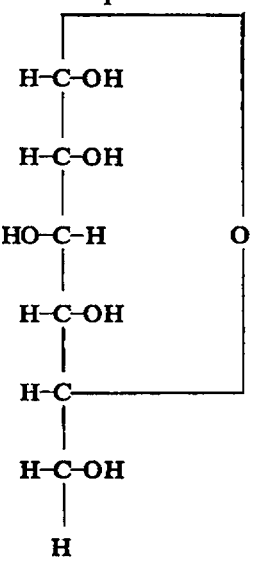
THE THREE DIMENSIONAL STRUCTURES OF PROTEINS

WALTER KAUZMANN
From the Frick Chemical Laboratory, Princeton University, Princeton

ABSTRACT The general nature of the problem of molecular structure in protein chemistry is discussed. As in all of organic chemistry, this problem has two aspects: the determination of the structural formula (a province of classical organic chemistry) and the determination of the molecular conformation. Recent progress in both of these areas has been great. Information on the amino acid sequences in proteins is rapidly accumulating (essentially a problem which concerns the structural formulae of proteins). Detailed knowledge of the conformations of proteins in general, and recently of myoglobin and hemoglobin in particular, has come from various kinds of x-ray crystallographic studies. Other physical tools (especially optical rotatory power and ultraviolet spectroscopy) give less detailed but still highly useful information about molecular conformation. The current state of our knowledge of the nature of the forces responsible for the molecular conformations of proteins is briefly reviewed.

The basic problem of chemistry is to determine the structure of matter and to relate the properties and behavior of matter to its structure. The protein chemist therefore has the job, first, of determining the arrangements of the atoms within protein molecules, and then of understanding why these arrangements make it possible for proteins to do the things that they do.

In organic chemistry the problem of structure determination has two aspects. When the organic chemist studies the structure of a molecule by the methods of classical organic chemistry, he tries to find the arrangement of the covalent bonds between the atoms within the molecule. The results of this stage of the investigation are expressed in the form of a *structural formula,* in which the symbols of atoms are printed onto a page with lines joining those atoms that are covalently bonded to each other. Thus the structural formula of α-glucose can be written as shown at the right: This formula includes information, not only about the atomic connections, but also about the stereochemical configurations of atoms about the asymmetric carbon atoms in the molecule.



Within the past twenty or thirty years, however, organic chemists have learned that the structural formula does not tell us all that we need to know about the structure of a molecule in order to understand its properties. Beyond the problem of the structural formula there is the problem of *molecular conformation;* that is, the problem of the actual spatial arrangement of the atoms on the molecule. How far is each atom in the molecule, not only from its immediate neighbors, but from every other atom in the molecule? How rigid is the structure? Is more than one conformation possible, and if so, what are the relative quantities of the different conformations?

The most direct method of attacking this second aspect of the problem of structure is by the method of X-ray crystallography. This method is, however, limited to the study of molecules in the crystalline state, and organic and physical chemists are also interested in developing other tools for investigating the structures of molecules, particularly in non-crystalline states. Since most of the chemical processes of living systems go on in solution, and not in the crystalline state, structural tools that can be used on solutions are especially interesting to biochemists.

Starting with the work of Fisher and Hofmeister in 1902, great progress has been made in the determination of the classical organic structural formulas of proteins. It is now established that proteins are copolymers formed from about twenty different kinds of L-amino acids by the elimination of water, as indicated in Fig. 1. In

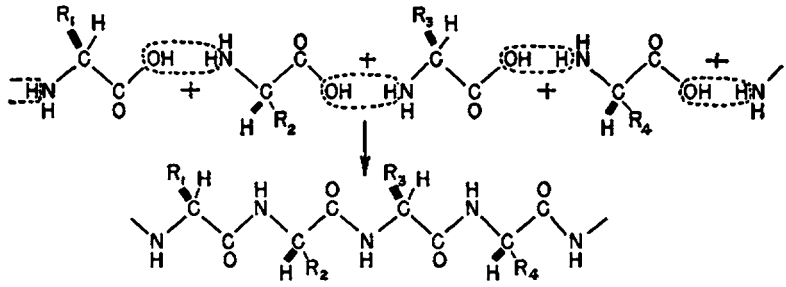


FIGURE 1 Coplymerization of L-amino acids to form the polypeptide backbone chain of a protein molecule. The side chains, R_1, R_2, R_3, R_4, . . . are those of the twenty-odd naturally occurring L-amino acids. The dashed loops indicate the formal removal of water molecules from the carboxyl and amino groups—an over-all reaction that actually takes place through a series of steps involving phosphorylation and group transfers.

order to be able to write the structural formula of a protein, then, it is necessary to determine the number of amino acid monomers that are incorporated into the protein molecule, and also the sequence in which they are joined together. Since the molecular weights of proteins range typically from about 10,000 to 100,000 or more, and since the average molecular weight of an amino acid residue (-NH-CHR-CO-) in a protein is about 100 to 120, protein molecules must contain something like 100 or more amino acid monomer units. Twenty years ago many chemists and bio-

chemists found it hard to believe that living organisms would be able to construct molecules as large as this with perfect reproducibility. They were prepared to find that different molecules of a given protein would show a fair amount of variability in their chemical structure. We now have reason to believe, however, that this variability, if it exists at all, can be quite slight, at least in certain proteins. When a cow makes the enzyme, ribonuclease (molecular weight 15,000, containing 124 amino acid residues per molecule), every molecule resembles every other molecule sufficiently closely that it is possible to speak of "the structural formula of ribonuclease." This formula, as recently revised by Smyth, Stein, and Moore (1), is shown in Fig. 2, where each of the symbols, ala, gly, ser, etc., should be taken to represent

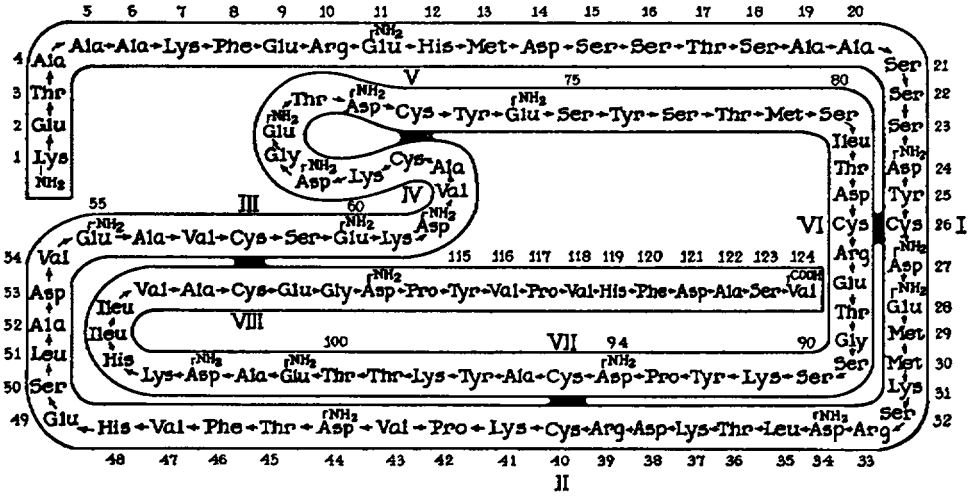


FIGURE 2 The sequence of amino acids in the polypeptide chain of bovine ribonuclease, as revised by Smyth, Stein, and Moore (1). The molecule contains four cystine—S—S—cross-linkages which are arranged in the manner shown. The folding of the polypeptide chain has been arranged so that these cross-linkages can be conveniently represented, and does not depict the actual chain conformation, which is as yet unknown. This figure originally appeared in *J. Biol. Chem.*, 1963, **238**, 227.

the bond structure of one of the natural L-amino acid residues (alanine, glycine, serine, etc.). Amino acid sequences have been completely determined for a number of other proteins, and those of still other proteins are now under active study. Sorm and Keil (2) have recently prepared a list of 19 proteins in which sequences of more than 15 amino acids are known. Complete sequences have been found for about 8 of these 19 proteins, and in 5 of these 8 the proteins contain more than 100 amino acid residues.

These sequence determinations must certainly be regarded as magnificent accomplishments of classical organic chemistry, but the structural formulae that they

produce do not by themselves explain the remarkable things that proteins are able to do in living systems, or even in the test tube. Nothing in Fig. 2 gives us a clue about the means by which ribonuclease is able to hydrolyze the phosphate ester linkages of ribonucleic acid so effectively and so specifically. In order to solve this problem it will be necessary to consider the second aspect of the problem of molecular structure—the problem of the molecular conformation of ribonuclease. We will have to know how the polypeptide chain in ribonuclease winds through space, and we will have to know the spatial positions of the amino acid side chains. The need for such detailed structural information in understanding the chemistry of small molecules only developed rather recently in the history of organic chemistry; for proteins this need is evident from the very start. A knowledge of the three dimentional structure of proteins will be central to the future development of protein chemistry and of biochemistry in general.

This problem is complex, but substantial progress is being made toward its solution. One very fruitful approach has been through the x-ray diffraction study of crystals of small molecules which are chemically related to proteins (simple amino acids and peptides). As a result of a great number of such studies Pauling, Corey, and Branson (3) were able to conclude that the peptide link in the polypeptide chain of a protein must possess certain definite geometrical characteristics, shown in Fig. 3. In particular the six atoms, C-NH-CO-C, involved in each peptide link must

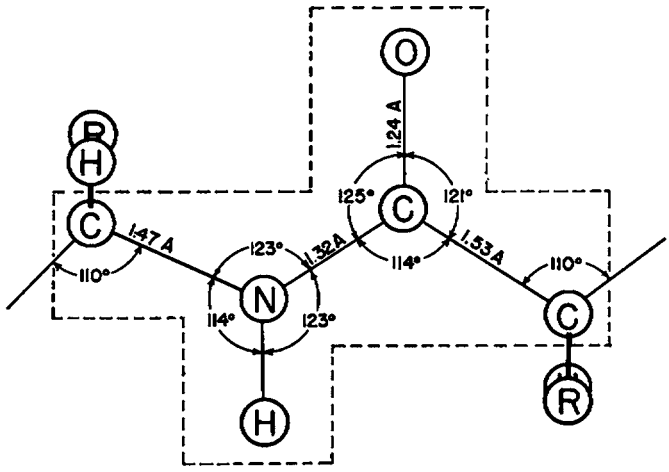


FIGURE 3 The geometry of the peptide linkage. Atoms enclosed in the dashed boundary are coplanar.

all lie in the same plane, and the bond lengths and the angular relationships between the bonds can be specified between rather rigid limits. A second conclusion that Pauling, Corey, and Branson could reach from these studies of crystal structures of

small, protein-like molecules was that there is a strong tendency for the amide hydrogen of one peptide group to form a hydrogen bond with the carbonyl oxygen of another peptide bond, and definite statements could also be made about the geometry of this bond.

Using the knowledge gained from the study of small molecules Pauling, Corey, Branson were able to show that polypeptide chains are capable of folding themselves into only a relatively few regular structural patterns. They were able to deduce several of these possible structures, and other workers were later able to show that several others are also possible. It is now known from the study of crystalline proteins and protein fibers that several of these structures actually exist in natural proteins as well as in synthetic polypeptides. Sketches of the most important of these structures are shown in Fig. 4.

In recent years a bold new application of x-ray diffraction to the problem of determining protein structure has been made by Perutz (4), Kendrew (5), and their collaborators, originally in the Cavendish Laboratory at Cambridge University. They have carried out analyses of the x-ray diffraction patterns of crystals of myoglobin and hemoglobin and have been able to deduce very detailed three dimensional structures for these proteins. A considerable amount of information could even be deduced about the structural formula of myoglobin in this way, thus offering the prospect of avoiding partially, if not completely, the classical organic chemist's approach to the determination of amino acid sequences in proteins. The structure of myoglobin that Kendrew has been able to obtain represents by far the most complete structural information that we now have about any protein molecule, and will provide the starting point for much future thinking about the principles of protein structure. A rough sketch of this structure is shown in Fig. 5; more elegant representations will be found in Kendrew's papers, especially in his article in the *Scientific American* (6). It is found that a large fraction (about 70 per cent) of the polypeptide chain of myoglobin exists in the form of a Pauling-Corey α-helix. The remainder of the chain exists in less regular folds, and portions of it even appear to have a certain degree of flexibility. This direct x-ray approach to the problem of protein structure holds great promise for the future, although it is extremely tedious to work out and there are serious technical limitations that have yet to be overcome for proteins in general. Other crystalline proteins are also under study. Recently preliminary reports have appeared showing that progress has been made in using x-ray diffraction to determine the structures of lysozyme (7) and ribonuclease (8).

Of course the x-ray crystallographic method is limited to the study of proteins in the crystalline state, and it is possible that the structure of a protein molecule in a crystal will differ from its structure in a living system or in solution. The protein crystals that have been used to date contain, however, a large amount of solvent (0.2 to 0.4 gm per gm of protein) and the protein molecules in crystalline myglobin and hemoglobin touch each other at only a few points. Therefore, it is somewhat un-

likely that the molecules in solution are greatly different in structure from those in the crystals.

Several tools are available in addition to x-ray diffraction for the study of the molecular structure of proteins. None of these other tools is capable of giving anything like the detailed structural information that can be obtained from x-ray diffraction studies on crystals, but many of them have the advantage that they can be applied to proteins in solution and to proteins that cannot be crystallized. Furthermore, they are much easier to use and to interpret than is x-ray diffraction. In general the procedure followed is to calibrate the property that is being studied by using simple molecular systems of known structure, and then to investigate the property in proteins.

Optical rotatory power has been particularly useful in this respect because it is unusually sensitive to molecular conformation. It was discovered many years ago that all proteins are levorotatory, and that when they are denatured in solution their optical rotations invariably become more negative. Studies by Yang and Doty with

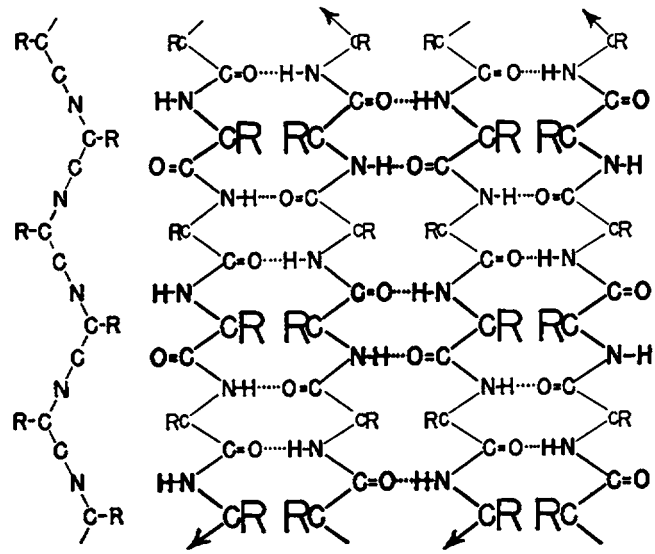
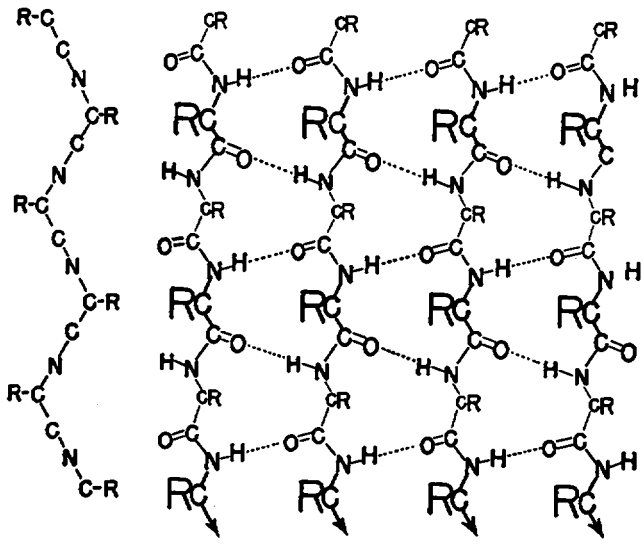
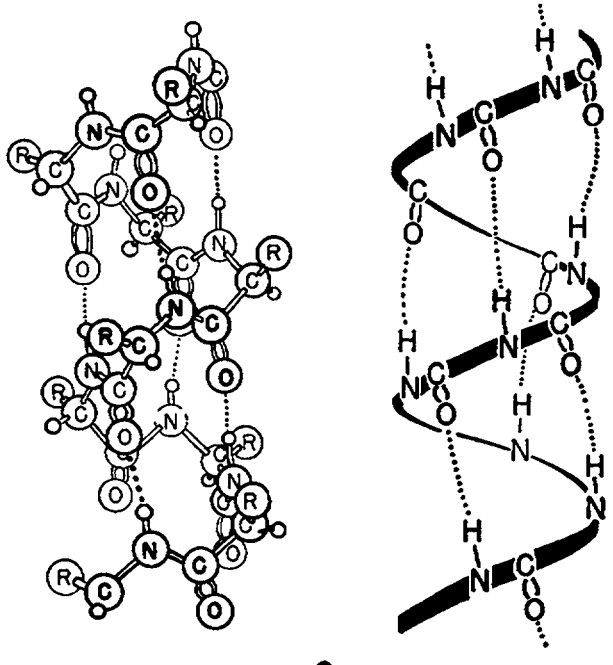


FIGURE 4 Some important conformations of polypeptide chains. A, the antiparallel pleated sheet. Polypeptide chain backbones run up and down the page, and alternate chains run in opposite directions as indicated by the arrows. The chains are held together in a grooved plane by C = O . . . H — N hydrogen bonds. B, the parallel pleated sheet. The polypeptide chains all run in the same direction. C, the right-handed α-helix. Right-hand sketch indicates the arrangement of peptide-peptide hydrogen bonds. (Adapted from drawings by Corey and Pauling.)



B



C

FIGURE 4B and C (see facing page for explanation).

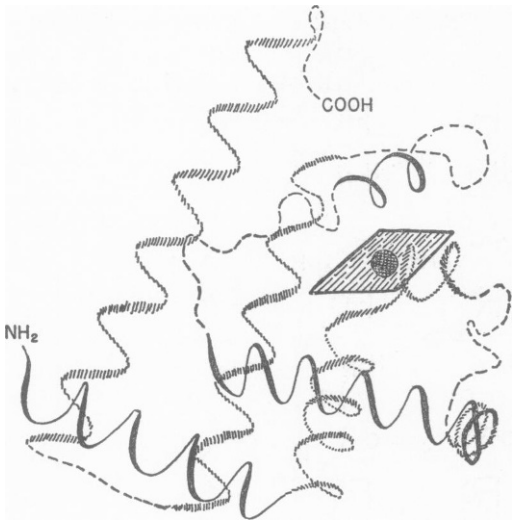


FIGURE 5 Rough sketch of three dimensional structure of the polypeptide chain of myoglobin (after Kendrew (6)). The dashed lines indicate non-helical portions of the chain. The shaded sphere and square represent the heme group, at which binding of oxygen occurs.

synthetic polypeptides synthesized from L-amino acids show that they are strongly levorotatory when they are dissolved in solvents that give them a randomly coiled structure, whereas they are weakly dextrorotatory when dissolved in solvents that are believed to confer an α-helical conformation onto the polypeptide chain (9). Since denatured proteins are believed to resemble in some degree the randomly coiled forms of poly L-amino acids, this observation would seem to suggest the view that native proteins are less levorotatory than denatured proteins because they contain some helically coiled structures within them.

Yang and Doty (9) also found that when the dependence of the optical rotation of proteins on the wave length of the light used in measuring the rotation was compared with that of randomly coiled and helical polypeptides, further correlations were suggested. The observed values of the optical rotation $[\alpha]_\lambda$, at wave length λ, were fitted to a semitheoretical equation proposed by Moffitt (10),

$$[\alpha]_\lambda M/100 = \frac{a_0\lambda_0^2}{\lambda^2 - \lambda_0^2} + \frac{b_0\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2}$$

where M is the average molecular weight of an amino acid residue in the protein or polypeptide, λ_o is a standard wave length, usually taken as 212 mμ, and a_o and b_o are empirical constants which both have the same dimensions as $[a]_\lambda$. The constant b_o was found by Doty and Yang to have the value of about $-600°$ for a polypeptide which is completely in the form of a right-handed helix, and it vanishes or is small when the polypeptide is in the random coil form. By fitting the observed variation of $[\alpha]_\lambda$ with λ for proteins to the above equation and assuming that the numerical value of b_o is a measure of the amount of helical structure present in the protein,

one can estimate the fraction of the polypeptide chain in the protein that is present in the helical form. When the method is applied to myoglobin the fraction of helix estimated in this way agrees with the fraction observed by Kendrew in his x-ray diffraction study of myoglobin. This method of studying the conformations of proteins has been reviewed in detail by Urnes and Doty (11). It is essentially an empirical approach, and we should not be surprised if its use in a structural problem as complex as that presented by the proteins may on occasion lead us to wrong answers. On the other hand it is certainly one of the most sensitive and convenient physical tools available to us at the moment for the detection of changes in the conformations of protein and polypeptide molecules, and on the whole it seems to be giving reasonable results.

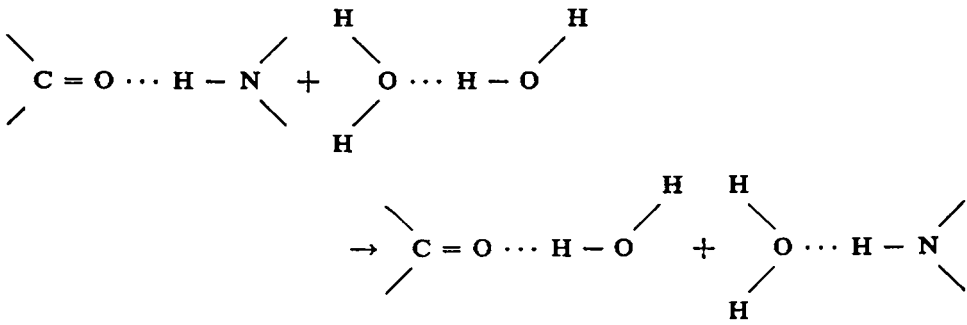
Recently promising advances have been made in the measurement of optical rotatory power (and also the closely related property of circular dichroism) in the far ultraviolet region of the spectrum, where the above equation is no longer valid (see Urnes and Doty (11)).

The absorption spectra of the various chromophoric groups present in proteins are influenced by the environments of the groups, and should therefore be capable of yielding some information on the nature of the surroundings of those groups. For instance, one might be able to determine whether these groups are immersed in a non-polar environment in the interior of the protein molecule, or whether they are exposed to the aqueous or other solvent that surrounds the molecule. The spectral shifts associated with tyrosine and tryptophan residues have been studied especially in this way. Recently much interest has developed in the absorption bands at 180 to 220 $m\mu$ ascribed to the peptide group. Pronounced hypo- and hyperchromic effects are observed in these bands when polypeptides and proteins undergo changes in conformation. This subject of spectral effects in proteins and polypeptides has just been thoroughly reviewed by Wetlaufer (12).

Some of the other physical tools that have been used for studying protein structure have been described elsewhere (13).

It is appropriate to discuss briefly the nature of the forces that must be responsible for the complex three dimensional structures that are beginning to be found in proteins. Why does the collection of atoms whose covalent bonds are described by diagrams of amino acid sequences like that shown in Fig. 2 fold itself up into the specific three dimensional structure that is present in a protein crystal and that is also undoubtedly present in solutions of native proteins? The large number of peptide linkages present in proteins makes it almost certain that hydrogen bonding involving the amide and carbonyl groups of the peptide link must be a factor here, and indeed, peptide-peptide hydrogen bonds of the kind postulated by Pauling and Corey have been found by Kendrew in myoglobin. It is important to realize, however, that native proteins are normally exposed to water, which is itself a very good hydrogen bond former. The replacement of a peptide-peptide hydrogen bond by peptide-

water hydrogen bonds should not be very difficult to accomplish,



The scant available evidence indicates that this reaction will proceed with an energy absorption of no more than about 1 kilocalorie (see Schellman (14)), and very possibly even less than this (see Levy and Magoulas (15), and Klotz and Franzen (16)).

About 25 to 40 per cent of the amino acid side chains of a typical protein are non-polar (for instance, the propyl group of valine, the butyl groups of leucine and isoleucine, and the benzyl group of phenylalanine). These groups have a low affinity for water, and it would be expected that proteins will tend to fold themselves into conformations in which the non-polar groups will be turned toward the interior of the molecule, and thus out of contact with water. Kendrew reports that the myoglobin molecule is indeed folded in just this way. This tendency of non-polar groups to adhere to each other in aqueous solutions has been referred to as "hydrophobic bonding" (17) or "apolar bonding " (18). This type of bond has the interesting property that the driving force responsible for its stability is entropic in character (Frank and Evans (19)); the properties of small non-polar molecules in aqueous solutions indicate that these molecules are encased in cages of ordered water molecules. The energy of the water molecules in these cages is somewhat lower than the energy of an equivalent amount of normal liquid water, which would normally tend to favor cage formation, and hence increase the solubility of non-polar molecules in water. Because the water in the cages is much more ordered than liquid water, however, cage formation is accompanied by a large decrease in entropy, and this tends to make the cages unstable. The entropy effect turns out to be much greater than the energy effect, and it is for this reason that hydrocarbons and other non-polar molecules are only slightly soluble in water. The same factors are presumably also responsible for the stability of hydrophobic bonds in proteins (13).

Klotz (18) has given a somewhat different interpretation to these observations and suggests that the non-polar groups of proteins induce a "crystallization" of water in the vicinity of protein molecules that is a major factor in stabilizing the native conformations of proteins.

Our knowledge and understanding of all these factors that determine the three

dimensional structures of proteins are developing at a rapid rate. Within the next ten years we should have a much clearer picture of how they operate, especially if the x-ray crystallographic approach continues to develop at the rate that it has for the last 10 years.

REFERENCES

1. SMYTH, D. G., STEIN, W. H., and MOORE, S., The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations, *J. Biol. Chem.*, 1963, **238**, 227.
2. SORM, F., and KEIL, B., Regularities in the primary structure of proteins, *Adv. Protein Chemistry*, 1962, **17**, 167.
3. PAULING, L., COREY, R. B., and BRANSON, H. R., The structure of proteins, *Proc. Nat. Acad. Sci.*, 1951, **37**, 205. PAULING, L., and COREY, R. B., The pleated sheet, a new layer configuration of polypeptide chains, *Proc. Nat. Acad. Sci.*, 1951, **37**, 251.
4. PERUTZ, M. F., Eighth Weizmann Memorial Lecture Series, Amsterdam, Elsevier Publishing Co., 1962, chapter 1
5. KENDREW, J. C., Myoglobin and the structure of proteins, *Science*, 1963, **139**, 1259; Side chain interactions in myoglobin, *in Brookhaven Symposia in Biology, No. 15*, Enzyme Models and Enzyme Structure, Upton, New York, Biology Department, Brookhaven National Laboratory, 1962, 216.
6. KENDREW, J. C., The three-dimensional structure of a protein molecule, *Scient. Am.*, 1961, **205**, 96.
7. BLAKE, C. C. F., FENN, R. H., NORTH, A. C. T., PHILLIPS, D. C., POLJAK, R. J., STANDORD, R. H., JR., MARSH, R. E., COREY, R. B., DICKERSON, R. E., REDDY, J. M., PINKERTON, M., and STEINRAUF, L. K., The structure of lysozyme, *Nature* 1962, **196**, 1173.
8. AVEY, H. P., CARLISLE, C. H., and SHUKLA, P. D., A tentative structure for ribonuclease based on x-ray study, *in Brookhaven Symposia in Biology, No. 15*, Enzyme Models and Enzyme Structure, Upton, New York, Biology Department, Brookhaven National Laboratory, 1962, 199.
9. YANG, J. T., and DOTY, P., The optical rotatory dispersion of polypeptides and proteins in relation to configuration, *J. Am. Chem. Soc.*, 1957, **79**, 761.
10. MOFFITT, W., Optical rotatory dispersion of helical polymers, *J. Chem. Physics*, 1956, **25**, 467. See also MOFFITT, W., and YANG, J. T., The optical rotatory dispersion of simple polypeptides, *Proc. Nat. Acad. Sci.*, 1956, **42**, 596.
11. URNES, P., and DOTY, P., Optical rotation and the conformation of polypeptides and proteins, *Adv. Protein Chemistry*, 1961, **16**, 401.
12. WETLAUFER, D. B., Ultraviolet spectra of proteins and amino acids, *Adv. Protein Chemistry*, 1962, **17**, 304.
13. KAUZMANN, W., Some factors in the interpretation of protein denaturation, *Adv. Protein Chemistry*, 1959, **14**, 1.
14. SCHELLMAN, J. A., The thermodynamics of urea solutions and the heat of formation of the peptide hydrogen bond, *Compt. rend. Lab. Carlsberg, serie chimique*, 1955, **29**, 223.
15. LEVY, M., and MAGOULAS, J. P., Effect of urea on hydrogen bonding in some dicarboxylic acids, *J. Am. Chem. Soc.*, 1962, **84**, 1345.
16. KLOTZ, I. M., and FRANZEN, J. S., Hydrogen bonds between model peptide groups in solution, *J. Am. Chem. Soc.*, 1962, **84**, 3461.
17. KAUZMANN, W., Denaturation of proteins and enzymes, *in* The Mechanism of Enzyme Action, (W. D. McElroy and B. Glass, editors), Baltimore, The Johns Hopkins Press, 1954, 70.

18. KLOTZ, I. M., Non-covalent bonds in protein structure, *in Brookhaven Symposia in Biology, No. 13*, Protein Structure and Function, Upton, New York, Biology Department, Brookhaven National Laboratory, 1960, 25.
19. FRANK, H. S., and EVANS, M. W., Free volume and entropy in condensed systems. III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes, *J. Chem. Physics*, 1945, **13**, 507.