

**PROTEINS: A THEORETICAL PERSPECTIVE OF
DYNAMICS, STRUCTURE, AND THERMODYNAMICS**

ADVANCES IN CHEMICAL PHYSICS

VOLUME LXXI

CHAPTER I

INTRODUCTION

Proteins are one of the essential components of living systems. Along with nucleic acids, polysaccharides, and lipids, proteins constitute the macromolecules that have important roles in biology. Nucleic acids, in the form of DNA and RNA, store and distribute the genetic information as needed. Of particular importance is the information that determines the sequences of amino acids that characterize the proteins. Proteins contribute to the structure of an organism and execute most of the tasks required for it to function. Proteins even form part of the complex mechanism by which they are synthesized. Polysaccharides, linear and branched-chain polymers of sugars, provide structural elements, store energy, and when combined with peptides or proteins, play an important role in antigenicity and, more generally, in cellular recognition. Lipids, which include molecules such as fatty acids, phospholipids, and cholesterol, serve as energy sources and are the most important components of the membrane structures that organize and compartmentalize cellular function.

In this volume we concentrate on globular proteins, the biological macromolecules with the greatest functional range. It is for these systems that the relation of function to structure and dynamics is best understood. Most chemical transformations that occur in living systems are catalyzed by enzymes; the globular proteins that have evolved for executing such specific tasks. As well as enhancing the rates of reactions, sometimes by eight or more orders of magnitude, globular proteins (e.g., repressors) inhibit certain reactions (e.g., the transcription of DNA) involved in the mechanism for the control of growth and differentiation. A breakdown of these control mechanisms can lead to unobstructed growth and the development of cancer. Other proteins (such as hemoglobin) serve to transport small molecules (such as oxygen), electrons, and energy to the appropriate parts of the organism. Antibody molecules are proteins that protect the organism by specifically recognizing and binding to foreign antigenic substances (such as viruses). Many proteins have structural roles; e.g., fibrous tissue is composed mainly of the protein collagen, and the major functional components of muscle, actin and myosin, are proteins.

Because of this wide range of protein functions and the need to develop

specialized proteins for each of them, the number of different proteins in an organism can be very large. The well-studied bacterium *Escherichia coli* contains about 3000 different kinds of proteins. Since many of them occur in multiple copies, there are a total of about 1 million protein molecules in a single bacterium. In human beings there are estimated to be on the order of 10^5 to 10^6 different proteins.

For most globular proteins, the biological function includes an interaction with one or more small molecules (a ligand, hormone, substrate, coenzyme, chromophore, etc.) or another macromolecule. Whether reactive or nonreactive systems are being considered, there can be important conformational alterations in the molecule that is bound and concomitant changes in the structure of the macromolecule to which the binding occurs. Such concerted conformational changes are the essential element for activity in some cases; in others, they play a less significant role. In hormone-receptor binding, for example, the structural changes induced in the receptors are fundamental to the transmission of information. Correspondingly, the conformational transition induced by ligand binding in hemoglobin is an integral part of the cooperative mechanism. Further, in many systems, small motions have been observed (e.g., the differences between the ligated and unligated structure of ribonuclease A) that appear to be involved in the function of the protein. Thus any attempt to understand the details of the activity of proteins requires an investigation of the dynamics of the structural fluctuations and their relation to reactivity and conformational change.

In addition to their biological importance, globular proteins are intrinsically interesting systems from the viewpoint of physical chemistry. They are long-chain polymers, but unlike most polymers they have a well-defined average structure. This structure is aperiodic (the "aperiodic crystal" of Schrödinger)¹ in the sense that it does not have regular repeats. Since the structure is determined by weak, noncovalent, interactions among the elements of the polypeptide chain, large fluctuations are expected. For a complete description of proteins, it is important, therefore, to know, in addition to the average structure, the form of the fluctuations that occur, to determine how they take place, and to evaluate their magnitudes and time scales.

Historically, hydrogen exchange experiments (i.e., the replacement of one isotope of hydrogen bound to an O, N, or S atom in the protein interior by another isotope from the solvent water) provided some of the earliest evidence for the existence of conformational fluctuations in proteins. More recently, a wide range of experimental methods (such as fluorescence quenching and depolarization, nuclear magnetic resonance relaxation, infrared and Raman spectroscopy, and X-ray and inelastic neutron scattering) have been used to study the motions in proteins. However, it is primarily the application of theoretical methods, particularly molecular dynamics simulations, that have

brought about a conceptual change in the pervading view concerning the nature of proteins.

Although to chemists and physicists it is self-evident that polymers such as proteins undergo significant fluctuations at room temperature, the classic view of such molecules in their native state had been static in character. This followed from the dominant role of high-resolution X-ray crystallography in providing structural information for these complex systems. The remarkable detail evident in crystal structures led to an image of biomolecules with every atom fixed in place. Tanford suggested that as a result of packing considerations "the structure of proteins must be quite rigid."² D. C. Phillips, who determined the first enzyme crystal structure, has written: "The period 1965-75 may be described as the decade of the rigid macromolecule. Brass models of DNA and a variety of proteins dominated the scene and much of the thinking."³ Molecular dynamics simulations have been instrumental in changing the static view of the structure of biomolecules to a dynamic picture. It is now recognized that the atoms of which biopolymers are composed are in a state of constant motion at ordinary temperatures. The X-ray structure of a protein provides the average atomic positions, but the atoms exhibit fluidlike motions of sizable amplitudes about these averages. Crystallographers have acceded to this viewpoint and have come so far as sometimes to emphasize the parts of a molecule they do not see in a crystal structure as evidence of motion or disorder. The new understanding of protein dynamics subsumes the static picture.

Knowledge of the average atomic positions allows discussion of many aspects of biomolecule function in the language of structural chemistry. However, recognition of the importance of fluctuations opens the way for more sophisticated and accurate interpretations of protein activity.

Simulations of proteins, as of many other systems (e.g., liquids), can, in principle, provide the ultimate details of motional phenomena. The primary limitation of simulation methods is that they are approximate. It is here that experiment plays an essential role in validating the simulation methods; that is, comparisons with experimental data serve to test the accuracy of the calculated results and provide criteria for improving the methodology. However, the experimental approaches to biomolecular dynamics are limited as to the information that can be obtained from them; e.g., if one is concerned with the time scale of motions, the frequency spectrum covered by experiments such as nuclear magnetic resonance (NMR) is incomplete, so that motional models that are able to rationalize the data can be inaccurate. When experimental comparisons indicate that the simulations are meaningful, their capacity for providing detailed results often makes it possible to examine specific aspects of the atomic motions far more easily than by making measurements. However, at the present stage of development, possible inaccuracies in the simulations must be kept in mind in evaluating and applying the results.

The present volume deals primarily with theoretical approaches to protein dynamics and thermodynamics. This rapidly developing field of study is founded on efforts to supplement our understanding of protein structure with concepts and techniques from modern chemical theory, including reaction dynamics and quantum and statistical mechanics. From a knowledge of the potential energy surface for a protein, the forces on the component atoms can be calculated and used to determine the phase space trajectory for the molecule at a given temperature. Such molecular dynamics simulations, which have successfully been applied to gases and liquids containing a large number of atoms, provide information concerning the thermodynamic properties and the time dependence of processes in the system of interest. More generally, statistical mechanical techniques are being used widely to characterize molecular motions and chemical reactions in condensed phases. The application of these methods to protein molecules is natural in that proteins contain many atoms, are densely packed, and function typically in a liquid environment.

Before focusing on the dynamical studies of biomolecules, it is useful to place this new field in perspective relative to the more general development of molecular dynamics. Molecular dynamics has followed two pathways which come together in the study of biomolecule dynamics. One of these, usually referred to as trajectory calculations, has an ancient history that goes back to two-body scattering problems for which analytic solutions can be achieved. However, even for only three particles with realistic interactions, difficulties arise. An example is provided by the simplest chemical reaction, $H + H_2 \rightarrow H_2 + H$, for which a prototype calculation was attempted by Hirschfelder, Eyring, and Topley in 1936.⁴ They were able to calculate a few steps along one trajectory. It was nearly 30 years later that the availability of computers made it possible to complete the calculation.⁵ Much has been done since then in applying classical trajectory methods to a wide range of chemical reactions.⁵⁻⁷ These classical studies have been supplemented by semiclassical and quantum-mechanical calculations in areas where quantum effects can play an important role.^{7,8} The focus at present is on more complex molecules, the redistribution of their internal energy, and the effect of this on reactivity.⁹

The other pathway in molecular dynamics has been concerned with physical rather than chemical interactions (in analogy to physisorption versus chemisorption) and with the thermodynamic and dynamic properties of a large number of particles, rather than detailed trajectories of a few particles. Although the basic ideas go back to van der Waals and Boltzmann, the modern era began with the work of Alder and Wainright on hard-sphere liquids in the late 1950s.¹⁰ The paper by Rahman¹¹ in 1964, on a molecular dynamics simulation of liquid argon with a soft sphere (Lennard-Jones) potential represented an important next step. Simulations of complex liquids followed; the now classic study of liquid water by Stillinger and Rahman was published in

1974.¹² Since then there have been many studies on the equilibrium and non-equilibrium behavior of a wide range of systems.^{13,14}

This background set the stage for the development of molecular dynamics of biomolecules. The size of an individual molecule, composed of 500 or more atoms for even a small protein, is such that its simulation in isolation can serve to obtain approximate equilibrium properties, as in the molecular dynamics of fluids, although detailed aspects of the atomic motions are of considerable interest, as in trajectory calculations. A basic assumption in initiating such studies was that potential functions could be constructed which were sufficiently accurate to give meaningful results for systems as complex as proteins or nucleic acids. In addition, it was necessary to assume that for these inhomogeneous systems, in contrast to the homogeneous character of even "complex" liquids such as water, simulations of an attainable time scale (10 ps in the initial studies) could provide a useful sample of the phase space in the neighborhood of the native structure. For neither of these assumptions was there strong supporting evidence in 1975. Nevertheless, the techniques of molecular dynamics were employed with the available potential functions in the first simulation of the internal atomic motions of a protein, the bovine pancreatic trypsin inhibitor (BPTI),¹⁵ which has played the role of the "hydrogen molecule" of protein dynamics.

In this volume we summarize first the elements of protein structure and provide a brief overview of the internal motions of proteins, their relation to the structural elements, and their functional role. We then outline the theoretical methods that are being used to study motional phenomena and thermodynamics. A description is given of the potential functions that determine the important interactions, and the various approaches that can be used to study the dynamics are outlined. Since the motions of interest involve times from femtoseconds to seconds or longer, a range of dynamical methods is required.

An important consideration in protein dynamics is the influence of solvents such as water on the functional integrity and structural stability of the biomolecular system. This influence is manifested in a variety of different phenomena, ranging from marked solvent effects on the rate of oxygen uptake in myoglobin to the stabilization of oppositely charged sidechain pairs on the surface of proteins. Although experimental data on protein-solvent interactions are being accumulated, our understanding of the structural, dynamic, and thermodynamic effects of water on biological systems is still incomplete. Some of the newer developments in the theory of aqueous solutions are described and it is shown how they can help to provide a fundamental understanding of solvated proteins.

Studies of the dynamics are of utility for determining thermodynamic properties as well as for providing information concerning the motions them-

selves. Of special interest is an understanding of the stability of proteins and the thermodynamics of their interactions with drugs and ligands. Theoretical methods are described for determining the free energies involved. Since the phenomena occur in the liquid state or some other condensed phase, it is necessary to be able to include the effect of solvent in going from the microscopic interactions to the macroscopic enthalpies, entropies, and free energies that are the experimental thermodynamic variables of interest. Such information, when augmented by the results of special techniques for the study of chemical reactions, leads naturally to an analysis of the reaction dynamics involving macromolecules.

The main body of this volume presents results that have been obtained in dynamical studies of proteins in vacuum, in solution, and in crystals. Because of the intense activity in this area, a selection has been made to provide a representative and coherent view of our present knowledge. Where possible, comparisons with experiment and the functional correlates of the motions are stressed. A description is given of specific experimental areas that are of particular importance for the analysis of dynamics or where the simulation results are providing information essential for the interpretation of the experimental data. We conclude with an outlook for future developments and applications in this exciting field.

A number of reviews on related material in protein dynamics have appeared. For reviews concerned primarily with theoretical work, the reader may read Careri et al.,^{16,17} Cooper,^{18,19} Weber,²⁰ Karplus,²¹ Karplus and McCammon,²² Levitt,²³ Levy,^{23a} McCammon and Karplus,^{24,25} McCammon, Pettitt and Karplus,^{26a} van Gunsteren and Berendsen,²⁷ and Welch et al.²⁸ Experimental work is reviewed in Campbell et al.,²⁹ Cusack,^{29a} Debrunner and Frauenfelder,³⁰ Dobson and Karplus,^{30a} Englander and Kallenbach,³¹ Gurd and Rothgeb,³² Karplus and McCammon,³³ Jardetsky,³⁴ Bennett and Huber,³⁵ Peticolas,³⁶ Ringe and Petsko,³⁷ Torchia,^{37a} Williams,^{38,39} Wagner and Wüthrich,^{39a} and Woodward and Hilton.⁴⁰ In addition, several volumes reporting on meetings devoted to protein dynamics have been published,^{41,42} as has an article in *Scientific American*.⁴³ An introductory description of the dynamics of proteins and nucleic acid has been presented by McCammon and Harvey.⁴⁴

CHAPTER II

PROTEIN STRUCTURE AND DYNAMICS—AN OVERVIEW

Since the function and dynamics of proteins are intimately related to their structure, we first provide a short survey of the structural elements of proteins. We then outline briefly the present state of our knowledge of protein dynamics and the role of the internal motions in protein function.

A. THE STRUCTURE OF PROTEINS

Much of what we now know about the structure of globular proteins comes from X-ray crystallographic studies.⁴⁵⁻⁴⁸ The first high-resolution structures of proteins were those of myoglobin and hemoglobin determined in the early 1960s by J. Kendrew and coworkers⁴⁹ and M. F. Perutz,⁵⁰ respectively, and the first structure of an enzyme, lysozyme, was reported by D. C. Phillips and coworkers in 1965.⁵¹ Since then a large number of protein structures have been determined. A recent listing (1987) of the Protein Data Bank at Brookhaven National Laboratories (where the results of many but not all protein structure determinations are deposited)⁵² includes the coordinates for about 200 different proteins. These results have made possible an analysis of the "anatomy" of protein structures.⁵³ We show in Fig. 1 a schematic view of the structures of several proteins that illustrate some of the structural motifs that have been found.

Each protein consists of a polypeptide chain that is made up of residues or amino acids linked together by peptide bonds. The polypeptide chain backbone, a portion of which is shown in Fig. 2, is composed of repeating units that are identical, except for the chain termini. Proteins vary widely in size, from 50 to 500 or so residues, corresponding to 1000 to 10,000 or so atoms. Approximately half of the atoms are hydrogens, which are not seen except in very high resolution X-ray and in neutron crystal structures; thus, most of the descriptions of proteins focus on the positions of the "heavy" atoms, C, N, O, and S. What distinguishes different proteins, other than the number of amino acids, is the sequence of amino acids in the polypeptide chain. There are 20

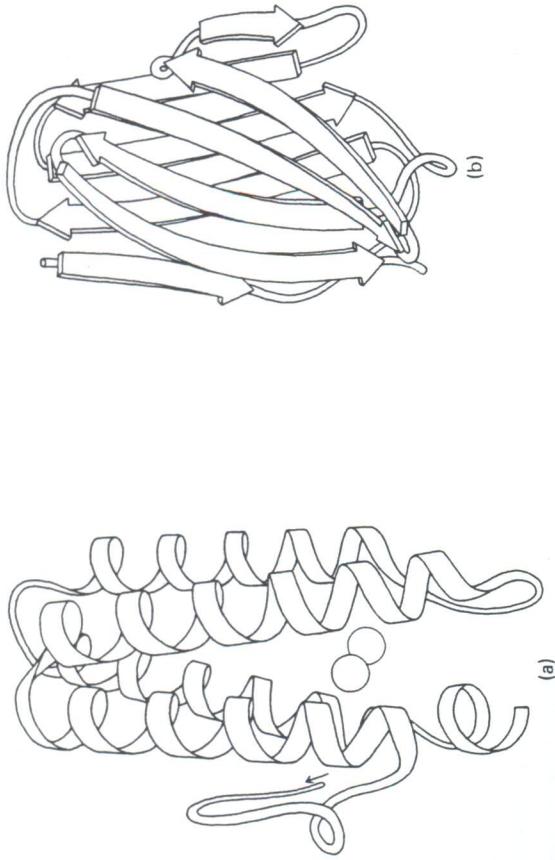
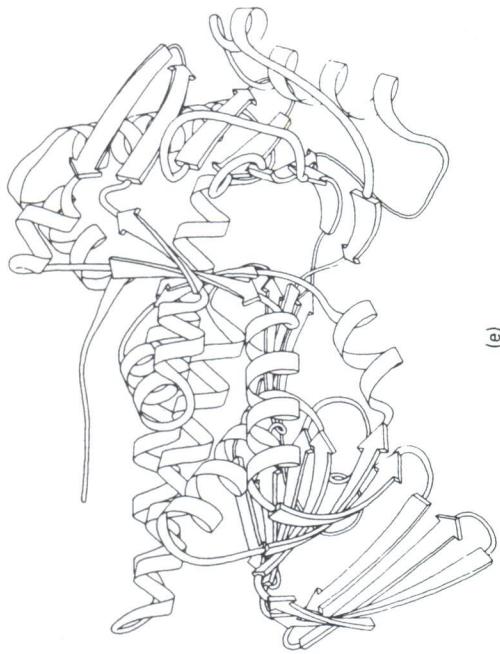


Figure 1. Schematic representations of protein structures: (a) myohemerythrin, an α -helical protein with antiparallel helices; (b) V₂ domain of an immunoglobulin, a β -sheet protein; (c) triose phosphate isomerase, a parallel $\alpha\beta$ protein with a central "beta barrel"; (d) carboxypeptidase, a parallel $\alpha\beta$ protein with a central β -sheet structure; (e) one domain of a complex protein structure with more than one domain. (From Ref. 53; courtesy of J. Richardson.)



(a)

(b)

(c)

(d)

(e)

commonly occurring amino acids that differ in their sidechains; they vary from the simplest, glycine, to the most complex, tryptophan (Fig. 3). It is the sequence of amino acids, referred to as the primary structure of the protein, that determines the native conformation, the structure that is stable under physiological conditions. The first protein amino acid sequence was determined by F. Sanger and coworker in 1953 for insulin.⁵⁴ It is generally believed that the native structure corresponds to a free-energy minimum, although there is no direct experimental or theoretical evidence for this. One suggestive result is that it is possible to denature (unfold) many proteins in

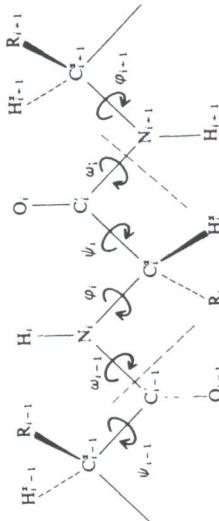


Figure 2. Polypeptide backbone with standard notation for mainchain atoms; the flexible dihedral angles ϕ and ψ and the more rigid, partially conjugated peptide bond angle ω are shown. The sidechains are indicated as R. [Adapted from C. R. Cantor and P. R. Schimmel, *Bioophysical Chemistry* (W. H. Freeman and Co., San Francisco, 1980).]

solution by increasing the temperature or lowering the pH and then to recover the native protein by returning the solution to normal values of temperature or pH.⁵⁵

From the analysis of many protein structures, it is found that the folding of portions of the polypeptide chain often has certain regularities, called elements of secondary structure. These can be defined in terms of the main-chain dihedral angles ϕ and ψ (see Fig. 2); it is not necessary to delimit the angle ω associated with the partially conjugated peptide bond because it is generally in the neighborhood of $180 \pm 5^\circ$. The most important secondary structural elements are the α -helix (Fig. 4a) and β -pleated sheets (Fig. 4b), both of which are regularly repeating structures with backbone hydrogen bonds that were predicted by Pauling and Corey⁵⁶ in the 1950s, prior to the first protein structure determination. The α -helix is a compact rodlike structure with 3.6 amino acids per turn, a rise of only 1.5 \AA per turn, and a $\text{C}=\text{O} \cdots \text{H}-\text{N}$ hydrogen bond between residues i and $(i+4)$. Wool (α -keratin) has an α -helix as its essential constituent and because it is so compact leads to the well-known extensibility of that fiber. Other helical structures [e.g., the 3_{10} helix with 3 residues per turn and an i -to- $(i+3)$ hydrogen bond] also occur in proteins. The β -pleated sheet structure is an extended structure with a displacement of approximately 3.47 \AA per residue (Fig. 4b). It can be regarded as a (degenerate) twofold helix in which the hydrogen bonds are between strands rather than within a strand, as for the α -helix. Pleated sheets can be formed with parallel or antiparallel orientations of adjacent strands. Silk, which is formed from antiparallel β -pleated sheets, is a very strong but rigid fiber because the strands are already extended to near their maximum length. Since globular proteins have a finite size with radii of gyration from 15 to 80 \AA or so, the secondary structural elements, such as α -helices and β -sheets, are limited in length. They often terminate in so-called turns, which have also been shown to have regular features that can be classified into a number of types (Fig. 4c).

When a large number of protein structures are examined, it is found that on the average, 25% of the amino acids are in helices, 25% in sheets, 25% in turns, and the remaining 25% in what are called random coil segments that have no simple regularity in their mainchain dihedral angles. A given protein structure may deviate widely from this set of averages, however; e.g., myoglobin has 85% of the residues in the α -helical configuration and the remainder in turns or random-coil sequences.

The overall spatial arrangements of the amino acid residues in proteins are referred to as the tertiary structure. In many cases this can be described approximately in terms of the packing together of secondary structural elements. Various motifs have been identified; these include the helix-turn-helix structure and the β strand- α helix- β strand structure, as well as others (see

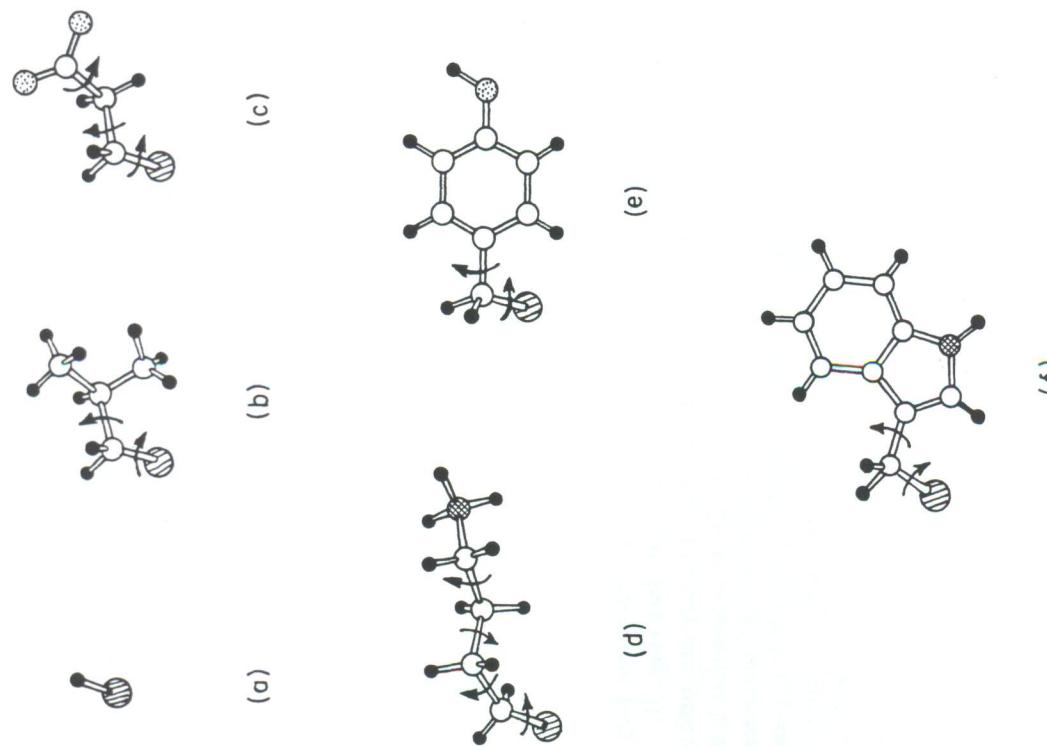
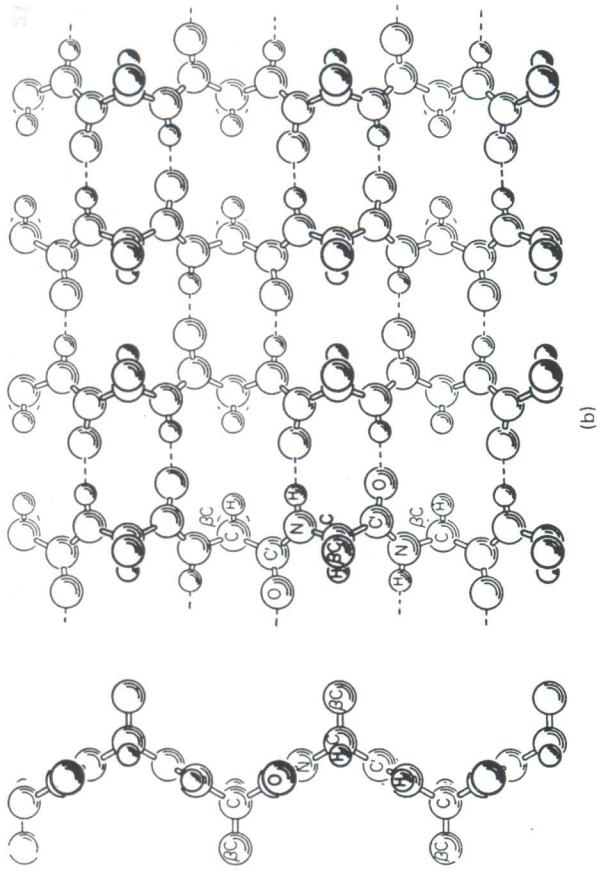
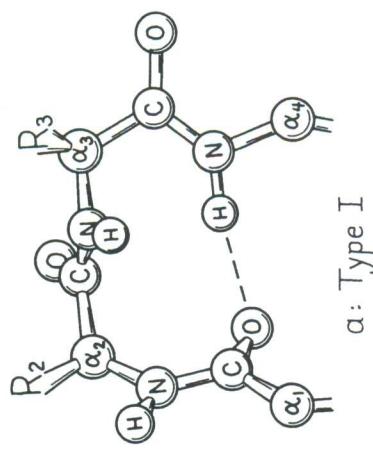
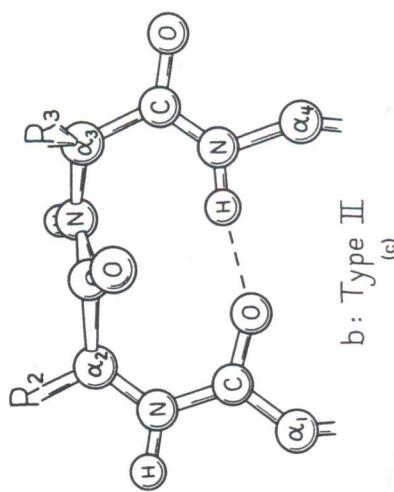
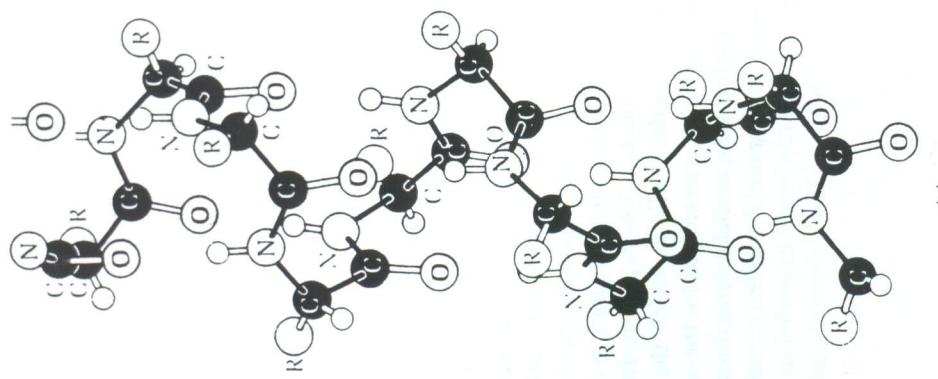


Figure 3. Some of the commonly occurring sidechains: (a) glycine (Gly, G); (b) leucine (Leu, L); (c) glutamic acid (Glu, E); (d) lysine (Lys, K); (e) tyrosine (Tyr, Y); (f) tryptophan (Trp, W). All atoms, including hydrogens (black dots), are shown; the α -carbons are indicated by lines, oxygens by dots, and nitrogens by crosshatching in the corresponding circles; open circles are other carbon atoms; and the flexible dihedral angles are indicated by arrows with the one nearest to the α -carbon called χ^1 , the next χ^2 , and so on.

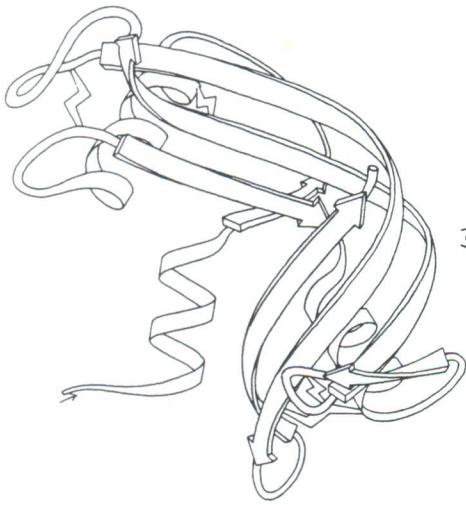


(b)

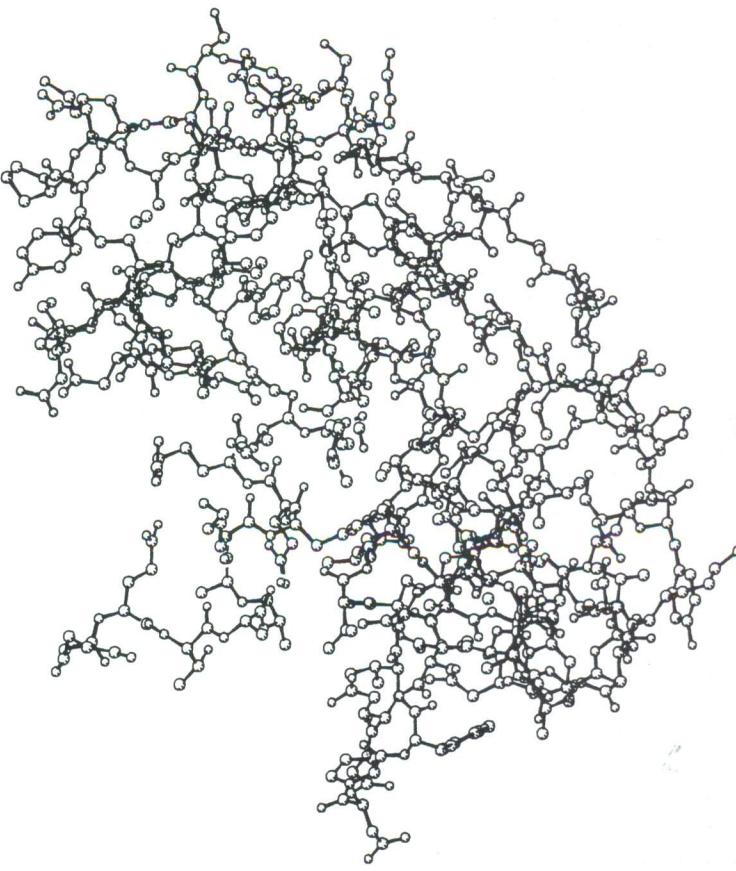
 α : Type I b : Type II

(a)

Figure 4. Protein secondary structural elements: (a) right-handed α -helix showing intra-chain hydrogen bonds as dotted lines ($\alpha_i, \phi \approx -60^\circ, \psi \approx -60^\circ$); (b) antiparallel β -pleated sheet showing interchain hydrogen bonds as dashed lines ($\beta_A: \psi \approx -120^\circ, \psi \approx 120^\circ$); (c) β -turns of types I and II, differing in the orientation of the central peptide group. [Part (a) is adapted from A. L. Lehninger, *Biochemistry* (Worth Publishers, Inc., New York, 1975); (b) from Ref. 8; and (c) from Ref. 53.]



(a)



(b)

Fig. 1). However, there also occur regions in proteins involving random-coil segments (e.g., some proteins have essentially no identifiable secondary structure) that require a more complex description. In such cases it may not be possible to do more than simply give a list of the mainchain dihedral angles (ϕ_i, ψ_i for all residues i).

Even in the absence of secondary structure, there are regularities in the amino acid distributions and their packing. In general, proteins are tightly packed systems (approximately the packing density of close-packed spherical atoms) with only a few voids (Fig. 5). Nonpolar sidechains of amino acids tend to be in the interior of the protein, often with clusters of aromatic and other nonpolar residues forming a stabilizing core. Most charged sidechains are on the surface of the protein, with clusters of such amino acids often concentrated in the active site. Polar residues, as well as the carbonyl and amide groups of the polypeptide chain, tend to be more uniformly distributed, with essentially all hydrogen-bond donors or acceptors located so that they form hydrogen bonds either with other parts of the protein or with the surrounding solvent.

In some cases proteins are divided into two or more domains (Fig. 1), each of which is like a globular protein but connected covalently to other domain(s) by the continuous polypeptide chain. Other proteins are oligomeric in that they are composed of several unconnected polypeptide chains (subunits) that usually, but not always, fold up independently and assemble to form the complete protein. The arrangement of the subunits relative to each other is referred to as the quaternary structure. Hemoglobin ($\alpha_2\beta_2$) (Fig. 6) and aspartate transcarbamoylase ($\alpha_6\beta_6$), where α and β refer to different types of subunits, are well-studied cases where different quaternary structures occur with significantly altered properties.

B. OVERVIEW OF PROTEIN MOTIONS

The general motional characteristics of globular proteins follow directly from their structural properties. The polypeptide chain of the protein has strong covalent bonding forces along the chain but relatively weak, noncovalent interactions between different parts of the chain that are packed together in the native structure; a few of the noncovalent interactions involve charged groups that form "salt links" whose interaction energy can approach that of a covalent bond. The only covalent interactions between different parts of the chain in globular proteins are disulfide bonds formed by oxidation of pairs of cystine sidechains, of which there are usually no more than 1 for every 20 or so residues. Fibrous proteins, such as elastin, have cross-links involving sidechains such as lysine.

The polypeptide chain of a protein has single bonds that permit internal torsional rotation to take place. This is true for the ϕ and ψ angles of each

Figure 5. Ribonuclease A: (a) schematic diagram; (b) diagram showing heavy atoms with small radii; (c) diagram showing all heavy atoms with van der Waals radii. [Part (a) was prepared by J. Richardson; (b) and (c) were prepared by A. Brünger from coordinates supplied by G. Petsko.]

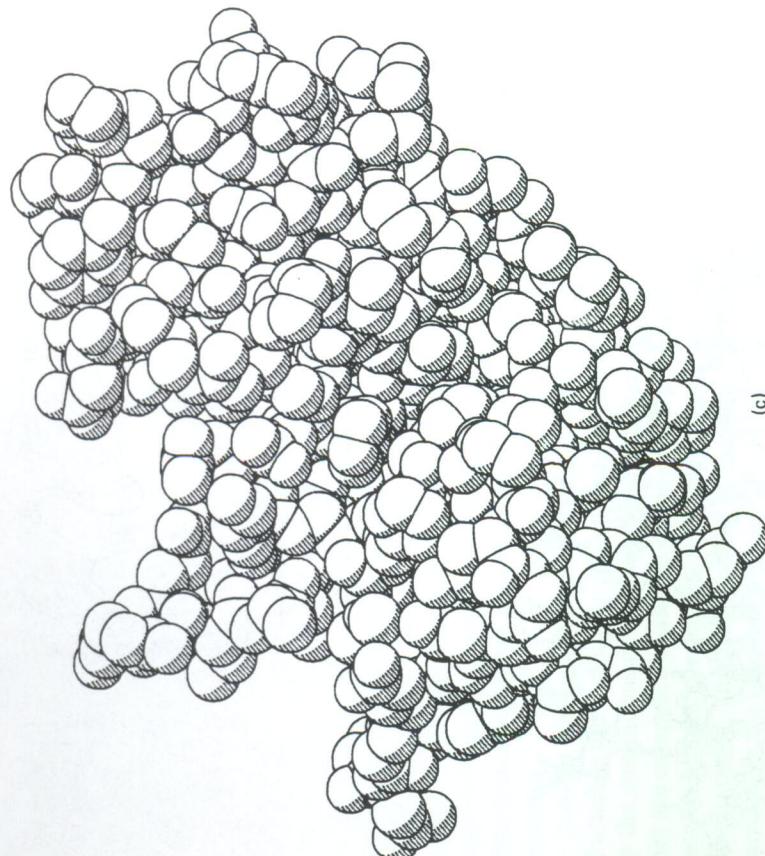


Figure 5. (Continued)

amino acid, with only the peptide groups torsional angle ω being relatively rigid with respect to twisting (Fig. 2). Also, all of the sidechains, except glycine, have one or more single bonds about which internal rotation can occur (Fig. 3).

At any given time, a typical protein exhibits a wide variety of motions; they range from irregular elastic deformations of the entire protein driven by collisions with solvent molecules to seemingly chaotic librations of interior groups driven by collisions with neighboring atoms in the protein. Considering only typical motions at physiological temperatures, the smallest effective dynamical units in proteins are those that behave nearly as rigid bodies because of their covalent bonding. Examples include the phenyl group in the sidechain of tyrosine (Fig. 3), the isopropyl group in the sidechains of valine or leucine (Fig. 3), and the amide groups of the protein backbone (Fig. 2). Except for the methyl rotations in the isopropyl group, these units display only relatively small internal motion, owing to the high energy cost associated with deformation.

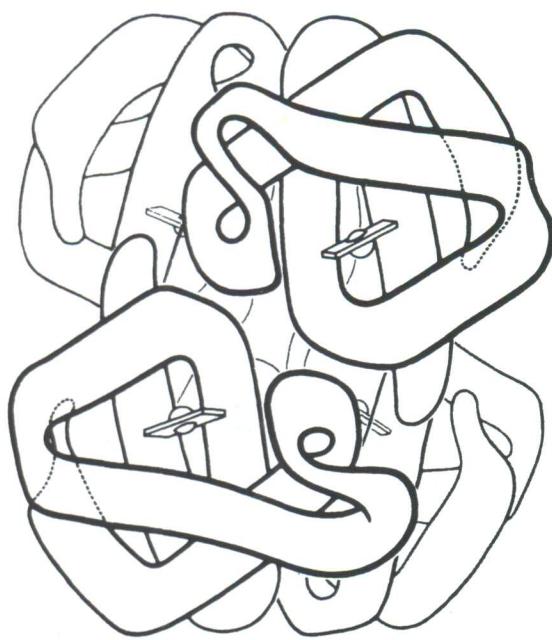


Figure 6. Diagram of the quaternary structure of the hemoglobin tetramer showing the four primarily helical subunits and the heme group in each. (Adapted from Ref. 263.)

tions of bond lengths, bond angles, or dihedral angles about multiple bonds. The important motions in proteins involve relative displacements of such groups associated with torsional oscillations about the rotationally permissive single bonds that link the groups together. High-frequency vibrations do occur within the local groups, but these are not of primary importance in the relative displacements.

Most groups in a protein are tightly engaged by atoms of the protein or of the surrounding solvent. At very short times ($\leq 10^{-12}$ s), the groups may display rattling motions in their cages, but such motions are of relatively small amplitude (≤ 0.2 Å). More substantial displacements of the groups occur over longer time intervals; these motions involve concomitant displacements of the cage atoms. Broadly speaking, such “collective” motions may have either a local or a rigid-body character. The former involves changes of the cage structure and relative displacements of neighboring groups, while the latter involves relative displacements of different regions of the protein but only small changes on a local scale.

The presence of such motional freedom implies that a native protein at room temperature samples a range of conformations. Most are in the general neighborhood of the average structure, but at any given moment an individual protein molecule is likely to differ significantly from the average struc-

ture. This in no way implies that the X-ray structure, which corresponds to the average in the crystal, is not important. Rather, it suggests that fluctuations about the average can play a significant role in protein function. In a protein, as in any polymeric system in which rigidity is not supplied by covalent cross-links, relatively large-scale fluctuations cannot be avoided. Thus, it is possible that they have a functional role as a consequence of evolutionary development.

Although the existence of fluctuations is now well established, our understanding of their biological role in most areas is incomplete. Both conformational and energy fluctuations with local to global character are expected to be important. In a protein, as in other nonrigid condensed systems, structural changes arise from correlated fluctuations. Perturbations, such as ligand binding, that produce tertiary or quaternary alterations, do so by introducing forces that bias the fluctuations in such a way that the protein makes a transition from one structure to another. Alternatively, the fluctuations can be regarded as searching out the path or paths along which transitions take place. In considering the internal motions of proteins, one must separate the dynamic from the thermodynamic elements; in the latter, the presence of flexibility is important (e.g., entropy of binding), while in the former the directionality and time scales play a role. Another way of categorizing the two aspects is that in thermodynamics, the equilibrium behavior is the sole concern, while in dynamics, the displacements from the average structure are the essential element. In certain cases, some features of the dynamics may be unimportant because they proceed on a time scale that is faster than the phenomenon of interest. An example might be the fast local relaxation of atoms involved in a much slower hinge-bending motion experienced in an enzyme active site; here only the time scale of the latter would be expected to be involved in determining the important rate process (e.g., product release), although the nature of the former would be of considerable interest. In other situations, the detailed aspects of the atomic fluctuations are a significant factor. This may be the case in the oxygen transport protein myoglobin, where local sidechain motions appear to be essential for the entrance and exit of ligands.

To summarize the available results concerning the dynamics of proteins and their functional role, we present in Table I some examples of the different categories of internal motions that have been identified. They cover a wide range of amplitudes (0.01 to 100 Å), energies (0.1 to 100 kcal/mol) and time scales (10^{-15} to 10^3 s). One expects an increase in one quantity (e.g., the amplitude of the fluctuations) to correspond to an increase in the others (e.g., a larger energy and longer time scale). This is often true, but not always. Some motions are slow because they are intrinsically complex, involving the correlated displacements of many atoms. An example might be partial-to-total un-

TABLE I
Internal Motions of Globular Proteins

I. Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)
(a) Atomic Fluctuations
1. Small displacements required for substrate binding (many enzymes) 2. Flexibility required for "rigid-body" motion (lysozyme, liver alcohol dehydrogenase, L-arabinose binding protein) 3. Energy "source" for barrier crossing and other activated processes 4. Entropy "source" for ligand binding and structural changes
(b) Sidechain Motions
1. Opening pathways for ligand to enter and exit (myoglobin) 2. Closing active site (carboxypeptidase)
(c) Loop Motions
1. Disorder-to-order transition covering active site (triose phosphate isomerase, penicilllopepsin) 2. Rearrangement as part of rigid-body motion (liver alcohol dehydrogenase) 3. Disorder-to-order transition as part of enzyme activation (trypsinogen-trypsin) 4. Disorder-to-order transition as part of virus formation (tobacco mosaic virus, tomato bush stunt virus)
(d) Terminal Arm Motion
1. Specificity of binding (λ -repressor-operator interaction)
II. Rigid-Body Motions (1 to 10 Å, 10^{-9} to 1 s)
(a) Helix Motions
1. Induction of larger-scale structural change (insulin) 2. Transitions between substrates (myoglobin)
(b) Domains (hinge-bending) Motions
1. Opening and closing of active-site region (hexokinase, liver alcohol dehydrogenase, L-arabinose binding protein) 2. Increasing binding range of antigens (antibodies)
(c) Subunit Motions
1. Allosteric transitions that control binding and activity (hemoglobin, aspartate transcarbamoylase)
III. Larger-Scale Motions (>5 Å, 10^{-7} to 10^4 s)
(a) Helix-coil transition
1. Activation of hormones (glucagon) 2. Protein folding transition

TABLE I—Continued.

(b) Dissociation/Association and Coupled Structural Changes
1. Formation of viruses (tomato bushy stunt virus, tobacco mosaic virus)
2. Activation of cell fusion protein (hemagglutinin)
(c) Opening and Distortional Fluctuations
1. Binding and activity (calcium-binding proteins)
(d) Folding and Unfolding Transition
1. Synthesis and degradation of proteins

folding transitions, in which the correlation of amplitude, energy, and time scale is expected to be approximately valid. However, in more localized events, often involving small displacements of a few atoms, the motion is slow because of a high activation barrier; examples are the aromatic ring flips in certain proteins. In this case the macroscopic rate can be very slow ($k \sim 1 \text{ s}^{-1}$ at 300°K), not because an individual event is slow (a ring flip occurs in $\sim 10^{-12} \text{ s}$), but because the probability that a ring has sufficient energy to get over an activation barrier that is the order of 16 kcal/mol is very small ($\sim 10^{-12}$).

In Table I we list various types of motions with their length and time scales and give specific examples as to where such motions are found to have functional roles. It can be seen that for all three of the somewhat arbitrary classes (local, rigid body, and larger-scale motions) the listed time scales vary over many orders of magnitude. This is due primarily, as already suggested, to the presence of activation barriers that can slow down even the simplest motion (e.g., atomic fluctuations in a double-well potential). What the table makes clear is that a great range of motional phenomena are found to have a functional role. In some cases (e.g., the atomic fluctuations required for larger “rigid-body” displacements, the sidechain oscillations that play a role in the entrance and exit of ligands in myoglobin, the allosteric transition in hemoglobin, the disorder-order transition in going from inactive trypsinogen to the active enzyme trypsin) there exist detailed theoretical and/or experimental studies of the motions involved. In many other cases, the role of the motion has been inferred only from structural studies that show two or more different conformations.

The richness of the motional phenomena that are involved in protein function, which is only hinted at in Table I, makes the field of macromolecular dynamics one of the most exciting and rapidly developing areas of chemical physics. It is our hope that the reader will come away from this volume with an understanding of the nature of protein motions, their functional role, and the methods used for studying them.