Navigating the Folding Routes

Peter G. Wolynes, Jose N. Onuchic, D. Thirumalai

To fold, a protein navigates with remarkable ease through a complicated energy landscape as it explores many possible physical configurations. This feat is beginning to be quantitatively understood by means of statistical mechanics and simplified computer models (*1*). Folded proteins are marvels of molecular engineering and it is hard to avoid thinking that all of their complex structural features play a role in their folding through an obligate multistep mechanism. A unique folding pathway, if it exists, could be elucidated with classical chemical experiments. A newer view holds that in the earlier stages a protein possesses a large ensemble of structures. The problem is not to find a single route but to characterize the dynamics of the ensemble through a statistical description of the topography of the free-energy landscape. Folding is easy if the landscape resembles a many-dimensional funnel leading through a myriad of pathways to the native structure. Only a few parameters should be needed to characterize statistically the topography of and routes down the folding funnel. Using experimental data, Onuchic *et al.* have estimated the extent, ruggedness, and slope of the folding funnel (*2*). Similar parameters characterize the energy landscape of simple computer models of proteins. These models of self-interacting necklaces of beads, often on lattices, lack most of the details of real proteins, but establishing a quantitative correspondence between the landscapes of computer models and real proteins makes it possible to use simulations to understand folding kinetics.

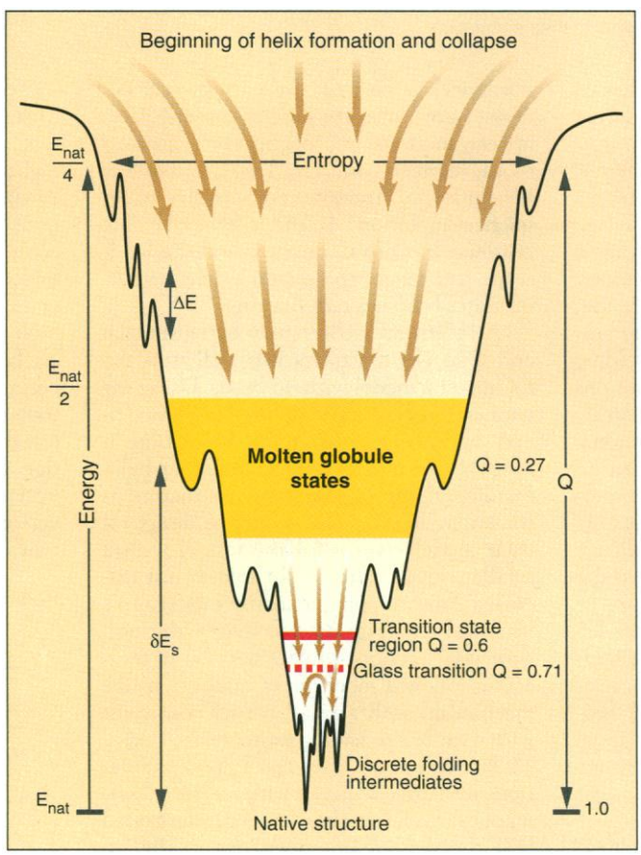

Fig. 1. Schematic of the folding funnel for a fast-folding 60-residue helical protein according to Onuchic *et al.* (*2*). The width of the funnel represents entropy, and depth, the energy. The flow of the molecule through the molten globule, folding bottleneck, or transition state ensemble and a glass transition region where discrete pathways emerge are indicated. The fraction of native contacts correctly made, *Q*, is indicated for each collection of states.

The extent of a protein energy landscape is huge. Before folding, each residue can take on about 10 different conformations; thus, a 60-residue protein can be in any of 10^{60} states. An unguided search, like a drunk playing golf, would take practically forever. A flat energy landscape (or golf course) is very unrealistic, but many years ago Bryngelson *et al.* pointed out that a difficult search also arises on a rugged energy landscape that might describe proteins (*1*). Such a landscape can arise because different segments of a typical heteropolymer come together without any guarantee that the many resulting individual interactions will not mutually conflict and thereby "frustrate" minimizing each other. This landscape will be rugged with many deep valleys corresponding to local minima. Transient trapping in these valleys slows the exploration of routes toward the most stable native structure. This trapping resembles the way a liquid becomes a glass when cooled, remaining fixed in one of many structures, unable to reconfigure to the lowest energy crystal state. At temperatures far above a glass transition, a rough landscape is easily traversed. At low temperatures, where the ground state has a significant chance to be thermally occupied, the search to find the deepest valley out of the many on the glassy energy landscape is incredibly slow. For a protein to be kinetically foldable, there must be a sufficient overall slope of the energy landscape so that the numerous valleys flow in a funnel toward the native structure. With such a slope, native structure elements are significantly more stable than expected by chance. Thus, the global energy minimum (native structure) is still thermodynamically stable above the glass transition temperature, where kinetic barriers for escaping glassy traps (misfolded structures) become too large. That the interactions of a kinetically foldable protein must have fewer conflicts than typically expected is known as the "principle of minimal frustration." Minimizing frustration or the ratio of glass to folding temperatures is equivalent to maximizing the "stability gap" between the native state and disordered collapsed structures measured in units of the ruggedness. The quantitative version of the minimal frustration principle has been used to infer energy functions useful for structure prediction (*3*) and to design proteins in machina (*4*). Many simulations of simple models have confirmed the principles of the energy landscape analysis (*5*), including a recent exhaustive study of ~200 short sequences (*6*).

Onuchic *et al.* look at real protein folding by developing a law of corresponding states (*2*). Studying phase transitions teaches that boiling is very similar for systems as different as water and methane because each system can be mapped onto the same pat of a universal phase diagram. Similarly, Onuchic *et al.* argue that the detailed local structural features like the hydrogen bonds of helices and side chain packing can be taken into account by finding appropriate values of the statistical parameters characterizing the free energy landscape. Helix formation and collapse solve part of the folding search locally so as to renormalize the effective number of degrees of freedom and change the details of single reconfiguration events. Nevertheless, the global features of the landscape of a real protein can be mapped onto those of a bead model.

P. G. Wolynes is in the School of Chemical Sciences, University of Illinois, Urbana, IL 61801, USA. J. N. Onuchic is in the Department of Physics, University of California at San Diego, La Jolla, CA 92093, USA. D. Thirumalai is at the Institute for Physical Sciences and Technology, University of Maryland, College Park, MD 20742, USA.

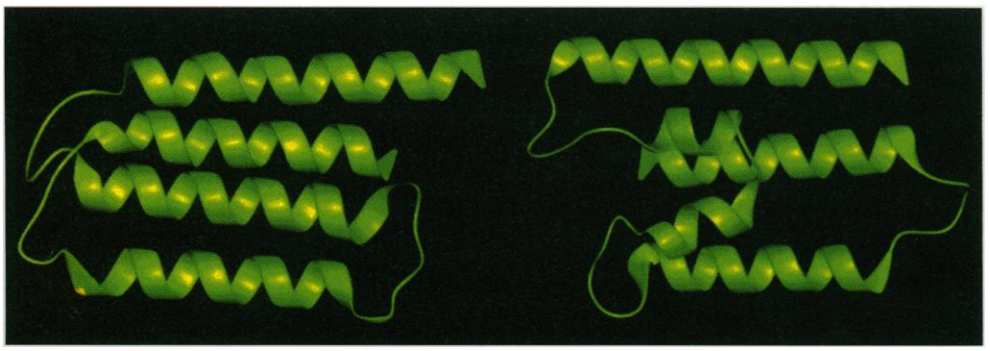

Fig. 2. (Left) Ribbon diagram for the native structure of a 96-residue helical protein that corresponds with a 46-residue bead model simulated by Thirumalai and Guo (*7*). **(Right)** A discrete native-like intermediate whose rearrangement is the slowest process in their simulation.

A theory of helix formation in collapsed polymers relates the measured amount of helix in the collapsed states to the conformational entropy associated with different chain topologies. The extent of the renormalized configurational energy landscape for a 60–amino acid helical protein is only a bit larger than that for models containing only 27 beads. At the folding temperature, using thermodynamics, the resulting entropy estimate yields the slope of the folding funnel. Finally, measurements on motions in the collapsed denatured proteins quantify the ruggedness and with the conformational entropy leads to values of a glass transition temperature. These estimates suggest real proteins resemble bead models in which only three kinds of residues are used to encode sequence in that both have a folding to the glass transition temperature ratio of 1.6.

A simulation of a three-letter code, 27-mer lattice model gives a picture of the folding mechanism and the folding funnel topography of a 60–amino acid helical protein. A caricature of the multidimensional funnel accurately representing the entropy and structural similarity involved in the search is shown in Fig. 1.

The funnel illustrates that a fast-folding helical protein has a collapsed molten globule band of states with roughly one-quarter of the number of native contacts correctly made. As one proceeds down the funnel, both entropy and energy decrease, but when roughly three-fifths of the native contacts are made, the incomplete compensation of entropy decrease by energy decrease leads to a very modest barrier ($\sim 3k_BT$) . Folding is thereby slowed by a thermodynamic bottleneck. The transition state or bottleneck region consists of a large ensemble of structures reflecting the multiple pathways of protein folding. After this bottleneck is crossed, the entropy still decreases until discrete kinetic intermediates appear, most having roughly three-quarters of the correct contacts. These native-like but misfolded structures are sensitive to sequence mutation. If the thermodynamic bottleneck is not too narrow and the landscape still rough, the search through intermediates becomes rate-limiting.

A recent off-lattice study of Thirumalai and Guo (*7*) illustrates this well into the folding of a model with 46 beads. Using the connection between simplified bead models and helical proteins, their simulation is roughly like that of a 96-residue four-helix bundle protein whose native structure is shown in Fig. 2. The sequence design of their model gives a folding funnel with a smaller slope than the fast folders just discussed. Now trapping in a rather native-like intermediate (Fig. 2) becomes a distinct, slow side reaction, while a fraction of molecules follow a more direct nucleation-like mechanism with a small barrier consistent with a finite size scaling estimate.

The correspondence theory puts simulations into direct contact with recent experiments. Several studies have demonstrated that the discrete intermediates in the refolding of cytochrome c can entirely disappear upon changing an individual residue's chemistry (*8*). In keeping with the energy landscape theory, these results show that the discrete intermediates often found are an epiphenomenon, their distinctness unimportant when trying to understand how folding is guided. Intermediates are relics of the landscape ruggedness whose features may fossilize some of the relevant guiding forces. Fersht's recent study of linear free energy relations for folding chymotrypsin inhibitor (*9*) suggests that different residues are between 30 and 70% folded in the ensemble of structures representing the transition state, reminiscent of the bottleneck in the three-letter-code funnel.

The small size of the actual thermodynamic barrier to folding is perhaps a bit surprising. Many studies on long time scales show very large activation barriers, but the landscape theory suggests that these arise from transient trapping. In vivo proteins are stable by several k_BT, so that folding may be largely a downhill run toward the final near native kinetic traps followed by a short search through them, as in many simulations. The near perfect compensation of entropy by enthalpy in the funnel suggests that proteins behave like fluids near a critical point. For theoreticians, a next step is to see how scaling and renormalization group ideas might be used to understand kinetics, especially for larger proteins. A more complete experimental characterization of the dynamics of partially folded proteins throughout their phase diagram, including very low temperatures and high pressures (*10*), is also needed to precisely quantify the glass transition. For experimentalists, the present perspective also shows that the guiding forces act in much less than a few milliseconds. A new generation of experiments using lasers to rapidly initiate folding (*11*) promises dramatic advances in direct measurement of the protein's energy landscape topography.

References and Notes

1. J. D. Bryngelson *et al.*, *Proteins: Struct. Funct. Genet.*, in press, and references therein.
2. J. N. Onuchic *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
3. R. A. Goldstein *et al.*, *ibid.* **89**, 4918 (1992).
4. E. I. Shakhnovich, *Phys. Rev. Let.* **72**, 3907 (1994).
5. J. D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3526 (1990); C. J. Camacho and D. Thirumalai, *ibid.* **90**, 6369 (1993); R. Miller *et al.*, *J. Chem. Phys.* **96**, 768 (1992).
6. A. Sali *et al.*, *J. Mol. Biol.* **235**, 1614 (1994).
7. D. Thirumalai and Z. Guo, *Biopolymers* **35**, 137 (1995).
8. G. Elove *et al.*, *Biochemistry* **33**, 6925 (1994); T. R. Sosnick *et al.*, *Nature Struct. Biol.* **1**, 149 (1994).
9. D. E. Otzen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10422 (1994).
10. J. Jonas and A. Jonas, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 287 (1994).
11. C. M. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11860 (1993).

12. We thank N. Socci, Z. Luthey-Schulten, and R. Cole for the preparation of the figures and W. Eaton for reading the manuscript. The work of the authors was supported by the National Science Foundation and the National Institutes of Health.