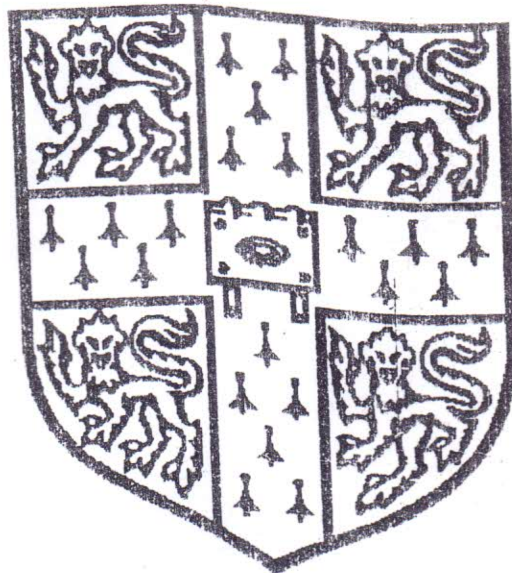
evel Chemistry 9254

10 JUL 1996

# University of Cambridge
## Local Examinations Syndicate



# BIOCHEMISTRY

## Option Booklet

September 1995

# BIOCHEMISTRY

## CONTENTS

1. Introduction - page 3
2. Proteins - page 6
3. Carbohydrates - page 35
4. Lipids - page 40
5. Nucleic acids - page 46

## SUMMARY

**2. Proteins**

- amino acids
- polypeptides and proteins
- protein structure: primary, secondary, tertiary and quaternary structures
- denaturation of proteins
- enzymes: relationship between enzymes and substrates; active sites; reversible inhibition; co-enzymes; lysozyme

**3. Carbohydrates**

- monosaccharides: α - and β - pyranose structure of glucose
- disaccharides and polysaccharides as condensation polymers: the nature of the glycosidic link; enzymic and acid hydrolysis of the glycosidic linkage
- structure and function of cellulose as a structural polymer and of starch and glycogen as storage polymers
- consequences of hydrogen bonding on the relative solubilities of monosaccharides and polysaccharides in water

**4. Lipids**

- biological functions of lipids
- simple structure and function of triglycerides; phosphogylcerides: simplified structure and function; formation of micelle and bilayers
- membranes: fluid mosaic model; active transport

**5. Nucleic acids**

- nucleotides and nucleic acids
- DNA and RNA: base pairing
- DNA and genetic information
- m-RNA and the triplet code
- ATP as an important example of a nucleotide: its importance in metabolic activity

## 1. Introduction

Biochemistry seeks to answer questions about the molecular structures and chemical processes of living organisms. The molecules found in living cells obey the same physical and chemical laws that apply to inanimate matter but, when operating together, these molecules give the cells extraordinary properties not shown by collections of non-living matter. It is useful to review some of these properties.

### (a) Characteristics of living matter

1. *Energy transformation.* Living organisms are able to obtain energy from their environment and then release this in a controlled and usable form.

2. *Growth.* Materials are taken from outside, modified and incorporated into the structure of the organism.

3. *Responsiveness.* Organisms react to various stimuli and many of the responses are to do with maintaining a constant internal environment.

4. *Movement.* All organisms are able to use energy to move themselves or some part of themselves.

5. *Reproduction.* All living organisms are able to reproduce themselves.

6. *Complexity.* Living cells contain many kinds of complicated molecules which are organised into very sophisticated internal structures to carry out the work of the cell.

The molecules of living organisms are compounds of mainly carbon, hydrogen, oxygen and nitrogen. Carbon and nitrogen are relatively scarce in non-living matter. In the atmosphere, these elements occur in simple molecules, such as carbon dioxide, water, oxygen and nitrogen. In the Earth's crust, they occur as carbonates and nitrates.

The variety of molecules built from these few elements in living cells is enormous. Even the simplest bacterium may contain 5000 different kinds of compound, including 3000 different proteins and 1000 different nucleic acids. In human systems, there may be as many as 5 million different proteins. However, it is remarkable that most cell macromolecules are composed of a few, fairly simple, building blocks.

Glucose molecules → starch and cellulose

20 different amino acids → proteins

Eight nucleotides → nucleic acids

The twenty amino acids that make up all proteins and the eight nucleotides that are used to make the nucleic acids are identical in all living species. Amino acids are synthesised only in plants which build up their own protein. Animals obtain amino acids by way of food chains and webs; ingested protein is hydrolysed into amino acids by enzymes in the digestive systems, and essential animal protein is built up within the organism, again controlled by the enzymes.

All organic macromolecules are ultimately derived from simple low relative molecular mass precursors obtained from the environment and then converted into increasingly complex molecules and structures as indicated below.

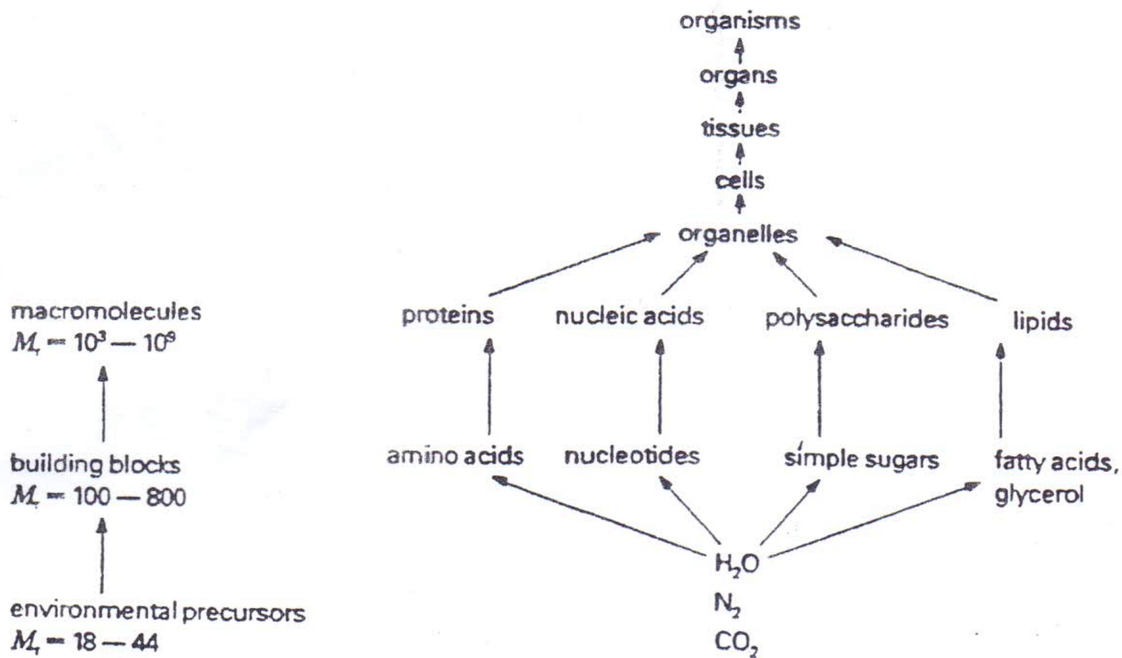



**Fig 1.1: Derivation of complex molecules and structures**

### (b) The structure of cells

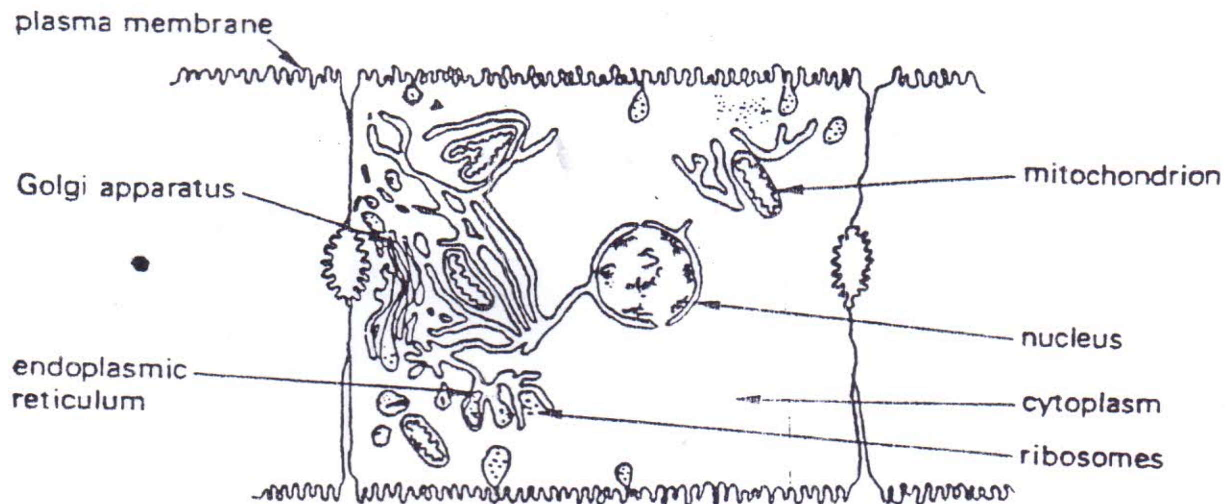



**Fig 1.2: Typical animal cell**

*The cell membrane* (plasma membrane) contains both lipids and proteins in equal amounts: the lipids are arranged in a bilayer (this will be investigated later in this text). The cell membrane is selectively permeable and contains a transport system, for $Na^+$ and $K^+$ ions, glucose, amino acids and other nutrients as well as certain enzymes.

In cells other than animal cells, the plasma membrane is protected by a *cell wall* which confers shape and rigidity on the cell and is very strong. The cell walls of plants are largely made of cellulose.

*The nucleus* is **the** control centre of the cell. It contains DNA organised into chromosomes. During cell division, the chromosomes undergo replication of their DNA, and so all the information needed for the development and other activities of the cell is passed on.

*The mitochondria.* There are usually several hundred mitochondria per cell and they are the sites of respiration, where carbohydrates, lipids and amino acids are oxidised to $CO_2$ and $H_2O$.

*The Golgi apparatus* functions in the secretion of cell products, such as proteins, to the exterior and it helps to form the cell membrane.

*The endoplasmic reticulum* consists of a system of membranes having inner compartments which interconnect to form channels. The rough-surfaced portion of the endoplasmic reticulum is studded with *ribosomes*. Proteins synthesised by the ribosomes cross the membrane of the endoplasmic reticulum and appear in the channels which form a transport system. Protein synthesis also takes place in unattached ribosomes.

## 2. Proteins

### (a) The function of proteins

Proteins are essential in living systems.

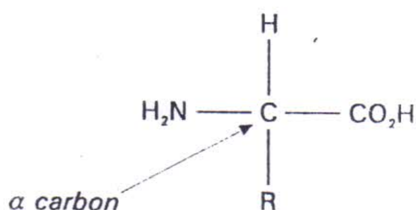
- *Structural proteins* provide the framework which defines the size and shape of cells.
- *Muscle fibres* are made of protein. Muscle produces mechanical force and movement. The two chief proteins in all muscle are actin and myosin; these are large enough to be photographed using an electron microscope.
- *Transport proteins* move metabolites around the cell or around the whole organism.
- Some *hormones* which control the level and type of cell activity are protein molecules.
- *Enzymes* catalyse metabolic processes which produce energy, build up new cell structures and destroy old ones. Enzyme catalysis is very effective; for example, enzymes can process up to one million metabolites in each second.

Notwithstanding their diverse functions, all proteins consist of long chains of amino acids. About twenty amino acids are used to make proteins. Each protein is determined by its amino acid sequence. The number of amino acids linked in different protein ranges from 44 in the hormone insulin to giant molecules containing more than 50 000 amino acids. Hence, an almost endless variety of protein molecules is possible.

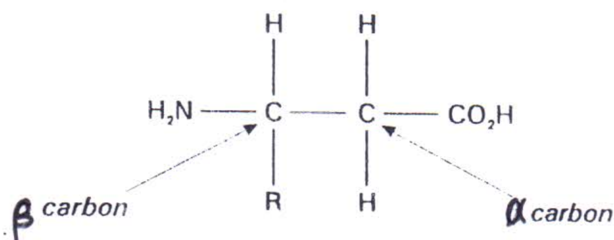
### (b) The monomers

All proteins are built up by linking together amino acids.

The common, naturally occurring, amino acids all have the same general structure. They are all α-amino acids having the general formula



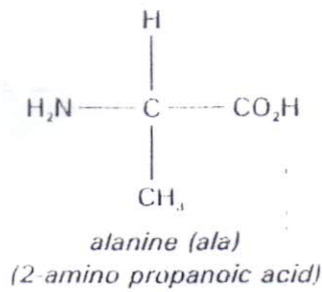

but differ with respect to R. α-amino acids have the amino group attached to the carbon next to the carboxyl group. If the amino group were attached to the next carbon atom, the molecule would be a β-amino acid.



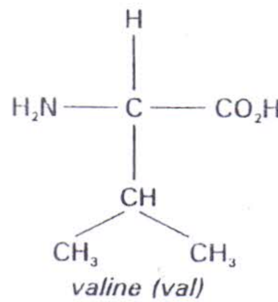

However, β-amino acids are rare in nature and are not found in proteins.

The R groups give the amino acids their respective characteristic properties. The amino acids can be classified according to the nature of the R group. For example:

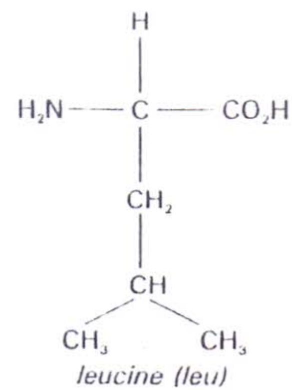
(i) *non-polar or hydrophobic R groups,*



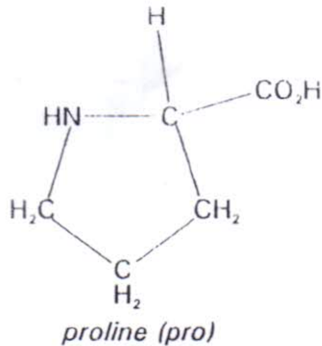

alanine (ala)
(2-amino propanoic acid)




valine (val)

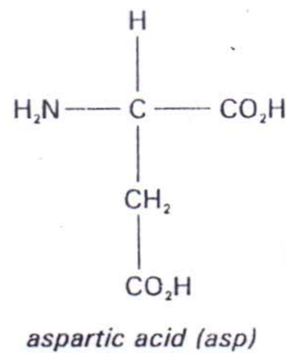



leucine (leu)

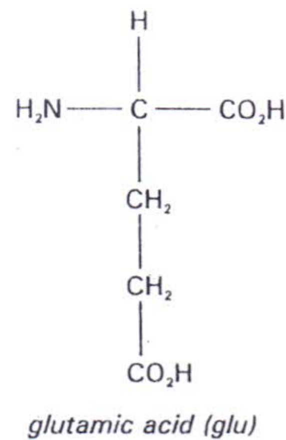



proline (pro)

(ii) *negatively charged R groups at pH 7 to 8,*

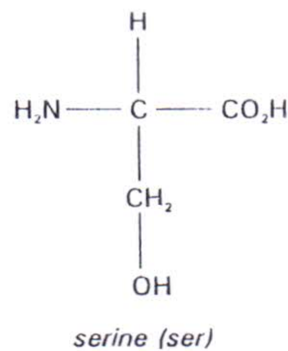



aspartic acid (asp)

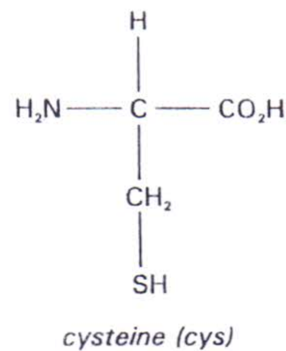



glutamic acid (glu)

(iii) *uncharged or hydrophilic R groups,*

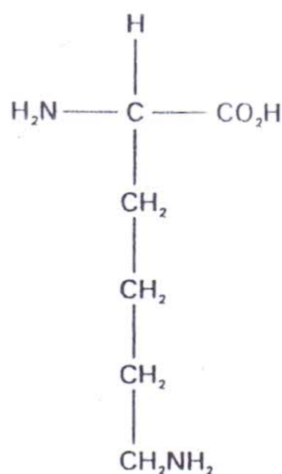



serine (ser)




cysteine (cys)

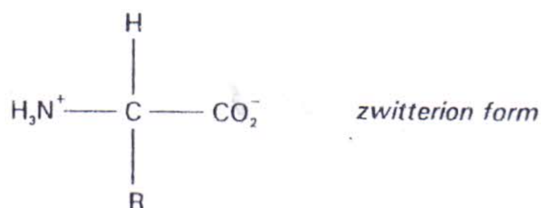
(iv) *positively charged R groups at pH 6 to 7.*



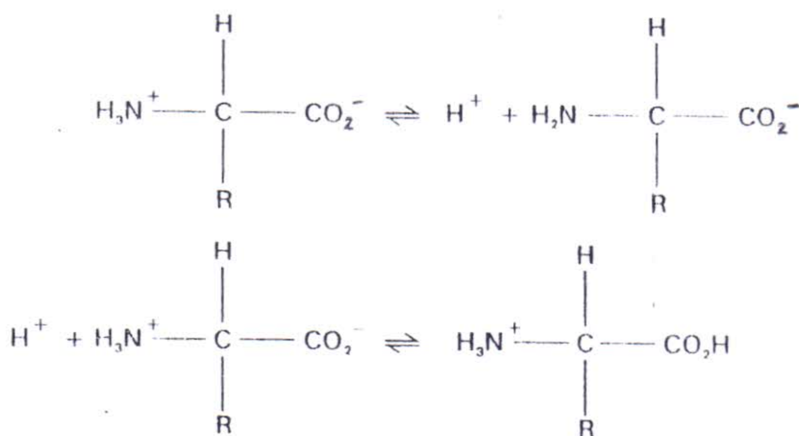

*lysine (lys)*

The α-carbon is the asymmetric centre in the α-amino acids, except for glycine. The two possible stereo isomers are labelled D- and L-. All naturally occurring amino acids, found in proteins, are of the L-configuration, though the actual direction in which they rotate plane-polarised light depends on the nature of the R group.

Crystalline amino acids have relatively high melting points or decomposing points, usually above 200 °C. They are much more soluble in water than in non-polar solvents. These and other pieces of evidence have led to the conclusion that amino acids occur in, and crystallise from, neutral aqueous solution as dipolar ions, called zwitterions. The electrostatic forces of attraction between oppositely charged groups account for the higher melting point of the crystal than expected if the crystal were merely stabilised by van der Waals' forces between the non-ionic form of the molecules.




In aqueous solution, a zwitterionic amino acid can act as either an acid (proton donor) or as a base (proton acceptor).



A simple α-amino acid, such as alanine is considered to be dibasic in its fully protonated form. It can donate two protons during its complete titration with a solution of a base.

(I) $H_3N^+CH(CH_3)CO_2H + OH^- \rightarrow H_3N^+CH(CH_3)CO_2^- + H_2O$

(II) $H_3N^+CH(CH_3)CO_2^- + OH^- \rightarrow H_2NCH(CH_3)CO_2^- + H_2O$

The following graph, Fig 2.1, shows the titration curve for an amino acid, such as alanine. The p$K$ values of the two stages of dissociation are far enough apart to see two clearly separate legs in the titration curve. The first leg corresponds to equation (I) above and the second leg corresponds to equation (II) above.

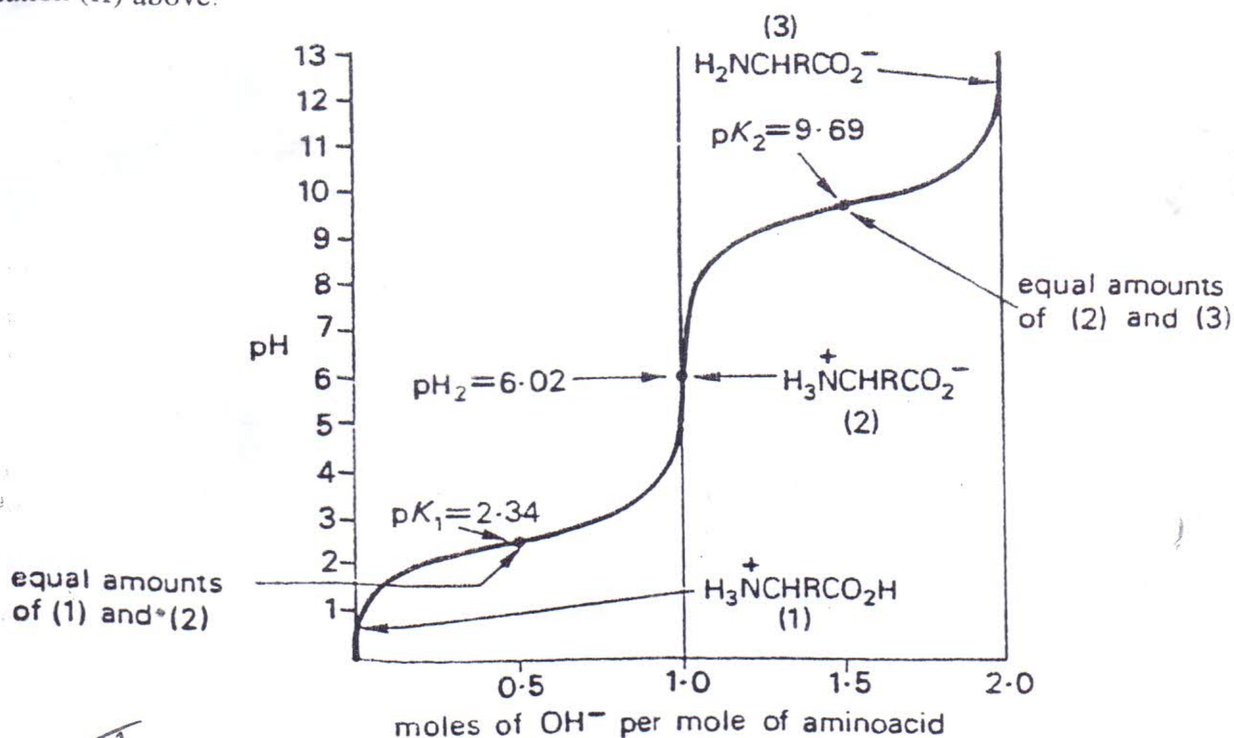



**Fig 2.1: Titration curve for an amino acid**

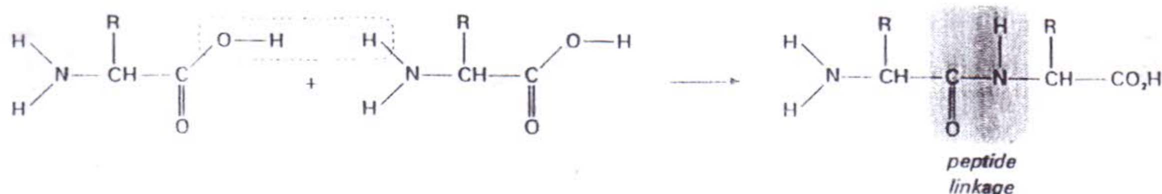
At pH 6.02, the slope of the titration curve reaches a maximum and this effectively divides the curve into two separate legs. There is no net electric charge on the molecule at this pH and the amino acid does not move in an electric field. This pH is called the isoelectric point, and the amino acid has least buffering effect at this point:

$$\text{isoelectric pH} = \tfrac{1}{2}(pK_1 + pK_2).$$

### (c) The peptide bond

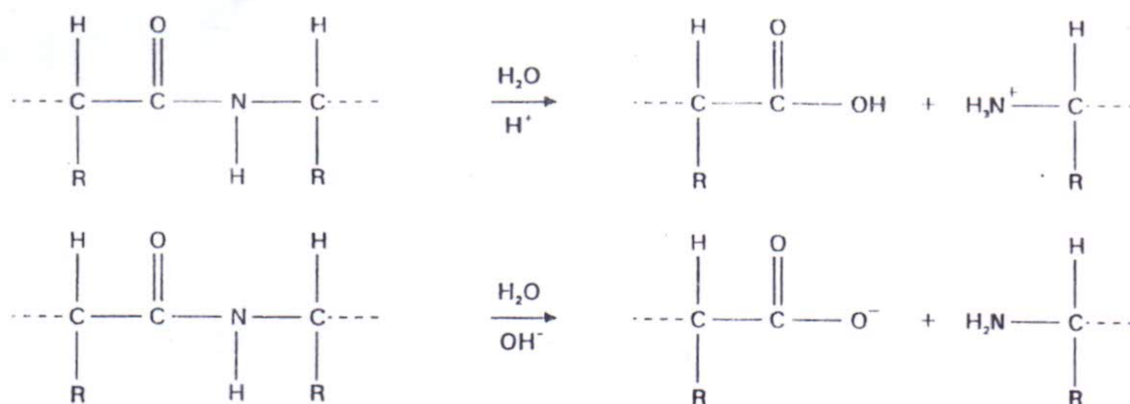
The peptide bond is an amide linkage and is formed by the elimination of a water molecule between the α-$NH_2$ of one amino acid and the α-$CO_2H$ group of a second amino acid. The amide bond is not only formed from the basic amine and the acidic carboxyl group, but also by the union of opposite poles of two zwitterions.

The resulting dipeptide is similar to a single amino acid and can form additional peptide linkages at either the -$CO_2H$ terminal or at the -$NH_2$ terminal. In nature, additional amino acids are added only to the -$CO_2H$ end of the molecule.



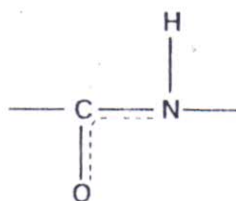

The resulting compounds are called peptides. Simple peptides, containing two or more amino acids residues linked covalently through peptide bonds, are formed by the hydrolysis of polypeptides. There is no precise dividing line between polypeptides and proteins, but, as a very general rule, polypeptides have relative molecular masses of 1500 or less and proteins have molecular masses of 5000 or more.

The peptide bond can be split by hydrolysis in the presence of a suitable enzyme, or by heating in acidic or in alkaline solution.



The presence of the peptide bond in a molecule can be demonstrated by the biuret test. Biuret is a simple compound, $NH_2CONHCONH_2$, containing the peptide bond. If copper(II) sulphate in alkaline solution is added to biuret, a purple colour is produced. Proteins and peptides (containing two or more peptide bonds) give a deep colouration with alkaline copper(II) sulphate. The intensity of the purple colour is proportional to the amount of polypeptide.

Peptides can be regarded as substituted amides and, as in amides, delocalisation of electrons occurs in the $-C{=}O$ and $C{-}N{-}$ (with H on N) bonds.



The C-N bond of the peptide linkage is, therefore, relatively rigid and cannot rotate freely. This has important implications for protein structure.

### (d) Separation of amino acids

The amino acids present in polypeptides and proteins can be separated by chromatographic techniques. The peptide bonds are readily hydrolysed by heating with either acids or alkalis. The usual procedure to effect complete hydrolysis is to heat a polypeptide with 6 mol $dm^{-3}$ hydrochloric acid at 100 °C to 120 °C for 10 to 24 hours in an evacuated tube. The amino acid composition of polypeptides and proteins can then be determined by paper chromatography or ion exchange chromatography.

*Paper chromatography.* Unglazed sheets of paper (similar to filter paper) are used in this technique. The cellulose of the paper fibres is hydrated and the water remains bonded to the cellulose, thus providing a stationary phase. The solvent (eluant) containing an amino acid mixture ascends the vertically held paper by capillary motion (or descends for descending chromatography). A distribution of the amino acids occurs between the flowing phase (eluant) and the stationary phase bonded to the paper fibres. At the end of the process, the different amino acids have moved different distances from the origin. The paper is dried, sprayed with ninhydrin solution and heated to locate the amino acids.

*Ion exchange chromatography.* In this method, solute molecules are separated by differences in their acid-base behaviour. A column is filled with granules of a synthetic resin containing fixed charged groups. The resins may have positively charged groups (anion-exchange resins) or negatively charged groups (cation-exchange resins) covalently bound and thus immobilised. Amino acids are usually separated on cation exchange columns filled with granules of a sulphonated poly(phenylethene) [poly(styrene)] resin. Sulphonic acids are strong acids and remain ionised at pH3. At pH3, amino acids are largely cations with net positive charge and the most basic amino acids (lysine, arganine and histidine) are bound the most firmly to the resin by electrostatic forces and the most acidic (glutamic acid and aspartic acid) are bound the least. Hence, some are displaced more readily than others by an ionic solution (e.g. aqueous sodium chloride) passed down the column and so appear more rapidly in the solution flowing out of the column (eluate). In this way, different amino acids are collected in many small fractions.

The entire procedure has been automated in an apparatus called an amino acid analyser. This method makes it possible to determine the type and amount of all the amino acids in a mixture in a little more than two hours, using only a fraction of a milligram of sample. By contrast, the first pure protein, β-lactoglobin, for which the complete amino acid composition was investigated took several years to analyse using the older techniques.

*Electrophoresis.* The differences in size and charge between amino acids and between proteins can be used to separate them by means of an electrical potential gradient.

The mixture of amino acids or proteins is dissolved in a buffer of pH chosen to yield the maximum separation of the compounds present. The mixture is then put on to filter paper or on to a gel and an electric field is applied. The amino acids or proteins then migrate in the field at a speed which depends on their charge-to-mass ratio. The position of the various compounds can be detected by colour tests.

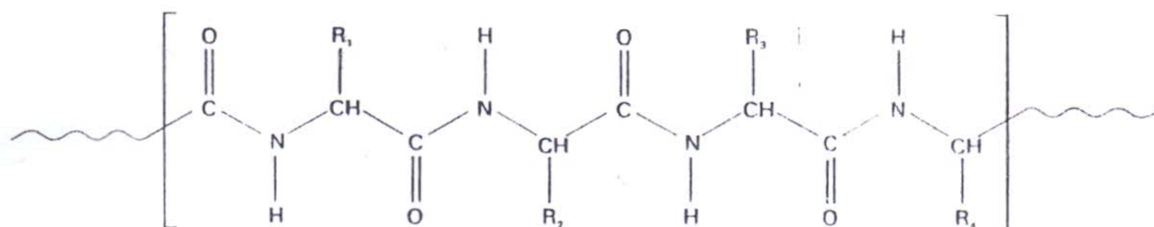
### (e) The primary structure of proteins

The *primary structure* of a protein is the sequence of the amino acids in the polypeptide chain or chains and it includes any covalent cross-linkages. It is the primary structure which determines what the protein is, how it folds, and its function(s).

On hydrolysis, simple proteins yield only amino acids. Relative molecular mass determinations indicate values ranging from about 5 000, arbitrarily the lower limit, to 1 million or more. Many

proteins having relative molecular masses above 36 000 contain two or more polypeptide chains. The individual polypeptide chains of most proteins contain between 100 and 300 amino acid residuės, e.g. ribonuclease, cytochrome c and myoglobin contain between 100 and 1555 amino acid residues. Some, however, have much longer chains.

In proteins, there is a repeating backbone of peptide bonds,

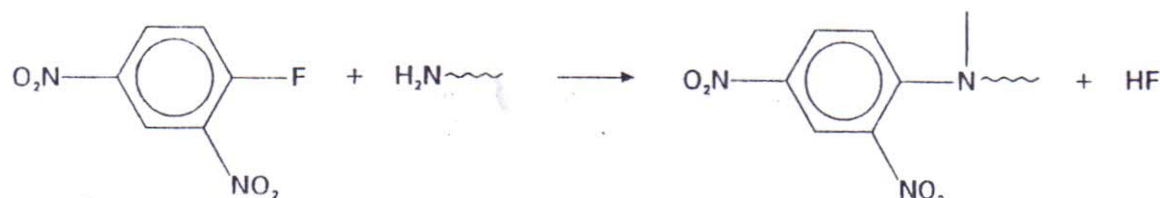



where $R_1$, $R_2$, $R_3$ etc represent the differing types of amino acid side chains. Since a protein may contain a single polypeptide chain which may contain several hundred amino acid residues, the number of ways in which these can be arranged in the chain is astronomical.

Amino acid analysis is only the initial stage in the determination of protein structure and reveals little about the protein and its properties. This is because complete hydrolysis of the protein indicates neither the sequence of the amino acids nor the shape of the protein and both of these are essential to its biological function.

Thé establishment of the sequence of amino acids in proteins can be established by the following method.

(i) The N-terminus (the end of the polypeptide having the free $NH_2$ group) is determined using fluorodinitrobenzene.




The reagent is yellow in colour and, after hydrolysis, it is necessary to pick out and identify any amino acid which moves as a yellow spot on the chromatography paper.

This technique identifies the position of a single amino acid and is known as 'end-group analysis'.

(ii) The polypeptide is then broken up into a number of simpler peptides by use of chemical hydrolysis or protolytic enzymes such as trypsin. These peptide fragments are then analysed or, if too large, broken down still further. Once two sets of peptide fragments have been obtained by two different procedures for cleavage of the original polypeptide and the amino acid sequence of the fragments has been established, it is possible to deduce the complete amino acid sequence from overlaps of the peptide sequences.

For example: fragments by cleavage method (1) Glu-Met-Leu-Gly-Arg
Ala-Gly
Tyr-Lys

fragments by cleavage method (2) Tyr-Lys-Glu-Met
Leu-Gly-Arg-Ala-Fly Gly

N terminus from A-Tyr.

Deduced sequence: Tyr-Lys-Glu-Met-Leu-Gly-Arg-Arg-Ala-Fly. Gly

A sufficient number of amino acid sequences in proteins have been determined to make it possible for several generalisations to be made.

(i) Two types of covalent bonds between amino acid residues are found in soluble proteins, the peptide bonds and the disulphide bonds of cysteine which hold two chains together. These latter are seen in insulin and in lysozyme (pages 29 to 31).

(ii) A protein has a unique sequence. There are, however, differences in amino acid sequences of the same protein in different species, even if the protein is very similar. These genetic variations in structure are usually confined to certain regions of the protein, e.g. cytochrome c varies slightly in man, duck, yeast and pumpkin. Presumably, these latter regions are less critical to the function of the protein and more variability in sequence can be tolerated.

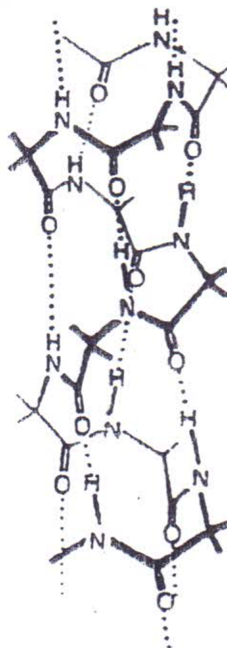
(iii) The sequences of amino acids within proteins appear nearly random. There are no sequences which are found to be repeated over and over in a large number of proteins. It has been shown, however, that large segments of proteins which are of the same evolutionary origin are very similar in structure.

**(e) Secondary structure of proteins**

The *secondary structure* of a protein is the term applied to the regular arrangement of sections of the polypeptide chains. X-ray diffraction methods have shown that the α-helix and the β-pleated sheet are common repeating structural patterns in proteins. These structures are both stabilised by hydrogen bonds.

The *α-helix* is a regular coiled configuration of the polypeptide chain. There are 3.6 amino acids per helical turn. The R groups on their respective α-carbon all point towards the outside of the helix and are perpendicular to the main axis of the helix. The oxygen of a given peptide carbonyl group is hydrogen-bonded to the nitrogen of the amino acid group associated with the fourth peptide linkage further down the chain. The hydrogen bonds in this structure are linear and, therefore, are maximally stable. Each bond is weak but there are so many of them that the total binding effect is strong.

The α-helix is flexible and elastic. The fibrous α-keratins are examples of the α-helix and, in hair and wool, three or seven such α-helices may be coiled around each other to form three-stranded or seven-stranded 'ropes' held together by disulphide cross linkages. This produces very strong fibres.



The α-helix is stabilised by hydrogen bonds (dotted lines).

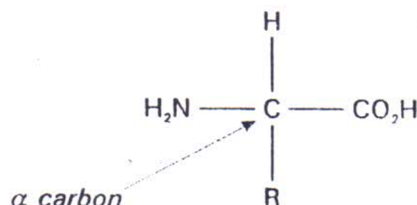
An α-helix may form with either L- or D-amino acids but not with a mixture of both. Also, a right-handed or a left-handed coil can be formed from the naturally occurring L-amino acids; however, the right-handed helix is more stable. All the α-helices found so far in proteins are right-handed.

The β-*pleated sheet* is a structure composed of 'side by side' polypeptides connected by hydrogen bonds. All the peptide linkages are involved in interchain hydrogen bonding and so the structure is very stable.

The β-sheet is formed most readily by polypeptides with repeating sequences of amino acids with compact side chains. The β-sheet is flexible but inelastic. One of the commonest forms of the β-sheet is β-keratin found in fibroin present in silk fibres. Muscle fibres, connective tissue, collagen, nails and feathers are largely β-pleated sheet structures. These are fibrous (or structural) proteins.

The β-pleated sheet is best described in terms of diagrams. Consider first Fig 2.2 (next page). This represents two protein strands, parallel and side-by-side to each other, one strand being in the foreground and the other in the background.

It will be remembered that the amino acids that join together to make a protein have the same general structure, i.e.




In Fig 2.2, each of the larger circles represents a carbon atom in the middle of the formula above. These circles are shown light for the background protein strand and dark for the foreground strand. Similarly, each of the smaller circles in Fig 2.2 represents a hydrogen atom as directly bonded to the middle C atom in the formula above; the light circles again 'belong' to the background protein and the dark circles to the foreground protein. An R group in the background protein is shown simply as R, either directly

above or directly below its C atom. For the foreground protein, the R groups are shown as R. The disjointed solid lines between adjacent large circles in the same sequence represent the peptide linkage, –CONH–, joining two amino acids into a protein. Finally, the dotted lines in Fig 2.2 represent hydrogen bonding between a >C=O group in one strand and an >NH group in the adjacent strand.

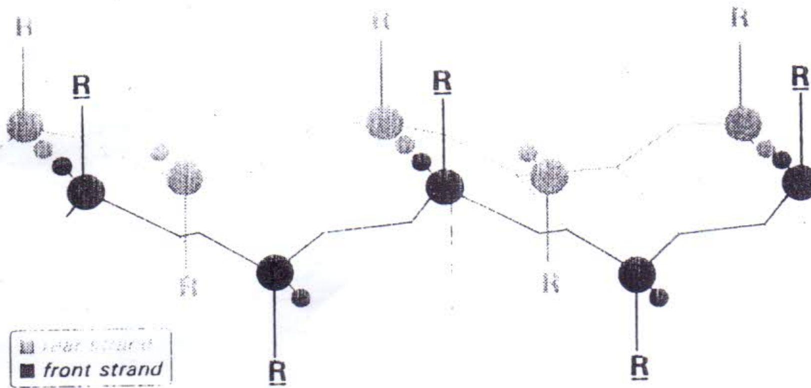



**Fig 2.2**

The dotted lines are interchain hydrogen bonds between the α-CO and –NH groups of the peptide linkages.

Fig 2.2 may be simplified as shown in Fig 2.3, which shows more directly the pleated nature of this type of protein conformation. Fig 2.3 shows three parallel protein strands. A number of points must be emphasised about this simplified diagram:

- each R group is covalently bonded to its C atom;
- the single H atoms bonded to C atoms have been omitted;
- the short *straight* lines now represent peptide linkages;
- the longer dashed lines merely show alignment of the protein strands - they do *not* represent hydrogen bonding. (Hydrogen bonding is shown by dotted lines, but only on the 'face' of the first pleat.)

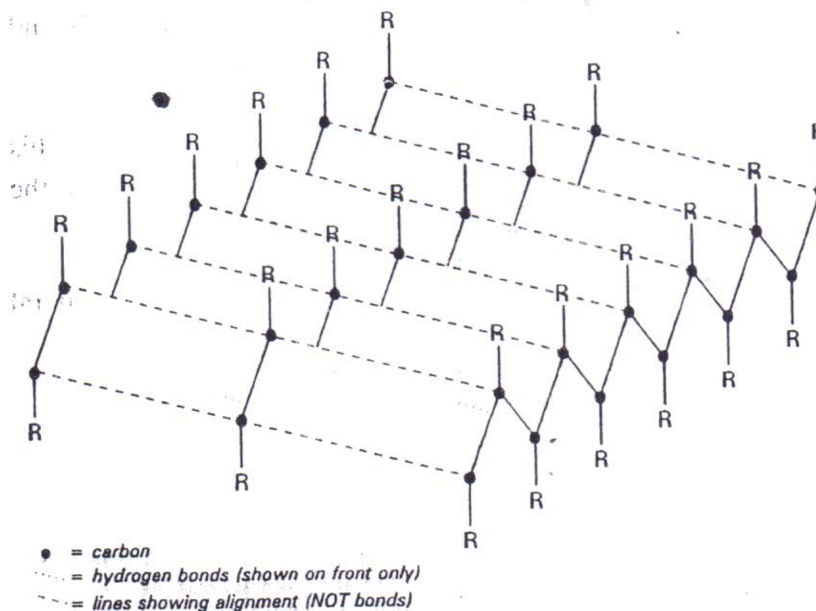



**Fig 2.3**

Diagram of three parallel chains in β-structure, showing the pleated sheet arrangement.

The dashed lines show the alignment of the chains. The dotted lines (shown on front face only for clarity) represent hydrogen bonds between the chains.

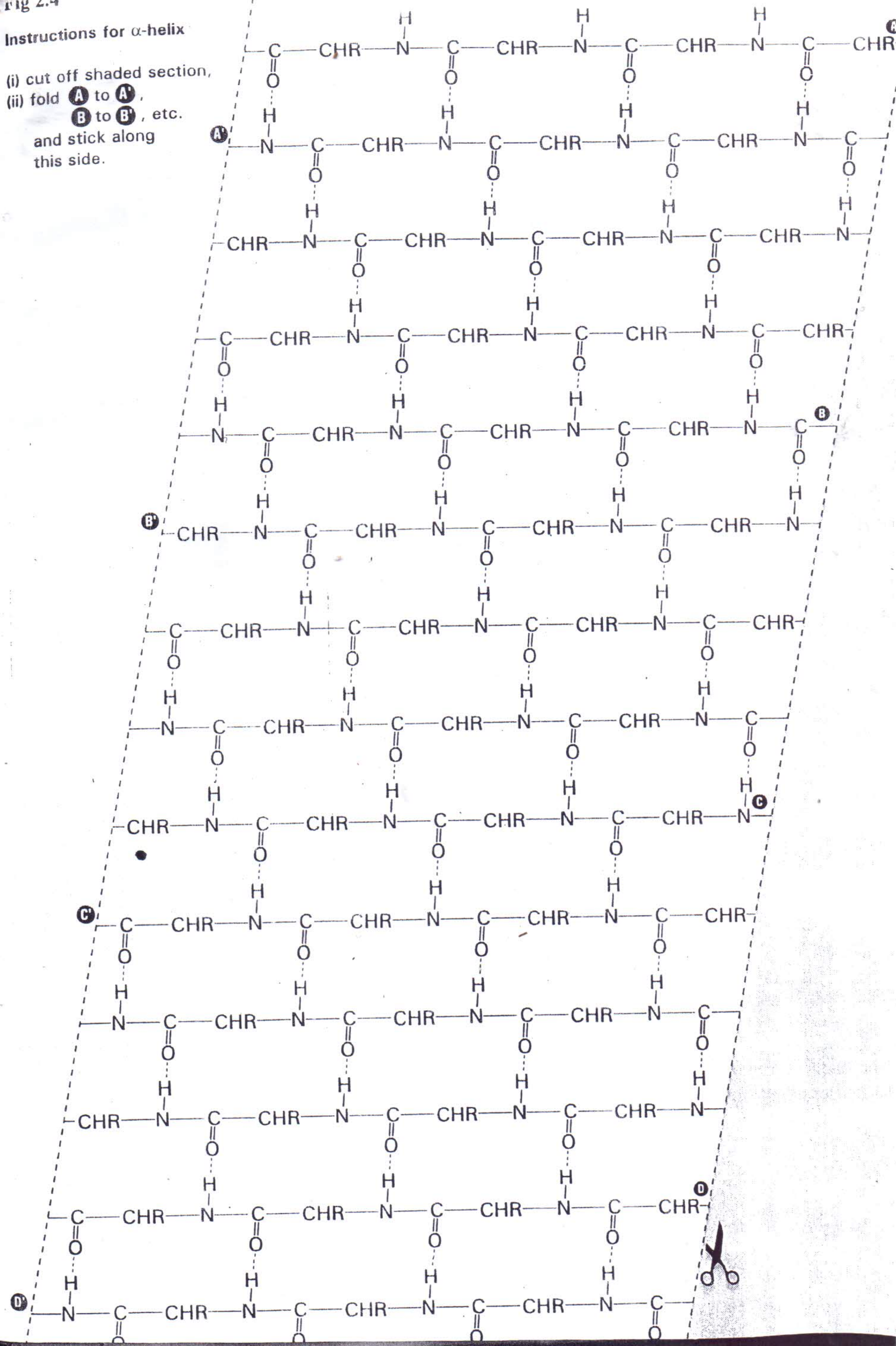
All the R groups project above or below the sheet.

It is also possible to illustrate the relationship between an α-helix and β-pleated protein structure by means of a diagram. The grid in Fig 2.4 is an alternative way of representing a β-pleated sheet. Each horizontal line represents $3^2/_3$ amino acid residues along a protein strand. These lines are shown straight in the grid for clarity but, as indicated by the 'disjointed' peptide linkages in Fig 2.2, a protein strand is not a straight line. Beacuse each horizontal line represents a protein strand, it can be recognised that the grid represents parallel protein strands that lie in a β-pleated conformation.

The relationship between a β-pleated sheet and an α-helix can now be demonstrated as follows.

Take a photocopy of the grid and roll the copy into a cylinder, i.e. the long sides of the copy being close together. Now adjust the cylinder slightly so that A joins to A', B joins to B', etc. The adjusted cylinder now represents a single continuous protein strand: this is the α-helix. The R groups in the α-helix are on the outside of the helix.

Fig 2.4



## (f) Tertiary structure of proteins

The discussion below applies chiefly to globular proteins. These proteins include enzymes, antibodies, some hormones and the globulins of blood serum (which are important in immunology), and are vital in maintaining the organisation of the protoplasm.

The *tertiary structure* is the overall three-dimensional shape of a protein. It is held together by hydrogen bonds and other weak interactions between the R groups (side chains) of the amino acid residues. The α-helix folds up into different patterns but resembles the folded tube of myoglobin which is illustrated on page 19. The folding is due to R group interactions:

(i) electrostatic attractions between charged groups;

(ii) hydrogen bonding between polar groups;

(iii) van der Waals' interactions between non-polar side chains.

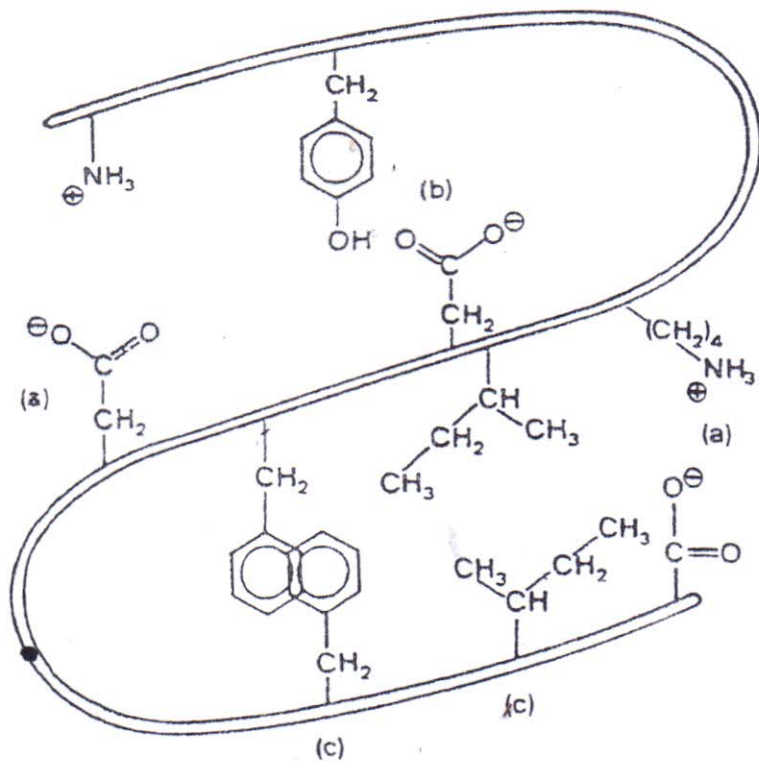
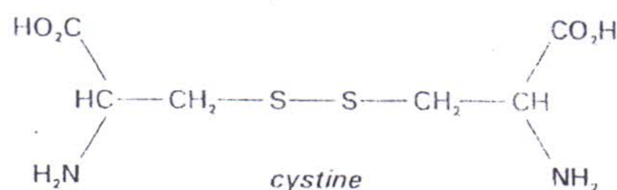



Fig 2.5

The polar R groups attract each other and the non-polar R groups attract each other. The process is similar to oil and water separating out.

The disulphide link in cystine also secures the folding of the α-helix. Cystine has two amino acid residues each of which is in a different part of the peptide strand.



[N.B. Cystine and cysteine should not be confused with each other. They are, however, chemically related - see the cysteine (cys) structure on page 7]

X-ray analysis shows that each type of globular protein has a distinctive folding pattern, but they all appear to exhibit three common features.

1. They show compact folding with little or no room for water molecules in the interior.

2. Nearly all of those R groups which are hydrophilic are on the outer surface of the molecules.

3. On the other hand, half or more of the R groups that are hydrophobic are internally located and, therefore, out of contact with water molecules.

These features help to explain the tendency of globular proteins to be soluble in water.

The diagram below shows the three-dimensional structure of myoglobin - the molecule which binds oxygen reversibly in muscle.

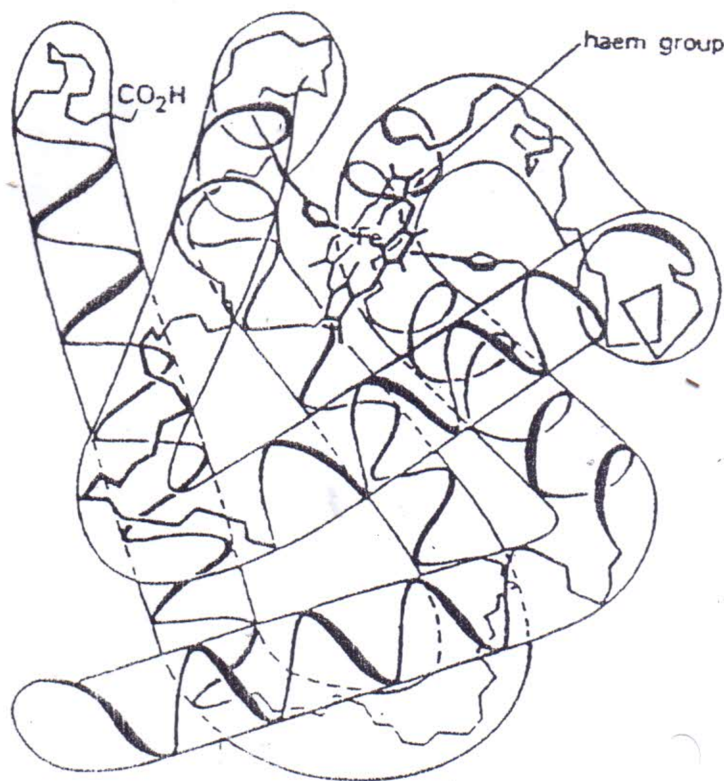



**Fig 2.6**

The diagram shows the helical and non-helical regions and the manner in which they fold around the haem group. Note that the folded molecule still has a high proportion of $\alpha$-helix.

Conditions which in any way disrupt the specific interactions that maintain the polypeptide chain in its preferred confirmation almost always lead to loss of biological activity.

### (g) Quaternary structure of proteins

Proteins containing two or more polypeptide chains are called oligomeric proteins and the individual chains are called the sub-units. A well investigated oligomeric protein is haemoglobin, the oxygen-carrying pigment of the red blood cells, which consists of four polypeptide chains fitting together tightly to form a compact, globular assembly of considerable stability.

Haemoglobin contains two α-sub-units and two β-sub-units (the two types of sub-unit have differing numbers of amino acid residues) and each of these sub-units is non-covalently linked a haem residue. It is the iron in the haem group which actually binds oxygen. Each sub-unit has an irregularly folded conformation in which lengths of pure α-helical regions are separated by bends. Both α- and β-sub-units have about 70 per cent α-helical regions.

In haemoglobin, there is very little contact between one α-sub-unit and the other, nor between the members of the pair of β-sub-units. There is a considerable number of R group contacts between an α-sub-unit and its neighbouring β-sub-unit: these contacts are of the same nature as those which stabilise a tertiary structure.

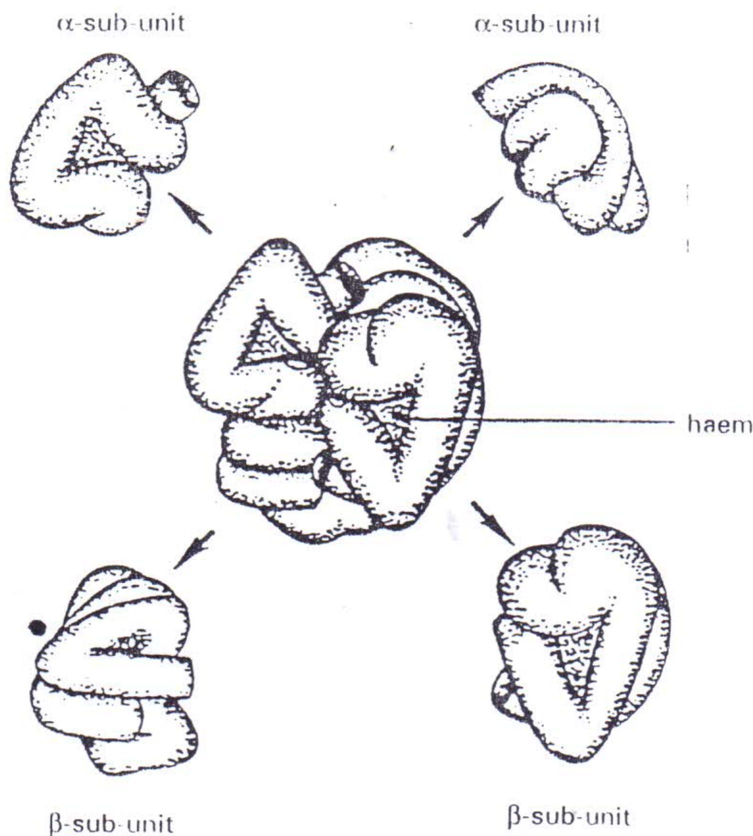



Fig 2.7

Quarternary structure refers to the characteristic manner in which the individually folded polypeptide sub-units are grouped.

### (h) Denaturation of proteins

The majority of proteins retain their biological activity only within a very limited range of temperature and pH. If either soluble or globular proteins are subjected to extremes of pH or high temperatures for only short periods, most of them undergo a physical change - denaturation, in which the most visible effect is a decrease in solubility. No covalent bonds of the polypeptide bonds are broken and the primary structure remains intact.

Most globular proteins undergo denaturation when heated above 60 °C to 70 °C. The coagulation of egg white on heating is an example of this process. The most significant consequence of denaturation is that the protein usually loses its characteristic biological activity, e.g. heating usually destroys the catalytic ability of enzymes. The heat disrupts the weak bonds which hold together the quaternary, tertiary and secondary structure of the protein and a more disordered arrangement results.

Changes in pH protonate or deprotonate the ionic R groups and so disrupt the electrostatic attractions which may be critical to the quaternary and tertiary structure of the protein. The protein is then denatured.

At a certain pH, there is a minimum of electrostatic repulsion between neighbouring molecules and they tend to coagulate and precipitate. Nearly all globular proteins show a solubility minimum with variation of pH.

Extreme heat or extremes of pH usually result in irreversible denaturation of most proteins: however, 8 mol $dm^{-3}$ aqueous urea is commonly used as an agent which reversibly denatures many proteins by competing for the intra-molecular hydrogen bonds which stabilise the structure.

### (i) Enzymes

Enzymes are globular proteins specialised to catalyse biological reactions. As catalysts, enzymes exhibit the general property of being able to accelerate the rate of a chemical reaction without being consumed in it. The enzyme performs its catalytic function by providing an alternative route with a lower activation energy for the reaction.

The reactants catalysed by an enzyme are referred to as its *substrate*.

Catalysts combine transiently with the reactants to produce a transition state having a lower energy of activation than the transition state of the uncatalysed reaction. When the reaction products are formed, the free catalyst is regenerated.

One of the most remarkable aspects of the catalytic power of enzymes is their specificity, so that only certain substrates are acted upon and only a single type of reaction takes place, without side reactions or by-products. These are important factors otherwise the cells would soon be swamped by unwanted chemicals. Although enzymes are very specific compared to man-made catalysts, they vary considerably in their degree of specificity. Some are almost absolutely specific for a given substrate and do not attack even very closely related molecules, such as the optical isomer, whereas others attack a whole class of molecules showing a common denominator of structure but at widely differing rates.

Another remarkable attribute of enzymes is their enormous catalytic power. Although enzymes are proteins, and thus relatively fragile molecules, they bring about their extraordinary catalytic effects in dilute aqueous solution at biological pH and moderate temperature, in sharp contrast to the rather extreme conditions often required to accelerate chemical reactions in the organic laboratory.

The enzyme carbonic anhydrase can catalyse 36 000 000 substrate molecules per minute, per enzyme molecule.

### (j) Enzyme kinetics

The general principles of reaction kinetics apply to enzyme-catalysed reactions, but with one very important feature not usually observed in non-enzymatic reactions, namely SATURATION with substrate - see Fig 2.8.

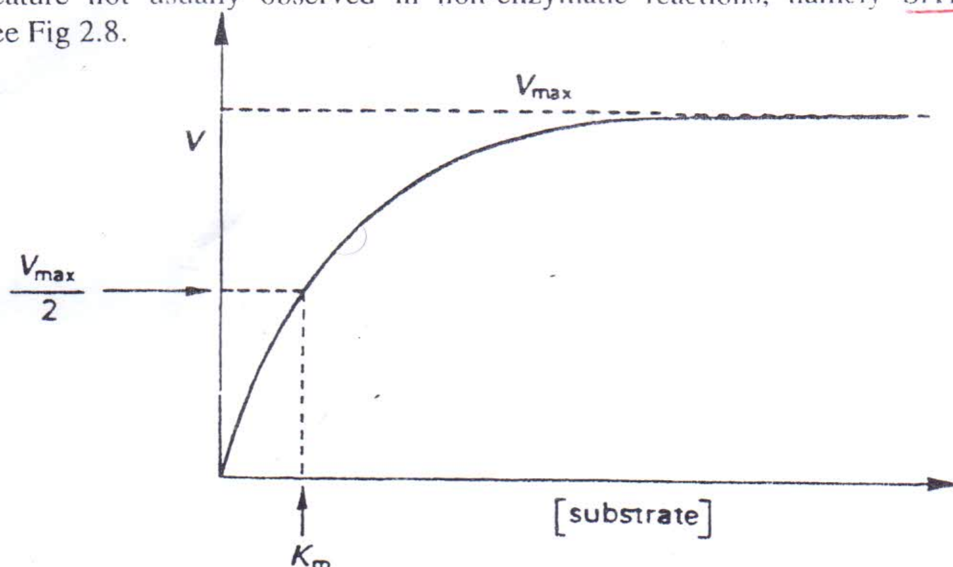



**Fig 2.8: Effect of substrate concentration on the rate of an enzyme-catalysed reaction.**

At low substrate concentration, the reaction velocity, $V$, is nearly proportional to the substrate concentration, and the reaction is, therefore, approximately first order with respect to the substrate. As the substrate concentration is increased, the rate increases to a lesser extent and is no longer nearly proportional to the substrate concentration: the reaction is now mixed order. With a further increase in the substrate concentration, the reaction rate becomes essentially independent of substrate concentration and approaches a constant rate. In this region, the reaction is essentially zero order with respect to the substrate and the enzyme is said to be saturated with its substrate. All enzymes show this saturation effect but they vary widely with respect to the substrate concentration required to produce it. Under intacellular conditions, enzymes are not necessarily saturated with their substrates.

This effect led investigators to the hypotheses that enzymes and substrate react reversibly to form a complex as an essential step of the enzyme-catalysed reaction and also that enzymes possess *active sites* where the substrate is adsorbed and reaction occurs.

Michaelis and Menton developed the general theory of enzyme-catalysed reactions and kinetics.

The theory assumes that the enzyme, $E$, first combines with the substrate, $S$, form an enzyme-substrate complex; this breaks down to form the free enzyme and the product $P$. Both reactions are considered to be reversible.

$$E + S \rightleftharpoons ES$$

$$ES \rightleftharpoons E + S$$

For reactions of this type, at very low $[S]$, $V$ increases almost linearly as a function of $[S]$. As $[S]$ increases further, $V$ increases less rapidly. Eventually, $V$ reaches a limiting value called $V_{max}$. $V_{max}$ is the limiting or maximum velocity at 'infinite' substrate concentration. The vaule of $[S]$ at which $V$ is equal to $V_{max}/2$ is called $K_m$ (the Michaelis constant). The Michaelis constant has the units mol $dm^{-3}$.

The value of $K_m$ for any given enzyme can be easily approximated from a series of simple experiments in which the initial reaction velocity is measured at different initial concentrations of the substrate with a fixed concentration of enzyme, and the values plotted.

It follows that $K_m$ is not a fixed value for a given enzyme; it may vary with the structure of the substrate, with pH and with temperature.

The maximum velocity $V_{max}$ varies widely from one enzyme to another for a given enzyme concentration. It also varies with the structure of the substrate, with pH and with temperature.

The Michaelis constant of an enzyme is important, not only in the mathematical description of enzyme kinetics, but also to the quantitative assay of enzyme activity in tissues and enzyme purification.

Some of these features of enzymatic catalysis are illustrated by glycolysis. Glycolysis is a biochemical pathway (anaerobic fermentation) by which many organisms extract chemical energy from various fuels in the absence of molecular oxygen. The phosphorylation of glucose is the initial stage in glycolysis. There is relatively little free glucose in cells, most intracellular glucose existing in the phosphorylated form. The phosphorylation of glucose to yield glucose-6-phosphate is catalysed by two types of enzyme, *hexokinase* and *glucokinase* which differ in the sugar specificity and affinity for glucose.

$$\text{ATP} + \text{glucose} \longrightarrow \text{glucose-6-phosphate} + \text{ADP}$$

Hexokinase is widely distributed and is normally employed by most cells. The hexokinase present in animal tissue is a regulatory enzyme: it is inhibited by its own product, glucose-6-phosphate. Whenever the cell has a high concentration of glucose-6-phosphate and requires no more for its energy demands, hexokinase is inhibited, thus preventing the formation of more glucose-6-phosphate.

Glucokinase, the other type of phosphorylating enzyme, has a much higher $K_m$ for glucose ($K_m = 2 \times 10^{-2}$ mol $dm^{-3}$) and this requires a much higher concentration of glucose to become fully active than hexokinase ($K_m = 1 \times 10^{-5}$ mol $dm^{-3}$) - see Fig 2.9.

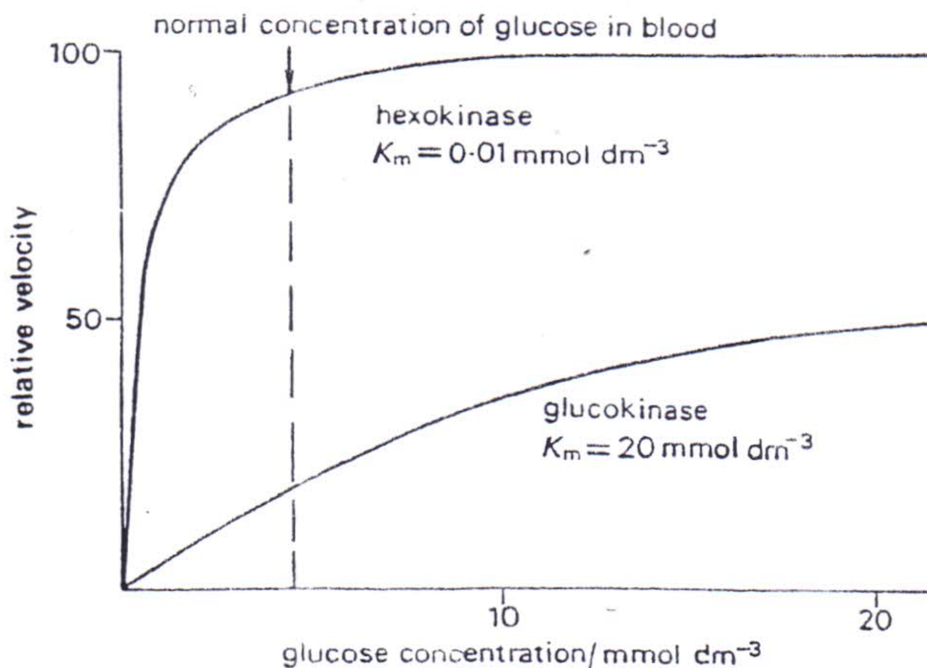



Fig 2.9

Glucokinase is present in liver, where it predominates over hexokinase, but is not present in muscle. It normally comes into play in emergencies when the blood glucose concentration is high. This enzyme is deficient in patients suffering from diabetes.

The catalytic properties and specificity of an enzyme are determined by the chemical groups in a region of the protein surface called the *active site*. The active site of a protein usually constitutes less than 5% of its surface area. The active site is always found in a cleft or crevice in the enzyme structure and has two distinct functions:

1. binding of the substrate, and
2. catalysis.

Generally, different amino acid side chains are involved in each function.

1. Enzymes exhibit remarkable specificity because of the precise fit between their binding site and the substrate. This is the (lock/enzyme)-(key/substrate) analogy. The binding between substrate and enzyme can involve ionic bonds, hydrogen bonds and van der Waals' interactions.

   Enzymes bind the substrates so that the atoms participating in the bond to be made or broken are orientated properly with respect to catalytic groups on the enzyme. The active site is stereospecific so that even optical isomers of a substrate do not fit.

   Diagrammatically (Fig 2.10):

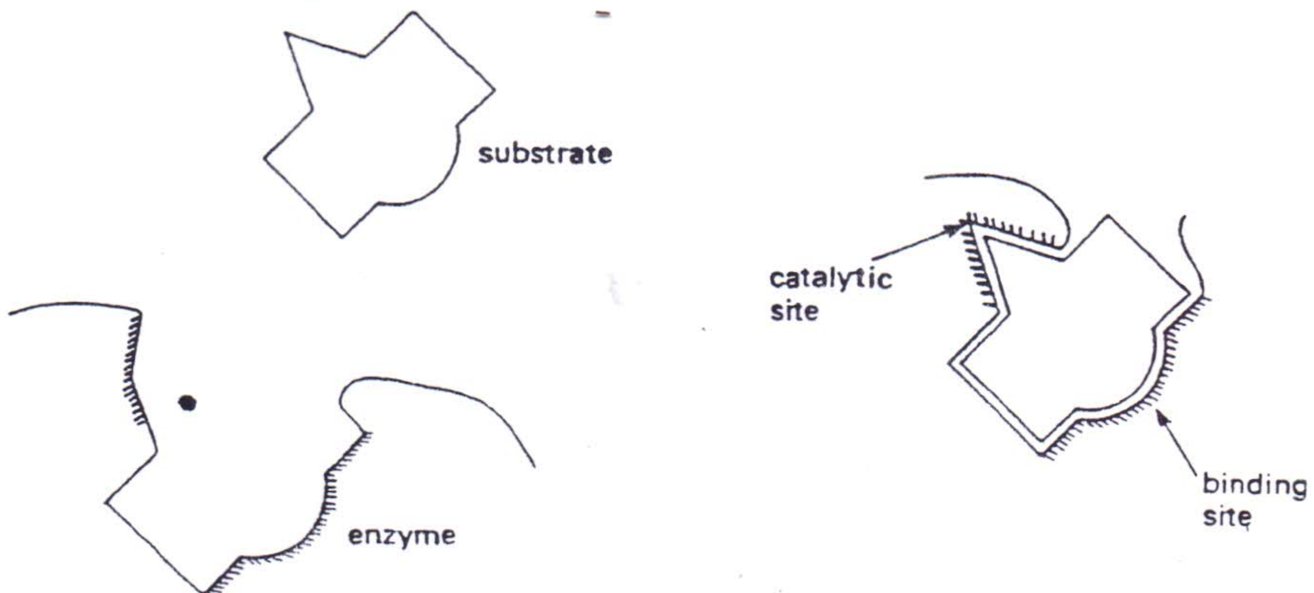



**Fig 2.10**

The enzyme chymotrypsin splits certain peptide linkages in proteins. It is specific for thos peptide linkages in which the carbonyl group is next to an aromatic amino acid residue such a phenylalanine. Its active site has a hydrophobic region for binding the substrate at it hydrophobic group (Fig 2.11).

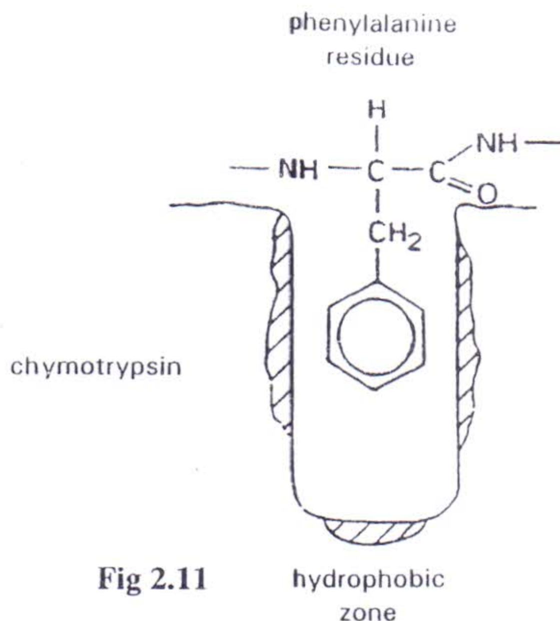



Fig 2.11

Trypsin is another proteolytic enzyme. It hydrolyses peptide linkages in which the carbox group is next to an amino acid residue with a basic group on its side chain, e.g. lysine ( arginine. The active site must provide a region which binds this type of side chain.

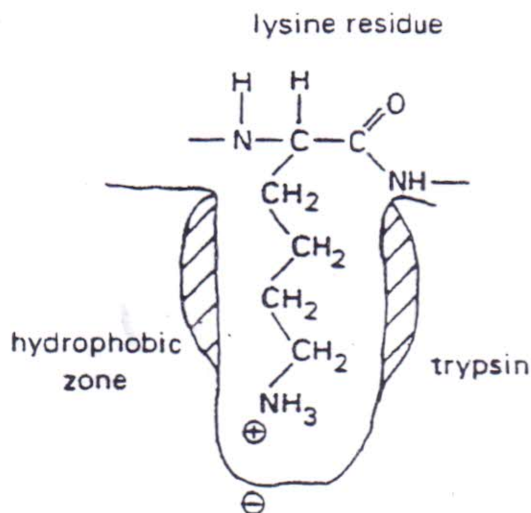
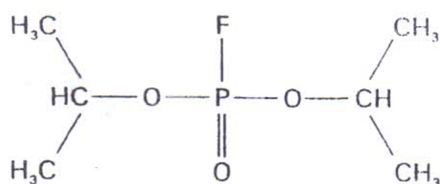



Fig 2.12

2. Only a few of the 20 amino acid side chains participate directly in catalysis. These are general amino acids with polar side chains, e.g. serine, histidine, lysine. These are reactive groups tl can catalyse the reaction of the bound substrate.

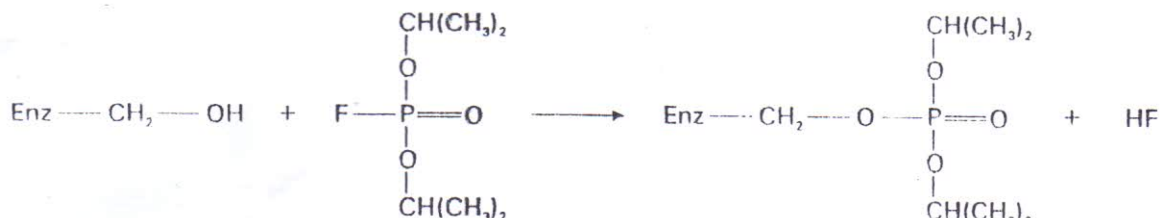
A method used to map active sites is to use a reagent capable of combining with different typ of functional groups in enzyme molecules in order to establish whether such groups a necessary for catalytic activity.

An example of such a reagent is di-(1-methylethyl)fluorophosphate (see next page).




*di-(1-methylethyl)fluorophosphate*

This reagent reacts with certain enzymes to yield inactive derivatives in which the -OH group of a specific serine residue is phosphorylated.




Di-(1-methylethyl)fluorophosphate also deactivates enzymes such as chymotrypsin. In this case, under mild conditions, only the serine residue at position 195 is phosphorylated and this deactivates the enzyme: this indicates that, although chymotrypsin contains a number of serine residues, it is the one on position 195 that is the catalytically reactive residue.

Less drastic inhibition of enzymes is described as *reversible inhibition*. These inhibitors are divided into two groups - (i) competitive, and (ii) non-competitive, inhibitors. These can be recognised experimentally by their effects on the reaction kinetics of the enzyme.

(i) A *competitive inhibitor* can combine with the free enzyme in such a way that it competes with the substrate for binding at the active site. The inhibitor resembles the substrate molecules sufficiently well to form some of the proper interactions of the binding site but it is not sufficiently similar to the substrate to take part in the reaction and thus be released as product (Fig 2.12).

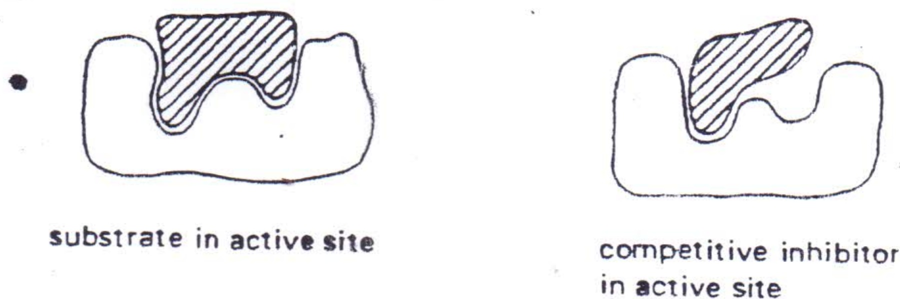



**Fig 2.12**

The percentage of competitive inhibition at a given inhibitor concentration can be decreased by increasing the substrate concentration (Fig 2.13).

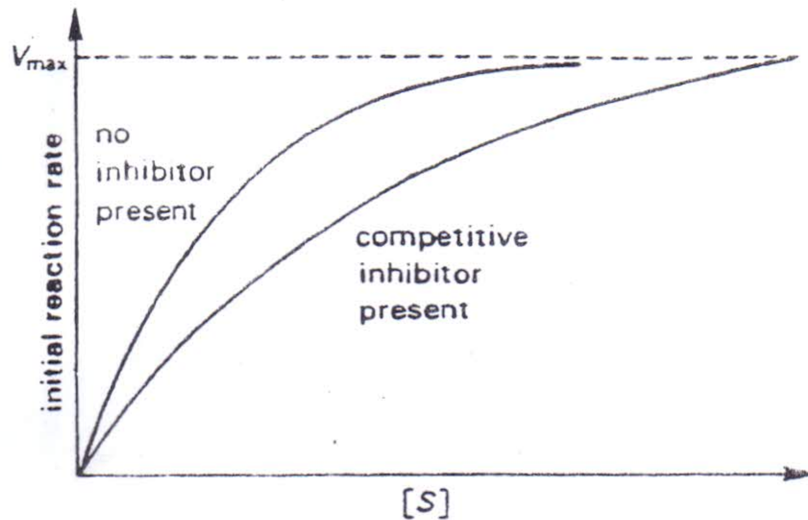



**Fig 2.13**

At high concentrations of the substrate, it is possible to reach $V_{max}$ even in the presence of the inhibitor; however, the effect is to increase the value of $K_m$.

The extent of competitive inhibition depends, therefore, on

1. the concentration of the inhibitor,
2. the concentration of the substrate,
3. the relative affinity of the active site for the inhibitor and substrate.

(ii) A *non-competitive inhibitor* can bind to a site on the enzyme other than the active site. This often deforms the enzyme so that it does not form the *ES* complex readily. The non-competitive inhibitor may also combine with the enzyme-substrate complex.

The inhibitor in this case is not shaped like the substrate and, by combining with some region in or near the active site, prevents access by the substrate; in this type of inhibition, there is no competition between the substrate and the inhibitor. Non-competitive inhibition can be recognised from plots of $V_0$ (initial reaction rate) against $[S]$ (substrate concentration), Fig 2.14.

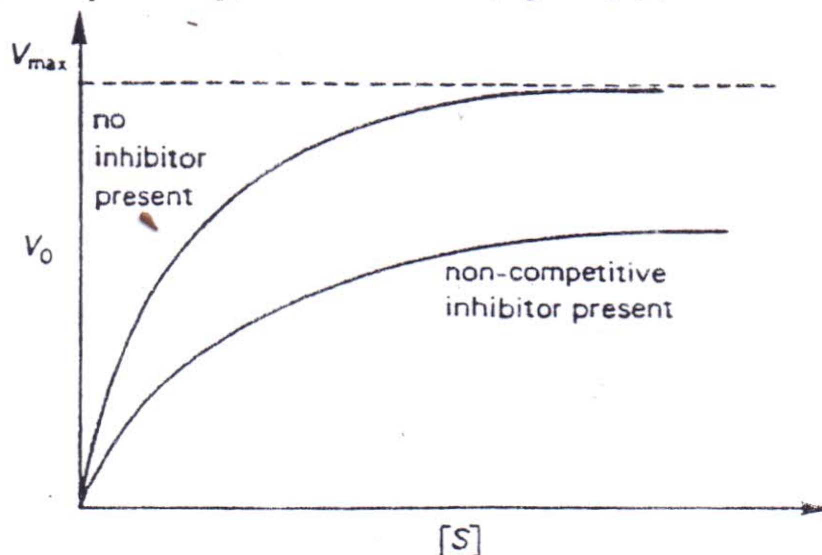


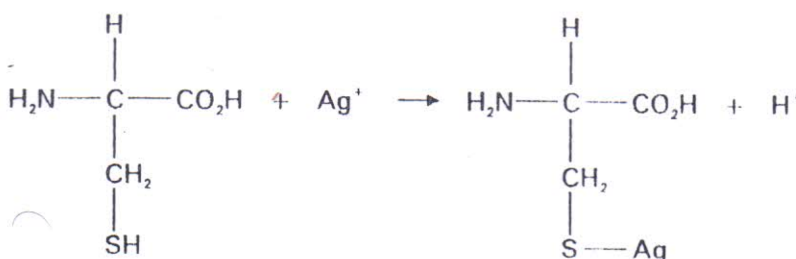

**Fig 2.12**

In this case, $V_{max}$ is decreased by the inhibitor and cannot be restored by increasing the substrate concentration.

Non-competitive inhibition is dependent on

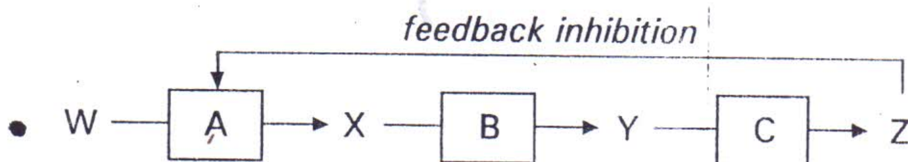
1. the concentration of the inhibitor,
2. the affinity of the enzyme for the inhibitor.

The most common type of non-competitive inhibition is given by reagents that can combine reversibly with some functional group of the enzyme which is essential for maintaining the catalytically active three-dimensional conformation of the enzyme molecule. Some enzymes possessing an essential -SH group are competitively inhibited by heavy metal ions such as mercury or silver.




This suggests that these -SH groups must be intact for the enzyme to retain its normal active conformation.

Enzymes are of considerable importance in the control of the various metabolic pathways and it is possible for the products of metabolism to inhibit the working of certain enzymes in the metabolic chain. If the concentration of a particular product, **Z**, becomes high, then it may act as a non-competitive inhibitor on the first enzyme, **A**, in the pathway.




This is referred to as *feedback control* of metabolism. The rate of the entire sequence is controlled by the concentration of the end product.

### (k) Cofactors and co-enzymes

Amino acid side chains cannot catalyse all the chemical reactions required by biological systems, e.g. no side chains are good electron acceptors. Therefore, small molecules or metal ions (*cofactors*) with additional chemical properties are required in addition to enzymes for the catalysis of some reactions. Some cofactors transiently bind to the enzyme during the reaction; others, called prosthetic groups, are permanently bound to an amino acid side chain in the enzyme's active site.

In enzymes requiring metal ions as cofactors, the metal ions serve as

1. the primary catalytic centre, or
2. a bridging group, to bind substrate and enzyme together through formation of a coordination complex, or
3. an agent stabilising the conformation of the enzyme protein in its catalytically active form.

Carbonic anhydrase is one of a class known as metalloenzymes, since at least one mole of zinc ions per mole of protein is an absolute requirement for its activity. Removal of zinc by other reagents that bind it more tightly completely inactivates carbonic anhydrase.

Some enzymes require cofactors as small as the chloride ion. This is true of the enzyme amylase which hydrolyses starch. It is significantly decreased in activity in the absence of sufficient $Cl^-$.

The term *co-enzymes* applies to organic molecules, often, but not always, derived from a vitamin, which are essential for activity of numerous enzymes. Vitamins are complex organic substances which are vital to the function of all cells and required in the diet of certain species. Vitamins are found bound in co-enzymes. One of the vitamin B compounds, pantothenic acid, is bound in co-enzyme A. Co-enzyme A acts as an acyl group carrier in enzymatic reactions in fatty acid oxidation, fatty acid synthesis, pyruvate (2-oxopropanoate) oxidation and biological acylations.

### (l) Lysozyme

Fleming first discovered lysozyme, in 1922, in nasal mucous and it was soon discovered in many tissues and secretions of the human body, in plants and most plentifully in white eggs. Fleming called the enzyme lysozyme because of its capacity to lyse, or dissolve, bacterial cell walls. Lysozyme is the first enzyme for which the three-dimensional structure has been established and its properties understood in atomic detail. The enzyme hydrolyses certain bonds between sugar residues of a cell-wall polysaccharide.

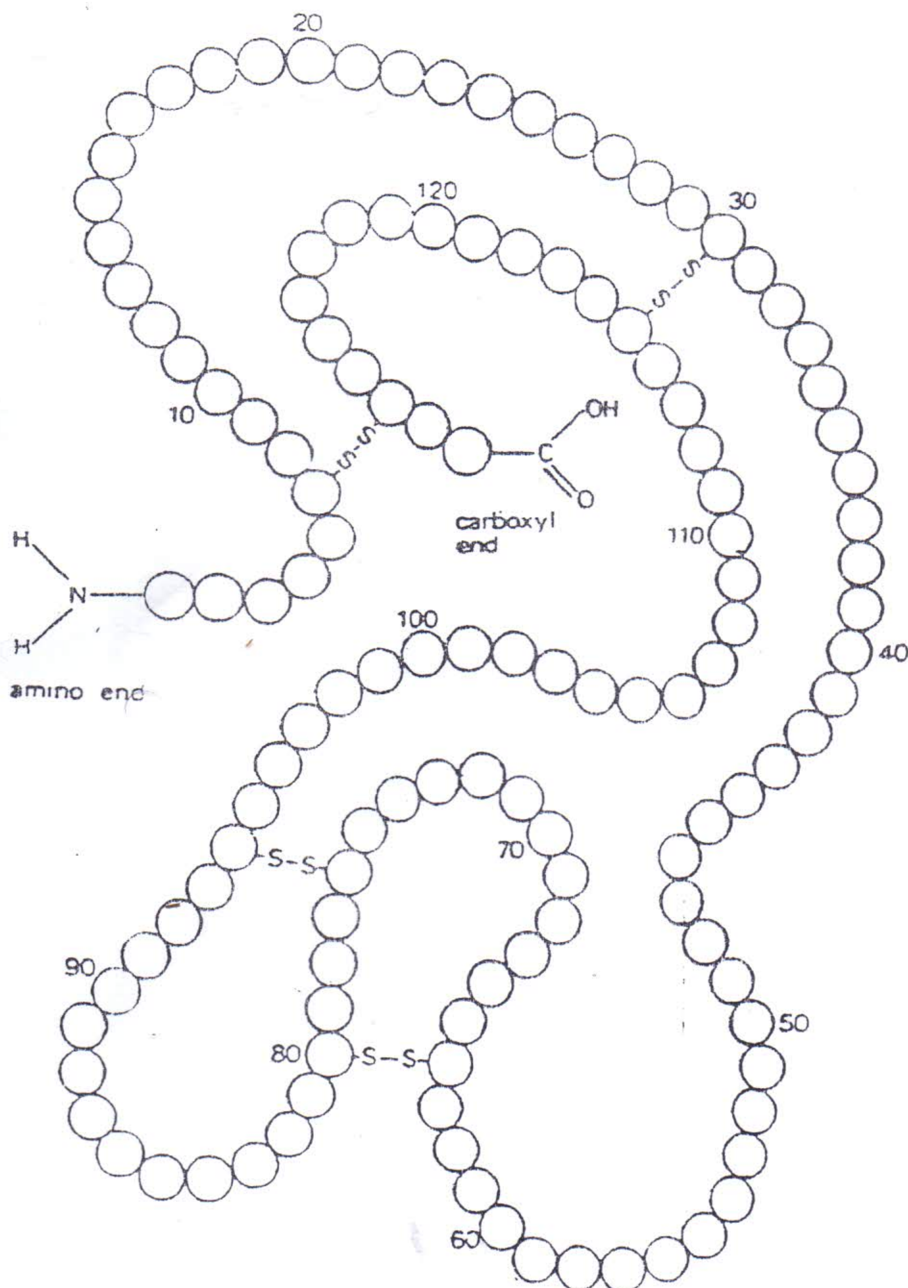



**Fig 2.15**

Like all enzymes, lysozyme is a protein and each molecule of lysozyme consists of a polypeptide chain containing 129 amino acid residues of 20 different kinds. The polypeptide chain is cross-linked in four positions by disulphide bonds between cysteine residues.

As with all enzymes, the properties are dependent on the tertiary structure of the molecule and, therefore, on which parts of the molecule are brought close together in the folded three-dimensional structure. This three-dimensional structure has been investigated using X-ray crystallography.

The molecule contains a large proportion of helices and a region of pleated sheet bound into the polypeptide chain by irregular folds.

The investigation of the lysozyme molecule suggests two generalisations about its conformation that agree with those arrived at from a study of myoglobin. Certain residues with acidic or basic side chains that ionise in water are all on the surface of the molecule, more or less readily accessible to the surrounding liquid. Conversely, most of the markedly non-polar and hydrophobic side chains are shielded from the surrounding liquid by more polar parts of the molecule.

in this way, the system consisting of the protein and the solvent attains a minimum energy.

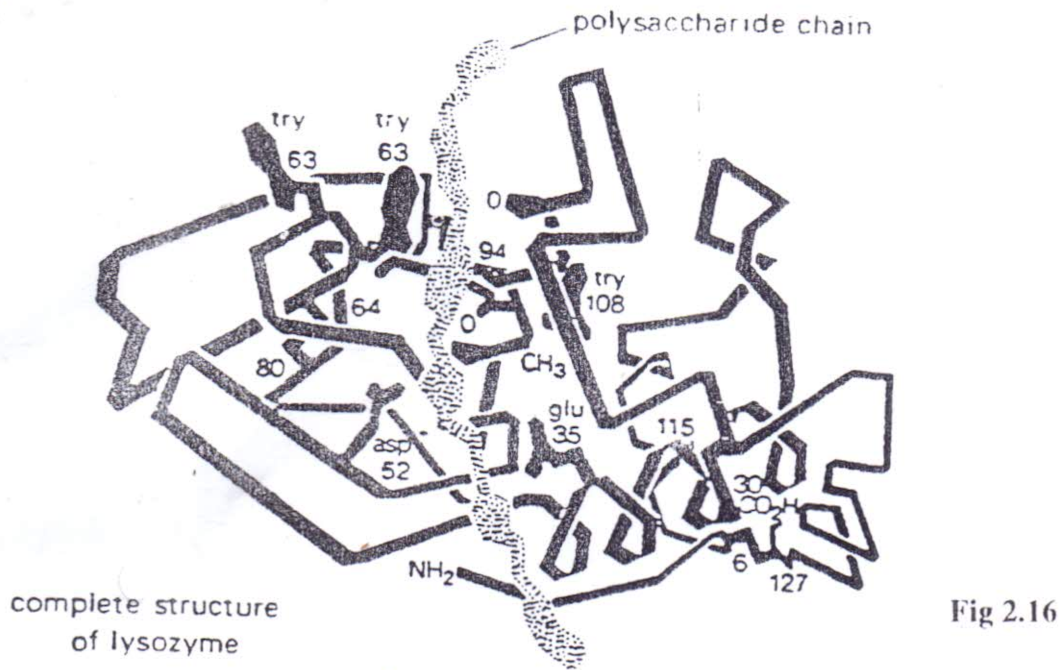



Fig 2.16

The folding of the lysozyme chain to conform to the above facts brings about the formation of a deep cleft running up one side of the molecule. This cleft forms the active site of the enzyme. High resolution X-ray crystallography has shown that a trisaccharide section of the polysaccharide chain is bound to the enzyme, filling the top half of the cleft. It is bound by hydrogen bonds and non-polar interactions. Electron density maps show that the enzyme conformation changes in such a way as to close the cleft slightly. From a knowledge of the position at which lysozyme breaks the polysaccharide chain, it has been possible to investigate the area of the active site which is most likely to act as the centre of catalytic activity. The side chains involved seem to be the carboxyl groups of residues 35 (glutamic acid) and 52 (aspartic acid). During hydrolysis of the bond between two sugar residues, the negative charge on Asp 52 stabilises the positively charged transition state by means of an ionic bond.

## Questions on Section 2: Proteins

1 What is meant by each of the terms *primary, secondary, tertiary* and *quaternary* as applied to the structure of proteins? Where possible, illustrate your answer with examples of the chemical bonding involved in each structure. [10]

2 The tertiary structure of globular proteins is often an α-helix folded so that most of the hydrophilic R groups occur on the outer surface with the hydrophobic R groups internally located.

(a) Describe the structure of the α-helix. Include in your answer a structure which shows the bonds that stabilise the helix. [4]

(b) In each case, give two examples of

(i) a hydrophilic R group,

(ii) a hydrophobic R group. [4]

(c) State and explain the physical property of the globular protein which follows from the hydrophilic R groups being on the outer surface. [2]

3 Proteins are formed from amino acids. Draw the structural formula of a section of a protein chain with the sequence:

alanine glycine alanine

[Glycine is aminoethanoic acid and alanine is 2-aminopropanoic acid.]

Label with an asterisk (*) any carbon atom which is chiral. [4]

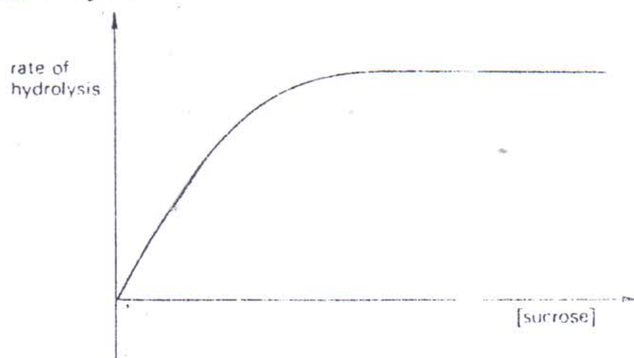
4 For each of the examples below, describe and explain the relevance of hydrogen bonding. In each case, draw clear diagrams of the bonds involved:

(a) the stabilisation of secondary structures of proteins; [5]

(b) the interaction of some R groups in amino acids. [2]

5 (a) Explain, with an equation, what is meant by the *enzymic hydrolysis* of a disaccharide. [4]

(b) The sketch below represents the effect of sucrose concentration on the initial rate of its hydrolysis using the enzyme invertase.




Explain why the hydrolysis is

(i) first order with respect to sucrose when the sucrose concentration is low,

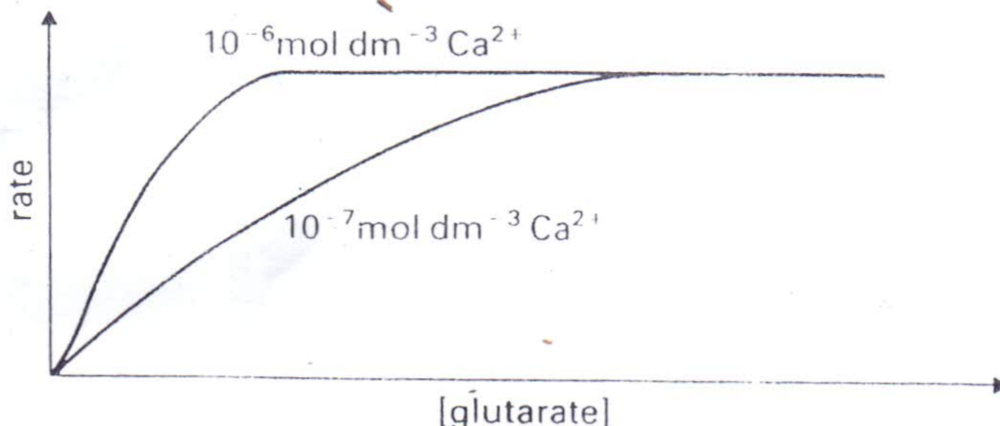
(ii) zero order when the sucrose concentration is high. [3]

(c) Explain how an enzyme can be inhibited from acting effectively in a biochemical process. [3]

6 (a) Describe and explain how the order of an enzyme-catalysed reaction depends on substrate concentration. [3]

(b) The Michaelis constant, $K_m$, is used as an indication of enzyme efficiency. Outline how the constant is determined experimentally. [4]

(c) The sketch below shows how the action of glutarate hydrogenase is affected by calcium ion concentration.




Comment on the shapes of the graphs and explain the action of the calcium ions in the process. [3]

(a) Explain what enzymes are and how they function in biochemical processes. [3]

(b) The different rates at which varying concentrations of glucose are converted into glucose-6-phosphate by the action of hexokinase are given in the table below.

| *concentration of glucose /* mol dm$^{-3}$ | *rate of phosphorylation/* mol min$^{-1}$ |
|---|---|
| 0.000 | 0.0 |
| 0.005 | $3.4 \times 10^{-5}$ |
| 0.010 | $4.6 \times 10^{-5}$ |
| 0.020 | $6.4 \times 10^{-5}$ |
| 0.050 | $8.8 \times 10^{-5}$ |
| 0.100 | $1.18 \times 10^{-4}$ |
| 0.200 | $1.44 \times 10^{-4}$ |
| 0.300 | $1.54 \times 10^{-4}$ |
| 0.400 | $1.60 \times 10^{-4}$ |
| 0.500 | $1.60 \times 10^{-4}$ |

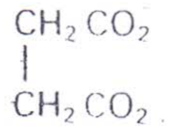
(i) Use graph paper to plot the variation of the rate of phosphorylation against glucose concentration.

(ii) Determine the Michaelis constant for the phosphorylation of glucose with hexokinase. State the units of this constant.

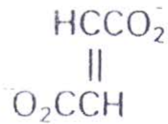
(iii) Explain how the shape of the graph can be interpreted in terms of biochemical activity.

[7]

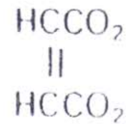
8 The enzyme succinate dehydrogenase recognises and acts on fumarate but not on maleate; it does, however, act on malonate.



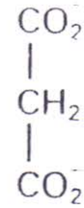

succinate




fumarate




maleate




malonate

In terms of the structures above, explain

(i) the difference in behaviour between fumarate and maleate, [3]

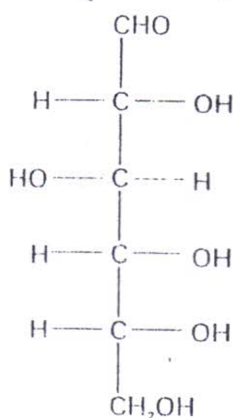
(ii) what sort of enzyme action takes place with malonate. [3]

## 3. Carbohydrates

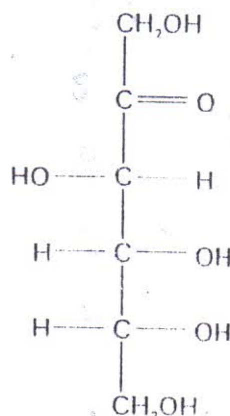
Carbohydrates are polyhydroxy aldehydes or ketones. Monosaccharides, simple sugars, consist of a single polyhydroxy aldehyde or ketone unit, and the most abundant monosaccharide is the six-carbon D-glucose. D-glucose is the major fuel for most organisms and the basic building block of the most abundant polysaccharides such as starch and cellulose. Polysaccharides contain many monosaccharide units formed in long linear or branched chains. Most polysaccharides contain recurring monosaccharide units of only a single kind or two alternating kinds, contrasting with proteins. Polysaccharides have two major functions, as a storage form of fuel and as structural elements.

### (a) Monosaccharides

Monosaccharides have the empirical formula $(CH_2O)_n$ where n > 3. The carbon skeleton of common monosaccharides is unbranched and, in a molecule, each carbon atom except one contains a hydroxyl group. If the carbonyl group is terminal, the monosaccharide is an aldehyde derivative (aldose); if the carbonyl group is in any other position, the monosaccharide is a ketone derivative (ketose).

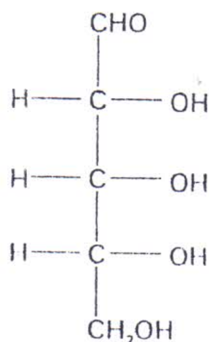



*D-glucose*
*aldohexose*




*D-fructose*
*ketohexose*

The hexoses are the most abundant monosaccharides but the five carbon aldopentoses are important components of nucleic acids, e.g. ribose.

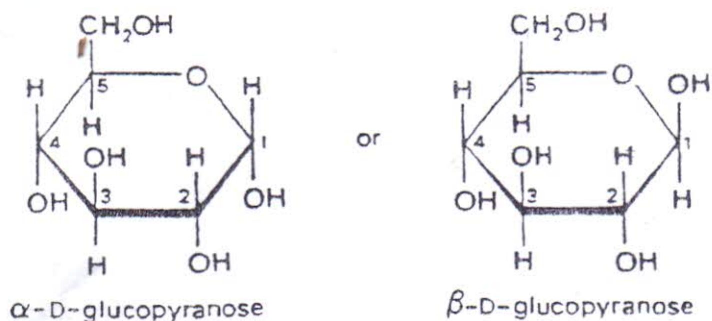



*D-ribose*

All simple monosaccharides are white crystalline solids, freely soluble in water (by formation of hydrogen bonds between the -OH groups and water molecules) and are insoluble in non-polar solvents.

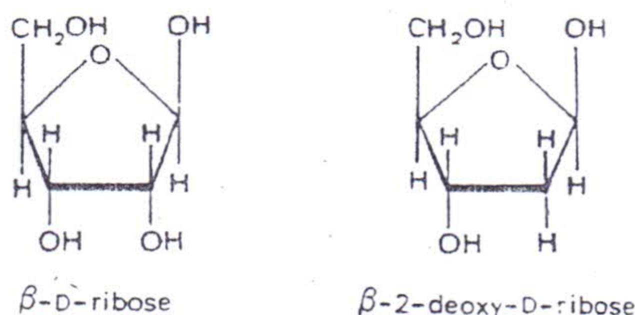
All monosaccharides contain one or more asymmetric (chiral) carbon atoms and, therefore, show optical isomerism.

When observed in aqueous solution, many monosaccharides appear to have one more optically active centre than is given by the open chain structure. D-glucose may appear in two different isomeric forms. α-D-glucose and β-D-glucose. They have the same elementary compositions but differ in physical and chemical properties. From investigations, it has been deduced that the α and β isomers of D-glucose are six-membered ring structures formed by the addition reaction of the -OH group at $C_5$ with the aldehyde group at $C_1$. The ring structures are called pyranose rings.




These two optical isomers, which differ only in the configuration about carbon atom 1, are called *anomers*.

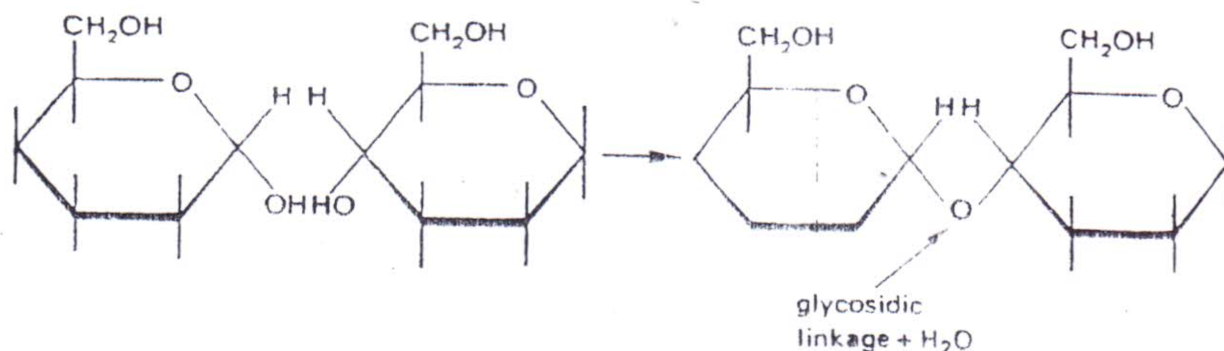
The pentose sugars form five-membered rings when they occur in nucleic acids or nucleotides. e.g.




The five-membered ring is approximately planar whereas the six-membered pyranose ring is puckered.

### (b) Disaccharides

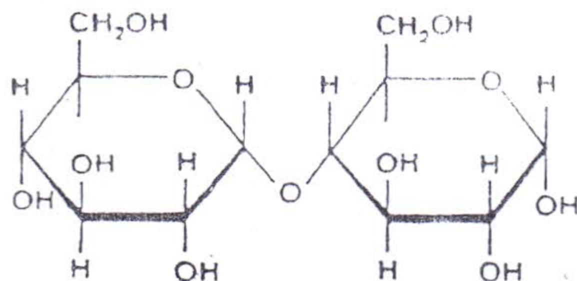
These consist of two monosaccharides joined by a glycosidic linkage. This linkage is formed by the reaction of the anomeric carbon atom with the -OH group of another monosaccharide, resulting in the elimination of a water molecule:




The most common disaccharides are maltose, lactose and sucrose.

Maltose is formed by the action of amylases on starch and contains two glucose residues. The -OH of the anomeric $C_1$ of the first glucose residue reacts with -OH group of $C_4$ of the second glucose residue.

Both of the glucose residues are in the pyranose form and the anomeric $C_1$ of the first residue is $\alpha$. The second glucose residue may have $\alpha$ or $\beta$ configuration at the anomeric carbon atom, and both forms are produced by enzyme action.

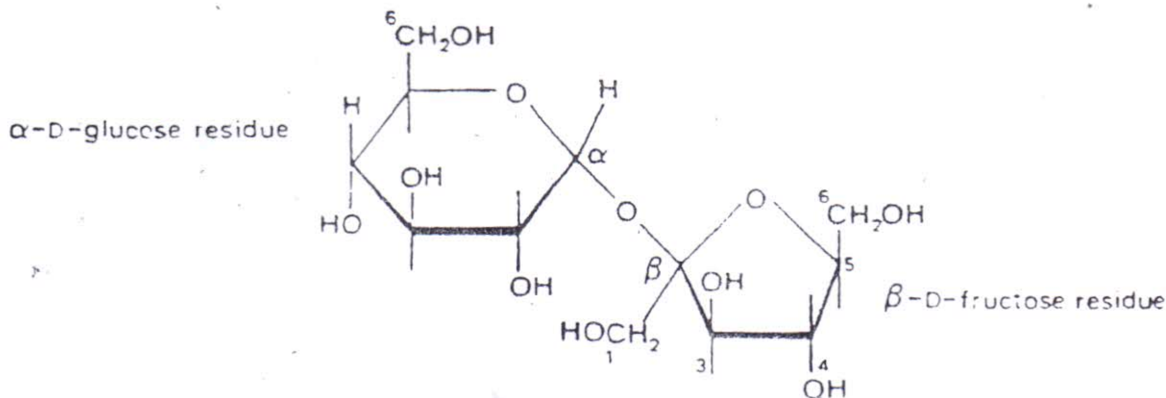



maltose (1 α — 4 link)

The second residue undergoes oxidising action and, therefore, is a reducing sugar. There must be a free aldehyde group for the saccharide to act as a reducing sugar.

Lactose occurs in milk but not otherwise in nature; it also has a free anomeric carbon atom and is, therefore, a reducing sugar.

Sucrose (cane sugar) is a disaccharide of fructose and glucose and is extremely abundant in the plant world.




sucrose 1α−2β link

It is unusual among the disaccharides in that it contains no free anomeric carbon atom: the two anomeric carbon atoms of the hexoses are linked and it does not, therefore, react as a reducing sugar.

**(c) Polysaccharides**

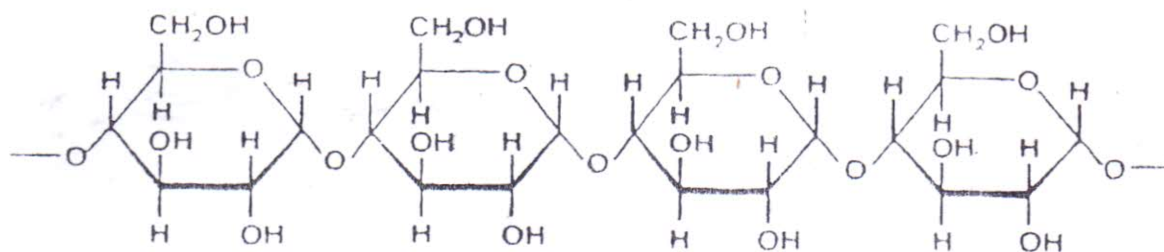
Polysaccharides, like proteins, are condensation polymers. Polysaccharides have high relative molecular masses and, on complete acid or enzyme hydrolysis, yield monosaccharides or monosaccharide derivatives. D-glucose is the most prevalent monosaccharide unit in polysaccharides.

*Storage polysaccharides*

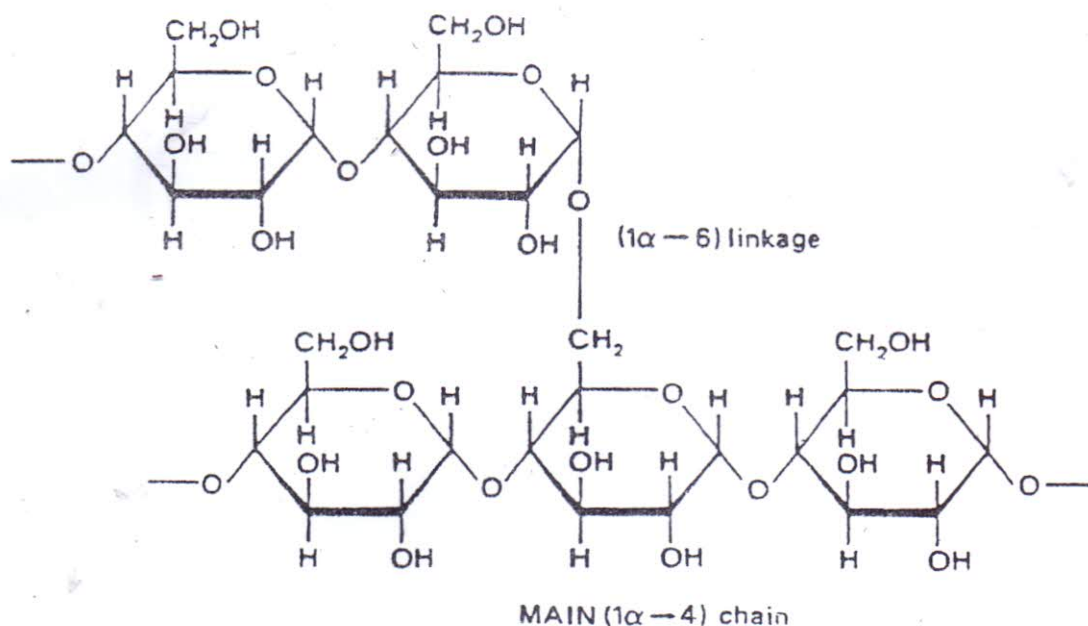
The most abundant polysaccharides of this type are *starch* in plants and *glycogen* in animals.

Starch occurs in two forms: α-amylose and amylopectin.

*α-amylose* consists of long unbranched chains in which all the D glucose units are bound in (1α - 4) linkages. The chains vary in relative molecular mass from a few thousand to 500 000. Amylose is not truly soluble in water but forms hydrated micelles which turn aqueous iodine blue. In these micelles, the polysaccharide chain is twisted into a helical coil.



By contrast, amylopectin is a highly branched molecule, the length of each branch averaging between 24 and 30 glucose residues. The backbone of glycosidic linkages is (1α - 4) but the branch points are (1α - 6) linkages. Amylopectin forms colloidal solutions which give a red-violet colour with iodine. Its relative molecular mass may be as high as 1 000 000.




Glycogen is the storage polysaccharide of animal cells, particularly in the liver and muscles. It is very similar to amylopectin but the branches occur more frequently, i.e. at every 8 to 12 glucose residues. Like amylopectin, the branches are (1α - 6) but its relative molecular mass is much higher, going up to several million.

Both starch and glycogen form useful storage molecules due to their relative insolubility and the fact that enzymes can easily hydrolyse these polymers to give soluble monomers which release a significant amount of energy on oxidation.

*Structural polysaccharides*

These polysaccharides serve as structural elements in cell walls and connective tissue, where they give shape, elasticity or rigidity to plant and animal tissues.

The most abundant structural polysaccharide in the plant world is cellulose. Like starch and glycogen, it is a polymer of D-glucose. It is a linear polymer with (1β - 4) linkages and forms insoluble molecules with relative molecular masses of approximately one million. The arrangement of these linear molecules into fibres in the cell walls of plants gives great tensile strength to their structure.

Cellulose is a major component of wood and cotton.

## Questions on Section 3: Carbohydrates

1 (a) The full structural formula of deoxyribose is given below. Draw the full structural formula of ribose. [1]

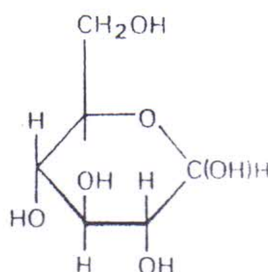
(b) Explain why monosaccharides have the following properties:
   (i) They are soluble in water.
   (ii) Their aqueous solutions rotate the plane of plane-polarised light.
   (iii) On heating a soild monosaccharide, a black solid remains and a colourless vapour is evolved. [3]

(c) The initial stage of glycolysis is the phosphorylation of glucose.
   (i) Copy and complete the equation below:

   glucose ................................ = glucose-6-phosphate + ......................................

   (ii) Draw the full structural formula of glucose-6-phosphate.
   (iii) Name an enzyme which catalyses the reaction in (i).
   (iv) Name a substance which reversibly inhibits this reaction. [5]

2 The glucose molecule may be representd by the strcture below which may be helpful in answering this question.




(a) By means of suitable structural formulae, explain why
   (i) glucose can form two different (1,4) disaccharides, [2]
   (ii) each of these can exist in two different forms. [2]

(b) Explain briefly the difference in structure and function between celluloses and starches derived largely from glucose units. [4]

(c) Explain why wood pulp is indigestible by humans. [2]

## 4. Lipids

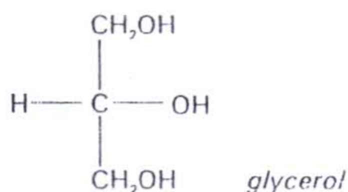
These are water insoluble organic compounds, which can be extracted from organisms by non-polar solvents (trichloromethane, and ethoxyethane). Lipids have a variety of biological functions:

1. structural components of cell membranes,
2. storage and transport forms of metabolic fuel,
3. protective surface coating for organisms,
4. insulation.

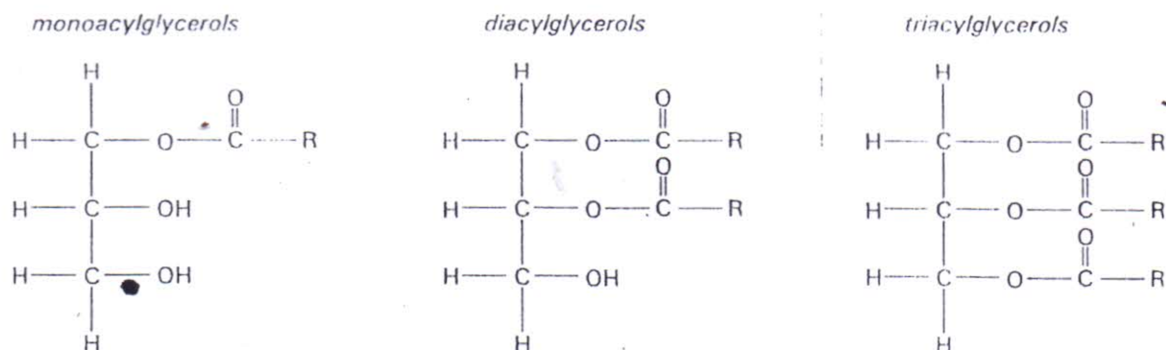
Some compounds classed as lipids have very considerable biological activity, e.g. hormones and vitamins.

### (a) Glycerides (acylglycerols)

Most lipids are esters and the greater proportion of these are esters of the trihydric alcohol, glycerol (propan-1,2,3-triol), with long-chain carboxylic acids.

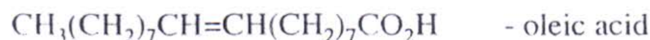



There are three -OH groups, so the following glyceride esters may form




R is a long hydrocarbon chain which may be saturated or unsaturated. Typical long chain/carboxylic acids (fatty acids) are:

$CH_3(CH_2)_{14}CO_2H$ - palmitic acid



$CH_3(CH_2)_7CH{=}CH(CH_2)_7CO_2H$ - oleic acid

Monoacylglycerols and diacylglycerols are formed during fat digestion but the most abundant lipids in living cells are the triacylglycerols. They are the main energy store of living cells as, mass for mass, they release more energy on oxidation than do carbohydrates.

Triacylglycerol deposits provide insulation for animals in cold climates.

Triacylglycerol may have different fatty acids esterified to each of the three carbon atoms of the glycerol and the melting point of the compound depends on these acids. The saturated fatty acids

produce fats which are solids at room temperature and unsaturated fatty acids produce oils which are liquids at room temperature.

The hydrolysis of the ester linkage can be catalysed by enzymes (lipases) or by heating with aqueous mineral acids and alkalis.

### (b) Phosphoglycerides

These are the second largest class of lipids and are major components in all cell membranes but are only present elsewhere in cells in very small amounts.

In these compounds, one of the hydroxyl groups of glycerol is esterified with phosphoric acid and the other hydroxyl groups are esterified with fatty acids.

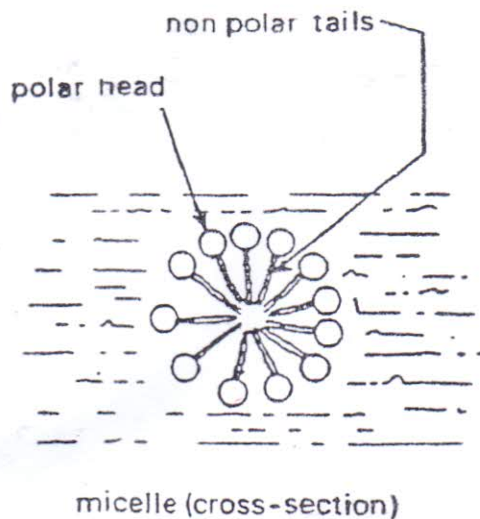
```
              CH2—O—C—R1
             /      ||
            /       O
R2—C—O—C—H
   ||       \       OH
   O         \      |
              CH2—O—P—O—X
                    ||
                    O
```

X is contributed by an alcohol

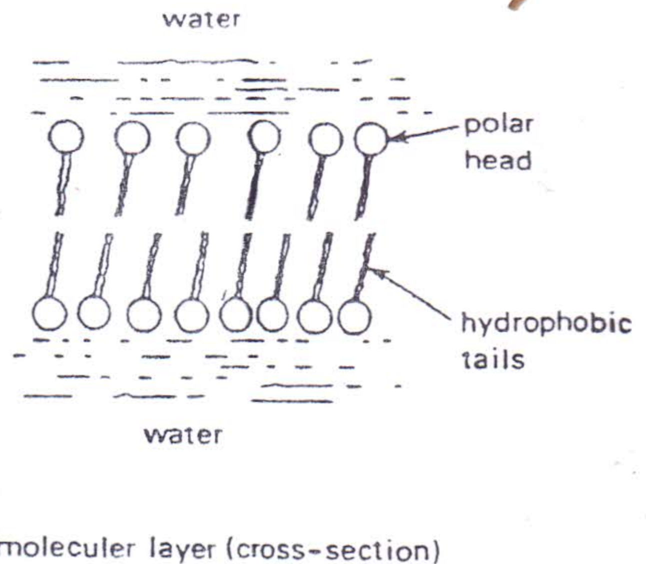
Phosphoglycerides contain a polar head and non-polar hydrocarbon tails and are referred to as polar lipids.

```
            X
            |
            O
            |                 } Polar head
       ⁻O—P=O
            |
            O
            |
CH2—CH—CH2
 |      |
 O      O
 |      |                     } Non polar tails
 C=O   C=O
 ~      ~
```

Phosphoglycerides are waxy solids, which are usually extracted using a trichloromethane-methanol mixture. With water, phosphoglycerides are only very slightly soluble to give a true solution, and a large amount is present as micelles or bimolecular layers (Fig 4.1)




micelle (cross-section)




bimoleculer layer (cross-section)

**Fig 4.1**

In these aggregates, the hydrocarbon tails are shielded from the aqueous phase and these structures are often described as forming due to *hydrophobic interactions.*

Triacylglycerols do not form bilayers.

*Membranes*

Specific lipids and enzymes associate with each other by hydrophobic interaction between the non-polar regions of the lipids and the proteins Most membranes contain approximately 40% lipid and 60% protein, the lipids being predominantly polar in nature. e.g. phosphoglycerides. The ratio of the polar lipids varies according to the type of membrane system, the organ and the species. The membrane proteins can be divided into two groups:

(a) the peripheral proteins which are only loosely attached to the membrane surface. These can be fairly easily removed by gentle extraction techniques.

(b) the integral proteins which constitute approximately 70% of the total membrane proteins. They are tightly bound to the lipid portion and can be removed only by very drastic treatment.

The most recent model of the membrane structure is the *fluid mosaic model* which suggests that the phospholipids are arranged in a bilayer forming a fluid core. In this bilayer, the phospholipids can move laterally giving the bilayer flexibility, high electrical resistance and impermeability to highly polar molecules. The fluid mosaic model suggests that the proteins are globular and that some of them are bedded into the bilayer and others completely span the membrane (Fig 4.2).

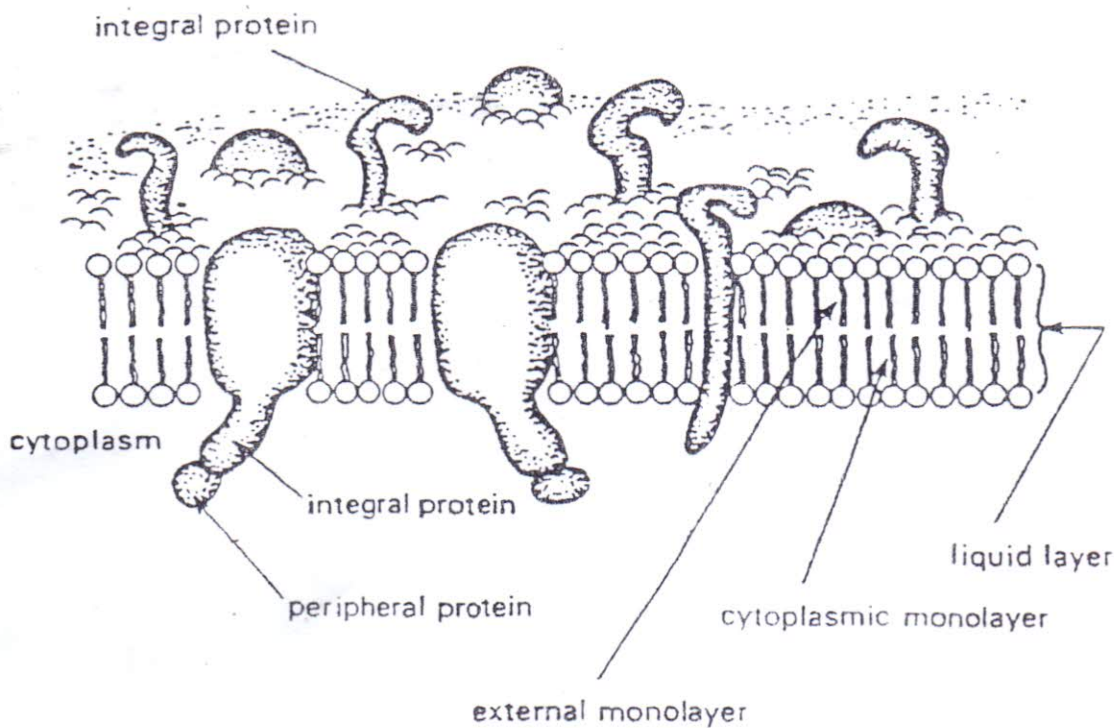



**Fig 4.2**

Many of the integral membrane proteins must be involved with the selective transport of compounds across the membrane.

Membrane proteins, both integral and peripheral, can be divided into two broad groups depending on their role in the membrane: (a) structural proteins, (b) dynamic proteins. Structural proteins are those which hold the membrane assembly together: they are often elongated and fibrous in shape and are found on the hydrophilic lipid surface. Dynamic proteins are those which participate in the cellular process at the membrane level. Three groups of dynamic proteins are generally present in all types of cells:

(i) *transport proteins* - those concerned with the passage of substances into and out of the cells;

(ii) *catalytic proteins* - those that participate as enzymes in reactions occurring at the membranes;

(iii) *receptor proteins* - found with some other substance (hormone, toxin, transmitter) on the exterior side of the membrane providing a signal that alters reactions occurring inside the cell or the membrane.

*Membrane Transport*

The passage of substances into and out of cells and their transport between the cytoplasm around sub-cellular organelles is governed by membranes. Membranes behave as semipermeable partitions - thus controlling the movement of materials across the membrane. Depending on the nature of the substance and the membrane involved, direct transport may occur by one of two general processes.

*Passive transport*

When a substance moves across the membrane by diffusion from a region of high concentration to a region of lower concentration without any energy expended, the movement is called passive transport. There are two types of passive transport: *(a)* simple diffusion, *(b)* facilitated diffusion.

Simple diffusion occurs without any interaction with a membrane's protein. The movement occurs through the lipid bilayer, through pores in the membrane caused by discontinuities in the lipid bilayer or through channels provided by membrane proteins that span the membrane and hence are not greatly affected by the fluidity of the membrane. Substances which move in this way are water, certain inorganic ions and some lipid-type substances.

Facilitated diffusion involves the use of a carrier molecule, usually a protein present in the membrane; the mobile carrier travels from one side of the membrane to the other and is, therefore, markedly influenced by the state of fluidity of the membrane.

It is conceivable that there are other methods of transport between these extremes such as a carousel type in which a carrier molecule rotates in the bilayer but, as with facilitated diffusion, little is known as to the manner in which the protein moves in the bilayer.

*Active transport*

When the passage of a substance moves against a concentration gradient (low to high), it requires the expenditure of energy by the cell. The energy is usually derived directly or indirectly from the hydrolysis of ATP. Like facilitated diffusion, the active transport mechanism requires the participation of membrane proteins. In some cases the transport is multimolecular requiring more than one membrane component and in some instances hormones may be required to stimulate one of the steps.

Possibly the most important and widely distributed active transport system in cells is responsible for the movement of $Na^+$ and $K^+$ across cell membranes. The system, referred to as the $Na^+$-$K^+$ pump, is responsible for maintaining a cell interior that is high in $K^+$ and low in $Na^+$ by moving $K^+$ into the cell and $Na^+$ out of the cell. Both exchanges are against the concentration gradients since the $K^+$ level of the plasma is lower and the $Na^+$ level is higher in the plasma than in the cell cytoplasm.

This ionic imbalance is important in the regulation of water content of the cell protein biosynthesis in the cell and in the functioning of nerve and muscle cells. The protein involved in the lipid bilayer is called $Na^+$-$K^+$-ATP ase, which spans the entire membrane.

A large percentage of the ATP produced in the body is used to maintain the $Na^+$-$K^+$ distribution in the tissues.

## Questions on Section 4: Lipids

1 (a) What are cell membranes and what is their simple function? [2]

(b) One of the structural components of membranes is the group known as phospholipids. These may be represented as:

(i) Explain what this drawing represents.

(ii) Copy this simple drawing of a benzene-water interface.

benzene
___
water

Indicate on your copy how two phospholipds would arrange themselves at this interface. [3]

(c) Sketch and label a small section of the bilayer formed by phospholipids. [1]

(d) What other major biochemical material, other than phospholipids, is found in membranes? [1]

(e) Explain what is meant by *active transport.* [3]

2 The major components of membranes are phospholipids and proteins. The phospholipids form a bilayer which is interspersed with protein molecules. Transport proteins extend av=cross the membranes and catalyse the transport of specific molecules and ions into and out of the cell. Transport can be passive or active.

(a) Give the full structural formula of a typical phospholipid. (If you are unable to draw a structural formula, marks can be scored by a simple labelled diagram of a phospholipid.) [3]

(b) Explain, using your answer to (a), how bilayers are constructed. State the nature of the chemical attractions at each end of the phospholipid. [3]

(c) Explain what is meant by *active transport.* Describe briefly one common example of active transport. [4]

## 5. Nucleic Acids

### (a) The monomers

Genetic information is stored and transmitted by the nucleic acids DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). The monomeric units of the nucleic acids are nucleotides, and each type of nucleic acid is distinguished from another by the sequence in the bases of its nucleotide monomers.

The monomeric units of DNA are described as deoxyribonucleotides and those of RNA are ribonucleotides. Each nucleotide contains three components:

(i) a nitrogenous heterocyclic base (a derivative either of pyrimidine or of purine),
(ii) a pentose sugar,
(iii) a phosphate group.

A nucleic acid is composed of a very long ribose-phosphate backbone with an organic base attached to each ribose (Fig 5.1)

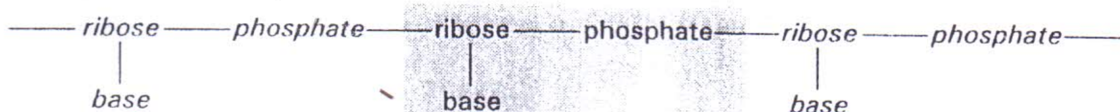



**Fig 5.1**

The unit shaded above is a *nucleotide*. This unit is not, strictly, a repeat unit because the bases attached to ribose may differ along the chain.

In both DNA and RNA, the heterocyclic base can be one of four different structures (Fig 5.2).

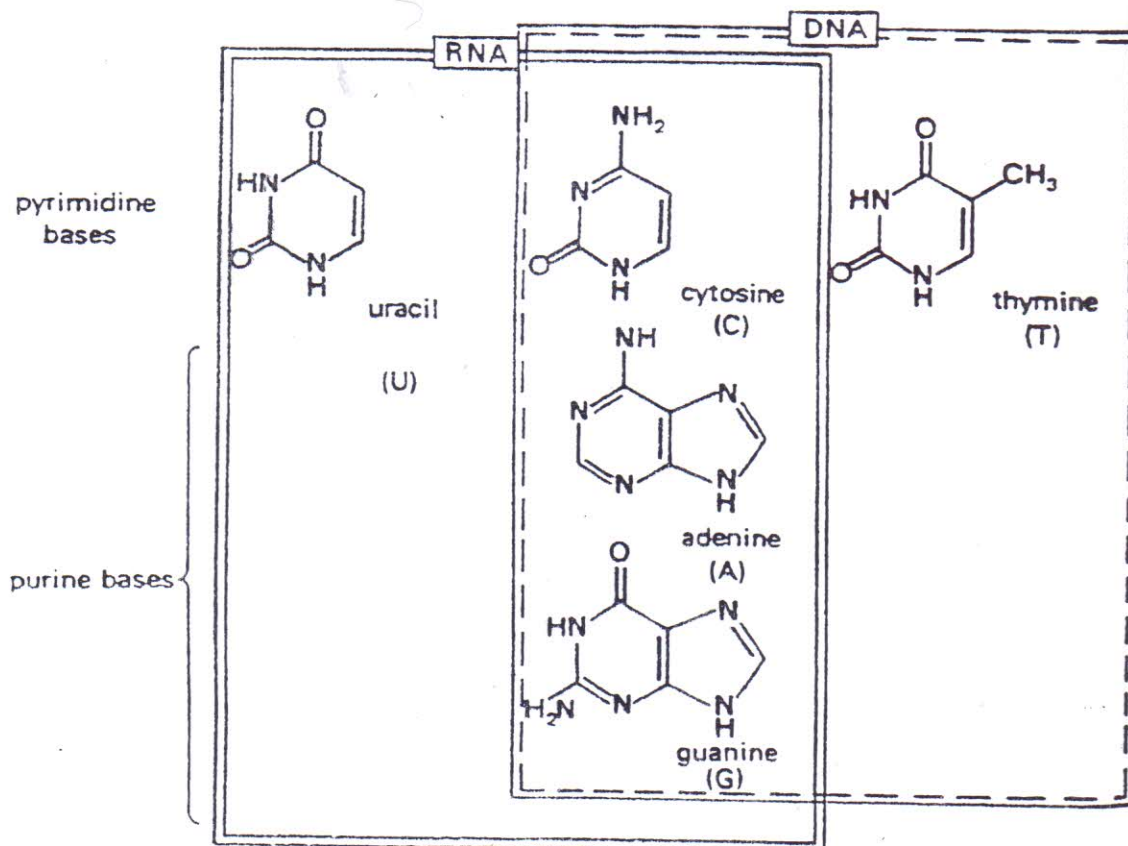
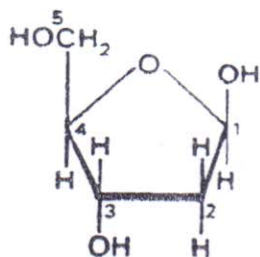



Fig 5.2

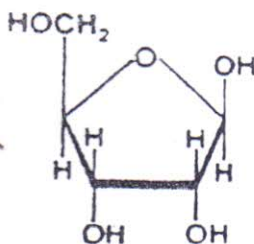
The heterocyclic bases found in RNA are the pyrimidines, uracil and cytosine, and the purines, adenine and guanine; those found in DNA are the pyrimidines, cytosine and thymine, and the purines, adenine and guanine.

The other difference in the chemical composition of the two nucleic acids is that the deoxyribonucleotides contain 2-deoxy-ribose as the pentose sugar,

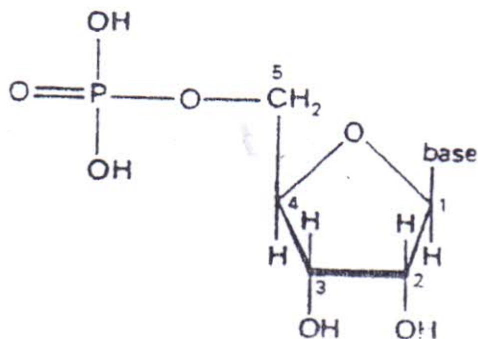



note: oxygen atom missing from the molecule at 2 position

and ribonucleotides contain D-ribose.



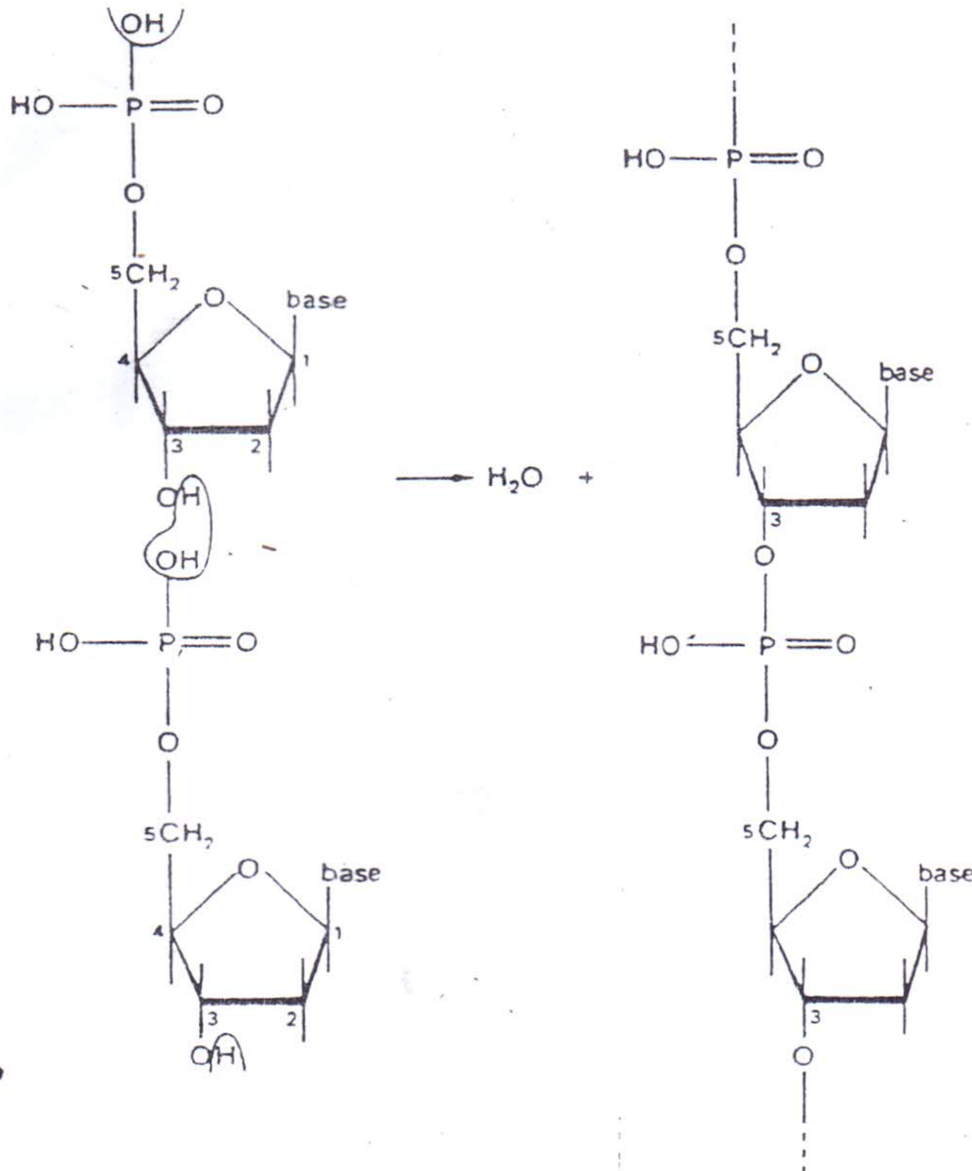

In the nucleotide unit, the base is joined to carbon 1 of the ribose sugar and the phosphate is present as an ester of the hydroxyl group on carbon 5.




a ribonucleotide

## (b) The polymers

The nucleotides polymerise to nucleic acids by using the hydroxyl group on carbon 3 of the ribose or deoxyribose to link to the phosphate group on carbon 5 of the next nucleotide. This forms another phosphate ester linkage.




## (c) DNA

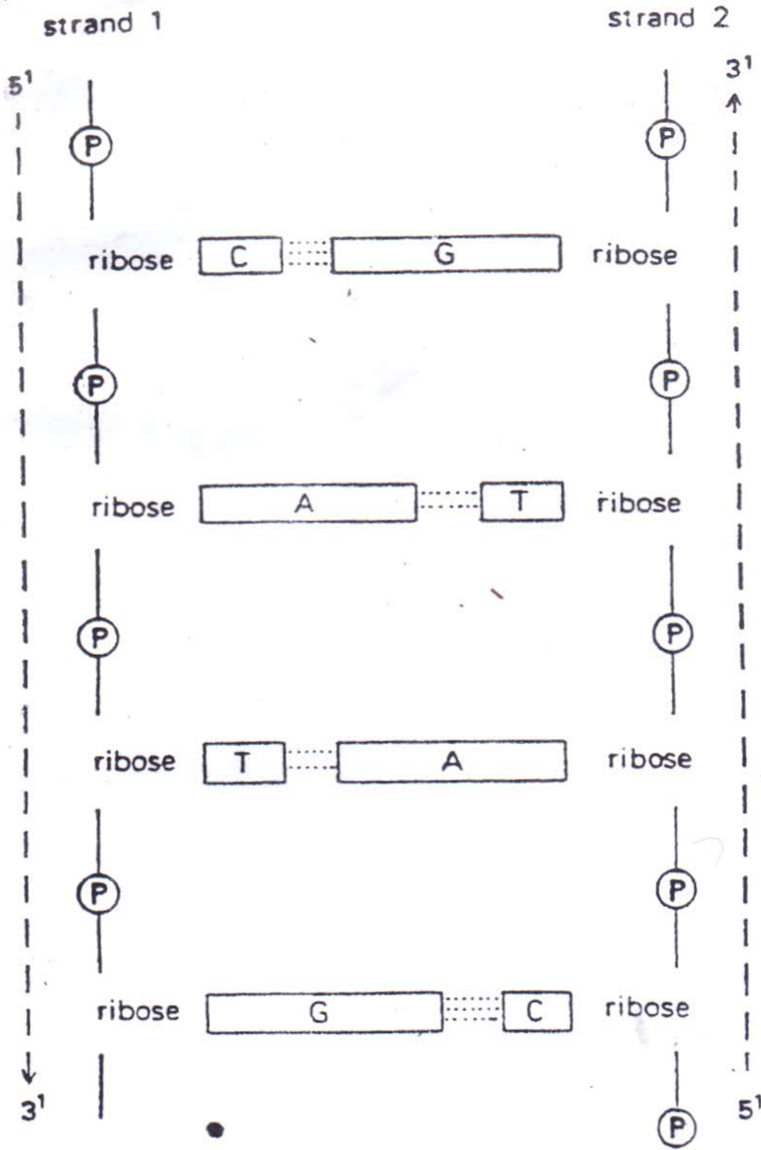
The molecules of DNA from different cells vary in the ratio of each of the four nucleotide monomers and in their relative molecular masses. However, it is found that mole ratios of the bases are always

$$\frac{\text{adenine}}{\text{thymine}} \quad \text{and} \quad \frac{\text{guanine}}{\text{cytosine}} = 1$$

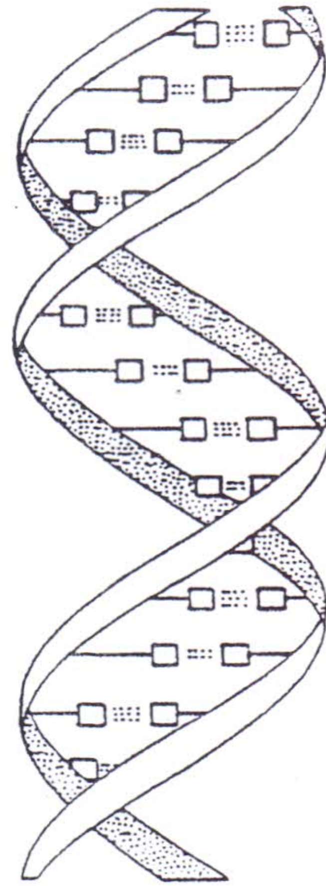
This led to the idea of base-pairing in DNA and thus a double helical structure for the molecule.

DNA molecules isolated from different organisms normally have two strands in a double helix. The two strands are antiparallel, the 3'-5' phosphate linkages running in opposite directions. The bases of each strand are arranged on the inside of the double helix, the bases of one strand being in the same plane as the bases in the other. Only certain base pairs fit inside the helical structure and form hydrogen bonds. The base-pairs which link up in DNA are A-T and G-C, and are referred to as being

complementary; the other possible pairs are either too large to fit into the helix or are too far apart to form stable hydrogen bonds (see Fig 5.3).




(i) Base pairing between the strands



(ii) Double helical structure

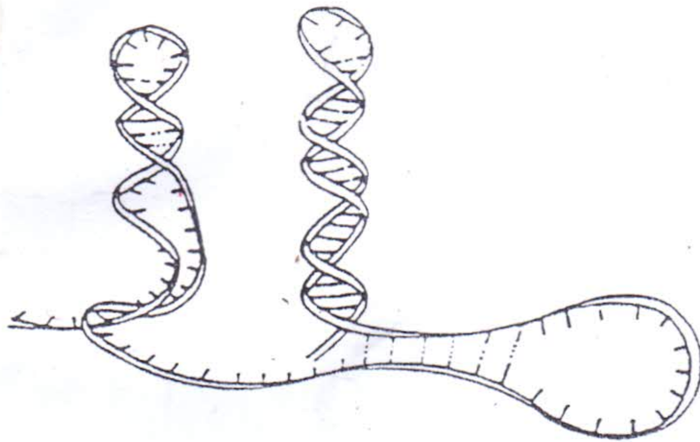
**Fig 5.3.**

In the double helix, the bases are stacked on top of each other like steps in a spiral staircase. Because the structure is only stable when the base pairs are complementary, the base sequence of one strand can determine the base sequence of the second strand. This is essential in the replication of DNA and the transmission of the information stored in the molecule.

By far the largest percentage of DNA present in cells is found in the nucleus, and a small percentage is found in the mitochondria.

### (d) RNA

RNA molecules are generally single stranded, in contrast to DNA, but do have some regions where base-pairing occurs (Fig 5.4)



Base-pairing that could occur by part of a molecule doubling back on itself.

**Fig 5.4**

The irregular, partially helical, structure has intra-strand hydrogen bonding between complementary antiparallel sequences of bases.

The base pairs found in RNA are primarily A-U and G-C.

There are three main types of RNA and they are all involved in protein synthesis:

(i) *ribosomal RNA* accounts for about 80% of the RNA in a cell and is associated with protein to form the structures which are called ribosomes. rRNA have $M_r$ values of up to 1 000 000.

(ii) *transfer RNA* accounts for up to 15% of the total RNA. tRNA are smaller molecules ($M_r$ 23 000 to 30 000) and act as carriers of amino acids during protein synthesis. There are many different tRNA molecules; each one is specific to a single amino acid. Since there is a high proportion of internal base-pairing in tRNA, the molecules show loops and are said to have a characteristic clover-leaf structure.

(iii) *messenger RNA* is complementary to a portion of the DNA in the nucleus. The molecules migrate from the nucleus to the cytoplasm during protein synthesis. mRNA is central to the transfer and assembly of the genetic information and is responsible for the sequence of amino acids (the primary structure) of the different proteins. mRNA molecules vary in size ($M_r$ 25 000 to 1 000 000) and have short half-lives. They break up when the protein strands have been completed.

## Replication of DNA

e double helical structure of DNA as suggested by Watson and Crick gave an insight into (i) the cept that DNA is the bearer of genetic information, (ii) a possible mechanism by which DNA lecules could be duplicated at cell division.

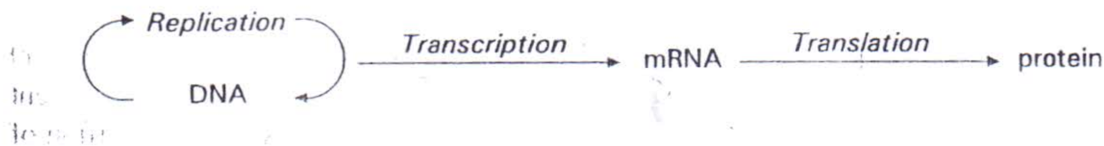
e DNA molecules in the nucleus are organised into separate chromosomes within the nucleus and h chromosome (in higher forms of life) appears to contain a single, long, continuous DNA molecule.

ving four differing types of base arranged in a long chain molecule enables information to be stored coded form, the code lying in the sequence of the bases. The code in DNA carries information cifying the order in which amino acids line up to produce proteins. Because of the manner in which phosphates are linked to the deoxyribose unit, the code can only be read in the correct direction.

e second important feature of the Watson-Crick hypothesis was that, on replication, each double ical DNA results in the formation of two daughter DNA molecules. Each single strand acts as a nplate on which a new strand is synthesised. Hence, each daughter DNA contains one of the strands he original, or parent, DNA. This mode of replication, in which one of the strands is conserved, is l to be semi-conservative. This was confirmed experimentally by Meselson and Stahl using the terial species *E. coli*, which had been grown in a medium containing $^{15}NH_4^+$ but lost density at a ear rate when grown in "normal" $^{14}NH_4^+$.

## Protein Synthesis

e basic ideas of the storage of information and the transmission of this information to control the cell be summarised as:




e two stages involved in producing a protein, using the information stored in the DNA, are transcription, (ii) translation.

*nscription of DNA*

this stage (Fig 5.5), mRNA that is complementary to a section of one strand of the chromosomal IA is made.

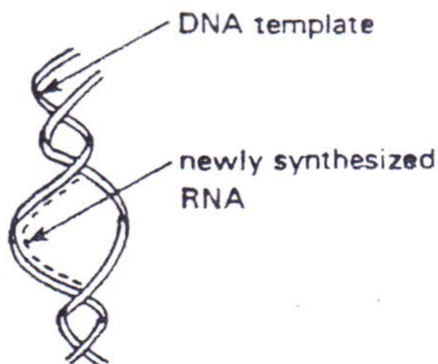



**Fig 5.5**

## (d) RNA

RNA molecules are generally single stranded, in contrast to DNA, but do have some regions where base-pairing occurs (Fig 5.4)

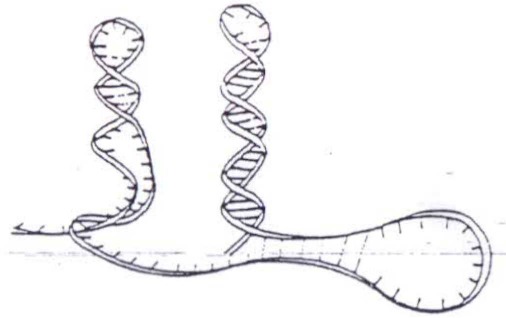



Fig 5.4

The irregular, partially helical, structure has intra-strand hydrogen bonding between complementary antiparallel sequences of bases.

The base pairs found in RNA are primarily A-U and G-C.

There are three main types of RNA and they are all involved in protein synthesis:

(i) *ribosomal RNA* accounts for about 80% of the RNA in a cell and is associated with protein to form the structures which are called ribosomes. rRNA have $M_r$ values of up to 1 000 000.

(ii) *transfer RNA* accounts for up to 15% of the total RNA. tRNA are smaller molecules ($M_r$ 23 000 to 30 000) and act as carriers of amino acids during protein synthesis. There are many different tRNA molecules; each one is specific to a single amino acid. Since there is a high proportion of internal base-pairing in tRNA, the molecules show loops and are said to have a characteristic clover-leaf structure.

(iii) *messenger RNA* is complementary to a portion of the DNA in the nucleus. The molecules migrate from the nucleus to the cytoplasm during protein synthesis. mRNA is central to the transfer and assembly of the genetic information and is responsible for the sequence of amino acids (the primary structure) of the different proteins. mRNA molecules vary in size ($M_r$ 25 000 to 1 000 000) and have short half-lives. They break up when the protein strands have been completed.



## (e) Replication of DNA

The double helical structure of DNA as suggested by Watson and Crick gave an insight into (i) the concept that DNA is the bearer of genetic information, (ii) a possible mechanism by which DNA molecules could be duplicated at cell division.

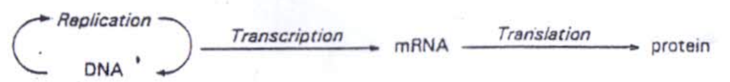
The DNA molecules in the nucleus are organised into separate chromosomes within the nucleus and each chromosome (in higher forms of life) appears to contain a single, long, continuous DNA molecule.

Having four differing types of base arranged in a long chain molecule enables information to be stored in coded form, the code lying in the sequence of the bases. The code in DNA carries information specifying the order in which amino acids line up to produce proteins. Because of the manner in which the phosphates are linked to the deoxyribose unit, the code can only be read in the correct direction.

The second important feature of the Watson-Crick hypothesis was that, on replication, each double helical DNA results in the formation of two daughter DNA molecules. Each single strand acts as a template on which a new strand is synthesised. Hence, each daughter DNA contains one of the strands of the original, or parent, DNA. This mode of replication, in which one of the strands is conserved, is said to be semi-conservative. This was confirmed experimentally by Meselson and Stahl using the bacterial species *E. coli*, which had been grown in a medium containing $^{15}NH_4^+$ but lost density at a linear rate when grown in "normal" $^{14}NH_4^+$.

## (f) Protein Synthesis

The basic ideas of the storage of information and the transmission of this information to control the cell can be summarised as:




The two stages involved in producing a protein, using the information stored in the DNA, are (i) transcription, (ii) translation.

*Transcription of DNA*

In this stage (Fig 5.5), mRNA that is complementary to a section of one strand of the chromosomal DNA is made.

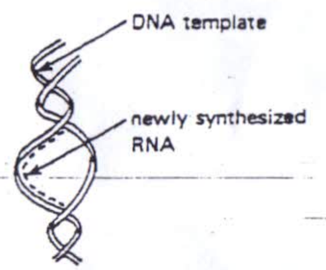



Fig 5.5

An enzyme called RNA polymerase catalyses the assembly of ribonucleotides from ribonucleotide triphosphates with DNA acting as a template. The base pairing is adenine/uracil and cytosine/guanine.

| e.g. | –A–A–C–T–G–G–T–A–C–C–G | *DNA base sequence* |
|---|---|---|
| | –U–U–G–A–C–C–A–U–G–G–C | *mRNA base sequence* |

The section of DNA that is transcribed to give the information needed to make one polypeptide chain is called a gene. One mRNA molecule may correspond to several genes.

Only one strand of the DNA double helix is transcribed to produce a particular mRNA and the mechanism of the selection of which strand is unknown.

After transcription, mRNA passes out of the nucleus into the cytoplasm where it becomes attached to the ribosomes and the next stage in protein synthesis takes place.

*Translation of mRNA*

Having transcribed the stored genetic information of the DNA molecule into an mRNA molecule, the next process is to translate the coded information in the sequence of bases in mRNA to produce the sequence of amino acids in the polypeptide chain. From mathematical considerations, it had appeared for a long time that each amino acid residue is coded by only a small number of consecutive nucleotides in the DNA chain. More than one nucleotide is required as there are only four bases in DNA but twenty amino acids present in proteins. Groups of two code letters in DNA (A, G, C and T) can yield only 16 different pairs ($4^2$), hence the code must contain more than two letters. The four bases taken three at a time give 64, ($4^3$), different triplets, and the triplet code is therefore a possibility.

Biochemical investigations of viruses and genetic studies on bacteriophages supported the theory that the coded information was in the form of a commaless triplet code (codon), i.e. it required no punctuation. It was next necessary to establish the direction of readout. Direct biological evidence was established for the 5'-3' direction of readout (reading from left to right, the residue on the left has a 5' phosphate group and that on the right has a free 3'-hydroxyl group.

*RNA codons and the amino acids for which they code:*

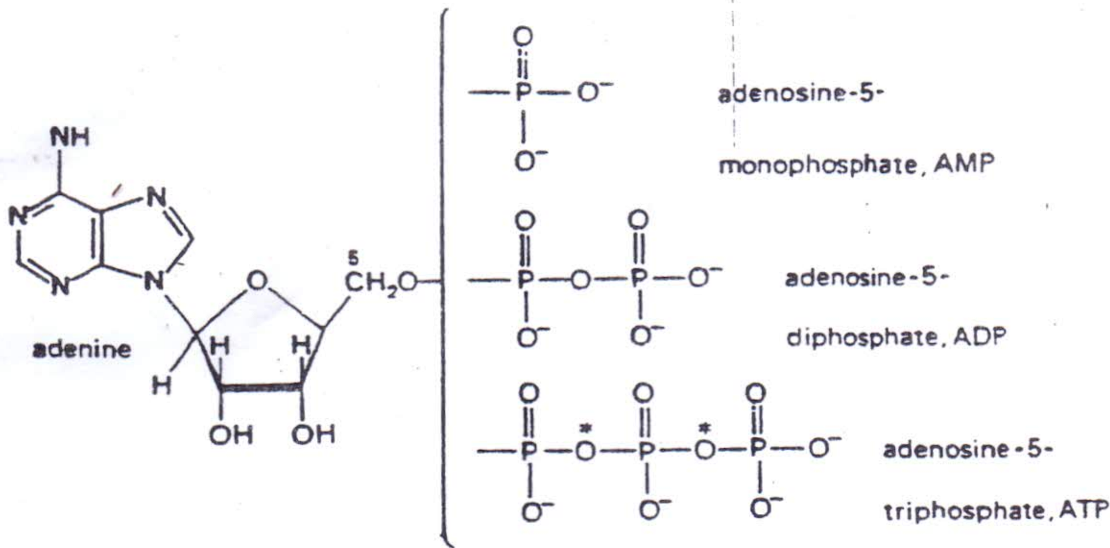
| First base of codon | Second base of codon: U | C | A | G | Third base of codon |
|---|---|---|---|---|---|
| U | UUU, UUC *Phe*<br>UUA, UUG *Leu* | UCU, UCC, UCA, UCG *Ser* | UAU, UAC *Tyr*<br>UAA, UAG *STOP*② | UGU, UGC *Cys*<br>UGA *STOP*②<br>UGG *Trp* | U<br>C<br>A<br>G |
| C | CUU, CUC, CUA, CUG *Leu* | CCU, CCC, CCA, CCG *Pro* | CAU, CAC *His*<br>CAA, CAG *Gln* | CGU, CGC, CGA, CGG *Arg* | U<br>C<br>A<br>G |
| A | AUU, AUC, AUA *Ile*<br>AUG *Met*① | ACU, ACC, ACA, ACG *Thr* | AAU, AAC *Asn*<br>AAA, AAG *Lys* | AGU, AGC *Ser*<br>AGA, AGG *Arg* | U<br>C<br>A<br>G |
| G | GUU, GUC, GUA, GUG *Val* | GCU, GCC, GCA, GCG *Ala* | GAU, GAC *Asp*<br>GAA, GAG *Glu* | GGU, GGC, GGA, GGG *Gly* | U<br>C<br>A<br>G |

① *AUG, as well as coding for Met, is used for initiation.*

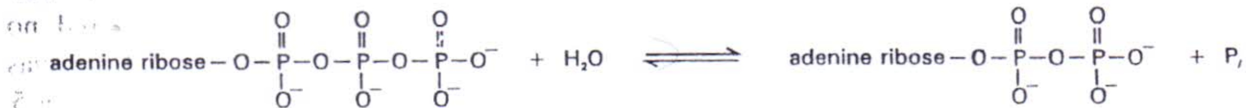
② *UAA, UAG, UGA do not code for any amino acids but are signals for the termination of the polypeptide chain.*

(f) A.T.P.

Adenosine triphosphate (ATP) is a nucleotide. It consists of a phosphate ester of the sugar ribose which is also attached to the base adenine. There are three types of phosphate ester with adenine. According to the number of phosphate residues present, they are monophospho-, diphospho- or triphosphonucleotides. All three types occur universally in cells.




The hydrolysis of the phosphoric anhydride bonds, asterisked in the formula above, results in a large release of energy and the formation of either adenosine diphosphate or monophosphate, together with inorganic phosphate ($P_i$).




There are a number of factors contributing to the products of hydrolysis being at a much lower energy than the reactants. A major one is that the high concentration of negative charge around the triphosphate group of ATP at pH 7.0. The negative charges repel each other strongly and some of this electrical stress is removed when the end phosphate group is released on hydrolysis.

The conversion of ADP to ATP is therefore an energy requiring process and an important one in metabolism.

Metabolic reactions can be divided into two classes; (i) those that tend to occur by themselves and are energy releasing, (ii) those that must be driven by the expenditure of energy.

It is possible to couple the energy yielded by the first class of reactions to drive the energy requiring reactions of the second class by using the interconversion of ATP and ADP as the link (Fig 5.6)

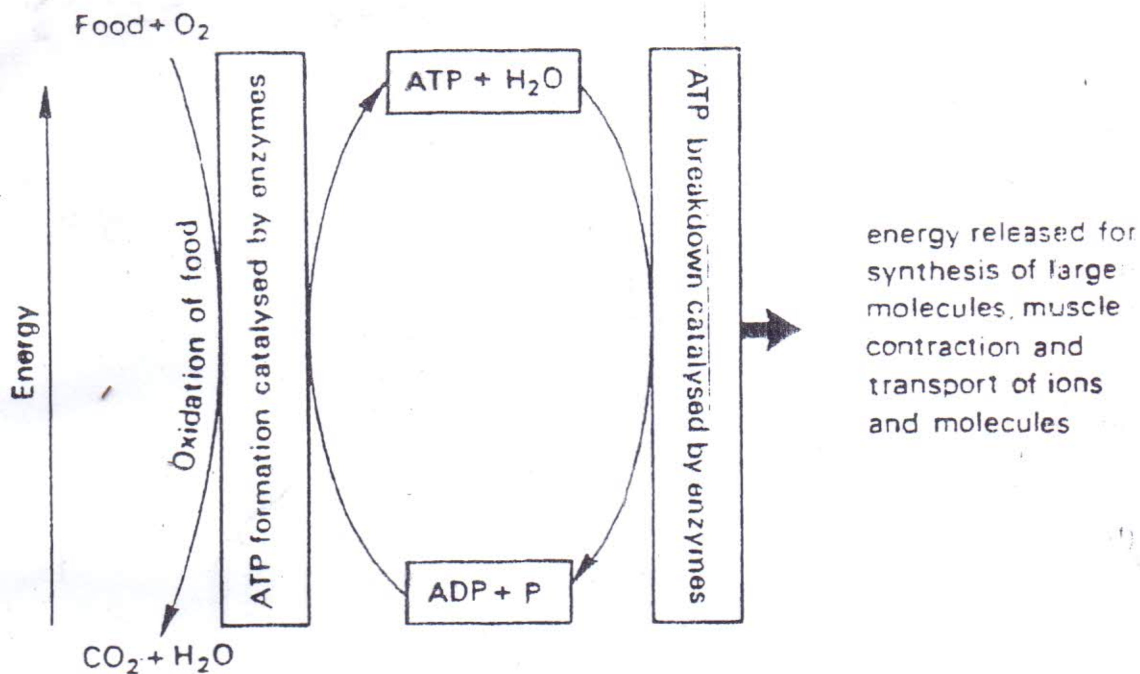


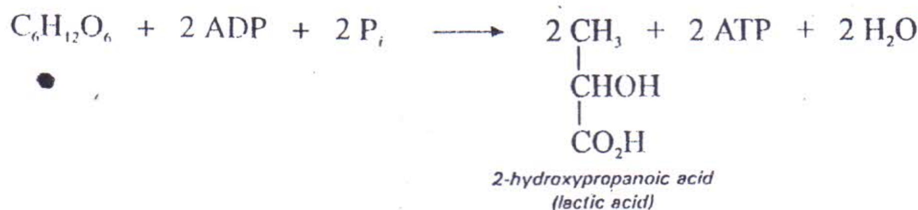

Fig 5.6

The reactions involved in the oxidation of food are energy releasing and drive the process

$$ADP + P_i \longrightarrow ATP$$

The ATP formed can be hydrolysed to provide the energy required for synthesis, muscle contraction and transport of substances across cell membranes. So, ATP is central to the transfer of chemical energy in living cells.

ATP can be generated by the breakdown of glucose and involves no overall oxidation or reduction. The various stages of glycolysis are quite complicated but the sum total of the reaction in, for example, rapidly contracting mammalian muscle is

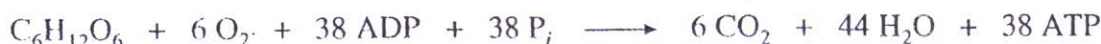


$$C_6H_{12}O_6 + 2\,ADP + 2\,P_i \longrightarrow 2\,CH_3CH(OH)CO_2H + 2\,ATP + 2\,H_2O$$

2-hydroxypropanoic acid (lactic acid)

The energy released by splitting glucose into two lactic acid molecules is enough to generate two ATP molecules. However, glycolysis releases only a small fraction of the chemical energy potentially available in the structure of the glucose molecule.

By contrast, the glucose can, under aerobic conditions, be oxidised completely to carbon dioxide and water, yielding a much larger amount of energy. Through a series of reactions (called the citric acid cycle) that follow on from glycolysis, a total of thirty-eight ATP molecules can be generated per molecule of glucose oxidised.



$$C_6H_{12}O_6 + 6\,O_2 + 38\,ADP + 38\,P_i \longrightarrow 6\,CO_2 + 44\,H_2O + 38\,ATP$$

## Questions on Section 5: Nucleic Acids

1 Give a concise account of each of the following:

(a) the chemical composition of nucleotides; [2]
(b) the structures and functions of nucleic acids; [5]
(c) the triplet code. [3]

2 (a) Describe, with the aid of a simple block diagram, the structure of two 'repeat units' of DNA (deoxyribonucleic acid). [3]

(b) Explain the principal roles of DNA and mRNA, and outline, in simple terms, how these roles are achieved. [7]

3 (a) (i) Draw a simple (block) structure of a section of a strand of RNA, showing three different nucleotides. [2]

(ii) Add to your structure the corresponding strand of double helix where the base-pairings are

adenine : uracil        guanine : cytosine [4]

In (a), use the following abbreviations: P = phosphate; S = ribose (sugar); A = adenine; C = cytosine; G = guanine; U = uracil.

(b) State what is meant by base-pairing and indicate the nature of the chemical bonds involved. [2]

(c) Explain briefly the biochemical role of mRNA molecules. [2]

4 Recent techniques have allowed DNA from various sources to be broken down by enzymes (called restriction endonucleases). These enzymes cut DNA molecules at specific places in the sequence of base pairs.

The various fragments that result from enzyme action are separated by gel electrophoresis carried out in a buffer of pH7 and constant potential difference.

(a) Describe, using a block diagram, the structure of the nucleotides which make up nucleic acids. State briefly how these are arranged in DNA [4]

(b) Describe how the enzyme functions in this example and state what sort of reaction occurs. [2]

(c) Explain, in terms of the structure of the nucleotide, why the DNA fragments move towards the positive electrode in the electrophoresis separation. [2]

(d) What factors determine how far the various DNA fragments move in the electrophoresis separation? [2]