Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 7129–7131, July 1997

Commentary

Contacting the protein folding funnel with NMR

José Nelson Onuchic

Department of Physics, University of California at San Diego, La Jolla, CA 92093-0319

Proteins control most functions in living organisms. Because their activity depends on their three-dimensional shape, they must fold into well defined structures in times that are relatively short biologically. The importance and beauty of these complex three-dimensional structures as well as the amazing fact that they form spontaneously makes protein folding a central problem in today's science.

Seeking a theoretical framework able to explain the large collection of early experimental results in this field, scientists adopted the idea that there is a "folding pathway," i.e., a series of obligate intermediates in the protein dynamical route toward the folded state (1–3). Experiments were focused on identifying and structurally characterizing such intermediates without questioning the pathway hypothesis itself. An alternative, richer viewpoint had its birth in the late 1980s and early 1990s (4–6). In this new view, folding does not occur by a single pathway but instead by an ensemble of routes toward the native state. Globally the energy landscape of a folding protein resembles a partially rough funnel. The local roughness reflects transient trapping at non-native configurations, but the overall shape of the funnel is determined by the guiding forces that have evolved to stabilize the native interactions more than alternatives (7).

A major goal then is to understand the kinetic consequences of the funneled landscape as well as the molecular forces used in its evolutionary design. A good understanding of how the funnel topography can explain a multiplicity of possible folding kinetic scenarios has been achieved by simulating minimalist models, composed of necklaces of beads of different kinds. Even though much was learned qualitatively from several off-lattice simulations, the majority of these numerical investigations use lattice models (see the reviews in refs. 8–11 and references therein for more details). In these models, the dominant energies, which give the landscape its funnel shape for well designed sequences, involve the hydrophobic energy of making contacts between appropriate residues. Thus the part of the energy that guides folding is strongly correlated with a single order parameter Q, measuring the fraction of native contacts present at any given configuration. For well designed sequences, the folding kinetics can be seen as a diffusion in an effective one-dimensional potential described by the free energy as a function of this coordinate Q (12).

At this point the question arises: in real proteins, what determines the funnel shape? Is it also such three-dimensional (tertiary) contacts? Setting up a correspondence between real proteins and these lattice models, one needs to consider additional degrees of freedom, particularly secondary structure formation but perhaps also side-chain orientation. One idea is that these degrees of freedom simply "renormalize" the entropy and energy scales of the funnel, so that it can be compared with the lattice models having only contact energies. In an initial attempt, a simple version of this renormalization was performed with the help of a theory of the helix coil transition in collapsed heteropolymers (13). After renormalization, the theory suggests the often studied 27-mer lattice model corresponds in the laboratory to a 60-aa helical protein. Simulations of the corresponding lattice model predicted many features of the folding kinetics that later were confirmed. For example, a transition state located midway between the native and unfolded state that is composed of an ensemble of delocalized nuclei is found (13, 14). Another prediction was that the collapsed "molten globule" phase is an ensemble of structures of varying topologies but biased toward the native topology. In each molten globule structure, roughly one-quarter to one-third of native contacts were made in the corresponding lattice model. A recent all-atom simulation by Bockzo and Brooks (15) of a three-helix bundle protein (fragment B of staphylococcal protein A) strongly supported this view. What evidence has been generated by experiments in the laboratory? The work of Balbach and collaborators in this issue provides a key experiment (16).

Several experiments already have addressed the nature of the transition state ensemble. These include studies on folding of cytochrome *c* (17–19), λ-repressor (20), apomyoglobin (21), CI2 (22), and ChY (23), which are in agreement with the landscape view described above. They are all consistent with a transition state characterized by an ensemble of delocalized nuclei that is located midway between the folded and unfolded state. The transition state by its nature is fleeting. Information about the energetics and structures of members of the transition state ensemble can only be inferred from protein engineering by invoking extra-thermodynamic free energy relations (24) that have a venerable history in organic chemistry but also are expected from energy landscape theory (8, 14). Theoretical work using the funnel concept has been used to understand the details of the transition state ensemble (25), but the comparison depends on specific energetic models. The structural features and the energetic features are intertwined.

The molten globule state is more accessible to direct structural study. In this issue, the Oxford group describes a great tool to be used toward a more quantitative measure of the role of tertiary contacts in defining these energy landscapes in the molten globule region (16). Using an NMR-based strategy, they develop a technique to identify and quantify the fraction of the contacts in the molten globule ensemble. They use the nuclear Overhauser effect (NOE), which depends on making contacts between residues in three-dimensional space, at least transiently. The difficulty in interpreting NOE measurements of the globule is that the molten globule has an NMR spectrum with very low resolution. The elegant trick in these experiments is to use the high resolution of the native spectrum to study contacts in the molten globule. By transferring the NOEs from the molten globule to the native state, they can see which of the specific contacts exist in this molten globule and at least a bit about how often they are made. This strategy now can be used to probe the tertiary structure contacts of any desired partially folded state that is thermodynamically stable under certain conditions.

Does the molten globule ensemble inferred from experiments look like that of lattice models? A popular picture underplays the diversity of molten globule states (26). Very low-resolution experiments suggest that the globule has most of the native secondary structure already formed and a similar

topology to the native state, but the side groups are not yet rigidly packed. This view seems to be supported by NMR experiments in apocytochrome b_{562} (27). Using a more tedious approach, the NMR spectrum of the molten globule of this protein could be assigned, and a single structure was inferred using the same algorithms as for fully folded proteins. These algorithms assume constraints inferred by NMR simultaneously apply to all members of the ensemble whose properties are being measured. As pointed out earlier (13, 14), the molten globule state even in lattice simulations is biased toward the native structure, and it is likely that the application of these algorithms to the lattice ensemble would give such an unique topology, too. The direct measurements of NOE's reported by Balbach and collaborators now can verify this hypothesis. Their molten globule for α-lactalbumin (A-state) does show a substantial presence of native contacts, but it also shows a measurable presence of non-native contacts. The NOE spectrum of the molten globule and native protein are not just scaled versions of each other. The existence of non-native contacts indicates that topologies different from the native protein exist at this molten globule state, in direct contradiction with the unique topology view. As in the lattice models, collapsed configurations exhibiting substantially different tertiary structure may contain some of the same tertiary contacts as the native, but they are transient. As an illustration, Fig. 1 shows a few molten globule configurations for an off-lattice minimalist model of a small β protein.

The qualitative features of the Oxford results are exactly what would be predicted by the funnel picture. Because the funnel shape of the landscape already induces the formation of native contacts much before the transition state is reached, the molten globule is expected to contain a certain amount of native contacts. The "center" of the funnel is indeed the native state. The funnel bias induces the partial formation of native contacts even for the early downhill events. In the corresponding states model for the α-helical fast folding proteins, the molten globule was estimated to have around one-quarter to one-third of native contacts. It will be interesting to see if this fraction quantitatively agrees with what is observed in their experiments. Several delicate points, however, regarding the detailed structural averaging involved in the NOE experiments need to be clarified to obtain a precise number, but these initial results seem to present no contradiction with the lattice simulations. They already give enough evidence to start to understand the basic nature of the contacts in the molten globule state, but we must look forward to further improvements to provide quantitative information.

A most exciting prospect suggested by the energy landscape theory and the funnel picture is that these static structural experiments on the molten globule eventually can make contact with the protein engineering studies on the transition state ensemble (22, 24). By using controlled denaturation, different thermodynamically stable molten globule ensembles can be probed via NMR. It is possible that they may directly overlap the transition state ensemble obtained at other particular thermodynamic conditions. We can foresee the day when a complete structural map of the ensembles occupied on a protein folding funnel becomes available.

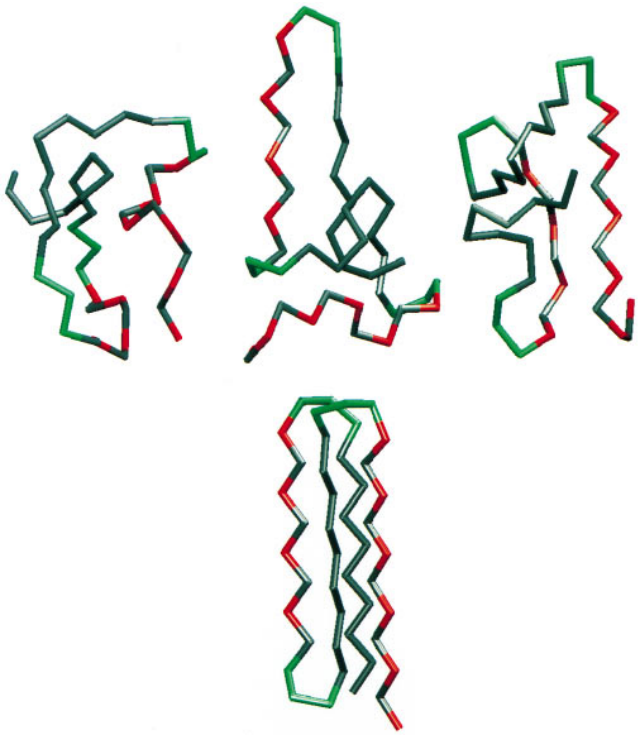

FIG. 1. An illustration of the molten globule for a minimalist model. (*Upper*) Three different configurations that are part of the molten globule ensemble for an off-lattice minimalist model of a small β-protein. (*Lower*) The native structure for this protein. Roughly one-quarter to one-third of the native contacts are made in each molten globule configuration but the three of them have different topologies. These configurations were generated using preliminary results for a minimalist model similar to the one described in refs. 11 and 28 that is being developed in collaboration with H. Nymeyer and A. Garcia. (This illustration was prepared with the help of Hugh Nymeyer.)

I would like to thank P. G. Wolynes for helpful editorial advice. Work at the University of California at San Diego was supported by the National Science Foundation (Grant MCB-9603839) and the University of California/Los Alamos Research (CULAR) Initiative.

1. Englander, S. W. & Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct.* **21,** 243–265.
2. Creighton, T. E. (1990) *Biochem. J.* **270,** 1–16.
3. Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59,** 631–660.
4. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
5. Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93,** 6902–6915.
6. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8721–8725.
7. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48,** 539–594.
8. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins* **21,** 167–195.
9. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995) *Protein Sci.* **4,** 561–602.
10. Mirny, L. A., Abkevich, V. & Shakhnovich, E. I. (1996) *Folding Design* **1,** 103–116.
11. Guo, Z. Y. & Thirumalai, D. (1995) *Biopolymers* **36,** 83–102.
12. Socci, N. D., Onuchic, J. N. & Wolynes, P. G. (1996) *J. Chem. Phys.* **104,** 5860–5868.
13. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3626–3630.
14. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. (1996) *Folding Design* **1,** 441–450.
15. Boczko, E. M. & Brooks, C. L. (1995) *Science* **269,** 393–396.
16. Balbach, J., Forge, V., Lau, W. S., Jones, J. A., van Nuland, N. A. J. & Dobson, C. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 7182–7185.
17. Jones, C. M., Henry, E. R., Hu, Y., Chan, C.-K., Luck, S. D., Bhuyna, A., Roder, H., Hofrichter, J. & Eaton, W. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11860–11864.
18. Mines, G. A., Pascher, T., Lee, S. C., Winkler, J. R. & Gray, H. (1996) *Chem. Biol.* **3,** 491–497.
19. Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. & Hofricher, J. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1779–1784.
20. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. (1997) *Nat. Struct. Biol.* **4,** 305–310.

21. Ballew, R. M., Sabelko, J. & Gruebele, M. (1996) *Nat. Struct. Biol.* **3,** 923–926.
22. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995) *J. Mol. Biol.* **254,** 260–288.
23. López-Hernández, E. & Serrano, L. (1996) *Folding Design* **1,** 43–55.
24. Fersht, A. R. (1997) *Curr. Opin. Struct. Biol.* **7,** 3–9.
25. Shoemaker, B. A., Wang, J. & Wolynes, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 777–782.
26. Ptitsyn, O. B. (1992) in *Protein Folding*, ed. Creighton, T. E. (Freeman, New York), p. 243.
27. Feng, Y., Sligar, S. G. & Wand, A. J. (1994) *Nat. Struct. Biol.* **1,** 30–36.
28. Honeycutt, J. & Thirumalai, D. (1992) *Biopolymers* **32,** 695–709.