

This paper is a summary of a session presented at the Ninth Annual Frontiers of Science Symposium, held November 7–9, 1997, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA.

Chemical physics of protein folding

CHARLES L. BROOKS III^{*}, MARTIN GRUEBELE[†], JOSÉ NELSON ONUCHIC[‡], AND PETER G. WOLYNES[†]

^{*}Department of Molecular Biology, The Scripps Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037; [†]School of Chemical Sciences, University of Illinois, Urbana, IL 61801; and [‡]Department of Physics, University of California at San Diego, La Jolla, CA 92093-0319

Protein molecules generally fold into precise three-dimensional shapes. While of direct interest to biochemists, the question of how folding occurs has attracted the interest of a much broader audience of scientists ranging from the traditional chemical scientists to computer scientists and physicists. Also changing over time has been the very meaning of the question of protein folding. To the descriptive scientist it may be sufficient to assert that folding occurs on a time scale no slower than protein biosynthesis, and that the information required to find the precise three-dimensional shape is contained in the one-dimensional sequence. Although exceptions to these generalizations recently have begun to emerge; in the study of prion-associated diseases (1), they have sufficient generality to allow us to treat the folding process as a black box for transcribing one-dimensional information into three-dimensional structures. It is, however, necessary to probe deeper into the mechanism if the prediction of protein structure from sequence and the design of truly novel protein-like molecules are to be achieved. These goals are of great practical significance in biology and medicine. The question of the mechanism of folding was once thought to be entirely analogous to the question of mechanism in intermediary metabolism or classical organic chemistry. In those problems the small number of participating species and the relatively specific routes by which they interconvert owing to the large scale of covalent energy barriers compared with thermal energies means that a small number of fairly discrete chemical steps can be isolated. This is the classic notion of a protein folding pathway with a series of discrete intermediates. Such discrete intermediates do occur in the late stages of protein folding, and to a great extent, the chemical kinetic details of these interconversions have been catalogued (2). However, to answer the practical questions of structure prediction and design, one must go a considerable distance beyond this phenomenology—a new viewpoint on folding is required.

This new viewpoint is that of the chemical physicists rather than the classical chemists (3). The chemical physics view brings the problem in much closer connection to the underlying forces and the underlying microscopic events. This view has required a new set of theoretical ideas, computational techniques, and major advances in experimental methodology.

Energy landscape theory provides the theoretical framework, asserting that a full understanding of the folding process requires a global overview of the landscape. The folding landscape of a protein resembles a partially rough funnel riddled with traps where the protein can transiently reside (Fig. 1). There is no unique pathway but a multiplicity of convergent folding routes toward the native state (3–7). Although we focus on developments from our groups in this paper, several other groups have participated in developing this new view.

The importance of a funnel landscape can be seen by contrasting random heteropolymeric molecules and proteins. Random heteropolymers have an underlying driving force to

collapse but do not adopt well-defined three-dimensional structures because of the conflict or frustration of different interaction energies. Instead, they exist in an ensemble of dissimilar low-energy structures and move among this multitude of states, jumping over barriers between adjacent minima, giving multiexponential kinetics. Proteins, on the other hand, have a single manifold of structurally similar collapsed low-energy structures. When the slope toward the native state is dominant over the ruggedness of the landscape, folding kinetics is exponential and fast. The essence of the funnel landscape idea is competition between the tendency toward the folded state and trapping because of ruggedness. This competition is measured by the ratio between the folding temperature (T_f) and the glass temperature (T_g). Good folding sequences fold rapidly on minimally frustrated landscapes with large values of T_f/T_g (3, 5, 7).

Minimally frustrated sequences not only fold fast at relevant temperatures but are robust folders, only weakly dependent on minor variations of the folding environment or to mutations. Robustness is essential for biology. Minor variations in pH, temperature, or mutations will affect the native configuration by favoring other low-energy structures, which in a funnel-like landscape are very similar to the original one. Energy landscape theory suggests a diversity of folding scenarios that has been observed by computer simulations of minimalist models (see, for example, references in refs. 3, 8, 11, and 15–18) allowing connections to studies of real proteins (19, 20).

A deeper understanding of how the landscape relates to particular properties of a protein's sequence or final folded topology requires detailed molecular calculations. Such calculations often use explicit atomic level models of the protein and surrounding solvent together with molecular dynamics and biased sampling methods to map the free energy landscape onto coordinates describing folding progress (21). These permit detailed atomic information about folding to be related to experimentally measurable quantities. Atomistic studies may be compared with general predictions of landscape theories.

In Fig. 1 we display the folding free energy profile for the α/β protein GB1 under conditions strongly favoring the native state (22). Folding of this protein is dominated by an initial collapse without significant formation of native interactions, followed by folding toward the native without significant change in size. This behavior contrasts with the mechanism seen for some small helical proteins that are even less frustrated (14, 23). These fold with commensurate reduction in size and gain in tertiary interactions. Thus, two apparently different mechanisms are used by proteins that differ in their final folded topology. Such behavior can be mimicked with minimalist lattice models by changing the balance of native and non-native (hydrophobic) attractive interactions (24).

Also directly connected to phenomenological models of folding is the distribution of stabilizing native interactions in the transition region of folding (20). For helical proteins (at 300 K), this is rather broad centering around 50% occupation of the native contacts. The interactions that stabilize the protein at this point in folding are distributed throughout the

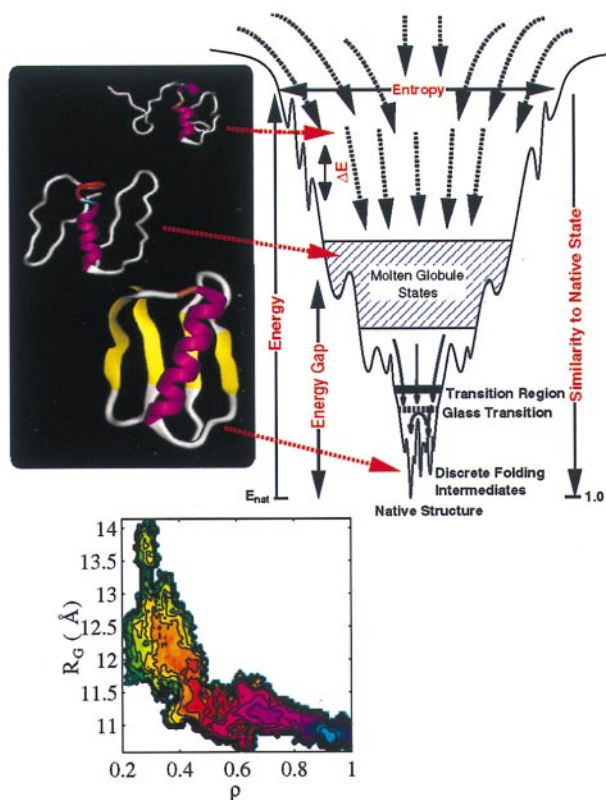


FIG. 1. The energy landscape for a folding protein. The major phenomenological parameters needed to capture this landscape include: the width of the funnel at small values of native similarity, indicating the entropy of denatured states; the roughness of the landscape, ΔE , which is related to the glass transition temperature, T_g ; the stability of the native state relative to the collapsed but non-native (molten) globule states, the energy gap. The ribbon diagrams of the α/β protein, segment B1 of streptococcal protein G (GB1) provide structures from ensembles of unfolded, molten globule, and native conformations. The folding landscape for GB1 is projected onto two coordinates, the radius of gyration, R_g , of the folding globule, and the fraction of native contacts, ρ , which indicates how close the folding protein is to the native. The free energy change as folding occurs is shown as a contoured surface: (native) state corresponds to the blue region and the most unfavorable unfolded state is represented by the green contours.

structure. For the α/β protein GB1, this distribution is more bimodal, with a larger number of native interactions formed with the highest and lowest probabilities compared to those that occur with 50% probability.

Simulations with detailed atomic models are extremely intensive numerically. The number and size of systems that can be studied is limited. Therefore, it is difficult for such simulations to provide a general view of folding dynamics. Here an intermediate level of model is sought giving the ability to directly simulate the folding process and broadly explore the connections between thermodynamics and kinetics (9, 18, 25–27).

Experiments also have been devised to explore the landscape. Methods such as dynamic NMR, protein engineering, laser initiated folding, and ultrafast mixing are being used (28–30). During the past 5 years it has become possible to study “burst phase kinetics,” which cannot be resolved within the dead time of conventional stopped-flow experiments. In some cases, e.g. apomyoglobin, substantial secondary structure and natively like tertiary contacts are observed to have already formed. Such burst phases could represent kinetics with a small or negligible activation barrier. Even in the case where

kinetics is limited by activation, much structure formation occurs by the subsequent downhill dynamics to the native state. These must be studied via fast folding experiments.

One experiment uses cold denaturation in supercooled water to prepare unfolded proteins (30). An infrared laser pulse then jumps the temperature by directly heating water 10–30° in 10 ns, placing the unfolded chain in an environment conducive to folding. The refolding is monitored by a series of UV pulses spaced 14 ns apart. The UV light excites tryptophan fluorescence, for which specific quenching mechanisms are introduced to serve as structural markers. Experiments on horse apomyoglobin show that the cold denatured state has some residual structure in the GH helices and loop; the N terminal A helix is largely destabilized at low temperature. The fast refolding experiment indicates that formation of the AGH core of apomyoglobin is an extremely efficient process (30). Refolding of phosphoglycerate kinase, a 415-residue two-domain enzyme, displays multiscale kinetics on a time scale from 10 to 6,000 microseconds, indicating the roughness of the free energy landscape. Experiments are beginning to build up a phase diagram of folding kinetics that can be used to test and refine theoretical models.

This work was supported by National Science Foundation Grant MCB-9603839 and National Institutes of Health Grants GM48807, RR12255, and GM44557.

- Harrison, P. M., Bamborough, P., Daggett, V., Prusiner, S. B. & Cohen, F. C. (1997) *Curr. Opin. Struct. Biol.* **7**, 53–59.
- Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660.
- Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48**, 545–600.
- Gō, N. (1983) *J. Stat. Phys.* **30**, 413–423.
- Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528.
- Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93**, 6902–6915.
- Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725.
- Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4**, 10–19.
- Guo, Z. & Thirumalai, D. (1996) *J. Mol. Biol.* **263**, 323–343.
- Šali, A., Shakhnovich, E. & Karplus, M. (1994) *J. Mol. Biol.* **235**, 1614–1636.
- Scheraga, H. A. (1992) *Protein Sci.* **1**, 691–693.
- Zwanzig, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9801–9804.
- Pande, V. S., Grosberg, A. Y. & Tanaka, T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 12972–12975.
- Boczek, E. M. & Brooks, C. L., III (1995) *Science* **269**, 393–396.
- Guo, Z. Y. & Thirumalai, D. (1995) *Biopolymers* **36**, 83–102.
- Mirny, L. A., Abkevich, V. & Shakhnovich, E. I. (1996) *Folding Design* **1**, 103–116.
- Friedrichs, M. S., Goldstein, R. A. & Wolynes, P. G. (1991) *J. Mol. Biol.* **222**, 1013–1034.
- Guo, Z. & Brooks C. L., III (1997) *Biopolymers* **42**, 745–757.
- Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3626–3630.
- Wolynes, P. G., Schulten, Z. L. & Onuchic, J. (1996) *Chem. Biol.* **3**, 425–432.
- Sheinerman, F. B. & Brooks C. L., III (1998) *J. Mol. Biol.* **278**, 439–455.
- Sheinerman, F. B. & Brooks C. L., III (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1562–1567.
- Guo, Z., Brooks C. L., III & Boczek, E. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10161–10166.
- Socci, N. D. & Onuchic, J. N. (1995) *J. Chem. Phys.* **103**, 4732–4744.
- Nymeyer, H., García, A. E. & Onuchic, J. N. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5921–5928.
- Shea, J.-E., Nochomovitz, Y. D., Guo, Z. & Brooks C. L., III (1998) *J. Chem. Phys.*, in press.
- Hardin, C., Luthey-Schulten, Z. A. & Wolynes, P. G. (1998) *Proteins Struct. Funct. Genet.*, in press.
- Fersht, A. R. (1997) *Curr. Opin. Struct. Biol.* **7**, 3–9.
- Eaton, W. A., Muñoz, V., Thompson, P., Chan, C. K. & Hofrichter, J. (1997) *Curr. Opin. Struct. Biol.* **7**, 10–14.
- Gruebele, M., Sabelko, J., Ballew, R. M. & Ervin, J. (1998) *Acc. Chem. Res.*, in press.