves, ice-cream, and other desserts.

4.4.2.5. Mammalian heteropolysaccharides: glycosaminoglycans

The extracellular space in animal tissues is filled with a gel-like material, **the extracellular matrix (ECM)**. The extracellular matrix holds the cells of a tissue together and provides a porous pathway for the diffusion of nutrients and oxygen to individual cells. The extracellular matrix consists of 3 major types of macromolecules: *fibrous components* (*collagen* and *elastin*), *proteoglycans* (carbohydrate-protein complexes with carbohydrate component represented by heteropolysaccharides glycosaminoglycans (mucopolysaccharides)) that have both structural and metabolic roles, and *glycoproteins* that have both structural and metabolic roles; the latter include providing linkage between matrix components and between cells and matrix components.

Proteoglycans (PGs) are the molecules with M.m. above tens millions that are characterized by a *core protein* covalently attached to sulphated heteropolysaccharides, called *glycosaminoglycans (GAGs)*. The core proteins are generally specific to each of the PGs types and show considerable variability in size. Similarly, there are various GAGs chains.

Glycosaminoglycans are the family of unbranched linear polymers composed of *repeating disaccharide units* (Fig. 4.43).

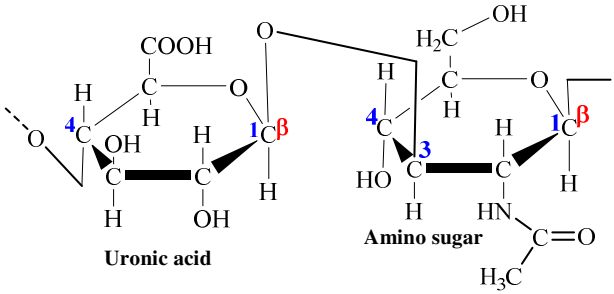


Fig. 4.43. The principle of glycosaminoglycan disaccharide unit structure

The type and number of these units determinates the properties of the PGs. Each disaccharide unit is formed by *uronic acid* (*glucuronic* or *iduronic*) and *aminosugar* (*N-acetylglucosamine* or *N-acetylgalactosamine)* (Fig. 4.43). In some glycosaminoglycans, one or more of the hydroxyls of the aminosugar is esterified with sulfate.

Combinations of sugars make up the disaccharide units, resulting in six major GAGs classes (Fig. 4.44) - chondroitin-4-sulphate (CS A) and chondroitin-6-sulphate (CS C), keratan sulphate (KS), dermatan sulphate (DS, also known as CS B), heparan sulphate (HS), heparin and hyaluronan (HA).

The types of glycoside bonds between the two monosaccharides in disaccaride unit and between disaccharide units in GAGs are given on Tab. 4.5.

Table 4.5.
The types of glycoside bonds in glycosaminoglycans

Mucopolysaccharide	*Two components of the disaccharide units*	*Linkages**	
Hyaluronic acid	D-glucuronic acid + N-acetyl-D-glucosamine	β-1 → 3	β-1 → 4
Chondroitin sulfate A	D-glucuronic acid + N-acetyl-D-galactosamine-4-sulfate	β-1 → 3	β-1 → 4
Chondroitin sulfate C	D-glucuronic acid + N-acetyl-D-galactosamine-6-sulfate	β-1 → 3	β-1 → 4
Dermatan sulfate (Chondroitin sulfate B)	L-iduronic acid + N-acetyl-D-galactosamine-4-sulfate	α-1 → 3	β-1 → 4
Keratosulfate	D-galactose + N-acetyl-D-glucosamine-6-sulfate	β-1 → 4	β-1 → 3

* Linkage of the first column represents the linkage involved between the two monosaccharides of the disaccharide unit whereas the linkage of the second column is the one involved between the repeating disaccharide units.

Chondroitin sulfates have such disaccharide unit as:

Glucuronic acid + Sulfated N-acetylgalactosamine

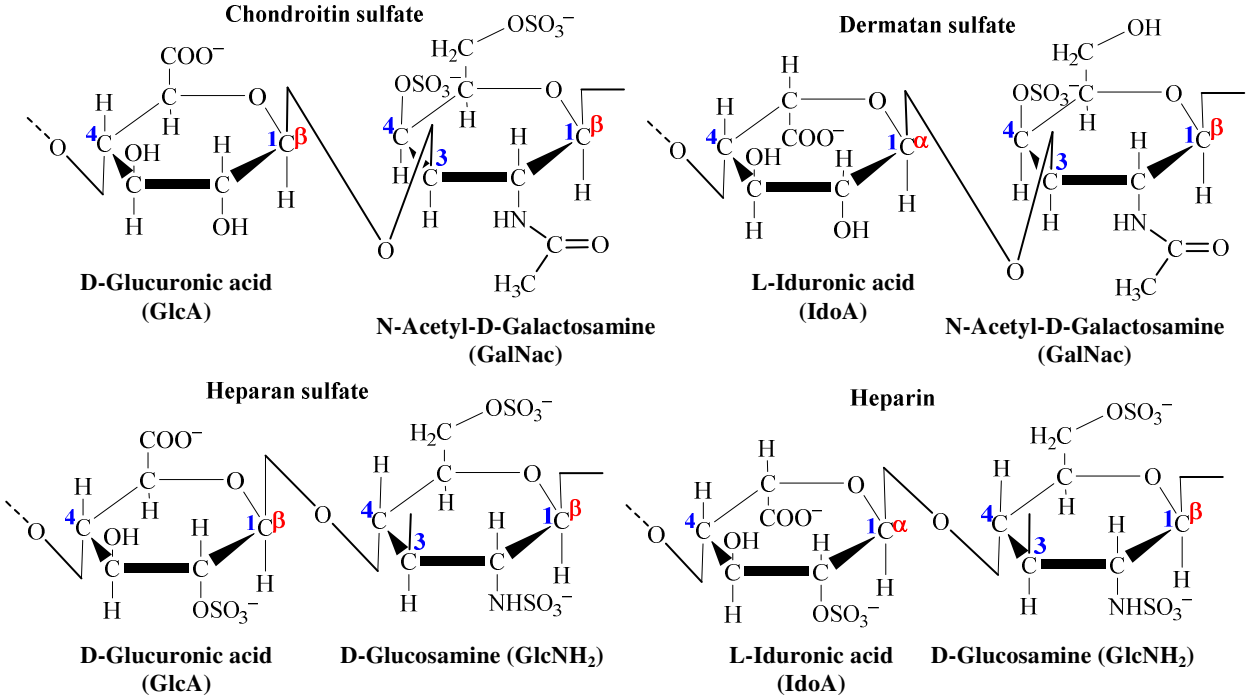
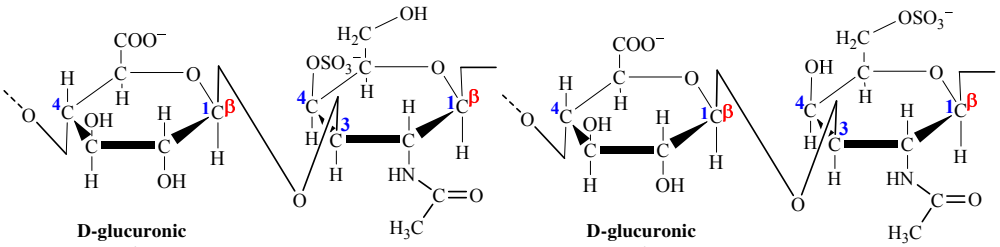


Fig. 4.44. The schematic structure of disaccharide units of some classes of glycosaminoglycans

The two glycosidic linkages involved in both types of chondroitin sulfate are the same - *β-(1→3)* (in disaccharide unit) and *β-(1→4)* (between disaccharide units):



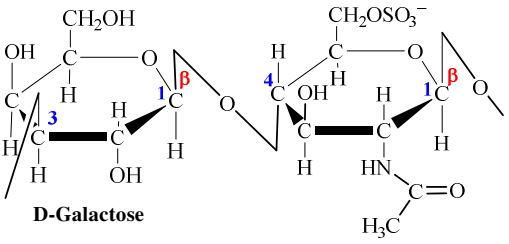
Chondroitin-4-sulfate **Chondroitin-6-sulfate**

They can form proteoglycan aggregates and are the most abundant GAG in the body. They are present in the ground substance of connective tissues distributed in cartilage, bone, tendons, cornea and skin. In cartilage, they bind collagen and hold fibers in a tight, strong network.

Keratan sulfates are the only GAGs which does not contain any uronic acid. Their disaccharide unit is:

D-Galactose + N-Acetylglucosamine

The two glycosidic linkages involved keratan sulfates forming are *β-(1→4)* (in disaccharide unit) and *β-(1→3)* (between disaccharide units):

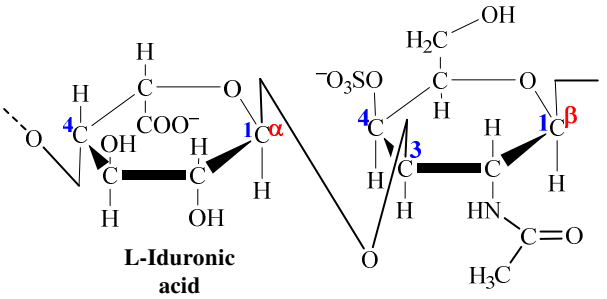


Sulfate content is variable and may be present on C-6 of either sugar. They are found in cornea, loose connective tissue and tendons.

Dermatan sulfates as a disaccharide unit have:

L-iduronic acid + N-acetylgalactosamine

The two glycosidic linkages involved dermantan sulfates forming are *α-(1→3)* (in disaccharide unit) and *β-(1→4)* (between disaccharide units):

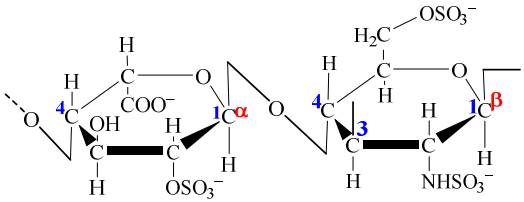


They present in skin, blood vessels and heart valves.

Heparin's disaccharide unit is:

Glucuronic or Iduronic acids + Glucosamine

Sulfate groups in heparin molecule are found on glucosamine and uronic acid:

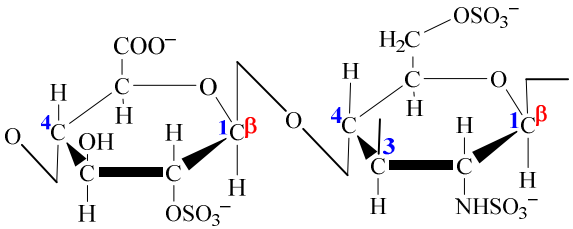


Unlike other GAGs that are extracellular, heparin is a component of intracellular granules of mast cells that line arteries, especially in liver, lungs and skin. It activates antithrombin III, which in turn inactivates thrombin, factor X and factor IX. It is also an anticoagulant used for taking blood for biochemistry studies as well as in suspected thromboembolic conditions to prevent intravascular coagulation.

Heparan sulfates have such disaccharide unit as:

D-glucuronate sulfate + N-sulfo-D-glucosamine

They are components of *basal laminae* (or *basement membrane* - thin extracellular matrix layer separating epithelial cells (cells lining body cavities and free surfaces) and cell surfaces:



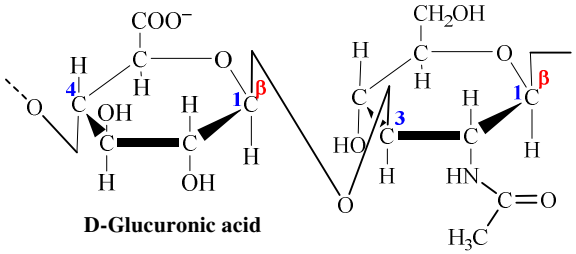
D-Glucuronate-2-sulfate **N-Sulfo-D-glucosamine-6-sulfate**

Heparan sulfates can act as receptors to help in the cell-cell interactions.

Hyaluronic acid (HA) as the disaccharide unit has:

Glucuronic acid + N-acetylglucosamine

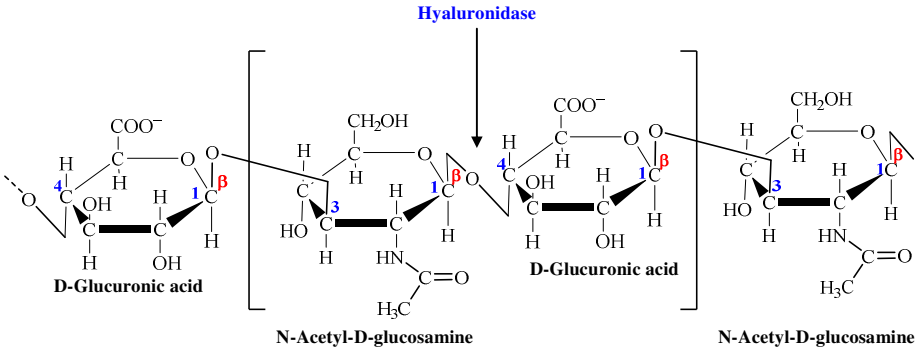
Two glycosidic linkages are invloved in hyaluronic acid formation - *β-(1→3)* (in disaccharide unit) and *β-(1→4)* (between disaccharide units):



Hyaluronic acid polymers are very large (100 - 10,000 kD). It is an acidic substance because the carboxyl groups are largely ionized at cellular pH, so it can displace a large volume of water.

In contrast to other GAGs, hyaluronic acid is unsulfated and isn't covalently attached to protein but forms non-covalently linked complexes with proteoglycans in the extracellular matrix. It present in connective tissue, tendons, synovial fluid of joints (where it serves as a lubricant and shock absorber), vitreous humor, umbilical cord and cartilage. In **rheumatic diseases** and arthritis hyaluronic acid is depolimerized and the viscosity of synovial fluid is decreased. It is also the only GAG found in bacteria.

Hyaluronic acid is hydrolyzed by enzyme **hyaluronidase** which is secreted by some pathogenic (disease-causing) bacteria and can hydrolyze the β-(1→4) glycosidic linkage of hyaluronate:



This enzyme cleaves hyaluronic acid to smaller fragments, rendering tissues more susceptible to invasion by the bacteria. This depolymerization effect allows any foreign bodies (such as pigments, pathogenic bacteria) to penetrate the tissue, since the cementing substance is being dissolved.

4.4.2.6. Proteoglycans as proteins conjugated with glycosaminoglycans

All of GAGs, except hyaluronic acid, are found covalently attached to proteins, forming **proteoglycans** (PGs), which consist of a core protein to which the linear GAG chains are covalently attached. Proteoglycans have a very high carbohydrate to protein ratio, often 95:5, and are found in the extracellular matrix and basal laminae.

Each core protein is *bound covalently* to glycosaminoglycan molecules, such as chondroitin sulfate, keratan sulfate, heparan sulfate, and dermatan sulfate (Fig. 4.45).

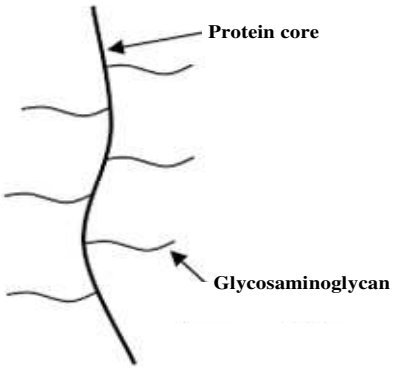
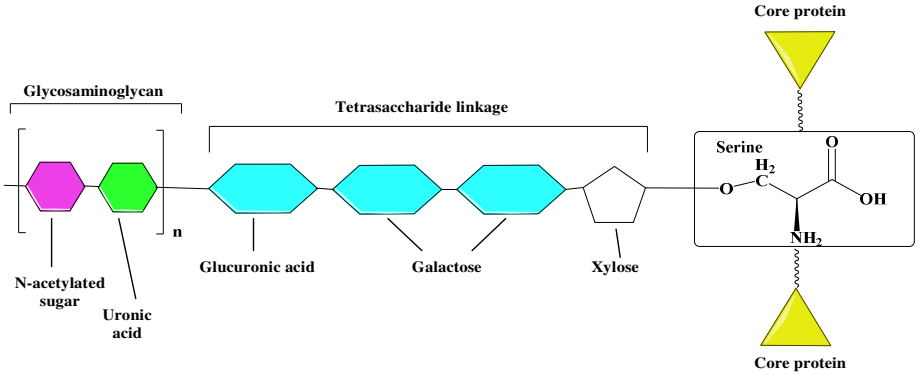


Fig. 4.45. The scheme of proteoglycan structure

A typical proteoglycan in human cartilage contains about 150 polysaccharide chains (each of Mr ≈20,000) covalently bound as side chains to each core protein.

The covalent attachments between glycosaminoglycans and core protein are O-glycosidic bonds between sugar residues and the OH-

groups of Ser residues in the protein (the core proteins of proteoglycans are rich in Ser and Thr residues which allow multiple GAG attachment):



The example of PGs is *aggrecan* (~200'000kD). It is the predominant PG in articular cartilage and plays a major role in normal joint function and in skeletal growth. A large compliment of chondroitin sulfates chains (approximately 100) and a smaller compliment of keratan sulfates chains (approximately 30) are attached to the protein core of the monomer.

4.4.2.7. Proteoglycan aggregates

The proteoglycan monomers associate with a molecule of hyaluronic acid to form **proteoglycan aggregates**. So proteoglycans agregates are composed of a very long strand of hyaluronic acid to which numerous molecules of core protein are bound non-covalently, at about 40 nm intervals (Fig. 4.46).

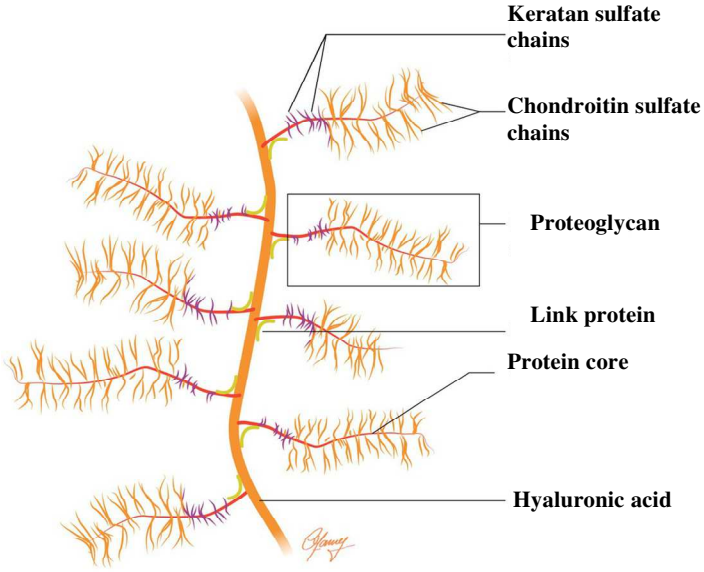


Fig. 4.46. The scheme of proteoglycans aggregate structure (by Sarraf, K., & Kader, D.)

As mentioned above, hyaluronic acid is atypical because it is not attached to a protein core and is not sulfated. But it is the most abundant and ubiquitous of the GAGs, and it plays an important role in binding to other PGs to form supramolecular complexes such as PG-aggregates.

The linkage between proteoglycan and hyaluronic acid is stabilized by a glycoprotein known as *link protein* that helps secure the PG monomers to the HA.

Properties of GAGs, PGs and PG aggregates. The combination of sulfate groups of the aminosugar residues and the carboxylate groups of the uronic acid residues gives the glycosaminoglycans a very high negative charge. Therefore all GAGs are negatively charged and so are able to attract Na^+ ions, creating an osmotic imbalance that results in the PG-GAG absorbing water from

surrounding areas. This absorption helps maintain the hydration of gel-like matrix; the degree of hydration depends on the number of GAG chains. Important mechanical functions of PGs include hydration of the matrix, stabilization of collagen networks, and the ability to resist compressive forces.

4.4.2.8. Murein as the heteropolysaccharide of the bacterial cell wall

Bacteria are surrounded by a rigid cell wall. Cell wall of gram positive and gram negative bacteria contains the *murein* that is a component of *peptidoglycan layer***.** The latter serves a structural role in the bacterial cell wall, giving the wall shape and structural strength, as well as counteracting the osmotic pressure of the cytoplasm.

Murein is a heteropolysaccharide formed by disaccharide units. Ever disaccharide unit contains (Fig. 4.47):

N-acetylglucosamine (NAG) + N-acetylmuramic acid (NAM)

They are linked by *β-(1→4)-glycoside bond.*

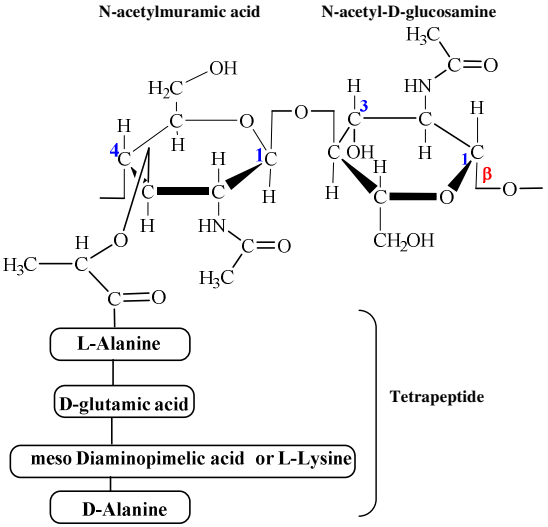


Fig. 4.47. The structure of murein linked with tetrapeptide

In peptidoglycan the lactyl group of each N-acetylmuramic acid residue of murein bonds to tetrapeptide side chain that contains L-alanine, D-glutamic acid, either meso-diaminopimelic acid or L-lysine (its decarboxylated product) and D-alanine (Fig. 4.47). Crosslinking between these tetrapeptides gives peptidoglycan its strong structure. Occurrence of D-amino acids in murein structure causes resistance of bacterial cell wall to proteases.

In gram (-) bacteria crosslinks between tetrapeptides are formed by *a direct link from the D-Ala residue above to the L-Lys in another peptide*, whereas **in gram (+) bacteria** *tetrapeptides are linked by five-glycine interbridges* (Fig. 4.48). These interbridges connect L-lysine of one tetrapeptide chain with D-alanine of another tetrapeptide. This crosslinking of tetrapeptide is formed via process called **transpeptidation**.

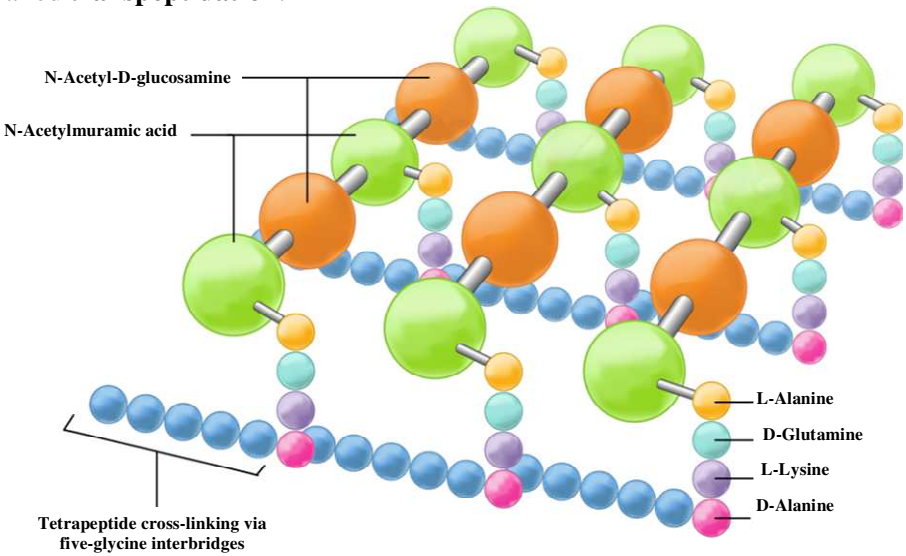


Fig. 4.48. Tetrapeptide crosslinking by five-glycine interbridges in peptidoglican structure of gram (+) bacteria cell wall

Therefore the tetrapeptide chain of one carbohydrate strand can be cross-linked to the tetrapeptide chain of another carbohydrate strand forming the 3D mesh-like layer.

The peculiarities of peptidoglycan layer structure in cell wall defines the differences in Gram Staining of gram (-) and gram (+) bacteria.

Gram staining is a method of staining bacteria. It is used to classify them into two groups: gram-positive, which show the stain, and gram-negative, which do not. The technique was developed in the 19th Century by Hans Christian Gram.

Gram staining differentiates bacteria by the chemical and physical properties of their cell walls by detecting peptidoglycan, which is present in the cell wall of Gram-positive bacteria. Gram-negative cells also contain peptidoglycan, but a very small layer of it that is dissolved when the alcohol is added. This is why the cell loses its initial color from the primary stain. *Gram-positive bacteria retain the crystal violet dye*, and thus are stained violet, while the Gram-negative bacteria do not:

After washing, a counterstain is added (commonly safranin or fuchsine) that will stain these *Gram-negative bacteria a pink color*. Both Gram-positive bacteria and Gram-negative bacteria pick up the counterstain. The counterstain, however, is unseen on Gram-positive bacteria because of the darker crystal violet stain (Fig. 4.49).

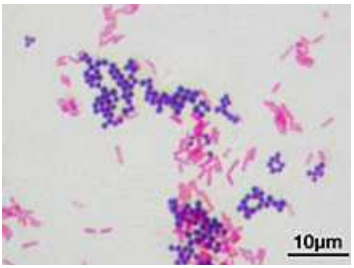


Fig. 4.49. A Gram staining of mixed *Staphylococcus aureus* (gram-positive cocci, in violet) and *Escherichia coli* (gram-negative bacilli, in pink)

Therefore, bacteria are classified as being either Gram-positive or Gram-negative based in differences in the structure of their peptidoglycan cell wall (Fig. 4.50).

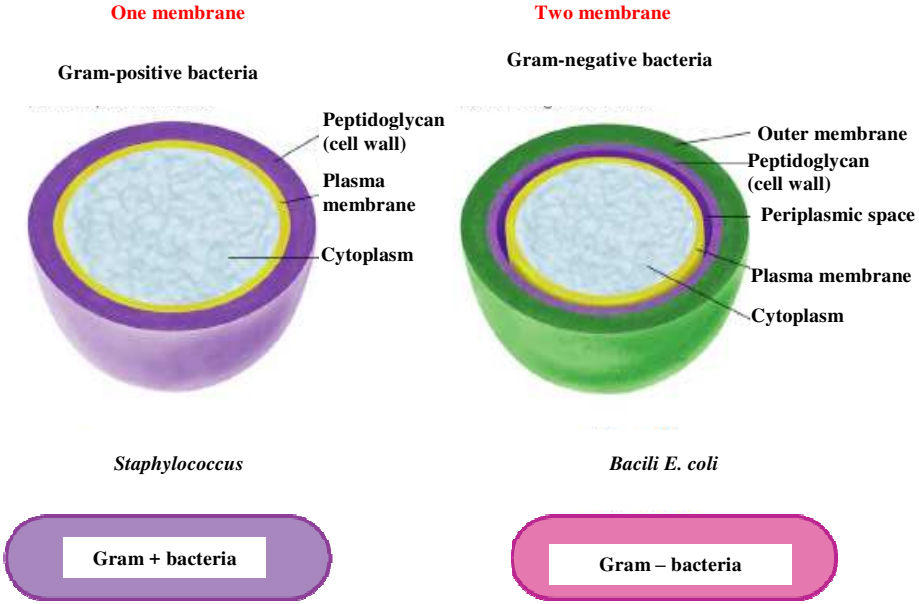


Fig. 4.50. The peculiarities of gram-positive and gram-negative cell wall structures

The peptidoglycan layer is substantially thicker in gram-positive bacteria (20 to 80 nanometers) than in gram-negative bacteria (7 to 8 nanometers). Peptidoglycan forms around 90% of the dry weight of Gram-positive bacteria but only 10% of Gram-negative strains (Fig. 4.50).

Murein and peptidoglycan are the targets for antibacterial medications. **Lysozyme** *cleaves β-(1→4)-glycosidic bond between N-acetylglucosamine and N-acetylmuramic acid in murein structure*, killing bacterial cells. Lysozyme is present in tears, presumably a defense against bacterial infections of the eye. It is also produced by

certain bacterial viruses to ensure their release from the host bacteria, an essential step of the viral infection cycle (Fig. 4.51).

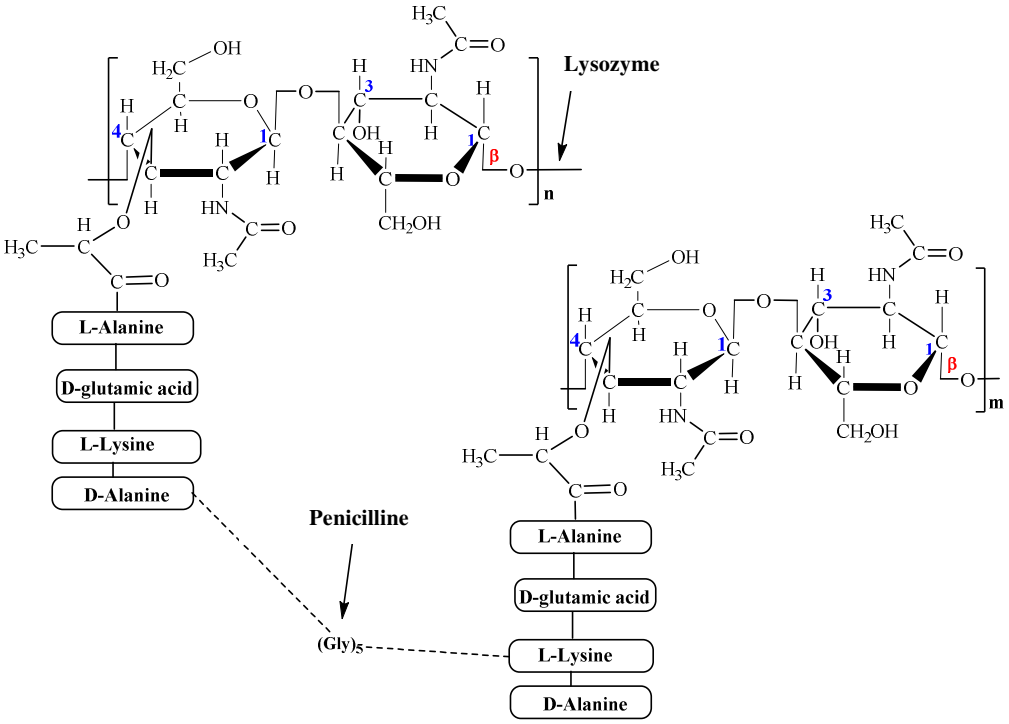


Fig. 4.51. Murein and peptidoglycan are the targets for antibacterial medications: the sites of lysozyme and penicillins action.

Several antibiotics, like penicillins, vancomycin, novobiocin, bacitracin etc. are known to *inhibit peptidoglycan crosslinking (transpeptidation)* and therefore destroy peptidoglycan and enhance osmo-sensitivity of growing cells.

The cell wall of Gram-positive bacteria has a thick peptidoglycan layer. As a result, it is very sensitive to the action of lysozyme and penicillin, or its derivatives. So penicillin is often the antibiotic of

choice for infections caused by Gram-positive organisms such as *Streptococcus pyogenes* which causes strep throat. This is almost always treated with some type of penicillin.

4.6. Glycoconjugates

Glycoconjugates are compounds that result from covalent linkages of carbohydrate molecules to either proteins or lipids. They have a profound effects on the functions of individual cells as weell as cell-cell interactions of multicellular organisms.

There are two classes of **carbohydrate-protein conjugates**:

- *glycoproteins* (have larger protein part);
- *proteoglycans* (have larger carbohydrate part).

Oligosaccharide parts of glycoproteins influence structure, folding and stability of protein; it may determine the lifetime of a protein (mark protein for age) as well as serve as markers to identify a cell type. When glycosylated proteins are at the cell surface they can modulate both cell-cell and cell-molecule interactions (e.g. hormone /receptor) and also can serve as *antigenic determinants* (in this way antibody recognizes the protein) on proteins (e.g. the difference between blood types is due to glycosylation of red blood cell proteins).

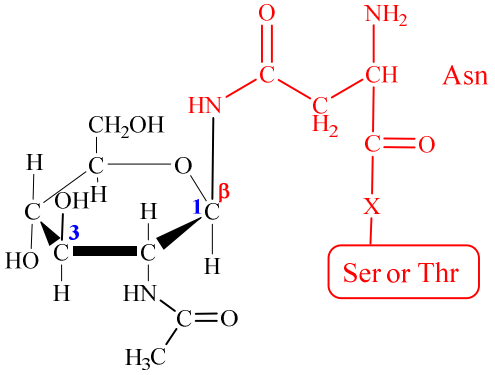
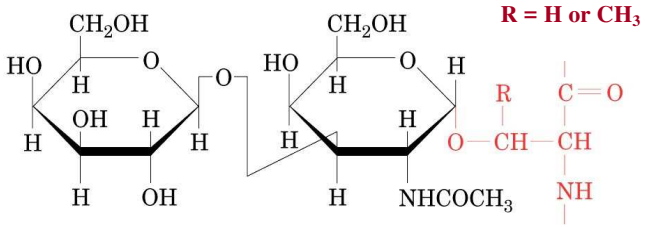
Glycolipids (oligosaccharide-containing lipid molecules) are found predominately on the outer surface of plasma membrane.

The process of addition of saccharides to proteins or lipids is called *glycosylation*. It is an enzyme-directed site-specific process that results in glycosylated proteins or lipids formation. Enzymes responsible for the course of glycosylation reactions are called *glycosyltransferases*. Glycosylation plays an important role in the synthesis of membrane and secreted proteins: majority of proteins synthesized in the rough ER undergo glycosylation. Glycosylation may also play a role in cell-cell adhesion (a mechanism employed by cells of the immune system).

Glycoprotein carbohydrate chains are highly diverse and classified into two groups (Tab. 4.6).

Table 4.6.

The comparative characteristic of N- and O-linked glycoproteins

N-linked glycoproteins (contain N-linked oligosaccharides and are formed in *N-glycosylation process*)	**O-linked glycoproteins** (contain O-linked oligosaccharides and are formed in *O-glycosylation process*)
 **N-acetyl-D-glucosamine**	 **β-Galactosyl-(1→3)-α-N-acetylgalactosaminyl-Ser/Thr**
N-glycosylation occurs *co-translationally in the ER at the consensus motif Asn–X–Ser/Thr of protein*, where X is any amino acid except proline	O-glycosylation occurs *post-translationally in the apparatus Golgi. The first monosaccharide is attached to polypeptide via its N-acetylgalactosamine and the hydroxyl group of side chain of amino acids serine or threonine*

Biosynthesis of N-linked glycoproteins begins with the addition of a large preformed oligosaccharide, containing 14 sugar residues - it attached via its N-acetylglucosamine (NAG) residue to the amide group of side chain of asparagine; subsequently certain sugar residues are removed and others are added, one at a time, in a defined order with each reaction catalyzed by a different enzyme. Following modifications of olygosaccharide part of N-glycozylated proteins occurs in apparatus Golgi	When O-linked glycoproteins are formed, each monosaccharide is added one at a time, and each sugar transfer is catalyzed by a different enzyme.
N-linked oligosaccharides contain *5 and more sugar residues*	O-linked oligosaccharides are short (1-4 monosaccharide residues)

4.6.1. N-linked Glycoproteins

All N-linked oligosaccharides have *a common pentasaccharide core* (Fig. 4.52).

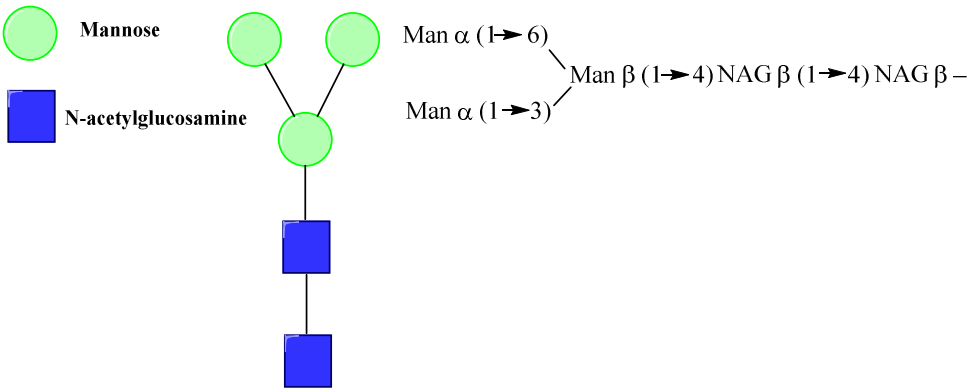


Fig. 4.52. The structure of pentasaccharide core of N-linked glycoproteins.

Therefore N-linked oligosaccharides contain 5 and more sugar residues. Modifications of the pentasaccharide core lead to a large number of possible structures which are classified into *high-mannose*, *complex* and *hybrid N-glycans* (Fig. 4.53).

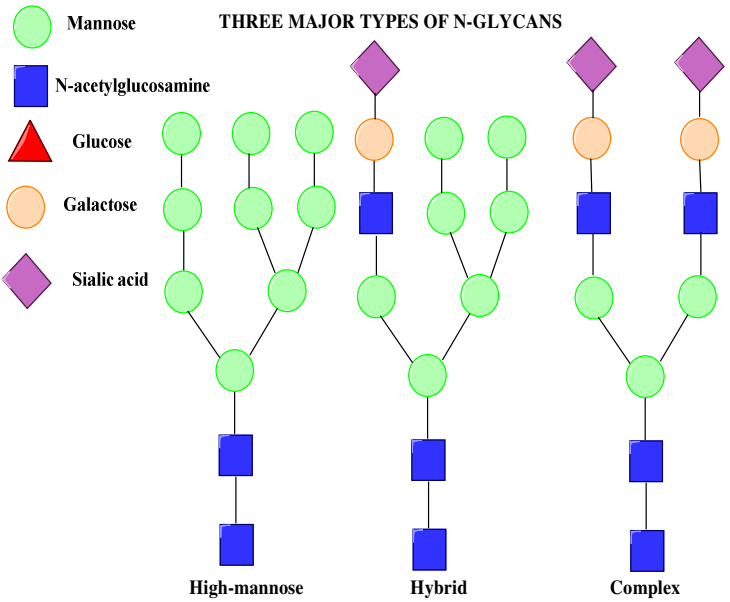


Fig. 4.53. The structures of high mannose, complex and hybrid N-glycans

High-mannose oligosaccharides have two N-acetylglucosamines with many mannose residues, often almost as many as are seen in the precursor oligosaccharides before it is attached to the protein mannose residues.

Complex oligosaccharides are so named because they can contain almost any number of the other multiple sugar types of saccharides, including more than the original two N-acetylglucosamines.

Hybrid oligosaccharides contain mannose residues on the one side of the branch, while on the other side N-acetylglucosamine initiates a complex branch. So these oligosaccharides have a peculiarities both high mannose and complex oligosaccharides.

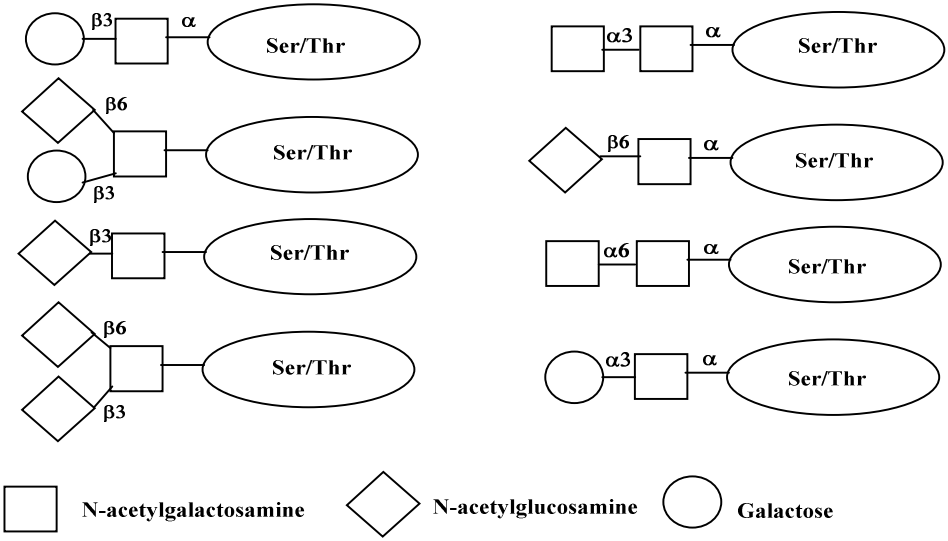
N-linked glycans serve as structural components of the cell wall and extracellular matrix and therefore contribute to cell-cell and cell-

matrix interactions and cell signaling. They can also modify protein properties such as stability to high temperature, pH, etc. and their solubility.

Within the immune system N-linked glycans of immune cell's surface help to direct cell migration - e.g. immune cells that migrate to the skin have specific sites of glycosylations that favor homing to that site. The glycosylation patterns on the various immunoglobulins (IgE, IgM, IgD, IgE, IgA, and IgG) bestow them with unique effector functions by altering their affinities for Fc and other immune receptors. Glycans may also be involved in "self" and "non self" discrimination, which may be relevant to the pathophysiology of various autoimmune diseases

4.5.2. O-linked glycoproteins

O-linked oligosaccharides are short (1-4 sugar residues) and are attached to polypeptides by the side chain hydroxyl group of amino acids serine or threonine in polypeptide chains:



While N-glycosylation is the most common glycosidic linkage, O-glycoproteins also play a key role in cell biology. This type of glycosylation is essential in the biosynthesis of mucins, a family of heavily O-glycosylated, high-molecular weight proteins that form mucus secretions. O-glycosylation is also critical for the formation of proteoglycan core proteins that are used to make extracellular matrix components. Additionally, antibodies are often heavily O-glycosylated.

4.5.3. The role of glycosylation reactions and glycosyltransferases in AB0 Blood group system functioning

The AB0 blood group antigens were discovered early in the 20th century by Karl Landsteiner and colleagues. They showed that humans could be divided into different groups according to the presence or absence of serum factors that would agglutinate red blood cells isolated from other humans. We now know that these serum factors are antibodies and that the corresponding antigens are glycan epitopes determined by the inheritance of genes that, for the most part, encode glycosyltransferases.

Human erythrocyte membranes contain various antigens of blood groups, which are represented by glycoproteins and glycolipids, which are also called **agglutinogens**, or hemagglutinogens (Fig. 4.54).

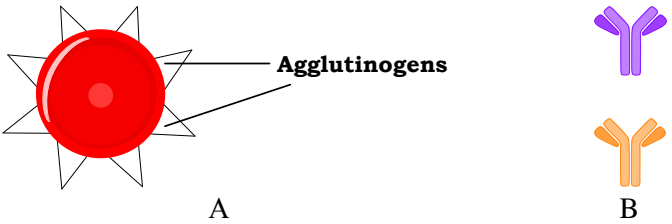


Fig. 4.54. Schematic representation of agglutinogens of membranes of erythrocytes (A) and agglutinins of blood plasma (B).

Each person has a specific set of agglutinogens, the most studied among them are *A*- and *B-agglutinogens*. Accordingly, in the presence of these antigens, four major blood groups are isolated (system AB0). People with group A have agglutinogen A, with group B - agglutinogen B, with group AB - have both agglutinogens, with group 0 - do not have any of these agglutinogens (Table 4.7).

The antigenic properties of agglutinogen A and B are determined by their carbohydrate part, or rather by the nature of the final carbohydrate residue.

Table 1.7.
The general characteristics of the AB0 system

Blood group	**Agglutinogens in the erythrocyte membrane**	**Agglutinins in plasma**	**Plasma agglutinates erythrocytes of type…**
0	None	Anti-A (α), Anti-B (β)	A, B, AB
A	A	Anti-B (β)	B, AB
B	B	Anti-A (α)	A, AB
AB	A, B	None	None

People with a group of 0 (I) without agglutinogens A and B contain *agglutinogen H*, the specificity of which is due to the presence of three terminal carbohydrate residues - N-acetylgalactosamine, β-galactose and L-fucose (Figure 4.55).

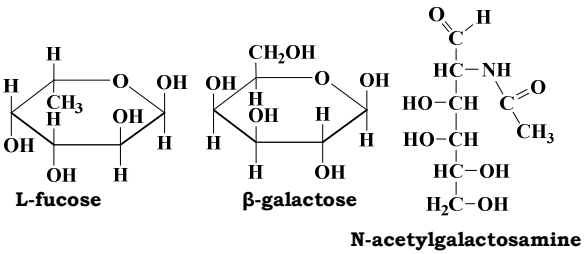


Fig. 4.55. Structural formulas of carbohydrates, the presence and localization of which in the composition of agglutinogen determines its appurtenance to the A, B or N type

In people belonging to group A (II), to these three residues is added another residue of N-acetylgalactosamine, which converts agglutinogen N into agglutinogen A; in humans with group B (III), agglutinogen H is converted into agglutinogen B by the addition of another galactose residue. Erythrocytes AB (IV) of the blood group have both types of agglutinogen - A and B (Fig. 4.56).

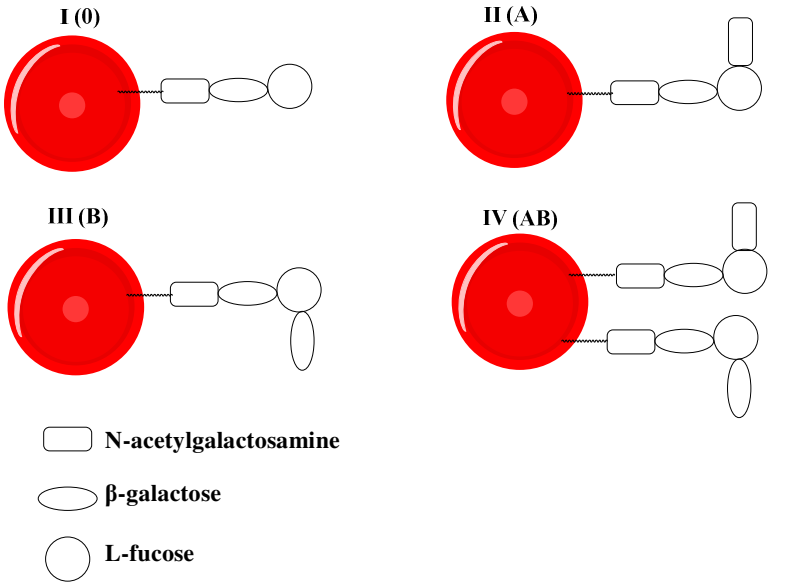


Fig. 4.56. Schematic representation of the peculiarities of the chemical composition of erythrocytes agglutinogen in people of different blood groups of the ABO system.

Such patterns in the antigenic composition of people with different blood groups are related to the presence or absence in their genomes genes encoding specific transferases - the enzymes responsible for glycosylation reactions.

In people of group A (II) there is a gene encoding a transferase catalyzing the attachment of the final N-acetylgalactosamine to the agglutinogen H, people with the group B (III) have a gene encoding a transferase for joining the final galactose, people with the AB (IV)

group, have both types of transferases. People with group 0 (I) do not have any of these transferases, therefore, the structure of agglutinogen H is retained.

Antibodies to agglutinogens of erythrocytes are called **agglutinins** (isogemaglyutinins) (Fig. 4.54). They are contained in blood plasma and are represented by γ-globulins. The most common of these are *α*- and *β-agglutinins* (or, respectively, anti-A- and anti-B-agglutinins). At the meeting of the same name agglutinogens and agglutinins (A and α, B and β), the hemagglutination reaction, which is accompanied by gluing, hemolysis and destruction of red blood cells, develops (Fig. 4.57).

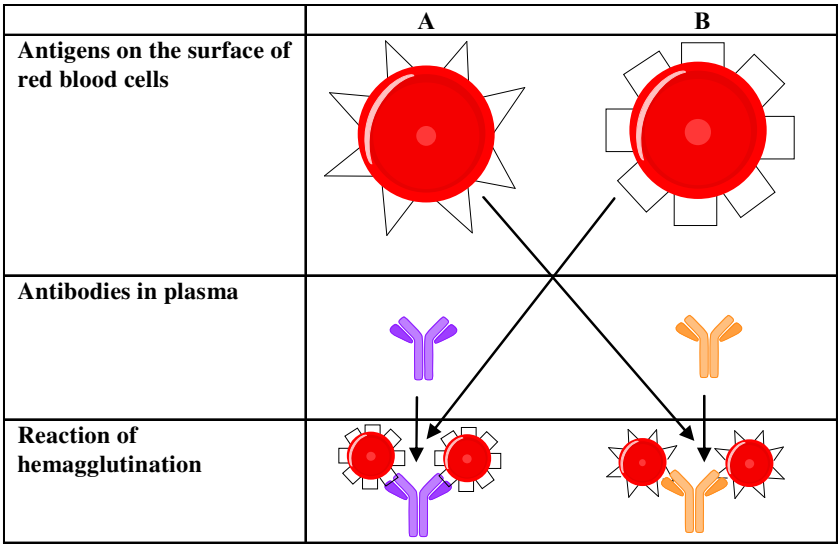


Fig. 4.57. Schematic representation of hemagglutination reaction

People with AB blood group are universal recipients; since they do not have circulating agglutinins, they can transfuse the blood of any group without the risk of developing a transfusion reaction due to the incompatibility AB0. People with a blood group 0 are universal donors; since they do not have antigens A and B, then the

blood of this group can be transfused to anyone without the risk of developing a transfusion reaction in the AB0 system.

4.6. Test questions

1. Inherited diseases, such as mucopolysaccharidoses, manifest in metabolic disorders of connective tissue, bone and joint pathologies. The sign of this disease is the excessive urinary excretion of the following substance:

A. Glycosaminoglycans
B. Amino acids
C. Glucose
D. Lipids
E. Urea

2. Which of these carbohydrates are resistant to digestion (are not hydrolyzed by the endogenous enzymes) in the human gastrointestinal tract?

A. Glycogen
B. Maltose
C. Chitin
D. Starch
E. Trehalose

3. Amylose and amylopectin are the components of:

A. Glycogen
B. Cellulose
C. Starch
D. Agar
E. Murein

4. Which of these polysaccharides are mammalian unbranched linear polymers composed of repeating disaccharide units?

A. Glycogen
B. Starch

C. Proteoglycans
D. Glycosaminoglycans
E. Cellulose

5. Which of these compounds aren't included in the dietary fibers composition?

A. Cellulose
B. Pectins
C. Glycosaminoglycans
D. Hemicelluloses
E. Inulin

6. Which of these carbohydrates is aldopentose?

A. Glucose
B. Galactose
C. Ribose
D. Ribulose
E. Glyceraldehyde

7. The carbon atom in molecule, which is connected with four different groups is called:

A. Reducing end
B. Non-reducing end
C. Chyral centre
D. Enantiomer
E. Hemiacetal group

8. Depending on the relative position of the -CH_2OH group and the -OH group on the anomeric carbon glucose molecules can be divided into:

A. D- and L-glucoses
B. "+" (clockwise) or "-" (anticlockwise)
C. α- or β-glucoses
D. Glucopyranose or glucofuranose
E. Open chain (acyclic) and closed chain (cyclic)

9. These hemostatic products (for example, Celox™) are currently used by different countries armies in the forms of bandages, powder, gauze or nasal tampon. They are obtained when processing raw materials containing carbohydrate:

A. Glycogen
B. Maltose
C. Chitin
D. Starch
E. Trehalose

10. Which of these compounds exist in bacterial cell wall and is the target at Gram staining?

A. Cellulose
B. Proteoglycans
C. Glycosaminoglycans
D. Peptidoglycans
E. Chitin

PART 5. STRUCTURE OF BIOLOGICAL MEMBRANE AND MEMBRANE TRANSPORT

5.1. Membrane structure.

5.1.1. Phospholipid bilayer and membrane lipids

Scientists began building molecular models of the membrane decades before membranes were first seen with the electron microscope (in the 1950s). In 1915, membranes isolated from red blood cells were chemically analyzed and found to be composed of lipids and proteins. Ten years later, two Dutch scientists reasoned that cell membranes must be phospholipid bilayers in which hydrophobic tails of the phospholipids are directed from water and hydrophilic heads - to water. In 1935 Hugh Davson and James Danielli suggested that the membrane might be coated on both sides with hydrophilic proteins. They proposed *a sandwich model*: a phospholipid bilayer between two layers of proteins. When researchers first used electron microscope to study cells (in the 1950s), the images seemed to support the Davson-Danielli model. By the late 1960s, however, many cell biologists recognized two problems with the model. First, inspection of a variety of membranes revealed that membranes with different functions differ in structure and chemical composition. A second, more serious problem became apparent once membrane proteins were better characterized. Unlike proteins dissolved in the cytosol, membrane proteins are not very soluble in water because they are amphipathic. If such proteins were layered on the surface of the membrane, their hydrophobic parts would be in aqueous surroundings. Taking these observations into account, S. J. Singer and G. Nicolson proposed in 1972 that membrane proteins reside in the phospholipid bilayer with their hydrophilic regions protruding. This molecular arrangement would maximize contact of hydrophilic regions of proteins and phospholipids with water in the cytosol and extracellular fluid, while providing their hydrophobic parts with a non aqueous environment.

In this *fluid mosaic model*, the membrane is a mosaic of protein molecules bobbing in a fluid bilayer of phospholipids.

Today is known that the cell membrane has mosaic structure. Mosaic is the art of creating images with an assemblage of small pieces of colored glass, stone, or other materials. The cell membrane is similar in that it is composed of many different, small units (many different proteins and phospholipids along with cholesterol).

Membranes lipids are:

- **Phospholipids**:
  - *Glycerophospholipids*;
  - *Sphingophospholipids*
- **Glycolipids**:
  - *Glyceroglycolipids*;
  - *Sphingoglycolipids*
- **Cholesterol**

Outer leaflet of membrane has **phosphatidylcholine**, **sphingomyelin** and inner leaflet has **phosphatidylethanolamine**, **phosphatidylserine** and **phosphatidylinositol**, therefore *plasma membrane is asymmetric*. Glycolipids localized only in outer leaflet, carbohydrate on surface. Cholesterol localized in both leaflets.

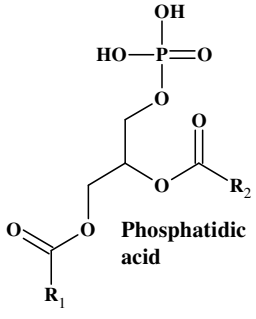
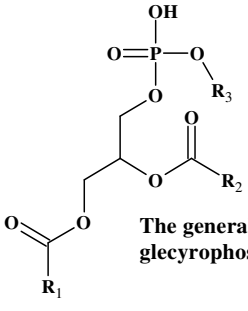
Glycerophospholipids are derivatives of *phosphatidic acid* (phosphatidate) and common constituents of cellular membranes. Phosphatidate has a glycerol backbone, hydroxyl groups of which near C1 and C2 are esterified by fatty acids, whereas hydroxyl near C3 is esterified by Pi (Tab. 5.1).

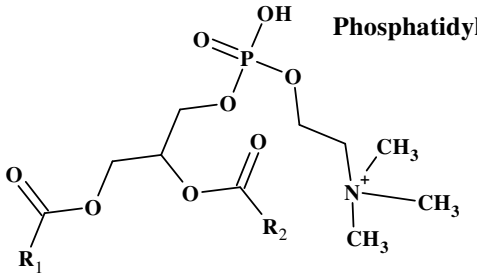
In most glycerophospholipids Pi is in turn esterified to OH of a polar head group (X): e.g., serine, choline, ethanolamine, glycerol, or inositol.

Each glycerophospholipid includes a polar region (glycerol, carbonyl O of fatty acids, Pi, and the polar head group (X)) and non-polar hydrocarbon tails of fatty acids (R1, R2).

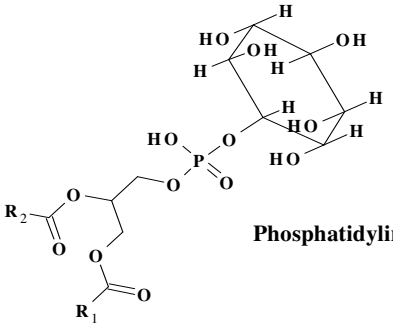
Sphingolipids are derivatives of the *alcohol sphingosine*, which has a long hydrocarbon tail, and a polar domain that includes an amino group (Tab. 5.2). This amino group can form an amide bond with a fatty acid carboxyl, to yield a *ceramide*. In most complex sphingolipids, a polar “head group" is esterified to the terminal hydroxyl of the sphingosine moiety of the ceramide.

Table 5.1.
Composition of membrane glycerophospholipids

 **Phosphatidic acid**	 **The general formula of glecyrophospholipids**

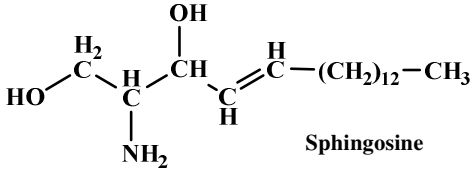


Phosphatidylcholine

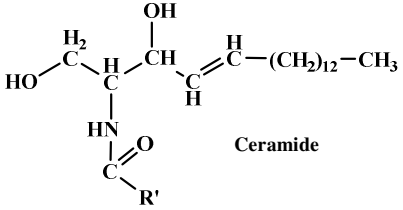


Phosphatidylinositol

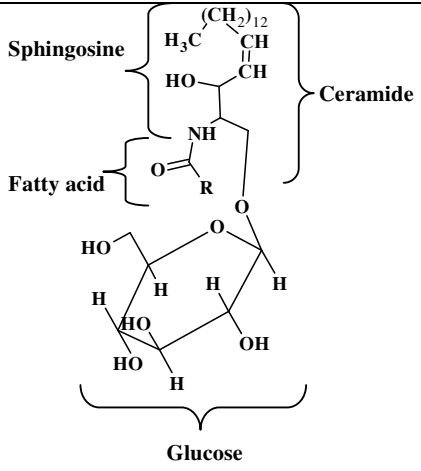
Table 5.2.
Composition of membrane sphingolipids

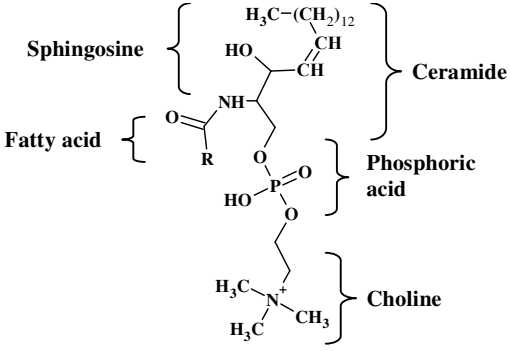


Sphingosine



Ceramide





Sphingomyelin has a phosphocholine or phosphoethanolamine head group. **Cerebroside** is a sphingolipid (ceramide) with a monosaccharide such as glucose or galactose as polar head group. **Ganglioside** is a ceramide with a polar head group that is a complex oligosaccharide, including the acidic sugar derivative sialic acid. Cerebrosides and gangliosides, collectively called glycosphingolipids, are commonly found in the outer leaflet of the plasma membrane bilayer, whereas their sugar chains extending out from the cell surface.

Amphipathic lipids in association with water form complexes in which polar regions are in contact with water and hydrophobic regions away from water. Depending on the lipid there are two possible molecular arrangements - various *micelle structures* (e.g., a spherical micelle is a stable configuration for amphipathic lipids with a conical shape, such as fatty acids) and *bilayer* (which is the most stable configuration for amphipathic lipids with a cylindrical shape, such as phospholipids).

All the molecules in the cell membrane (proteins, phospholipids, cholesterol) are a fluid and move around independent of each other, unless they are attached to the cytoskeleton or other proteins in oligomeric complexes. Fluidity enhanced by cholesterol and unsaturated lipids. The speed of *lipids lateral movement* is about 10^7 times per second, the speed of *rotation* is faster (Fig. 5.1).

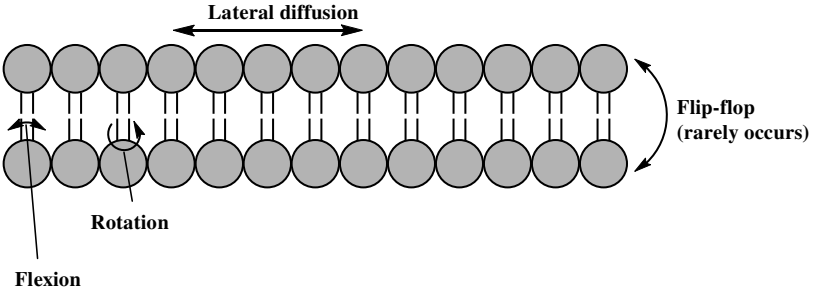


Fig. 5.1. The ways of lipid movements in biological membranes

Flip-flop of lipids (from one half of a bilayer to the other) is normally very slow – about once per month, as flip-flop would

require the polar head-group of a lipid to traverse the hydrophobic core of the membrane. So the two leaflets of a bilayer membrane tend to differ in their lipid composition.

Some membranes contain enzymes - *flippases* - which actively transport particular lipids from one monolayer to the other and in this case the speed of flip-flop enhances to about once per 1-2 minutes.

Cholesterol, an important constituent of cell membranes, has a rigid ring system and a short branched hydrocarbon tail (Fig. 5.2).

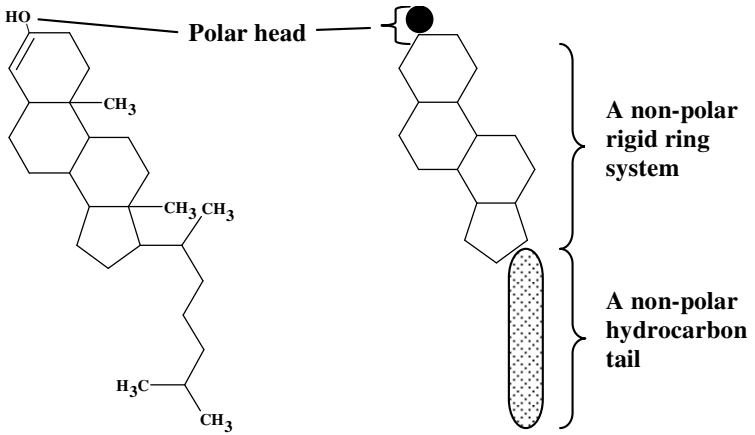
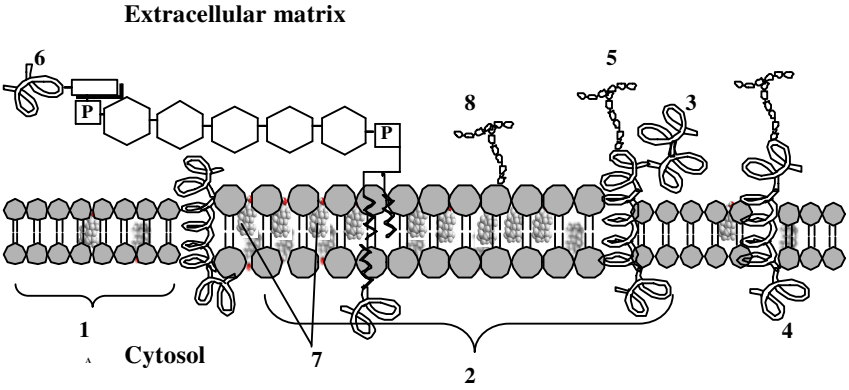


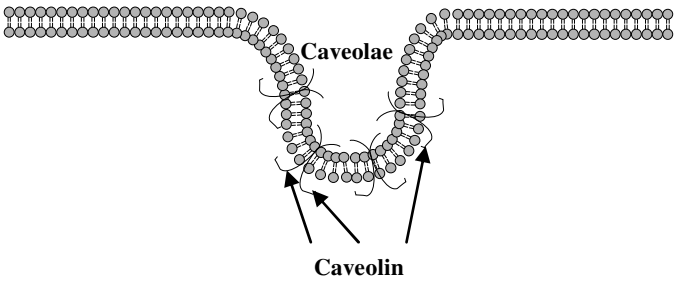
Fig. 5.2. Cholesterol structure

Cholesterol is largely hydrophobic. But it has one polar group, a hydroxyl, making it amphipathic. Cholesterol inserts into bilayer membranes with its hydroxyl group oriented toward the aqueous phase and its hydrophobic ring system adjacent to fatty acid chains of phospholipids. The OH group of cholesterol forms hydrogen bonds with polar phospholipid head groups. Interaction with the relatively rigid cholesterol decreases the mobility of hydrocarbon tails of phospholipids. Phospholipid membranes with a high concentration of cholesterol have a fluidity intermediate between the liquid crystal and crystal states.

Cholesterol and sphingolipids (sphingomyelin and glycolipids) in membranes form clusters - **lipid rafts**. Rafts are highly-ordered versus phospholipid bilayer:



Lipid raft structure:
1 – lipids near lipid raft
2 – lipid raft area
3 – transmembrane protein in lipid raft
4– transmembrane protein near lipid raft
5 – carbohydrate part of the glycoprotein
6 – GPI-anchored protein
7 – cholesterol
8 – carbohydrate part of the glycolipid



Visualize lipid rafts with fluorescent probe Laurdan sensitive to rigidity of phospholipid bilayer (false color).

Lipid rafts are complex sphingolipids tend to separate out from glycerophospholipids and co-localize with cholesterol in membrane microdomains called lipid rafts.

Membrane fragments assumed to be lipid rafts are found to be resistant to detergent solubilization, which has facilitated their isolation and characterization. Proteins involved in cell signaling often associate with lipid raft domains. Otherwise soluble signal proteins often assemble in complexes at the cytosolic surface of the plasma membrane in part via insertion of attached fatty acyl or isoprenoid lipid anchors into raft domains. Integral proteins may concentrate in raft domains via interactions with raft lipids or with other raft proteins. Some raft domains contain derivatives of phosphatidylinositol that bind signal proteins with pleckstrin homology domains.

Caveolaes are invaginated lipid raft domains of the plasma membrane that have roles in cell signaling and membrane internalization. Caveolin is a protein associated with the cytosolic leaflet of the plasma membrane in caveolae. Caveolin interacts with cholesterol and self-associates as oligomers that may contribute to deforming the membrane to create the unique morphology of caveolae.

5.1.2. Membrane proteins

Membrane proteins differ by their functions. There are:

- *channel proteins* that forms tubular structures which allow passage of molecules through membrane;
- *carrier proteins* which molecules combine with a substance to be transported and therefore assist passage of molecules through membrane;
- *cell recognition proteins* that provide cells with unique chemical compounds and help body to recognize foreign substances;

- *receptor proteins* which bind to a ligand and cause cell to respond to this ligand;
- *proteins-enzymes* that carry out metabolic reactions.

Membrane proteins also differ by their localization in membrane. There are integral proteins, proteins which link to the surface of the plasma membrane via a covalently attached hydrophobic anchor as well as peripheral proteins.

Integral proteins have domains that extend into the hydrocarbon core of the membrane. Intramembrane domains have largely hydrophobic surfaces, which interact with membrane lipids. Residues with aliphatic side-chains (leucine, isoleucine, alanine, and valine) predominate in the middle of the bilayer.

A *membrane-spanning α-helix* is the most common structural motif found in integral proteins. If a hydrophobic part of integral protein is represented by *one transmembrane α-helix*, topology studies are expected to confirm location of N- and C-termini on opposite sides of membrane. If *two transmembrane α-helices* are predicted, N- and C-termini should be on the same side. The segment between the α-helices should be on the other side.

Integral membrane proteins insert into lipid bilayer and can dissociate with reagents that disrupt hydrophobic interactions – *detergents* (e.g. *octylglucoside*, *sodium dodecylsulfate (SDS)*). **Detergents** are *amphipathic* (both hydrophobic and hydrophilic) compounds that solubilize integral transmembrane proteins.

Some proteins bind to membranes **via a covalently attached hydrophobic anchor** that inserts into the bilayer. For example, some proteins may link to the cytosolic surface of the plasma membrane *via a covalently attached fatty acid* (e.g., palmitate or myristate) or *an isoprenoid group*. Palmitate is usually attached via an ester linkage to the thiol of a cysteine residue. Palmitated protein may be released from plasma membrane to cytosol via *depalmitoylation*, hydrolysis of the ester link. An isoprenoid such as a farnesyl residue is attached to some proteins via a thioether linkage to a cysteine thiol.

Peripheral proteins have a lot of polar amino acids; therefore they are able to form hydrogen and (or) ionic bonds with polar groups of lipids and proteins localized in membrane. Such proteins

can be easily separated from membrane by influence of salt solutions or via pH change. Such procedures lead to loosing in protein charge and, as a result – destruction of above-mentioned non-covalent bonds.

Proteins are much larger than lipids and move more slowly, but some membrane proteins do drift. And some membrane proteins seem to move in a highly directed manner, perhaps driven along cytoskeletal fibers by motor proteins connected to the membrane proteins' cytoplasmic regions. However, many other membrane proteins seem to be held immobile by their attachment to the cytoskeleton or to the extracellular matrix.

5.1.3. Membrane carbohydrates

Membrane **carbohydrates** are usually short, branched chains of fewer than 15 sugar units. Some are covalently bonded to lipids, forming molecules called **glycolipids**. However, most are covalently bonded to proteins, which are thereby **glycoproteins**. The carbohydrates on the extracellular side of the plasma membrane vary from species to species, among individuals of the same species, and even from one cell type to another in a single individual. The diversity of the molecules and their location on the cell's surface enable membrane carbohydrates to function as markers that distinguish one cell from another. For example, the four human blood types designated A, B, AB, and O reflect variation in the carbohydrate part of glycoproteins on the surface of red blood cells.

5.2. Membrane transport

There are two main types of membrane transport: passive (diffusion) and active (Fig. 5.3.)

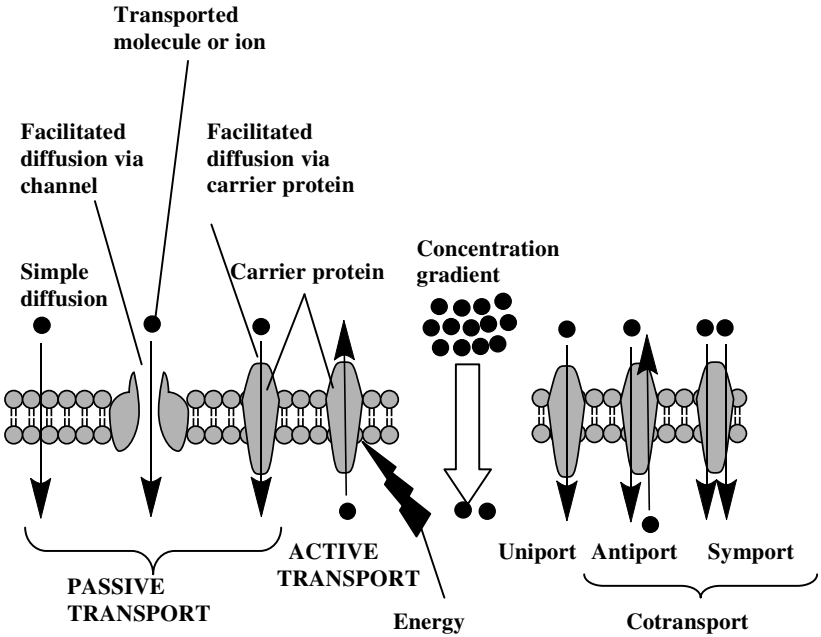


Fig. 5.3. The scheme of passive and active membrane transport

5.2.1. Passive transport

Diffusion is the random, passive (happens all by itself without transferring energy to the molecules) movement of molecules from a location where they are higher in to a location where they are lower in concentration (Fig. 5..3).

Small hydrophobic molecules like steroids, oxygen molecules, and carbon dioxide have no charge and therefore do not stick to anything, and are small enough to squeeze between phospholipids of the membranes. So they can go through the phospholipid bilayer by **simple diffusion** by which molecules move across membrane from high to low concentration without any help from proteins (Fig. 5.4).

Small hydrophilic molecules (ions, monomers and other small organic molecules) will cross the membrane by **facilitated diffusion**, by which molecules have net movement from high to low concentration with assistance from a membrane proteins.

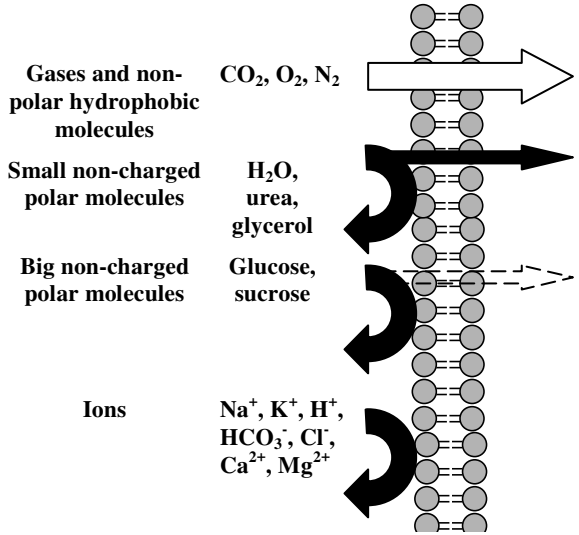
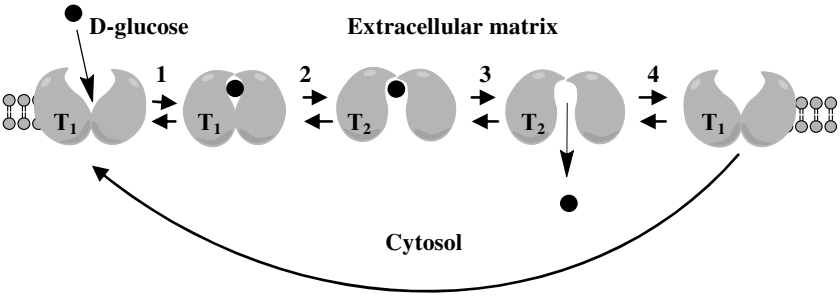


Fig. 5.4. The permeability of biological membranes for different types of molecules and ions

These proteins can form *channels*, can be *carrier proteins*, which by definition change conformation (shape) up binding of solute or can be *ionofores* (Fig. 5.3).

Molecules which cross the membrane by facilitated diffusion *can go through the membrane faster than normal diffusion process*. The *velocity of transport is saturable* in facilitated diffusion**.**

The uptake of glucose into erythrocyte is a good example of facilitated diffusion. It is rapidly moved across the membrane down its concentration gradient with the help of a **carrier protein** that belongs to *permeases* family**.** Permeases are multipass transmembrane proteins used to facilitate the diffusion of specific molecules across biological membrane. Permeases for glucose are proteins *GLUTs:*



Some ions can pass through membrane by **ion channels.** These structures have two discrete states: *open* (conducting) and *closed* (nonconducting). Some channels have also *inactivated* state (open but nonconducting).

Ligand-gated channels open when a ligand – atom or molecule – binds to a specific site on a protein. Examples of such molecules are represented by receptor for glutamate (NMDA-R), nicotinic acetylcholine receptor, ryanodine and inositol-1,4,5-triphosphate receptors, receptors for glycine, γ-amino butyric acid (GABA), serotonin, ATP etc. Most of their ligands are extracellular, whereas cADP-ribose, inositol-1,4,5-triphosphate and sometimes Ca^{2+} serve as intracellular ligands. For instance, *nicotinic acetylcholine receptor (AchR)* consists of a pentamer of protein subunits, with two binding sites for this neurotransmitter (Fig. 5.5). Binding of acetylcholine alter the receptor's configuration and cause an internal pore to open. This pore allows Na^+ ions to flow down their electrochemical gradient into the cell. The AchR also responds to nicotine, and so is called the “nicotinic” acetylcholine receptor – nAChR. *CORBA* snake toxins inhibit AchR.

Mechanic-gated channels can open upon mechanical (physical) stress. In this case it is caused by vibrations from sound entering your ear as the receptor in blue gets pulled open when the cilia vibrate.

Acetylcholine-gated cation channel consists of 5 subunits, every of whic has 4 transmembrane domains

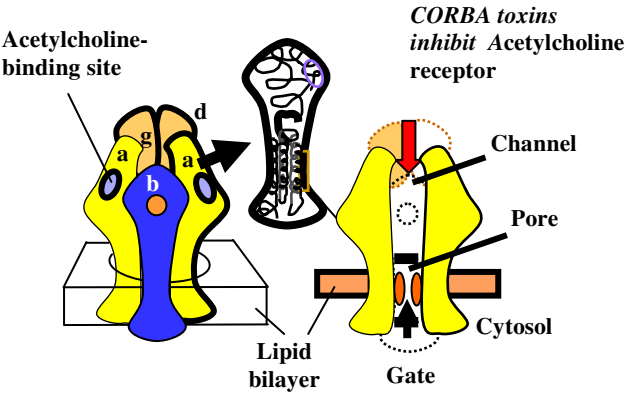


Fig. 5.5. Schematic representation of nicotinic acetylcholine receptor structure

Light-gated channels can open and close in response to light. Channel rhodopsins belong to subfamily of proteins that function as light-gated ion channels. They serve as sensory photoreceptors in unicellular green algae, controlling phototaxis: movement in response to light. Expressed in cells of other organisms, they allow the light-induced depolarization of cells and enable light to control electrical excitability, intracellular acidity, calcium influx, and other cellular processes.

Voltage-gated channels can open in response to changes in membrane potential, subsequently open and inactivate, are specific for a particular ion, have common domain structure and are regulated by external signals.

Voltage-gated sodium channel has one α- and one β-subunit. α-subunit is responsible for pore and β-subunit modify channel

function but is not essential to create the pore. α-subunit (260kda) contains 4 internal repeats (similar amino acid sequence) - domains 1-4. Each repeat contains of 6 transmembrane α-helical structures (S1-S6): 5 hydrophobic segments (S1, S2, S3, S5, S6) which form selectivity filter and pore and 1 highly positive charged segment (S4) which is voltage sensors of the channel (Fig. 5.6).

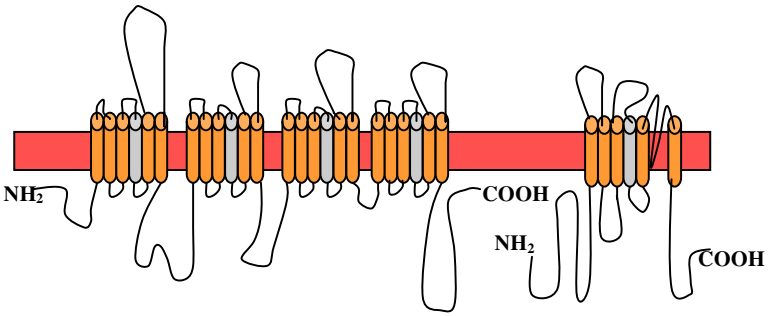
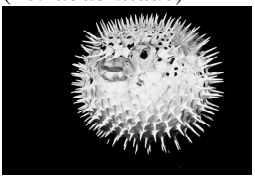
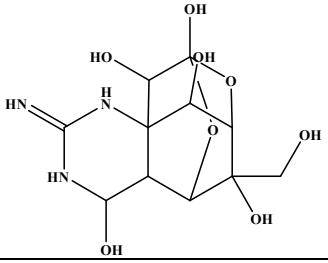
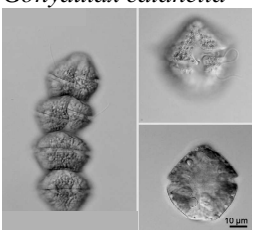
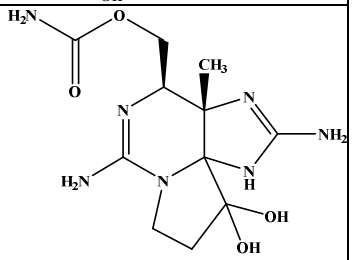


Fig. 5.6. The structures of α-subunits of voltage-gated sodium and potassium channels

Voltage-gated Na^+ channel binds specific neurotoxins. Both *tetrodotoxin* from puffer fish (10ng lethal dose) and *saxitoxin* from marine protozoa are *Na^+ channel blocking toxins,* whereas *batrachotoxin* from colombian frog and *veratridine* from lilies are *soluble in fats activators of Na^+ channel,* that blocks inactivation and therefore causes channels to open at more negative potentials and to stay open much longer than usual.

Tetrodotoxin from puffer fish is found mainly in the liver and gonads (Tab. 5.1). It binds to the outside of the sodium channel on a one-to-one basis.

Table 5.1.
Voltage-gated Na^+ channel blockers

Toxin	Organisms	Formula
Tetrodotoxin	Puffer fish (*Tetraodontidae)* 	
Saxitoxin	marine protozoa *Gonyaulax catanella* 	

Saxitoxin from algae bio-accumulates to bivalves such as oyster and mussels. Red-tide (algal bloom) could be dangerous. They have sodium channel blockers too. When saxitoxin accumulated in mussels or oysters and consumed by human, paralytic shell-fish poisoning resulted.

Sodium channel blockers can cause suffocation when the nervous system controlling respiration is blocked. At low dose, paralytic effects have been observed in patients intoxicated with these toxins.

Voltage-gated potassium channel is 70 kDa protein that has 4 α- and 4 β-subunits. Each subunit has S1-S6 membrane-spanning α helices which are homologous to one of the repeated units (domens 1-4) of voltage-gated sodium channel (Fig. 5.6). S5 and S6 form the actual pore of the K^+ channel, selective filter and S4 is voltage sensors of the channel.

K^+-ion channel blockers are tetraethyl ammonium (TEA), phalloidyn from *Amanita phalloides,* tityustoxin and Charybdotoxin

from scorpions (Tityus serrulatus and Leiurus quinquestriatus), dendrotoxins from Dendroaspis angusticeps, tarantula toxins (hanatoxins) and alkaloid toxins from plants.

Voltage-gated calcium channels in neurons are mostly responsible for the entry of calcium into the presynaptic ending following depolarization (and subsequent exocytosis of neurotransmitter), in heart are coupling to excitation contraction and in all excitable secretory cells (adrenal medulla, pancreas) cause entry of calcium which induces secretion.

They were initially divided into 2 classes – HVA (high voltage activated) and LVA (low voltage activated) Ca^{2+}-channels. HVA Ca^{2+}-channels are further divided into L, N, P/Q and R-types channels, while LVA Ca^{2+}-channels consist of only T-type channels. R-type is occasionally classified as IVA (intermediate-voltage-activated) channels.

There are a lot of *chloride ion channels* which have different structure, gating mechanisms and functions. *Extracellular ligand-gated (ELG) Cl^--channels* are pentamers (each subunit has 4 transmembrane domains) and function as receptors (postsynaptic GABA and Glycine receptors). *CFTR (cystic fibrosis transmembrane conductance regulator)-channel* belong to ABC (ATP-binding cassette) transporters, has 2 transmembrane domains (each of them consist of 6 α-helical segments) and its gating controlled by ATP and/or phosphorylation by cAMP- , cGMP- dependant kinases. There are also *voltage gated chloride channels (CLC), nucleotide/volume sensitive chloride channels (CLNS1A), chloride intracellular channels (CLIC*) – they have no membrane spanning domains and regulate electrolyte composition and acidification of intravascular spaces) and *calcium activated (CLCA) chloride channels*.

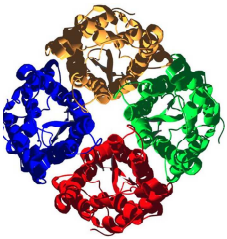
Often chloride ion channels are permeable to other anions, but chloride happens to be the most abundant anion. Chloride channels regulate cell volume, membrane potential, resting potential, depolarization, signal propagation, transport processes.

Some diseases associated with chloride ion channels - cystic fibrosis (genetic disorder in which gland secretions are abnormally thick), myotonia congenita (genetic neuromuscular disorder with mutations in the chloride ion channels of muscle cell plasma

membranes) etc. *Cystic fibrosis* has many affects: abnormally thick and sticky gland secretions, many organs in the body are affected by clogging mucous (the lungs are most affected). Air passages are blocked and the mucous also serves as a growth environment for bacteria giving rise to respiratory infections. The decrease in oxygen levels also causes pulmonary arteries to constrict leading to high blood pressure and increases the strain on the heart. 98% of all patients die from cardiopulmonary complications. 1 out of 1500 Caucasian children are affected.

Aquaporines (AQPs**),** also known as *water channels*, were found by Peter Agre (1992; 2003 Nobel Prize in Chemistry was awarded to him for the discovery of aquaporines). For many years, scientists assumed that water leaked through the cell membrane, and some water does. The presence of water channels increases membrane permeability to water. Many human cell types express them, as do certain bacteria and many other organisms, such as plants for which it is essential for the water transport system. Aquaporines selectively conduct water molecules in and out of the cell, while preventing the passage of ions and other solutes. Some of them, known as *aquaglyceroporines*, also transport other small uncharged solutes, such as glycerol, CO_2, ammonia and urea across the membrane, depending on the size of the pore. However, the water pores are completely impermeable to charged species, such as protons, a property critical for the conservation of the membrane's electrochemical potential.

Aquaporines are integral membrane pore proteins. Aquaporines form tetramers in the cell membrane, with each subunit acting as a water channel:



Aquaporine tetramer

Each subunit acts as a water channel

The different aquaporines contain differences in their peptide sequence, which allows for the size of the pore in the protein to differ between aquaporines. Each monomer consists of *six membrane-spanning segments (1-6)* arranged in two hemi-pores, which fold together to form the hourglass-shaped channel. There are also *five interhelical loop regions (A – E)* that form the extracellular and cytoplasmic vestibules (Fig. 5.7).

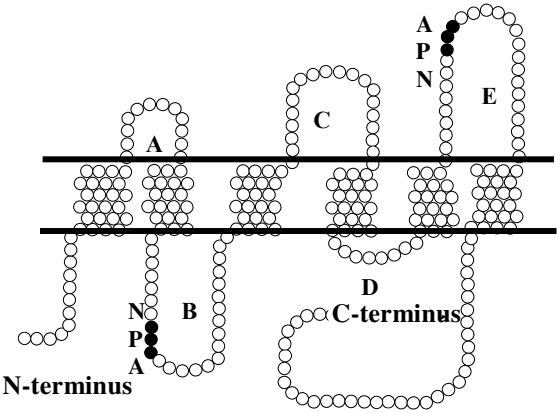


Fig. 5.7. The structures of aquaporine subunit

Loops B and E are hydrophobic loops that contain the highly, although not completely conserved, asparagine–proline–alanine (NPA) motif, which overlap the middle of the lipid bilayer of the membrane forming a 3-D 'hourglass' structure where the water flows through.

There are 11 different variants of these channels in the human body and more may still be discovered. They are responsible for many reactions in the body and one major function is done in the kidneys. *AQP1* is a widely expressed water channel, whose physiological function has been most thoroughly characterized in the kidney. It is found in the basolateral and apical plasma membranes of the proximal tubules, the descending limb of the loop

of Henle. *AQP2* is found in the apical cell membranes of the kidney's duct principal cells and in intracellular vesicles located throughout the cell. 70% of water is reabsorbed by AQP1 from primary urine into blood and 10% of water is reabsorbed by AQP2.

AQP2 is the only aquaporine regulated by *vasopressin* (the same as *antidiuretic hormone* (ADH)). This regulation can occur in two ways: short-term regulation (minutes) through trafficking of AQP2 vesicles to the apical region where they fuse with the apical plasma membrane and long-term regulation (days) through an increase in AQP2 gene expression. A deficiency in the vasopressin may be affected by *diabetes insipidus* and show an increase of urine excretion to 10 – 15 liters a day.

Ionophores are common growth enhancers in livestock feed, and are used in veterinary medicine as a coccidiostat in poultry. They are lipid-soluble molecules usually synthesized by microorganisms to transport ions across the lipid bilayer of the cell membrane. There are two main kinds of ionophores: *mobile ion carriers* that bind to a particular ion, shielding its charge from the surrounding environment, and thus facilitate its crossing the hydrophobic interior of the lipid membrane (ex. Valinomycin) and *channel formers* that introduce a hydrophilic pore into the membrane, allowing ions to pass through while avoiding contact with the membrane's hydrophobic interior (ex. Gramicidin). Ionophores disrupt the transmembrane ion concentration gradients required for the proper functioning and survival of microorganisms, and thus have antibiotic properties. They are produced naturally by a variety of microbes and act as a defense against competing microbes. Many antibiotics, particularly the macrolide antibiotics, are ionophores that exhibit high affinities to Na^+ or K^+.

Gramicidin is a polypeptide with alternating L- and D-amino acids, sharing the general formula: formyl-L-**X**-Gly-L-Ala-D-Leu-L-Ala-D-Val-L-Val-D-Val-L-Trp-D-Leu-L-**Y**-D-Leu-L-Trp-D-Leu-L-Trp-ethanolamine (X and Y depend upon the type of gramicidin molecule). It is obtained from the soil bacteria species *Bacillus brevis*. Gramicidin's bactericidal activity is a result of increasing the permeability of the bacterial cell membrane, allowing Na^+ ions to

travel through unrestricted and thereby destroying the ion gradient between the cytoplasm and the extracellular environment.

Valinomycin is obtained from the cells of several *Streptomyces*. It consists of enantiomers D- and L-valine, D-hydroxyvaleric acid and L-lactic acid. It functions as a potassium-specific transporter and facilitates the movement of K^+ ions through lipid membranes "down" an electrochemical potential gradient.

5.2.2. Active Transport

Active transport is the movement of a substance across a cell membrane against its concentration gradient (from low to high concentration) with assistance from a membrane channel protein and an energy source (ATP). Small, hydrophilic molecules will cross the membrane by active transport. Active transport uses ATP hydrolysis directly (*primary active transport*) or indirectly (*secondary active transport*).

5.2.2.1. Primary active transport

Active transporters are membrane proteins specifically bind and move the molecules across the membrane to a unique direction using ATP hydrolysis as an energy source. Most of the enzymes that perform this type of transport are **transmembrane ATPases**, which classified according to their protein sequence homology and structures. P-type of ATPases (from ***p****hosphorylation*) are represented by Na^+/K^+-ATPase, proton-potassium pump (H^+/K^+-ATPase) and Ca^{2+}-ATPase, they all are sensitive to vanadate inhibition. V-type (from ***v****acuole type*) ATPases regulate H^+ gradients and evoke acidification of lysosomes, endosomes, Golgi, and secretory vesicles. F-type (from energy coupling ***f****actor*) ATPases example is H^+-ATPase/ATP-synthase which generate ATP energy from moving the proton across inner mitochondrial membrane. There are F_1 and F_0 subcomplexes in this enzyme: F_1 generates ATP, F_0 lets H^+ go through the membrane and may bind olygomycin that is ATP synthase blocker. There are also **ABC**

transporters (from A<u>TP</u>-***b***inding ***c***assette) - proteins for active transport of hydrophobic organic chemicals.

P-type ATPases, also known as E1-E2 ATPases**,** are a large group of evolutionarily related ion pumps that are found in bacteria, archaea and eukaryotes. They named P-type ATPases because they catalyze auto- (or self-) phosphorylation of a key conserved aspartate residue within the pump molecule. So they form a high-energy aspartyl-phosphoranhydride intermediate in the reaction cycle, and they interconvert between at least two different conformations, denoted by E1 and E2. The E1-E2 notation stems from the initial studies on this family of enzymes made on the Na^+,K^+-ATPase, where the sodium form and the potassium form are referred to as E1 and E2, respectively (Fig. 5.8).

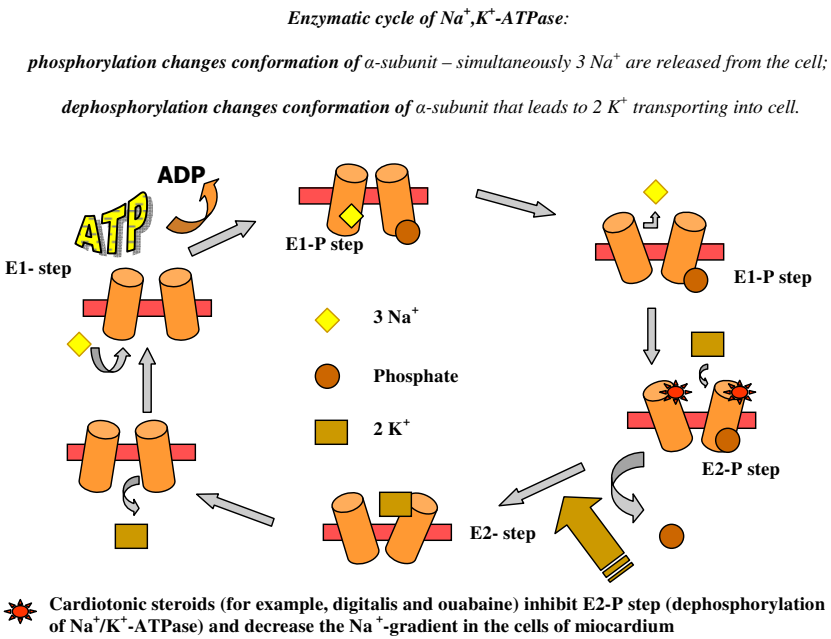


Fig. 5.8. The principle of P-type ATPases function

Na^+/K^+-pump is α2β2 tetramer. The α-subunit has 12 transmembrane domains (TMDs). The β-subunit is glycosylated and has one TMD. It actively transports 3 Na^+ ions out and 2 K^+ ions into the cell using a single ATP molecule (Fig. 5.8). This causes a higher concentration of sodium outside and a higher concentration of potassium inside the cell. Such an unequal transport of charge generates a voltage across the membrane or a membrane potential. The outside will become relatively positive and the inside relatively negative since $3Na^+$ (+3 charges) go out for every $2K^+$ (+2 charges) coming in. The result is an electrochemical gradient - a combination of chemical concentration gradient and electrical gradient.

Certain steroids derived from plants are potent inhibitors of Na^+/K^+-pump. So, cardiotonic steroid digitalis from the dried leaf of the foxglove plant inhibits the dephosphorylation of the E2-P form of the ATPase and increases the force of heart muscle contraction because inhibition of Na^+/K^+-pump causes higher level of Na^+ inside the cell, slower extrusion of Ca^{2+} by the Na^+/Ca^{2+}-exchanger and increase in the intracellular level of Ca^{2+} that enhances the ability of cardiac muscle to contract.

The *plasma membrane Ca^{2+}-ATPase (PMCA)* is a transport protein in the plasma membrane of cells that serves to remove calcium from the cell with a stoichiometry of one Ca^{2+} ion removed for each molecule of hydrolyzed ATP. It is vital for regulating the amount of Ca^{2+} within cells.

The *sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA)* transfers two Ca^{2+} ions from the cytosol of the cell to the lumen of the sarcoplasmic (endoplasmic) reticulum at the expense of ATP hydrolysis during muscle relaxation.

The *gastric hydrogen potassium ATPase* or *H^+/K^+-ATPase* is the proton pump of the stomach and, as such, is the enzyme primarily responsible for the acidification of the stomach contents. The H^+/K^+-ATPase is found in parietal cells, which are highly specialized epithelial cells located in the inner cell lining of the stomach called the gastric mucosa. Parietal cells possess an extensive secretory membrane system and the H^+/K^+-ATPase is the major protein constituent of these membranes. The H^+/K^+-ATPase transports one hydrogen ion (H^+) from the cytoplasm of the parietal

cell in exchange for one potassium ion (K^+) retrieved from the gastric lumen (Fig. 5.9).

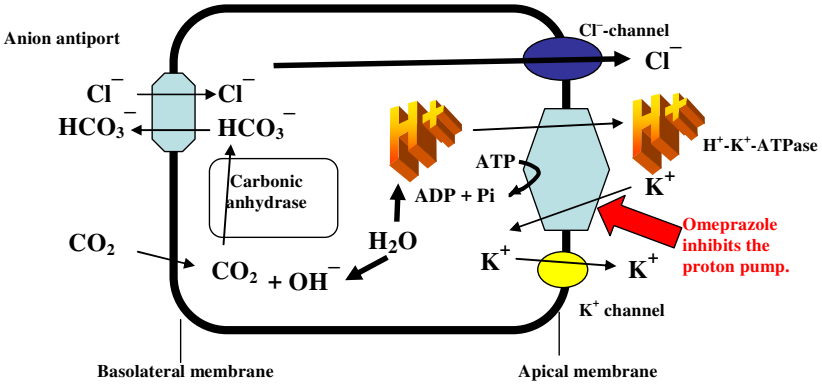


Fig. 5.9. The scheme of H^+/K^+-ATPase function

As ion pump the H^+/K^+-ATPase is able to transport ions against a concentration gradient using energy derived from the hydrolysis of ATP. Like all P-type ATPases, a phosphate group is transferred from adenosine triphosphate to the H^+/K^+-ATPase during the transport cycle. This phosphate transfer powers a conformational change in the enzyme that helps drive ion transport. H^+/K^+-ATPase is an electroneutral antiport: K^+ is removed by K^+-channel and concurrently Cl^- channel removes Cl^- to the same direction.

Two drug categories are commonly used to inhibit H^+/K^+-ATPase activity. Histamine receptor H_2 antagonists like *cimetidine* (Tagamet) inhibit the signaling pathway that leads to histamine-dependent activation of the ATPase. Proton pump inhibitors like *omeprazole* (Prilosec) directly bind to and inactivate the H^+/K^+-ATPase.

ABC (ATP-binding cassette) transporters pass wide variety of substrates across extra- and intracellular membranes, including metabolic products, lipids, sterols and drugs**.** Their molecules have 6 or 12 trans-membrane helices with 2 ATP binding

sites, drug or ligand- binding sites yet to be clearly identified. Examples are multidrug-resistance protein (MRPs), p-glycoproteins (P-gp - bile salt export pump, multidrug-resistance glycoproteins (MDRs) etc.) that cause active transport and one transporter which accomplishes facilitate diffusion - chloride channel CFTR.

Multidrug resistance proteins and glycoproteins are chemical pumps which use ATP energy to actively remove the hydrophobic drugs out of the cells and cause development of resistance to one drug also makes the cell less sensitive to a range of other compounds. After ATP hydrolysis, changes in their structures facilitate the movement of hydrophobic drugs.

5.2.2.2. Secondary active transport (cotransport, coupled transport)

A single ATP-powered pump that transports a specific solute can indirectly drive the active transport of several other solutes in a mechanism called **cotransport**. A substance that has been pumped across a membrane can do work as it moves back across the membrane by diffusion, analogous to water that has been pumped uphill and performs work as it flows back down. Another transport protein, a **cotransporter,** separate from the pump, can couple the "downhill" diffusion of this substance to the "uphill" transport of a second substance against its own concentration (or electrochemical) gradient. Therefore secondary active transport needs in special carrier proteins which can use passive flow of one ion (e.g., Na^+) to pull in another ion against its concentration gradient.

The two main forms of secondary active transport are antiport and symport. In **antiport** two species of ion or other solutes are pumped in opposite directions across a membrane. An example is the *sodium-calcium exchanger* or antiporter, which allows three sodium ions into the cell to transport one calcium ion out. By **symport** the two species move in the same direction across the membrane. An example is the *glucose symporter SGLT1*, which co-transports one glucose (or galactose) molecule into the cell for every two sodium ions it imports into the cell. This symporter is located in the small intestines, trachea, heart, brain, testis, and prostate (Fig. 5.10).

Glucose uptake from lumen to the capillaries using;

- **glucose transporter (permease, facilitated diffusion),**
- **glucose-sodium symport (secondary-active transport)**
- **Na^+-K^+-ATPase (primary-active transport)**

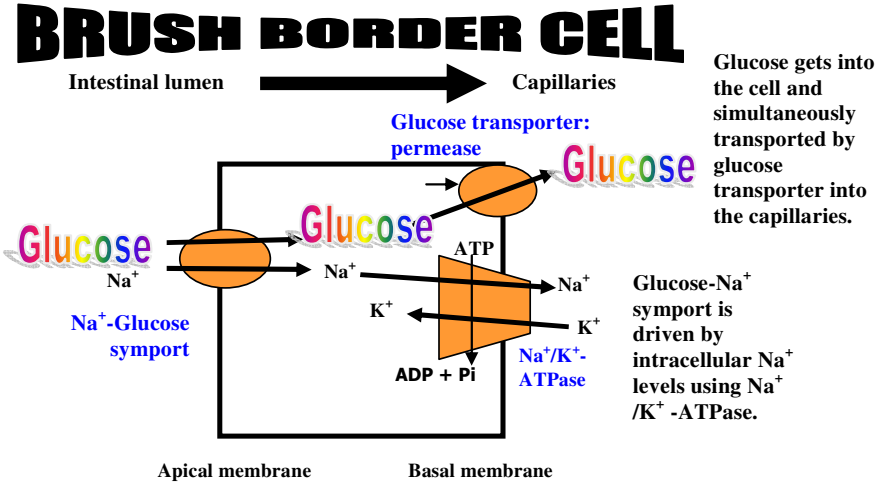


Fig. 5.10. The scheme of SGLT1-dependent cotransport

Primary and secondary active transporters work coordinately in animal cells. They generate membrane potential, generate proton gradient, maintain acidity, etc.

5.2.2.3. Cytosis as active transport

Large molecules, such as proteins and polysaccharides, as well as larger particles, generally cross the membrane in bulk by mechanisms that involve packaging in vesicles. Like active transport, these processes require energy.

The cell secretes certain biological molecules by the fusion of vesicles with the plasma membrane; this process is called **exocytosis**. A transport vesicle that has budded from the Golgi apparatus moves

along microtubules of the cytoskeleton to the plasma membrane. When the vesicle membrane and plasma membrane come into contact, specific proteins rearrange the lipid molecules of the two bilayers so that the two membranes fuse. The content of the vesicle then spill to the outside of the cell, and the vesicle membrane becomes part of the plasma membrane.

Many secretory cells use exocytosis to export products. For example, the cells in the pancreas that make insulin secrete it into the extracellular fluid by exocytosis. Other example, neurons (nerve cells) use exocytosis to release neurotransmitters that signal other neurons or muscle cells.

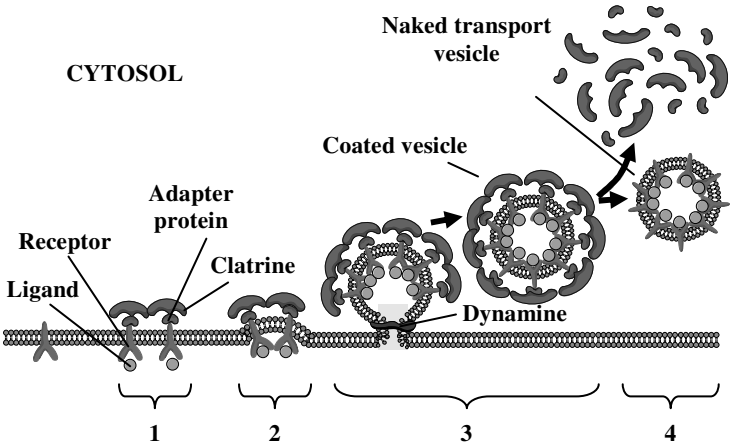
In **endocytosis**, the cell takes in biological molecules and particulate matter by forming new vesicles from the plasma membrane. Although the proteins involved in the process are different, the events of endocytosis look like the reverse of exocytosis. A small area of the plasma membrane sinks inward to form a pocket. As the pocket deepens, it pinches in, forming a vesicle containing material that had been outside the cell.

There are three types of endocytosis: phagocytosis ("cellular eating"), pinocytosis ("cellular drinking"), and receptor-mediated endocytosis.

In *phagocytosis*, a cell engulfs a particle by wrapping pseudopodia around it and packaging it within a food vacuole. The particle is digested after the food vacuole fuses with a lysosome containing hydrolytic enzymes.

In *pinocytosis*, the cell "gulps" droplets of extracellular fluid into tiny vesicles. It is not the fluid itself that is needed by the cell, but the molecules (salts, monomers, small molecules) dissolved in the droplets. These small molecules are transported by proteins out of the vesicles and into the cytosol. Because any and all included solutes are taken into the cell, pinocytosis is nonspecific in the substances it transports.

Receptor-mediated endocytosis enables the cell to acquire bulk quantities of specific substances, even though those substances may not be very concentrated in the extracellular fluid. This type of endocytosis needs in membrane receptors to which specific ligands bind (Fig. 5.11).



1 – Ligand binding and coated assembly formation
2 – Bud formation
3 – Vesicle formation
4 – Uncoating

Fig. 5.11. The scheme of receptor-mediated endocytosis

Then the receptor proteins cluster in regions of the membrane called *coated pits*, which are lined on their cytoplasmic side by a fuzzy layer of *coat protein clathrin*. *Adaptin* (4 types) binds clathrin and receptors, acting as a bridge. Next, each coated pit forms a coated vesicle containing the ligand molecules. There are relatively more bound molecules inside the vesicle, but other molecules are also present. *Dynamin* forms a ring around the bud (GTPase). After the ingested material is liberated from the vesicle, the emptied receptors are recycled to the plasma membrane by the same vesicle. *Hsp70 chaperone* and *auxillin* uncoat the vesicle.

Human cells use receptor-mediated endocytosis to take cholesterol for membrane synthesis and fornation of other steroids. Cholesterol travels in blood in particles called *low-density lipoproteins (LDLs)*, a complex of lipids and a protein. LDLs bind to LDL receptors on plasma membranes and then enter the cells by endocytosis (LDLs thus act as ligands, a term for any molecule that binds specifically to a receptor site on another molecule). In humans

with *familial hypercholesterolemia*, an inherited disease characterized by a very high level of cholesterol in the blood, LDLs cannot enter cells because the LDL receptor proteins are defective or missing. Consequently, cholesterol accumulates in the blood, where it contributes to early atherosclerosis, the buildup of lipid deposits within the walls of blood vessels. This buildup causes the walls to bulge inward, thereby narrowing the vessels and impeding blood flow.

Vesicles not only transport substances between the cell and its surroundings but also provide a mechanism for rejuvenating or remodeling the plasma membrane.

5.3. Test questions

1. Which ion channel has such characters: blocked by a specific neurotoxin (tetrodotoxin from puffer fish); has an α-subunit, which contains 4 internal repeats - each repeat contains 5 hydrophobic segments (S1, S2, S3, S5, S6). Segment S4 is highly positive charged and forms voltage sensor of the channel.

A. The Voltage-Gated Sodium Channel;
B. The Voltage-Gated Potassium channel;
C. CFTR;
D. The Voltage-Gated Calcium Channel;
E. Acetylcholine-gated cation channel

2. Glucose is a water soluble molecule. By what route does it pass through the cell membrane?

A. Transferred by the carrier proteins
B. Diffuses across the membrane
C. Passes between the phospholipid molecules
D. Enters in a vesicle
E. It can't pass through the cell membrane

3. Which of the following is the difference between active transport and facilitated diffusion?

A. Active transport involves membrane proteins, but facilitated does not.
B. Active transport requires energy from the cell, but facilitated diffusion does not.
C. Facilitated diffusion involves membrane proteins, but active transport uses active proteins.
D. Facilitated diffusion requires energy from the cell and active transport does not.
E. -

4. What is the function of a P-glycoprotein?

A. They are involved in the anchoring of cells to an extracellular matrix.
B. They break down the basement membranes round blood vessels prior to angiogenesis.
C. They are responsible for the resistance shown by some cancer cells to anticancer drugs.
D. They induce metastasis.
E. -

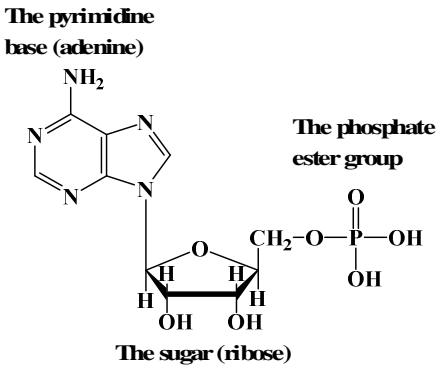
5. **Which structure realizes facilitated diffusion?**

A. Na+, glucose-cotransporting protein;
B. P-glycoproteins;
C. MRPs
D. H+-ATPase;
E. CFTR;

PART 6.
CLASSIFICATION, STRUCTURE AND BIOLOGICAL ROLE OF HETEROCYCLIC COMPOUNDS

Heterocyclic compounds are the organic compounds contain one or more aromatic or non-aromatic rings with at least one atom (called *heteroatom*) in the ring being an element other than carbon (C), most frequently oxygen (O), nitrogen (N), or sulfur (S). These compounds belong to different classes of molecules.

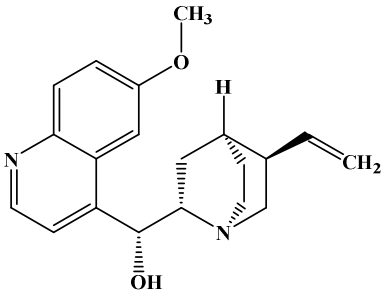
Two thirds of all organic compounds are aromatic heterocycles. Heterocyclic compounds include many of the biochemical material essential to life. For example, nucleic acids **-** the chemical substances that carry the genetic information controlling inheritance – are built from nucleotides that contain heterocyclic compounds - purine and pyrimidine bases (Fig. 6.1).



Fig, 6.1. The structure of adenine nucleotide – adenosine-5'-monophosphate (AMP)

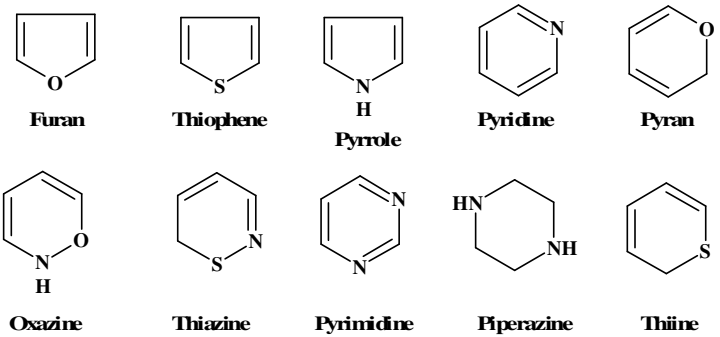
In addition, many naturally occurring pigments, vitamins, and antibiotics belong to this class of organic compounds. Most pharmaceuticals are heterocycles. In particular, *quinine*, whose

formula is shown in Fig. 6.2, is used for malaria treatment for 400 years.



Fig, 6.2. The structure of quinine

Heteroatom can be represented by N, O, S, B, Al, Si, P, Sn, As, Cu, but more often - by N, O, or S (Fig. 6.3).



Fig, 6.3. The structures of the most widespread heterocyclic compounds

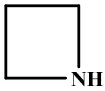
6.1. Heterocyclic compounds classification

There are two main ways of heterocyclic compounds classifying:

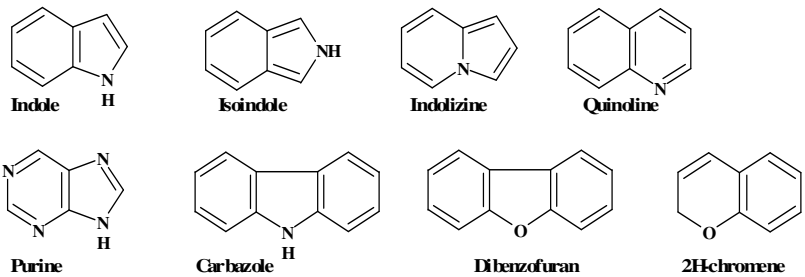
- By ring size
- Aromatic/non aromatic

When classification based on the ring size is used, heterocyclic compounds can be divided into **three-, four-, five- and six-membered** ones (Tab. 6.1). Seven-, eight- or more-membered heterocyclic compounds are less common.

Table 6.1.
The examples of three-, four-, five- and six-membered heterocyclic compounds

Three-membered	Four-membered
 Ethylene oxide oxirane  Ethyleneimine aziridine  Ethylene sulfide thiirane	 Oxetane  Azetidine  Thietane
Five-membered	**Six-membered**
 Furan  Pyrrole  Thiophene	 Pyridine

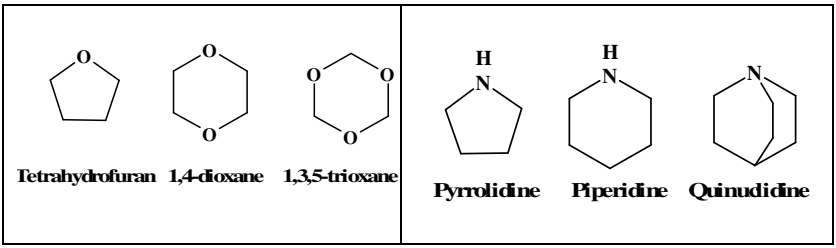
In addition there are **the fused rings-containing heterocyclic compounds** (they have two ore more rings joined at two adjacent atoms) (Fig. 6.4).



Fig, 6.4. The formulas of the heterocyclic compounds with fused rings in their structures

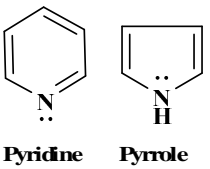
Nonaromatic heterocyclic compounds have physical and chemical properties that are typical of the particular hetero atom (Fig. 6.5):

- *tetrahydrofurane* and *1,4-dioxane* are typical **ethers**
- *1,3,5-trioxane* behaves as **an acetal**
- *pyrrolidine* and *piperidine* are typical **secondary amines**
- the bicyclic compound *quinuclidine* is **a tertiary amine**.

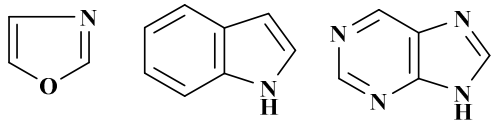


Fig, 6.4. The structures of some nonaromatic heterocyclic compounds

Aromatic heterocyclic compounds include such compounds as *pyridine*, where nitrogen replaces one of the CH groups in benzene, and *pyrrole*, in which the aromatic sextet is supplied by the four electrons of the two double bonds and the lone pair on nitrogen:



Other aromatic heterocycles contain more than one hetero atom, and still others contain fused aromatic rings. Examples which we will treat in more detail later include *oxazole*, *indole* and *purine*:



Oxazole **Indole** **Purine**

6.2. Nomenclature of heterocyclic compounds

Systematic nomenclature system is used for naming 3-10-membered monocyclic heterocyclics of various degree of the unsaturation containing one/more heteroatoms. It specifies the following characters such as:

- the ring size (3-, 4-, 5-, 6-, 7-, 8-membered);
- the type of heteroatom;
- the position of heteroatom;
- one/more heteroatom etc.

The commonly used names for heterocyclic compounds are their **trivial names**.

Systematic nomenclature of heterocyclic compounds: the Hantzsch-Widman system. The compound formed by replacing a carbon by a hetero atom is named by an appropriate prefix (Tab. 6.2):

- **aza** for nitrogen (N),
- **oxa** for oxygen (O),
- **thia** for sulfur (S).

Table 6.2.
The principle of heterocyclic compounds names formation based on heteroatom type

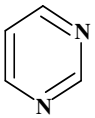
Heteroatom	Prefix
O	oxa
S	thia
N	aza

Saturated and unsaturated monocyclic rings are named according to ring size (Tab. 6.3).

Table 6.3.
The principle of heterocyclic compounds names formation based on the ring size

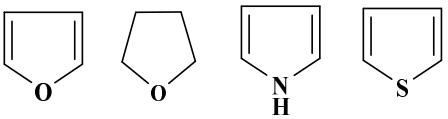
	Ring Size	3	4	5	6
Suffix	**Unsaturated**	irene	ete	ole	ine
	Saturated	irane	etane	olane	inane

When **two or more the same heteroatoms** are present, the prefixes **"di", "tri"**, etc. are used:



Pyrimidine
1,3-diazine

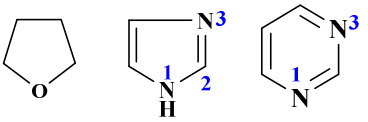
When **a maximally unsaturated ring includes a saturated atom**, its location may be designated by a **"#*H*"** prefix to avoid ambiguity. The examples of systematic and trivial names of some heterocyclic compounds with one heteroatom are given on Fig. 6.5.



Tetrahydrofuran **Furan** **Pyrrole** **Thiophene**
Oxolane **Oxole** **1*H*-azole** **Thiole**

Fig, 6.5. Trivial (black) and systematic (blue) names of some heterocyclic compounds with one heteroatom

In a monocyclic compounds **that contain only one heteroatom** numbering starts at this atom. When **more than one heteroatom is present**, the ring is numbered **to give the substituents or other heteroatoms the lowest number possible** with the ending**:**

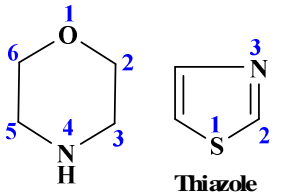


Furan **Imidazole** **Pyrimidine**
Oxole **1,3-diazole** **1,3-diazine (not 1,5-diazine!!!)**

If the heteroatoms are different they are named in the order of "O, S, N," etc. **by combining the prefixes** (Tab. 6.2) in the order of preference:

oxygen is considered to be older, then sulfur and then nitrogen

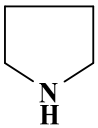
Therefore, in the presence of oxygen and nitrogen in the ring, numbering begins with oxygen):



1,4-Oxazine **Thiazole**
1,3-Thiazole

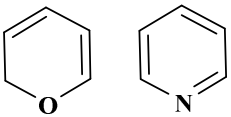
Exceptions and modifications of nomenclature of the Hantzsch-Widman system. **Saturated 5-membered heterocycles with nitrogen**

as heteroatom should use respectively the traditional "**olidine**" suffix:



Pyrrolidine
Azolidine

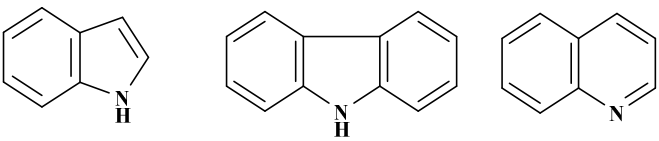
Established use of **oxine** and **azine** for other compounds prohibits their use for pyran and pyridine respectively:



Pyran
***2H*-Pyran**

Pyridine

Trivial names are also often used **for fused aromatic rings-contained heterocyclic compounds** (Fig. 6.6).



Indole
2,3-Benzopyrrole

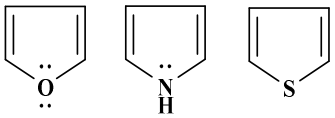
Carbazole
Dibenzopyrrole

Quinoline
1-azanaphtalene
benzo[b]pyridine

Fig, 6.6. Trivial (black) and systematic (blue) names of some fused aromatic rings-contained heterocyclic compounds

6.3. Five-membered heterocyclic compounds with one heteroatom

The main examples of five-membered heterocyclic compounds are aromatic heterocycles pyrrole, furan and thiophene:

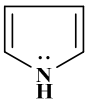


Furan **Pyrrole** **Thiophene**

6.3.1. Physical properties of furan, pyrrole, thiophene

Thiophene

Furan



Pyrrole

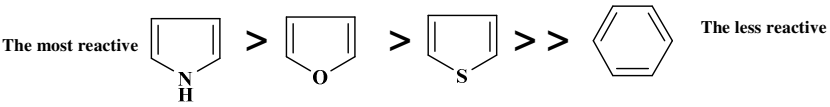
At room temperature, **thiophene** is a colorless liquid with a mildly pleasant odor reminiscent of benzene, with which thiophene shares some similarities. Like benzene, thiophene forms an azeotrope with water.

Furan is typically derived by the thermal decomposition of pentose-containing materials, cellulosic solids especially pine-wood. Furan is a colorless, flammable, highly volatile liquid with a boiling point close to room temperature. It is toxic and may be carcinogenic.

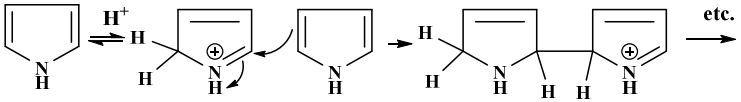
Pyrrole is a heterocyclic aromatic organic compound. Substituted derivatives are also called *pyrroles*. *Porphobilinogen* is a trisubstituted pyrrole, which is the biosynthetic precursor to many natural products.

6.3.2. Chemical properties of furan, pyrrole, thiophene

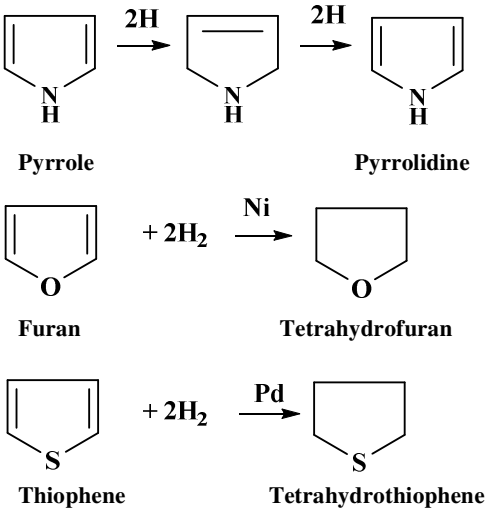
The most typical reaction of furan, pyrrole and thiophene is electrophilic substitution. All three heterocycles are much more reactive than benzene, the reactivity order being:



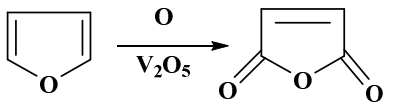
1. Interaction with mineral acids: Pyrroles are polymerized by even mineral acids, probably by such mechanism as the following:



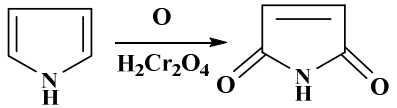
2. Reduction reactions for furan, pyrrole and thiophene follow the scheme**:**



3. Reactions of oxidation – their examples for furan and pyrrole are given below:



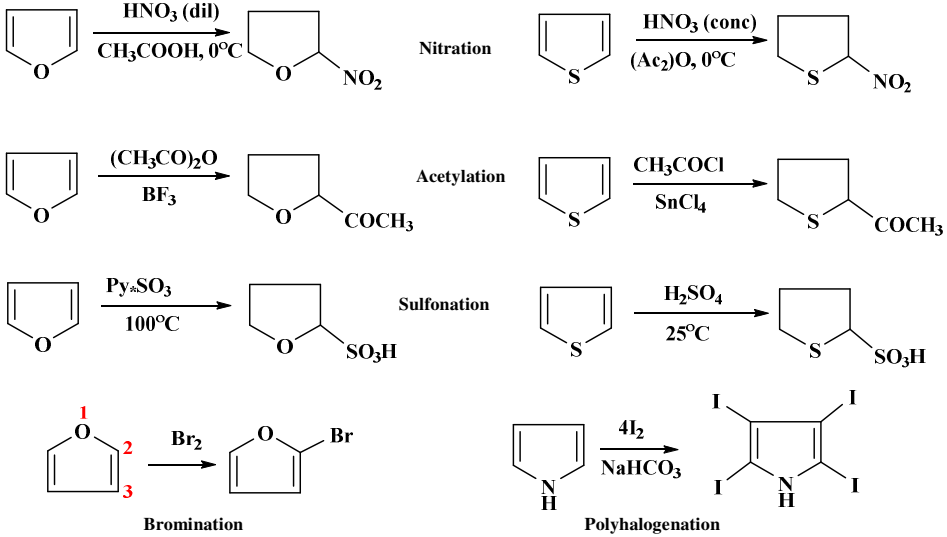
Furan **Maleinic anhydride**



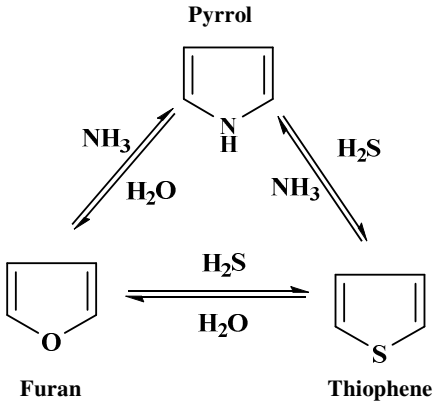
Pyrrole **Maleinmide**

4. Reactions of electrophilic substitution are typical for all of these aromatic heterocycles:

For furan: **For thiophene:**



5. Reciprocal transformation of furan, pyrrole and thiophene (Yurie`s cycle reactions):



For identification of pyrrole and furan the method coloring of a pine chip is used. Couples of pyrrole painted a pine chip soaked in hydrochloric acid in the red colour and furan - in the green colour. Qualitative reaction on thiophene is indophenin`s reaction: a mixture of izathine with concentrated sulfuric acid painted in the blue colour.

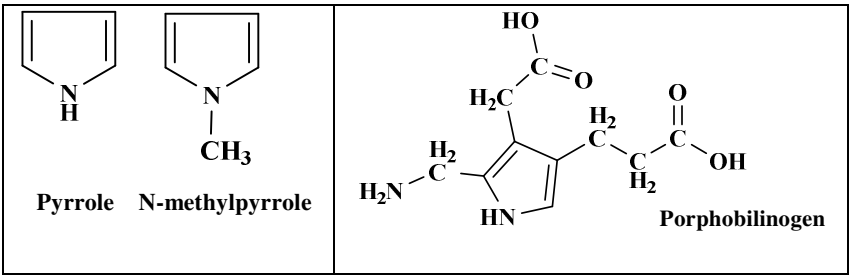
6.3.3. The important derivatives of pyrrole, furan and thiophene

6.3.3.1. Pyrrole derivatives.

The main pyrrole derivatives are:

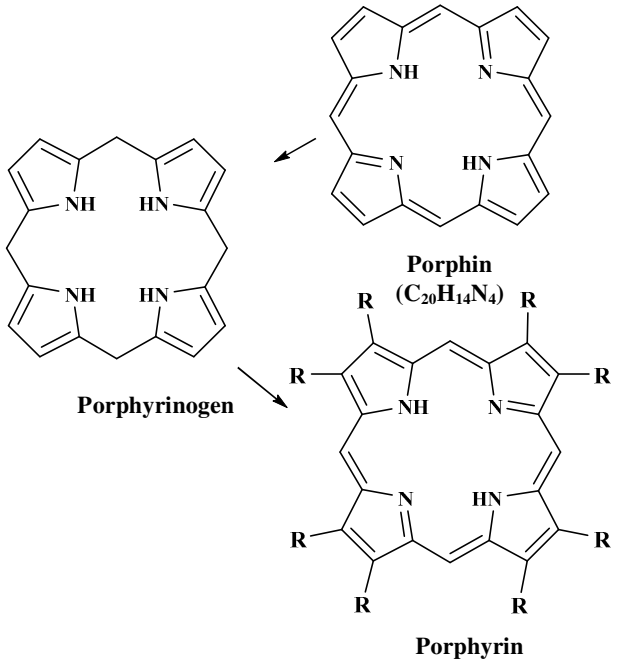
- *porphobilinogen and tetrapyrrole compounds* (porphin, porphyrinogens, porphyrins)
- *indole and its derivatives*.

Porphobilinogen and tetrapyrrole compounds (porphin, porphyrinogens, porphyrins). **Pyrrole**, or pyrrol, is a heterocyclic aromatic organic compound with a five-membered ring. Substituted derivatives are also called **pyrroles** - for example, *N-methylpyrrole*:



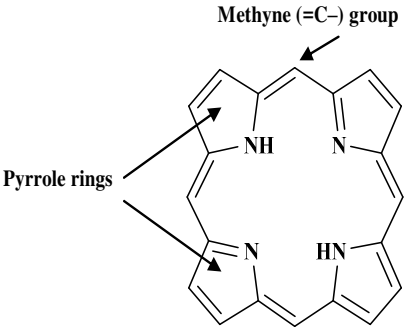
Porphobilinogen is a trisubstituted pyrrole, which is the biosynthetic precursor to many natural products.

Porphin, sometimes spelled *porphine*, is the parent chemical compound for types of biochemically significant compounds called **porphyrinogens** and **porphyrins (**Fig. 6.7).



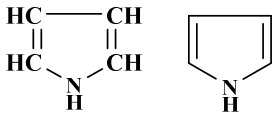
Fig, 6.7. The common structures of porphin, porphyrinogen and porphyrin

Porphin molecule essentially consists of four pyrrole rings joined together by four methyne (=CH—) groups to form a larger macrocycle ring:

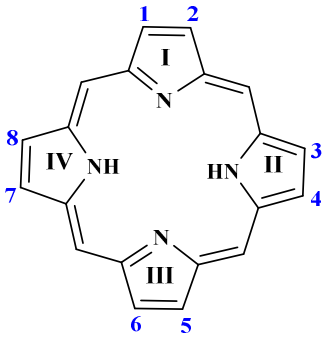


Porphin
($C_{20}H_{14}N_4$)

Porphyrins are porphin derivatives containing the substituents in the positions 1, 2, 3, 4, 5, 6, 7, 8:



Pyrrole



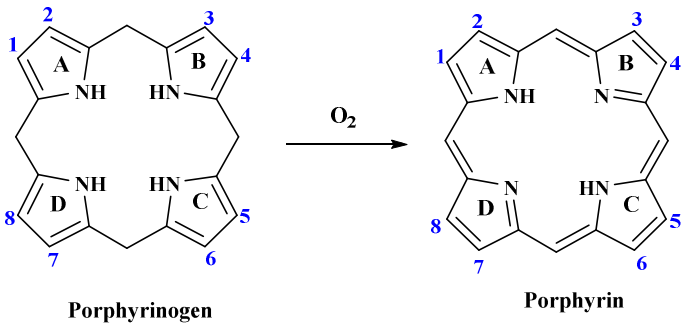
Porphin
($C_{20}H_{14}N_4$)

There are next types of porphyrins:

- *uroporphyrins*
- *coproporphyrins,*
- *protoporphyrins,* etc.

Each class of porphyrins contains several isomers, denoted by the letters of the Latin alphabet (III, IX, etc.).

Porphyrinogens are metabolic precursors of porphyrins**.** *Colorless porphyrinogens* are converted into *colored (red) porphyrins* via enzymatic or non-enzymatic (under the influence of oxygen of air) oxidation. Therefore *porphyrinogens have four saturated methylene (–CH_2–) groups* that couples four pyrrole rings, whereas *four pyrrole rings of porphyrins are linked by four methyne (–CH=) groups,* and *two pyrrole rings have no NH groups* (Fig. 6.8).



Fig, 6.8. Scheme of porphyrin formation and the main differences between porphyrinogen and porphyrin.

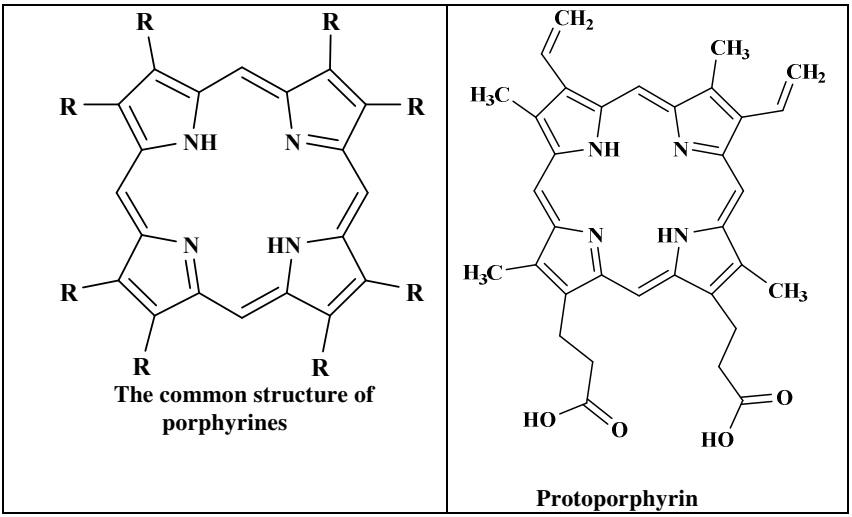
There are next types of porphyrinogens:

- *uroporphyrinogens,*
- *coproporphyrinogens,*
- *protoporphyrinogens,* etc..

Protoporphyrins are tetrapyrroles containing the following side chains (Fig. 6.9):

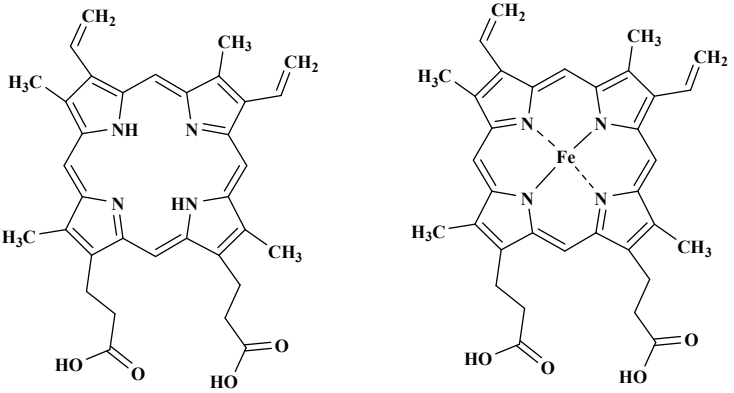
- methyl (4) – 1, 3, 5 and 8 locations

- propionic acid (2) – 2 and 4 locations
- vinyl (2) – 6 and 7 location



Fig, 6.9. The particularities of protoporphyrin structure

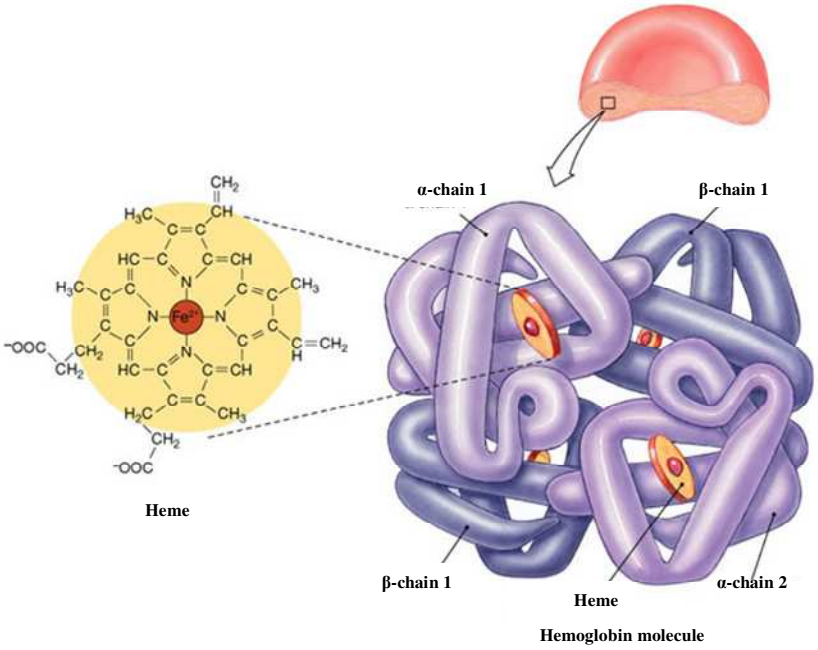
Four central nitrogen atoms in biological *protoporphyrin* compounds can act as *ligands* to bond metal ion (for example, *iron* ion) in the center of the molecule (Fig. 6.10).



Fig, 6.10. The structures of free (left) and bound (right) protoporphyrins

When there is no metal ion bound to nitrogens in the center, such compounds are called **free protoporphyrins**. If nitrogens are bonded to metal in the center, **protoporphyrins** are **bound**. Complex of protoporphyrin IX with iron (Fe^{2+}) is called **heme**. Protoporphyrin IX with magnesium ion (Mg^{2+}) is the main part of the **chlorophylls**.

Heme is found in myoglobin, hemoglobin, cytochromes, enzyme catalase. **Hemoglobin** (also spelled haemoglobin and abbreviated Hb) is the iron- containing oxygen-transport hemoprotein in the red blood cells of vertebrates (Fig. 6.11).



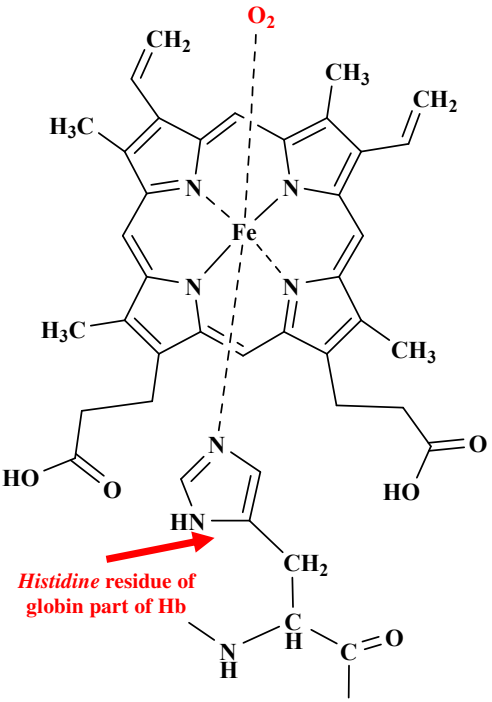
Fig, 6.11. The structures of hemoglobin A and heme.

It carries O_2 from the lungs to other organs and tissues, where it releases O_2 for use by cells, and binds CO_2 to bring it to the lungs.

The hemoglobin molecule contains the protein part (**globin**) and the colored non-protein part (**heme**).

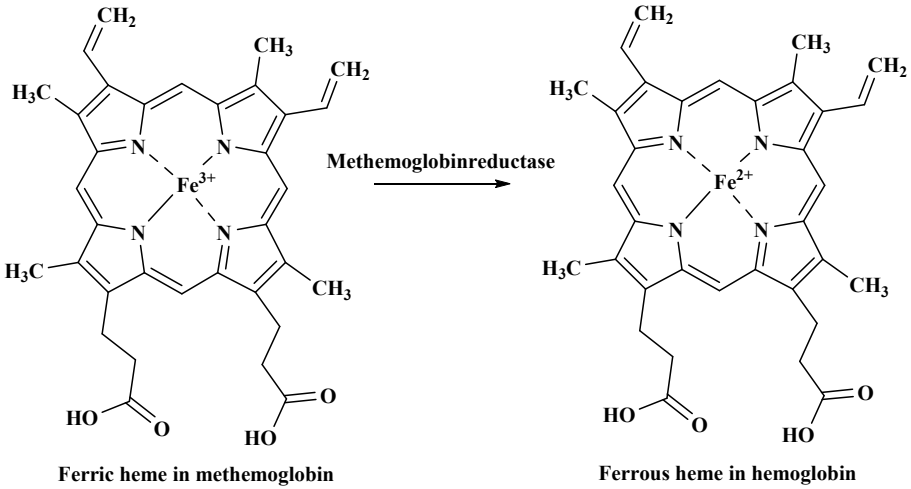
Globin of **hemoglobin A** (HbA, also known as *adult* hemoglobin or HbA1) consists of two α- and two β-subunits (or chains), each of which is connected with heme (Fig. 6.11). Consequently, *the entire hemoglobin molecule contains 4 protein subunits and 4 heme structures*.

A heme group consists of an **iron (Fe) ion** held in a heterocyclic ring of protoporphyrin IX. In this structure Fe ion coordinates with four nitrogens of pyrrol rings, which all lie in one plane. Iron also bounds to the *histidine* residue of globin below the porphyrin ring as well as can reversibly bind O_2 above the porphyrin ring with **oxyhemoglobin** forming (Fig. 6.12).



Fig, 6.12. The linkage between heme and globin part in hemoglobin molecule and mechanism of oxyhemoglobin formation

Iron ion in hemoglobin may be either in Fe^{2+} or in Fe^{3+} state (i.e. iron presences in the ferric form instead of the usual ferrous form), but **methemoglobin** *(hemoglobin with Fe^{3+})* cannot bind oxygen. The enzyme *methemoglobin reductase* converts methemoglobin (Fe^{3+}) to hemoglobin (Fe^{2+}) (Fig. 6.13).

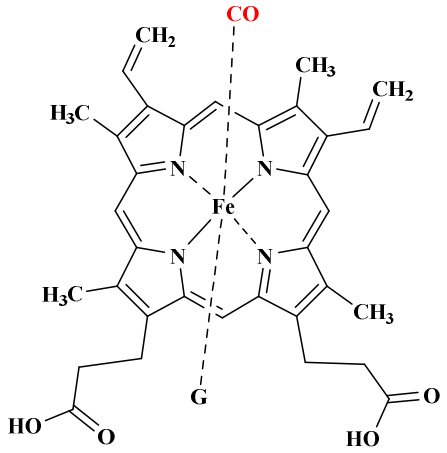


Ferric heme in methemoglobin **Ferrous heme in hemoglobin**

Fig, 6.13. Methemoglobin and methemoglobin reductase-dependent reaction

Methemoglobinemia (congenital or acquired) occurs when red blood cells (RBCs) contain methemoglobin at levels higher than 1%. This results in a decreased availability of oxygen to the tissues. One of the causes of methemoglobinemia is methemoglobin reductase deficiency

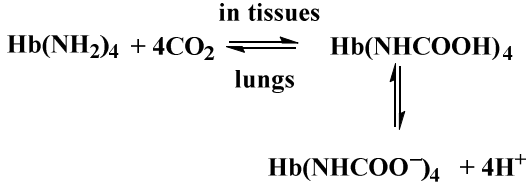
Carbon monoxide (CO) binds to hemoglobin at the same sites as O_2 (**to Fe^{2+} ion**), but approximately 200 times more tightly. Complex CO with hemoglobin is called **carboxyhemoglobin** (HbCO) (Fig. 6.14).



Fig, 6.14. Carboxyhemoglobin structure

Normally, O_2 would bind to hemoglobin in the lungs and be released in areas with low oxygen partial pressure (e.g. active muscles). When CO binds to hemoglobin, it cannot be released as easily as O_2. The slow release rate of CO causes an accumulation of CO-bound hemoglobin molecules. Because of this, fewer hemoglobin molecules are available to bind and deliver O_2, thus causing the gradual suffocation associated with carbon monoxide poisoning.

Carbhemoglobin is a complex of hemoglobin with carbon dioxide (CO_2), and is one of the forms in which CO_2 exists in the blood. 10% of CO_2 is carried in blood this way (85% carried in blood as bicarbonate [hydrogen carbonate], 5% to 7% carried as free CO_2, in solution, or plasma). Hemoglobin can bind from one to four molecules of CO_2. Carbon dioxide molecules form an amide linkage to the four terminal-amine groups of the four protein chains in the deoxy form of the molecule (Fig. 6.15). Thus, one hemoglobin molecule can transport four carbon dioxide molecules back to the lungs, where they are released.



Fig, 6.15. Scheme CO_2 binding to the one subunit of Globin. All four Globin subunits (2α, 2β-chains) in general have 4 terminal NH_2-groups and therefore can bind four CO_2 molecules

Besides adult hemoglobin (HbA), there are embryonic, fetal hemoglobins as well as HbA2, glycosylated hemoglobin and different types of hemoglobins with mutations.

Embryonic hemoglobins are found in embryos of up to 8 weeks of gestation. They are:

- *Gower 1* (ζ2ε2)
- *Gower 2* (α2ε2)
- *Hemoglobin Portland I* (ζ2γ2)
- *Hemoglobin Portland II* (ζ2β2).

Fetal hemoglobin (the predominant hemoglobin from the eighth week to term of birth) - **Hb F** (α2γ2). The major transition from fetal to adult hemoglobin synthesis occurs in the perinatal period - "around the time of birth". HbF and embryonic Hb have a higher affinity for oxygen than HbA, facilitating oxygen transfer from the maternal to the fetal circulation

Hemoglobins that exist in blood erythrocytes in norm after birth:

- **Hb A** (sometimes called Hb A1) (α2β2) – the most common with a normal amount over 95%
- **Hb A2** (α2δ2) – δ chain synthesis begins late in the third trimester and, in adults, it has a normal range of 1.5–3.5%
- **Hb F** (α2 γ2) - less than 2% of hemoglobin.
- **Glycosylated** (or glycated) **hemoglobin** (**HbA1c)** is the form of hemoglobin bound with glucose. The binding of glucose to amino acids of globin takes place spontaneously

(without the help of an enzyme). Its content correlates with the maintenance of glucose in plasma of blood, so determination of this index is used for early diagnosis of diabetes mellitus and assessment of the effectiveness of its therapy.

Hemoglobinopathies are diseases caused by mutation in one of globin chains. For example, in *sickle-cell anemia* the gene defect is a mutation of a single nucleotide in gene of the β-chain of globin, which results in glutamic acid being substituted by valine at position 6. Hemoglobin with this mutation is called **hemoglobin S (HbS)** and is opposed to the normal adult HbA.

Thalassemias are decreasing or absence of normal globin subunits production, often through mutations in regulatory genes. *Beta-thalassemia* occurs when your body can't produce β-chains of globin, whereas *alpha-thalassemia* occurs when the body can't make α-chains of globin.

Chlorins and **bacteriochlorins** are other derivates of pyrrol. The most studied chlorin is **chlorophyll**. The basic structure of chlorophyll molecule is a modified porphyrin ring, co-ordinated to a central atom. This is very similar in structure to the heme group found in hemoglobin, except that in heme the central atom is iron, whereas in chlorophyll it is **magnesium** (Fig. 6.16).

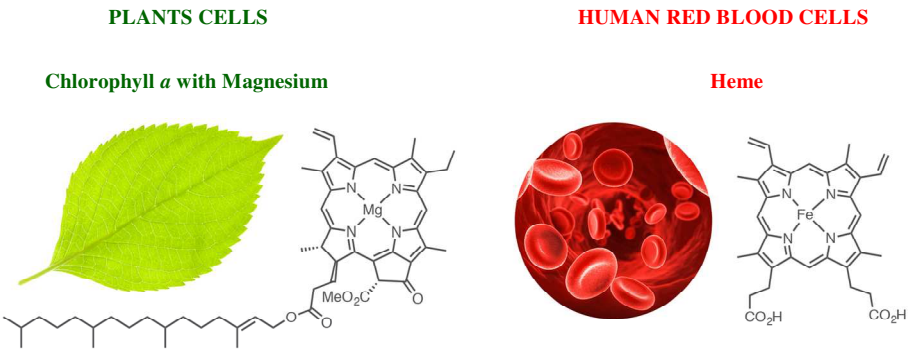


Fig. 6.16. Plant chlorophyll and human heme as the porphyrins

Chlorophyll is involved in photosynthesis, the process by which light energy is converted to chemical energy through the synthesis of organic compounds.

Vitamin B_{12} contains **a cobalt ion** at the centre of its molecule. The core of the molecule is a **corrin ring**, which is similar to the porphyrin ring found in heme and chlorophyll. Corrin ring has various attached side groups.

In structure of Vitamin B12 cobalt ion coordinates with four nitrogens of pyrrole rings, which all lie in one plane. Cobalt also bounds to nitrogen of 5,6-dimethylbenzimidazole group below the corrin ring as well as can reversibly bind different ligands above the corrin ring (Fig. 6.17).

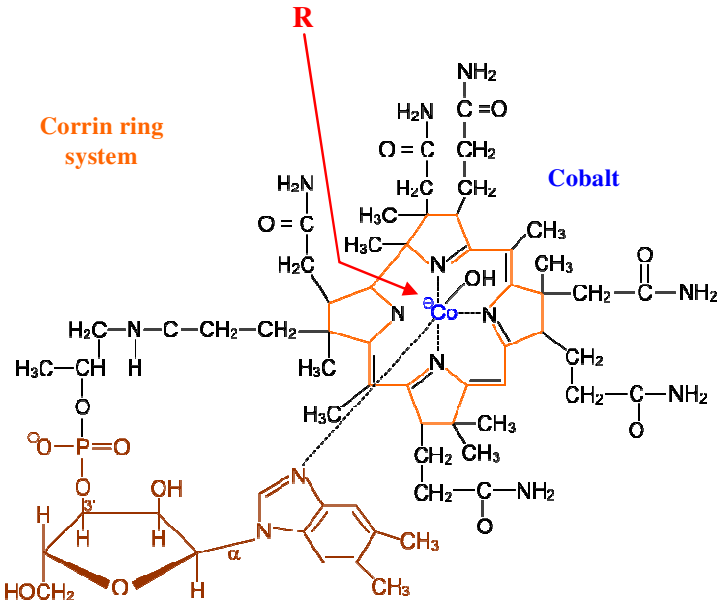


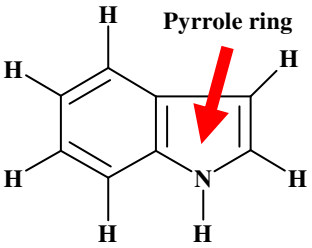
Fig. 6.17. Structure of Vitamin B12

These ligands can be:

- a methyl group (–CH_3) – in **methylcobalamin**
- a 5'-deoxyadenosine (here the C5′ atom of the deoxyribose forms the covalent bond with cobalt) – in **adenosylcobalamin**
- a cyanide group (–CN) – in **cyanocobalamin**
- hydroxyl group (–OH) – in **hydroxocobalamin.**

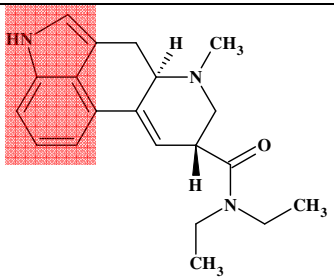
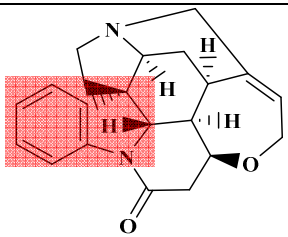
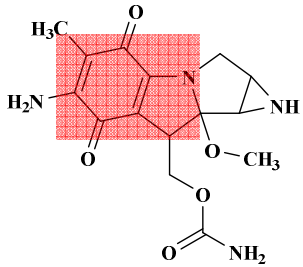
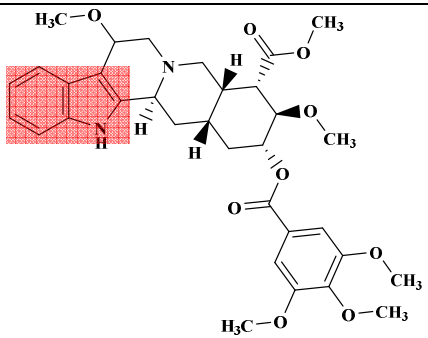
The other nitrogen of the 5,6-dimethylbenzimidazole is linked to a five-carbon sugar (ribose), which in turn connects to a phosphate group, and thence back onto the corrin ring via one of the seven amide groups attached to the periphery of the corrin ring.

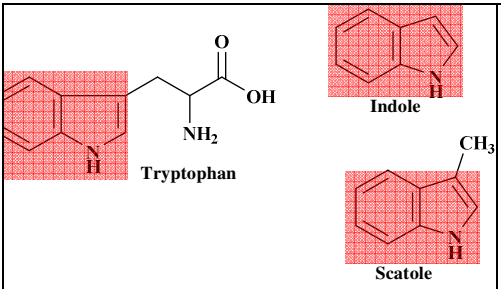
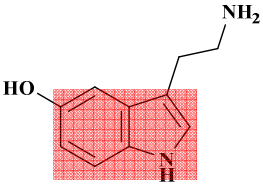
Indole and its derivatives. **Indole** is an aromatic heterocyclic organic compound. It has a bicyclic structure, consisting of a six-membered benzene ring fused to a five-membered nitrogen-containing pyrrole ring:



The most important indole derivatives are such compounds as tryptophan, skatole, strychnine, reserpine, mitomycin C and lysergic acid (LSD) (Tab. 6.4).

Table 6.4. The most important indole derivatives

	**Lysergic acid diethylamide (LSD) -** is a psychedelic drug. Its psychological effects may include altered awareness of the surroundings, perceptions, and feelings as well as sensations and images that seem real though they are not
	**Strychnine** – the neurotoxin which acts as an antagonist of glycine recep tors
	**Mitomycin C** is a chemotherapeutic agent by virtue of its antitumour activity
	**Reserpine** is the antipsychotic and antihypertensive drug

	**Tryptophan –** the amino acid **Indole** and its homologous **skatole** (3-methylindol) are endogenous toxins formed in the intestine when bacterial enzymes break down the amino acid tryptophan
	Another derivative of indole - **serotonin -** is also formed from tryptophan and is an important neurotransmitter in the central nervous system

6.3.3.2. Furan and thiophene derivatives

The most important derivatives of furan and thiophene as well as some information about their using are given on Fig. 6.18 and 6.19, respectively.

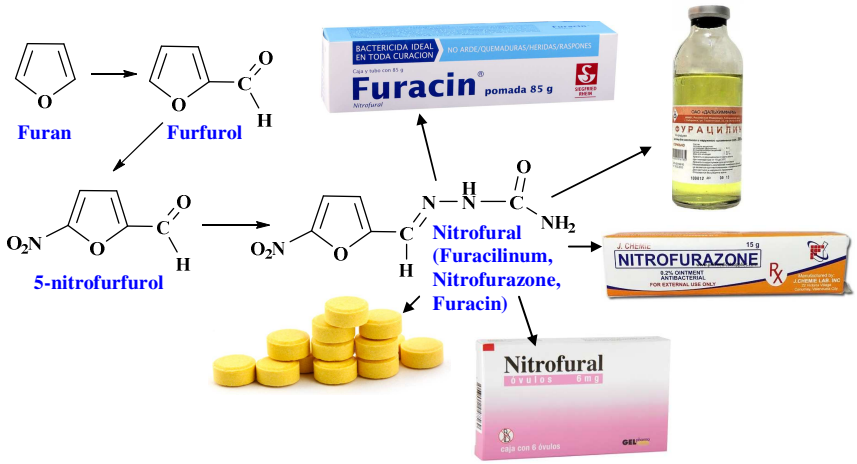


Fig. 6.18. The most important furan derivatives

Furfurol is used in the pharmaceutical industry for the synthesis of pharmaceuticals including a bactericidal compound **furacilin** (**nitrofural**).

Tetrahydrothiofene is a part of **Biotin (vitamin H).** In medicine, thiophene derivatives are used not extensively - in particular, they are contained in the preparate "**Ichthyol**", which is a complex mix of sulphuric shales.

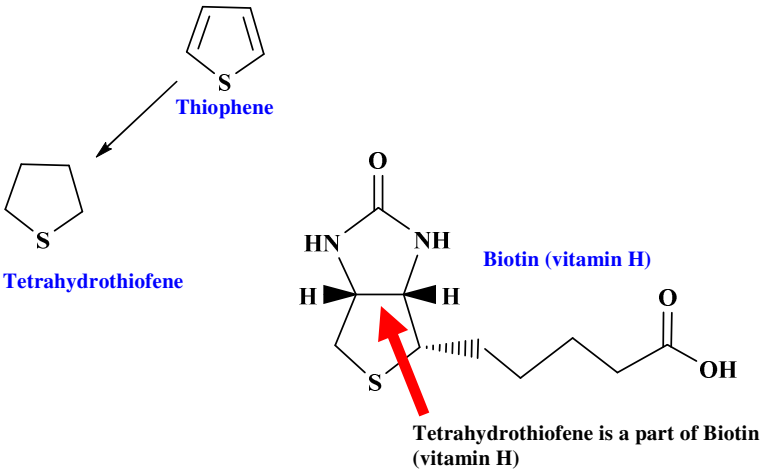


Fig. 6.19. The most important thiophene derivatives

6.4. Five-membered heterocyclic compounds with two or more heteroatoms

The main representatives of five-membered heterocyclic compounds with two or more heteroatoms are such diazoles as *pyrazole*, *imidazole* and *thiazole* as well as *triazoles* and *tetrazoles* (Fig. 6.20).

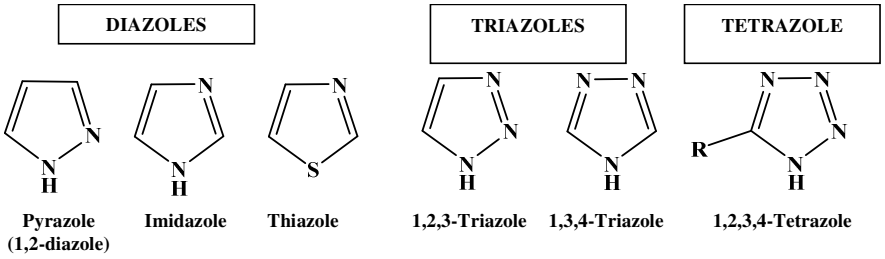


Fig. 6.20. The most important five-membered heterocyclic compounds

6.4.1. Pyrazole derivatives

These compounds don't occur in nature, but many drugs have been synthesized on the basis of pyrazole such as **antipyrine**, **aminopyrine** (or *aminophenazone*), **analgine** (or *metamizole*) and **phenylbutazone** (or *butadion*) (Fig. 6.21).

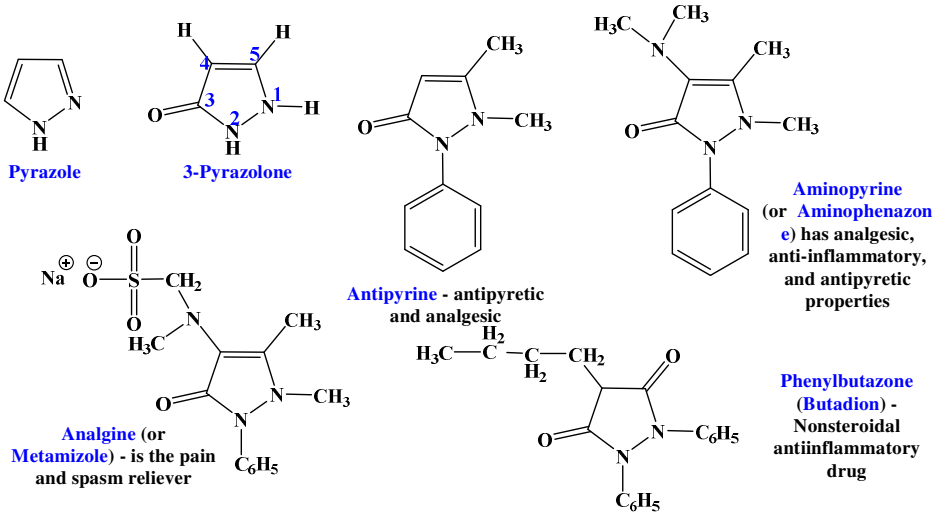


Fig. 6.21. The structures of some drugs-synthetic derivatives of pyrazole

6.4.2. Imidazole derivatives

Imidazole derivatives include some natural compounds (such as *histidine* – one of the essencial amino acids, biogenic amine *histamine*, formed via histidine decarboxylation, *carnosine* that is dipeptide, made up of the amino acids beta-alanine and histidine) and also some synthetic nitroimidazole antibiotics (*azomycin*, *metronidazole*, etc.) (Fig. 6.22, 6.23).

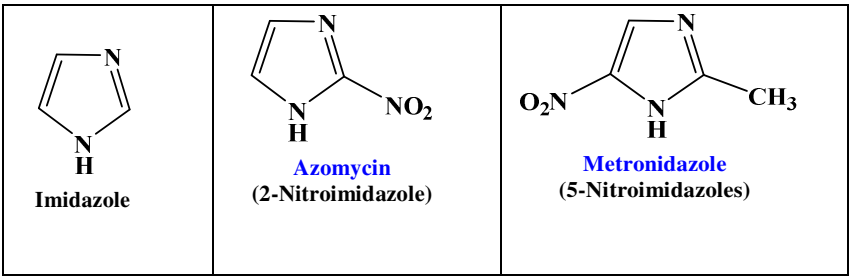


Fig. 6.22. The structures of some nitroimidazole antibiotics - synthetic derivatives of pyrazole

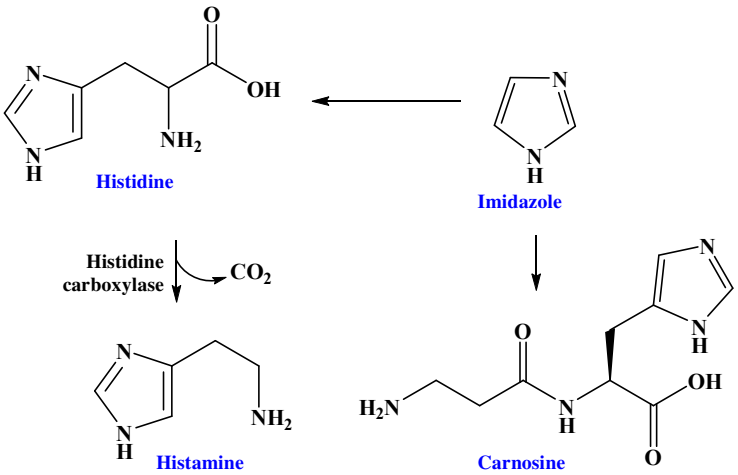


Fig. 6.23. The structures of some natural derivatives of pyrazole

Histamine is a transmitter of local immune response, a regulator of the gut physiological function as well as a neurotransmitter in the brain and spinal cord. **Carnosine** is a dipeptide (beta-alanyl-L-histidine), in high concentrations is found in the brain and muscles where it is needed for these tissues functions

6.4.3. Thiazole derivatives

The structure of thiazole is found in the composition of some natural substances such as **thiamine (vitamin B1)** and *sulfanilamide* medical product called **norsulfazol**. A saturated heterocycle, *thiazolidine*, is a fragment of the structure of natural and synthetic **penicillin antibiotics** (Fig. 6.24).

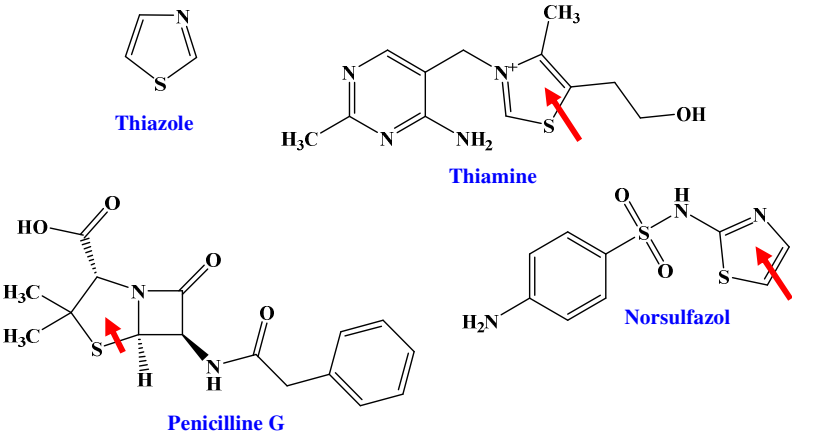
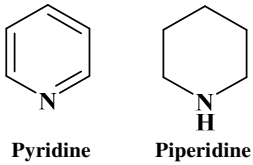


Fig. 6.24. The structures of some thiazole derivatives

6.5. Six-membered heterocyclic compounds with one heteroatom

Pyridine is the main six-membered heterocyclic compound. Other representative of this group – **piperidine** – is a saturated derivative of pyridine:

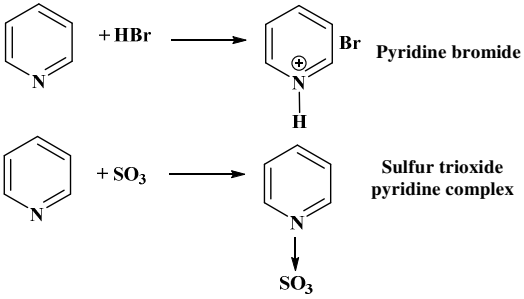


Pyridine **Piperidine**

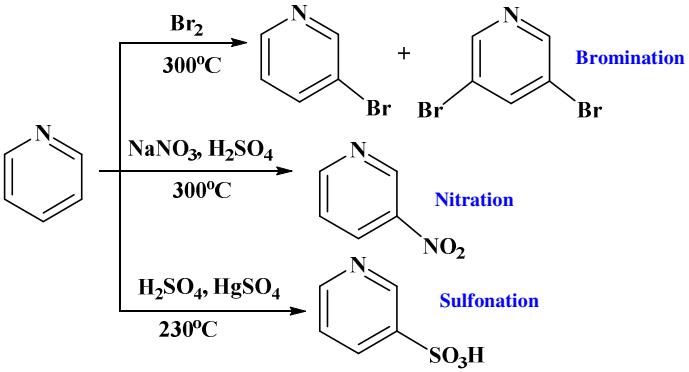
6.5.1. Chemical properties of pyridine

Typical pyridine reactions can be divided into three groups:

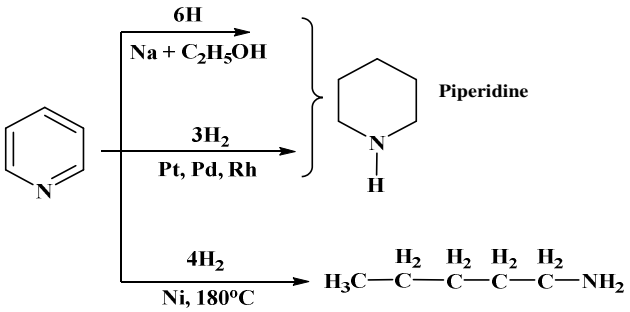
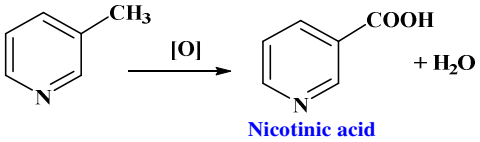
1) Reactions which followings with participation of heteroatom:



2) **Reactions of substituting for hydrogen atoms of pyridine ring -** the reactions of *nitration*, *sulphonation* and *halogenation*:



3) Reactions of reduction and oxidation:

Reduction	
Oxidation	

6.5.2. Pyridine and piperidine derivatives

The main pyridine derivatives are **vitamin PP** (**nicotinic acid** and **nicotinamide**, both of which exibit activity of this vitamin)**, vitamin B6 (pyridoxine), isoniazid** and **ftivazid** – antibiotics used for tuberculosis treatment, and **cordiamine** - stimulant which mainly affects the respiratory cycle and is used as a medical countermeasure against tranquilizer overdoses (Fig. 6.25). Other compounds such as *quinoline* and *isoquinoline*, as well as their derivatives – antibiotic *nitroxoline* and alkaloids *papaverine*, *morphine*, *codeine*, *quinine*, respectively – have pyridine ring as a part of their fused rings (Fig. 6.26).

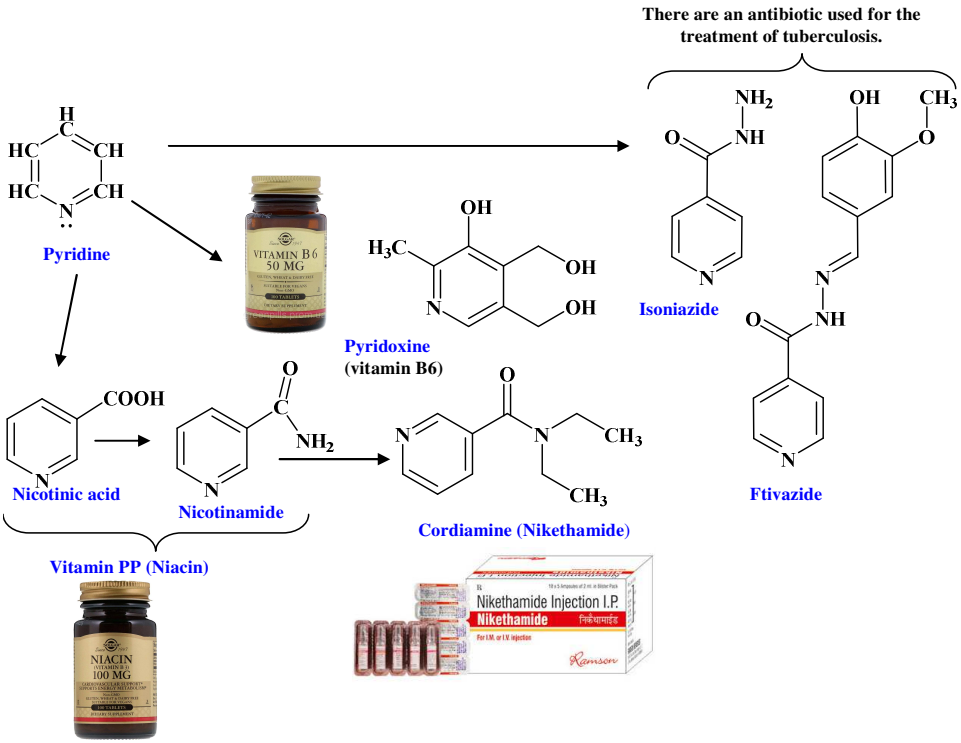


Fig. 6.25. The structures of some pyridine derivatives

Nitroxoline molecules are capable to bind the ions of some metals (Co^{2+}, Cu^{2+}, Bi^{3+}, Zn^{2+}, Fe^{2+}) that are necessary for the life of intestinal bacteria, into strong complexes and remove them from the body. Papaverine is an opium antispasmodic alkaloid used in the **treatment** of visceral spasm, vasospasm (especially those involving the intestines, heart, or brain).

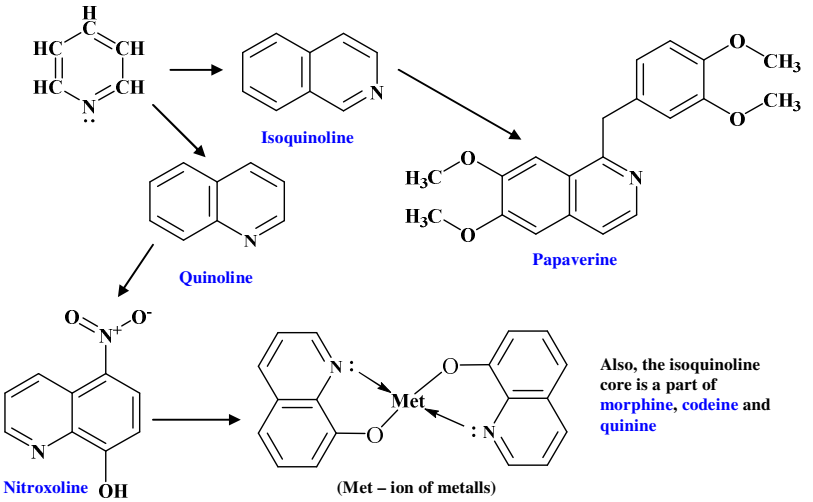
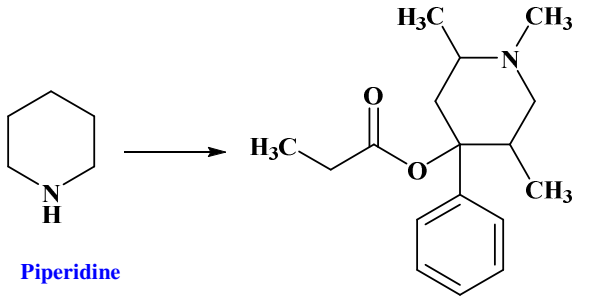


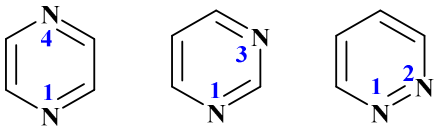
Fig. 6.26. The structures of some pyridine derivatives with fused rings

The famous piperidine derivative is opioid analgesic **trimeperidine** (or **promedol**):



6.6. Six-membered heterocycles with two heteroatoms

The main representatives of this group of heterocyclic compounds are *pyrazine*, *pyrimidine* and *pirydazine*:



These ring systems, particularly which of pyrimidine, occur commonly in natural products. The most important biological functions are possessed by such pyrimidine derivatives as *barbituric acid* as well as *pyrimidine and purine bases* (Fig. 6.27, 6.28).

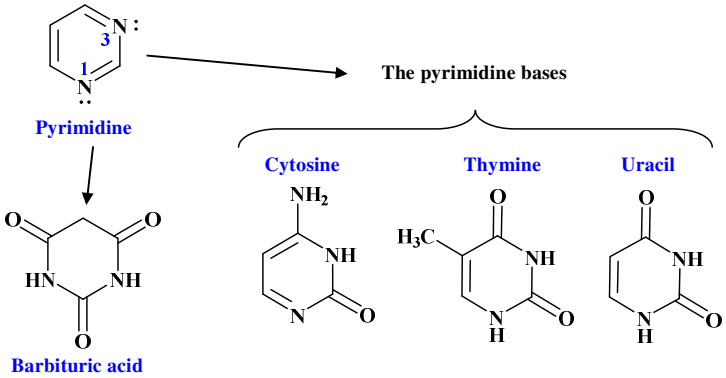


Fig. 6.27. The structures of pyrimidine bases and barbituric acid that are pyrimidine derivatives

Barbituric acid is the parent substance for the synthesis of *barbiturate drugs* – compounds that behave as central nervous

system depressants (act like Tranquilizers). The first barbiturate drug which started to use in medicine in 1903 was **barbital** (**Veronal**), and the second one, **Phenobarbital**, was first marketed in 1912.

The **pyrimidine bases cytosine, thymine**, and **uracil** are especially important because they in a couple with purine derivatives adenine and guanine are components of nucleic acids.

Purine is a heterocyclic aromatic organic compound that consists of a pyrimidine ring fused to an imidazole ring (Fig. 6.28). **Purines (adenine and guanine)** together with pyrimidine bases (cytosine, thymine and uracil) play crucial roles as parts of DNA and RNA. Purines are also significant components in a number of other important biomolecules, such as **ATP, GTP, cyclic AMP, NADH,** and **coenzyme A**.

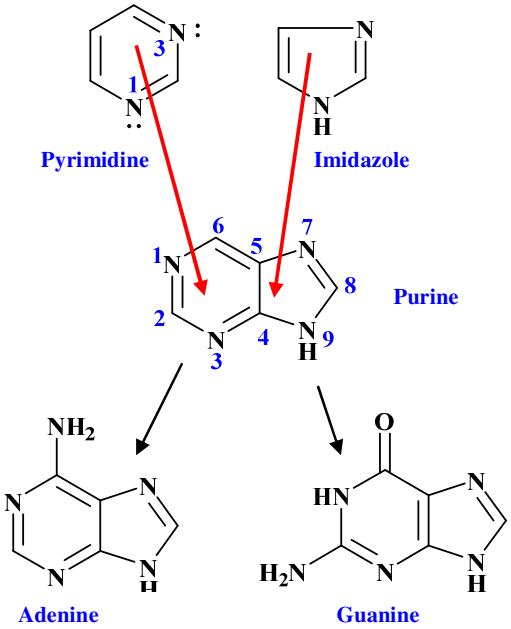


Fig. 6.28. The structures of purine bases that are pyrimidine and imidazole derivatives

Besides purine bases, there are also other derivatives of pyrimidine and imidazole – for example, *purine metabolites* (**hypoxanthine, xanthine** and **uric acid**) and **methylxanthines**. The latter are the central nervous system stimulants found in coffee bean (**caffeine**), in tea (**theophyline**) and cacao (**theobromine**) (Fig. 6.29).

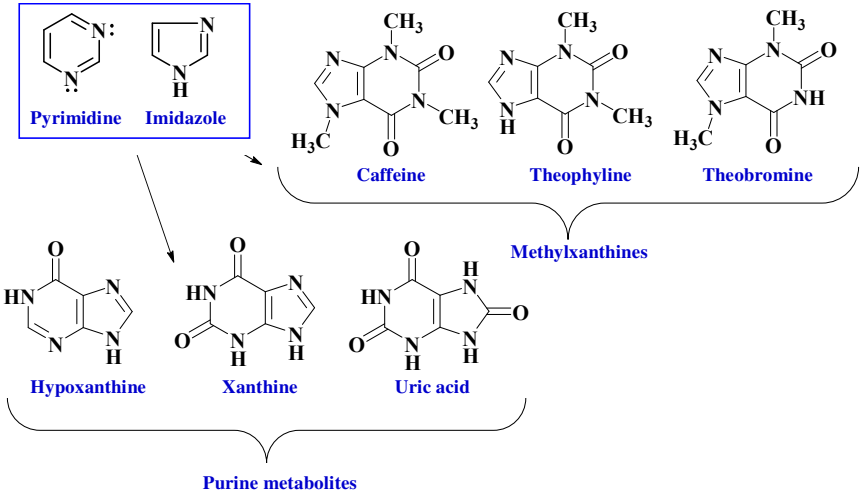


Fig. 6.29. The structures of purine metabolites and methylxanthines that are pyrimidine and imidazole derivatives

6.7. Test questions

1. Which type of hemoglobin consists of 2α and 2γ subunits?

A. HbA2;
B. HbAIC;
C. HbF;
D. Gower 1;
E. HbS

2. Hemoglobin with Fe^{3+} is called _____

A. oxyhemoglobin;
B. carbhemoglobin;
C. methemoglobin;
D. carboxyhemoglobin
E. fetal hemoglobin

3. A genetic disease that results in the production of an abnormal ratio of hemoglobin subunits is called _____

A. methemoglobinemia
B. hemoglobinopatia
C. talassemia
D. erythrocytosis
E. folate-dependent anemia

4. Which of these atoms can never act as a hetero atom?

A. Sulfur
B. Nitrogen
C. Oxygen
D. Carbon
E. Cuprum

5. Pyridine belongs to:

A. Three-membered heterocyclic compounds
B. Four-membered heterocyclic compounds
C. Five-membered heterocyclic compounds
D. Six-membered heterocyclic compounds
E. Seven-membered heterocyclic compounds

6. Pyrolle belongs to:

A. Three-membered heterocyclic compounds
B. Four-membered heterocyclic compounds
C. Five-membered heterocyclic compounds
D. Six-membered heterocyclic compounds
E. Seven-membered heterocyclic compounds

7. Which of these heterocycles is a fused rings-containing heterocyclic compounds?

A. Pyrolle
B. Pyridine
C. Furan
D. Indol
E. Tiophene

8. Porphobilinogen, porphin, porphyrinogens and porphyrins are derivatives of:

A. Indole
B. Purine
C. Pyrolle
D. Pyridine
E. Pyrimidine

9. Vitamin B12 is derivative of:

A. Indole
B. Purine
C. Pyrolle
D. Pyridine
E. Pyrimidine

10. Nicotinic acid is derivative of:

A. Indole
B. Purine
C. Pyrolle
D. Pyridine
E. Pyrimidine

PART 7. NITROGEN-CONTAINING BASES, NUCLEOSIDES, NUCLEOTIDES AND NUCLEIC ACIDS

7.1. Nitrogen-containing bases, nucleosides and nucleotides

Five nitrogen-containing bases are nucleotide components. Three of them are derivatives of pyrimidine, a monocyclic base with a six-membered ring, and two are derivatives of purine, a bicyclic base with fused five- and six-membered rings (Fig. 7.1).

Pyrimidines are represented by:

- cytosine (in DNA, RNA);
- uracil (in RNA);
- thymine (in DNA).

Purines are:

- adenine (in DNA, RNA);
- guanine (in DNA, RNA).

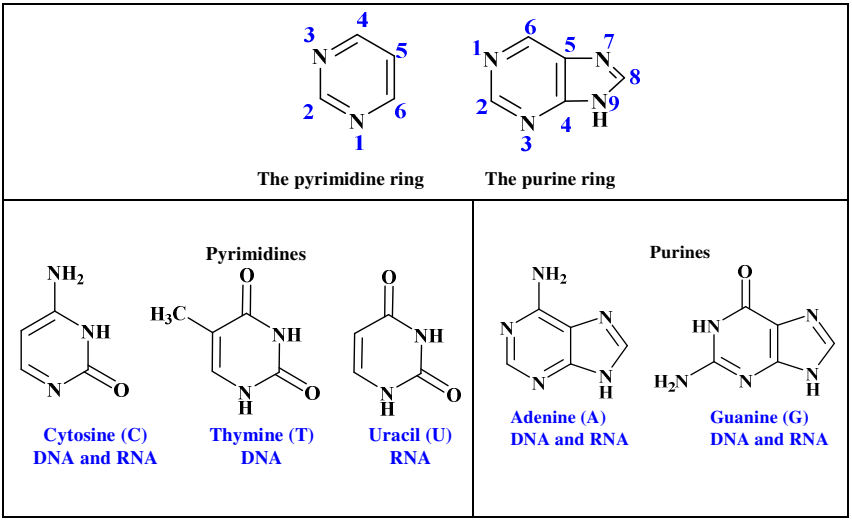


Fig. 7.1. Nitrogen-containing purine and pyrimidine bases

Nucleosides are compounds formed when bases are covalently linked to the 1' position of a pentose sugar ring via a glycosidic bond (Fig. 7.2).

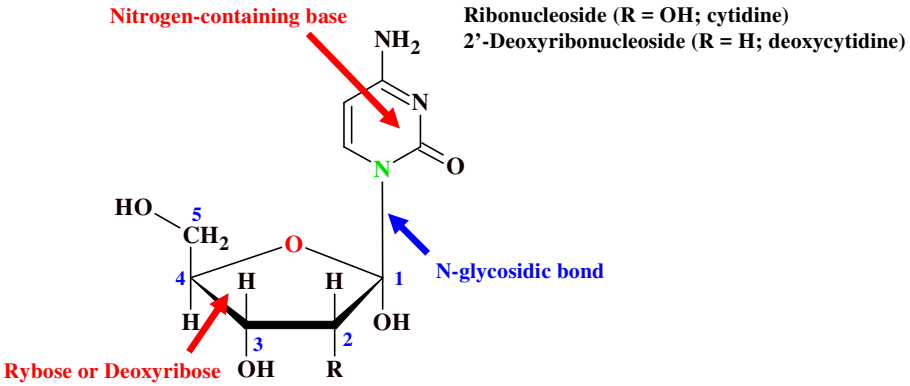
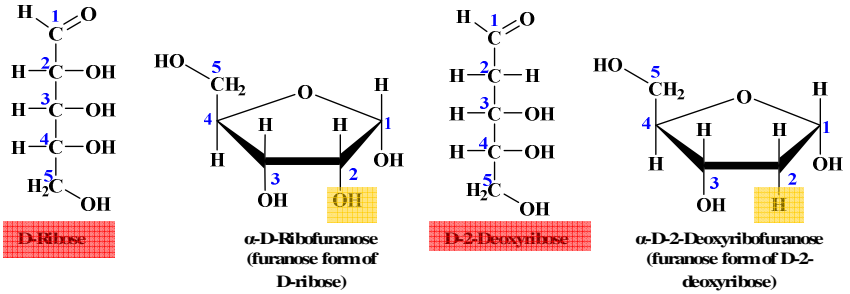


Fig. 7.2. Nucleosides structure

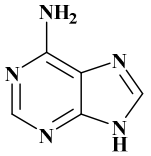
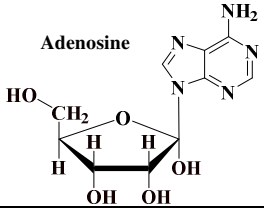
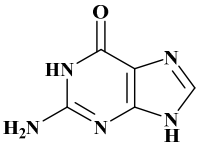
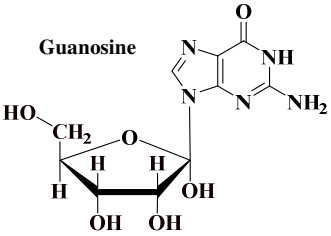
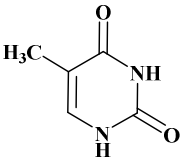
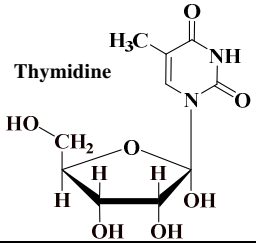
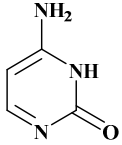
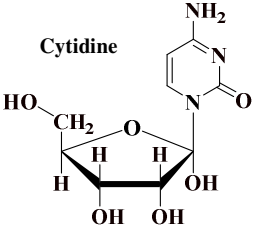
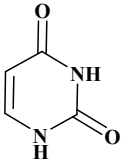
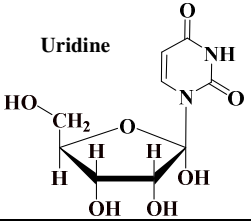
Nucleosides names are formed by adding *-idine* to the root name of a pyrimidine or *-osine* to the root name of a purine (Tab. 7.1).

The sugar unit of a nucleoside is either the pentose ribose or the 2-deoxyribose:



Nucleosides that have ribose called **ribonucleosides**. Nucleosides that have 2-deoxyribose called **deoxyribonucleosides**.

Table 7.1. Nitrogen-containing bases and nucleosides names

Nitrgenous base	Nucleoside
 **Adenine**	 **Adenosine**
 **Guanine**	 **Guanosine**
 **Thymine**	 **Thymidine**
 **Cytosine**	 **Cytidine**
 **Uracil**	 **Uridine**

Nucleotides are the nucleoside phosphates that consist of:

- nitrogen-containing base
- pentose (ribose or deoxyribose)
- one or more phosphate groups (Fig. 7.3)

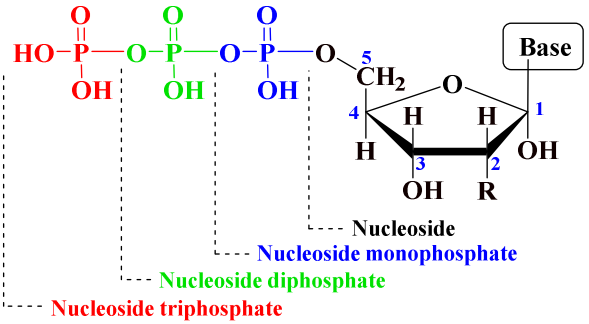


Fig. 7.3. Structures of nucleoside triphosphate, diphosphate and monophosphate

Nucleotides that have deoxyribonucleosides called *deoxyribonucleotides*. Nucleotides that have ribonucleosides called *ribonucleotides* (Fig. 7.4). Most nucleotides are ribonucleotides.

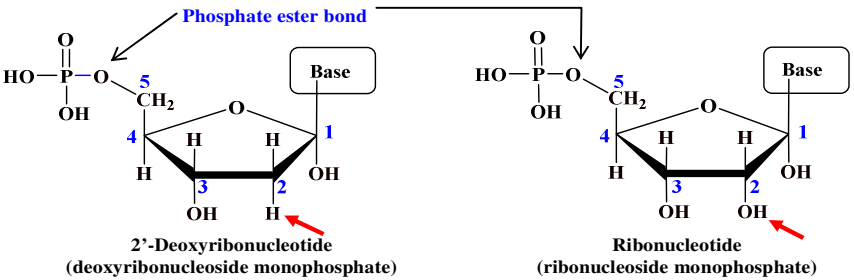


Fig. 7.4. Nucleotides structure

Nucleotides names are given on Tab. 7.2.

Table 7.2.
Nitrogen-containing bases, nucleosides and nucleotides names

BASES	NUCLEOSIDES	NUCLEOTIDES
Adenine (A)	Adenosine	Adenosine 5'-triphosphate (ATP)
	Deoxyadenosine	Deoxyadenosine 5'-triphosphate (dATP)
Guanine (G)	Guanosine	Guanosine 5'-triphosphate (GTP)
	Deoxyguanosine	Deoxyguanosine 5'-triphosphate (dGTP)
Cytosine (C)	Cytidine	Cytidine 5'-triphosphate (CTP)
	Deoxycytidine	Deoxycytidine 5'-triphosphate (dCTP)
Uracil (U)	Uridine	Uridine 5'-triphosphate (UTP)
Thymine (T)	Thymidine/ Deoxythymidie	Thymidine/deoxythymidine 5'-triphosphate (dTTP)

Nucleotides have next functions in our body:

- *Nucleoside 5'-triphosphates* are carriers of energy.
- *Cyclic nucleotides* are signal molecules and regulators of cellular metabolism and reproduction (Fig. 7.5).

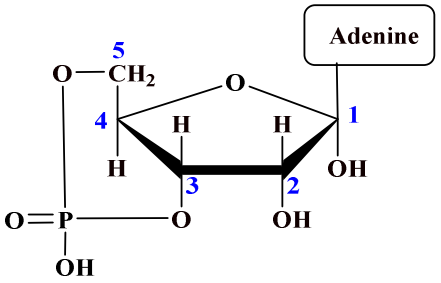


Fig. 7.5. Structure of cyclic adenosine 3',5'-monophosphate (cAMP)

- ATP is central to energy metabolism.

- GTP drives protein synthesis.
- CTP drives lipid synthesis.
- UTP drives carbohydrate metabolism.
- NADP, NAD, FMN, FAD are coenzymes of redox enzymes

Nucleic acids are the polymers of ribonucleotides (RNA) or deoxyribonucleotides (DNA). *3',5'-phosphodiester bond* link nucleotides together to form polynucleotide chains.

There are several important differences between RNA and DNA (Fig. 7.6):

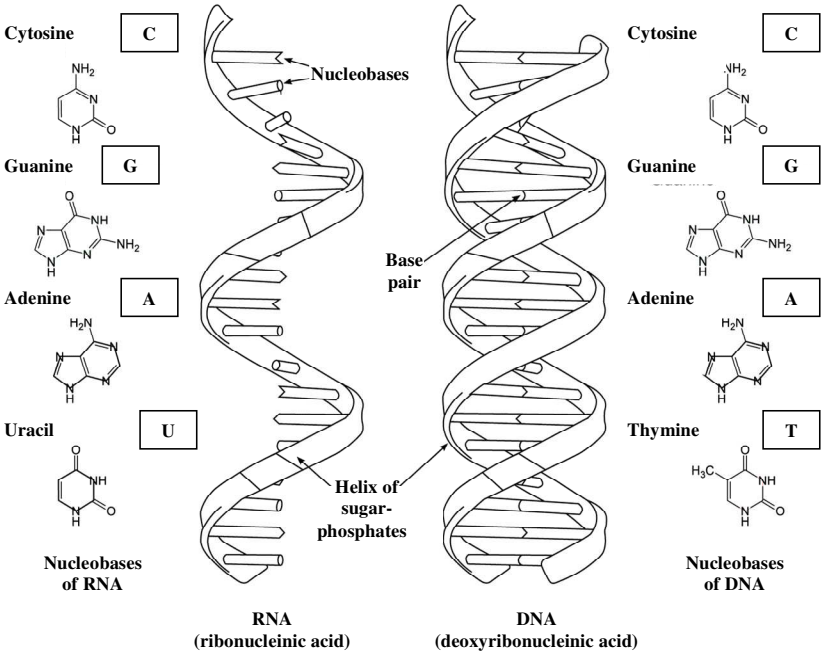


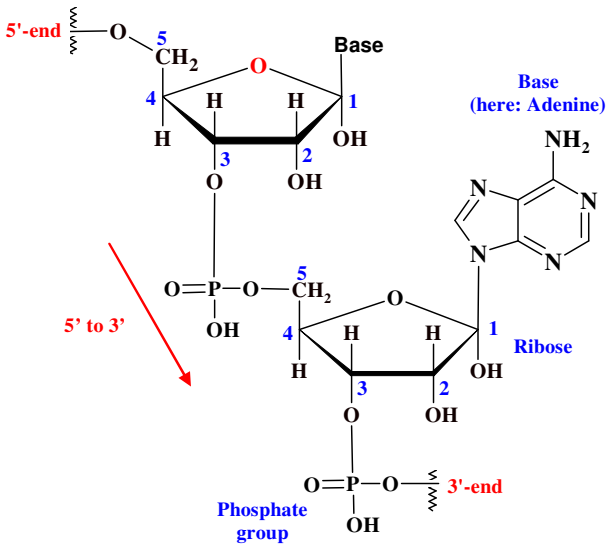
Fig. 7.6. Structure of cyclic adenosine 3',5'-monophosphate (cAMP)

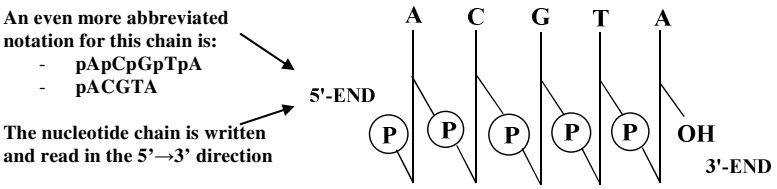
- the pentose sugar in RNA is ribose, in DNA it's deoxyribose;
- in RNA, uracil replaces the base thymine (U pairs with A) – therefore, DNA has A, G, C, T, whereas RNA – A, G, C, U;
- RNA is single stranded while DNA is double stranded;
- RNA molecules are much smaller than DNA molecules.

7.2. Deoxyribonucleinic acid

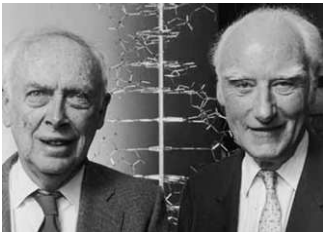
7.2.1. The primary structure of Nucleic Acid

The primary structure of any nucleic acids is the nucleotide sequence in polydeoxynucleotide chain. *3',5'-phosphodiester bond* link nucleotides together to form polynucleotide chains:

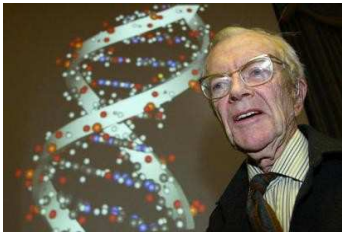




James Watson, Francis Crick and Maurice Wilkins in 1962 took the Nobel Prize in Physiology and Medicine for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material:



James Watson



Maurice Wilkins

Francis Crick

7.2.2. Secondary structure of DNA - double helix structure

It was discovered by Watson and Crick at 1953. **The secondary structure is defined as the relative spatial position of all the atoms of nucleotide residues.** DNA has two separate strands (is *double strande*, *dsDNA*). These strands are antiparallel (5'→3'direction) and have complementary sequences (Fig. 7.7).

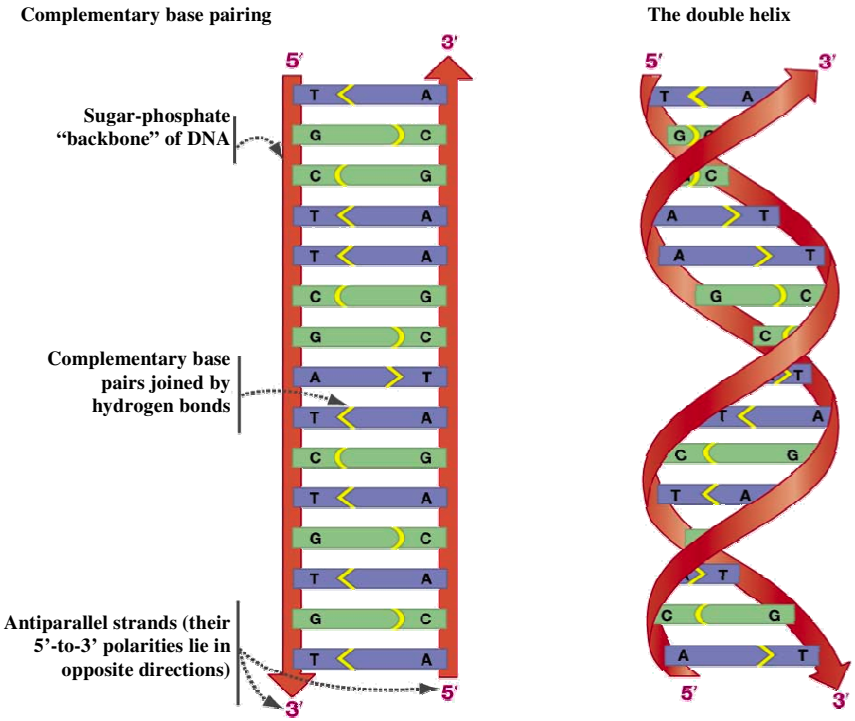


Fig. 7.7. The secondary structure of DNA

In DNA double helix, the two strands of DNA are held together by *hydrogen bonds*. These hydrogen bonds are formed between the complementary nitrogen bases: A and T (two bonds); G and C (three bonds) (Fig. 7.8). Stable configuration also is maintained by *base stacking force (hydrophobic interaction).* Sugar-phosphate backbones (negatively charged) line outside whereas base pairs (stack one above the other) localized inside. **Chargaff's rules:** A=T, G=C, Purines = Pyrimidines.

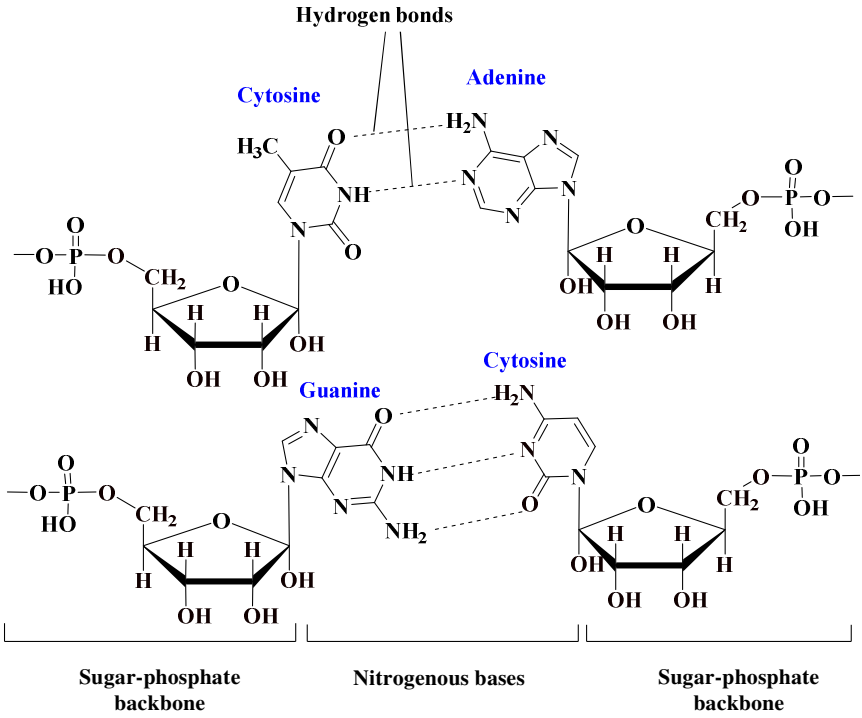
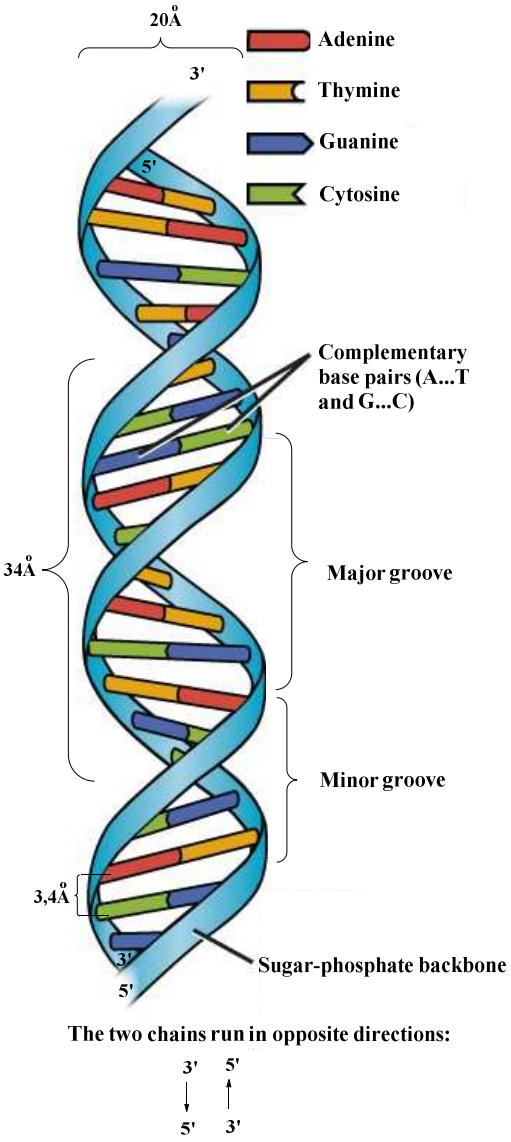


Fig. 7.8. The hydrogen bonds formation and sugar-phosphate backbone and base pairs localization in the secondary structure of DNA

There are next parameters of DNA helix:

- it is the antiparallel right-handed double helix;
- **its diameter** is about 2 nm;
- **the distance between two base pairs:** 0.34 nm (Fig. 7.9);
- each **turn of the helix** involves 10 bases pairs, 3.4 nm.
- it has **two groves** (Fig. 7.10): minor and major ones.



.

Fig. 7.9. Some parameters of DNA helix

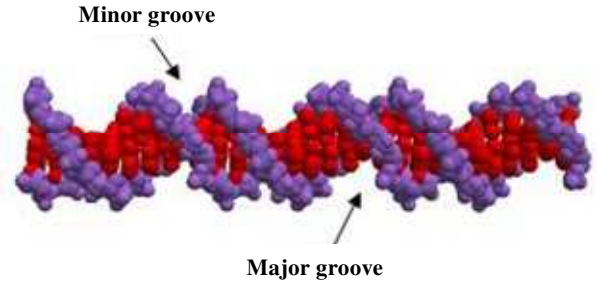


Fig. 7.10. Minor and major groove of DNA helix

Small molecules like drugs bind in the minor groove, whereas proteins like transcription factors can interact with the major grooves.

7.2.3. Packing of Prokaryotic and eukaryotic DNA

The length of *E. coli* DNA is about 1.6 million nm, whereas *E. coli* cell is only 2000 nm long. So **DNA is compact and folded forming the circular structures *in prokaryotic cells*** (Fig. 7.11).

The linear ***eukaryotic DNA*** appears in a highly ordered form called **chromosomes** during metaphase, whereas shows a relatively loose form of **chromatin** in other phases.

The basic unit of chromatin is **nucleosome**. Nucleosome is composed of **DNA** and very basic proteins, rich in arginine and lysine, called **histones** - H1, H2A, H2B, H3 and H4. Two molecules of each histones H2A, H2B, H3 and H4 form **histone core** (Fig. 7.12).

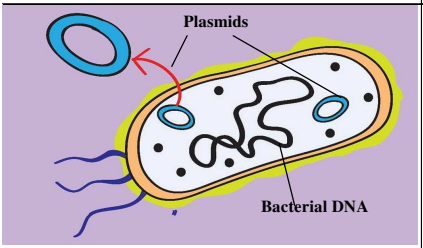
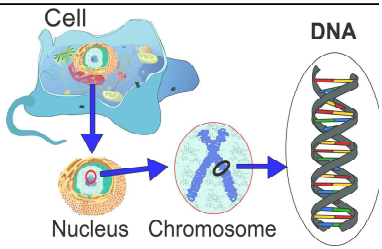
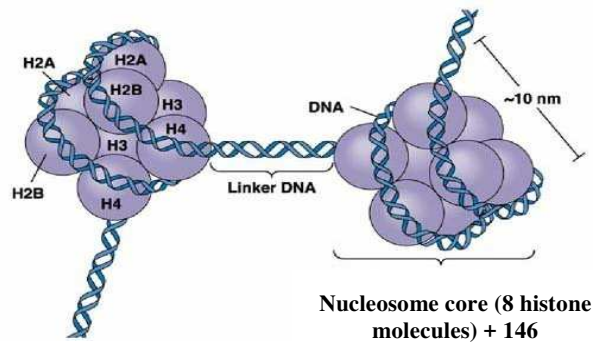
PROCARIOTS	EUCARIOTS
	
– A circular DNA molecule – Naked – no associated proteins – Plasmids often present – One chromosome only	– Contains a linear DNA molecule – Assosiated with histon proteins – No plasmids – Two ore more different chromosoms

Fig. 7.11. The particularities of prokaryotic and eukaryotic DNA packing



Nucleosome core (8 histone molecules) + 146 nucleotide pairs of DNA

Fig. 7.12. The nucleosome structure

DNA chain (about 146 pairs of nucleotides long) makes a 1,75 turns around histone core. In addition, histone H1 associated with **linker** or **binder DNA** (about 50 bp) is also involved in the

nucleosome chain organization. Such association of DNA with histones forms a structure of effectively compacting DNA

Nucleosomal chain is the first level of chromatin organization; the others of them are given on Fig. 7.13. The packing of DNA into chromosomes is very important because chromosome is a compact form of the DNA that protects DNA from damage. DNA in a chromosome can be transmitted efficiently to both daughter cells during cell division. Moreover, changes in chromosome organization regulate the gene expression.

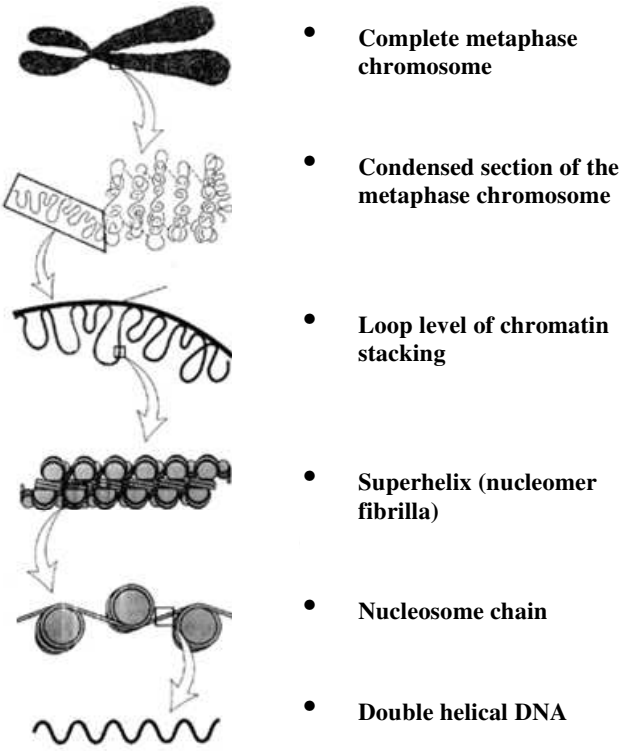


Fig. 7.13. The levels of chromatin organization

7.3. Ribonucleinic acid

There are next RNA types:

- **messenger RNA (mRNA)** is the carrier of genetic information about translated protein from DNA to ribosomes;
- **transfer RNA (tRNA)** transports amino acids to ribosomes to synthesize protein;
- **ribosomal RNA (rRNA)** are the components of ribosomes;
- **heterogeneous nuclear RNA (hnRNA)** is mainly the precursors of mRNA (pre-mRNA) that converts into mature mRNA via *processing* and *splicing*;
- **small nuclear RNA (snRNA)** take parts in pre-mRNA splicing;
- **small nucleolar RNA (snoRNA)** play a key role in the processing of rRNA molecules;
- **small cytoplasmic RNA (scRNA)** regulates the targeting of *de novo* synthesized proteins to subcellular compartments;
- **catalytic RNA** or **ribozyme -** RNA molecules that are capable of catalyzing specific biochemical reactions, similar to the action of protein enzymes;
- **small interfering RNA (siRNA)** – is a class of double-stranded RNA molecules, 20-25 pairs of nucleotide in length that interfere with the expression of a specific gene and degrade mRNAs.

RNA molecules are largely single stranded but they can have double stranded regions.

7.3.1. Messenger RNA (mRNA)

It comprises only about 5% of the RNA in the cell. Eukaryotic mRNA has 2 typical parts in its structure (Fig. 7.14):

- **cap** – linkage of *7-methylguanosine* to the 5'-terminal residue.

- **poly A tail** – attachment of an adenylate polymer (poly A, 20~250 nucleotides) at the 3' terminal to enhance stability of mRNA

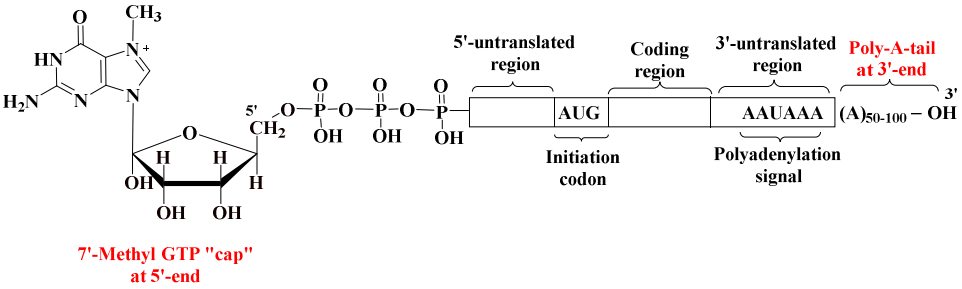


Fig. 7.14. The structure of mRNA

Pre-mRNA processing. mRNA is formed as pre-mRNA (hnRNA) (Fig. 7.15).

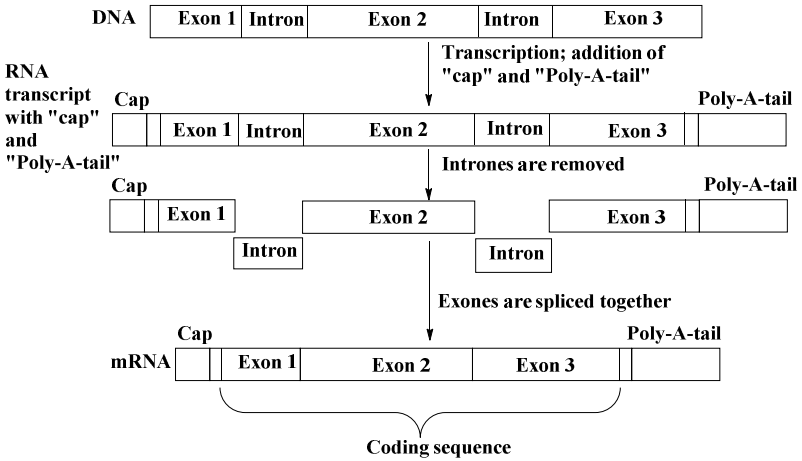


Fig. 7.15. Processing and splicing as the mechanisms of hnRNA transformation into mature mRNA

hnRNA *doesn't include the cap and Poly-A-tail* as well as *has introns* (intervening sequences) *and exones* (coding sequences). Transformation of pre-mRNA into mature mRNA is called *processing*.

In processing Poly(A)-tail is added to the 3' end of the pre-mRNA and a 5' cap is added to the 5' end of it. Also the removing of all introns from the pre-mRNA occurs – this process is known as *splicing*

7.3.2. Transfer RNA (tRNA)

These RNA make up 15% of the RNA in the cell. There are at least 20 types of tRNA in one cell **–** each amino acid has at least one unique tRNA which carries it to the ribosome. Aminoacyl tRNA molecules are the substrates of protein synthesis.

The primary structure of tRNA. tRNA has 74~95 bases and is the smallest of three major RNA. It has modified bases:

- *pseudouridine (ψ);*
- methylguanosine;
- dihydrouridine (D);

It also has *the 3'-terminal sequence* that is always *CCA.*

The secondary structure of tRNA has the form of **cloverleaf** (Fig. 7.16) with:

- **Acceptor stem** with 3'-terminal CCA;
- **Four loops:**
  - **D loop**
  - **Anticodon loop**
  - **TψC loop**
  - **Variable loop**

The paired regions of rRNA are called **the "stems".**

Acceptor stem with 3'-terminal CCA binds amino acid (forming CCA-amino acid). The amino acid loaded onto the tRNA by *aminoacyl tRNA synthetases* to form *aminoacyl-tRNA*, is covalently bonded to the 3'-hydroxyl group on the CCA tail. D loop acts as a recognition site for aminoacyl-tRNA synthetase, an enzyme involved in the aminoacylation of the tRNA molecule.

Anticodon loop recognizes amino acid codon on the mRNA. TψC loop is a specialized region on the tRNA molecule which acts as a special recognition site for ribosome to form a tRNA-ribosome complex during protein biosynthesis or translation. Variable loop plays a role in aminoacyl tRNA synthetases recognition of tRNA.

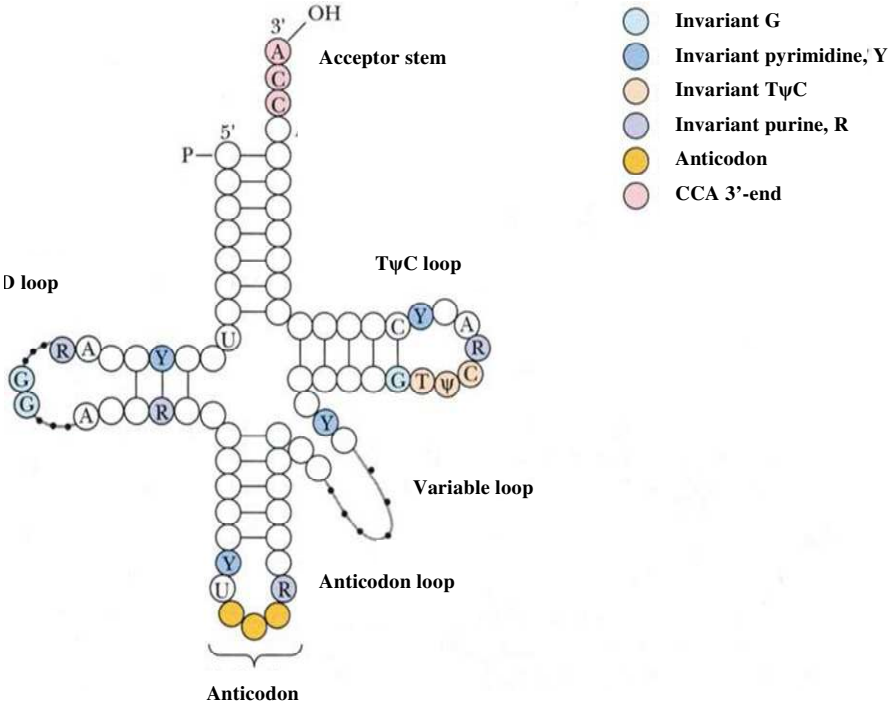
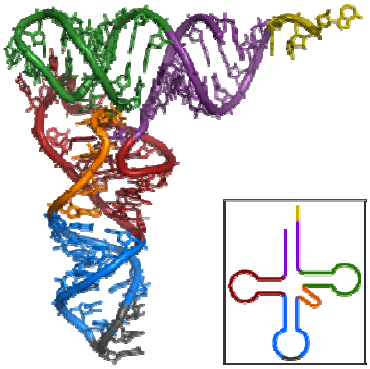


Fig. 7.16. The secondary structure of tRNA: the cloverleaf

Tertiary structure of tRNA. All tRNAs have a similar **L-shaped 3D-structure** that allows them to fit into the P- and A-sites of the ribosome:



Pre-tRNA processing. In all organisms, tRNAs are transcribed in a pre-tRNA. In bacteria, multiple tRNAs are often transcribed as a single RNA. The first step in their processing is the release individual pre-tRNAs from this RNA. In eukaryotes, each pre-tRNA is transcribed as a separate transcript. Then the pre-tRNA processing occurs that has five steps (Fig. 7.17):

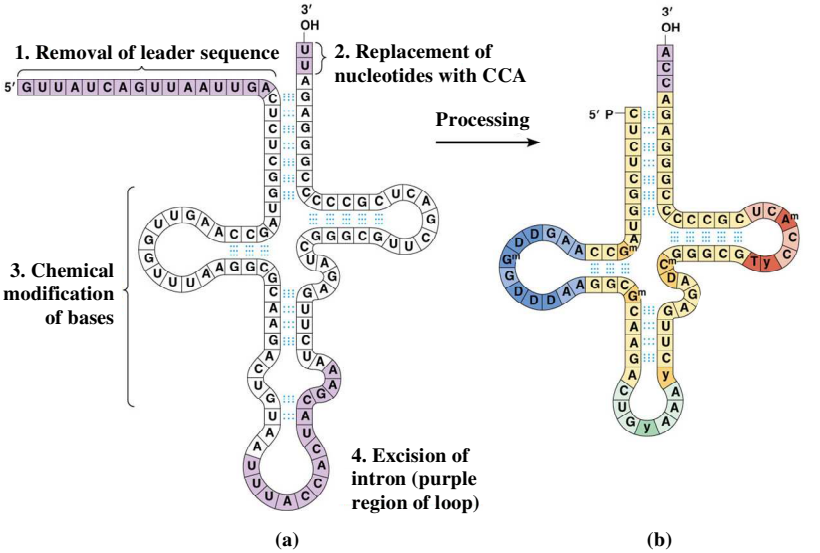


Fig. 7.17. pre-tRNA processing: a – primary transcript (precursor) for yeast tyrosine tRNA; b – mature tRNA, secondary structure

1. The 5′ end of the pre-tRNA, called the *5′ leader sequence*, is cleaved off.

2. The 3′ end of the pre-tRNA is cleaved off. In all eukaryote pre-tRNAs, but in only some bacterial pre-tRNAs, a CCA sequence of nucleotides is added to the 3′ end of the pre-tRNA. Some bacteria pre-tRNAs already have the CCA encoded in their transcript.

3. Chemically modification of nucleotides in pre-tRNA: conversion of adenine (A) to pseudouridine (ψ), adenine to inosine (I), and uridine to dihydrouridine (D).

4. Some eukaryotic pre-tRNAs have introns that have to be spliced out.

7.3.3. Ribosomal RNA (rRNA)

RNA of this type makes up 80% of the RNA in the cell. **Ribosomes** have about 2/3 RNA and about 1/3 protein. rRNA serves as a scaffold for ribosomal proteins. rRNA has a characteristic secondary structure due to many intramolecular H-bonds.

rRNA types of pro- and eukaryotes are given on Tab. 7.3.

Table 7.3. rRNA types of pro- and eukaryotes

Eukaryotic rRNAs	Prokaryotic rRNAs
5S* rRNA	5S rRNA
28S rRNA	23S rRNA
18S rRNA	16S rRNA
5,8S rRNA	

* - S represents Svedberg units, they represent measures of sedimentation rate that correlates with M.m. of RNA

Pre-rRNA processing. Four rRNAs in eukaryotes are at first transcribed as two long precursor molecules: the first of them contains the pre-rRNA that will be processed into the 5S rRNA, and the second is the precursor of the 28S, 5,8S and 18S rRNAs. Enzymes then cleave these precursors into subunits corresponding to each rRNA (Fig. 7.18).

In bacteria, there are only three rRNAs, and all of them are transcribed as a part of one long precursor molecule that is cleaved into the individual rRNAs. Some of the bases of pre-rRNAs then are methylated for added stability.

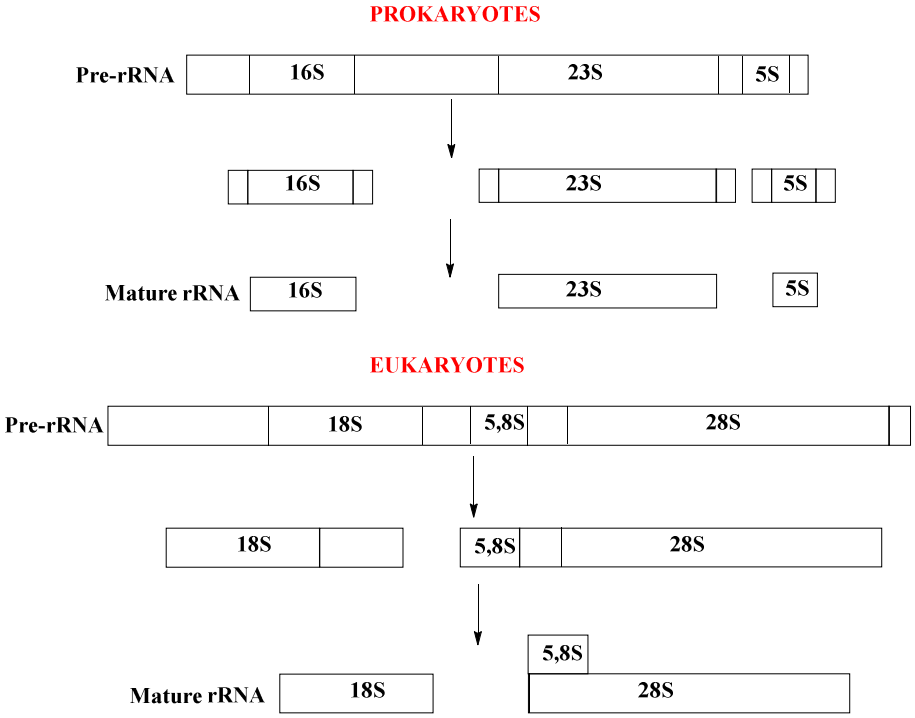


Fig. 7.18. pre-rRNA processing in eukaryotes and prokaryotes.

7.4. The general properties of DNA and RNA

Nucleic acids – both DNA and RNA - are amphiphilic normally acidic molecules (the latter - because of phosphate). They are

insoluble in organic solvents, but can be precipitated by ethanol. These substances own optical absorption – they are able to specific UV absorption at 260 nm due to aromatic groups. This property can be used to identify nucleic acid. They also have thermal stability, but curtain temperature can cause disassociation of double-stranded DNA (dsDNA) into two single-stranded DNAs (ssDNAs).

7.4.1. Denaturation, renaturation and hybridization of DNA

DNA may be the subject to denaturation, renaturation and hybridization.

In **denaturation** hydrogen bonds are broken, 3-D structure is destroyed, the double helix turns into single strand irregular coil. As a result, the value of 260 nm absorption is increased by 30%-40% (**hypochromic effect**) and biological functions of DNA are lost (Fig. 7.19).

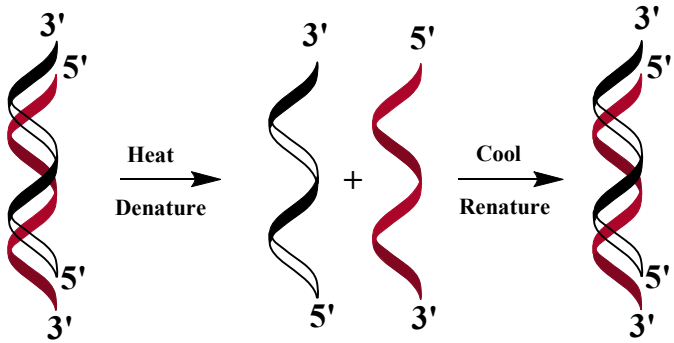


Fig. 7.18. Heat denaturation and renaturation of DNA

Hypochromic effect is due to the fact that stacked bases in a double helix absorb less ultra-violet light.

Renaturation of DNA occurs, when the denatured DNA solution is slowly cooling down (at Annealing) – at such conditions the single strand DNA can reform a double strands helix to recover its biological functions.

At lowing down denaturing temperature ssDNAs from different sources (or ssDNA and RNA) which has the complementary bases will be repaired into double strands with forming of hybrid DNA or DNA-RNA. This process is called **molecule hybridization**.

7.4.2. Heat denaturation and melting temperature of DNA (T_m)

When DNA is heated to certain temperature, the absorption value at 260 nm is increased sharply which indicates that the helix of dsDNA is separated into ssDNAs. **T_m (melting temperature of DNA)** is the temperature at which 50% of DNA in a sample has denatured from double-stranded DNA (dsDNA) to single-stranded DNA (ssDNA) (Fig. 7.19).

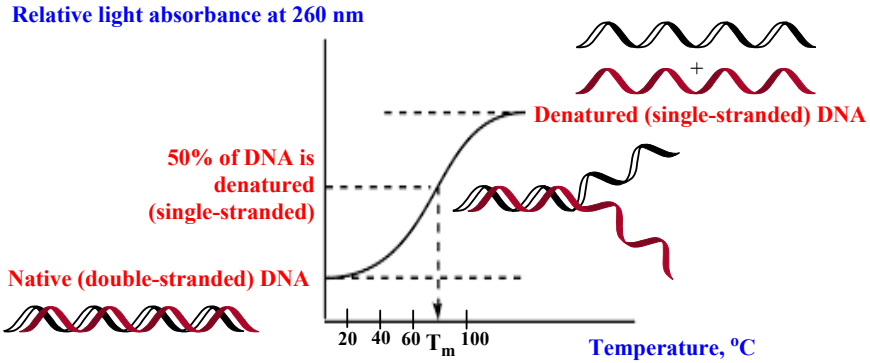


Fig. 7.19. The determination of DNA melting temperature

The main factor that affects T_m is G-C content. There are three hydrogen bonds between G-C pair and two between A and T. So a double-stranded DNA rich in G and C needs more energy to be broken than one that is rich in A and T, meaning higher melting temperature(Tm). Therefore, the more G-C content, the higher T*m* value (Fig. 7.20).

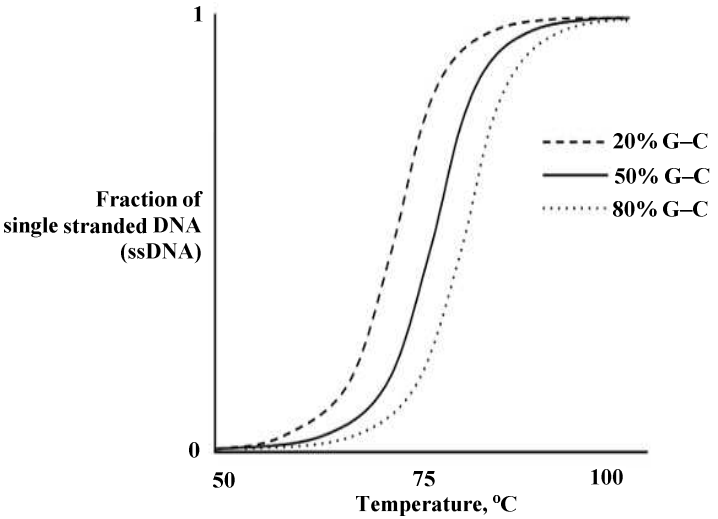


Fig. 7.20. The influence of G-C content on DNA melting temperature

7.5. Test questions

1. Purines are derivatives of:

A. Indole
B. Thiophene
C. Pyrolle
D. Pyridine
E. Pyrimidine

2. Compounds formed when nitrogenous bases are covalently linked to the 1' position of a pentose sugar ring via a glycosidic bond are called:

A. Purines
B. Pyrimidines
C. Nucleosides

D. Nucleotides
E. Nucleic acids

3. Adenosine and cytidine are:

A. Purines
B. Pyrimidines
C. Nucleosides
D. Nucleotides
E. Nucleic acids

4. Compounds that consist of: nitrogen-containing base, pentose (ribose or deoxyribose) and one or more phosphate groups

A. Purines
B. Pyrimidines
C. Nucleosides
D. Nucleotides
E. Nucleic acids

5. Which type of linkage binds nucleotides together to form one polynucleotide chain?

A. Glycosidic bond
B. Peptide bond
C. Hydrogen bond
D. 3',5' phosphodiester bond
E. electrostatic interactions

6. Which type of linkage stabilizes the primary structure of any nucleic acid?

A. Glycosidic bond
B. Peptide bond
C. Hydrogen bond
D. 3',5' phosphodiester bond
E. electrostatic interactions

7. The secondary structure of DNA is called:

A. α-helix

B. β-sheat
C. double helix
D. cloverleaf
E. collagen helix

8. The basic unit of chromatin composed of DNA and very basic proteins histones is called:

A. Nucleotide
B. Nucleoside
C. Nucleosome
D. Nucleoporine
E. Chromosome

9. "Cap" and "poly-A-tail" are typical in:

A. mRNA
B. tRNA
C. rRNA
D. DNA
E. hnRNA

10. Which types of nucleic bases are rich in modified bases such as pseudouridine, methylguanosine and dihydrouridine?

A. mRNA
B. tRNA
C. rRNA
D. DNA
E. hnRNA

RECOMMENDED LITERATURE AND SOURCES

1. Biologically important classes of bioorganic compounds. Biopolymers and their structural components: Theoretical course of biological and bioorganic chemistry, Module 1 / A.O. Syrovaya, E.R. Grabovetskaya, N.M. Tkachuk et al. – Kharkov: KhNMU. – 2013.– 183 p.
2. Campbell Neil A., Reece Jane B., Simon Eric Jeffrey Essential Biology. Pearson/Benjamin Cummings: 2006. – 462 p
3. Carey Francis A. Advanced Organic Chemistry / Francis A. Carey, Richard A. Sundberg // Paperback, 5th Edition. – 2007. – 1199 p.
4. CHAPTER 3 STRUCTURES AND FUNCTIONS OF NUCLEIC ACIDS **https://fdocuments.in/document/chapter-3-structures-and-functions-of-nucleic-acids.html**
5. Culav Elizabeth M, Clark C Heather, Merrilees Mervyn J. Connective Tissues: Matrix Composition and Its Relevance to Physical Therapy // *Physical Therapy*, 1999. 79(3): 308–319.
6. Fatty Acids in Foods and their Health Implications,Third Edition (Food Science and Technology) / edited by Ching Kuang Chow. CRC Press Taylor @ Francis group: 2008. - 1298 p
7. Frederick A. Bettelheim, William H. Brown, Mary K. Campbell, Shawn O. Farrell. Introduction to General, Organic and Biochemistry. Cengage Learning: 2009. – 864 p
8. Harper's Illustrated Biochemistry 31st Edition / Victor Rodwell, David Bender, Kathleen Botham, Peter Kennelly, P. Anthony Weil. McGraw-Hill Education / Medical: 2018. – 800 p
9. HETERO CYCLIC COMPOUNDS https://fdocuments.in/document/hetero-cyclic-compounds.html

10. Heterocyclic Compounds https://basicschemistry.blogspot.com/2013/07/heterocyclic-compounds.html
11. Hydroxy- and oxoacids. Heterofunctional compounds of benzene series. Metabolites and parent structures of medicines: methodical instructions for 1st year students' self-work in Biological and Bioorganic Chemistry (module 1) / compiled by A.O. Syrovaya, L.G. Shapoval, V.N. Petiunina et al. – Kharkiv: KhNMU, 2014. – 25 p.
12. Kantharaj G. R. Molecular Biology for Masters http://mol-biol4masters.masters.grkraj.org/html/Co_and_Post_Translational_Events4-Glycosylation_of_Proteins.htm
13. Kevin A. Boudreaux Fundamentals of Organic Chemistry. Organic and Biochemistry for Today. - 29 pgs https://www.angelo.edu/faculty/kboudrea/index_2353/Chapter_08_2SPP.pdf
14. Lehninger Principles of Biochemistry David L. Nelson, Albert L. Lehninger, Michael M. Cox W. H. Freeman: 2008 1158 p
15. Quin Louis D. Fundamentals of Heterocyclic Chemistry: Importance in Nature and in the Synthesis of Pharmaceuticals / Louis D. Quin, John Tyrell // Hardcover, 1st Edition. – 2010. – 327 p.
16. Sarraf, K., & Kader, D. (2017). Knee oral core topics. In P. Banaszkiewicz & D. Kader (Eds.), Postgraduate Orthopaedics: The Candidate's Guide to the FRCS (Tr & Orth) Examination (pp. 292-338). Cambridge: Cambridge University Press. doi:10.1017/9781316091685.020
17. Smith Michael B. March's Advanced Organic Chemistry. Reactions, mechanisms, and structure / Michael B. Smith, Jerry March // Hardcover, 6th Edition. – 2007. – 2384 p.
18. Synelnyk T.B. METHODICAL POINTING from the course "BIOLOGICAL MEMBRANES AND REGULATION OF

METABOLISM" (part 1. Biological membranes structure and membrane transport) Kyiv-2013. 17 pgs https://biology.univ.kiev.ua/images/stories/Kafedry/Biochimiya/Biblioteka/Synelnyk_BIOLOGICAL_MEMBRANES_AND_REGULATION_OF_METABOLISM_eng.pdf
19. Virtual Textbook of Organic Chemistry https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/intro1.htm#contnt
20. Zurabyan S.E. Fundamentals of bioorganic chemistry / S.E. Zurabyan – M.: Geotar-Med. – 2003. – 320 p.