

# Proteins: Paradigms of complexity

Hans Frauenfelder\*

Center for Nonlinear Studies, MS B258, Los Alamos National Laboratory, Los Alamos, NM 87545

**Proteins are the working machines of living systems. Directed by the DNA, of the order of a few hundred building blocks, selected from 20 different amino acids, are covalently linked into a linear polypeptide chain. In the proper environment, the chain folds into the working protein, often a globule of linear dimensions of a few nanometers. The biologist considers proteins units from which living systems are built. Many physical scientists look at them as systems in which the laws of complexity can be studied better than anywhere else. Some of the results of such studies will be sketched.**

“The history of physics is also a history of concepts. For an understanding of the phenomena, the first condition is the introduction of adequate concepts.”

Pauli to Heisenberg

**D**uring the past few decades, the general attitude of many physicists has undergone a sea change. It used to be that physicists loved simple systems, tried to understand them in the simplest terms, and often looked down on fields like chemistry and biology, where complexity reigned. No longer. Now many physicists are studying complex nonlinear systems and discover to their surprise how beautiful the problems are and how rewarding the interaction with biologists and chemists can be. Here I try to give a brief description of what proteins are, what they do, how complex they are, and why they are nearly ideal systems for the study of complexity. Whether this complexity can be called “self-organized” is a question of semantics.

## Complexity

What is complexity? Which systems are complex? What are the crucial concepts in complex systems?

A system can be called complex if it can assume a large number of states or conformations and if it can carry information. One often hears even biologists talk about “astronomically large numbers.” Astronomically large numbers are actually very small compared with biological numbers. They are of the order of  $10^{200}$  or  $\log n_{\text{astro}} \approx 200$ . Consider now DNA. It is built from four different units (bases) and may contain  $10^9$  bases. The number of conceivable DNA is therefore  $\log n_{\text{bio}} \approx 10^8 \gg \log n_{\text{astro}}$ . The number of possible protein is of the order of  $\log n_{\text{prot}} \gg 200$ . Even the number of states that an individual protein can assume is very large. Biological systems clearly also carry information. Hence proteins, and in general biological systems, are complex.

## Proteins

Proteins are built from 20 different amino acids (1, 2). Directed by the DNA, of the order of a few hundred of these building blocks are linked together into a linear polypeptide chain. The order in which the different amino acids are inserted determines structure, function, and dynamics. In the proper solvent, the chain folds into a compact structure that is often globular and that has linear dimensions of a few nanometers. Proteins perform essentially all functions in biological systems.

The textbook picture of a protein is clear: The folded structure is unique; each atom is in its proper place. The pictures obtained by x-ray diffraction techniques appear to support this—at first sight—appealing situation. Such proteins would be aptly char-

acterized by Schrödinger’s words, “aperiodic crystals” (3). Reality, however, is different. Proteins are dynamic and not static systems (4), and they must perform motions to execute their functions. Motions are possible only if a given protein can assume a large number of somewhat different conformations, for instance with open and closed channels. Actually, the motions involve the atoms not just of the protein itself but also of the hydration shell, a layer of water surrounding the protein. The structure and dynamics of the protein and the hydration shell can be characterized by the energy or conformation landscape.

## The Energy Landscape

The energy landscape is a construct in  $3N$  dimensions, where  $N$  is the number of atoms in the protein and the hydration shell (5, 6). The energy landscape contains valleys and saddle points between valleys. We call each valley a conformational substate. A substate describes the structure of the entire protein, because it characterizes the positions of all atoms. Transitions between substates correspond to protein motions. Unfortunately, it is difficult to visualize the landscape, because it lives in a hyperspace. One- or two-dimensional cross sections can give a misleading impression. One difference between such a representation and the complete landscape is the path between two substates. In the low-dimensional cross section, it may appear that the protein has to overcome many saddles, whereas in reality only one or two steps may be necessary.

One goal of the physics approach to proteins is the exploration of the energy landscape. In no protein is the entire landscape known. This state is not surprising if one contemplates how many years it took to determine the energy levels of complex nuclei or atoms—systems that are far simpler than proteins. Nevertheless, a number of features have emerged, mainly from studies of myoglobin (5, 7). One important feature is that the energy landscape is organized in a hierarchy, with valleys within valleys within valleys. In other words, the substates are organized in a series of tiers. The different tiers are distinguished by the (average) size of the barriers separating them. At the top of the hierarchy, in tier 0, are the taxonomic substates. They are small in number and are different enough that their properties can be studied individually. Myoglobin, for instance, has three taxonomic substates, called  $A_0$ ,  $A_1$ , and  $A_3$ . At physiological temperatures, the three substates interconvert rapidly and are in thermal equilibrium. The equilibrium can be shifted by external agents, for instance pH, lactate, or pressure. Each taxonomic substate contains a very large number of substates of tier 1, or statistical substates. Different statistical substates have, in general, different rates for a particular reaction and slightly different wavelengths of some transitions. At high temperatures, transitions among substates of tiers 0 and 1 are faster than, say, micro- or nanoseconds. At low temperatures, say below 100 K, transi-

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Self-Organized Complexity in the Physical, Biological, and Social Sciences,” held March 23–24, 2001, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

\*E-mail: frauenfelder@lanl.gov.

tions among the substates of tier 0 and 1 are essentially absent, and the existence of substates can be recognized, for instance, by the facts that reactions become nonexponential in time (8) and that “holes” can be burned into inhomogeneous spectral lines (9). Each statistical substate contains substates with lower barriers. They are small in number, can be called “few-level substates,” and may be similar to such levels in glasses. Transitions between such substates can occur even in the millikelvin region.

Some features of the energy landscape of myoglobin are thus clear, but the details are far from being known. Moreover, the connections among substates, structure, and dynamics are far from understood.

### The Energy Landscape and Function

One connection of the energy landscape to function is obvious: Transitions among substates are protein motions, and protein motions are essential for protein function. The connection is deeper, however. As stated earlier, myoglobin has three taxonomic substates. Consider  $A_1$  and  $A_0$ .  $A_1$  dominates at high pH,  $A_0$  at low pH. The function of  $A_1$  is described in every textbook;

it is the storage of dioxygen (1). It turns out that  $A_0$  may have a very different function, namely that it is involved in nitrite reactions (10). The “simple” protein myoglobin thus may actually be an allosteric enzyme, and the taxonomic substates may be intimately involved in this function. This recognition may open the way to search for such allosteries in other proteins and to find out whether protein networks are involved. But what is the role of the other tiers in function? Low-temperature experiments (8) prove that different substates of tier 1 perform the same binding reaction but with different rates. Thus tier 0 may determine the function, tier 1 the reaction rate. The function of lower tiers is not yet known.

### Final Remark

This brief sketch should make it clear that proteins are truly complex systems and that the complexity can be described through the energy landscape. The complexity has arisen through evolution. The structure and function of proteins are coded in the DNA. Within the living system, proteins are part of a complex proteins network (11), and the complex interactions in the network may control the actual function. Can this be called self-organized?

1. Stryer, L. (1995) *Biochemistry* (Freeman, New York).
2. Fersht, A. (1999) *Structure and Mechanism in Protein Science* (Freeman, New York).
3. Schrödinger, E. (1944) *What Is Life?* (Cambridge Univ. Press, Cambridge, U.K.).
4. Linderstrøm-Lang, K. U. & Schellman, J. A. (1959) *Enzyme* **1**, 443–510.
5. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991) *Science* **254**, 1598–1603.
6. Frauenfelder, H. & McMahon, B. H. (2000) *Ann. Phys. (Leipzig)* **9**, 655–667.
7. Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E. T., Sauke, T. B., Shyamsunder, E. & Young, R. D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5000–5004.
8. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. (1975) *Biochemistry* **14**, 5355–5373.
9. Friedrich, J. (1995) *Methods Enzymol.* **246**, 226–259.
10. Frauenfelder, H., McMahon, B. H., Austin, R. H., Chu, K. & Groves, J. T. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 2370–2374.
11. Fields, S. (2001) *Science* **291**, 1221–1224.