

1. PROTEINS¹⁻⁵⁾

We have already stated that a protein is a linear chain, built from 20 amino acids. (A few rare ones occur occasionally, but we will not discuss them.) The chain contains of the order of 100-200 amino acids. Of particular interest are the globular proteins, which act, for instance, as enzymes (catalysts). These proteins, in the proper solvent, fold into the final protein, as sketched in Fig. 1.1.

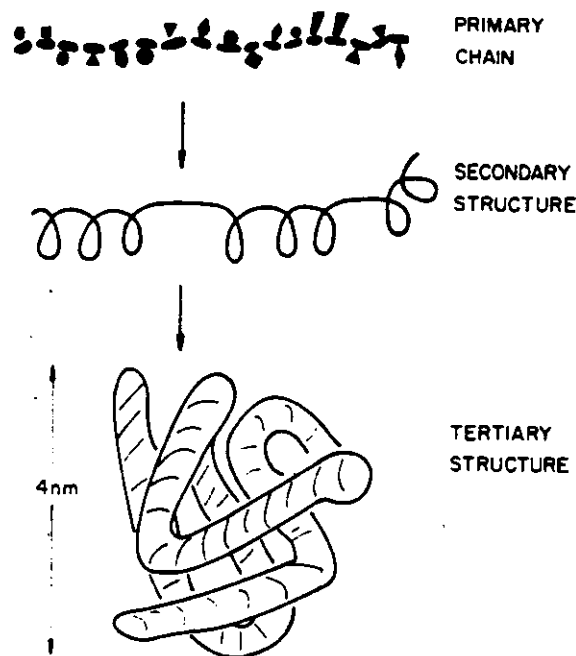


Fig. 1.1 The linear polypeptide chain (primary sequence) folds into the final tertiary structure.

In this chapter we will describe the building blocks of proteins and make some remarks about protein structure.

-
1. Lehninger, Chapters 3-7.
 2. Vol'kenshtein, Chapter 5.
 3. Stryer, Part I.
 4. H. Neurath, The Proteins, Academic Press, multi-volume.
 5. G. E. Schulz and R. H. Schirmer, Principles of Protein Structure, Springer, New York, 1979.

1.1 Amino Acids, the Building Blocks

The building blocks from which proteins are constructed are the amino acids. To a physicist, such complex organic molecules are often a nightmare. We suggest forgetting at the beginning that they are organic molecules and simply consider them building blocks with specified properties. As they are used over and over again, their properties become familiar and they lose their intimidating character.

All amino acids are built around a carbon atom. Two of the four ligand positions (valences) connect the amino acid to the other building blocks in the chain, one (the "sidechain") produces the characteristic features of a particular amino acid, and the fourth carries a hydrogen atom. The general structure thus is as shown in Fig. 1.2. R is the residue that determines the

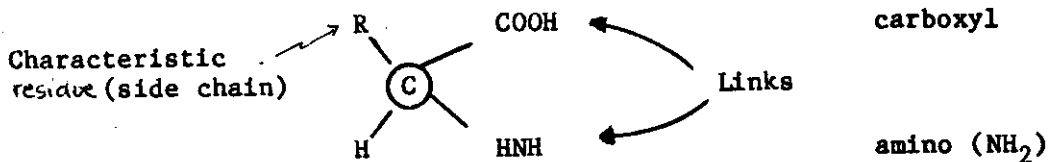


Fig. 1.2 General structure of an amino acid.

important specific properties of the amino acid. The individual building blocks in a protein are linked together by peptide bonds. The bond has covalent character and is so strong that it is practically never broken by thermal effects. In the formation of peptide bonds, water is eliminated and a bond is formed: $\text{COOH} \dots \text{HNH} \rightarrow -\text{CO}-\text{NH}- + \text{H}_2\text{O}$. The resulting polypeptide chain thus looks as indicated in Fig. 1.3. It consists of a backbone and two

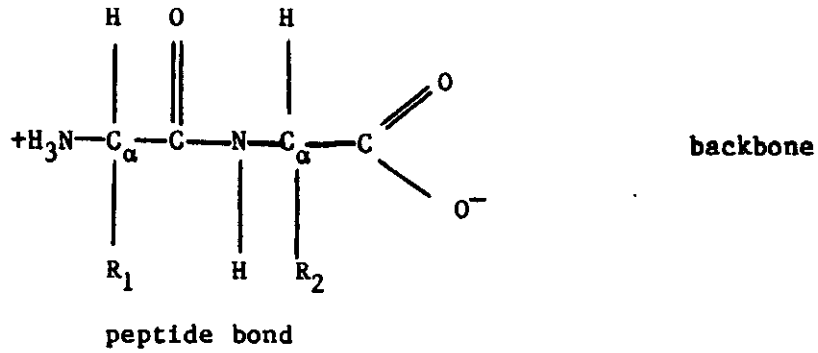


Fig. 1.3 A polypeptide chain is formed from individual amino acids through peptide bonds.

radicals per amino acid residue. The backbone is non-specific; the entire specificity is in the radicals. In addition to the components shown, proteins contain two endgroups,


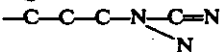
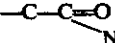
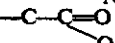

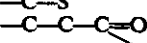
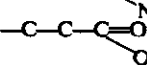

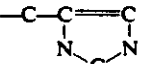
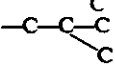
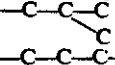






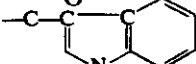
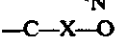

the terminal carboxyl $-\text{CO}_2^-$ denoted by $-\text{C}$

the terminal amino $-\text{NH}_3^+$ denoted by $-\text{N}$

The twenty common amino acids are given in Table 1.1.

Table 1.1

Properties of protein building blocks. The length (L) is for the side chain only. The molecular weight is for the entire amino acid - subtract 17.9 to obtain molecular weight of residue. The polarity indicates whether the amino acid is nonpolar (NP) or polar with a net positive, negative, or neutral charge at pH = 6.

Amino Acid	Symbol	Molecular weight (amu)	L (nm)	Polarity	Side chain (X = benzene)	one-letter symbol
Alanine	ALA	89	0.28	NP		A
Arginine	ARG	174	0.88	+		R
Asparagine	ASN	132	0.51	0		N
Aspartic Acid	ASP	133	0.50	-		D
Cysteine	CYS	121	0.43	0		C
Glutamine	GLN	146	0.64	0		Q
Glutamic Acid	GLU	147	0.63	-		E
Glycine	GLY	75	0.15	0		G
Histidine	HIS	155	0.65	+		H
Isoleucine	ILE	131	0.53	NP		I
Leucine	LEU	131	0.53	NP		L
Lysine	LYS	146	0.77	+		K
Methionine	MET	149	0.69	NP		M
Phenylalanine	PHE	165	0.69	NP		F
Proline	PRO	115		NP		P
Serine	SER	105	0.38	0		S
Threonine	THR	119	0.40	0		T
Tryptophan	TRP	204	0.81	NP		W
Tyrosine	TYR	181	0.77	0		Y
Valine	VAL	117	0.40	NP		V

The side chains are shown schematically in Fig. 1.4.

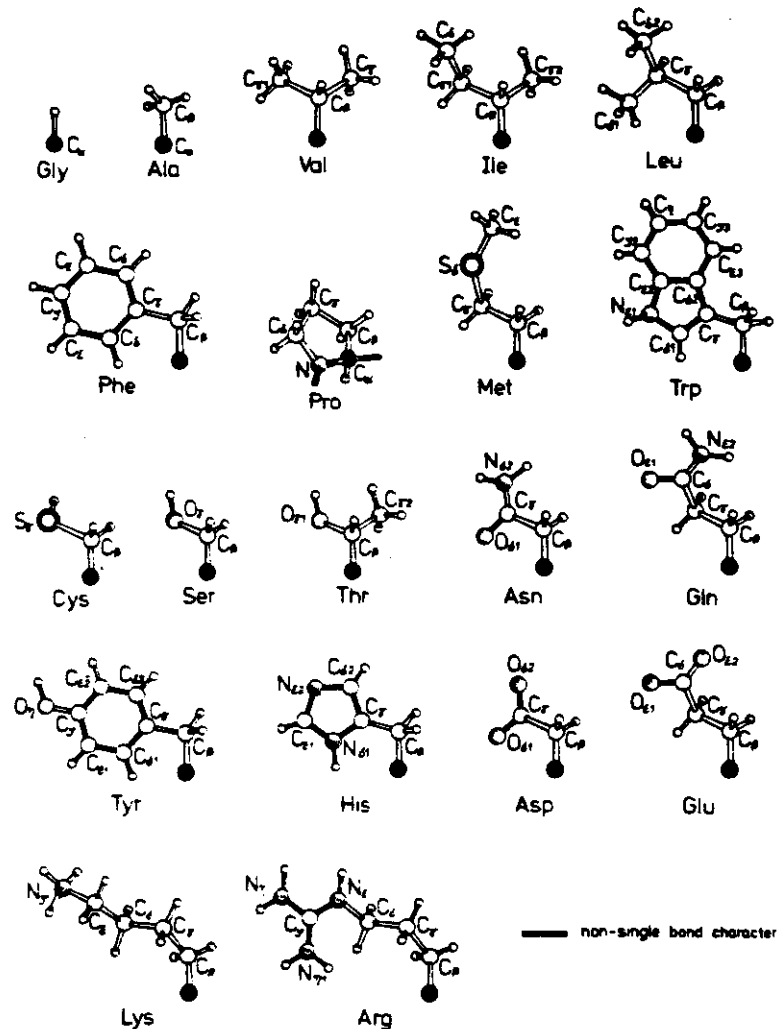


Fig. 1.4 (After Schulz and Schirmer) The 20 standard amino acid side chains. For proline part of the main chain is inserted. All other side chains are shown as they emerge from the C_α -atom in the main chain. The residue names are given as three-letter symbols. Atom names are those given in the IUPAC-IUB recommendations of 1969. The main chain in Pro is indicated by solidly drawn bonds. All C_α -atoms are black.

In general, one classifies the amino acids into four groups:

- I. Amino acids with nonpolar (hydrophobic) radicals. These amino acids are less soluble in water than the polar amino acids.
- II. Amino acids with uncharged polar radicals. The polar R groups of these amino acids can hydrogen-bond with water and they are thus more soluble in water than those of group I.
- III. Amino acids with negatively charged (acidic) R groups. The members of this class possess a negative charge at pH 6-7.
- IV. Amino acids with positively charged (basic) R groups. The basic amino acids have a positive charge at pH 7. Histidine is a borderline case; at pH 7, only about 10% have a positive charge; at pH 6, the percentage is about 50.

An excellent discussion of the properties of each amino acid can be found in Schulz and Schirmer.

All amino acids can exist in two different forms, related to each other by a mirror reflection (Fig. 1.5).

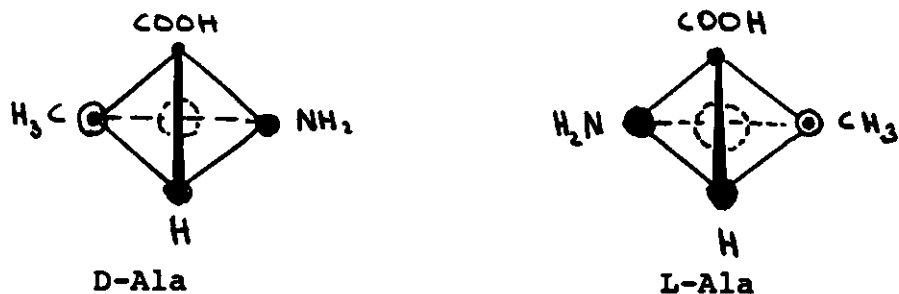


Fig. 1.5 D and L forms of alanine.

D and L stand for dextro (right) and levulo (left). The two forms rotate the polarization of light in different directions. In nature, only the L-amino acids are biologically active. In the laboratory, both forms can be produced. Some experiments, probably wrong, provide food for speculation. Did the L dominance come about by accident or through a deeper cause? Neutrinos and beta particles emitted in the weak interaction have a handedness. Could this be the agent to produce predominantly left-handed amino acids? Could one of the two forms be energetically favored? (6)

In describing a sequence, the various amino acids are designated either by a three-letter or by a one-letter abbreviation. The abbreviations are listed in Table 1.1.

1.2 The Primary Sequence

Proteins consist of about 100 to 1000 amino acids bound together to form a polypeptide chain. Investigation of the primary sequence already reveals fascinating features. Consider for instance the primary sequence of cytochrome c, an electron transport enzyme. (R. E. Dickerson, Scientific American, April 1972; reprinted in The Chemical Basis of Life). The primary sequence of this protein has been determined in a large number of cases. Table 1.2 gives part of the sequence for these. (The complete table can be found in Dickerson or in M. O. Dayhoff, Atlas of Protein Sequence and Structure.)

6. For a review see L. Keszthelyi, "Origin of the Asymmetry of Biomolecules and Weak Interaction". Origins of Life 8, 299-340 (1977).

Table 1.2 Partial Sequence of Cytochrome

	1									2									3									4															
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9				
Human										G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Rhesus Monkey										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Horse										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Pig, Bovine, Sheep										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Dog										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Gray Whale										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Rabbit										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Kangaroo										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Chicken, Turkey										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Penguin										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Pekin Duck										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Pigeon										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Snapping Turtle										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Rattlesnake										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Bullfrog										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Tuna Fish										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Dogfish										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Lamprey										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Fruit Fly										(G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Screw-worm Fly						G	V	P	A	G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Silkworm Moth						G	V	P	A	G	D	V	E	K	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Tobacco Horn Worm Moth						G	V	P	A	G	N	A	E	N	G	K	K	I	F	I	H	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L		
Wheat	A	S	F	S	E	A	P	P	G	N	P	D	A	G	A	K	I	F	F	K	T	K	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L		
Neurospora crassa					G	F	S	A	G	D	S	K	R	G	A	N	L	F	F	K	T	R	C	A	E	C	H	G	E	G	G	N	L	T	Q	K	I	G	P	N	L		
Baker's Yeast					T	E	F	K	A	G	S	A	K	K	G	A	T	L	F	F	K	T	R	C	E	L	C	H	T	V	E	K	G	G	P	H	K	V	G	P	N	L	
Candida krusei					P	A	P	F	E	Q	G	S	A	K	K	G	A	T	L	F	F	K	T	R	C	A	E	C	H	T	I	E	A	G	G	P	H	K	V	G	P	N	L

COMMON																																																
										G		G		F		C		CH																														
ALTER-NATIVES	A	S	F	S	G	V	P	A	-	D	V	E	K	N	K	K	I	V	Q	K	A	Q						T	V	E	K	G	G	K	H	T	N											
		P	T	E	F	S	P		-	N	A	K	N	S	I	D	A	A	T	L	K	T	R	S	E			G	C	D	A	A	A	P	O	V	A											
			A	P	A	K	Q		-	P	A												E	L				E	G	N	N	L	T	Q	I													
					E				-	S								T	T									I	G		G				I													

We notice that this electron-transfer protein has changed very little in the course of about 10^9 years. Cyt c is essentially the same in bacteria, fungi, plants, and in humans. In particular, some positions are invariant, they are occupied by exactly the same amino acid in every sequence. A mutation in such a position is lethal.

The table can be used to draw a family tree: The differences between any two chains are counted. Man and rhesus monkey, for instance, differ by one residue. The difference table then provides us with a family tree. Table and family tree are shown on the next page, in Fig. 1.6.

In the following section, we will discuss briefly a few simple proteins, to make the general discussion somewhat more specific.

1.3 Some Typical Proteins

Proteins perform an amazing number of functions and we only mention a major subdivision here between structural and functional proteins. Structural proteins are more uniform and use repeated units, as in a mass-produced building. Functional proteins are tailor-made for a specific job, such as storing energy or matter (oxygen) or catalyzing a particular reaction. The basic building blocks are the same, but they are used differently. Moreover, non-protein subunits are sometimes incorporated into functional proteins. The non-polypeptide part is called the prosthetic group, the protein without it apoprotein. As shown in Fig. 1.1, a protein first folds into a secondary and then into a tertiary structure. The geometry of the amino acids permit only a small number of secondary structures. Of these, two are most often used in nature, the alpha helix and the antiparallel beta pleated sheet. Their basic structures are shown in Fig. 1.7.

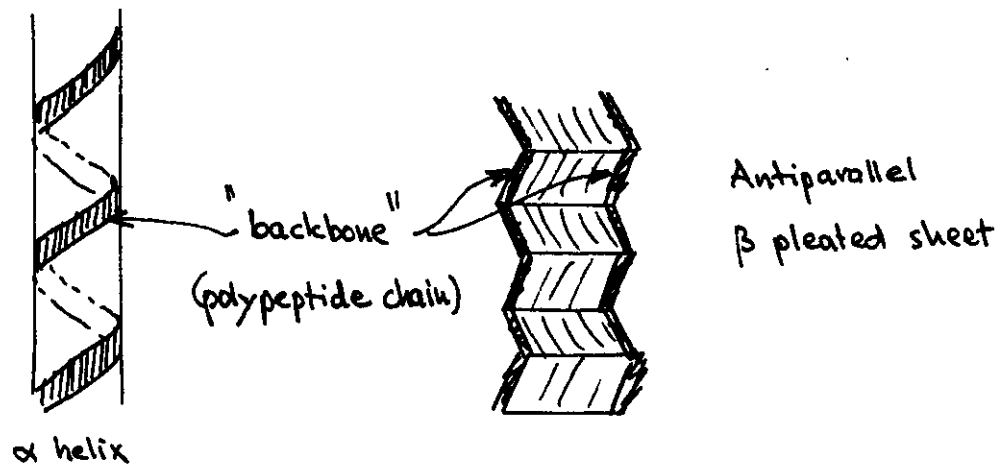


Fig. 1.7 Alpha helix and antiparallel beta pleated sheet.

Globular proteins are largely built from pieces of alpha helices, interrupted by groups that permit bending and twisting, and from beta pleated sheets. Structural proteins also use both units, but in a periodic arrangement. We sketch in the following some examples that we will use again in later discussions.

1.3.1. Heme Proteins. When nature discovers a good trick, it is used over and over again. In heme proteins, nature uses a particular organic molecule, the heme group, embeds it into different proteins ("globins"), and with the protein modifies the properties of the heme group so that the entire system performs a given task. A heme protein thus looks schematically as sketched in Fig. 1.8. Before discussing the heme group and some heme proteins in more detail, we list in Table 1.3 a number of heme proteins with their functions.

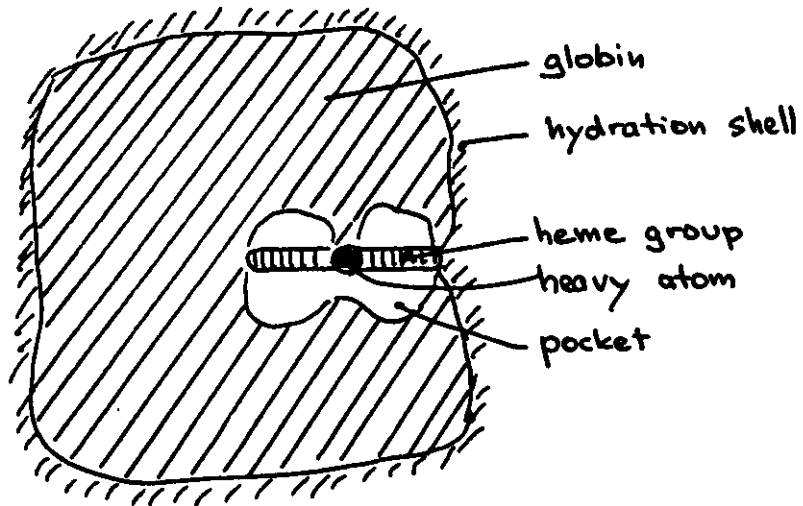


Fig. 1.8 A cross section through a heme protein.