found in *Scapharca i.* hemoglobin. Further lowering of the pH produces changes in the g_z region and the spectrum acquires the typical features described for hemoproteins in which the Fe-Im bond is strongly weakened or broken. The pK of this transition is 6.7. No obvious kinetic counterpart of the co-operative oxygen equilibrium binding curves is found in *Nassa m.* myoglobin. No co-operativity and no quickly reacting species are observed in CO association or dissociation kinetic experiments. Oxygen dissociation displays biphasic time courses. The pattern is not affected by the initial pO_2 and is observed also in oxygen pulse experiments. This shows that two sites, with different oxygen affinity, are operative in the molecule. In contrast with what observed in other hemoproteins, the kinetic control of co-operativity in oxygen binding appears to depend on an increase in the 'on' rate of the high affinity site, rather than on the usual decrease of the 'off' constant. This is another peculiarity of this dimeric myoglobin. The pH dependence of the two rates is characterised by two distinct pK values: 6.7 for the fast and 8.5 for the slow phase. Arrhenius plots show the same activation energy for the two kinetic components, indicating that the difference in their rate constants is of entropic origin.

As a final point, the amino acid sequence of *Nassa m.* globin shows 63% identity with the globin of *Busycon c.* and 46% with that of *Cerithidea r.*, both Prosobranchia molluscs like *Nassa m.* In contrast, sequence conservation is less than 20% when the comparison is done with the myoglobins of the sea snails belonging to the Opistobranchia sub-class, like *Aplysia l.*, while molluscs of this sub-class have myoglobins with a very high sequence homology. A comparison of *Aplysia l.* myoglobin aminoacid sequence with the primary structure of *Scapharca i.* dimeric hemoglobin and with that of human myoglobin shows percent conserved residues similar to what observed comparing *Aplysia l.* myoglobin and *Nassa m.* myoglobin sequences, as if the separation between two sub-classes of sea-snails and between these and obviously more distant organisms, such as the Arcid clam *Scapharca i.* and the mammal human, was about the same.

Acknowledgements. The results discussed here have been obtained in collaboration with: Massimo Coletta, University of Camerino; Giulietta Smulevich, Anna Rita Mantini and Mauro Poli, University of Florence; Marina Piscopo and Rosanna del Gaudio, University of Naples; Alessandro Desideri, University of Messina; Francesca Polizio, Second University of Rome 'Tor Vergata'; Paolo Ascenzi, University of Turin.

Proteins - Paradigms of complex systems

H. Frauenfelder

Los Alamos National Laboratory, Physics Division P-21, MS M715, Los Alamos (New Mexico 87545, USA), Fax + 1(505)6654507

Introduction

The past decade has seen an emerging interest of physicists in complexity. Complex systems range from glasses and spin glasses to the brain. To study the physics of complexity calls for a system that is large enough to be truly complex, yet small enough so that it can be understood and managed experimentally. Proteins satisfy these conditions and their exploration, over wide ranges in time, temperature, and external conditions provide an insight into some of the salient features of complexity. The most conspicuous aspect is the existence of a rough energy landscape; proteins exist in a dynamic equilibrium among a very large number of slightly different structures. This characteristic appears to be the dominant property that marks all complex systems. The experiments with proteins are beginning to expose some crucial attributes of the energy landscape and they may form the beginning of a quantitative physics of complexity.

Complex systems

Complexity, in one form or another, is all around us. Music is complex and so are languages, economies, societies, and the brain. For many years, physical scientists shied away from complex systems and were content or even smug in concentrating their work on 'simple' systems. Within the last few decades, even physicists have started looking at complex systems and have begun to search for unifying concepts and laws. Such a search can start with two questions:

1) which are the best systems in which to study complexity? and
2) how do we study complexity?

The 'simplest' complex systems are probably glasses and spin glasses. Glasses, of course, have been known for a few thousand years and have been studied in great detail[1]. Despite the large number of investigations, many problems remain unsolved. Spin glasses,

dilute alloys in which for instance iron atoms are embedded in gold, have also been explored for a considerable time, but still are not fully understood[2]. After glasses and spin glasses, biomolecules may well be the 'next-simplest' complex systems. In particular proteins may be the nearly ideal systems in which to study complexity.

Proteins

Proteins are the machines of life[3]. They are built from 20 different building blocks, amino acids. Of the order of one to two hundred amino acids are hooked together in the cell to form a linear polypeptide chain. This chain then folds, either spontaneously or with the help of other biomolecules (chaperones), into a compact, approximately globular molecule. The arrangement of the amino acids in the primary chain determines the final tertiary structure and also the function of the working protein. Structures and function of many proteins are described in detail in many biochemistry texts[3].

While the detailed structure of the protein is of the utmost importance for its function, two other properties are crucial for the discussion of complexity, namely aperiodicity (or disorder) and frustration.

Crystals are nearly perfectly periodic and this periodicity has important consequences for their properties, as can be seen in any text on condensed matter physics. Schrödinger[4] called proteins 'aperiodic crystals' before their structure was known and, with this name, he captured one of the essences of proteins. Proteins are truly aperiodic and this aperiodicity is caused by the aperiodic arrangement of the different amino acids in the primary sequence.

While aperiodicity is immediately obvious, frustration[5] is harder to understand and recognise. Two examples explain the concept. Consider first three spins, in which any two of them like to be anti-parallel. The system then is frustrated; the two arrangements ↑↑↓ and ↑↓↑ have equal energies and the result is a two-well potential for the energy of the system. The second example is NH_3. Here the N atom can be on either side of the three hydrogens and the result is again a double-well potential for the energy of the system. More generally, competing conditions lead to frustration. In proteins, different side chains may compete for the same space in the folded protein and proteins thus are frustrated. We can guess that most (or all) complex systems are frustrated. Certainly economies and societies are always subject to competing pressures and to frustration.

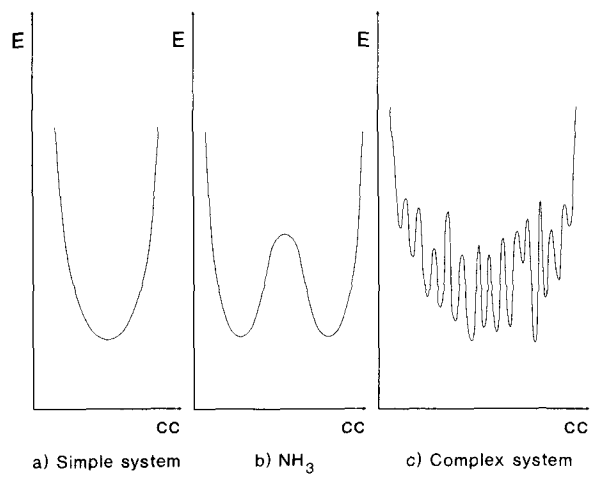

Figure 1. Energy level and energy landscape for a simple and for complex systems. cc is a conformational coordinate.

Aperiodicity (or disorder) and frustration together lead to a rough energy landscape, the property that may well be the most important unifying concept for complex systems. Much of the rest of this brief review is devoted to explaining this concept in more detail.

Even before delving into the concept of the energy landscape we can provide some arguments why proteins are good systems for the study of complexity:

1) Proteins are aperiodic and frustrated.

2) Proteins are large enough to be truly complex, but still small enough that we can hope for a thorough understanding (in time).

3) Proteins are the result of about 4 Gy of R&D and we hence can start from highly sophisticated systems, with a clear function. Since one of the goals is to understand the relation between complexity and function, this feature makes biomolecules better subjects than glasses or spin glasses.

4) Proteins can be modified nearly at will by genetic engineering. The effect of specific modifications on properties can therefore be studied.

5) Many proteins contain superb spectroscopic probes. If needed, additional probes can be inserted at exact locations.

6) The field is limitless; in principle, more than 10^{200} different amino acid combinations can be formed.

The approach

Progress in physics has often been based on analogies. The history of physics also teaches us which steps have led to an understanding of a particular field. In atomic physics, for instance, somewhat simplified, the main steps comprised the elucidation of the structure (Rutherford) and the energy levels (Balmer, Planck, Bohr), and the discovery of the dynamical laws (Heisenberg, Schrödinger). The advances in condensed matter, nuclear, and particle physics followed similar paths. Can these lessons from the physics of 'simple' systems help in discovering the concepts and laws of complexity? It turns out that some progress along the lines structure – energy – dynamics has indeed been made.

The structure of proteins

X-ray diffraction is the classical tool to study structures; it yields the electron distribution of the system under

investigation. Max von Laue, the discoverer of the tool, believed that it would be impossible to determine the structure of biomolecules[6]. Max Perutz[7] and John Kendrew[8], however, succeeded and solved the structure of two proteins, hemoglobin and myoglobin. Since then, structure determination has become a major industry and a large number of proteins are now known. Typical computer-produced pictures can be found in any text on biochemistry[3,9].

Texts actually give a somewhat misleading impression, because they depict proteins as rigid and completely determined structures, with each atom fixed at a well-defined position. As will become clear later, the X-ray structures are averages and deviations from these averages are critical for functions.

A point well known to protein crystallographers, but often surprising to physicists, is the fact that the average position of essentially every atom is known. It is therefore, at least in principle, possible to relate the detailed structure to properties.

The energy landscape

In a simple system such as an atom or a crystal, the ground state is unique and can be labelled unambiguously by a simple number, the ground state energy. In a frustrated, but not disordered system such as the three spins or NH_3 discussed earlier, the 'energy landscape' becomes a double well (fig. 1b) and the wells can be labelled by the structure of the system. In a complex system, however, the ground state must be described by a rough energy landscape, as sketched in figure 1c. Such a system can assume a large number of different structures and each specific structure corresponds to a minimum or energy valley in figure 1c. Figure 1c actually represents a one-dimensional cross section through a hyper-space; the complete description of the energy landscape requires a space of high dimensions.

The appearance of an energy landscape in proteins is unavoidable[10]. Frustration and aperiodicity together cause the existence of many minima that are only slightly different in energy. The first clear indication for the existence of a very large number of protein conformations occurred in studies of the binding of carbon monoxide and dioxygen to myoglobin after photodissociation at low temperatures[11]. Rebinding after flash photolysis was observed optically. Below about 160 K, only one rebinding process was seen, but the time course was highly non-exponential. The non-exponentiality was explained by assuming that the protein could assume a large number of somewhat different conformations, each with a different rebinding rate. Each of these conformations corresponds to a particular valley in figure 1c.

Proteins are machines; to work they usually must be able to assume at least two states, for instance a charged and a neutral one. In each of these states, there will be an energy landscape with a large number of minima. To distinguish them from the different states, they are called **conformational substates (cs)**. Thus each valley in figures 1c represents a conformational substate.

At low temperatures (below about 180 K), a given protein molecule will be frozen into a particular cs. At high temperatures, a given protein will jump from cs to cs - it will move. If the motion is rapid enough, the observed rebinding will become exponential in time.

Conformational substates

One of the first questions usually asked is: How can the existence of cs be reconciled with the apparently well-defined X-ray structures of proteins? The answer comes from a detailed consideration of the Debye-Waller factor[12]. X-ray diffraction provides two pieces of information on each non-hydrogen atom in a protein, the average position and the Debye-Waller factor which gives the mean-square displacement (msd). In a 'common' crystal, the msd is small and essentially the same for each atom. In a protein crystal, however, the msd is large and different for different atoms[13]. The Debye-Waller factor consequently supports the concept of cs by implying that each protein possesses a structure that differs somewhat from the average.

Other experiments also support the concept of cs[14]. Particularly convincing are laser hole burning experiments[15,16]. In a simple system such as an atom or a nucleus, spectral lines are homogeneous. If hit by a much narrower laser line, the entire line will react as a single line. In a protein, lines will be inhomogeneously broadened because the line position will be somewhat different in each molecule. With a narrow laser line, a hole can be burned into such an inhomogeneous line, demonstrating the existence of cs.

Organisation of the energy landscape

The experiments sketched so far provide a powerful argument for the existence of cs in proteins and consequently for the existence of a rough energy landscape, but they tell us little about its properties. The situation then is similar to that in atomic physics before Balmer, when spectral lines were seen, but no regularities were known. Some experiments, however, already give us an inkling about the organisation of the cs in one protein, myoglobin. Two results may be important for all complex systems:

1) Substates appear to fall into two classes, taxonomic and statistical, as indicated in figure 2. Taxonomic substates are few in number. They are so different that they can be characterised and described individually. They can not however, be separated, physically or chemically and at ambient temperatures they inter-convert rapidly. Each taxonomic cs can exist in a very large number ($>10^4$) of statistical substates. These show only minor

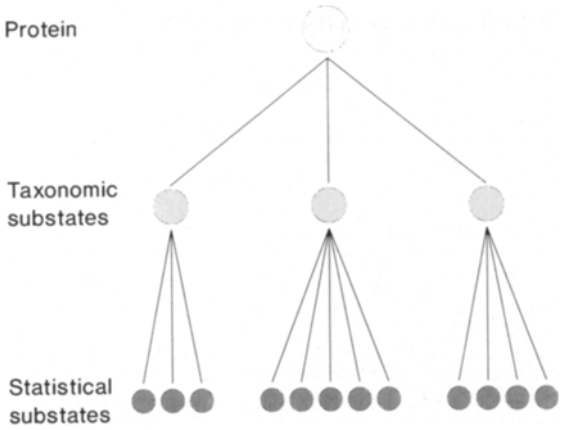

Figure 2. Taxonomic and statistical conformational substates.

differences in their properties and cannot be identified individually. Their properties must be described by distributions.

2) The substates appear to be organised in a hierarchy[17]. The split into taxonomic and statistical cs shown in figure 2 is only the first step in such a hierarchy. The statistical cs are most likely again arranged in a hierarchy, characterised by decreasing barriers between cs.

These two properties are only the beginning of an insight into the energy landscape of proteins. To really understand the organisation, far more experimental and theoretical work will be needed. How many tiers (levels) exist in a given protein? How does the energy landscape depend on the primary sequence and on the tertiary structure? How do genetic modifications (landscaping) affect the landscape? How can the taxonomic and the statistical cs be described in detail?

Another set of questions concerns the generalisation of the concepts to other complex systems. Does the separation into taxonomic and statistical substates also hold for more complex systems? Languages, for instance, appear to show such a separation. English, German, and Chinese are rather different, but within each, there exist very large numbers of dialects. If the separation is general, what law causes it?

Dynamics

Proteins perform a vast assortment of functions, from storage and transport of matter, energy, and information, to catalysis. In all of these, protein motions are involved. What are the concepts and laws that govern both the functions and the motions? In most of the processes, we can simplify the description by introducing two types of coordinates, reaction and conformation. Consider for instance the binding of dioxygen to hemoglobin. The dioxygen will move from the outside to the inside of the hemoglobin and then bind at an iron atom close to the center of the hemoglobin. The reaction coordinate describes the motion of the dioxygen through the protein. The X-ray structure of the hemoglobin[7], however, shows no opening through which the dioxygen could enter or leave. The protein must fluctuate and create large enough channels for the entrance and exit. Such fluctuations correspond to conformational motions. These motions, in turn, correspond to jumps between conformational substates and are thus transitions along conformational coordinates. Reaction and conformational motions are strongly coupled; the reactions are governed by the fluctuations, and the conformational motions can be caused by reactions that produce protein quakes[17]. Both reactions and conformational motions raise questions that are not yet treated in texts[18,19] and that may be representative for complex systems.

1 Allen, R., The Physics of Amorphous Solids. Wiley, New York 1983.
2 Fischer, K.H., and Hertz, J. A., Spin Glasses. Cambridge University Press, Cambridge 1991.
3 Stryer, L., Biochemistry. W.H. Freeman, San Francisco 1988.
4 Schroedinger, E., What is Life? Cambridge University Press, Cambridge 1944.
5 Toulouse, G., Comments Phys. *2* (1977) 115.
6 Von Laue, M., Röntgenstrahlinterferenzen. Akademische Verlagsanstalt, Leipzig 1948.
7 Perutz, M. F., Science *140* (1963) 863.
8 Kendrew, J. C., Science *139* (1963) 1259.
9 Dickerson, R. E., and Geis, I., Hemoglobin: Structure, Function, Evolution, and Pathology. Benjamin/Cummings, Reading, MA, 1983.
10 Frauenfelder, H., Sligar, S. G., and Wolynes, P. G., Science *254* (1991) 1598.
11 Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H., and Gunsalus I. C., Biochemistry *14* (1975) 5355.
12 Frauenfelder, H., Int. J. Quantum Chem. *35* (1989) 711.
13 Frauenfelder, H., Petsko, G. A., and Tsernoglou, D., Nature *280* (1979) 558.
14 Frauenfelder H., Parak, F., and Young, R.D., A. Rev. Biophys. biophys. Chem. *17* (1988) 451.
15 Kohler, W., and Friedrich, J. F., J. chem. Phys. *90* (1989) 1270.
16 Jankowich, R., and Small, G. J., Science *237* (1987) 618.
17 Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E. T., Sauke, T. B., Shyamsunder, E., and Young, R. D., Proc. natl Acad. Sci. USA *82* (1985) 5000.
18 Frauenfelder, H., and Wolynes, P. G., Science *229* (1985) 337.
19 Frauenfelder, H., Chu, K., Nienhaus, G. U., and Young, R.D., in: The Barrier Crossing Problem, p. 120. Eds G. R. Fleming and P. Haenggi. World Scientific Pub. Singapore 1993.