Chapter 3
Chemical Composition of the Cell

Life on Earth evolved in the water, and all life still depends on water. At least 80% of the mass of living organisms is water, and almost all the chemical reactions of life take place in **aqueous solution.** The other chemicals that make up living things are mostly **organic macromolecules** belonging to the four groups **proteins, nucleic acids, carbohydrates** or **lipids.** These macromolecules are made up from specific **monomers** as shown in the table below. Between them these four groups make up 93% of the dry mass of living organisms, the remaining 7% comprising small organic molecules (like vitamins) and inorganic ions.

Table.3.1. Chemical Composition of the Cell

Group name	monomers	polymers	% dry mass
Proteins	amino acids	polypeptides	50
nucleic acids	nucleotides	polynucleotides	18
carbohydrates	monosaccharides	polysaccharides	15

Group name	components	largest unit	% dry mass
lipids	fatty acids + glycerol	Triglycerides	10

3.1. Carbohydrates

Carbohydrates have the general molecular formula $(CH_2O)n$, and thus were once thought to represent "hydrated carbon". However, the arrangement of atoms in carbohydrates has little to do with water molecules. Starch and cellulose are two common carbohydrates. Both are **macromolecules** with molecular weights in the hundreds of thousands. Both are **polymers** (hence "**polysaccharides**"); that is, each is built from repeating units, **monomers**, much as a chain is built from its links. The monomers of both starch and cellulose are the same: units of the sugar **glucose**.

3.1.1. Monosaccharides

Three common sugars share the same molecular formula: $C_6H_{12}O_6$. Because of their six carbon atoms, each is a **hexose**.

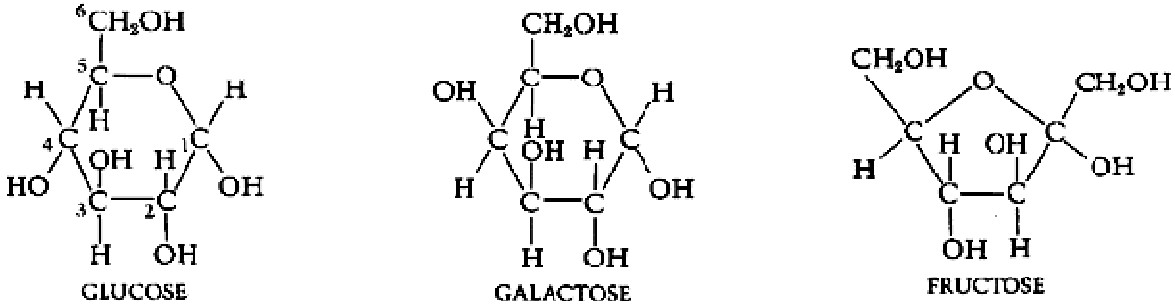


Fig.3.1. Examples of Monosaccharides

They are:

1. **Glucose**, "blood sugar", the immediate source of energy for cellular respiration
2. **Galactose**, a sugar in milk (and yogurt), and
3. **Fructose**, a sugar found in honey.

Although all three share the same molecular formula ($C_6H_{12}O_6$), the arrangement of atoms differs in each case. Substances such as these three, which have identical molecular formulas but different structural formulas, are known as **structural isomers**. Glucose, galactose, and fructose are "single" sugars or **monosaccharides**. Two monosaccharides can be linked together to form a "double" sugar or **disaccharide**.

3.1.2. Disaccharides

Three common disaccharides:

1. **Sucrose** — common table sugar = glucose + fructose
2. **Lactose** — major sugar in milk = glucose + galactose
3. **Maltose** — product of starch digestion = glucose + glucose

Although the process of linking the two monomers is rather complex, the end result in each case is the loss of a hydrogen atom (H) from one of the monosaccharides and a hydroxyl group (OH) from the other. The resulting linkage between the sugars is called a **glycosidic bond**.

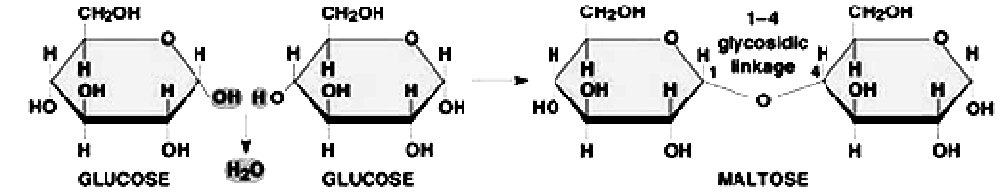


Fig.3.2. Formation of Carbohydrates through glycosidic bonds

The molecular formula of each of these disaccharides is

$$C_{12}H_{22}O_{11} = 2\ C_6H_{12}O_6 - H_2O$$

All sugars are very soluble in water because of their many hydroxyl groups. Although not as concentrated a fuel as fats, sugars are the most important source of energy for many cells. Carbohydrates provide the bulk of the calories (4 kcal/gram) in most diets, and starches provide the bulk of that. Starches are polysaccharides.

3.1.3. Polysaccharides

Polysaccharides are polymers composed of several molecules of monosaccharides. Common examples include Starch, glycogen, cellulose etc.

(a) Starches

Starches are polymers of glucose. Two types are found:

1. **Amylose** consists of linear, unbranched chains of several hundred glucose residues (units). The glucose residues are linked by a **glycosidic bond** between their #1 and #4 carbon atoms.
2. **Amylopectin** differs from amylose in being highly branched. At approximately every thirtieth residue along the chain, a short side chain is attached by a glycosidic bond to the #6 carbon atom (the carbon above the ring). The total number of glucose residues in a molecule of amylopectin is several thousand.

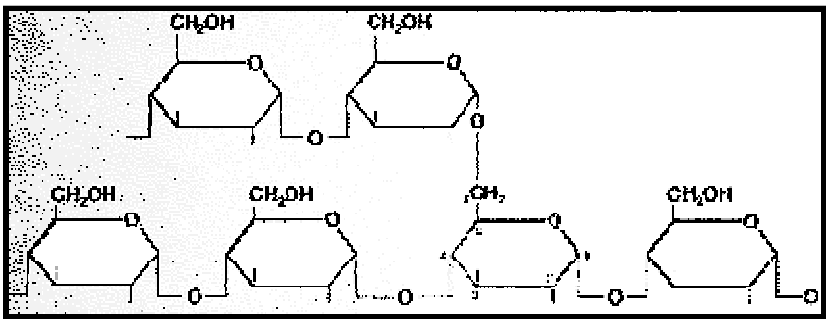


Fig.3.3a. Structure of amylose

Starches are insoluble in water and thus can serve as storage depots of glucose. Plants convert excess glucose into starch for storage. The image shows starch grains (lightly stained with iodine) in the cells of the white potato. Rice, wheat, and corn are also major sources of starch in the human diet. Before starches can enter (or leave) cells, they must be digested. The hydrolysis of starch is done by amylases. With the aid of an **amylase** (such as pancreatic amylase), water molecules enter at the 1 -> 4 linkages, breaking the chain and eventually producing a mixture of **glucose** and **maltose**. A different amylase is needed to break the 1 -> 6 bonds of amylopectin.

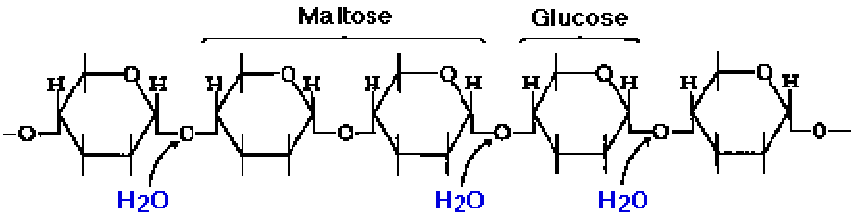


Fig.3.3b. Structure of amylopectin

(b) Glycogen

Animals store excess glucose by polymerizing it to form **glycogen**. The structure of glycogen is similar to that of amylopectin, although the branches in glycogen tend to be shorter and more frequent.

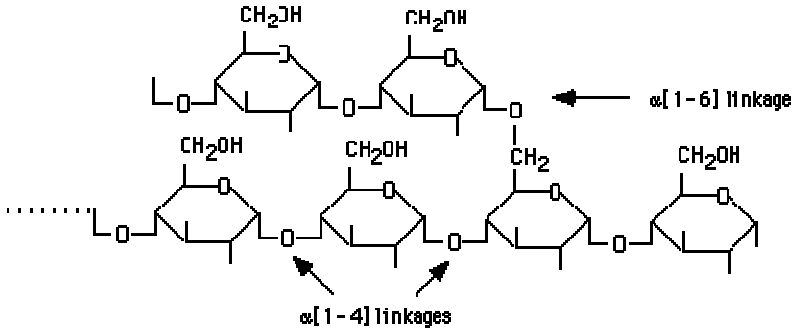


Fig.3.3c. Structure of glycogen

Glycogen is broken back down into glucose when energy is needed (a process called glycogenolysis). In **glycogenolysis**, phosphate groups — not water — break the 1 -> 4 linkages and the phosphate group must then be removed so that glucose can leave the cell. The liver and skeletal muscle are major depots of glycogen. There is some evidence that intense exercise and a high-carbohydrate diet ("carbo-loading") can increase the reserves of glycogen in the muscles and thus may help marathoners work their muscles somewhat longer and harder than otherwise. But for most of us, carbon loading leads to increased deposits of fat.

(c) Cellulose

Cellulose is probably the single most abundant organic molecule in the biosphere. It is the major structural material of which plants are made. Wood is largely cellulose while cotton and paper are almost pure cellulose.

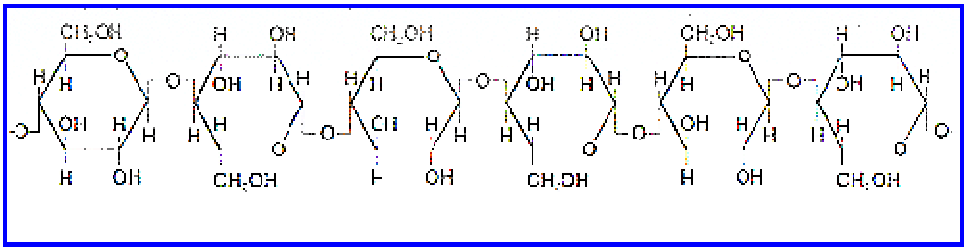


Fig.3.3d. Structure of Cellulose

Like starch, cellulose is a polysaccharide with glucose as its monomer. However, cellulose differs profoundly from starch in its properties.

1. Because of the orientation of the **glycosidic bonds** linking the glucose residues, the rings of glucose are arranged in a flip-flop manner. This produces a long, straight, rigid molecule.
2. There are no side chains in **cellulose** as there are in **starch**. The absence of side chains allows these linear molecules to lie close together.
3. Because of the many **-OH groups**, as well as the oxygen atom in the ring, there are many opportunities for **hydrogen bonds** to form between adjacent chains.

The result is a series of stiff, elongated fibrils — the perfect material for building the cell walls of plants.

3.2. Lipids

Fat molecules are made up of four parts: a molecule of **glycerol** (on the right) and three molecules of **fatty acids**. Each fatty acid consists of a **hydrocarbon chain** with a **carboxyl group** at one end. The **glycerol molecul**e has three **hydroxyl groups**, each able to interact with the carboxyl group of a fatty acid. Removal of a water molecule at each of the three positions forms a **triglyceride**. The three fatty acids in a single fat molecule may be all alike (as shown here for **tristearin**) or they may be different.

They may contain as few as 4 carbon atoms or as many as 24. Because fatty acids are synthesized from fragments containing two carbon atoms, the number of carbon atoms in the chain is almost always an even number. In animal fats, 16-carbon (palmitic acid) and 18-carbon (stearic acid - shown here) fatty acids are the most common

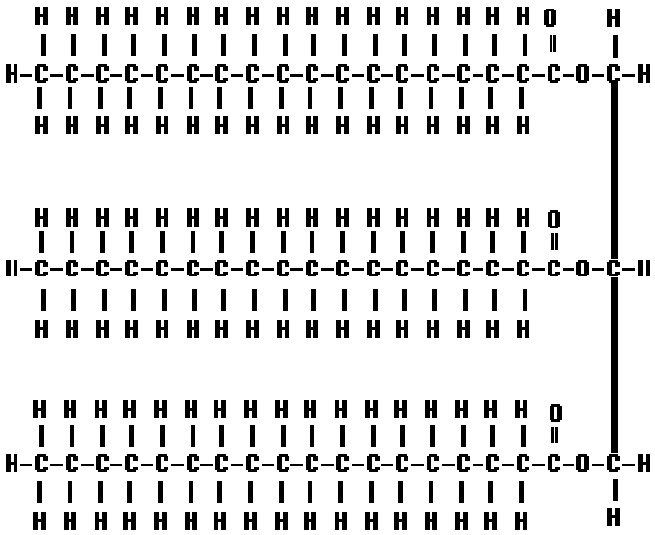


Fig.3.4a. Tristearin molecule

3.2.1. Unsaturated Fats

Some fatty acids have one or more double bonds between their carbon atoms. They are called unsaturated because they could hold more hydrogen atoms than they do. **Monounsaturated** fats have a single double bond in their fatty acids. **Polyunsaturated** fats, such as **trilinolein** shown here, have two or more.

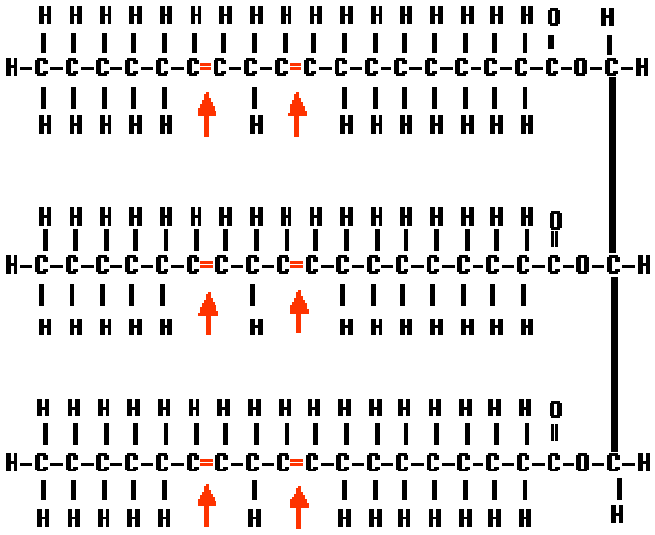


Fig.3.4b. Trilimolen molecule

Double bonds are rigid and those in natural fats introduce a kink in the molecule. This prevents the fatty acids from packing close together and as a result, unsaturated fats have a lower melting point than do saturated fats. Because most of them are liquid at room temperature, we call them **oils**. Corn oil, canola oil, cottonseed oil, peanut oil, and olive oil are common examples.

3.2.2. Trans Fatty Acids

Because the most abundant (and least expensive) source of fat is from plant oils but many cooking applications, particularly baked products, need solid fats the food industry uses **hydrogenated** oils for things like **shortening** and **margarine**.

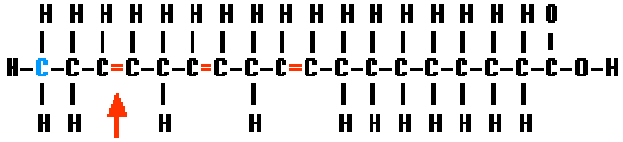


Fig.3.4c. Linolenic Acid, an omega-3 Fatty Acid

In **hydrogenation**, plant oils are exposed to hydrogen at a high temperature and in the presence of a catalyst. Two things result: Some double bonds are converted into single bonds and other double bonds are converted from **cis** to **trans** configuration. Both these effects straighten out the molecules so they can lie closer together and become solid rather than liquid.

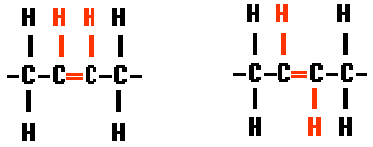


Fig.3.4d. Cis- and Trans- Configurations

3.2.3. Omega Fatty Acids

One system for naming unsaturated fatty acids is to indicate the position of the **first** double bond counting from the opposite end from the carboxyl group. That terminal carbon atom is called the **omega** carbon atom. Thus a monounsaturated fatty acid with its single double bond after carbon #3 (counting from and including the omega carbon) is called an omega-3 fatty acid. But so is a **polyunsaturated** fatty acid, such as **linolenic acid**, if its first double bond is in that position.

3.3. Proteins

Proteins are **macromolecules**. They are constructed from one or more unbranched chains of **amino acids**; that is, they are **polymers**. A typical protein contains 200–300 amino acids but some are much smaller (the smallest are often called **peptides**) and some much larger (the largest to date is **titin** a protein found in skeletal and cardiac muscle; it contains 26,926 amino acids in a single chain!).

3.3.1. Amino Acids

Amino acids are the building blocks (monomers) of proteins. 20 different amino acids are used to synthesize proteins. The **shape** and other properties of each protein is dictated by the precise sequence of amino acids in it.

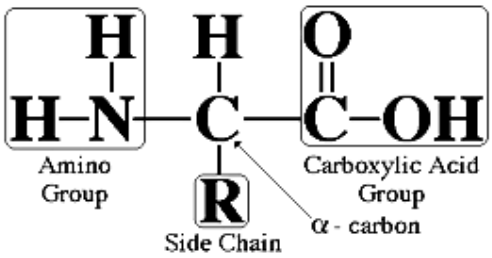
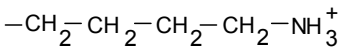
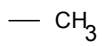
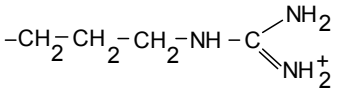
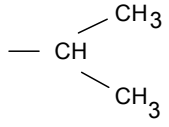
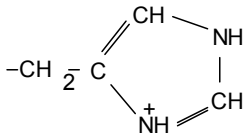
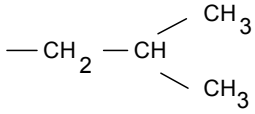
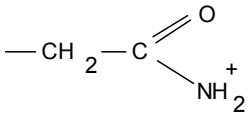
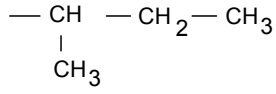
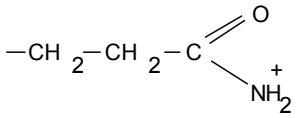
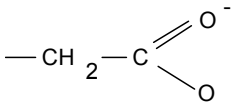
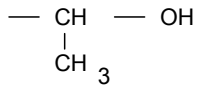
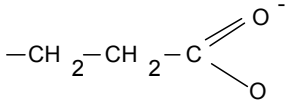
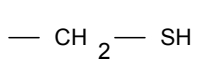
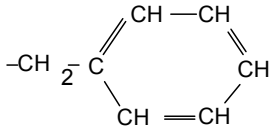
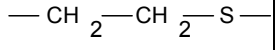
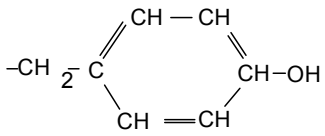
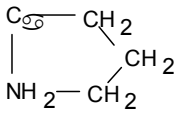
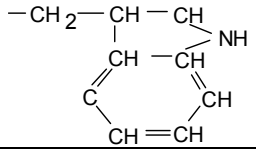


Fig.3.5a. Structure of Amino Acid

Each amino acid consists of an **alpha carbon atom** to which is attached a hydrogen atom, an amino group (hence "amino" acid), a carboxyl group (-COOH). This gives up a proton and is thus an acid (hence amino "acid") and one of 20 different "R" groups. It is the structure of the R group that determines which of the 20 it is and its special properties.

Table 3.2. The Twenty Amino Acid R-Groups

	Simple R groups		**Basic R groups**
Glycine Gly G		Lysine Lys K	
Alanine Ala A		Arginine Arg R	
Valine Val V		Histidine His H	
Leucine Leu L		Asparagine Asn N	
Isoleucine Ile I		Glutamine Gln Q	
	Hydroxyl R groups		**Acidic R groups**
Serine Ser S		Aspartate Asp D	
Threonine Thr T		Glutamate Glu E	
	Sulphur R groups		**Ringed R groups**
Cysteine Cys C		Phenylalanine Phe F	
Methionine Met M		Tyrosine Tyr Y	
	Cyclic R group		

Proline Pro P		Tryptophan Trp W	

Humans must include adequate amounts of 9 amino acids in their diet. These "essential" amino acids cannot be synthesized from other precursors. However, **cysteine** can partially meet the need for **methionine** (they both contain **sulphur**), and **tyrosine** can partially substitute for **phenylalanine**. The Essential Amino Acids are Histidine, Isoleucine, Leucine, Lysine, Methionine (and/or cysteine), Phenylalanine (and/or tyrosine), Threonine, Tryptophan and Valine. Two of the essential amino acids, **lysine** and **tryptophan**, are poorly represented in most plant proteins. Thus strict vegetarians should ensure that their diet contains sufficient amounts of these two amino acids.

3.3.2. Polypeptides

Polypeptides are chains of amino acids. Proteins are made up of one or more polypeptide molecules. The amino acids are linked covalently by **peptide bonds**. The image shows how three amino acids linked by peptide bonds into a **tripeptide**. One end of every polypeptide, called the **amino terminal** or **N-terminal**, has a free amino group. The other end, with its free carboxyl group, is called the **carboxyl terminal** or **C-terminal**. The sequence of amino acids in a polypeptide is dictated by the **codons** in the messenger RNA (mRNA) molecules from which the polypeptide was **translated.** The sequence of codons in the mRNA was, in turn, dictated by the sequence of codons in the DNA from which the mRNA was **transcribed.** The schematic below shows the **N-terminal** and the **C-terminal**. Proteins are made up of one or more polypeptide molecules.

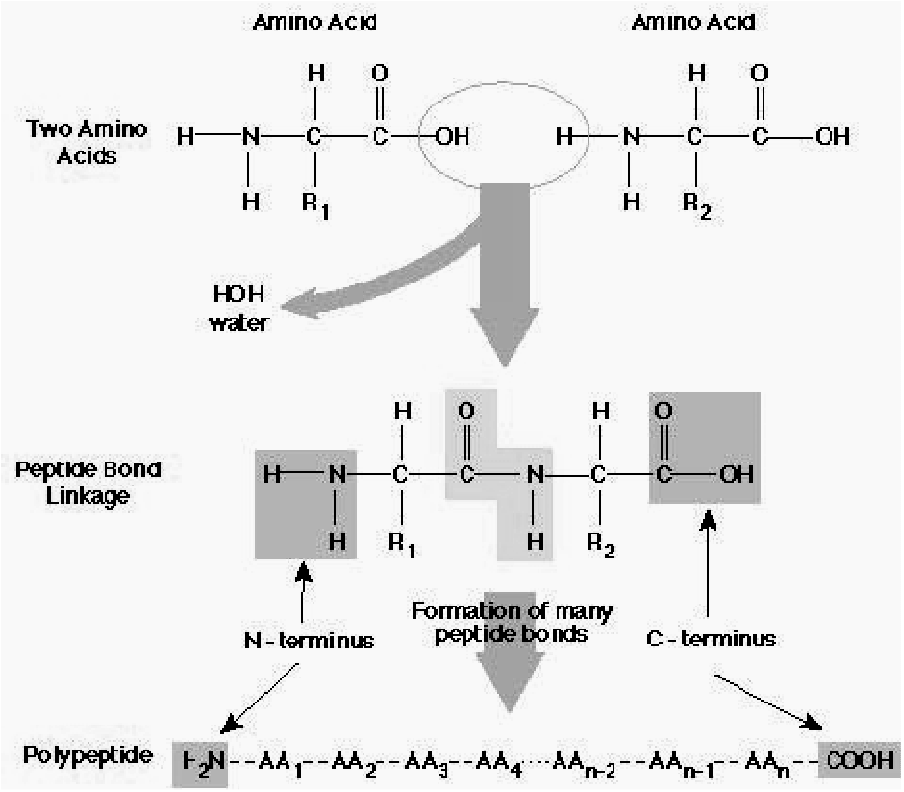


Fig.3.5b. Formation of polypeptide chain through a peptide bonding

3.3.3. Functions of Proteins

Every function in the living cell depends on proteins.

1. Motion and locomotion of cells and organisms depends on **contractile proteins**. [Examples: Muscles]
2. The catalysis of all biochemical reactions is done by **enzymes**, which contain protein.
3. The structure of cells, and the **extracellular matrix** in which they are embedded, is largely made of protein. [Examples: **Collagens**] (Plants and many microbes depend more on carbohydrates, e.g., cellulose, for support, but these are synthesized by enzymes.)
4. The transport of materials in body fluids depends of proteins. [e.g. Blood]
5. The **receptors** for **hormones** and other **Signalling molecules** are proteins.
6. Proteins are an essential nutrient for heterotrophs.
7. The t**ranscription factors** that turn genes on and off to guide the differentiation of the cell and its later responsiveness to signals reaching it are proteins. and many more — proteins are truly the physical basis of life.

3.3.4. Modifications of Protein Structure

When proteins are first synthesized, a process called **translation**, they consist of a linear assembly of the various **amino acids**, of which only 20 are normally used. Later, "post-translational" steps can alter some of the amino acids by **covalent attachment** of a variety of sugar residues to form **glycoproteins**. A **phosphate groups**, on **Tyr** for example. The adding of phosphate groups (by **kinases**) and their removal (by **phosphatases**) are crucial to the control of the function of many proteins. A Sulphate groups (SO_4^{2-}) can also be covalently attached to Tyr residues.

(a) Circular Proteins

Some bacteria, plants, and animals (but not humans) cut one or more peptides out of certain of their translated proteins and link the free ends together to form a circular protein. The details of how this is done are not yet known, but with a free amino group and one end and a free carboxyl at the other (the groups that form all **peptide bonds**) there is no chemical difficulty to overcome. The advantage of circular proteins seems to be great resistance to degradation (e.g., no free end for **peptidases** to work on).

(b) Inteins

Another, very rare, post-translational modification is the later removal of a section of the polypeptide and the splicing together (with **a peptide bond**) of the remaining **N-terminal and C-terminal** segments. The portion removed is called an **intein** (a "protein **intron**") and the ligated segments are called **exteins** ("protein **exons**"). Genes encoding inteins have been discovered in a variety of organisms, including some "true" bacteria such as *Bacillus subtilis,* several mycobacteria, several blue-green algae (cyanobacteria). Some Archaea such as *Methanococcus jannaschii*, and *Aeropyrum pernix* and a few eukaryotes, e.g., budding yeast (*Saccharomyces cerevisiae*). None has been found in the genomes of higher eukaryotes like *Drosophila, C. elegans,* or the green plant Arabidopsis.

3.3.5. Protein Structure

The function of a protein is determined by its shape. The shape of a protein is determined by its **primary structure** (sequence of amino acids). The sequence of amino acids in a protein is determined by the sequence of nucleotides in the gene

(DNA) encoding it. The function of a protein (except when it is serving as food) is absolutely dependent on its **three-dimensional structure**. A number of agents can disrupt this structure thus **denaturing** the protein. Changes in **pH** (alters **electrostatic interactions** between charged amino acids) and changes in salt concentration (does the same). Changes in temperature (higher temperatures reduce the strength of **hydrogen bonds**) and the presence of reducing agents (break S-S bonds between **cysteines**). None of these agents breaks **peptide bonds**, so the primary structure of a protein remains intact when it is denatured. When a protein is denatured, it loses its function. Examples: A denatured **enzyme** ceases to function and A denatured **antibody** no longer can bind its antigen. Often when a protein has been **gently** denatured and then is returned to normal **physiological conditions** of temperature, pH, salt concentration, etc., it spontaneously regains its function (e.g. enzymatic activity or ability to bind its antigen).
This tells us that the protein has spontaneously resumed its native three-dimensional shape and its ability to do so is **intrinsic**; no outside agent was needed to get it to refold properly. However, there are: enzymes that add sugars to certain amino acids, and these may be essential for proper folding. Proteins, called **molecular chaperones**, that may enable a newly-synthesized protein to acquire its final shape faster and more reliably than it otherwise would.

(a) Chaperones

Although the three-dimensional (tertiary) structure of a protein is determined by its primary structure, it may need assistance in achieving its final shape. As a polypeptide is being synthesized, it emerges (N-terminal first) from the ribosome and the folding process begins. However, the emerging polypeptide finds itself surrounded by the watery cytosol and many other proteins. As hydrophobic amino acids appear, they must find other hydrophobic amino acids to associate with. Ideally, these should be their own, but there is the danger that they could associate with nearby proteins instead — leading to aggregation and a failure to form the proper tertiary structure. To avoid this problem, the cells of all organisms contain molecular chaperones that stabilize newly-formed polypeptides while they fold into their proper structure. Most (~80%) newly-synthesized proteins are stabilized by molecular chaperones that bind briefly to their surface until they have folded properly. The chaperones use the energy of **ATP** to do this work.

(b) Chaperonins

However, some proteins (~20%) are so complex that a different group of chaperones - called chaperonins - are needed. Chaperonins are hollow cylinders into which the newly-synthesized protein fits while it folds. The inner wall of the cylinder is lined with hydrophobic amino acids which stabilize the hydrophobic regions of the polypeptide chain while it folds safely away from the watery **cytosol** and other **proteins** outside. Chaperonins also use ATP as the energy source to drive the folding process. As mentioned above, high temperatures can denature proteins, and when a cell is exposed to high temperatures, several types of **chaperonins** swing into action. For this reason, these chaperonins are also called **heat-shock proteins** (HSPs). Protein aggregation is the cause of some disorders such as **Alzheimer's disease**, **Huntington's disease**, and **prion diseases** (e.g., "mad-cow" disease). Perhaps, a failure of chaperones is involved. If so, perhaps ways can be found to treat these diseases by increasing the efficiency of chaperones. Despite the importance of chaperones, the rule still holds: the final shape of a protein is determined by only one thing: the precise sequence of amino acids in the protein. And the sequence of amino

acids in every protein is dictated by the sequence of nucleotides in the gene encoding that protein. So the function of each of the thousands of proteins in an organism is specified by one or more genes. There are some cases where a protein can exist in more than one conformation; that is, a given p**rimary structure** can give rise to two different **tertiary structures.**

(c)Primary Structure

The primary structure of a protein is its linear sequence of amino acids and the location of any disulfide (-S-S-) bridges.

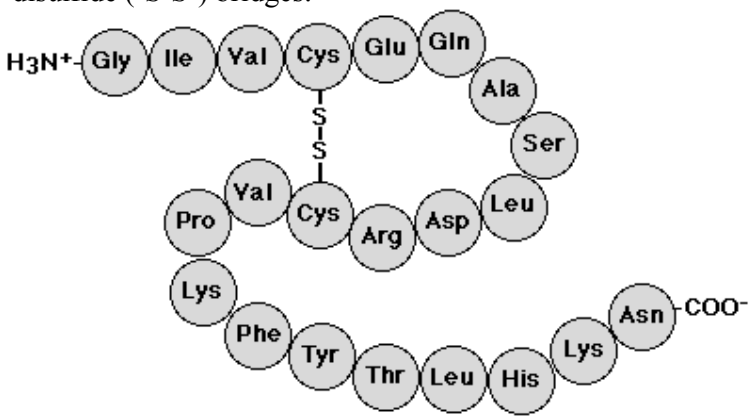


Fig.3.6a. Primary Structure of Polypeptide

Note the **amino terminal** or "**N-terminal**" (NH_3^+) at one end; **carboxyl terminal** ("**C-terminal**") (COO^-) at the other.

(d) Secondary Structure

Most proteins contain one or more stretches of amino acids that take on a characteristic structure in 3-D space. The most common of these are the **alpha helix** and the **beta conformation**.

(i) Alpha Helix

The **R groups of the amino** acids all extend to the outside. The **helix** makes a complete turn every 3.6 amino acids. The helix is right-handed; it twists in a clockwise direction. The carbonyl group (-C=O) of each **peptide bond** extends parallel to the axis of the helix and points directly at the -N-H group of the peptide bond 4 amino acids below it in the helix. A **hydrogen bond** forms between them [**-N-H·····O=C-**].

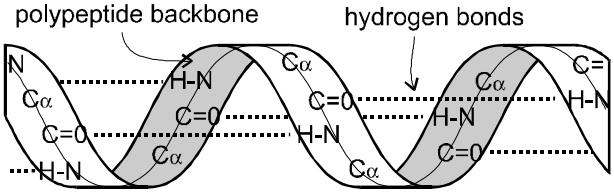


Fig.3.6b. The α-helix *Secondary Structure of Proteins*

(ii) Beta Conformation

This consists of pairs of chains lying side-by-side and stabilized by hydrogen bonds between the carbonyl oxygen atom on one chain and the **-NH group** on the adjacent chain. The chains are often "anti-parallel"; the **N-terminal to C-terminal** direction of one being the reverse of the other.

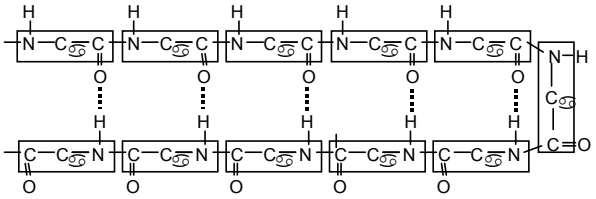


Fig.3.6c. The β-sheet *Secondary Structure of Proteins*

(e) Tertiary Structure

Tertiary structure refers to the three-dimensional structure of the entire **polypeptide chain.** The function of a protein (except as food) depends on its tertiary structure. If this is disrupted, the protein is said to be **denatured**, and it loses its activity. A **mutation** in the gene encoding a protein is a frequent cause of altered tertiary structure. The mutant versions of proteins may fail to reach their proper destination in the cell and/or be degraded. Examples: Most cases of **cystic fibrosis** are caused failure of the mutant CFTR protein to reach its destination in the plasma membrane. **Diabetes insipidus** is caused by improper folding of mutant versions of V2 - the **vasopressin** (ADH) receptor or **aquaporin**. **Familial hypercholesterolemia** is caused by failure of mutant **low-density lipoprotein (LDL) receptors** to reach the plasma membrane. **Osteogenesis imperfecta** is caused by failure of mutant Type I **collagen** molecules to assemble correctly. Mutant proteins may aggregate forming insoluble, nonfunctional deposits. This is particularly likely if the mutation causes **hydrophobic R groups** to be displayed at the surface of the molecule rather than in its interior.

Examples: An inherited form of **Alzheimer's disease** is characterized by insoluble deposits, called amyloid, of a mutant protein in the brain. **Bovine spongiform encephalopathy** (**BSE**) ("mad cow") disease and the human version - **Creutzfeldt-Jakob disease** (**CJD**) - are characterized by **amyloid deposits** in the brain of a mutant version of the **prion protein**. The normal protein has lots of alpha helical regions and is soluble. In the mutant version, the alpha helix is converted into beta conformation and the protein becomes insoluble. Curiously, tiny amounts of the mutant version can trigger the alpha-to-beta conversion in the normal protein. Thus the mutant version can be **infectious**. There have been several cases in Europe of people ill with Creutzfeldt-Jakob disease that may have acquired it from ingesting tiny amounts of the mutant protein in their beef. A number of other proteins altered by a point mutation in the gene encoding them, e.g., **fibrinogen, lysozyme, transthyretin** (a serum protein that transports **thyroxin** and **retinol** (vitamin A) in the blood) can form insoluble amyloid deposits in humans.

The many hydrogen bonds that can form between the polypeptide **backbones** in the beta conformation suggests that this is a stable secondary structure potentially available to all proteins and so a tendency to form insoluble aggregates is as well. This may account for the large investment in the cell in **chaperones** and **proteasomes** as well as the crucial importance of particular amino acid side chains in maintaining a globular, and hence soluble, tertiary structure.

The tertiary structure of many proteins is built from several **domains**. Often each domain has a separate function to perform for the protein, such as: binding a small ligand (e.g., a peptide in the molecule shown here), spanning the plasma membrane (transmembrane proteins), containing the catalytic site (enzymes), DNA-binding (in transcription factors) and providing a surface to bind specifically to another protein.

In some (but not all) cases, each domain in a protein is encoded by a separate exon in the gene encoding that protein.

In the **histocompatibility molecule**, three domains α_1, α_2, and α_3 are each encoded by its own exon. Two additional domains being a transmembrane domain and a cytoplasmic domain are also encoded by separate exons. (β_2-microglobulin, "β_2m", is NOT a domain of this molecule. It is a separate molecule that binds to the three alpha domains (red line and circle) by **noncovalent forces** only. The complex of these two proteins is an example of **quaternary structure**.) A correspondence between exons and domains is more likely to be seen in recently-evolved proteins. Presumably, "**exon shuffling**" during evolution has enabled organisms to manufacture new proteins, with new functions, by adding exons from other parts of the genome to encode new domains.

(f) Quaternary Structure

This structure is found in proteins containing more than one polypeptide chain, and simply means how the different polypeptide chains are arranged together. The individual polypeptide chains are usually globular, but can arrange themselves into a variety of quaternary shapes. e.g.:

(i) Haemoglobin, the oxygen-carrying protein in red blood cells, consists of four globular subunits arranged in a tetrahedral (pyramid) structure. Each subunit contains one iron atom and can bind one molecule of oxygen.

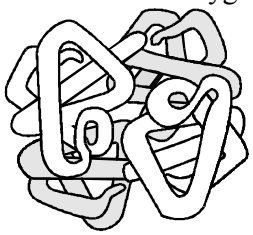


Fig.3.6d. Haemoglobin

(ii) Immunoglobulins, the proteins that make antibodies, comprise four polypeptide chains arranged in a Y-shape. The chains are held together by sulphur bridges. This shape allows antibodies to link antigens together, causing them to clump.

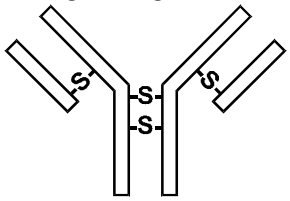


Fig.3.6e. Immunoglobulins

(iii) Actin, one of the proteins found in muscles, consists of many globular subunits arranged in a double helix to form long filaments.

Fig.3.6f. Actin

(iv) Tubulin is a globular protein that polymerises to form hollow tubes called microtubules. These form part of the cytoskeleton, and make cilia and flagella move.

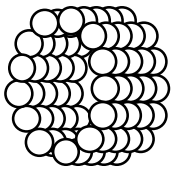


Fig.3.6g. Tubulin

These four structures are not real stages in the formation of a protein, but are simply a convenient classification that scientists invented to help them to understand proteins. In fact proteins fold into all these structures at the same time, as they are synthesised.

The final three-dimensional shape of a protein can be classified as **globular** or **fibrous**.

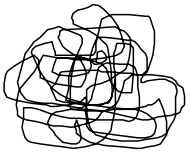


Fig.3.6h. Globular structure

Fig.3.6i.. Fibrous (or filamentous) structure

The vast majority of proteins are **globular,** including **enzymes, membrane proteins, receptors, storage proteins**, etc. Fibrous proteins look like ropes and tend to have structural roles such as **collagen** (bone), **keratin** (hair), **tubulin** (cytoskeleton) and **actin** (muscle). They are usually composed of many polypeptide chains. A few proteins have both structures: the muscle protein myosin has a long fibrous tail and a globular head, which acts as an enzyme.

3.4. Glycoproteins

Glycoproteins have **carbohydrate** attached to them. The attachment is a **covalent** linkage to the hydroxyl (-OH) group of the **R group** of **serine** or **threonine** - called "**O-linked**" in both cases or to the amino group ($-NH_2$) in the R group of **asparagine -** called "**N-linked**". The carbohydrate consists of short, usually branched, chains of **sugars** and nitrogen-containing **amino sugars**. Sugars are very **hydrophilic** thanks to their many -OH groups. Their presence makes glycoproteins far more hydrophilic than they would be otherwise and are often essential for the proper folding of the protein into its **tertiary structure**. Most of the proteins exposed to the watery surroundings at the surface of cells are glycoproteins.

This image below shows the primary structure of **glycophorin A**, a glycoprotein that spans the plasma membrane ("Lipid bilayer") of human **red blood cells**. Each RBC has some 500,000 copies of the molecule embedded in its plasma membrane. Fifteen carbohydrate chains are "O-linked" to serine (Ser) and threonine (Thr) residues. One carbohydrate chain is "N-linked" to the asparagine (Asn) at position 26. Two **polymorphic** versions of **glycophorin A**, which differ only at residues 1 and 5, occur in humans. These give rise to the MN blood groups. The M allele encodes Ser at position 1 (Ser-1) and Gly at position 5 (Gly-5).The N allele encodes Leu-1 and Glu-5

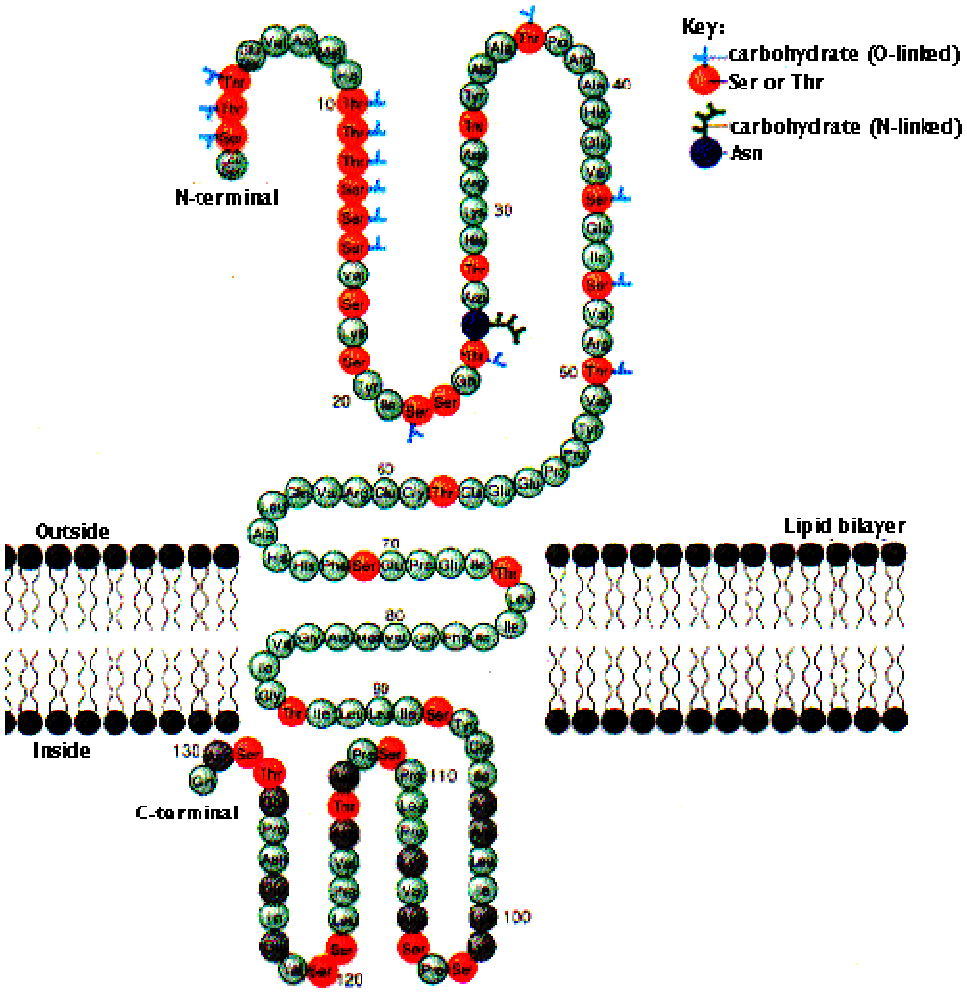


*Fig.3.7. Primary structure of **glycophorin A***

3.5. Nucleotides

Nucleic acids are linear, unbranched polymers of nucleotides. Nucleotides consist of three parts: A five-carbon sugar (hence a **pentose**). Two kinds are found:

1. **Deoxyribose**, which has a hydrogen atom attached to its #2 carbon atom (designated 2')
2. **Ribose**, which has a hydroxyl group atom there

Deoxyribose-containing nucleotides, the **deoxyribonucleotides**, are the monomers of **DNA**. Ribose-containing nucleotides, the **ribonucleotides**, are the monomers of **RNA**. A nitrogen-containing ring structure called a **base**. The base is attached to the 1' carbon atom of the pentose. In **DNA**, four different bases are found:

1. two **purines**, called **adenine** (**A**) and **guanine** (**G**)
2. two **pyrimidines**, called **thymine** (**T**) and **cytosine** (**C**)

RNA contains:

1. The same purines, **adenine** (**A**) and **guanine** (**G**).
2. RNA also uses the pyrimidine **cytosine** (**C**), but instead of thymine, it uses the pyrimidine **uracil** (**U**).

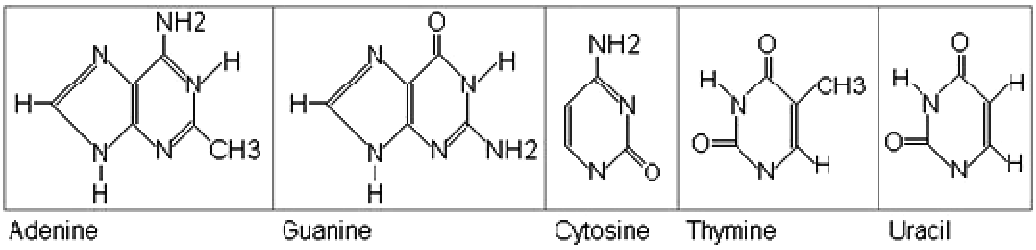


Fig.3.8a. Structure of the Nitrogenous Bases of DNA and RNA

The combination of a base and a pentose is called a **nucleoside**.

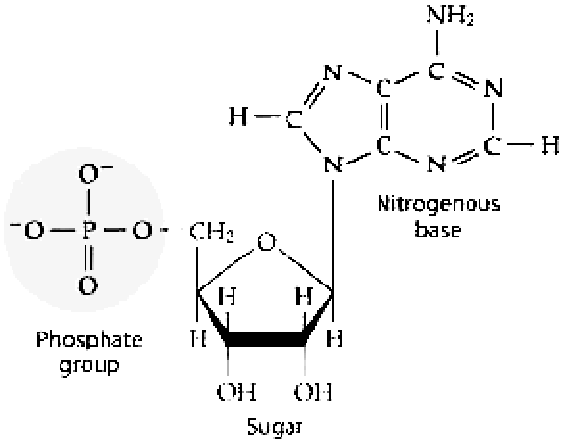


Fig.3.8b. Structure of a Nucleotide

One (as shown in the figure), two, or three **phosphate** groups. These are attached to the 5' carbon atom of the pentose. Both DNA and RNA are assembled from **nucleoside triphosphates**. For **DNA**, these are **dATP**, **dCTP**, **dGTP**, and **dTTP**. For **RNA**, these are **ATP**, **CTP**, **GTP**, and **UTP**. In both cases, as each nucleotide is attached, the second and third phosphates are removed.

Table3.3. The nucleosides and their mono-, di-, and triphosphates

	Base	Nucleoside	Nucleotides		
DNA	Adenine (A)	Deoxyadenosine	dAMP	dADP	dATP
	Guanine (G)	Deoxyguanosine	dGMP	dGDP	dGTP
	Cytosine (C)	Deoxycytidine	dCMP	dCDP	dCTP
	Thymine (T)	Deoxythymidine	dTMP	dTDP	dTTP
RNA	Adenine (A)	Adenosine	AMP	ADP	ATP
	Guanine (G)	Guanosine	GMP	GDP	GTP
	Cytosine (C)	Cytidine	CMP	CDP	CTP
	Uracil (U)	Uridine	UMP	UDP	UTP

3.6. ATP (Adenosine Triphosphate)

ATP is a **nucleotide** that performs many essential roles in the cell. It is the major **energy currency** of the cell, providing the energy for most of the energy-consuming activities of the cell. It is one of the monomers used in the synthesis of **RNA** and, after conversion to deoxyATP (dATP), **DNA**. It regulates many biochemical pathways. In mammals, it also functions outside of cells. Its release from damaged cells can elicit pain, and its release from the stretched wall of the urinary bladder signals when the bladder needs emptying!

When the third phosphate group of ATP is removed by **hydrolysis**, a substantial amount of **free energy** is released. The exact amount depends on the conditions, but we shall use a value of 7.3 kcal per mole.

$$\boxed{\mathbf{ATP + H_2O \rightarrow ADP + P_i}}$$

ADP is adenosine diphosphate. P_i is inorganic phosphate. For this reason, this bond is known as a "high-energy" bond and is depicted in the figure by a wavy line. (The bond between the first and second phosphates is also "high-energy".) (But please note that the term is **not** being used in the same sense as the term "bond energy". In fact, these bonds are actually weak bonds with low bond energies.)

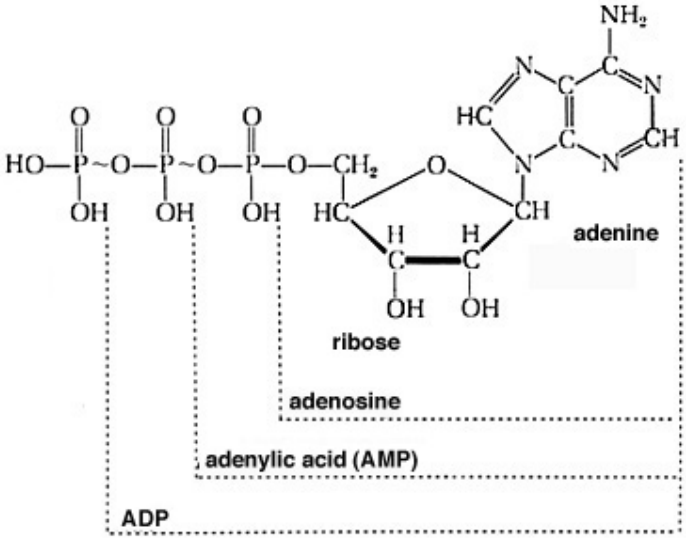


Fig.3.9. Structure of a ATP