85-290 (1985), p. 612; J. S. Rinehart, *Geysers and Geothermal Energy* (Springer-Verlag, New York, 1980). Seismicity can also create geysers [G. D. Marler and D. E. White, *Geol. Soc. Am. Bull.* **86**, 749 (1975)].
9. C. R. Faust and J. W. Mercer, *Water Resour. Res.* **15**, 23 (1979).
10. ———, *ibid.*, p. 31. We used a revised version of the Faust and Mercer algorithm for heat and mass transport provided by D. O. Hayba (written communication).
11. It would be difficult to determine an appropriate capillary-pressure function for this system; relevant data are limited. The assumption of fluid-rock thermal equilibrium has a minor effect on our simulations. For the simulation depicted in Figs. 3 and 4, cyclic temperature variations range from ~2°C at 190 m in depth to ~10°C at 10 m in depth. About 7×10^5 kJ per cycle is exchanged with the rock (relative to a throughflow rate of $\sim 7 \times 10^6$ kJ per cycle). Models that explicitly calculate fluid-rock heat exchange would tend to predict lesser values, depending on the assumed fracture spacing.
12. The injection of pure heat over much of the range of simulated rates (0.0025 to 5.0 MW per 0.9 m^2) seems physically implausible; focused steam upflow and condensation are a more likely heat-input mechanism. The geyser cycles shown in Figs. 3 and 4 involved the injection of 2.5 MW, which could be supplied by <1 kg s^{-1} (0.893 kg s^{-1}) of steam at 215°C (2800 kJ kg^{-1}). Nearly identical geyser cycles were obtained when saturated steam was injected at the base of the column. Basal heat input rates as low as 0.0025 MW, corresponding to <0.001 kg s^{-1} of steam injection, were sufficient to induce periodic behavior (Fig. 6C), albeit with drastically reduced eruption magnitudes.
13. For the simulations in Figs. 2 and 3, the mass discharge is about 11,000 kg per eruption, or 30,000 kg per hour, similar to that estimated for Old Faithful, Yellowstone (24,000 kg per hour; R. A. Hutchinson, oral communication). The time-averaged heat discharge is 5.5 MW, also similar to the Old Faithful rate, calculated on the basis of an estimated steady recharge rate of 6 kg s^{-1} (*2*) from a liquid reservoir at 215°C (*14*). The simulated heat discharge is much larger than the heat input at the base of the fracture zone (2.5 MW) because about 3.0 MW is obtained from recharge water at 100°C.
14. R. O. Fournier, *Annu. Rev. Earth Planet. Sci.* **17**, 13 (1989).
15. The period is also affected by the choice of relative permeability functions. Compared to the linear functions used here, "fracture-flow" functions increase eruption frequency and Corey-type functions decrease frequency. The fracture-flow functions assume that relative permeability to steam increases rapidly as liquid saturation decreases [M. L. Sorey, M. A. Grant, E. Bradford, *Water Resour. Res.* **16**, 767 (1980)], whereas with Corey-type functions steam is relatively immobile at moderate to high liquid saturations [A. T. Corey, *Soil Sci. Soc. Proc.* **21**, 7 (1957)].
16. There is field evidence for a deep root to some geyser activity. Enthalpy-chloride and enthalpy-silica relations indicate that the water discharged from Old Faithful, Yellowstone, has boiled adiabatically from temperatures >200°C (equivalent to depths of ≥180 m) (*14*). Further, instruments placed in a geysering geothermal well in Guatemala indicated that boiling occurs to depths of at least 800 m during geyser-like discharge periods [D. E. Michels, *Proc. Workshop Geotherm. Reservoir Engl.* **16**, 255 (1991)].
17. Under unusual conditions, it is conceivable that the permeability of the surrounding matrix would be affected. For example, if essentially all recharge to the geyser conduit occurs through thin (2-mm), subhorizontal, widely spaced (50-m) fractures, and a vertical strain of 10^{-6} is wholly accommodated by this fracture set, the fracture permeability of the matrix ($\sim 10^{-11}$ m^2) will vary by ~10%.
18. S. A. Rojstaczer and D. C. Agnew, *J. Geophys. Res.* **94**, 12403 (1989).
19. We ignored atmospheric loading effects and tested whether geyser periodicity was influenced by the magnitude of atmospheric pressure at the land surface. We varied atmospheric pressure between 0.6 and 1.0 bar, holding other parameters constant, and found no significant change in eruption frequency (Fig. 6E).
20. Permeability increases associated with seismicity have been documented in the region near the Loma Prieta earthquake (10/17/89, **M** 7.1) [S. A. Rojstaczer and S. C. Wolf, *Geology* **20**, 211 (1992)]. Silver and Valette-Silver (*7*) proposed that the response of Calistoga "Old Faithful" to Loma Prieta and other remote earthquakes is due to changes in permeability or pore volume in the rock matrix adjacent to the geyser conduit.
21. D. P. Hill *et al.*, *Science* **260**, 1617 (1993).
22. We thank J. D. Bredehoeft, R. O. Fournier, P. A. Hsieh, and two anonymous referees for helpful reviews of earlier versions of the manuscript, and D. R. Jones for drafting the figures.

27 July 1993; accepted 3 September 1993

Formation of a Molten Globule Intermediate Early in the Kinetic Folding Pathway of Apomyoglobin

Patricia A. Jennings and Peter E. Wright*

Hydrogen exchange pulse labeling and stopped-flow circular dichroism were used to establish that the structure of the earliest detectable intermediate formed during refolding of apomyoglobin corresponds closely to that of a previously characterized equilibrium molten globule. This compact, cooperatively folded intermediate was formed in less than 5 milliseconds and contained stable, hydrogen-bonded secondary structure localized in the A, G, and H helices and part of the B helix. The remainder of the B helix folded on a much slower time scale, followed by the C and E helices and the CD loop. The data indicate that a molten globule intermediate was formed on the kinetic folding pathway.

Molten globules have been postulated as universal folding intermediates formed on the kinetic folding pathways of all proteins (*1*). The term "molten globule" describes a partially folded state with a molecular volume intermediate to that of the native and the unfolded states, a relatively high content of secondary structure, and few fixed tertiary interactions (*1–3*). Two-dimensional (2D) NMR studies of these intermediate species (*4–6*) have been used to establish the presence of elements of hydrogen bonded secondary structure in which amide protons are protected from exchange. Although equilibrium molten globules have been detected for a few proteins under appropriate conditions of pH, temperature, and ionic strength (*2, 3, 7–9*), it has been unclear whether this species actually participates in kinetic folding events.

Apomyoglobin (apoMb) (Fig. 1) is one of the few proteins studied that forms an equilibrium molten globule. The molten globule state is populated at low pH and temperature (*8, 10, 11*), is compact, and contains substantial secondary structure (~35% helix) that appears to be largely associated with folding of the A, G, and H helices (*4*). To investigate the kinetic folding pathway(s) and the role of the molten globule in folding, we have undertaken stopped-flow circular dichroism (CD) and hydrogen exchange pulse-labeling studies of the apoMb refolding reaction.

ApoMb is well suited for studies of folding mechanisms. The absence of the heme group makes folding unimolecular, the protein is stable under various conditions, and the NMR spectrum of the carbon-monoxy complex of the holoprotein has been assigned (*12*). Although the structure of apoMb has not been determined, all available data (*4, 10, 13, 14*) indicate that it has a compact hydrophobic core that resembles the tertiary structure of the holoprotein. The differences between the apo and holoproteins appear to be largely confined to the F helix region, which abuts the heme and provides the proximal histidine ligand (*4, 14*).

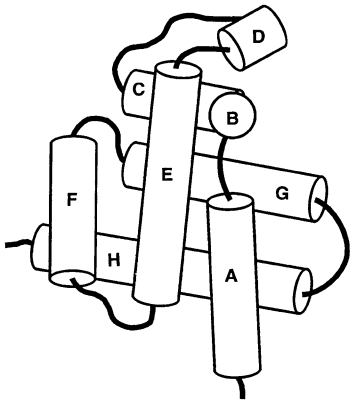

Fig. 1. Scheme of Mb with the eight α helices (A to H) indicated.

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

*To whom correspondence should be addressed.

Circular dichroism provides a sensitive probe of apoMb folding. The urea-induced unfolding reaction of apoMb is cooperative with a midpoint of ~2.2 M urea (Fig. 2). The unfolding transition is fully reversible, as judged by the complete recovery of the native helical content when the urea was diluted from 6 to 0.8 M and by the ability of the refolded apoprotein to bind the heme group quantitatively. In a representative kinetic trace obtained by monitoring the change in ellipticity as a function of refolding time (Fig. 2A), the ellipticity at the endpoint of the refolding reaction was the same as that observed for the native apoprotein in 0.8 M urea. The transient was well fitted by a single exponential decay with a relaxation time of 943 ± 20 ms. The stopped-flow CD measurements indicate that the protein was completely refolded within 4 s and that a substantial fraction (64%) of ellipticity at 222 nm was regained in a burst phase within the dead time of the instrument (5 ms). The burst phase was kinetically uncoupled from subsequent folding events by a difference in rate of two to three orders of magnitude; thus, the initial amplitude can be taken as a measure of the concentration of the species formed within the first 5 ms (*15*, *16*). Measurement of the burst phase ellipticity as a function of final denaturant concentration (Fig. 2B) shows that the folding of this species is cooperative. However, the transition is distinct from that observed for unfolding of the native protein in both amplitude and urea dependence, demonstrating that the burst phase species is an intermediate and not a fractional population of the native protein. A two-state fit (*15*) to the data of Fig. 2B gives an apparent stability of 2.5 ± 0.5 kcal/mol for the intermediate relative to the unfolded protein. The CD spectrum of the burst phase intermediate (Fig. 2C) shows the formation of helix.

To investigate further the kinetic folding pathway of apoMb, rapid mixing procedures coupled with amide proton exchange were used to label structured intermediates, with subsequent analysis of the data by 2D NMR methods (*17*, *18*). The refolding of apoMb was studied in the absence of heme (*19*); the protein was then reconstituted (*20*) and NMR spectra were recorded. The intensities of assigned (*12*) crosspeaks in both homonuclear double quantum and NOESY spectra (*21*) were measured to determine the extent of amide proton exchange at variable refolding time periods. Results obtained from both data sets are in

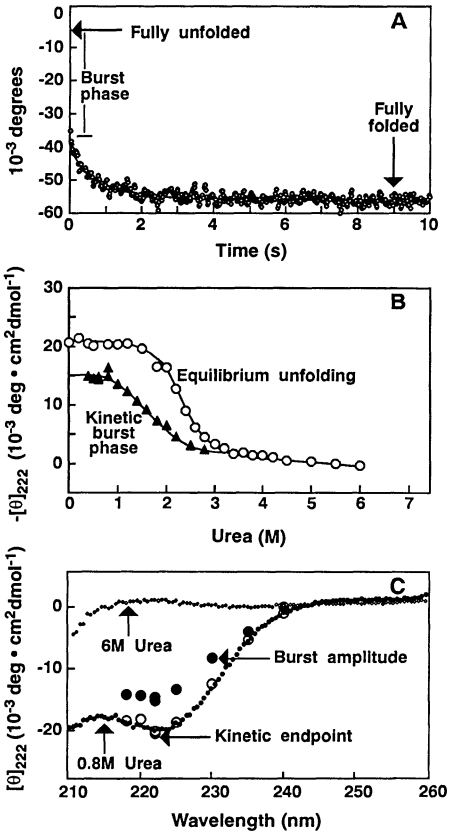

Fig. 2. (**A**) Refolding kinetics of apoMb at pH 6.1 and 5°C monitored by stopped-flow CD spectroscopy at 222 nm (*33*). The ellipticity of the unfolded protein (left arrow) was obtained by linear extrapolation of data collected above 4 M urea in equilibrium CD experiments. The ellipticity of the folded protein (right arrow) in 0.8 M urea was determined from equilibrium experiments. The solid line is a single exponential fit of the data. (**B**) The equilibrium unfolding of apoMb (○) monitored by the change in ellipticity at 222 nm as a function of urea concentration at pH 6.1 and 5°C. The burst phase amplitudes $[\Theta]_0$ observed in refolding experiments are also indicated (▲). (**C**) CD spectra of the burst phase intermediate (●) and refolded (○) protein generated by varying the wavelength of detection during stopped-flow refolding experiments. Spectra of the native protein in 0.8 M urea (lower curve) and unfolded protein in 6.0 M (upper curve) urea under equilibrium conditions are also shown.

Table 1. Amide proton protection data.

Residue† (helix)	Rate (apoMb)	Partner*	Region involved
L9 (A7)	F (<5 ms)	5–9	
V10 (A8)	F	6–10	
L11 (A9)	F	7–11	
W14 (A12)	F	10–14	
V17 (A15)	F	13–17	A3-A15
L29 (B10)	F	25–29	
I30 (B11)	F	26–30	
F33 (B14)	F	29–33	B6-B14
K102 (G4)	F	102‡	
L104 (G5)	F	100–104	
I107 (G8)	F	103–107	
A110 (G11)	F	106–110	
I112 (G13)	F	108–112	
H113 (G14)	F	109–113	
V114 (G15)	F	110–114	
L115 (G16)	F	111–115	G1-G16
K133 (H10)	F	129–133	
F138 (H15)	F	134–138	
I142 (H19)	F	138–142	
A143 (H20)	F	139–143	H6-H20
E18 (A16)	M (~1 s)	14–18	A16
I28 (B9)	M	24–28	
R31 (B12)	M	27–31	B9-B12
K34 (B15)	S (~2.5 s)	30–34	B15
H36 (C1)	S	32–36	
T39 (C4)	S	35–39	
L40 (C5)	S	36–40	C1-C5
F43 (CD1)	S	40–43	CD1
H64 (E7)	S	60–64	
V66 (E9)	S	62–66	
V68 (E11)	S	64–68	
L69 (E12)	S	65–69	
T70 (E13)	S	66–70	
L72 (E15)	S	68–72	
A74 (E17)	S	70–74	
I75 (E18)	S	71–75	
L76 (E19)	S	72–76	
K77 (E20)	S	73–77	E3-E20

*The hydrogen bonding partners were determined in the neutron diffraction study of the holoMbCO complex (*34*). †Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. Both residue number and helix position are given for clarity. ‡The backbone amide proton of K102 is involved in a hydrogen bond with a water molecule and is completely buried in the hydrophobic interior of the protein (*34*).

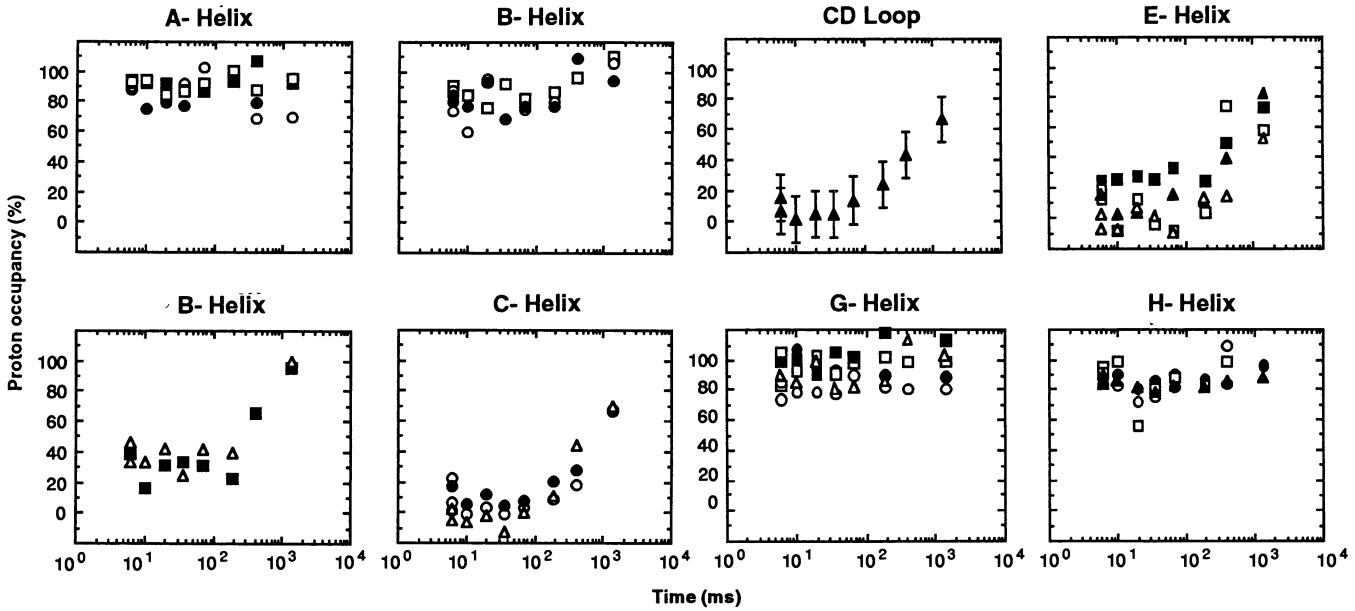

Fig. 3. Representative data for the time course of protection of backbone amide protons from exchange with solvent deuterons. Proton occupancies as a function of the refolding time before application of the labeling pulse are plotted. Representative experimental errors are indicated for the CD loop. Separate determinations of the peak heights for duplicate experiments differed on average less than 20%.

excellent agreement. Thirty-eight amide protons, well distributed throughout the protein, exchange sufficiently slowly under the experimental conditions to be used as probes of the folding process (Table 1). A substantial fraction of the available probes become protected (Fig. 3) at the shortest delay achievable under the conditions of the pulse labeling experiments (6.1 ms).

The amide protons of residues from the A, G, and H helices are fully protected within 6.1 ms after exposure to refolding conditions (Fig. 3 and Table 1). The amide protons of L29, I30, and F33 in the B helix also exhibit complete protection within 6.1 ms. Additional residues in the B helix (I28 and R31) become fully protected within 1 second after initiation of refolding. The amide protons of the residues at the COOH-terminal ends of helices A (E18) and B (K34) show protection kinetics slightly slower than other amides in these helices, possibly reflecting helix fraying before the formation of the native fold (6, *22*). Finally, the amide protons of the C and E helices become highly protected from exchange only after ~2 s and demonstrate very similar protection kinetics (Fig. 3). The observation of protection profiles for the amide protons in the A, G, H, and B helices which are similar to one another yet are kinetically distinct from those observed for the rest of the protein provides clear indication of the regions of the protein that participate in the cooperatively folded kinetic intermediate observed by CD.

Several residues in the B and E helices show partial protection (~20 to 30%) from exchange at the earliest time point (Fig. 3). Partial protection can arise in several ways as follows: the residues are moderately protected via a loose association with the folded subdomain, a folding phase occurs in the same time frame as the length of the labeling pulse (*23*), or there is a bifurcation of the folding pathway such that a proportion of protein molecules folds completely and a second subset folds more slowly (*24*). This latter mechanism seems unlikely since amide protons in the C helix and C-D loop and much of the E helix show no protection at the earliest time points (*25*). A more likely explanation involves a dynamic process in which helix B and part of the E helix become associated with the folded subdomain in a folding phase that is in the same time frame as the labeling pulse (*23*).

The kinetic pathway for folding of apoMb involves an intermediate species that strongly resembles the equilibrium molten globule state previously characterized at low pH and temperature. The stopped-flow CD studies indicate that the first detectable event in the folding of apoMb, occurring within the 5 ms dead time, is the formation of an intermediate containing substantial helix. This burst-phase intermediate has a helix content (35%) equal to that of the equilibrium molten globule (*4*, *10*, *11*). The hydrogen exchange pulse-labeling experiments also show formation of a transient intermediate in less than 6.1 ms. A substantial fraction of the available probes, localized in the A, G, and H helices and a region of the B helix, are completely protected at this time. The same amide protons are protected in the equilibrium molten globule state (*4*), providing strong evidence for the similarity of the two species. Although similar hydrogen exchange protection patterns are observed in both the equilibrium and kinetic folding intermediates, the protection factors are greater in the transient species. It is important to note that the stability of the equilibrium and kinetically populated species can vary due to the necessary differences in solution conditions. Indeed, it has recently been shown that the apoMb equilibrium molten globule state is ~1 kcal/mol more stable at neutral pH, corresponding to the conditions used for our kinetic studies, than at pH 4.2, where it is maximally populated (*26*).

Hydrogen exchange measurements for several proteins under native conditions indicate that the most slowly exchanging amide protons are both involved in hydrogen bonds and are buried in hydrophobic cores (*23*). Despite concerns of non-native hydrogen bond formation in folding intermediates (*27*), available evidence supports protection from exchange as a result of native-like secondary structure formation (*28*). In the case of apoMb, the complementary data from stopped-flow CD and NMR pulse labeling experiments provide particularly compelling evidence that the molten globule folding intermediate contains helix and that the amide protons are protected because of stabilized secondary structure formation. The helix content (~30%) estimated by assuming that all protected amide protons in the folding intermediate are in helical structures is in excellent agreement with the helicity measured by CD (~35%). This species is coop-

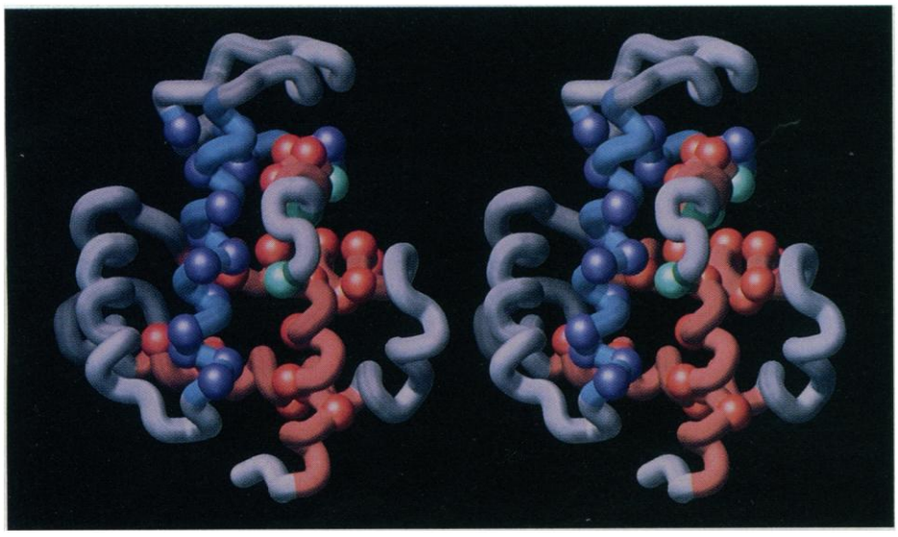

Fig. 4. Stereoview of the backbone of Mb (*34*) showing the regions protected within 6.1 ms (red), 1 s (green), and 2.5 s (blue). The gray coloring indicates regions where amide protons in the native protein exchange too rapidly to be used as probes of folding. Colored spheres indicate the positions of the amide proton probes. The colored backbone indicates the regions inferred to participate in formation of helical hydrogen bonds. The orientation is identical to that of the schematic in Fig. 1.

eratively folded (Fig. 2B), a probable result of both helix formation and interactions between helices. These conclusions are supported by NMR and CD studies of model peptides derived from the sequence of myoglobin, which indicate that the H helix can form in isolation but is only marginally stable under folding conditions and provides little protection of amide protons from exchange. Further stabilization of helical secondary structure in the G and H sequences appears to require formation of long-range tertiary hydrophobic interactions within a compact state (*29*).

The hydrogen exchange data offer an exquisitely detailed view of the folding pathway of apoMb. The locations of the secondary structural elements that are progressively stabilized during folding are indicated in Fig. 4. The A, G, and H helices fold first and are stabilized within a compact molten globule intermediate formed in less than 5 ms. The amide protons of residues L29, I30, and F33 in the B helix are also protected on this time scale. With the exception of the COOH-terminal end (K34), the remainder of the B helix folds within 1 s, apparently by assembly onto the already folded A-G-H helical core. The C helix, C-D loop, and E helix fold even more slowly, the estimated time for complete protection of amide protons from exchange being ~2.5 s. Association of the B helix with the A-G-H intermediate appears to be necessary prior to packing and stabilization of the C and E helices. Thus, the dominant pathway of folding of apoMb is

$$U \rightarrow A{\cdot}G{\cdot}H \rightarrow A{\cdot}B{\cdot}G{\cdot}H \rightarrow A{\cdot}B{\cdot}C{\cdot}CD{\cdot}E{\cdot}G{\cdot}H \rightarrow N$$

No information is available on the folding kinetics of the D and F helices since, in native holoMb, neither of these contains slowly exchanging amide proton probes. The kinetic intermediate has many of the characteristics of a molten globule: secondary structure sufficiently stable to allow complete protection of amide protons from exchange, and formation of a hydrophobic core and loose hydrophobic interactions. While no direct information is available on the molecular volume of the kinetic intermediate relative to the native and unfolded states, we may nevertheless infer that this species closely resembles what is commonly recognized as a molten globule.

Myoglobin has been the subject of several theoretical studies that explore the stabilization of the fold and have been extrapolated to give predictions of the folding pathway (*30*). These studies assume that the initial steps in folding involve the formation of helices, followed by assembly of helical substructures. Although none of the predicted folding pathways are in complete agreement with our experimental results, several groups have postulated formation of the G–H helix pair at an early stage of folding based on considerations of buried surface area. While recent studies of model peptides clearly indicate that the G–H helical hairpin is not independently stable (*29*), it may well be formed transiently prior to formation of the A–G–H molten globule intermediate observed on the kinetic folding pathway.

A generalized model of protein folding has been proposed in which folding proceeds by way of a molten globule intermediate (*2, 3, 7–9* and references therein). Until now, the principal evidence in support of this hypothesis has come from optical studies of α-lactalbumin refolding (*16*), but detailed structural characterization of the observed kinetic folding intermediate has not yet been realized. Although all of the proteins studied previously by hydrogen exchange pulse-labeling methods form folding intermediates that contain stable, hydrogen bonded secondary structure (*17, 18, 31, 32*), it was not established unambiguously that these intermediates were structurally equivalent to the equilibrium molten globule. Our data, however, provide strong evidence in support of the molten globule folding model by clearly establishing a close structural similarity between the earliest detectable kinetic folding intermediate and the equilibrium molten globule state.

REFERENCES AND NOTES

1. O. B. Ptitsyn, *Vestn. Akad. Nauk SSSR* **5**, 57 (1973).
2. H. Christensen and R. H. Pain, *Eur. Biophys. J.* **19**, 221 (1991).
3. K. Kuwajima, *Proteins Struct. Funct. Genet.* **6**, 87 (1989).
4. F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **249**, 1544 (1990).
5. M.-F. Jeng, S. W. Englander, G. A. Elöve, A. J. Wand, H. Roder, *Biochemistry* **29**, 10433 (1990); J. Baum, C. M. Dobson, P. A. Evans, C. Hanley, *ibid.* **28**, 7 (1989).
6. M.-F. Jeng and S. W. Englander, *J. Mol. Biol.* **221**, 1045 (1992).
7. O. B. Ptitsyn, *FEBS Lett.* **285**, 176 (1991).
8. _____, R. H. Pain, G. V. Semisotnov, E. Zerovnik, O. I. Razgulyaev, *ibid.* **262**, 20 (1990).
9. R. L. Baldwin, *Ciba Found. Symp.* **161**, 190 (1991).
10. Y. V. Griko, P. L. Privalov, S. Y. Venyaminov, V. P. Kutyshenko, *J. Mol. Biol.* **202**, 127 (1988).
11. F. M. Hughson and R. L. Baldwin, *Biochemistry* **28**, 4415 (1989).
12. Y. Theriault *et al.*, in preparation.
13. C. L. Brooks, *J. Mol. Biol.* **227**, 375 (1992); M. J. Cocco and J. T. J. Lecomte, *Biochemistry* **29**, 11067 (1990); S. R. Anderson, M. Brunori, G. Weber, *ibid.* **9**, 4723 (1970).
14. M. J. Cocco, Y.-H. Kao, A. T. Phillips, J. T. J. Lecomte, *Biochemistry* **31**, 6481 (1992).
15. K. Kuwajima, Y. Hiraoka, M. Ikeguchi, S. Sugai, *ibid.* **24**, 874 (1985).
16. M. Ikeguchi, K. Kuwajima, M. Mitani, S. Sugai, *ibid.* **25**, 6965 (1986).
17. J. B. Udgaonkar and R. L. Baldwin, *Nature* **335**, 694 (1988).
18. H. Roder, G. A. Elöve, S. W. Englander, *ibid.*, p. 700.
19. Sperm whale myoglobin (Sigma) was purified chromatographically [B. C. Mabbutt and P. E. Wright, *Biochim. Biophys. Acta* **832**, 175 (1985)]. The apoprotein was prepared by 2-butanone extraction [F. W. J. Teale, *ibid.* **35**, 543 (1959)] and its concentration was determined [S. C. Gill and P. H. von Hippel, *Anal. Biochem.* **182**, 319 (1989)]. Pulse-labeling experiments were performed at 5°C in a rapid-mixing device (BioLogic model QFM5). ApoMb (fully protonated) was unfolded in 6M urea and 10 mM acetate, pH 6.1, at 5°C. Refolding was initiated by rapid dilution (1:7.5)

into acetate buffer (10 mM, pH* 6.1, in D_2O) for variable time periods before being pulse-labeled (acetate-Mes-tris; I = 0.2 M, pH* 10.2, in D_2O); the final pH* was 10.2. The final pH* of all steps was checked in independent experiments. Labeling was quenched after 20 ms by dilution into buffer (acetate-Mes-tris; I = 0.25 M, pH* 1.9, in D_2O) to final pH* 5.6. This solution was injected into a reservoir that contained a 1.1 molar excess of bovine hemin (Sigma) to apoprotein.

20. Reconstituted protein was concentrated (Amicon) to ~2 ml, equilibrated with CO and reduced with sodium dithionite. Sodium dithionite was removed by successive cycles of concentration and redilution in CO-equilibrated potassium phosphate buffer (50 mM) in an Amicon concentrator. The protein was concentrated to 3 to 4 mM and equilibrated for 12 hours at 4°C and 4 hours at 35°C before the data were collected.
21. Double quantum and NOESY spectra were collected for each sample on a 600-MHz Bruker AMX spectrometer. A total of 32 transients of 4095 (2Q) or 8192 (NOESY) real data points were recorded for each of 512 t_1 increments. Data were processed with zero filling to 4K complex points and sine bell multiplication in both dimensions with the FTNMR program (Hare Research, Woodinville, WA). Data at each time point were normalized to three nonexchangeable protons and to control experiments in which fully folded or fully unfolded protein was exposed to the labeling pulse (*18*).
22. A. A. Schreier and R. L. Baldwin, *J. Mol. Biol.* **105**, 409 (1976).
23. S. W. Englander and L. Mayne, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243 (1992).
24. G. A. Elove and H. Roder, *Protein Refolding* (American Chemical Society, Washington, DC, 1991), pp. 51–63.
25. Changing the pH or the duration of the labeling pulse may discriminate between a sequential and a parallel folding pathway (*17, 23, 24*). Experiments were carried out in which the length of the labeling pulse was varied after a fixed (10.2 ms) folding time. The proton occupancies remained constant when the labeling pulse was lengthened (two to three times longer). Variation of the pulse pH is not feasible for apoMb due to the instability of the apoprotein at pH >10.2, and the tendency to aggregate at pH <9.5.
26. D. Barrick and R. L. Baldwin, *Biochemistry* **32**, 3790 (1993).
27. T. E. Creighton, *Curr. Biol.* **1**, 8 (1991).
28. R. L. Baldwin and H. Roder, *ibid.*, p. 218.
29. J. P. Waltho, V. A. Feher, G. Merutka, H. J. Dyson, P. E. Wright, *Biochemistry* **32**, 6337 (1993); H.-C. Shin *et al.*, *ibid.*, p. 6356.
30. O. B. Ptitsyn and A. A. Rashin, *Biophys. Chem.* **3**, 1 (1975); T. J. Richmond and F. M. Richards, *J. Mol. Biol.* **119**, 537 (1978); F. E. Cohen, T. J. Richmond, F. M. Richards, *ibid.* **132**, 275 (1979); F. E. Cohen, M. J. E. Sternberg, D. C. Phillips, *Nature* **286**, 632 (1980); M. Gerritsen, K.-C. Chou, G. Nemethy, H. A. Scheraga, *Biopolymers* **24**, 1271 (1985); D. Bashford, F. E. Cohen, M. Karplus, I. D. Kuntz, D. L. Weaver, *Proteins Struct. Funct. Genet.* **4**, 211 (1988); G. Chelvanayagam, Z. Reich, R. Bringas, P. Argos, *J. Mol. Biol.* **227**, 901 (1992).
31. J. B. Udgaonkar and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8197 (1990); M. Bycroft, A. Matouschek, J. T. Kellis Jr., L. Serrano, A. R. Fersht, *Nature* **346**, 488 (1990); A. Matouscheck, J. T. Kellis Jr., L. Serrano, M. Bycroft, A. R. Fersht, *ibid.*, p. 440; M. S. Briggs and H. Roder, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2017 (1992); J. Lu and F. W. Dahlquist, *Biochemistry* **31**, 4749 (1992).
32. G. A. Elove, A. F. Chaffotte, H. Roder, M. E. Goldberg, *Biochemistry* **31**, 6876 (1992); S. E. Radford, C. M. Dobson, P. A. Evans, *Nature* **358**, 302 (1992); P. Varley *et al.*, *Science* **260**, 1110 (1993). These papers also report formation of burst phase intermediates observed by stopped-flow CD but, in contrast to that formed by apoMb, these intermediates do not protect amide protons from exchange.
33. In stopped-flow CD experiments (Aviv 62DS Circular Dichroism Spectrometer equipped with a Biologic SFM-3 stopped-flow system), apoMb was unfolded in 6 M urea at pH 6.1 and refolded by dilution (1:7.50) into acetate buffer (10 mM, pH 6.1).
34. X. Cheng and B. P. Schoenborn, *J. Mol. Biol.* **220**, 381 (1991).
35. Supported by NIH grant DK-34909 (P.E.W.) and NIH postdoctoral award GM14541 (P.A.J.). We thank C. R. Matthews for the use of the stopped-flow CD spectrometer (supported by BRS Shared Instrument Grant RR04953 from the National Institutes of Health), H. J. Dyson for discussions and for review of the manuscript, L. Tennant for technical assistance, and D. Barrick, M. Cocco, G. Elöve, F. Hughson, B. Jones, C. Mann, D. Morikis, and Y. Theriault for discussions.

20 April 1993; accepted 23 August 1993

Detection of Transient Protein Folding Populations by Mass Spectrometry

Andrew Miranker, Carol V. Robinson, Sheena E. Radford, Robin T. Aplin, Christopher M. Dobson

Hydrogen-deuterium exchange measurements are becoming increasingly important in studies of the dynamics of protein molecules and, particularly, of their folding behavior. Electrospray ionization mass spectrometry (ESI-MS) has been used to obtain the distribution of masses within a population of protein molecules that had undergone hydrogen exchange in solution. This information is complementary to that from nuclear magnetic resonance spectroscopy (NMR) experiments, which measure the average occupancy of individual sites over the distribution of protein molecules. In experiments with hen lysozyme, a combination of ESI-MS and NMR was used to distinguish between alternative mechanisms of hydrogen exchange, providing insight into the nature and populations of transient folding intermediates. These results have helped to detail the pathways available to a protein during refolding.

Measurements of hydrogen-deuterium exchange have provided considerable insight into protein and peptide stability, protein dynamics including local fluctuations, antibody epitope mapping, allosteric interactions, and protein folding (*1, 2*). The advent of ESI-MS has enabled the determination of accurate molecular weights of picomolar quantities of proteins, typically to within 1 dalton for proteins with molecular weights up to about 20,000 (*3*). This level of accuracy and the ability to observe intact protein molecules have made the observation of hydrogen exchange of proteins by electrospray ionization feasible (*4*). In this report, we show that a combination of ESI-MS and NMR can provide crucial information on the nature and populations of species on the folding pathway of a protein, information that hitherto was not possible to obtain with either technique in isolation.

Experiments were performed with hen egg-white lysozyme, the native structure of which has been characterized in great detail by a variety of experimental and theoretical methods (*5, 6*). The refolding kinetics of this enzyme have been extensively studied under equilibrium conditions in which a cooperative two-state model is well established (*7*). Furthermore, non-equilibrium refolding studies of lysozyme by pulsed hydrogen-exchange labeling and two-dimensional (2D) NMR have already revealed many details of the folding process (*8*). These details include the observation that the two structural lobes of the native protein are also distinct folding domains and that one, the predominantly α helical domain, folds more rapidly than the other, largely a β sheet domain, in the majority of molecules.

A comparison of the ESI-MS spectra of protonated lysozyme and a sample in which all labile hydrogens had been exchanged for deuterium [deuterated lysozyme (*9*)] is shown in Fig. 1. Although both samples had been dissolved in protonated solvent before the spectra were run, the deuterated protein shows an increase in mass of about 80 daltons (*10*). Because lysozyme contains approximately 250 exchangeable hydrogens, these 80 sites represent hydrogens protected from exchange by the compact structure of the native protein under these conditions. Studies with NMR have shown that the majority of these protected hydrogens arise from backbone amide groups (*6, 11*).

Although amide hydrogen exchange can be measured in a consistent manner by ESI-MS and NMR, the nature of these two methods is complementary. The NMR method monitors the average exchange at individual sites, permitting interpretation

A. Miranker, S. E. Radford, C. M. Dobson, Oxford Centre for Molecular Sciences and New Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QT, United Kingdom.
C. V. Robinson and R. T. Aplin, Oxford Centre for Molecular Sciences and Dyson Perrins Laboratory, Oxford University, South Parks Road, Oxford OX1 3QY, United Kingdom.