Zeitschrift: Helvetica Physica Acta
Band: 57 (1984)
Heft: 2

Artikel: From atoms to biomolecules
Autor: Frauenfelder, Hans
DOI: https://doi.org/10.5169/seals-115505

Helvetica Physica Acta, Vol. 57 (1984) 165–187 0018–0238/84/020165–23$1.50 + 0.20/0

From atoms to biomolecules

By Hans Frauenfelder, Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801 USA

(30. XI. 1983)

The steps leading from individual atoms to biomolecules show many of the characteristic features that belong to the theme of the present symposium "Atoms and Molecules: From the Individual Properties to the Collective Phenomena." In this survey I will sketch how biomolecules are built from atoms and describe some aspects of the physics of biomolecules. No better description of the beauty of the field has been given than the one Schrödinger wrote in 1943, long before the remarkable details of biomolecules began to emerge. Schrödinger calls biomolecules "aperiodic crystals." Comparing ordinary periodic crystals with biomolecules he said [1]: "Yet, compared with the aperiodic crystal, they are rather plain and dull. The difference in structure is of the same kind as that between an ordinary wallpaper in which the same pattern is repeated again and again in regular periodicity and a masterpiece of embroidery, say a Rafael tapestry, which shows no dull repetition, but an elaborate, coherent, meaningful design traced by the great master."

The study of biomolecules involves mathematics, physics, chemistry, biochemistry and biology and the borders between these fields become meaningless. The overlap of the fields presents some difficulties but, more important, makes the investigation of biomolecules challenging and exciting. Since the present survey is addressed to many different audiences, it is necessarily superficial. However, a number of excellent texts can be consulted for more complete treatments [2-6].

1. Biomolecules and life

The chain from atoms to organisms in Fig. 1 consists of a number of clearly distinguishable systems. The complexity increases with increasing number of atoms. At the present time it is impossible to predict the behavior of an organism starting from the individual properties of the atoms. An understanding can only be reached by breaking the chain in pieces and studying for instance how the properties of biomolecules depend on the properties of its building blocks. In the present survey, we discuss the construction of biomolecules from individual building blocks and some simple functions of biomolecules.

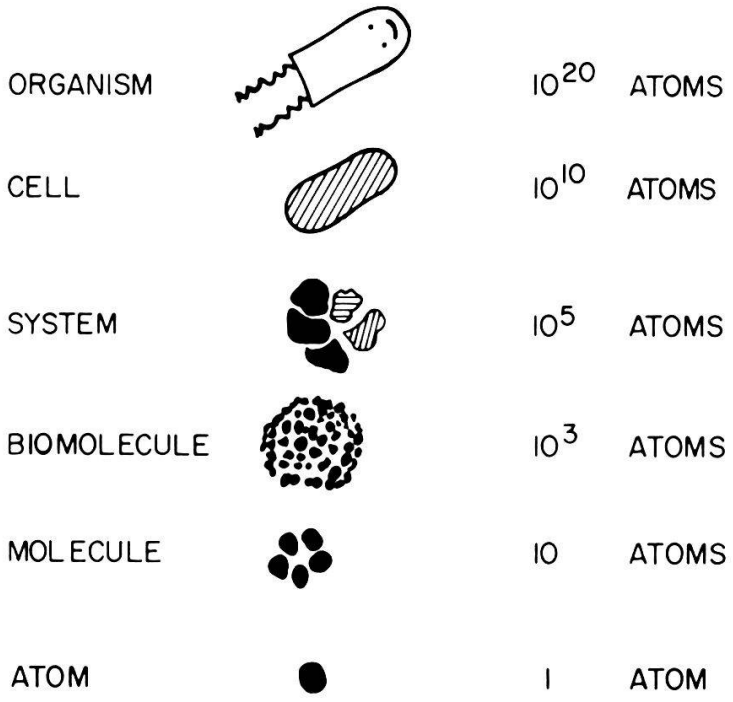

Figure 1
From atoms to organisms – a survey.

Two types of biomolecules are responsible for most biological phenomena, *nucleic acids* and *proteins*. We can call the first the legislative, the second the executive. The information is stored and transmitted by nucleic acids and they direct the construction of proteins as shown schematically in Fig. 2. The proteins perform the functions that are responsible for nearly all aspects of life; they are miniature machines.

The information is stored in the form of "three-letter words" on a very long linear unbranched DNA molecule (deoxiribonucleic acid). The DNA molecule is

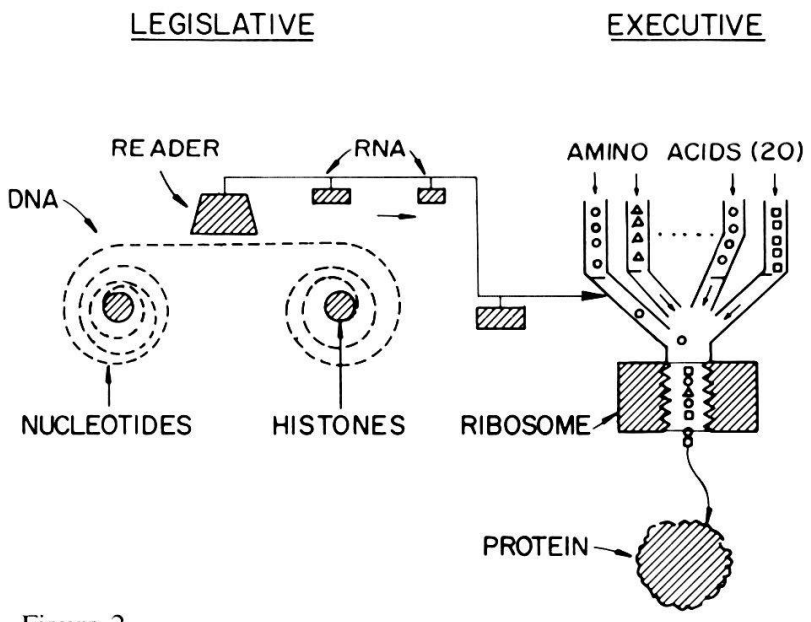

Figure 2
Biomolecules: Legislative and executive. Nucleic acids store and transport information and direct the assembly of proteins. Proteins, assembled from amino acids (AA), are the machines of life. The information is stored on DNA and transported by RNA.

wound around protein molecules (histones), probably for easier storage and access. The arrangement thus is similar to the storage of computer information on tapes that are wound on reels. The information on the DNA needed for the construction of a particular protein is read and transcribed onto a RNA molecule (ribonucleic acid) and transported to a ribosome, the factory where protein assymbly takes place. The protein is also built as a linear chain, but the building blocks of nucleic acids and proteins are different: Nucleic acids are built from four different nucleotides, proteins from twenty different amino acids. The RNA instructs the ribosome in which order the amino acids must be connected to form the primary sequence of the protein. The instruction involves a translation from the DNA and RNA language (three-letter words from an alphabet of four letters) to the protein language (twenty amino acids). When the primary chain emerges from the ribosome, it folds into the functionally active three-dimensional structure.

2. Nucleic acids and proteins

2.1. *The building blocks*

Nucleic acids, the information carriers, are built from four different building blocks, the *nucleotides*, which form the four letters of the genetic alphabet. A nucleotide consists of a base, a sugar, and one or more phosphate groups, as indicated schematically in Fig. 3a. Sugar and phosphate group are the same for all nucleotides and the four letters are distinguished by four different bases. In DNA, which stores the information, the bases are adenine (A), cytosine (C), guanine (G), and thymine (T). In RNA, which transfers information, thymine is replaced by uracil (U). Here we are not interested in the molecular structure of the four bases, but simply represent them by the four letters A, C, G, and U.

Proteins, the executive, are constructed from 20 building blocks, the *amino acids*. Each amino acid consists of a backbone and a side chain. The backbone is the same in all amino acids, but the side chains are different. One particular amino acid is shown in Fig. 3b. The twenty different amino acids are listed in Table I, together with the standard abbreviation and some of their properties.

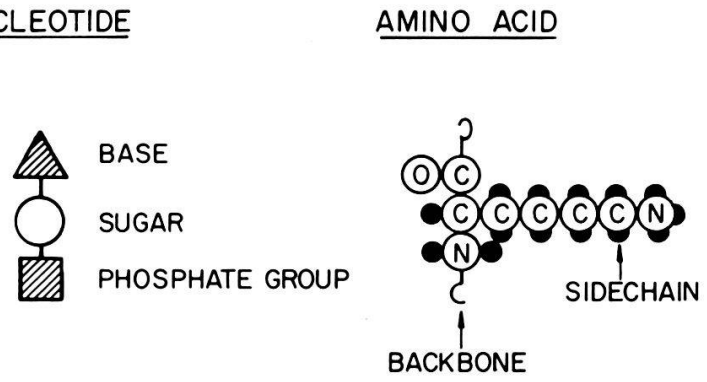

Figure 3
The building blocks of nucleic acids and proteins. The nucleic acids have four different bases, the amino acids 20 different side chains.

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

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

2.2. *The primary structure*

Nucleic acids and proteins are essentially linear systems as indicated in Fig. 4. The building blocks are linked by covalent bonds and the arrangement of the building blocks, the primary sequence, is crucial for the function.

2.3. *Language and translation*

The information is stored in DNA and transmitted by RNA in the form of three-letter words. Each three-letter word codes for a particular amino acid. The

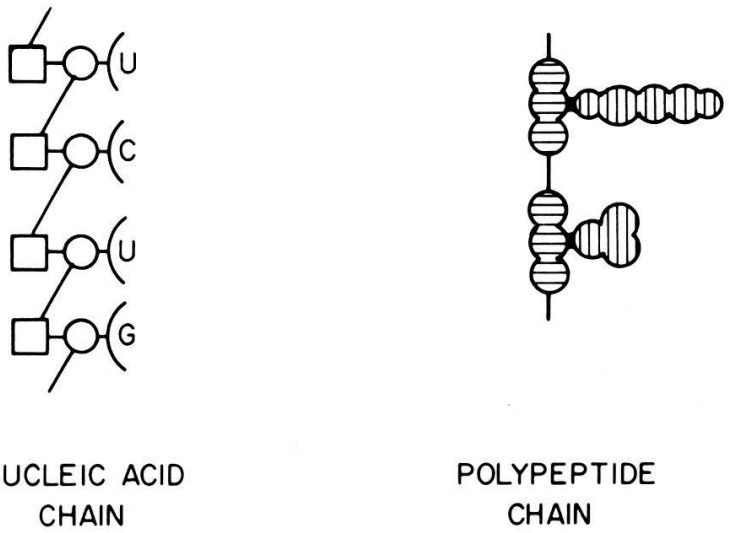

Figure 4
Nucleic acids and proteins are both linear chains; the individual building blocks are connected by covalent bonds.

translation, called the genetic code, is given in Table II. A particular sequence, for instance . .UCU–UAC–ACG . . . determines a unique primary sequence of amino acids, in this case . . Ser–His–Ala . . . . The genetic code is degenerate; most amino acids can be "called " by more than one word.

2.4. *Folding and tertiary structure*

When the linear chain comes off the ribosome, as indicated in Fig. 2, it folds into the tertiary structure. It is likely that folding involves intermediate steps (secondary structure) as shown in Fig. 5. The final three-dimensional structure is determined by the order of the amino acids in the primary sequence.

Table II
Genetic code

	Second position				
First	U	C	A	G	Third
	PHE	SER	TYR	CYS	U
	PHE	SER	TYR	CYS	C
U	LEU	SER	Stop	Stop	C
	LEU	SER	Stop	TRP	G
	LEU	PRO	HIS	ARG	U
	LEU	PRO	HIS	ARG	C
C	LEU	PRO	GLN	ARG	A
	LEU	PRO	GLN	ARG	G
	ILE	THR	ASN	SER	U
	ILE	THR	ASN	SER	C
A	ILE	THR	LYS	ARG	A
	MET	THR	LYS	ARG	G
	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
G	VAL	ALA	GLU	GLY	A
	VAL	ALA	GLU	GLY	G

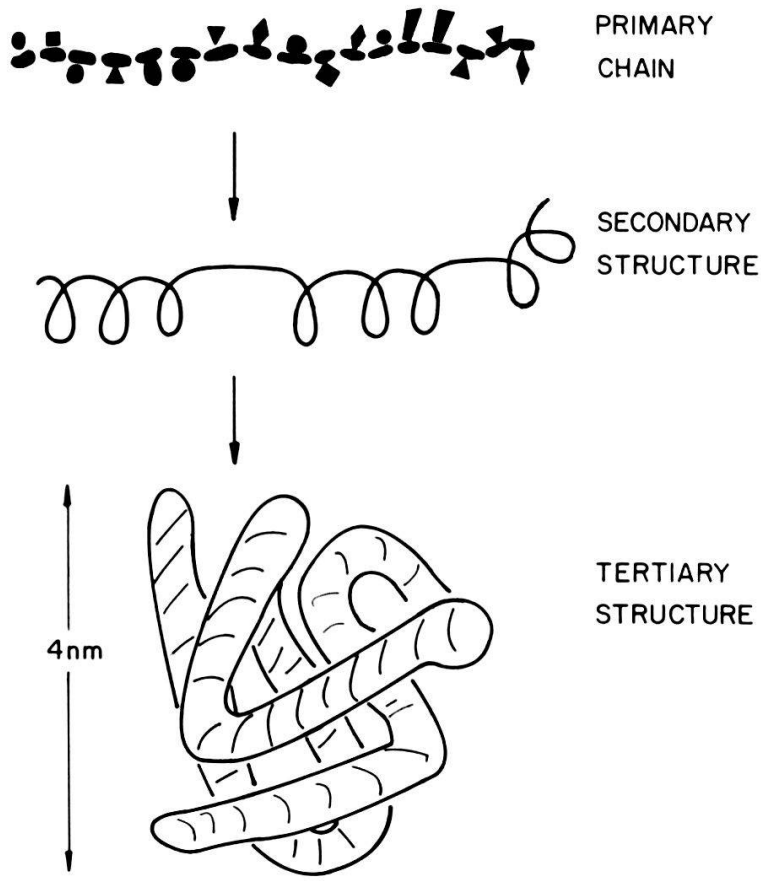

Figure 5
The linear polypeptide chain (primary sequence) folds into the final tertiary structure.

The folded protein is nearly closed-packed and it represents in a most beautiful way the aperiodic crystal described so eloquently by Schrödinger. Computer graphics permits an impressive representation of proteins and nucleic acids. The computer-produced pictures are based on structures determined by X-ray diffraction.

A particular protein usually performs one well defined function. Hemoglobin, for instance, transports oxygen in the blood. Two fundamental questions thus arise:

Given the primary sequence, what is the tertiary structure?

Given the tertiary structure, what is the function?

If these two questions are solved, a revolution in biology, medicine, and pharmacology will ensue; we will be able to custom-design a particular protein and perform feats that we can at present only dream about. We are, however, very far from a solution of these problems. Even the first question, which appeared a few years ago to be close to a computer solution, is probably much more difficult than it appears. Very small changes in protein structure can have major effects on function; a meaningful answer to the first question must place the atoms very close to the correct positions; otherwise the structure is not functional. Because the two fundamental questions are so difficult we must be more modest and begin with solvable problems. We can try to understand the connection between structure and function in the opposite direction: How is a particular known function performed?

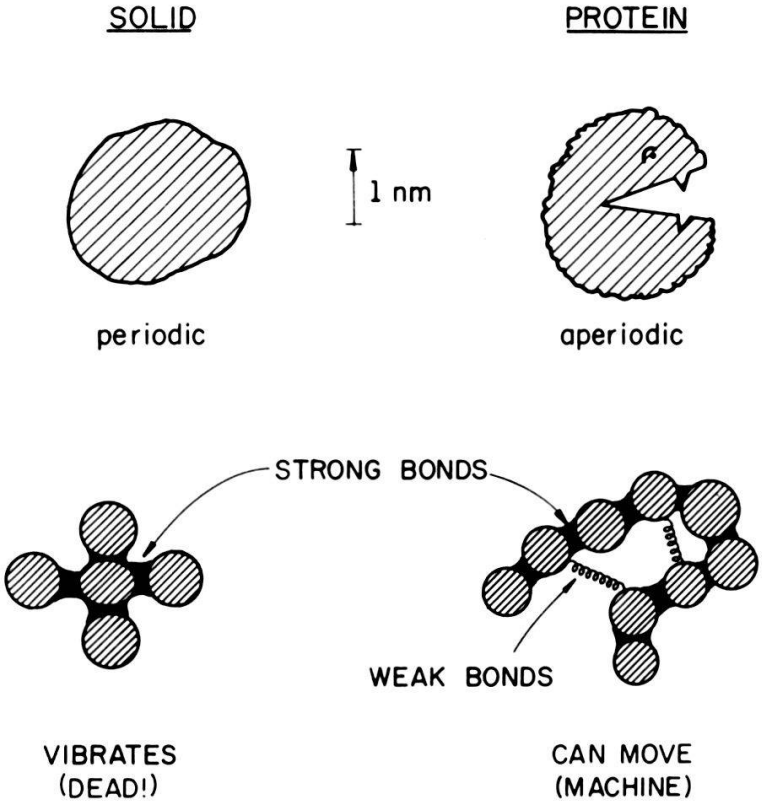

Figure 6
Solids and biomolecules. Solids are periodic, biomolecules aperiodic. In a solid, each atom is strongly bonded in all directions; in a protein, bonds are strong (covalent) only along the backbone. An atom in a biomolecule can (usually) only vibrate; atoms in biomolecules can move.

2.5. *Solids and biomolecules*

Before discussing functional aspects of a protein, we compare a typical solid and a globular protein of about the same size, as sketched in Fig. 6. While there are some similarities in the physical behavior, there exist some profound differences that are important for the function of proteins:

(i) The solid is periodic, the protein aperiodic. The aperiodicity permits the protein to perform particular functions, but makes the detailed description more difficult.

(ii) In a solid, the "strong" forces that hold the atoms together are essentially equally strong in all three directions. In a protein, however, the bonds are "strong" (covalent) along the backbone, but the cross connections are "weak" (hydrogen bonds, disulfide bridges, Van der Waals forces). A solid is "dead" and an individual atom can, as a rule, only vibrate around its equilibrium position. In contrast, the weak bonds in a biomolecule can be broken by thermal fluctuations. A biomolecule can therefore execute large motions, it can breathe and can act as a miniature machine.

(iii) A solid is spatially homogeneous, apart from surface effects and from small defects. A biomolecule, in contrast, is inhomogeneous; some regions behave like solids and others more like liquids.

(iv) A solid cannot be modified on an atomic or molecular scale at a particular point; modifications are either periodic or random. In contrast, a protein can be changed at any desired place at the molecular level: Through

genetic engineering, the primary sequence is modified at the desired location and this modification leads to the corresponding change in the protein. We have stated in Section 2.3 that the partial sequence .. UCU–UAC–ACG ... corresponds to the amino acid sequence .. Ser–His–Ala .... If the letter A in the second codon (word) is changed to G, His is replaced by Arg, as can be understood with Table II. This change actually occurs in hemoglobin Zürich, with consequences for the carriers.

The four differences between solids and biomolecules already indicate that the physics of biomolecules possesses new, exciting (and difficult) features. Two more aspects make it even clearer that the field is enormously rich:

(v) The number of "possible" biomolecules is incredibly large. Consider a medium-sized protein, constructed from 150 amino acids. Since there exist 20 amino acids, the number of possible combinations is $(20)^{150} \approx 10^{200}$. If we produce one copy of each combination and fill the entire universe, we need 10^{100} copies of our universe to store all combinations. This example demonstrates that biomolecular problems cannot be solved by random experiments.

(vi) A protein with a given primary sequence can fold into a very large number of slightly different conformational substates. Each individual building block can, on the average, assume 2–3 different configurations with approximately equal energy. The entire protein thus possesses about $(2–3)^{150}$ states of approximately equal energy.

The properties alluded to in (i) to (vi) above imply that biomolecules are complex many-body systems. Their size indicates that they lie at the border between classical and quantum systems. Since motion is essential for their function, as suggested in Fig. 6, collective phenomena play an important role. Moreover, we can expect that many of the features involve nonlinear processes. Function, from storing information, energy, charge, and matter, to transport and catalysis, is an integral characteristics of biomolecules. The physics of biomolecules is a rich field. It stands now where nuclear, particle, and condensed matter physics were around 1930. We can expect exciting progress in the next few decades.

3. Physics of heme proteins

Experimental studies have to be performed on a particular system. Atomic physics made great progress, from the Balmer relation past the Bohr atom, the Schrödinger and Dirac equations to quantum electrodynamics because the hydrogen atom is simple enough to make accurate calculations and unambiguous experiments possible, yet complex enough to show nontrivial behavior. The choice of good systems among biomolecules is more difficult. The number of possible systems is unbelievably large and nature probably has constructed far more than 10^7 proteins. The system selected should be "simple" enough so that progress in understanding is possible, yet complex enough that contact to biochemistry and biology can be made.[1] *Heme proteins* satisfy the conditions reasonably well. They

[1]) It is important to point out that work in biological physics requires close interaction among physicists, chemists, biochemists, and biologists. A physicist working alone easily gets lost, works on uninteresting and/or improperly prepared systems, and asks the wrong questions.

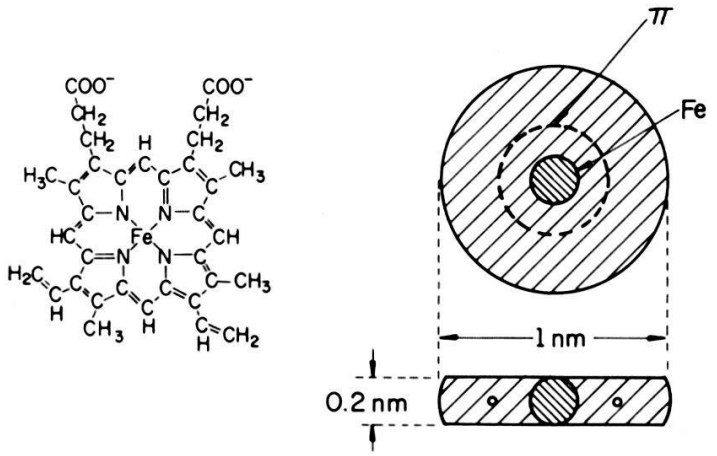

Figure 7
The heme group as seen by chemists and by physicists. π indicates the pi-electron ring.

have been studied by many different tools in great detail, their structure and function can be described and related, and they perform many different biological functions. In heme proteins, nature has taken one particular organic molecule, the heme group, and modified its behavior through the protein structure so that the entire system can perform a wide variety of tasks. Thus in studying heme proteins, the physics of structure and of function can be attacked in a systematic way.

3.1. *Heme proteins*

The *heme group*, shown in Fig. 7a, consists of an organic part and an iron atom. For the *chemist*, the organic part, protoporphyrin, is made up of four pyrrole groups, linked by methene bridges to form a tetrapyrrole ring. Four methyl, two vinyl, and two propionate side chains are attached to the tetrapyrrole ring. The side chains can be arranged in 15 different ways, but only one arrangement, protoporphyrin IX, is commonly found in biological systems. The iron atom binds covalently to the four nitrogens in the center of the protoporphyrin ring. The iron can form two additional bonds, one on either side of the heme plane. (For details, consult references [7–9].)

For *physicists*, who are usually afraid of any molecular structure with more than two atoms, the heme group can be shown as a disk, about 1 nm in diameter and 0.2 nm thickness (Fig. 7b). The disk has an iron atom with two free valences in the center and a one-dimensional electron ring (pi electrons) surrounds the iron.

In *heme proteins*, the heme group is embedded in the protein, as sketched in Fig. 8. In most cases, one of the free iron bonds is linked covalently to a particular side chain of the protein and the other side is free to accept for instance an oxygen molecule. In some cases, both iron bonds are used for connections to the protein.

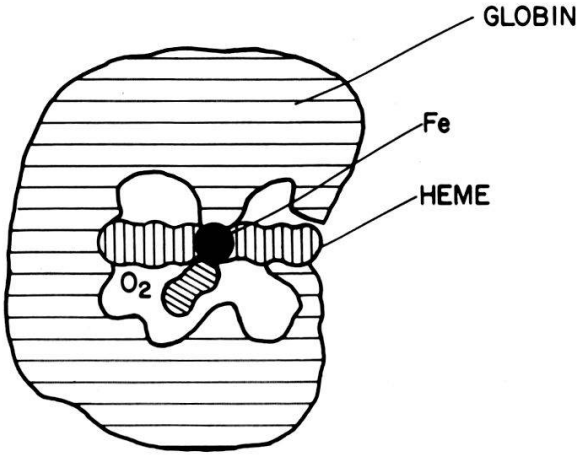

Figure 8
A cross section through oxymyoglobin.

Heme proteins perform a wide variety of tasks. The best known heme protein is *hemoglobin*, the oxygen carrier in blood [10–13]. Myoglobin stores and transports oxygen in muscles. Cytochrome c transports electrons, cytochrome P450 detoxifies substances, peroxidases oxidize with H_2O_2. Chlorophyll, where the iron atom is replaced by Mg, is essential in the transformation of light to chemical energy. We will often use *myoglobin* (Mb) as example. Mb consists of 153 amino acids and has a molecular weight of about 17,800 dalton. Its dimensions are about 3 nm × 4 nm × 4 nm. Since its function is, as far as we know, relatively simple and its structure is well known, we can consider it as the "hydrogen atom of biology." It is easily obtainable (mainly from sperm whales), cheap, and even physicists have a difficult time destroying it. Mb thus is a nearly ideal protein for detailed physical studies.

3.2. *The average (static) protein structure*

Proteins are aperiodic systems. Without a knowledge of their three-dimensional structure, a deep understanding is impossible. In solids and molecules, the static structure is determined by X-ray diffraction. The same technique is used for proteins, but the problem is much more difficult. How difficult can be seen from the quote by von Laue [14]: "Die Elektronenverteilung in ihnen zu bestimmen, ist von vorneherein aussichtslos." It needed the patience and perseverence of Max Perutz to solve the problem.

In principle, the structure, i.e. the electron distribution in a protein can be determined without a protein single crystal. Consider the scattering of X-rays from a single oriented protein molecule as shown in Fig. 9. The electron distribution of the protein molecule determines the scattering intensity $I(\theta)$ where θ is the scattering angle. If $I(\theta)$ is measured for a number of orientations of the protein molecule, the charge distribution can be found by Fourier transformation. The principle is the same as in the structure determination of particles and nuclei [15]. A major obstacle in all these experiments is the phase problem; In order to invert directly, the scattering amplitude is required, but $I(\theta)$ is proportional to the absolute square of the amplitude. The scattering phase must be determined by a separate experiment. We will not discuss this problem here, but refer to the

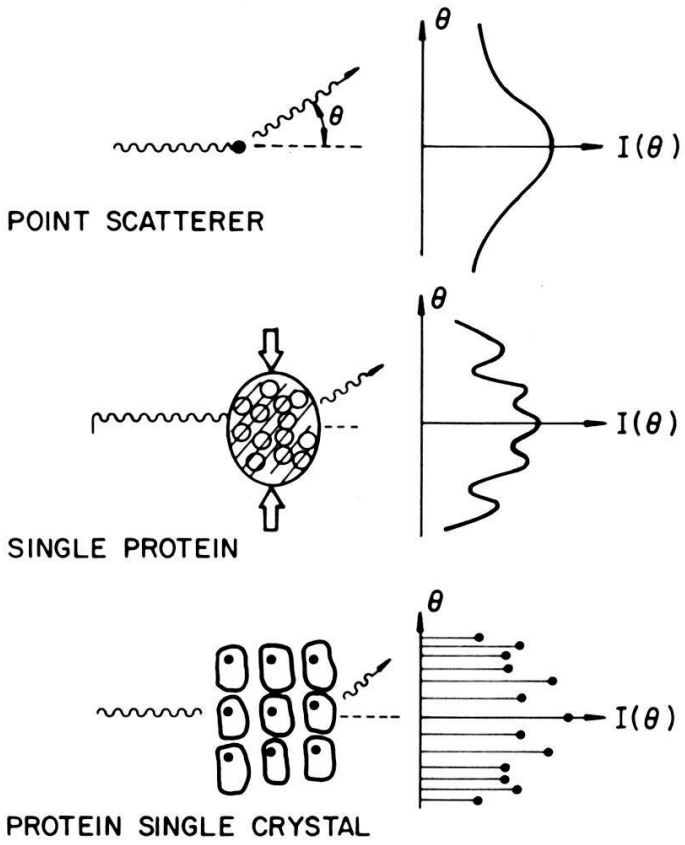

Figure 9
Scattering from a point scatterer, a single oriented protein, and a protein single crystal.

literature [6, 16–19] and only mention that the Mössbauer effect also can lead to a solution of the phase problem [20].

The determination of the structure of a protein by using a single protein suffers from three fatal shortcomings: It is not possible to select and orient a single protein, the scattering intensity from a single proton is too small, and the protein would be radiation damaged. All three problems are overcome by using protein single crystals, as sketched in Fig. 9. The protein molecules are oriented in the crystal, the crystal can be handled and positioned correctly, the coherent interference from various proteins increases the intensity, and radiation damage is minimized. The scattering intensity $I(\theta)$ is no longer a smooth function of θ; the interference from the very large number of proteins leads to the appearance of discrete "spots" (Laue–Bragg pattern). After the phase problem is solved, the analysis of the diffraction pattern leads to the determination of the electron density. The protein structure thus is found. One remark is in order here. The *position* of the spots in the diffraction pattern depends on the lattice parameters of the entire protein crystal; the electron distribution within each unit cell determines the *intensity* of the spots. The protein structure consequently is obtained from the spot intensities.

Myoglobin and hemoglobin were the first two globular proteins whose atomic structure was determined by X-ray diffraction: myoglobin by Kendrew and hemoglobin by Perutz [21, 22]. At present, the atomic structure of a large number of proteins is known. Initially, the best tools to visualize the structures were

painstakingly built models. At present, computer graphics permits representations of astonishing beauty in which even small details can be studied [23, 24].

X-ray diffraction is well suited to look at the heavier atoms (for instance C, N, O) in proteins but it is difficult to see the very important hydrogen atoms. Neutron protein diffraction fills the gap and permits direct location of the hydrogen atoms [25, 26].

3.3. *Ligand binding to heme proteins*[2])

Biomolecules, in contrast to ordinary systems, can be studied in two different ways: Their physical properties can be investigated just as if they were ordinary molecules or crystals, but their biological processes can also be used as probes. The second approach is in the spirit of the present symposium because collective phenomena play an essential role in biological processes.

The binding of small molecules such as oxygen or carbon monoxide to heme proteins is one of the simplest biological phenomena [27]. We describe here studies of the binding process and sketch how such investigations can yield rich information concerning the static and dynamic properties of proteins. The method is sketched in Fig. 10. The system to be studied, for instance MbCO (myoglobin with carbon monoxide bound to the heme iron) is placed in a cryostat. A laser flash breaks the bond between the iron and the ligand. The ligand molecule moves away from the binding site, but later rebinds. Photodissociation and rebinding can be observed optically because MbCO and Mb, or MbO_2 and Mb, have different optical absorption spectra as shown in Fig. 10. (This fact is even known to physicists: arterial and venous blood have a different color!) The behavior of the binding process as a function of external parameters permits many conclusions concerning the heme protein.

In our own work, we have investigated the binding of O_2 and CO to many heme proteins in the time range from about 100 ps to 1 ks (13 orders of magnitude in time), the temperature range from 2 to 330 K, the pressure range up to 200 MPa (~2 kbar), and in solvents of various viscosities [28–32]. When biochemists or biologists learn that studies are performed at very low temperatures they usually ask "why?". Evolution has adjusted the processes at physiological temperatures so that they mesh smoothly. To study the elementary steps, the overall process must be slowed down and separated into components; such a separation occurs as the temperature is lowered.

The number of investigations of ligand binding to heme proteins is very large [27]; we select here mainly the ones from our own group. The important results of our flash photolysis experiments can be summarized as follows. Before photodissociation, the ligand is bound to the heme iron as shown in Fig. 8. At temperatures below about 180 K, the ligand molecule remains within the heme pocket after photodissociation and rebinds from there. We denote this geminate process by "$\mathcal{I}$." Above about 180 K, a ligand can either rebind directly or first move into the protein matrix. The ligands that move into the matrix diffuse around there, return to the pocket and then bind. We call this geminate return trip "matrix process, $\mathfrak{M}$." Above about 220 K, ligands in the matrix can either return to the

[2]) The term "ligand" is used by biochemists to denote small molecules that bind to proteins or nucleic acids.

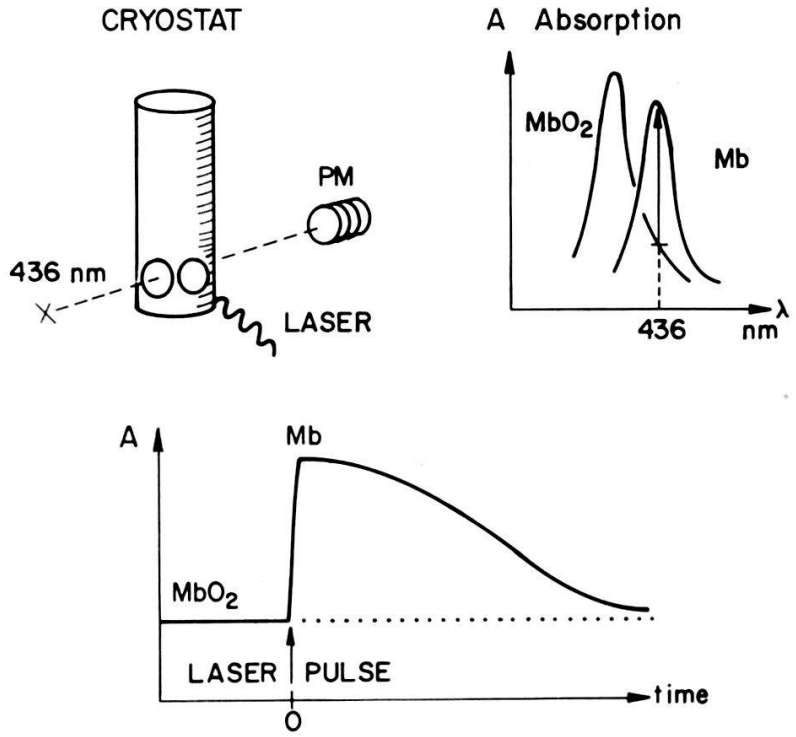

Figure 10
Flash photolysis. The protein sample (MbO_2) is placed in a cryostat, the absorbance is monitored at a suitable wavelength, and the MbO_2 is photodissociated by a laser flash. Because MbO_2 and Mb differ in the optical absorbance, photodissociation and rebinding can be monitored.

pocket or move out into the solvent. Proteins that have lost their ligand entirely acquire one again through second-order binding of a ligand present in the solvent. The rate of the "solvent process, $\mathfrak{S}$" is proportional to the ligand concentration in the solvent. In myoglobin, the ligand process dominates above about 280 K. Each of the three processes, $I, \mathfrak{M}$, and $\mathfrak{S}$, yields interesting results. We discuss the phenomena deep inside the protein in Section 3.4, and return to $\mathfrak{M}$ and $\mathfrak{S}$ in Section 3.6.

3.4. *Low-temperature phenomena*

Below about 180 K, the photodissociated ligand, for instance CO, remains within the heme pocket and rebinds from there. We denote the protein state with CO in the heme pocket by B and call the bound state (MbCO) A. Fig. 11 shows the heme and part of its environment in the two states. His denotes the distal histidine, the amino acid that is linked covalently to the fifth bond of the iron atom. In state A the heme is planar, the iron has spin 0 and is very close to the mean heme plane. In state B, the heme is domed, the iron has spin 2 and lies about 0.5 Å out of the mean heme plane. The CO molecule is essentially free in the heme pocket: The CO stretching vibration after photodissociation at temperatures below 20 K is close to that of the free CO [33]. The change to spin 2 even below 5 K has been proven with a superconducting susceptometer [34]. Figure 11 shows that the binding step $B \rightarrow A$ requires a motion of the iron into the mean heme plane, a spin change $2 \rightarrow 0$ of the iron, a flattening of the heme group, and a simultaneous approach of the CO molecule. The entire process can be described in the vastly simplified diagram Fig. 11b with B represented by a shallow and A by a deep well. If Fig. 11 is correct, two phenomena should be observable: At

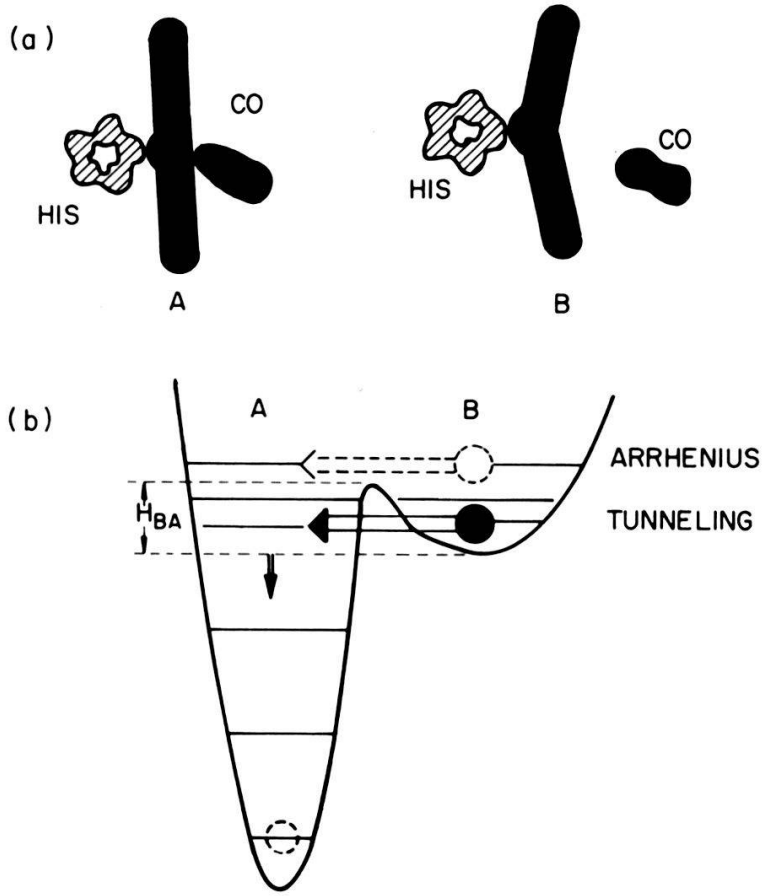

Figure 11
(a) The two states involved in ligand binding inside heme proteins. In *A*, CO is bound, in *B* CO is free in the heme pocket. (b) Schematic representation of the potential that describes the binding $B \rightarrow A$.

high temperatures, the CO molecule should bind by overcoming the barrier in a classical Arrhenius motion, at low temperatures it should tunnel through the barrier. (Of course the two processes are not different, they are two aspects of the same phenomenon.) At high temperatures, the binding rate consequently should depend exponentially on temperature,

$$k_{BA} = A_{BA} \exp(-H_{BA}/RT). \tag{1}$$

In the low-temperature limit, the binding rate should become temperature independent. Indeed, the experiments, shown for one case in Fig. 12, exhibit the expected behavior: Above about 20 K, the average binding rate coefficient $\bar{k}_{BA}$ increases exponentially in agreement with equation (1); below about 10 K, $\bar{k}_{BA}$ is essentially constant (29). The heme pocket turns out to be an excellent laboratory to study quantum-mechanical tunnelling [35, 36]. Tunnelling is not only characterized by a temperature-independent rate for $T \rightarrow 0$, it should also show a characteristic dependence on the mass of the tunnelling system. Using a Fourier-transform infrared spectrometer and following binding through the observation of a change in the CO stretching vibration on binding, we have also verified the isotope dependence of the transition $B \rightarrow A$ at 20 K [37]. While the existence of molecular tunnelling in heme proteins is established, a great many questions, particularly concerning the connection between structure and binding, remain unsolved and will require more experiments and more detailed theories.

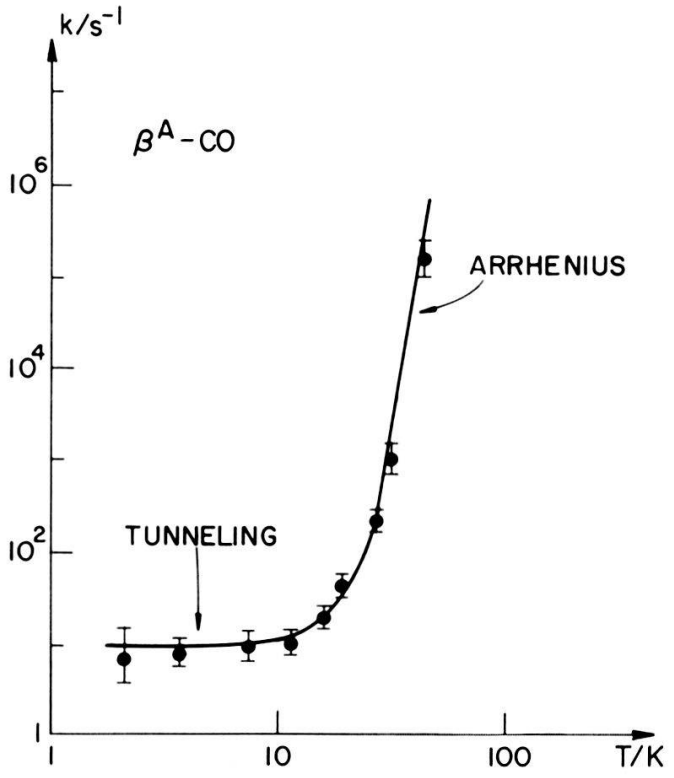

Figure 12
Tunnelling and classical Arrhenius transitions in the low-temperature binding of CO to the separated beta chain of hemoglobin.

The unambiguous observation of molecular tunnelling was actually not the first surprise revealed by the low-temperature studies on ligand binding in heme proteins. The time dependence of the transition $B \rightarrow A$ presented the first puzzle. To understand the puzzle, consider again Fig. 11. We assume that photodissociation promotes all protein molecules into state B at the time $t = 0$. Each protein molecule rebinds a ligand with a rate coefficient k_{BA}. The fraction $N(t)$ of protein molecules that have not rebound a ligand at the time t after photodissociation is given by the standard exponential law,

$$N(t) = \exp(-k_{BA}t). \tag{2}$$

The experimental data found in hundreds of experiments [38, 39, 28–32] disagree with the simple law equation (2). A typical set of curves for the binding of CO to Mb, plotted as $\log N(t)$ versus $\log t$, is given in Fig. 13. These curves demonstrate that the rebinding process is not exponential in time, but approximately follows a power law,

$$N(t) \cong (1 + t/t_0)^{-n}. \tag{3}$$

Here, n and t_0 are temperature-dependent parameters; the values of these parameters for the binding of CO to Mb at 60 K are $n = 0.17$, $t_0 = 0.16$ s. Figure 13 shows that at 60 K, rebinding becomes observable at about 10^{-3} s; after 10^3 s, $N(t) = 0.2$ so that 80% of the Mb molecules have rebound a CO. The extrapolation to $N(t) = 10^{-3}$ predicts a time of 7×10^{16} s or 2×10^9 years! At 60 K, even during a time comparable to the age of the universe not all CO will rebind. The existence of such "endless" phenomena was observed 150 years ago, when

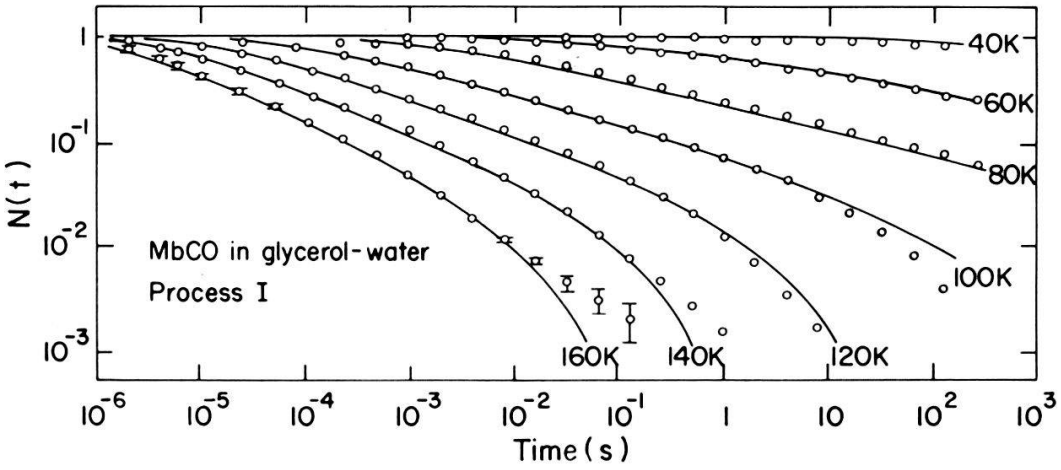

Figure 13
Time dependence of the binding of CO to Mb between 40 and 160 K. $N(t)$ is fraction of Mb molecules that have not rebound a CO molecule at the time t after photodissociation. (After ref. 28.)

Wilhelm Weber followed a suggestion by Gauss and studied the torsion of silks threads [39, 40]. Kohlrausch somewhat later investigated the torsion of glass fibers and the speed of discharge of a Leyden jar [41]. In both cases he found fractal time dependences of the form t^{-n}, similar to the relation equation (3).

The simplest (and only) explanation that we have been able to find for the nonexponential time dependence can be described with Fig. 11. If the barrier height H_{BA} between wells B and A is not unique, but different in different protein molecules, the rate coefficients k_{BA} will also be different for different molecules of the same protein. Denote with $g(H_{BA})\,dH_{BA}$ the probability of finding a barrier with height between H_{BA} and $H_{BA}+dH_{BA}$. The function $N(t)$ then is no longer given by equation (2), but by the generalization

$$N(t)=\int dH_{BA} g(H_{BA}) \exp(-k_{BA}t), \tag{4}$$

where k_{BA} is related to H_{BA} by the Arrhenius relation equation (1) if the tunnel effect can be neglected. Equation (4) must be inverted to find the distribution function $g(H_{BA})$ from the experimentally determined $N(t)$. Unfortunately the inversion, an incomplete inverse Laplace transform, is highly unstable. Nevertheless, to the desired accuracy it can be performed [42, 28] and a few typical results are shown in Fig. 14. To a nuclear physicist, the distributions look like beta spectra and suggest that a statistical phenomenon must be hidden somewhere. To a biochemist, the curves show that the distribution is very different in different proteins and can be used to characterize and understand different ligand-protein combinations. In the following section, we will describe our explanation for the distribution $g(H_{BA})$; in Section 3.6, we will discuss functional implications.

3.5. *Proteins as a state of matter*

Figures 13 and 14 and equation (3) raise some fundamental questions: Why is the time dependence of ligand binding to a protein at low temperatures fractal? What causes the distribution in barrier height? The answer to these questions leads to the realization that a protein is a state of matter different from solid, liquid, or gaseous. A protein combines properties of the three states and in some

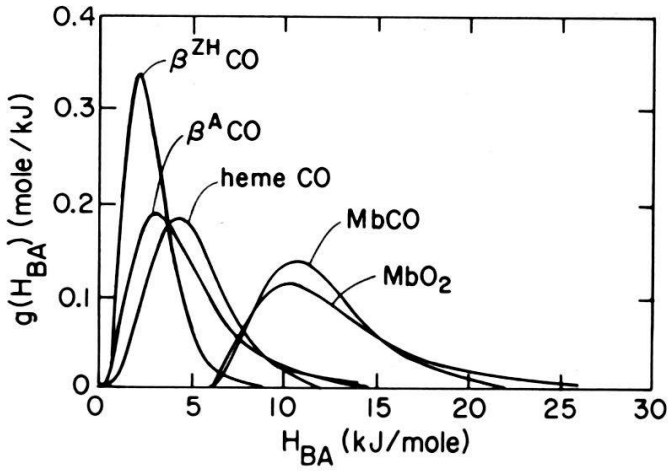

Figure 14
Activatation energy distributions for various heme proteins. β^A is the separated beta chain of ordinary human hemoglobin, β^{ZH} is the beta chain of human hemoglobin Zürich.

aspects resembles an amorphous solid or a glass. In addition, however, a protein has a purpose and consequently an organization that is superimposed over the simpler features. To describe the situation, we return to the construction of a protein as sketched in Section 2 and in Figs. 5 and 6.

The crucial point is that a protein cannot be in a unique state of lowest energy, its ground state must be very highly degenerate. We can make this statement plausible in two different ways. Consider a protein in a given conformation, a geometrical arrangement of all atoms that permits the protein to execute a particular function. Some of the hydrogen bonds that stabilize the protein can be shifted without changing the overall binding energy appreciably. Many of the side groups can be rotated without affecting the overall energy. The protein will thus be able to assume a large number of *conformational substates*, all with essentially equal energy and all performing the same biological function. The number of such states is difficult to estimate, but it is very large. Moving from one substate to another requires breaking and reforming one or more hydrogen bonds, or rotating part of a side chain past other atoms. Two substates will consequently be separated by an energy barrier. The complete description of the structure of the protein in any substate requires specification of the coordinates of all atoms in a space of about 10^4 dimensions. We simplify the description by using only *one* conformational coordinate; the energy of the protein as function of this coordinate is sketched in Fig. 15. The representative point is one particular well means that the protein is in that particular conformational substate, with all 10^4 coordinates having the corresponding values.

The second way to make the existence of substates plausible starts with folding. The construction of a particular protein on the ribosome takes only a short time. The folding process is, however, so complex that it is impossible for a particular protein molecule to reach the optimal state within, say, one second. (In fact, no computer calculations have yet succeeded in reproducing a folding path.) A particular protein must therefore be satisfied in reaching a state that approximates the best reasonably well.

The existence of substates can explain the observed distribution of barrier heights. Assume that each substate implies a particular barrier height. At low

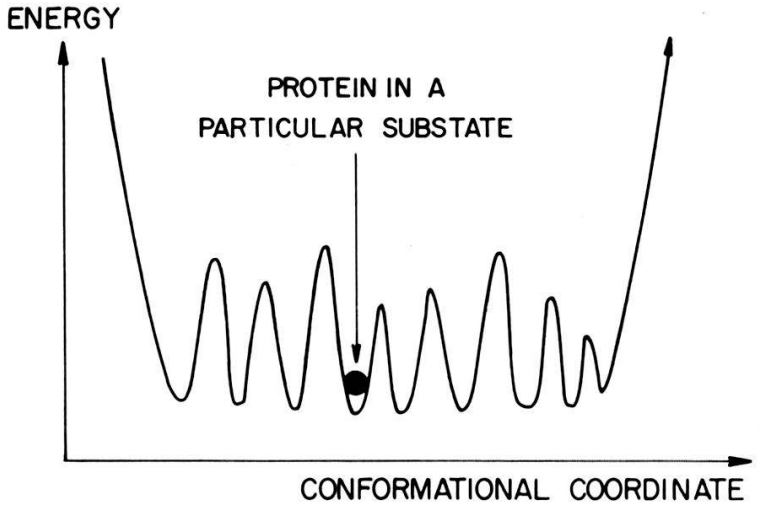

Figure 15
Protein energy versus configurational coordinate. A large number of conformational substates (potential minima) have essentially the same energy. The dot represents a protein in a particular substate.

temperatures, transitions among the substates will be very slow or absent: Each protein molecule is frozen into a particular substate with a corresponding barrier height and the observed rebinding rate will be given by a distribution as indicated in equation (4). At high temperatures, however, transitions among the substates can be very fast and the transition $B \rightarrow A$ occurs with an average rate $\bar{k}_{BA} = \int dH_{BA} k(H_{BA}) g(H_{BA})$.

The postulate of conformational substates implies that many parts of the protein can undergo considerable motion and induced some biochemists to ask: "If the protein is indeed flexible and does not exist in a unique structure, why does X-ray diffraction give us such beautiful structures?" This question led us to look at the information contained in X-ray diffraction with the result that more evidence for substates became apparent [43]. To understand the basic physical concepts, consider again Fig. 9b. Each spot in a Laue diagram receives contributions from all atoms in each protein. To form the spot, the wavelets from equivalent atoms in different unit cells interfere constructively, as shown for two atoms in Fig. 16a. If all equivalent atoms sit in exactly the right position, interference is maximal. If, however, the atoms do not occupy the "ideal" position, but are spread out over a linear distance $\langle x^2 \rangle^{1/2}$ as indicated in Fig. 16b, the contribution to the total intensity of the relevant Laue spot is reduced by the *Debye-Waller factor* [44–46]

$$T = \exp(-8\pi^2 \langle x^2 \rangle \sin^2 \theta / \lambda^2). \tag{5}$$

Here λ is the wavelength of the X-rays, $\langle x^2 \rangle$ the mean square displacement of the atom, and θ the scattering angle as shown in Fig. 16a. Initially, equation (5) was derived by Debye and by Waller in order to correct for the thermal motion of the atoms. In a good approximation, equation (5) will also apply if the atoms are displaced for some other reason. Protein crystals were known to have large values of $\langle x^2 \rangle$ but these were usually ascribed to crystal disorder. The observation of nonexponential time dependence at low temperatures and the resulting postulate of the existence of conformational substates leads to a different interpretation of the large values of $\langle x^2 \rangle$ in proteins (43): The large fluctuations seen are not caused

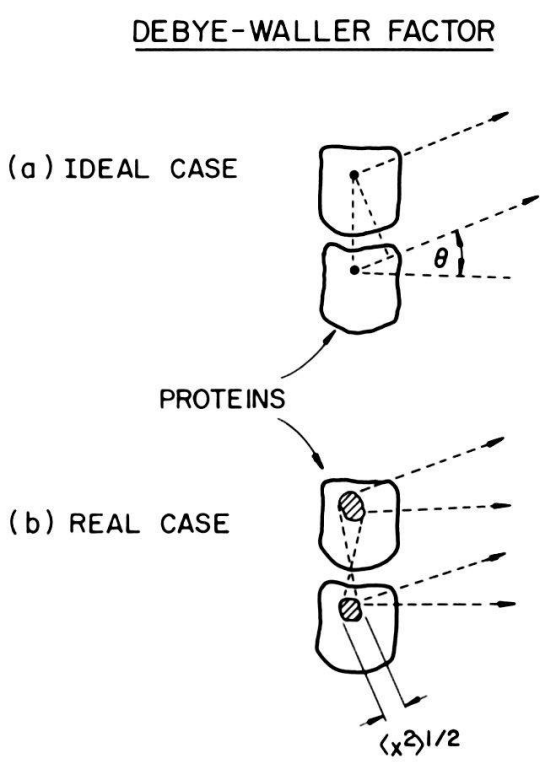

Figure 16
Debye–Waller factor. In the ideal case, top, equivalent atoms in different proteins (different unit cells) are at the same position. The contribution to a particular Laue spot is maximal. In the real case (bottom) atoms are randomly displaced over a volume with linear dimensions $(\langle x^2 \rangle)^{1/2}$ and the contribution to the intensity is reduced by the Debye–Waller factor.

by random disorder, but are integral and essential properties of biomolecules, necessary for their function. Indeed, detailed studies of the values of $\langle x^2 \rangle$ provide significant information [43, 47]. Similar studies in other proteins reinforce this conclusion [48, 49].

Figures 17 and 18 provide two examples of the type of information that can be extracted from the Debye–Waller factor. Figure 17 shows a particular residue, Lys 147. Table I indicates that lysine has a long polar side chain; in Lys 147, the side chain extends from the protein into the solvent. The diagram at the bottom of Fig. 17 demonstrates that the mean-square displacement is small near the backbone and increases towards the solvent. Even though X-ray diffraction does not give information about motion, the implication is clear: the side chain of Lys 147 moves. In Fig. 18, $\langle x^2 \rangle$ averaged over the atoms NCC of the backbone is plotted versus residue number. The open circles refer to 300 K, the solid dots to 80 K. In a solid all atoms would have the same $\langle x^2 \rangle$; $\langle x^2 \rangle$ would be about 0.01 Å^2 at 300 K and about 0.003 Å^2 at 80 K. Figure 18 shows that myoglobin behaves very differently from a solid: $\langle x^2 \rangle$ is not uniform, in most places it is much larger than 0.01 Å^2, and the temperature dependence differs in different parts. We can interpret Fig. 18 by saying that myoglobin contains solid-like and semi-liquid parts. In the solid-like regions, the mean-square displacement is small, in the semi-liquid ones it can be large. The detailed discussion suggests that the different behavior has biological significance [43, 47].

3.6. *Protein dynamics*

We have stated in the previous section that proteins are dynamic and not static systems. At physiological temperatures they move and breathe and the

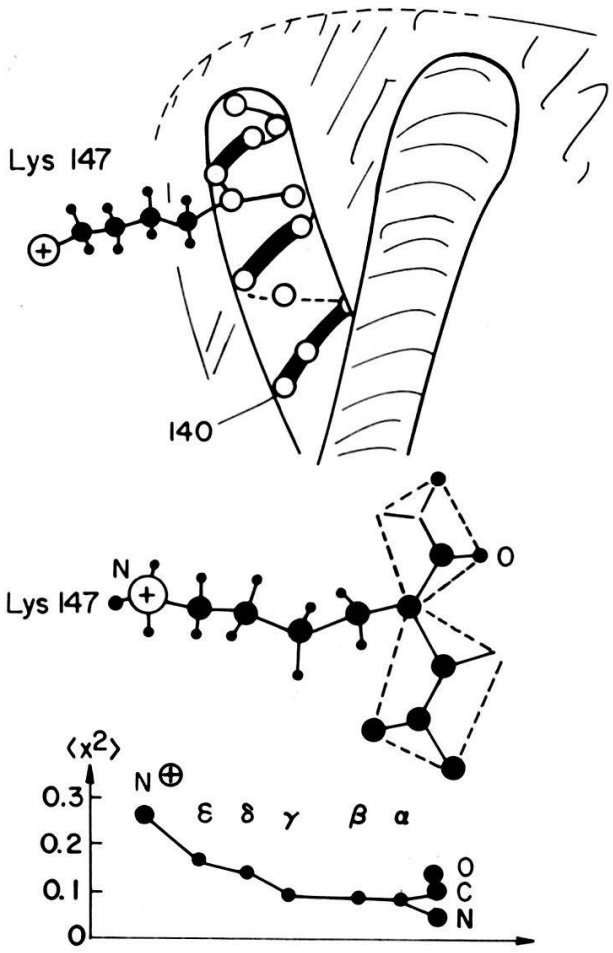

Figure 17
The mean-square displacements of the atoms of the residue Lys 147 in myoglobin. (Data from ref. 43.) Top: The position of residue Lys 147.

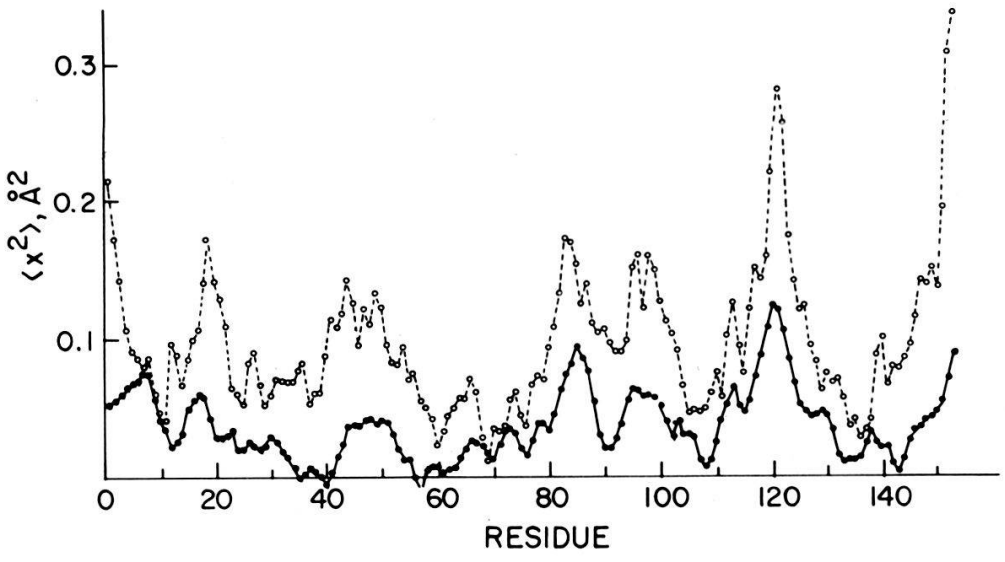

Figure 18
Values of the mean-square displacement $\langle x^2 \rangle$ for the backbone of Mb. The points represent averages over the atoms (N, C_α, C) of each backbone, plotted versus residue number. Open circles at 300 K, dots at 80 K. (After ref. 47.)

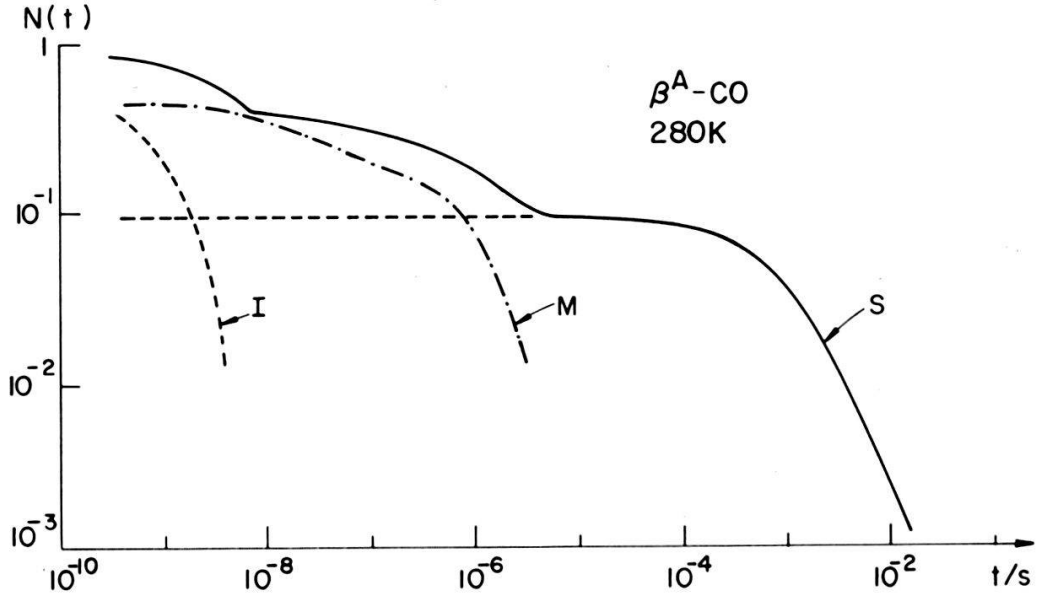

Figure 19
Rebinding of CO to the separated beta chain of human hemoglobin at 280 K, The solid line represents the experimental data. I is the internal rebinding from the pocket, $\mathfrak{M}$ the rebinding from the matrix, and $\mathfrak{S}$ the binding from the solvent.

motion is important for the biological function [50–54]. The various possible modes, their time scales and their functional importance are so complex that we are only at the beginning of this field. Every possible tool is required and theory and experiment must work hand-in-hand. We will not discuss the theoretical approaches here, but state only that theoretical studies of protein dynamics, particularly by Karplus and his collaborators, have made great progress and that contact between theory and experiment has been established [55–58].

The binding of CO and O_2 to heme proteins is again a phenomenon where the importance of protein dynamics can be studied in detail [59]. Perutz first pointed out that the static (average) structure of hemoglobin does not have an opening or channel big enough the let dioxygen enter or leave [60]. The flash photolysis experiments discussed in Section 3.3, extended to about 330 K, provide us with a clearer picture of where dynamic features are essential. The prominent features of the rebinding of CO to a single chain (β^A) of the hemoglobin tetramer are shown in Fig. 19. The overall curve for $N(t)$ can be separated into the three processes I, $\mathfrak{M}$, and $\mathfrak{S}$ introduced in Section 3.3. The matrix process S is slowest; it is exponential in time and proportional to the CO concentration in the solvent. The fastest process, I, can be obtained by extrapolation of its very well studied behavior at low temperatures [28]. The extrapolation fits the data well. Above about 260 K, process I is no longer nonexponential in time (see Section 3.4), but has become exponential. After subtraction of I and $\mathfrak{S}$ from the experimental $N(t)$, the matrix process $\mathfrak{M}$ is left. It is independent of the CO concentration in the solvent. The time dependence is approximately given by $t^{-1/2}$ over a considerable range in time. The interpretation of the three processes has already been given in Section 3.3 and we repeat the essential features for the solvent process: A CO molecule coming from the solvent enters the protein matrix through the hydration shell, diffuses (or percolates) through the protein matrix to the heme pocket, shuttles around in the heme pocket, and finally binds

covalently at the heme iron. Protein dynamics is essential in all steps of this surprisingly complex process. This fact can for instance be ascertained by studying the rates of the various processes as a function of the viscosity of the external solvent [30]; the observed dependence can be understood in terms of protein dynamics [61].

We have now moved from atoms to dynamic proteins, from individual properties to collective phenomena, involving a large number of atoms in a well-orchestrated way. The survey is superficial, only a few selected topics have been touched and a great deal of material has been omitted. Moreover, while we already know and understand a great deal, we really know only an extremely small fraction of what could be known and understand even less. It is, however, likely that the understanding of biomolecules will progress rapidly and that future experiments will yield many exciting and unexpected results.

Acknowledgements

Much of the work discussed here was performed with the collaborators whose names are listed on the various publications. I should like to thank them for many discussions and interactions. Part of the work was supported by Grant PHS GM 18051 from the Department of Health and Human Services and by Grant PCM82–09616 from the National Science Foundation.

REFERENCES

[1] ERWIN SCHRÖDINGER, *What is Life*? Cambridge University Press 1944, 1967.
[2] R. E. DICKERSON and I. GEIS, *The Structure and Action of Proteins*. Benjamin/Cummings, 1969.
[3] L. STRYER, *Biochemistry*, 2nd Ed. W. H. Freeman, 1981.
[4] J. D. WATSON, *Molecular Biology of the Gene*. 3rd. Ed. Benjamin, 1976.
[5] C. R. CANTOR and P. R. SCHIMMEL, *Biophysical Chemistry*, 3 Vol. W. H. Freeman, 1980.
[6] *Biophysics*, W. HOPPE, W. LOHMANN, H. MARKL, and H. ZIEGLER, Eds. Springer, 1983.
[7] *Porphyrins and Metalloporphyrins*, K. M. SMITH, Ed. Elsevier, 1975.
[8] *The Porphyrins*, 7 Vols., D. DOLPHIN, Ed. Academic Press, 1979.
[9] *Iron Porphyrins*, 3 Vols., A. B. P. LEVER and H. B. GRAY, Eds. Addison-Wesley, 1983.
[10] M. F. PERUTZ, Scientific American *211* (11) 2 (1964).
[11] M. F. PERUTZ, Scientific American *239* (6) 92 (1978).
[12] M. WEISSBLUTH, *Hemoglobin*, Springer, 1974.
[13] R. E. DICKERSON and I. GEIS, *Hemoglobin*: *Structure, Function, Evolution and Pathology*. Benjamin/Cummings, 1983.
[14] M. VON LAUE, *Röntgenstrahlinterferenzen*, Akademische Verlagsgesellschaft, 1948.
[15] H. FRAUENFELDER and E. M. HENLEY, *Subatomic Physics*, Chapter 6, Prentice Hall, 1974.
[16] W. L. BRAGG and M. F. PERUTZ, Proc. Roy. Soc. London *A225*, 315 (1954).
[17] G. H. STOUT and L. H. JENSEN, *X-ray Structure Determination*. Macmillan, 1968.
[18] T. L. BLUNDELL and L. N. JOHNSON, *Protein Crystallography*. Academic Press, 1976.
[19] J. D. DUNITZ, *X-ray Analysis and the Structure of Organic Molecules*. Cornell University Press 1979.
[20] R. L. MÖSSBAUER, F. PARAK, and W. HOPE, in *Mössbauer Spectroscopy II*, U. Gonser, Ed. Springer, 1981, p. 5.
[21] J. C. KENDREW, Science *139*, 1259 (1963).
[22] M. F. PERUTZ, Science *140*, 863 (1963).
[23] R. J. FELDMANN, Annu. Rev. Biophys. Bioeng. *5*, 477 (1976).
[24] R. LANGRIDGE, T. E. FERRIN, I. D. KUNTZ, and M. L. CONNOLLY, Science *211*, 661 (1981).
[25] B. P. SCHOENBORN and A. C. NUNES, Annu. Rev. Biophys. Bioeng. *1*, 529 (1972).

[26] A. A. KOSSIAKOFF, Annu. Rev. Biophys. Bioeng. *12*, 159 (1983).
[27] E. ANTONINI and M. BRUNORI, *Hemoglobin and Myoglobin in Their Reactions with Ligands*. North Holland, 1971.
[28] R. H. AUSTIN, K. W. BEESON, L. EISENSTEIN, H. FRAUENFELDER, and I. C. GUNSALUS, Biochemistry *14*, 5355 (1975).
[29] N. ALBERDING, R. H. AUSTIN, K. W. BEESON, S. S. CHAN, L. EISENSTEIN, H. FRAUENFELDER, and T. M. NORDLUND, Science *192*, 1002 (1976).
[30] D. BEECE, L. EISENSTEIN, H. FRAUENFELDER, D. GOOD, M. C. MARDEN, L. REINISCH, A. H. REYNOLDS, L. B. SORENSEN, and K. T. YUE, Biochemistry *19*, 5147 (1980).
[31] W. DOSTER, D. BEECE, S. F. BOWNE, E. E. DIIORIO, L. EISENSTEIN, H. FRAUENFELDER, L. REINISCH, E. SHYAMSUNDER, K. H. WINTERHALTER, and K. T. YUE, Biochemistry *21*, 4831 (1982).
[32] D. D. DLOTT, H. FRAUENFELDER, P. LANGER, H. RODER, and E. E. DIIORIO, Proc. Natl. Acad. Sci. USA *80*, 6239 (1983).
[33] J. O. ALBEN, D. BEECE, S. F. BOWNE, W. DOSTER, L. EISENSTEIN, H. FRAUENFELDER, D. GOOD, J. D. MCDONALD, M. C. MARDEN, P. P. MOH, L. REINISCH, A. H. REYNOLDS, E. SHYAMSUNDER, and K. T. YUE, Proc. Natl. Acad. Sci. USA *79*, 3744 (1982).
[34] H. RODER, J. BERENDZEN, S. F. BOWNE, H. FRAUENFELDER, T. B. SAUKE, E. SHYAMSUNDER, AND M. B. WEISSMAN, Proc. Natl. Acad. Sci. USA (1984).
[35] V. I. GOLDANSKII, Chemica Scripta *13*, 1 (1978–79); Nature *279*, 109 (1979).
[36] H. FRAUENFELDER, in *Tunnelling in Biological Systems*, B. Chance et al., Eds. Academic Press, 1979, p. 627.
[37] J. O. ALBEN, D. BEECE, S. F. BOWNE, L. EISENSTEIN, H. FRAUENFELDER, D. GOOD, M. C. MARDEN, P. P. MOH, L. REINISCH, A. H. REYNOLDS, and K. T. YUE, Phys. Rev. Letters *44*, 1157 (1980).
[38] R. H. AUSTIN, K. BEESON, L. EISENSTEIN, H. FRAUENFELDER, I. C. GUNSALUS, and V. P. MARSHALL, Science *181*, 541 (1973).
[39] B. B. MANDELBROT, *The Fractal Geometry of Nature*. W. H. Freeman, 1982.
[40] W. WEBER, Götting. Gel. Anz., p. 8 (1835), Annalen der Physik und Chemie (Poggendorf) *34*, 247 (1835).
[41] R. KOHLRAUSCH, Annalen der Physik und Chemie (Poggendorf) *2–72*, 353 (1847); *2–91*, 56, 179 (1854).
[42] R. H. AUSTIN, K. BEESON, L. EISENSTEIN, H. FRAUENFELDER, I. C. GUNSALUS, and V. P. MARSHALL, Phys. Rev. Letters *32*, 403 (1974).
[43] H. FRAUENFELDER, G. A. PETSKO, and D. TSERNOGLOU, Nature *280*, 558 (1979).
[44] P. DEBYE, Verh. Deut. Physik. Ges. *15*, 738 (1913).
[45] I. WALLER, Z. Physik *17*, 398 (1923); Ann. Physik *83*, 152 (1927).
[46] B. T. M. WILLIS and A. W. PRYOR, *Thermal Vibrations in Crystallography*, Cambridge University Press, 1975.
[47] H. HARTMANN, F. PARAK, W. STEIGEMANN, G. A. PETSKO, D. RINGE PONZI, and H. FRAUENFELDER, Proc. Natl. Acad. Sci. USA *79*, 4967 (1982).
[48] P. J. ARTYMIUK, C. C. F. BLAKE, D. E. P. GRACE, S. J. OATLEY, D. C. PHILLIPS, and K. J. E. STERNBERG, Nature *280*, 563 (1979).
[49] G. A. PETSKO and D. RINGE, Annu. Rev. Biophys. Bioeng. *13* (1984).
[50] G. CARERI, P. FASELLA, and E. GRATTON, CRC Crit. Rev. Biochem. *3*, 141 (1975).
[51] F. R. N. GURD and T. M. ROTHGEB, Adv. Protein Chem. *33*, 73 (1979).
[52] P. DEBRUNNER and H. FRAUENFELDER, Annu. Rev. Phys. Chem. *33*, 283 (1982).
[53] *Mobility and Function in Proteins and Nucleic Acids*. Ciba Foundation Symp. 93, Pitman, London, 1983.
[54] *Structure and Dynamics: Nucleic Acids and Proteins*. E. Clementi and R. H. Sharma, Eds. Adenine Press, 1983.
[55] M. KARPLUS and J. A. MCCAMMON, CRC Crit. Rev. Biochem. *9*, 293 (1981).
[56] M. KARPLUS and J. A. MCCAMMON, Ann. Rev. Biochem. *53*, 263 (1983).
[57] J. A. MCCAMMON and M. KARPLUS, Acc. Chem. Res. *16*, 187 (1983).
[58] J. A. MCCAMMON, *Reports of Progress in Physics* (1983).
[59] H. FRAUENFELDER, in *Structure, Dynamics, Interactions and Dynamics of Biological Macromolecules*. C. Hélène, Ed., Reidel 1983, p. 227.
[60] M. F. PERUTZ and F. S. MATTHEWS, J. Mol. Biol. *21*, 199 (1966).
[61] W. DOSTER, Biophys. Chem. *17*, 97 (1983).