Proteins as Paradigms of Complex Systems

P. W. Fenimore[a], Hans Frauenfelder[†a], Robert D. Young[b]

[a]Center for Nonlinear Studies, MSB258, Los Alamos National Laboratory, Los Alamos, NM 87545;

[b]Department of Physics and Astronomy, Northern Arizona University, Flagstaff AZ 86011-6010

ABSTRACT

The science of complexity has moved to center stage within the past few decades. Complex systems range from glasses to the immune system and the brain. Glasses are too simple to possess all aspects of complexity; brains are too complex to expose common concepts and laws of complexity. Proteins, however, are systems where many concepts and laws of complexity can be explored experimentally, theoretically, and computationally. Such studies have elucidated crucial aspects. The energy landscape has emerged as one central concept; it describes the free energy of a system as a function of temperature and the coordinates of all relevant atoms. A second concept is that of fluctuations. Without fluctuations, proteins would be dead and life impossible. A third concept is slaving. Proteins are not isolated systems; they are embedded in cells and membranes. Slaving arises when the fluctuations in the surroundings of a protein dominate many of the motions of the protein proper.

Keywords: Complexity, proteins, energy landscape, fluctuations, slaving.

1. INTRODUCTION

Complexity is difficult to define, but it is clear that the important systems in our world, from biology to the Internet and social networks, are complex. Do such systems share common properties and are they governed by universal concepts and laws? Proteins, or biomolecules in general, are good choices for studying complexity. Here we use results from one particular protein, myoglobin (Mb), the hydrogen atom of biology. We show that essential properties of complex systems, namely a large number of states, the existence of multiple energy, entropy, and time scales, the presence of fluctuations and relaxations, and their sensitivity to the environment, can all be explored by using Mb. The lessons learned from Mb can be used to understand protein families other than heme proteins, as will be discussed in another contribution.[1]

The physicists approach to biological systems is shaped by past experience. In many fields of physics, the exploration involved structure, energy levels, and dynamics. Of course, progress is never linear and structure, energy levels, and dynamics are nearly always intertwined. In the H atom, the steps involved the Balmer series, the Bohr atom, and quantum mechanics. In proteins, the three steps can also be recognized, but a fourth one, function, is unique to living systems. We begin by sketching the structural aspects of proteins.

Proteins are constructed from twenty different building blocks, amino acids[2] [3]. All amino acids have the same backbone, N - C_α - C. Sidechains, attached to the C_α, differ and give the amino acids their characteristic chemical and physical properties. Directed by the DNA, of the order of a few hundred amino acids are linked covalently by peptide bonds. In the proper environment the linear polypeptide chain folds into the tertiary structure, which is often globular. Textbooks usually give the impression that the tertiary structure is static and unique, much like an aperiodic crystal.[4] This impression is wrong. Proteins are

---

[†] Frauenfelder@lanl.gov. phone +1 505 665 2547. fax +1 505 665 2659.

dynamic systems that can assume a large number of different conformations and that move rapidly from conformation to conformation.

Myoglobin is built from 153 amino acids and its size is about 2x3x4 nm^3. The polypeptide chain encloses a porphyrin group with an iron atom at its center. Mb occurs mainly in muscles where it gives rise to the red color. A major role of Mb is the storage of dioxygen, but it is also involved in the enzymatics of nitrogen oxide.[5]

Fig. 1 gives two views of Mb.[6] Fig 1a shows Mb as visualized by most biochemists and crystallographers. In this view, Mb indeed looks like an aperiodic crystal. Every atom is in its proper place and no hole is visible that would permit the entry or exit of small molecules like O_2. Fig. 1b presents a physicist's view, a schematic cross section through Mb. The porphyrin ring, the iron atom, and some cavities are recognizable. O_2, NO, and CO bind at the iron atom, denoted by Fe.
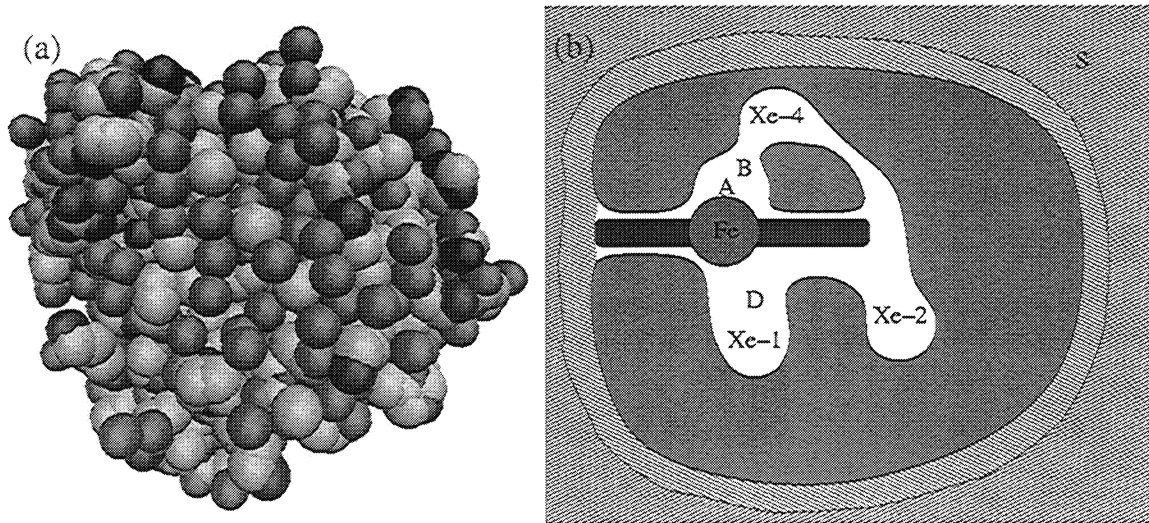

Fig. 1. Two views of myoglobin (Mb) from roughly the same orientation. (a) An atomistic view of Mb. The positions of the atoms are deduced from X-ray measurements. H atoms are not shown. The atoms are nearly close-packed, leaving no room for entry or exit of the ligand in this static structure. (b) A schematic cross section through Mb. The folded polypeptide chain (green) is surrounded by a hydration shell (inner yellow), a thin layer of water molecules, less than about 0.7 nm thick. The red ball is an Fe atom, the blue region is the phorphyrin, and the outer gold region is the bulk solvent (S). The Fe atom is tethered to the peptide on the bottom. Sites A, B, and D are explained in the text.

2. THE ENERGY (CONFORMATION) LANDSCAPE[7] [8]

2.1 Proteins exist in many conformations

Fig. 1a gives a misleading picture of proteins, because a protein can assume a large number of somewhat different conformations or conformational substates (CS). The protein molecules in a protein sample thus all have the same composition, but have a wide range of different structures (conformations). At low temperature, each protein molecule remains fixed in a particular CS; at ambient temperatures it fluctuates among all accessible CS. Clear evidence for the existence of CS came from flash photolysis studies with carboxymyoglobin, MbCO.[9] In these experiments, CO is initially bound to the heme iron in the position denoted by A in Fig. 1b. The sample is placed in a cryostat and the Fe=CO bond is broken by a laser flash. The CO moves into the heme pocket to a position denoted by B in Fig.1b. At temperatures below about 160 K, the CO cannot escape from the heme pocket and rebinds to the iron. Surprisingly, the rebinding is highly nonexponential in time and can be approximated by a power law,

$$N(t) = (1 + kt)^{-n} .\qquad\qquad (1)$$

Here N(t) is the fraction of Mb molecules that have not rebound a CO at the time t after the flash. The rate coefficient k and the exponent n change smoothly with temperature. The nonexponential time dependence could be caused either by a homogeneous ensemble with all protein molecules being identical and having the same nonexponential rebinding, or by an inhomogeneous ensemble with each Mb molecule having a different rebinding rate. Multiple-flash experiments, the poor man's single molecule technique, prove the second alternative to be correct.[9] The observed kinetics, Eq. (1), can then be explained by postulating that the enthalpy barrier, H_{BA}, between B and A is not unique, but is given by a distribution. Many experiments support this explanation.[10][11] The existence of different barriers in different Mb molecules leads directly to the concepts of CS and of an energy landscape. The names came later, but the concepts are already clear in Fig. 2, adapted from ref. 9. In this figure, different groups of substates are labeled by their barrier heights H_{BA}. The barriers between CS are represented by the spikes.
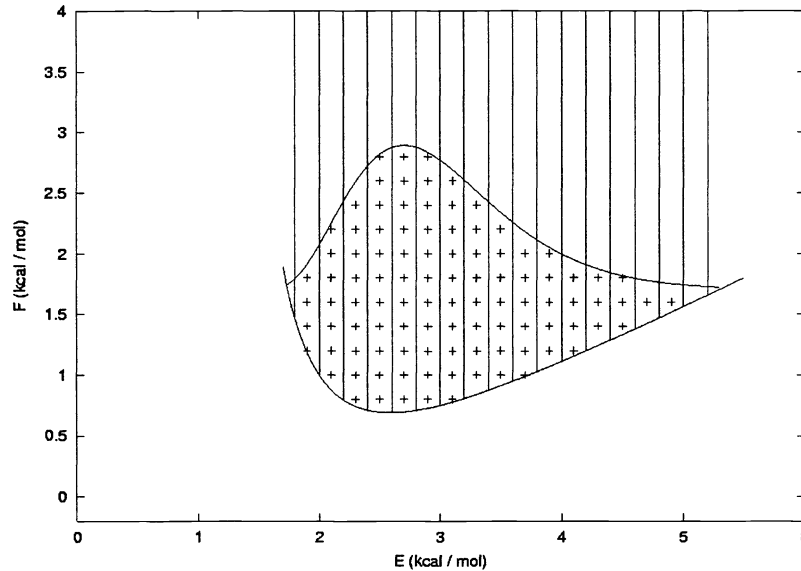

Fig. 2. Conformational substates and energy landscape.[9] The figure, taken from an early paper, presents only a one-dimensional cross section. Barriers between substates are shown as spikes. The upper curve is Boltzmann distributed at 200K and the symbols fill the enclosed area.

Fig. 2 is far too simple because it presents the CS in a one-dimensional plot. In reality, the CS and the energy landscape live in a high-dimensional space. A particular substate CS is characterized by the coordinates of each atom of the protein, the hydration shell, and part of the surrounding.[12][13] These coordinates can be looked at as describing a vector c in the conformational hyperspace. An attempt at explaining the concept is made in Fig.3, where the hyperspace is projected onto two dimensions. A few of the more than 3000 conformational axes are drawn. Each point represents one particular CS. The energy landscape at a given temperature is given by the free energy as a function of the vector c. An ensemble of Mb molecules is a swarm of points in the conformation space, similar to a mosquito cloud in real space. The number of CS is not known, but it most likely far exceeds 10^{+10}. Because proteins are dynamic systems, the swarm is not stationary, but proteins move from CS to CS at ambient temperatures, just like mosquitos. The swarm of substates is not featureless, but is organized hierarchically.[14] At present, the organization is only imperfectly known, but some general features are sketched in Fig. 4. Three different types of CS can be distinguished, taxonomic, statistical, and local.
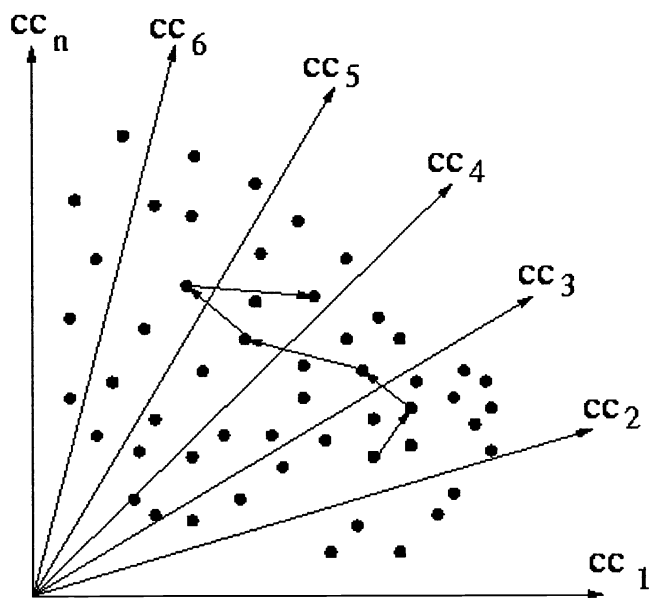
Fig. 3. A projection of the hyperspace characterizing a protein in a given conformation. Each point represents a vector **c** describing one particular conformation or conformational substate (CS). The axes cc_1 through cc_n are roughly 3000 coordinates for myoglobin.
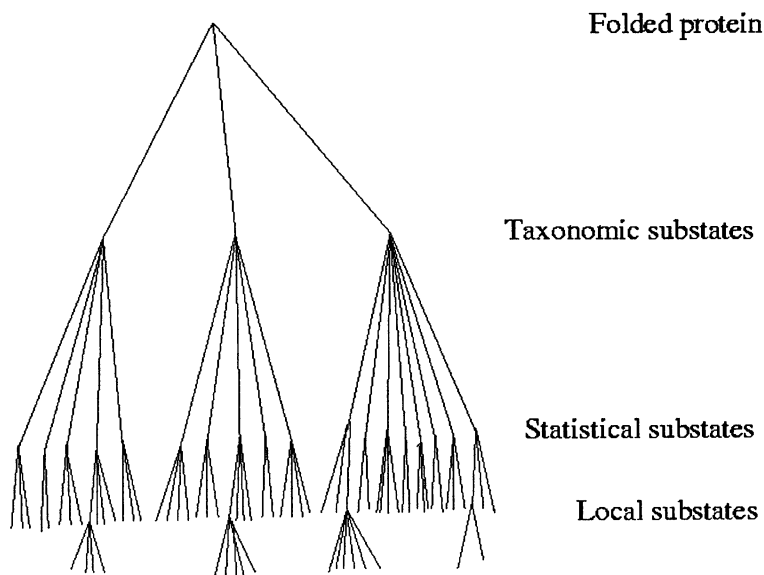

Fig. 4. The energy landscape of Mb, including the hydration shell and a part of the solvent, is organized hierarchically. Three different types of CS are known at present: taxonomic, statistical, and local.

1.2 Taxonomic substates (isomers)

In Mb, the top tier shown in Fig. 4 consists of three (possibly four) different conformations, denoted A_0, A_1, (A_2), and A_3. They are characterized by the C-O stretch frequencies of the bound CO[15] and are called "taxonomic" because they can be characterized individually. The structures of A_0 and A_1, in particular, have been determined by X-ray diffraction.[16] The main difference between the two, in addition to smaller changes, is the position of histidine 64. This amino acid is inside the heme pocket in A_1, but extends partly into the solvent in A_0. The two CS also have different functions: A_1 stores oxygen, while A_0 catalyzes NO.[5] The structure of A_3 and its function are not known. We expect that essentially all proteins will possess taxonomic substates. Prions may be a particularly dramatic example; antibodies may also use them.[17]

1.3 Statistical substates

Each taxonomic substate rebinds CO nonexponentially in time at low temperatures.[18] Thus a very large number of CS reside within each taxonomic substate. We call them "statistical" because they cannot be studied and described individually. The low-temperature data imply that statistical substates within a given taxonomic substate perform the same function, but with different rates.

1.4 Local substates

Within each statistical substate a protein can still assume slightly different conformations. We call them. "local substates". These substates have mainly been studied by laser hole burning.[19] The connection of these CS with structure and function is not known. It is also not known if they are important for function.

3. PROTEIN DYNAMICS

Proteins are dynamic systems. Within a given CS they can vibrate, but they can also jump from CS to CS. Some of these jumps depend strongly on external parameters such as temperature, pH, salts, and other solvent properties. In a glassy solvent, only motions among the local substates occur with measurable rates well below the glass temperature T_g. Near T_g very slow motions among the statistical substates set in. At ambient temperature, these motions are of the order of 10^{10} s^{-1} in a liquid solvent. In a rigid environment such as a crystal or a glass, most of the motions are blocked.

Some characteristic features of proteins are similar to those of super-cooled liquids and glasses.[20] Protein properties can usually not be described by single values; distributions must be used. The temperature dependence of some motions do not follow the Arrhenius relation,

$$k(T) = A \exp[-H/k_B T], \tag{2}$$

but can be described by the Vogel-Tammann-Fulcher relation,

$$k(T) = A_v \exp[-H_v/k_B(T-T_0)]. \tag{3}$$

Here k_B is the Boltzmann constant and H and H_v are activation enthalpies. T_0 is the temperature where $k(T)$ extrapolates to zero. As already pointed out, the time dependence of many motions is nonexponential. There are, however, differences between glasses and proteins. Glasses and glass-forming liquids are not influenced or governed by the environment, but the effect of the solvent and the hydration shell on protein dynamics is profound. We will now turn to this aspect of protein dynamics.

4. SLAVING[21]

4.1 Protein motions

Proteins are machines that interact with the environment. We can expect different types of motions, some independent of the environment and some governed by it. To get some insight into these possibilities, we turn again to Mb and to Fig.1b. We select a number of processes and characterize them by the temperature dependence, $k(T)$, of their rate coefficients. The dielectric fluctuation rate of the solvent[22] (for example glycerol/water, 3/1 v/v.), essentially the tumbling rate of water molecules, is denoted by k_{diel}. Other rate coefficients are measured by following a CO molecule after photodissociation [9] [23] and can be described by referring to Fig. 1b. We consider three processes, the covalent binding step B$\rightarrow$ A with average rate coefficient k_{BA} , the passage of a CO molecule from D to A (k_{DA}), and the exit of the CO from Mb (k_{exit}). We also select transitions between the taxonomic substates A_0, A_1, and A_3, denoted by k_{01} and k_{13}.[18] Rather than plotting the rate coefficients, we show in Fig. 5 log (k/k_{diel}) versus $10^3/T$. Fig. 5 displays two very different classes of motions, slaved and non-slaved. In 3/1 v/v glycerol/water, k_{diel} changes from about 10^{11} s^{-1} to 10^{-4} s^{-1} between 300 K and 160 K, 15 orders of magnitude. The slaved motions, namely the exit of CO and the transitions between the taxonomic substates follow the solvent fluctuations over the entire temperature range where they can be observed. The temperature dependence of these motions cannot be fitted by the Arrhenius law, Eq. (2). If one tries such a fit, the preexponential factor becomes unphysically

large. The empirical V-T-F relation, Eq. (3), however, fits the data well. The nonslaved motions, in this case the covalent binding and the motion of the CO molecule through the protein, do not follow k_{diel} and hence are not coupled to bulk solvent fluctuations. They can be fitted to the Arrhenius relation, Eq. (2), with reasonable values of the preexponential factor A. The nonslaving can be appreciated intuitively. The core of the protein is solid-like[24] and shields the motion of the CO from the solvent dynamics. The slaving, however, leads to a problem, to be discussed next.
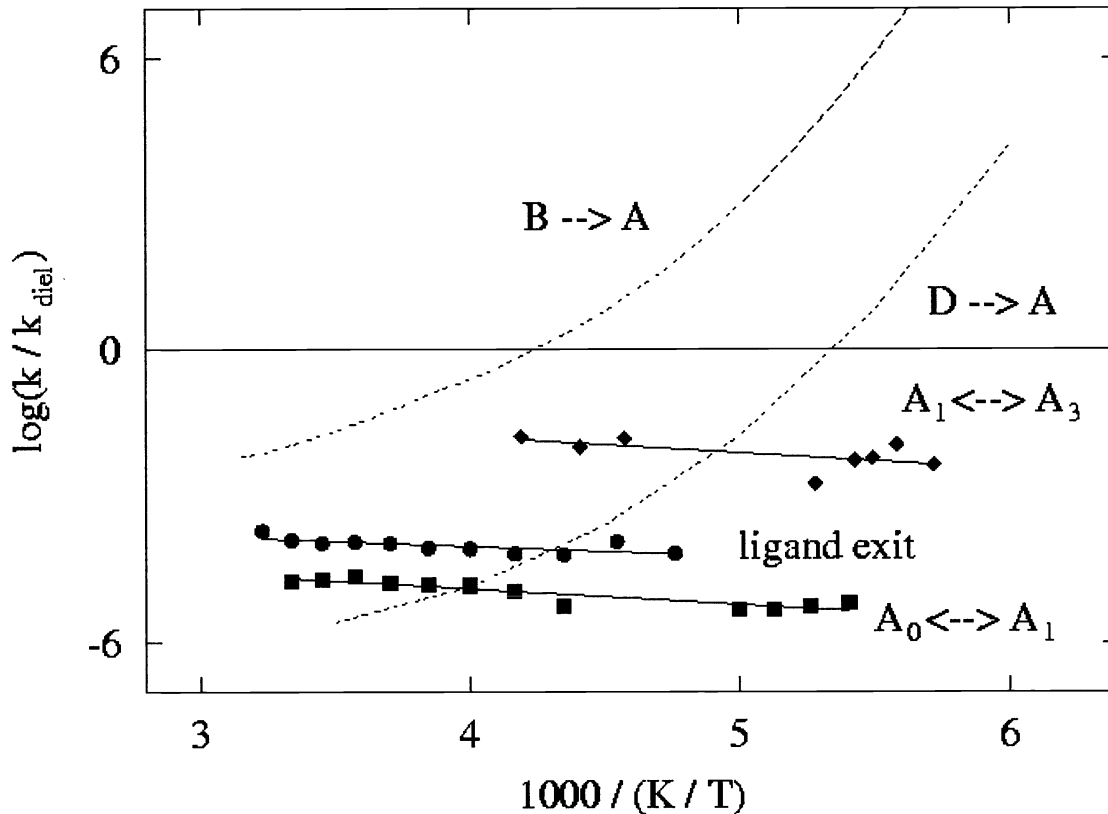

Fig. 5. Slaved and nonslaved motions.The plot shows the ratio of the rate of various processes in proteins to the Debye relaxation rate of the solvent. The slaved motions, for instance the fluctuations between the taxonomic substates A_0 and A_1 or A_1 and A_3, or the exit into the solvent, follow the dielectric relaxation rate in the solvent over many orders of magnitude. Nonslaved motions, such as the transitions from D to A, or the rate of covalent binding of a CO at the heme, B to A, are not affected by the solvent fluctuations.

4.2 Entropic control of slaved motions.
The rate coefficient for slaved motions can be written as

$$k_{slaved} = k_{diel}(T)/n(T), \qquad (4)$$

where the coefficient $n(T)$ depends only weakly on T compared to k_{slaved} and k_{diel}. Why can $n(T)$ be as large as 10^5 or, in other words, what causes the hierarchy of motions seen in Fig. 5 if there are no internal enthalpy barriers? If the barriers that slow for instance the motion of the distal histidine in and out of the heme pocket are not enthalpic, they must be entropic. How can this feature be explained and described? We propose the model shown in Fig. 6.
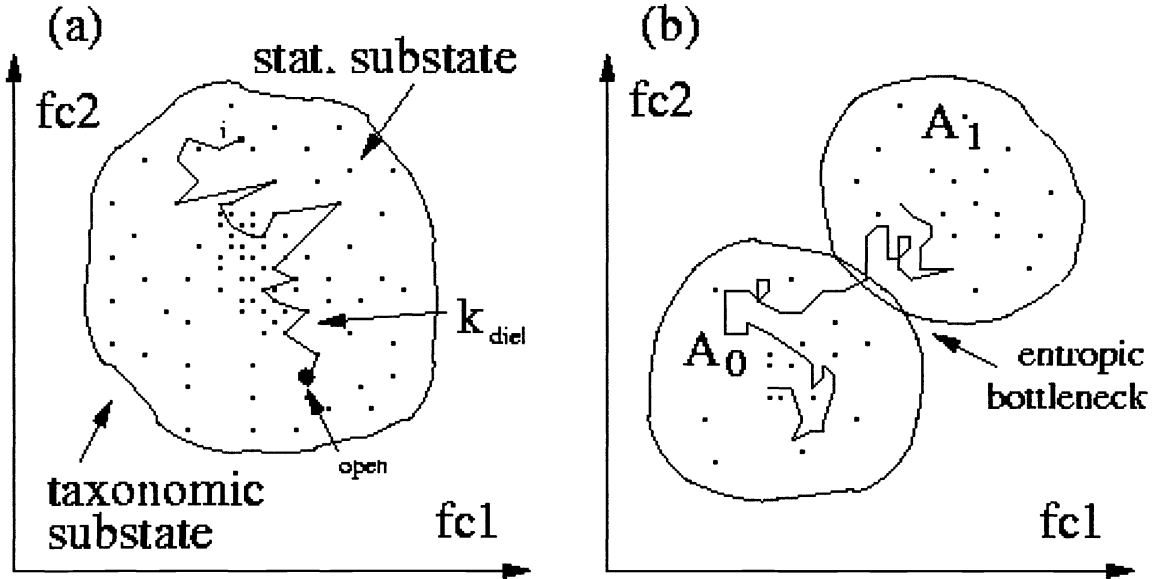
Fig. 6 (a) Two-dimensional cross section through the energy landscape of Mb. The functional coordinates fc1 and fc2 describe the transitions between selected substates.[25] Each protein in a given taxonomic substate i performs a random walk in the hyperspace of the protein conformations. A small subset, denoted by "open" represents protein substates that have an open path between the protein interior and the solvent. A protein in a substate i makes a random walk until it reaches the open substates so that a ligand can enter or exit. (b) Two taxonomic substates, for instance A_0 and A_1. The transition between the two taxonomic substates can occur only through an entropic bottleneck.

The figure represents a two-dimensional cross section through the hyperspace of all statistical substates in a taxonomic substate. The functional coordinates fc1 and fc2 describe the transitions between selected substates.[25] A protein in a given taxonomic substate i performs a random walk in the hyperspace of the protein conformations[26] till it reaches the subset, denoted by "open" of the substates in which the protein has an open path between the inside and the solvent A ligand can then enter or exit. For simplicity we assume that each individual step between adjacent statistical substates occurs with the rate coefficient k_{diel}. Similarly the transition between two taxonomic substates involves random walk through an entropic barrier.[27] This model explains the large value of n(T) in Eq. (4) and the nonexponential time dependence of many processes.

4.3 Fluctuations are crucial

Because the slaved motions crucial for function follow k_{diel} it is obvious that fluctuations in the solvent are essential for protein function. Moreover, the fluctuations in the solvent dominate the protein fluctuations. Fluctuations in the enthalpy H, the volume V, and the electric dipole moment M can be responsible for the slaving. The fluctuations are given by

$$<(\Delta H)^2> = k_B T^2 m C_P, \qquad <(\Delta V)^2> = k_B T V B_T, \qquad <(\Delta M)^2> = k_B T V \varepsilon_0 \chi. \qquad (5)$$

Here C_P is the heat capacity at constant pressure, B_T the isothermal bulk compressibility, χ the polarizability, ε_0 the permittivity coefficient of vacuum, and m and V are the mass and volume of the fluctuating system.. The susceptibilities of the solvent are larger than those of proteins. One can therefore expect that the solvent fluctuations indeed control the protein fluctuations. It is not yet clear which of the fluctuations dominate.

5. CONCLUSIONS

The experiments sketched here, and the corresponding computations and theories, demonstrate that proteins, in particular myoglobin, possess the crucial properties that permit studies of complexity: the

existence of a multitude of hierarchical states; the presence of multiple time, energy, entropy, and length scales, and the interaction of the complex system with its environment. Of course we are only at a beginning. Remember that the elucidation of even such a "simple" system as the hydrogen atom took a very long time, involving Balmer, Bohr, de Broglie, Schroedinger, Pauli, Heisenberg, Dirac, and many others. Proteins are much more complex than the hydrogen atom. Moreover, they are involved in the functions of life. We can therefore expect that it will take a long time till proteins are understood at the level of atoms, nuclei, and solids. But it is an exciting challenge for physicists to work in collaboration with biologists, biochemists, and chemists to come to a quantitative description and understanding of proteins and their interactions and functions. Understanding fluctuations ("noise") will be crucial in this endeavor.

REFERENCES

[1] B. H. McMahon, P. W. Fenimore and M. Labute. "Noise in biological systems." In this issue.

[2] L. Stryer, *Biochemistry.* W. H. Freeman, New York, 1995.

[3] C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland, New York, 1999.

[4] E. Schroedinger, *What is Life?* Cambridge University Press, 1944.

[5] H. Frauenfelder, B. H. McMahon, R. H. Austin, K. Chu, and J. T. Groves. "The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin." *Proc. Natl. Acad. Sci. USA* **98**, 2370-2374, 2001.

[6] I. Schlichting, J. Berendzen, G. N. Phillips, R. M. Sweet. "Crystal structure of photolysed carbonmonoxy-myoglobin." *Nature* **371,** 808-812, 1994. See also <http://www.rcbs.org/>.

[7] H. Frauenfelder, S. Sligar, and P. G. Wolynes. "The energy landscapes and motions of proteins." *Science* **254**, 1598-1603, 1991.

[8] H. Frauenfelder, A. R. Bishop, A. Garcia, A. Perelson, P. Schuster, D. Sherrington, and P. J. Swart, eds., *Landscape Paradigms in Physics and Biology. Concepts, Structures, and Dynamics*, North-Holland, Amsterdam, 1997.

[9] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, "Dynamics of Ligand Binding to Myoglobin." *Biochemistry* **14**, 5355-5373, 1975.

[10] H. Frauenfelder, F. Parak, and R. D. Young. "Conformational substates in proteins." *Ann. Rev. Biophys. Biophys. Chem.* **17**, 451-479, 1988.

[11] G. U. Nienhaus and R. D. Young. "Protein Dynamics." *Encyclopedia of Applied Physics* **15**, 163-183, 1996.

[12] F. H. Stillinger. "A topographic view of supercooled liquids and glass formation." *Science* **267**, 1935-1939, 1995.

[13] P. G. Debenedetti and F. H. Stillinger. "Supercooled liquids and the glass transition." *Nature* **410**, 259-267, 2001.

[14] A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young. "Protein states and proteinquakes." *Proc. Natl. Acad. Sci. USA* **82**, 5000-5004, 1985.

[15] 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. "Infrared spectroscopy of photodissociated carboxymyoglobin at low temperatures." *Proc. Natl. Acad. Sci. USA* **79**, 3744-3748, 1982.

[16] F. Yang and G. N. Phillips Jr. "Crystal structures of CO-, deoxy- and met-myoglobins at various pH values." *J. Mol. Biol.* **256**, 762-774, 1996.

[17] L. C. James, P. Roversi, and D. S. Tawfik. "Antibody multispecificity mediated by conformational diversity." *Science* **299**, 1362-1367, 2003.

[18] J. B. Johnson, D. C. Lamb, H. Frauenfelder, J. D. Mueller, B. McMahon, G. U. Nienhaus, and R. D. Young. "Ligand binding to heme proteins. VI. Interconversion of taxonomic substates in carbonmonoxymyoglobin." *Biophys. J.* **71**, 1563-1573, 1996.

[19] D. T. Thorn-Leeson, D. A. Wiersma, K. Fritsch, and J. Friedrich. "The energy landscape of myoglobin: An optical study." *J. Phys. Chem. B* **101**, 6331-6340, 1997.

[20] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach, A. H. Xie, and R. D. Young. "Glassy behavior of a protein." *Phys. Rev. Letters* **62**, 1916-1919, 1989.

[21] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak. "Slaving: Solvent fluctuations dominate protein dynamics and functions." *Proc. Natl. Acad. Sci. USA* **99**, 16047-16051, 2002.

[22] J. R. Huck, G. A. Noyel, and L. J. Jorat. "Dielectric properties of supercooled glycerol-water solutions." *IEEE Transactions on Electrical Insulation*, **23**, 627-637, 1988.

[23] T. Kleinert, W. Doster, H. Leyser, W. Petry, V. Schwarz, and M. Settles. "Solvent composition and viscosity effects on the kinetics of CO binding to horse myoglobin." *Biochemistry* **37**, 717-733, 1998.

[24] H. Frauenfelder, G. A. Petsko, and D. Tsernoglou. "Temperature-dependent X-ray diffraction as a probe of protein structural dynamics." *Nature* **280**, 558-563, 1979.

[25] A. E. Garcia and G. Hummer, "Conformational dynamics of cytochrome c: correlation to hydrogen exchange." *Proteins: Struct. Funct. Genetics* **36**, 175-191, 1999.

[26] W. Nadler and K. Schulten, "Theory of Mössbauer spectra of proteins fluctuating between conformational substates." *Proc. Natl. Acad. Sci. USA* **81**, 5719-5723, 1984.

[27] H.-X. Zhou and R. Zwanzig. "A rate process with an entropy barrier" *J. Chem. Phys.* **94**, 6147-6152 (1991).