# BIOMOLECULES: WHERE THE PHYSICS OF COMPLEXITY AND SIMPLICITY MEET

How can a protein fold itself properly in an instant, and how could functional proteins possibly have evolved in the brief span of life on Earth? Addressing such questions, we learn a lot about the physics of complexity.

Hans Frauenfelder and Peter G. Wolynes

Are we moving toward a time when no new and exciting problems appear in physics? Would the vaunted "theory of everything" mean the end of creative physics? A similar scenario was played out at the end of the last century, when some great physicists declared that only minor problems remained to be solved.

As it was a century ago, it may now again be that the most exciting time lies just beyond the horizon. Physicists for a long time have boasted that they undertook only the "simple" problems, leaving complicated, messy ones for the biologists and chemists. But the world is complex. We can therefore ask if a "physics of complexity" is possible and if we can extract new physical concepts and laws from complex systems.

We can easily say what simple systems are. Every physicist has tackled the harmonic oscillator, the Bohr atom and the Schrödinger equation with a square-well potential. Atoms, crystals and even nuclei and subnuclear particles are, in some sense, simple systems. But what is complexity? Whatever "complexity" means, most people agree that biological systems have it.

Thus it seems that the serious, detailed study of biological systems with an eye toward general physics principles is a likely route to a useful science of complexity. Among the biological problems approached in this way are the organization of societies of organisms, morphogenesis and the development of individual organisms, and the intercellular signaling employed in the nervous and immune systems.[1] But it is the study of biological macromolecules that has seen the most intense interplay between theory and experiment. Proteins, nucleic acids and biological membranes are the simplest systems that can be regarded as components of the living world.

**Hans Frauenfelder,** now on the staff of the Los Alamos National Laboratory, is a professor emeritus of physics, biophysics and chemistry, of the University of Illinois, Urbana–Champaign. **Peter Wolynes** is a Lycan Professor and a Center for Advanced Study Professor of chemistry, physics and biophysics at the University of Illinois, Urbana–Champaign.

Thus they have given us a most valuable wellspring and proving ground for new physical ideas. In this article we hope to show how the study of protein dynamics has revealed some principles of the physics of complexity and why biomolecular physics shows promise of revealing still more.

Diversity is the first prerequisite of complexity. A complex system should show many significantly different states. Although a quiescent fluid being heated from below may have many atomic configurations, they are all so similar on the macroscopic level that we do not see the system as complex. But proteins, as we shall see, have a great diversity of conformational states with substantively distinct molecular behavior.

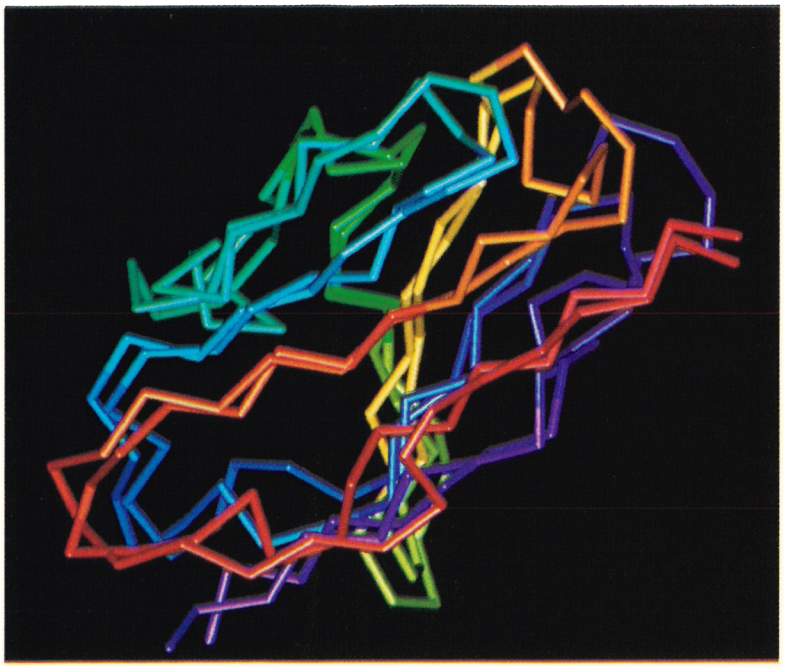
The differentiation of states is a hallmark of biology, and of complexity in general. The traditional physicist is apt to recoil from the need to characterize these states, asserting that all investigations are either physics or philately. The description of an individual protein state may well be the province of the traditionally trained biologist. But physical and mathematical thinking can address the questions of how to characterize large numbers of different states and why biomolecules generally have classifiable states.

Another characteristic of complex systems is "contingency," that is to say, the dependence of the present state on the vagaries of its past history. The quest to understand which features do and which do not depend on chance or antecedents reveals general principles.

## Biological molecules

These characteristics of diversity, differentiation and contingency are important features of the physics of biomolecules. A protein molecule is a long chain of amino acids that can fold itself into different configurations. Work in the last several decades has characterized many aspects of the diversity of such conformational states.[2] In the process we have learned much about how states are differentiated and organized. Contingency appears significantly in the spatial and temporal properties of proteins, both on laboratory and evolutionary time scales.

Several features of the biomolecular world make it



**Predicted and observed** folded structures of an immunoglobulin protein molecule, indicated by the two adjacent multicolored strands, are impressively similar. The structure was determined by x-ray diffraction, and the prediction was based on spin-glass methods. The spectral color sequence, starting with red, denotes the sequence of component amino acids that make up this protein chain. (Adapted from ref. 14.) **Figure 1**

especially easy to investigate the issues of complexity. The diversity of biomolecules themselves is the first help. Biomolecules on Earth today have resulted from the complex process of biological evolution. Evolution has created many variant molecules, some of which perform similar functions in different organisms. One can compare the sequences in which components are arrayed on biomolecular chains in many quantitative statistical ways, and a huge database of such sequence information already exists. At the same time genetic engineering makes it possible to create other variants at will and to obtain inexhaustible quantities of reproducibly characterized samples. Once the principles of biomolecular physics are better understood, it may also be possible to create useful biomolecules completely *de novo*.

Another help is that biomolecules are "mesoscopic" in size. They are small enough that one can generate and study large ensembles of them, and yet each one is large enough to have numerous differentiable sites that can be probed separately by powerful spectroscopic methods. X-ray diffraction and nuclear magnetic resonance spectroscopy demonstrate that proteins organize themselves in such a way that we can give at least an approximate spatial location to these differentiable sites in a folded protein.

Finally, the dynamics of biomolecules cover an extraordinarily wide range of time scales, from femtoseconds to teraseconds.

Theory has also contributed to the development of concepts of complexity in biomolecules. The faster processes in proteins can be studied by computer simulation.[3] (See the article by Martin Karplus in PHYSICS TODAY, October 1987, page 68.) Simulations can greatly enhance the interpretation of laboratory experiments, but on present computers the time range that can be simulated is limited to a few nanoseconds. Because many important biomolecular processes occur on time scales longer than that, we need other theoretical approaches.

Considerable inspiration has come from the study of disordered systems in condensed matter physics.[4] Mathematical techniques borrowed from that area have been valuable, but biomolecular systems raise different issues that lead to an emphasis on kinetics. An important tool is the statistical characterization of the energy landscape on the global scale.

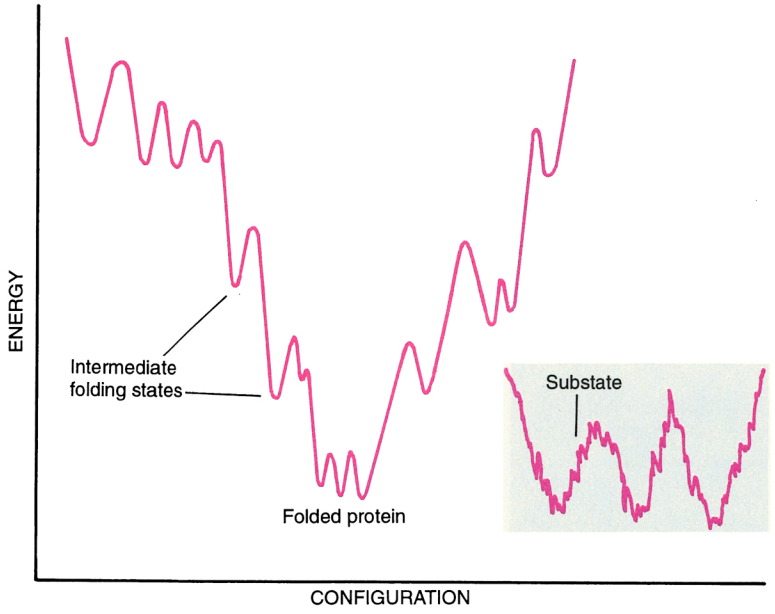
## The energy landscape

The path from proteins to the discovery of physical laws of complex systems appears daunting. How can simple concepts emerge from such complicated systems? But physics teaches us how to approach the task. The approach that has worked for atoms, solids, nuclei and elementary particles may also work for biomolecules: Study the structure, investigate the energy levels and determine the dynamics.

The structure of many proteins is known, and their salient features can be described simply. A typical protein links about a hundred amino acids in a well-defined sequence. This polypeptide chain is called the primary structure. Under the proper circumstances, the chain folds into a close-packed tertiary structure—the working protein—which is typically globular. (See the article by Jerome Karle in PHYSICS TODAY, June 1989, page 22.) Figure 1 shows an immunoglobulin molecule in its "native" folded configuration, together with the configuration predicted by methods we will discuss in this article.

The static picture suggested by figure 1 is misleading when we consider the energy levels. In simple systems such as atoms, nuclei or particles, the ground state is nondegenerate, and the excited states can be labeled primarily by their energies above the ground state, plus auxiliary quantum numbers such as spin and parity. In complex systems such a description is no longer possible: The ground state of a folded protein is highly degenerate, and in place of specific energy levels, we must speak of an "energy landscape."

The energy of a protein is a function of the topological arrangement of the atoms. It is described by a hypersurface in a space of a few thousand coordinates, with a very large number of energy valleys (or craters) separated by mountains and ridges. (See figure 2.) Each valley in the hypersurface describes the protein in a particular conformation. Proteins perform specific functions. Even in the simplest cases they have two different states, for




**Folding-energy landscape** for a protein molecule, depicted schematically in one-dimensional cross section. The folding begins as the molecule descends the funnel-like multidimensional landscape that guides it toward the configurations of lowest energy at the bottom. Rough and smooth features coexist: The funnel is created by the smooth guiding forces superimposed on the rougher small-scale features of the heteropolymer. Intermediate folding states lurk on the slopes of the funnel, and even in the bottom region of completed folding there are a large number of substates, as shown in the blow-up. **Figure 2**

instance a charged state and a neutral state. Each such state has an energy landscape with many valleys. By analogy to magnetic states and substates, we say that every valley characterizes a "conformational substate."

## Excitement and relaxation on the landscape

For simple systems in a single phase, the energy landscape is described by way of elementary excitations from the ground state. The most familiar such excitations are phonons: the small-amplitude harmonic motions in solids. Some excitations involve large-amplitude rearrangements of the ground state. These are the so-called defects. Low-energy defects are usually local distortions; more extended dislocations have higher energies. Defects play a crucial role in long-time-scale processes such as diffusion, because the formation and transport of defects is thermally activated.

Just like crystalline solids, proteins and nucleic acids exhibit these conventional excitations. The harmonic vibrations of protein chains have been studied experimentally by neutron diffraction and computer simulations. The proteins also have local defects, which can be seen in careful x-ray diffraction studies.

In a complex system, many low-energy states separated by high barriers and globally quite different from one another can nonetheless have nearly the same energy. Unlike the ground states familiar to solid-state, nuclear and particle physicists, these low-energy states are not related by any symmetry. The situation is much like that in spin glasses.[4] (See the Reference Frame columns on spin glasses by Philip Anderson in PHYSICS TODAY in between January 1988 and March 1990.)

In addition the local excitations like crystal defects, proteins are also subject to phonons and other global excitations. The description of global excitations requires a statistical and global view of the energy landscape. These excitations involve a large range of energy scales.

The relaxation of a biomolecule from an excited nonequilibrium state toward equilibrium exhibits characteristics that distinguish complex systems from simple ones. Because the many local minima in figure 2 are surrounded by barriers of different heights, the typical relaxation process is not simply exponential in time. It can often be described by the stretched exponential function

$$\Phi(t) = \exp(-[\kappa(T)\,t]^{\beta})$$

where $t$ is time and $T$ is temperature. (See the article by Harvey Scher, Michael Shlesinger and John Bendler in PHYSICS TODAY, January 1991, page 26). Many relaxation phenomena in glasses and spin glasses can be described by this equation with $\beta$ less than 1. Pressure- and temperature-jump experiments by Frauenfelder's group at the University of Illinois have shown that relaxation phenomena in proteins also can be described by stretched exponentials. The characteristic rate coefficient $\kappa(T)$ changes with temperature in a distinctly non-Arrhenius fashion, because deeper minima become more occupied as the temperature is lowered. The behavior found in proteins is the same as the phenomenological law for glasses:[5]

$$\kappa(T) = \kappa_0 \exp(-(E/kT)^2)$$

where $k$ is the Boltzmann constant. (Arrhenius's law, by contrast, would have $\kappa$ go like $\exp(-E/kT)$.) With decreasing temperature, the rate coefficient decreases ever more rapidly. Such behavior is characteristic of a glass transition in which the "transition temperature" depends on the characteristic observing time. Proteins exhibit similar transitions near 200 K.

## A simple model

Simple models often help one understand the experimental data and design new experiments. Bernard Derrida's random-energy model[6] illustrates many features of rough energy landscapes of the kind shown in figure 2. In his model each of a set of discrete states is given an energy chosen randomly from a Gaussian distribution. The statistical independence of different nearby states leads to an extraordinarily rough energy surface in which minima can be surrounded by high barriers. This statistical independence means that the states that just happen to have the lowest energy need not resemble one another by any measure. That is actually a very good model of the folding landscape for the conformational states of a heteropolymer with a random sequence.

Many dynamical and thermodynamic aspects of the random-energy model can be obtained with elementary

statistical considerations. The nonexponential relaxation, for example, is due to the variety of barrier heights. The model also exhibits a phase transition remarkably like glass transitions observed in the laboratory.

The transition is basically an entropy crisis: The number of thermally available states decreases so fast at the transition temperature that the system freezes into one of a small number of states. The deep minima individually have very little entropy, and they are thermodynamically few in number. Even a mesoscopic system may still have a countable number of states.

The frozen states are not unique. The exact state into which the system freezes would be different in each preparation. Although these states have nearly the same energy (on the macroscopic scale), their precise energies in Derrida's model depend sensitively on which precise example of the random-energy landscape is chosen. This is an example of what one calls a "non-self-averaging" property. That is to say, motions between low-energy states of a protein molecule with a given sequence of amino acids are quite different from the "mean" behavior of proteins of the same composition averaged over all possible sequences. (See the box at right.)

## Simple and complex landscape features

We will sketch two areas of research: protein folding and the behavior of folded proteins. They both involve the energy landscape of the protein, but on different length scales. We first ask how biomolecules obtain their functioning three-dimensional structure. The one-dimensional information contained in the sequence of amino acids is apparently sufficient for a protein molecule to organize itself into its folded working shape by moving on its energy landscape into one of a rather limited (but still very large) set of configurations. (See the article by Hue Sun Chan and Ken Dill in PHYSICS TODAY, February 1993, page 24.) Max Delbrück, a physicist who became the founder of molecular biology, once remarked that this transformation of one-dimensional into three-dimensional information was a paradox almost tantamount to a new law of physics, "one that nobody could have pulled out of quantum mechanics without first having seen it in operation."

Early in the 1960s Cyrus Levinthal asked how long a protein would take to fold if it tried every configuration at random. Because of the enormous entropy of the chain, this time could be $3^{100}$ times as long as it takes to make a simple conformational move. That's much harder than trying for a hole in one on a golf course while blindfolded. A major focus of biomolecular theorists in recent years has been the characterization of those aspects of the energy landscape that would make self-organization possible. The consensus emerging from these studies is that the energy landscape of folding proteins must be *both* simple and complex if it is to reconcile the conceptual paradoxes with the experimental reality.

In 1987 Wolynes and his student Joseph Bryngelson suggested that the random-energy model could describe features of the energy landscape for misfolded protein states.[7] The misfolding minima would act as traps that slow down the molecule's self-organization. They showed (as did Eugene Shakhnovich and Alexander Gutin in Russia) that escape from these traps would become progressively slower as the temperature is lowered, becoming cosmologically large at the ideal glass transition temperature of the model,[7] just as it does in Levinthal's paradox.

Thus the random-energy model suggests that blind golfing and stochastic mountain climbing are computationally equivalent. The energy landscape that guides folding must somehow then be simpler than the worst possible case in the random-energy model. One can construct a smoother energy landscape by minimizing the "frustration" of the energy terms.[7,8] (See the box on page 62.) That is to say, instead of choosing the amino acid sequence randomly, one can select a sequence for which the energy contributions add up progressively to stabilize the native structure.

In other words, Bryngelson and Wolynes were suggesting that biological proteins, unlike random heteropolymers, satisfy a principle of minimal frustration. This idea has its roots in many earlier notions, especially the work of Nobuhiro Go in Japan. He pointed out that the

### The Random-Energy Model

Imagine the histogram of the energy levels of a random heteropolymer. We can approximate this energy distribution by a Gaussian probability function

$$P(E) = \frac{\exp(-E^2/2\sigma^2)}{\sqrt{2\pi}\sigma}$$

The Gaussian approximation is good because the energy is a sum of many conflicting terms. At a temperature $T$, the distribution of occupied states, given by this Gaussian times the Boltzmann factor $e^{-E/kT}$, is strongly peaked around a mean energy

$$\langle E \rangle = -\frac{\sigma^2}{2kT}$$

In the thermodynamic limit only a small range of energies around this mean energy is occupied. From Boltzmann's identification of the entropy $S$ with the logarithm of the number of states, we get

$$S = k \log W_0 P(E) = S_0 - \langle E \rangle^2/2\sigma^2$$

where $W_0$ is the total number of configurations of the system and $S_0$ is the corresponding entropy. Notice that the entropy dramatically decreases with decreasing temperature and vanishes at a temperature given by

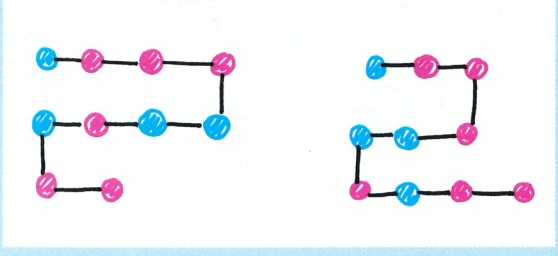
$$T_0 = \frac{\sigma}{2\sqrt{S_0}}$$

The meaning of the entropy crisis at $T_0$ is that one is now sampling the very edges of the distribution, where the histogram deviates strongly from the Gaussian. Only a few states will be occupied; exactly which ones depends on the randomness of the landscape and the thermal history. Thus many properties of protein folding are not self-averaging. In a biopolymer these states depend sensitively on the sequence of molecular components. One must bear in mind that below the dynamic transition temperature (which is higher than $T_0$), a biopolymer is a nonequilibrium system with residual entropy, whose annealing may require geological time!

## Frustration

One can understand the concept of frustration, first introduced by Philip Anderson and by Gérard Toulouse in 1977, by considering three spins interacting antiferromagnetically. Assume that two antiparallel spins (↑↓) have a lower interaction energy than if they were parallel (↑↑). The third spin is said to be "frustrated" because it will always be parallel to one and antiparallel to the other (↑↓↑ or ↑↑↓). The energy landscape of the three spins thus has two minima not related by symmetry.

Frustration arises in folded proteins, for instance, because side chains have to compete for positions that minimize their respective energies.[9] Heteropolymers with random sequences also exhibit frustration on the folding length scale.[10] Imagine a polymer with pairwise short-range interactions between the amino acids that may be either attractive or repulsive. They might, for example, be hydrophobic or hydrophilic. When two segments of a string of such residues in a heteropolymer are brought together by folding, there will be both positive and negative contributions to the configuration energy. It is very hard, if not impossible, to organize the string locally so that the various phobias (and philias) are all accommodated. Thus they are frustrated! (The sketch below shows two folded configurations of a frustrated polymer. Balls of a given color prefer the proximity of their own kind, but no folding can make them all happy at once.) In general, when segments are brought together, they give random contributions to the energy of configuration. Even a simple translation of one part of the string with respect to another can give a wildly different energy. That's why random heteropolymers generally have rough energy landscapes that resemble the random-energy model.



Proteins differ from random heteropolymers in that their sequences have been selected so that there are low-energy configurations in which many interactions can cooperatively be satisfied. That is rather like Daniel Mattis's model of disordered ferromagnets, in which the interactions are chosen at random, but in a cunning way that avoids frustration.

structural themes of proteins have an inherent harmony, in which appropriate secondary structures (for example, $\alpha$ helices and $\beta$ sheets) can be built up into supersecondary structures and then into complete tertiary folds without the substructures being dismantled.[11] To have a completely unfrustrated energy function the amino acids would have to be arrayed in such a way that every contact made by folding the protein is energetically favorable or, at worst, neutral.

Obviously, if protein structures were arbitrary, this frustration-free limit could be achieved only if there were an infinite variety of amino acids. But there are only 20 kinds of amino acids; so some degree of frustration and energy-landscape complexity is inevitable. A minimally frustrated landscape will have a set of very-low-lying minima that have many structural characteristics in common. For real proteins, the average of these configurations will presumably look much like the structure determined by x-ray diffraction. The lower-lying excitations will look like progressive unfoldings of the configurations corresponding to those energy minima. The energy landscape of a protein would therefore resemble that shown in figure 2.

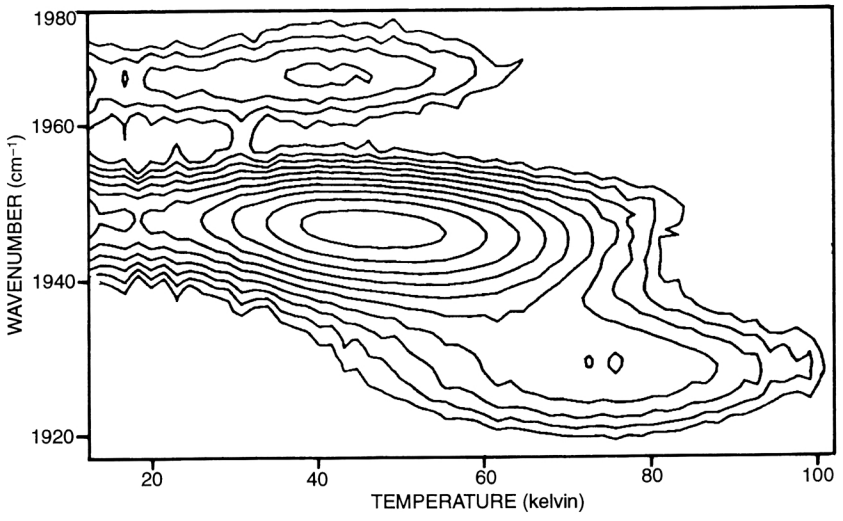
Folding on such a landscape can occur via many paths, and there will be a funnel of states all leading down to the low-energy native conformations. Thus the folding process resembles the crystallization of a solid from a melt. The guiding forces provided by the harmonious interactions circumvent the Levinthal paradox, just as it is circumvented by water freezing to form ice crystals.

A variety of computer simulation studies have confirmed that an energy landscape having the combined features of complex and simple systems can give rise to rapid folding. Peter Leopold, Mauricio Montal and Jose Onuchic at the University of California, San Diego, have shown that guiding forces for particular sequences designed using a 20-letter code (to simulate the 20 different amino acids) can lead to fast folding, and they have delineated much of the structure of the folding funnel.[12] A sequence without a folding funnel cannot fold spontaneously. Shakhnovich (now at Harvard) and coworkers have shown that while simple toy-model proteins with random sequences described by a two-letter code will generally not fold in short times, more realistic model proteins designated by a larger code will fold especially quickly when the sequences are specified to be largely unfrustrated.[13]

The idea that the folding-energy landscape has both simple and complex features has important experimental consequences. The guiding forces will be most active in the early stages of folding, because that's when the density of states is quite large. On the other hand, in the last stages of folding, when the entropy is much reduced, a glass transition could well intervene, and in fact many experiments have observed the very large activation energies characteristic of glassy systems appearing in the last stages of protein folding. That suggests considerable residual randomness in the sequences of naturally occurring proteins. Apparently evolution has implemented the principle of minimal frustration only to the extent necessary for obtaining protein folding on physiologically relevant time scales.

The low-entropy part of the energy spectrum displays sequence-specific details analogous to the non-self-averaging features of spin glasses. Therefore the final stages of folding will depend on the specific sequence of amino acids, whereas earlier folding stages should be robustly insensitive to details of sequence. So the study of late stages of folding will not teach us much about the part of the energy landscape that gives rise to self-organization, but will be crucial to understanding how mutations affect viability and function. Folding-pathway mutants have been found for a large number of naturally occurring proteins, and they may appear in several pathologies, including Alzheimer's disease.

The principle of minimal frustration asserts that proteins live on the border between the simple and complex in the folding-energy landscape. This principle can be put to several practical uses. For purposes of predicting protein structure Wolynes and coworkers Richard Goldstein and Zan Luthey-Schulten[14] used it to in-




**Temperature-derivative spectroscopy** of the protein myoglobin with an attached CO ligand. Temperature was slowly raised as infrared spectra were taken near the stretching frequencies of the bound CO. Displayed are contours of the temperature derivative of the absorbance as a function of infrared wavenumber. Three peaks show up, indicating that the bound CO can have three different well-defined frequencies. That shows the existence of three taxonomic substates. But rebinding takes place over an extended range of temperatures, indicating that the taxonomic substates are further divided into statistical sub-substates. (Adapted from ref. 16.) **Figure 3**

vestigate the appropriate forms of energy functions in protein folding. Put more quantitatively, the principle states that the energy gap between the folded protein structure and an ensemble of misfolded structures should be as large as possible when measured in units of the mean-square fluctuations of misfolded-structure energies. Applying it to a sample set of proteins whose sequences and structures are known lets us determine the optimal interaction potentials between amino acids. We can then use these potentials to calculate the molecular dynamics of folding for new sequences and to test possible folding structures for new sequences. Figure 1 is an example of successful prediction of protein structure by means of such algorithms.

While the structure predictor's problem is to find interactions from known data, the protein designer must use whatever interaction potentials are provided by nature to find sequences that yield particular structures. Once again the minimal-frustration principle leads to sequences of weakly frustrated heteropolymers. Model studies of this kind have been undertaken by Shakhnovich and others. Designing by finding minimally frustrated sequences is not a difficult computational problem. Even though the number of minimally frustrated sequences is very small, a simple Monte Carlo method easily finds them. Thus one avoids another paradox in protein folding, posed by the astronomer Fred Hoyle, champion of steady-state cosmology. How, asks Hoyle, could proper protein sequences have evolved in the finite time available in a Big Bang universe?

## Complexity in folded proteins

The properties of folded proteins as complex systems have emerged mainly from experimental studies of myoglobin, the molecular biologist's "hydrogen atom." Mb is a relatively small protein, consisting of about 150 amino acids. Mb stores $O_2$ and other small ligands such as CO and NO, which bind at an iron atom located in a heme group inside the protein.

The first unambiguous evidence for conformational substates in Mb appeared in low-temperature flash photolysis experiments done by Frauenfelder and coworkers at Illinois 20 years ago.[15] In such experiments, photons hit an MbCO target and break the bond between the CO and the iron atom. The CO either rebinds or moves out of the protein to rebind later. When the Illinois group studied the rebinding of CO to Mb, they found that the rebinding time at any temperature between 4 K and 160 K extended over many orders of magnitude. The Illinois group postulated that each Mb molecule was frozen into a particular substate with a corresponding activation enthalpy $H$ for the intramolecular rebinding of CO. If $g(H)\,dH$ denotes the probability of finding an Mb molecule with its barrier in the interval $dH$, the probability of remaining unbound at time $t$ is

$$N(t) = \int g(H) \exp(-\kappa(H)t)\, dH$$

This expression, with $\kappa(H)$ given by the Arrhenius law, adequately describes the rebinding on time scales from microseconds to kiloseconds, in the temperature range from 40 K to 160 K.

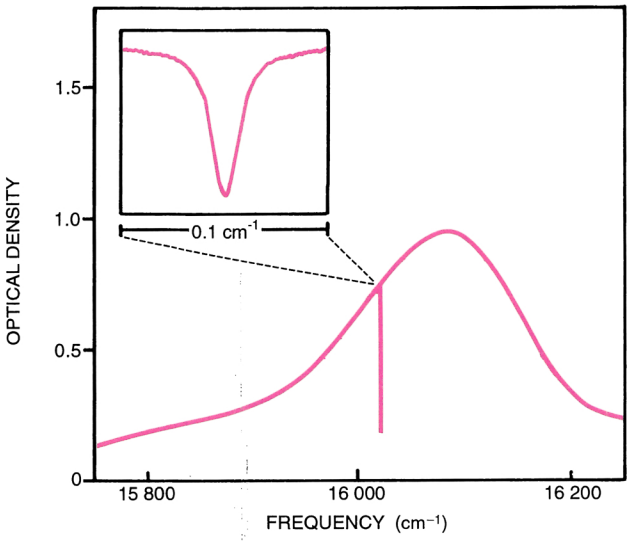
The appearance of a distribution $g(H)$ in place of a unique enthalpy barrier $H$ is characteristic of complex systems. It implies a rough energy landscape even for the folded protein, as shown in the inset of figure 2.

Figure 3 shows the results of a direct measurement of the distribution $g(H)$, carried out by Joel Berendzen and David Braunstein at Illinois by means of temperature-derivative spectroscopy.[16] An MbCO sample, originally photolyzed at 12 K in a Fourier-transform infrared spectrometer, was ramped up in temperature while spectra were taken. The differences between successive spectra, shown in figure 3, give the fraction of CO molecules with stretching frequency $\nu$ that rebind at temperature $T$. The broad range of temperatures over which CO rebinds is evidence for a distribution $g(H)$.

The data of figure 3 also provide evidence that the energy landscape is arranged in a hierarchy. Mb can assume three different substates, characterized by different stretching frequencies of the bound CO, as shown by the three contour peaks in the figure. Because these substates can be described individually, we call them "taxonomic substates." Each one contains a very large number of sub-substates that can only be described by distributions; these are the so-called statistical substates. There is evidence for additional tiers in this hierarchy of substates, so that a crude classification is emerging.

In addition to kinetics experiments, there is a variety of other approaches to the elucidation of the energy and structure landscapes: The Debye–Waller factors of individual atoms in proteins provide evidence for structural distributions.[17] Mössbauer studies by Peter Debrunner's group at Illinois and by Fritz Parak and coworkers in Germany support the substate concept and provide evidence for a glass-like transition near 200 K.

Laser hole-burning studies are particularly unambi-




**Burning a narrow hole** in a broad protein spectral line can be done with a spectrally narrow laser beam. That's because the observed protein line is an inhomogeneous superposition of the intrinsically much narrower lines of many different substates of slightly different structure. The laser light makes the narrow spectral hole by exciting only a small subset of the protein molecules. (Courtesy of Josef Friedrich, University of Bayreuth, Germany.) **Figure 4**

guous. Protein spectral lines should be inhomogeneous, because the environment of the central "chromophore" is different in different substates. So one gets a superposition of lines. A number of groups have shown that a narrow laser line can burn a "hole" into a protein spectral band by exciting only a small subset of the conformational substates.[18] (See figure 4.) Another characteristic feature of glassy systems, namely the anomalous specific heat exhibited below 1 K, has also been shown to occurs in proteins.[19]

So folded proteins share fundamental properties with other complex systems: a rough and hierarchical energy landscape; the absence of equilibrium below some glass transition temperature; two-level states; distributed observables; and relaxation, transport and reaction processes with nonexponential time dependence and non-Arrhenius temperature dependence. Probing with x rays, neutrons and nuclear magnetic resonance can yield protein structure in atomic detail, and proteins can be modified almost at will. Therefore proteins may well serve as the paradigm for complex systems, where we can explore in depth the connections between structure, energy landscape, dynamics and function.

## A laboratory for complexity

The study of biomolecules has been an important contributor to the emerging physics of complexity. In many ways proteins are at the border between the simple and the complex. Aspects of both can be seen in biomolecular behavior. Many of the phenomenological features of folded protein dynamics display complexity. Experiments and theory have, however, shown us how to summarize some of this complexity in new laws, for example, the non-Arrhenius behavior of rates and the nonexponential dynamics of relaxation.

The complex-system viewpoint has also allowed us to understand that some properties of proteins are expected to be peculiar to each individual molecule and to appreciate why the individuality of systems should be a general feature of biology. The simplicity of proteins reflected in the folding-energy landscape emerges as such an important feature: The organized behavior that allows spontaneous folding is a product of evolutionary selection.

The interplay of biomolecular physics and evolutionary theory and experiment is likely to be a major theme of future work. In the short term, understanding biomolecules from a physical viewpoint should have important practical consequences in medicine and other applied life sciences. As our understanding of protein folding increases, the prediction of protein structures will become more routine and accurate, making it possible to design drugs for numerous diseases. Many diseases are themselves the result of errors in protein dynamics caused by mutations. But in the long run, the main impact of biomolecular physics is likely to be on the foundations of the physics of complexity. That will help us understand the even greater complexity that lies beyond the biomolecular level.

## References


1. See articles in *Biologically Inspired Physics*, L. Peliti, ed., Plenum, New York (1992), and *Spin Glasses and Biology*, D. L. Stein, ed., World Scientific, Singapore (1992).
2. H. Frauenfelder, S. Sligar, P. G. Wolynes, Science **254**, 1598 (1991). H. Frauenfelder, F. Parak, R. D. Young, Annu. Rev. Biophys. Chem. **17**, 451 (1988).
3. J. A. McCammon, S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge U. P., New York (1987).
4. M. Mezard, G. Parisi, M. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore (1987). K. Binder, A. P. Young, Rev. Mod. Phys. **58**, 8011 (1986).
5. I. Ibsen *et al.*, Phys. Rev. Lett. **62**, 1916 (1989).
6. B. Derrida, Phys Rev. B **24**, 2613 (1981); D. Gross, M. Mezard, Nucl. Phys. B **240**, 431 (1984).
7. J. D. Bryngelson, P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
8. J. D. Bryngelson, P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989). E. I. Shakhnovich, A. M. Gutin, Europhys. Lett. **9**, 569 (1989).
9. D. L. Stein, Proc. Natl. Acad. Sci. USA **82**, 3670 (1985).
10. T. Garel, H. Orland, Europhys. Lett. **6**, 307 (1988). E. I. Shakhnovich, A. M. Gutin, Europhys. Lett. **8**, 327 (1989).
11. N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
12. P. Leopold, M. Montal, J. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
13. E. I. Shakhnovich, A. M. Gutin, Proc. Natl. Acad. Sci. USA **90**, 7195 (1993); Protein Eng. **6**, 793 (1993).
14. R. Goldstein, Z. Luthey-Schulten, P. G. Wolynes, Proc. Natl. Acad. Sci. USA **87**, 4918 (1992); **89**, 9029 (1992); **90**, 9949 (1993).
15. R. H. Austin, K. Beeson, L. Eisenstein, H. Frauenfelder, I. Gunsalus, Biochemistry **14**, 5355 (1975).
16. J. Berendzen, D. Braunstein, Proc. Natl. Acad. Sci. USA, **87**, 1 (1990).
17. H. Frauenfelder, G. Petsko, D. Tsernoglou, Nature **280**, 558 (1979).
18. W. Kohler, J. Friedrich, J. Chem. Phys. **90**, 1270 (1989). R. Jankowich, G. J. Small, Science **237**, 618 (1987).
19. V. Goldanskii, Yu. F. Krupyanskii, V. N. Fleurov, Dokl. Akad. Nauk SSSR **272**, 23 (1983). ■