# STRUCTURE AND FUNCTION OF BIO-MOLECULES

## Table of contents

# 1. INTRODUCTION

The molecules that form the building blocks of living organisms obey the same laws of nature as all other "chemical molecules". However, such molecules are different in a sense that they have a **function**. They can be seen as highly efficient "tools" and/or "machines" or as dedicated "building materials". Certain biomolecules even have the ability to replicate and repair themselves. As far as we know there are no biomolecules without function. However, in a number of cases this function is not known. Many biological molecules form complex and highly ordered structures. This order is maintained using energy from the surrounding. The chemistry of such molecules is nevertheless an important basis for understanding how biomolecules can fulfill their diverse functions. All biological phenomena have a molecular and therefore a chemical basis (Table 1.1).

**Table 1.1** Distribution of the most important elements

| Earth (crust) | | Human body | |
|---|---|---|---|
| O | 47.0 % | H | 63.0 % |
| Si | 28.0 % | O | 25.5 % |
| Al | 7.9 % | C | 9.5 % |
| Fe | 4.5 % | N | 1.4 5 |
| Ca | 3.5 % | Ca | 0.31 % |
| Na | 2.5 % | P | 0.22 % |
| K | 2.5 % | Cl | 0.08 5 |
| Mg | 2.2 % | K | 0.06 5 |
| Ti | 0.46 % | S | 0.05 % |
| H | 0.22 % | Na | 0.03 % |
| C | 0.19 % | Mg | 0.01 % |

Major elements of the organic materials: O, C, N, H, P, S

- Elements that form stable covalent bonds
- Elements from the middle of the PSE ➨ medium strength electronegativity

little tendency to form ions

- Elements from the top of the groups ➡ double bonds possible

light atoms, i.e. strong bonds

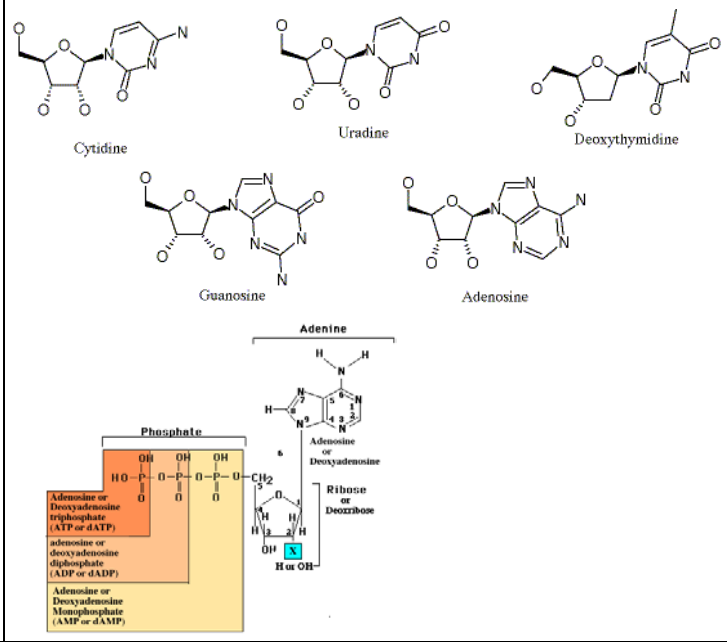
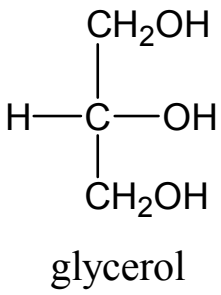
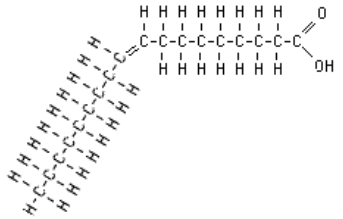
*Carbon atoms can link to each other and form 3D structures*

Ions: $Na^+$, $K^+$, $Mg^{2+}$, $Ca^{2+}$, $Cl^-$

Trace elements: Mn, Fe, Co, Cu, Zn, B, Al, V, Mo, I, Si, Sn, Ni, Cr, F, Se

The number of biological molecules is extremely large. However, the number of building blocks used to make these molecules is surprisingly small (Table 1.2). The molecules tend to have similar function in all living organisms. Especially the higher organisms tend to use building blocks derived from food to build their own macromolecules (metabolism).

**Table 1.2** Examples of building blocks used in biological molecules.

| | | |
|---|---|---|
| Components of the nucleic acids |  | |
| Some components of lipids |  glycerol |  Oleic acid, a monounsaturated fatty acid. Note that the double bond is *cis*; this is the common natural configuration. |

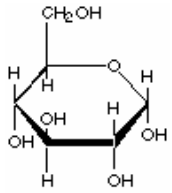
| | |
|---|---|
| A monosaccharide |  <br>α –Glucose |

**Table 1.3** Typical Biopolymers

| **Monomers** | **Polymers** |
|---|---|
| **AMINO ACIDS** (Lysine) | Proteins<br>Peptide Hormones<br>Neurotransmitters<br>Toxic alkaloids |
| **NUCLEOTIDES** (Adenin) | Nucleic Acids<br>ATP<br>Coenzymes<br>Neurotransmitters |
| **FATTY ACIDS** (Palmitic acid) | Membrane lipids<br>Fats<br>Waxes |
| **SUGARS** (Glucose) | Cellulose<br>Starch<br>Fructose<br>Mannose<br>Sucrose<br>Lactose |

The building blocks are used to form typical biopolymers such as proteins (amino acids), polysaccharides (monosaccharides), DNA/RNA (mononucleotides), and lipids (molecular aggregates) (Table 1.3). The function of these biopolymers tends to be the same in all living organisms. In addition, the basic building blocks can be modified to fulfill other functions. For example, 20 amino acids are used to build the proteins. Besides those more than 150 amino acids are known, which fulfill other functions in the living organism. Most of these are chemical descendents of the 20 basic ones. Hybrid-biopolymers are also known. Examples are the glycoproteins, which contain either one or several polypeptide chains (amino acids) and one or several glycosilation structures (saccharides).

*Tasks / functions of biopolymers*

| | |
|---|---|
| Polynucleotides | Storage and transfer of genetic information |
| Proteins | Realization of the genetic information, catalytic and transport functions, building material, etc. (many others) |
| Polysaccharides | Storage of energy, structures (building material for the cell walls, etc.) |
| Lipids | Storage of energy, structures (building material for cell membranes, etc.) |

Only polynucleotides and proteins store information in their structure.

# 2. PROTEINS

To a large extent, cells are made of protein, which constitutes more than half of their dry weight (Table 2.1.). Proteins are the most versatile class of molecules in living organisms. Amongst their functions are: catalysis (enzymes), transport, storage (casein), contraction (muscles), protection (antibodies), attack (toxins), hormones (insulin, growth hormone) and structure (collagen). All proteins contain C, H, N, O some S, P, Fe, Zn, Cu. Acidic hydrolysis yields 20 different α-amino acids (L-form), which are encoded by the **genetic code** and which constitute the building blocks of the proteins in all living organisms. In addition, approximately 150 amino acids are known, which are not encoded by the genetic code and which are sometimes posttranslatorial modifications of the above mentioned 20 α-amino acids. However, βγ and δ-amino acids are also known as well as some D-type amino acids. Typical posttranslatorial modifications of amino acids in proteins include: acetylation of the N-terminus (reduces degradation), hydroxylation of proline residues transformation of glutamate into γ-carboxyglutamate or the addition of sugar residues and lipid molecules to modify the final hydrophobicity of the molecule. Phosphorylation and methylation of specific amino acid residues have a signaling effect in several metabolic pathways. These modifications generally are reversible. Many proteins are also cleaved and trimmed after synthesis (trypsinogen → trypsin).

**Table 2.1.** Approximate Chemical Compositions of a Typical Bacterium and a Typical Mammalian Cell

| Percent of Total Cell Weight | | |
|---|---|---|
| Component | *E. coli* Bacterium | *Mammalian Cell* |
| $H_2O$ | 70 | 70 |
| Inorganic ions ($Na^+$, $K^+$, $Mg^{2+}$, $Ca^{2+}$, $Cl^-$, etc.) | 1 | 1 |
| Miscellaneous small metabolites | 3 | 3 |
| Proteins | 15 | 18 |
| RNA | 6 | 1.1 |
| DNA | 1 | 0.25 |
| Phospholipids | 2 | 3 |
| Other lipids | - | 2 |
| Polysaccharides | 2 | 2 |
| Total cell volume: | $2 \times 10^{-12}$ cm $^3$ | $4 \times 10^{-9}$ cm $^3$ |
| Relative cell volume: | 1 | 2000 |

Proteins, polysaccharides, DNA, and RNA are macromolecules. Lipids are not generally classed as macromolecules even though they share some of their features; for example, most are synthesized as linear polymers of a smaller molecule (the acetyl group on acetyl CoA), and they self-assemble into larger structures (membranes). Note that water and protein comprise most of the mass of both mammalian and bacterial cells.

## *Amino Acids*

**Amino acids** are bi-functional compounds containing both a carboxylic acid group (-COOH) and a basic amino group ($-NH_2$) attached to the same carbon atom (Fig. 2.1). They are building blocks of proteins. They are linked together by a peptide bond (see later).

**Fig. 2.1** Amino Acid Structure

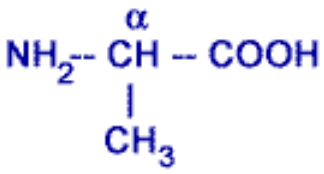
Each protein species contains one or several polypeptide chains of defined amino acid sequence. In addition both the size (5'000 – 1'000'000 g/mol) and **conformation** are defined. While the amino acid sequence is determined by the genetic code, the 3D-structure of a protein can at present not be predicted. The structures of some of the more important and more accessible proteins have been determined, e.g. by X-ray diffraction studies. While simple proteins contain only polypeptide chains, complex proteins (**lipoproteines, phophoproteines, metalloproteines, glycoproteines**) contain also non-peptide structures (**prosthetic groups**). Proteins can be **soluble** (globular proteins: antibodies, hormones, transport proteins like serum albumin) or **insoluble** (structure proteins: myosin, fibrinogen). Proteins can assemble to **supramolecular complexes** (for example multienzyme complexes such as fatty acid synthetase). The biological activity of a protein depends on its three-dimensional structure. This structure (**native structure**) is under physiological conditions rather stable, however, exposing a protein to "extreme" conditions (high temperatures low/high pH) can **denature** the molecule (loss in solubility / biological activity). Denaturing can be reversible in some cases, but irreversible in others.

### Types of Amino Acids

- According to **structure**

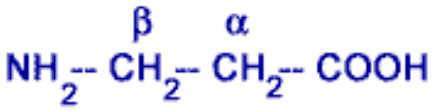
According to molecular structure, amino acids can be divided into three classes:

Amino acids having ($-NH_2$) group attached to the alpha carbon atom are called $\alpha$- amino acids.



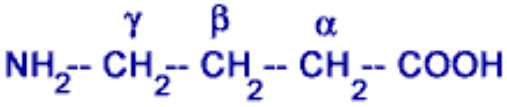
α – Amino acid

Amino acid having (-$NH_2$) group attached to the β- carbon atom aware called β- amino acids.



β – Amino acid

Amino acids having (-$NH_2$) group attached to the γ-carbon atom are called γ-amino acids.



γ – Amino acid

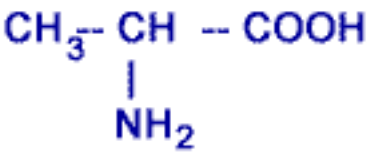
- According to **nature**

According to nature there are three classes of amino acids

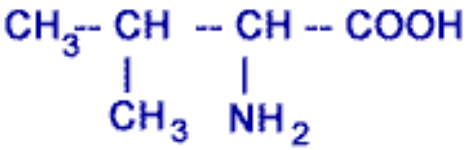
1. NEUTRAL AMINO ACIDS

Amino acids containing one acid group (-COOH) and one basic amino group(-$NH_2$) are known as neutral amino acids.

Formulas



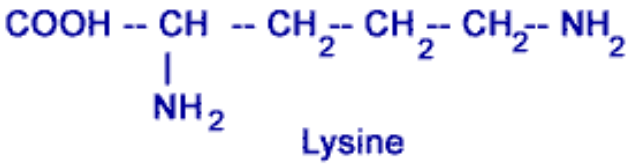
Alanine



Valine

2. BASIC AMINO ACIDS

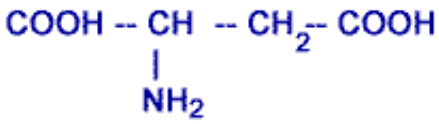
Amino acids containing a one acidic group (-COOH) and more than one basic amino group (-$NH_2$) are known as basic amino acids.

Formula 


Lysine

3. ACIDIC AMINO ACIDS

Amino aces containing one basic group (-NH2) and moor than one acidic group (-COOH) are known as acidic amino acids.

Formula 


Aspartic acid

Except for the simplest amino acid, i.e. glycine, all amino acids have at least one chiral C-atom and therefore two optical isomers (threonine and isoleucine have, e.g., two optical C-atoms). The basis for the nomenclature is D/L-glycerine aldehyde. *Protein forming amino acids are all of the L-type*. D-type amino acids are also known in nature but do not occur in proteins. In the case of amino acids with two asymmetric carbons, the naturally in proteins occurring form is called the L-form, its mirroring image the D-form. The other two possible stereoisomers are called *diastereomeric forms* (allo forms). A *meso form* has two mirroring asymmetric carbons. The D/L-character of the amino acids has little consequence in most chemical reactions. However, **the reaction with a second substance, which itself also contains asymmetrical C-atoms be differ considerably**.

Properties of Amino Acid Side Chains (R-groups) (Fig 2.2)

**Charged Residues** are seldom buried in the interior of a folded protein. They are normally found on the surface of the protein where they interact with water and with other important biological molecules. Note that these groups can be important in the recognition (binding) of oppositely charged groups on molecules that interact with proteins.

**Polar Residues** are both buried as well as on the surface of the protein. They either form hydrogen bonds with other polar residues in the protein or with water. For example, the OH group of Serine can both donate as well as accept a hydrogen bond:

**Non-polar residues** do not interact favorably with water. The central core of most proteins is composed almost exclusively of non-polar residues, stabilized by numerous van der Waals interactions. However, a significant number of non-polar residues are also found on the surface of the protein.

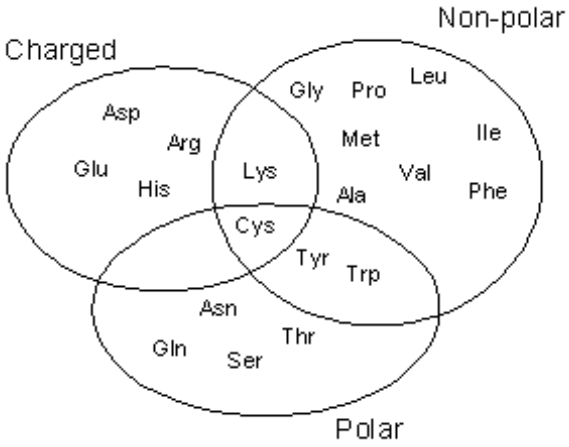



**Fig 2.2** Properties of Amino Acid Side Chains (R-groups)

Amino acids are grouped by the chemical properties of the side chain (Fig. 2.4). The same amino acid can fall into multiple groups (Table 2.2).

**size** - for example, affecting how well the side chain fits in a binding site

**hydrophobicity**- for example, determining whether the side chain will be buried in a protein interior or membrane as opposed to interacting with the solvent (water)

**H-bonding potential** - for example, providing geometrically specific interactions or chemical catalysis

**charge** - for example, providing for long-range electrostatic interactions

**polarity** - providing water solubility

**chemical structure** - i.e., aliphatic, aromatic, sulfur-containing, aliphatic hydroxyl-containing, basic, acidic, amide derivative

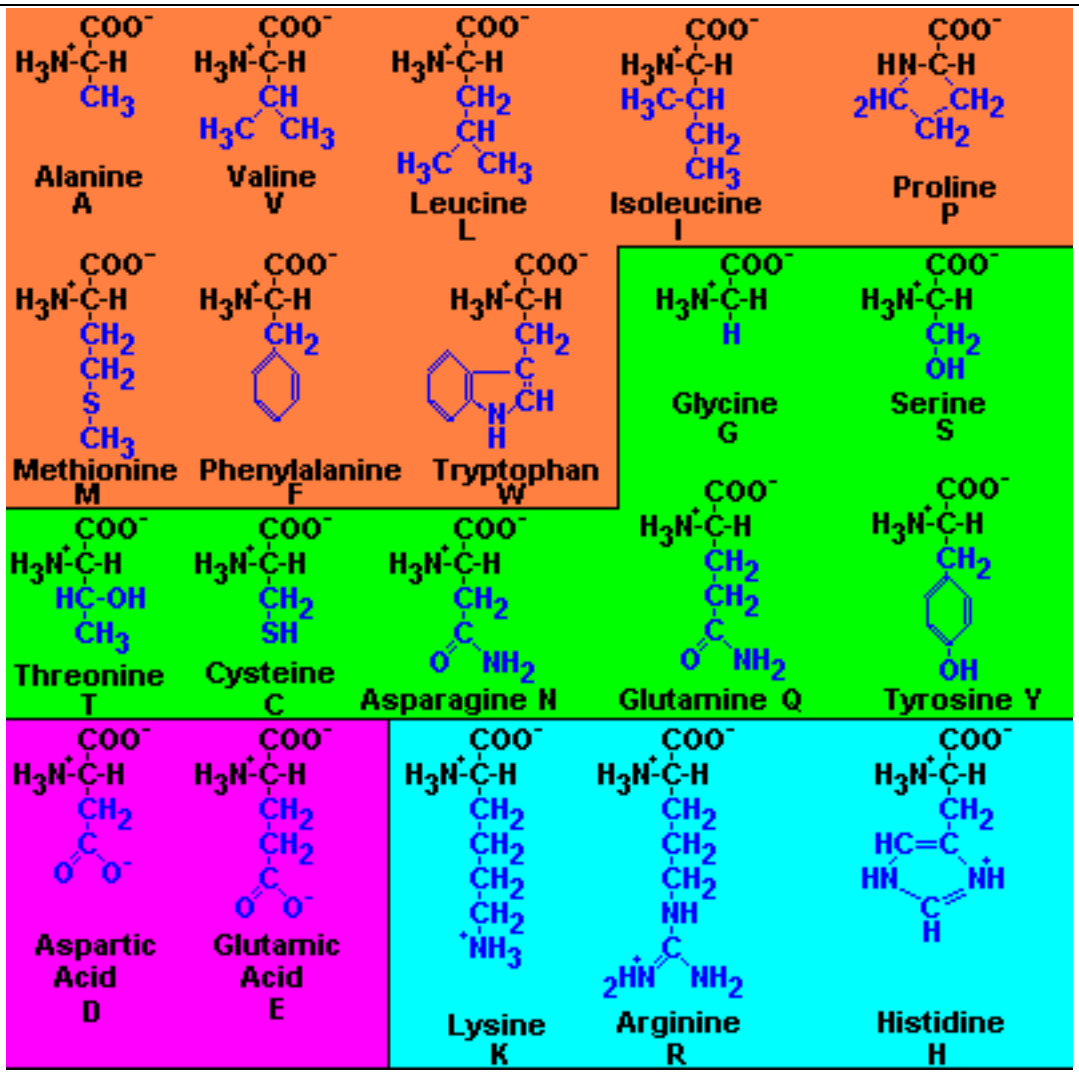


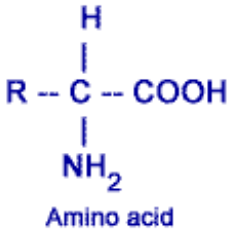

**Fig. 2.4** The 20 amino acids found in proteins

| **Table 2.2** Amino Acid Classification | |
|---|---|
| *Aliphatic Amino Acids*<br>glycine<br>alanine | *Hydrophobic*<br>valine - beta-branched<br>leucine<br>isoleucine - beta-branched, second chiral center |
| *a-imino acid*<br>proline - containing a pyrrolidine ring | *Aromatic Amino Acids*<br>- all are considered hydrophobic. However, tyrosine and tryptophan also can form hydrogen bonds.<br>phenylalanine |

| | |
|---|---|
| | tyrosine<br>tryptophan - containing an indole group |
| *Sulfur-containing*<br>- both are considered hydrophobic. Cysteine can form disulfide bonds with another cysteine.<br>cysteine - containing a thiol group<br>methionine - containing a thioether group | *Aliphatic hydroxyl-containing*<br>serine<br>threonine - beta-branched, second chiral center |
| *Basic*<br>lysine - containing a primary amine<br>arginine - containing a guanidino group<br>histidine - containing an imidazole group | *Acidic*<br>aspartate<br>glutamate |
| *Amide Derivatives*<br>(of the acidic amino acids)<br>asparagine<br>glutamine | |

Zwitter ion

Amino acids have a zwitterionic character under physiological conditions (Fig. 2.3).

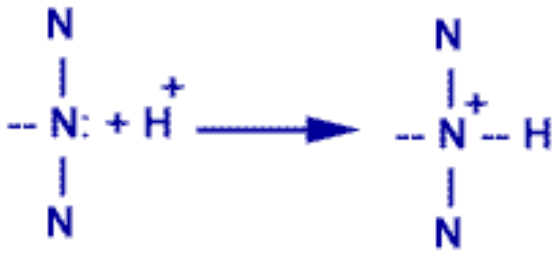



Amino acid

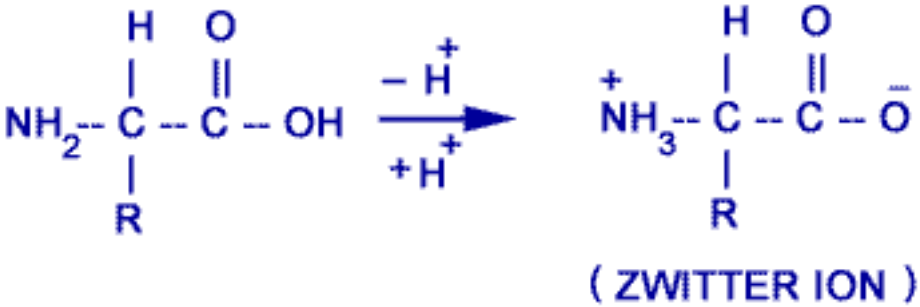
This uncharged electrically neutral molecule as shown above is not correct.

The (COOH) group being acidic will ionize in water as follows:


On the other hand the -$NH_2$ group with a lone pair on electron is a Lewis base and capable of accepting a proton.




Remember that the carboxylic acid group can donate a proton and the amino group is basic and can accept a proton. So, an amino acid can react with itself by the acid group donating a proton to the amino group, thus making one end of the molecule negative and the other end positive. The resulting ion is dipolar charged but overall electrically neutral. This is called *Zwitter ion*. Due to this reason amino acids are amphoteric i.e. they donate or accept proton.




One of the values of this change is that it increases the solubility of amino acids in water. The amino acid zwitterions can still join to one another by dehydration reactions. The oxygen in the

departing water molecule still comes from the carboxylate group, but now both of the hydrogens come from the amino group.

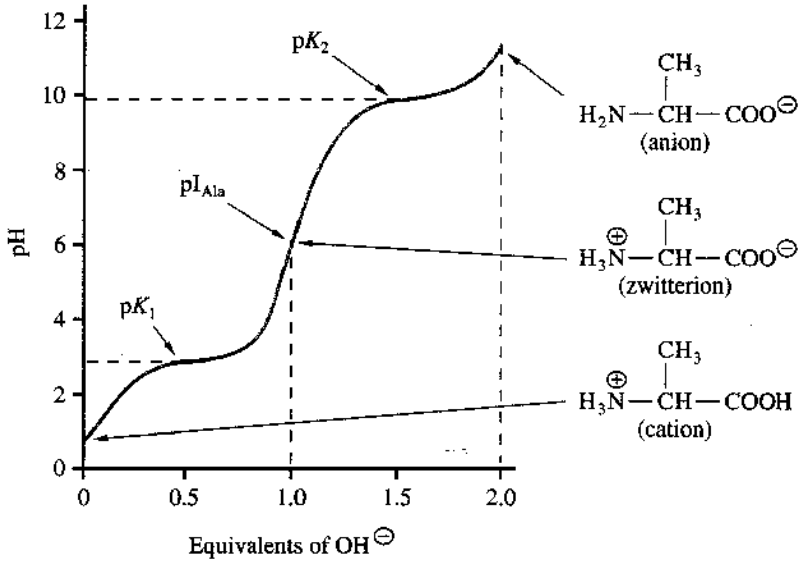



**Fig. 2.3** Titration curve for Alanine

| Amino acid | Abbreviation | $pK_1$ α-COOH | $pK_2$ α-NH2 | $pK_3$ side chain |
|---|---|---|---|---|
| Glycine | Gly | 2.34 | 9.60 | |
| Alanine | Ala | 2.34 | 9.69 | |
| Asparagine | Asn | | | |
| Isoleucine | Ile | | | |
| Leucine | Leu | 2.36 | 9.60 | |
| Methionine | Met | | | |
| Phenylalanine | Phe | | | |
| Proline | Pro | | | |
| Serine | Ser | 2.21 | 9.15 | |
| Threonine | Thr | 2.63 | 10.43 | |
| Tryptophane | Trp | | | |
| Valine | Val | | | |
| Glutamine | Gln | 2.17 | 9.13 | |
| Aspartate | Asp | 2.09 | 9.82 | 3.86 |
| Glutamate | Glu | 2.19 | 9.67 | 4.25 |
| Histidine | His | 1.82 | 9.17 | **6.00** |
| Cysteine | Cys | 1.71 | 10.78 | **8.33** |
| Tyrosine | Tyr | 2.20 | 9.11 | 10.07 |
| Lysine | Lys | 2.18 | 8.95 | 10.53 |
| Arginine | Arg | 2.17 | 9.04 | 12.48 |

Essential Amino Acids

Amino acids that can not be synthesized by our body are known as essential amino acids. They are necessary for growth and transmission of impulses in the nervous system.They must be supplied to our body through our diet . Their deficiency results in many diseases. About 10 amino acids are essential.

For example ; argenine, valine, lysine , phenol-alanine, leucine, isoleucine, threonine, methionine, tryptophan, histadine.

Role of amino acids in the human body

Amino acids have a very vital role in our body. When we eat protein rich food, many enzymes act on protein and hydrolyze it in to amino acids. The blood receive it by absorption through the walls of intestine and carries them to cells where one of the following reactions take place:

- The amino acid may be converted into body protein.
- Oxidation may take place to provide energy.

If our diet is low in carbohydrates or fats, body proteins may be transformed into either of these or used to make hormones and other body requirements.

## *The peptide bond*

Amino acids in proteins are linked together through an acid- amide group type of bond known as peptide bond. The peptide bond is formed between two amino acid molecules when amino group of one amino acid is linked with the carboxylic group to the other amino acid molecule by the elimination of water molecule (Fig. 2.4).

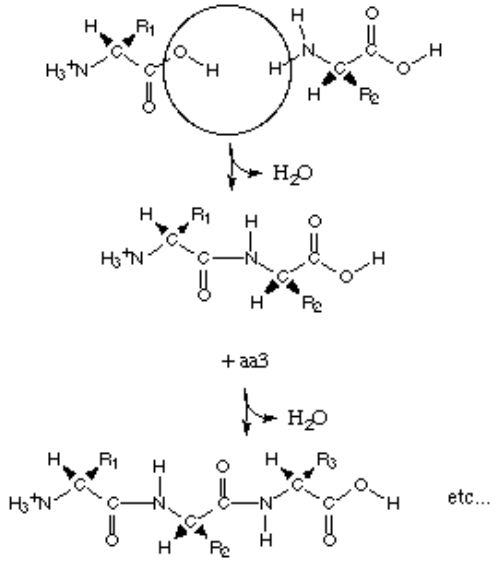



**Fig. 2.4** Peptide bond formation

The C-N-bond has some "double bond" quality. There is no free rotation around the C-N-bond and this bond is shorter than expected for a singly bond. All four atoms of the peptide group and the two α-C-atoms are in the same plane. The structure is rigid and the hydrogen of the substituted amino group is almost always placed in trans position in respect to the oxygen atom of the carbonyl group (exception: amino acid before a proline). In comparison, the freedom of rotation about the neighboring C-C and C-N bonds is high, as expected for atoms linked by a single bond.

## *Protein conformation*

The **conformation** of a molecule corresponds to its orientation in space (**configuration** by comparison concerns the arrangement of the atoms in a molecule). The conformation of a protein incorporates the so-called secondary, tertiary and sometimes quaternary structures (Fig. 2.5). The function of a protein arises from its conformation. Denatured or "stretched out" polypeptide chains normally do not have the same (or any) biological activity.

| | |
|---|---|
| Primary structure | AA-sequence |
| Secondary structure | α-helix/β-sheet |
| Tertiary structure | 3D-structure (computer simulations) |
| Quaternary structure | several polypeptide chains |

**Conformation**

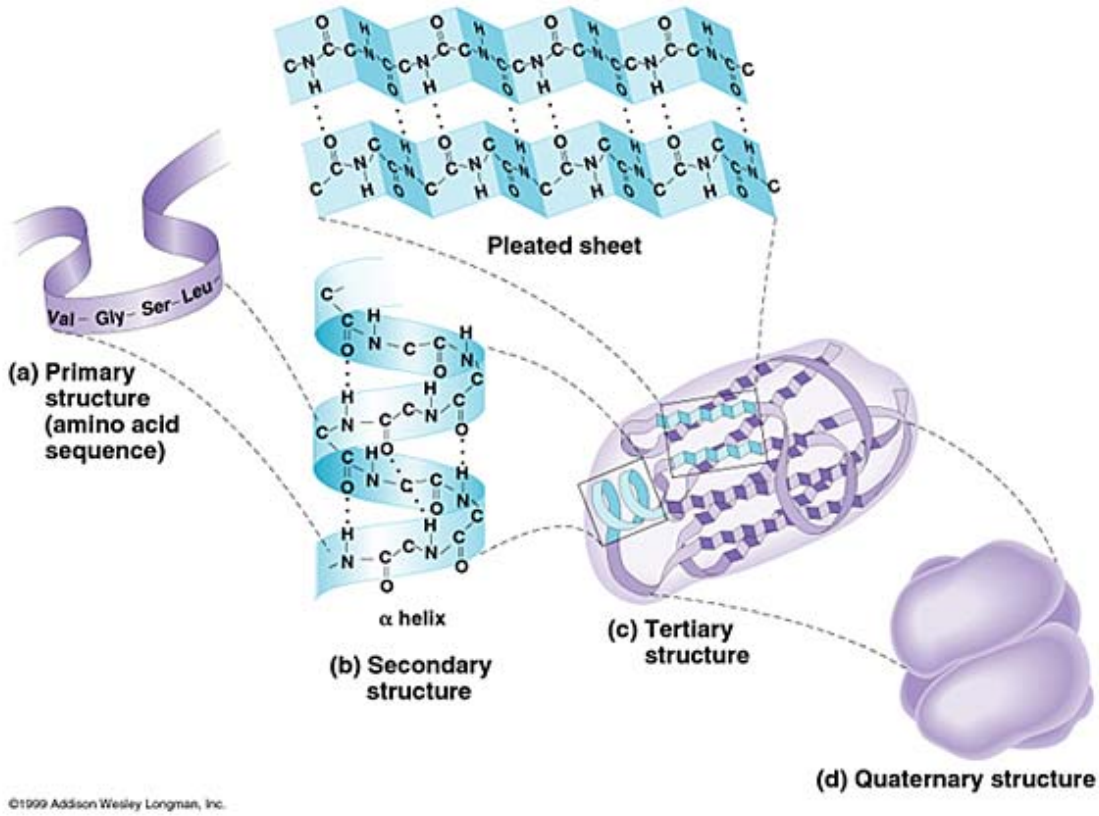


**Fig. 2.5** Four levels of organization of a protein.

Proteins tend to have a complex three-dimensional structure. If one considers the number of atoms and bonds in a typical protein it is amazing to note that normally each protein has a single (sometimes a few) structure. If the protein is reversibly denatured and the denaturing agent removed afterwards, the native structure is often reestablished quite quickly. Various effects contribute to the stabilization of the native structure under physiological conditions. Hydrophobic interactions seem to be most important, but H-bridges and electrostatic ones also contribute. In fact, the side chains of 11 out of the 20 protein-forming amino acids can participate in hydrogen bonding. Especially in extracellular proteins we find disulfide bridges between cysteine residues as stabilizing agents. These can also be used to covalently link several polypeptide chains. Cross-links derived from lysine side chains are also possible (e.g. in collagen fibers). The native structure of a protein is not absolutely rigid. Especially when studying enzymes, we will see that some adjustment ("induced fit") is still possible and in fact often necessary, if a protein is to fulfill its biological function.

At present it is not possible to deduce the three-dimensional structure of a protein from its amino acid sequence. However, some rules have been deduced. In the case of water-soluble (globular) proteins, the structure is very compact. Polar and charged amino acids are located at the outside of the molecule, while hydrophobic (non-polar) amino acids are found mainly inside the protein molecule. Even when the protein is "water-soluble", the interior can exclude water molecules to a high degree. In fact, many biocatalytic reactions require a highly hydrophobic environment in order to take place. The "active site" of an enzyme is therefore typically found in a "catalytic cleft", i.e. burrowed deeply in the hydrophobic interior of the enzyme.

## *The secondary structures: α–helix and β-sheet*

Two recurring motives (**secondary structures**) have been observed in proteins, the **α-helix** and the **β-pleated sheet**. The α-helix (Fig. 2.6) is a rod like structure and perhaps the simplest stable form a polypeptide chain can take. The α-helix content of proteins can be between 0 % and 100 %. In the case of L-amino acids, an α-helix, which turns clockwise is more stable. The helix is stabilized by hydrogen bridges between the NH and the CO groups of the main chain. An α-helix is most stable in the case of small amino acids (alanine) and in the absence of charges and bulky side chains. Proline cannot form an α-helix and thus interrupts the chain.

**Fig. 2.6 An α-helix is a common structure formed by parts of the polypeptide chain in proteins.** (A) Every peptide bond in an α helix is hydrogen-bonded t a neighboring peptide bond (B) A perfect α helix is shown in outline (C) Ribbon model of an α helix.

The β-sheet (Fig. 2.7) is also stabilized by H-bridges, this time between NH and CO groups of different β-stands (alignment either in parallel or anti-parallel). The polypeptide chains (β-strands) are nearly fully extended. The axial distance between neighboring amino acids is 3.5 Å as compared to the 1.5 Å in the α-helix. The residues can be located above or below the sheet. Lysine and alanine are the amino acids most suited to forming a β-sheet.

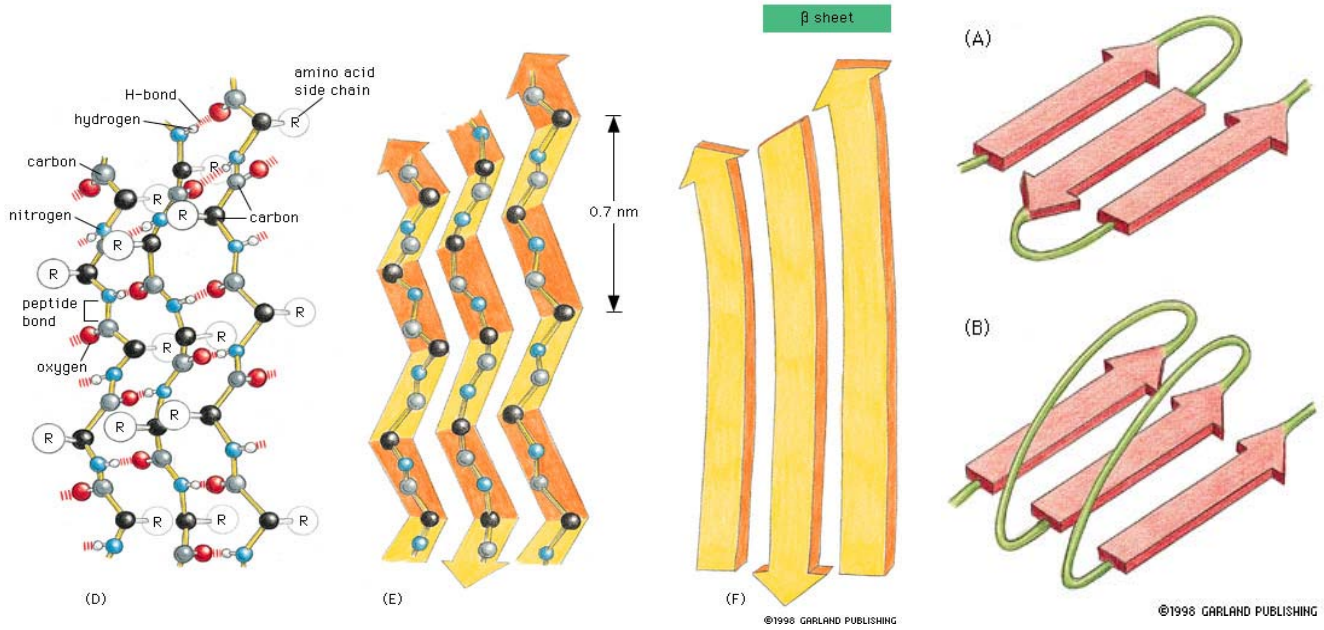



**Fig. 2.7. A β-sheet is a common structure formed by parts of the polypeptide chain in globular proteins.** (A-B) Chains can be parallel or antiparallel in a β-sheet (D-E) A perfect antiparallel β-sheet is shown in detail, with the amino acid side chains denoted R (F) Ribbon model of a β-sheet.

| α-helix | β-sheet |
|---|---|
| simplest stable form | volumetric side groups cause problems |
| ca 3.6 amino acids per turn | chains can be parallel or antiparallel |
| can contain either | common structure in globular proteins (antibodies) |
| L- or D-amino acids | |
| the right handed form is more | |
| stable for L -amino acids | |
| some amino acids are more | |
| suited to helix formation (alanine, leucine, phenylalanine, tryptophane, methionine, histidine, glutamine, valine) | |
| proline and hydroxyproline interrupt the helix | |

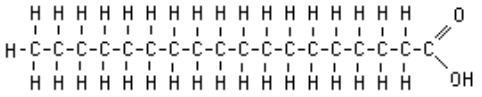
# 3. LIPIDS

Lipids contain large hydrophobic structures (hydrocarbon skeleton). They are mostly non-soluble in water. Important biologicals such as fatty acids, fats (oils), terpenes (steroids, cholesterol), prostaglandins, but also complex molecules such as phosphoglycerides, sphingolipides, and hybrid molecules such as lipoproteins or glycolipids, belong in this group. Certain fat-soluble vitamins (A, E, K) and hormones are also found in this class. Lipids serve as structural elements, e.g., in the cell membrane. They store and transport metabolic energy in chemical form. They also form protective surface and fatty layers. In addition, they play an important role in cell-cell recognition and in the immune response.

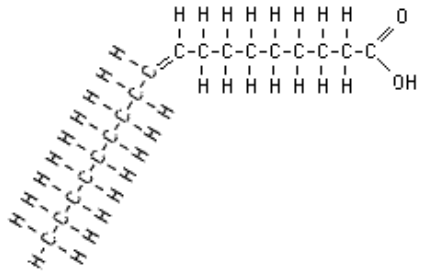
## *Fatty acids*

Fatty acids (Table 3.1.) contain long hydrocarbon chains and a terminal (under physiological conditions deprotonated) carboxylic acid end group. The hydrocarbon chain can be fully saturated or contain one or several double bonds (unsaturated fatty acid). A few fatty acids with triple bounds have also been found. In nature, free fatty acids are rarely observed. Fatty acid esters (lipids) are much more common.

**Table 3.1.** Some Examples of Fatty Acids



Stearic acid, a saturated fatty acid



Oleic acid, a monounsaturated fatty acid. Note that the double bond is *cis*; this is the common natural configuration.

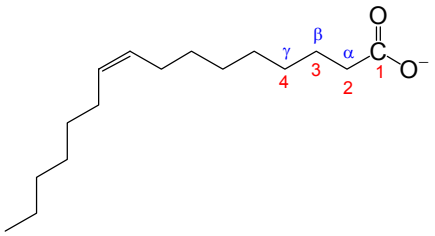
The fatty acid nomenclature (Table 3.2.) allows to compile in the name the most relevant information concerning the molecule, i.e. the length/number of C atoms and the presence and location of the double (triple bonds). For example, the most common saturated fatty acid, trivial name "palmitate", is officially called 16 : 0, n-hexadecanoate. Oleate, an unsaturated fatty acids, would be described as 18 : $1\Delta^9$, cis-$\Delta^9$-Octadecenoate. The configuration of the double bond is usually cis. Trans-double bonds are always marked as such, for example, 18 : $1\Delta^{9trans}$, in the case of Elaidinacid. Unsaturated fatty acids have lower melting points than saturated ones.

Most fatty acids carry an even number of carbon atoms; 16 and 18 are especially popular. Double bonds are often found in the C9 and C10 region. When a fatty acid contains more than one double bond, these are almost never conjugated. Mammals are not able to synthesize fatty acids with more than one double bond. They do need these fatty acids, however, e.g., for the production of prosta-

**Table 3.2.** Fatty Acid Nomenclature

Abbreviated notation for a 16-C fatty acid with one cis double bond between carbons 9 &10 is **16:1 cis $\Delta^9$.**

**Some Examples:**
14:0 myristic acid
16:0 palmitic acid
18:0 stearic acid
18:1 cis$\Delta^9$ oleic acid
18:2 cis$\Delta^{9,12}$ linoleic acid
18:3 cis$\Delta^{9,12,15}$ a-linonenic acid
20:4 cis$\Delta^{5,8,11,14}$ arachidonic acid
20:5 cis$\Delta^{5,8,11,14,17}$ eicosapentaenoic acid




glandins and therefore have to rely on their diet to provide such molecules. Linoleate (cis, cis-Δ9, Δ12-Octadecadienolate) and Linolenate (all-cis-Δ9,Δ12,Δ15-Octadecatrienolate), which are very common in certain plants, are the so-called **essential fatty acids**.

## *Hormones derived from fatty acids*

**Arachidonate**, a 20:4 fatty acid derived from the essential fatty acid linolenate is the precursor of several classes of signal molecules, including prostaglandins, protacyclins, thromboxanes and leukotrines (Fig. 3.1.). Prostaglandins are short-lived local hormones fashioned by reductases and isomerases. They are 20 carbon fatty acids containing a 5-carbon ring. Depending on the cell, they may stimulate inflammation, regulate the blood flow, control ion transport across membranes, modulate synaptic transmission or induce sleep. All prostaglandins lower the blood pressure and cause the contraction of smooth muscles. Aspirin inhibits the first step in the synthesis of prostaglandins by irreversibly inhibiting prostaglandin synthase (Fig. 3.2.).

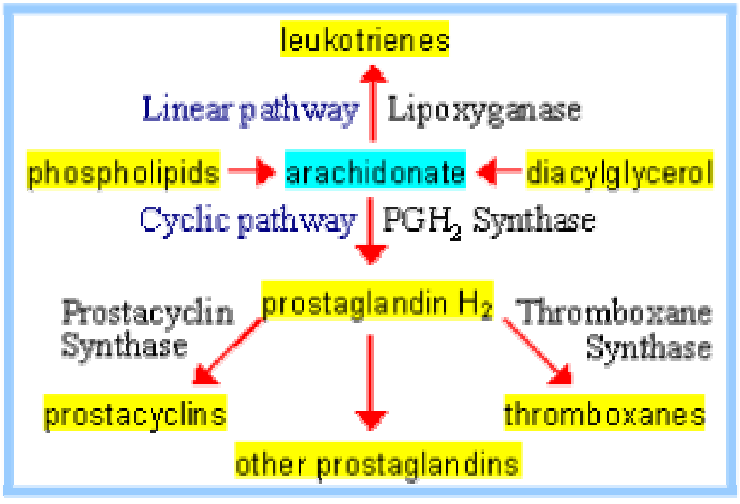



**Fig. 3.1.** *Arachidonate* – a major precursor of several classes of signal molecules.

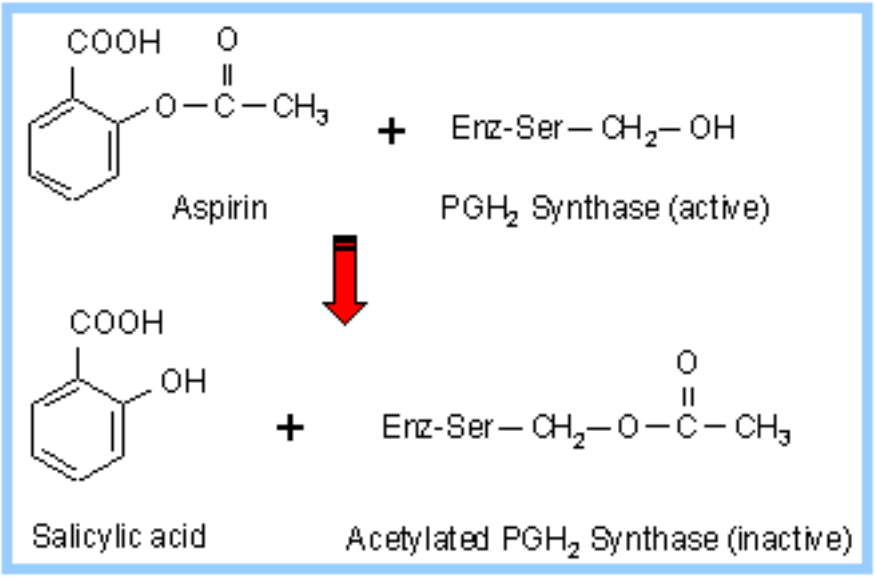



**Fig. 3.2.** Inhibition of Prostaglandin Synthase by Aspirin.

## *Neutral Fats (complex lipids)*

Neutral fats are fatty acid esters of glycerol (Triacylglycerols). Simple triglycerols contain three times the same fatty acid (Fig. 3.3.). Triacylgylcerols are the most common type of lipid. They constitute an important storage form of chemical energy in both plants and animals, since they are both reduced and anhydrous. The complete oxidation of 1 g of fatty acid yields about 9 kcal, as compared to about 4 kcal in the case of proteins of carbohydrates. In addition, carbohydrates tend to bind water. Consequently, one gram of nearly anhydrous fat stores more than six times as much energy as a gram of hydrated glycogen (a polysaccharide).

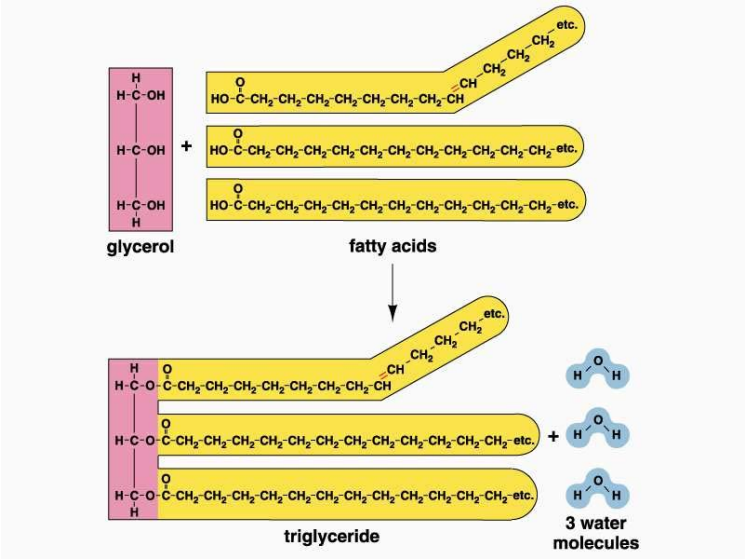



**Fig. 3.3.** Formation of a triglyceride.

## *Other complex lipids*

While neutral fats are mainly used for energy storage, other types of complex lipids are used to form, e.g. cellular membranes. Typically these lipids contain a polar "head group" in addition to the hydrophobic moiety of the molecule (Fig. 3.4.). Such lipids can either be based on a glycerol ester or on the sphingosine structure. In the **phosphoglycerides** (important components of most cellular membranes) one OH-group of the glycerol forms an ester with phosphoric acid, the two others carry fatty acids. One example for a phosphoglyceride is lecithin. **Sphingolipids** are also important components of plant and animal cells. These structures are especially common in the brain and in the nerve cells. The molecules contain a fatty acid, sphingosin (or a derivative thereof) and a polar group. The cells of higher animal contain mainly sphingomyeline, which contain phosphorylethanolamine or phosphorylcholin as polar group. Glycosphingolipids contain instead a neutral sugar molecule (mono- and disaccharides) as polar group. These molecules are therefore uncharged. The most complex type of glycosphingolipids are the **gangliosides**, which contain in addition one or more molecules of sialic acid (negatively charged). Approximately 6 % of the lipids found in the gray matter of the brain are gangliosides. Glycosphingolipids perform several important functions in living organisms, including receptors for neurotransmitters, nerve impulses, blood group specificity, tissue immune reactions.

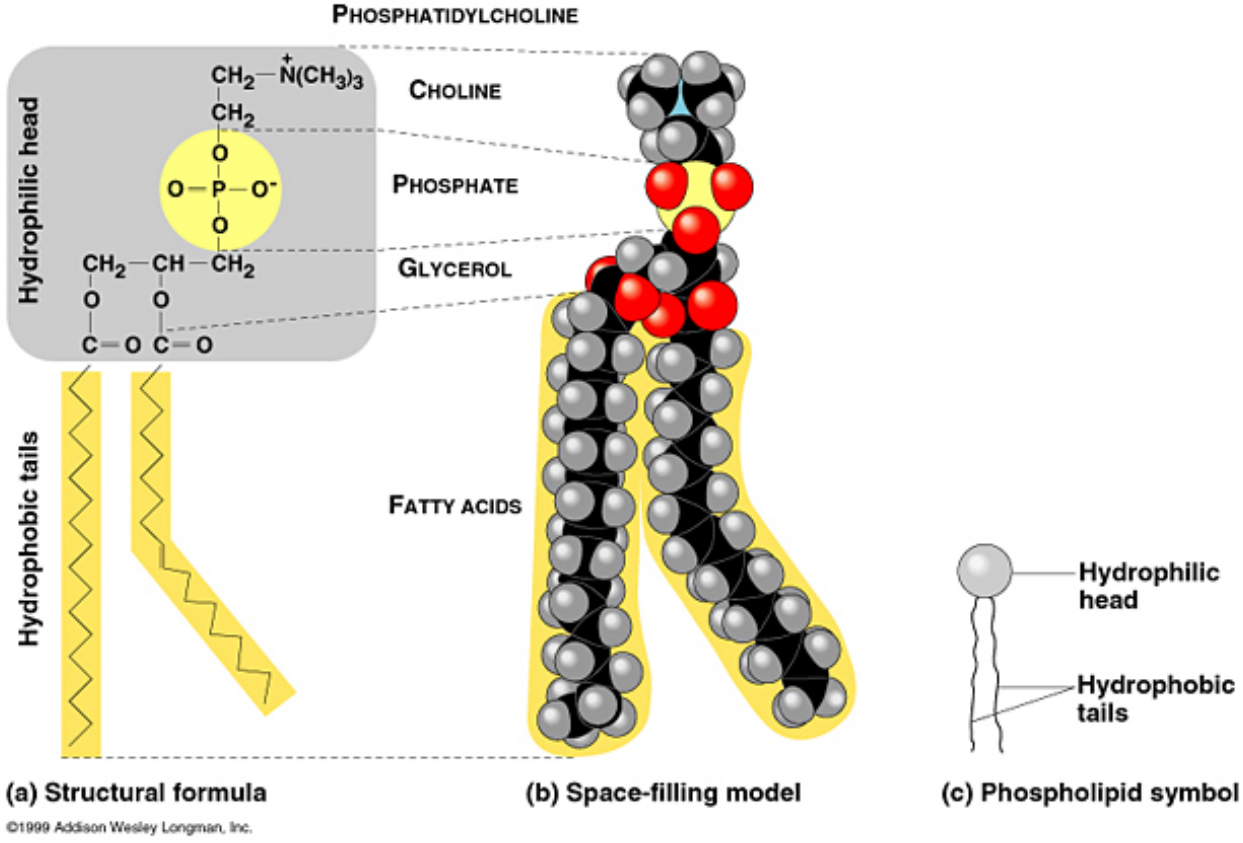



**Fig. 3.4.** Complex Lipids

## *Micelles, lipid double layers*

When an amphiphilic molecule such as a deprotonated fatty acid is placed into water in a concentration above the "critical micelle concentration", only a small number of the molecules is truly dissolved on a molecular level. The majority of the molecules form so-called micellar aggregates , in which the hydrophobic chains avoid direct contact with the water molecules (Fig. 3.5.). Only the polar head groups come into contact with the water under these circumstances. For sterical reasons (e.g. **two** bulky fatty acid tails) polar lipid such as a phosphoglyceride are less prone to form micelles. Instead extended lipid bilayers are spontaneously formed in order to minimize the contact between the hydrophobic moieties and the water molecules. These structures can be seen as an analog to the cell membrane. Cell membranes (see chapter on membrane processes) have been shown contain up to 75% lipids (rest proteins), mostly phosphoglycerides, which are arranged in a bilayer.

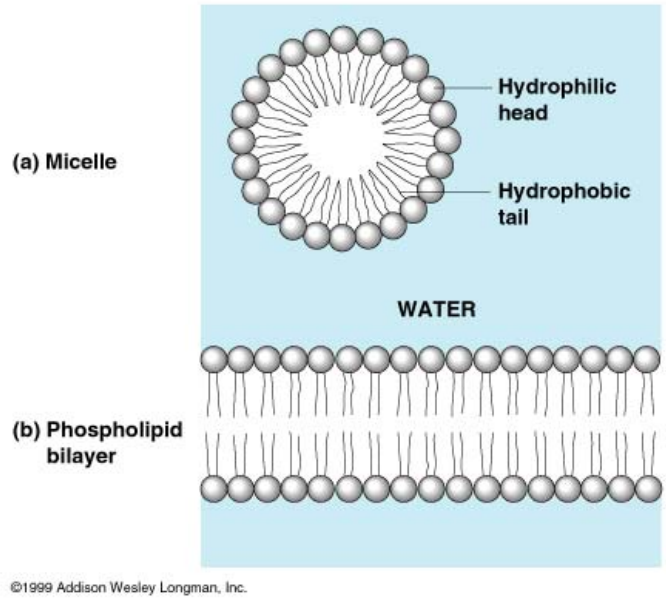


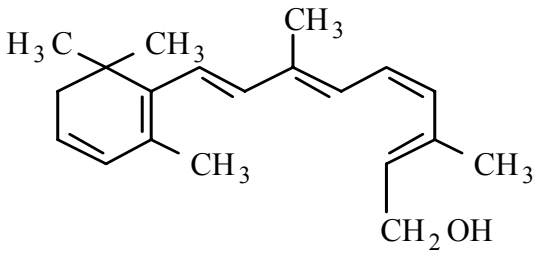
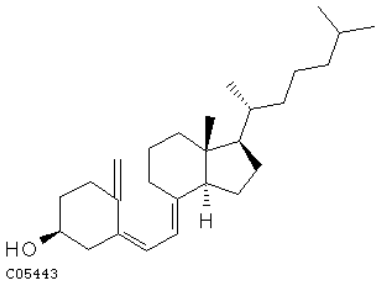

**Fig. 3.5.** a. Diagram of a section of a micelle formed from ionized fatty acid molecules. Most phospholipids do not form micelles.

b. Diagram of a section of a bilayer membrane formed from phospholipid molecules.

## *Simple lipids*

Simple lipids are hydrocarbons, which do not contain fatty acid derived structures. In nature these are less common than the complex lipids discussed above, but also perform many important biochemical functions. Some of these lipids are highly active even at low concentration (hormones). The most important class in this group are the terpenes (including e.g. the steroids and certain fat-soluble vitamins) (Table 3.3.). All terpenes are based on the isoprene structure. They can be linear, cyclic or both. Especially in plants, terpenes with an intensive and characteristic smell/taste are found (menthol, camphor). Squalene, a triterpene, is an important precursor of cholesterol (itself a steroid precursor), β-carotene an important precursor of vitamin A (Fig. 3.6.). Vitamins E and K are also terpenes.

**Table 3.3.** Some Examples of Terpenes

| | |
|---|---|
|  |  |
| Vitamin A (Retinol) | Vitamin $D_3$ (cholecalciferol) |

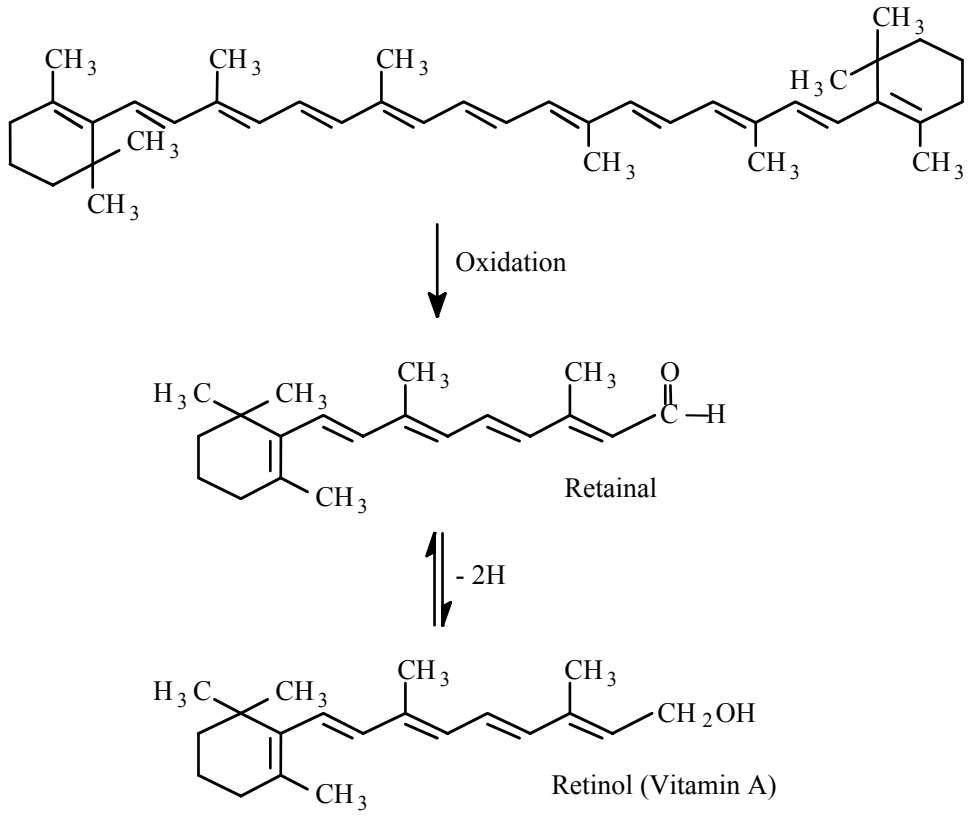



**Fig. 3.6.** Synthesis of Vitamin A from β-carotene

**Steroids** are all derivatives of a tetracyclic fully saturated hydrocarbon. Many steroids with varied function and structure are known. They differ in number and position of the double bond, in type, number and position of the functional group(s) as well as in configuration of the ring system and the functional groups. A well-known example of this group is cholesterol, the precursor of the male and female sexual hormones (androgens and estrogens) and progesterone.

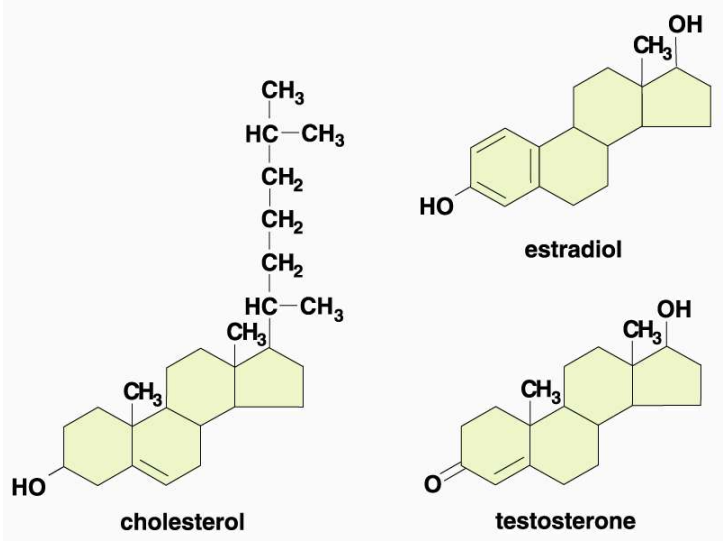



**Fig. 3.7.** Examples of Steroids

# 4. CARBOHYDRATES (SACCHARIDES)

The basic units of the carbohydrates are the **monosaccharides** (simple sugars). The formula is often $(CH_2O)n$, hence the name **Carbohydrate**. The carbon chain is linear (no branches) and each carbon atom (safe for the carbonyl / ketyl group) carries an OH-group. All monosaccharides are white and water-soluble. Most taste sweet. Monosaccharides can be considered as polyhydroxyaldehydes (**aldoses**) and polyhydroxyketones (**ketoses**). A monosaccharide consists of a single aldose or ketose unit. **D-Glucose** is for example the most common monosaccharide in nature. It is both an important nutrient (under normal condition the brain uses exclusively glucose for energy), but also the building block of the polysaccharides **starch** and **cellulose**. **Oligosaccharides** contain up to 10 monosaccharide units. **Polysaccharides** are made from long chains or networks of monosaccharides units. Carbohydrates have several important functions in nature. One is the short (glucose) to midterm (starch glycogen) storage of energy, another the provision of building materials for cell walls (structure) such as cellulose. Carbohydrates serve as metabolic intermediates and the addition of sugar units plays an important role in the posttranslatorial modification of proteins (fine-tuning of function). Carbohydrate units on cell surfaces play an important role in cell-cell communication. ATP is a phosphorylated sugar derivative, the same is true for many coenzymes. Finally, sugars are important structuring elements in the backbone of both DNA and RNA.

## *Monosaccharides*

The simplest sugar unit is the monosaccharide, a single aldose or ketose unit (Fig. 4.1.). The smallest possible monosaccharide unit is the **triose**, in particular glycerinaldehyd (aldotriose) and dihydroxy acetone (ketotriose)

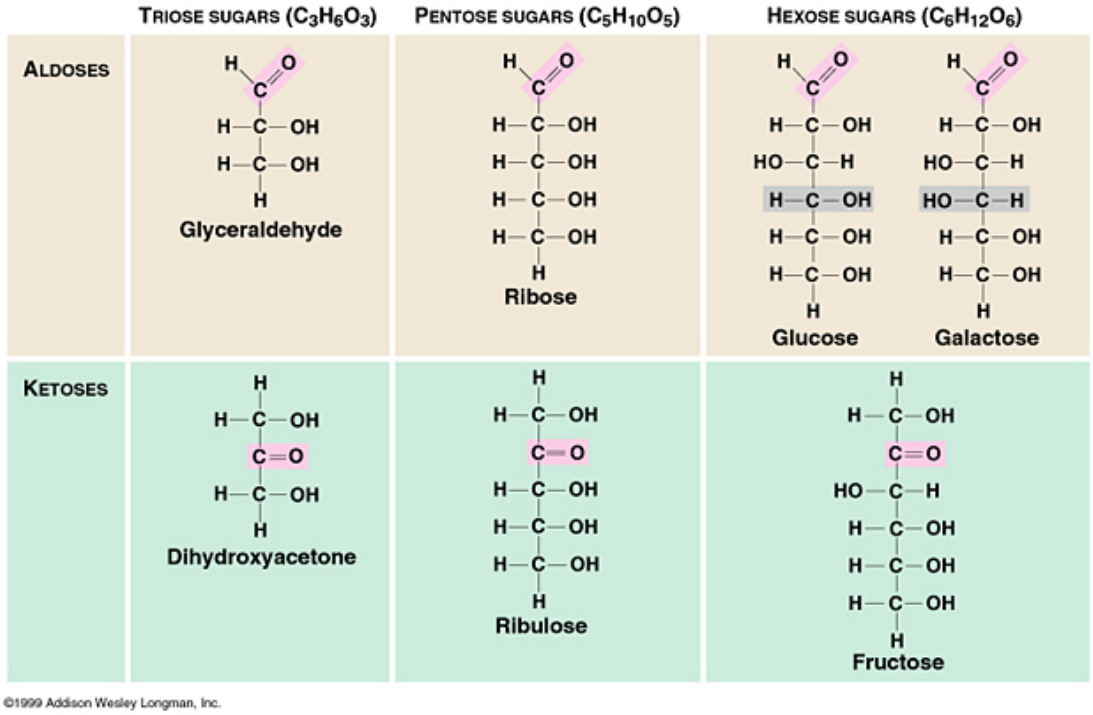


**Fig. 4.1.** Structure of Monosaccharides

Larger monosaccharides are named according to their number of carbon atoms, i.e.: Tetrose (C4), Pentose (C5), Hexose (C6), Heptose (C7), etc. All monosaccharides (safe for dihydroxyacetone, which is therefore not optically active) contain one or more chiral C-atoms (Fig. 4.2.). If the molecule contains more than one chiral C-atom, it is the one, which is farthest away from the carbonyl group, which determines if the molecule has **D or L configuration**. Most sugars found in nature are of the D-configuration. D-Glucose is the most common monosaccharide in nature. L sugars are rare; some examples include L-Fucose, L-Rhamnose, L-Sorbose.

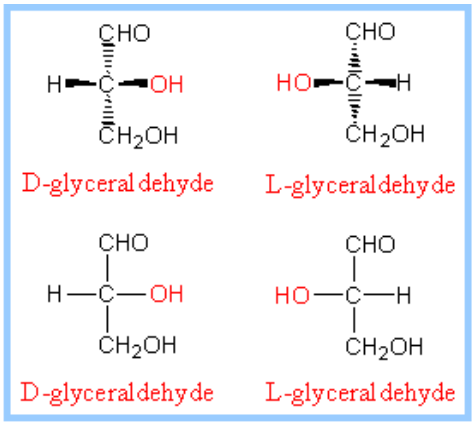



**Fig. 4.2.** D and L configuration of monosaccharides

Two sugars that differ only in the configuration of a single asymmetric carbon are called **epimers**, e.g. D-Glucose and D -Mannose (C2) or D-Glucose and D-Galactose (C4). Two sugars that are mirror images are called **enantiomers**, while **diastereomers** are sugars that differ at a carbon atom without being mirror images (Fig. 4.3 and Fig. 4.4).

Many of the large sugar units (> C5) tend to form ring structures in solution. The ring is formed by the reaction of the aldehyde or keto group with one of the OH-groups, through the formation of intramolecular hemiacetals and hemiketals (Fig. 4.5). Six-membered rings are called **pyranoses** five-membered rings **furanoses**. The open and the ring form are in equilibrium with each other. Depending on the type of monosaccharide, in aqueous solution these ring forms can be the preferred structure in solution (D-glucose in water is to more than 99 % in the ring form). Since the reaction creates an asymmetric C-atom, two ring forms can be produced (**anomeric forms**), which differ in their physico-chemical behavior.

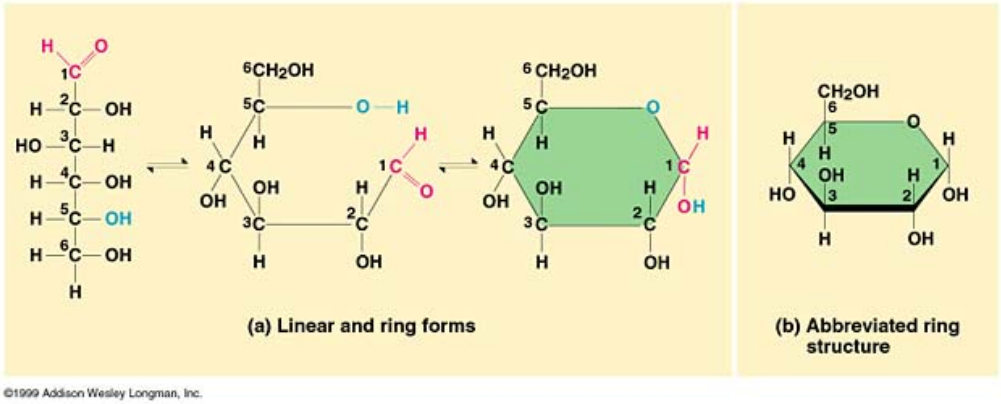


**Fig. 4.5.** Formation of ring structures of sugars in solutions.

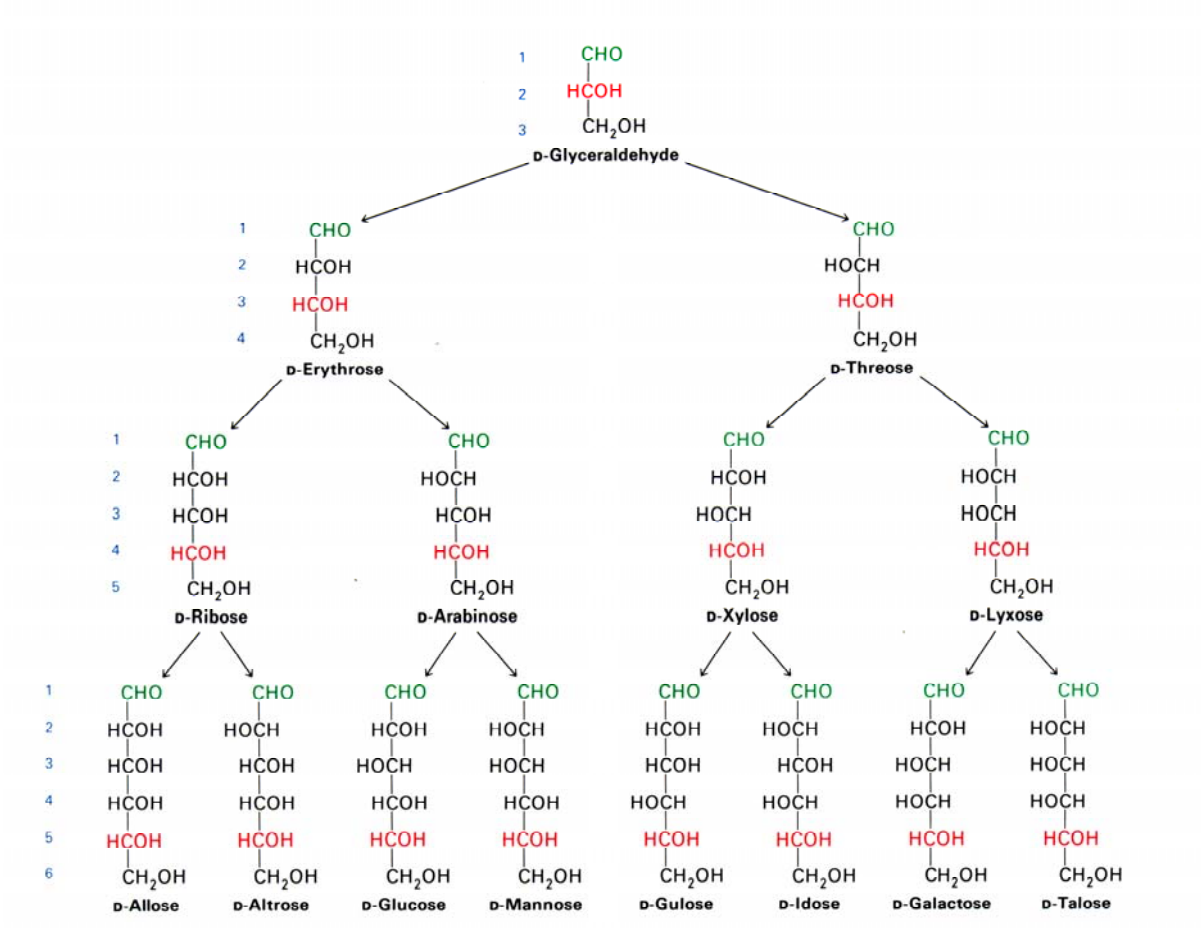



**Fig. 4.3.** Stereochemistry of the D-aldoses

In water two **anomeric** forms exist of glucose, α-D-glucopyranose (33 %) and β-D-glucopyranose (66 %). The $[\alpha]_D^{20}$ is at first +112.2° or = 18.7°, later it becomes $[\alpha]_D^{20}$ = 52.7° (**mutarotation**).

| | α-D-Glucose | β-D-Glucose |
|---|---|---|
| Melting point | 146 | 150 |
| Solubility in water | 82.5 | 178 |
| Oxidation rate for glucose oxidase | 100 | < 1.0 |

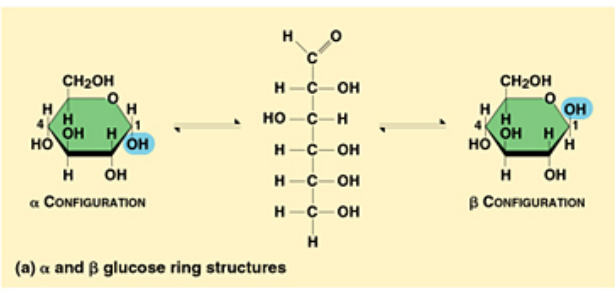


**Fig. 4.6.** Anomeric forms of glucose.

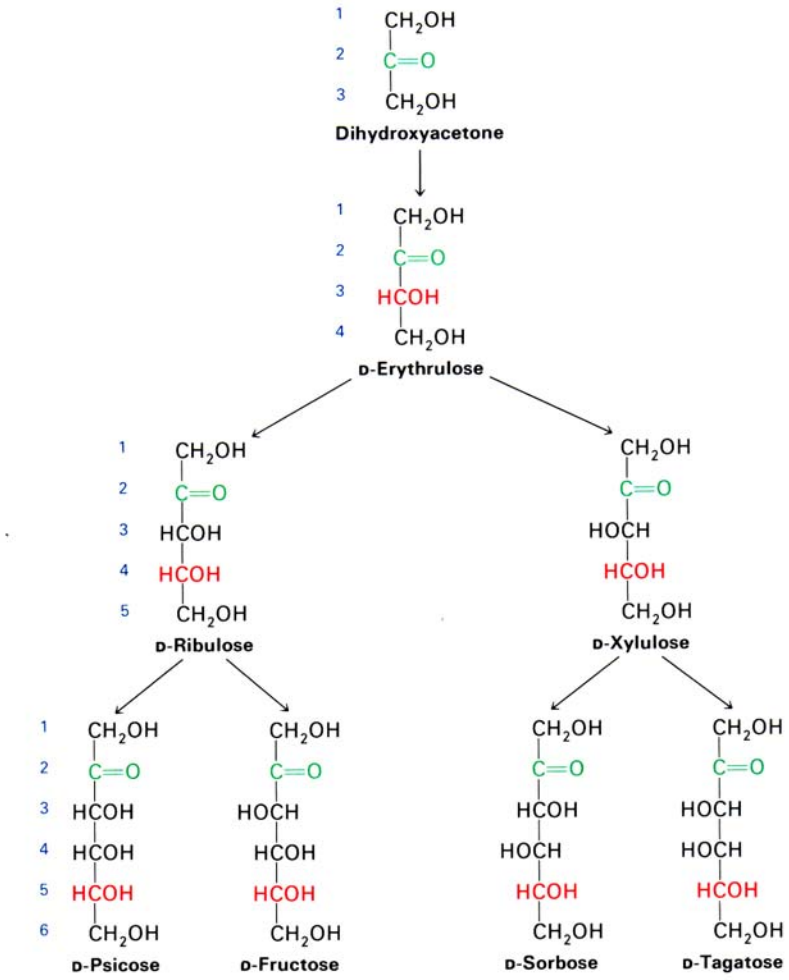



**Fig. 4.4.** Stereochemistry of the D-ketoses

A six-membered pyranose ring cannot be planar. Instead a (stable) **chair** or a (less stable) **boat form** is assumed (Fig. 4.5.). The substituents on the ring can then be either in an axial or an equatorial position. Bulky groups are less crowded in the equatorial position. This contributes to the stability of glucopyranose (all OH-groups in the equatorial position). Furanoses on the other hand may take the **"envelope" form** to reduce the strain within the molecule.

## *The glycosidic bond*

Sugar units can form mixed acetales with other alcohols (e.g. **also the OH-groups of other sugar molecules**). The created linkage is called a glycosidic bond (Fig. 4.8.), in particular an O-glycosidic bond (N-glycosidic bonds link sugars to amino groups). The two monosaccharide units in a disaccharide are typically linked via such an O-glycosidic bonds (Fig. 4.9.). The glycosidic bond is stable towards bases and can be hydrolyzed by heating at low pH or by glycosidases.

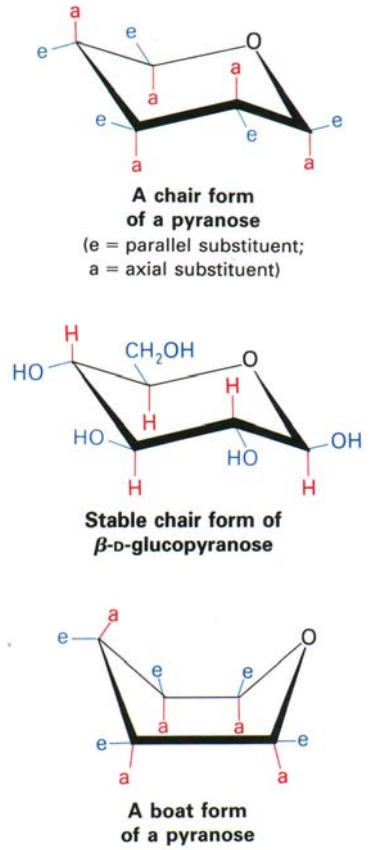



**Fig. 4.7.** Chair and Boat form of a six-membered pyranose ring

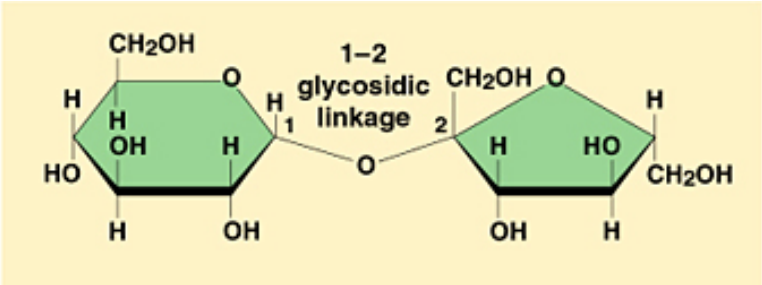



**Fig. 4.8.** The glycosidic bond

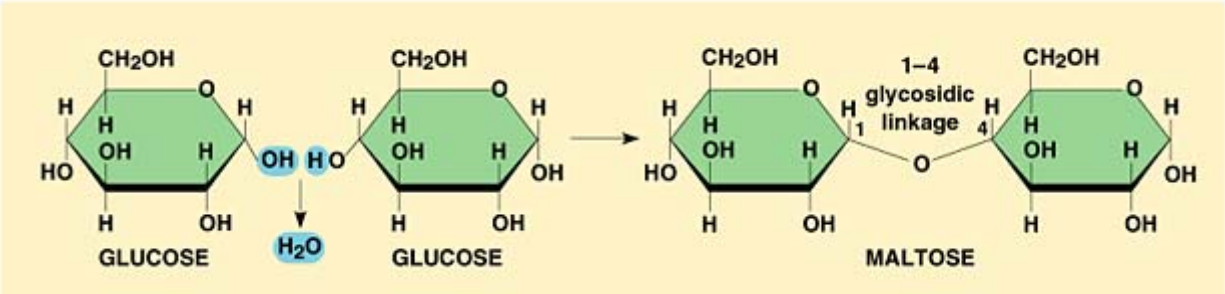
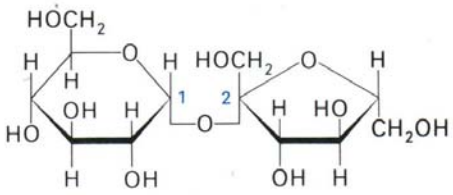


**Fig. 4.9.** Formation of maltose.

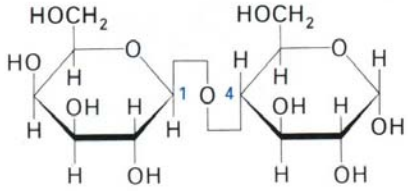
## *Disaccharides and polysaccharides (Glycanes)*

**Disaccharides** consist of two sugar units joined by an O-glycosidic bond. In nature **sucrose** (glucose/fructose), **lactose** (galactose/glucose) and **maltose** (glucose/glucose) are the most common disaccharides (Fig. 4.10.).

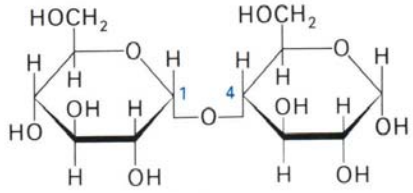
Since the involved C-atoms are chiral, two types of linkages (called again α and β) are possible. The nomenclature of the glycosidic bond indicates the type (α̃β) and the number of the involved C-atoms in the monosaccharides, for example the name of the maltose α-form, a disaccharide made from two glucose units would be: O-α-D-glucopyranosil-(1 → 4)-β-D- glucopyranose (Fig. 4.10.). Since the second unit (free anomeric C-atom) can still be oxidized, it is called the reducing end of the sugar (saccharide).



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



**Lactose**
**(β-D-Galactopyranosyl-(1→4)-α-D-glucopyranose)**



**Maltose**
**(α-D-Glucopyranosyl-(1→4)-α-D-glucopyranose)**

**Fig. 4.10.** Disaccharides and there nomenclature.

Most carbohydrates are found in nature as part of polysaccharides. **Homopolysaccharides** contain one type of monosaccharide, while **heteropolysaccharides** contain different types of monosaccharide units. **Starch** (plants) and **glycogen** (animals) are composed of glucose units (Fig. 4.11 and Fig. 4.12). They are important storage molecules for chemical energy. Starch consists of **Amylose**

and **Amylopectin** (Fig. 4.12). **Dextran** is a storage polysaccharide found in yeast and bacteria. Cell walls of bacteria and plants are often constructed from polysaccharides.

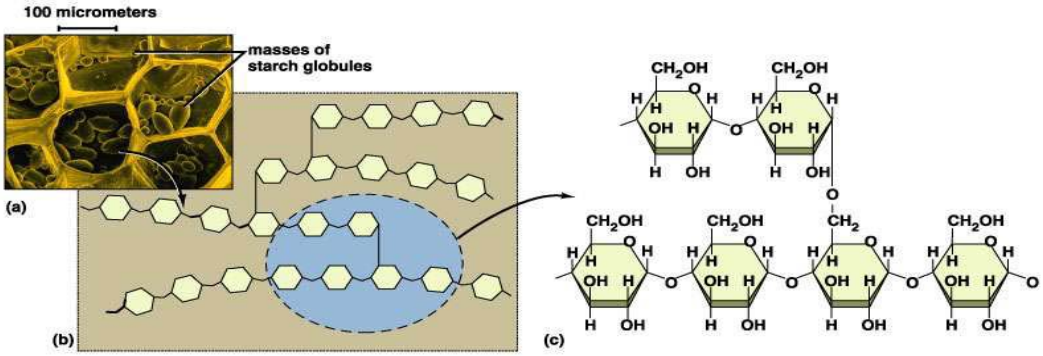



**Fig. 4.11.** Starch, an important storage molecule for chemical energy.

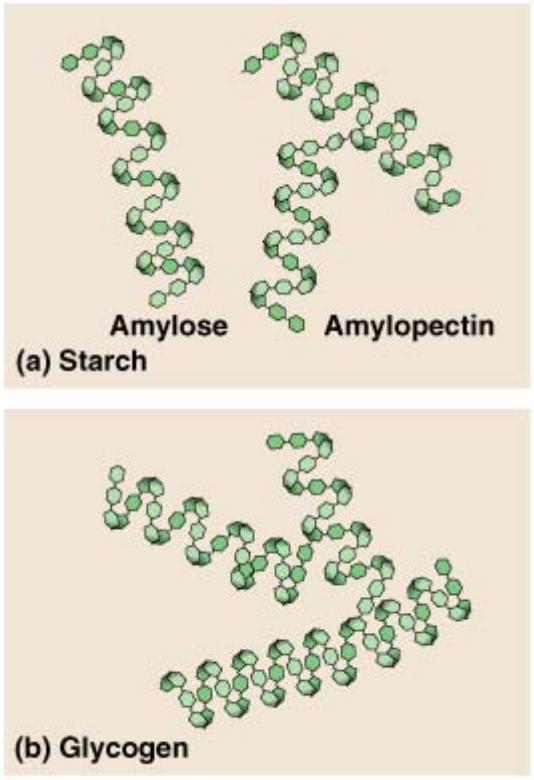



**Fig. 4.12.** Starch (plants) and Glycogen (animals) are homopolysaccharides composed of glucose.

Examples include **cellulose** (most abundant organic compound in the biosphere) and **chitin**. Polysaccharides attached to the cell membrane mediate important information such as cell cell recognition.

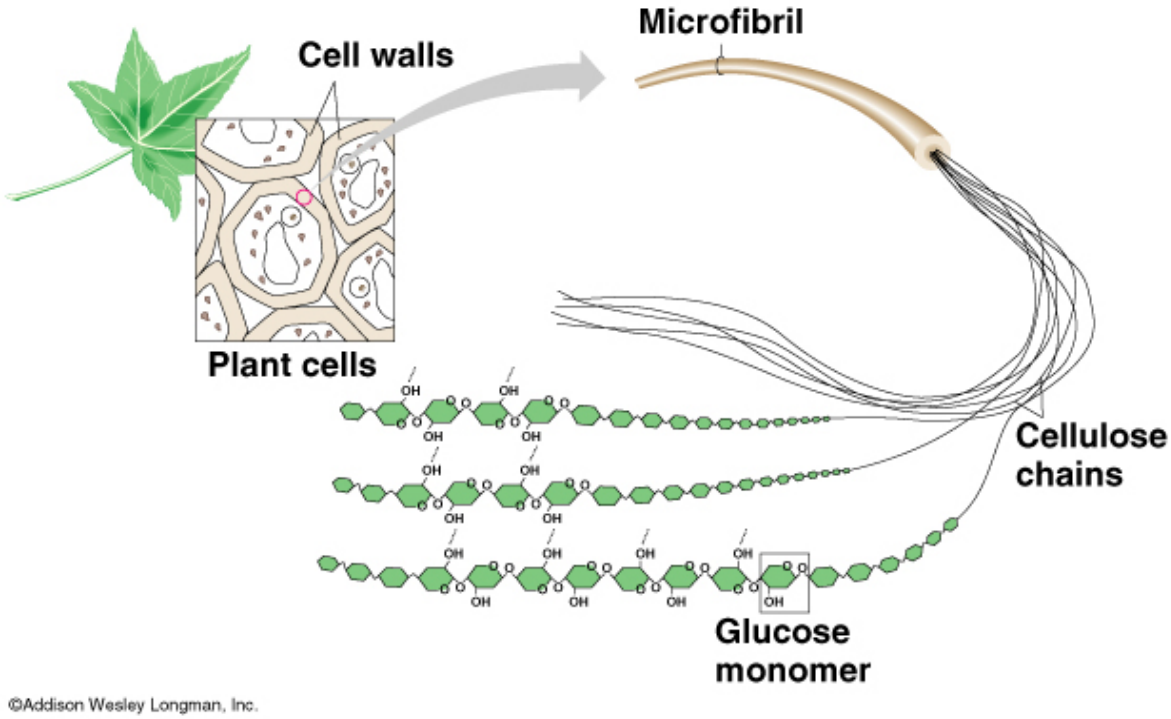



**Fig. 4.13.** Cellulose, component of cell walls of plants.

## *Glycosaminoglycans*

Glycosaminoglycans are found on the cell surface and in the extracellular matrix of animals. Many glycosaminoglycanes are made of disaccharide repeating units containing a derivative of an amino sugar, such as glucoseamine or galactoseamine. At least one of the sugars carries a negatively charged carboxylate of sulfate group. An example is the anticoagulant **heparin** (Fig. 4.14). Cartilage is **proteoglycan** (Fig. 4.15.), i.e. a protein containing one or more covalently linked glycosaminoglycans.

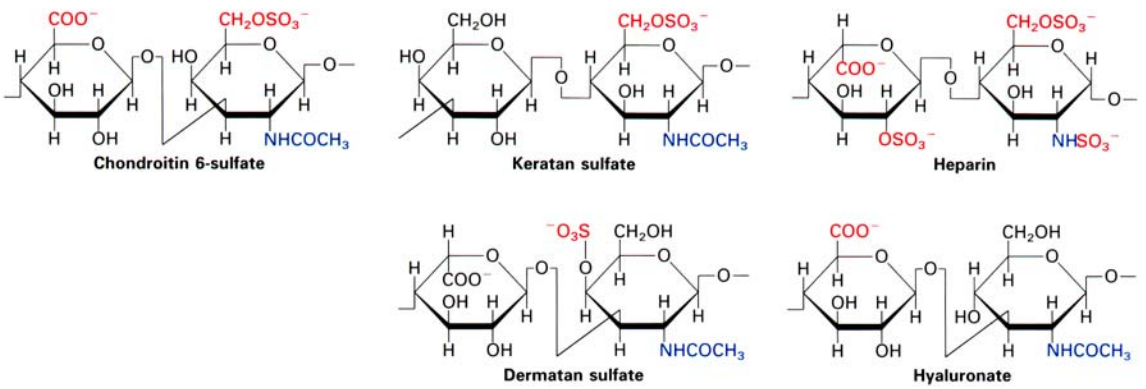


**Fig. 4.14.** Structural formulas of the repeating disaccharide units of some major glycosaminoglycans.

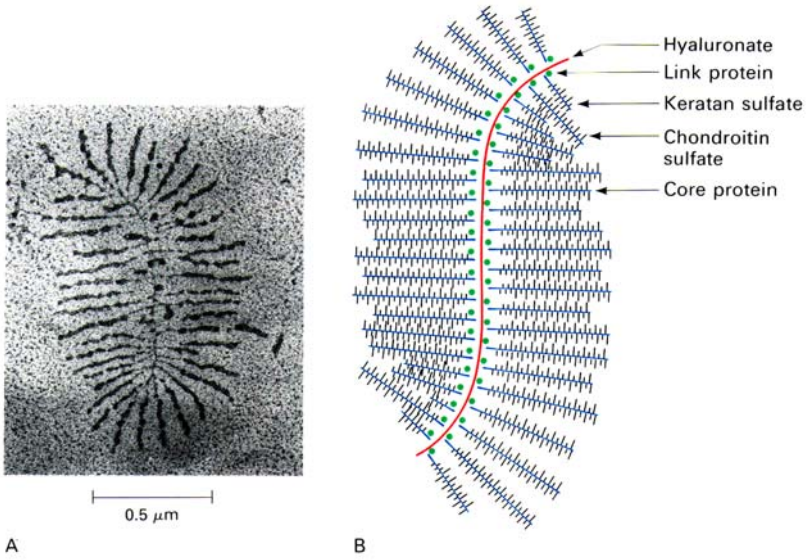



**Fig. 4.15.** (A) Electron micrograph of a proteoglycan aggregate from bovine fetal epiphyseal cartilage. Proteoglycan monomers arise laterally at regular intervals from the opposite sides of an elongated central filament of hyaluronate (Courtesy of Dr. Joseph Buckwalter and Dr. Lawrence Rosenberg). (B) Schematic diagram.

## *Glycoproteins*

Many of the proteins produced by eukaryotic cells contain sugar residues. These proteins are called glycoproteins. In some cases the sugars account for 80 % of the molecular mass of the protein (e.g. in the case of some mucoproteins). The sugars are attached either to the N-atom in the side chain of asparagine (N-glycosilation site) (Fig. 4.16.) or to the O-atom in the side chain of either a serine or a threonine residue (O-glycosilation site).

N-linked oligosaccharides contain a common pentasaccharide core consisting of three mannose and two GlcNAc units. Additional units may be attached to this core in a great variety and the structure of the final carbohydrate entities is complex and species and cell dependent. The final structure is rich in information and decisive for many biological processes. Bacteria do not glycosilate proteins at all. This has certain consequences for molecular biotechnology (recombinant proteins), since the carbohydrate residues have important functions (signaling).

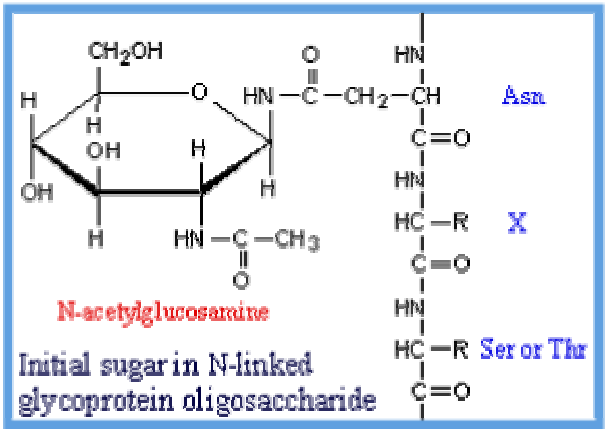



**Fig. 4.16.** N-acetylglucosamine linked to an asparagines residue by an glycosidic bond.

Glycosilation takes place in the **endoplasmatic reticulum** (N-linked sugars, core pentasaccharides are added to certain asparagines) and in the **golgi apparatus** (O-linked sugars, further modifications of N-linked sugars). The golgi apparatus is divided into several compartments, which contain different enzymes. The enzyme composition of the golgi apparatus may also differ considerably according to cell type. Carbohydrate modification (terminal glycosilation) takes place in several of the golgi compartments. In the cis-golgi, e.g., three mannose units are removed from all proteins destined to be excreted or membrane proteins. In the medial golgi two more mannose units are removed and two GlcNAc and a fucose are added in some cells. Finally, in the trans golgi, another GlcNAc is added followed by galactose and **sialic acid**.

**The final structure of a glycoprotein can therefore be said to depend on the genetic code but also on the cell type in which it is expressed** (Fig. 4.17.).

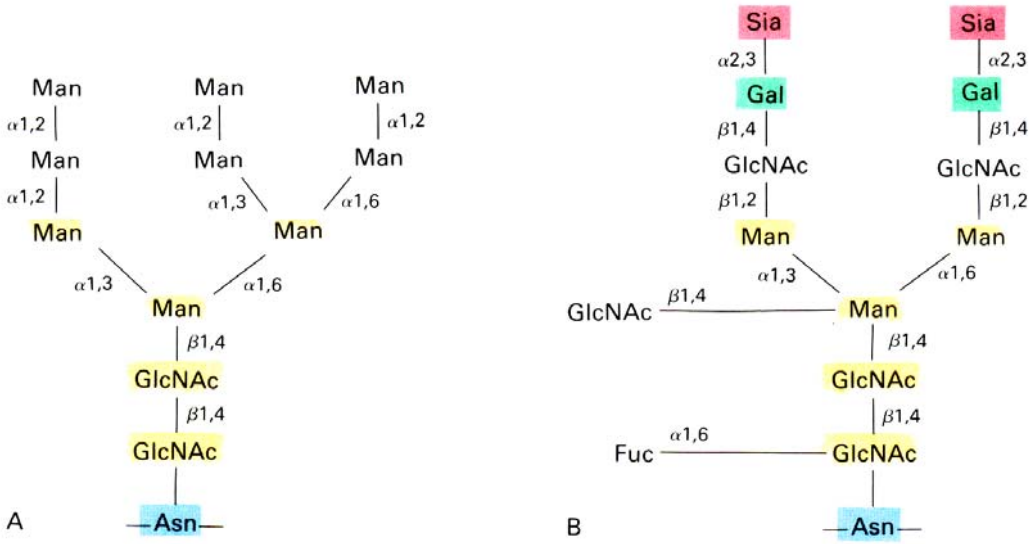


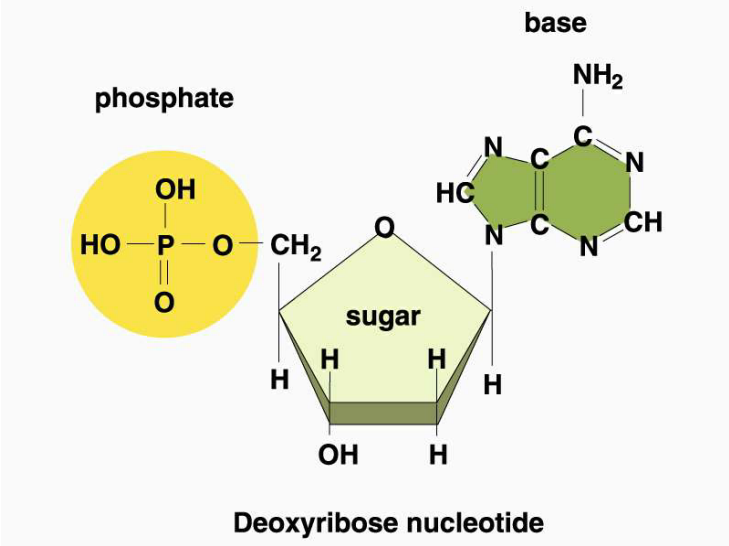

**Fig. 4.17.** Structure of an asparagine-linked oligosaccharide unit in (A) human immunoglobulin and (B) porcine thyroglobulin.

# 5. POLYNUCLEOTIDES

## *Polynucleotides*

A **nucleotide** (Fig. 5.1.) consists of a heterocyclic base, a pentose (ribose, desoxyribose) unit and a phosphate group. The bases can have a purine (Adenine, Guanine) or pyrimidine structure (Uracil, Cytosine, Thymine). The sugar is linked to the base via an □-N-glycosidic bond. All important nucleosides occur inside the cell not only as monophosphates but also as di- and triphosphates. A molecule consisting of only the base and the sugar unit is called a **nucleoside** (Fig. 5.2.). Nucleotides are perhaps best known as the monomeric units of DNA and RNA, the carriers of genetic information in the cell. However, nucleotides also have many other important biological functions.

**Fig.5.1**.



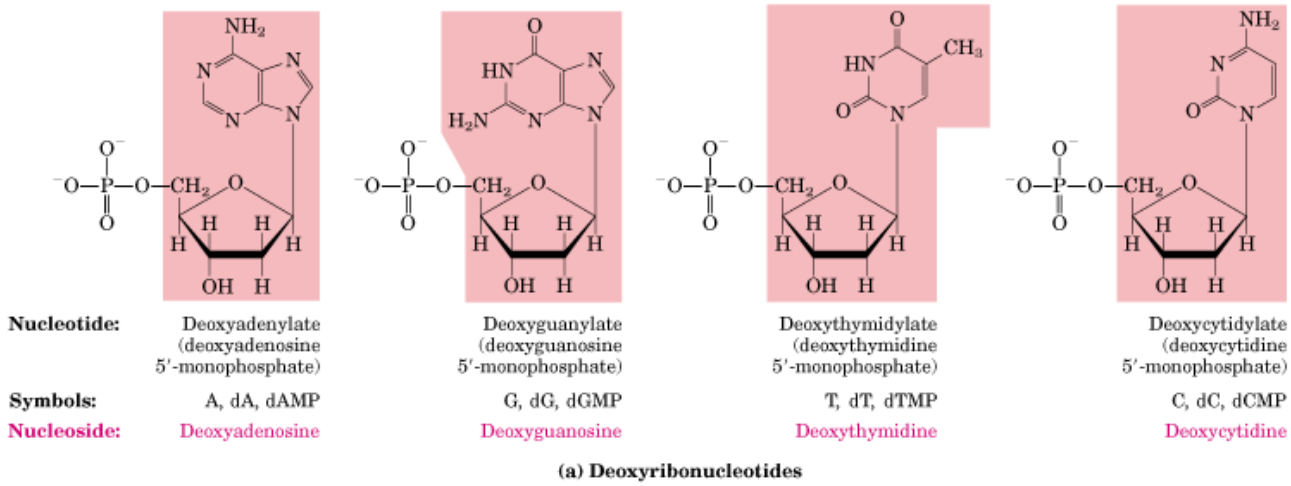




**Fig. 5.2.** Structure of the major deoxyribonucleosides.

Adenosine triphosphate (**ATP**) is perhaps the most important one amongst these molecules, but others such as GTP, CTP and UTP are also known. ATP transfers phosphate and pyrophosphate groups in many biochemical reactions and is the most important form of "chemical energy" in the metabolism (Fig. 5.3.). The other triphosphates are used for similar purposes in specific metabolic reactions.

**FAD** and **NAD(P)** (Fig. 5.4.) provide reductive power to many biochemical reactions. Cyclic AMP (**cAMP**) (Fig. 5.5.) and cyclic GMP (**cGMP**) are important intracellular messenger (G-protein receptor, hormone interaction) in signal cascades triggered by hormones such as insulin (primary messenger). Some coenzymes such as Coenzyme A (**CoA**) (Fig. 5.6.) are also derivatives of nucleotides and nucleosides. Coenzyme A (CoA) is the universal carrier of acyl groups in biochemistry, e.g. in fatty acid oxidation, fatty acid synthesis, pyruvate oxidation and other biological acylations. Vitamin $B_6$ also contains a nucleotide group.

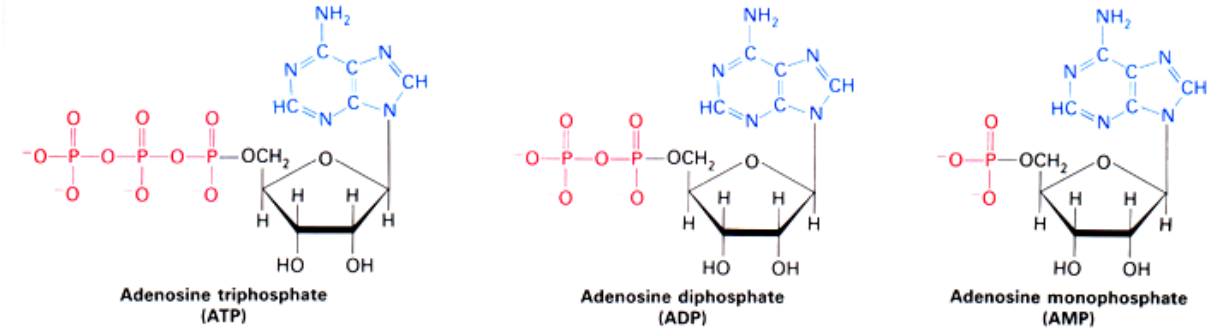
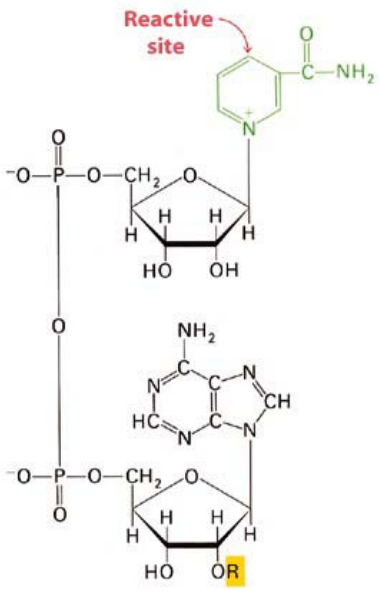


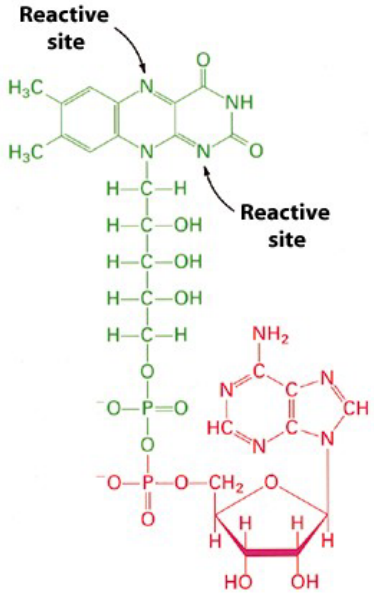

**Fig. 5.3.** Structures of ATP, ADP and AMP (Adenosine consists of adenine linked to ribose).

**Fig. 5.4.**

**N**icotinamide **A**denine **D**inucleotide (**NAD**)




**F**lavin **A**denine **D**inucleotide (**FAD**)




## *Ribozymes*

**More recently it has been discovered that RNA can have enzymatic functions ("ribozymes").**

This multitude of biological functions is by some scientists interpreted as a relict of the "**RNA-World**", i.e. a time when most biological functions were in fact carried out by nucleotides. By the

time the even more versatile proteins had started to perform some of these function (especially catalysis), certain nucleotide-based cofactors had already adapted extremely well to their function. Consequently, the new enzymes continued to make use of these factors, even though the "conventional" interaction through base pairing was no longer possible.

**Fig. 5.5.**

**A**denosine 3',5'-**cyclic M**ono**P**hospate (**cyclic AMP, cAMP**)

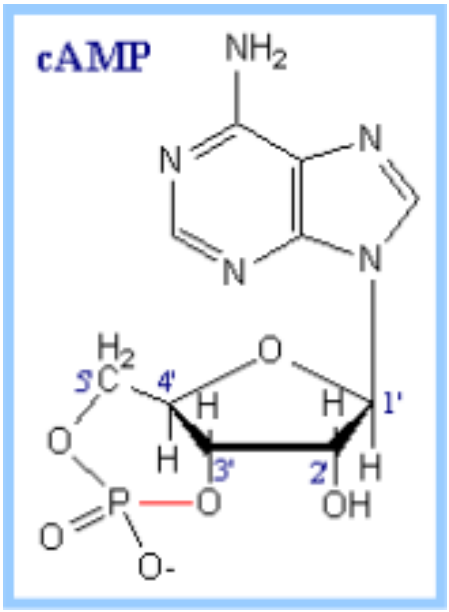
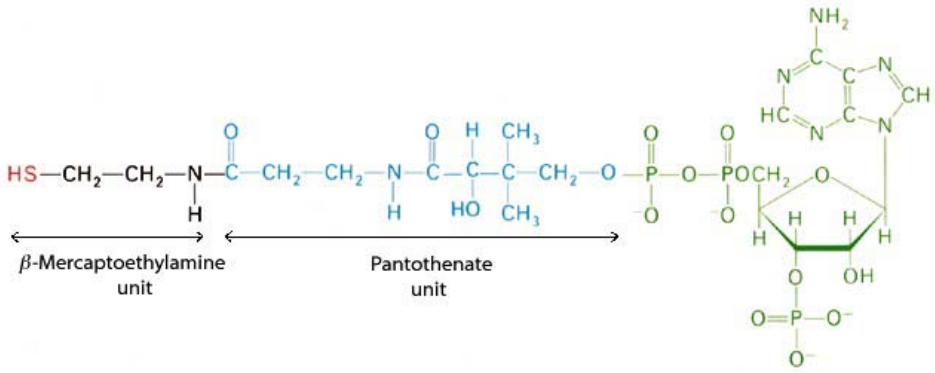



**Fig. 5.6.**

**Coenzyme A**




Especially transfer (tRNA) and ribosomal (rRNA) RNA show some interesting analogies to modern proteins. Even though they make use of only four standard building blocks, there is extensive processing and modification of both the sequence and the chemistry of the monomeric units after the initial transcription (Fig. 5.7.). Controlled methylation is, e.g., used to enhance hydrophobicity or to prevent certain unwanted base pairings. In addition, these RNA molecules carry the biologically

relevant information not in their sequence but in their *structure*, which is well defined and important for the function (ribozymes) (Fig. 5.8.).

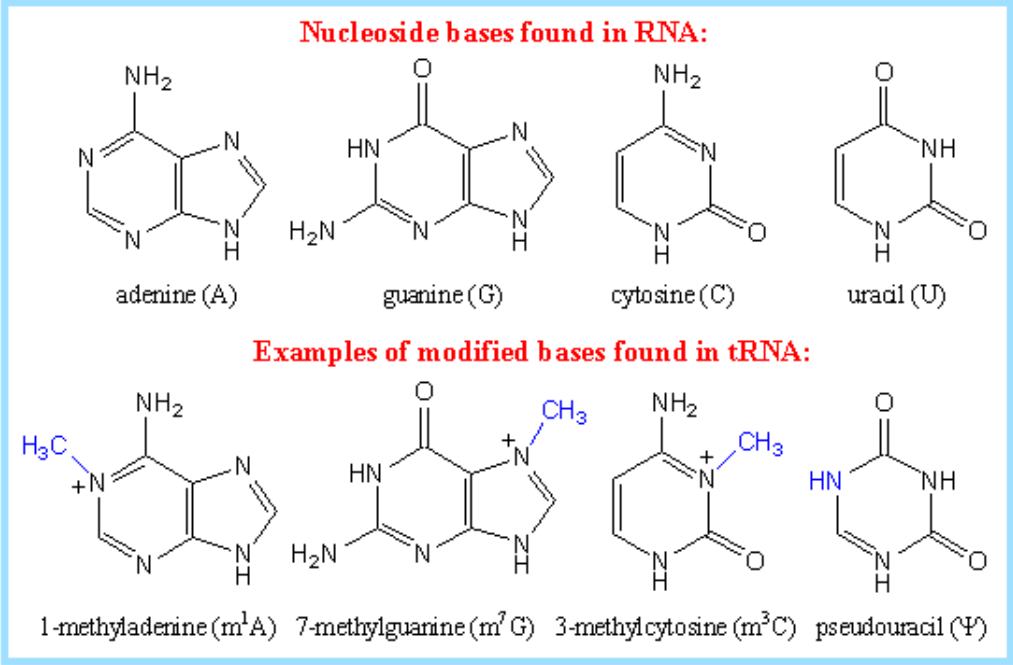



**Fig. 5.7.** Structures of modified basis found in tRNA.

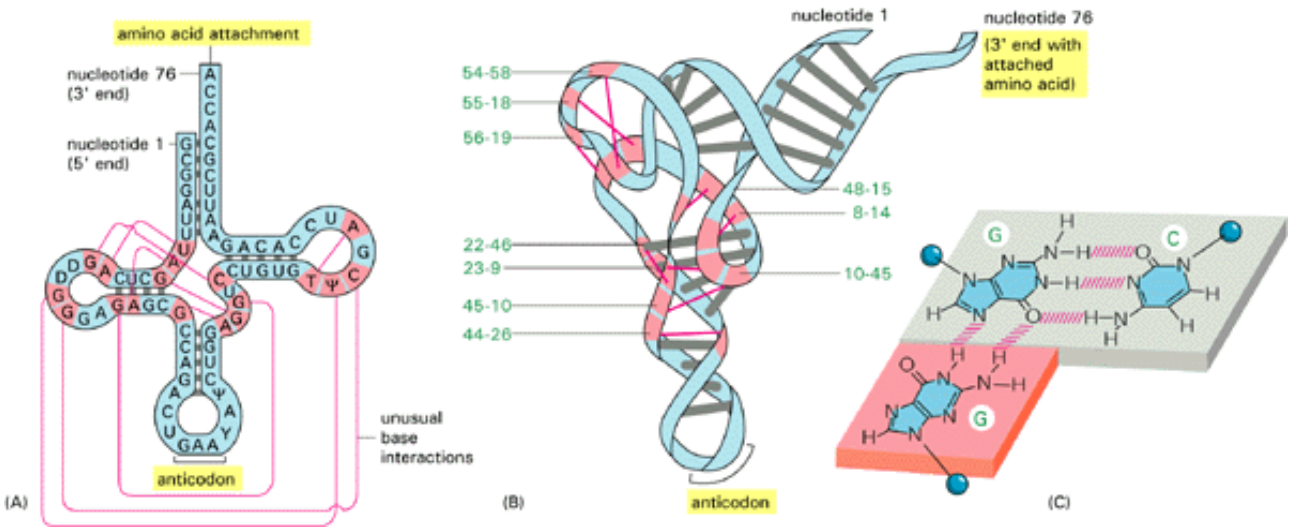



**Fig. 5.8.** Phenylalanine tRNA of yeast. (A) The molecule is drawn with a cloverleaf shape to show the complementary base-pairing that occurs in the helical regions of the molecule. (B) The actual shape of the molecule, based on x-ray diffraction analysis, is shown schematically. (C) One of the unusual base-pair interactions. Here one base forms hydrogen-bond interactions with two others ; several such “base triples” help fold up this tRNA molecule.

The ribosomal RNA molecules of the ciliated protozoan *Tetrahymena* are an example of RNA molecules that function as catalysts. They are initially synthesized as a large precursor from which

one of the rRNAs is produced by an RNA-splicing reaction. The surprise came with the discovery that his splicing can occur in vitro in the absence of protein. It was subsequently shown that the intron sequence itself has an enzymelike catalytic activity that carries out the two-step reaction illustrated in Figure 5.9.

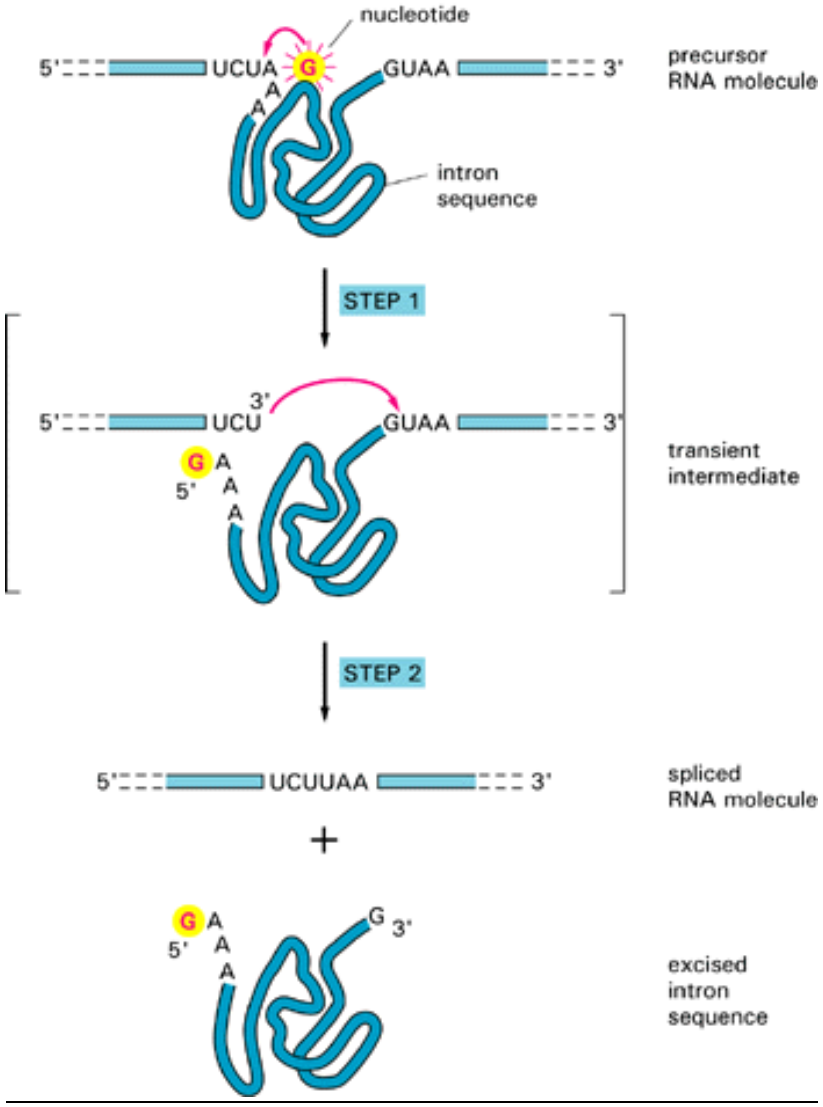



**Fig. 5.9.** A self-splicing RNA molecule. The diagram shows the self-splicing reaction in which an intron sequence catalyzes its own excision from a Tetrahymena ribosomal RNA molecule.

## *The DNA/RNA molecule*

The difficulty that geneticists had in accepting DNA as the substance of genes is understandable, considering the simplicity of its chemistry. A DNA chain is a long, unbranched polymer composed of only four types of subunits (Fig. 5.10.). These are the deoxyribonucleotides containing the bases adenine (A), cytosine (C), guanine (G), and thymine (T). The nucleotides are linked together by covalent phospho-diester bonds that join the 5' carbon of one deoxyribose group to the

3’ carbon of the next. The four kinds of bases are attached to this repetitive sugar-phosphate chain almost like four kinds of beads strung on a necklace.

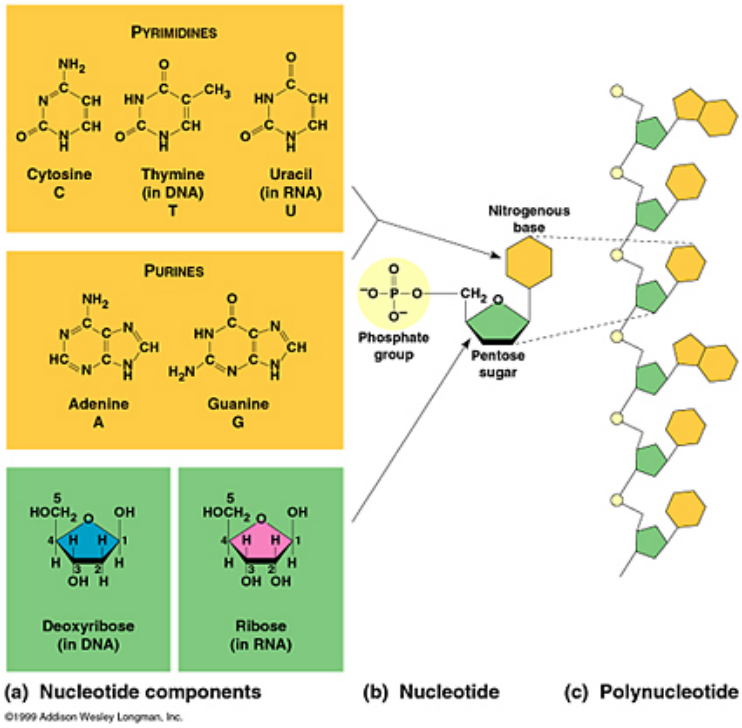



**Fig. 5.10.** Structure of the DNA/RNA molecule.

For sterical reasons, base pairing in DNA (and RNA) molecules occurs always between a purine and a pyrimidine base (Fig. 5.11.).

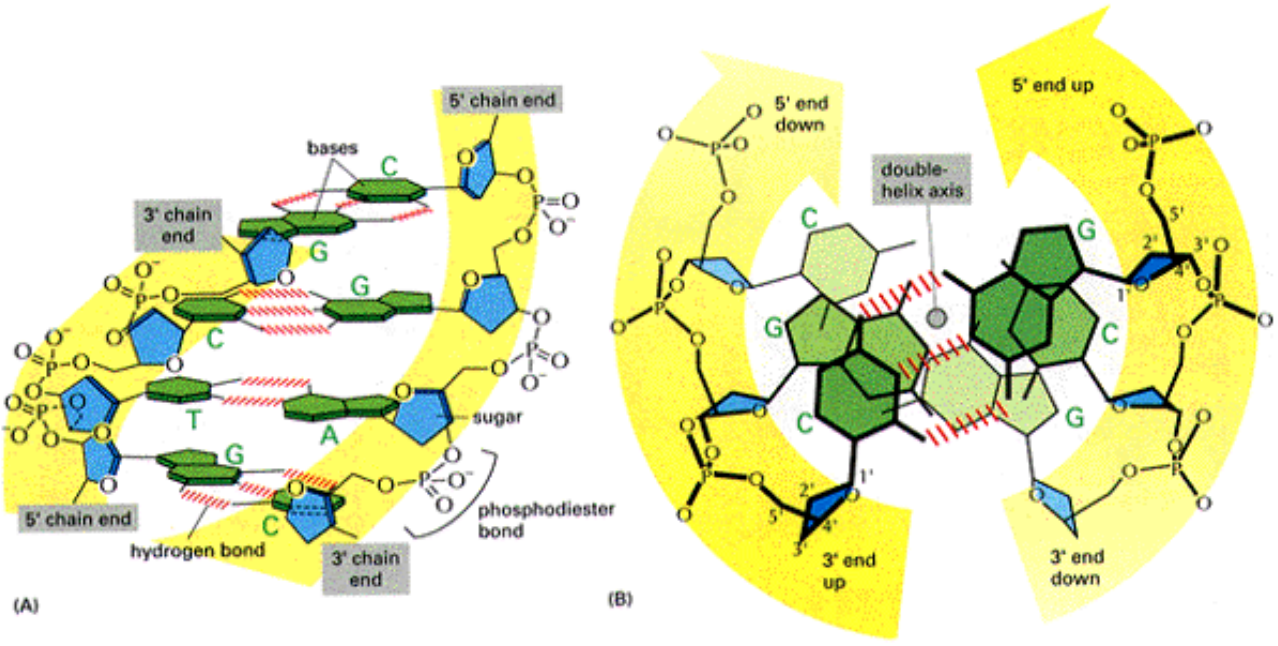



**Fig. 5.11. The DNA double helix.** (A) A short section of the helix viewed from its side. (B) The helix viewed from an end. Note that the two DNA strands run in opposite directions and that each base pair is held together by either two or three hydrogen bonds.

## *The genetic code*

The rules by which the nucleotide sequence of a gene is translated into the amino acid sequence of a protein, the so-called genetic code, were deciphered in the early 1960s. The sequence of nucleotides in the mRNA molecule that acts as an intermediate was found to be read in serial order in groups of three. Each triplet of nucleotides, called a codon, specifies one amino acid. The genetic code is almost universal. All sixty-four codons have been deciphered (Fig. 5.12.). Sixty-one triplets correspond to particular amino acids, whereas three code for chain termination. Because there are twenty amino acids and sixty-one triplets that code for them, it is evident that the code is highly redundant (*degenerate)*.

| 1st position (5' end) | 2nd position U | 2nd position C | 2nd position A | 2nd position G | 3rd position (3' end) |
|---|---|---|---|---|---|
| U | Phe | Ser | Tyr | Cys | U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | STOP | STOP | A |
| | Leu | Ser | STOP | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

**Fig. 5.12.** The genetic code.

Note: Given the position of the bases in a codon, it is possible to find the corresponding amino acid. For example, the codon 5'AUG3' on mRNA specifies methionine, whereas CAU specifies histidine. UAA, UAG and UGA are termination signals. AUG is part of the initiation signal, in addition to coding for methionines in the polypeptide chain.

For a given sequence there are three possible reading frames (Fig. 5.13.). In almost every case only one of these reading frames will produce a functional protein. Since there are no punctuation signals except at the beginning and end of the RNA message, the reading frame is set at the initiation of the translation process and is maintained thereafter.

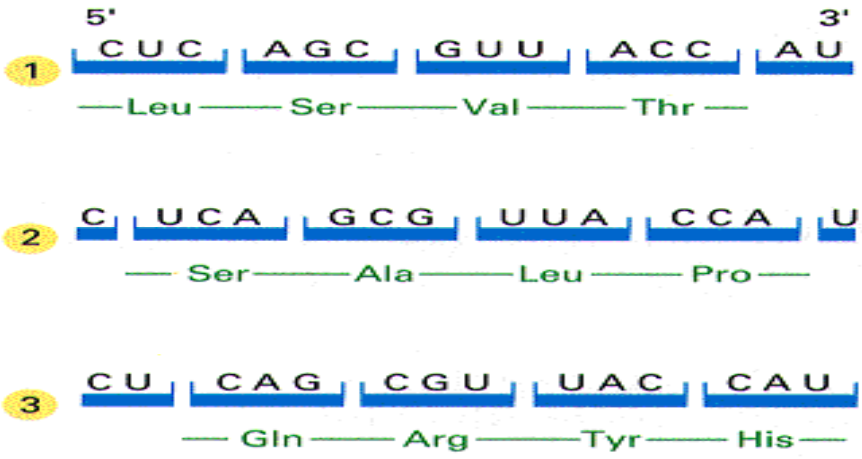



**Fig. 5.10.** The three possible reading frames for a given sequence. In the process of translating a nucleotide sequence (blue) into an amino acid sequence (green), the sequence of nucleotides in an mRNA molecule is read from the 5' to the 3' end in sequential sets of three nucleotides. In principle, therefore, the same RNA sequence can specify three completely different amino acid sequences, depending on the "reading frame".

## *Flow of the genetic information*

The synthesis of proteins involves copying specific regions of DNA (the genes) into polynucleotides of a chemically and functionally different type known as ribonucleic acid, or RNA (Fig. 5.10), a process known as DNA transcription. RNA retains all of the information of the DNA sequence from which it was copied, as well as the base-pairing properties of DNA. RNA molecules are relatively short compared to DNA molecules since they are copied from a limited region of the DNA – enough to make one or a few proteins. RNA transcripts that direct the synthesis of protein molecules are called messenger RNA (mRNA) molecules, while other RNA transcripts serve as transfer RNA (tRNAs) or form the RNA component of ribosomes (rRNA) or smaller ribonucleoprotein particles.

The codon recognition process by which genetic information is transferred from mRNA via tRNA to protein depends on the same type of base-pair interactions that mediate the transfer of genetic information from DNA to DNA and from DNA to RNA (Fig. 5.14.). But the mechanism of ordering the tRNA molecules on the mRNA are complicated and require a ribosome, a complex of more than 50 different proteins associated with several structural RNA molecules (rRNAs). Each ribosome is a large protein-synthesizing machine on which tRNA molecules position themselves so as to read the genetic message encoded in an mRNA molecule.

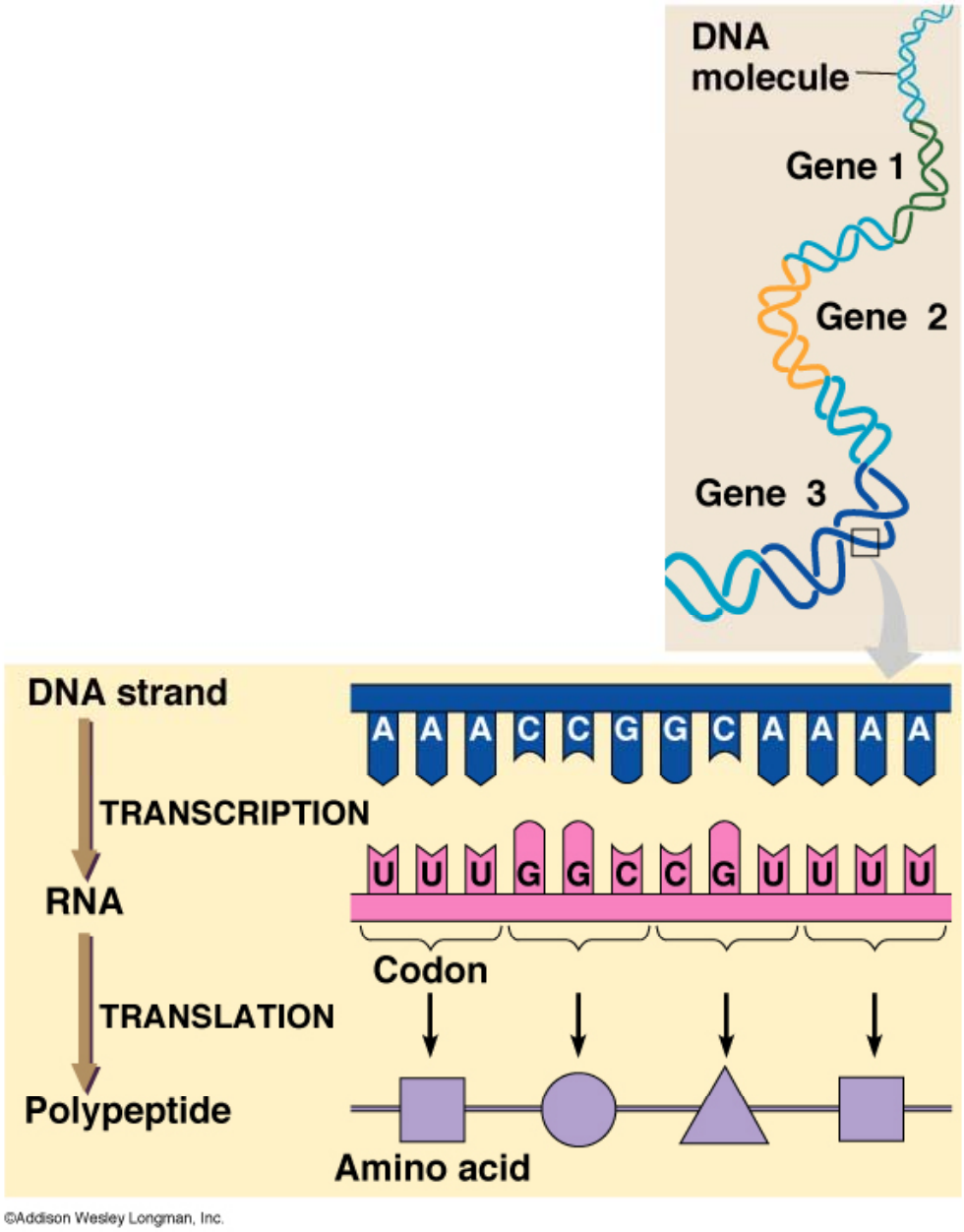


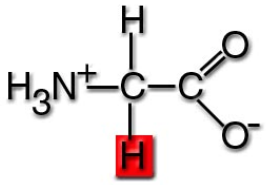

**Fig. 5.14.** Diagram of the flow of the genetic information.

# 6. RESUME

## LES PROTEINES

Les protéines jouent des rôles clés dans pratiquement tous les processus biologiques. Presque tous les catalyseurs des systèmes biologiques sont des protéines appelées *enzymes*. Les protéines déterminent ainsi *le profil des transformations chimiques* dans les cellules. Elles médient un grand nombre d'autres fonctions telles que le transport et la mise en réserve, les mouvements coordonnés, le support mécanique, la protection immune, l'excitabilité, l'intégration du métabolisme et le contrôle de la croissance et de la différenciation.

Les unités structurales de base des protéines sont des *acides aminés*.

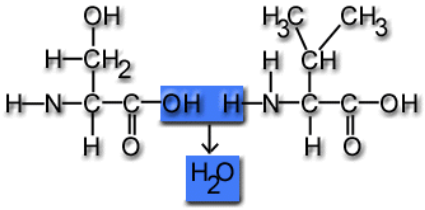



GLYCINE (Gly)

Toutes les protéines de toutes les espèces – des bactéries à l'Homme – sont construites à partir du même groupe de 20 acides aminés. Les chaînes latérales de ces modules élémentaires peuvent être groupées comme suit :

1. chaînes latérales aliphatiques - glycine, alanine, valine, leucine, isoleucine et proline
2. chaînes latérales aliphatiques hydroxylées – sérine et thréonine
3. chaînes latérales aromatiques – phénylalanine, tyrosine et tryptophane
4. chaînes latérales basiques – lysine, arginine et histidine
5. chaînes latérales acides – acide aspartique et acide glutamique
6. chaînes latérales amides – asparagine et glutamine
7. chaînes latérales soufrées – cystéine et méthionine

De nombreux acides aminés sont unis par des *liaisons peptidiques* pour former une chaîne polypeptidique.



-réaction de polymérisation des acides aminés

Une liaison peptidique unit le groupe α-carboxyle d'un acide aminé à la fonction amine du suivant. Des liaisons disulfures croisées peuvent être formées par des résidus cystéines. Les protéines ont une séquence unique d'acides aminés qui est déterminée génétiquement. Le déterminant critique de la fonction biologique d'une protéine est sa *conformation* qui est la disposition tridimensionnelle des atomes d'une molécule. Trois conformations régulièrement répétées de chaînes polypeptidiques sont connues : l'hélice α et le feuillet plissé β .

La séquence des acides aminés d'une protéine détermine sa séquence tridimensionnelle. La forte tendance des résidus hydrophobes à fuir l'eau favorise le reploiement des protéines solubles. Les protéines sont stabilisées aussi bien par de nombreuses liaisons hydrogènes et interactions de van der Waals que par des interactions hydrophobes.

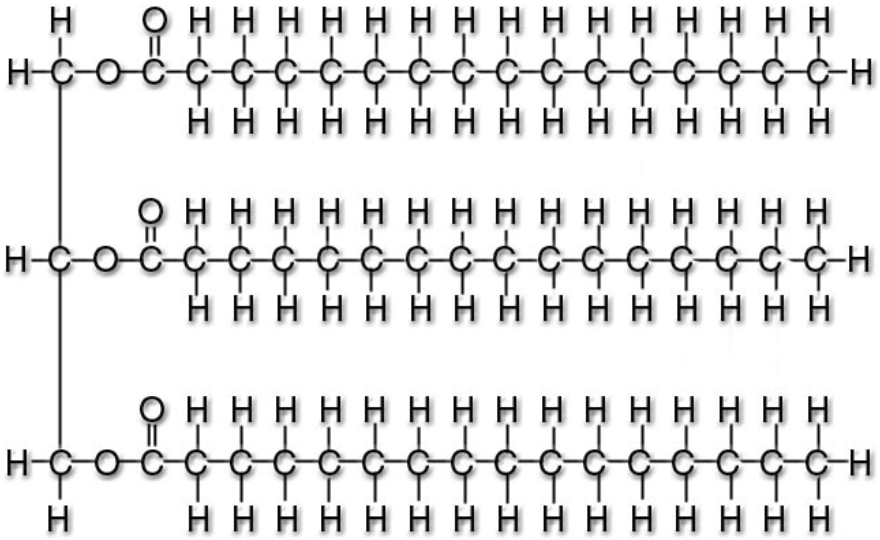
La *structure* d'une protéine se présente comme suit :

- structure primaire : séquence d' acides aminés.
- structure secondaire : interactions entre les liaisons peptidiques. La protéine se replie en hélice
- structure tertiaire interaction entre les différents radicaux. La protéine a une forme tridimensionnelle.
- structure quaternaire : interaction entre les différentes unités de la protéine, ce qui va lui donner une conformation définitive ( globulaire, spiralée etc.)

## LES LIPIDES

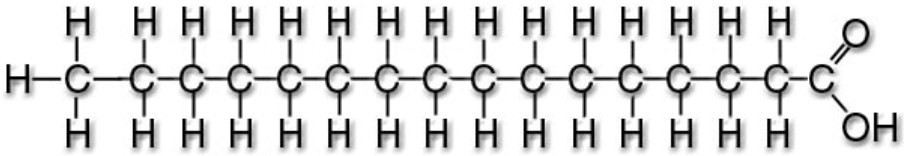
On peut distinguer 3 catégories de lipides : les triglycérides (graisses neutres), les phospholipides et les stéroïdes.

*Les lipides simples*: les triglycérides: graisses neutres (réserves énergétiques à long terme) formées d'un Glycérol + 3 Acides gras.



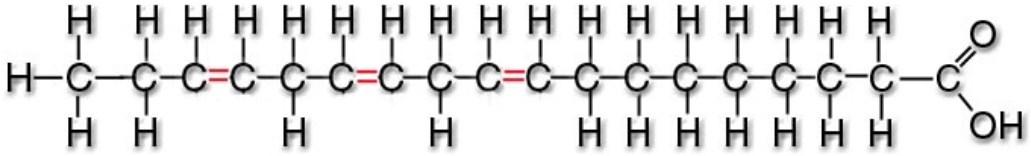
Triaglycérol (graisse)

Les acides gras peuvent être *saturés ou insaturés*, identiques ou différents. Il n' y a pas de double liaison entre les atomes de carbone d'un acide gras saturé.



Acide gras saturé $CH_3(CH_2)_{14}COOH$

ex. d'acide gras saturé : C16= acide palmitique

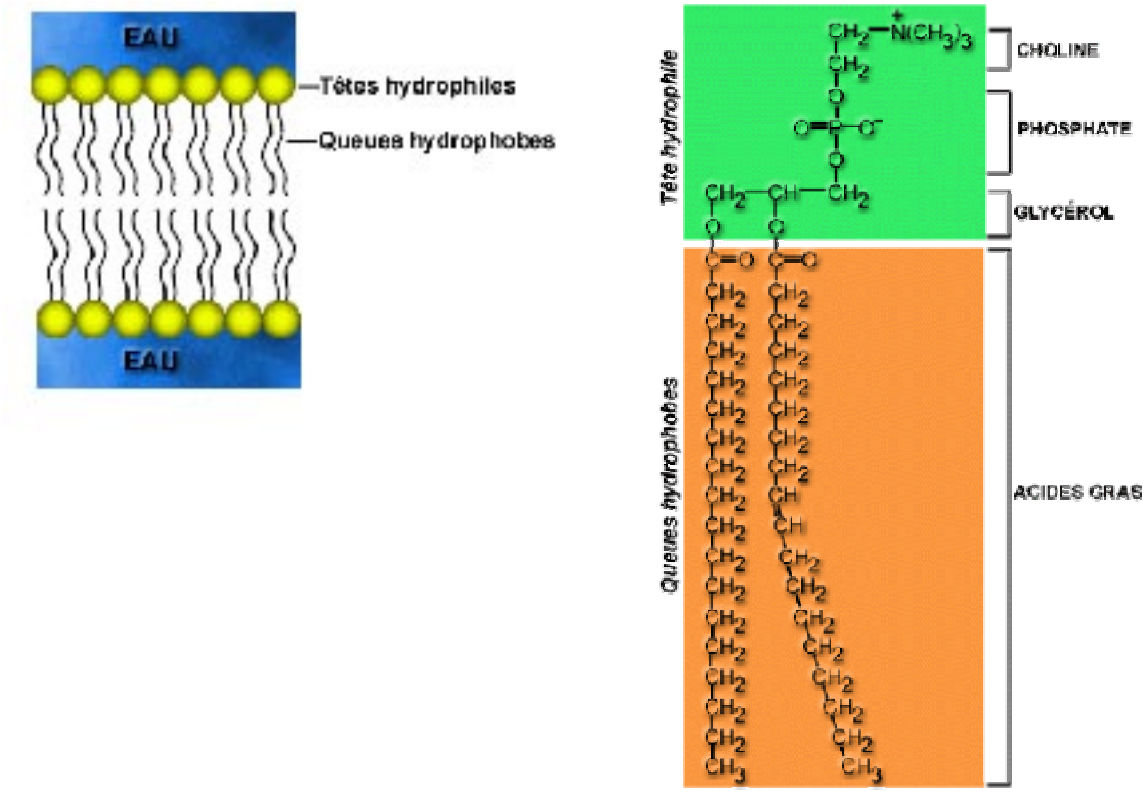


Acide gras insaturé $CH_3CH_2(CH{=}CHCH_2)_3(CH_2)_6COOH$

ex. d'acide gras insaturé : C18= acide linoléique (une double liaison)

Les acides gras insaturés (AGI) se retrouvent surtout chez les végétaux alors que les acides gras saturés (AGS) se rencontrent dans les graisses animales (lard, saindoux....) . Il semble que les AGS soient à l'origine de certains troubles cardio-vasculaires tel l'athérosclérose, car ils ont tendance à favoriser la formation de cholestérol qui peut se déposer dans la paroi des artères et les rendre moins flexibles, voire même les boucher. D'où le conseil de consommer des graisses végétales plutôt que des graisses animales.

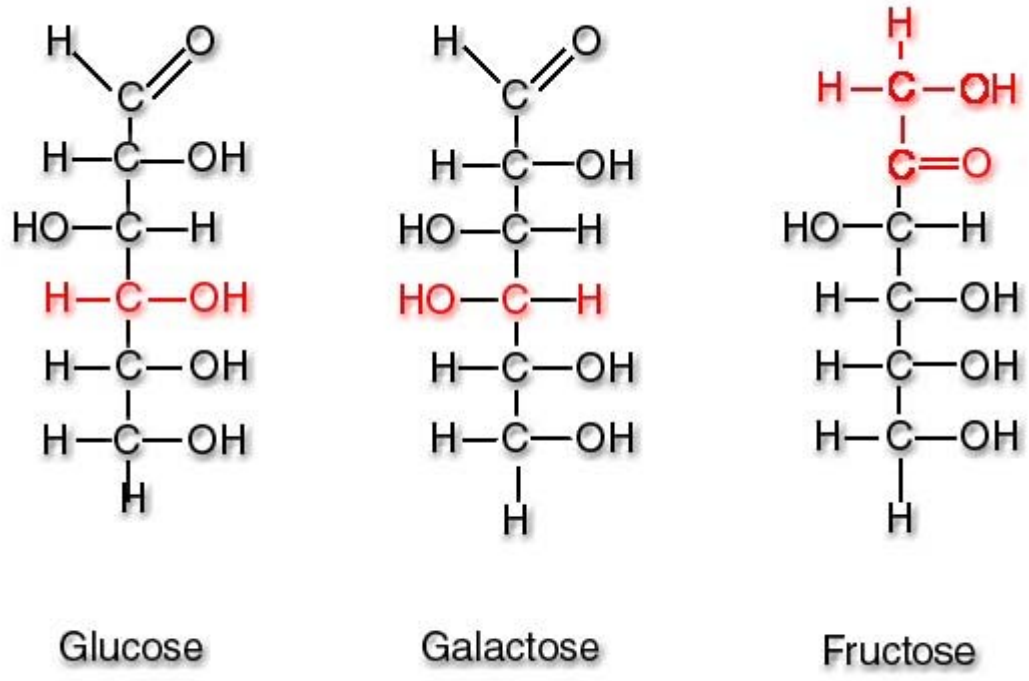
*Les phospholipides* : Ont la même structure que les triglycérides, mais en plus un acide gras porte un groupement phosphaté à un bout. Ces groupements peuvent s'ioniser et être chargés d'où un bout polaire très soluble dans l'eau (bout hydrophile) et un bout non polaire insoluble dans l'eau (bout hydrophobe). Ce sont des molécules amphipatiques. Les phospholipides sont des constituants fondamentaux de la *membrane cellulaire*. Si on met de l'huile dans l'eau, les 2 corps ne se mélangent pas. Les molécules de lipides vont former des micelles avec leur bout hydrophobe en position centrale.




*Les stéroïdes* : Ce sont des lipides complexes, comportant un noyau stérol. Ce sont des composés  importants sur le plan biologique car ils font partie de la membrane cellulaire dont ils assurent la cohésion. De plus, le noyau stérol se retrouve dans le cholestérol et dans un très grand nombre d'hormones : œstrogène, progestérone, testostérone, cortisol, etc.

## LES GLUCIDES

Les *glucides* sont des aldéhydes ou des cétones possédant deux ou plusieurs groupes hydroxyles. Les *aldoses* sont des glucides avec une fonction aldéhyde (comme le glycéraldéhyde et le glucose), tandis que les *cétoses* contiennent un groupe cétone (comme la dihydroxyacétone et le fructose).

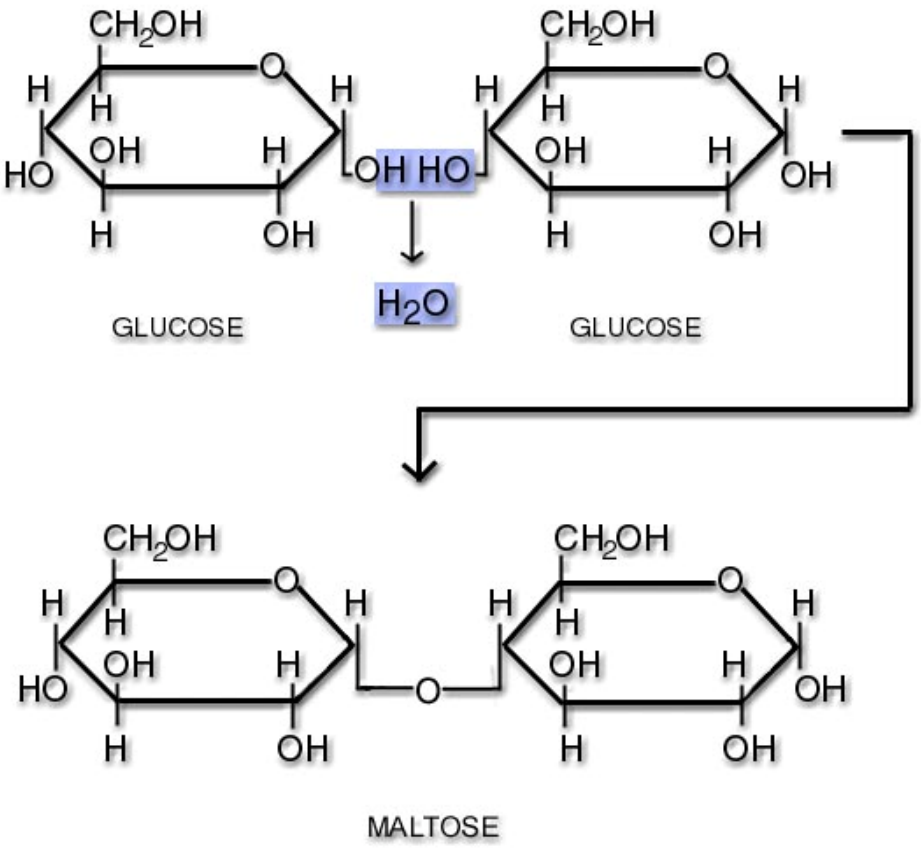



**Ces 3 glucides sont des isomères et ont des propriétés différentes**

Un sucre appartient à la *série D* lorsque la configuration absolue de son carbone asymétrique le plus éloigné du groupe aldéhyde ou cétone est la même que celle du D-glycéraldéhyde. Presque tous les glucides naturels appartiennent à la série D. L'aldéhyde en C-1 de la forme linéaire du glucose réagit avec le groupe hydroxyle en C-5 pour former *un cycle pyranose* à 6 atomes. Le groupe cétone en C-2 de la forme linéaire du fructose réagit avec le groupe hydroxyle en C-5 pour former *un cycle furanose* à 5 atomes. Un nouveau centre d'asymétrie est créé au niveau de l'atome de carbone anomérique lors de la formation de ces cycles. Le groupe hydroxyl lié au carbone anomérique est au-dessous du plan du cycle dans *l'anomère α,* alors qu'il est au-dessus du cycle dans *l'anomère β.* Tous les atomes des cycles ne sont pas dans le même plan. Au contraire, les cycles pyranoses adoptent habituellement la *conformation chaise* et les cycles furanoses la *conformation enveloppe*.

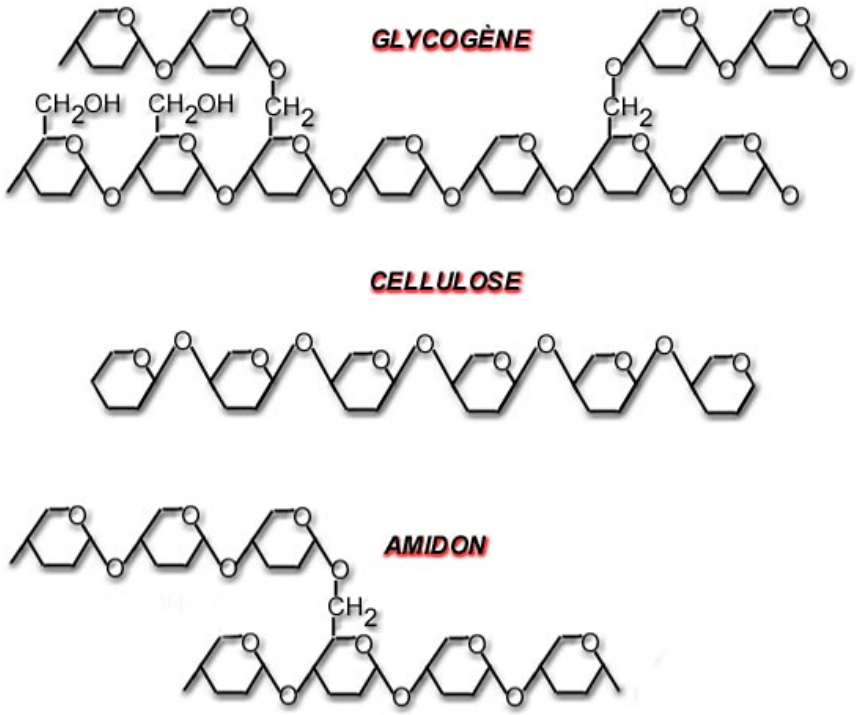
Les glucides sont unis à des alcools ou des amines par des *liaisons glycosidiques* par l'intermédiaire de l'atome de carbone anomérique. Par exemple, les liaisons *O-glycosidiques* lient des oses à d'autres oses dans les disaccharides et les polysaccharides. Des liaisons *N-glycosidiques* lient des oses à des purines ou des pyrimidines dans les nucléotides, le RNA ou le DNA.

Le saccharose, le lactose et le maltose sont des *disaccharides*.




-réaction de condensation du maltose

Le *saccharose* (le sucre de table) est formé de glucose et de fructose unis par une liaison α-glycosidique entre les carbones anomériques. Le *lactose* (du lait) est formé de galactose uni au glucose par une liaison β-1,4. Le *maltose* (de l'amidon) est formé de deux glucoses unis par une liaison α-1,4. L'amidon est une forme polymérique du glucose rencontrée chez les végétaux ; le *glycogène* joue un rôle identique chez les animaux. La plupart des unités glucoses de l'amidon et du glycogène sont unies par des liaisons α-1,4. Le glycogène a plus de points de ramifications formées par des liaison α-1,6 que n'en a l'amidon, ce qui rend le glycogène plus soluble. La *cellulose*, polymère structural majeur des parois cellulaires végétales, est constituée d'unités glucoses unies par des liaison β-1,4. Ces liaisons β donnent de longues chaînes non ramifiées qui forment des fibres très résistantes à la traction. Par contre, les liaisons α dans l'amidon et le glycogène conduisent à des hélices ouvertes, ce qui les adapte bien à leur rôle de réserves d'énergie mobilisables.




Les surfaces cellulaires et les matrices extracellulaires, chez les animaux, contiennent des polymères d'unités disaccharidiques répétitives appelés *glycosaminoglycanes*. Dans chaque unité répétitive, l'une des composantes est un dérivé de la glucosamine ou de la galactosamine. Ces glucides hautement anioniques ont une forte densité de groupes carboxyles ou sulfates. Les protéines portant des glycosaminoglycanes liés par covalence sont appelées *protéoglycanes*.

Les unités oligosaccharidiques des protéines membranaires intégrales sont liées soit à l'atome d'oxygène de la chaîne latérale des résidus sérines ou thréonines, soit à l'atome d'azote de l'amide de la chaîne latérale des résidus asparagines. Les oligosaccharides N-liés contiennent un noyau commun constitué de trois résidus mannoses et deux résidus N-acétylglucosamines. D'autres sucres s'attachent à ce noyau pour former des structures diverses. Les glucides sont importants dans l'adressage moléculaire et dans la reconnaissance entre cellules.

## *LES NUCLEOTIDES*

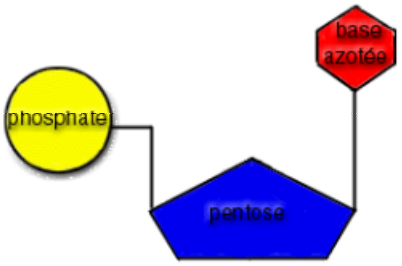
*Structure de base* : le nucléotide : Base -Sucre- Phosphore

- Base:

bases puriques : adénine (A), guanine (G),

bases pyrimidiques : cytosine (C), uracyl (U), thymine (T)

- Sucre : ribose ou un désoxyribose
- Phosphore : groupement phosphate




NUCLÉOTIDE

*Types d'acides nucléiques* :

**L'acide désoxyribonucléique : ADN**

L' ADN est *un polymère d'unités désoxyribonucléotidiques*. Un *nucléotide* est constitué d'une base azotée, d'un ose et d'un ou de plusieurs groupes phosphates. L'ose d'un désoxyribosecléotide est le *désoxyribose*. Le base azotée est un dérivé de la *purine* ou de la *pyrimidine*. Les purines du DNA sont l'adénine (A) et la guanine (G), et les pyrimidines sont la thymine (T) et la cytosine (C). Les bases des molécules de DNA recèlent *l'information génétique*, tandis que les oses et les groupes phosphates jouent un rôle structural.

- Base : adénine, guanine, cytosine, thymine (A ||||||||| T G ||||||||| C)
- Sucre : désoxyribose
- Groupement phosphaté

L' ADN est formé de 2 brins enroulés en double hélice selon le modèle de Watson et Crick. Les brins sont unis comme le zip d'une fermeture éclair par des bases selon le mode A-T ou G-C.

**L'acide ribonucléique : ARN** formé d'un seul brin.

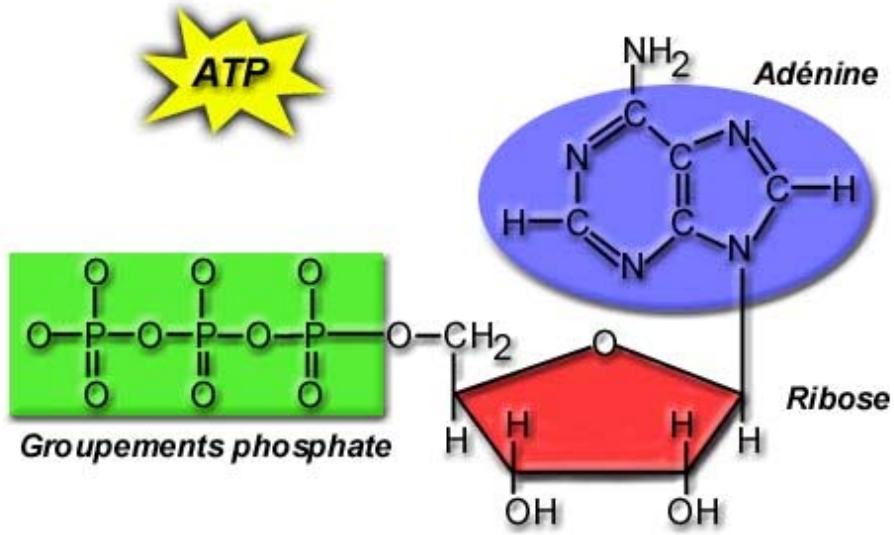
- Base : adénine, guanine, cytosine, uracile (A ||||||| U G ||||||| C)
- Sucre : ribose
- Groupement phosphaté

L'ADN est le support de l'information génétique (gènes et chromosomes) et est avec l'ARN, à la base de la synthèse des protéines.

**L'ATP= Adénosine Triphosphate** (forme d'énergie utilisable par la cellule).

- Adénine + ribose =Adénosine
- 3 groupements phosphates : triphosphate

*ATP --> ADP + P + Énergie*




Un *nucléoside* est constitué d'une base purique ou pyrimidique liée à un ose. Un *nucléotide* est un ester phosphate de nucléoside. Par exemple, *le désoxy-adénosine 5'-triphosphate (dA*TP) est un précurseur activé dans la synthèse de l'ADN.

# PROBLEMES

1.

(a) La tropologie, protéine musculaire de 70 kDa, est constituée de deux brins en hélice α enroulés l'un autour de l'autre. Quelle est la longueur de la molécule ?

(b) Supposez qu'un segment de 40 résidus d'une protéine se reploie en une structure β antiparallèle à deux brins avec un coude en épingle à cheveux de quatre résidus. Quelle est la plus longue dimension de ce motif ?

2.

(a) Classifier les aminoacides suivantes : Glu, Cys et Gly.

(b) Former un tripeptide (glutinative) avec les aminoacides.

(c) La glycine est un résidu aminoacide très conservé dans l'évolution des protéines. Pourquoi ?

3. Expliquer l'origine du terme Carbohydrate (nom donné par les Anglo-saxons aux glucides)?

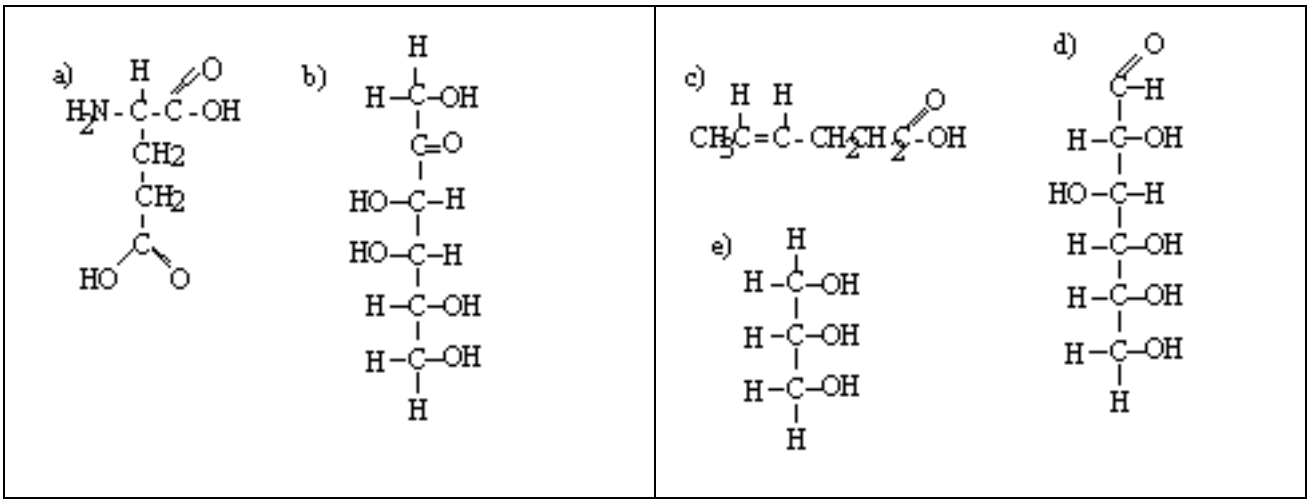
4. Donner la projection Fischer du D-glucose, D-fructose, D-galactose, D-ribose et 2-désoxy-D-ribose?

5. Donner la formule α-D-glucose et β-D-glucose ?

6. Donner la formule du lactose (disaccharide formé par du galactose uni au glucose par une liaison glycosique β-1,4) ?

## Autoévaluation

1. Les questions 1 à 7 se réfèrent aux diagrammes suivants:




Laquelle des molécules serait le monomère d'une protéine?

- a
- b
- c
- d
- e

2 Quel est le groupement radical de l'acide aminé illustré?

- $CH_3$
- $CH_2$-$CH_2$-COOH
- COOH
- C=OH
- $NH_2$

3 Quelle molécule est le glucose?

- a
- b
- c
- d
- e

4 Laquelle est un acide gras?

- a
- b
- c
- d
- e

5 Quelle molécule serait la moins soluble dans l'eau?

- a
- b
- c
- d
- e

6 Les molécules b et d sont des exemples de?

- isomères
- isotopes
- acides organiques
- substances hydrophobes
- dissacharides

7 Quelles sont les deux molécules qui seraient le produit de l'hydrolyse d'une graisse neutre?

- a et b
- b et c
- a et e
- c et e
- c et d

8 Lequel des énoncés suivants est vrai en ce qui concerne les phospholipides?

- Ils sont tous hydrophobes
- Ils sont tous hydrophiles
- Ils sont en partie hydrophile et en partie hydrophobes
- Ils sont tous composés d'acides aminés polaires
- Aucun énoncé est vrai

9 La structure suivante est celle d'une molécule composée de C, N et H. Les atomes d'H n'ont pas été inclus. Quelle serait la formule chimique complète pour cette molécule?

- C5H11N
- C5H9N
- C5H10N
- C5H12N
- C5H6N

10 Des composés ayant le même contenu atomique mais ayant une structure différente et des propriétés chimiques différentes sont des?

- Isotopes
- Isomères
- Composés ioniques
- Composés polaires covalents
- Composés non-polaires covalents

11 Si deux sucres à 5 carbones sont combinés pour former un disaccharide, combien de molécules d'eau seront formées?

- 5
- 2
- 3
- aucune
- 1

12 Les mammifères entreposent le sucre sous forme de?

- amidon
- cellulose
- kératine
- glycogène
- glucose

13 La conformation d'une protéine dépend de plusieurs types de liaisons et de l'interaction entre des groupes fonctionels. Quel type de liaison n’est pas affecté lorsqu'une protéine est dénaturée par la chaleur?

- liaison hydrogène
- Liaison ionique
- Interaction hydrophobe
- Liaison peptidique
- Aucune de ces réponses

14 Le processus par lequel une protéine est faite à partir de l'information provenant de l'ARNm est la:?

- transciption
- conscription
- dénaturation
- contraction musculaire
- traduction

15 L'hémoglobine a une structure?

- Primaire
- Secondaire
- Tertiaire
- Quaternaire
- Linéaire seulement