Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 2099–2100, March 1993

Commentary

From independent modules to molten globules: Observations on the nature of protein folding intermediates

Jeffrey Skolnick, Andrzej Kolinski, and Adam Godzik

Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037

The study of Morjana *et al.* (1) on a partially unfolded intermediate of protein disulfide isomerase (PDI) raises a number of important points concerning the nature of folding intermediates of globular proteins. PDI is believed to possess two internally homologous domains (2). When subjected to guanidine hydrochloride denaturation, circular dichroism, and fluorescence spectroscopy indicate the presence of a partially folded equilibrium intermediate containing about 40% of the native state's secondary structure. Consistent with this amount of secondary structure, these authors suggest that the intermediate may result from unfolding of one of the two structural domains. Based on the fact that the far-UV spectrum is not similar to the native state, these authors argue that this equilibrium intermediate is not a molten globule, a folding intermediate state whose properties have been the subject of much attention and debate (3–5). While one might have expected that all intermediates have similar characteristics, as the PDI study points out, there are many types with rather diverse physical properties; the origins of these differences are not well understood. What differentiates them are the amount of native secondary structure and the degree of side-chain packing specificity. They range from fragments whose properties are essentially identical to the native molecule (6) to those where only a small portion of the molecule has loosely contacting native-like secondary structure (7). In what follows, we summarize what is known about the various types of folding intermediates, discuss what experimental information is required to better characterize them, and describe recent theoretical advances that may provide a fuller understanding of the protein folding process.

Perhaps the most expected intermediates are those where a domain or subdomain folds independently and adopts both the native secondary and tertiary structure within this folding unit; i.e., it constitutes a native-like folding module (8–10). PDI may be such a case. However, the best characterized of these native-like modules is the equilibrium folding intermediate seen on urea denaturation of the α-subunit of tryptophan synthase (6, 11) whose native conformation is a single-domain eight-member α/β-barrel (12). This intermediate is located at the N terminus and is composed of six of the eight β-strands plus associated helices. Moreover, the N-terminal fragment folds independently, with properties very similar to the equilibrium intermediate of the entire molecule. Possible insights into the origin of the stability of this intermediate may be obtained from the recent computer simulations of Godzik *et al.* (13). In these simulations, the local secondary-structure biases are assumed to be consistent with the folded state, but the folding pathway is predicted. These simulations correctly predict the structure of the equilibrium intermediate and indicate that the N-terminal fragment is substantially more stable than the corresponding C-terminal fragment. The equilibrium intermediate basically possesses the secondary structure of the native molecule, but the helices at the fragment interface oscillate and at times cover the exposed hydrophobic core of the molecule. Thus, they act as a cork to impede further assembly. These simulations predict that many, but not all, of the side-chain contacts are identical to those found in the native state. It would be very interesting to see whether this prediction can be verified by experiment.

More generally, it appears that the formation of secondary structure found in the native state and the adoption of well-defined patterns of side-chain contacts characteristic of folded proteins are interrelated but separate processes (3–5). Indeed, in the molten-globule state, the molecule is compact, with a volume about 20% larger than native, and has the native-like secondary structure but poorly defined tertiary contacts (10). A key point is whether the entire molecule adopts native-like secondary structure or whether a portion remains unfolded. In both cases, it is unknown whether or not there is a continuous or discrete spectrum of backbone conformations.

We first describe the situation when most, if not all, of the native-state secondary structure is present as an intact globule. One example is provided by barnase (14, 15), which prior to the rate-limiting step in folding exhibits an intermediate containing most of the secondary structure present in the native state. A *de novo*-designed four-helix bundle of DeGrado and coworkers (16, 17) is another example. The molecule contains a leucine core and is exceptionally resistant to thermal denaturation. It has well-defined secondary structure, but, as assessed by its ability to bind a hydrophobic dye, has poorly defined tertiary packing (18). Moreover, there is evidence that both right- and left-turning bundles are isoenergetic (T. M. Handel, S. A. Williams, and W. F. DeGrado, personal communication). Thus, this molecule has secondary-structure specificity but lacks side-chain packing specificity. Additional insights into the properties of this molecule are provided by a recent series of *de novo* folding simulations of this sequence that provide independent predictions in very good accord with experiment (19). Both right- and left-turning bundles are predicted to be isoenergetic. For each bundle type, the folded conformations lack uniquely defined tertiary interactions and move within a set of three distinct roughly isoenergetic structures. This characterization of the molten-globule state awaits experimental verification. Additional *de novo* simulations on a modified sequence, containing aromatic, isoleucine, and valine substitutions at 14 of the 73 sites (17), predict the existence of a native four-helix bundle that is unique, low-temperature, and right-turning. It has well-defined tertiary contacts and a 20% smaller volume than in the molten-globule state, in spite of the fact that bulkier aromatics are in the hydrophobic core. On the basis of the degree of agreement with experiment, we conclude that there are now theoretical tools available that will permit more detailed study of molten globules.

Finally, there are folding intermediates, which we term fragmentary globules, where only a portion of the molecule adopts native-like secondary structure, and the tertiary contacts between these elements are poorly defined. Examples include the two compact states of apomyoglobin existing at neutral and acidic pH (7, 20), the A form of α-lactalbumin (3, 21), the N- and C-terminal

helices in a cytochrome *c* intermediate (22–24), and a ribonuclease A intermediate (25, 26).

Given the wide variety of possible folding intermediates, more careful experimental characterization is required. It is essential to establish the amount and identity of the secondary structure present. Compactness alone is a poor criterion, as both intact and fragmentary globules can have similar hydrodynamic properties. One must determine whether or not the side-chain packing is well-defined. For native-like modules, studies of their folding kinetics should be undertaken to ascertain whether these intermediates themselves have kinetic intermediates that are fragmentary and/or intact globules. Furthermore, the cooperative process responsible for side-chain fixation needs to be better understood. *De novo* studies of designed molecules offer the possibility of exploring the packing requirements for side-chain fixation. For example, in the *de novo* proteins studied by Raleigh and DeGrado (17), one can engineer a variety of mutants that span the range of the 14 residue substitutions and attempt to ascertain which subset of these substitutions provides for well-defined side-chain packing. The newly developed *de novo* models of protein folding should prove very useful to guide experiment. Thus, in the next few years, the prognosis for making major progress in the understanding of protein folding is quite good.

1. Morjana, N. A., McKeone, B. J. & Gilbert, H. F. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2107–2111.
2. Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. & Rutter, W. J. (1985) *Nature (London)* **317**, 267–270.
3. Kuwajima, K. (1989) *Proteins* **6**, 87–103.
4. Kim, P. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660.
5. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990) *FEBS Lett.* **262**, 20–24.
6. Matthews, C. R. (1990) in *Protein Folding*, eds. Gierasch, L. M. & King, J. (AAAS, Washington, DC), pp. 192–197.
7. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990) *Science* **249**, 1544–1548.
8. Kim, P. & Baldwin, R. (1982) *Annu. Rev. Biochem.* **51**, 459–489.
9. Ptitsyn, O. B. & Rashin, A. (1973) *Dokl. Akad. Nauk SSSR* **213**, 473–475.
10. Ptitsyn, O. B. (1986) *J. Protein Chem.* **6**, 273–292.
11. Matthews, C. R. (1991) *Curr. Opin. Struct. Biol.* **1**, 28–31.
12. Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W. & Davies, D. R. (1988) *J. Biol. Chem.* **263**, 17857–17861.
13. Godzik, A., Skolnick, J. & Kolinski, A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2629–2633.
14. Matouschek, A., Serrano, L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 819–836.
15. Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 837–852.
16. Handel, T. M. & DeGrado, W. F. (1992) *Biophys. J.* **61**, A265.
17. Raleigh, D. P. & DeGrado, W. F. (1992) *J. Am. Chem. Soc.* **114**, 10079–10081.
18. Handel, T. M., Williams, S. A. & DeGrado, W. F. (1993) *Science*, in press.
19. Godzik, A., Kolinski, A. & Skolnick, J. (1993) *J. Comp. Aided Mol. Design*, in press.
20. Griko, Y. V., Privalov, P. L., Venyaminov, S. Y. & Kutyshenko, V. P. (1988) *J. Mol. Biol.* **202**, 127–138.
21. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989) *Biochemistry* **28**, 7–13.
22. Jeng, M.-F., Englander, S. W., Elove, G. A., Wand, A. J. & Roder, H. (1990) *Biochemistry* **29**, 10433–10437.
23. Briggs, M. S. & Heinrich, R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2017–2021.
24. Baldwin, R. L. & Roder, H. (1991) *Curr. Biol.* **1**, 218–220.
25. Udgaonkar, J. B. & Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699.
26. Udgaonkar, J. B. & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8197–8201.