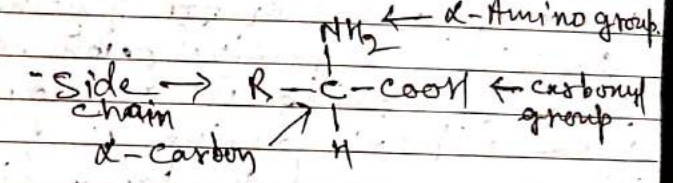
Unit - II : Amino Acids, Peptides and Proteins

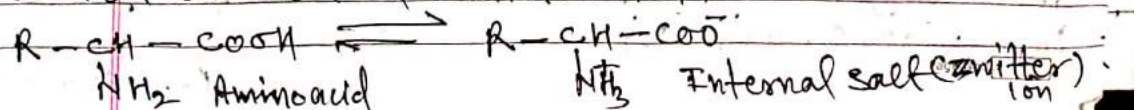
Amino acids :→ These are biologically important organic compounds which contain both amino ($-NH_2$) and carboxylic groups ($-COOH$) in same molecule. Depending upon the position of the amino group i.e. α, β, γ, δ etc. on the carbon chain w.r.t. the carboxyl group, they are classified as α-, β-, γ-, δ- etc. amino acids respectively. out of these α-amino acids (i.e. NH_2 group is linked to the carbon atom next to $-COOH$ group i.e. α-carbon) are the most important because they are the constituents of proteins which are very important biomolecules. The general formula of α-amino acids is



Twenty six different α-amino acids have been isolated by the hydrolysis of various proteins. out of these 20 amino acids occur in almost all proteins while the remaining 6 are found in special tissues. These amino acids differ from one another in the nature of the side chain group 'R'. Therefore properties of side chain groups R determine the properties of amino acids and properties of proteins they constitute.

Zwitter ion structure of α-amino acids :→

In α-amino acids, both acidic ($-COOH$) and basic ($-NH_2$) groups are present in the same molecule. Therefore neutralization reaction takes place between these groups and an internal salt (dipolar) is formed which is called zwitter ion.



Evidences in favour of Dipolar str (Zwitterion): ②

The absence of free amino and carbonyl groups and the existance of dipolar ionic structure is supported by the following facts:

① They are soluble in water but insoluble in common organic solvents like benzene, alcohol etc. It shows ionic nature of amino acids

② They have high dipole moments (1.0 – 1.5 D) which indicates ionic character of amino acids. They also raise the dielectric constant of water when dissolved in it which is a property of polar molecules.

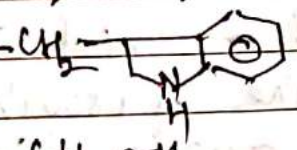
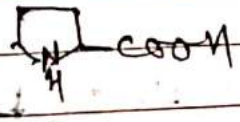
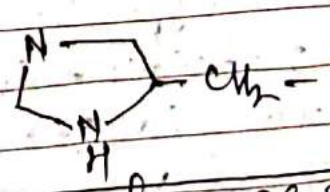
③ The K_a (acidity constant) and K_b (basicity constant) values for amino acids are very low e.g the value of K_a and K_b for glycine are 1.6×10^{-10} and 2.5×10^{-12} respectively. whereas K_a for normal carbonylic acids is of the order of 10^{-5} while K_b for most aliphatic amines is of the order of 10^{-4}. These of K_a and K_b for amino acids are in accordance with the Zwitterion str. in which the acid centre is an ammonium ion ($\overset{+}{N}H_3$) instead of carbonylic gp and the basic centre is the carbonylate ion (COO^-) instead of a free amino group.

④ Infrared and Raman spectra of neutral solutions of α-amino acids indicate the presence of $-\overset{+}{N}H_3$ and $-COO^-$ and not $-NH_2$ and $-COOH$ groups.

Structure, names, abbreviations and one letter code of twenty most commonly occuring amino acids is shown in the table. They differ in the structure of the side chain (R)

α-Aminoacid.

Saatvik

S. No.	Amino acid	Three letter symbol	one letter code	Side chain (R)	Isoelectric point
1.	Glycine	Gly	G	$-H$	6.0
2.	Alanine	Aln	A	$-CH_3$	6.1
3.	Valine*	Val	V	$-CH(CH_3)_2$	6.0
4.	Leucine	Leu	L	$-CH_2CH(CH_3)_2$	6.0
5.	Isoleucine	Ileu	I	$-CH(CH_3)-CH_2-CH_3$	6.0
6.	Phenyl*-alaline	Phe	F	$-CH_2 \cdot C_6H_5$	5.9
7.	Methionine*	Met	M	$-CH_2-CH_2-S-CH_3$	5.7
8	Tryptophan*	Trp	W		5.9
9.	Serine	Ser	S	$-CH_2OH$	5.7
10	cysterine	Cys	C	$-CH_2SH$	5.1
11	Glutamine	Gly	Q	$-CH_2-CH_2CONH_2$	5.7
12.	Threonine*	Thr	T	$-CH(OH)-CH_3$	5.7
13.	Tyrosine*	Tyr	Y	(Complete Str.)	5.6
14.	Proline	Pro	P		6.3
	Acidic Amino acids:-				
15.	Aspartic acid	Asp	D	$-CH_2-COOH$	3.0
16.	Asparagine	Ang	N	$-CH_2-CONH_2$	5.4
17.	Glutamic acid	Gly	E	$-CH_2-CH_2-COOH$	3.2
	Basic amino acids:				
18	Lysine*	Lys	K	$-CH_2(CH_2)_3 \cdot NH_2$	9.7
19.	Arginine	Arg	R	$-CH_2-CH_2-CH_2-NH-C(=NH)-NH_2$	10.8
20.	Histidine	His	H		7.6

* → indicates essential aminoacids.

Scanned with CamScanner

Classification of α-Amino acids :→

α-Amino acids can be classified as under:

(A) Classification based on requirement in the body :→ out of the 20 amino acids required for protein synthesis, human body can synthesize only 10 and the rest 10 are taken in the form of food or diet. Based on this, there are two type of amino acids (1) Essential amino acids (2) Non-essential amino acids

(1) Essential amino acids or Indispensable amino acids:- Those amino acids which cannot be synthesized by the body as known as essential amino acids. These amino acids must be supplied to body in the form of food or diet. These amino acids are required for the growth of the body and their deficiency causes disease such as kwashiorkor.

The essential α-amino acids are: valine, leucine, isoleucine, ~~tryptophan~~ phenylalanine, methionine, tryptophan, threonine, lysine, arginine and histidine. These have been marked by asterisks (*) in the table.

(2) Non-essential amino acids or Dispensable amino acids:- Those amino acids which can be synthesised by the body itself and need not to be supplied in diet are known as non-essential amino acids. These are ten in no. Their names are: Glycine, Alanine, Serine, Cysterine, Glutamine, Tyrosine, Proline, Aspartic acid, Asparagine and Glutamic acid.

(B) Classification according to neutral, acidic and basic character :→ Based upon this criterion, amino acids are of three types :-

(1) Neutral amino acids (2) Acidic amino acids (3) Basic amino acids

(1) ~~Amino~~ Neutral amino acids :→ Those amino acids which contain equal no. of amino ($-NH_2$) and

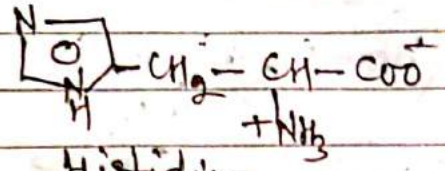
Scanned with CamScanner

Carbonylic (-COOH) groups are neutral amino acids. (5)

eg. $CH_2(+NH_3)-COO^-$ Glycine; $CH_3-CH(+NH_3)-COO^-$ Alanine; $HOCH_2-CH(+NH_3)-COO^-$ Serine

(2) Basic amino acids:→ Those amino acids which contain more no. of $-NH_2$ groups than the $-COOH$ groups are known as basic amino acids eg.

$H_2N-CH_2-CH_2-CH_2-CH(+NH_3)-COO^-$ Lysine


Histidine

(3) Acidic amino acids:→ Those amino acids which contain more no. of $-COOH$ gps than the $-NH_2$ gp. are acidic amino acids. eg.

$HOOC-CH_2-CH(+NH_3)-COO^-$ Aspartic acid; $HOOC-CH_2-CH_2-CH(+NH_3)-COO^-$ Glutamic acid

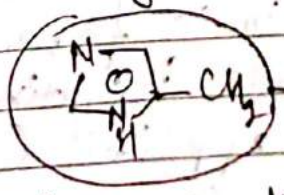
(c) Classification based upon nature of side chain (R):→ According to this classification, amino acids have been divided into four categories:-

(1) Amino acids having acidic side chain:→ Those amino acids which contain more no. of $-COOH$ group than $-NH_2$ groups in the side chain belong to this category. eg.

$HOOC-CH_2-CH(+NH_3)-COO^-$ (acidic) Aspartic acid; $HOOC-CH_2-CH_2-CH(+NH_3)-COO^-$ (acidic) Glutamic acid

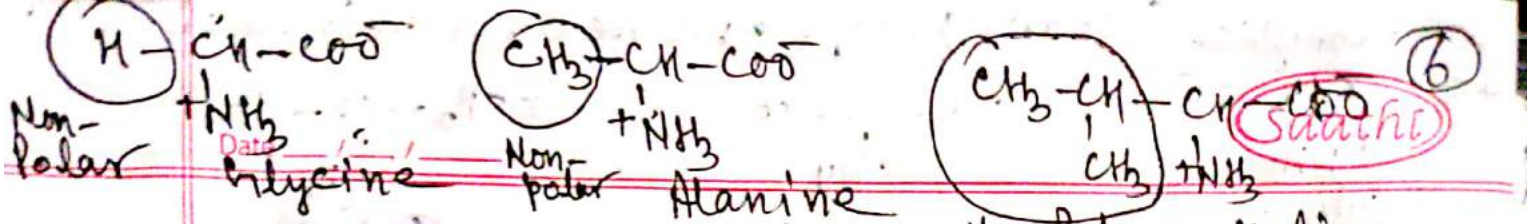
(2) Amino acids having basic side chain:→ Those amino acids having more no. of $-NH_2$ or $=NH$ gps than $-COOH$ gps in the side chain belong to this category

eg. $H_2N-CH_2-CH_2-CH_2-CH_2-CH(+NH_3)-COO^-$ (Basic) Lysine


(Basic) Histidine

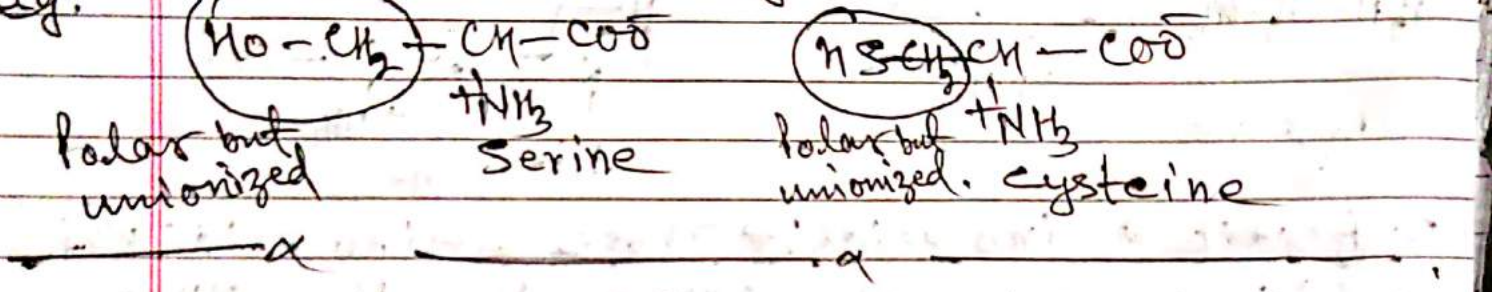
(3) Amino acids having non-polar side chain:→ Such amino acids have H, CH_3, $CH(CH_3)_2$ etc. side chains which are non-polar in nature. eg.

Scanned with CamScanner



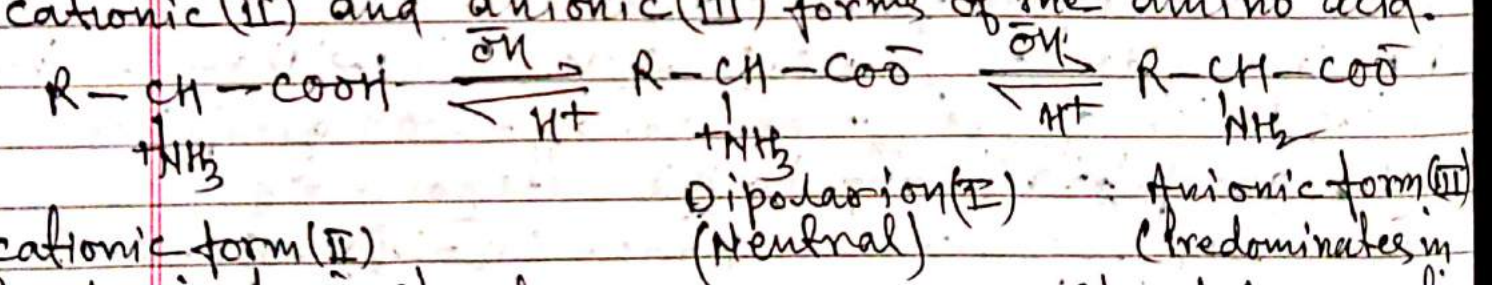
(4) <u>Amino acids having polar and unionized side chains</u>- Amino acids having groups like -OH and -SH on the side chain belong to this category.

eg.



<u>Isoelectric Point (I.P.)</u> :→ Isoelectric point of an amino acid is the pH at which there is no net migration of amino acids under the influence of an applied electric field i.e. electrophoresis.

In an aqueous solution of an amino acid, there exists an equilibrium between the dipolar ion (I) cationic (II) and anionic (III) forms of the amino acid.



The exact position of equilibrium depends upon the pH of solution. If solution is strongly acidic (i.e. low pH), the amino acids exists predominantly in the cationic form (II) and thus migrates towards the cathode (negative electrode) when the electric field is applied (Electrophoresis). On the other hand, in strongly basic solution, it predominantly exists in the anionic form (III) and thus migrates towards anode (positive electrode) when

the electric field is applied (electrophoresis). (7)

At some intermediate value of pH, the concentration of the cationic form (II) and the anionic form (III) are minimum and equal to each other. In other words at this pH, the concentration of the neutral dipolar ionic form (I) is the maximum and hence there will be no net migration of the amino acid in the electric field (i.e. on electrophoresis).

This pH at which there is no net migration of the amino acid under the influence of the applied electric field is called the Isoelectric point (PI)

Thus at a pH lower than PI, there will be a net migration of the amino acid towards cathode because the amino acid exists predominantly in the cationic form (II). On the other hand, at a pH higher than PI, there will be a net migration of the amino acid towards anode because the amino acid mainly exists in the anionic form (III).

The isoelectric point of an amino acid depends upon its structure and hence each amino acid has a characteristic isoelectric point. The isoelectric points of different amino acids are shown in the table.

Proteins :- (Greek, proteois meaning first)

Proteins are extremely complex natural polymers. They are essential for growth and maintenance of life. They are present in almost all living cells and are found in almost every part of every plant and animal. In human beings, they are the main constituents of muscles, skin, hair, tendons, arteries and connective tissues. All proteins contain the same elements i.e. C, H, O, N and S. Some of these may also contain P, I and traces of metals such as Fe, Cu, Zn and Mn.

Scanned with CamScanner

Structure of Proteins :→ Proteins are made up of amino acids. Amino acids are linked to each other by peptide bonds. These polypeptide chains acquire many structural modifications to attain final str. for protein. The structures of proteins are quite complex. Their complete structures are usually discussed at four different levels.

① Primary Structure ② Secondary Structure
③ Tertiary Structure ④ Quaternary structure.

① **Primary Structure** :→ The primary proteins may have one or more polypeptide chains. The primary structure of a protein refers to the covalent structure including disulphide linkages of each polypeptide chain. It simply refers to the sequence in which the various amino acids present in a protein are linked to one another.

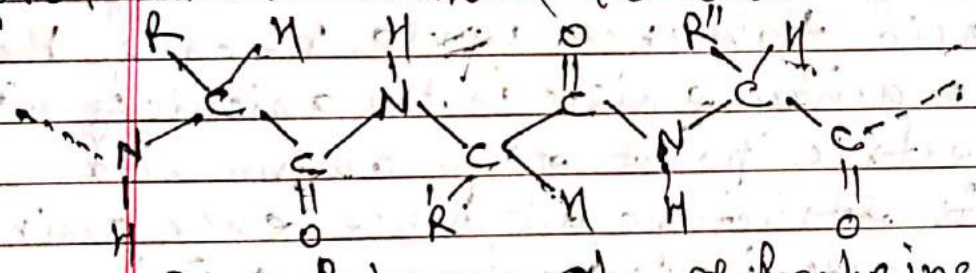


Fig: Primary str. of proteins

The primary str. of proteins depends upon
(i) No. of amino acids present in the protein
(ii) No. of moles of each amino acid present in the protein
(iii) Sequence in which the different amino acids are present in the polypeptide chain.

The first ever primary str. of a protein i.e. insulin was determined by British chemist Frederic Sanger and for this work he was awarded the Nobel Prize in 1958.

② **Secondary Structure** :→ The primary structure gets coiled or folded because of H-bonding and hence assumes a new configuration (Shape). This new configuration which the polypeptide chain assumes as a result of H-bonding is called

Scanned with CamScanner

Secondary structure. Depending upon the size of R-groups, the secondary structures are of two types:- (a) α-Helix Structure (b) β-Pleated Sheet structure.

(a) α-Helix Structure :→ The α-helix str. was postulated by Linus pauling in 1951. If the size of the R groups is quite large, the H-bonds are formed between the C=O of one amino acid residue and the N-H of the fourth amino acid residue in the chain. This causes the polypeptide chain to coil up into a spiral structure called right handed α-helix structure as shown below:-

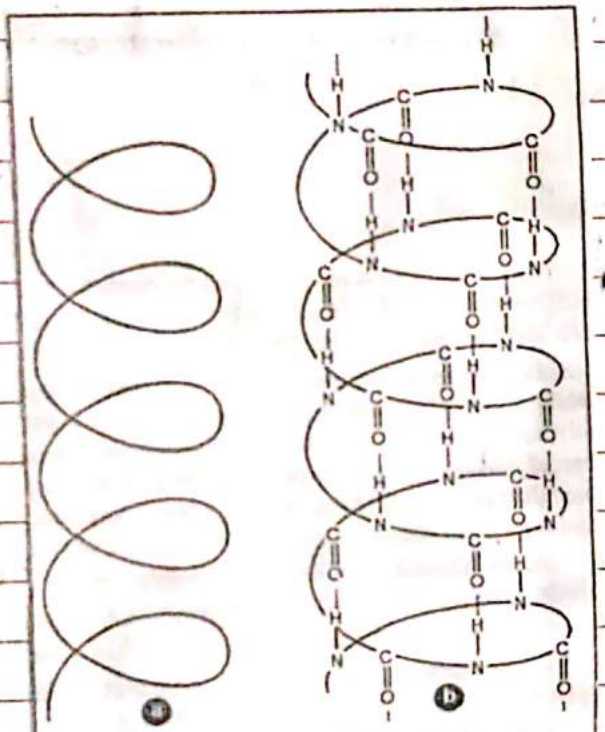


FIGURE (a) A right handed α-Helix
(b) Hydrogen bonding in α-Helix

The α-helix is also known as 3.6_{13} helix since each turn of the helix has approximately 3.6 amino acids and a 13-membered ring is formed by H-bonding. The distance between two successive turns is 5.4 A°.

The α-helix may be right handed or left handed. It has been found that all α-amino acids except glycine are optically active and have L-configuration and for them right handed helix is more stable. Thus the helical str. of proteins is always right handed i.e. the polypeptide chain turns

Scanned with CamScanner

in the clockwise direction along the helical str. just as the threads of a screw or bolt are right handed.

Many fibrous proteins such as α-keratin in hair, nail, wool, skin, beaks, claws and myosin in muscles have α-helix str. though globular proteins also contain segments of α-helix.

(b) <u>β-Pleated sheet Structure</u>:→ The secondary str. of proteins was proposed by Linus Pauling in 1931. The simplest possible arrangement of the peptide chains is the one in which the peptide chains are fully extended to form flat zig-zag structure.

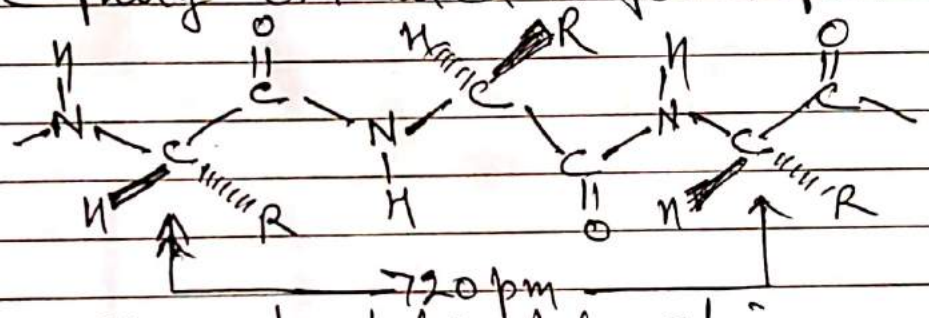


Fig: Extended peptide chain

These chains lie side by side to form flat sheet. Each chain is held by H-bonds to the two neighbouring chain as shown below:-

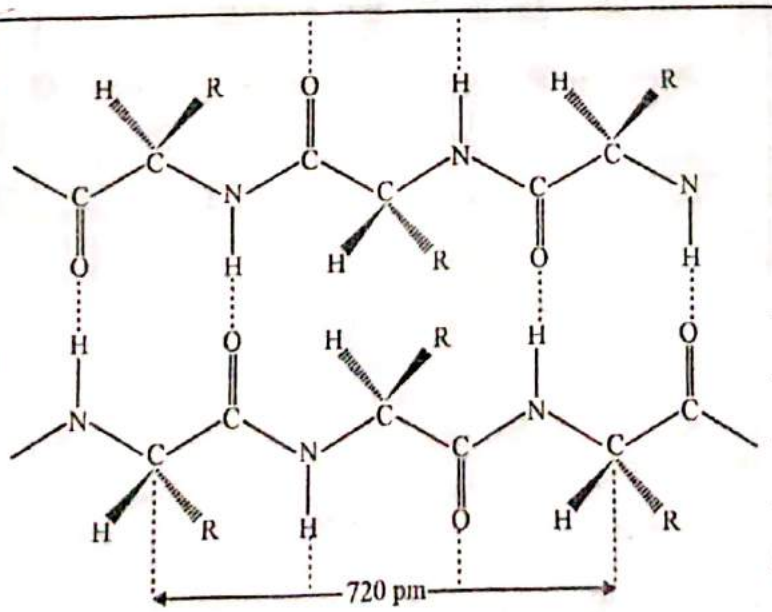


FIGURE . Hypothetical flat sheet structure of a protein. The chains are fully extended and adjacent chains head in opposite directions.

Because of the crowding between the sidechain R groups, no ~~protein~~ natural protein is likely to have this str. Therefore, it is an ideal str. However, if the size of the group R is small to medium, the polypeptide chains contract a little in order to make room for them. This contraction results in pleated sheet in which the polypeptide chains are held together by H-bonds with somewhat shorter distance between alternate amino acid residues. This is called β-pleated sheet structure as shown below:-

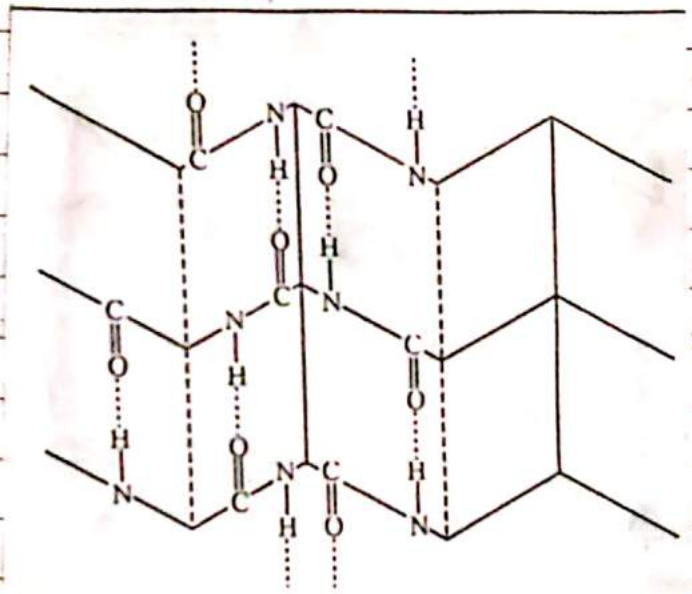


FIGURE 1 β-Pleated sheet structure of proteins.

③ Tertiary Structure :→ Tertiary structure of a protein refers to its complete three dimensional structure. In other words, tertiary structure refers to the manner in which the entire protein molecule folds up in the three dimensional space to produce a specific shape (tertiary str.). At normal pH and temperature, each protein will take up a shape (tertiary str.) that is energetically most stable. This shape is specific to a given amino acid sequence (primary str.) and is called the native shape of the protein.

(a) Tertiary str. of fibrous proteins :→ In these proteins, several α-helices (sec. structure) are coiled about

each other to form molecular ropes or rods. Thus the tertiary structure of fibrous proteins is rope-like or rod like. The str. of collagen triple helix is shown in below:—

(b) Tertiary structure of Globular Proteins:→

FIGURE [redacted]. Tertiary structure of collagen triple helix

unlike fibrous proteins, globular proteins do not have the same secondary structure throughout the length of the molecule. Parts of the molecule may have the α-helical str., while the other parts may have the β-pleated sheet str. and still other parts may be random coils. These different segments of the protein then fold up to the entire molecule a spherical shape. This folding results through the interaction of the various side chain groups of the constituent amino acids. These interactions involve several types of forces such as vander waals interactions, disulphide bridges, ionic or salt bridges and hydrogen bonding as shown below.

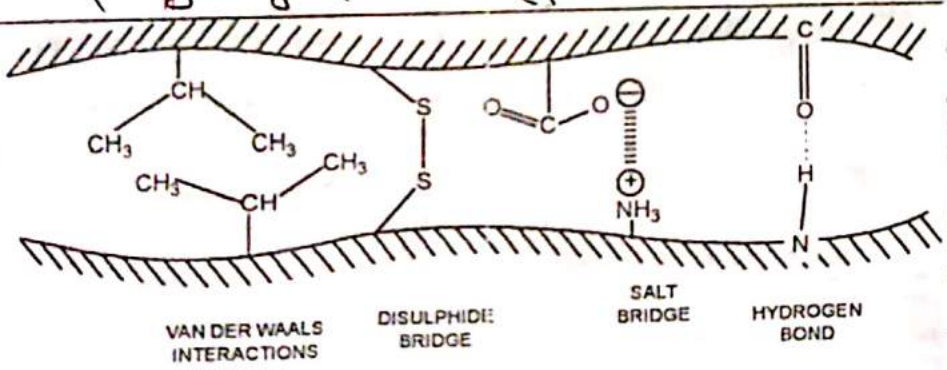


FIGURE [redacted]. Forces which stabilize the tertiary structure.

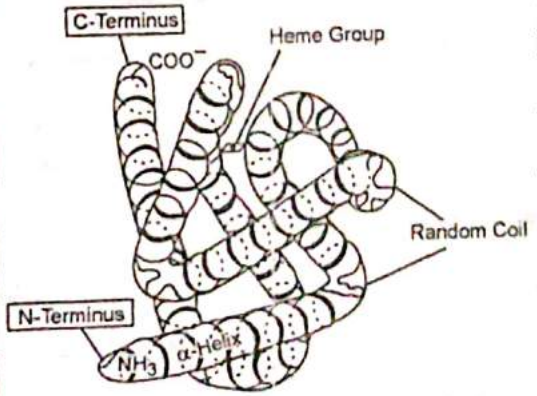


FIGURE [redacted]. Tertiary structure of myoglobin.

The tertiary structures of proteins is determined by their x-ray differaction studies. M.F. Perutz and J.C. Kendrew of Cambridge University, England were the first to determine the str. of myoglobin and haemoglobin and were award Nobel Prize in 1962 in Chemistry. A simplied representation of tert. str. of myoglobin is given above in the fig.

Scanned with CamScanner

(4) **Quaternary Structure of Proteins** :→ There are certain proteins which exist as assemblies of two or more polypeptide chains called subunits or protomers. These subunits may be identical or different and are held together by non-covalent forces such as hydrogen bonds, electrostatic interactions and Vander waals' interactions. The quaternary str. refers to the no. of subunits and their arrangements in an aggregate protein molecule.

eg. Haemoglobin possesses quaternary str. It contains four polypeptide chains or subunits, two identical alpha chains (each containing 141 amino acid residues) and two identical β-chains (each containing 146 amino acid residues). These 4 subunits lie more or less at the vertices of a regular tetrahedron as shown in the fig. Each polypeptide chain carries a heme (iron-protoporphyrin) group at its end.

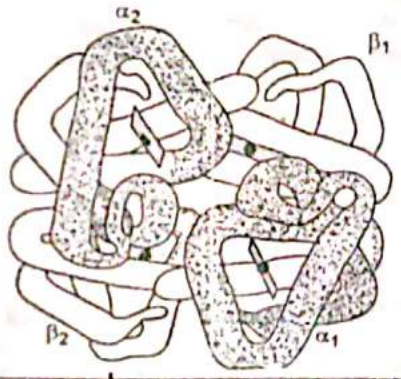


Fig: Quaternary str. of Haemoglobin

Determination of Primary Structure of ~~Protein~~ Peptides:-

The sequence in which amino acids are arranged in a protein chain is known as primary structure. The joining of various amino acids through peptide bonds results in the formation of a peptide chain. The sequence of amino acids in the peptide chain or protein molecule is unique and any change in the sequence will result in denaturation of protein.

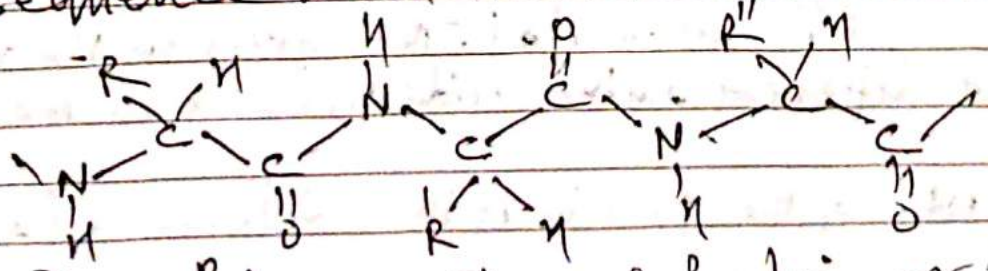


Fig: Primary str. of Protein or Peptide.

Scanned with CamScanner

Following steps are involved in the determination of primary str. of proteins (14)

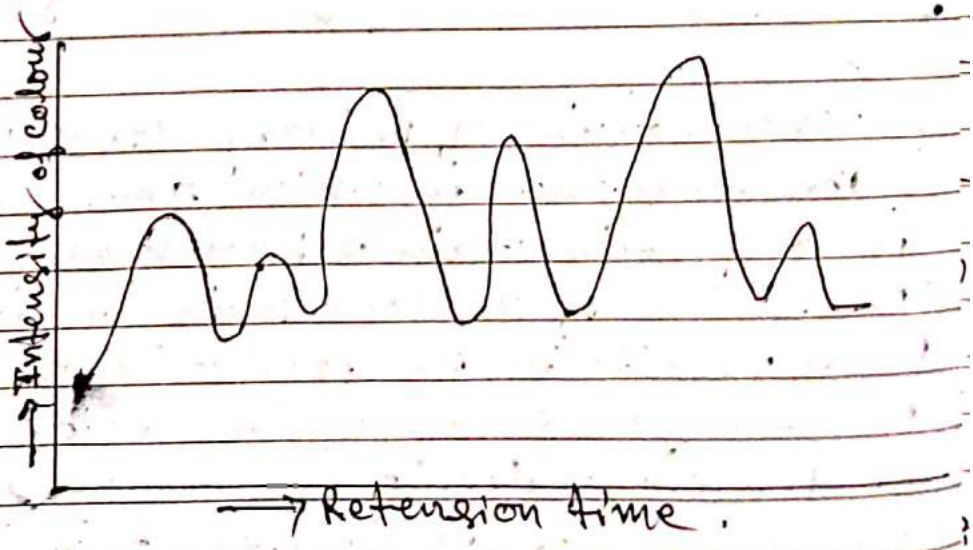
Saathi

Step I : Determination of amino-acid composition :→

The protein is completely hydrolysed to its constituent amino acids either by heating with 6N HCl at 373-393K for 20-24 hours or by enzymes. Alkaline hydrolysis is generally not used since it causes racemerization. The resulting mixture of amino acids is separated either by ion-exchange chromatography or by gas chromatography of their methyl esters.

In ion-exchange chromatography, the mixture of amino acids is dissolved in an aqueous buffer and allowed to pass down the column. As each amino acid leaves the chromatographic column, it is mixed with ninhydrin which produces purple colour with amino acids. The purple colour is monitered by a spectrophotometer. The intensity of the colour is proportional to the amount of the amino acid present. A graph is plotted between the intensity of the colour and the retension time (i.e. the time of elution), when a chromatogram consisting of a series of peaks is obtained.

The positions of peaks are characteristic of individual amino acids and the areas under the peaks correspond to their relative amounts. From the weight of each amino acid, the no. of moles of each amino acid is calculated. This gives us the relative numbers of different amino acids present in the given protein.



The mol. wt. of the protein is then determined either by

Scanned with CamScanner

a chemical method or any one of the physical methods such as osmotic pressure, light scattering method, ultracentrifugation, X-ray diffraction etc. From the mol. wt., the mol. formula and the no. of moles of each amino acid residue actually present in the protein molecule is then calculated. (15)

The whole process is automated ie. carried out automatically by a commercial instrument called the amino acid analyser. This technique is so sensitive that only $10^{-5} - 10^{-7}$ g of the peptide is required to carry out amino acid analysis.

Step-2 :- Sequencing of amino acids :→ The sequence in which amino acids occur along a polypeptide chain is determined by a combination of Terminal Residue Analysis and partial hydrolysis as discussed below :-

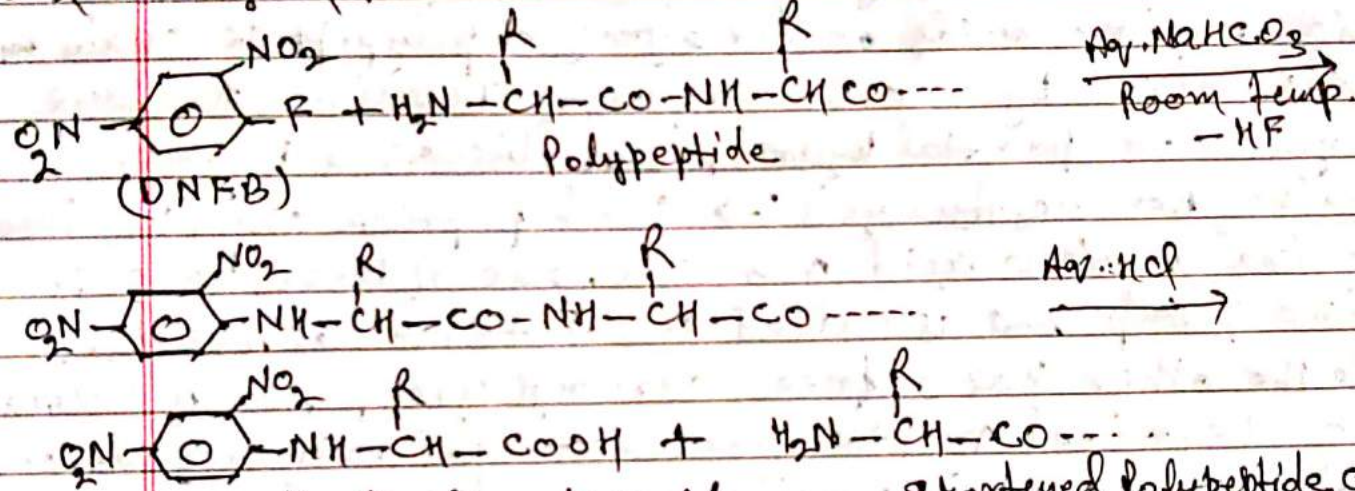
Terminal Residue Analysis :→ Each peptide chain has two terminal amino acid residues. one of these has a free amino group and is called the N-terminal residue while the other has a free carbonyl group and is called the C-terminal residue. These terminal amino acids are identified by certain specific reagents. In this process, the given polypeptide is treated with a suitable reagent which reacts either with the N-terminal or the C-terminal amino acid residue to form the tagged peptide. The tagged peptide is then subjected to partial hydrolysis which releases the tagged amino acid which is then identified. The resulting peptide (shortened by one amino acid) is tagged and subjected to partial hydrolysis again to yield the second tagged amino acid which is again identified. This process is repeated no. of times till the entire peptide chain is degraded and all the amino acids identified. Depending upon the nature of the termini, two types of reagents have been developed and their use has been discussed below :-

Scanned with CamScanner

Determination of N-terminal Amino acid :→

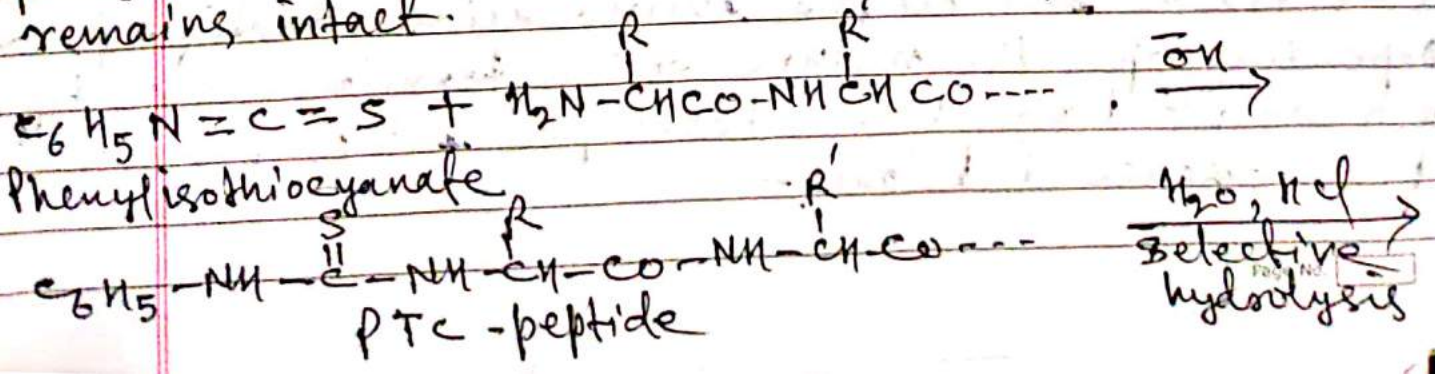
Following methods may be used :-

① **Sanger's method** :- In this method, peptide is treated with 2,4-Dinitrofluorobenzene (DNFB) called Sanger's reagent in the presence of mildly basic solution (aq. $NaHCO_3$) at room temperature. The $-NH_2$ group (N-terminal) of peptide chain reacts with -F atom of Sanger's reagent. The resulting product is hydrolysed with dil. HCl. It results cleavage of bond connecting N-terminal amino acid with rest of the polypeptide chain. The N-terminal amino acid attached with DNFB is separated and analysed.

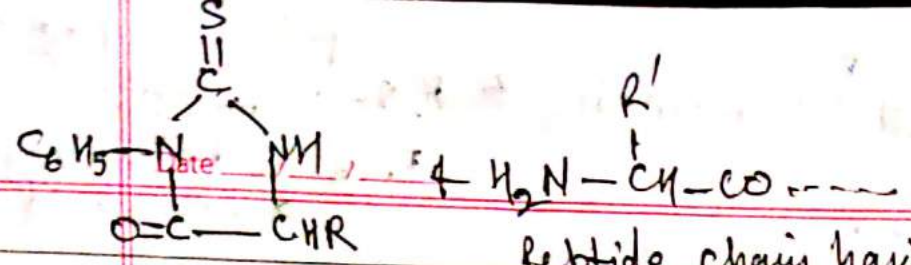


This process is repeated till all amino acids of peptide chain are identified.

② **Edman's Method** :→ In this method, the polypeptide chain is heated with Phenyl isothiocyanate called Edman's reagent in the presence of dil. alkali. The product so obtained is hydrolysed with dil. HCl, which selectively removes the N-terminal amino acid as the phenylthiohydantoins (PTH) but rest of the peptide chain remains intact.



Scanned with CamScanner



Phenylthiohydantoin (PTH)

Peptide chain having one amino acid residue less than the parent peptide.

The phenylthiohydantoin thus produced is separated from the degraded peptide and identified chromatographically by comparing it with phenylthiohydantoins obtained from known amino acids. This leads to identification of N-terminal amino acid.

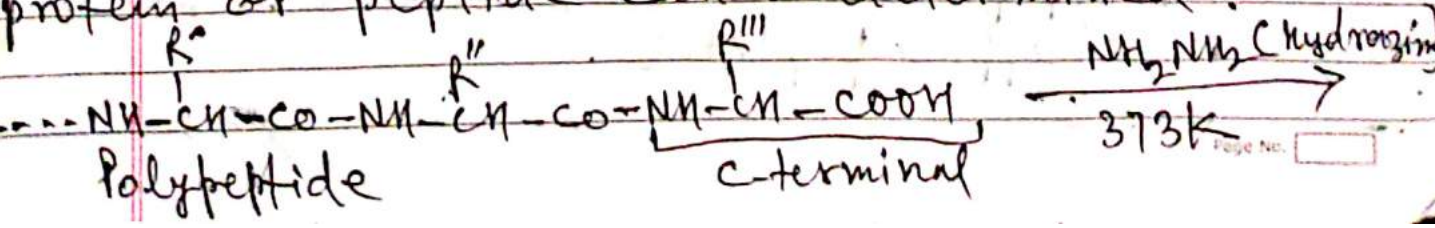
This process is then repeated on the degraded peptide to identify the new N-terminal amino acid residue. Theortically, this process can be repeated over and over again till all the amino acids in the peptide chain are identified but practically this process cannot be extended by 30 residues.

This process has now been automated i.e. is carried out automatically by a machine called protein sequenator.

(B) Determination of C-Terminal amino acids:→

Following methods may be used:

(1) Hydrazinlysis method:- In this method, the polypeptide is heated with anhydrous hydrazine at 373K when all amino acid residues except the C-terminal one is converted into amino acid hydrazides. The mixture of products thus obtained is subjected to chromatography over a column of a strong cation exchange resin. On elution, the strongly basic hydrazides are retained but the free amino acid is eluted. By the identifying the free amino acid, the C-terminal amino acid residue of a protein or peptide can be determined.



$$H_2N-\overset{R'}{CH}-CONH.NH_2 + H_2N-\overset{R''}{CH}-CO-NH.NH_2 + H_2N-\overset{R'''}{CH}-COOH$$

Hydrazides. Free c-terminal aminoacid.

(2) Enzyme Method :→ In this method, the polypeptide with c-terminal residue is subjected to selective hydrolysis with the enzyme carbonypeptidase. The enzyme selectively removes the c-terminal amino acid and the degraded peptide is also produced.

$$\cdots NH-\overset{R'}{CH}-CO-NH\overset{R''}{CH}CO-NH\overset{R'''}{CH}-COOH \xrightarrow[H_2O]{\text{carbonypeptidase}}$$

Polypeptide

$$\cdots NH\overset{R'}{CH}CO-\overset{R''}{CH}COOH + H_2N-\overset{R'''}{CH}-COOH$$

Degraded peptide c-Terminal aminoacid.

The free amino acid is identified and the degraded peptide is again treated with the enzyme to yield a second free amino acid and a further shortened peptide. This process is repeated over and over again till the sequence of all the amino acids residues in the protein is worked out.

This process is a time consuming process as it requires hundred of steps to sequence a polypeptide chain. Therefore in actual practice, a polypeptide chain is subjected to partial hydrolysis with an acid or an enzyme and the smaller peptides so formed i.e. dipeptides, tripeptides, tetrapeptides etc. are identified by the terminal residue analysis. When enough of these smaller peptides have been identified, it is possible to find out the full sequence of amino acid residues in the entire polypeptide chain by matching a no. of such overlapping peptides.

Synthesis of Simple Peptides (upto dipeptides) :→

(A) Classical Method :→

Date __/__/____ Saathi

When two different amino acids are allowed to react in presence of a dehydrating agent, besides the desired reaction between $-NH_2$ gp. of one amino acid and $-COOH$ gp. of the other amino acid, there are equal chances that the reaction may also occur between the $-COOH$ and $-NH_2$ groups of two molecules of the same amino acid, thereby giving a mixture of dipeptides.

eg. When glycine and alanine are heated together in presence of a dehydrating agent, a mixture of four different dipeptides is formed whose separation is difficult.

$$\text{Glycine} + \text{Alanine} \xrightarrow[-H_2O]{} \text{Gly-Gly} + \text{Ala-Ala} + \text{Gly-Ala} + \text{Ala-Gly}$$

To overcome this problem, the $-$amino gp. is first protected before carbonyl gp is activated followed by deactivation. Thus the following steps are involved in peptide synthesis.

(1) Protection of the amino group.
(2) Activation of the carbonyl group of the N-Protected amino acid.
(3) Formation of the peptide bond
(4) Removal of the protecting groups.

(1) Protection of the amino group :→ A protecting group is a group which otherwise makes a reactive group unreactive and thus prevents the molecule from undergoing an unwanted reaction.

The two most widely used groups to protect the amino group are :-

(i) Benzyloxycarbonyl or carbobenzyloxy group or benzyl chloroformate, $C_6H_5CH_2-O-CO-$

(ii) tert-Butoxycarbonyl group, $(CH_3)_3C-O-CO-$

Scanned with CamScanner

Benzyloxycarbonyl group: is denoted by 'Z' according to IUPAC system. It can be introduced by treating the amino acid with benzyloxycarbonyl chloride in the presence of NaOH at 278 K.

$$C_6H_5-CH_2-O-\overset{O}{\overset{\|}{C}}-Cl + H_3\overset{+}{N}-\overset{R}{\overset{|}{C}H}-COO^- \xrightarrow[-HCl]{NaOH,\ 278K}$$

Benzyloxycarbonyl chloride α-Aminoacid

$$\underbrace{C_6H_5-CH_2-O-\overset{O}{\overset{\|}{C}}}_{\text{Protection of } NH_2 \text{ gp (Z)}}-NH-\overset{R}{\overset{|}{C}H}-COOH$$

Benzyloxycarbonylamino acid (or Z-amino acid).

(ii) tert-Butoxycarbonyl gp: It is denoted by Boc. It is introduced by treating the amino acid with di-tert-butyl carbonate in presence of triethylamine. e.g.

$$(CH_3)_3C-O-C(=O)-O-C(=O)-O-C(CH_3)_3 + \overset{+}{N}H_3-\overset{R}{\overset{|}{C}H}-COO^- \xrightarrow{(C_2H_5)_3N}$$

Di-tert-butyldicarbonate α-amino acid

$$\underbrace{(CH_3)_3C-O-\overset{O}{\overset{\|}{C}}}_{\text{Protection of } NH_2 \text{ gp.}}-NH-\overset{R}{\overset{|}{C}H}-COOH + (CH_3)_3C-O-\overset{O}{\overset{\|}{C}}-OH$$

tert-butoxyformic acid

tert-Butoxycarbonyl amino acid or (Boc-amino acid)

(2) Activation of the carboxylic group of N-protected amino acid: → It is done to convert -OH or -COOH gp. into a better leaving group. Two methods which are commonly used are:-

(1) Conversion of the carboxyl group into the acid chloride:- This is done by treating the N-protected amino acid with thionyl chloride or Phosphorus pentachloride.

$$C_6H_5CH_2-O-\overset{O}{\overset{\|}{C}}-NH-\overset{R}{\overset{|}{C}H}-COOH \xrightarrow[-SO_2,\ -HCl]{SOCl_2}$$

$$C_6H_5CH_2-O-\overset{O}{\overset{\|}{C}}-NH-\overset{R}{\overset{|}{C}H}-COCl$$

Benzyloxycarbonylamino acid chloride

Scanned with CamScanner

(ii) Conversion of carbonyl group into p-nitrophenyl (reactive ester) :→ This is done by treating the N-protected amino acid with p-nitrophenol in the presence of a dehydrating agent such as N,N'-dicyclohexylcarbodiimide (DCC) under mild conditions.

$$C_6H_5CH_2O-CO-NH-\underset{R}{CH}-COOH + HO-C_6H_4-NO_2 \xrightarrow[-H_2O]{DCC}$$

Benzyloxycarbonyl amino acid p-Nitrophenol

$$C_6H_5-CH_2O-CO-NH-\underset{R}{CH}-CO-O-C_6H_4-NO_2$$

(3) Formation of peptide bond :→ After the protection of the amino group and activation of the carbonyl grp. of the N-protected amino acid, it is allowed to react with same or different free amino acid when a peptide is formed. e.g.

$$C_6H_5CH_2O-CO-NH-\underset{R}{CH}-COCl + \overset{+}{H_3N}-\underset{R'}{CH}-COO^- \xrightarrow[-HCl]{263K,\ THF}$$

Benzyloxycarbonylamino acid α-Amino acid

$$C_6H_5CH_2O-CO-NH-\underset{R}{CH}-CO-NH-\underset{R'}{CH}-COOH$$

A Z-protected dipeptide.

(4) Removal of the protecting group :→ The final step in the synthesis of peptides is the removal of the protecting groups. Depending upon the nature of the protecting groups used, the following methods are commonly used :—

(i) Removal of Benzyloxycarbonyl group (Z) : This can be easily removed by catalytic hydrogenation (H_2/Pd) or by hydrolysis with HBr in cold CH_3COOH.

$$C_6H_5CH_2OCO-NH\underset{R}{CH}CO-NH-\underset{R'}{CH}-COOH \xrightarrow[\text{or } HBr/CH_3COOH]{H_2/Pd\ (-C_6H_5CH_3)}$$

$$[HOOC-NH-\underset{R}{CH}CO-NH-\underset{R'}{CH}-COOH] \xrightarrow{-CO_2} \overset{+}{H_3N}-\underset{R}{CH}CO-NH-\underset{R'}{CH}-COO^- \text{ (Dipeptide)}$$

($-C_6H_5CH_2Br$)

Scanned with CamScanner

(iv) Removal of t-butoxycarbonyl group :→ This group can be removed by treating the protected dipeptide with HBr|CH_3COOH or by brief treatment with a strong acid such as trifluoroacetic acid (CF_3COOH).

$$(CH_3)_3C-O-CO-NH-\overset{R}{CH}-CO-NH-\overset{R'}{CH}-COOH \xrightarrow[\text{or } CF_3COOH]{HBr|CH_3COOH}$$

Boc protected dipeptide

$$H_2N-\overset{R}{CH}-CO-NH-\overset{R'}{CH}-COOH + (CH_3)_3C-OH + CO_2$$

A-dipeptide.

Example :→ Let us consider the synthesis of a simple dipeptide i.e. glycylalanine.
Following steps are involved:

(i) Protection of the amino group of glycine :→

$$C_6H_5CH_2OCOCl + {}^{+}H_3N-CH_2-COO^{-} \xrightarrow[-HCl]{} C_6H_5CH_2OCO-NHCH_2-COOH$$

Benzyloxycarbonyl chloride, Glycine → N-Benzyloxycarbonyl glycine.

(ii) Activation of the N-protected amino acid :→

$$C_6H_5CH_2OCO-NHCH_2-COOH \xrightarrow[-HCl,\ -SO_2]{SOCl_2} C_6H_5CH_2OCO-NHCH_2-COCl$$

N-protected and C-activated glycine

(iii) Formation of the Peptide bond :→

$$C_6H_5CH_2OCO-NHCH_2COCl + {}^{+}H_3N-\overset{CH_3}{CH}-COO^{-} \xrightarrow[-HCl]{}$$

Alanine

$$C_6H_5CH_2OCO-NHCH_2CO-NH-\overset{CH_3}{CH}-COOH$$

N-Benzyloxycarbonylglycylalanine

(iv) Removal of the protecting group :→

$$C_6H_5CH_2OCO-NHCH_2CO-NH\overset{CH_3}{CH}-COOH \xrightarrow{H_2|Pd}$$

$${}^{+}H_3N-CH_2CO-NH-\overset{CH_3}{CH}-COO^{-} + C_6H_5CH_3 + CO_2$$

Glycylalanine (Gly-ala).

Scanned with CamScanner

Drawbacks of classical method of Synthesis of Peptides

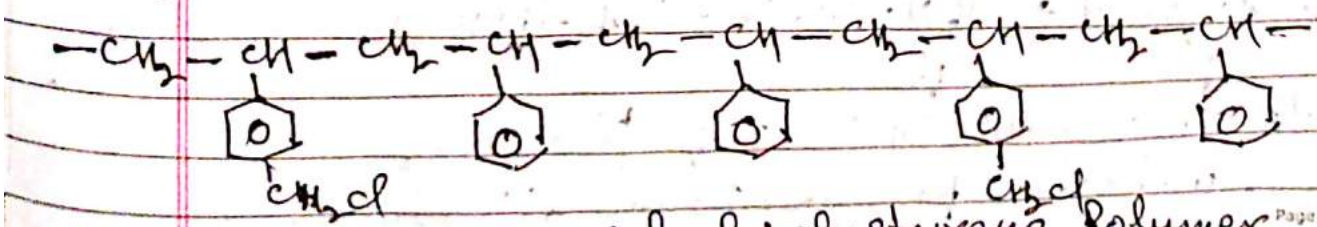
The classical peptide synthesis is suitable for the synthesis of oligopeptides containing 2-10 amino acids. But the synthesis of polypeptides containing large no. of amino acid residues is very difficult by this method due to the following reasons.

(i) Time consuming :→ This method requires several steps involving protection, activation, condensation and deprotection for each new peptide bond to be formed. Further the new peptide formed in each cycle must be isolated and purified. Therefore this method is very tedious and time consuming.

(ii) Low yield :- Since the process involves a no. of steps, the overall yield goes on decreasing after each extension. eg. if the yield of a dipeptide is 80% then the yield of the tripeptide will be 80% of 80 = 64% and that of tetrapeptide would be 80% of 64 = 51.2%. If a decapeptide is to be synthesised, the overall yield would be 13.4% and yield of still higher peptides would be still lower.

Merrifield Solid-Phase Peptide Synthesis :→ To overcome the difficulties encountered in the classical peptide synthesis R.B. Merrifield (Rockefeller University, USA), in 1964 devised a brilliant method for peptide synthesis called the Solid-Phase Peptide Synthesis.

In this method, the peptide synthesis is carried out on the surface of an insoluble solid support. The solid support consists of a polystyrene polymer cross-linked with about 2% p-divinylbenzene. Further about 5% of the benzene rings of this polymer carry chloromethyl (-CH_2Cl) substituents).



chloromethylated polystyrene Polymer

This chloromethylated polystyrene polymer is porous in nature and has gel like structure. It is used in form of small beads through which the various reagents and solvent used can easily percolate. In chemical reactions, it is represented as $\boxed{\text{Polymer}}-CH_2-Cl$

The following steps are involved in this method:-

Step I :→ The desired amino acid is first of all protected at its N-terminal by Boc (tert-butoxycarbonyl chloride) or Z (Benzyloxycarbonyl (gp)) group. N-protected amino acid is then reacted with solid polymer i.e. $\boxed{\text{Polymer}}-CH_2-Cl$ in the presence of triethylamine $(C_2H_5)_3N$ as shown below:-

$$Boc-NH-\underset{R}{CH}-COOH + Cl-CH_2-\boxed{\text{Polymer}} \xrightarrow[-HCl]{(C_2H_5)_3N}$$

$$Boc-NH-\underset{R}{CH}-\underset{\text{ester linkage}}{\overset{O}{\overset{\|}{C}}-O}-CH_2-\boxed{\text{Polymer}}$$

After the ester linkage has been formed, the excess reagents are removed by washing with suitable solvents.

Step II :→ The Boc protecting group is then removed by treatment with trifluoroacetic acid (CF_3COOH). The salt of the polymer bound amino acid thus formed is converted into the free acid by treatment with a solution of excess of triethylamine and is purified by washing.

$$Boc-NH\underset{R}{CH}CO-OCH_2-\boxed{\text{Polymer}} \xrightarrow[\text{(ii)}(C_2H_5)_3N]{\text{(i) }CF_3COOH} H_2N\cdot\underset{R}{CH}CO-OCH_2-\boxed{\text{Polymer}}$$

Polymer bound amino acid.

Step III :→

A second Boc-protected amino acid is then coupled with the above polymer bound amino acid by means of dicyclohexylcarbodimide (DCC). The excess reagents and by-products are washed away.

$$Boc-NH-\underset{R'}{CH}\cdot COOH + H_2N-\underset{R}{CH}\cdot CO-OCH_2-\boxed{\text{Polymer}} \xrightarrow[-H_2O]{DCC}$$

$$Boc-NH-\underset{R'}{CH}CO-NH-\underset{R}{CH}CO-OCH_2-\boxed{\text{Polymer}}$$

Boc-protected polymer bound dipeptide.

Scanned with CamScanner

Step-IV The Boc-protecting group is removed as in step III.

$$Boc-NH-\underset{R'}{CH}CO-NH-\underset{R}{CH}CO-OCH_2-\boxed{Polymer} \xrightarrow{CF_3COOH}$$

$$H_2N-\underset{R'}{CH}.CO-NH\underset{R}{CH}CO-O-CH_2-\boxed{Polymer}$$

Polymer bound dipeptide

Step III and IV are repeated to add another amino acid. In this way 100 of amino acids can be linked to one another in a specific order to synthesize the desired polymer bound peptide. The last acid to be attached is the N-terminal amino acid.

Step V : At the end of the synthesis, the polypeptide is removed from the polymer by treatment with Anhyd. HF.

$$H_2N.\underset{R''}{CH}CO-[NH\underset{R'}{CH}CO]_n-NH\underset{R}{CH}CO-OCH_2-\boxed{Polymer} \xrightarrow{H^+}$$

$$H_2N-\underset{R''}{CH}CO-[NH\underset{R'}{CH}CO]_n-NH\underset{R}{CH}-COOH + HO-CH_2-\boxed{Polymer}$$

Polypeptide

This solid phase technique has now been automated i.e. computer controlled peptide synthesising machines are now available in which coupling and removal of protecting groups ~~are now available~~ with different amino acids is automatically carried out as many times as desired.

Advantages of Solid Phase Synthesis :→

(i) This technique requires less time as compared to classical method of Protein synthesis. eg. Synthesis of enzyme (bovine pancreatic ribonuclease) containing 124 amino acids units involving 369 chemical reactions and 11931 steps were synthesised in 6 weeks only and an overall yield of 17% was obtained.

(ii) yields are extremely high as compared to that of classical method.

(iii) Purification of each step of coupling is not necessary since insoluble polymer bound amino acid is thoroughly washed with suitable solvents to remove excess of reagent.

Scanned with CamScanner