CHAPTER 8

CHEMISTRY OF AMINO ACIDS

CHAPTER AT A GLANCE

After reading this chapter the learner would be able to understand and appreciate the followings:

- Structure of amino acids
- Classification of amino acids
- Physical, electrical and chemical properties of amino acids.
- Derivatives of standard amino acids
- Non-standard amino acids
- Biological functions of amino acids

- **Amino Acids** are the basic **structural units (monomers) of proteins.** Although over 300 different amino acids are known to occur in cells and tissues, only 20 of them universally occur in the proteins of all forms of life. These 20 amino acids are called **protein amino acids or stander amino acids** or **magic 20.** The others are called non-standard or non-protein amino acids.
- **Asparagine** was the 1st amino acid discovered and isolated from protein of Asparagus in 1806 and the last of the 20 amino acids to be discovered was **threonine.**
- In recent years twonew amino acids have been added to this list namely **selenocysteine** and **pyrrolysine**. Both are directly incorporated into proteins during translation. Interestingly they are coded by stop codons; stop codon UGA can code for selenocysteine, while pyrrolysine can be coded by stop codon UAG. Selenocysteine is an unusual amino acid containing trace element selenium in place of sulfur atom of cysteine. Selenocysteine is usually found in the active sites of certain enzymes **(selenoproteins)**, e.g. glutathione peroxidase, glycine reductase, thioredoxin reductase etc.

STRUCTURE OF AMINO ACIDS

1. Amino acids are organic compounds containing an amino group and a carboxylic group. Hence, they are also called **amino carboxylic acids.** As both carboxyl and amino groups are attached to the same carbon (a-carbon atom), amino acids are also called **α-amino acids.**
2. **Amino acids are substituted methanes** with 4 substituent groups occupying the four valence positions of the carbon. The four groups of each a-amino acid consists of an **amino group** (-NH2), a **carboxyl group**, (-COOH), a hydrogen atom and a variable **side chain** or **R- group** bonded to a central carbon atom called **α-Carbon atom.**

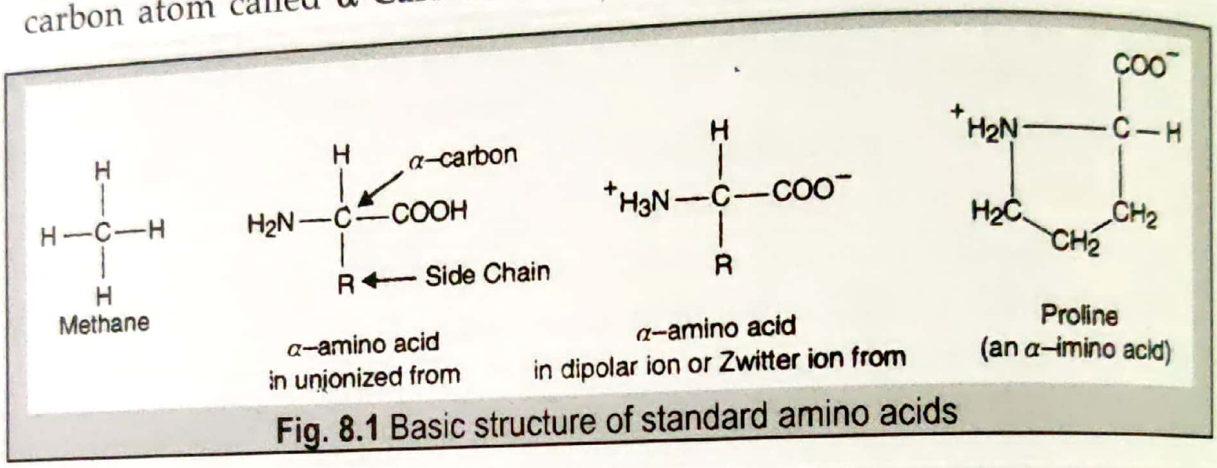


Fig. 8.1 Basic structure of standard amino acids

3. Out of 20 standard amino acids nineteen are **α-amino acids** andproline is the exceptional **α-imino acid** having imino group (-NH-). In proline, the R-group is bonded to both the nitrogen and a-carbon atom, resulting in a cyclic structure. Thus, proline has a secondary rather than primary amino group.
4. The R-group can be hydrogen (in glycine), a methyl group (in alanine), hydroxyl methyl (in serine) or an aliphatic group or an aromatic group or a heterocyclic group.
5. **Nomenclature** - The standard amino acids have been assigned three letter abbreviations and one letter symbols, e.g., for Glycine, Gly is the abbreviation and G is the symbol.

CLASSIFICATION OF AMINO ACIDS

The amino acids are classified into various ways.

1. Classification based on their incorporation in proteins
2. Classification based on the structure of the side chain (R group)
3. Classification based on the polarity of the side chain (R group)
4. Classification based on reaction in solution
5. Classification based on the position of amino group
6. Classification based on metabolism
7. Classification based on nutritional requirements

(1) **Classification of Amino Acids Based on their incorporation in proteins.**

Based on incorporation of amino acids in protein, they are classified into two groups :

(i) Proteogenic amino acid.

(ii) Non-proteogenic amino acid.

(i) Proteogenic amino acid : The amino acids which are used for synthesis of protein are called proteogenic or proteinogenic amino acids. Again they are of two types.

(a) Major amino acids (b) Rare amino acids.

(a) Major amino acids : The major amino acids are involved in synthesis of vast number of proteins. They are 20 in numbers listed below.

(b) Rare amino acids : are the derivatives of the 20 major amino acids. E.g. (i) Hydroxyproline, (ii) hydroxylsine, (iii) aminocitric acid, (iv) aspargine (v) glutamine.

(ii) Non-proteogenic amino acid : The amino acids which are not incorporated into the proteins or not participate in protein synthesis are called non-proteogenic amino acids. They occur in cells either in a free state as metabolites or as part of non-protein compounds. Example - ornithine, citrulline, y aminobutric acid, P-alanine, pantothenic acid, thyroxine and many others.

(2) **Classification Based on the structure of side chain :** On the basis of the nature of their side chain, amino acid classified into following categories.

A. **Aliphatic amino acids :**

a. **Monoamino monocarboxylic acids :**

- Simple amino acids : Glycine, Alanine (Fig. 8.2).
- Branched chain amino acids : Valine, Leucine, Isoleucine (Fig. 8.3).
- Hydroxy amino acids : Serine, Threonine (Fig. 8.4).
- Sulfur-containing amino acids : Cysteine, Methionine (Fig. 8.5).
- Amino acids with amide group : Asparagine, Glutamine (Fig 8.6).

b. **Monoamino dicarboxylic acids :** Aspartic acid, Glutamic acid (Fig. 8.7).

c. **Dibasic monocarboxylic acids :** Lysine, Arginine (Fig. 8.8).

B. **Aromatic amino acids :** Phenylalanine. Tyrosine (Fig. 8.9).

C. **Heterocyclic amino acids :** Tryptophan (Fig. 8.10). Histidine (Fig. 8.11).

D. **Imino acid :** Proline (Fig. 8.11).

E. **Derived amino acids :**

a. **Derived amino acids found in proteins :** After the synthesis of proteine, some of the amino acids are modified, e.g. hydroxy proline (Fig. 8.12) and hydroxy lysine are important components of collagen. Gamma carboxylation of glutamic acid residues of proteins is important for clotting process (Fig. 8.12). In ribosomal proteins and in histones, amino acids are extensively methylated and acetylated.

b. Derived amino acids not seen in proteins : (Non-proteins amino acids) : Some derived amino acids are seen free in cells, e.g. Omithine (Fig. 8.12), Citrulline, Homocysteine. These are produced during the metabolism of amino acids. Thyroxine may be considered as derived from tyrosine.

c. Non-alpha amino acids : Gamma amino butyric acid (GABA) is derived from glutamic acid. Beta alanine, where amino group is in beta position, is a constituent of pantothenic acid (vatimin) and coenzyme A.

Special Groups in Amino Acids :

In the figures, special groups are boxed. Arginine contains **guanidinium (guanidine)** group; Phenylalanine **(benzene);** Tyrosine **(phenol);** Tryptophan (indole); Histidine **(imidazole);** and Proline **(pyrrolidine)** (Table 3.1). Proline has a secondary amino group, and hence it is an imino acid.

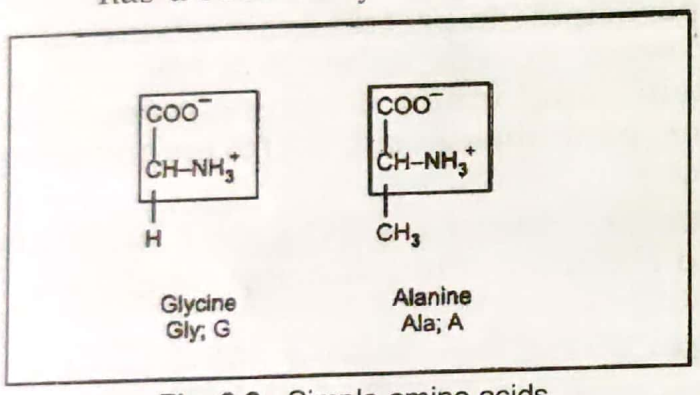


Fig. 8.2 : Simple amino acids

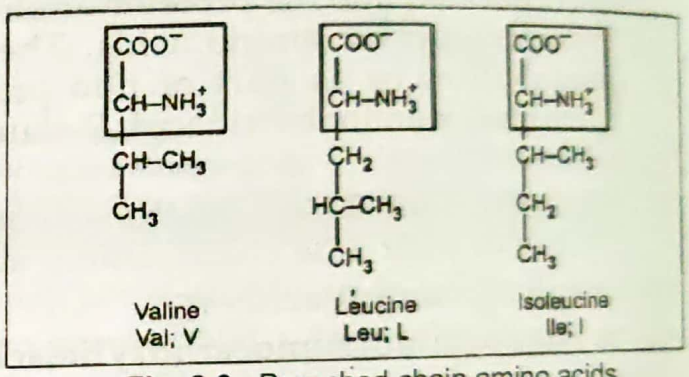


Fig. 8.3 : Branched chain amino acids

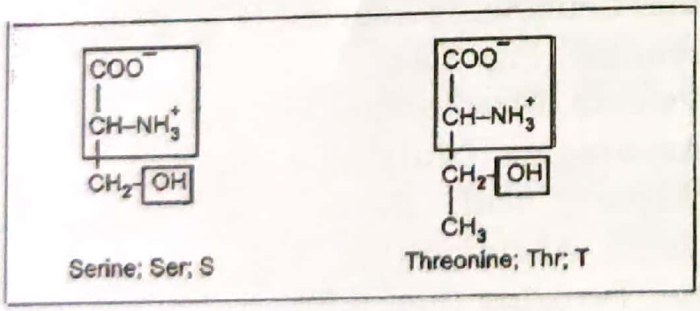


Fig. 8.4 : Hydroxy amino acids

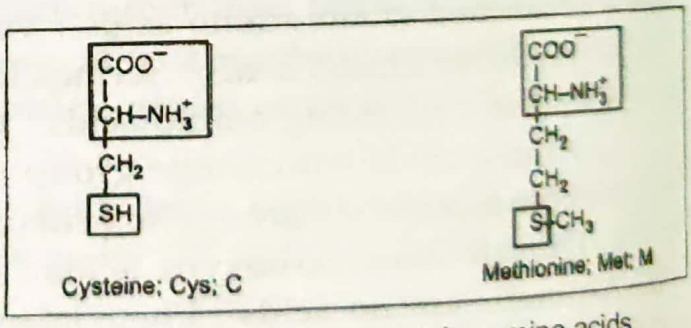


Fig. 8.5 : Sulfur-containing amino acids

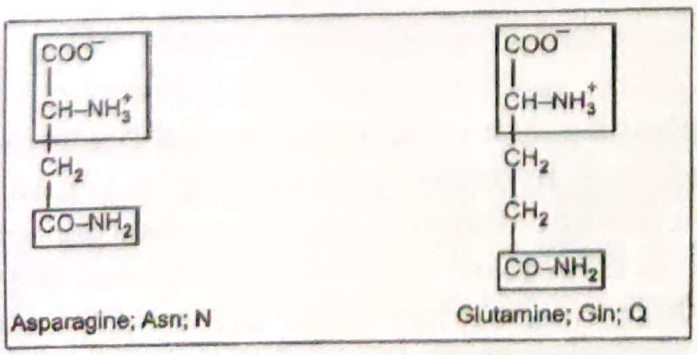


Fig. 8.6 : Amino acids with amide groups

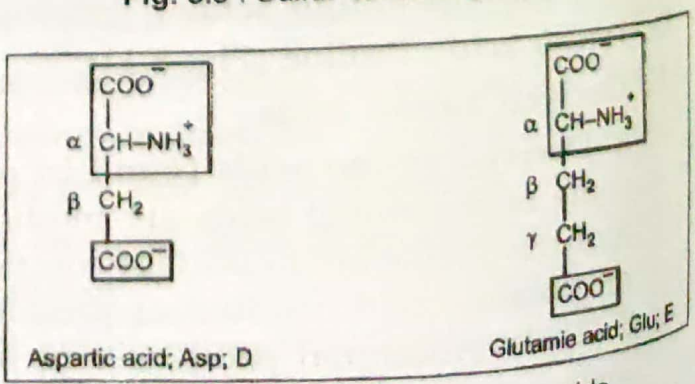


Fig. 8.7 : Dicarboxylic amino acids

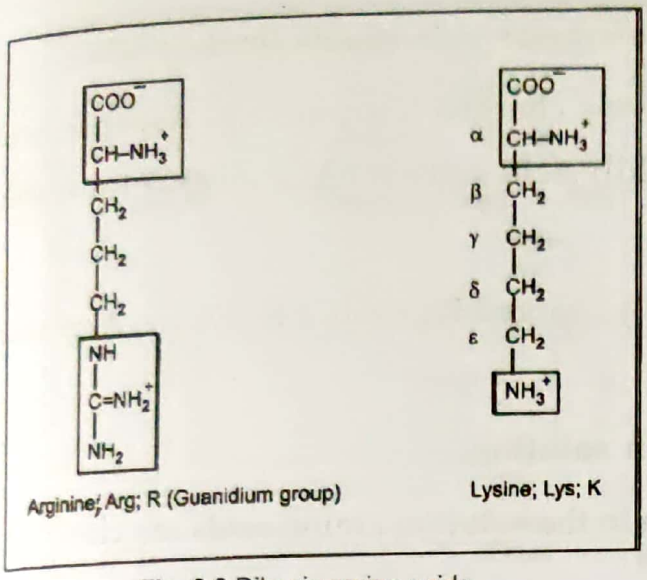


Fig. 8.8 Dibasic amino acids

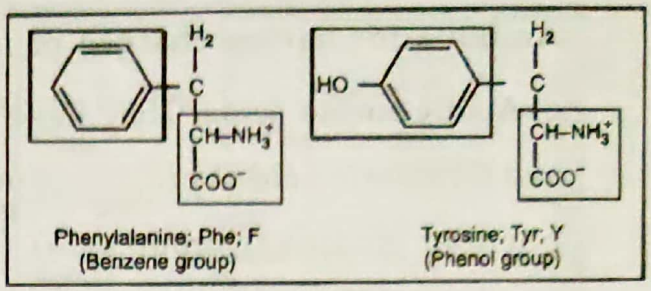


Fig. 8.9 Aromatic amino acids

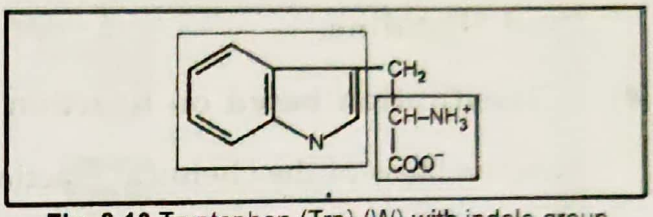


Fig. 8.10 Tryptophan (Trp) (W) with indole group

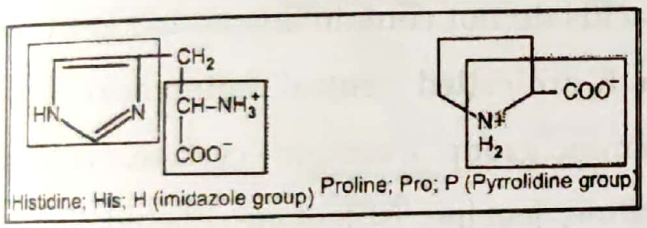


Fig. 8.11 Histidine and proline

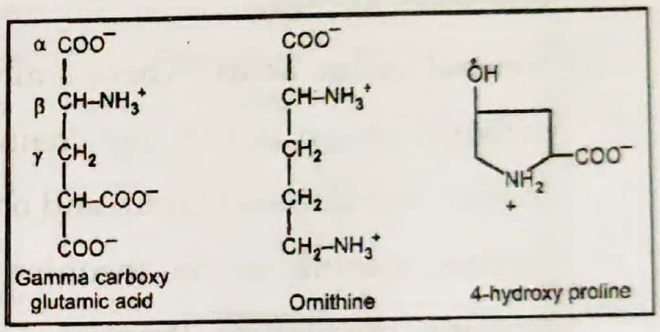


Fig. 8.12 Some derived amino acids

(3) **Classification based on the polarity of the side chain.**

On the basis of polarity of the R groups, amino acids are of following types –

A. **Amino acids having non-polar (hydrophobic) side chains:** These amino acids are hydrophobic (water repellant) and lipophilic. They include, **proline, aromaticamino acids** (phenylalanine and tryptophane) and **aliphatic amino acids** (alanine, valine, leucine, isoleucine and methionine). The parts of protein made up of these amino acids will be hydrophobic in nature and occupy the interior of protein.

B. **Amino acids having uncharged or non-ionic polar side chains (hydrophilic):** This group includes amino acids having R group as (a) hydrogen group (-H) group as in **glycine**, (b) hydroxyl group (-OH) as in **serine** and **threonine**, (c) sulf-hydryl group (-SH) as in **cysteine**, (d) amide group (-CO-NH2) as in **asparagine** and **glutamine.**

C. **Amino acids having charged or ionic polar side chains (hydrophilic):**

a. **Acidic amino acids:** They have –vely charged R group, e.g. **Aspartic acid** and **Glutamic acid**. Tyrosine is mildly acid since it has a weakly ionisable phenolic hydroxyl group.

b. **Basic amino acids:** They have +vely charged R group, e.g. Lysine, **Arginine** and **Histidine.**

(4) **Classification based on Reaction in solution.**

On the basis of the chemical reaction in the solution amino acids are classified into three groups :

(i) **Neutral amino acids :** These amino acids do not contain any amino group or carboxyl group in the side chain or R are called neutral amino acid. They contain one carboxyl group and one amino group. Example - cystine, cysteine, glycine, alanine, serine, theonine, valine, leucine, isoleucine, phenylalanine, tyrosine, methionine, thyroxin, tryptophan, proline, hydroxyproline etc.

(ii) **Acidic amino acids :** These amino acids contain additional carboxylic groups in the side chain. Example - Aspartic acid, glutamic acid, aminocitric acid etc.

(iii) **Basic amino acids :** These amino acids contain an additional amine group in the side chain. Example - Lysine, arginine, histidine, etc.

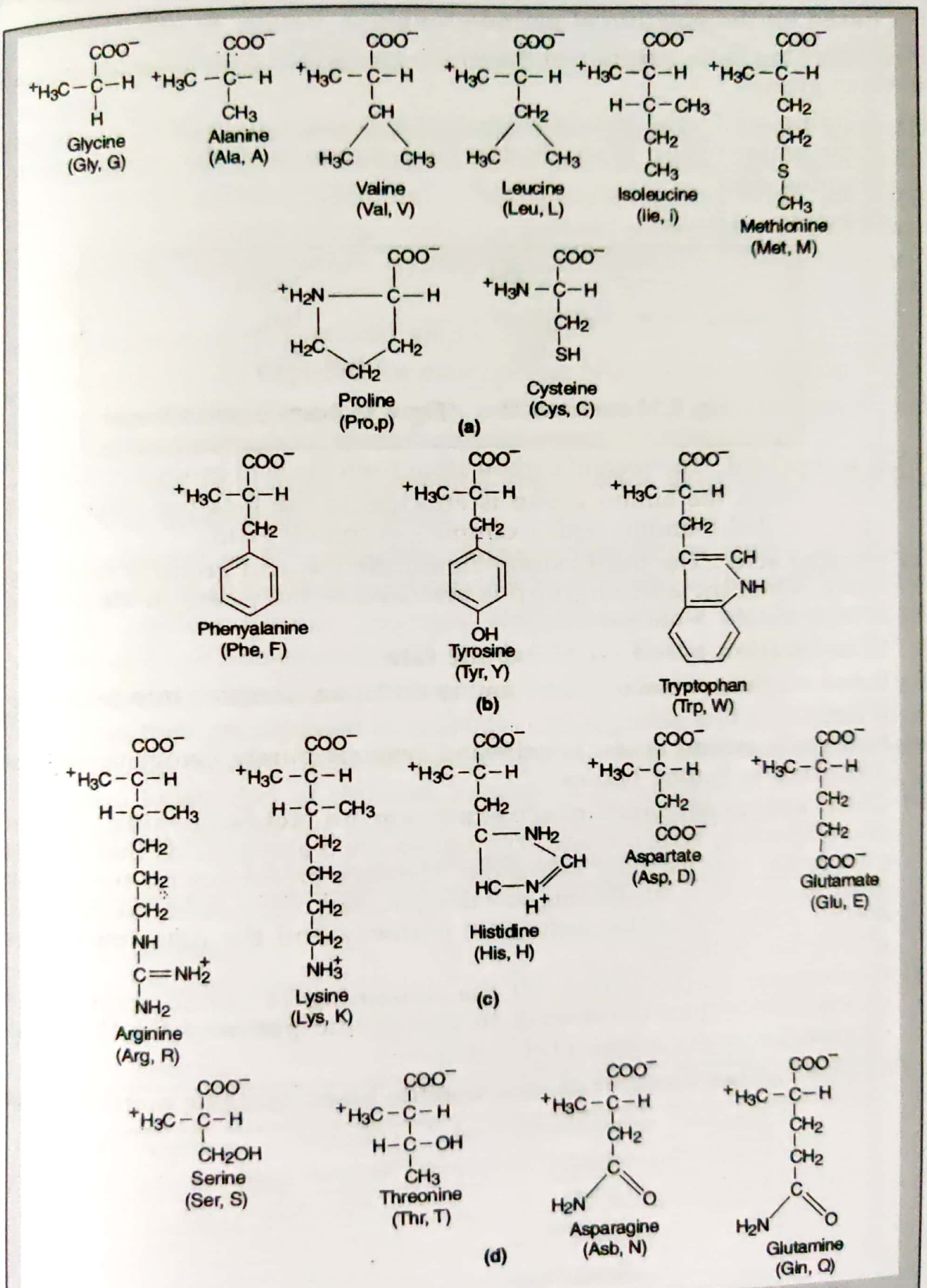


Fig. 8.13 The 20 standard amino acids (a) Hydrophobic, aliphatic R groups, (b) hydrophobic, aromatic side groups (c) Polar charged R groups (d) Polar, uncharged R groups

5. **Based on the position of amino group :**

According to the position of the amino group (NH_2), amino acids are classified into three groups :

(a) **α-amino acid :** The carbon atom next to acid group is called ot-carbon atom. If the amino group is attached to the a-carbon atom the amino acid is called α-amino acid.

Example - Alanine

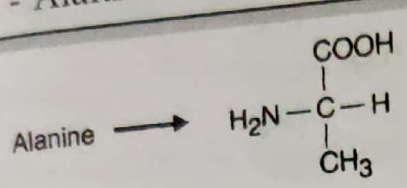


Fig. 8.14 α-aminoacid

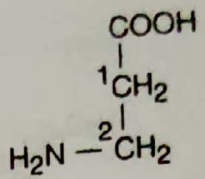


Fig. 8.15 β-amino propionic acid

(b) **β-amino acid :** The second carbon atom from the acid group is called P-carbon atom. When the amino group is attached to the β-carbon atom, the amino acid is called P-amino acid. Example - Propionic Acid.

(c) **γ-amino acid :** The third carbon atom from the acid group is caleld γ-carbon atom. When the amino group is attached to the γ-carbon atom the amino acid is called γ-amino acid.

(6) **Classification based on Metabolic fate.**

Based on their metabolic fate, amino acids are classified into following three types:

(a) **Ketogenic amino acids: Leucineand lysineis** purely ketogenic because it is converted to ketone bodies.

(b) **Both ketogenic and glucogenic amino acids:** Lysine, isoleucine, phenylalanine, tyrosine and tryptophan are 5 amino acids that are partially glucogenic and partially ketogenic. They are precursors for the synthesis of glucose as well as fat. During metabolism, part of the carbon skeleton of these amino acids will enter ketogenic pathway and the other part enter into glucogenic pathway.

(c) **Glucogenic amino acids:** All the remaining 14 amino acids are purely glucogenic as they enter only to glucogenic pathway for the formation of glucose or glycogen. (Table-8.1).

TABLE-8.1: Classification of amino acids based on their metabolic fate.

Ketogenic (lamino acid)	Both glucogenic and ketogenic (5 amino acids)	Glucogenic (14 amino acids)
Leucine	Lysine	Glycine, Alanine, Valine
	Isoleucine	Serine, Threonine, Cysteine
	Phenylalanine	Methionine, Asparagine, Glutamine
	Tyrosine	Aspartic acid, Glutamic acid
	Tryptophan	Arginine, Histidine, Proline

(7) **Classification of Amino Acids on Nutritional Requirements.**
On the basis of their synthesis in animal body, amino acids are classified into 4 groups :

(a) **Essential** or **Indispensable amino acids:** The carbon skeleton of some amino acids can't be synthesized is animal body, and therefore must be supplied in their diet and are called **essential** or **indispensableamino acids.** For example - in human essential amino acids are **valine, isoleucine, phenylalanine, methionine, leucine, lysine, tryptophan** and **threonine.** (Remember as VIP, MLL, TT).

Biological importance of Essential amino acids:

1. These amino acids required for maintaining N-equilibrium in man.
2. Normal growth and optimal health will not occur, if one such amino acid is deficient in the diet.
3. Quantity and proportion of essential amino acids determine the **biological value** (nutritive value) of proteins.

(b) **Non-essential or Dispensable amino acids:** These amino acids can be synthesized in animal body and may not be supplied in the diet. For example- *alanine, arginine, tyrosine, asparagin, glutamine, proline,* etc. In plants all the amino acid are non-essential.

(c) **Semi-essential amino acids:** These amino acids are synthesized at a slower rate than their requirement in the body, e.g. **arginine** and **histidine**. Growing individuals require them in food. Arginine and histidine are essentially required in the diet of children, teenagers, pregnant women and lactation mother. But they are not essential for the adult individual.

TABLE-8.2: Essential Amino Acids.

1. Valine	6. Lysine
2. Isoleucine	7. Tryptophan
3. Phenylalanine	8. Threonine
4. Methionine	9. Histidine (Semi-essential)
5. Leucine	10. Arginine (Semi-essential)

TABLE-8.3: List of Non-essential Amino Acids.

1. Alanine	7. Glycine
2. Asparagine	8. Hydroxyproline
3. Aspartic acid	9. Hydroxy lysine
4. Cysteine	10. Proline
5. Glutamic acid	11. Serine
6. Glutamine	12. Tyrosine

(d) **Conditionally essential amino acids:** These amino acids are normally nonessential, but become essential during times of physiological stress or in case of moderate to chronic illness. In these conditions, a person may lose the ability to manufacture enough nonessential amino acids and thus require supplementation or to be taken in food. Some of the conditionally essential amino acids are: **glycine, arginine, cysteine, glutamic acid, tyrosine** and **proline.**

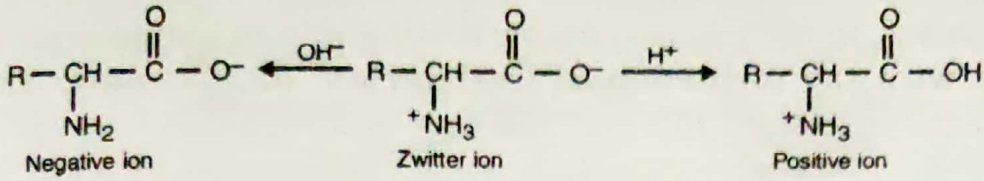
PROPERTIES OF AMINO ACIDS

A. PHYSICAL PROPERTIES

1. **Texture:** Amino acids are colourless, *crystalline solids* vary from slender needles (tyrosine) to thick hexagonal plates (cysteine).
2. **Solubility:** Most of the amino acids are usually soluble in water, dilute acids and bases. They are slightly soluble in alcohol and insoluble in organic solvents like ether.
3. **Melting Point:** Amino acids are generally melt at higher temperature (generally above 2000C) anfd often result in decomposition.
4. **Taste:** Amino acids may be **sweet** (Gly, Ala& Val), **tasteless** (Leu & Tyr) or **bitter** (Arg & Ile).
5. **Absorption of UV radiation:** Aromatic amino acids like tryptophan, tyrosine, histidine and phenylalanine absorb UV-light.
6. **Optical Properties:** All amino acids except glycine are *optically active* i.e. they can rotate the plane polarized light. As a result they exist in two optical isomers i.e. *l* **(-) forms** and or **d (+) forms**. Most naturally occurring amino acids are L-amino acids due to the presence of one asymmetric α-carbon atom. But **threonine** and **isoleucine** are the amino acids having two chiral (asymmetric) carbon atoms each. Hence, they exist in 4 optical isomers.

   D-amino acids are rare in nature, but are found in small peptides of bacterial cell walls and certain antibiotics. Recently two free D-amino acids found in mammals i.e. **D-serine** in fore brain and **D-aspartate** in brain and PNS.
7. **Zwitter ions or dipolar ions:**

The name zwitter is derived from the German word which means "hybrid". Zwitter ion (or) dipolar ion is a hybrid molecule containing positive & negatively ionic groups. In neutral solution (pH = 7.0), amino acids are predominantly zwitterions where the a-amino group (- NH_3^+) is protonated and the α-carboxylic group is dissociated (-COO). This is due to shifting of the proton from carboxyl group to amino group of the self-molecule at normal pH of cellular levels.



B. ELECTRICAL PROPERTIES

1. Amphoteric nature or ampholytes:

Amino acids (zwitterions) are **ampholytes** (amphoteric electrolytes) i.e. they act as both acid (proton donor) and base (proton acceptor). In acidic solution (e.g. pH = 1), only the α-amino group is ionized and the amino acids carry positive charge and migrate towards cathode. But in alkaline solution (e.g. pH = 11), only the α-carboxylic group is ionized, so the amino acids carry negative charges and move towards anode.

2. Isoelectric point or isoelectric pH : The pH at which an amino acid bears no net electric charge (i.e. exist in form of zwitter ion and there is no anion or cation) is called isoelectric point or pI. The pI can be calculated as the midpoint between pK_1 and pK_2 values.

$$pI = \frac{1}{2}(pK_1 + pK_2)$$

3. pK values: All amino acids possess at least two weakly acidic functional groups. However, -COOH is several thousand times stronger acid than $-NH_3^+$. The pK value is the pH at which these acidic groups are half dissociated. The pK of -COOH group and $-NH_3^+$ group are respectively **pK_1** and **pK_2**. In all the 20 standard amino acids, the value of pK_1 varies from 1.8 - 2.9 and the pK_2 value varies from 8.84 - 10.78.

C. CHEMICAL PROPERTIES

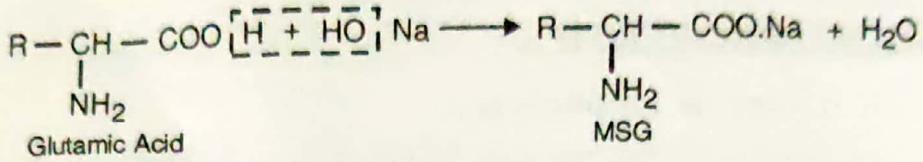
Amino acids contain three groups, viz., NH_2, COOH and R-group. Hence, they show three types of reactions. Of these, reactions of NH_2, COOH group are common to all amino acids, whereas the reactions of the R-group are specific and are different for different amino acids.

I Chemical properties of amino acids due to -COOH group:

(a) **Decarboxylation and amine formation:** The amino acids will undergo decarboxylation (loss CO_2) to form the corresponding "amines" when heated in presence of barium hydroxide. Thus important amines are produced from amino acids.

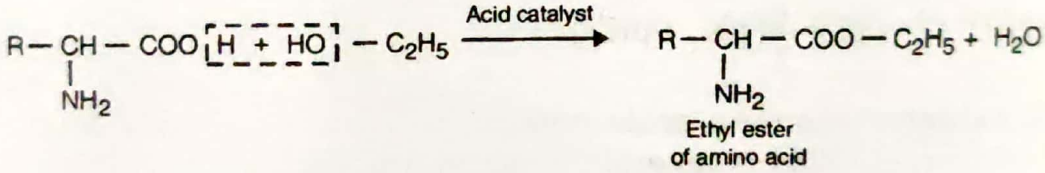
- Histidine → Histamine + CO_2
- Tyrosine → Tyramine + CO_2
- Tryptophan → Tryptamine + CO_2
- Lysine → Cadaverine + CO_2
- Glutamic acid → Gamma Amino Butyric Acid (GABA) + CO_2

(b) Reaction with Alkalies (Salt formation): Amino acids react with alkalies to form "Salts". In the process, the carboxyl group of amino acids can release a H^+ ion with the formation of Carboxylate (COO^-) ions. These may be neutralized by cations like Na^+ and Ca^{+2} to form Salts.

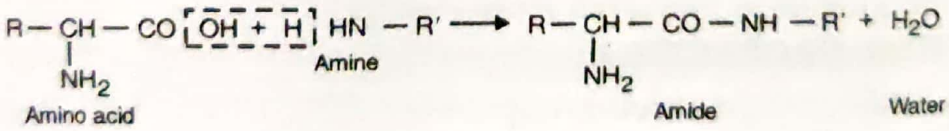


For example, glutamic acid reacts with NaOH to form **monosodium glutamate** (MSG), which is the sodium salt of glutamic acid, is one of the most abundant non-essential amino acids found naturally in tomatoes, cheese and other food. It is extensively used as a flavor enhancing agent in the preparation of Chinese food, soups, canned vegetables and processed meats.

(c) Reaction with Alcohols (Esterification): The -COOH groups of amino acids react with alcohols to form "Esters". The esters are volatile and can be separated easily and identified by gas chromatography.



(d) Reaction with Amines:

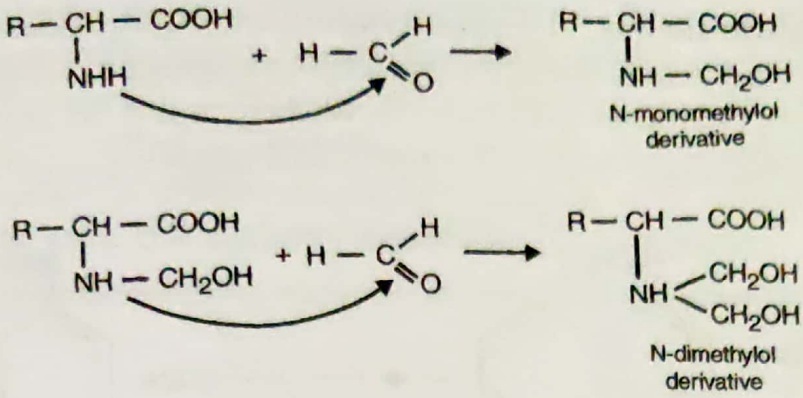


Amino acid reacts with Amines to form "Amides".

II. Chemical properties of amino acids due to -NH_2 group:

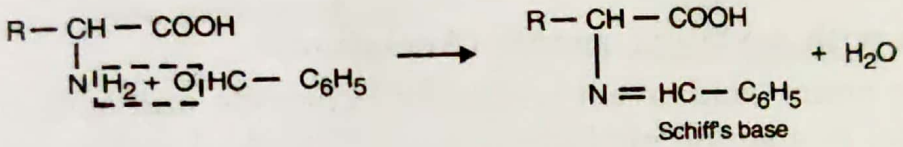
(a) Reaction with Mineral acids (Salt formation): When the amino acids are treated with mineral acids (like HCl), it forms "Acid Salts".

(b) Reaction with Formaldehyde: When the amino acid reacts with two molecules of Formaldehyde it forms "N-dimethylol derivative" (Hydroxy-methyl derivative). This reaction is done in two steps. These derivatives are insoluble in water and resistant to attack by microorganisms.



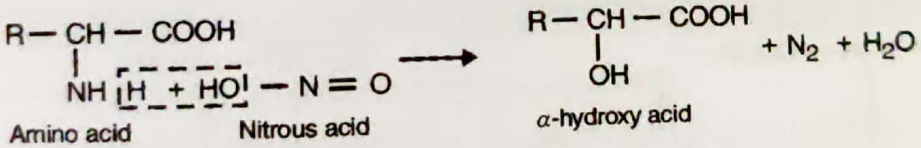
(c) Reaction with Benzaldehyde:

When the amino acid reacts with benzaldehyde, it gives "Schiff's base".



(d) Reaction with Nitrous acid (Van slyke reaction):

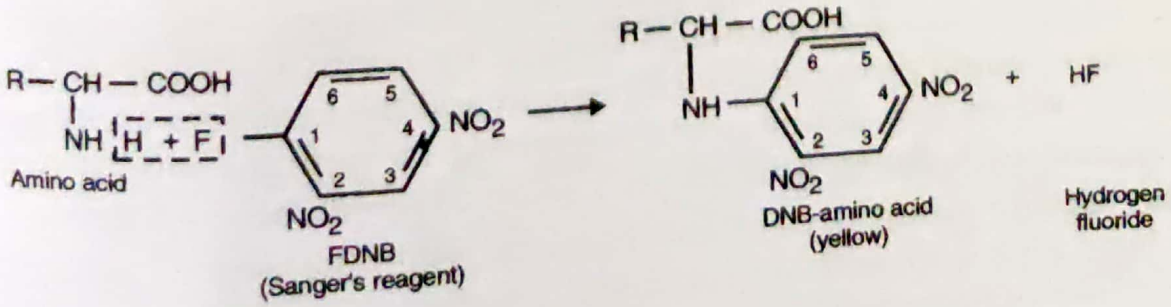
Amino acids react with nitrous acid (HNO_2) to produce the corresponding "α-hydroxyl acids" with the liberation of N_2. The imino acids proline and hydroxyproline do not respond to this reaction.



(e) Reaction with Sanger's reagent:

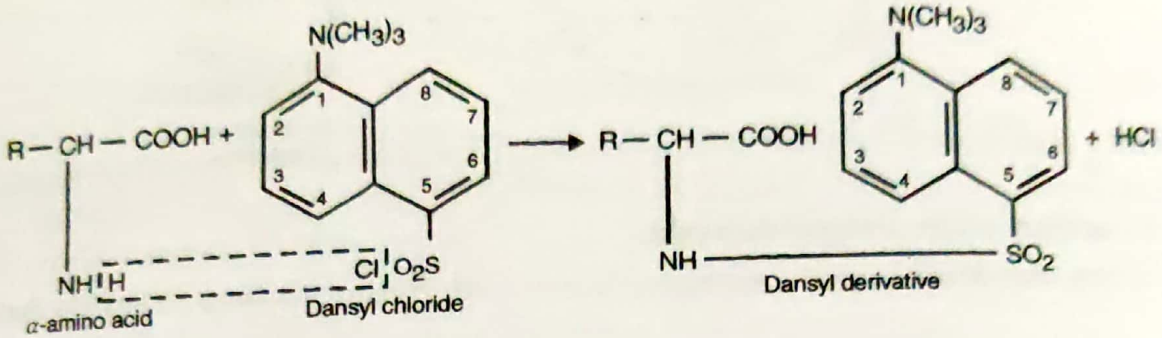
"1-flouro-2,4-dinitrobenzene" is called Sanger's reagent (FDNB). In mildly alkaline solution, -NH2 group of α-amino acids react with Sanger's reagent to produce Yellow colored derivative, **DNP-amino acid** (Dinitrophenyl amino acid).

With the help of this reaction, Frederick Sanger (1953) determined the amino acid sequence of insulin.



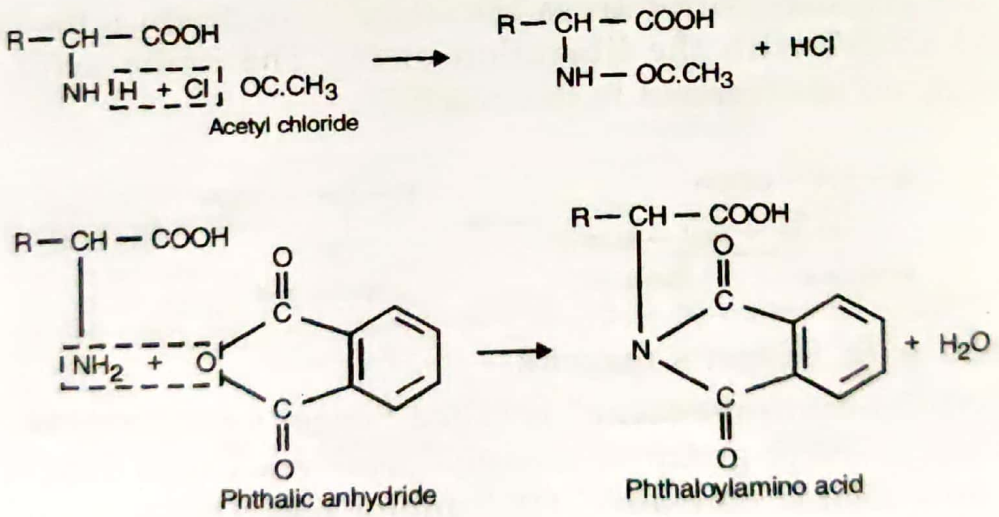
(f) Reaction with Dansyl Chloride:

Dansyl chloride means "1-Dimethylaminonapthaline-5-sulphonyl chloride". When the amino acid reacts with dansyl chloride reagent, it gives a "Flourescent Dansyll derivative".Dansyl chloride is widely used to modify amino acids; specifically, protein sequencing and amino acid analysis.



(g) Reaction with acylating agents (Acylation):

When the amino acids react with acetyl chloride and acid anhydride (Pthalic anhydride) in alkaline medium it gives "pthaloyl amino acid".



III. Reaction shared by both $-NH_2$ & -COOH groups:

a. Dehydration reaction (peptide bond formation) : The peptide or amide bond (CO-NH) is a covalent linkage formed by removal of one water molecule between the α-amino group of one amino acid and the α-carboxyl group of another amino acid. When two amino acids are joined by peptide bond it is known as dipeptide, and each amino acid unit is known as amino acid residue.

The peptide group is rigid and planner, while the peptide bond has partial double bond character with no rotational freedom. A **peptide** is a molecule that consists of two or more amino acid residues, e.g., **dipeptide** (2 amino acid residues),

tripeptide (3 amino acid residue) etc. A **polypeptide** usually contains 20-4000 residues. The **backbone** of the polypeptide includes the repeating sequence of α-carbon atoms and peptide bonds while the **side chains** of different amino acids responsible for functioning of the protein. A polypeptide with more than 100 residues is generally called a **protein**. A protein may consist of one or more polypeptide. A peptide is a polar molecule starting with the amino terminus (N-terminus) and ends with the **carboxy terminus** (C-terminus).

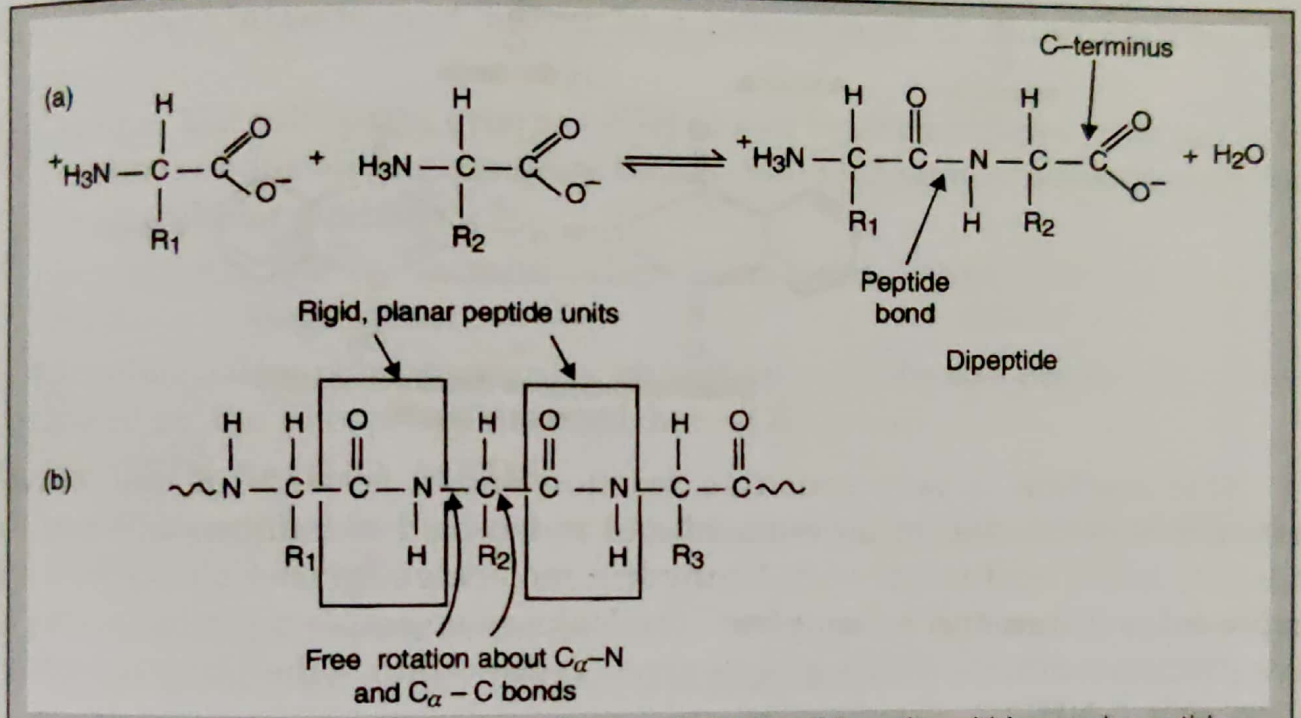
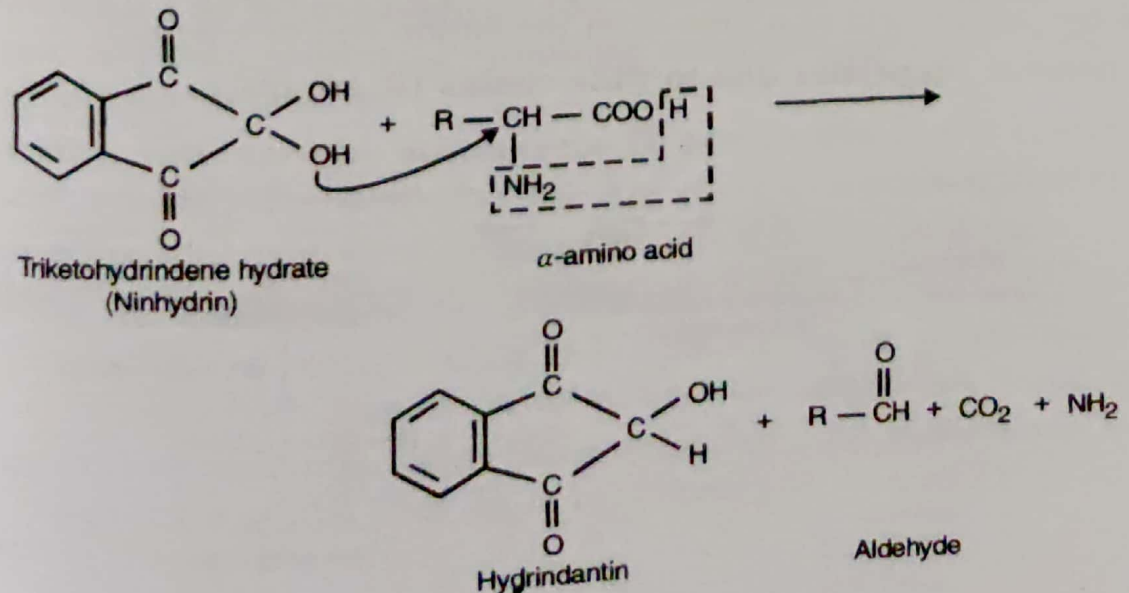


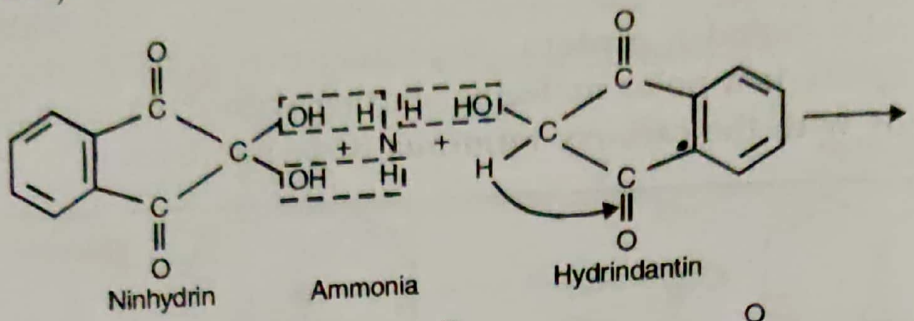
Fig. 8.5 (a) Formation of a peptide bond, (b) Peptide units within a polypeptide

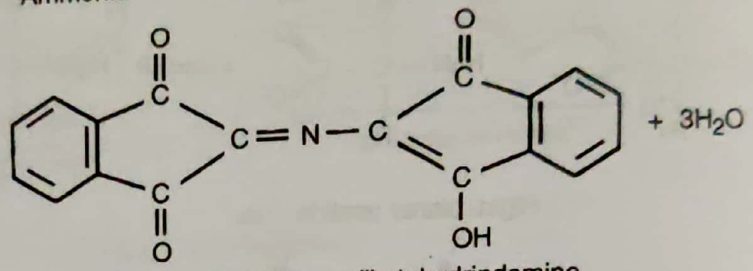
b. Ninhydrin reaction (Oxidative deamination):

Step1: Ninhydrin (= triketohydrindene hydrate) is a powerful oxidizing agent and causes oxidative decarboxylation of α-amino acids producing CO_2, NH_3 and an aldehyde with one less carbon atom than the parent amino acid.



Step2 : The reduced ninhydrin **(hydrindantin)** then reacts with the liberated NH_3 and a mole of ninhydrin, forming Blue-colored Ruheumann's complex (diketohydrin).





Diketohydrindylidene-diketohydrindamine
(Ruheman's purple)

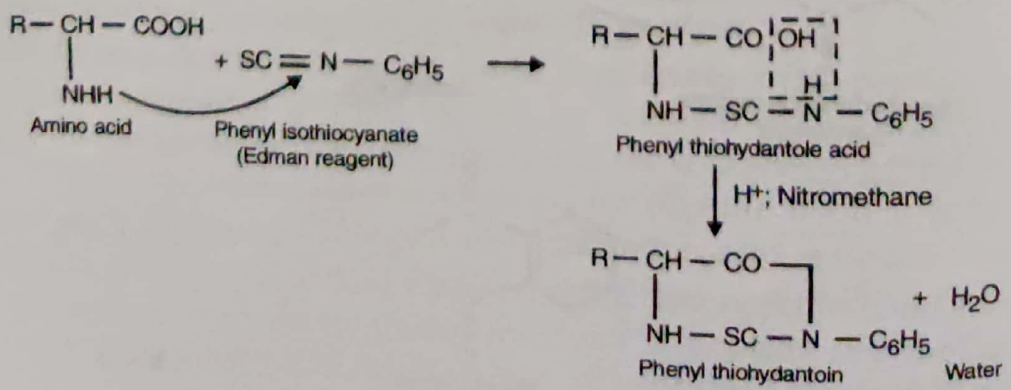
This reaction is very sensitive reaction and it is used for qualitative and quantitative estimation of amino acid and imino acid identification.When Amino acids (or) Imino acid reacts with Ninhydrin molecule it gives Color. When it gives **Purple color (Rhumann's Complex)** -the Unknown sample is Amino acids (Which have primary amine - (NH_2) or it is gives Yellow color - the Unknown sample is Imino acid **(-NH-).**

c. Reaction with Edman's reagent :

Edman's reagent is **"phenylisothiocyanate"**. When amino acids react with Edman's reagent it gives *"phenyl thiohydantoic acid"* finally it turns into a *"Phenyl thiohydantoin (PTH)-amino acid"* (Edman's derivative).

Pahr Edman devised this technique of **Edman degradation** to sequentially remove one amino acid at a time from the amino end of a peptide.

IV. Chemical properties due to Side chains (R groups)



DERIVATIVES OF STANDARD AMINO ACIDS

Some polypeptides chain unusual or rare amino acid residues which are derived by the post transcriptional modification from one of the 20 amino acids. The rare amino acids are not coded by DNA. Some of them are as follows:

1. **N-formylmethionine.** It is the 1st amino acid of prokaryotic protein.
2. **4-hydroxyproline** and **5-hydroxylysine** found in collagen fibers in connective tissue.
3. **γ-carboxyglutamic acid** found as a constituent of blood clotting factor prothrombin.
4. **Cystine.** The sulfhydryl groups (-SH) of two cysteine resides oxidized to form a cystine residue having disulfide bond (-S-S-). Cystine residues mostly found in extracellular proteins.
5. **Acetylation:** The N- termini when acetylated, makes the proteins more resistance to degradation.
6. **Phosphorylation:** Phsophoserine, phosphothreionine and phosphotyrosine are formed by the phosphorylation of their -OH group.

NON-PROTEIN AMINO ACIDS

Over 150 modified amino acids are known that are not components of proteins. A few examples are as follows:

(a) **D-amino acids** present in the peptides of bacterial cell wall are defensive in nature as they are not hydrolyzed by peptidases.

(b) **Ornithine** and **citrulline**, participate in Urea cycle.

(c) **γ-aminobutyric acid** (GABA) is a neurotransmitter derived from decarboxylation of glutamate.

(d) **Dopamine** is also a neurotransmitter derived from tyrosine.

(e) **Thyroxin,** a thyroid hormone, is a product of tyrosine.

(f) **Histamine** is a decarboxylation product of histidine. It is a vasodilator, causes allergic response, powerful stimulant of gastric secretion and constrict bronchial smooth muscles.

(g) **Azaserine, valinomycin, atinomycin D, gramicidin S** and **tyrocidine** are the useful components of antibiotics. Azaserine acts as an anti-cancer drugs.

(h) **Cycloserine,** a derivative of serine is an anti-tuberculous drug.

BIOLOGICAL SIGNIFICANCE OF AMINO ACIDS

(1) **Building blocks of** proteins: Only L- L-amino acids are polymerized to form proteins, though both D-amino acids and non-L-amino acids found in nature.

(2) **Biological buffers:** Amino acids being amphoteric, act as buffers in solutions, resisting changes in pH. They do so by donating H^+ ions as pH increases and accepting H^+ as pH decreases.

3) **Nitrogen storage:** Asparagine and glutamine are amide derivatives of aspartic acid and glutamic acid. They serve as storage of nitrogen.
4) **Proline** forms bands or kinks in polypeptide chains.
5) **Cysteine** links chains together by forming disulfide bonds.
6) **Histidine** found in the active site of enzymes where it causes making and breaking of bonds.
7) **Thearomatic rings** of Phe, Tyr and Trp help in electron transfer.
8) **Formation of glucose:** Some amino acids form glucose by losing amino group.
9) **Formation other compounds:**

(i) **tyrosine** produces the hormones thyroxin and adrenaline and the skin pigment melanin, **(ii) glycine** forms heme, and

(iii) **tryptophan** produces vitamin nicotinamide and plant hormone Indole Acetic acid (IAA). The coenzyme A, a vitamin pantothenic acid, coenzyme glutathione and alkaloids are some other compounds formed by amino acids.

10) **Antibiotics:** The non-protein amino acids are useful compounds of antibiotics e.g. Azaserine, valinomycin etc.
11) **Genetic defects:** Inborn errors in the metabolism of amino acids cause several disorders e.g. Phenylketonuria.
12) **L-amino acids** and their derivatives help in nerve transmission, cell growth and biosynthesis of porphynine, purines, pyrimidines and Urea.

MODEL QUESTIONS

I. Very Short Answer Type Questions (1 Mark)

1. **How many amino acids universally occur in the proteins of all forms of life?**
Ans. 20 standard amino acids.

2. **Name the first and the last discovered amino acids out of 20 standard ones.**
Ans.Asparagine – the first amino acid discovered; Threonine – the last amino acid discovered.

3. **Name the two newly discovered amino acids.**
Ans. Selenocysteine and pyrrolysine are two newly discovered amino acids that are cotranslationally inserted into proteins and known as 21st and 22nd amino acids in the genetic code.

4. **Which amino acids can absorb UV light?**
Ans. Tryptophan and tyrosine, and to a very lesser extent phenylalanine, absorb UV light.

5. **Name two sulfur containing amino acids.**
Ans. Cysteine (with sulfhydryl group) and methionine (with thioether group) are the two sulfur containing amino acids.

CHAPTER 9

CHEMISTRY OF PROTEINS

CHAPTER AT A GLANCE

After reading this chapter the learner would be able to understand and appreciate the followings:

- Protein and polypeptide chain
- Biological significance of proteins
- Protein structure
- Bonds stabilizing protein structure
- Levels of organization in proteins
- Ramachandran Plot
- Denaturation and renaturation of proteins
- Classification of proteins
- Simple and Conjugate proteins
- Properties of proteins

John J. Berzelius (1838) first coined the term 'protein' (Gr. proteios = of the first rank or primary) to stress the importance of this class of polymers. Proteins are the most abundant (10%-15%) organic constituents of a living cell. **Collagen** is the most abundant protein in the animal world and **RuBisCO** is the most abundant protein in the whole biosphere.

BIOLOGICAL SIGNIFICANCE OF PROTEINS

With respect to their biological role, proteins are divided into several groups. They are structural proteins, enzymes, hormones, transporters, defence proteins, receptors, contractile proteins, storage protens etc. (See Table 9.1)

STRUCTURE OF PROTEINS

Proteins are the macromolecules composed of one or more polypeptide chains, each of which is a mixed polymer of a-amino acid residues joined end-to-end by peptide bonds.

Table 9.1 : Types of Proteins, Examples of Functions.

Type	Examples	Occurrence / Function
Structural	Collagen	Component of connective tissues, bone, tendors, cartilage
	Keratin	Skin, feathers, nails, hair, horn.
	Elastin	Elastic connective tissue (ligaments).
	Viral coat proteins	'Wraps up' nuclei acid of virus
Enzymes	Trypsin	Catalyses hydrolysis of protein
	Ribulose bisphosphate carboxylase	Catalyses carboxylation (addition of CO_2) of ribulose bisphosphate photosynthesis.
	Glutamine synthestase	Catalyses synthesis of the amino acid glutamine from glutamic acid ammonia
Hormones	Insulin	Help to regulate glucose metaboolism.
	Glucagon ACTH	Stimulates growth and activity of the adrenal cortex.
Respiration Pigment	Haemoglobin	Transports O_2 in vertebrate blood
	Myoglobin	Stores O_2 in muscles.
Transport	Serum albumin	Transport of fatty acids and lipids in blood.
Protective	Antibodies	Form complexes with foreign proteins.
	Fibrinogen	Forms fibrin in blood clotting.
	Thrombin	Involved in blood clotting mechanism.
Contractile	Myosin	Moving filaments in myefibrils of muscle.
	Actin	Stationary filaments in myefibrils of muscle.
Storage	Ovalbumin	Egg white protein
	Casein	Milk protein
Toxins	Snake vesom	Enzymes
	Diptheria toxin	Toxin made by diptheria bacteria.

On the basis of number of polypeptide chain, proteins can be monomeric or multimeric.

(i) **Monomeric** protein consists of single polypeptide chain, e.g., lysozyme, myoglobin.

(ii) The **oligomeric** or **multimeric** protein consists of 2 or more polypeptide chains, each of which is called a protomere or subunit. For example, Rubisco consists of 24 polypeptides, hemoglobin (Hb) is a tetrameric consists of two α-chains and two ß- chains, immunoglobulins consists of 2 L-chains and 2 H-chains etc.

Each polypeptide is a polymer of amino acids. The left end of the polypeptide has first amino acid with free α-amino group which is referred to as **N-terminal (amino terminal) end** and the right end has last amino acid with free a-carboxylic group which is called as **C-terminal (carboxyl terminal) end.**

A polypeptide chain is synthesized on the ribosome as a linear sequence of amino acids. Just after the synthesis, the newly synthesized (nascent) polypeptide folds into a specific three dimensional shape called conformation. The conformation adopted by polypeptide to perform the biological activity is called **native conformation.**

Previously, it was thought that proteins fold spontaneously to attain their native states. Recent studies revealed that **chaperone proteins** accelerate the folding process of nascent polypeptides into their native conformations. The deficiencies of chaperone proteins cause diseases due to incorrect folding of proteins. For example, in Alzheimer disease the amyloid plaques develop due to protein clumping in brain cells.

BONDS STABILIZING PROTEIN STRUCTURE

Proteins are the polymers of amino acids. Amino acids are joined together by special type of covalent bond (peptide bond) to form linear structures called polypeptides. The polypeptides are then folded into specific structures to form the functional conformation of the protein. The folding of proteins into specific shapes and conformations are assisted and stabilized by many types of bonds in them. Some of them are strong bonds whereas others are weak interactions. Important types of bonds involved in stabilizing protein structure and conformation are as follows:

A. Covalent bonds

1. Peptide bonds
2. Disulfide bonds
3. Desmosine bridges

B. Non-covalent bonds

1. Hydrogen bonds
2. Ionic or electrostatic bonds or salt bridges
3. Vander Walls forces(= London dispersion forces)
4. Hydrophobic interactions

1. Peptide bond (= amide bond). The primary structure of protein is stabilized by peptide bonds. **Peptide bond** or **amide bond** (CO-NH) is a covalent bond formed by removal of one water molecule between the carboxylic group of one amino acid and the amino group of another amino acid. When two amino acids are joined by peptide bond it is called dipeptide, and each amino acid unit is known as amino acid residue.

Peptide bond is a strong covalent bond with high bond dissociation energy. Peptide bond formation is catalyzed by the enzyme peptidyltransferase during the translation process of protein synthesis. Peptidyltransferase is a ribozyme; it is a part of the ribosomal RNA (rRNA) of large subunit of ribosomes. In prokaryotes the 23S rRNA and in eukaryotes the 28S rRNA acts as the peptidyltransferase enzyme.

The peptide group (CO-NH) is rigid and planner. In the late 1930s, Linus Pauling and Robert Corey from their X-ray crystallography studies of several amino acids and dipeptides suggested that the peptide group has a rigid and planner structure, and the peptide bond (carbon-nitrogen bond) has partial (~40%) double bond character.

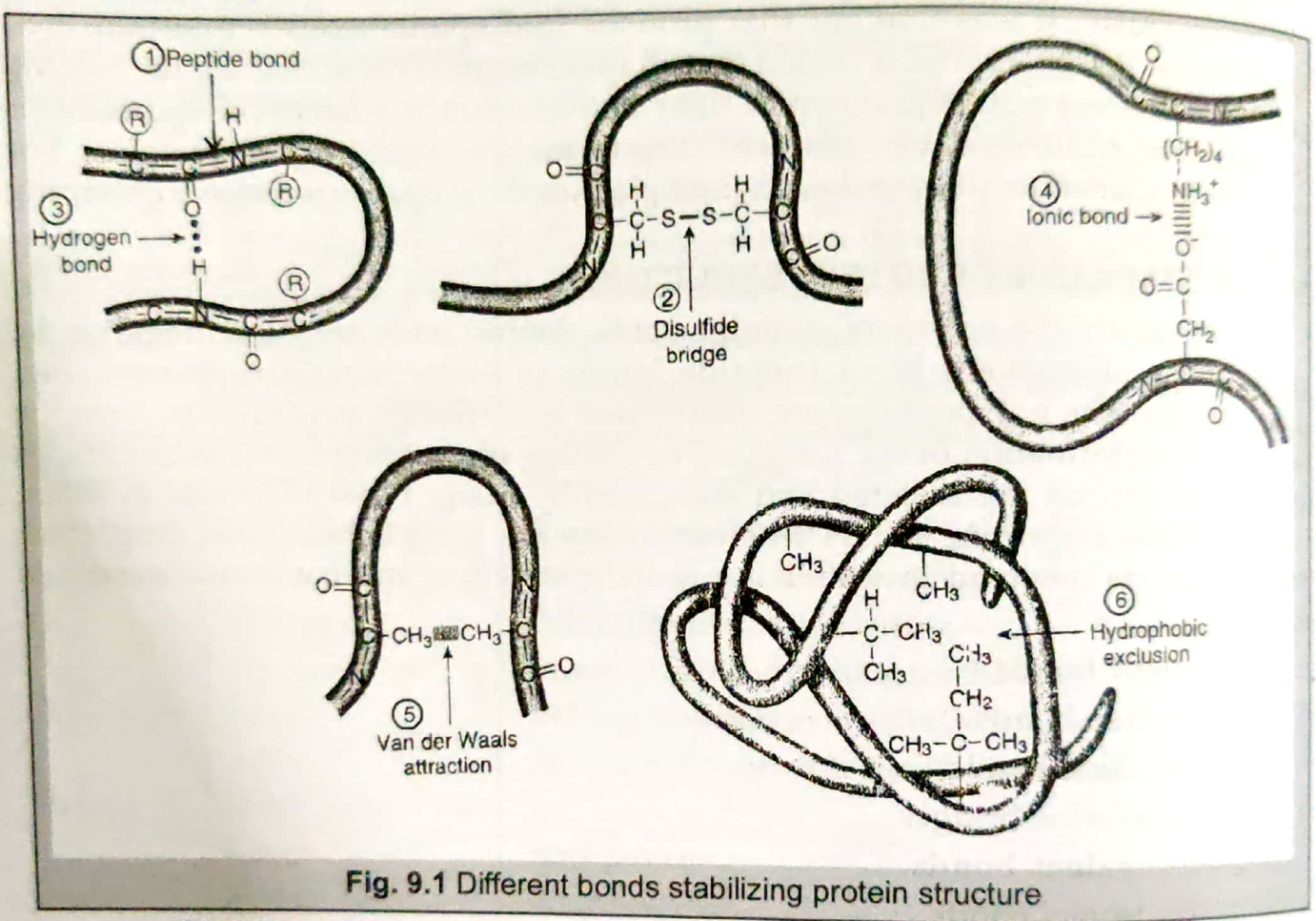


Fig. 9.1 Different bonds stabilizing protein structure

TABLE - 9.2 Various bonds contributing to protein structure.

Bonds in Proteins

Primary structure	Secondary structure	Tertiary structure	Quaternary structure
Covalent bonds 1. Peptide bond 2. Di-sulphide bond	1. Hydrogen bond 2. Peptide bond 3. Di-sulphide bond	1. Hydrophobic bonds 2. Hydrogen bonds 3.Ionic bonds 4. Vander Walls force 5. Di-sulphide bonds	1. Hydrophobic bonds 2. Hydrogen bonds 3.Ionic bonds 4. Vander Walls force 5. Di-sulphide bonds

2. Disulfide bonds. Some proteins contain disulfide bonds. **Disulfide bond** is a covalent bond formed from two sulfhydryl groups (-SH) of two cysteine residues in a protein. The resulting disulfide is called **cystine** residue. The cysteine (Cys or C, a sulfur containing amino acid) contain a highly reactive sulfhydryl group (-SH) in its side chain (R group). The sulfhydryl is highly polar and highly reactive. If two molecules of cysteine line up alongside each other, the neighboringsulfhydryl groups can be oxidized. This reaction results in the formation of a permanent covalent connection between two cysteine residues called disulfide bond(-S-S-).

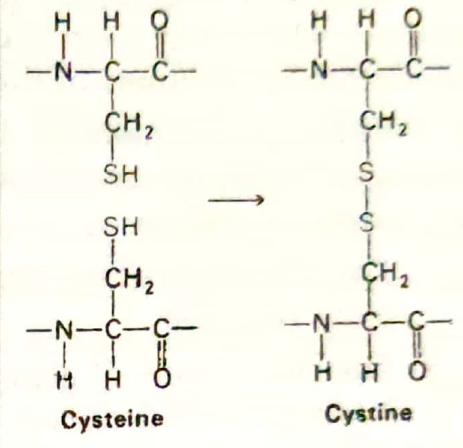


Fig. 9.2 Formation of a disulfide bond (-S-S) from the sulfhydryl groups (-SH) of two cysteine residues, the product is a cystine residue.

A disulfide bond may be formed between the cysteine residues of the same polypeptide chain or different polypeptide chain of a functional protein. Intracellular proteins usually lack disulfide bonds, whereas extracellular proteins often contain several. Disulfide bonds stabilize the tertiary structures of the protein.

3. Desmosine bridges. Desmosine cross link is formed by four lysine groups combine around the pyridinium nucleus with exposed carboxyl groups. It is found in elastin; they are formed of lysyl oxidase.

4. Hydrogen bonds. A hydrogen atom normally forms a covalent bond with only one atom at a time. However, a covalently bonded hydrogen atom form an additional bond called hydrogen bond. **Hydrogen bond** is a weak electrostatic attraction between an electronegative atom (the acceptor atom) and a hydrogen atom covalently bonded to another atom (the donoratom such as Oxygen, Nitrogen or Fluorine) of same or different molecules of their close vicinity. The hydrogen atom is closer to the donor (D) than to the acceptor (A):

$$D - H + A \rightleftharpoons D - H \cdots A$$

The covalent bond between the donor and the hydrogen atom must be dipolar, and due to high electronegativity, donor atom (O and N) attract the shared electron of hydrogen more towards it. Thus the hydrogen atom attachto donor atom gets a partial positive charge called δ^+ whereas the donor atom will get a partial negative charge called δ^-. Consequently, the neighboring acceptor atom (oxygen of –C = O or nitrogen atom of $-NH_2$ group) having nonbonding electrons attract the δ^+ charge of hydrogen atom.

The entire length of polypeptide chain contains CO and NH groups. The CO **group of each amino acid is hydrogen bonded to the NH group of the amino acid that is situated four residues ahead in the linear sequence** (Fig.9.8). Thus, all the main chain CO and NH groups are hydrogen bonded. The secondary structures (alpha helix and beta plates) are stabilized by the formation of hydrogen bonds.

Residues with polar R groups generally occur on the surface of globular proteins, where they form hydrogen bonds primarily to water molecules. Elsewhere, the aminoacyl residues of the backbone form hydrogen bonds with one another.

In fact, side chains of 11 of the 20 basic amino acids also can participate in hydrogen bonding. According to their hydrogen bonding potentialities, they can be grouped as follows:

- The side chain of tryptophan and arginine can serve as hydrogen bond donors only. (Fig.9.3)
- The side chains of asparagine, glutamine, serine, and threonine can serve as hydrogen-bond donors and acceptors.
- The hydrogen-bonding capabilities of lysine, aspartic and glutamic acids, tyrosine and histidine vary with pH. These groups can serve as both acceptors and donors over a certain range of pH, and as acceptors or donors (but not both) at other pH values. (Fig.9.3).

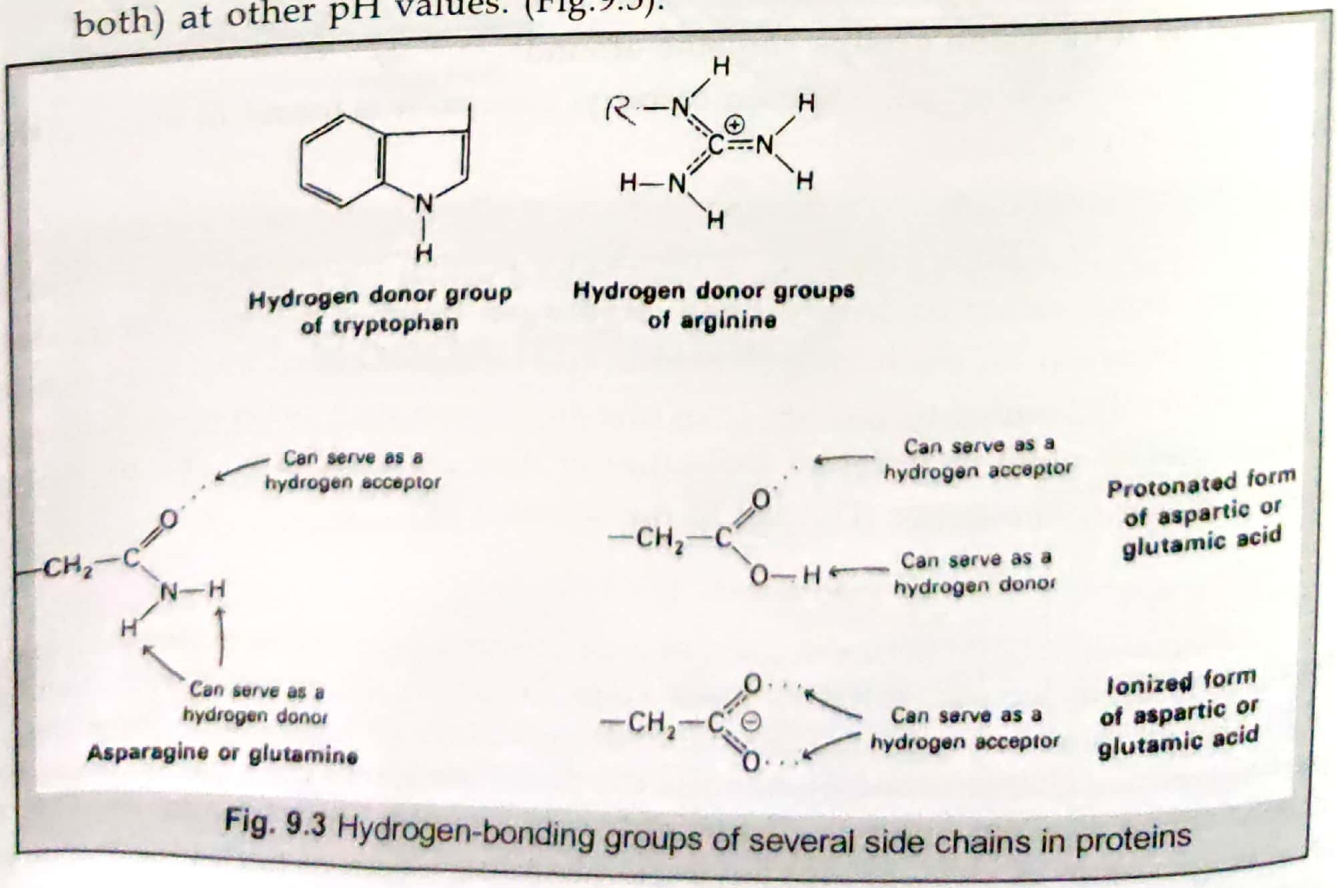


Fig. 9.3 Hydrogen-bonding groups of several side chains in proteins

Hydrogen bonds are very weak bonds. Hydrogen bonds have only 1/20 the strength of a covalent bond. Hydrogen bonds are strongest, when the three atoms (the donor, the hydrogen atom and the acceptor) all lie in a straight line. Occurrence of hydrogen bonds in high frequency makes a considerable contribution towards the molecular stability of proteins. **Hydrogen bonds are involved in stabilizing the secondary, tertiary and quaternary structures of proteins.**

5. Ionic bonds or electrostatic interactions or salt bonds. It is a chemical bond formed between the 2 ions of opposite charges. In proteins, ionic bonds are formed between the ionized acidic or basic groups of amino acids. The R groups of certain amino acids contain additional acidic (-COO-) or basic ($-NH_3^+$) groups. These R groups can ionize to produce charged groups at certain pH. Acidic R groups will be negatively chargedsince they release H^+ ions. Basic R groups will be +vely charged since they accept the H^+ ions from the medium. After the ionization of the side chain (R groups) as mentioned above, the amino acids in the protein chain can attract or repel each other based on their charges. The attraction of oppositely charged R groups results in the formation of ionic bonds. Since polar residues also participate in ionic interactions, the presence of salts such as KCl can significantly decrease ionic interactions between surface residues.

Ionic bonds are weak bonds and they are very fragile in an aqueous medium. Even a change in pH may breakdown the ionic bonds. This is the reason for the denaturation of proteins in the acidic or basic medium. **Tertiary and quaternary structures of proteins are stabilized by ionic bonds.**

6. Van der Waals interactions (= London dispersion forces). These are weak forces or attractions between atoms due to oppositely polarized electron clouds. When two atoms approach one another closely, they create a nonspecific, weak attractive force that produces a van der Waals interaction, named after a Dutch physicist **Johannes Dideric van der Waals** (1837-1923). All types of molecules, both polar and nonpolar, exhibit **van der Walls interaction**. Attraction decreases rapidly with increasing distance and is effective only when atoms are quite close to one another. However, if atoms get too close together that their electron orbitals overlap, they become repelled by the negative charges in their outer electron shells.

The distance at which attractive force is maximal and the repulsive force is minimal is termed the van der Waals contact distance. Each type of atom has a

characteristic van der Waals radius at which it is in van der Waals contact with other atoms. The optimum contact distance between the two atoms is the sum of their van der Waals radii.

7. Hydrophobic interactions. The force that causes the hydrophobic molecules or non-polar portions of molecules to aggregate rather than to dissolve in water is called the hydrophobic bond or, more precisely, the **hydrophobic interaction.** This is not a separate bonding force; rather, it is the result of the energy required to insert a non-polar molecule into water. A nonpolar molecule cannot form hydrogen bonds with water, so it distorts the usual water structure, forcing water to make a cage of bonds around it, but not with it. Conversely, nonpolar molecules bond together comfortably through van der Walls interactions. The result is a very powerful tendency for hydrophobic molecules to bond together and not dissolve in water.

Some R groups in the amino acids are nonpolar, e.g. alanine, valineleucine, isoleucine and methionine. The nonpolar R groups are hydrophobic and they try to stay away from water. In a long polypeptide chain, there may be many such nonpolar amino acids which may be adjacent to each other or separated by polar R groups. In an aqueous environment (inside the cell) the linear polypeptide will fold into such a shape that the hydrophobic amino acids come in close contact with each other and tend to orient towards the inner side of the protein and they try to exclude the water due to its hydrophobicity. By this method, the peptide chain of a globular protein will fold into a spherical shape in the aqueous environment. The hydrophilic residues form a shell over the hydrophobic moieties. The hydrophilic shell makes the protein soluble in aqueous medium.

Although hydrophobic interactions do not form true bonds they are of great importance in the formation of tertiary and quaternary structures.

LEVELS OF PROTEIN STRUCTURE

Variation in number and order of amino acids sequence reflect the structure of a protein. There are four levels of protein structural organization of proteins. They are: Primary, Secondary, Tertiary and Quaternary levels of protein structure. Recent studies revealed two additional levels of protein organization i.e. Super-secondary structures or motifs and domains.

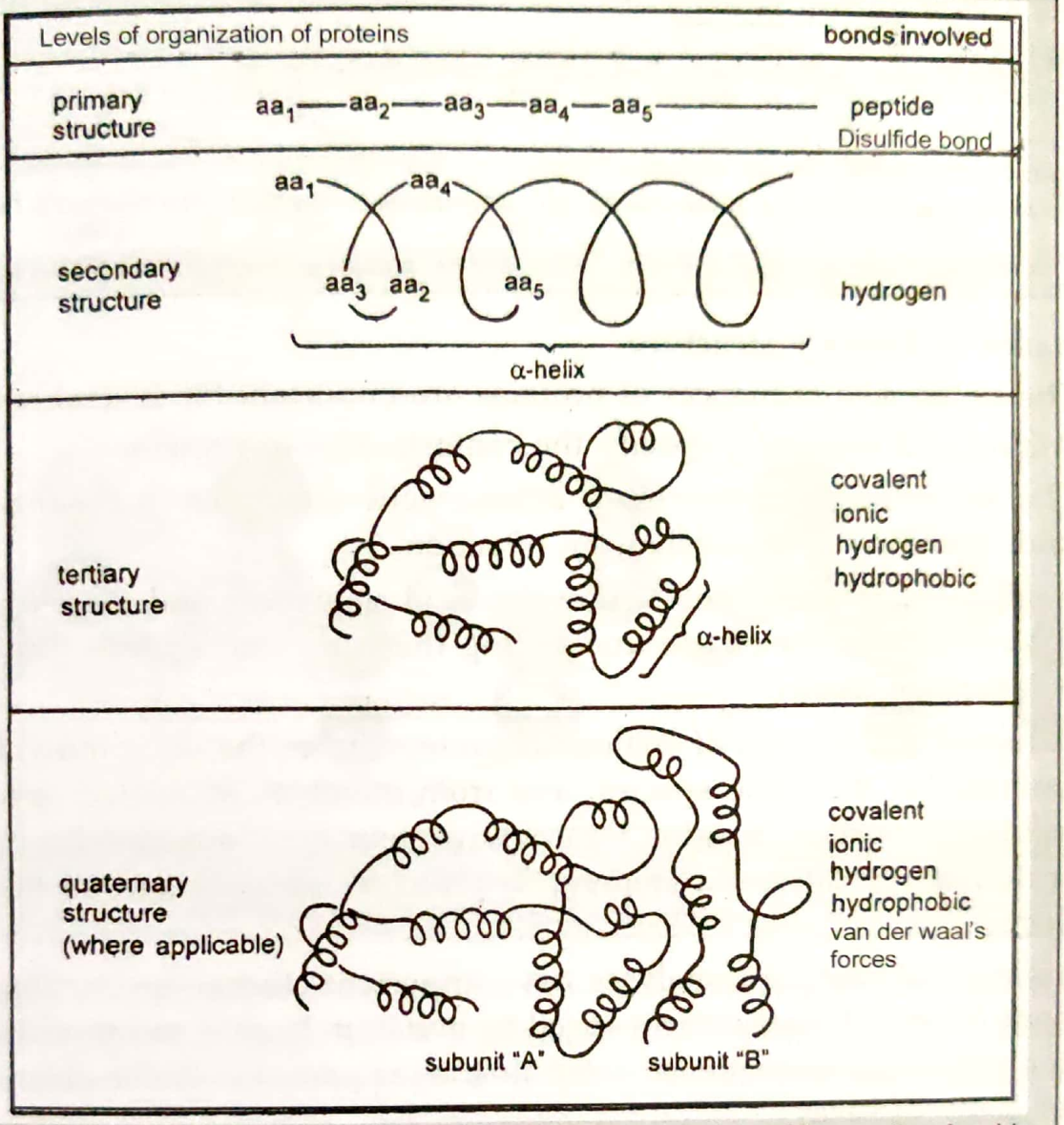


Fig. 9.4 Diagrammatic representation of the four levels of forces involved in generating protein conformation.

A. Primary Structure (1° structure):

Primary structure of a protein refers to the **sequence of amino acid residues** of its polypeptide chain(s) which read in N-terminus → C- terminus direction. It is the 1st level of organization of protein determined by the codons of mRNA or cistron of DNA. The 10 structure is stabilized by the **peptide bonds** as well as and **disulfide bonds** between cysteine residues, if there are any.

Frederick Sanger (1953) first determined the 1° structure of bovine insulin. Insulin consists of two chains, α- chain (21 amino acids long) and ß-chain (30 amino acids long), joined together by two disulfide bonds (Fig.9.5). Now, the 1° structure of a polypeptide is determined by an automated device called **spinning cup sequenator**, developed by PehrEdman and Geoffrey Begg.

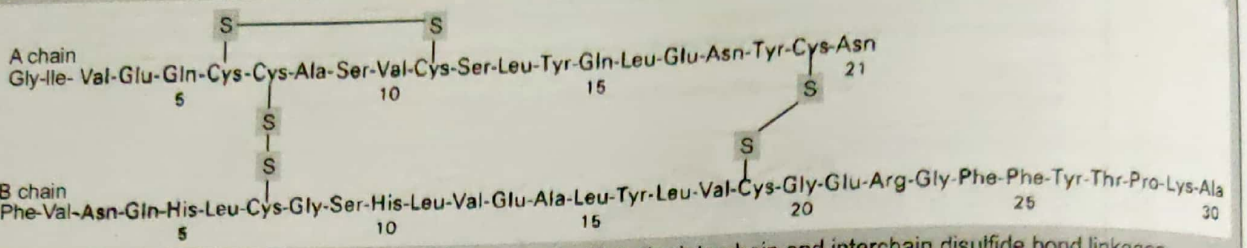


Fig. 9.5 The primary structure of bovine insulin. Note the intrachain and interchain disulfide bond linkages

Importance of Primary structure

The amino acid sequences of proteins are important for several reasons:

1. Amino acid sequences specify the conformation of proteins.
2. The knowledge of protein's amino acid sequence is essential for an understanding of its mechanism of action.
3. Analysis of relations between amino acid sequences and three-dimensional structures of proteins are uncovering the rules that govern the folding of polypeptide chains.
4. Sequence comparison of analogous proteins from the same individual, from members of the same species, and from members of related species, have yielded important insights into how protein functions and have indicated evolutionary relationships among the proteins and the organisms that produce them.
5. Amino acid sequence analyses have important clinical applications because many inherited diseases are caused by mutation leading to amino acid change in a protein. Thus sequence determination is part of valuable diagnostic tests for many diseases.

Prediction of conformation from amino acid sequence

In late 1930s, Pauling and Corey demonstrated that the α-carbon of adjacent amino acid residues are separated by 3 covalent bonds, arranged as Cα - C - N - Cα. They also demonstrated that the peptide C - N bond is somewhat shorter (1.32 Å or 0.132 nm), which is between C - N single bond (1.42 Å) and a C = N double bond (1.27 Å). This indicated a **resonance** or partial sharing of two pair of electrons between the carbonyl oxygen and amide nitrogen (Fig. 9.6). The four atoms of the peptide groups (C, O, N, H) are lie in a single plane, in such a way that the oxygen atom of carbonyl group and the hydrogen atom of the amino group are always trans (opposite) to each other. From these findings Pauling and Corey concluded that the **peptide C -N bonds (ω-bonds) cannot rotate freely because of their partial double-bond character.** In contrast, the rotation is permitted only about the C_α - C bonds **(ω-bonds)** and N - Cα bonds **(ϕ-bonds)** which are pure single

bonds. The backbone of a polypeptide chain can thus be pictured as a series of rigid planes, with consecutive planes sharing a common point of rotation at Ca(Fig.9.6).The rigid peptide bonds restrict the range of possible conformations for a polypeptide chain.

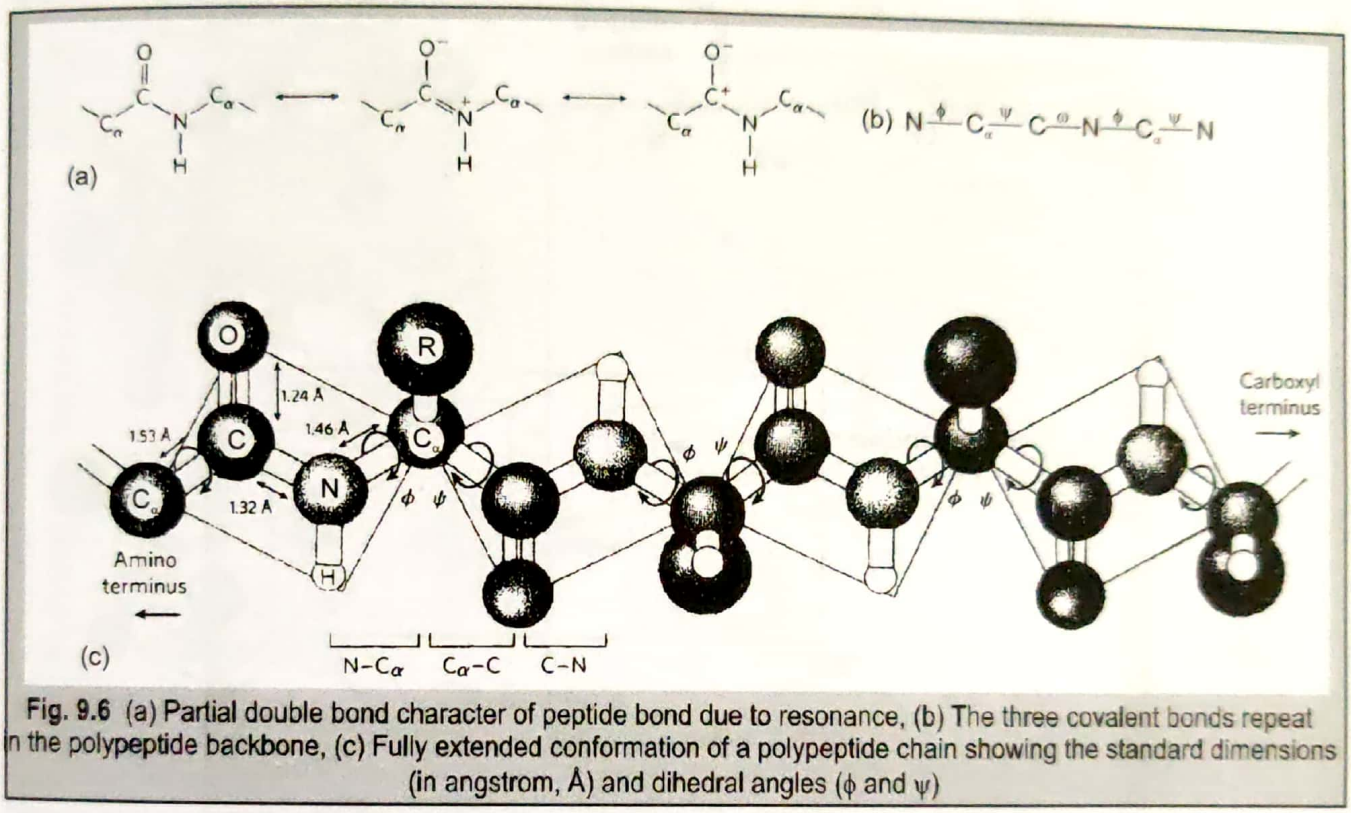


Fig. 9.6 (a) Partial double bond character of peptide bond due to resonance, (b) The three covalent bonds repeat in the polypeptide backbone, (c) Fully extended conformation of a polypeptide chain showing the standard dimensions (in angstrom, Å) and dihedral angles (ϕ and ψ)

Peptide conformation is defined by three dihedral angles (also called torsion angles) called **ϕ(phi), ψ (psi), and ω (omega),** reflecting rotation about each of the three repeating bonds in the peptide back bone. A dihedral angle is the angle at the intersection of two planes. The phi angle (ϕ) refers to rotations about the N - Casingle bonds of the main chain; psi angle (ω) refers to rotations about the Cα- C single bonds. In a fully stretched-out polypeptide chain, ϕ = ω =180°. In principle, ϕ and ψ can have any value between -180° and +180°, but many values are prohibited by steric interference between atoms in the polypeptide backbone and amino acid side chains. The conformation in which both ϕ and ψ are 0° is prohibited for this reason.

Ramachandran Plot

When all phi and psi angles are specified, the conformation of the main chain or backbone of a polypeptide can be completely defined. **G.N. Ramachandran** (1963) recognized that a residue in a polypeptide chain cannot have any pair of values of ϕ and ψ. Due to steric hindrance, certain combinations are not possible. The possible allowed ranges of ϕ and ψ angles can be predicted and visualized when ψ angle is plotted versus ϕ anglewhich iscommonly known as **Ramachandran**

plot. Ramachandran plot illustrate the combinations of phi and psi dihedral angles that are permitted in a peptide back bone and those that are not permitted due to steric constraints.

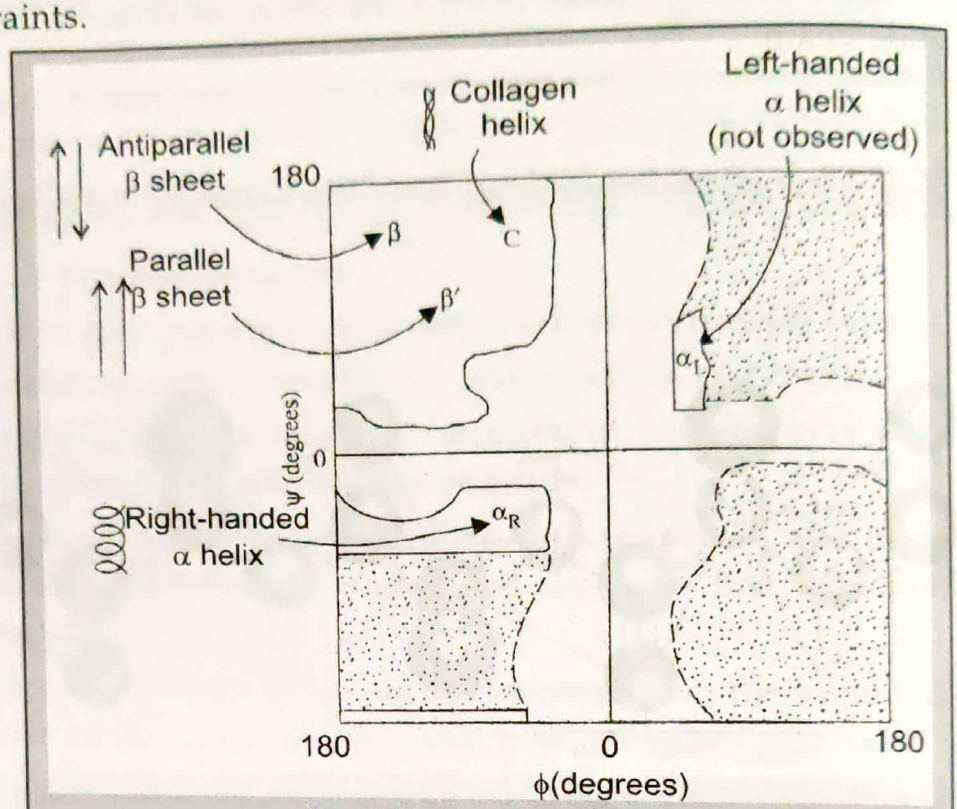
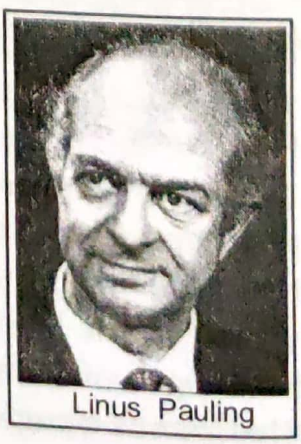


Fig. 9.7 A Ramachandran plot.
The screened regions show allowed values of f and y for L-alamine resideus. Additional conformation accessible to glycine (dotted regions) because it has a very small side chain.

B. Secondary (2°) structure:

A protein's **secondary structure (2° structure)** refers to the local spatial arrangement of backbone atoms in a selected segment of polypeptide chain without considering the conformations of side chains or its relationship to other segments. The type of 20 structure of a polypeptide depends upon its amino acid composition. In 1951, **Linus Pauling** and **Robert Corey** proposed two common types of secondary structures called,a-helix (alpha helix) and ß-pleated sheet(beta pleated sheet). The α-helix formation is favoured by alanine, leucine, glutamate and methionine residues, whereas ß-sheet is favoured by valine, isoleucine and tyrosine residues. A third type of secondary structure known as triple helix is found in collagen.


Linus Pauling


Robert Corey

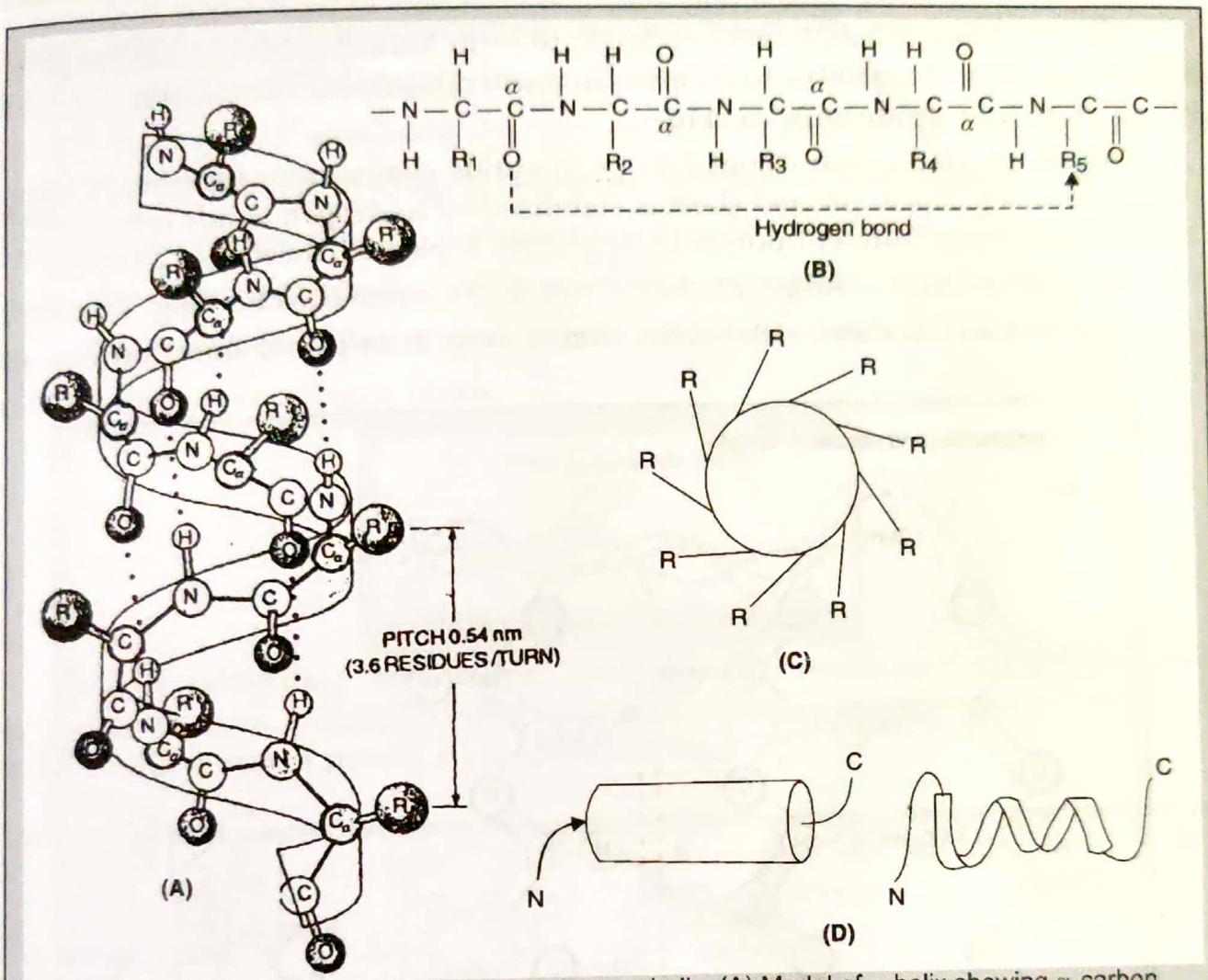


Fig. 9.8 The folding of the polypeptide chain into an α-helix. (A) Model of α-helix showing α-carbon atoms (c_α) along the polypeptide backbone; (B) in the α-helix, the CO group of residue n is hydrogen bonded to the NH group of residue (n + 4); (C) Cross-sectional view of α-helix showing the positions of side chains (R groups) on the outside of the helix; (D) Representation of α-helices by cylinder or coil in topology diagram.

(i) α- Helix : The backbone atoms of a polypeptide chain tightly coiled in a right-handed (clockwise) manner to form many rod-like structures at intervals called α-helices. For example, the single polypeptide chain of myoglobin contains 8 helices. On the outside of helix the side chains extend outward in a helical manner. The length of each helix usually varies from 1.7-4.0nm. In an α-helix, 3.6 amino acid residues present per turn covering a distance (pitch) of 0.54 nm (5.4Å). The α-helix is stabilized by **hydrogen bonds** between the CO group of one amino acid with the NH group of fourth amino acid away. Thus all the main chain CO and NH groups are hydrogen bonded. Glycine and proline are often called helix breakers because of their inability to form hydrogen bonds.

Two or more α-helices may entangle to form **helical cables** or **helical coiled coils** as found in keratin in hair, myosin and tropomyosin in muscle, fibrin in blood clots and epidermin in skin.

(ii) ß- pleated sheet: About 2-15 polypeptide chains come together to form a ß-pleated sheet. The ß- pleated sheet is stabilized by hydrogen bonds between CO- and NH groups in different polypeptide chains. ß-pleated sheet is of 2 types -

Parallel ß-sheet - Adjacent chains run in the same direction e.g. ß-keratin.

Antiparallel ß-sheet - Adjacent chains own in opposite direction e.g. silk fibroin.

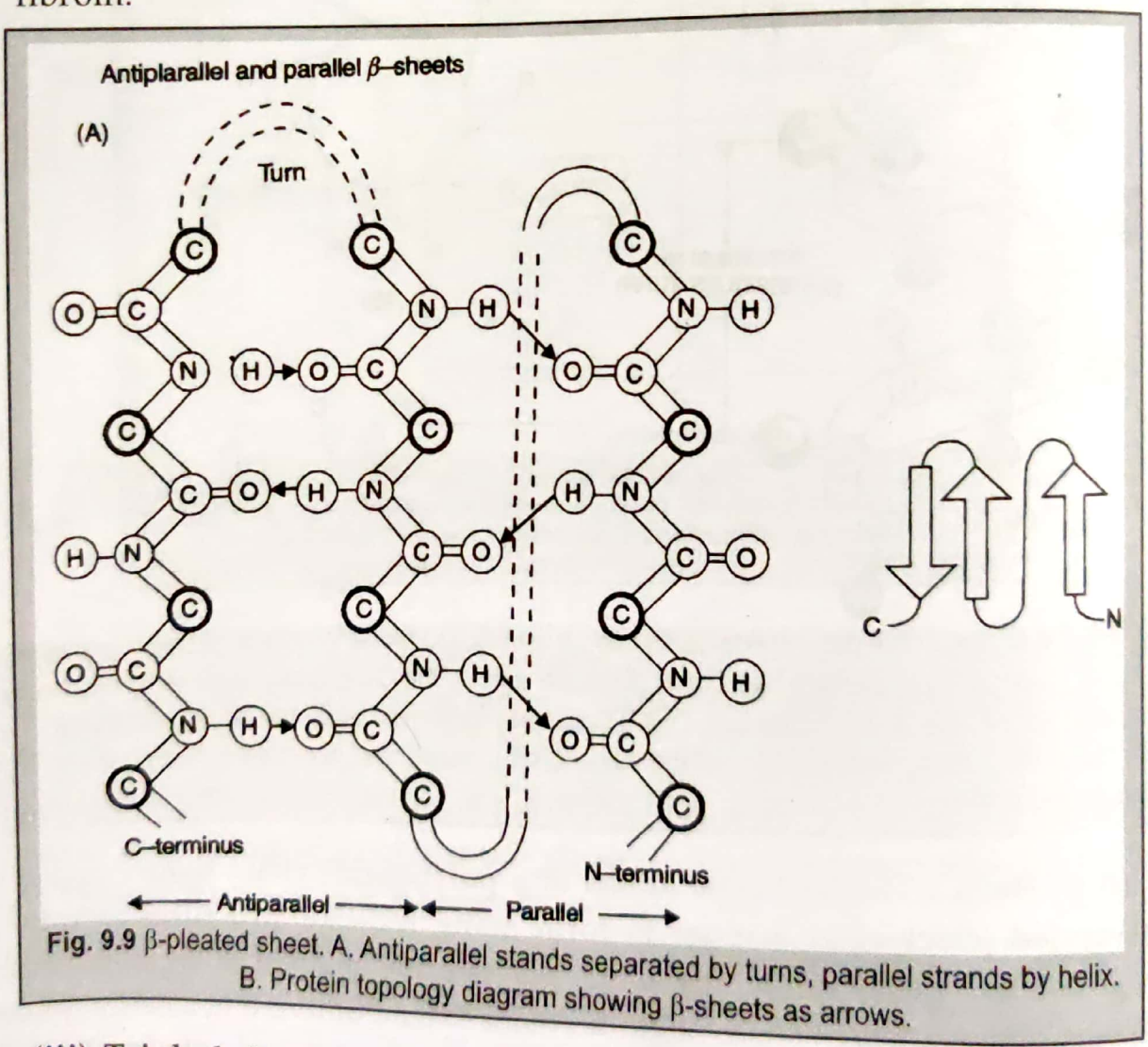


Fig. 9.9 β-pleated sheet. A. Antiparallel stands separated by turns, parallel strands by helix. B. Protein topology diagram showing β-sheets as arrows.

(iii) Triple helix : Collagen, the most abundant protein in human body, **exhibits elongated tripe-helical structure.** A collagen molecule consists of three polypeptide chains that wrap around one another in a right-handed fashion to form triple-helix (Fig. 9.10). The triple helical structure of a collagen molecule is stabilized by both non-covalent interchain hydrogen bonding and covalent bonds

formed between the three polypeptide chains. **Collagen is rich in proline, hydroxy-proline, lysine and hydroxy-lysine residues.** The hydroxy-proline residues play an important role in stabilizing the triple-helical structure of collagen molecule by providing additional hydrogen bondings, whereas hydroxy-lysine and lysine residues are responsible for covalent cross-linking between adjacent colalgen molecules. The neighbouring collagen molecules associate through the covalent cross-linking to form collagen fibres (fibrils). This cross-linking is necessary for high tensile strength of collagen fibres.

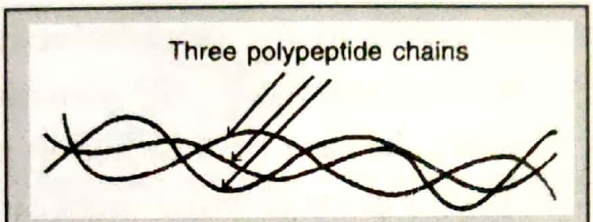


Fig. 9.9 10 Triple-helix of collagen

Super secondary structures: Two or more secondary structures often aggregate to from a complex structural unit called supersecondary structure or motif. Some common motifs are as follows (Fig.9.11).

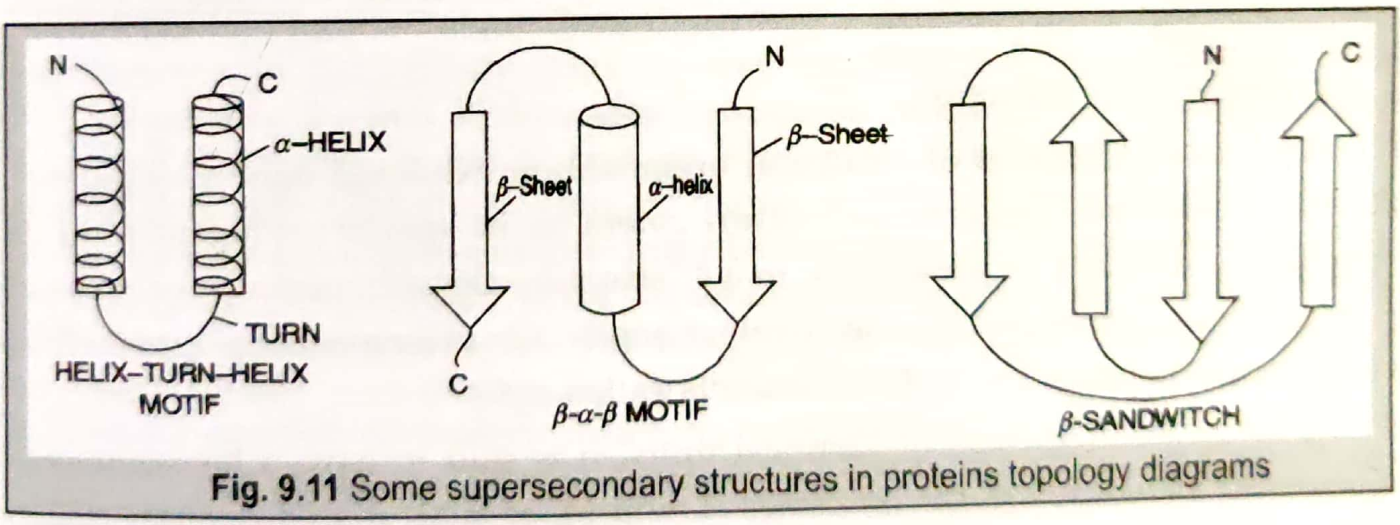


Fig. 9.11 Some supersecondary structures in proteins topology diagrams

Biological significance of secondary structure of proteins

β-amyloid proteins (molecular weight 4.3 kD), deposited in brains of patients with **Alzheimer's disease,** are composed of twisted β-pleated sheet fibrils which are identical to that of silk fibroin. Alzheimer's disease is a **neuropsychiatric disorder** frequency seen at old age.

Brittle bone syndrome (also known as **osteogenesis imperfacta),** an inherited collagen disease, is characterized by abnormal fragility of bones, twisted

spine (humpback) and delayed wound healing. This syndrome results from **defective formation (maturation) of collagen fibres** (Type-1 collagen). The infants born with osteogenesis imperfecta may have multiple fractures and do not survive.

In **scurvy** (vitamin C deficiency), the cross-linking and maturation of collagen, necessary for high tensile strength and proper functioning of connective tissues and for development of bone matrix and dentine, is impaired (see Chapter Vitamins).

C. Tertiary (3°) Structures:

Protein tertiary structure refers to the **intramolecular folding** of an entire polypeptide chain thereby forming a compact 3-D structure which has a specific shape (either globular or fibrous). In case of a largepolypeptide, that consists of more than ~200 residues form two or more globular units called **domains**. (A domain is a compact, globular, structurally independent unit that connects with other such unit by peptide backbone).

The 3° structure is stabilized by hydrogen bonds, ionic bonds, hydrophobic interactions, Vander Walls force, and London dispersion forces and disulfide bonds if present.

D. Quaternary (4°) structure:

It is the fourth level of structural organization exhibited only in oligomeric proteins. A protein's quaternary structure refers to the **spatial arrangement** of its **polypeptide subunits or protomers**. In a 4° structure the subunits may or may not be identical, and stabilized by non-covalent bonds. For example, haemoglobin (Hb) is an oligomeric protein with four subunits called $\alpha_2ß_2$.

If the protein consists of identical units, it is said to have a **homogeneous quanternary structure** e.g. the isozymes H_4 and M_4 of **lactic dehydrogenase** (LDH) and the enzyme **phosphorylase.** If the units are dissimilar, the protein is said to have a **heterogeneous quaternary structure** e.g. **haemoglobin** which consists of two α chains and β chains.

The primary structure of a polypeptide determines the secondary, tertiary and quaternary structure of a protein. Primary and secondary structures are fitted upto higher level tertiary and quaternary structures of a three-dimensional arrangement of amino acids in space. The tertiary and quaternary structures give most protein their functional properties.

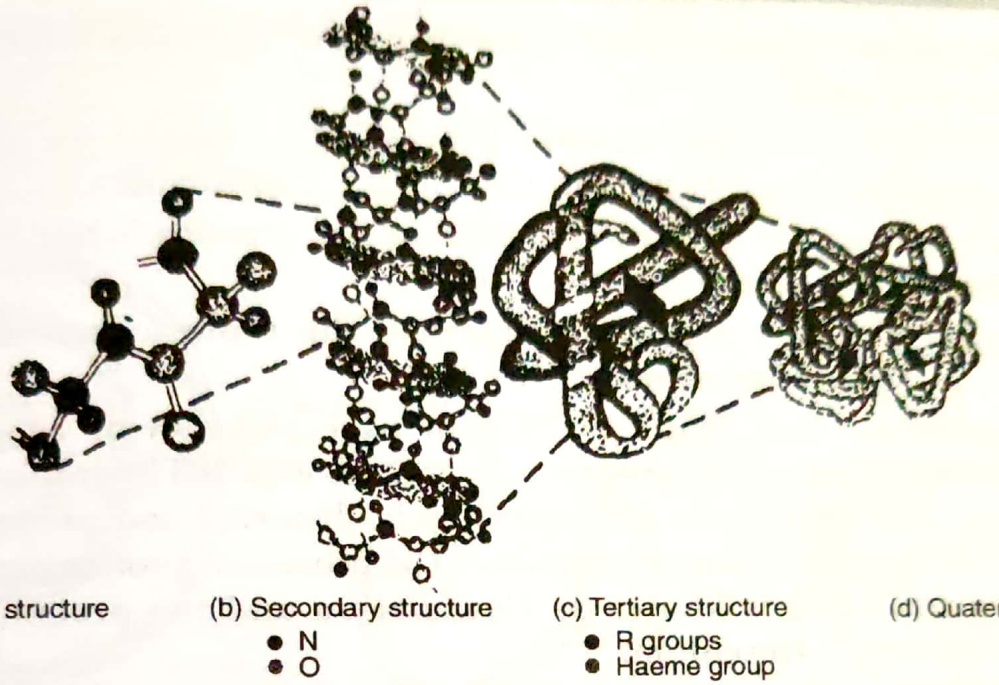


Fig. 9.12 Primary, secondary, tertiary and quaternary structures of haemoglobin

DENATURATION AND RENATURATION OF PROTEINS

Protein Denaturation:

All proteins have a unique three-dimensional structure (tertiary or quaternary structure) and are biologically active (or functional) in their native conformation. The phenomenon of a **loss of three-dimensional structure** (i.e. native conformation of protein) sufficient to cause **loss of biological function** is called **protein denaturation**. Any partial unfolding or change in 3-D structure and function that brings a native state of a protein into distorted or random coil is called **denaturation**. But, the separation of subunits in a 4° structure in called **dissociation**. Denaturation occurs due to partial or complete unfolding (i.e. uncoiling) and disorganization of secondary, tertiary and quaternary structures of proteins. Protein denaturation does not involve loss of primary structure, i.e. denaturation is not accompanied by breaking of peptide bonds. Protein denaturation results from disruption of mainly non-covalent bonds that are responsible for stabilizing all the four levels of structural organization of protein molecule.

Characteristics of Protein denaturation:

1. Protein denaturation results in **loss of biological activity** (e.g. catalytic activity, antigenic property etc.)
2. **Loss of solubility**. A denatured protein becomes insoluble in solution in which it was originally soluble.
3. Denaturation **alters physical properties of proteins** such as electrophoretic mobility, viscosity, surface tension, etc.

4. Denatured proteins **cannot be crystallized.**
5. The **viscosity** of denatured protein (solution) **increases** while its surface **tension decreases.**
6. Denaturation is an **irreversible process** and most of the protein once denatured cannot revert back to their native conformation.
7. Increase in ionizable and sulfhydryl groups of proteins, which is due to loss of hydrogen and disulfide bonds.
8. Denatured protein is more **easily digested**. This is due to increased exposure of peptide bonds to enzymes.
9. Denaturation often leads to protein **precipitation** or **coagulation**, a consequence of protein aggregate formation as exposed hydrophobic surfaces associate. The aggregates are often highly distorted and precipitated from their solutions due to loss of solubility. The process of protein precipitation at isoelectric pH (pI) is known as **flocculation** and the protein precipitate obtained is called **flocculum.**

Agents of Denaturation

Proteins are denatured by variety of conditions such as high temperature, variation in pH and ionic concentrations; addition of detergents etc. Various physical and chemical agents cause protein denaturation.

Physical agents: Heat, X-rays, UV rays, high pressure and violent shaking

Chemical agents: Acids, alkalies, organic solvents, urea, detergents (sodium dodecyl sulfate), heavy metals (Pb, Hg, Cd, Cu etc.), trichloroacetic acid, picric acid, salicylate, guanidine hydrochloride, phosphotungstic acid etc.

Types of Denaturation

Denaturation can be two types: **irreversible denaturation** and **reversible denaturation** (= renaturation). Many proteins undergo irreversible denaturation. For example, albumin cannot be re-natured by removing the physical agent. In reversible denaturation, denatured proteins are re-natured when the physical agent is removed.Ribonuclease is a good example of reversible denaturation.

Protein Renaturation:

When the normal condition is established smaller denatured proteins refold spontaneously into its native conformation. This is called **renaturation** but larger protein can rarely renaturate (fold spontaneously) to its native state.Certain globular proteins (Hb, ribonuclease etc.) denatured by heat, extreme pH or denaturing reagents regain their native structure and biological functions when the denaturing agent is removed. For example, immunoglobulin chains are denatured when treated with urea. When urea is removed by dialysis, the subunits are reassociated and biological activity of immunoglobulin is regained.

Renaturation process was first of all demonstrated by **Anson and Mirsky** (1931) for hemoglobin and by **Christian Anfinsen** (1950s) for ribonuclease. Anfinsen's experiment first demonstrated that the amino acid sequence of a polypeptide chain contain all the information required to fold the chain into its native, three-dimensional conformation.

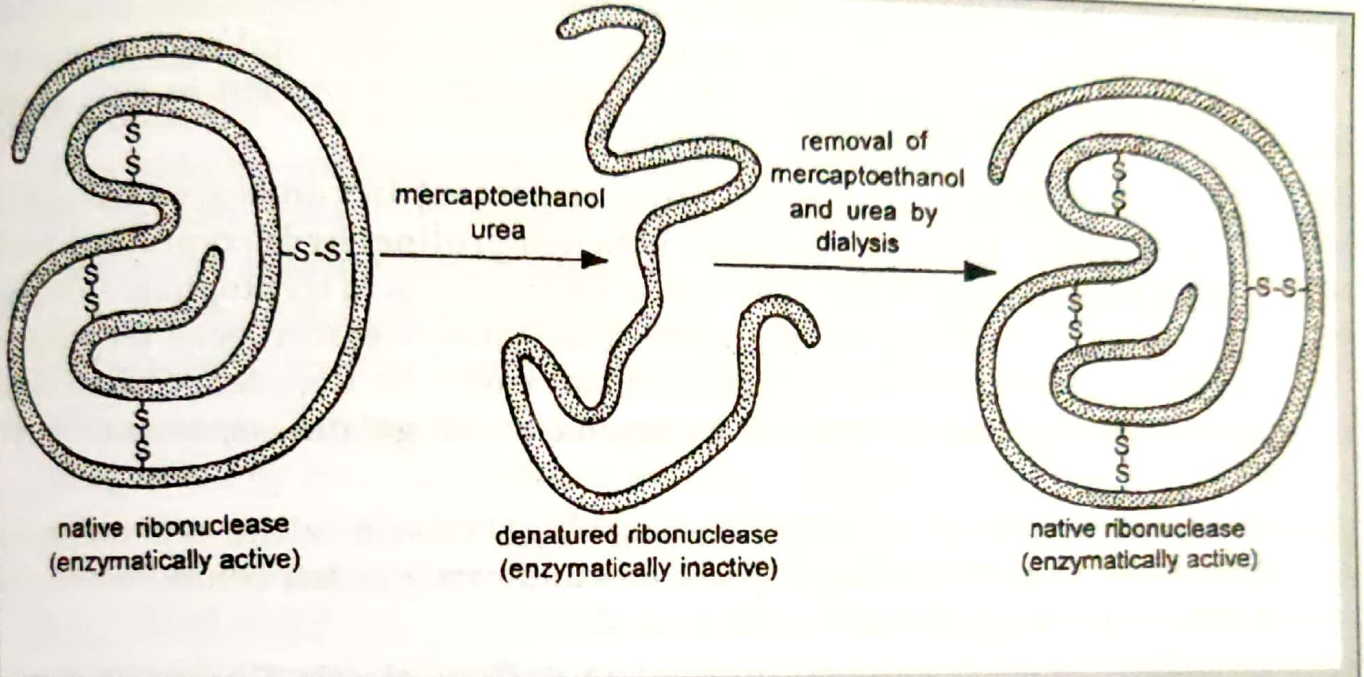


Fig. 9.13 Summary of the Anfinsen establishing that denatured proteins may refold spontaneously backinto their active conformation.

CLASSIFICATION OF PROTEINS

Proteins have been classified into several ways, but none of them are found satisfactory. The broader classification of proteins is given below:

A. Classification based on composition and solubility
B. Classification based on functions
C. Classification based on shape and size
D. Classification based on nutritional value

A. Classification Based on Composition and Solubility

It is the most commonly employed classification of proteins. According to this classification, proteins are classified into three major groups with many subgroups:

- Simple proteins
- Conjugated or complex proteins
- Derived proteins

1. Simple proteins

They are composed of only amino acids and on hydrolysis yield only amino acids. The simple proteins have been divided into two main subgroups: Fibrous protein and Globular proteins.

(a) **Fibrous proteins:** These proteins are elongated **fiber-like structure** and present in **animals only.** They are also known as **scleroproteins.** They are insoluble in all common solvents like water, dilute acid, alkali, organic solvents and neutral salt solution. Most of the fibrous proteins resist digestion by proteolytic enzymes. Some of the fibrous proteins are as follows :

(i) *Keratins :* These proteins are present in exoskeletal structures like hairs, nails, hooves, horns, claws, etc. of mammals. Human hairs and nails contain **α-keratin** rich in cysteine (about 14%). **β-keratins** are present in silk, spider-web, etc. and are deficient in cysteine.

(ii) *Collagens :* Collagens are fibrous proteins **present in connective tissues** like bones, cartilage, teeth, etc. Collagen is rich in **proline, hydroxyproline, lysine and hydroxylysine** and deficient in essential amino acid **tryptophan.** Collagens yield gelatin on strong boiling with acid-water. **Gelatin** lacks tryptophan. Gelatin is a water-soluble protein and can easily be digested. Collagen and gelatin are digested by enzyme **collagenase** and **gelatinase,** respectively in GIT.

(iii) *Ossein :* It is present in bones and cartilage. Ossein is the organic matrix (composed of mainly collagens) of bones and cartilage left behind when their mineral salts are dissolved in dilute acids.

(iv) *Elastins :* These proteins are **present in yellow elastic fibrous tissues** of tendons, blood vessels, uterine muscles, skin, bladder and elastic ligaments. Elastins are digested by enzyme **elastase** in small intestine.

(v) *Spongin :* It is present in **sponges.** Spongin is a type of collagen like protein that forms the fibrous skeleton of sponges, a kind of animals belonging to phylum Porifera.

(vi) *Fibroin :* It is present in **silk fibres.** Fibroin is **rich in glycine, alanine and resine** (about 85% of the total amino acids). Silk is a natural proteinous substance produced by insects known as *Bombyx mori* (silk worms).

Biological Significance of Fibrous Proteins:

- α-Keratins are constituents of **exoskeletal structures** such as hairs, nails, hooves, horns, claws and epidermal layer of skin.
- Collagen is the **most abundant protein** in human body. It has structural and functional role in bones, cartilage, teeth, skin, tendons, blood vessels, basement membranes, etc. A number of diseases, known as collagen diseases, result from inherited defect in collagen synthesis and maturation.
- Gelatin, prepared from collagen, is used as diet in **convalescents.**
- Since elastins are connective tissue proteins with **rubber-like-properties,** elastic fibres are highly stchable. During **pregnancy,** large quantity of elastin is produced in uterus which makes the uterine muscles more stretchable.

(b) **Globular proteins:** These proteins are **spherical** or **oval** in shape and **soluble** in various solvents. They are further classified into following subgroups on the basis of their solubility –

(i) **Albumins:** They are **soluble in water** and **coagulated by heat.** These proteins are present in both animals and plants, e.g. lactalbumin (of milk), egg albumin or ovalbumin (of egg white), plasma albumin, legumelin (of legumes), leucosin (of cereals), etc. Albumins have low amount of glycine.

(ii) **Globulins:** These are insoluble in water but **soluble in dilute salt solutions.** They are also coagulated by heat and can be precipitated from their solution by half saturation with ammonium sulfate. They are present in animals and plants, e.g. **ovaglobulin** (of egg yolk), **lactoglobulin** (of milk), **serum globulins** (e.g. immunoglobulins, ceruloplasmin, plasminogen, transferrin, C-reactive protein etc.), ferritin, **legumin** (of peas), **amandin** (of almonds) etc.

(iii) **Protamines:** They are soluble in water, dilute acids and alkalies. They are not coagulated by heating. They are rich in basic amino acids arginine and lysine, and so are strongly basic. They do not contain cysteine, tryptophane and tyrosine. Protamines are found in association with nucleic acids in sperms, e.g. **salmine** (of salmon sperm), **clupeine** (of herring sperm), **scrombine** (of mackerel sperm) etc. When mixed with insulin, protamines increase the duration of insulin action. Protamine zinc insulinate is a common commercial preparation of insulin. Protamines have isoelectric pH 7.4.

(iv) **Histones:** They are soluble in water, dilute acids and alkalies, but insoluble in dilute ammonia water. Like protamines, histones are **rich in basic amino acids** arginine and histidine. They are not readily coagulated by heat. Like protamine, histones are found in salt combination with nucleic acids. Histones help in packaging of nuclear DNA in eukaryotic cells.

(v) **Prolamines:** They are **soluble in 70-80% ethanol**, but insoluble in water and absolute alcohol. They are **rich in proline**, but **deficient in lysine**. Prolamines are plant proteins mainly found in cereals, e.g. **hordein** (from barley), **avenin** (from oat), **gliadin** (from wheat), **zein** (from maize), **secalin** (from rye) etc.

(vi) **Glutelins:** These are water insoluble, but easily soluble in dilute acids and alkalies. They are heat coagulable. Glutelins are plant proteins, e.g. **glutenin** (from wheat) and **oryzenin** (from rice) etc.

(vii) **Lectins:** They are proteins / glycoproteins, which have at least one non-catalytic domain that exhibits reversible binding to specific monosaccharides or oligosaccharides. They can bind to carbohydrate moieties on the surface of eryshrocytes and agglutinate the erythrocytes. They are found in plant tissues, animals and micro-organisms.

Biological Significance of some Globular Proteins:

- **Gliadin and glutenin,** plant proteins present in wheat, cause coeliac disease in genetically predisposed inviduals.
- **Plasma albumin** maintains electrolyte and water balance in body. Decrease in plasma albumin is seen in nephrotic syndrome, liver disease, Kwashiorkor, etc. and leads to **oedema.**
- **Immunoglobins** - IgA, IgD, IgM, IgE and IgG are γ-globulins are responsible for **humoral immune response.** Immunoglobulins are also called **antibodies.**
- **Ceruloplasmin** along with **transferrin** plays an important role in iron absorption and transport. **Ferritin** serves to store iron.
- Haemoglobin transports oxygen from lungs to various tissues and serves as ICF buffer in RBCs.
- Egg albumin is a storage pretien.
- Protamines and histones are constituents of **chromatin.**

2. Conjugated proteins

They are composed of protein conjugated with some non-protein moiety called **prosthetic group.** The prosthetic groups may be carbohydrate, nucleic acid, lipid, metal, FAD, retinol etc. On hydrolysis, conjugated proteins yield amino acids and non-protein moiety. Conjugated proteins are following types:

(a) Glycoproteins: These are the proteins attached to carbohydrates (prosthetic group). Hydroxyl groups of serine or threonine and amide groups of asparagine and lutamine form linkages with carbohydrate residues. Blood group substances, complements, plasma albumins, ovalbumins, ?-globulins, Castle's intrinsic factor, etc. are the examples of glycoproteins. When carbohydrate content is more than 10% of the molecule, the viscosity is correspondingly increased; they are sometimes called mucoproteins.

(b) Mucoproteins: Mucoproteins are **composed of simple proteins and carbohydrates** such as hyaluronic acid, chondroitin, etc. **Mucin** (of saliva and gastric juice), **α-ovomucoid** and **β-ovomucoid** (of see white), **pregnanediol** (from human pregnancy urine), **gonadotropic hormones** (e.g. FSH, hCG, LH), **oroso-mucoid,** etc. are the examples of mucoproteins. Tendomucoid, osseomucoid and chondroproteins are mucoproteins present in tendons, bones and cartilage, respectively. Mucin and oroso-mucoid contain approximately 15% and 40% carbohydrate, respectively.

(c) Nucleoproteins : Nucleoproteins are **composed of basic proteins** (e.g. histones and protamines **and nucleic acids** (DNA and RNA) as prosthetic group. Nucleohistones and nucleoprotamines are the examples of Nucleoproteins which are present in cell nuclei and are chief constituent of chromatin. Ribosomes are composed of ribonucleoproteins (RNA plus proteins). Synthesis of Nucleoproteins increases during cell division.

(d) **Chromoproteins :** Chromoproteins are **coloured proteins** and are composed of simple proteins combined with coloured prosthetic group. The examples of various chromoproteins are given below.

(i) *Haemoglobin* : It is **composed of heme** (prosthetic group) and **globin** (protein moiety). Hb, a **heme-protein,** is a respiratory protein present in RBCs. Hb is red in colour. **Myoglobin** (in red muscles), **helicorubin** (in mollusks) and **chlorocruorin** (in marine worms) are some other examples of coloured heme-proteins.

(ii) *Cytochromes :* Cytochromes (heme-proteins) are **electron carrier proteins** composed of heme (prosthetic group) and protein moiety. Cytochromes are componentes of mitochondrial electron transport chain. **Catalase** and **peroxidase** are also heme containing conjugated proteins and catalyze decomposition of H_2O_2.

(iii) *Rhodopsin :* It is **composed of retinal** (prosthetic group) and **opsin** (protein moiety). Rhodopsin is a **visual pigment** present in rod cells of retina of eyes and participates in vision in dim light. Rhodopsin is purple in colour, hence also named **visual purple.**

(iv) *Flavoproteins :* These proteins contain **FAD and FMN as prosthetic group** and are **yellow in colour.** For example, succinate dehydrogenase, a flavoenzyme, is **composed of FAD** (prosthetic group) and **apoprotein** (protein moiety). FAD and FMN are coenzyme forms of vitamin B_2 (riboflavin), which is yellow in colour.

(e) **Phosphoproteins :** These proteins are **compsed of phosphoric acid** (prosthetic group) and **protein moiety. Casein** (of milk) and **ovovitellin** (of egg yolk) are the two well-established examples of phosphoproteins. These proteins contain about 1% phosphorus.

(f) **Metalloproteins :** These proteins contain **metal ions as prosthetic group. Carbonic anhydrase** (with Zn), **ceruloplasmin** (with Cu), **transferrin** (with Fe), **ferritin** (with Fe), **glutathione peroxidase** (with Se), **superoxide dismutase** (with Zn and Mo), etc. are the examples of metalloproteins. Metal ions are structural as well as functional part of these proteins.

(g) **Lipoproteins :** Lipoproteins are **composed of apoproteins and lipids** (prosthetic group). On the basis of their density (ratio of proteins and lipids), lipoproteins have been sub-classified in **chylomicrons,** very low density lipoproteins **(VLDL).** Low density lipoproteins **(LDL)** and high density lipoproteins **(HDL).**

Biological Significance of Conjugated Proteins :

- **Mucin** (a mucoprotein) of saliva is required for **oral lubrication,** which is necessary for **mastication,** clear speech and functioning of soft oral tissues against each other.
- Plasma oroso-mucoid is suggested to be an **indicator of acute inflammation.**
- Mucoproteins like tendomucoid, osseomucoid and chondroproteins are important constituents of the ground substance of tendons, bones and cartilage, respectively.
- Cytochromes take part in **mitochondrial electron transport** during oxidative phosphorylation.
- **Haemoglobin** takes part in **oxygen transport** in blood and is **main ICF buffer of RBCs.**
- Flavoenzymes **(flavoproteins)** participate in **oxidation-reduction reactions.**
- Milk protein **casein** is a storage protein and good source of phosphorus.
- Ceruloplasmin, transferrin and ferritin (all metalloproteins) participate in iron absorption, transport and storage.
- **Lipoproteins serve to transport lipids.** Estimation of serum VLDL, LDL and HDL has diagnostic importance in coronary heart disease, hypertension and atherosclerosis.
- Metalloenzymes - peroxidase, catalase and superoxide dismutase are important enzymes of **antioxidant defence system** and protect the cell from oxidative damage.
- **Intrinsic factor,** a glycoprotein, produced by gastric parietal cell is required for **intestinal absorption of vitamin B_{12}.**

3. Derived proteins

They **do not exist in nature**. They are **derived from the denaturation or degradation of natural or native proteins.** The **primary derived proteins** include denatured and coagulated proteins that are produced by agents such as heat, acids, alkalies etc., while the **secondary derived proteins** are the hydrolytic products of proteins. Progressive hydrolysis of proteins results in smaller and smaller chains: Protein → peptones → peptides → amino acids.

B. Classification Based on Functions :

1. Catalytic proteins, e.g. enzymes.
2. Structural proteins, e.g. collagen, elastin.
3. Contractile proteins, e.g. myosin, actin.
4. Transport proteins, e.g. haemoglobin, myoglobin, albumin, transferrin.

5. Regulatory proteins or hormones, e.g. ACTH, insulin, growth hormone.
6. Genetic proteins, e.g. histones.
7. Protective proteins, e.g. immunoglobulins, interferons, clotting factors.

C. **Classification Based on Shape :**

Globular Proteins : They are spherical or oval in shape. They are easily soluble, e.g. albumins, globulins and protamines.

Fibrous Proteins : The molecules are elongated or needle shaped. Their solubility is minimum. They resist digestion. Collagen, elastin and keratins are examples.

D. **Classification Based on Nutritional Value :**

Nutritionally Rich Proteins : They are also called as **complete proteins or first class proteins.** They contain all the essential amino acids in the required proportion. On supplying these proteins in the diet, children will grow satisfactorily. A good example is **casein** of milk.

Incomplete Proteins : They **lack one essential amino acid.** They cannot promote body growth in children; but may be able to sustain the body weight in adults. Proteins from **pulses are deficient in methionine,** while proteins of **cereals lack in lysine.** If both of them are combined in the diet, adequate growth may be obtained.

Poor Proteins : They **lack in many essential amino acids** and a diet based on these proteins will not even sustain the original body weight. Zein from corn lacks tryptophan and lysine.

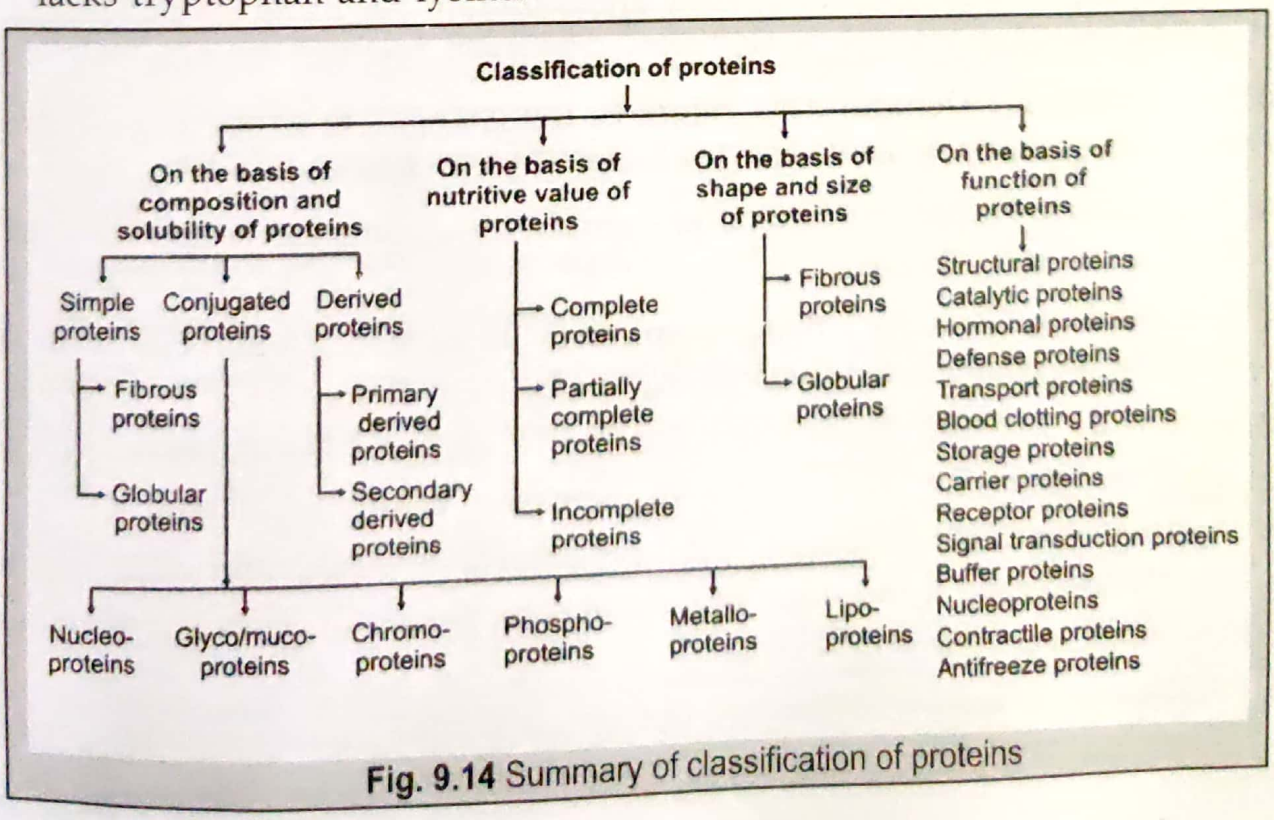


Fig. 9.14 Summary of classification of proteins

5.7 Immunoglobulins : Structure and function

Antibodies, the antigen-binding *glycoproteins* are synthesized exclusively by B-cells and in billions of forms, each with a different amino acid sequence and a different antigen binding site. Collectively called *immunoglobulins* (Ig), they are among the most abundant protein components in the blood, constituting about 20% of the total protein components in the blood plasma. Antibodies (Ab) are present on the B-cell membrane and also are secreted by plasma cells.

5.7.1 Basic structure of antibody molecule

The simplest antibodies are Y-shaped molecules with two identical antigen-binding sites, one at the tip of each arm of the Y. Because of their two antigen-binding sites, they are described as *bivalent*.

Antibody has a common structure of 4 polypeptide chains. It is a heterodimer and consists of two identical **light (L) chains** (each containing about 220 amino acids residues, about 25000 MW) and two identical **heavy (H) chains** (each usually containing about 440 amino acids residues, about 50,000 MW). Each light chain is bound to heavy chain by disulfide bridges and other non-covalent linkages. Thus, antibody is a dimer of H—L chain.

All species studied have the two major classes of light chains: κ and λ. Any one individual of a species produces both types of light chain. However, in any one immunoglobulin molecule, the light chains are always either both κ or both λ, never one of each. While there are two types of light chains, the immunoglobulins of virtually all species have been shown to consist of five different types of heavy chains- α, γ, δ, ε and μ. These five different types of heavy chains are called *isotypes*. The heavy-chains of a given antibody molecule determine the class of that antibody: IgM (μ), IgG (γ), IgA (α), IgD (δ) or IgE (ε). Each class can have either κ or λ light chains. Any individual of a species makes all heavy chains, but in any one antibody molecule, both heavy chains are identical. Thus an antibody molecule of the IgG class could have the structure κ2γ2 with two identical κ light chains and two identical γ heavy chains. Alternatively, it could have the structure λ2γ2 with two identical λ light chains and two identical γ heavy chains.

Immunoglobulin heavy chain isotypes

Isotype	*Heavy chain*
IgM	μ
IgD	δ
IgG	γ
IgA	α
IgE	ε

Minor differences in the amino-acid sequences of the α and the γ heavy chains led to further classification of the heavy chains into subclasses. In humans, there are two subclasses of α heavy chains (α_1 and α_2) and four subclasses of γ heavy chains (γ_1, γ_2, γ_3 and γ_4).

Both light and heavy chains have a variable sequence at their N-terminal ends but a constant sequence at their C-terminal ends. Light chains have a **constant region** (C_L) about 110 amino acids long and a **variable region** (V_L) of the same size. The variable region (V_H) of the heavy chains (at their N-terminus) is also about 110 amino acids long, but the heavy-chain constant region (C_H) is about three to four times longer (330 or 440 amino acids), depending on the class. It is the N-terminal ends of the light and heavy chains that come together to form the antigen-binding site.

The diversity in the variable regions of both light and heavy chains is for the most part restricted to three small *hypervariable regions* (each ~10 amino acid residues long) in each chain called **complementarity determining regions** (CDR); the remaining parts of the variable region, known as **framework regions**, are relatively constant. Proceeding from either the V_L or V_H amino terminus, these regions are called CDR1, CDR2 and CDR3. The CDR3 is the most variable of the CDRs.

Deduction of Ab structure

Porter and Edelman elucidated the basic structure of the Ig molecule but their approaches were different. Porter found that proteolytic treatment with the enzyme *papain* split the Ig molecule (molecular mass 150,000Da) into three fragments of about equal size. Two of these fragments were found of equal size, each of molecular mass 45,000Da called **Fab fragments** (fragment antigen binding) because they had Ag binding activity. Fab fragments are considered to be univalent, possessing one binding site each and being in every way identical to each other. The third fragment of molecular weight 50000 called the **F_c fragment** (fragment crystallizable). It cannot bind antigen. At about the same time, Edelman discovered that when γ-globulin was extensively reduced by treatment with mercaptoethanol, the molecule fell apart into four chains: two identical light chains with a molecular mass of about 53,000Da each and two others of about 22,000 Da each. The large molecules were designated **heavy** (H) **chains** and the smaller ones, **light** (L) **chains**. On the basis of these results, the structure of immunoglobulin molecules, was proposed.

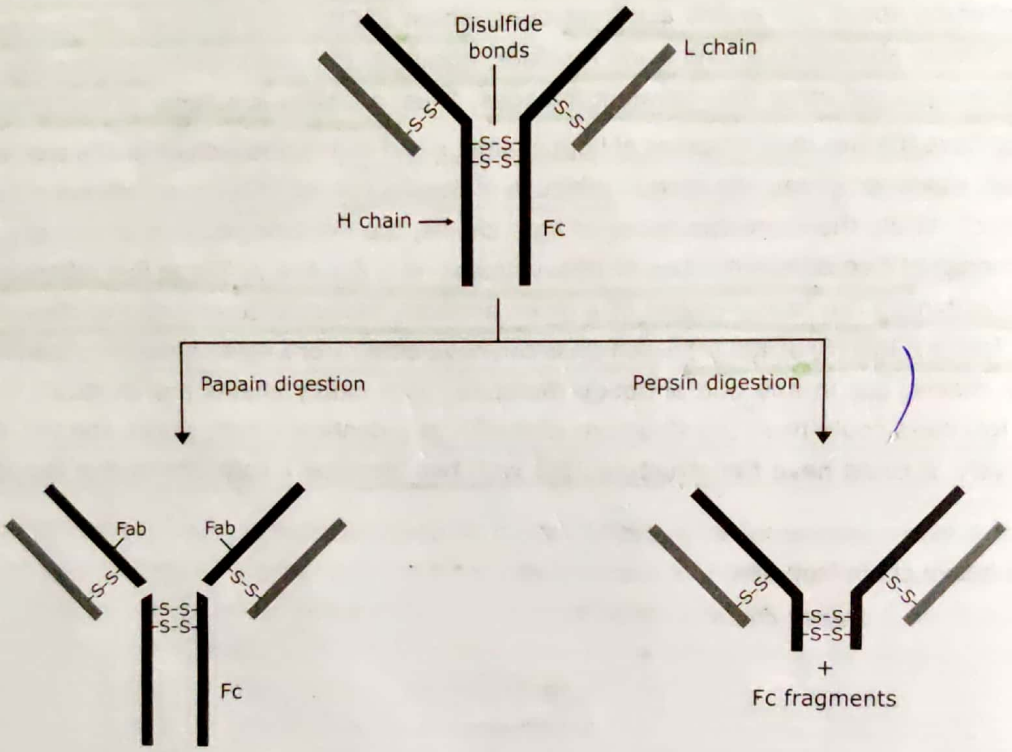


Figure 5.12 Prototype structure of IgG, showing chain structure and interchain disulfide bonds. The fragments produced by various treatments are also indicated. Light (L) chains are depicted in gray and heavy (H) chains in black.

Another researcher, Nisonoff used *pepsin* to digest the Ig. It generated a single 100,000 MW fragment composed of two Fab-like fragments and designated $F(ab')_2$. Unlike Fab-fragments, $F(ab')_2$ fragment was able to precipitate antigens. The F_C fragment was not recovered from pepsin digestion, because it had been digested into multiple fragments.

Immunoglobulins domains

The structure of the immunoglobulin molecule is determined by the primary, secondary, tertiary and quaternary organization of the protein. The *primary structure* i.e. the amino-acid sequence, accounts for the variable and constant regions of the heavy and light chains. The *secondary structure* is formed by folding of the extended polypeptide chain back and forth upon itself into an *antiparallel β-pleated sheet*. The chains are then folded into a tertiary structure of compact globular domains, which are connected to neighboring domains by continuations of the polypeptide chain that lie outside the β-pleated sheets. Finally, the globular domains of adjacent heavy and light

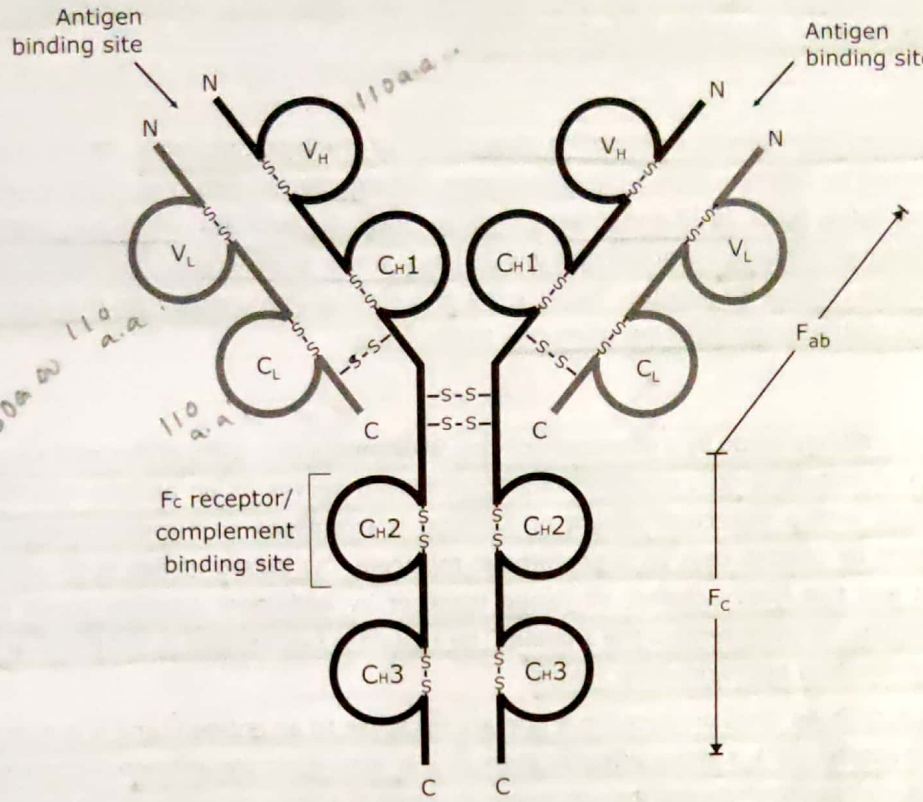


Figure 5.13 Schematic diagram of structure of immunoglobulins.

polypeptide chains interact in the *quaternary structure* forming functional domains that enable the molecule to specifically bind antigen and at the same time, perform a number of other biological effector functions. Both light and heavy chains are made up of repeating segments—each about 110 amino acids long and each containing one intrachain disulfide bond. These repeating segments fold independently to form compact functional units called immunoglobulin (Ig) domains. A light chain consists of one variable (V_L) and one constant (C_L) domain. Similarly heavy chain consists of one variable (V_H) and three or four constant (C_H) domains. Most heavy chains have three constant domains (C_H1, C_H2, and C_H3), but those of IgM and IgE antibodies have four (C_H1, C_H2, C_H3 and C_H4). V_L and C_L domains of light chain pair with the variable (V_H) and first constant (C_H1) domain of the heavy chain to form the antigen-binding region.

The γ, δ and α-heavy chains contain an extended peptide sequence between the C_H1 and C_H2 domains that has no homology with the other domains. This region, called the **hinge region**, is rich in proline and cysteine residues and is flexible, giving IgG, IgD and IgA flexibility between the two Fab arms of the Y-shaped antibody molecule. The hinge region varies in length from 10 to more than 60 amino acid residues in different isotypes.

Immunoglobulin can be expressed either as *secreted* immunoglobulin or as *membrane bound* immunoglobulin. The carboxyl-terminal domain in secreted immunoglobulin differs from membrane-bound immunoglobulin. Secreted immunoglobulin has a hydrophilic amino acid sequence at the carboxyl-terminal end. In membrane-bound immunoglobulin, the carboxyl terminal domain contains three regions: An extracellular hydrophilic spacer sequence, a hydrophobic transmembrane sequence and a short cytoplasmic tail.

5.7.2 Different classes of immunoglobulin

In mammals, there are five *classes* of antibodies, IgA, IgD, IgE, IgG, and IgM, each with its own class of heavy chain–α, δ, ε, γ, and μ respectively. In addition, there are a number of subclasses of IgG and IgA immunoglobulins; for example, there are four human IgG subclasses (IgG1, IgG2, IgG3, and IgG4), having γ_1, γ_2, γ_3, and γ_4 heavy chains, respectively. Each class (and subclass) has the characteristic properties of its own. Except for their variable

regions, all the immunoglobulins within a class have about 90% homology in their amino acid sequences, but only 60% homology exists between classes (e.g. IgG and IgA).

IgG

IgG is the most abundant Ig in serum constituting about 80% of the total serum Ig. There are four human IgG subclasses, distinguished by differences in γ-chain sequence and numbered according to their decreasing average serum concentrations: IgG1, IgG2, IgG3 and IgG4. Except for the IgG3 subclass, which has a rapid turnover, with a half-life of 7 days, the half-life of IgG is approximately 23 days, which is the longest half-life of all immunoglobulin isotypes. The IgG isotype (except for subclass IgG2) is the only class of immunoglobulin that can pass through the placenta, enabling the mother to transfer her immunity to the fetus.

IgM

IgM is the first class of antibody made by a developing B-cell. It accounts of 5–10% of the total serum Ig. IgM is also found on the surface of mature B-cells together with IgD where it serves as an antigen-specific B-cell receptor (BCR). Monomeric IgM, with a molecular weight of 180,000 is expressed as a membrane-bound antibody on B-cells. IgM is secreted by plasma cells as a pentameric molecule, composed of five such units, each of which consists of two light and two heavy chains, all joined together by additional disulfide bonds between their F_c portions and by a polypeptide chain termed the **J-chain**. The J-chain is synthesized in the B-cell or plasma cell. The half-life of the IgM molecule is approximately 5 days.

IgM is the first immunoglobulin class produced in a primary response to an antigen, and it is also the first Ig to be synthesized by the neonate. IgM is more efficient than IgG in activating complement. Complement activation requires two F_c regions in close proximity, and the pentameric structure of a single molecule of IgM fulfills this requirement.

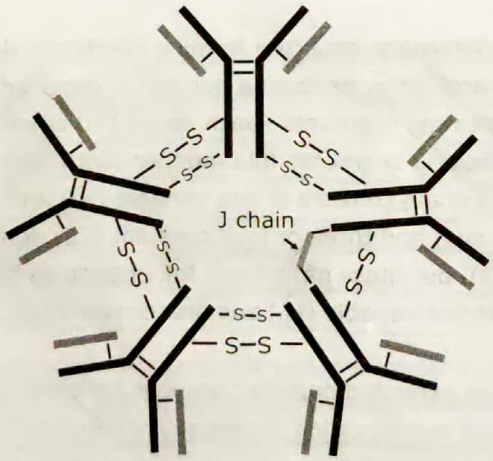


Figure 5.14 IgM forms pentamer where five immunoglobulins are covalently linked together with disulfide bonds. Because each monomer has two antigen binding sites, a pentameric IgM has 10 binding sites. A polypeptide called J (joining) chain links the heavy chains of monomeric units of polymeric IgM. The linkage is by disulfide bonds between the J-chain and the carboxyl-terminal cysteines of IgM heavy chains.

IgA

IgA constitutes 10–15% of the total Ig in serum. It is the predominant Ig class in external secretions such as breast milk, saliva, tears and mucus of the bronchial, genitourinary and digestive tracts. The IgA class of immunoglobulins contains two subclasses: IgA_1 (93%) and IgA_2 (7%). Serum IgA has a half-life of 5.5 days. The IgA present in serum is predominantly monomeric. The IgA of external secretions, called *secretory IgA*, consists of a dimer or tetramer with a *J-chain* polypeptide (15 kDa) and a polypeptide chain called *secretory component*. Secreted IgA is transported through epithelial cells into the intestinal lumen (a process known as *transcytosis*) by an IgA specific Fc

receptor called the *poly-Ig receptor*. It is an integral membrane glycoprotein. The secreted, dimeric IgA containing the J-chain binds to the poly-Ig receptor on mucosal epithelial cells. This complex is endocytosed into the epithelial cell and actively transported in vesicles to the luminal surface. Here the poly-Ig receptor is proteolytically cleaved, its transmembrane and cytoplasmic domains are left attached to the epithelial cell and the extracellular domain of the receptor, which carries the IgA, molecule, is released into the intestinal lumen. The soluble IgA-associated component of the receptor is called the **secretory component**.

The daily production of secretory IgA is the maximum among other Ig. It is the major immunoglobulin found in the **colostrum** of milk in nursing mothers, and it may provide the neonate with a major source of protection against pathogens during the first few weeks after birth.

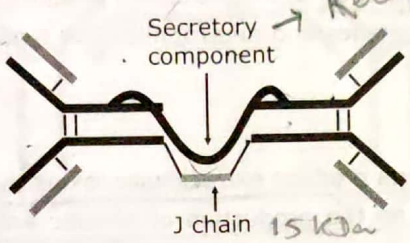


Figure 5.15 IgA (dimer). It is composed of at least two monomeric IgA molecules, covalently linked through the J-chain and secretory component.

IgE

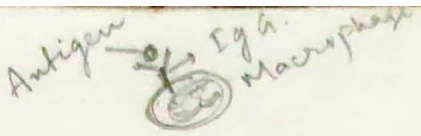
It mediates the immediate hypersensitivity reactions that are responsible for the symptoms of hay fever, asthma, hives and anaphylactic shock. IgE binds to F_C receptors on the membranes of blood basophils and tissue mast cells. Cross-linkage of receptor-bound IgE molecules by antigen (allergen) induces degranulation of basophils and mast cells; as a result a variety of pharmacologically active mediators giving rise to allergic manifestations. IgE mediated degranulation is necessary for anti-parasitic defense.

IgD

It constitutes 0.2% of the total Ig in serum. IgD together with IgM, is the major membrane bound Ig expressed by mature B-cells. It is thought to function in the activation of B-cells by Ag. No biological effector function has been identified for IgD.

Table 5.6 Characteristics of different antibody classes.

	IgG	IgA	IgM	IgD	IgE
Molecular weight ($\times 10^{-3}$)	150	150, 400	900	180	190
"Y" units/molecule	1	1-2	5	1	1
Percentage of total Ig in blood	~80	10-15	5-10	<1	<1
Crosses Placenta	+	-	-	-	-
Enters secretions	+	+ +	-	-	-
Agglutinates particles	+	+	+ + +	-	-
Allergic reactions	+	-	-	-	+ + + +
Activate complement	++	-	++++	-	-
Half life (days)	~23	~5.5	~5	~2.8	~2
Additional protein subunit	-	J and S	J	-	-
Binds to macrophage/neutrophils	+	-	-	-	-
Binds to mast cells	-	-	-	-	+
Percentage carbohydrate	3	7	7-10	12	11



5.7.3 Action of antibody

In addition to binding antigen, antibodies participate in a broad range of effector functions—that will result in removal of the antigens and death of the pathogens. Because these effector functions result from interactions between heavy-chain constant regions and other serum proteins or cell-membrane receptors, not all classes of antibodies have the similar functional properties. Some major antibody mediated effector functions are given below.

Opsonization

Opsonization is the process by which microorganisms or other foreign particles are coated with antibody and/or complement, and thus prepared for *recognition* and *ingestion* by phagocytic cells. Opsonizing antibodies, IgG, bind to F_c receptors on the surface of macrophages and neutrophils. This binding provides the phagocyte a method to capture antigens.

Toxin/viral neutralization

Some Gram positive and negative bacteria produce extracellular toxins that contribute to their pathogenic effects. Immunity against such toxin depends on the production of specific antibodies that inactivate the toxins. This process is called *toxin neutralization.* Once neutralized, the toxin-antibody complex is either unable to attach to receptor sites on host target cells and is unable to enter the cell, or it is ingested by macrophages. An antibody capable of neutralizing a toxin is called *antitoxin.* Similarly, antibody like IgG, IgM, and IgA can bind to some viruses during their extracellular phase and inactivate them. This antibody mediated viral inactivation is called *viral neutralization.* Viral neutralization prevents a viral infection due to the inability of the virus to bind to and enter its target cell.

Activation of complement

Complement is a set of plasma proteins that can be activated either by binding to certain pathogens or by binding to the antibody. Complement activation is often described as a series of cascading enzymatic events, leading to the generation of specific complement components that cause *opsonization* and *phagocytosis* of infectious agents as well as direct lysis of the invading organisms. IgM and most IgG subclasses (IgG1, IgG2, IgG3) can activate the complement system.

Immune complex formation

The interaction of antigen with antibodies may result in a variety of consequences, including **precipitation** and **agglutination**. Precipitation results from the interaction of a *soluble* antibody with a *soluble* antigen to form an *insoluble* complex. Antibodies that thus aggregate soluble antigens are called **precipitins**.

The reactions of antibody with a multivalent antigen that is particulate (i.e. an insoluble particle) results in the cross-linking of the various antigen particles by the antibodies. This cross-linking eventually results in the clumping or agglutination of the antigen particles by the antibodies.

Antibody dependent cell-mediated cytotoxicity

The function of NK cell is to destroy malignant cells and cells infected with microorganisms. They recognize their targets with the help of binding with antibodies. They can bind to antibodies that coat infected or malignant cells; thus the antibody bridges the two cell types. This process is called Antibody-Dependent Cell-mediated Cytotoxicity (ADCC), and can result in the death of the target cell.

5.7.4 Antigenic determinants on immunoglobulins

Since antibodies are glycoproteins, they can themselves function as potent immunogens to induce an antibody response. The antigenic determinants, or epitopes on Ig molecules fall into three major categories which are located at characteristic portions of the molecule.

Isotype

An antibody class is determined by the constant region sequence of the heavy chain. The five human isotypes, designated IgA, IgD, IgE, IgG and IgM, exhibit structural and functional differences. Isotypic determinants are constant-region determinants that collectively define each heavy-chain class and sub-class within a species. Each isotype is encoded by a separate constant-region gene, and all members of a species carry the same constant-region genes. Different species inherit different constant-region genes and therefore express different isotypes. Therefore, when an antibody from one species is injected into another species, the isotype determinants will be recognized as foreign, inducing an antibody response to the isotypic determinants on the foreign antibody.

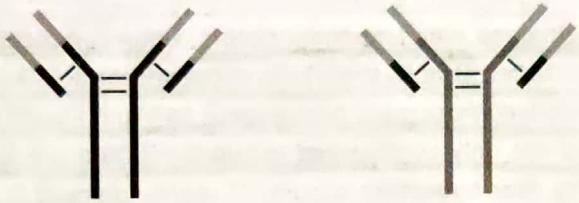


Figure 5.16 Isotype - an antibody class that is determined by the constant-region sequence of the heavy chain.

Allotype

It is based on *genetic differences among individuals*. It depends on the existence of *allelic forms* of the same gene. These alleles encode minor amino-acid differences, called allotypic determinants, that occur in some but not all members of a species. As a result of allotype, a heavy or light chain constituent of any immunoglobulin can be present in some members of a species and absent in others. This situation contrasts with that of immunoglobulin classes or subclasses, which are present in all members of a species.

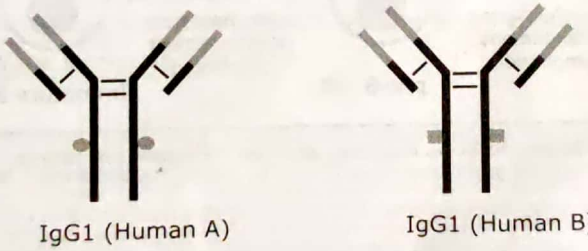


Figure 5.17 Allotype - alternative allelic forms of the same isotype.

Idiotype

The unique amino-acid sequence of the variable domains of a given antibody can function not only as an antigen-binding site but also as a set of antigenic determinants. The idiotypic determinants are generated by the conformation of the heavy- and light-chain variable regions. Each individual antigenic determinant of the variable region is referred to as an *idiotype*.

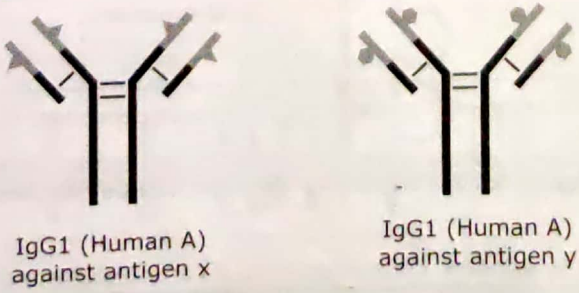


Figure 5.18 Idiotype – a set of antigenic determinants (idiotopes) characterizing a unique antibody.