CONTENTS

- Importance
- Amino Acids
- Nonstandard Protein Amino Acids
- Nonprotein Amino Acids
- Peptides
- Chemical Bonds Involved in Protein Structure
- Characteristics of Chemical Bonds
- Protein Configuration
- Dynamics of Globular Proteins Structure
- Prediction of Secondary and Tertiary Protein Structures
- Cleavage of a Protein
- Biological Roles of Proteins

Two Examples of the Thousands of Biological Structures Composed Predominantly of Protein

(*a*) **Lenses of eyes**, as in this net-casting spider, which focus light rays. *and (b) Mantis wildlife Films)*
(*b*) **feathers,** which are adaptations in birds for thermal insulation, flight, and sex recognition. *Courtesy: (a) Frans Lanting*

CHAPTER

9

Proteins–I

General Structure

IMPORTANCE

The name protein (*proteios*[G] = pre-eminent or first) was first suggested, in 1838, by a Swedish chemist Berzelius to a Dutch chemist Mulder, who referred it to the complex organic nitrogenous substances found in the cells of the living beings. They are most abundant intracellular macro-molecules and constitute over half the dry weight of most organisms. "Proteins occupy a central position in the architecture and functioning of living matter. They are

Jöns Jacob Berzelius (LT, 1779-1848) was Professor of Chemistry at Stockholm, Sweden. He is famous for developing the modern system of chemical symbols. He isolated several elements and is also credited with the discovery of selenium (Se). He developed the concept of atomic weight. With Humphrey Davy, he laid the foundations of electrochemistry. He is also credited with publishing the earliest book on Biochemistry entitled 'Lectures in Animal Chemistry' in 1806.

Gerardus Johannes Mulder's (LT, 1802-1880) comment about proteins reads as follows:

> "There is present in plants and animals a substance which ... is without doubt the most important of the known substances in living matter, and without it, life would be impossible on our planet. This material has been named Protein."

intimately connected with all phases of chemical and physical activity, that constitute the life of the cell. Some proteins serve as important structural elements of the body, for example, as hair, wool and collagen, an important constituent of connective tissue ; other proteins may be enzymes, hormones or oxygen- carriers. Still other proteins participate in muscular contraction, and some are associated with the genes, the hereditary factors."(White, Handler and Smith, 1964).

They are, therefore, essential to cell structure and cell function. The proteins with catalytic activity (enzymes) are largely responsible for determining the phenotype or properties of a cell in a particular environment. The total hereditary material of the cell or genotype dictates which type of protein the cell can produce. In fact, the proteins have built into their structure the information that instructs them in "what to do" (catalytic activity), "where to go" in the cell (intracellular organization) and "when and how to perform" (control of function through bacterial interactions with other proteins, small activators or inhibitors).

The bacterium *Escherichia coli* is estimated to contain about 5,000 different types of compounds which include some 3,000 different kinds of proteins and 1,000 nucleic acids. In humans, there may be 1,00,000 different kinds of proteins each with a unique structure. None of the proteins of *E.coli* is indentical with any of the human proteins. Thus, in about 1.5 million species of living organisms, there are probably 10^{10} to 10^{12} different kinds of protein molecules and about 10^{10} different kinds of nucleic acids.

The constituent elements of proteins are carbon, hydrogen, oxygen, nitrogen and very rarely sulfur also. In certain complex proteins, phosphorus occurs as well. The elemental composition of proteins in plants and animals presents a great deal of variation (Table 9-1).

Table 9–1. Elemental composition of plant and animal proteins

Protein	*Carbon*	*hydrogen*	*Oxygen*	*Nitrogen*
Green plants	**54**	**7**	**38**	**0.003**
Mammals	**21**	**10**	**62**	**3.0**

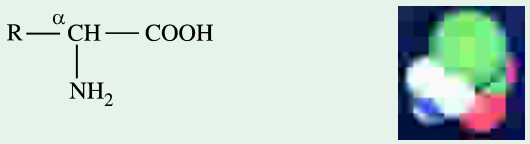
Most animal proteins contain from 0.5 to 2.0 per cent sulfur. Insulin is, however, a notable exception to this in possessing about 3.4% sulfur.

AMINO ACIDS

Structure

All proteins are macromolecules because of their very high molecular weights. These are the polymers, *i.e.*, chain-like molecules produced by joining a number of small units of amino acids called monomers. The amino acids are, therefore, regarded as '**building blocks of proteins**'. The general formula of an amino acid along with its space-filling model in presented in Fig, 9-1.

The term **macromolecule** was coined by Schaudinger in 1920s



(a) General formula

(b) Space-filling model. The large green sphere represents the R group.

Fig. 9–1. General structure of an amino acid

Each amino acid is a nitrogenous compound having both an acidic carboxyl (— COOH) and a basic amino (— NH_2) group. R stands for the side chains that are different for each amino acid. R can be as simple as a hydrogen atom (H) or a methyl group (— CH_3) or a more complex structure. The first carbon is the part of the carboxyl group. The second carbon, to which is attached the amino group, is called the α-*carbon*. The α-carbon of most amino acids is joined by covalent bonds to 4 different groups. Thus, the α-carbon in all the amino acids is asymmetric except in glycine where the α-carbon is symmetric. Because of this asymmetry, the amino acids (of course, except glycine) exist in two optically active forms : those having — NH_2 group to the right are designated as D-forms and those having — NH_2 group to the left as L-forms (Fig. 9–2). However, the two amino acids, threonine and isoleucine have two asymmetric carbon atoms each and thus have $2^n = 2^2 = 4$ optical isomers. At pH 7.0, both the carboxyl and amino groups are ionized.

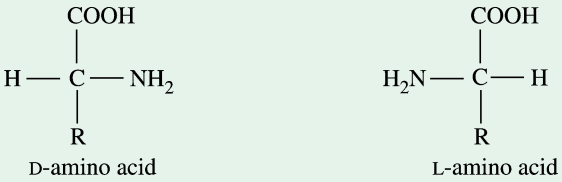


Fig. 9–2. Two optical isomers of amino acids

Specific Rotation

It is interesting to note that *the amino acids found in the proteins belong to the L-series*. Many of the naturally occurring L-amino acids rotate the plane of polarized light to the left (*i.e.*, they are levorotatory) while others rotate the plane of polarized light to the right (*i.e.*, they are dextrorotatory). Table 9–2. shows the specific rotation of some L-amino acids.

Table 9–2. Specific rotation of some amino acids isolated from proteins

Amino acid	*Specific rotation,* $[\alpha]_D^{25^\circ}$
L-alanine	– 86.2
L-histidine	– 38.5
L-phenylalanine	– 34.5
L-threonine	– 28.5
L-serine	– 7.5
L-alanine	+ 1.8
L-glutamic acid	+ 12.0
L-isoleucine	+ 12.4
L-arginine	+ 12.5
L-lysine	+ 13.5

Thus, it is evident that the symbols D and L do not identify the property of light rotation, *i.e.*, D-isomers can be either dextrorotatory (*d*) or levorotatory (*l*); similarly, L-isomers can be either (*d*) or (*l*). However, to minimize confusion, the symbols *d* and *l* are usually not used nowadays. Moreover, the **DL nomenclature** has limitations because it describes the asymmetry of only one carbon atom in a compound and many biomolecules contain two or more asymmetric carbon atoms. The R and S classification or **RS notation** of isomers, introduced in 1956 by Robert Cahen, Christopher Ingold and Vladimir Prelog and being currently used in chemistry, is more useful for defining the asymmetry

of biomolecules because its accounts for all asymmetric carbons in an isomer. If any atom (other than H) or group on the asymmetric carbon is on the right side, that asymmetric carbon is designated as R (from *rectus*^{L} = right) ; conversely, if any atom (other than H) or group is on the left side, the asymmetric carbon is then designated as S (from *sinister*^{L} = left). As an illustration, isoleucine (with two asymmetric carbon atoms, C–2 and C–3) will have four stereoisomers. And the configuration of the biological isomer, L-isoleucine is more completely described as (2 S), (3 S)–isoleucine (Fig. 9–3).

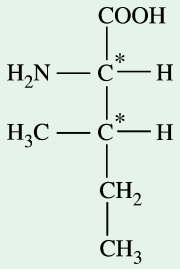


Fig. 9–3. Fisher projection formula of L-isoleucine of (2S), (3S)-isoleucine

The two asymmetric carbon atoms are indicated by an asterisk each.

Distribution in Proteins

The distribution of the 20 amino acids is not uniform in all proteins. Nearly 40% by weight of fibroin and 25% by weight of collagen are accounted for by glycine. Fibroin is all rich in alanine (30% by weight). Serine and threonine predominate in casein and phosvitin. Collagen (in connective tissue), gliadin (in wheat) and zein (in corn) are rich in proline. Human serum albumin with 585 amino acid residues has only one tryptophan moiety.

The **pulse** are notable as they lack S-containing amino acid, methionine (Met) but contain good amount of the basic amino acid, lysine (Lys); whereas **cereals** lack lysine but have sufficient quantity of methionine. When combined, these make good the deficiency of each other through mutual supplementation and are therefore better utilized in human body.

Location in Proteins

Amino acids with uncharged polar side chains are relatively hydrophilic and are usually on the outside of the proteins, while the side chains on nonpolar amino acids tend to cluster together on the inside. Amino acids with acidic or basic side chains are very polar, and they are nearly always found on the outside of the protein molecules.

Physical Properties

Regarding their physical characteristics, the amino acids are colourless crystalline substances. The crystal form may vary from slender needles (tyrosine) to thick hexagonal plates (cystine). They may be either tasteless (tyrosine), sweet (glycine and alanine) or bitter (arginine). It is interesting to note that (S)-phenylalanine is bitter in taste whereas its enantiomer (R)-phenylalanine is sweet. They have high melting points (above 200°C) and often result in decomposition. Amino acids are soluble in polar solvents such as water and ethanol but they are insoluble in nonpolar solvents such as benzene and ether. Some amino acids like tryptophan, tyrosine, histidine and phenylalanine absorb ultraviolet rays at 260–290 mμ. This property enables the identification of not only these amino acids but also the proteins which contain them.

The order of hydrophilicity and hydrophobicity of various amino acids is as follows :

Hydrophilic (= polar) amino acids :

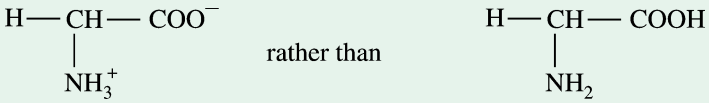
Tyr > Ser > Asp > Glu > Asn > Gln > Arg

Hydrophobic (= apolar) amino acids :

Phe > Ala > Val > Gly > Leu > Cys

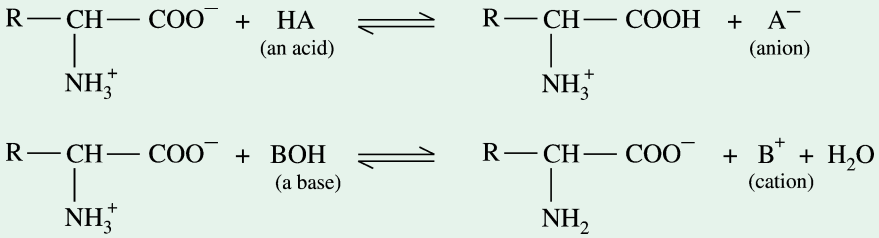
Although over 100 amino acids have been shown to be present in various plants and aminals, only 20 of them (L-isomers) are found as constituent of most proteins. These 20 amino acids of proteins are often referred to as **standard, primary or normal amino acids**, to distinguish them from others. The almost universal use of these 20 amino acids for the synthesis of various protein molecules is "one of nature's enigmatic rules." These have, therefore, rightly been called as the "**magic 20**", a phrase coined by Francis Crick.

Electrochemical Properties. All molecules possessing both acidic and basic groups might exist as uncharged molecules or in ionic form or as a mixture of both. Calculations have revealed that an aqueous solution of most amino acids, glycine for example, can have only one part of uncharged molecules in one lakh parts of the ionic form. Hence, it is more accurate to write the structure of glycine as



In the ionic form, the proton migrates from the carboxyl to amino group, thus producing carboxylate (COO^-) and ammonium (NH_3^+) ions.

Amino acids react with both acids and bases. Hence, they are **amphoteric** in nature.



Thus, in acid solution, the COO^- ion acquires a proton and the amino acid becomes an ammonium salt of the acid. Conversely, in alkaline solution, the NH_3^+ ion loses a proton and the amino acid becomes the anion of a salt. These reactions are of reversible nature and depend on the pH of the medium. Henceforth, amino acids serve as buffers and tend to prevent pH change when an acid or a base is added.

The α–COOH and α–NH_2 groups of amino acids are ionized in solutions at physiological pH, with the deprotonated carboxyl group (—COO^-) bearing a negative charge and the protonated amino group (—NH_3^+) a positive charge. An amino acid in its dipolar state is called a *zwitterion*. The dissociable α-COOH and α-NH_3^+ groups are responsible for the two characteristic pK values (pK_1 for —COOH and pK_2 for —NH_3^+) of α-amino acids. An amino acid with a third dissociable group in its side chain (*i.e.*, R group) exhibits an additional pK_R value (refer Table 9–3). However, pK values depend on temperature, ionic strength and the microenvironment of the ionizable group.

Table 9–3. Some properties associated with the standard amino acids

Amino acid	*Mr*	pK_1 (—COOH)	pK_2 ($—NH_3^+$)	p*KR* (*R group*)	*pI*	*Hydropathy index**	*Occurrence in proteins (%)†*
Nonpolar R groups							
Alanine	89	2.34	9.69		6.01	1.8	9.0
Valine	117	2.32	9.62		5.97	4.2	6.9
Leucine	131	2.36	9.60		5.98	3.8	7.5
Isoleucine	131	2.36	9.68		6.02	4.5	4.6
Proline	115	1.99	10.96		6.48	– 1.6	4.6
Phenylalanine	165	1.83	9.13		5.48	2.8	3.5
Tryptophan	204	2.38	9.39		5.89	– 0.9	1.1
Methionine	149	2.28	9.21		5.74	1.9	1.7
Polar, uncharged R groups							
Glycine	75	2.34	9.60		5.97	– 0.4	7.5
Serine	105	2.21	9.15	13.60	5.68	–0.8	7.1
Threonine	119	2.11	9.62	13.60	5.87	– 0.7	6.0
Tyrosine	181	2.20	9.11	10.07	5.66	– 1.3	3.5
Cysteine	121	1.96	8.18	10.28	5.07	2.5	2.8
Asparagine	132	2.02	8.80		5.41	– 3.5	4.4
Glutamine	146	2.17	9.13		5.65	– 3.5	3.9
Negatively-charged R groups							
Aspartate	133	1.88	9.60	3.65	2.77	– 3.5	5.5
Glutamate	147	2.19	9.67	4.25	3.22	– 3.5	6.2
Positively-charged R groups							
Lysine	146	2.18	8.95	10.53	9.74	– 3.9	7.0
Arginine	174	2.17	9.04	12.48	10.76	– 4.5	4.7
Histidine	154	1.82	9.17	6.00	7.59	– 3.2	2.1

* A scale combining hydrophobicity and hydrophilicity ; can be used to predict which amino acid will be found in an aqueous environment (-values) and which will be found in a hydrophobic environment (+ values).

† Average occurrence in over 200 proteins

Note that in nature, the commonest amino acid is alanine and the rarest amino acid is tryptophan.

(Adapted from Klapper MH, 1977)

As an instance, ionization of the amino acid, **alanine** (Fig. 9–4) is discussed.

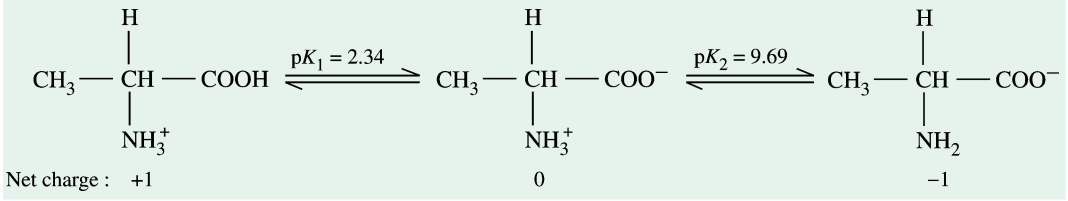


Fig. 9–4. The two dissociations of L-alanine

Titration involves the gradual addition or removal of protons. Fig. 9–5 shows the titration curve of the diprotic form of alanine. Each molecule of added base results in the net removal of one proton from one molecule of amino acid. *The plot has two distinct stages*, each corresponding to the removal of one proton from alanine. At low pH values, molecules of alanine bear a net charge of + 1 because both functional groups are protonated ; for example, at pH 0.35, 99% molecules are positively-charged and are present as $^{+}H_3N$—$CH(CH_3)$—COOH. At the midpoint in the first stage of titration, in which the —COOH group of alanine loses it proton, equimolar concentrations of proton-donor, $^{+}H_3N$—$CH(CH_3)$—COOH and proton-acceptor, $^{+}H_3N$—CH(CH3)—COO^{-} species are present. At the midpoint of a titration, the pH is equal to the pK_a of the protonated group being titrated. For alanine, the pH at the midpoint is 2.34, thus its —COOH group has a pK_a of 2.34.

The **pH and pK_a** are simply convenient notations for proton concentration and equilibrium constant for ionization, resepectively. The pK_a is a measure of the tendency of a group to give up a proton, with that tendency decreasing tenfold as the pK_a increases by one unit.

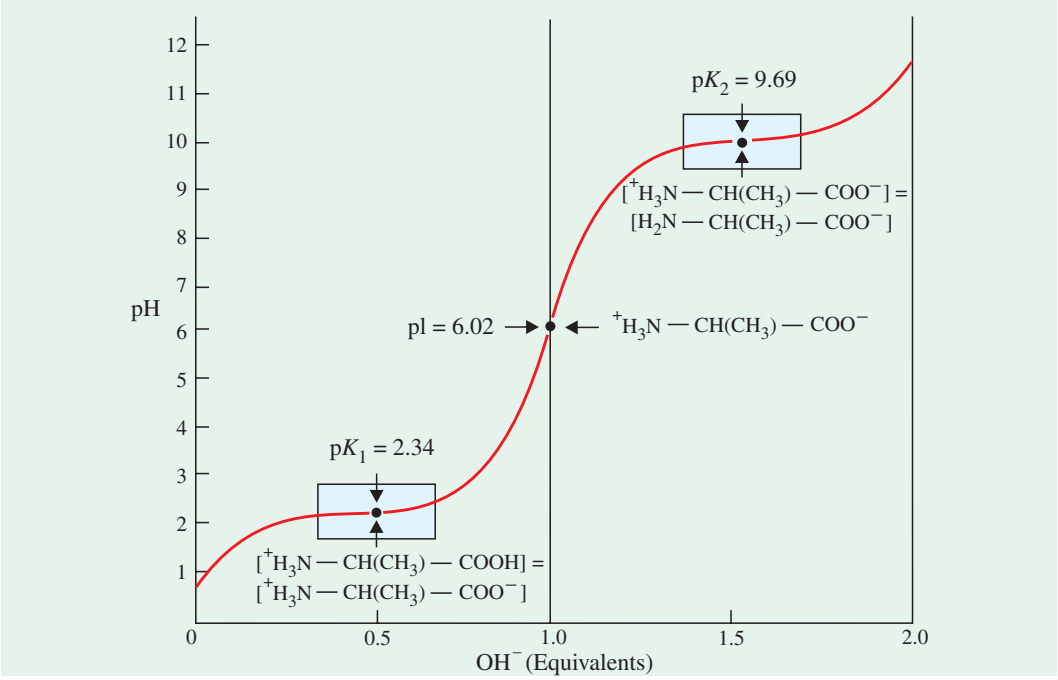


Fig. 9–5. The titration curve of alanine

The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centred about pK_1 = 2.34 and pK_2 = 9.69, indicate the regions of maximum buffering power.

As the titration proceeds, another important point is reached at pH **6.02.** Here there is a point of inflexion, at which removal of the first proton is essentially complete, and removal of the second has just begun. At this pH, the alanine is present as its dipolar form, fully ionized ($^{+}H_3N$—$CH(CH_3)$—COO^{-}) but with no *net* electric charge.

The second stage of titration corresponds to the removal of a proton from the $-NH_3^+$ group of alanine. The pH at the midpoint of this stage is **9.69** which is equal to the pK_a for the $-NH_3^+$ group. The titration is complete at a pH of **11.69**, at which point the molecules of alanine bear a net charge of −1 because both functional groups are deprotonated. And 99% of the molecules are negatively charged and are present as H_2N—$CH(CH_3)$— COO^{-}. It may, however, be noted that the characteristic pH value of 6.02, at which the average net charge is zero (called the isoelectric point or isoelectric pH), is midway between the two p*Ka* values [(2.35 + 9.69)/2].

The study of titration curve provides three useful informations :

1. It gives a quantitative measure of the pK_a of each of the two ionizing groups, 2.34 for the —COOH group and 9.69 for the —NH_3^+ group in case of alanine.
2. It gives information that alanine, say for example, has *two* regions of buffering power : one centred around the first pK_a of 2.34 and the second centred around pH 9.69.
3. It also provides information regarding the relationship between the net electric charge of the amino acid and the pH of the solution.

Classification

Three systems of classifying amino acids are in vogue.

A. On the basis of the composition of the side chain or R group. Threre are 20 different amino acids which regularly appear in proteins. These possess a side chain which is the only variable feature present in their molecules. The other features such as α-carbon, carboxyl group and amino group are common to all the amino acids. The common component of an amino acid appears in Fig. 9–6.

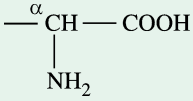


Fig. 9–6. Common component of an amino acid

Based on the composition of the side chain, the twenty amino acids, whose structure is shown in Fig. 9–7, may be grouped into following 8 categories (Fairley and Kilgour, 1966) :

(*i*) *Simple amino acids.* These have no functional group in the side chain, *e.g.*, glycine, alanine, valine, leucine and isoleucine.

(*ii*) *Hydroxy amino acids.* These contain a hydroxyl group in their side chain, *e.g.*, serine and threonine.

(*iii*) *Sulfur-containing amino acids.* These possess a sulfur atom in the side chain, *e.g.*, cysteine and methionine.

(*iv*) *Acidic amino acids.* These have a carboxyl group in the side chain, *e.g.*, aspartic acid and glutamic acid.

(*v*) *Amino acid amides.* These are derivatives of acidic amino acids in which one of the carboxyl group has been transformed into an amide group (—$CO.NH_2$), *e.g.*, asparagine and glutamine.

(*vi*) *Basic amino acids.* These possess an amino group in the side chain, *e.g.*, lysine and arginine.

(*vii*) *Heterocyclic amino acids.* These amino acids have in their side chain a ring which possesses at least one atom other than the carbon, *e.g.*, tryptophan, histidine and proline.

(*viii*) *Aromatic amino acids.* These have a benzene ring in the side chain, *e.g.*, phenylalanine and tyrosine.

The classification given above is only a practical one and can conveniently be followed. It does not, however, strictly delimit the various categories. For example, tryptophan may also be included under aromatic amino acids and similarly, histidine under basic amino acids.

B. On the basis of the number of amino and carboxylic groups. McGilvery and Goldstein (1979) have classified various amino acids as follows :

I. Monoamino-monocarboxylic amino acids :

1.	Unsubstituted	—	Glycine Alanine, Valine, Leucine, Isoleucine
2.	Heterocyclic	—	Proline
3.	Aromatic	—	Phenylalanine, Tyrosine,Tryptophan
4.	Thioether	—	Methionine

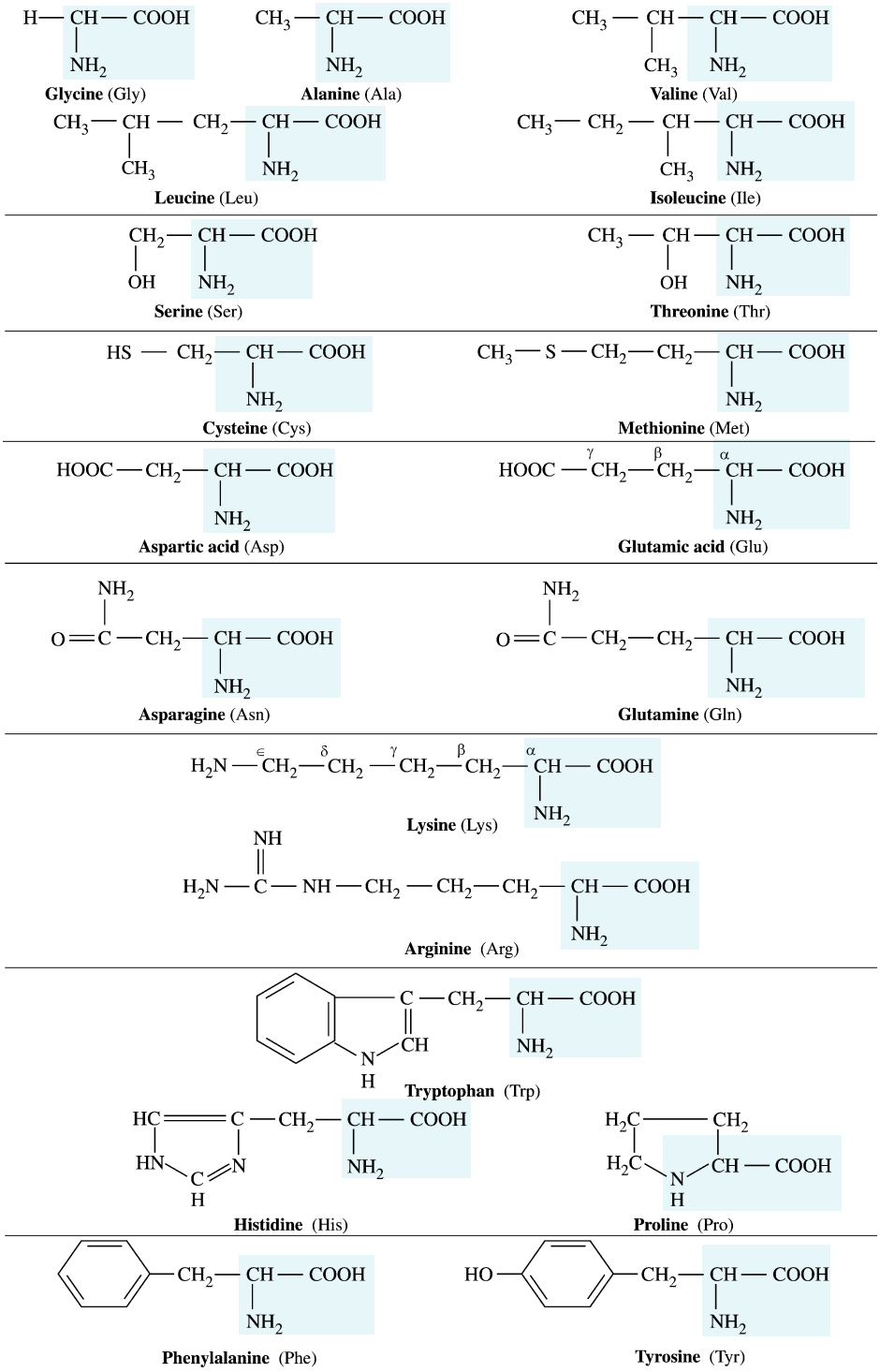


Fig. 9–7. The twenty amino acids ("Magic 20") found in proteins
The standard three-letter abbreviation for each amino acid is written in bracket.
The common component is shown in a blue enclosure.

5. Hydroxy — Serine, Threonine
6. Mercapto — Cysteine
7. Carboxamide — Asparagine, Glutamine

II. Monoamino-dicarboxylic amnino acids : Aspartic acid, Glutamic acid

III. Diamino-monocaryboxylic amino acids : Lysine, Arginine, Histidine

C. On the basis of polarity of the side chain or R group. A more meaningful classification of amino acids is, however, based on the polarity of the R groups present in their molecules, *i.e.*, their tendency to interact with water at biological pH (near pH 7.0). The R groups of the amino acids vary widely with respect to their polarity from totally nonpolar or hydrophobic (water-hating) R groups to highly polar or hydrophilic (water-loving) R groups. This classification of amino acids emphasizes the possible functional roles which they perform in proteins. The system recognizes following 4 categories :

I. Amino acids with nonpolar R groups. The R groups in this category of amino acids are hydrocarbon in nature and thus hydrophobic. This group includes five amino acids with aliphatic R groups (alanine, valine, leucine isoleucine, proline), two with aromatic rings (phenylalanine, tryptophan) and one containing sulfur (methionine).

1. *Alanine* (α-aminopropionate). It was first isolated in 1888 from silk fibroin where it occurs in abundance, along with glycine and serine. *It is the parent substance of all the amino acids except glycine*. The various amino acids may be derived from alanine by replacement of one or two H atoms of the methyl group present on α-carbon atom. Alanine is the least hydrophobic of the 8 nonpolar (=hydrophobic) amino acids because of its small methyl side chain.

2. *Valine* (α-aminoisovalerate). It is widely distributed but rarely occurs in amounts exceeding 10%. It is a branched chain amino acid and can be derived from alanine by the introduction of two methyl groups in place of two H atoms of the methyl group present on α-carbon atom.

3. Leucine (α-aminoisocaproate). Its presence in proteins was shown, first of all the amino acids, by Proust in 1819. It was first isolated from cheese, but later was obtained in purer form from hydrolysates of wool. Leucine is generally more prevalent than valine or isoleucine in proteins. It is also a branched chain amino acid and is the next higher homologue of valine. As such, it has much in common with valine from the viewpoint of reactivity and function. It is also one of the few amino acids that are sparingly soluble in water.

4. *Isoleucine* (α-amino-β-methylvalerate). It is an isomer of leucine and is also a branched chain amino acid. It was discovered somewhat late (in 1904) by Paul Ehrlich (LT, 1854-1915) because it was difficult to separate from leucine owing to their similar chemical composition and properties. This amino acid is one of the 5 common ones that have more than one asymmetric carbon atoms. It has 2 asymmetric carbon atoms and thus occurs in 4 stereoisomeric forms, but the natural form obtained from protein hydrolysates is the L-*erythro* isomer, where erythro signifies that the orientation of the methyl and amino groups is similar to that found in erythrose, with the groups concerned on the same side of the carbon chain.

Valine, leucine and isoleucine are quite alike chemically and possess a branched carbon chain. As most animals cannot synthesize the branched carbon chain, these amino acids are, therefore, indispensable or essential in the diet.

5. *Proline* (2-pyrrolidinecarboxylate). *It is present in almost all proteins that have been studied*, but zein from corn and gelatin are relatively high in proline content. It is a cyclized derivative of glutamic acid. Its α-amino group is not free but is substituted by a portion of its R group to yield a cyclic structure. The secondary amino group in proline is held in a rigid conformation. This reduces the structural flexibility of the protein because of the maximal steric hindrance of its side chain. Unlike other amino acids, proline is very soluble in ethanol. Moreover, it does not give many of the characteristic amino acid tests which are generally based on the presence of the unsubstituted α-amino group.

6. *Phenylalanine* (α-amino-β-phenylpropionate). It is one of the two common, clearly benzenoid amino acids. It is of widespread distribution and closely resembles tyrosine in structure. It also cannot be formed in animals because of its aromatic ring. Interestingly, (S)-phenylalanine is bitter whereas its stereoisomer, (R)-phenylalanine is sweet.

7. *Tryptophan* (α-amino-β-3-indolepropionate or β-indolylalanine). It was discovered in 1901 in the laboratory of F.G. Hopkins, one of the pioneers in biochemistry. *Tryptophan is the most complex amino acid found in proteins*. It is a heterocyclic amino acid and is a derivative of indole (with one N atom). Although widespread, it is usually limited in quantity. Tryptophan is the only amino acid of proteins which is nearly completely destroyed upon acid hydrolysis.

8. *Methionine* (α-amino-β-methylmercapto-butyrate). It is the only common amino acid possessing an ether linkage. Cereals have sufficient quantity of methionine whereas pulses lack in it. It is a methylation product of homocysteine. Apart from its role as a protein constituent and as an essential amino acid, methionine is also important as a donor of active methyl groups.

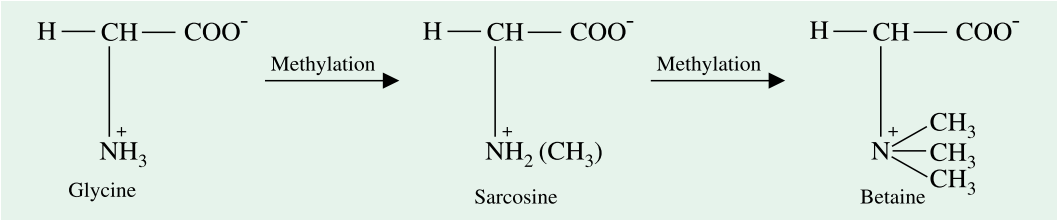
Recently (1996), **homocysteine** has been regarded in the same category as cholesterol, in inducing heart attacks. In 1980s, a study on 15,000 doctors in the U.S. called "physicians health study", showed that those with homocysteine higher than 12% had a risk of heart attack that was three-and-a-half times higher than those with lower levels. In fact, homocysteine, though not a well-known name, badly clogged the arteries. However, researchers have pointed out that high homocysteine levels are easier to treat than high cholesterol levels and can be effected by consuming abundant amounts of a B-vitamin, folic acid.

It may, however, be seen that **proline** (and also its 4-hydroxylated derivate, **hydroxyproline**) do not contain a primary amino group (—NH_2). The amino group in them is, in fact, utilized in ring formation. Both proline and hydroxyproline are actually the *substituted α-amino acids* as they have a secondary amino group $\left(\begin{matrix}C\\C\end{matrix}>NH\right)$ in their molecule (By contrast, all other 19 protein amino acids are designated as *α-amino acids*). Since they are an integral part of the protein molecule and exhibit similar reactions, they are usually inadvertently referred to as amino acids.

Some authors use the term **"imino acids"** for proline and hydroxyproline which is not correct, because an imine is characterized by a $>CH = NH$ and not a $\begin{matrix}C\\C\end{matrix}>NH$.

II. Amino acids with polar but uncharged R groups. The R groups of these amino acids are more soluble in water *i.e.*, more hydrophilic than those of the nonpolar amino acids because they contain functional groups that form hydrogen bonds with water. This category includes 7 amino acids, *viz.*, glycine, serine, threonine, tyrosine, cysteine, asparagine and glutamine. The polarity of these amino acids may be due to either a hydroxyl group (serine, threonine, tyrosine) or a sulfhydryl group (cysteine) or an amide group (asparagine, glutamine). The R group of glycine, a single hydrogen atom, is too small to influence the high degree of polarity of the α-amino and α-carboxyl groups.

9. *Glycine or glycocoll* (α-aminoacetate). It is the simplest amino acid and the only one lacking an asymmetric α-carbon atom. Where it is present in a protein, the minimal steric hindrance of the glycine side chain allows much more structural flexibility than the other amino acids. Glycine represents the opposite structural extreme in comparison to proline. It is, therefore, unique in being optically inactive. It is one of the first to be isolated from proteins and has a characteristic sweet taste, hence so named (*glykos*^{G} = sweet). Glycine is present in abundance in scleroproteins. This occurs in especially large amounts in the protein gelatin (25% of the total amino acids) and in silk fibroin (40%). In addition, many nonprotein compounds also contain glycine as a component part, for instance glutathione (a common tripeptide), sarcosine (= N-methylglycine) and hippuric acid (= benzoylglycine). Methylation of glycine yields sarcosine which is found in many peptides. Further methylation produces betaine, a substance which exists exclusively in the zwitterionic or salt form (for the same reason, zwitterionic formulae are sometimes called as betaine structures).



Glycine, the nonsymmetric amino acid is sometimes considered nonpolar. However, glycine's small R group (a hydrogen atom) exerts essentially no effect on the hydrophilicity of the molecule.

10. *Serine* (α-amino-β-hydroxypropionate; derived from the word, *serum*). It was first obtained from silk protein, sericin. It is found in all proteins but occurs in unusually high proportions in silk proteins, fibroin and sericin. This contains an alcoholic hydroxyl group which participates in ester formation. Esters with phosphoric acid have physiological significance as components of nucleotides and proteins. The unesterified serine residue appears to conduct special functions in many enzymes. The hydroxyl group on serine (and also on threonine) makes it much more hydrophilic and reactive than alanine and valine.

11. *Threonine* (α-amino-β-hydroxybutyrate). It was the last of the common amino acids to be discovered in proteins (Meyer and Rose, 1936). Its very name points towards its relationship with the sugar threose. Threonine is the next higher homologue to serine. It has two asymmetric carbon atoms and thus occurs in 4 stereoisomeric forms, namely D and L threonine and D and L allothreonine. The natural form is the L-*threo* isomer, hence its name. The *erythro* (unnatural) form of threonine is commonly referred to as allothreonine. Threonine is less abundant than serine in most proteins.

12. *Tyrosine* (α-amino β-(*p*-hydroxypyhenyl) propionate). It is the other aromatic amino acid normally found as a component of proteins. It was first isolated in 1857 from cheese and hence so named (*tyros*^{G} = cheese). The phenolic group of tyrosine is weakly acidic and loses its proton above pH 9. It is sparingly soluble in water. It is destroyed during acid hydrolysis of proteins but may be isolated after enzymic hydrolysis. Some proteins (*e.g.*, protamines) contain almost no tyrosine.

13. *Cysteine* (α-amino-β-mercaptopropionate). It is the sulfur analogue of serine and is one of the most reactive amino acids found in proteins. It contains sulfhydryl (SH) group which is quite reactive and esp., easily dehydrogenated. When it is dehydrogenated (*i.e.*, oxidized), two molecules join to form the amino acid cystine (see reaction on page 153). Fibrous proteins such as keratin from hair are especially rich in cystine (12%). *Cystine and its reduction product cysteine are together counted as one of the twenty amino acids.*

Cystine, from which the name 'cysteine' was later derived, was isolated in 1843 from urinary stones and was hence so named (*cystos*^{G} = bladder).

14. *Asparagine* (β-amide of α-aminosuccinate). It was the first amino acid to be discovered in 1806 and was first isolated in 1813 from a plant, *Asparagus*, for which it was so named. This is the β-amide of aspartic acid and has been isolated from proteins after enzymic hydrolysis. Asparagine has long been known as a constituent of plant tissues.

15. *Glutamine* (γ-amide of α-aminoglutarate). It is the homologue of asparagine. This is the γ-amide of glutamic acid and has been isolated from proteins after enzymic hydrolysis. Free glutamine is found in many animal and plant tissues. It is even less stable toward hydrolysis than is asparagine. It occurs widely in proteins.

Amino acids of categories I and II are jointly referred to as *neutral amino acids* because each one of them contains one acidic and one basic group.

III. Amino acids with negatively charged (= acidic) R groups. These are monoamino-dicarboxylic acids. In other words, their side chain contains an extra carboxyl group with a dissociable proton. The resulting additional negative charge accounts for the electrochemical behaviour of proteins. The two amino acids which belong to this category are aspartic and glutamic.

16. *Aspartic acid* (α-aminosuccinate). Its presence in protein was discovered by Ritthausen in 1868. It is the parent compound of asparagine.

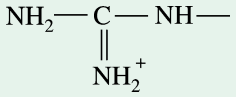
17. *Glutamic acid* (α-aminoglutarate). It is the homologue of aspartic acid. Its presence in protein was discovered also by Ritthausen in 1866. It was found in wheat gluten, hence so named. It is the parent compound of glutamine and occurs widely in proteins.

Both aspartic and glutamic acids in almost all proteins are linked to other amino acids only through their α-carboxyl and α-amino groups, not through their side chain carboxyl groups. Aspartic and glutamic acids are usually called aspartate and glutamate respectively to emphasize that their side chains are nearly always negatively charged at physiological pH.

IV. Amino acids with positively charged (=basic) R groups. These are diamino-monocarboxylic acids. In other words, their side chain contains an extra amino group which imparts basic properties to them. Lysine, arginine and histidine belong to this category.

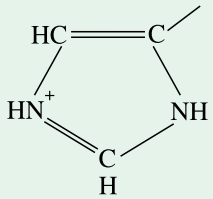
18. *Lysine* (α, ε-diaminocaproate). It has a second amino group at ε-position. Lysine is generally abundant in animal proteins but present in limited amounts in plant proteins, such as those of corn and wheat. Ionic charges are provided by protonation of the amino group of the e-carbon. It is nutritionally important since it is not synthesized by higher animals. Pulses contain good amount of lysine whereas cereals lack in it.

19. *Arginine* (α-amino-δ-guanidinovalerate). It is abundant in highly basic proteins of the cell nucleus (histones) and in sperm proteins, such as the protamines from salmon and herring sperm. In the latter, arginine may be as much as 80% of the total amino acids. It is found generally in all proteins, although in lesser amounts. Arginine is unique in possessing the guanidinium group, and due to which it is more strongly basic than lysine. Protonation of the guanidinium group provides ionic charges.



Gaunidinium group (charged form) of arginine

20. *Histidine* (α-amino-β-imidazolopropionate). ***Histidine is the last entry in the list of "magic 20".*** It contains a weakly basic imidazolium R group with p*K′* value of 6.0, and, therefore, is less than 10% protonated at pH 7. In many enzyme proteins, it functions as a proton donor or acceptor. Histidine is the only amino acid which has a proton that dissociates in the neutral pH range. It is this property which allows certain histidine residues to play an important role in the catalytic activities of



Imidazolium group (charged form) of histidine

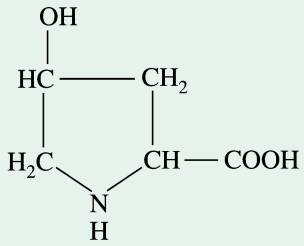
some enzymes. Histidine occurs in limited quantities in most proteins, but hemoglobin, protamines and histones contain relatively large amounts. It is also a component in the simple peptides carnosine and anserine. Histidine's basic properties are clearly marginal. The two ring nitrogen atoms have a relatively weak affinity for an H^+ and are only partly positive at neutral pH.

Carnosine ($C_9H_{14}O_2N_4$) is an optically active crystalline dipeptide ; m.p. 260°C ; found in muscle tissue.

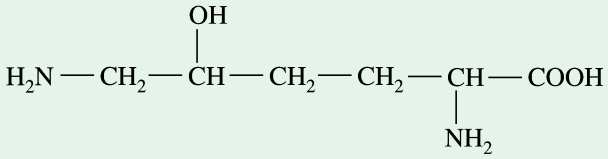
NONSTANDARD PROTEIN AMINO ACIDS

In addition to the above-mentioned twenty standard amino acids which are building blocks of

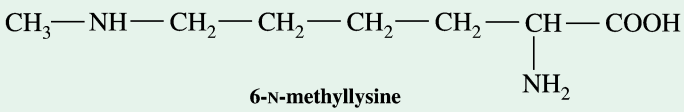
proteins and have a wide range of distribution in proteins, several other amino acids exists. These have a limited distribution but may be present in high amounts in a few proteins and hence deserve mention. As an example, **hydroxyproline** has a limited distribution in nature but constitutes as much as 12% of the composition of collagen, an important structural protein of animals. Similarly, **hydroxylysine** is also a component of collagen, where it accounts for about 1% of the total amino acids. **N-methyllysine** is found in myosin, a contractile protein of muscle. Another important nonstandard or less common amino acid is **γ-carboxyglutamate**, which is found in the blood-clotting protein, prothrombin as well as in certain other proteins that bind Ca^{2+} in their biological function.



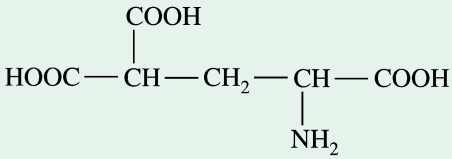
L-hydroxyproline
(erythro-4-hydroxy-L-proline)



L-hydroxylysine
(erythro-5-hydroxy-L-lysine)

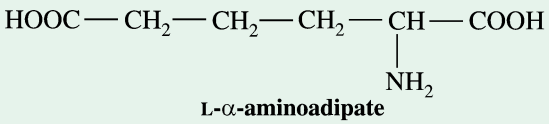


6-N-methyllysine

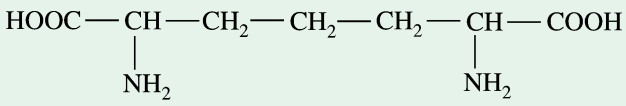


γ-carboxyglutamate

In protein from corn, **α-aminoadipate** has been detected. It has one CH_2 group more than glutamic acid. Another amino acid, **α-ε-diaminopimelate** has been found in bacterial protein.



L-α-aminoadipate



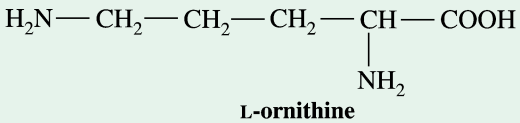
L, L-α, ε-diaminopimelate

NONPROTEIN AMINO ACIDS

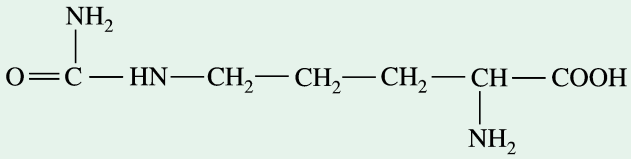
There are some 300 additional amino acids which are never found as constituents of proteins but which either play metabolic roles or occur as natural products.

Among the important nonprotein amino acids, which play metabolic roles, are L-ornithine, L-citrulline, β-alanine, creatine and γ-aminobutyrate. **L-ornithine** and **L-citrulline** occur in free state in the animal tissues and are metabolic intermediates in the urea cycle. L-ornithine possesses one CH_2 group less than its homologue, lysine. **β-alanine**, an isomer of alanine, occurs free in nature and also

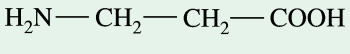
as a constituent of an important vitamin pantothenic acid and of coneyzme A. It is also found in the naturally occurring peptides, carnosine and anserine. The quaternary amine **creatine**, a derivative of glycine, plays an important role in the energy storage process in vertebrates where it is phosphorylated and converted to creatine phosphate. Lastly, γ-aminobutyrate is found in free form in the brain.



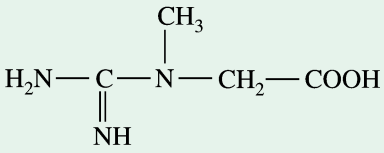
L-ornithine



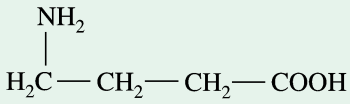
L-citrulline



β-alanine

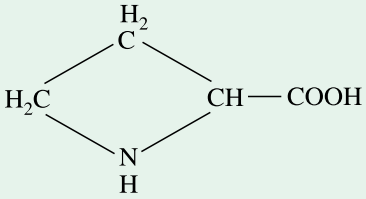


Creatine

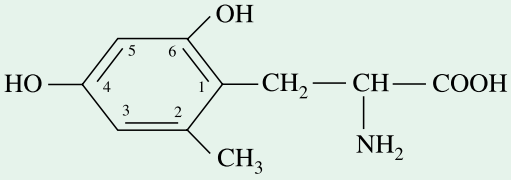


γ-aminobutyrate

Higher plants are especially rich in nonprotein amino acids. These nonprotein amino acids are usually related to the protein amino acids as homologues or substituted derivatives. They have a limited distribution, sometimes to a single species even. Thus, **L-azetidine-2-carboxylic acid**, a homologue of proline, accounts for 50% of the nitrogen present in the rhizome of Solomon's seal, *Polygonatum multiflorum*. **Orcylalanine** is found in the seed of cornocockle, *Agrostemma githago*. It may be considered as a substituted phenylalanine. Furthermore, in the toxic polypeptides of *Amanita phalloides*, in addition to **hydroxyleucine**, ***allo*-threonine** is also found.



L-azetidine-2-carboxylic acid

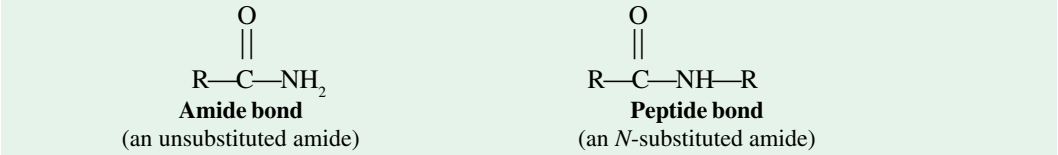


Orcyl-L-alanine
(4, 6-dihydroxy-2-methyl phenyl–L-alanine)

PEPTIDES

Peptide Bond

The amino acid units are linked together through the carboxyl and amino groups to produce the primary structure of the protein chain. The bond between two adjacent amino acids is a special type of amide bond, in which the hydrogen atom of amino ($-NH_2$) group is replaced by an R radical. Such a substituted amide bond is known as the **peptide bond**. And the chain, thus formed, by linking



together of many amino acid units is called a **peptide chain** (Fig. 9–8).

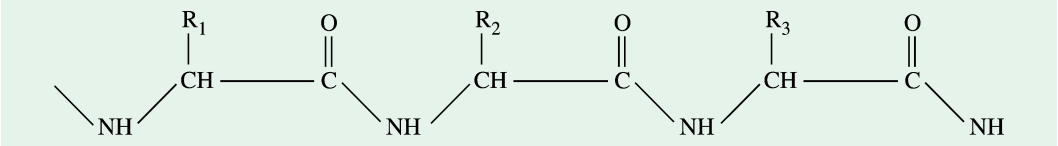


Fig. 9–8. Peptide chain

The characteristic structure of the peptide bond is shown is Fig. 9–9.

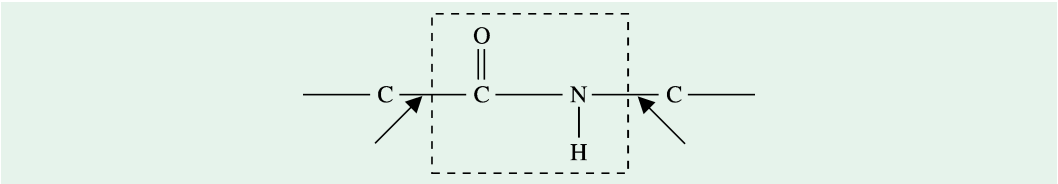


Fig. 9–9. Peptide bond

The peptide bond is shown enclosed in the dashed box. The four atoms (C, O, N, H) of the peptide bond form a rigid planar unit. There is no freedom of rotation about the C—N bond. On the contrary, the 2 single bonds (shown with arrows) on either side of the rigid peptide unit, exhibit a high degree of rotational freedom.

Each peptide chain is of considerable length and may possess from 50 to millions of amino acid units. Depending on the number of amino acid molecules composing a chain, the peptides may be termed as a **dipeptide** (containing 2 amino acid units), **a tripeptide** (containing 3 amino acid units) and so on. If a peptide is made up of not more than 10 amino acids, it is called an **oligopeptide ;** beyond that it is a **polypeptide**. Polypeptides when they are made up of over 100 amino acids are, sometimes, called as **macropeptides.** *Strictly speaking, the proteins are polypeptides with more than 100 amino acids.* All naturally-occurring important peptides, however, possess a shorter individual name, such as glutathione etc. Proteins differ widely in amino acid content. Various types of proteins in an organism may have varied amounts of a particular amino acid. Some amino acids are in abundance in one protein, may be in meagre amounts in others and may even be lacking in the rest. Tryptophan, for instance, lacks in certain proteins. However, *most of the proteins contain all the 20 amino acids.* As the number and manner in which the amino acids are grouped is highly variable, the number of proteins approaches almost to infinity.

Sometimes, the word **'proteinoids'** is used for short polypeptides containing up to 18 amino acids (Fried GH, 1990).

It is analogous to the indefinite number of words that can be formed with the 26 letters of English alphabet. While the words have to be restricted in length, there is no such restriction regarding the number of amino acids that may form a protein. According to an estimate given by Erlene B. Cunningham (1978), if each protein molecule were to consist of only 250 amino acid residues, the utilization of all the 20 different monomers would permit the formation of 10325 different protein molecules ! In reality, protein molecules often contain more than 250 aminoacyl units, and hence there is possibility for an even greater number of different protein molecules.

N- and C-terminals

Each amino acid in the chain is termed a *residue*. The two ends of the peptide chain are named as amino terminal and carboxyl terminal or simply as an N-terminal and C-terminal respectively. These two terminal groups, one basic and another acidic, are the only ionizable groups of any peptide chain

except those present in the side chain. The terminal amino acid with the free amino group is called as the *N-terminal amino acid* and the one with the free carboxyl group at the other end as *C-terminal amino acid*.

Representation of Peptide Chain

To fix the convention for representation of peptide structures in mind, it is helpful to imagine a rattlesnake moving from left to right across the page. The C-terminal residue forms its fangs and the N-terminal residue its rattle (Fig. 9–10).

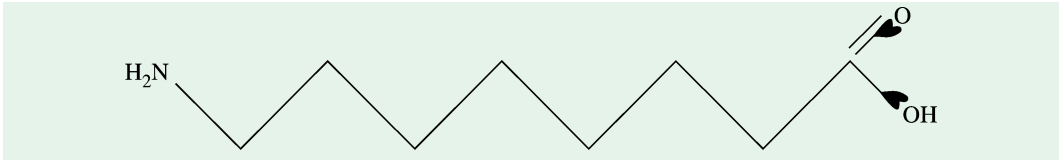


Fig. 9–10. Mnemonic device for peptide chain

Naming of Peptide Chain

In naming a polypeptide, the convention is that the N-terminal residue (which is shown at the left hand part of the structure) is written first and the C-terminal residue in the formation of each peptide

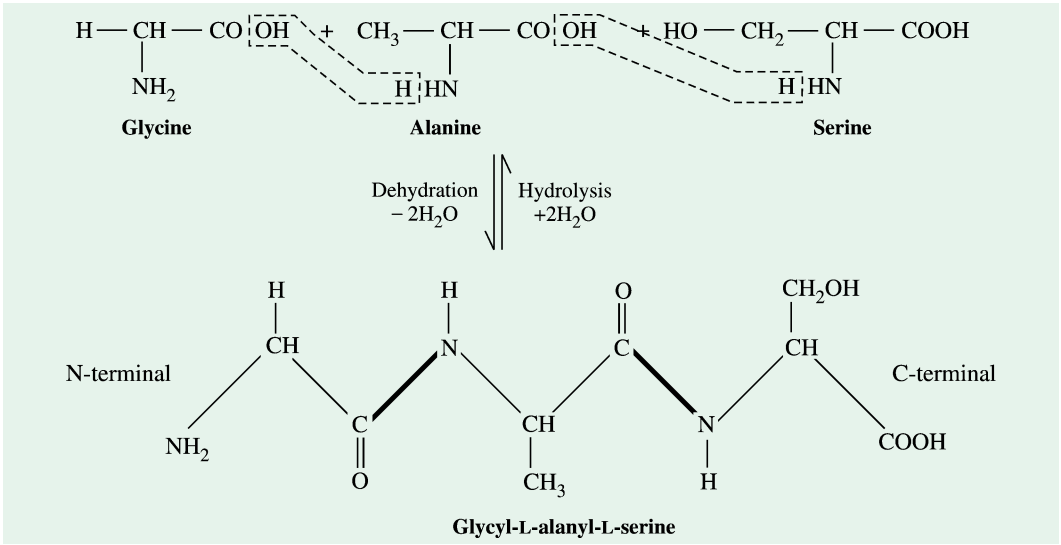


Fig. 9–11. Construction of a tripeptide chain from three different amino acids

The formation of each peptide bond (indicated by bold lines) involved splitting of a molecule of water.

Thus, the process of synthesis of proteins may be deemed as essentially a dehydration synthesis.

the end. The names of various intermediary amino acid residues are written in the same sequence as they are placed. Further, the names of all the amino acid residues, except the last one, are written by adding the suffix *-yl* because all these are the acyl groups. The name of the last amino acid, however, is written as such. For example, a tripeptide containing glycine, alanine and serine (structure shown in Fig. 9–11) is named as glycyl-L-alanyl-L-serine and abbreviated as Gly-Ala-Ser. If the sequence of amino acid in such a tripeptide is not known, the abbreviation would be (Gly, Ala, Ser), the parenthesis and commas indicating that only the compostiton of the tripeptide is known.

Similarly, glutathione or GSH (a tripeptide containing glutamic acid, cysteine and glycine) is named as γ-glutamyl-cysteyl-glycine and abbreviated as Glu-Cys-Gly. Glutathione is a naturally occurring and widely distributed polypeptide.

Determination of the Amino Acid Sequence of a Polypeptide

This can be explained by taking the example of a **dodecapeptide** whose composition was found to be Ala_2, Arg, Glu, Gly, Leu, Lys_2, Phe, Tyr_2, Val. It was determined that the N-terminal amino acid

of the dodecapeptide was valine and the C-terminal amino acid, leucine. Hydrolysis of the dodecapeptide by trypsin yielded four peptides whose structures were determined and found to be those given in A to D.

Tyr-Glu-Lys	Phe-Gly-Arg	Val-Lys	Ala-Tyr-Ala-Leu
A	**B**	**C**	**D**

Since valine was the N-terminus and leucine the C-terminus of the dodecapeptide, it is apparent that peptide C must represent the amino acid sequence at the N-terminal end, and peptide D the amino acid sequence at the C-terminal end of the dodecapeptide. To establish the order of the A and B peptides in the interior of the dodecapeptide, another sample of the dodecapeptide was hydrolyzed by *chymotrypsin*, the four peptides formed were sequenced, and their structures were found to be those given in **E** to **H**.

Ala-Leu	Glu-Lys-Ala-Tyr	Val-Lys-Phe	Gly-Arg-Tyr
E	**F**	**G**	**H**

The sequences Gly-Arg-Tyr in peptide **H** and Glu-Lys-Ala-Tyr in peptide **F** clearly establish that peptide **B** must precede peptide **A** in the dodecapeptide. Hence, the structure of the dodecapeptide is unambiguously determined to be that as shown belew :

Val-Lys-Phe-Gly-Arg-Tyr-Glu-Lys-Ala-Tyr-Ala-Leu.

Stereochemistry of Peptide Chains

All proteins are made of amino acids of L-configuration. This fixes the steric arrangement at the α-carbon atom. The dimensions of the peptide chain are known exactly. These have been depicted in Fig. 9–12.

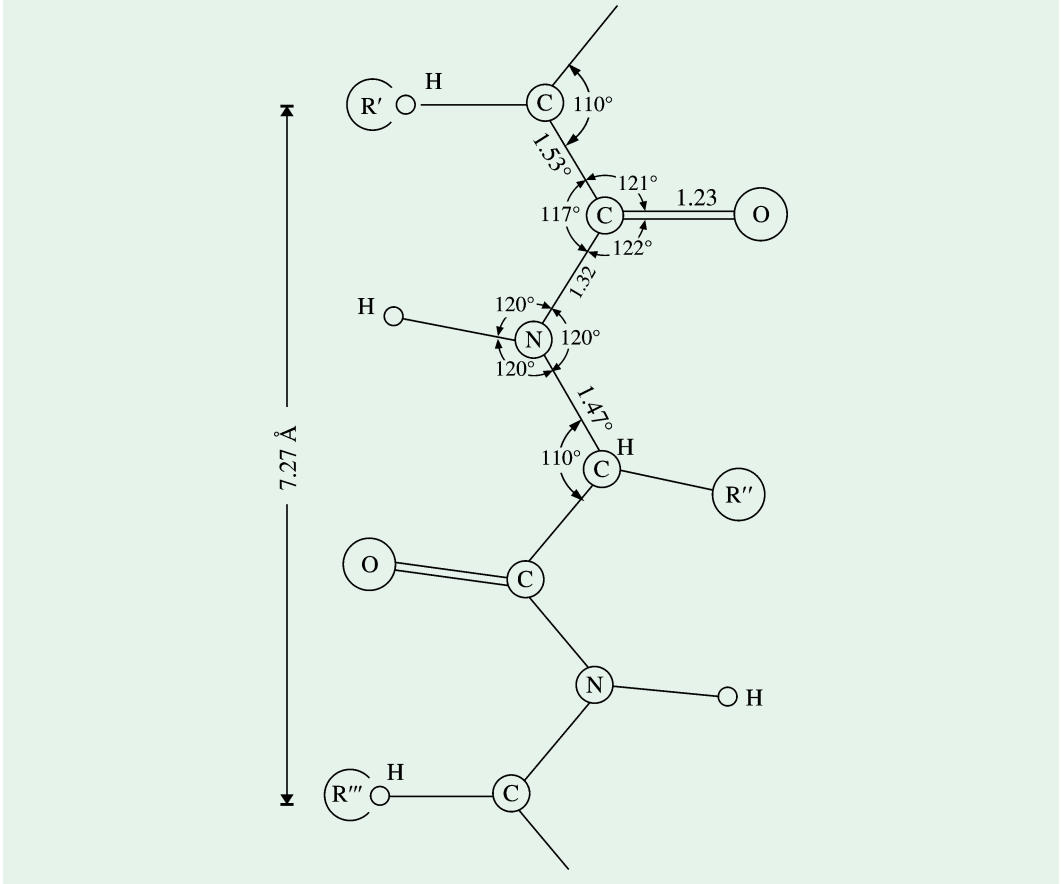
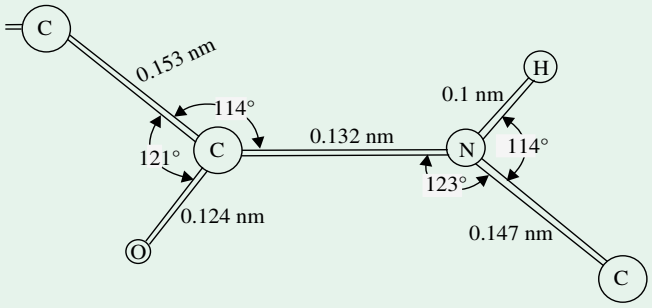


Fig. 9–12. Dimensions within the peptide chain

The peptide bond, which is an imide (substituted amide) bond, has a planar structure. The 6 atoms within the plane are related to each other by bond lengths and angles that vary little from amino acids residue to amino acid residue. Only 3 of these bonds are part of the peptide chain *per se :* the α-carbon to carbonyl carbon bond, the C–N bond, and the imide nitrogen to α-carbon bond. Since the double bond character of the C–N bond limits rotation about it, only the first and the last allow rotation. The rotation angles ϕ and ψ establish the relative positions of any 2 successive amide planes along the polypeptide chain. The α-carbon atoms can be thought of as shiwel centres for the adjacent amide planes.



Biological Roles

Peptides participate in a number of biological activities.

1. They serve as intermediates in the formation of **proteins.**

2. They appear as constituents in a group of compounds called **alkaloids.** Majority of these have been isolated from fungi, although they are also found in higher plants. Ergotamine is a peptide alkaloid from rye ergot and has pronounced pharmacological properties. The four components of this alkaloid are lysergic acid, alanine, proline and phenylalanine.

3. Many of them possess **antibacterial activities** and are usually present in fungi and bacteria. Penicillin G with 3 components (valine, cysteine and phenylacetic acid) is a common antibiotics.

4. Certain other peptides serve as **growth factors.** Folic acid, a water-soluble vitamin, is a noteworthy example of it (see Chapter 34). Another group of peptides serving as growth factor for a variety of microorganisms is streptogenins.

5. Higher animals do synthesize certain peptides serving as **hormones** (see Chapter 31).

6. Certain peptides like glutathione participate in controlling the **oxidation-reduction potential** of the cell. This may also serve as a key intermediate in electron-transfer systems.

7. A direct correlation has been found to exist between the amount of peptides in the urine of the patients and their **mental state of disturbance.** A group of Norwegian doctors have found an excess of peptides in urine specimens from patients with psychiatric disturbances. The peptides have been shown to induce in animals some of the conditions for the development of psychiatric disorders which lead to mania, depression or schizophrenia. The urine tests can, thus, indicate if a person is suffering from mental illness. The cause of this hypersecretion of peptides is not yet well established. It may exist from birth in organic genetic derangements or may be induced in healthy people by environmental factors.

CHEMICAL BONDS INVOLVED IN PROTEIN STRUCTURE

Given a full assortment of amino acids, a cell can synthesize all of its protein components. *Protein synthesis is a multiple dehydration process* (refer Fig. 9–11). The net structure of a protein becomes possible as a result of linking together of various amino acid units. The union of these amino acids to each other forming a chain and also among various amino acid residues of different chains involves various types of chemical bonds (Fig. 9–13). These are described below.

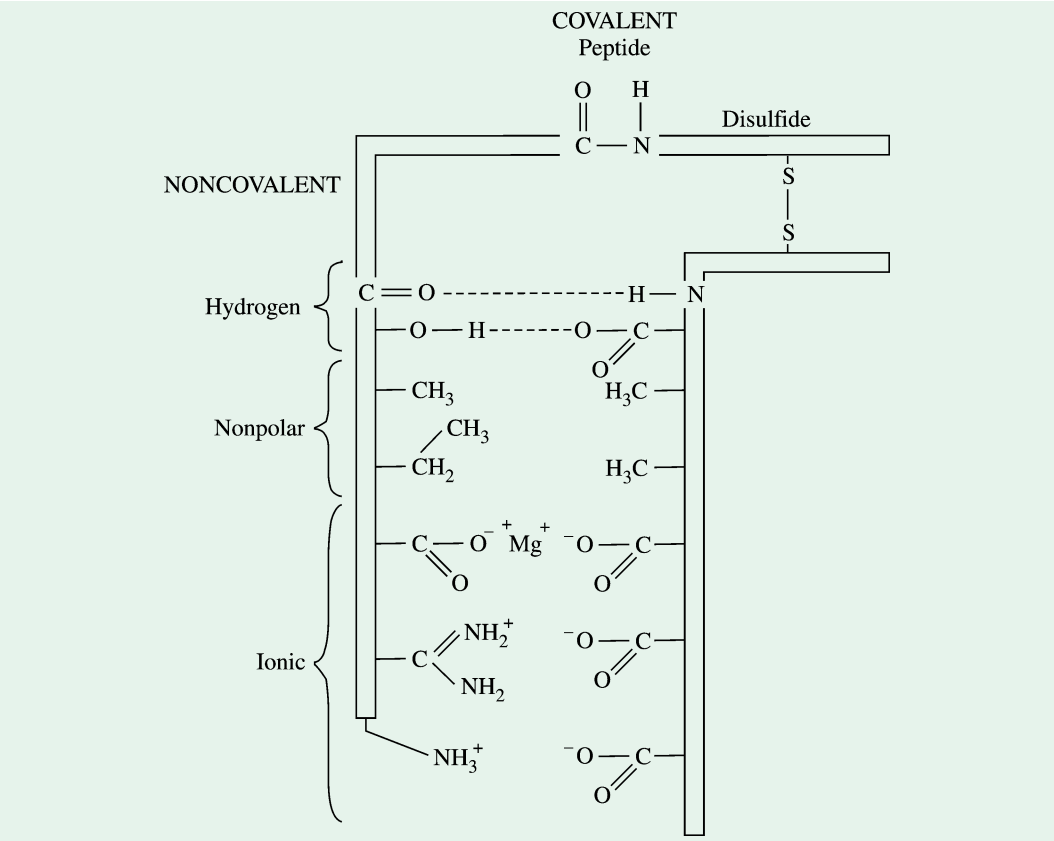


Fig. 9–13. Types of chemical bonds involved in protein structure

(*After Hartman and Suskind, 1969*)

A. Primary Bond

The principal linkage found in all proteins is the covalent **peptide bond,** —CO—NH— (**Fischer,** 1906). It is a specialized amide linkage where C atom of —COOH group of one amino acid is linked with the N atom of —NH_2 group of the adjacent amino acid. *Peptide bond is, in fact, the backbone of the protein chain.*

B. Secondary Bonds

Many of the properties of proteins, however, do not coincide with the linear chain structure, thus indicating that a variety of bonds other than the peptide exist in them. These secondary bonds, as they are called, hold the chain in its natural configuration. Some of the secondary bonds commonly found in proteins are listed below :

1. **Disulfide Bond** (—S—S—). In addition to the peptide bond, a second type of covalent bond found between amino acid residues in proteins and polypeptides is the disulfide bond, which is formed by the oxidation of the thiol or sulfhydryl (—SH) groups of two cysteine residues to yield a mole of **cystine**, an amino acid with a disulfide bridge (Fig. 9–15). In generalized form, the above reaction may be written as :

$$2R—SH + 1/2\ O_2 \rightleftharpoons R—S—S—R + H_2O$$

A disulfide bond is characterized by a bond strength of approximately 50 kilocalories per mole and a bond length of about 2Å between the two sulfur atoms. Hence, disulfide bond formation between 2 cysteine residues located some distance apart in the polypeptide chain requires that the polypeptide chain be folded back on itself to bring the sulfur groups close together. Although disulfide bridges are very strong, when compared to the strength of noncovalent bonds they are very short-range since, as

with all covalent bonds, even a slight extension breaks them completely. They, therefore, only stabilize the tertiary structure when it has reached something approximating to its final form.

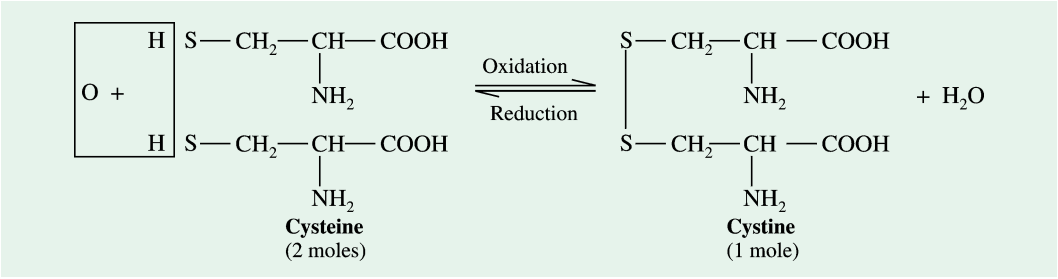


Fig. 19–14. Formation of cystine (= dicysteine)

Ocytocin (Fig. 9–15), a hormone stimulating the contraction of smooth muscles especially during childbirth, is an example where an internal disulfide bond is present between two cysteine units separated from each other in the peptide chain by 4 other amino acid units.

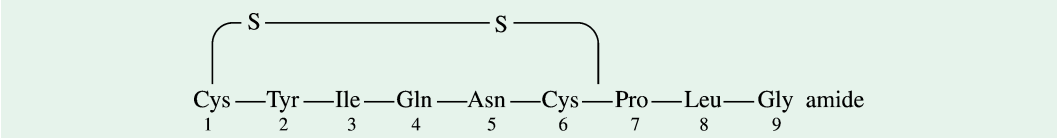
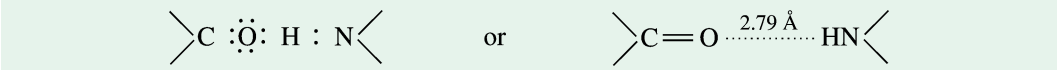


Fig. 9–15. Structure of ocytocin

Insulin (Fig. 9–16) is another excellent example where two peptide chains are linked together by 2 disulfide bonds. The presence of an internal disulfide bond in the glycyl (or A) chain between residues 6 and 11 is noteworthy.

These two chemical bonds, namely, peptide and disulfide, are relatively stable. Both these bonds, collectively or individually, maintain the linear form (or the *primary structure*) of the protein molecule.

2. Hydrogen Bond (>CO.....HN<). When a group containing a hydrogen atom, that is covalently-bonded to an electronegative atom, such as oxygen or nitrogen, is in the vicinity of a second group containing an electronegative atom, an energetically favourable interaction occurs which is referred to as a hydrogen bond. The formation of a hydrogen bond is due to the tendency of hydrogen atom to share electrons with two neighbouring atoms, *esp*., O and N. For example, the carbonyl oxygen of one peptide bond shares its electrons with the hydrogen atom of another peptide bond. Thus,



An interaction sets in between a C=O group and the proton of an NH or OH group if these groups come within a distance of about 2.8 Å. This secondary valence bond is symbolized by a dotted line, ... (The solid line, however, represents the normal covalent bond). The strength of the hydrogen bond is only 5 to 8 kilocalories per mole and is maximal when the bond is linear (Fig. 9–17). Hydrogen bonding between amides or peptides of the type depicted in Fig. 9–17 (*c*) plays an important role in stabilizing some conformations of the polypeptide chain.

Silk fibroin, composed mainly of glycine, alanine and serine units, is an example of the presence of hydrogen bonds involving the imide (>NH) and carbonyl (>C=O) groups of the peptide bonds. Here the hydrogen bonds link the vicinal peptide chains (refer Fig. 9–26).

In other proteins like **keratin of wool**, however, the hydrogen bonds link the side chains so that a single peptide chain is held in a coiled or helical form (refer Fig. 9–24).

Since the binding energy of a hydrogen bond amounts to only 1/10th of that of a primary valence, the hydrogen bonds are relatively weak linkages but many such bonds collectively exert considerable

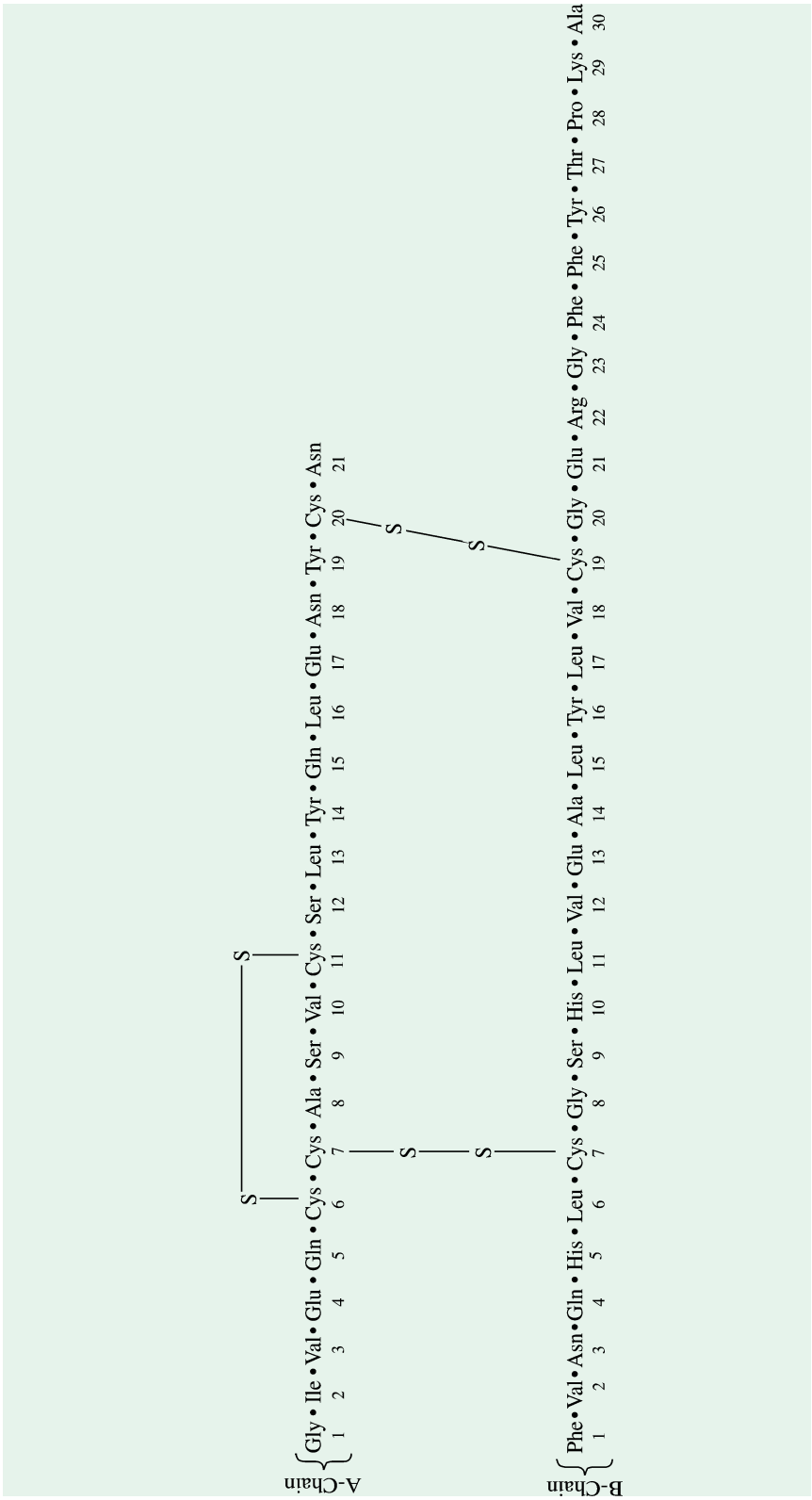


Fig. 9–16. Structure of bovine insulin

Numbers refer to specific amino acid residues. Glycine and phenylalanine are the N-terminal amino acids and the asparagine and alanine, the C-terminal amino acids. The A-chain of the insulin of man pig, dog, rabbit and sperm whale are identical. The B-chains of the cow, pig dog, goat and horse are identical. The amino acid replacements in the A-chain usually occur in positions 8, 9 and 10 (for details, refer Chapter : Animal Hormones)

force and help in maintaining the helical structure (or the *secondary structure*), characteristic of many proteins (**Mirsky** and **Pauling,** 1936).

3. Nonpolar or **Hydrophobic Bond.** Many amino acids (like alanine, valine, leucine, isoleucine, methionine, tryptophan, phenylalanine and tyrosine) have the side chains or R groups which are essentially hydrophobic, *i.e.*, they have little attraction for water molecules in comparison to the strong hydrogen bonding between water molecules. Such R groups can unite among themselves with elimination of water to form linkages between various segments of a chain or between different chains. This is very much like the coalescence of oil droplets suspended in water.

The association of various R groups in this manner leads to a relatively strong bonding. It also serves to bring together groups that can form hydrogen bonds or ionic bonds in the absence of water. Each type linkage, thus, helps in the formation of the other; the hydrophobic bonds being most efficient in this aspect. The hydrophobic bonds also play important role in other protein interactions, for example, the formation of enzyme-substrate complexes and antibody- antigen interactions.

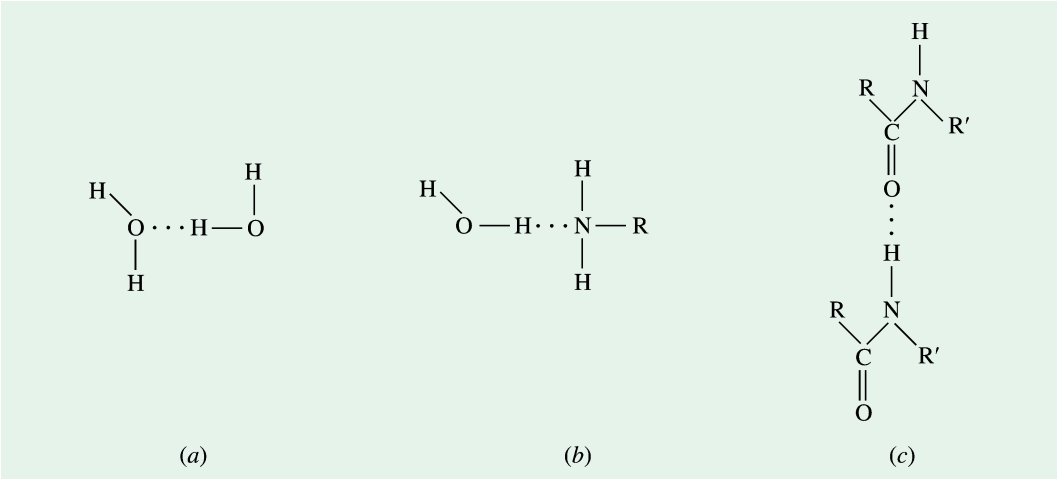


Fig. 9–17. Some common examples of hydrogen bonding

The hydrogen bonds (represented by a dotted line) have been shown between (*a*) two water molecules, (*b*) water and an amine, and (*c*) two amide groups.

4. Ionic or **Electrostatic Bond** or **Salt linkage** or **Salt bridge.** Ions possessing similar charge repel each other whereas the ions having dissimilar charge attract each other. For example, divalent cations like magnesium may form electrostatic bonds with 2 acidic side chains.

Another instance of ionic bonding may be the interaction between the acidic and basic groups of the constituent amino acids shown at the bottom of Fig. 9–13. The R groups of glutamic acid and aspartic acid contain negatively charged carboxylate groups, and the basic amino acids (arginine, histidine, lysine) contain positively charged amino groups in the physiological pH range. Thus, these amino acids contribute negatively charged and positively charged side chains to the polypeptide backbone. When two oppositely charged groups are brought close together, electrostatic interactions lead to a strong attraction, resulting in the formation of an electrostatic bond. In a long polypeptide chain containing a large number of charged side chains, there are many opportunities for electrostatic interaction. Intramolecular ionic bonds are rather infrequently used in the stabilization of protein structure but when they are so used, it is often with great effect. In fact, ionized groups are more frequently found stabilizing interactions between protein and other molecules. Thus, ionic bonds between positively charged groups (side chains of lysine, arginine and histidine) and negatively charged groups (COO^- group of side chain of aspartic and glutamic acids) do occur.

These ionic bonds, although weaker than the hydrogen bonds, are regarded as responsible for maintaining the folded structure (or the *tertiary structure*) of the globular proteins.

CHARACTERISTICS OF CHEMICAL BONDS

Table 9–4 lists some characteristics of the 2 types of chemical bonds : covalent and non cavalent. The strength of a bond or the **bond strength** can be measured by the energy required to break it, here in the table given in kilocalories per mole (kcal/mole). One kilocalorie is the quantity of energy needed to raise the temperature of 1,000 g of water by 1°C. An alternative unit in wide use is the kilojoule, kJ which is equal to 0.24 kcal. Individual bonds vary a great deal in strength, depending on the atoms involved and their precise environment, so that *the values are only a rough guide*. The **bond length** is the centre-to-centre distance between the two interacting atoms.

Table 9–4. Characteristics of chemical bonds

Bond Type	*Length (nm)**	*Strength (kcal/mole)†*	
		In vacuum	*In water*
Covalent	0.15	90	90
Noncovalent			
Ionic	0.25	80	3
Hydrogen	0.30	4	1
van der Wall's attraction (per atom)	0.35	0.1	0.1

* The length given here for a hydrogen bond is that between its two nonhydrogen atoms.

† Note that the aqueous environment in a cell will greatly weaken both the ionic and the hydrogen bonds between nonwater molecules.

PROTEIN CONFIGURATION

To describe a complicated macromolecule like protein, the biochemists have, for convenience, recognized 4 **basic structural levels of organization** of proteins based on the degree of complexity of their molecule (Fig. 9–18). These structural levels were first defined by Linderström–Lang and are often referred to as primary, secondary, tertiary and quarternary. Three of these structural levels (primary, secondary and tertiary) can exist in molecules composed of a single polypeptide chain, whereas the fourth (*i.e.*, quarternary) involves interactions of polypeptides within a multichained protein molecule. In mathematical term, these are also depicted as 1°, 2°, 3° and 4° respectively. The basic *primary structure* of a protein is relatively simple and consists of one or more linear chains of a number of amino acid units. This linear, unfolded structure or the polypeptide chain often assumes a helical shape to produce the *secondary structure*. This, in turn, may fold in certain specific patterns to produce the twisted three-dimensional or the *tertiary structure* of the protein molecule. Finally, certain other proteins are made up of subunits of similar or

IRVING GEIS

(LT, 1908 – 1997)

Irving Geis is well known for his lucid visualizations of molecular structures, particularly proteins and nucleic acids. These have appeared in *Scientific American* for the past thirty five years and in major chemistry, biology, and biochemistry textbooks. He is a co-author with R.E. Dickerson, Director of the Molecular Biology Institute of UCLA, of 3 books 1. Chemistry. Matter and the Universe 2. The Structure and Action of Proteins 30 Hemoglobin: Structure, Function, Evolution and Pathology.

In addition to drawing. painting, and writing, Irving Geis is a frequent lecturer at universities and medical schools on protein structure and function.

A recent Guggenheim fellowship made possible the assembly and cataloging of his drawings and paintings into The Geis Archives of molecular structure.

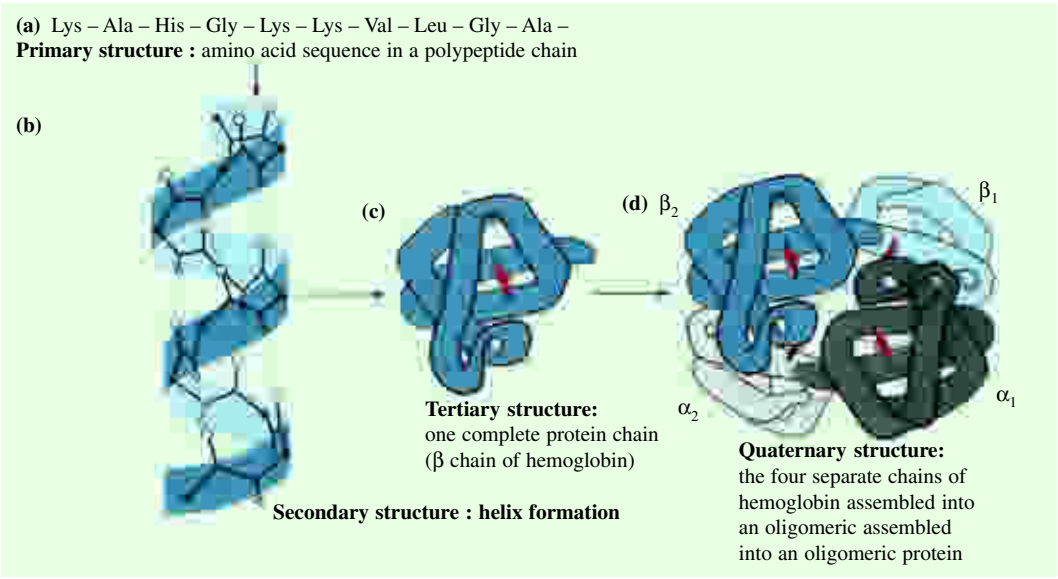


Fig. 9–18. The structural hierarchy in proteins

dissimilar types of the polypeptide chains. These subunits interact with each other in a specific manner to give rise to the so-called *quaternary structure* of the protein. This, in fact, defines the degree of polymerization of a protein unit.

1. Primary Structure : Amino Acid Sequence

The *primary structure* of a protein refers to the number and sequence of amino acids, the constituent units of the polypeptide chain. The main mode of linkage of the amino acids in proteins is the peptide bond which links the α-carboxyl group of one amino acid residue to the α-amino group of the other. The proteins may consist either of one or of more peptide chains.

Rigid and Planar Peptide Bond

Linus Pauling and Robert Corey, in the late 1930s, demonstrated that the α-carbons of adjacent

LINUS CARL PAULING (LT, 1901-1994)

Linus Pauling (with α-his helix ball-and-stick model in the photograph) , the son of a German father and English mother, was an American physical chemist. Although, he had initially an undistinguished school education, Pauling obtained his doctorate degree in 1925 from California Institute of Tech (Caltech). He was one of the exceptional men of his times and carried the message of Quantum Mechanics to the New World. His pioneering application of principles of Quantum Physics to explain chemical properties, his investigations of molecular structures of proteins and, above all, his work on the nature of chemical bond rank among the outstanding pieces of chemical research of the century. To the graduate students of Chemistry in the 1940s and 50s, his book, '*The Nature of Chemical Bond*' (1939) became a Bible. At age 30, Pauling had become a full professor in 1931 and Chairman of Chemistry and Chemical Engineering some 6 years later. With Robert Corey (LT, 1867-1971), he investigated amino acids and polypeptides. Around 1940s, he and his colleagues conducted researches on antibody-antigen reactions. Pauling was the first one to devise a scale for comparing the electronegativity of different elements.

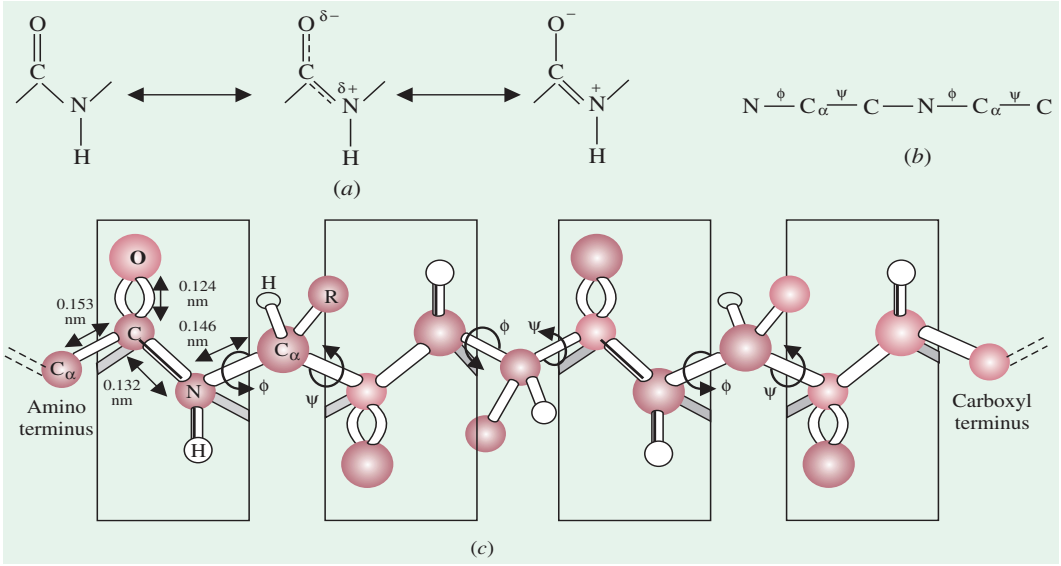
Pauling had his fair share of awards, which included, besides **Nobel Chemistry Prize (1954),** the prestigious Davy Medal of the Royal Society (1947), the Willard Gibbs Medal (1946) and the Presidential

Award for Merit for his distinguished work during World War II in the Explosives Division of the Natural Research Commission (USA).

Soon after war, however, Pauling spoke bitterly against the nuclear arms race and advocated multilateral disarmament and an end to atomic testing. Following the publication of his famous book, '*No More War !*' (1958), he then sent, in Jan'58, a petition signed by 11,021 scientists to the United Nations, urging an end to the testing of nuclear weapons. For all this, he earned the **1962-Nobel Peace Prize,** which was presented to him on October 10, 1963, the day a US-Soviet partial nuclear test ban treaty came into force. He also won the International Lenin Peace Prize. His political activities took a heavy toll of his time and energy. The result was that he did not produce scientific work of significance after 1951. Nevertheless, during his career in science spanning more than 60 years, Pauling published several books and more than 1,000 scientific papers. He also established **Linus Pauling Institute of Science and Medicine** at Palo Alto.

However, Pauling's fame had been tarnished in his later years by his strong advocacy of vitamin C (ascorbic acid) as a 'wonder drug' and helpful in maintaining youthness for a longer period. He also published a book, *'Vitamin C and the Common Cold'* (1971) wherein he claimed that by ingesting 1,000 mg daily of vitamin C (instead of the 60 mg daily minimum as recommended by the U.S. National Research Council), a person would catch 45% fewer colds and suffer 60% fewer days of illness. His findings were bitterly criticized experimentally by Terene Anderson of the University of Toronto in Canada. But this should not minimize the magnitude of his earlier contributions. Till today, he holds the distinction of being **the only person to have won two Nobel Prizes on his own.** His life fell into 3 distinct phases : the first evoked reverences, the second love and the third ridicule. Truly, Pauling ranks among the most versatile scientists of more than one generation of 20th century.

amino acids are separated by three covalent bonds, arranged C_α—C—N—C_α. They also demon strated that the amide C—N bond in a peptide is somewhat shorter (1.32— or 0.132 nm) than the C—N bond in a simple anine (1.49 Å or 0.149 nm) and that the atoms associated with the bond are coplanar. This indicated a **resonance** or partial sharing of two pairs of electrons between the carbonyl oxygen and the amide nitrogen [Fig. 9–19 (*a*)]. The oxygen has a partial negative charge and the nitrogen a partial positive charge, setting up a small electric dipole. The 4 atoms of the peptide group (C, H, O, N) lie in a single plane, in such a way that the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen are *trans* to each other. *Virtually, all peptide bonds in proteins occur in trans configuration.* From these studies, Pauling and Corey concluded that the amide C—N bonds are unable to rotate freely because of their **partial double-bond character.** The backbone of a polypeptide chain can thus be separated by substituted methylene groups —CH(R)— [Fig. 9–19 (*c*)]. The rigid peptide bonds limit the number of conformations that can be assumed by a polypeptide chain.



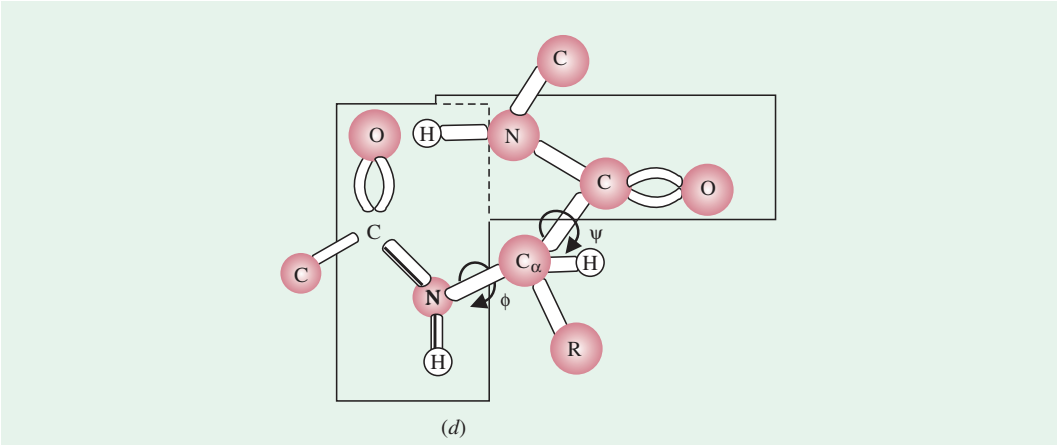


Fig. 9–19. The details of the planar peptide bond

(*a*) **The planar peptide group.** Note that the oxygen and hydrogen atoms are on opposite sides of the C—N bond. This is *trans* configuration.

(*b*) **The three bonds between the sequential C_α carbons in a polypeptide chain.** The N—C_α and C_α—C bonds can rotated, with bond angles designated ϕ and ψ respectively.

(*c*) **Limited rotation around two (N—C_α and C_α—C) of the three types of bond in a polypeptide chain.** The third type *i.e.*, C—N bonds in the planar polypeptide groups, which make one-third of all the backbone bonds, are not free to rotate.

(*d*) **The two coplanar peptide bonds flanking an α carbon.** By convention, ϕ and ψ are both defined as 0° when the two peptide bonds flanking an α carbon are in the same plane. In a fully stretched out polypeptide chain, ϕ = ψ = 180°.

However, rotation is permitted about the bond between the nitrogen and α-carbon atoms of the main chain (N—C_α) and between the α carbon and carbonyl carbon atoms (C_α—C). By convention, the degree of rotation at the N—C_α bond is called *phi* (ϕ) and that between C_α—C bond is called psi (ψ). Again, by convention, both ϕ and ψ are defined as 0° in the conformation in which the two peptide bonds connected to a single a carbon are in the same plane [Fig. 9–19 (*d*)]. In principle, ϕ and ψ can have any value between –180° and + 180°, but many values of ϕ and ψ are prohibited by steric interference between atoms in the polypeptide backbone and the amino acid side chains. The conformation in which ϕ and ψ are both 0° is prohibited for this reason.

Ramachandran Plot

Gopalasamudram Narayana Ramachandran (1922–) is an Indian biophysicist and crystallographer who, along with Gopinath Kartha, worked out the triple helical structure of collagen.

The conformation of the main polypeptide chain can be completely determined if the values ϕ and ψ for each amino acid residue in the chain are known. In a fully stretched polypeptide chain, ϕ = ψ = 180°. G.N. Ramachandran (1963) recognized that an amino acid residue in a polypeptide chain cannot have just *any* pair of values of ϕ and ψ. By assuming that atoms behave as hard spheres, allowed ranges of ϕ and ψ can be predicted and visualized in steric contour diagram called **Ramachandran plots.** Such a plot for poly-L-alanine (or *any amino acid except glycine and proline*) shows three separate allowed ranges (the screened regions in Fig. 9–19). One of them contains ϕ–ψ values that generate the antiparallel β sheet, the parallel Beet and the collagen helix. A second region has ϕ–ψ values that produce the right-handed α helix : a third, the left-handed α helix. Though sterically allowed, left-handed a helices are not found in proteins because they are energetically much less favoured.

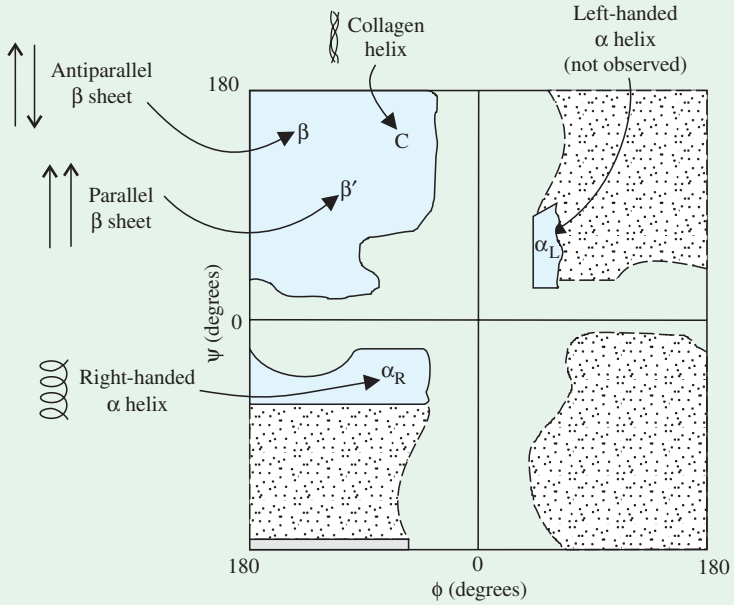


Fig. 9–20. A Ramachandran plot

The screened regions show allowed values of ϕ and ψ for L-alanine resideus. Additional conformations are accessible to glycine (dotted regions) because it has a very small side chain.

(*After Lubert Stryer, 1995*)

For **glycine**, these three allowed regions are larger, and a fourth appears (shown as dotted in Fig. 9–19) because a hydrogen atom causes less steric hindrance than a methyl group. *Glycine enables the polypeptide backbone to make turns that would not be possible with another residue.*

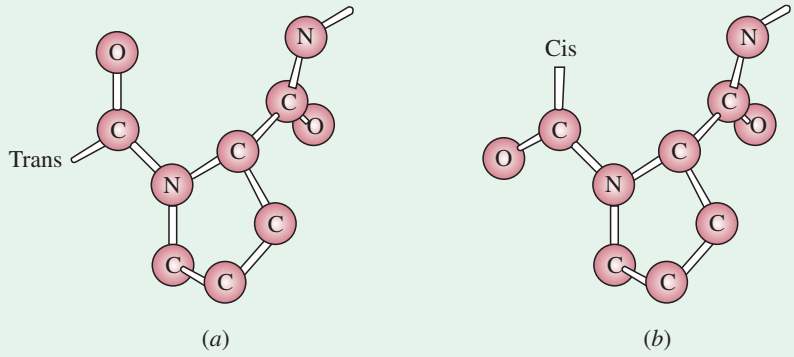


Fig. 9–20.(a) Formation of *trans* (A) and *cis* (B) peptide bonds between proline and its preceding residue in the polypeptide chain

The presence of a 5-membered ring in proline, which locks ϕ (the angle between N and C_α) at about – 65°, is its another distinctive feature.

Proline, too, is special [Fig. 9–20(a)]. The 5-membered ring of proline prevents rotation about the N—C_α bond, which fixes ϕ at about – 65°. Hence, a proline residue has a markedly restricted range of allowed conformations. The residue on the N-terminal side of a proline is also constrained because

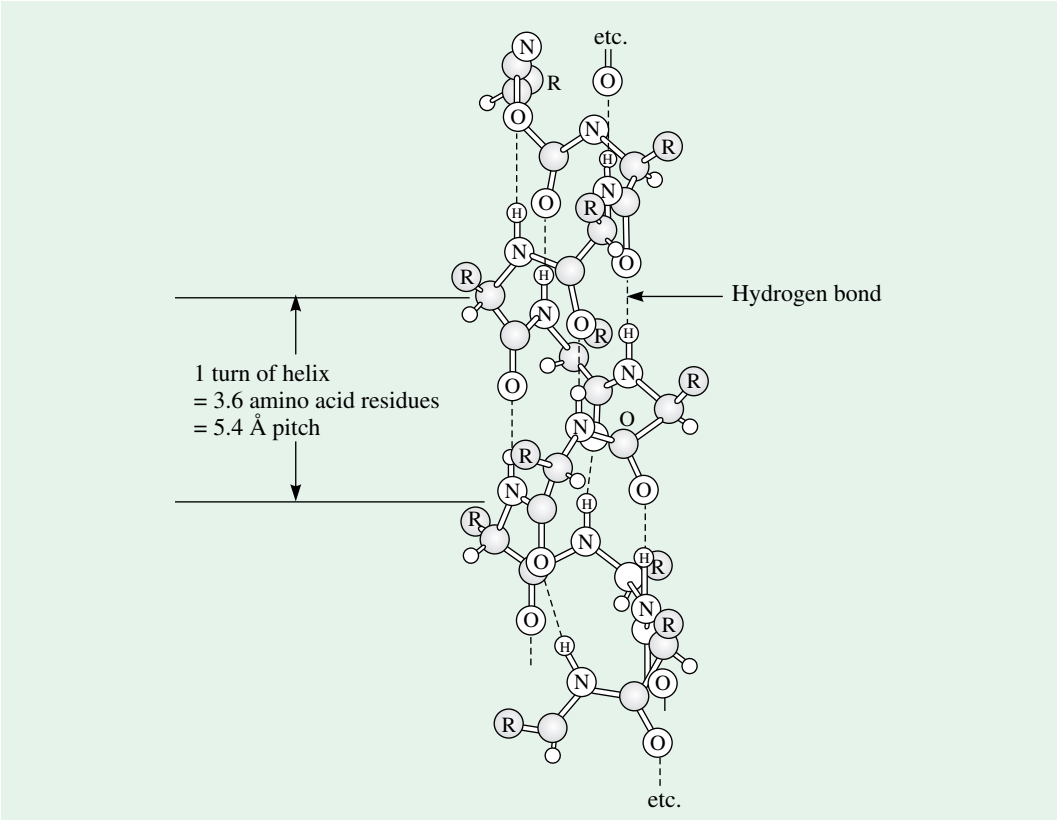


Fig.9–21. Portion of a right-handed α helix showing its three-dimensional (or '3-D') structure

All the shaded balls represent carbon atoms or residues (R) of amino acids.

(*Redrawn from Linus Pauling and Robert Corey, 1955*)

of steric hindrance imposed by the 5-membered ring. Proline also disfavours α-helix formation because it lacks an amide H atom for hydrogen bonding.

2. Secondary Structure : Helix Formation or Local Folding

If the peptide bonds were the only type of linkage present in proteins, these molecules would have behaved as irregularly coiled peptide chains of considerable length. But the globular proteins, however, do show some regular characteristic properties, indicating the presence of a regular coiled structure in these molecules. This involves the folding of the chain which is mainly due to the presence of hydrogen bonds. Thus, folding and hydrogen bonding between neighbouring amino acids results in the formation of a rigid and tubular structure called a **helix**. This constitutes the *secondary structure* of proteins, which refers to the steric or spatial relationship of amino acids that are near to each other in the amino acid sequence.

Based on the nature of hydrogen bonding (whether intramolecular or intermolecular), Pauling and Corey (1951) identified two regular types of secondary structure in proteins : alpha helix (α-helix) and beta pleated sheet (β-pleated sheet).

α-Helix

Early x-ray diffraction studies conducted by William Astbury in 1930s, of fibrous proteins such as hair and wool, showed a major **periodicity** or repeat unit of 5.0 to 5.5 Å, indicating some regularity in the structure of these proteins. A minor repeat unit of 1.5 Å was also observed. With x-ray diffraction studies, Pauling and Corey (1951) found that a polypeptide chain with planar peptide bonds would

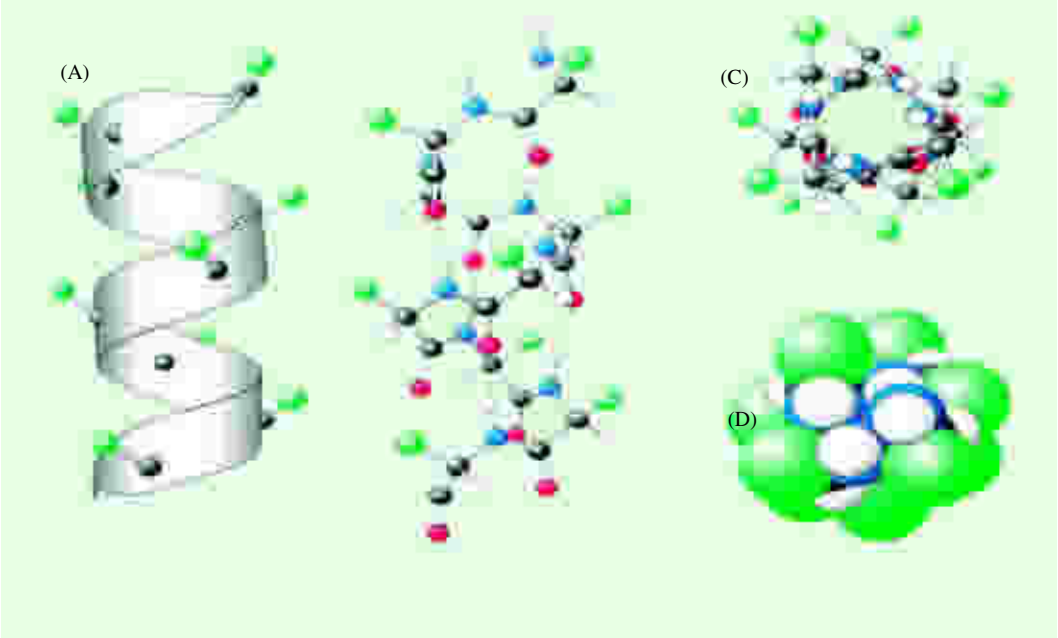


Fig. 9–22. Structure of the α helix (A)

A ribbon depiction with the α-carbon atoms and side chains (green) shown. (B) A side view of a ball-and-stick version depicts the hydrogen bonds (dashed lines) between NH and CO groups. for (C), consult back of figure (D) A space-filling view of part C shows the tightly packed interior core of the helix.

form a right-handed helical structure by simple twists about the α-carbon-to-nitrogen and the α-carbon-to-carboxyl carbon bonds. They called this helical structure as **α-helix.** The helix is so named because of the mobility of α-carbon atoms.

The α-helix is a rodlike structure. The tightly coiled polypeptide main chain forms the inner part of the rod, and the side chains extend outward in a helical array (Figs. 9–21 and 9–22). The α-helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain. 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. 9–23). Thus, all the main chain CO and NH groups are hydrogen-bonded.

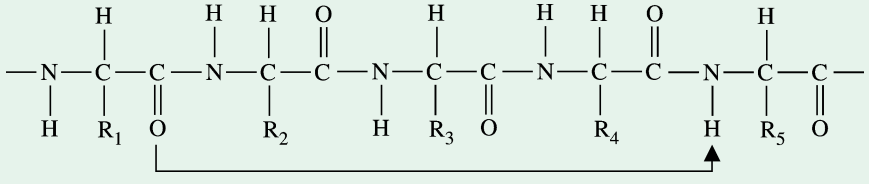


Fig. 9–23. Diagram showing that in the α-helix, the CO group of residue *n* is hydrogen-bonded to the NH group of residue (*n* + 4)

It is, thus, apparent that the α-helical structure depends on the intramolecular (= intrachain) hyrogen bonding between the NH and CO groups of peptide bonds. The hydrogen bonding occurs spontaneously and, as a result, a polypeptide can assume a rod-like structure with well-defined dimensions (Fig. 9–24). The α-helix (or α-conformation, as it is also called) has a pitch of 5.4 Å (= 0.54 nm) and contains 3.6 amino acids per turn of the helix, thereby giving a rise per residue of 5.4/3.6 = 1.5 Å (= 0.15 nm), which is the *identity* period of α-helix. The amino acid residues in an α-helix have conformations with $\phi = -60°$ and $\psi = -45°$ to $-50°$.

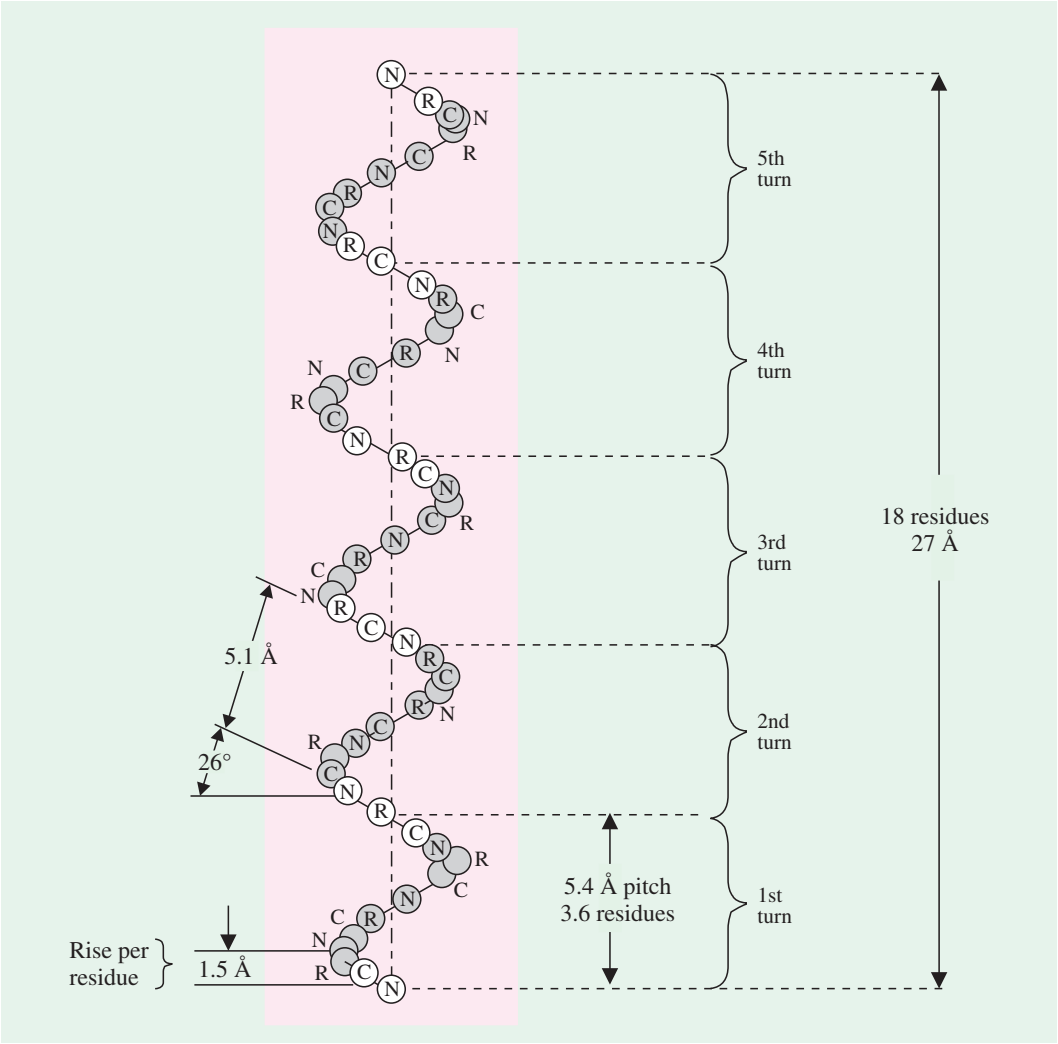


Fig. 9–24. Average dimensions of an α-helix

Note the average dimensions of an α-helix. The letters C and N represent the carbon and nitrogen atoms respectively in the peptide bonds. The letter R represents an α-carbon with a side chain attached. The shaded circles represent atoms back of the plane of the paper while the white circles indicate atoms above the plane.

(*Redrawn from Linus Pauling and Robert Corey, 1955*)

Fig. 9–25. Two types of helices

Right-and left-handed helices are related to each other in the same way as right and left hands.

A helix can be right-handed (clockwise) or left-handed (anticlockwise) (refer Fig 9–25) ; *α-helices of known polypeptides* (*i.e.*, *L-amino acids*) *are right-handed*. Biologically functional proteins do not usually exhibit cent per cent α-helical structure. Some have a high percentage of their residues in α-helical structures, *e.g.*, myoglobin and hemoglobin ; others have a low percentage, *e.g.*, chymotrypsin and cytochrome *C*. A very interesting feature of the α-helix, besides the periodicity, is the fact that the carbonyl group (>CO═O) of every peptide bond is in a position to form a hydrogen bond with the >N—H group of the peptide bond in the next turn of the helix, thereby contributing to the ***stability*** of the α-

helix. Long polypeptides chains in the α-helical conformation, forming structures are reminiscent of a multistranded rope. The protofibril of hair, for example, is 20 Å in diameter and appears to be made up of 3 right-handed helices wound about each other to form a left-handed supercoil.

The model-building experiments have shown that an α-helix can form with either D- or L-amino acids. However, the residues must all be of one stereoisomer ; a d-amino acid will disrupt any regular structure consisting of L-amino acids, and vice versa. The preferential formation of such a helix over many other possible conformations in nature is due, in part, to the fact that it (α-helix) makes optimal use of internal hydrogen bonds.

Many globular proteins contain short regions of such α-helices, and those portions of a transmembrane protein that cross the lipid bilayer are usually α-helices because of the constraints imposed by the hydrophobic lipid environment. In aqueous environment, an isolated α-helix is usually not stable on its own. Two identical α-helices that have a repeating arrangement of nonpolar side chains, however, will twist around each other gradually to form a particularly stable structure known as a coiled coil. Long rod-like *coiled coils* are found in many fibrous proteins, such as intracellular α-keratin fibres that reinforce skin and its appendages.

α-helix occurs in the protein **α-keratin**, found in skin and its appendages such as hair, nails and feathers and constitutes almost the entire dry weight of hair, wool, feathers, nails, claws, quills, scales, horns, hooves, tortoise shell, and much of the outer layer of skin. The basic structural unit of α-keratin usually consists of 3 right-handed helical polypeptides in a left-handed coil that is stabilized by crosslinking disulfide bonds (Fig. 9–26).

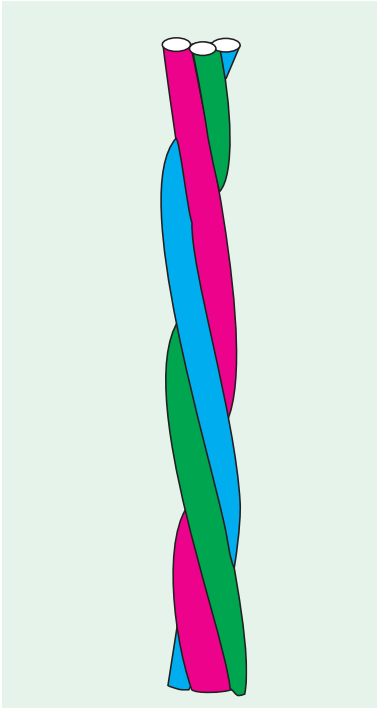


Fig. 9–26. Schematic of the 3 α-helical coils in keratin

Destabilization of α-helical conformation by certain amino acid residues (Table 9–5) can occur in various ways :

Table 9–5. Amino acids affecting α-helical structure

Destabilize α-helix		*Create bends in α-helix*
Aspartic acid	Glycine	Proline
Glutamic acid	Serine	Hydroxyproline
Lysine	Isoleucine	
Arginine	Threonine	

1. A prolyl reisdue has its α-N atom in a rigid ring system and cannot participate in α-helical structure ; instead, it creates a sharp bend in the helix.
2. A sequence of aspartyl and/or glutamyl residues can destabilize α-helical structure because the negatively-charged side chains repel one another (electrostatic repulsion), and the forces of repulsion are greater than those of hydrogen bonding.
3. A cluster of isoleucyl residues, because of steric hindrance imposed by their bulky R groups, also disrupts helical conformation.
4. Glycine, with a small hydrogen atom as an R group, is another destabilizer. The lack of a side chain on glycine allows for a great degree of rotation about the amino acid's α carbon ; hence, conformations other than a helical bond angles are possible.

β-Pleated Sheet

Pauling and Corey (1953) identified a second type of repetitive, minimum-energy or stable conformation, which they named **β-pleated sheet** (β because it was the second structure they elucidated, the α-heilx having been the first). The formation of β-pleated sheets depends on ***intermolecular*** (= interchain) hydrogen bonding, although intramolecular hydrogen bonds are also present. The pleated sheet structure is formed by the parallel alignment of a number of polypeptide chains in a plane, with hydrogen bonds between the >C = O and —N—H groups of adjacent chains. The R groups of the constituent amino acids in one polypeptide chain alternately project above and below the plane of the sheet, leading to a two-residue repeat unit. The β sheet structures are quite common in nature and are favoured by the presence of amino acids, glycine and alanine. Silk and certain synthetic fibres such as nylon and orlon are composed of β-structures.

The β-pleated sheet differs markedly from the rodlike α-helix :

1. A polypeptide chain in a β-pleated sheet, called a β-strand, has *fully extended conformation*, rather than being tightly coiled as in the α-helix.
2. The *axial distance* between adjacent amino acids in β-pleated sheets is 3.5 Å, in contrast with 1.5 Å for the α-helix.
3. β-sheet is stabilized by hydrogen bonds between NH and CO groups in *different* polypeptide strands, whereas in the α-helix, the hydrogen bonds are between NH and CO groups in the same strand.

There are two types of β-pleated sheet structures. If the N-terminal ends of all the participating polypeptide chains lie on the same edge of the sheet, with all C-terminal ends on the opposite edge, the structure is known as a **parallel β-pleated sheet.** In contrast, if the direction of the chains alternates so that the alternating chains have their N-terminal ends on the same side of the sheet, while their C-terminal ends lie on the opposite edge, the structure is known as the **antiparallel β-pleated sheet**. In other words, the hydrogen-bonded neighbouring polypeptides are aligned in the same N-to-C terminus direction in the parallel pleated sheets and in the opposite N-to-C direction in the antiparallel pleated sheets. Both parallel and antiparallel β-pleated sheets have similar structures, although the repeat period is shorter (6.5 Å 0.65 nm) for the parallel conformation in comparison to antiparallel conformation (7 Å or 0.7nm). Pleated-sheet structures can also be formed from a single polypeptide chain if the chain repeatedly folds back on itself. Although a β-pleated sheet structure is usually associated with structural proteins, it is also known to occur in the 3-dimensional structures of certain globular proteins, *e.g.*, the enzymes lysozyme and carboxypeptidase A.

Silk fibroin is one example of a protein that has the antiparallel pleated sheet structure. It is a member of a class of fibrillar proteins called α-keratins. Silk fibroin (Fig. 9–27) is composed mainly

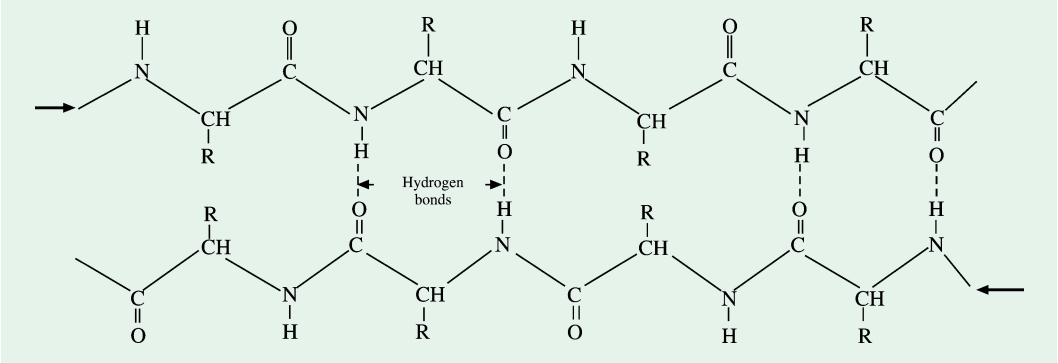
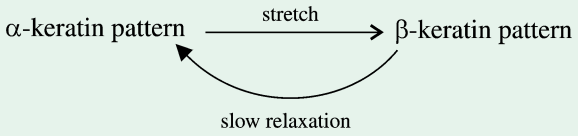


Fig. 9–27. A portion of two chains of silk fibroin

Note that the two chains runs in opposite direction.

of glycine, alanine and serine units linked together by the peptide bonds. Glycine constitutes approximately 45% of the total amino acid residues, and alanine plus serine compose another 42%. Hence, the R groups extending above and below the plane of the pleated sheet are small and allow the pleated sheets to stack. Here the two chains which run in opposite directions are linked by hydrogen bonds (shown by dotted lines). It may be noted that each of the two chains can also form hydrogen bonds with still other protein and so on. Very large protein aggregates are, thus, formed provided the R groups are relatively small. If the R groups are too large, the hydrogen bonding may not occur as the chains are held far apart for union. An interesting phenomenon occurs when hair or wool (α-keratins) is treated with moist heat and stretched because β-keratin structures with parallel β pleated sheets are produced.



Random Coil

Besides the α-helix and β-pleated sheet structures of proteins, which were recognized by Pauling and Corey in 1950s, there also exist a third type of secondary structure in proteins called **random coil**. When a polypeptide contains adjacent bulky residues such as isoleucine or charged residues such as glutamic acid and aspartic acid, repulsion between these groups causes the polypeptide to assume a random coil configuration (Fig. 9–28).

Thus, we see that the R groups distributed along the polypeptide backbone determine the secondary structure adopted by different portions of the polypeptide (α-helix or β-pleated sheet), or the lack of a well-defined structure (random coil).

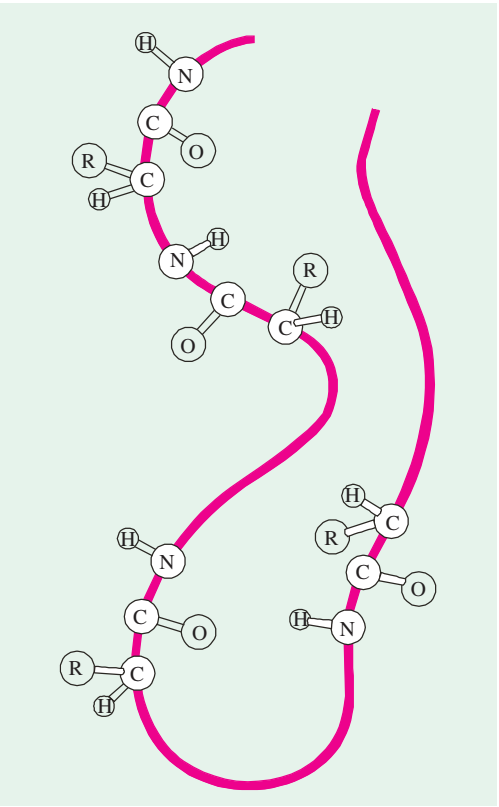


Fig. 9–28. A random coil

The backbone of the polypeptide is drawn, but only a few amino acid residues are shown.

(*Adapted from Francisco J. Ayala and John A. Kiger Jr., 1980*)

Other Secondary Structures

The α-helix and β-conformation are the major repetitive secondary structures easily recognized in a wide variety of proteins. The **γ-helix**, however, is an example of a highly hydrogen-bonded structure that is insufficiently stable to be used in proteins because of the lack of interatomic contacts. In addition to the α-helix and β-pleated sheet structures, other repetitive structures also exist in one or a few specialized proteins. Two such structures (β-turn and collagen triple helix) are described below :

β Turn or β Bend or Hairpin Bend

Most proteins have compact, globular shapes owing to reversals in the direction of their polypeptide chain. β turn (Fig. 9–29) is often found where a polypeptide chain abruptly reverses direction. β turns usually connect the ends of two adjacent segments of an antiparallel β pleated sheet, hence their name. The structure is a tight turn (~180°) involving 4 amino acids. The essence of this hairpin turn is that the CO

group of residue *n* of a polypeptide is hydrogen-bonded to the NH group of residue *n* + 3. In other words, the hydrogen bond between the peptide groups of the first and fourth amino acid residues is involved in bend formation. Glycine and proline residues often occur in β turns ; the former because it is small and flexible and the latter because peptide bonds involving the imino nitrogen of proline readily assume the *cis* configuration, a form that is particularly amenable to a tight turn. β turns are often found near the surface of a protein. They are also called as **reverse turns.**

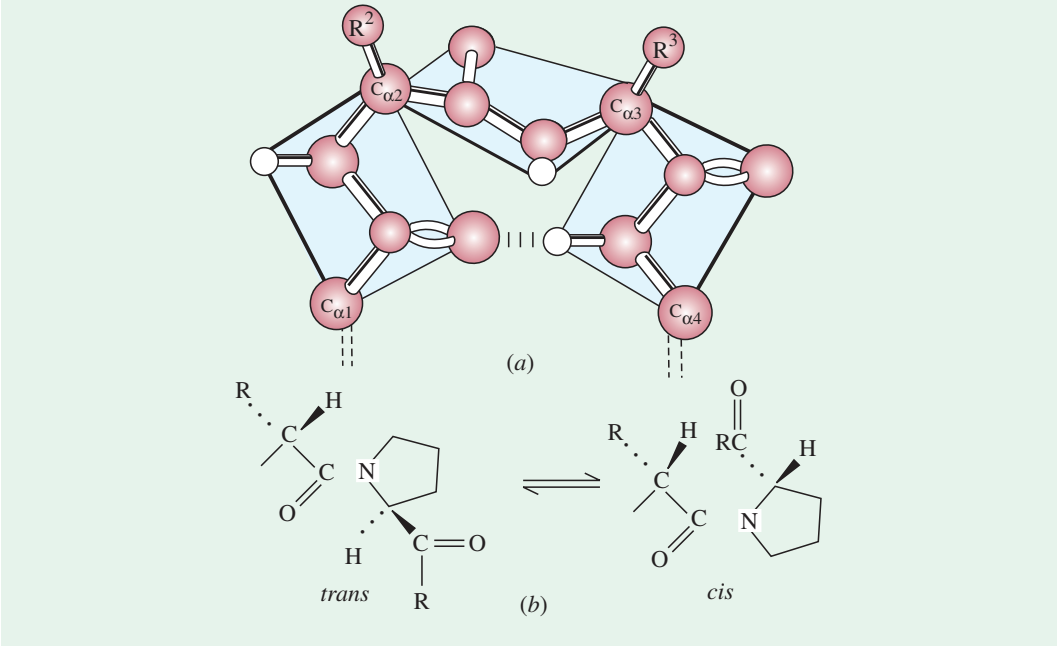


Fig. 9–29. Structure of a β turn or hairpin bend

(*a*) Hydrogen bonding in a tetrapeptide involving first and fourth amino acid residues. Here the NH and CO groups of residue 1 of the tetrapeptide are hydrogen-bonded, respectively, to the CO and NH groups of residue 4.

(*b*) The *trans* and *cis* isomers of a peptide bond involving the imino nitrogen of proline.

Collagen Triple Helix

Collagen is the most abundant protein of mammals. It is the principal structural element of the human body and makes up 25–33% of all the body protein. It is found in connective tissues such as tendons, cartilage, the organic matrix of bones and the cornea of the eye. This extracellular protein contains three helical polypeptide chains, each nearly 1,000 residues long. The amino acid sequence of collagen is remarkably regular : *nearly every third residue is glycine.* Proline is also present to a much greater extent than in most other proteins. Furthermore, collagen contains 4-hydroxyproline (Hyp), which is rarely found elsewhere. The percentage composition of predominant amino acids found in collagen is : Gly (35%), Ala (11%) and Pro + Hyp (25%). The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly–X–Pro or Gly–X–Hyp, where X can be any amino acid. This repeating tripeptide sequence adopts a left-handed helical structure with 3 residues per turn.

Collagen is a rod-shaped molecule, about 3,000 Å (= 300 nm) long and only 15 Å (=1.5 nm) in diameter. The helical motif of its 3 chains is entirely different from that of the α helix. Hydrogen bonds within a strand (*i.e.*, intrachain hydrogen bonds) are absent. Instead, each of these 3 helices is stabilized by steric repulsion of the pyrrolidone rings of the proline and hydroxyproline residues. The 3 strands wind around each other in a cable fashion (Fig. 9–30), to form a **superhelix**. The superhelical

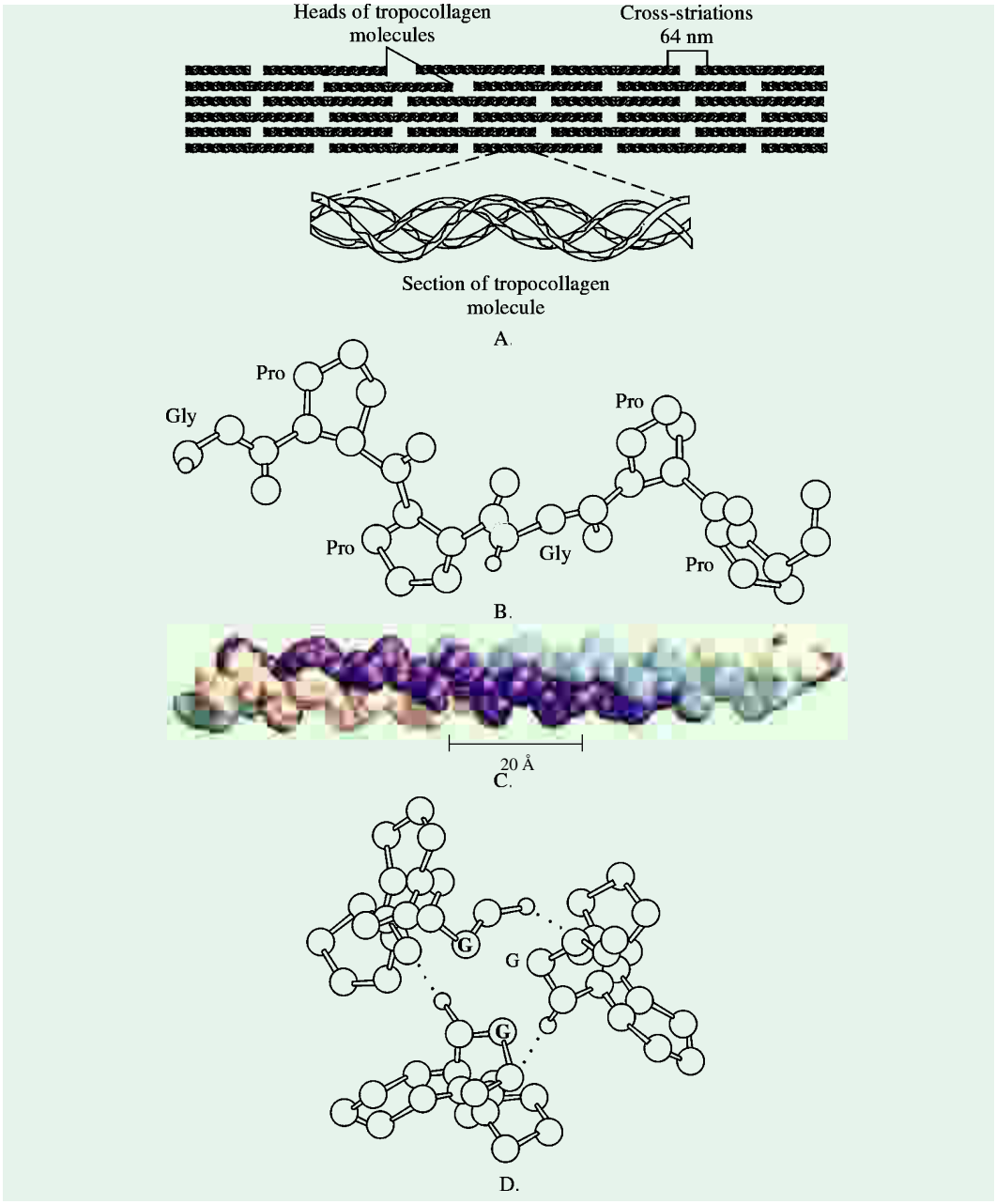


Fig. 9–30. The structure of collagen

A. **Collagen fibres and a section of tropocollagen molecule.** The basic collagen monomer is a triple helix, composed of 3 helical α chains. The collagen monomers before aligned in rows in which the molecules in one row are staggered relative to those in the neighbouring rows.

B. **Conformation of a single strand of the collagen triple helix.** The sequence shown here is Gly-Pro-Pro-Gly-Pro-Pro.

C. **Space-filling model of the collagen triple helix.** The three strands are shown in differnt shades.

D. **Cross section of a model of the triple-stranded helix of collagen.** Each strand is hydrogen-bonded to other two (... denotes a hydrogen bond). The α-carbon atom of a glycine residue in each strand is labelled G. Every third residue must be glycine because there is no space near the helix axis (centre) for a larger amino acid residue. Note that the pyrrolidone rings are on the outer side.

twisting is right-handed (cf α-keratin). The rise (axial distance) per residue in this superhelix is 2.9 Å, and there are nearly 3 residues per turn. The three strands are stabilized by the interchain formation of hydrogen bonds between the > C = O group of one chain and the >N—H group of another chain. Proline does not have a hydrogen atom attached to its nitrogen when participating in a peptide bond and therefore cannot participate in this interchain hydrogen-bonding. This 3-stranded superhelix is known as the **collagen triple helix**. The amino acid residue on either side of glycine is located on the outside of the cable, where there is room for the bulky rings of Pro and Hyp residues. The tight wrapping of the collagen triple helix provides great tensile strength with no capacity to stretch. *Collagen fibres can support up to 10,000 times their own weight and are said to have greater tensile strength than a steel wire of equal cross section.*

The collagen fibrils consist of recurring 3-stranded polypeptide units called **tropocollagen** (MW 3,00,000), arranged head-to-tail in parallel bundles (Fig. 9–29). In some collagens, all three chains are identical in amino acid sequence and thus are homotrimers ; while in others, two chains are identical and the third differs and thus are heterotrimers. The heads of adjacent molecules are staggered, and the alignment of the head groups of every fourth molecule produces characteristic cross-striations 640 Å (= 64 nm) apart. A series of complex covalent cross-links are formed within and between the tropocollagen molecules in the fibril, leading to the formation of strong mature collagen. The rigid, brittle character of the connective tissue in older people is the result of an accumulation of covalent cross-links in collagen as they age. The electron micrograph of human collagen fibres in given in Fig 9–31.

Fig. 9-31. An electron micrograph 0.2̄ μm of human collagen fibers showing their characteristic banding pattern

The bands repeat along the fibre with a periodicity of 64 to 70 nm

(Courtesy : Jesome Gross and Francis O. Schmit)

The food product **gelatin** is derived from collagen. Although it is a protein, it has little nutritional value because collagen lacks significant amounts of many amino acids that are essential in the human diet.

Human genetic defects involving collagen illustrate the close relationship between amino acid sequence and three-dimensional structure in this protein. **Osteogenesis imperfecta** results in abnormal bone formation in human babies. **Ehlers-Danlos syndrome** is characterized by loose joints. Both can be lethal and both result if a different glycine residue in each case is replaced by a cysteine (in case of osteogenesis imperfecta) or a serine (in case of Ehlers-Danlos syndrome) residue. These seemingly small replacements have a catastrophic effect on collagen function because they disrupt the Gly–X–Pro (or Hyp) repeat that gives collagen its unique helical structure.

Table 9–6 summarizes the important differences between the structures of α-keratin of hair and collagen of bones.

Table 9–6. Structural differences between α-keratin and collagen

α-keratin	*Collagen*
1. Found in skin and its appendages such as hair, nails and feathers.	1. Found in connective tissues such as tendons, cartilage, bones and the cornea of the eye.
2. The polypeptide chain is a right-handed helix.	2. The unique polypeptide chain is a left-handed helix.
3. Hydrophobic amino acids (Phe, Ile, Val, Met and Ala) predominate in the helix.	3. Gly, Pro, Hyp and Ala predominate in the helix.
4. The 3 helical strands wrap together into a superhelix, called protofibril.	4. The 3 helical strands wrap together into a repeating superhelical structure, called tropocollagen.
5. The superhelical twisting is left-handed.	5. The superhelical twisting is right-handed.
6. The covalent cross-links, between polypeptide chains within the triple-helical ropes and between adjacent ones, are contributed by disulfide bonds.	6. The covalent links are contributed by an unusual type of covalent link between two Lys residues that creates a nonstandard amino acid residue called lysinonorleucine.

Elastin

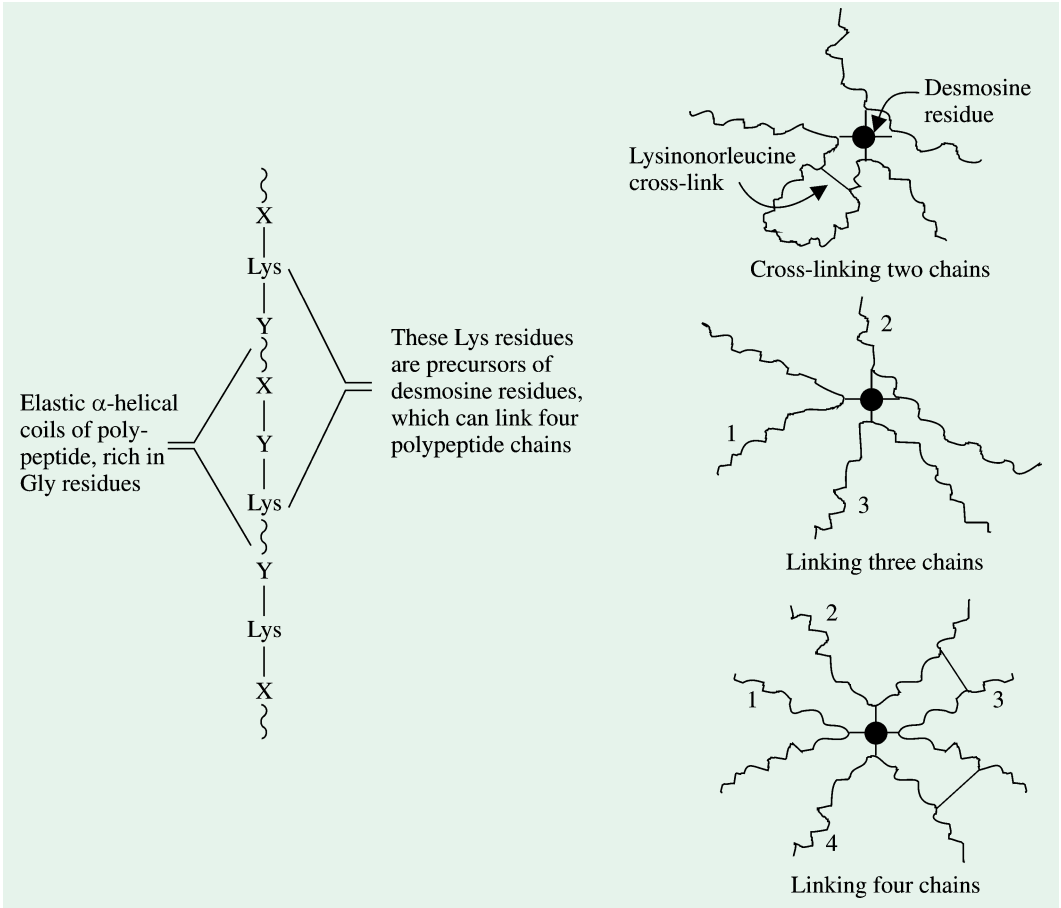


Fig. 9–32. Tropoelastin molecules and their linkage to form a network of polypeptide chains in elastin
As shown, desmosine residues can link two, three or four tropoelastin molecules. In addition, other types of linkages such as lycinonorleucine are also present.

(*Adapted from Lehninger, Nelson and Cox, 1993*)

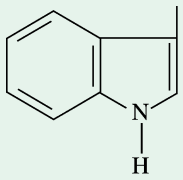
Elastin is found in elastic connective tissues such as ligaments and blood vessels. It resembles collagen in some of its properties, but differs in other. The polypeptide subunit of elastin fibrils is **tropoelastin** (MW 72,000), containing about 800 amino acid residues. Tropoelastin differs from tropocollagen in having many lysine but few proline residues. Also, the helix which it forms is quite different from a helix and the collagen helix.

Tropoelastin consists of α helical portions of polypeptides rich in Gly residues, separated by short regions containing Lys and Ala residues. The helical regions stretch on applying tension but regain their original length when tension is released. The regions containing Lys residues form covalent cross-links. Four lysine side chains come together and are enzymatically converted into **desmosine** and a related compound, **isodesmosine** ; these amino acids are found exclusively in elastin. Like collagen, elastin also contains lysinonorleucine. These nonstandard amino acids are capable of joining tropoelastin chains into arrays that can be stretched reversibly in all directions (Fig. 9–32).

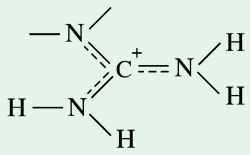
Hydrogen-bonding Potentiality of Proteins

All reversible molecular interactions in biological systems are mediated by 3 kinds of forces: electrostatic bonds, hydrogen bonds and van der Waals bonds. We have already seen that the hydrogen bonds between main-chain NH and CO groups work in forming α helices, β sheets and collagen fibrils. In fact, side chains of 11 of the 20 fundamental amino acids also can participate in hydrogen-bonding. These have been separated into 3 groups, according to their hydrogen- bonding potentialities :

1. **Tryptophan and Arginine.** Their side chains can serve *as hydrogen bond donors only*.

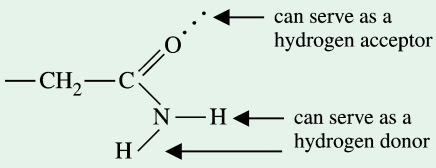


Hydrogen donor group of tryptophan

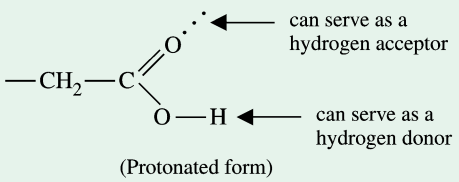


Hydrogen donor groups of arginine

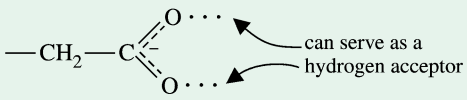
2. **Asparagine, Glutamine, Serine and Threonine.** Like the peptide unit itself, their side chains can serve *as hydrogen bond donors and acceptors.*
3. **Lysine, Aspartic acid, Glutamic acid, Tyrosine and Histidine.** The hydrogen-bonding capabilities of lysine (and the terminal amino group), aspartic and glutamic acids (and the terminal carboxyl group), tyrosine and histidine vary with pH. These groups can *serve as both acceptors and donors* within a certain pH range and *as acceptors* or *donors* (but not both) at other pH values (Fig. 9–33).



Asparagine or **Glutamine**



(Protonated form)



(Ionized form)

Aspartic acid or **Glutamic acid**

Fig. 9–33. Hydrogen-bonding groups of several side chains in proteins

3. Tertiary Structure : Folding of the Chain or Overall Folding

If the globular proteins consisted only of a small helix, these molecules would have been elongated structures with considerable length and a small cross-sectional area (*i.e.*, a large axial ratio). But as we now know about the existence of globular proteins, the helix must, therefore, possess many other types of bonds placed at regular intervals. These additional bonds include disulfide, hydrogen, hydrophobic and ionic. In such globular proteins (including enzymes, transport proteins, some peptide hormones and immunoglobulins), polar groups because of their hydrophobicity are most often located on the molecule's exterior and nonpolar R groups in the interior, where their interactions create a hydrophobic environment. The *tertiary structure*, thus, involves the folding of the helices of globular proteins. It refers to the spatial arrangement of amino acids that are far apart in linear sequence and to the pattern of disulfide bonds. The dividing line between secondary and tertiary structure is, hence, a matter of taste. X-ray crystallographic studies have revealed the detailed 3-'D' structures of more than 300 proteins.

Myoglobin (Mb)

Myoglobin (*myo*^{G} = muscle ; *globin*^{G} = a type of protein) is a relatively small, oxygen- binding heme protein, found in muscle cells. It has the distinction of being the first globular protein to have its 3-'D' structure elucidated by x-ray diffraction studies. This was accomplished by John C. Kendrew at a resolution of 6 Å in 1957, 2 Å in 1959 and 1.4 Å in 1962 (Fig. 9–34). Myoglobin molecule (Fig. 9–35) contains a single polypeptide chain of 153 amino acid residues and a single prosthetic iron-porphyrin (or heme) group, identical with that of hemoglobin. The heme group is responsible for the deep red-brown colour of myoglobin (and also of hemoglobin). Myoglobin is especially abundant in the muscles of diving mammals such as the whale, seal and porpoise, whose muscles are so rich in this protein that they are brown. Storage of oxygen by muscle myoglobin permits these animals to remain submerged for long periods. The function of myoglobin is to bind oxygen in the muscles and to enhance its transport to the mitochondria, which consume oxygen during respiration.

Between 1912 and 1915, **William Henry Bragg** and his son, **William Lawrence Bragg,** developed the technique of **x-ray diffraction** by determining the crystalline structure of NaCl. As the joint 1915 Nobel Prize recipients in physics, the Braggs became the only father-son combination to receive the award and W. Lawrence, who was 25 years old at the time, the youngest scientist so honoured.

Myoglobin (MW = 16,700) is an extremely compact macromolecule with oblate, spheroid shape and leaves little empty space in its interior. Its overall molecular dimensions

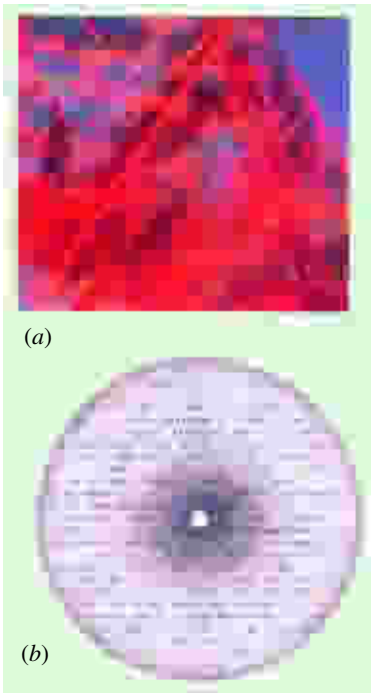


Fig. 9.34. Sperm whale myoglobin crystal and x-ray photograph

(*a*) **Crystal of myoglobin.**(*b*) **An X-ray diffraction pattern (or photograph) of a single crystal of sperm whale myoglobin**. The pattern of spots is produced as a beam of x-rays is diffracted by the atoms in the protein crystal, causing the x rays to strike the film at specific sites. Information derived from the position and intensity (darkness) of the spots can be used to calculate the positions of the atoms in the protein that diffracted the beam. The intensity of each diffraction maximum (the darkness of each spot) is a function of the myoglobin crystal's electron density. The photograph contains only a small fraction of the total diffraction information available from a myoglobin crystal.

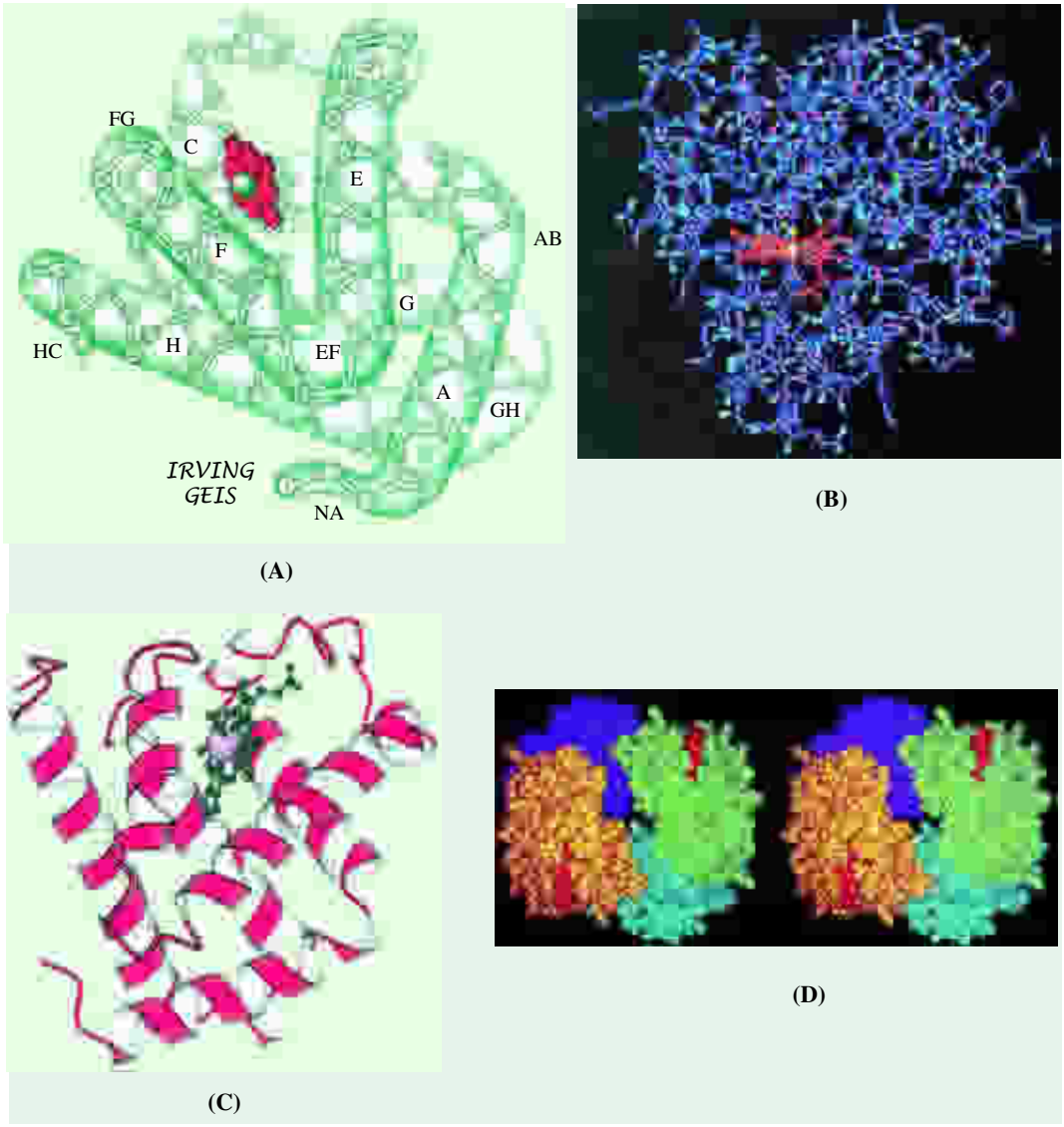


Fig. 9–35. The structure of sperm whale myoglobin

(A) **The tertiary structure**. Its 153 C_α positions are numbered from the N-terminus and its eight helices are sequentially labeled a through H. The last half of the EF corner is now regarded as a turn of helix and is therefore disignated the F′ helix. the heme group is shown in red. Most of the amino acids are part of α helices. the nonhelical regions occur primarily as turns, where the polypeptide chain changes direction. The position of the heme is indicated in red.

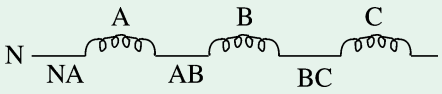
(B) **The ball- and stick model.** The three-dimensional structure of myoglobin shows the position of all of the molecule's atoms other than hydrogen and reveals many interactions between the amino acids. The heme group is indicated in red.

(C) **Ribbon model.** A schematic view shows that the protein consists largely of α helices. The heme group is shown in black and the iron atom is shown as a purple sphere

(D) **A stereo, space-filling model showing the quaternary structure of hemoglobin.** The model shows the α_1, α_2, β_1, and β_2 subunits as colored yellow, green, light blue, and dark blue, respectively, Heme groups are red. The protein is viewed along its molecular twofold rotation axis which relates the $\alpha_1\beta_1$ protomer to the $\alpha_2\beta_2$ protomer.

(*Courtesy : (A) Irving Geis (B) Ken Eward*)

are 45 × 35 × 25 Å. The sausagelike outline of the tertiary structure of myoglobin chain is folded into an odd, irregular form. The backbone of the molecule is made up of 8 almost-straight α-helical segments, designated from the N-terminus as A through H (Fig 9–36). The first residue in helix A is designated A_1, the second A_2 and so forth. Interspersed among the helical segments are 5 nonhelical regions, each identified by the two helical segments it joins, *e.g.*, AB is located between helical sections A and B. There are also two other nonhelical regions : two residues at the N-terminus (named NA_1 and NA_2) and five residues at the C-terminus (named HC_1 through HC_5). The longest helical segment has 23 amino acid residues and the shortest only 7. All the helical segments are of α-type and right-handed ; there being no β-structure (refer Table 9–7).



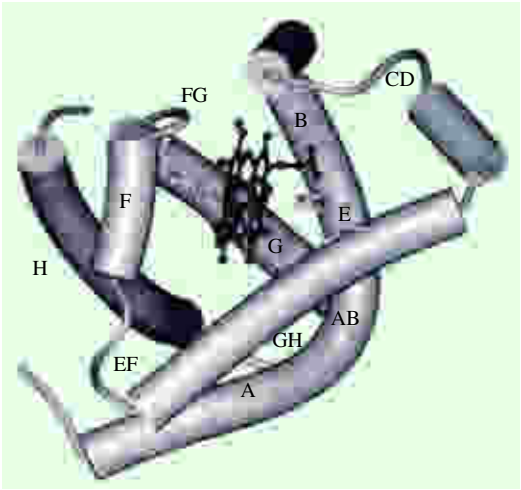


Fig. 9–36. The eight α helical segments of myoglobin molecule.

The eight α helical segments (shown here as cylinders) are labeled A through H. Nonhelical residues in the bends that connect them are labeled AB, CD, EF, and so forth, indicating the segments they interconnect. A few bends, including BC and DE, are abrupt and do not contain any residues; these are not normally labeled. (The short segment visible between D and E is an artifact of the computer representation.) The heme is bound in a pocket made up largely of the E and F helices, although amino acid residues from other segments of the protein also participate.

Table 9–7. Approximate amounts of α-helix and β-structure in some single-chain proteins*

Protein	*Total residues*	*Residues, %*	
		α-helix	*β-structure*
Cytochrome C	104	39	0
Ribonuclease	124	26	35
Lysozyme	129	40	12
Myoglobin	153	78	0
Chymotrypsin	247	14	45
Carboxypeptidase	307	38	17

* Portions of polypeptide chains, that are not accounted for by α-helix or β-structure, consist of bends, reverse turns and irregularly-coiled stretches.

(*Adapted from Cantor and Schimmel, 1980*)

Of the 153 amino acid residues, 121 (*i.e.*, 79%) are present on the helical regions and the remaining 32 amino acid residues are distributed over the nonhelical areas. The nonhelical areas possess various types of bonding such as hydrogen and nonpolar linkages. The flat heme group is tightly but noncovalently bound to the polypeptide chain. A notable feature of whale myoglobin is the absence of a disulfide bridge, since both cysteine and cystine residues are lacking. The electron density map of sperm whale myoglobin is presented in Fig. 9–37.

Other important features of the myoglobin molecule are listed below :

1. The molecule is very compact and leaves so little space in its interior as to accommodate only 4 water molecules.
2. All the polar R groups of the molecule, except two, are located on its outer surface and all of them are hydrated.
3. Most of the hydrophobic R groups are located in the interior of the molecule. Hydrophobic R groups of helices E and F form the sides of a pocket into which the hydrophobic heme group fits. The porphyrin ring of the heme is largely hydrophobic, except for the 2 propionic acid side chains which stick out of the pocket and into the aqueous environment.
4. Each of the 4 proline residues occurs at a turn. Other turns or bends contain serine, threonine and asparagine.
5. All the peptide bonds of the polypeptide chain are planar, with the carbonyl and amide groups being *trans* to each other.
6. The heme group is flat and rests in a crevice in the molecule. The iron atom in the centre of the heme group has two coordination bonds that are perpendicular to the plane of the heme group. One bond is attached to the R group of histidine (93), whereas the other bond is the site to which an O_2 molecule is bound. Upon oxygenation, the iron atom descends into the heme plane and the oxygen bound to myoglobin is stabilized by hydrogen-bonding to the imidazole ring of His E_7 (Fig. 9–38). Reversible oxygenation requires that the iron atom be in the ferrous state (Fe^{2+}) and myoglobin with or without oxygen bound to the Fe^{2+} of heme is called oxyhemoglobin and deoxyhemoglobin respectively.
7. The inside and outside are well defined. There is little empty space inside. *The interior consists almost entirely of nonpolar residues* such as Leu, Val, Met, and Phe. On the contrary, Glu, Asp, Gln, Asn, Lys and Arg are absent from the interior of the

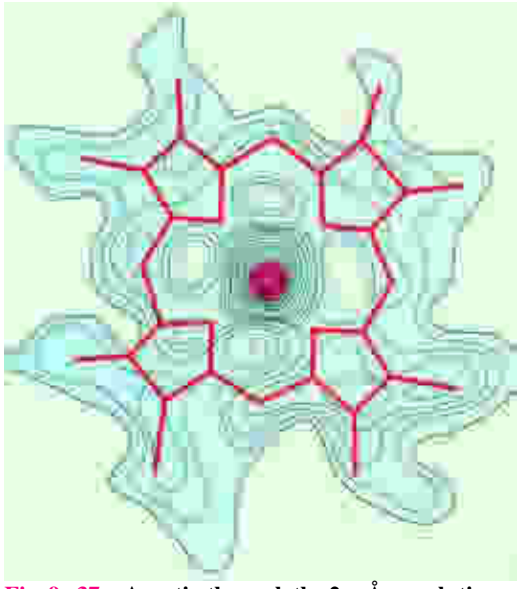


Fig. 9 –37. A sectin through the 2.o-Å- resolution electron density map of sperm whale myoglobin

The heme group is represented by red. The large peak the centre of the map represents the electron-dense Fe atom.

(After Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC and Shore VC,

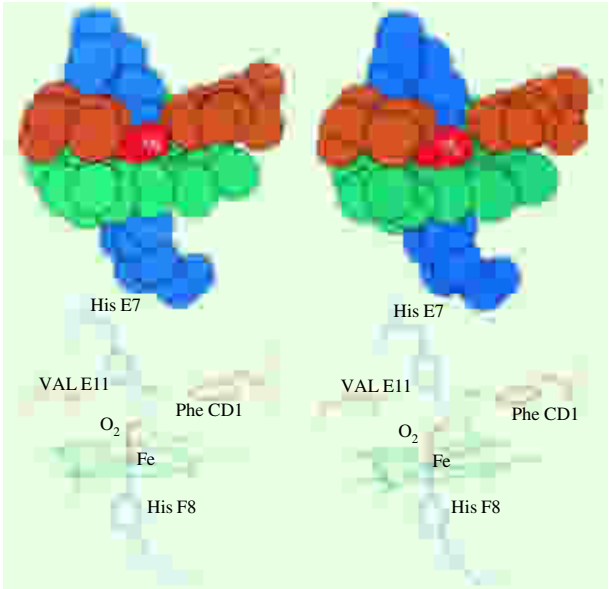


Fig. 9–38. Stereo drawings of the heme complex in oxymyoglobin

In the upper drawing, atoms are represented as spheres of van der Waals radii. the lower drawing shows the corresponding skeletal model with a dashed line representing the hydrogen bond between the distal His and the bound O_2.

(After Phillipas SEV, 1980)

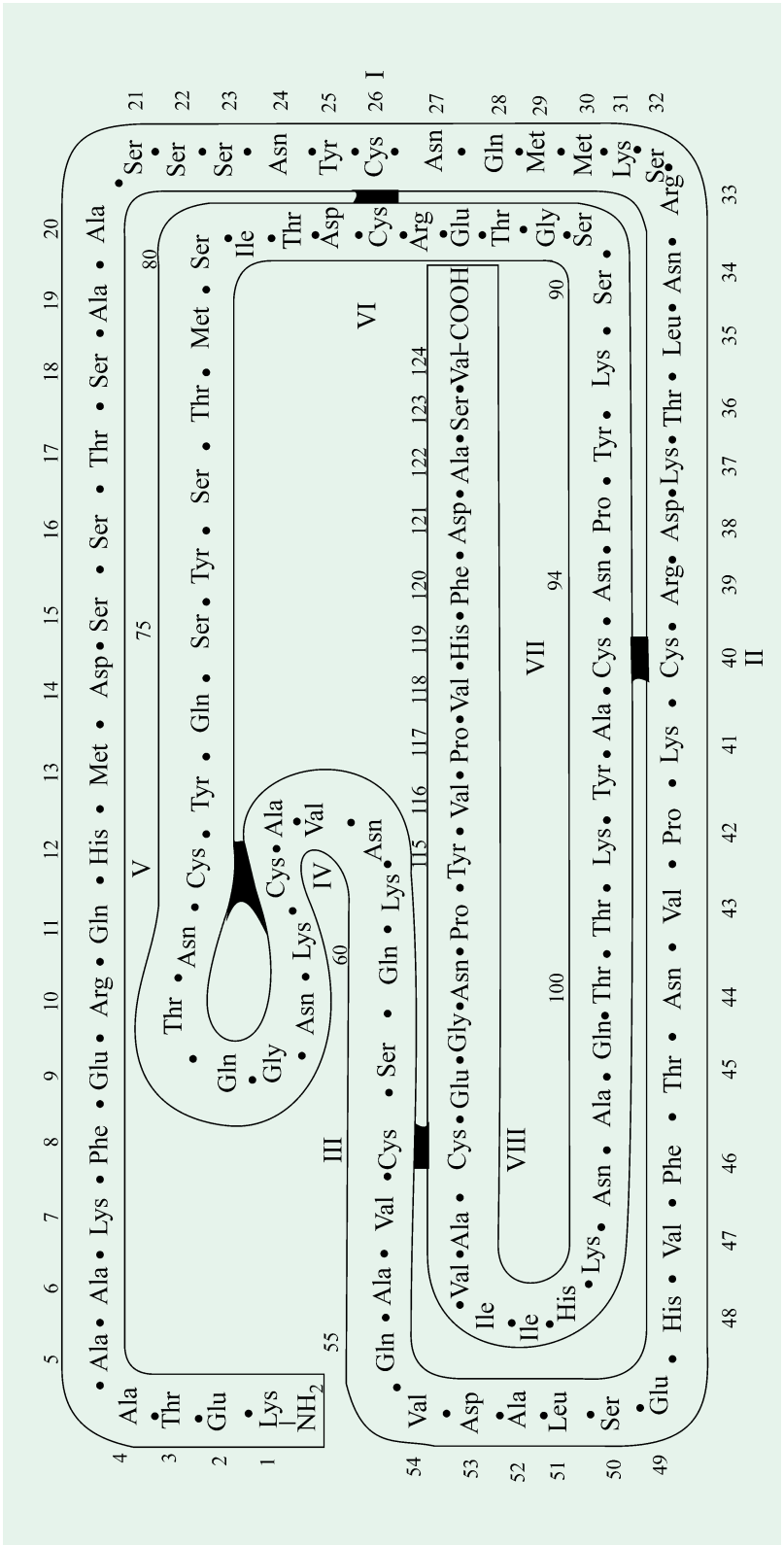


Fig. 9–39. The amino acid sequence of the enzyme ribonuclease from bovine

This enzyme molecule consists of 124 amino The 4 dark blocks between cysteines represent 4 disulfide bonds or bridges. Note that this illustration is diagrammatic; the polypeptide is actually folded to a give a complex 3-dimensional configuration.

(*After Smyth DG, Stein WH and Moore S, 1963*)

protein. Residues with polar and nonpolar parts within them (Thr, Tyr, Trp) are so oriented that their nonpolar portions point inward. The only polar residues inside myoglobin are two histidines which have a critical function at the binding site. The exterior of the protein has both polar and nonpolar residues.

8. Myoglobin without its heme prosthetic group is designated *apomyoglobin*. The main function of apomyoglobin molecule is to provide a hydrophobic environment for the heme group and a properly oriented imidazole group to occupy the 5th coordination position of the iron. The presence of the hydrophobic environment and the proximal histidine enable the heme to combine reversibly with O_2 and prevent the oxidation of Fe^{2+} to Fe^{3+}.

Ribonuclease (RNase)

Ribonuclease (MW = 13,700) is another noteworthy example of a protein with tertiary structure. Two groups of workers (Christian Anfinsen and his associates, 1950, at the National Institute of Health ; William Stein and Stanford Moore, 1958, at Rockfeller Institute) have elucidated the complete structure of this pancreatic protein. Ribonuclease molecule (Fig. 9-39) consists of 124 amino acid units contained in a single polypeptide chain with 4 disulfide bridges. In establishing the amino acid sequence, the RNase was subjected to hydrolysis either by pepsin or by trypsin (for further details, refer Fig. 17–10 on page 300).

Carboxypeptidases

Carboxypeptidases are made by the exocrine cells of the pancreas as their inactive zymogens, *procarboxypeptidases*. Synthesis of these enzymes as inactive precursors protects the exocrine cells from destructive proteolytic attack. Carboxypeptidases are exopeptidase enzymes that catalyze the hydrolysis of (or cleave) proteins either from N- or C-terminus, *i.e.*, inwardly from only one end of

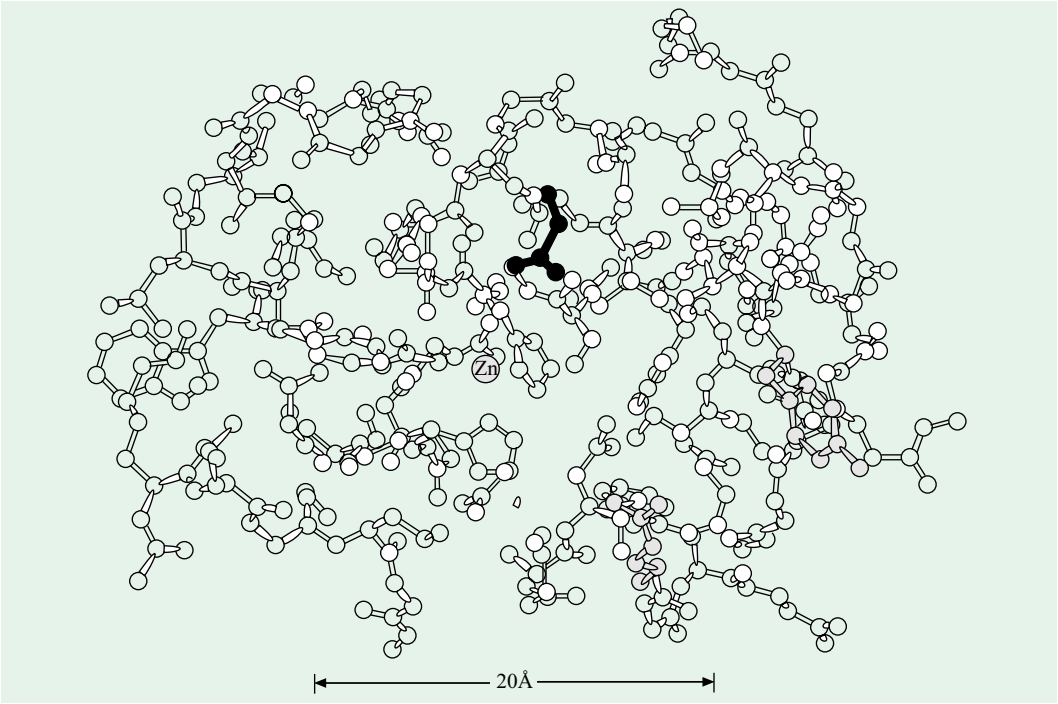


Fig. 9–40. The structure of carboxypeptidase A

The three shaded amino acids are Glu 270 (upper), Tyr 248 (middle) and Arg 145 (lower). Also shown is the zinc ion (Zn^{2+}) located in a groove near the surface of the molecule.

(*After William N. Lipscomb, 1971*)

the chain. (By contrast, endopeptidases are enzymes that cleave internal peptide bonds.). Different forms of carboxypeptidases are usually *substrate-specific*. *Carboxypeptidase A*, the enzyme used initially in C-terminal residue determinations, readily hydrolyzes (or cleaves) a C-terminal residue with an aromatic or a bulky aliphatic side chain. *Carboxypeptidase B* cleaves C-terminal lysine or arginine residues while *carboxypeptidase C* cleaves a C-terminal proline residue. However, *carboxypeptidase Y* has the capacity to cleave all C-terminal amino acids.

Carboxypeptidase A (MW = 34,600) is a zinc-containing proteolytic enzyme. Its 3-'D' structure (Fig. 9–40) was accomplished at a resolution of 2 Å by William Lipscomb in 1967. This enzyme is a globular protein which contains 307 amino acid residues in a single polypeptide chain. Carboxypeptidase A molecule is a compact ellipsoid with dimensions of $50 \times 42 \times 38$Å. This molecule is considerably larger than myoglobin and shows more variety in its secondary structures. Of the 307 amino acid residues, 38% are involved in the formation of 8 helical segments (designated A through H) and 17% are involved in β pleated sheet structure that extends through the molecule, forming a core. Helix D is connected to helix E by a segment containing 52 amino acids. The folding of this segment is very complex and contains no recognizable regular secondary structure. A tightly bound zinc ion (which is essential for enzymatic activity) is located in a groove near the surface of the molecule, where it coordinates with 2 histidine side chains, a glutamate side chain and a water molecule. A large pocket near the zinc ion accommodates the side chain of the terminal residue of the peptide substrate.

Fig. 9-41 gives the ribbon model of the structure of bovine carboxypeptidase A, along with its central β sheet.

Fig. 9–41. Bovine carboxypeptidase A showing its central β sheet

(Photo Courtesy : Irving Geis, 1986)

Two aspects of catalytic mechanism of carboxypeptidase A are noteworthy :

(A) **Binding of substrate :** For instance, the binding of glycyltyrosine (a slowly-hydrolyzed substrate) is accompanied by a large structural rearrangement of the active site. The phenolic hydroxyl group of tyrosine 248 moves a distance of 12 Å from the surface of the molecule to the vicinity of the terminal carboxylate of the substrate. Consequently, the active site cavity closes so that the water molecule extrudes from it.

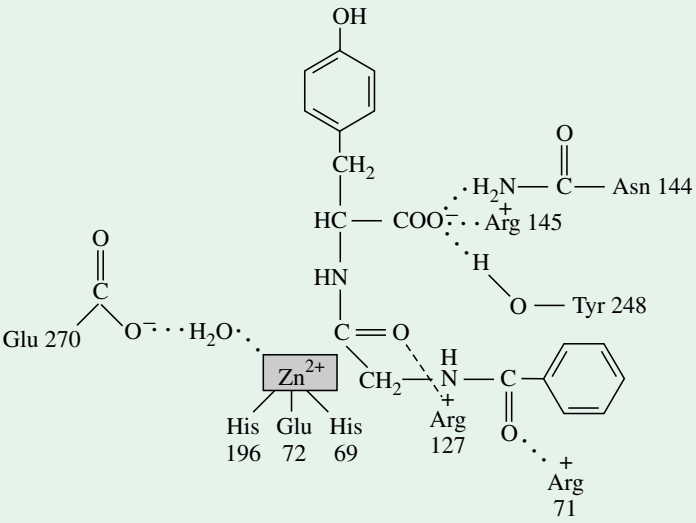


Fig. 9-42. Mode of binding of N-benzoylglycyltyrosine to carboxypeptidase A

(*After Christianson DW and Lipscomb WN, 1989*)

How does carboxypeptidase A recognize the terminal carboxylate group of a peptide substrate? Three interactions explain this (Fig. 9–42). The induced-fit movement of Tyr 248 enables its phenolic OH group to become hydrogen-bonded to the terminal carboxylate group. Further, the guanidium of Arg 145 moves 2 Å in this transition so as to enable it to form a salt bridge with the terminal COO^-, which is hydrogen-bonded to a side chain Asn 144. Besides, the terminal side chain of the substrate sits in a hydrophobic pocket of the enzyme (this is the reason why carboxypeptidase A requires an aromatic or bulky nonpolar residues at this position). Two interactions involving the penultimate amino acid residue are also noteworthy. The carbonyl oxygen of the scissile peptide bond interacts with the guanidium group of Arg 127. Finally, Tyr 248 is also hydrogen-bonded to the peptide NH of penultimate residue.

(B) **Activation of bound water :** Proteins and peptides are stable at neutral pH in the absence of a protease because water does not readily attack peptide bonds. In fact, carboxypeptidase A also activates a water molecule. This is accomplished by the bound zinc ion with the help of adjacent carboxylate group of Glu 270. Thus, *the zinc-bound water behaves muck like an OH^- ion*. The activation of water by zinc ion involves two steps :

1. First step in catalysis is the attack of activated water molecule on the carbonyl group of the scissile peptide bond (Fig. 9–43). Specifically, the nucleophilic oxygen atom of activated water attacks the carbonyl carbon atom. Concurrently, Glu 270 accepts a proton from the water. A negatively-charged tetrahedral intermediate is, thus, formed. This intermediate is stabilized by electrostatic interactions with Zn^{2+} and the positively-charged side chain of Arg 127.

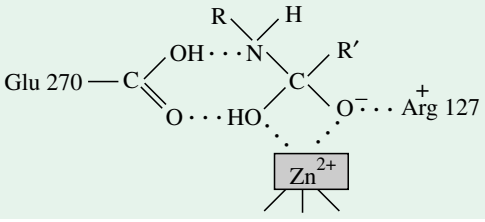


Fig. 9 -43. Proposed tetrahedral transition state in the peptide bond hydrolysis by carboxypeptidase A

2. The final step is the transfer of a proton from the COOH group of Glu 270 to the peptide NH group. The peptide bond is simultaneously cleaved and the reaction products diffuse away from the active site. In fact, the bound substrate is surrounded on all sides by catalytic groups of the enzyme. And the substrate-induced structural changes in the active site promote catalysis in 3 ways :
   (*a*) activation of H_2O by Zn^{2+},
   (*b*) proton abstraction and donation by Glu 270, and
   (*c*) electrostatic stabilization by Arg 127.

It is, thus, apparent that a substrate could not enter such an array of catalytic groups (nor could a product leave) unless the enzyme were flexible.

4. Quaternary Structure : Protein-Protein Interactions or Multichain Association

A fourth degree of complexity in protein structure has recently been recognized to be of great value in many proteins. Some globular proteins consist of 2 or more interacting peptide chains. Each peptide chain in such a protein is called a *subunit*. These chains may be identical or different in their primary structure. This specific association of a number of subunits into complex large-sized molecules is referred to as the *quaternary structure*. In other words, quaternary structure refers to the spatial arrangement of subunits and the nature of their contact. The same forces (disulfide, hydrogen, hydrophobic and ionic bonds) involved in the formation of tertiary structure of proteins are also involved here to link the various polypeptide chains.

Tobacco Mosaic Virus (TMV)

TMV, with 158 amino acid residues, is an instance of the protein-protein interactions. TMV particles are rod like, tubular structures measuring 3,000 Å length and 180 Å diameter. The tube is made up of a single RNA molecule (30,000 Å long) which is coiled round the central axis. The RNA molecule is a right-handed helical filament and is embedded in a protein matrix or the coat which is made up of 2,130 identical molecules of a single protein that interact to form a cylinder enclosing the RNA genome. The amino acid sequence of this coat protein (Fig. 9–44) was elucidated in 1962 independently by H. Fraenkel-Conrat and H. Wittmann. The coat protects the RNA molecule from disintegration by various environmental changes.

AcSer-Tyr-Ser-Ile-Thr-Thr-Pro-Ser-Gln-Phe-Val-Phe-Leu-Ser-Ser-Ala-Trp-Ala-Asp-Pro-
1 5 10 15 20
Ile-Glu-Leu-Ile-Asn-Leu-Cys-Thr-Asn-Ala-Leu-Gly-Ans-Gln-Phe-Gln-Thr-Gln-Gln-Ala
25 30 35 40
Arg-Thr-Val-Val-Gln-Arg-Gln-Phe-Ser-Glu-Val-Trp-Lys-Pro-Ser-Pro-Gln-Val-Thr-Val-
45 50 55 60
Arg-Phe-Pro-Asp-Ser-Asp-Phe-Lys-Val-Tyr-Arg-Tyr-Asn-Ala-Val-Leu-Asp-Pro-Leu-Val-
65 70 75 80
Thr-Ala-Leu-Leu-Gly-Ala-Phe-Asp-Thr-Arg-Asn-Arg-Ile-Ile-Glu-Val-Glu-Asn-Gln-Ala-
85 90 95 100
Asn-Pro-Thr-Thr-Ala-Glu-Thr-Leu-Asp-Ala-Thr-Arg-Arg-Val-Asp-Asp-Ala-Thr-Val-Ala
105 110 115 120
Ile-Arg-Ser-Ala-Ile-Asn-Asn-Leu-Ile-Val-Glu-Leu-Ile-Arg-Gly-Thr-Gly-Ser-Tyr-Asn-
125 130 135 140
Arg-Ser-Ser-Phe-Glu-Ser-Ser-Ser-Gly-Leu-Val-Trp-Thr-Ser-Gly-Pro-Ala-Thr
145 150 155 158

Fig. 9–44. Amino acid sequence of the coat protein or tobacco mosaic virus (TMV)

Note that the serine in position 1 is acetylated (AcSer). Also note that there are no extended repetitions of an amino acid sequence, that there are no clusters of either polar or nonpolar amino acids, and that the frequencies of occurrence of different amino acids vary widely.

Hemoglobin (Hb)

Hemoglobin (*hemo*^{G} = blood ; *globin*^{G} = a protein, belonging to the myoglobin-hemoglobin family), the oxygen transporter in erythrocytes, constitutes about 90% of the protein of red blood cells. It is a tetrameric protein, *i.e.*, it contains 4 polypeptide chains. The four chains (of which two are of one kind and two of another) are held together by noncovalent interactions. Each chain contains a heme group and a single oxygen-binding site. *Hemoglobin A*, the principal hemoglobin in adults, consists of two alpha (α) chains and two beta (β) chains. Adults also have a minor hemoglobin (~ 2% of the total hemoglobin) called hemoglobin A_2, which contains two delta (δ) chains in place of two β chains of hemoglobin A. Thus, the subunit composition of hemoglobin A is $\alpha_2\beta_2$, and that of hemoglobin A_2 is $\alpha_2\delta_2$. HbA and HbA_2 are the postnatal forms of hemoglobin.

Prior to birth, embryonic and fetal hemoglobins are used as oxygen carrier. The first hemoglobin to appear in embryonic development has the subunit composition $\zeta_2\,\varepsilon_2$: the two zeta (ζ) chains are analogous to the a chain and the two epsilon (ε) chains are analogous to the β chains. When, after about 6 weeks, ζ chain production ceases, the tetramer $\alpha_2\,\varepsilon_2$ appears, *i.e.*, ζ is replaced by α. A third embryonic hemoglobin, $\zeta_2\,\gamma_2$ has also been identified where ε is replaced by γ. These last two hemoglobins represent transition phases leading to the appearance of *fetal hemoglobin* (*HbF*), whose tetrameric composition is $\alpha_2\,\gamma_2$. The α and ζ chains contain 141 residues each. The β, γ and δ chains contain 146 residues each and have homologous (similar but not identical) amino acid sequences. The γ and δ chains differ from the β chain at 39 and 10 amino acid residues respectively.

Hemoglobin (MW = 64,500) provides an example of the interaction of unidentical protein subunits. Since hemoglobin is 4 times as large as myoglobin, much more time and effort were required to solve its 3- 'D' structure, finally achieved by Max Perutz, John C. Kendrew and their colleagues in 1959. They determined the 3- 'D' structure of adult hemoglobin (HbA) of horse, which is very similar to that of human hemoglobin (HbA). The x-ray analysis has revealed that the hemoglobin molecule is nearly spherical with a diameter of 55 Å (5 .5 nm).

Human hemoglobin protein (Fig. 9–45) consists of 4 polypeptide chains of two types, two α-chains and two β-chains. The polypeptide portion is collectively called as *globin*. The α-chain has valine at the N-terminal and arginine at the C-terminal whereas in the β-chain, valine is situated at the N-terminal and histidine at the C-terminal. Each a chain is in contact with both β chains. In contrast, there are few interactions between the two α-chains or between the two β–chains. Each chain has a heme prosthetic group in a crevice near the exterior of the molecule. The heme groups are involved in the binding of oxygen. The α-chain has 141 residues and the β-chain, which is more acidic, has 146 residues. The protein, thus, has a total (141 × 2 + 146) 574 amino acid residues. Each of the 4 chains has a characteristic *tertiary structure*, in which the chain is folded. Like myoglobin, the α- and β-chains of hemoglobin contain several segments of α-helix, separated by bends. The α- and β-chains are held together as a pair by ionic and hydrogen bonds. The 2 pairs are then joined to each other by additional ionic bonds, hydrogen bonds and the hydrophobic forces. Thus, the 4 polypeptide chains fit together almost tetrahedrally to produce the

The designations α and β, which are common labels in Biochemistry, are used simply to identify two different polypeptide chains and are not to be confused with α and β secondary structures or a and β amino acids.

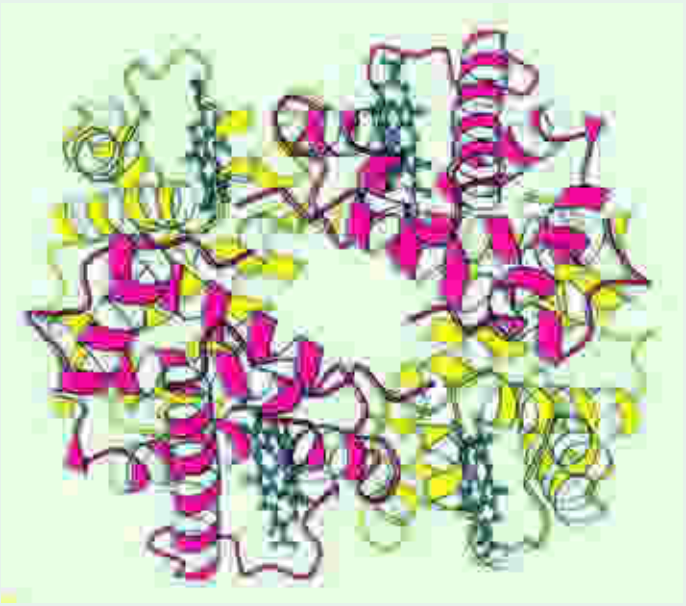


Fig. 9–45. Quaternary structure of hemoglobin molecule, showing interaction of 4 polypeptide chains
Hemoglobin, which is composed of 2 α chains and 2 β chains, functions as a pair of αβ dimers. The structure of the two identical α subunits (red) is similar to but not identical with that of the two identical β subunits (yellow). The molecule contains four heme groups (black with the iron atom shown in purple).

characteristic *quaternary structure*. The hemes are 2.5 nm apart from each other and tilted at different angles. Each heme is partly burried in a pocket lined with hydrophobic R groups. It is bound to its polypeptide chain through a coordination bond of the iron atom to the R group of a histidine residue. The sixth coordination bond of the iron atom of each heme is available to bind a molecule of oxygen. Myoglobin and the α and β chains of hemoglobin have nearly the same tertiary structure. Both have well over 70% α-helical nature, both have similar lengths of α-helical segments and the bends have about the same angles.

Similarity in 3- 'D' Structure of Hemoglobin and Myoglobin

The 3-'D' structures of myoglobin and the α and β chains of human hemoglobin are strikingly similar (Fig. 9–46). The 8 helices in each chain of hemoglobin are virtually superposable on those of myoglobin. This close resemblance in the folding of their main chains was unexpected because their amino acid sequences are rather different. In fact, these 3 chains are identical at only 24 or 141 positions.

The amino acid sequence of hemoglobins from more than 60 species, ranging from lamprey eels to humans, are known. A comparison of these sequences shows considerable variations at most positions. However, 9 positions have the same residues in most species studied so far (Table 9–7). These highly conserved 9 residues are especially important for the functioning of the hemoglobin molecule. For example, the invariant ***F8 histidine*** is directly bonded to the heme iron ; proline ***C2*** may be essential because it defines one end of the C helix ; and tyrosine ***HC2*** stabilizes the structure by forming a hydrogen bond between the H and F helices.

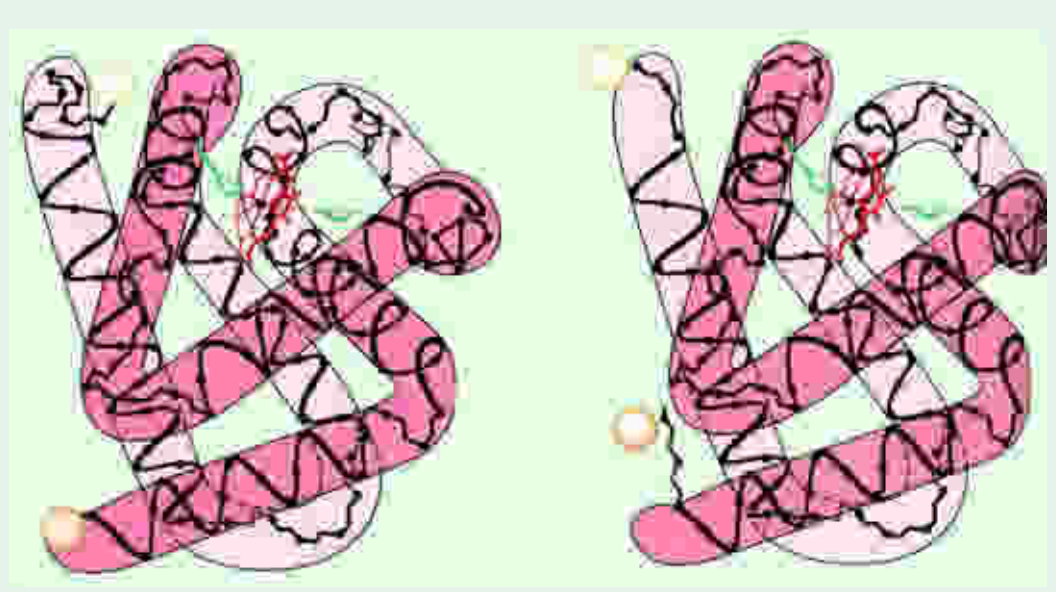


Fig. 9–39. Comparison of the conformation of the main chain of myoglobin and the β chain of hemoglobin A_1

Proximal His F8, distal His E7 and Val E11 side chains are shown. other amino acids of polypeptide chain are represented by α-carbon positions only; the letters M, V and P. refer to the methyl, vinyl and propionate side chains of the heme. The overall structures are very similar, except at NH_2- terminal and COOH-terminal ends.

(*Courtesy: Fersht A, 1977*)

Table 9–8. The nine highly conserved (or invariant) amino acid residues in hemoglobins

Position	*Amino acid*	*Role*
F8	Histidine	Proximal heme-linked histidine
E7	Histidine	Distal histidine near the heme
CD1	Phenylalanine	Heme contact
F4	Leucine	Heme contact
B6	Glycine	Allows close approach between B and E helices
C2	Proline	Helix formation
HC2	Tyrosine	Cross-links the F and H helices
C4	Threonine	Uncertain
H10	Lysine	Uncertain

(*Adapted from Lubert Stryer, 1993*)

The amino acid residues in the interior of hemoglobin are strikingly nonpolar but vary considerably. However, the change is always of one nonpolar residue for another as from alanine to isoleucine. The nonpolar core functions in binding the heme group and in stabilizing the 3-'D' structure of each subunit. On the contrary, *the surface amino acid residues are highly variable.* Indeed, few are consistently positively or negtatively charged.

Hemoglobin as an Allosteric Protein

The influence that the binding of an O_2 molecule to hemoglobin has on the subsequent oxygenation of the other heme groups of the molecule is an example of an interaction at one site on a protein affecting another site located in a distinctly different region of the same molecule. Such interactions are generally referred to as *allosteric interactions*. In addition to hemoglobin, many enzymes are allosteric proteins. *A quaternary structure is usually characteristic of allosteric proteins.*

Hemoglobin is a much more intricate and sentient molecule than is myoglobin. Hemoglobin, in addition to transporting oxygen, also transports CO_2, a waste product of metabolism, to the lungs to be respired. The ability of hemoglobin to bind H^+ (another waste product of metabolism) is also an important physiological function of the macromolecule, since it is essential for the maintenance of physiological pH. Also, the oxygen-binding properties of hemoglobin are regulated by interactions between separate, nonadjacent sites. *Hemoglobin is an allosteric protein, whereas myoglobin is not*. This difference is expressed in 3 ways.

1. The binding of O_2 to hemoglobin enhances the binding of extra O_2 to the same hemoglobin molecule. In other words, *oxygen binds cooperatively to hemoglobin*. In contrast, the binding of O_2 to myoglobin is not cooperative.
2. *The affinity of hemoglobin for O_2 depends on pH* (*i.e.*, *pH-dependent*), whereas that of myoglobin is independent of pH.
3. The oxygen affinity of hemoglobin is further regulated by organic phosphates such as 2,3-bisphosphoglycerate (BPG), whereas that of myoglobin is not. Thus, *hemoglobin has a lower affinity for oxygen than does myoglobin.*

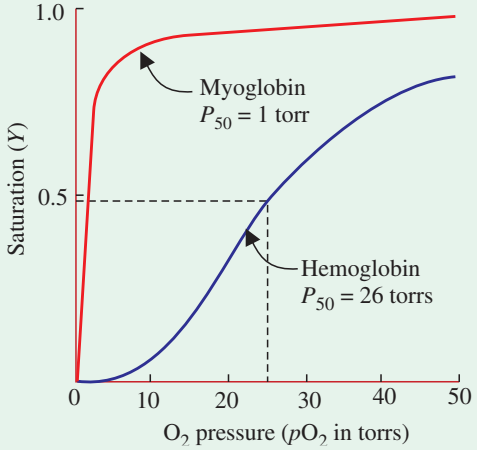


Fig. 9–47. Oxygen dissociation curves of myoglobin and hemoglobin in whole blood

Cooperativity in Oxygen-binding of Hemoglobin

The differences between hemoglobin and myoglobin as oxygen carriers are readily apparent when their oxygen dissociation curves are compared. An **oxygen dissociation curve** is a plot of the binding sites *i.e.*, heme groups with oxygen (expressed as *Y*) versus the partial pressure of oxygen (expressed as pO_2 in torr*) to which the protein solution is exposed, *i.e.*, $Y = [HbO_2] / [HbO_2] + [Hb]$. The oxygen dissociation curves of

Torr is a unit of pressure equal to that exerted by a column of mercury 1 mm high at 0°C and standard gravity; named after Evangelista Torricelli (LT, 1608-1647), the inventor of the mercury barometer.

myoglobin and hemoglobin (Fig. 9–47) differ in two ways:

A. For any given pO_2, Y is higher for myoglobin than for hemoglobin. This means that *myoglobin has a higher affinity for O_2 than does hemoglobin*. Oxygen affinity can be characterized by a quantity P_{50}, which is the partial pressure of oxygen at which 50% of sites are filled (*i.e.*, at which $Y = 50$). For myoglobin, P_{50} is typically 1 torr, whereas for hemoglobin, P_{50} is 26 torrs.

B. Secondly, *the oxygen dissociation curve of myoglobin is hyperbolic, whereas that of hemoglobin is sigmoidal* (*S-shaped*).

Because myoglobin has only a single heme group, the phenomenon of cooperative binding is not possible, as evidenced by its hyperbolic curve. It has been calculated that the cooperative binding of oxygen by hemoglobin enables it to deliver 1.83 times as much O_2 under typical physiological conditions as it would if the sites were independent.

The Bohr Effect : H+ and CO2 promote the release of O2.

Hemoglobin not only furnishes O_2 to tissues but also transports the waste products of metabolism ($H^+ + CO_2$). Since the same biomolecule is responsible for both transport systems, it is not surprising that there is a regulatory interplay between the two functions. It has been observed that increased concentrations of CO_2 and H^+ (*i.e.*, lowering of pH) decrease the O_2 affinity of hemoglobin ; conversely, increased concentrations of O_2 lower the affinity for CO_2 and H^+ (Fig. 9–48). By contrast, increased concentrations of CO_2 and H^+ have almost no effect on myoglobin's O2 affinity. Increasing the concentrations of CO_2 (at constant pH) also lowers the O_2 affinity. In actively metabolizing tissue, such as contracting muscle, much CO_2 and acid are produced. The presence of higher levels of H^+ and CO_2 in the capillaries of such metabolically active tissue promotes the release of O_2 from oxyhemoglobin. This important mechanism for meeting the higher O_2 needs of metabolically active tissues was discovered by Christian Bohr in 1904. This metabolic interdependence of the two transport functions of hemoglobin is known as Bohr effect.

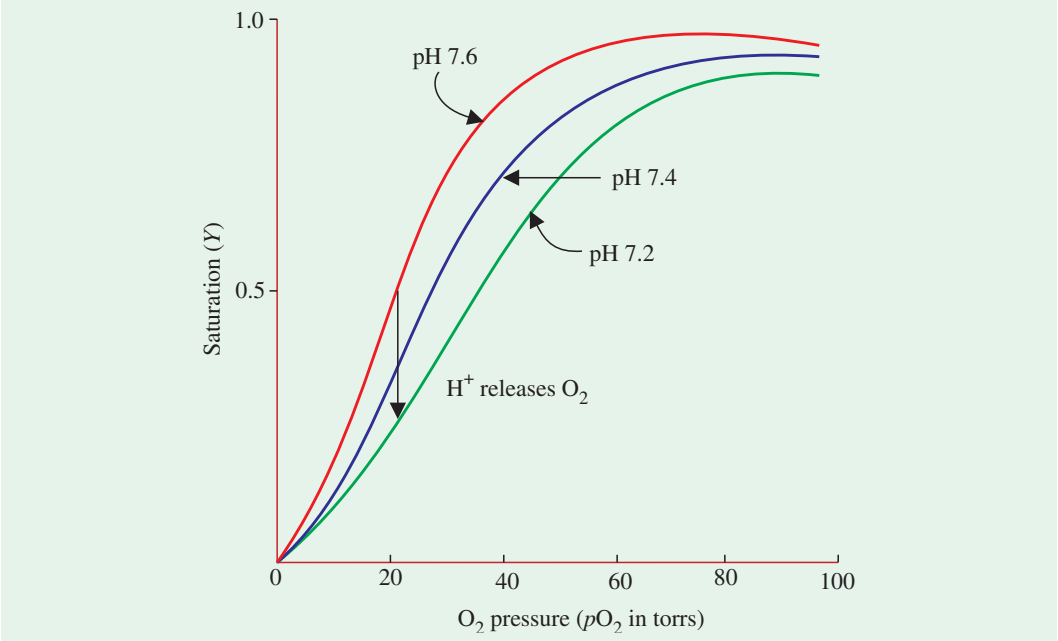


Fig. 9–48. Effect of pH on the oxygen affinity of hemoglobin

The three O_2 dissociation curves of hemoglobin at different pH have been shown. Lowering the pH from 7.5 to 7.2 results in the release of O_2 from oxyhemoglobin.

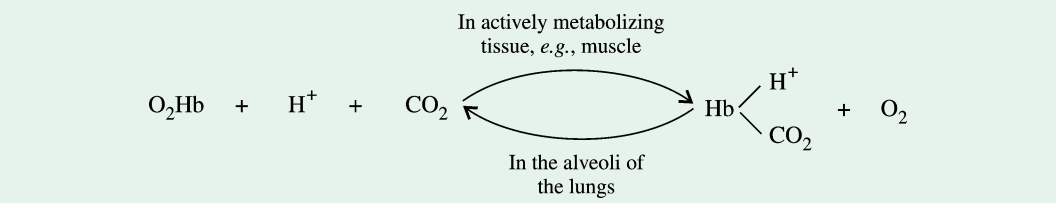


Fig. 9–49. Summary of the Bohr effect

J.B.S. Haldane, in 1914, discovered the reciprocal effect which occurs in the alveolar capillaries of the lungs. The high concentration of O_2 there unloads H^+ and CO_2 from hemoglobin, just as the high concentrations of H^+ and CO_2 in active tissues drives off O_2 (Fig. 9–49).

BPG as a Hemoglobin Regulator

Joseph Bancroft, as early as 1921, observed that the O_2 affinity of hemoglobin within erythrocytes is lower than that of hemoglobin in free solution. Much later in 1967, Reinhold Benesch and Ruth Benesch showed that an anionic organic phosphate present in human red cells, that is 2, 3-*bisphosphogylcerate, BPG* (also known as 2,*3-diphosphoglycerate, DPG*) binds to hemoglobin and thereby lowering the O_2 affinity of hemoglobin. In the absence of BPG, the P_{50} of hemoglobin is 1 torr, like that of myoglobin. In its presence, P_{50} becomes 26 torrs (Fig. 9–50). Thus, *BPG lowers the O_2 affinity of hemoglobin by a factor of 26, which is essential in enabling hemoglobin to unload oxygen in tissue capillaries*. BPG reduces the oxygen affinity of hemoglobin by binding in the central cavity of deoxyhemoglobin and not to the oxygenated form. As oxygenation occurs, the accompanying conformational changes in hemoglobin make the central cavity too small to accommodate BPG, which is then expelled.

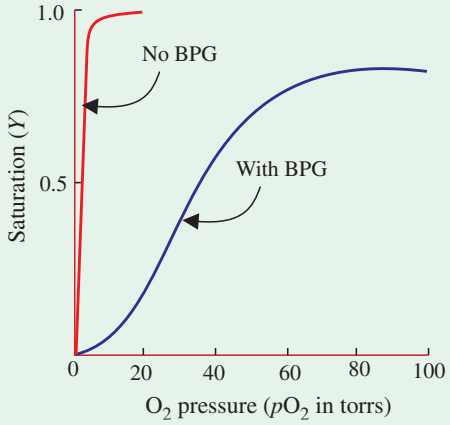


Fig. 9–50. BPG regulation of O_2 affinity of hemoglobin
Note that 2,3-bisphosphoglycerate (BPG) decreases the oxygen affinity of hemoglobin molecule.

Oxygen Affinity of Fetal Hemoglobin

Hemoglobin F has a higher O_2 affinity under physiological conditions than does hemoglobin A (Fig. 9–51). The higher O_2 affinity of hemoglobin F optimizes the transfer of oxygen from the maternal to the fetal circulation. Hemoglobin F is oxygenated at the expense of hemoglobin A on the other side of the placental circulation. In fact, hemoglobin F does not bind BPG as strongly as does hemoglobin A because the γ chain (analogous to the β chain of hemoglobin A) has a seryl residue in the H21 position instead of a positively charged histidine. Because of this particular amino acid difference, hemoglobin F has a higher affinity for oxygen than does hemoglobin A. In the absence of BPG, the O_2 affinity of hemoglobin F is actually lower than that of hemoglobin A. Thus, we see that different

forms of a protein, called **isoforms** or **isotypes**, in different tissues have a clear-cut biological advantage, as beautifully illustrated by diverse forms of hemoglobin.

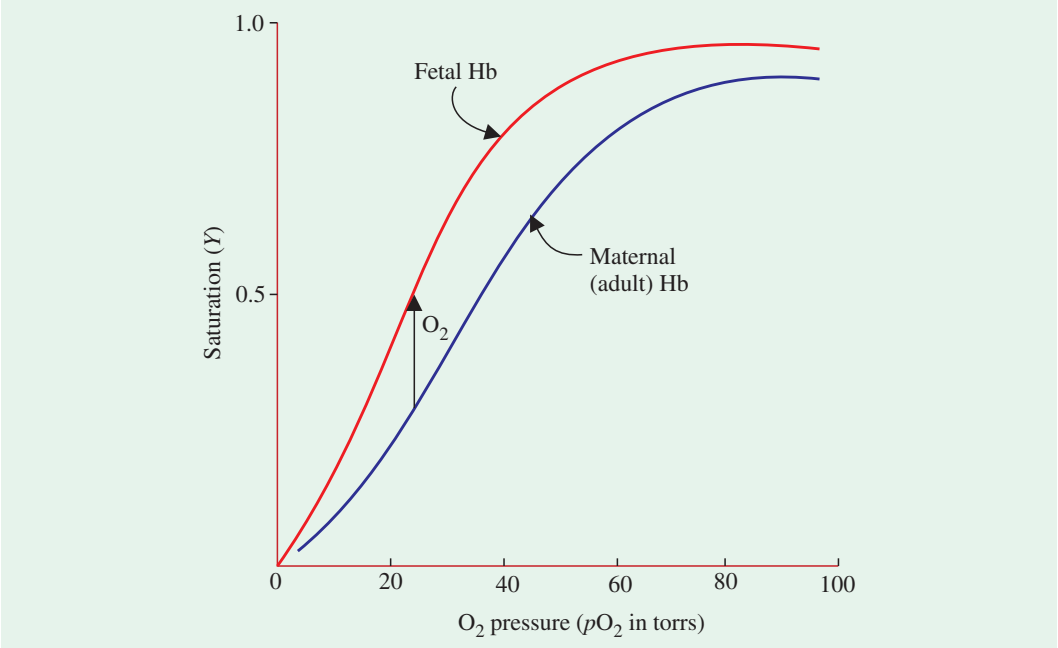


Fig. 9–51. Oxygen dissociation curves for fetal hemoglobin (HbF) and adult hemoglobin (HbA)
In the presence of BPG, the O_2 affinity of fetal hemoglobin is higher than that of maternal hemoglobin. The arrow represents transfer of oxygen from maternal oxyhemoglobin to fetal deoxyhemoglobin.

Quaternary Structural Changes in Hemoglobin (= Oxygenation of Hemoglobin)

Hemoglobin can be dissociated into its constituent chains. The isolated α chain has a high O_2 affinity, a hyperbolic dissociation curve and O_2-binding property which is indifferent to pH, CO_2 concentration and BPG level. β chains by themselves readily associate to form a tetramer (B_4). Like the α chain and myoglobin molecule, B_4 lacks the allosteric properties of hemoglobin and has a high oxygen affinity. In short, the allosteric properties of hemoglobin arise from interactions between its subunits.

Hemoglobin undergoes a major conformational change on binding oxygen, and as such oxy- and deoxyhemoglobin differ markedly in quaternary structure. In the quaternary structure of deoxyhemoglobin, there are 8 additional electrostatic interactions (salt linkages), not found in oxyhemoglobin, making deoxyhemoglobin the more rigid molecule of the two (Fig. 9–52). Six of these 8 interactions are between chains. The C-terminal residue of the 4 chains of deoxy-hemoglobin are also involved in salt linkages. As a consequence of oxygenation, a hemoglobin molecule undergoes conformational changes which disrupt 8 salt linkages and the cooperative binding observed with oxygen is the result of these structural changes.

In the oxyhemoglobin molecule, the distance between the iron atoms of the β chains decreases from 40 to 33 Å, thus making the molecule more compact. During the transition phase from oxy- to deoxyhemoglobin, large structural changes take place at two of the four contact regions (the $\alpha_1\,\beta_2$ contact and the identical $\alpha_2\,\beta_1$ contact) but not at the others (the $\alpha_1\,\beta_1$ contact and the identical $\alpha_2\,\beta_2$ contact). In fact, the $\alpha_1\,\beta_2$ contact region is designated to act as a switch between two alternative

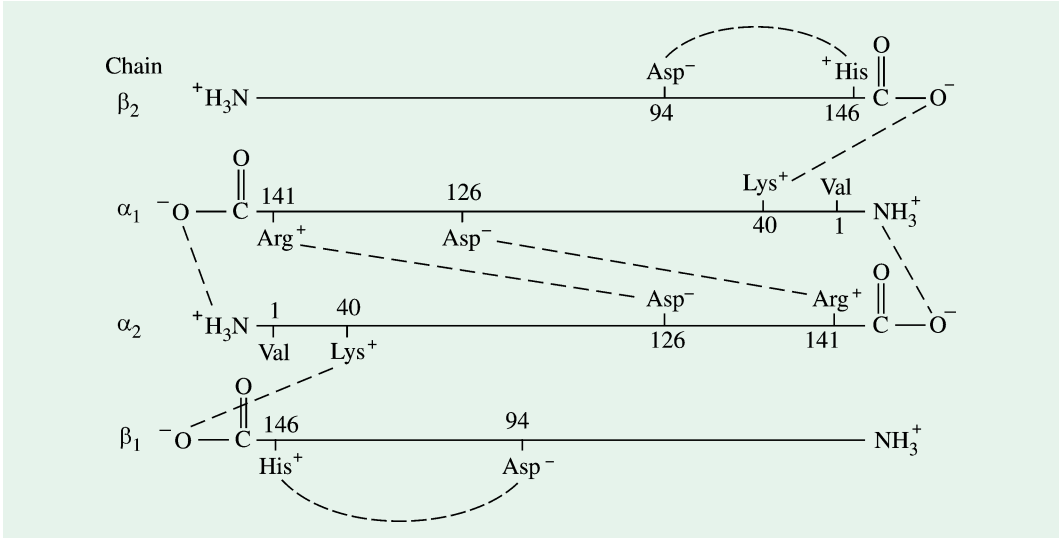


Fig. 9–52. **Schematic representation of 8 electrostatic interactions (shown by dotted lines) that occur between α and β chains of deoxyhemoglobin**

(*Adapted from Frank B. Armstrong, 1989*)

structures. In oxyhemoglobin, the C-terminal residues of all 4 chains have almost complete freedom of rotation. By contrast, in deoxyhemoglobin, these terminal groups are anchored. Deoxyhemoglobin is a tauter, more constrained molecule than oxyhemoglobin because of the presence of additional salt links. The quaternary structure of deoxyhemoglobin is termed the T (tense or tout) form ; that of oxyhemoglobin, the R (relaxed) form.

The **designations R and T** are generally used to describe alternative quaternary sturctures of an allosteric protein, the T form having a lower affinity for the substrate.

The above-mentioned conformational changes take place at some distance from the heme. But changes also take place at the heme group itself on oxygenation. In deoxyhemoglobin, the iron atom is about 0.4 Å out of the porphyrin plane toward the proximal histidine, so that the heme group domed (convex) in the same direction (Fig. 9–53). On oxygenation, the iron atom moves into the plane of the porphyrin to form a strong bond with O_2, and the heme becomes more planar. Besides, His F8, because of its bonding to the iron atom, also moves toward the heme plane. This slight displacement of the histidine residue initiates a sequential series of conformational changes that ultimately results in the disruption of some of the subunit interactions in the quaternary structure of deoxyhemoglobin.

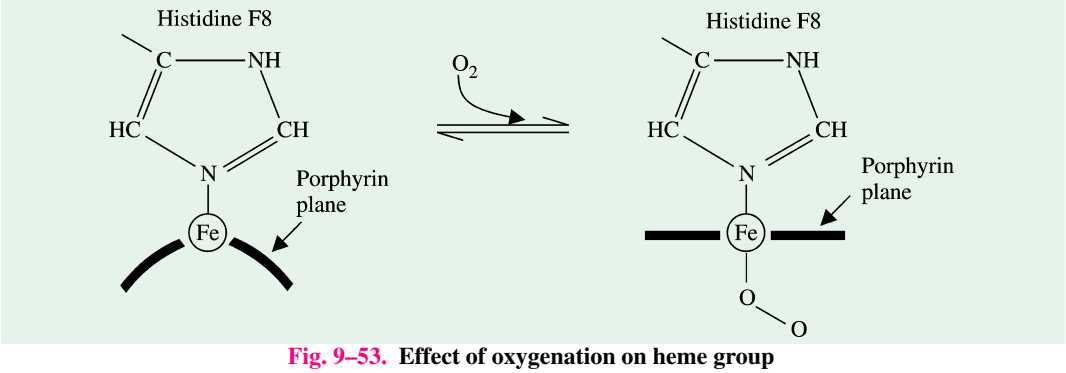


Fig. 9–53. **Effect of oxygenation on heme group**

The iron atom moves into the plane of the heme group on oxygenation. The proximal histidine (F8) is pulled along with the iron atom and becomes less tilted.

Mechanism of Allosteric Interactions of Hemoglobin

The structures of fully deoxygenated and fully oxygenated hemoglobin provide insight into how the binding of O_2, CO_2, H^+ and BPG influence each other. How does allosteric mechanism take place? How does hemoglobin switch from the deoxy to the oxy structure when it binds successive O_2 molecules? Two models – the sequential model and the concerted model – have been proposed to explain the allosteric changes in multi-subunit protein. Both these models are based upon the ideas that the proteins are flexible in conformation and that a protein subunit can exist in only two states.

A. **Sequential Model.** Developed by Daniel E. Koshland Jr. in 1966, the sequential model, in its simplest form, makes 3 ***assumptions*** **:**

1. Only two conformational states, T and R, are accessible to any subunit.
2. The binding of the ligand switches the conformation of the subunit to which it is bound but not that of its neighbours.
3. The conformational change elicited by the binding of substrate in one subunit can increase or decrease the binding affinity of the other subunit in the same molecule. A T subunit with an R neighbour has higher affinity for ligand than does a T subunit with a T neighbour because the subunit interfaces are different in TR and TT.

The sequential model (Fig. 9–54), in its essence, proposes that *symmetry in not conserved in allosteric transitions and that the subunits change conformation one at a time*. The T state is symbolized by a square, and the R state by a circle. Deoxyhemoglobin is in the T_4 state. The binding

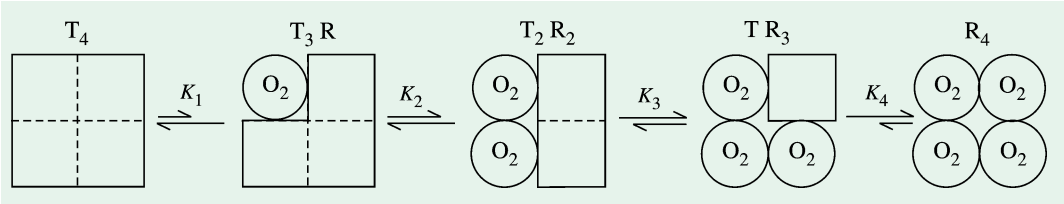


Fig. 9–54. Simple sequential model for a tetrameric allosteric protein

The binding of a ligand to a subunit changes the conformation of that particular subunit from the T (square) to the R (circle) form. This transition affects the affinity of the other subunits for the ligand.

of O_2 to one of the subunits changes its conformation from T to R, but leaves the other subunits in the T form. The oxygen-binding affinity of unoccupied sites in RT_3 is higher than in T_4 because some salt links have been broken on binding the first O_2. R_2T_2 and R_3T, which have higher oxygen affinities than does RT_3, are formed when the second and third O_2 bind respectively. Finally, R_4 is produced on binding the fourth O_2. Thus, according to the sequential model, the affinity of hemoglobin for successive oxygen molecules increases because fewer salt bridges need be broken. The model can be better understood in terms of a ***postage-stamp analogy*** (refer Fig. 9–55). Two perforated edges (*i.e.*, sets of salt bridges) must be torn to remove the first stamp. Only one perforated edge must be torn to remove the second stamp, and one edge again to remove the third stamp. The fourth stamp is then free.

Fig. 9–55. Postage-stamp analogy of the simple sequential model

B. Concerted Model (or **Symmetry Model).** Developed by Jacques Monod, Jeffries Wyman and Jean-Pierre Changeux (hence, also called **MWC model**) in 1965, the concerted model takes a different view of allosteric interactions. The essence of this elegant and incisive model (Fig. 9–56) is that *symmetry is conserved in allosteric transitions* and that *all the subunits change conformation together.* This model also is based on 3 ***assumptions***:

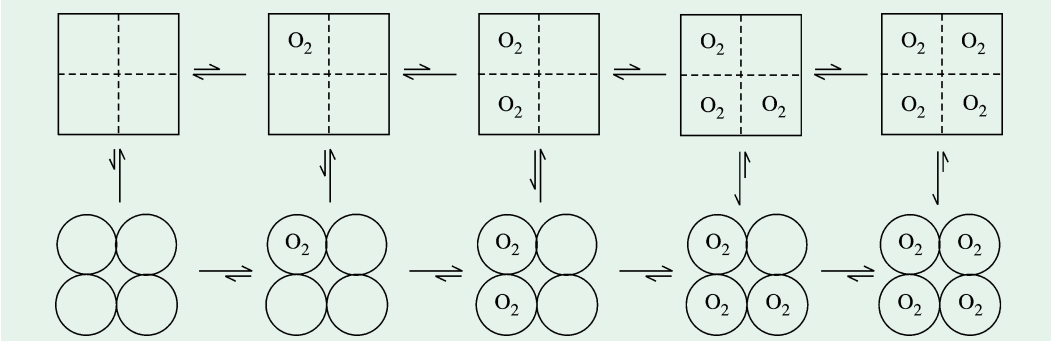


Fig.9–56. Concerted model (or MWC model) for a tetrameric allosteric protein

The squares denote the T form, and the circles denote the R form.

1. The protein interconverts between two conformations, T and R. All the subunits of a particular molecule must be in the T form, or all must be in R form. Hybrids such as TR are forbidden.
2. Ligands bind with low affinity to the T form, and with high affinity to the R form.
3. The binding of each ligand increases the probability that all subunits in that molecule are in the R form. The allosteric transition is said to be *concerted* because subunits change in unison from T to R or vice versa. Stated in terms of the postage-stamp analogy, either all four perforated edges or none are broken.

It is, thus, apparent that the two models offer contrasting views regarding the mode of cooperative interactions in multi-subunit protein. The models differ in many points through:

1. The symmetry model assumes that the active and inactive states are in equilibrium, whereas sequential model assumes that the transition from inactive state to active state in *induced* by substrate binding.
2. The symmetry model states that if one subunit changes its form, the other subunit necessarily changes to that form also because, unless symmetry is preserved, the two subunits cannot interact. By contrast, the sequential model allows interaction between subunits in *different* states.
3. The symmetry model assumes that binding of the first molecule necessarily enhances binding of a second molecule of the same type, whereas the sequential model allows *either an increase or decrease* of affinity for the second molecule.
4. The symmetry model has few intermediate states, whereas the sequential model has *more* intermediate states.

In fact, the actual allosteric mechanism of hemoglobin is more complex than visualized by either the sequential or the concerted model. These models should be regarded as limiting cases. Actual allosteric processes combine, of course in varying degrees, elements of both.

The *protein-protein interactions*, thus, serve many important functions :

(*a*) protect nucleic acids of virus particles from destruction,

(*b*) conserve genetic information, and

(*c*) construct the various enzyme-complexes found in the cells.

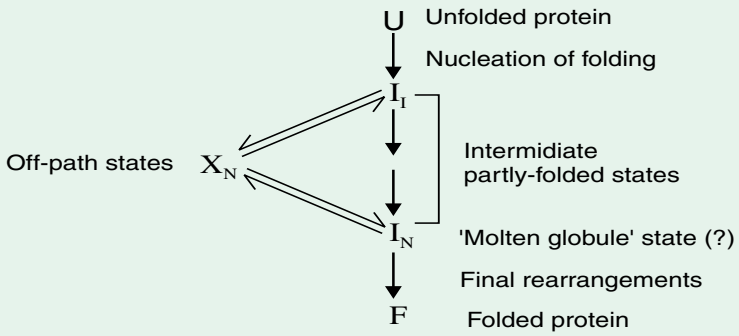
The foregoing discussion and also the evidences presently available confirm the view that "the primary structure of the protein, which dictates the limitations of secondary and tertiary structure, also delineates the correct quaternary structure necessary for biological activity of the molecule." (Hartman and Suskind, 1969).

DYNAMICS OF GLOBULAR PROTEIN STRUCTURE

The precise pictures produced by x-ray diffraction give the impression of a rigid and inflexible structure for the globular proteins. Recent researches, however, indicate that globular proteins fold *via* complex kinetic pathways and that even the folded structure, once acquired, is a dynamic structure. Some of the dynamic aspects of globular proteins are described below :

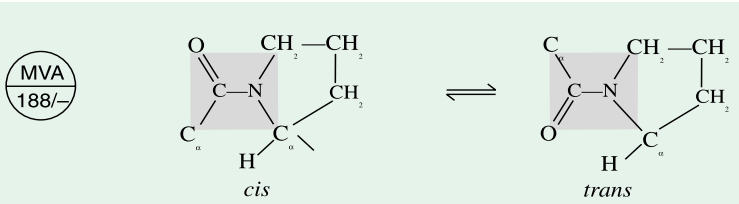
1. Kinetics of Protein Folding

The folding of globular proteins from their denatured states (or conformations) is a remarkably rapid process and often completes in less than a second. This viewpoint was first expressed in 'Levinthal's paradox', first propounded by Cyrus Levinthal in 1968. A rough estimate reveals that about 10^{50} conformations are possible for a polypeptide chain such as ribonuclease (124 residues). And if the molecule tries a new conformation every 10^{-13} second, it would still take about 10^{30} years to try a significant fraction of them. In actuality however, ribonuclease folds experimentally *in vitro* in about 1 minute. To explain this anomaly, it has been suggested that folding takes place through a series of intermediate states. These observations led to the **path way model of folding,** as depicted below :



The nucleation step is critical because it is much more difficult to begin an α helix than to extend it (note that at least 4 residues must fold properly to make the first stabilizing H bond). On the contrary, it is now recognized that nucleation may begin at many places and that all of these partially folded structures will be 'funneled' by energy minimizations toward the final state. The funneling model proposes that there is not just one but *many* possible paths from the denatured state to the folded state, and each path leads downhill in energy. During the descent toward the free energy minimum, there may be pauses corresponding to metastable intermediates, as incorporated in the pathway model. An important folding intermediary for many proteins appears to be what is called as the "molten globule" state. It is a compact structure in which much of the secondary and tertiary folding has occurred, but the internal hydrophobic residues have not yet settled into their final packing.

Evidence suggest for the presence of "off-path" states– those in which some key element is incorrectly folded. However, the cell has ways to assist incorrectly-folded proteins to find their proper conformations. Larger complex proteins may never achieve 100% folding success ***in vivo*** but the cell can identify those which are incorrect and dispose of them. One such common folding error occurs *via* the incorrect *cis-trans* isomerization of the amide bond adjacent to a proline residue :



Unlike other peptide bonds in proteins, for which the *trans* isomer is highly favoured (by a factor of

about 1,000), proline residues favour the *trans* form in the preceding bond by a factor of only 4. Hence, there are more chances that the 'wrong' isomer (in terms of the functional protein conformation) will form first; this will later be converted to the correct configuration involving chain rearrangement enzymatically to speed up *in vivo* folding.

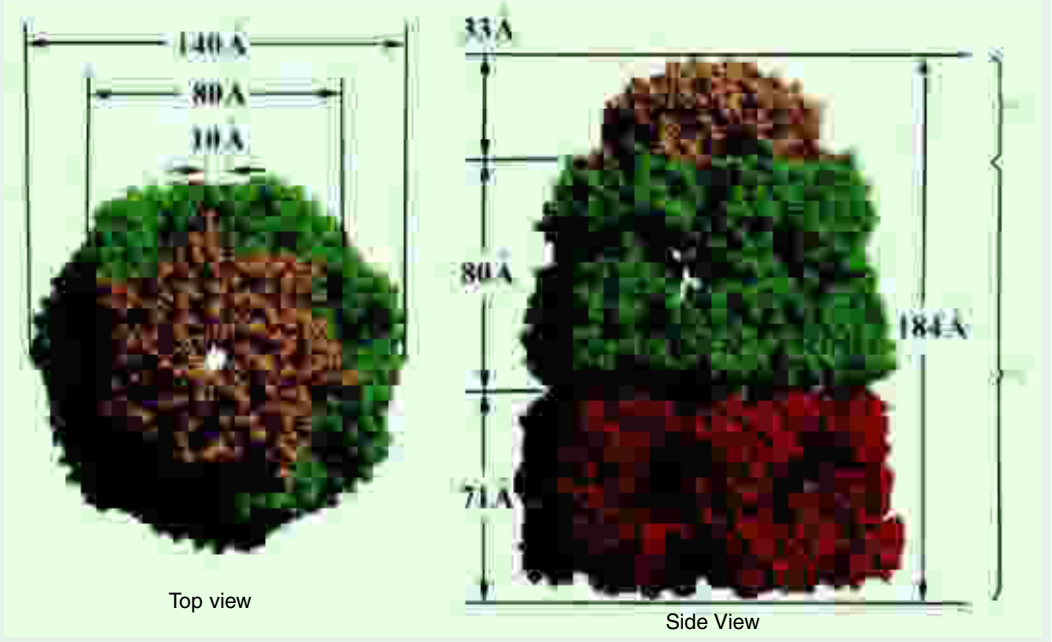
2. Kinetics of Disulfide Bond Formation

If a protein is folded from a state in which disfulfide bonds have been disrupted and must hence be re-formed, the process is more complicated and slower– often taking many minutes. Some disulfide bonds that are missing in the native structure are formed in intermediate stages of the folding. Obviously, the protein can utilize a number of alternative pathways to fold but ultimately finds both its proper tertiary structure and the correct set of disulfide bonds. This process is aided *in vivo* by enzymatic catalysis of —S—S— bond rearrangement.

3. Chaperonins

It has been discovered that protein folding and assembly *in vivo* sometimes requires the aid of special proteins called *chaperonins* or *molecular chaperones*. As the name signifies, the function of these chaperones is to keep the newly-formed protein away from improper folding or aggregation. Improper folding may correspond to being trapped in a deep local minimum on the energy scenario. Aggregation is often a danger because the protein, released from the ribosome in an unfolded state, will have groups exposed. These will be tucked inside in normal folding, but when exposed they stand the chance of making hydrophobic interactions with *other* polypeptide strands and thereby aggregating.

Many chaperonin systems have been discovered but the GroEL-ES complex from *E. coli* is the best-studied one. The structure of this enormous complex has recently been elucidated by x-ray diffraction [Fig. 9–57(*a*)]. It consists of 2 basic portions – GroEL and GroES. GroEL is made up of 2 rings, each consisting of 7 protein molecules; the centre of each ring is an open cavity, accessible to the solvent at the ends. Either cavity can be 'capped' with GroES, which is again a 7-membered ring of smaller subunits.



(a)

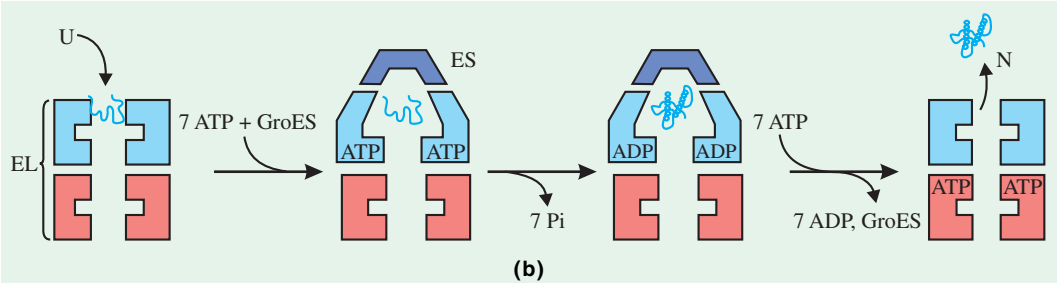


Fig. 9–57. The GroEL–GroES chaperonin complex

(*a*) X-ray diffraction structure of the El-ES $(ADP)_7$ complex

(*b*) A schematic of the function of GroEL-ES complex

The unfolded protein enters into a sort of cavity lined with hydrophobic residues. Then the cavity changes, presenting a hydrophilic lining. This releases the protein from the walls and it folds and is then released. Note that ATP is required, probably to drive the process in one direction.

(Adapted from (a) Xu Z, Horwich AL and Sigler. PB, 1997 and (b) Netzer and Hartl, 1998)

It is argued that basically, the cavities provide 'shelters' in which nascent protein chains can be 'incubated' until they have folded properly. The GroEL-ES complex does not stipulate the folding pattern *i.e.*, upto the protein itself to do. But insulation from the environment prevents chances of aggregation or misfolding. The cycle experienced by a protein molecule is schematically shown in [Fig. 9–57(*b*).]

The conformational changes in GroEL are diagrammatically represented in (Fig 9–58).

The reconstructions of the GroEL and GroEL–GroES complexes, based on high resolution electron micrographs, have been presented in (Fig. 9–59).

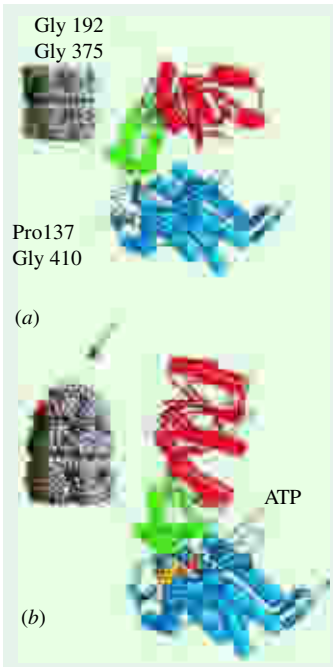


Fig. 9–58. Conformational change in GroEL

(*a*) The drawing on the left shows a surface view of the two rings that make up the GroEL chaperonin. The drawing on the right shows the tertiary nstructure of one of the subunits of the top GroEL ring. The polypeptide chain can be seen to fold into three domains

(*b*) When a GroES ring (arrow) binds to the GroEL cylinder, the apical domain of each GroEL subunit of the adjacent ring undergoes a dramatic rotation of approximately 60° with the intermediate domain (shown in green) acting like a hinge. The effect of this shift in parts of the polypeptide is a marked elevation of the GroEL wall and enlargement of the enclosed chamber.

(Adapted from Xu, Z, Horwich AL and Sigler PB, 1997)

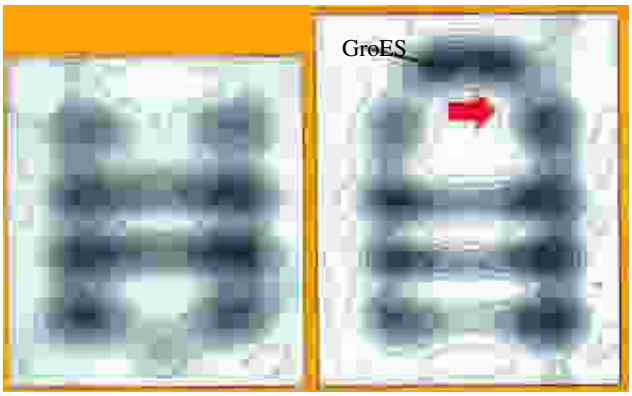


Fig. 9–59. GroEL-GroES complex

Reconstructions of the GroEL and GroEL-GroES complex, based on high resolution electron micrographs, taken of specimens that had been frozen in liquid ethane and examined at – 170°C

The GroEL complex with GroES appears as a dome on one end of the cylinder. It is evident that the binding of the GroES is accompanied by a marked change in conformation of the apical end of the proteins that make up the top GroEL ring (arrow), which results in a marked enlargement of the upper chamber. *(Adapted from Chens e tal, 1994)*

However, processing *via* chaperonins is not a universal phenomenon as only a small fraction of the proteins made in *E. coli* (or any cell) are processed *via* chaperonins. Some proteins are too large to be accomodated within the cavity whereas others fold safely on their own.

4. Motion within Globular Proteins

Evidences accumulated indicate that the folded globular protein molecules are not static and they continually undergo a wide variety of internal motions. These motions are due to the interactions of protein molecules with their environment. The resulting motions can be roughly grouped into 3 classes (Table 9–9).

Table 9–9. Motions within globular protein molecules

Class	*Type of Motion*	*Approximate Amplitude (nm)*	*Range Time(s)*
1	Vibrations and oscillations of individual atoms and groups	0.2	10^{-15}–10^{-12}
2	Constructed motions of structural elements, like α helices and groups of residues	0.2–1	10^{-12}–10^{-8}
3	Motions of whole domains; opening and closing of clefts	1–10	$\geq 10^{-8}$

Class 1 motions occur even within protein molecules in crystals and account, in part, for the limits of resolution obtainable in x-ray diffraction studies. In **class 2** and **class 3**, the motions are larger in magnitude and slower in rate. These are more likely to occur in solution. Some of the motions, like the opening and closing of clefts in molecules, are probably involved in the enzymatic functions. Binding or release of a small molecule from a protein depends on the time required to open or close a shaft. Likewise, the protein 'gates' that pass molecules and ions through membranes rapidly change from open to closed states.

5. Prions

Until very recently, scientists believed that the diseases could be transmitted from one organism to another *via* viruses or microorganisms. After all, DNA and RNA were the carriers of genetic information. But now evidences have gathered to point out that some diseases are transmitted by a protein and nothing more. Table 9–10 lists some such diseases (called **prion diseases**) along with their host and geographic distribution.

In 1900, a fatal neurodegenerative disease called **Kuru** was reported only in a tribe of Papua New Guinea. Carleton Gajdusek, a virologist at the US National Institute of Health, discovered and understood the basis of the disease and the mode of its transmission. He pointed out that a peculiar ritual of the tribal people of eating the brain of dead relatives was the main mode of transmission of this disease. The frequency of Kuru gradually declined since the practice was abandoned. Gajdusek was awarded 1976 Nobel prize in medicine or physiology for this basic work, along with Baruch Blumberg, a virologist of the Institute of Cancer Research, Philadelphia.

But Stanley B. Prusiner, an Americal biochemist, was not convinced about the nature of the infectious agent which was described by Gajdusek and others as 'unconventional slow virus'. However, Prusiner was against the so-called virus theory because of the following 3 reasons :

(*a*) The infective agent is extremely resistant to UV and ionizing radiations whereas all viruses are sensitive.

(*b*) The infectivity of the scrapie agent is not changed by treatment with nucleases (DNAase and RNAase enzymes). That means the infective agent does not carry DNA or RNA as

Table 9–10. Prion disease in different species

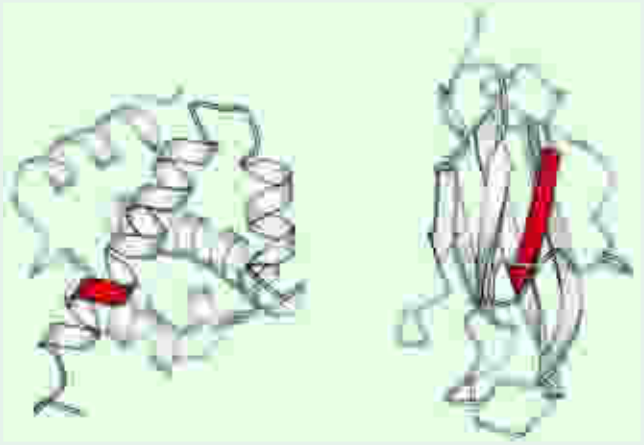
Host	*Disease*	*Geographic distribution*	*First clinical observation*
Sheep	Scrapie	Cosmopolitan except Australia, New Zealand and some European countries	1730
Goat	Scrap	—	—
Man	Kuru ('Laughing death')	Papua New Guinea	1900
	Creutzfeldt-Jacob disease	Cosmopolitan	1920
	Gertsmann-Straussler–Scheinker syndrome	Cosmopolitan	1926
	Fatal familial insomnia	—	—
Mink	Transmissible mink encephalopathy	North America; Europe	1947
Mule deer	Chronic wasting disease	North America	1967
Cattle	Bovine spongiform encephalopahy, BSE (= Mad cow disease)	U.K., Ireland and some other European countries	1985

genetic material which is the most essential component of a virus particle.

(*c*) Absence of any agent-specific antibody titre, strong tendency to aggregate and hydrophobicity go against the nature of a true virus.

Backed by these evidences, Prusiner and his colleagues in 1982, extracted the infectious material from hamster brains, and they gave the first blow to the virus theory. Prusiner suggested that the infectious agent, causing certain degenerative disorders of central nervous system (CNS) in animals and more rarely in humans, is a small proteinaceous infectious particle, which he called *prion* (pronounced as 'preeon') and the protein believed to be responsible for infection was called *prion-related protein* or *PrP*. Soon he found that there were 2 isoforms of PrP. The first one is the normal cellular **prion-related protein, PrP^{c}** (the superscript '*c*' denotes cellular), which is the nonpathological form; the PrP^{c} may act as an acetyl-choline receptor inducer and plays an important role in the transmission of nerve signal. The other isoform of PrP is infective and called **prion-related protein scrapie, PrP^{sc}** (the superscript 'sc' denotes scrapie, which is now used to refer to all infectious form of prion causing scrapie- like diseases in animals and humans). It is this form, in which the disordered N-terminal portion appears to fold into a β-sheet, that wrecks havoc with the nervous system. It is postulated that when PrP^{sc} is ingested into the body, it induces the conversion of PrP^{c} in the recipient to PrP^{sc}; thus the disease is transmitted. It is something like a bad guy, who converts a good guy into one of its 'bad' form. How this conversion is catalyzed is unknown, but it strongly suggests that PrP^{c} represents an especially stable "off-path" folding of the type hypothesized in the preceding section. Fig. 9–60 presents the computer-generated images of the two forms of human prion protein.

In 1988, Prusiner and his team reported that human prion diseases can certainly be inherited, *i.e.*, they could be heritable and communicable. Later studies on transmission of human prion to transgenic mice threw more light on prion diseases in 1994. When researchers tried to infect transgenic mice (carrying the human PrP gene) with human prion, there was no development of CNS dysfunction as expected. The mice could become susceptible to human prions only after removal of mouse PrP gene which is called **'gene knockout'** in genetics terminology.



(a) Normal prion protein, PrPc

(b) Abnormal prion protein, Pr P^{sc}

Fig. 9–60. Computer-generated images of the two forms of human prion protein.

Unfortunately, prion diseases have already played havoc among cattle population in some European countries. **Bovine spongiform encephalopathy (BSE),** or more commonly known as **mad cow disease,** took its toll in cattle population in Britain and elsewhere. Epidemiological studies revealed that the most probable cause of the epidemic was the inclusion of ruminant-derived meat and bone meal (which utilizes sheep brain, spinal cord etc.) in the cattle feed. Things became worse as reports of 12 persons afflicted with Creutzfeldt-Jakob disease came in. It was reported that these new cases occurred due to consumption of BSE-infected beef. These reports sent a shock wave throughout the world, apprehending the possible impact of prion diseases.

The recognition of the relationship of PrP to diseases won Prusiner the Nobel prize in physiology or medicine in 1997. His discovery of a new biological principle of infection goes by the name **'prion theory'.**

PREDICTION OF SECONDARY AND TERTIARY PROTEIN STRUCTURES

A. Prediction of Secondary Structure

The protein amino acids arrange in innumerable ways, with the help of a variety of chemical bonds, to produce a definite pattern of secondary structure in proteins, which may be either an α helix, or a β sheet or a turn. Acquisition of any one, of these 3 forms (α helix, β sheet, turn) by a protein depends upon the frequency of occurrence of particular amino acid residues in these secondary structures. Table 9-11. lists the relative frequencies (P_α, P_β, P_t) of amino acids in producing secondary structures, *i.e.*, α helix, β sheet and turn, respectively.

A perusal of the table (on the next page) reveals that Met, Glu, Ala, and Leu residues tend to be present in α helices, whereas Val, Ile, Phe and Tyr tend to be present in β strands. Pro, Gly, Asp, and Ser have a propensity (or inclination) to lie in turns.

Table 9–11. **Relative frequencies of amino acid residue occurrence in secondary structures of different globular proteins.**

Amino acid	*α helix* (P_α)	*β sheet* (P_β)	*Turn* (P_t)	
Alanine (Ala)	**1.29**	0.90	0.78	
Cysteine (Cys)	**1.11**	0.74	0.80	
Leucine (Leu)	**1.30**	1.02	0.59	
Methionine (Met)	**1.47**	0.97	0.39	Favour
Glutamic acid (Glu)	**1.44**	0.75	1.00	α-helices
Glutamine (Gln)	**1.27**	0.80	0.97	
Histidine (His)	**1.22**	1.08	0.69	
Lysine (Lys)	**1.23**	0.77	0.96	
Valine (Val)	0.91	**1.49**	0.47	
Isoleucine (Ile)	0.97	**1.45**	0.51	
Phenylalanine (Phe)	1.07	**1.32**	0.58	Favour
Tyrosine (Tyr)	0.72	**1.25**	1.05	β sheets
Tryptophan (Trp)	0.99	**1.14**	0.75	
Threonine (Thr)	0.82	**1.21**	1.03	
Glycine (Gly)	0.56	0.92	**1.64**	
Serine (Ser)	0.82	0.95	**1.33**	Favour
Aspartic acid (Asp)	1.04	0.72	**1.41**	turns
Asparagine (Asn)	0.90	0.76	**1.28**	
Proline (Pro)	0.52	0.64	**1.91**	
Arginine (Arg)	0.96	0.99	0.88	

* Note that arginine shows no significant preference for any of the structures.

(*Adapted from Creighton TE, 1992*)

There are some obvious reasons for these preferences :

1. The α helix can be regarded as the default conformation. Branching at the β carbon atom (as in **Val**, **Thr** and **Ile**) tends to destabilize α helices because of steric clashes. These residues are easily accomodated in β strands, where their side chains project out of the plain containing the main chain.
2. **Ser, Asp** and **Asn** tend to disintegrate α helices because their side chains contain H bond donors or acceptors in close proximity to the main chain, where they compete for the main-chain NH and CO groups.
3. **Pro** tends to disrupt both α helices and β strands because it lacks an NH group and because its ring structure restricts its ϕ value to near -60 degrees.
4. **Gly** readily fits into all structures and hence does not favour helix formation, in particular.

It is worthmentioning that the conformational preferences of amino acid residues are not toppled over all the way to one structure, as seen in Table 9–11. As an instance, glutamic acid, one of the strongest helix formers, prefers α helix to β strand by only a factor of (1.44/0.75) about 2. The preference ratios of most other amino acid residues are smaller; for example, for methionine it is

(1.47/0.97) about 1.5 and for alanine it is (1.29/0.90) about 1.25. Indeed, some penta- and hexapeptide sequences have been found to adopt one structure in one protein and an entirely different structure in another protein (refer Fig. 9–51). Hence, some amino acid sequences do not uniquely determine secondary structure of proteins. Tertiary interactions between residues that are far apart in the sequence may be decisive in specifying the secondary structure of some segments.

It is now possible to predict the protein secondary structure with moderate accuracy. Based on the relative frequency values of the constituent amino acids, P.Y. Chou and G.D. Fasman (1974) have framed certain rules for the prediction of globular protein secondary structures. Following are the 3 **Chou-Fasman rules** for prediction :

1. Any segment of 6 residues or more, with $(P_\alpha) \geq 1.03$, as well as $(P_\alpha) > (P_\beta)$, and not including Pro, is predicted to be α helix.
2. Any segment of 5 residues or more, with $(P_\beta) \geq 1.05$, and $(P_\beta) > P_\alpha$, is predicted to be β sheet.
3. Examine the sequence for tetrapeptides with $(P_\alpha) < 0.9$, $(P_t) > (P_\beta)$. They have a good chance of being turns. The actual rules for predicting β turns are more complex, but this method will work in most cases.

B. Prediction of Tertiary Structure

Prediction of tertiary structure of proteins is much more difficult because the higher-order folding depends so critically on specific side chain interactions, often between residues far removed from one another in the sequence. However, recent recognition of overall patterns in tertiary folding has given some success. For example, predictions of secondary structure have been used to predict an α/β barrel structure for the enzyme **tryptophan synthase**, that is in excellent harmony with x-ray result.

The method for predicting the tertiary structure of proteins depends on the fact that, in their spontaneous folding, proteins are seeking a free energy minimum. A random-coil chain is allowed, in computer simulation, to undergo a large number of small permutations in its configuration, through rotation about individual bonds. The computer programme keeps track of the total energy, in terms of possible interactions, and seeks an energy minimum. This approach is still in its infancy.

CLEAVAGE OF A PROTEIN

Certain proteolytic enzymes and chemical reagents cleave protein between specific amino acid residues (Table 9–12). The enzyme *trypsin*, for example, cuts on the carboxyl side of lysine or arginine, whereas the chemical *cyanogen bromide* splits peptide bonds next to methionine residues. Since these enzymes and chemicals cleave at relatively few sites in a protein molecule, they tend to create relatively large and relatively few peptides. If such a mixture of peptides is separated by biochemical procedures (chromatography, electrophoresis etc), a characteristic pattern or **peptide map** would be obtained which will be diagnostic of the protein from which the peptides were generated. This peptide map is, sometimes, referred to as the **protein's fingerprint.**

Table 9–12. Some reagents frequently used to cleave peptide bonds in proteins*

Reagent	*Amino acid 1*	*Amino acid 2*
Enzymes		
Trypsin	Lys or Arg	Any
Chymotrypsin	Phe, Trp or Tyr	Any
V8 protease	Glu	Any
Chemicals		
Cyanogen bromide	Met	Any
2-nitro-5-thiocyano-benzoate	Any	Cys

* The specificity for the amino acids on either side of the cleaved bond is indicated. The carboxyl group of amino acid 1 is released by the cleavage; this amino acid is to the left of the peptide bond as normally written.

BIOLOGICAL ROLES OF PROTEINS

Proteins are of utmost significance to biological systems. These are most critical to life and perform various functions. Some of their roles are given below.

1. Many proteins act as **catalysts**, thus usually enhancing the rate of chemical reactions to such extents as needed by the living cells.

2. The fibrous proteins serve as **components of the tissues** holding the skeletal elements together. Collagen is a structural unit of connective tissues.

3. The nucleoproteins serve as **carriers of genetic** characters and hence govern inheritance of traits.

4. Proteins also perform **transport functions.** Many compounds enter the cells and accumulate inside at much higher concentrations than expected from diffusion alone. These changes require the input of energy and are usually termed active transport. The mechanism of active transport involves proteins either as catalysts or as adsorbents or as both.

5. Various protein **hormones** are known. These regulate the growth of plants and animals, besides controlling many other physiological functions.

6. Under conditions of non-digestion and no chances for denaturation, the proteins accumulate inside the cells and produce **toxicity**. Venoms of snakes and insects are injected by biting into the blood. Certain of the foreign proteins present in venom are actually enzymes. These enzymes attack body tissue causing destruction of blood cells leading ultimately to death.

In other cases, the toxicity from foreign proteins results from responses by the affected animals. Allergic reactions exemplify this category. These reactions occur when an animal is exposed to a foreign protein to which it has been sensitized by prior exposure. The toxic disorders include skin blisters, swelling of limbs and respiratory congestion, leading sometimes to death.

7. Blood plasma, which is obtained after removal of the blood cells by centrifugal action, is essentially a solution of proteins in water. It is used for the **treatment of shock** produced by serious injuries and operations.

8. Interferon (IF or **IFN)** is a generic term which applies to a number of (about over 20) related low molecular weight, regulatory glycoproteins produced by many eukaryotic cells in response to numerous inducers : a virus infection, double-stranded RNA, endotoxins, antigenic stimuli, mitogenic agents, and many parasitic organisms capable of intracellular growth (*Listeria monocytogenes*, chlamydiae, rickettsias, protozoa). They are effective in treating viral diseases and cancer, and in eliminating its side effects. The most widely-studied property of interferons is their ability to 'interfere' (hence, their nomenclature) with the replication of viruses. *They are usually species-specific but virus-nonspecific.* Interferon was discovered in 1957 in London by the British virologist Alick Issacs and a visiting scientist from Switzerland Jean Lindenmann, both of the National Institute for Medical Research, London. They found an agent responsible for such viral interference ; a protein released by cells exposed to a virus, that enables other cells to resist viral infection. They called it interferon (IF). Since then, several different classes of interferons have been identified. Human interferons have been classified into 3 classes, depending on which type of they are produced by :

(*a*) Fibroblasts (F) or alpha interferons (IFN-α)

(*b*) Leukocytes (L) or beta interferons(IFN-β)

(*c*) Immune type (I) or gamma interferons (IFN-γ)

In 1978, a Tokyo metropolitan medical team has cultivated this "wonder substance" using cells from the placenta taken at the time of birth (or parturition). Large-scale production of this protein is possible because the human cells reproduce in tissue culture. Because of their antiviral and antiproliferative activities, interferons have been considered important therapeutically in treating viral

diseases such as hepatitis, encephalitis, cancer and even common cold. The IFs may not be a magic bullet like penicillin against all viruses, yet it may be a very useful addition to the armamentarium already available against viruses in general, and cancer in particular.

9. Peptides from humans called *defensins* have been found to be **antibiotic** in nature. Produced by the immune system, these cells smother and kill the invading pathogens. They are secreted by the epithelial cells lining the moist body surface of mammals and serve as the body's own disinfectants.

10. Another group of peptides called *endorphins* are found in the brain and are involved in the **suppression of pain**, creation of euphoric highs and feelings of joy. Candace Pert, a codiscoverer of the body's opiates, says "They reveal the harmony as well as the economy inherent in nature. Nothing is wasted". She adds "Once a molecule is found effective, it tends to be used over and over again in the ladder of life, in newer and ever more exciting configurations."

11. The frog's secretion *adenoregulin*, a 33-amino-acid-long peptide, works on the receptors in the brain which handle adenosine. Adenosine is a fundamental component in all human cell fuel. And the frog peptide seems to **enhance the binding of adenosine** by subtly altering the receptor in the brain.

12. The American scientists at the Chicago Medical School and the Harvard Medical School, in 1982, have isolated about 30 g of a glycopeptide, known as *S factor*, from 4.5 tones of urine from healthy males. The S factor is composed of alanine, glutamic acid, diaminopimelic acid, and muramic acid. The S factor acts **as a soporific,** *i.e.*, as a sleep-promoting agent and has little side effects. The researchers infused a very light concentration of the S factor into the brain of the rabbit and it was found that it induced a 50% increase in what is known as *slow wave sleep*— a deep dream-free sleep that occurs in animals and humans after sleep derivation and is normal as judged by various criteria.

REFERENCES

See list following Chapter 11.

PROBLEMS

1. (*a*) Tropomyosin, a 70-kd muscle protein, is a two-stranded α-helical coiled coil. Estimate the length of the molecule ? (*b*) Suppose that a 40-residue segment of a protein folds into a two-stranded antiparallel β structure with a 4-residue hairpin turn. What is the longest dimension of this motif ?

2. Glycine is a highly conserved amino acid residue in the evolution of proteins. Why ?

3. Identify the groups in protein that can form hydrogen bonds or electrostatic bonds with an arginine side chain at pH 7.

4. The shape of hair is determined in part by the pattern of disulfide bonds in keratin, its major protein. How can curls be induced ?

5. Proteins are quite stable. The lifetime of a peptide bond in aqueous solution is nearly 1000 years. However, the $\Delta G^{\circ\prime}$ of hydrolysis of proteins is negative and quite large. How can you account for the stability of the peptide bond in light of the fact that hydrolysis releases much energy ?

6. For an amino acid such as alanine, the major species in solution at pH 7 is the zwitterionic form. Assume a pK_a value of 8 for the amino group and a pK_a value of 3 for the carboxylic acid and estimate the ratio of the concentration of neutral amino acid species (with the carboxylic acid protonated and the amino group neutral) to that of the zwitterionic species at pH 7.

7. All L amino acids have an *S* absolute configuration except L-cysteine, which has the *R*

configuration. Explain why L-cysteine is designated as the *R* absolute configuration.

8. Translate the following amino acid sequence into one-letter code : Leu-Glu-Ala-Arg-Asn-Ile-Asn-Gly-Ser-Cys-Ile-Glu-Cys-Glu-Ile-Ser-Gly-Arg-Glu-Ala-Thr.

9. Would you expect Pro–X peptide bonds to tend to have *cis* conformations like those of X–Pro bonds ? Why or why not ?

10. A protein was purified to homogeneity. Determination of the molecular weight by molecular exclusion chromatography yields 60 kd. Chromatography in the presence of 6 M urea yields a 30-kd species. When the chromatography is repeated in the presence of 6 M urea and 10 mM β-mercaptoethanol, a single molecular species of 15 kd results. Describe the structure of the molecule.

11. The three-dimensional structure of biomolecules is more conserved evolutionarily than is sequence. Why is this the case ?

12. The sequences of three proteins (A, B, and C) are compared with one another, yielding the following levels of identity :

	A	B	C
A	100%	65%	15%
B	65%	100%	55%
C	15%	55%	100%

Assume that the sequence matches are distributed relatively uniformly along each aligned sequence pair. Would you expect protein A and protein C to have similar three-dimensional structures ? Explain.

13. Would a homopolymer of alanine be more likely to form an α helix in water or in a hydrophobic medium ? Explain.

14. Using the data in Table 9–3, calculate the *average* amino acid residue weight in a protein of typical composition. This is a useful number to know for approximate calculations.

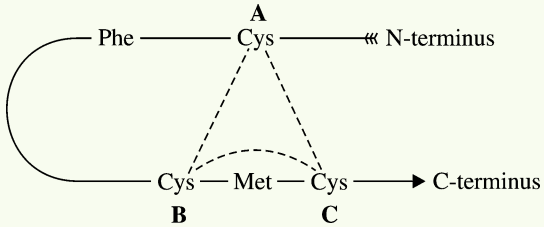
15. The melanocyte-stimulating peptide hormone α-*melanotropin* has the following sequence :

Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val

(*a*) Write the sequence using the one-letter abbreviations.

(*b*) Calculate the molecular weight of α-melanotropin, using data in Table 9-3. Why is this result not *exactly* correct at neutral pH ?

16. A protein has been sequenced after destruction of —S—S— bonds. It is known to contain 3 Cys residues, located as shown below. However, only one of these is a free —SH; two are involved in an S—S bond.



The only methionine and the only aromatic amino acid (Phe) in this protein are in the positions indicated. Cleavage of the *intact* protein (with —S—S— bridge intact) by either cyanogen bromide or chymotrypsin does *not* break the protein into two peptides. Where is the —S—S— bridge (AB, BC, or AC) ?

17. *Apamine* is a small protein toxin present in the venom of the honeybee. It has the sequence

CNCKAPETALCARRCQQH

(*a*) It is known that apamine does not react with iodoacetate. How many disulfide bonds are present ?

(*b*) Suppose trypsin cleavage gave two peptides. Where is (are) the S—S bond(s) ?

18. In the protein *adenylate kinsase,* the C-terminal region is α-helical, with the sequence

Val-Asp-Asp-**Val-Phe**-Ser-Gln-**Val**-Cys-Thr-His-**Leu**-Asp-Thr-**Leu**-Lys-

The hydrophobic residues in this sequence are presented in boldface type. Suggest a possible reason for the periodicity in their spacing.

19. Consider a small protein containing 101 amino acid residues. The protein will have 200 bonds about which rotation can occur. Assume that three orientations are possible about each of these bonds.

(*a*) Based on these assumptions, about how many *random-coil* conformations will be possible for this protein ?

(*b*) The estimate obtained in (*a*) is surely too large. Give one reason why.

20. It has been postulated that the normal (noninfectious) form of prion differs from the infectious form only in secondary/tertiary structure.

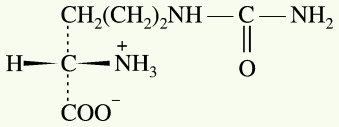
(*a*) How might you show that changes in secondary structure occur ?

(*b*) How might you check for changes in quaternary structure ?

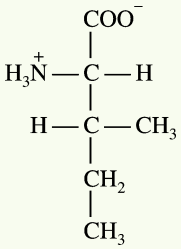
(*c*) If this model is correct, what are the implications for structural prediction schemes like that of Chou and Fasman ?

21. It is observed that chloride ion acts as a negative allosteric effector for hemoglobin. Suggest a possible explanation for why this should be so.

22. Is citrulline isolated from watermelons (shown below) a D- or L-amino acid ? Explain.



23. The structure of the amino acid isoleucine is :



(*a*) How many chiral centres does it have ?

(*b*) How many optical isomers ?

(*c*) Draw perspective formulas for all the optical isomers of isoleucine.

24. Lysine makes up 10.5% of the weight of ribonuclease. Calculate the minimum molecular weight of ribonuclease. The ribonuclease molecule contains ten lysine residues. Calculate the molecular weight of ribonuclease.

25. What is the approximate molecular weight of a protein containing 682 amino acids in a single polypeptide chain ?

26. Pepsin of gastric juice (pH ≈ 1.5) has a pI of about 1, much lower than that of other proteins. What functional groups must be present in relatively large numbers to give

pepsin such a low pI ? What amino acids can contribute such groups ?

27. One method for separating polypeptides makes use of their differential solubilities. The solubility of large polypeptides in water depends upon the relative polarity of their R groups, particularly on the number of ionized groups; the more ionized groups there are, the more soluble the polypeptide. Which of each pair of polypeptides below is more soluble at the indicated pH ?

(*a*) $(Gly)_{20}$ or $(Glu)_{20}$ at pH 7.0

(*b*) $(Lys–Ala)_3$ or $(Phe–Met)_3$ at pH 7.0

(*c*) $(Ala–Ser–Gly)_5$ or $(Asn–Ser–His)_5$ at pH 6.0

(*d*) $(Ala–Asp–Gly)_5$ or $(Asn–Ser–His)_5$ at pH 3.0

28. William Astbury discovered that the x-ray pattern of wool shows a repeating structural unit spaced about 0.54 nm along the direction of the wool fiber. When he steamed and stretched the wool, the x-ray pattern showed a new repeating structural unit at a spacing of 0.70 nm. Steaming and stretching the wool and then letting it shrink gave an x-ray pattern consistent with the original spacing of about 0.54 nm. Although these observations provided important clues to the molecular structure of wool, Astbury was unable to interpret them at the time. Given our current understanding of the structure of wool, interpret Astbury's observations.

29. A number of natural proteins are very rich in disulfide bonds, and their mechanical properties (tensile strength, viscosity, hardness, etc.) are correlated with the degree of disulfide bonding. For example, glutenin, a wheat protein rich in disulfide bonds, is responsible for the cohesive and elastic character of dough made from wheat flour. Similarly, the hard, tough nature of tortoise shell is due to the extensive disulfide bonding in its α-keratin. What is the molecular basis for the correlation between disulfide-bond content and mechanical properties of the protein ?

30. When wool sweaters or socks are washed in hot water and/or dried in an electric dryer, they shrink. From what you know of α-keratin structure, how can you account for this ? Silk, on the other hand, does not shrink under the same conditions. Explain.

31. In the following polypeptide, where might bends or turns occur ? Where might intrachain disulfide cross-linkages be formed ?

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Ile–Ala–His–Thr–Tyr–Gly–Pro–Phe–Glu–Ala–Ala–Met–Cys–Lys–Trp–Glu–Ala–Gln–
19 20 21 22 23 24 25 26 27 28
Pro–Asp–Gly–Met–Glu–Cys–Ala–Phe–His–Arg–

32. Both myoglobin and hemoglobin consist of globin (protein) bound to a heme prosthetic group. One heme group binds one O_2. Why is the oxygen saturation curve (saturation versus pO_2) of myoglobin a rectangular hyperbola while that of hemoglobin is sigmoidal ?

33. Using known endo- and exopeptidases, suggest a pathway for the complete degradation of the following peptide :

His-Ser-Lys-Ala-Trp-lle-Asp-Cys-Pro-Arg-His-His-Ala

34. How do depilatory creams remove hair ?

35. Which of the following characteristics are associated with myoglobin, hemoglobin, both of them or neither of them ?

(*a*) majority of structure in α-helical conformation

(*b*) oxygen carrier

(*c*) carbon dioxide carrier

(*d*) not an allosteric protein

(*e*) heme group(s) in a polar crevice(s)
(*f*) protoporphyrin IX
(*g*) sigmoidal oxygen dissociation curve
(*h*) tertiary structure
(*i*) quaternary structure
(*j*) blocked N-terminal residue

36. What is homocysteine ?
37. Why do some people have curly hair while in others hair grows straight ?
38. Why does our skin become wrinkled in old age ?
39. What does the blood test for homocysteine signify ?
40. Why do some wounds leave scars while some do not ?
41. Why don't we feel pain when cutting hair / nails ?