Picture of Proteins by X-Ray Diffraction

GOPINATH KARTHA

Roswell Park Memorial Institute, Buffalo, New York 14203

Received June 21, 1968

During the past two decades research in the biological sciences appears to have undergone a subtle change, the focus shifting from the cellular to the molecular level. Proteins form a large and varied class of molecules of biological significance, and it is now recognized that many of the biological properties and specificities of these molecules are best understood in terms of their physical organization at the atomic level.

Laue's discovery[1] of the diffraction of X-rays by crystals initiated the use of a powerful tool for determining the three-dimensional arrangement of molecules. At its present stage of development, for medium-sized molecules of molecular weight as high as 1000, this technique, though elaborate and time consuming, provides virtually complete stereochemical information, provided single crystals measuring a few tenths of a millimeter can be prepared.

The situation is not so favorable in regard to protein structure studies. The molecular weights of proteins as determined by physical methods are much higher, varying from millions in the case of virus proteins to thousands in the case of even the smallest protein to be studied. Such giant molecules have less tendency to pack in specific long-range repetitive patterns to produce crystals of the high quality needed to give a sufficient number of diffraction data. The covalent distance between atoms in molecules is of the order of 1.5 Å, and in the case of crystals of molecules of medium complexity X-ray diffraction data are usually measurable to a spacing of 1 Å or smaller. However, even with the best protein crystals, the diffraction pattern fades off rapidly for spacings less than 2 Å. The construction of the picture of a protein molecule with all atoms resolved solely from diffraction data of the quality available is out of the question. Hence we must supplement information from X-ray diffraction studies with what we know from other sources in order to deduce the detailed structure of a protein.

What Is a Protein?

Proteins are naturally occurring long-chain polymers which represent polymerization of α-amino acids. Twenty-two different amino acid units may appear in proteins, these differing chiefly in the identity of the "R" side chain. These side groups vary in complexity from a single hydrogen atom in the case of glycine to large groups containing 17 or 18 atoms in arginine and tryptophan.

Thus the main chains of all proteins have the same kind of backbone structure. They differ in length from protein to protein and in the proportions and arrangements of the different kinds of amino acid units. Individual protein molecules may contain over 100 amino acid units, and the number of possible permutations and combinations is huge.

As a first step in determining the organization of a protein, one needs to know the number of amino acid units of each kind in the protein chain and their sequence of arrangement. Beginning with the pioneering work of Martin and Synge,[2] chemical methods of amino acid analysis and sequence determination have been developed which provide the desired information, the so-called *primary* structure, for a number of proteins.[3-5]

A protein molecule may sometimes contain more than one chain and also non-amino acid groups as, for example, in hemoglobin. In some proteins, in ribonuclease for instance, the chains may be looped or cross-linked by cystine disulfide bridges.

To understand the function of the protein we need not only its *primary structure* but also the three-dimensional arrangement of the main and side chains. These arrangements are sometimes referred to as *secondary* and *tertiary* structures. The importance of secondary and tertiary structure is obvious from the fact that many of the physical and biological properties of molecules are drastically changed by treatments which are too mild to break the covalent bonds which hold the primary structure together.

Proteins can be divided into two broad classes, the *fibrous* and the *globular* proteins. The fibrous proteins are insoluble, and their function is mainly structural as a building material in hair, bone, teeth, muscle, tendon, etc. The globular proteins are more extensively distributed in the body, are soluble, and are important in metabolic processes, being capable of many elaborate and specific chemical functions. From the crystallographer's point of view, these are also two quite distinctive forms, as shown in Figure 1 by the quality of the X-ray diffraction patterns produced by members of these two classes.

It is immediately obvious from the two patterns that globular proteins form excellent three-dimensional crystals which give thousands of sharp reflections of a wide range of intensity and which can be measured with great accuracy. The fibrous proteins, on the other

(1) M. V. Laue, *Sitzb. Math. Physik. Klasse Bayer, Akad. Wiss. München.*, 303 (1912).

(2) A. J. P. Martin and R. L. M. Synge, *Biochem. J.*, **35**, 1358 (1941).

(3) F. Sanger and H. Tuppy, *ibid.*, **49**, 463 (1951).

(4) D. G. Smyth, W. H. Stein, and S. Moore, *J. Biol. Chem.*, **238**, 227 (1963).

(5) R. E. Canfield, *ibid.*, **238**, 2968 (1963).

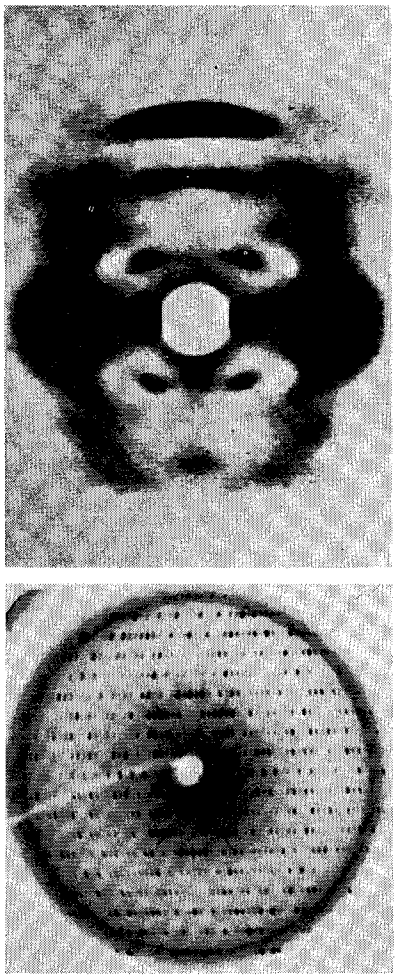

Figure 1. (a, top) X-Ray diffraction pattern from fibrous protein collagen. The fiber axis is at 75° to the X-ray beam. (b, bottom) X-Ray diffraction pattern from a crystal of a modification of hemoglobin. The picture shows only one section of the diffraction pattern. The total pattern is composed of many similar sections.

hand, show only disoriented one-dimensional periodicity which at best gives rather streaky and diffuse diffraction patterns.

Structure of Fibrous Proteins

Paradoxically, in spite of the meager data available from X-ray diffraction studies, early success was obtained in investigations of the structures of fibrous proteins. This is because in a well-oriented fiber the protein chains may be arranged parallel to the fiber axis or as a helix with the helical axis coinciding with the fiber axis. The intensity distribution in the medium-scattering range of the fiber diffraction pattern is mostly determined by the arrangement of the backbone chain. Hence, knowledge of a few parameters such as the repeat distance along the fiber axis and the pitch of the helix gives some idea of the organization of the chains in the molecule. Detailed crystallographic studies on simple amino acids and peptides[6-8] have

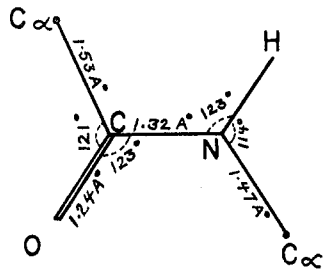

Figure 2. The atomic parameters in peptide bonds from studies of small peptides.

provided valuable information regarding the stereochemistry of the peptide bond connecting the amino acid residues and the basic principles that underlie stable protein conformations. These studies showed that the peptide bonds (–C(=O)NH–) joining the α-carbon atoms are remarkably planar and that within the peptide unit bond lengths and angles are not much affected by changing the R groups at the α carbon atom. The stability of these structures is enhanced by hydrogen bonding between the CO and NH groups of the main chain with an NH---O distance of 2.8 ± 0.1 Å (Figure 2).

The facts that the peptide links are planar, that maximum use is made of the hydrogen-bonding capabilities of NH–OC pairs, and that the residues usually occur in the *trans* configuration, together with the usual van der Waals radii of atoms, enabled enumeration of principles that formed the basis of the search for stereochemically acceptable models of secondary structure. These, along with energy considerations, severely restrict the number of possible conformations that the backbone is capable of assuming. The variable parameters are few, and the X-ray diffraction theory of helical structures[9] in conjunction with other methods enables one to select the most probable model. The application of these principles is exemplified in detecting the arrangements of the chains in the keratin[10] and collagen[11] groups of proteins. The probable secondary structure for the collagen group of proteins was proposed to have the form of a triple helix on the basis not only of X-ray diffraction and infrared dichroism studies on collagen fibers but also of the fact that this protein contains an unusually large percentage of glycyl, prolyl, and hydroxyprolyl residues in its primary structure. That the nonintegral α helix, with 3.7 residues per turn, and the pleated-sheet structures proposed for fibrous proteins on the basis of these principles have in fact been observed in globular protein studies lends further support to the validity of these structural principles.

Structure of Globular Proteins

The situation is somewhat different in globular proteins, as these are akin to a rolled-up ball of string with no pronounced chain direction. As such, globular

(6) J. Donohue, *J. Am. Chem. Soc.*, **72**, 949 (1950).
(7) D. P. Shoemaker, J. Donohue, and C. S. Lu, *Acta Cryst.*, **6**, 241 (1953).
(8) L. Pauling and R. B. Corey, *Proc. Roy. Soc.* (London), **B141**, 21 (1953).
(9) W. Cochran, F. H. C. Crick, and V. Vand, *Acta Cryst.*, **5**, 581 (1952).
(10) L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Natl. Acad. Sci. U. S.*, **37**, 205 (1951).
(11) G. N. Ramachandran and G. Kartha, *Nature*, **176**, 593 (1955).

proteins take up a large variety of conformations which are decided not only by interactions between the main chain atoms but also by interactions between side groups and with the solvent around the protein molecules. The number of possible arrangements is very much larger, and success in predicting plausible and satisfactory models solely from stereochemical and energetic considerations has been small. However, the comparative abundance and excellence of the X-ray diffraction data from crystals of globular proteins make them amenable to the laborious but detailed and powerful techniques developed by X-ray crystal structure analysts.

Protein Crystallography

Many proteins in equilibrium with their mother liquor under suitable conditions form good three-dimensional crystals in which one, or a few, molecules are repeated in three dimensions forming large regular arrays. For many proteins, crystals measuring 1 mm or more can be grown with little difficulty. These crystals usually contain 40% or more by weight of solvent, which is lost on drying with accompanying breakdown of three-dimensional periodicity. Further, these crystals deteriorate on long irradiation by X-rays. However, techniques of mounting crystals[12] and measuring X-ray intensities which minimize these effects have been developed.

Once protein crystals are available, studies of unit cell geometry, molecular weight, symmetry of arrangement within the unit cell, etc., are fairly simple and straightforward. The detailed atomic arrangement inside the cell is, however, a different problem.

The relative intensities of X-ray scattering from the crystal planes depend on the details of the electron density distribution at every point in the unit cell of the crystal. It is necessary to use this information to build up the image of the protein molecule. For a typical protein crystal, with an X-ray beam of suitable wavelength, one can obtain thousands of well-defined diffraction intensity maxima. The proper experimental geometry needed to measure each of these diffraction maxima can be calculated from the well-known Bragg law[13] of reflections of X-rays from crystal planes and the indices *hkl* of these planes.

The Fourier series formulation of the electron density distribution was proposed by Bragg[14] as a standard tool for analysis of X-ray diffraction results from complex crystals. In this formulation, the electron density, ρ, at point x, y, z in a three-dimensionally periodic crystal is represented by eq 1, where $\mathbf{F}_{hkl}$ is the structure

$$\rho(x,y,z) = \sum_h \sum_{k=-\alpha}^{\alpha} \sum_l \mathbf{F}_{hkl} \exp\,[-2\pi i \times (hx + ky + lz)] \quad (1)$$

factor of reflections from planes of Miller indices *hkl*.

(12) M. V. King, *Acta Cryst.*, **7**, 601 (1954).
(13) W. L. Bragg, *Proc. Cambridge Phil. Soc.*, **17**, 43 (1913).
(14) W. L. Bragg, *Proc. Roy. Soc.* (London), **A123**, 537 (1929).

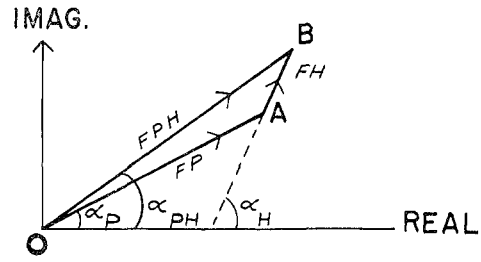

Figure 3. Vector **AB** added to **OA** gives resultant **OB**. Knowing both magnitude and direction of vector **AB**, but only magnitudes of **OA** and **OB**, it is possible to obtain (except for a sign ambiguity) the angle difference between **AB** and the other two vectors.

The summation is to extend over all diffraction maxima.

The structure factor $\mathbf{F}_{hkl}$ for a given plane is in general a complex number and can be represented by an amplitude factor and a phase factor, as in eq 2. The

$$\mathbf{F}_{hkl} = |F_{hkl}| \exp\,[2\pi i \alpha(hkl)] \quad (2)$$

amplitude factor $|F_{hkl}|$ of the reflection is related to the integrated intensity of the diffraction maximum in a direct manner, and the experimental measurement of the X-ray reflections enables these to be obtained for the various terms on the right-hand side of eq 1. However, no experimental method has yet been designed that gives directly information regarding the phase angle $\alpha(hkl)$ in eq 2. As such, eq 1 is of little use for direct synthesis of the electron density distribution and ultimately the picture of the protein molecule until the missing information regarding the phases of the reflections is supplied. It is this missing information (the well-known *phase problem* of X-ray crystallography) that makes the journey from the crystal to the magnified picture of the unit cell contents at atomic resolution neither automatic nor straightforward.

Solution of the Phase Problem in Protein Crystallography

Even though the first diffraction patterns[15] of globular proteins were taken a quarter of a century ago, and in spite of the many ingenious methods that have been used in attempting to interpret the data in the absence of phase information, there was little progress in the use of X-ray diffraction information from these crystals by means of eq 1 until development and application of the isomorphous series method[16] allowed the phase angles of protein reflections to be evaluated.

As in phase-contrast microscopy, the strategy for obtaining information regarding phase differences is to convert them in some manner into amplitude differences that can be experimentally measured. The principle is illustrated in Figure 3.

As early as 1927, Cork, in his study of alums,[17] had shown that amplitude changes caused by isomorphous substitution of atoms at known position by other atoms

(15) J. D. Bernal and D. Crowfoot, *Nature*, **133**, 794 (1934).
(16) D. W. Green, V. M. Ingram, and M. F. Perutz, *Proc. Roy. Soc.* (London), **A225**, 287 (1954).
(17) J. M. Cork, *Phil. Mag.*, **4**, 688 (1927).

of different scattering power lead to phase information of real structure amplitudes. Extension of this method to noncentrosymmetric structures[18] is not difficult except that at least three isomorphous crystals are necessary for unambiguous evaluation of the phases. Harker[19] has given a geometrical construction for the evaluation of phase angles from the measured amplitudes of the three isomorphs and knowledge of the replaceable atom configurations. In these, the equivalent of the vector $\mathbf{F}_H$ of Figure 4, whose amplitude and phase are known, is the change in structure factor caused by the addition of heavy atoms in known positions. In such a case, the vector $\mathbf{F}_H$ is computed using eq 3, where $\mathbf{f}(j)$ is the scattering factor of atom j at

$$\mathbf{F}_H(hkl) = \sum_j \mathbf{f}(j) \exp [2\pi i(hx_j + ky_j + lz_j)] \quad (3)$$

x_j, y_j, z_j and the summation extends over all atoms that cause the amplitude change. Knowing the types and positions of these atoms, one can calculate their contribution $\mathbf{F}_H$ to every reflection hkl both in magnitude and phase using eq 3.

Recently, anomalous scattering effects from replaceable atoms, which are usually heavy atoms, have also been valuable in obtaining information regarding protein phases. These effects arise from the fact that, in general, the atomic scattering factor $\mathbf{f}(j)$ in eq 3 is a complex number although the complex component is negligible in most cases. However, for suitable wavelengths and atoms, the correction

$$\mathbf{f}(j) = f_0(j) + \Delta f'(j) + if''(j) \simeq f'(j) + if''(j) \quad (4)$$

becomes apprecible and gives a scattering component for $\mathbf{F}_H$ which has a phase advance of $\pi/2$ over the component that is in phase with the incident beam. This out-of-phase component has the effect of making the intensities from the front and back surfaces of an atomic plane different, as shown in Figure 4.

The differences between the amplitudes $\mathbf{F}(hkl)$ and $\mathbf{F}(\bar{h}\bar{k}\bar{l})$ are small, but, if accurately measured, they can provide valuable information regarding protein phase angles.[20,21] In fact, it can be shown that the information one obtains from an isomorphous heavy atom derivative and these "Bijovet" differences, caused by anomalous scattering, are complementary in nature, and, between them, lead to an unambiguous evaluation of the protein phase angle. This complementarity of isomorphous and anomalous scattering information from heavy-atom-derivative crystals of proteins has been much exploited[22,23] and is of considerable value in protein structure investigations.

Even though other more direct methods, not dependent on the availability of heavy-atom isomorphous protein crystals, have been put forward[24] and may in due course turn out to be of considerable value, the successes in the actual solution of unknown protein structures have been, until now, mostly based on the use of isomorphous and anomalous scattering methods.

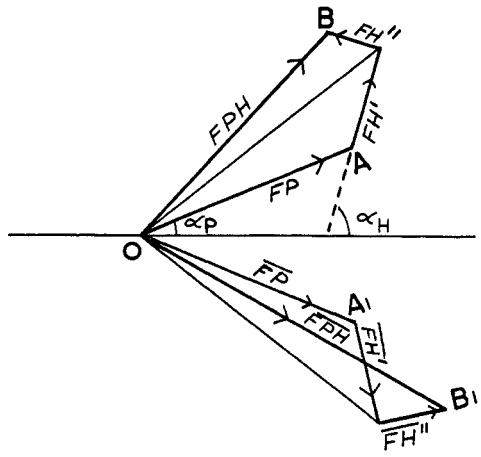

Figure 4. When the heavy-atom scattering has an anomalous component $\mathbf{F}_H''$, the structure amplitudes $|F_{PH}|$ and $|F_{\overline{PH}}|$—from hkl and $\bar{h}\bar{k}\bar{l}$ reflections—may be unequal. This difference, though usually small, if accurately measured provides information on the phases of protein reflections.

For the application of multiple isomorphous replacement and anomalous scattering (Miraas) methods for protein phase determination, it is essential to grow protein crystals into which a few heavy atoms have been incorporated at specific sites without materially affecting the rest of the structure. The presence of a large volume of solvent between the protein molecules in the crystal makes this possible, and, with luck, it is occasionally possible to obtain satisfactory heavy-atom-derivative crystals suitable for protein phase determination by Miraas techniques.

From suitable heavy-atom-derivative crystals, diffraction data, including intensity differences between Bijovet pairs of reflections, are measured to the desired resolution; these structure amplitudes, together with those of the parent protein, constitute the experimental basis for synthesizing the protein image. The first step in the process of phase evaluation is to determine if specific heavy-atom binding has indeed taken place and, if so, to locate the positions of the heavy atoms in the unit cell of the crystal.

For determining the positions of the heavy atoms, Patterson maps providing interatomic vectors are usually employed. These maps are computed as a two- or three-dimensional Fourier summation similar to eq 1, the coefficients being derived from measured diffraction amplitudes only. It has been shown[22,23] that the use of $(|F_{PH}| - |F_P|)^2$ or $(|F_{PH}| - |F_{\overline{PH}}|)^2$ or some suitable combination of these as coefficients in the series gives maps that greatly facilitate location of the heavy atom. These heavy-atom parameters in the different derivative crystals are referred to the same common origin[22,25] and refined, prior to calculation of their scattering contributions, $\mathbf{F}_H$, for phase evaluation of each observed reflection.

(18) C. Bokhoven, J. C. Schoone, and J. M. Bijovet, *Acta Cryst.*, **4**, 275 (1951).
(19) D. Harker, *ibid.*, **9**, 1 (1956).
(20) J. M. Bijovet, *Nature*, **173**, 888 (1954).
(21) A. C. T. North, *Acta Cryst.*, **18**, 212 (1965).
(22) G. Kartha and R. Parthasarathy, *ibid.*, **18**, 745, 749 (1965).
(23) B. W. Mathews, *ibid.*, **20**, 82 (1966).
(24) M. G. Rossmann and D. M. Blow, *ibid.*, **16**, 39 (1963).
(25) D. C. Phillips, *Advan. Structure Res.* **2**, 75 (1966).

Errors in measured amplitudes, in estimates of atomic parameters, and in the assumption of isomorphism of derivative crystals all introduce errors in the estimates of protein phase angles. It is essential to reduce the effect of these errors[26] in the final electron density maps in order to obtain the maximum amount of information with as little noise as possible.

X-Ray Image of Proteins

As soon as estimates for the phase angles of the reflections have been obtained, it is possible to compute a point-by-point representation of the electron density distribution in the unit cell by means of eq 1. Even for a low-resolution image of the protein this computation involves a great deal of numerical work; however, the availability of the present generation of fast, large memory computers makes this feasible. Unlike the situation in the study of small molecules, unfortunately, the electron density maps of proteins so obtained present problems of interpretation. This is caused not only by the errors inherent in the estimated phase angles but also by the resolution available in computing the protein map, which is not sufficient to resolve atomic detail. In fact, in the very coarse resolution maps that are initially obtained, even features that eventually turn out to be regions of highest electron density[27] under adequate resolution may fail to show up.

In the absence of a large amount of secondary structure it may be necessary to proceed to a resolution of ~2.5 Å before the tertiary structure of the protein is revealed with sufficient clarity in the computed image. At such a resolution the image, together with a knowledge of the primary structure from chemical sequence work, can give us a good picture of the protein molecule. In the absence of any knowledge of the amino acid sequence, the resolution and quality of the X-ray image needed to show all the important details of the structure with clarity and certainty may be at, or even just beyond, the limit of most protein crystallographic studies.

Further, the side chains of amino acid units often take up different conformations in different molecules in the crystal, and the image obtained by diffraction studies is a weighted average of all conformations. This tends to wash out the details of long side chains that extend into the solvent from the surface of the protein molecule. However, at resolutions near 2.5 Å, or better, characteristic shapes in the X-ray image are shown by different amino acid residues, and this, in conjunction with partial chemical information, may settle most of the features of the three-dimensional organization of a protein molecule of average complexity.

What Are Protein Molecules Like?

The number of globular proteins whose structures are known in detail is still too small for safe generalizations regarding their topology or appearance. The solution of the myoglobin structure[28] by Kendrew and coworkers in 1958 in Cambridge remained a unique success until 7 years later the structure of the enzyme lysozyme was determined by Phillips[29] and colleagues in London. Following the report on the enzyme ribonuclease[30] in early 1967, in the United States, the tertiary structures of the enzymes chymotrypsin,[31] carboxypeptidase, and papain appeared in rapid succession, and studies of other proteins such as cytochrome *c* and carbonic anhydrase are well under way. We hope to have good information regarding the folding of the chains in half a dozen or more proteins before the end of this decade.

The protein chain of myoglobin (mol wt ~17,000) is mostly folded as segments of α helices, of different lengths and axial directions, connected at their ends by short sections of chains of no regular arrangement. More than 75% of the residues are in the α-helical conformation. The over-all shape of the molecule is that of an oblate ellipsoid, with a pocket containing the heme group. The function of myoglobin is not to catalyze any specific chemical reaction but to assist in transport of oxygen. The oxygen is attached to the nonprotein heme group which contains an iron atom in the center. Transport of the whole complex is facilitated by having a coat of hydrophilic residues around the surface of the molecule, while the orientation of the hydrophobic side chains toward the inside creates an oily environment of low dielectric constant which enables the heme group to accept the oxygen molecule with ease. The tertiary structure and function of hemoglobin which carries oxygen in blood is basically the same, though here additional complications arise because the molecule consists of four myoglobin-like subunits.

Both lysozyme and ribonuclease (Figure 5) are enzymes which catalyze specific biological reactions inside the living cell. They have molecular weights in the neighborhood of 14,000. For both enzymes, the primary sequences have been chemically established[4,5] and were available to aid the interpretation of the electron density maps. The function of lysozyme is to break down the polysaccharides composing the bacterial cell walls, while ribonuclease catalyzes the hydrolysis of ribonucleic acid by breaking down the phosphodiester linkage in the polynucleotide sequence.

Both enzyme molecules consist of a single protein chain looped at four positions by cystine disulfide bridges. Compared to myoglobin and hemoglobin, both contain much smaller, but different, percentages of residues in helical conformation. They also show sections of polypeptide chains that double back on

(26) D. M. Blow and F. H. C. Crick, *Acta Cryst.*, **12**, 794 (1959).
(27) G. Kartha, *Nature*, **214**, 234 (1967).
(28) J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, *ibid.*, **181**, 662 (1958).
(29) C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *ibid.*, **206**, 757 (1965).
(30) G. Kartha, J. Bello, and D. Harker, *ibid.*, **213**, 862 (1967).
(31) B. W. Mathews, P. B. Sigler, R. Henderson, and D. M. Blow, *ibid.*, **214**, 652 (1967).

Figure 5. The three-dimensional representation of the folding of the main chain in bovine pancreatic ribonuclease crystals from X-ray diffraction studies. The phosphate group, shown shaded, binds in the cleft of the molecule. Binding of nucleotide inhibitors shows that this cleft is also part of the active site.

themselves forming the antiparallel pleated-sheet structure proposed by Pauling and Corey in 1951 for the β form of keratin. There is also a clustering of hydrophobic residues in some regions of both enzymes, but in general their structures are much more open than for myoglobin. These two enzymes also show a similarity in molecular topology near the active site, each such site lying in a cleft on one side of the molecule in which substrates are held in an environment suitable for the catalysis of a particular reaction. The substrate binding site in ribonuclease is rich in positively charged groups suitable for binding the negatively charged parts of the substrate. However, generalization of these characteristics, for example, to infer the presence of a cleft on one side of all enzyme molecules, does not seem to be warranted. The structure of chymotrypsin—which catalyzes the breakage of peptide and ester bonds—does not show any pronounced pair of jaws for chewing up its substrate molecules.

Conformation in Crystals and Solutions. Are They the Same?

It is pertinent to ask if the conformations one finds in crystals are similar to those existing in the biological environment of the cell. At present, while it is not possible to give a definite answer in the affirmative, it seems very likely that a protein molecule does possess an individuality which it preserves basically unaltered when it passes from the solvent to the crystalline environment under normal conditions.

Many lines of evidence support this view. Proteins, while crystallizing in general, carry with them 40–50 wt %, or more, of solvent. These solvent molecules do not show signs of extensive ordered arrangement between the protein molecules, although the latter do form an ordered crystal lattice. The regions of contact between protein molecules in crystals seem to be small and few. Hence, one can argue *a priori* that the protein molecule in crystals should behave roughly in the same way as in highly concentrated solutions, and that no drastic change in conformation is brought about by going from the dissolved to the crystalline state. Support for this view is found in the fact that many physical measurements on the molecules in solution, such as optical rotatory dispersion, give results similar to those one obtains from X-ray studies of protein crystals.

The ribonuclease studies provide compelling evidence in support of this view. For example, crystalline ribonuclease has the same activity toward, and specificity for, its substrate as in solution.[32] The similarity in the over-all arrangement of the protein chains in ribonuclease A and ribonuclease S,[33] in spite of the fact that the former is crystallized from an alcoholic medium and the latter from a strong ammonium sulfate medium, is another argument against the possibility of large conformational changes due to environmental variations.

(32) J. Bello and E. F. Nowoswiat, *Biochim. Biophys. Acta*, 105, 325 (1965).

(33) H. W. Wyckoff, K. D. Hardman, N. M. Allewell, I. Tadashi, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, 242, 3984 (1967).

One way of investigating conformational relationships in solution is by chemical investigation of residues near the site of enzyme activity. Both for ribonuclease and for chymotrypsin the results obtained by such studies are in harmony with what is seen in the X-ray images of the molecules in the crystalline state. For example, in ribonuclease A, the two reactive histidines at positions 12 and 119 are implicated at the active site by the chemical evidence[34] of Crestfield, *et al.* The interesting synthetic S peptide studies of Hofmann[35] and coworkers on the binding and active sites of ribonuclease and the arrangement of the tyrosine residues as deduced by Shugar[36] and others are in reasonable agreement with the results found in the crystalline state. From all these, it seems a reasonable hypothesis that crystallographic studies do give pictures of protein molecules that represent quite well molecular conformations in the biological environment.

What Are the Future Possibilities?

Once the structure of a protein is worked out in laborious detail by crystallographic study in a given crystalline form, it is possible to investigate small but very significant variations in the structure with limited additional effort. For example, the binding of small substrates and inhibitor molecules and the results of specific chemical modifications may be easily studied, if the modified form can be coaxed to crystallize in a form basically similar to that of the parent protein. The method of looking at the differences in electron densities before and after modification is a powerful tool in the study of interactions of small molecules with proteins. The studies of azide myoglobin[37] and arsenated ribonuclease[38] show that even quite small changes in electron density can be clearly detailed in a large protein molecule. The type of study is also helpful in establishing the details of the binding of substrate and inhibitor molecules on an enzyme and thus in suggesting features of the mechanism of enzyme action.[39]

When two similar molecules crystallize in forms which are not isomorphous, study of the modification by the difference electron density method is not straightforward. However, knowing the details of molecular arrangement in one crystalline form can simplify search for possible arrangements of the same molecule in different crystalline forms either by model building and packing considerations or by more general mathematical fitting techniques applied in real or reciprocal space.

Evolution and Protein Conformation

When one contemplates the great size of a three-dimensional protein molecule in comparison with the small region which is its active center and the modest size of the molecule on which it operates, he cannot but wonder why nature is so lavish in the design of these giant molecules and whether the same job could possibly have been done with equal efficiency by a much smaller molecule. For example, in the design of the myoglobin molecule, is it absolutely essential to have a molecule of weight around 17,000 to transport efficiently one molecule of oxygen from one part of the cell to the other? Is it possible that only a small part of the molecule is really essential for protecting, controlling, and performing the specific function for which the molecule is designed and that most of the molecule is just useless appendage?

It seems that, during the millions of years of evolution, molecules performing basically similar functions, but having different evolutionary histories, have undergone subtle changes which enable each molecule to work in harmony with its own specific surroundings. Of the multiple mutations causing changes in the primary sequence of amino acids, the molecule has chosen those unique combinations most suited to its specific environment. Mutations that are biologically ineffective, either because they are inefficient in performing their chemical functions or because their chemical sequences cannot fold up into the conformations needed for this specific function, are soon rejected.

However, even though in consequence many proteins from different biological sources which perform similar chemical functions have large variations in primary structure and crystallize in different forms, it is conceivable that the nature of the active center and the over-all tertiary structure do remain similar. Though almost one-third of the residues differ in ribonuclease from the pancreas of cows and rats, detailed examination shows[40] that the main changes in the residues are mostly in regions away from the active site. It seems possible that, in spite of the large differences in the primary structure of the two ribonucleases, the rat enzyme may be folded into a configuration somewhat akin to that of cows in order to achieve the same function.

It is too early to speculate on the question of how the protein molecule folds up into its active three-dimensional configuration after it is synthesized as a one-dimensional chain. The phenomenon of reversible unfolding into a random coil and subsequent spontaneous folding back into a configuration very similar to that of the native protein,[41] at least in the case of some proteins, strongly suggests that the primary sequence does contain all the information needed to fold the molecule into its native configuration, which is pre-

(34) A. M. Crestfield, W. H. Stein, and S. Moore, *J. Biol. Chem.*, **238**, 2413, 2421 (1963).
(35) K. Hofmann, F. M. Finn, M. Limetti, J. Montibeller, and G. Zanetti, *J. Am. Chem. Soc.*, **88**, 3633 (1966).
(36) D. Shugar, *Biochem. J.*, **52**, 142 (1952).
(37) L. Stryer, J. C. Kendrew, and H. C. Watson, *J. Mol. Biol.*, **8**, 96 (1964).
(38) G. Kartha, J. Bello, and D. Harker, "Structural Chemistry and Molecular Biology," W. H. Freeman, San Francisco, Calif., 1968, p 29.
(39) C. C. F. Blake, L. N. Johnson, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Proc. Roy. Soc.* (London), **B167**, 378 (1967).
(40) J. J. Beintema and M. Gruber, *Biochim. Biophys. Acta*, **147**, 612 (1967).
(41) C. B. Anfinsen, *Harvey Lectures*, **Ser. 61**, 95 (1965–1966).

sumably also thermodynamically the most stable one. However, the facts that widely different primary structures fold up to similar tertiary structures and that theoretical attempts at predicting the later from the primary sequence solely as an energy minimization problem have not yet met with appreciable success in any practical case make this an extremely interesting and challenging problem. Whether the conformation that the protein finally adopts is indeed a unique energy minimum or whether it is only one of the many local minima into which the protein chain can be coaxed by gentle prodding from one minimum to another is one of the as yet unanswered but hotly discussed questions in biology today.

My thanks are due to the National Science Foundation, the National Institutes of Health, and the Roswell Park Memorial Institute for support and to Mrs. C. Vincent and J. C. Wallace for preparing some of the diagrams in this article. I wish to record my indebtedness to Professor David Harker for continued encouragement and advice and to Dr. Jake Bello whose ingenuity and skill provided us with all the crystalline material used in obtaining the results on ribonuclease discussed here. It is also my pleasure to thank Professor M. F. Perutz, who first showed me how to take X-ray diffraction photographs of globular proteins, and Professor G. N. Ramachandran, who initiated me in the application of X-ray diffraction methods to fibrous proteins.

Additions and Corrections

Volume 1, 1968

F. H. Westheimer: Pseudo-Rotation in the Hydrolysis of Phosphate Esters.

Page 77. The sentence beginning on the next to last line should read as follows: Recently Frank and Usher[87] have found that the hydrolysis of **34** proceeds with production of methanol, whereas that of **33** proceeds with formation of acetoin and dimethyl phosphate; they have explained these results by the pseudo-rotation hypothesis and the "preference rules"[7] here reviewed.

Hiroshi Tanida: Solvolysis Reactions of 7-Norbornenyl and Related Systems. Substituent Effects as a Diagnostic Probe for Participation.

Page 243. In the formula at the bottom of the right column, $R^+ \rightleftharpoons$ **37** should read $R^+ \leftrightharpoons$ **37**.

Page 244. In the upper chart, <00.2% should read <0.02%.