

# I Amino acids and proteins.

## I.1 Introduction

$\alpha$ -Keratin is a structural material - The essential constituent of hair. Insulin is a hormone which organizes the metabolism of glucose in our body. Invertase is an enzyme which catalyses the hydrolysis of sucrose into glucose and fructose. All the three molecules have one thing in common amongst themselves. All of them are poly peptides or proteins. Proteins are an extremely important class of biomolecules.

The analyses of a vast number of proteins from all conceivable sources have shown that all proteins ~~and~~ or poly peptides are composed of a combination of ~~a~~ very simple building blocks bearing the general formula  $\text{H}_2\text{N}-\text{CH}_2-\text{CO}_2\text{H}$ .

There are  $\alpha$ -amino acids consisting of a  $\text{CO}_2\text{H}$  group and a  $\text{NH}_2$  group attached to the same ' $\alpha$ ' carbon atom which also contains a side chain R. All amino acids except glycine where  $R=H$ , are optically active since the ' $\alpha$ ' carbon is an asymmetric centre.

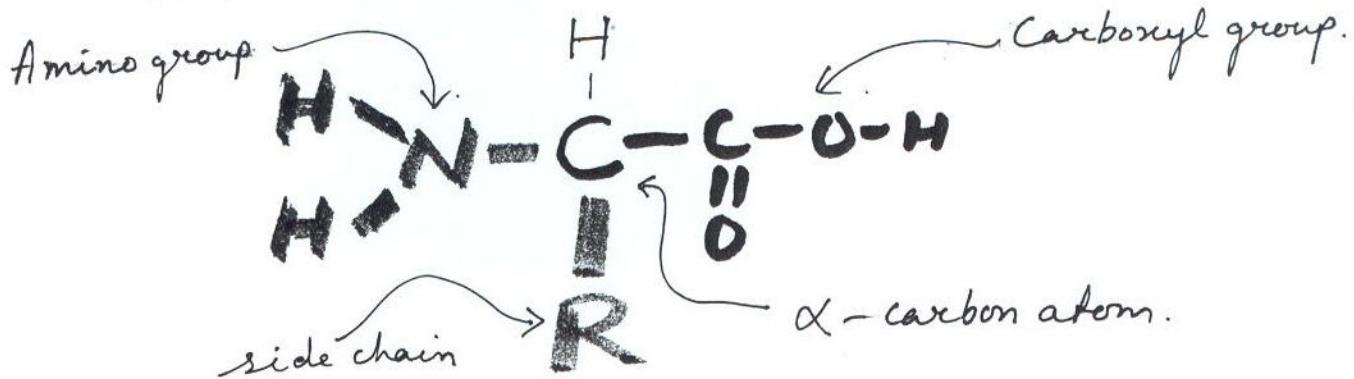
There are 20 different ~~are~~ commonly occurring amino acids ~~of which~~ of which 10 are called essential amino acids since they cannot be synthesized in our body and have to be obtained by breaking down (by hydrolysis) proteins in our diet. The other 10 amino acids may be synthesized.

within our body if the essential amino acids are present in proper abundance. A balanced diet therefore must consist of a proper dose of proteins. ~~not all~~ Milk proteins contain all the essential amino acids in the desired proportion and thus it constitutes ~~add~~ a balanced diet for infants. However, proteins from other sources do not have a proper distribution of amino acids. For example, proteins in corn, rice and wheat are rich in methionine but poor in lysine, tryptophan and threonine. Bean and pulses have proteins which are deficient in methionine but rich in the other amino acids like lysine etc. This explains why <sup>our</sup> staple diet mostly consists of rice or "roti" (wheat or corn bread) and dal (pulses).

Apart from the 20 common amino acids, there are 6 more variants which are ~~used in~~ found in special tissues.

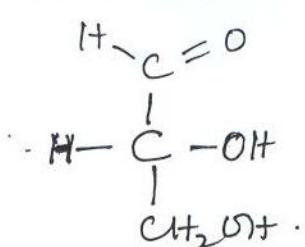
~~This~~ These amino acids differ only with respect to the side chain 'R' and the properties of amino acid side chains determines the properties of the proteins they constitute.

### I.2 General features of amino acids:

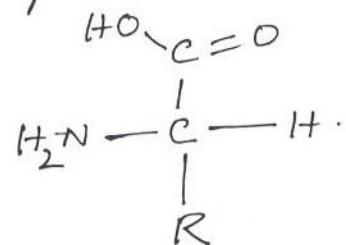


(a) The asymmetric  $\alpha$ -carbon atom and the CORN rule.

except for glycine where  $R = H$  all amino acids are optically active since the  $\alpha$ -carbon atom is asymmetric. Of the two possible enantiomers of any amino acid only one is found in ~~biologys~~ nature. Just as for carbohydrates only the D-series configuration occurs in nature, for amino acids only the L-series configuration is observed (see Fischer projections below).



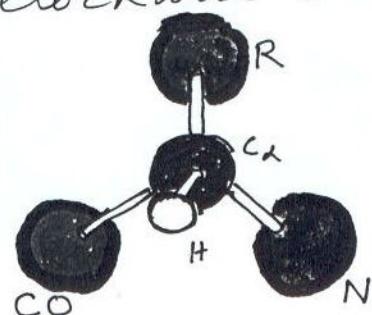
D-Glyceraldehyde.



L- $\alpha$ -Amino acid.

A useful method for identifying the biologically active L-configuration of amino acids is by the application of the ~~CORN~~ CORN rule. If we look along the  $H-C_\alpha$  bond of an L-amino acid keeping the  $\alpha$ -carbon away from us, we will find that the  $\text{CO}_2\text{H}$ ,  $\text{R}$  and  $\text{NH}_2$  (CORN) groups form a clockwise orientation.

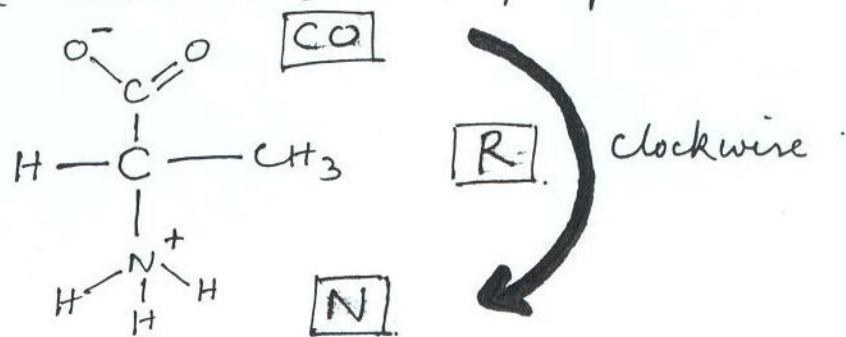
The CORN  
Rule.



(b) Amino acids exist as zwitterions.

The presence of a basic  $\text{NH}_2$  and an acidic  $\text{CO}_2\text{H}$  group on the same carbon atom confers some unusual properties on the amino acids. For example their melting points are fairly high — around  $300^\circ\text{C}$ , though e.g. Their ~~most~~ esters have melting points of around  $100^\circ\text{C}$ . Further, even though amino acids are organic compounds, they are far more soluble in water and other polar solvents than in non polar solvents.

All these ~~for~~ observations can be explained in terms of an ionic salt like structure of amino acids ~~is~~ wherein the basic  $\text{NH}_2$  group is protonated by the nearby  $\text{CO}_2\text{H}$  group. In the solid state, therefore, amino acids have an  $\text{NH}_3^+$  group and a  $\text{CO}_2^-$  group. Molecules with charged groups of opposite polarity are called zwitterions or dipolar ions. They may also be referred to as inner salts. The figure below, gives the zwitterionic structure of ~~the~~ the L- $\alpha$ -amino acid alanine (Fischer projection).



I.3

**(a) Acid-base properties of amino acids in aqueous solutions.**

(a) The  $pK$  values of groups.

All amino acids are soluble both in acids as well as ~~base~~ in bases. Whether the solvent is acidic, basic or plain neutral water, it should be remembered that amino acids in aqueous solution always contain charged groups. Let us examine this in greater detail on the basis of acid base equilibria.

Any ~~of the~~ amino acid is associated with at least two  $pK_a$  values — one corresponding to the  $\text{CO}_2\text{H}$  group and the other corresponding to the  $\text{NH}_2$  group. ~~For~~ For the  $\text{NH}_2$  group, actually, the  $pK_a$  value corresponds to the  $K_a$  of the ~~the~~ conjugate acid  $\text{NH}_3^+$ . These are usually referred to as  $pK_1$  and  $pK_2$  respectively. In addition, the side chain may also contain acidic groups such as  $-\text{CO}_2\text{H}$  ~~as~~ and  $-\text{C}_6\text{H}_4\text{OH}$  or basic groups such as ~~as~~  $-\text{NH}_2$ . The  $K_a$  for acidic or basic side chain group is given as  $pK_R$ . The  $pK_1$ ,  $pK_2$  and  $pK_R$  (where pertinent) of the 20 common amino acids are listed in table I.1.

~~[The three letter and one letter abbreviations of the amino acids are also provided. The~~

amino acids are classified in terms of the nature (nonpolar, polar uncharged and charged) of the side groups. The amino acids marked with a red star (\*) are essential amino acids.]

The  $p_{K_1}$ ,  $p_{K_2}$  and  $p_{K_R}$  values not only reflect the charge distribution of acid base behaviour of amino acids, they also tell us about the distribution of charges in the molecule when dissolved in a buffer of a particular  $p_H$ :

Any weak acid exists as a mixture of the unionized acid and its conjugate base in its aqueous solution.



$$\text{Where. } K_a = \frac{[A^-][H_3O^+]}{[HA]} \quad \text{or.} \quad \frac{[A^-]}{[HA]} = \frac{K_a}{[H_3O^+]}$$

$$\text{Thus } \log\left(\frac{\text{Conjugate base}}{\text{acid}}\right) = \log K_a - \log[H_3O^+] \\ = p_H - p_{K_a}$$

$$\text{If } p_H > p_{K_a} \Rightarrow \log\left(\frac{\text{Conjugate base}}{\text{acid}}\right) > 0 \Rightarrow \frac{\text{Conjugate base}}{\text{acid}} > 1 \\ \Rightarrow [\text{conjugate base}] > [\text{acid}]$$

The opposite will be true if  $p_H < p_{K_a}$ .

Thus if  $p_H > p_{K_1}$  (of carboxyl group), the conjugate base concentration will be greater and the  $\text{CO}_2^-$  group and thus carry a negative charge. But if  $p_H < p_{K_1}$ , then the  $\text{CO}_2^+$

group will not ionize and remain uncharged.

Analogously, if  $p_H > p_{K_2}$  (of conjugate acid- $NH_3^+$ ) then the concentration of the conjugate base  ~~$NH_2^-$~~ , the uncharged  $-NH_2$  group in this case will dominate. But if  $p_H < p_{K_2}$  then the  $-NH_3^+$  group will not undergo ionization and will carry a positive charge.

(b) Charge distribution at the physiological  $p_H$ .

The normal  $p_H$  of blood serum is 7.4. A perusal of the table I-1 will show that the  $p_{K_1}$  values of the  $\alpha\text{-CO}_2\text{H}$  groups of the amino acids lie in ~~the range~~ a small lie within a small range around 2.2 and the  $p_{K_2}$  values of the  $\alpha\text{-NH}_2$  groups lie in a small range around ~~8~~ 9. Clearly, at a  $p_H$  above 3.5, the  $\alpha\text{-CO}_2\text{H}$  groups will exist almost totally as  $\text{CO}_2^-$  and at a  $p_H$  below 8, the  $\alpha\text{-NH}_2$  groups will exist totally as  $-NH_3^+$ .

Hence in the blood at a physiological  $p_H$  around 7.4, all  $\alpha\text{-CO}_2\text{H}$  groups will carry a negative charge and exist as  $\text{CO}_2^-$ , ~~and~~ all  $\alpha\text{-NH}_2$  groups will carry a positive charge and exist as  $-NH_3^+$ . ~~and amino acids will exist as zwitterions.~~ Depending on the exact  $p_H$  and  $p_K$  values, ~~there~~ however, a particular amino acid may have a net positive or a net negative charge.

~~It should also be apparent that at a  $p_H$  below 2, almost all  $NH_2$  groups will~~

## The uncharged acid

the concentration of positively charged ~~or negatively~~ charged molecules <sup>of a particular amino acid.</sup> may be greater or less than the concentration of negatively charged molecules.

It should be obvious that below a  $p_{\text{I}}$  of around 2 almost all molecules will carry a negative charge and above a  $p_{\text{I}}$  of around 10 almost all molecules will have a positive charge.

### (c) Electrophoresis and isolectric points

If a gel consisting of a buffered solution of an amino acid dispersed in e.g. agar agar or polyacryl amide is subjected to an electrical potential then according to the depending on the charges on the amino acid molecules, they will move towards the positive ~~side~~ or the negative electrodes (anode and cathode respectively). This phenomenon is called electrophoresis.

If a given amino acid in a particular buffer has a higher concentration of positively charged molecules we will observe a net movement of amino acids towards the cathode (cataphoresis) or the negative electrode. On the other hand if the negatively charged molecules are at a higher concentration we will observe a net movement towards the anode (anaphoresis).

For every amino acid there exists a buffer  $p_{\text{I}}$  when the concentrations of positively and negatively charged species are exactly

equal. This  $p_H$  value is called ~~as~~ the isoelectric point <sup>$p_I$</sup>  for the given amino acid. ~~In a buffer with  $p_H$  equal to its isoelectric point there will be no net movement of amino acids during electrophoresis.~~ In the table I.1 the  $p_I$  values are also provided

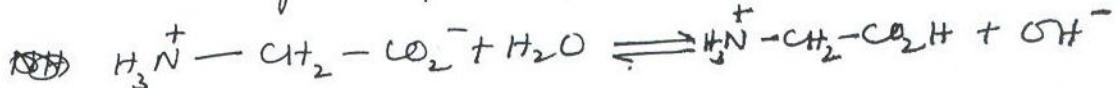
(d) ~~Calculation of isoelectric points of amino acids without acidic or basic side chains~~

Amino Let us take the case of an amino acid glycine. It has a  $p_{K_1}$  (for  $\alpha\text{-CO}_2^+$ ) = 2.35 and a  $p_{K_2}$  (for  $\alpha\text{-NH}_3^+$ ) = 9.78. If we dissolve glycine in water it will hydrolyze as an acid



with  $p_{K_a} = p_{K_2} = 9.78$ .

It will also hydrolyze as a base



$$\text{with } p_{K_b} = p_{K_w} - p_{K_1} = 14 - 2.35 = 11.65$$

The  $p_H$  of the solution will therefore be given as.  
(recall  $p_H$  of a solution of a salt of a weak acid with a weak base).

$$\begin{aligned} p_H &= \frac{1}{2} (p_{K_w} + p_{K_a} - p_{K_b}) = \frac{1}{2} (14 + 9.78 - 11.65) \\ &= \frac{1}{2} (12.13) = 6.06 \end{aligned}$$

This  $p_H$  is also referred to as the isoionic point.

Note that if we substitute the values of  
 $p_{K_a} = p_{K_2}$  and  $p_{K_b} = p_{K_w} - p_{K_1}$  in the relation  
 $p_H = \frac{1}{2} (p_{K_w} + p_{K_a} - p_{K_b})$ .

we have a much simpler and easy to remember relation

$$p_H = \frac{1}{2} (p_{K_1} + p_{K_2})$$

This also naturally gives the same value of the isoionic point

$$p_H = \frac{1}{2} (2.35 + 9.78) = \frac{1}{2} (12.13) = 6.06 .$$

We also know that at  $p_H = 6.06$  the  $[H_3O^+]$  and  $[OH^-]$  values are very small and we can easily invoke the charge balance relation to predict that the two hydrolysis reactions proceed to the same extent. This means:

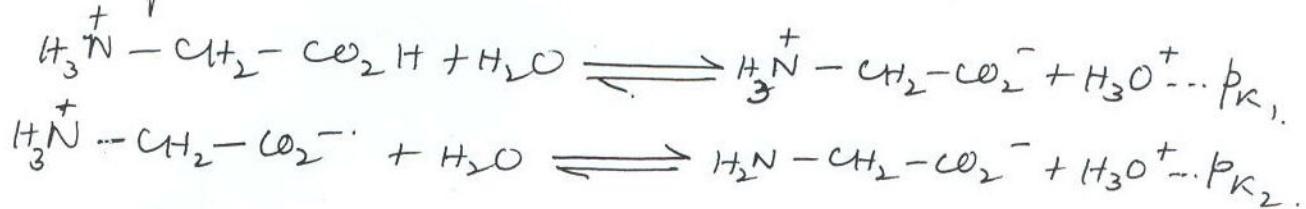
"~~At~~ A solution of glycine in water has a  $p_H = 6.06$  and because of equal degree of hydrolysis of glycine zwitterion (both as acid as well as base), we have the concentration of the negatively charged form  $[NH_2-CH_2-CO_2^-]$  equal to that of the positively charged form  $[H_3N^+-CH_2-CO_2H]$ ".

We can therefore easily extrapolate that if glycine is dissolved in a buffer of  $p_H 6.06$ , the concentrations of the cationic and anionic forms of glycine will be equal. In other words.

"The isoionic point is equal to the isoelectric point of amino acids and has a value.  
~~It~~  $P_I = \frac{1}{2} (p_{K_1} + p_{K_2})$ ".

(11)

There is another way of looking at this situation. Let us consider the titration of the cationic form of glycine with a strong base. The cationic form may be treated as a dibasic acid.



At the half/neutralization point for the first ionization

$$p_H = p_{K_1} \quad \text{and. } [\text{H}_3\overset{+}{\underset{3}{\text{N}}}-\text{CH}_2-\text{CO}_2\text{H}] = [\text{H}_3\overset{+}{\underset{3}{\text{N}}}-\text{CH}_2-\text{CO}_2^-]$$

At the first end point .

$p_H = \frac{p_{K_1} + p_{K_2}}{2}$  and. the situation corresponds to an aqueous solution of  $\text{H}_3\overset{+}{\underset{3}{\text{N}}}-\text{CH}_2-\text{CO}_2^-$  with equal concentrations of  $\text{H}_3\overset{+}{\underset{3}{\text{N}}}-\text{CH}_2-\text{CO}_2\text{H}$  and  $\text{H}_2\text{N}-\text{CH}_2-\text{CO}_2^-$  (isolectric or isoionic point).

(We may further observe that at half neutralization for the second ionization

$$p_H = p_{K_2} \quad \text{and. } [\text{H}_3\overset{+}{\underset{3}{\text{N}}}-\text{CH}_2-\text{CO}_2^-] = [\text{H}_2\text{N}-\text{CH}_2-\text{CO}_2^-]$$

The relative concentrations of the different species will be the same if we externally manipulate the  $p_H$  of the solution by using a buffer.

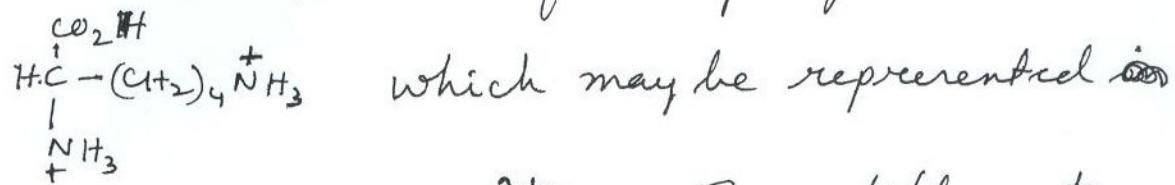
We may use this approach to understand calculate the iso electric point for amino acids containing acidic or basic side chain, as solved examples.

(e) Calculation of isoelectric points of amino acids containing acidic or basic side groups.

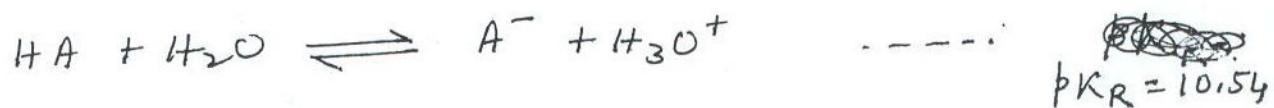
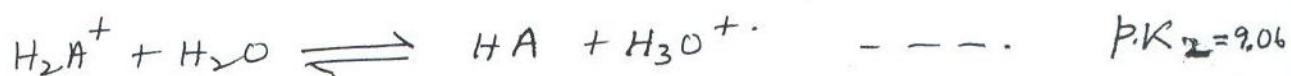
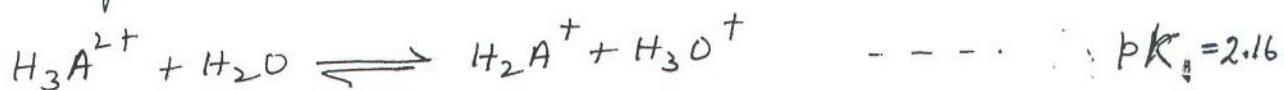
Example:

The  $\text{pK}_1$  ( $\alpha\text{-CO}_2\text{H}$ ),  $\text{pK}_2$  ( $\alpha\text{-NH}_3^+$ ) and  $\text{pK}_R$  ( $\epsilon\text{-NH}_3^+$ ) for the amino acid lysine are respectively 2.16, 9.06 and 10.54. Calculate the isoelectric point of lysine.

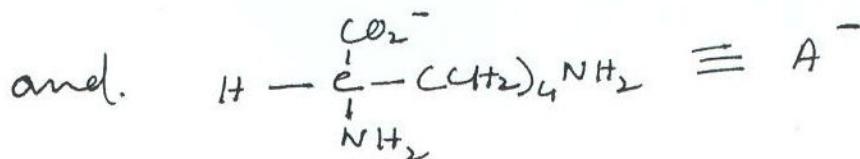
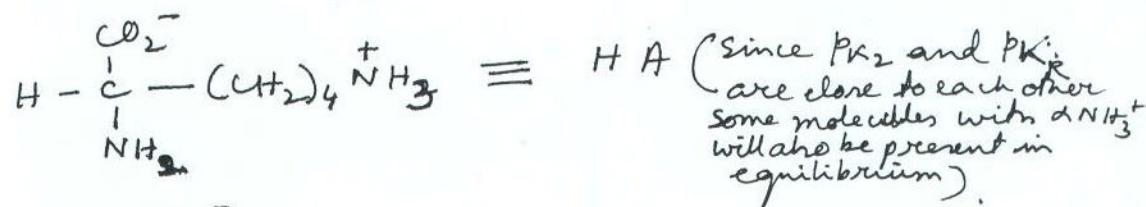
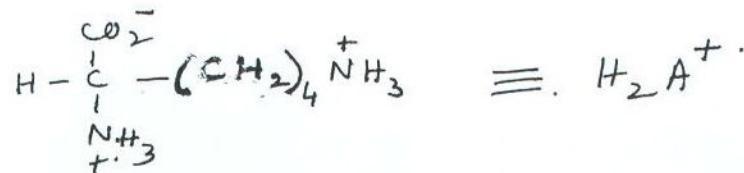
Solution: The ~~cationic~~ form of lysine is



symbolically as  $\text{H}_3\text{A}^{2+}$ . The different dissociation equilibria and their  $\text{pK}_a$  values are given as.



Where.



We now have.

$$\text{at } p_H = p_{K_1}, \quad [H_3A^{2+}] = [H_2A^+].$$

at  $p_H = \frac{p_{K_1} + p_{K_2}}{2}$  we have  $H_2A^+$  along with  $[H_3A^{2+}] = [HA]$  because of hydrolysis of  $H_2A^+$  as acid as well as base.

$$\text{at } p_H = p_{K_2}, \quad [H_2A^+] = [HA].$$

and at  $p_H = \frac{p_{K_2} + p_{K_R}}{2}$  we have  $HA$  along with

$[H_2A^+] = [A^-]$ . because of hydrolysis of  $HA$  as acid as well as base

Hence isolectric point is at

$$\boxed{p_H = \frac{p_{K_2} + p_{K_R}}{2}} = \frac{9.06 + 10.54}{2} = 9.8 \text{ Answer.}$$

For amino acids with basic side chain  
isolectric point is at  $\boxed{\frac{p_{K_2} + p_{K_R}}{2}}$ .

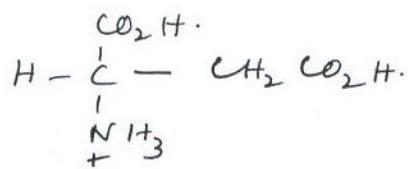
### Example 2

The  $p_{K_1} (\alpha\text{-CO}_2H)$ ,  $p_{K_2} (\alpha\text{-NH}_3^+)$  and  $p_{K_R} (\beta\text{-CO}_2H)$  of aspartic acid are respectively 1.99, 9.9 and 3.90. Calculate the ~~not~~ isoelectric point of aspartic acid.

Solution

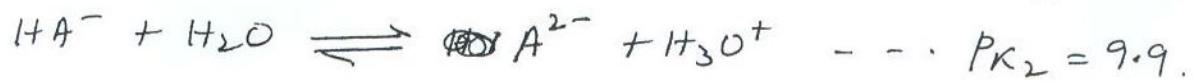
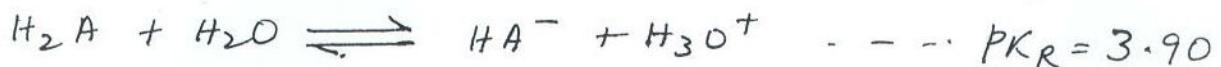
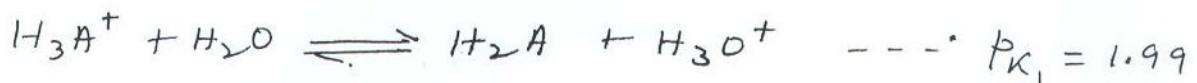
(4)

The fully protonated cationic form of ~~lysine~~ aspartic acid is

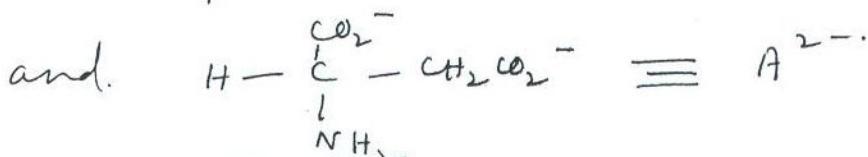
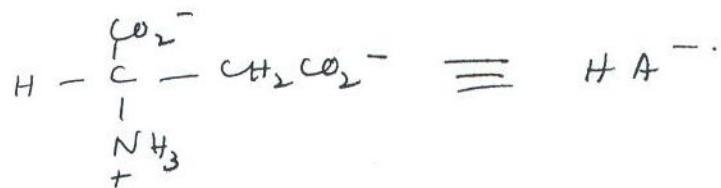
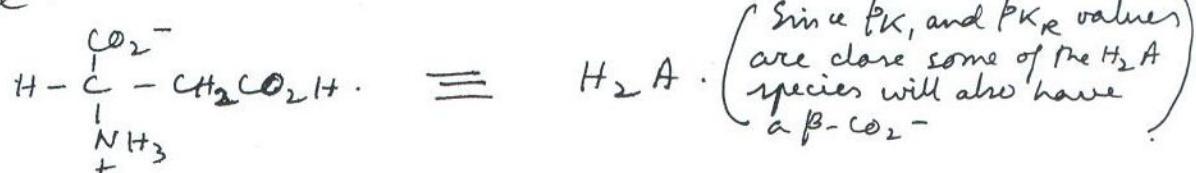


This may be represented as  $\text{H}_3\text{A}^+$ .

The different dissociation equilibria and the corresponding  $\text{pK}_a$  values are as follows.



where -



We now have

$$\text{at } \text{p}_\text{H} = \text{pK}_1 \quad [\text{H}_3\text{A}^+] = [\text{H}_2\text{A}]$$

at  $\text{p}_\text{H} = \frac{\text{pK}_1 + \text{pK}_R}{2}$  we have  $\text{H}_2\text{A}$  along with  $[\text{H}_3\text{A}^+] = [\text{HA}^-]$  since  $\text{H}_2\text{A}$  hydrolyses both as an acid as well as base.

Hence isoelectric point is at .

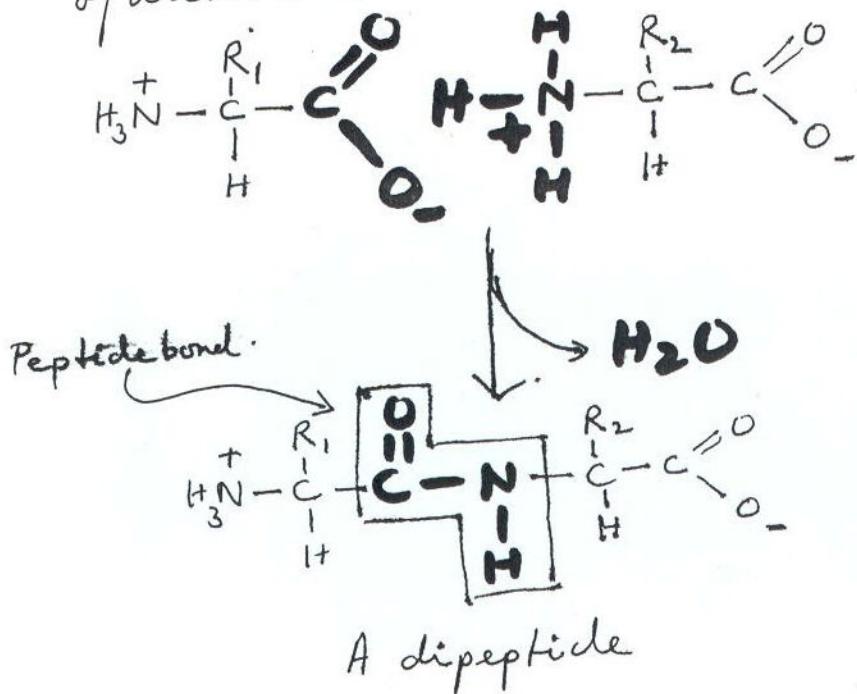
$$p_H = \frac{pK_1 + pK_R}{2} = \frac{1.99 + 3.90}{2} = 2.94. \text{ Answer.}$$

For amino acids with acidic side chain. the isoelectric point is at  $\frac{[pK_1 + pK_R]}{2}$

I. 4. Peptide bonds and <sup>Polypeptides</sup> ~~The structure of proteins~~

(a) The peptide bond.

Amino acids may be joined together by an amide linkage called Peptide bond. It may formally be described Peptide bond formation thus may formally be described as a condensation of two amino acids involving the elimination of water as shown below .



An important aspect of the peptide bond is that C, O, N and H atoms must lie on the same plane in the ground state. This is because there is conjugation of the lone pair on N with  $\text{C}=\text{O}$ . Polymers composed of two, three, a few (3-10), and many amino acid residues (alternatively called peptide units) are known, respectively, as dipeptides, tripeptides, olopeptides, and polypeptides. In general these substances may also be referred to as peptides.

Proteins are molecules that consist of one or more polypeptide chains.

(b) The structure of proteins and poly.

(b) Polypeptides and its amino acid residues.

Polypeptides are linear polymers; that is each amino acid residue is linked to its neighbours in a head to tail fashion rather than forming a branched chain. The great variety in the structure, property and function of polypeptides and proteins is provided by the variation in the side chains of the different amino acid residues and their sequence in the peptide chain. The sequence of amino acids is coded in our genetic material the DNA, which directs the synthesis ~~of proteins~~ of proteins.

(c) Classification of side chains.

The side chains of the 20 different amino acids vary considerably in their physicochemical

properties such as polarity, acidity, basicity, aromaticity, bulk, conformational flexibility, ability to cross link, ability to hydrogen bond and chemical reactivity. It is therefore important that we understand the characteristics of these side chains by examining the structures of the amino acids listed in Table I.1.

### (i) Amino acids with nonpolar side chains

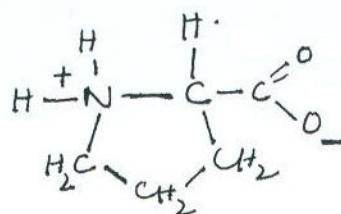
These side chains have a variety of shapes and sizes.

Glycine has the smallest side chain H

Alanine, valine, leucine and isoleucine have aliphatic hydrocarbon side chains ranging in size from a methyl group for alanine to isomeric butyl groups for leucine and isoleucine.

Methionine has a thiol ether side chain  $-CH_2CH_2S-CH_3$ , which resembles the n-butyl group (S has a size similar to  $CH_3$ ).

Proline is a very special case and is the only amino acid that does not have a primary  $-NH_2$  group. It is a cyclic secondary amino acid which can provide conformational constraints and are often responsible for hairpin bends or turns in the amino acid peptide chain.



Proline:

Phenylalanine and tryptophan are amino acids with bulky, aromatic and non polar side groups.

(ii) Amino acids with uncharged polar side chains

There are six amino acids which belong to this class. ~~and so~~ They have hydroxyl, amide or thiol groups.

Serine and threonine bear hydroxyl R-groups of different sizes.

Asparagine and glutamine have side chains of different sizes containing the amide ( $\text{CONH}_2$ ) (you must not confuse these groups with the peptide bonds).

Tyrosine- ~~Tryptophan~~ has a phenolic side chain.

~~Tryptophan~~ Tyrosine, tryptophan and phenylalanine having aromatic side chains are mainly responsible for the UV absorbance and fluorescence exhibited by proteins.

Cysteine has a thiol group and can form disulphide bridges with another cysteine in a different peptide chain (to link two chains) or in the same peptide chain (to provide cross links). ~~The resulting of these~~, ~~so~~ earlier bridged combinations was referred to as Cystine (instead of cysteine). The stiffness and insolubility of hair and nails are because of disulphide linkages between the peptide chains of Keratin through the cysteine residues.

Table I. I.

(1) (2) (3)

Name	Structural formula	$\text{pK}_1$ $\alpha\text{-COOH}$	$\text{pK}_2$ $\alpha\text{-NH}_3^+$	$\text{pK}_{\text{R}}$ Side chain	PI (isoelectric point)
Three letter symbol and one letter symbol				,	

Amino acids with non polar side chains .

Glycine Gly G	$\text{COO}^-$ $\begin{array}{c}   \\ \text{C} \\   \\ \text{H}_2 \\   \\ \text{NH}_3^+ \end{array}$	2.35	9.78	6.06
Alanine Ala A	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H-C-CH}_3 \\   \\ \text{NH}_3^+ \end{array}$	2.35	9.87	6.11
Valine Val V	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H-C-CH} \\   \\ \text{CH} \\   \\ \text{NH}_3^+ \end{array}$	2.29	9.74	<del>6.02</del>
Leucine Leu L	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H-C-CH}_2-\text{CH} \\   \\ \text{CH} \\   \\ \text{NH}_3^+ \end{array}$	2.33	9.74	<del>6.04</del>
Isoleucine Ile I	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H-C-C-CH}_2-\text{CH}_3 \\   \\ \text{H} \\   \\ \text{NH}_3^+ \end{array}$	2.32	9.76	6.04

(20) (MVA)

Name	Structural formula	$\text{pK}_1$ $\alpha\text{-COOH}$	$\text{pK}_2$ $\alpha\text{-NH}_3^+$	PR Side chain	PI gزو-electric point.
Three letter symbol and one letter symbol					
Methionine*	$\text{COO}-$	2.13	9.28		5.70
Met M	$\begin{array}{c} \text{H}-\overset{\text{l}}{\underset{\text{l}}{\text{C}}}-\text{CH}_2-\text{CH}_2-\text{S}-\text{CH}_3 \\   \\ \text{NH}_3^+ \end{array}$				
Proline	$\text{COO}-\overset{\text{H}_2}{\underset{\text{C}}{\text{C}}}-$	1.95	10.64		6.30
Pro P	$\begin{array}{c} \text{CH}_2 \\   \\ \text{C} \\ / \quad \backslash \\ \text{H} \quad \text{N} \\   \\ \text{H}_2 \end{array}$				
Phenylalanine*	$\text{COO}-$	2.20	9.31		5.76
Phe F	$\begin{array}{c} \text{H}-\overset{\text{l}}{\underset{\text{l}}{\text{C}}}-\text{CH}_2-\text{C}_6\text{H}_4 \\   \\ \text{NH}_3^+ \end{array}$				
Tryptophan*	$\text{COO}-$	2.46	9.41		5.94
Trp W	$\begin{array}{c} \text{H}-\overset{\text{l}}{\underset{\text{l}}{\text{C}}}-\text{CH}_2-\text{C}_6\text{H}_4-\text{NH}_3^+ \\   \\ \text{H} \end{array}$				

Amino acids with uncharged polar side chains.

Serine	$\text{COO}-$	2.19	9.21		5.7
Ser S	$\begin{array}{c} \text{H}-\overset{\text{l}}{\underset{\text{l}}{\text{C}}}-\text{CH}_2\text{OH} \\   \\ \text{NH}_3^+ \end{array}$				

(21) 11/16/14

Name	Structural formula	$\text{pK}_1$ $\alpha\text{-COOH}$	$\text{pK}_2$ $\alpha\text{-NH}_3^+$	$\text{pK}_R$ Side chain	$\text{pI}$ Dielectric point
Three letter symbol and one letter symbol					
Threonine *	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{C}-\text{CH}_3 \\   \quad   \\ \text{OH} \quad \text{NH}_3^+ \end{array}$	2.09	9.10		<del>5.60</del>
Thr.					5.60
T					
Asparagine	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{C}(=\text{O})\text{NH}_3^+ \end{array}$	2.14	8.72		5.43
Asn					
N					
Glutamine	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{CH}_2-\text{C}(=\text{O})\text{NH}_3^+ \end{array}$	2.17	9.13		5.65
Gln					
Q					
Tyrosine	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{C}_6\text{H}_4\text{OH} \\   \\ \text{NH}_3^+ \end{array}$	2.20	9.21	10.46 (phenol)	6.33
Tyr					
Cysteine	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2\text{SH} \\   \\ \text{NH}_3^+ \end{array}$	1.92	10.70	8.37 (sulfhydryl)	5.14
Cys					
C					

(22)

14/20

Three letter Symbol and One letter Symbol	Structural formula	$\text{pK}_1$ $\alpha\text{-COOH}$	$\text{pK}_2$ $\alpha\text{-NH}_3^+$	$\text{pK}_R$ Side Chain	PI Groeder point.
---	--------------------	---------------------------------------	---	-----------------------------	----------------------

Amino acids with charged polar side chains.

Lysine *	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H}-\text{C}-\text{(CH}_2)_4-\text{NH}_3^+ \\   \\ \text{NH}_3^+ \end{array}$	2.16	9.06	10.54	9.8
				$(\text{E-NH}_3^+)$	
Arginine *	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H}-\text{C}-\text{(CH}_2)_3-\text{NH}-\text{C}^{\text{NH}}_2 \\   \\ \text{NH}_3^+ \quad " \quad \text{NH}_3^+ \end{array}$	1.82	8.99	12.48	10.74
				$(\text{guani-} \text{dino})$	
Histidine *	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H}-\text{C}-\text{CH}_2-\text{C}_\text{H}=\text{N} \\   \\ \text{NH}_3^+ \quad \text{H} \end{array}$	1.80	9.33	6.04	7.680
				$(\text{imida-} \text{zole})$	
Aspartic acid	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H}-\text{C}-\text{CH}_2-\text{C}^{\text{O}} \\   \\ \text{NH}_3^+ \quad \text{O}^- \end{array}$	1.99	9.90	3.90	2.940
Asp				$(\beta\text{-COOH})$	
D					
Glutamic acid	$\text{COO}^-$ $\begin{array}{c}   \\ \text{H}-\text{C}-\text{CH}_2-\text{CH}_2-\text{C}^{\text{O}} \\   \\ \text{NH}_3^+ \quad \text{O}^- \end{array}$	2.10	9.47	4.07	3.080
Glu				$(\gamma\text{-COOH})$	
E					

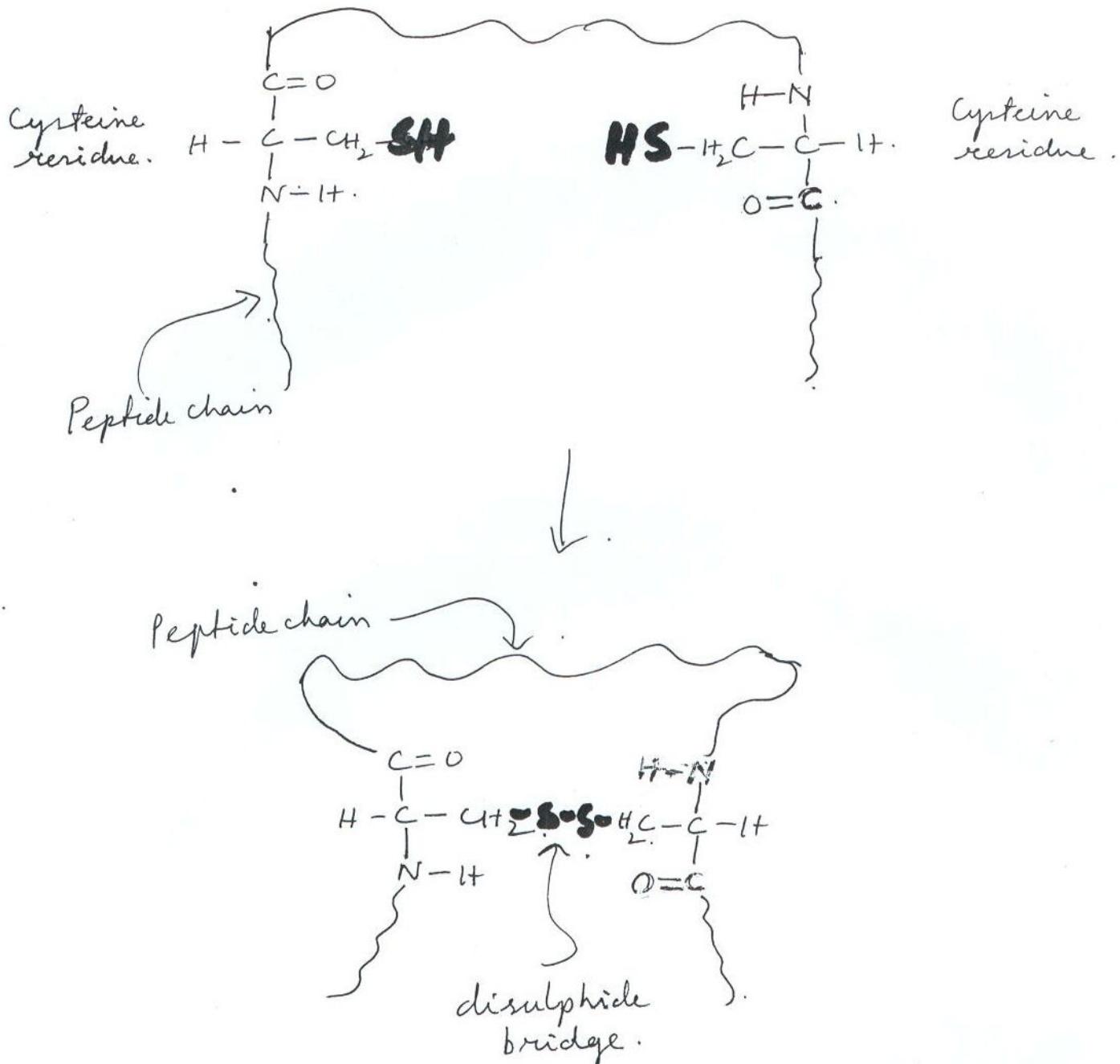


Figure: The cystine residue consists of two disulphide linked cysteine residues.

### (iii) Amino acids with charged polar side chains

The side chains in the five amino acids belonging to this class may be positively or negatively charged.

Lysine, which carries a butylammonium side chain, arginine, which carries a

guanidino group and histidine which has an imidazolium moiety, all have basic side groups, which are positively charged at physiological pH values.

Of these, histidine has  $pK_a = 6.0$  and is very special in the sense that at the acidic end of the physiological pH range, they are partially charged and at the basic end they are uncharged. The histidine side chain therefore, often participates in the catalytic reactions of enzymes.

The acidic amino acids, aspartic acid and glutamic acid, are negatively charged above pH 3; in their ionized state, they are often referred to as aspartate and glutamate. Asparagine and glutamine are respectively the amides of aspartic acid and glutamic acid.

#### F.5 The Structure of proteins

##### (a) The primary structure -

The primary structure of proteins implies the sequence of amino acids present in the polypeptide chain (or chains).

##### (b) The secondary structure -

The long, flexible peptide chains of proteins are folded into relatively regular conformations called the secondary structure. This is largely a consequence of hydrogen bonding, which can be formed between the C=O and N-H groups of different peptide bonds. The formation of such hydrogen bonding between amide groups within the same chain, causes the peptide

chain to coil up into a spiral structure called the  $\alpha$ -helix. ~~Wool~~ This type of structure is adopted by fibrous proteins such as those present in wool (keratin) and hair (collagen). These proteins are elastic since on application of tension the hydrogen bonds break up and the coil stretches out like a spring. On releasing the tension the hydrogen bonds are formed again and the short end coiled structure is restored.

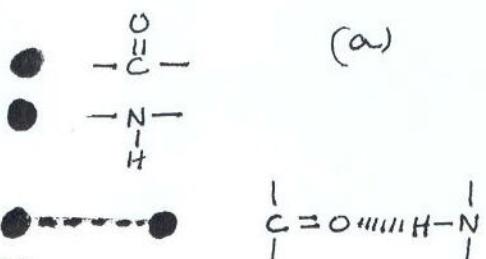
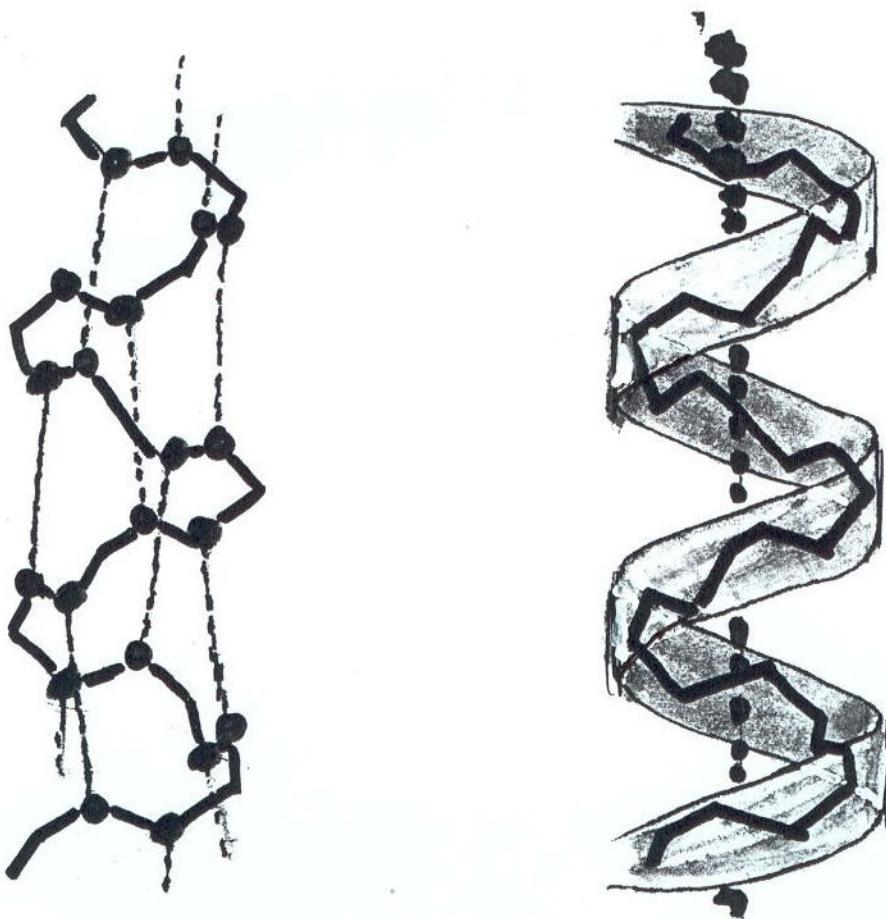
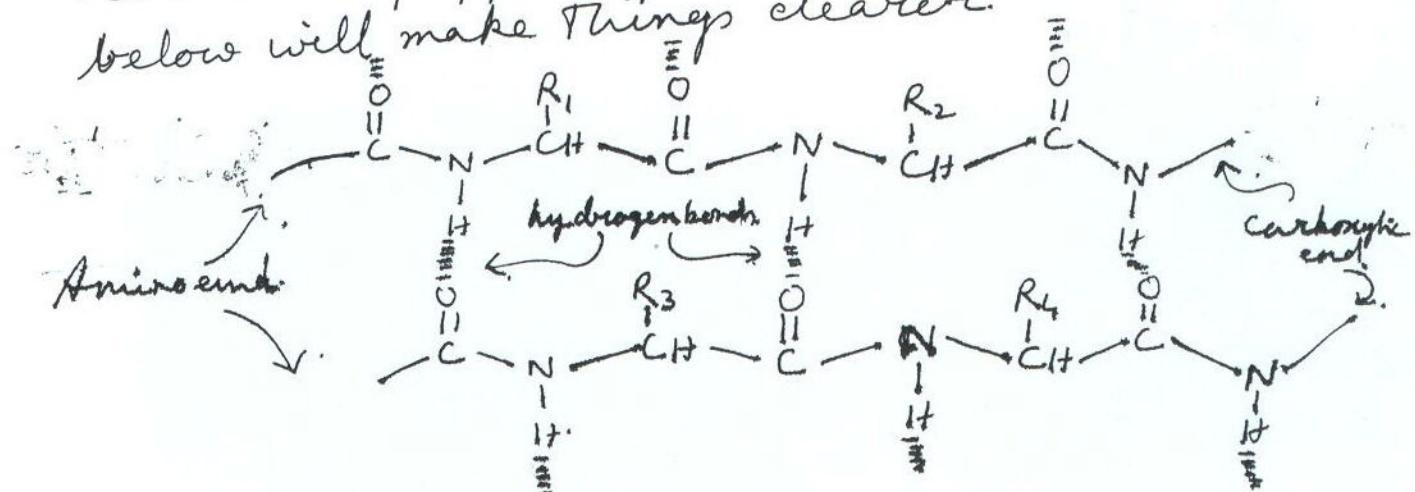
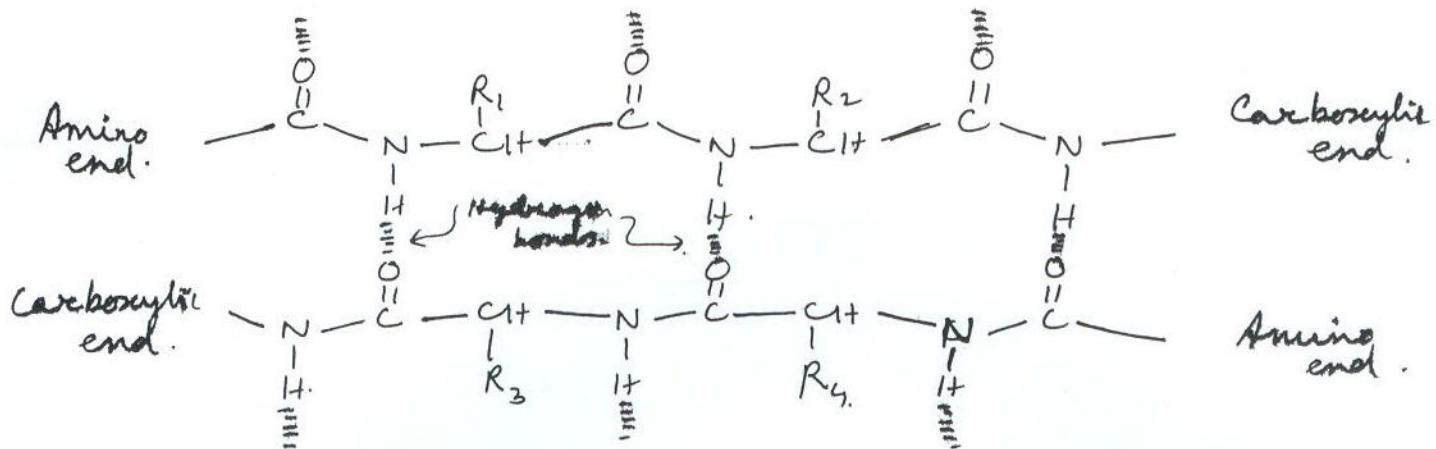


Figure  
 (a) Hydrogen bonding in  $\alpha$ -helix    (b) The helical structure.

Silk is also made up of the fibrous protein fibroin. But silk is not elastic. On the other hand silk fabrics have their characteristic mechanical properties such as easy bending and folding. This is because the secondary structure of proteins such as fibroin ~~too~~ belongs to another class called the ~~peptides~~  $\beta$ -pleated or sheet structure. In such cases, ~~position~~ the peptide chains are not coiled and the hydrogen bonding takes place between ~~too or more~~ the C=O and N-H groups of two or more different chains (or different regions of the same chain). Several sheet proteins may then be stacked upon one another giving a three dimensional structure. The peptide chains involved in  $\beta$ -sheet formation may be parallel to each other (both having the same direction of amino to carboxylic ends) or antiparallel. To each other (the amino to carboxylic directions being oppositely oriented). The figure below will make things clearer.



Parallel  $\beta$ -sheet arrangement



### (C) The tertiary structure:

In a typical polypeptide the chain is folded up in a complex fashion giving a definite native ('shape') to the whole molecule. The 'shape' is sensitive to the temperature and  $p_H$  of the solvent in which the protein is dissolved and even slight changes in  $p_H$  may cause the peptide chain to become permanently changed. This phenomenon is called the denaturation of proteins. A ~~common~~ example is the "splitting" of milk at low  $p_H$  ~~is~~ accompanied by increase in temperature when milk proteins coagulate and separate out as 'paneer'. ~~Sometimes~~ Some times the cooling to the native state and uncoiling due to addition of certain reagents is a reversible process.

The  $p_H$  sensitivity as well as attainment of the definite native shape are because of interaction of the side chains among themselves and with the solvent (water).

Normally in the polypeptide chain

There are several regions with  $\alpha$  L (helical) or  $\beta$  (sheet) secondary structures. The subsequent folding up of the chain consisting of regions of secondary structures is described by the Tertiary structure of proteins.

The various forces that stabilize the tertiary structure of proteins are

- (i) hydrogen bonding involving  $C=O$ ,  $N-H$ ,  $O-H$  and other polar groups.
- (ii) Ionic bonding or salt bridge formation involving side chain  $CO_3^{2-}$  (eg in aspartate or glutamate residues) and  $NH_3^+$  (eg in lysine, arginine or histidine residues).
- (iii) disulphide bridges involving cysteine residues
- (iv) Hydrophobic interactions involving the hydrophobic side chains of residues such as leucine or tryptophan etc.

According to the most popular theory regarding the primary driving force responsible for the folding of a protein into its native tertiary structure is hydrophobic interactions. The hydrophobic groups tend to get away from the solvent and tend to group up.

This is sometimes also described as hydrophobic bonding. ~~to the~~

Globular proteins provide excellent examples of hydrophobic bonding. In such proteins the peptide chain folds up in the form of a globule with the hydrophobic residues grouped up into the inner core and the ~~groups with polar~~ residues with polar side chains are exposed to the solvent by opening up towards the outer surface.

The figure below explains some of these aspects of tertiary structure of proteins.

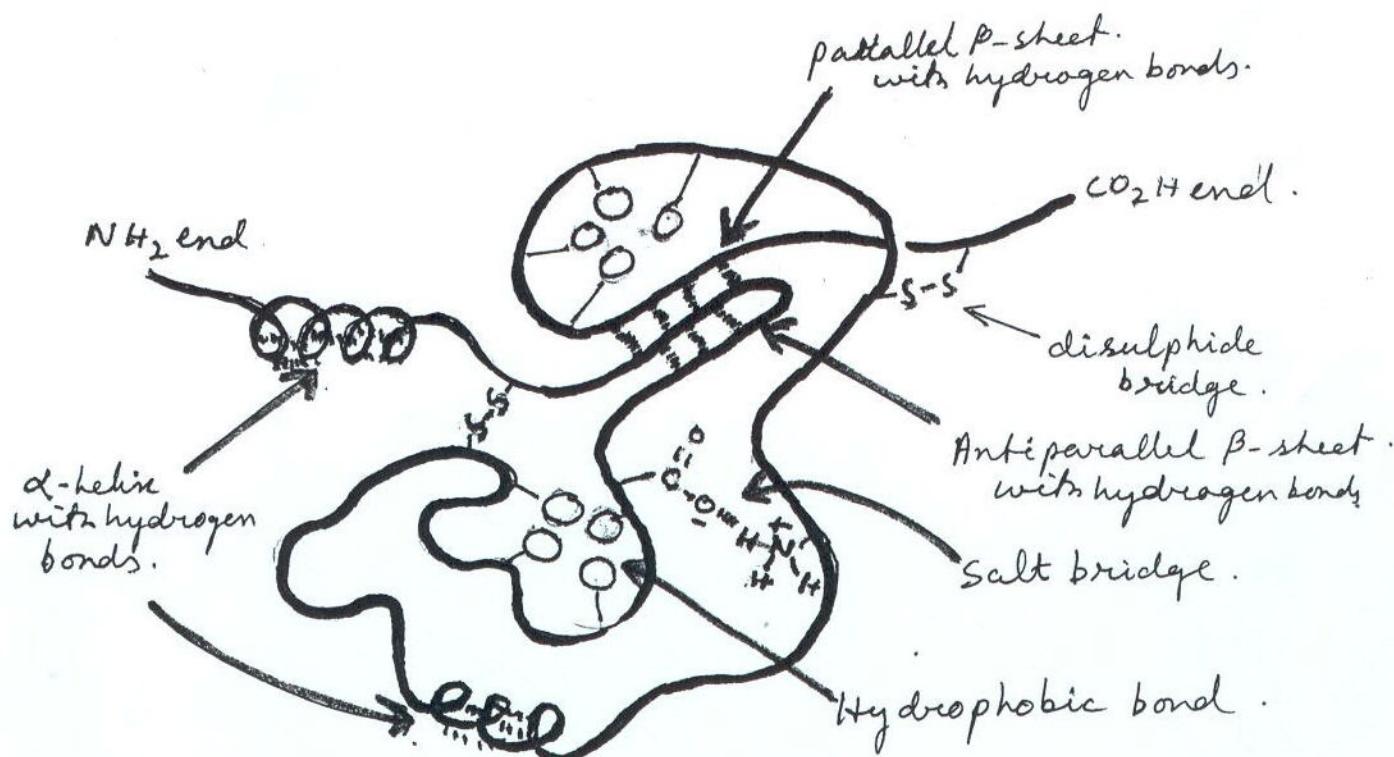


Figure.

Tertiary structure of a protein — a schematic diagram showing the forces involved.

(d) The quaternary structure of proteins

Many proteins have more than one peptide chain where each peptide chain has its own primary, secondary and tertiary structures. The arrangements of these different chains with respect to each other in the complete protein is described as the quaternary structure of the protein.

For example in the protein haemoglobin the quaternary structure describes how the two  $\alpha$ -chains and the two  $\beta$ -chains are grouped around the haeme porphyrin unit.