

SECTION I

PROTEIN STRUCTURE

INTRODUCTION -- D. W. Von Endt

PROTEINS IN GENERAL

We're all wet it seems, since water accounts for the major portion of human cellular components, as it does in virtually all animals. The remainder of the cells consist primarily of protein. Proteins are interesting compounds; thousands of kinds exist in individual cells. They form the true "thread of life", since that they are responsible for the many aspects of cellular function. The unique role played by proteins in living organisms has been recognized for a long time. Fruton and Simmonds (1958) quoted the Dutch chemist G.J. Mulder who began the first systematic investigation of these materials during the middle of the 19th century. In 1844-1851 he stated: "There is present in plants as well as animals a substance which...performs an important function in both. It is one of the very complex substances....It is without doubt the most important of the known components of living matter, and it would appear that, without it, life would not be possible. The substance has been named protein."

Proteins form cellular enzymes which convert food into energy. They transfer this energy to places in the cell where it is used, replicate other proteins, and are largely responsible for the ability of an organism to grow and rejuvenate itself. Proteins are responsible for animals being able to transfer energy into movement; they are responsible for sight, hearing and breathing; and form the vital interface between the organism and its environment.

These remarkable molecules belong to that class of high molecular weight biopolymers which are formed from amino acid, monomer, building blocks bound in peptide linkage. Each type of protein is composed of a unique combination of amino acids bonded "head-to-tail", providing the protein with its function. While other biopolymers such as cellulose and the synthetic polymers contain only a few types of sub-units, proteins consist of more than twenty individual amino acids which are linked in a

specific sequence, in characteristic proportion, in each protein. Consequently, it is possible to construct many more different kinds of proteins than polysaccharides, for instance. This fact allows nature to tailor each protein to a specific chemical task, and lends complexity to a study of these compounds.

The molecular weight of proteins vary from about 13,000 to many millions. Most proteins, because of their molecular size, are not diffusible through membranes such as cellophane, are actually of colloidal dimensions and exhibit colloid properties. This characteristic is important to an explanation of the properties of gelatin based adhesives. Some of the smaller proteins are soluble in water alone, while some of the more complex ones require the presence of small amounts of salts and/or acids or bases in order to dissolve. On the other hand, the structural proteins (mainly collagen) of skin, bone and teeth are sufficiently complex and large as to be insoluble in virtually all solvents and are dissolved only by reagents which cause considerable structural alteration.

Native proteins tend to be sensitive to heat, changes in pH, surfactant activity and various chemical reagents. When subjected to any of these events, proteins will undergo a series of alterations in a number of their properties. These alterations are known generally as denaturation. The amino acids making up proteins may also possess free ionic or charged groups (either a "+" or a "-" charge); and the mixture of charged amino acids making up a protein lends a net charge (slightly positive or slightly negative, depending upon the number and type of amino acids) to the whole protein molecule. The charged nature of proteins allows them to form reactions with ionic reagents, in some cases, resulting in insoluble compounds. This characteristic is useful in the tanning of skin. The charged nature of proteins also allows them to migrate in an electric field. This phenomenon accounts for one way in which a mixture of proteins may be separated. The amino acids of proteins also contain weakly ionizable atomic arrangements which are responsible for many of the intermolecular bonding forces between protein molecules, and between them and other non-protein components. This characteristic accounts for some of the toughness of connective tissue and the tenacity of glue.

BASIC PROTEIN STRUCTURE

Elementary Composition In classical methods of organic chemistry, the

first step in studying the structure of a compound usually consisted of an analysis to determine the proportion of the various chemical elements it contained. This was how Mulder began his systematic study of proteins. In their elementary composition, proteins contain about 50-55% carbon, 6-7% hydrogen, 20-23% oxygen and 12-19% nitrogen. Many proteins also contain sulphur (usually in the range of 0.2-3.0% although there may be significantly higher amounts in some proteins) and 0-6% phosphorous (some proteins contain much more). It is the element nitrogen that is most characteristic of proteins, and nitrogen analysis is frequently used to give an estimate of protein content. It has become common practice to consider that proteins contain 16% nitrogen and that the protein content of materials can be estimated by multiplying the percent of nitrogen by 6.25. Since nitrogen varies by 12-19%, and not all of the nitrogen may be from proteins; this factor provides only a rough approximation. An elementary analysis of protein provided the first quantitative indication that the word protein represented substances of large molecular weight. For instance, the structural protein tropocollagen, found in skin and bone, has a molecular weight of about 300,000; and silk fiber has a molecular weight of 350,000. Because these are such large molecules, chemists have centered their attention on the relatively low molecular weight amino acids that are obtained when a protein is subjected to hydrolysis (usually by heating in a mineral acid solution).

Amino Acids With the exception of the amino acids proline and hydroxyproline (which are imino acids), all the amino acids are α -amino carboxylic acids (from which the name "amino acid" is derived) in which the nitrogen is attached to the carbon α to (next to) the carboxyl group. The rest of the atoms forming the molecule can be represented by the symbol "R-" and are the source of the structural variations leading to the different amino acids. This can be seen in Figure 1 where three amino acids are pictured. The simplest of all amino acids is glycine, in which two hydrogens, an amino group (NH_2) and an acidic carboxyl group (COOH) are attached to the same carbon. Serine and aspartic acid also are included in the Figure to illustrate some of the different "R" groups capable of substituting for one of the hydrogens. To further illustrate the diversity of structure of the amino acids, a listing of those 19 which are commonly found in connective tissue, along with their structures and a general classification

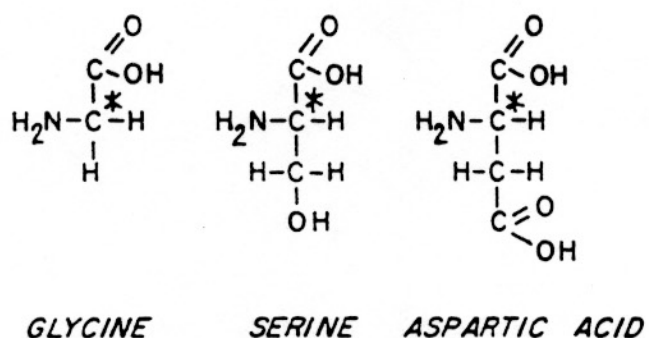


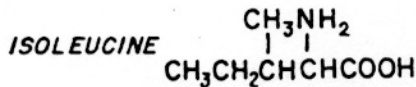
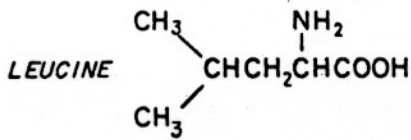
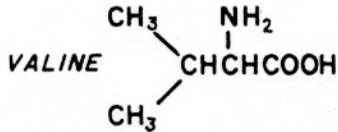
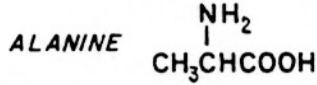
Figure 1. A Fischer projection of glycine, serine and aspartic acid. The carbon atom which is starred is the one at which substitution occurs to form different amino acids.

system based on their "R" groups, is presented below as Figure 2. Here it can be seen that the side chains ("R" groups) of some amino acids contain only carbon and hydrogen bonded in a straight line, (e.g., alanine) while others vary the theme by containing a "branch" (e.g., isoleucine). Still others vary by containing carbon-hydrogen rings (e.g., phenylalanine), or complex rings involving atoms other than carbon and hydrogen (e.g., tryptophan). Further combinations include the addition of a second amine group in lysine, a second carboxyl group in aspartic and glutamic acids, a hydroxyl group in serine and threonine, sulphur in cysteine, and a combination of ring and hydroxyl group as in hydroxyproline. This tremendous diversity of structure, as has been mentioned previously, allows for the diversity of structure and function seen in the many types of protein formed from the amino acids.

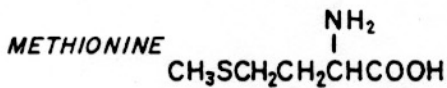
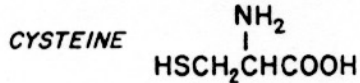
Simple amides may be formed by reacting a carboxylic acid with ammonia resulting in the splitting out of water. By analogy, an amide bond also may be formed between the carboxyl group of one amino acid and the amino group of another. This special amide bond is called the peptide bond. Since amino acids contain both an amine and a carboxyl group, a continuation of these reactions at either the leftover free amino or free carboxyl group could build up long peptide chains. The basic structure of proteins consists of such chains. Figure 3 illustrates these events using

ALIPHATIC I. NEUTRAL

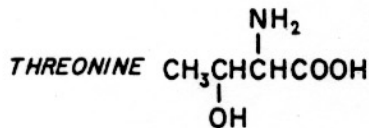
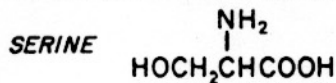
HYDROCARBON SIDE-CHAIN



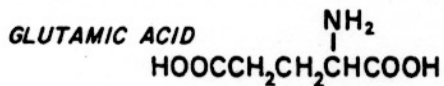
SULFUR-CONTAINING



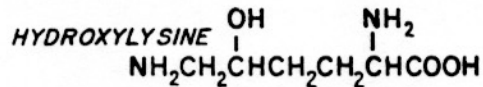
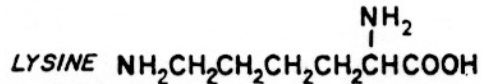
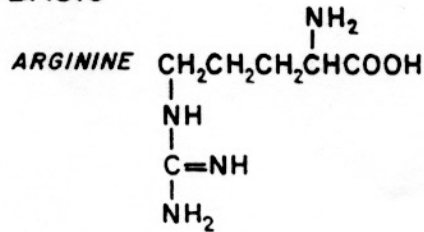
HYDROXYL-CONTAINING



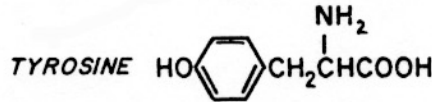
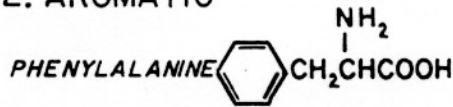
ACIDIC



BASIC



2. AROMATIC



3. HETEROCYCLIC

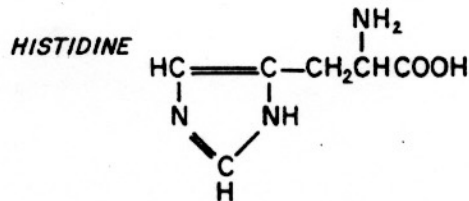
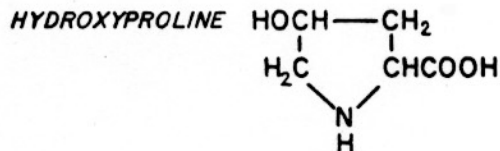
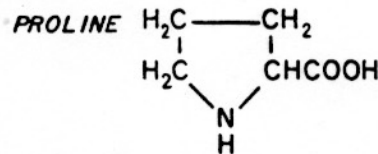


Figure 2. The names and structures of the major amino acids found in connective tissue.

as examples an acid and an amine to form the dipeptides glycyl-alanine and alanyl-glycine.

With so many varied amino acids to draw from, it is clear that a polypeptide has the possibility of containing a very large number of positional isomers. The situation is even more complicated in the case of the larger proteins in which the molecular weights run into the hundreds of thousands.

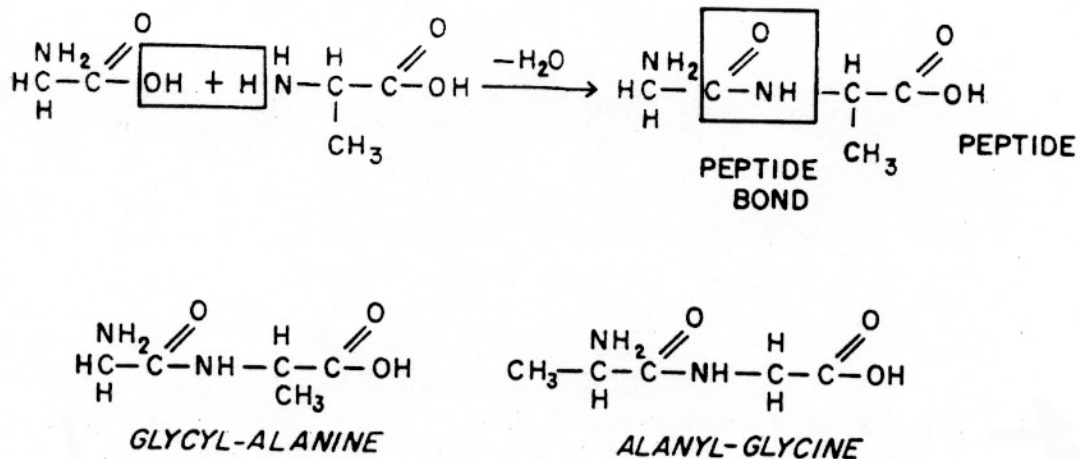


Figure 3. The formation of the peptide bond between amino acids, and an illustration of the simple dipeptides glycyl-alanine and alanyl-glycine.

Fibrous Proteins The proteins with which we are concerned here are the structural fibrous proteins. Their function is to provide strength in tissues such as bone, skin, hair, and silk. With such generally similar functions, it would seem that these proteins should have approximately the same composition and structure, while other proteins with different functions would possess different structures. This concept is true only in the general sense that structural proteins are fibers or sheets of fibers. Some of the variations in the amino acid content of proteins of differing function is illustrated in Table I where it may be seen that ^{the blood} ~~muscle~~ protein, ^{albumin} for instance, differs significantly ^{in composition} from both bone protein and the milk protein, casein.

Proteins are a chemical illustration of the architectural concept

that "form follows function". The covalent bonding of the atoms which form the amino acids, and the amino acids which form the protein, totally determine the function of the protein. Some proteins exist as fibers (as does collagen) whose shape and strength rests on its covalent, primary structure. Some fibrous proteins coil about themselves to form a helical secondary structure. This self-coiling is determined entirely by the primary structure. For instance, many proteins are not fibrous, but have a more rounded secondary structure which can be gently "un-wound" under appropriate conditions to form the long, linear polypeptide basic unit. When these are returned to an environment which simulates their native physiological milieu, they will reform spontaneously to their original secondary structure as well as to their original function.

It may be noticed in Table I that the composition of gelatin, animal glue and collagen are similar. This is because both gelatin and glue are composed of disrupted native collagen and share the composition of the parent. The data in this Table also indicate that glycine and alanine, the two smallest amino acids, account for almost half the total of the amino acids of which collagen is composed. Collagen exists in nature as a tight coil, or helix, with the bulky amino acid "R" groups facing toward the outside of the helix. The presence of very small "R" groups (in the case of glycine, a hydrogen atom) throughout most of the molecule of collagen allows for the formation of the helix. Collagen also contains significant amounts of proline and hydroxyproline. These molecules contain a rigid five-membered ring which forces the protein chain to assume a bend each time one of these amino acids is encountered in the chain. This process is also thought to aid in the formation of the helix. Egg albumin and casein, on the other hand, are not fibrous proteins like collagen and their amino acid content is different in that all the amino acids are represented in a more equal manner. It may be noticed that both albumin and casein contain a significant percentage of glutamic acid, which lends ionic character to these proteins and (along with other molecular qualities) are one reason these two proteins are used as adhesives and pigment and ground binders.

TABLE I The Amino Acid Composition of Some Proteins in Per Cent

	Gelatin	Glue	Collagen	Albumin	Casein
Glycine	27.5	33.5	25.3	3.1	2.0
Alanine	11.0	11.3	10.5	6.7	3.2
Valine	2.6	2.0	2.7	7.1	7.2
Leucine	3.3	2.3	3.9	9.2	9.2
Isoleucine	1.7	1.2	1.7	7.0	6.1
Serine	4.2	3.5	4.2	8.2	6.3
Threonine	2.2	1.8	2.5	4.0	4.9
Cystine	-	-	-	1.0	0.3
Cysteine	-	-	-	1.4	-
Methionine	0.9	5.0	0.8	5.2	2.8
Aspartic Acid	6.7	4.7	7.1	9.3	7.1
Glutamic Acid	11.4	7.2	11.9	16.5	22.4
Lysine	4.5	2.7	4.1	6.3	8.2
Hydroxylysine	1.0	0.5	1.1	-	-
Arginine	8.8	4.7	9.2	5.7	4.1
Histidine	0.8	0.5	1.0	2.4	3.1
Phenylalanine	2.2	1.3	2.9	7.7	5.0
Tyrosine	0.3	0.1	0.6	3.7	6.3
Tryptophan	-	-	-	1.2	1.2
Proline	16.4	12.8	14.7	3.6	10.6
Hydroxyproline	14.1	9.4	14.1	-	-

REFERENCES

The following is a partial list of references consulted for this section. The relevant portions of each volume may be used for further readings on the subjects which they cover.

- Bourne, G.H., ed. (1972) The Biochemistry and Physiology of Bone, Vol. 1, 2nd Edition. New York: Academic Press.
- Cantarow, A. and Schepartz, B. (1957) Biochemistry. Philadelphia: W.B. Saunders.
- Fruton, J.S. and Simmonds, S. (1958) General Biochemistry, 2nd Edition. London: J. Wiley & Sons.
- Reed, R., ed. (1966) Science for Students of Leather Technology. London: Pergamon Press.
- Reed, R. (1972) Ancient Skins, Parchments and Leathers. London: Seminar Press.
- Von Endt, D.W. (1979) Techniques of Amino Acid Dating. In R. Humphrey and D. Stanford (eds.) Pre-Llano Cultures of the Americas: Paradoxes and Possibilities. Washington, D.C.: Anthropological Society of Washington.