

Stepwise helix formation and chain compaction during protein folding

Heinrich Roder*

Basic Science Division, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111; and Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104

One of the persistent problems in protein folding is understanding the relative importance of secondary and tertiary structure formation. Elements of secondary structure are stabilized primarily by interactions among amino acids that are close in sequence, whereas tertiary structure is supported by interactions among more distant segments of the chain. Therefore, it is reasonable to assume that folding is a hierarchical process with secondary structure preceding tertiary structure formation (1).

In fact, fluctuating elements of secondary structure often persist under denaturing conditions where the chain is otherwise disordered and devoid of specific tertiary interactions. Numerous observations of helical structure in protein fragments and synthetic peptides further support the conclusion that secondary structure formation is governed by the local amino acid sequence context. Conversely, secondary structure elements are marginally stable in isolation and require stabilizing tertiary interactions to become persistent. Under physiological conditions, hydrophobic interactions among apolar side chains favor collapse of the polypeptide chain into a compact conformation, suggesting that solvent-induced chain collapse is a critical early step in folding. Such a hydrophobic collapse is likely to be accompanied by formation of hydrogen-bonded secondary structure to avoid the energetically unfavorable burial of polar backbone atoms (2). In a recent issue of PNAS, Uzawa *et al.* (3) presented evidence that compact structural ensembles rich in secondary structure can appear early in folding. They followed the formation of helical secondary structure and the decrease in chain dimensions throughout the folding reaction of apomyoglobin, covering the time range from $\approx 300 \mu\text{s}$ to 1 s.

Important structural events occur on the microsecond time scale, which cannot be accessed by conventional kinetic techniques. Temperature jump and other rapid perturbation methods have shown that isolated helices and β -hairpins form and decay over a time window ranging from $\approx 100 \text{ ns}$ to several microseconds (4, 5). Recent advances in rapid mixing techniques combined with structurally informative spectroscopic probes made it possi-

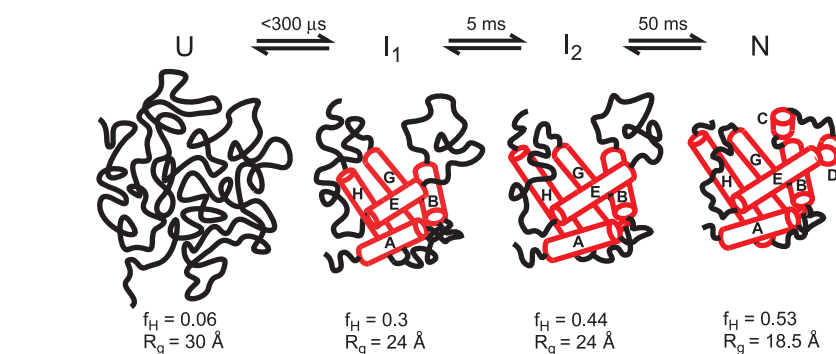


Fig. 1. Sequential four-state mechanism involving two partially structured ensembles, I_1 and I_2 , in addition to the acid-unfolded (U) and native (N) states. Cartoons of the different states are consistent with the fraction of helical residues (f_H) and radius of gyration (R_g) reported in ref. 3. Cylinders indicate the approximate position of α -helices in the native state and possible arrangement of core helices in the intermediates consistent with amide protection data (17).

ble to resolve conformational events on the submillisecond time scale preceding the rate-limiting step in the folding of globular proteins (6–9). CD spectroscopy in the far-UV region provides valuable information on secondary structure content, but low inherent sensitivity and various artifacts have limited the resolution of stopped-flow CD measurements to the 10-ms time range (10). Akiyama *et al.* (11) were able to extend the time resolution down to the 400- μs range by coupling an efficient turbulent mixer with a CD spectrometer. Their continuous-flow measurements of CD spectral changes in the far-UV region revealed the formation of (helical) secondary structure during the second and third (final) stages of folding of cytochrome *c*. More recently, Akiyama *et al.* (12) designed a mixing device with a dead time of $\approx 160 \mu\text{s}$ for continuous-flow small-angle x-ray scattering (SAXS) measurements by using synchrotron radiation. This heroic experiment (both in terms of technical difficulty and amounts of protein consumed) made it possible to follow the changes in size and shape associated with the various stages of folding of cytochrome *c* (9, 13).

Uzawa *et al.* (3) relied on their capability of measuring secondary structure and molecular dimensions with submillisecond time resolution to explore the folding of horse apomyoglobin. On extracting the heme, the apo-form of myoglobin retains its tightly packed globular structure ($R_g = 18.5 \text{ \AA}$) along with seven α -helical seg-

ments (labeled A–E, G, and H; the F helix is disordered in the absence of the heme). By combining continuous-flow CD measurements with conventional stopped-flow data, the authors followed the progress of helix formation during refolding of acid-denatured apomyoglobin over a time window from 300 μs to 1 s. The changes in the CD signal at 222 nm are consistent with a stepwise increase in helix content in three kinetic phases, including a major process occurring within the instrumental dead time and two observable phases with time constants of 5 and 50 ms, respectively. The SAXS measurements show a major collapse of the chain within the 300- μs dead time, followed by a further decrease in chain dimensions during the final (50-ms) folding phase.

Uzawa *et al.* (3) interpret their findings in terms of a four-state folding mechanism (Fig. 1), based on their observation of three distinct folding phases. Fig. 1 also shows cartoons of the various states populated during apomyoglobin folding consistent with the size and secondary structure information obtained in the present study, as well as more detailed information on the location of stable helical segments obtained in previous NMR-based studies (14, 15). The acid-unfolded state, U, contains

See companion article on page 1171 in issue 5 of volume 101.

*E-mail: roder@fccc.edu.

© 2004 by The National Academy of Sciences of the USA

only a small amount of α -helical structure detectable by CD and is substantially expanded compared to the native state. However, detailed NMR studies (16) show clear deviations from the behavior expected for a random coil, including evidence for fluctuating residual secondary structure and restricted backbone motions. The high helix content of the I_1 intermediate ($\approx 60\%$ of the native level) indicates that it contains helical segments in regions outside the A, G, and H helix core observed initially by hydrogen exchange labeling (14). Likely candidates are the B and E helices, which are partially solvent-protected in an ensemble of early intermediates (15). The five folded helices correspond to peaks in the hydrophobicity plot of the myoglobin sequence, whereas the still unfolded C and D helices are among the most polar regions of the protein. Recent amide protection data on mutant proteins indicate that the early intermediate contains primarily native-like helix-helix interactions (17). The shape information extracted from the SAXS measurements at early folding times further supports a bipartite structure with a compact core comprising the bulk of the chain and some disordered segments (3). The I_2 intermediate formed during a later stage of folding is characterized by a further increase in helix content consistent with an increase in the length of existing helices. However, the overall size and shape consistent with a partially hydrated conformation containing some disordered loops remains unchanged at this stage.

Uzawa *et al.* (3) clearly demonstrate that major structural events occur within the first few hundred microseconds of initiating the apomyoglobin folding reaction. Secondary structure formation and chain compaction appear to be coupled during the early stages of folding. This conclusion is consistent with continuous-flow fluorescence results by Jamin *et al.*

(18) indicating that a partially structured equilibrium state with properties similar to the I_1 intermediate exhibits two-state folding/unfolding kinetics with a folding time of $\approx 50 \mu\text{s}$ in the absence of denaturant. Cytochrome *c*, the only other protein whose submillisecond folding kinetics has been followed by CD and SAXS, also shows a large decrease in overall dimensions on a 100- μs time range (8, 9, 12), but the helix content of the early intermediate is lower than that of apomyoglobin (11). Thus, the exact balance between chain compaction and secondary structure is likely to vary considerably from one protein to another.

How unique is the folding scheme shown in Fig. 1? A sequential mechanism with two on-pathway intermediates appears to be the simplest kinetic mechanism consistent with the experimental data. However, it is likely that this simple scheme will have to be expanded to account for structural heterogeneity at the early (I_1) stage observed in hydrogen exchange and T jump experiments (15, 19). Uzawa *et al.* (3) argue that a mechanism involving two intermediates along parallel pathways can be ruled out, because this would predict more rapid accumulation of native-like species than observed. Similar reasoning was previously used to argue against off-pathway mechanisms on the basis of MS analysis of hydrogen-deuterium exchange data (17, 20). However, a rigorous distinction between on- and off-pathway mechanisms is possible only if the experimental probe used to follow the time course can discriminate between intermediate and native states, and accumulation of intermediate and native states is kinetically coupled. Given these stringent conditions, there are very few examples of bona fide on-pathway intermediates (21, 22). In the case of apomyoglobin, the first condition is fulfilled, because the native state has a characteristic signature in the SAXS

and mass-based H/D exchange data and can be clearly distinguished from intermediates. However, the second condition is not satisfied because the early intermediates appear on a much shorter time scale than the fully folded state and, in the case of I_1 , are kinetically not resolved. Thus, a rapid preequilibrium is established between U and I_1 , and it is not possible to determine whether the intermediate lies on or off the main folding pathway. Nevertheless, the sequential mechanism offers a physically plausible scenario involving stepwise acquisition of helical structure and monotonous decrease in molecular size (Fig. 1) and is thus structurally more appealing than some of the formally possible alternatives.

The folding pathway of apomyoglobin (Fig. 1) is a striking example of a complex folding mechanism where a series of conformational ensembles with increasing structural organization precede the final (rate-limiting) step in the acquisition of the native structure. The earliest intermediate observed already shows clear indications of a partially structured core giving rise to nonspherical contributions to the x-ray scattering profile, a strong helical far-UV CD component and solvent-shielded amide protons in a cluster of interacting α -helices (3, 17). These properties are far from random, arguing against the notion that rapid conformational changes reflect a random collapse of the chain driven by nonspecific hydrophobic interactions (23). For all but the simplest proteins, the rate-limiting transition state for folding cannot be reached in a single, concerted step, resulting in transient accumulation of partially structured conformations. Rapid secondary structure formation and concomitant compaction of the chain, guided by a few critical tertiary contacts, may limit the conformational space to be searched during folding, providing the bias toward the native structure necessary for efficient folding on a relatively smooth energy surface (24).

- Baldwin, R. L. & Rose, G. D. (1999) *Trends Biochem. Sci.* **24**, 26–33.
- 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.
- Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I. & Fujisawa, T. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 1171–1176.
- Callender, R. H., Dyer, R. B., Gilmanshin, R. & Woodruff, W. H. (1998) *Annu. Rev. Phys. Chem.* **49**, 173–202.
- Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. & Hofrichter, J. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
- Takahashi, S., Yeh, S.-R., Das, T. K., Chan, C.-K., Gottfried, D. S. & Rousseau, D. L. (1997) *Nat. Struct. Biol.* **4**, 44–50.
- Shastry, M. C. R., Luck, S. D. & Roder, H. (1998) *Biophys. J.* **74**, 2714–2721.
- Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10115–10117.
- Shastry, M. C. R. & Roder, H. (1998) *Nat. Struct. Biol.* **5**, 385–392.
- Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Nagamura, T. (1987) *FEBS Lett.* **221**, 115–118.
- Akiyama, S., Takahashi, S., Ishimori, K. & Morishima, I. (2000) *Nat. Struct. Biol.* **7**, 514–520.
- Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1329–1334.
- Hagen, S. J. & Eaton, W. A. (2000) *J. Mol. Biol.* **297**, 781–789.
- Jennings, P. A. & Wright, P. E. (1993) *Science* **262**, 892–895.
- Nishimura, C., Dyson, H. J. & Wright, P. E. (2002) *J. Mol. Biol.* **322**, 483–489.
- Yao, J., Chung, J., Eliezer, D., Wright, P. E. & Dyson, H. J. (2001) *Biochemistry* **40**, 3561–3571.
- Nishimura, C., Wright, P. E. & Dyson, H. J. (2003) *J. Mol. Biol.* **334**, 293–307.
- Jamin, M., Yeh, S. R., Rousseau, D. L. & Baldwin, R. L. (1999) *J. Mol. Biol.* **292**, 731–740.
- Gulotta, M., Gilmanshin, R., Buscher, T. C., Callender, R. H. & Dyer, R. B. (2001) *Biochemistry* **40**, 5137–5143.
- Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J. & Wright, P. E. (1999) *Protein Sci.* **8**, 45–49.
- Walkenhorst, W. F., Green, S. M. & Roder, H. (1997) *Biochemistry* **36**, 5795–5805.
- Capaldi, A. P., Shastry, R. M. C., Kleanthous, C., Roder, H. & Radford, S. E. (2001) *Nat. Struct. Biol.* **8**, 68–72.
- Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002) *J. Mol. Biol.* **324**, 359–371.
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins* **21**, 167–195.