Protein Science (1999), *8*:45–49. Cambridge University Press. Printed in the USA.

Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate

VICKIE TSUI, CARLOS GARCIA, SILVIA CAVAGNERO, GARY SIUZDAK, H. JANE DYSON, AND PETER E. WRIGHT

Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037

(RECEIVED April 24, 1998; ACCEPTED August 24, 1998)

Abstract

Folding of apomyoglobin is characterized by formation of a compact intermediate that contains substantial helicity. To determine whether this intermediate is obligatory or whether the protein can fold directly into the native state via an alternate parallel pathway, we have combined quench-flow hydrogen-exchange pulse labeling techniques with electrospray ionization mass spectrometry. The mass spectra of apomyoglobin obtained at various refolding times suggest that apomyoglobin indeed folds through a single pathway containing an obligatory intermediate with a significant hydrogen-bonded secondary structure content.

Keywords: amide proton protection; hydrogen-deuterium exchange; myoglobin; protein folding

The mechanism by which proteins fold into their unique native structures remains a central problem in structural biology. With the rapid progress being made in the sequencing of many genomes, understanding the mechanisms of protein folding becomes more than theoretical enlightenment. Insight into the general principles of protein folding and structure prediction is essential to the process of utilizing this vast new storehouse of information.

An important method for studying the folding pathway of proteins utilizes the quench-flow hydrogen exchange pulse labeling technique coupled with NMR for identification of protected amide protons (Roder, 1989). Following the initial work of Katta and Chait (1991), Miranker et al. (1993) described a variation of this technique, in which amide proton hydrogen-exchange during the folding of hen lysozyme was monitored by electrospray ionization mass spectrometry (ESI-MS) as well as by NMR. NMR monitors the average exchange at individual sites and allows the folding of local regions to be probed at the single residue level, whereas ESI-MS distinguishes the populations of protein molecules with different masses and can therefore be used to detect the existence of multiple pathways. On the basis of the mass spectrometry data, Miranker et al. (1993) concluded that a detectable fraction of the hen lysozyme formed a native-like structure with a time constant of 5 to 10 ms, while the rest of the protein folded sequentially, the folding of the α domain being followed by that of the β domain, with a time constant of 350 ms.

Apomyoglobin is extremely well suited for studies of folding mechanisms. Its native structure is characterized by eight helices (A–H) and a partially helical loop connecting helices C and D, as shown in Figure 1. The kinetic folding pathway of apomyoglobin has been studied in detail by hydrogen exchange pulse labeling and stopped flow methods (Jennings & Wright, 1993). A well-defined intermediate with highly protected protons in helices A, G, and H was found to form within the first 6 ms of refolding. This was shown to be very similar to a compact equilibrium molten globule at low pH and temperature; this species has recently been extensively characterized at equilibrium by NMR (Eliezer et al., 1998). The data of Jennings and Wright (1993) also indicated that several residues in the B and E helices showed partial protection from exchange at the earliest time point. This partial protection could arise for residues that are loosely packed in the burst phase intermediate, such that their amides are weakly protected from exchange. Alternatively, it is possible that partial protection might result from rapid folding of a fraction of the apomyoglobin molecules into the fully native state along a parallel pathway that does not involve the A-G-H intermediate, i.e., by a mechanism similar to that invoked for lysozyme (Miranker et al., 1993).

Several theoretical studies based on statistical mechanics and Monte Carlo simulations of folding have also suggested that this type of multiple-pathway mechanism may be operative in most or

Reprint requests to: Dr. Peter E. Wright or Dr. H. Jane Dyson, Department of Molecular Biology MB2, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037; e-mail: wright@scripps.edu; dyson@scripps.edu.

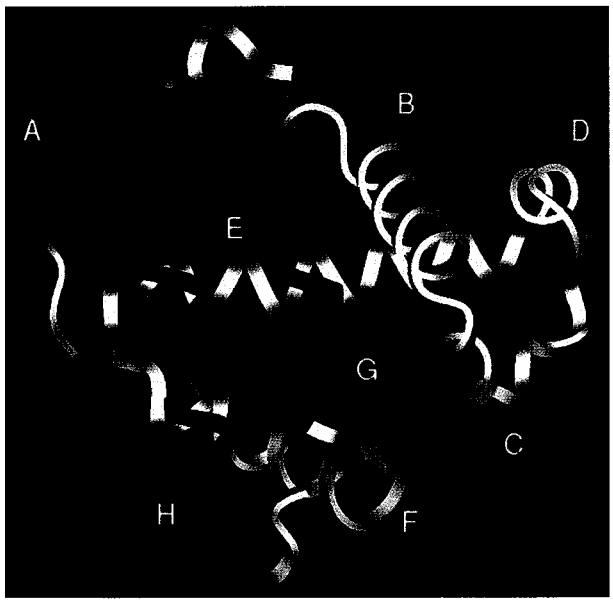

Fig. 1. Structure of native myoglobin (Kuriyan et al., 1986) with the heme group omitted for clarity. Helices A–H are labeled, with the A, G, and H helices colored in red. These three helices are well formed in the A-G-H folding intermediate of apomyoglobin.

even all proteins (Bryngelson & Wolynes, 1987; Shakhnovich & Gutin, 1993; Sali et al., 1994a, 1994b). In this paper, we address this issue for apomyoglobin. Whether the majority of the apomyoglobin molecules fold by very similar pathways that include the obligatory formation of the A-G-H intermediate, or whether the protein exhibits multiple pathway behavior, is a question that must be addressed to make a proper interpretation of the folding data. Here we describe quench-flow mass spectrometry measurements that establish unequivocally that apomyoglobin folds via a single dominant pathway incorporating an obligatory intermediate, with no detectable parallel pathway by which the protein is folded directly and rapidly into its native state. It should be noted that these experiments do not address directly the separate issue of whether the intermediate is on- or off-pathway. It appears from recent studies of the folding kinetics of apomyoglobin (Jamin & Baldwin, 1996; Cavagnero et al., 1998; Jamin & Baldwin, 1998) that the intermediate is in fact on-pathway. In the present paper, we address the separate issue of the existence of multiple folding pathways.

Results and discussion

The measured mass of the ^{15}N-labeled apomyoglobin (17,549 Da) is in close agreement with the theoretical molecular weight of 17,549 Da. The measured mass of fully-deuterated protein obtained from a sample unfolded in acid and allowed to refold in D_2O is 17,808 Da. This compares with a theoretical mass of 17,830 Da, counting all backbone amide protons and all exchangeable side-chain protons [lysine (3 exchangeable protons per residue), arginine (5), histidine (2), serine (1), threonine (1), asparagine (2), glutamine (2), tyrosine (1), and tryptophan (1)]. Aspartic acid and glutamic acid are assumed to be deprotonated under the conditions of the experiment. The total expected mass for the fully-deuterated ^{15}N protein is somewhat higher than that observed, most likely due to a small amount of exchange of highly labile side-chain deuterons in the short interval between addition of the carrier buffer for mass spectrometry (which contains H_2O) and the actual mass spectrometry measurement.

The mass spectra of apomyoglobin observed after various refolding times in the quench-flow apparatus are shown in Figure 2. If the protein is folded from D_2O with no time allowed for folding in H_2O, a single peak with a mass of 17,630 Da is expected, corresponding to the labeling pattern expected of the unfolded protein *U*. At the earliest time point (6 ms), a single peak is also observed, centered at 17,606 Da, corresponding to the intermediate state *I* (expected mass 17,604 Da). The mass spectra indicate that most, if not all, of the apomyoglobin is present as the intermediate after 6 ms refolding. There is no indication of the presence of residual unfolded protein at this time point; the small peak at ~17,650 Da is due to the presence of a small amount of residual salt (likely 2× Na) in the samples and appears in all of the mass spectra in Figure 2. In addition, the fully-folded native state (*N*) (expected mass 17,589 Da) is not detectable at the earliest refolding time, 6 ms. These results alone provide conclusive evidence that a single folding pathway is operative.

Detectable populations of fully folded native protein *N* begin to appear between 37 and 105 ms refolding times, and the steady increase in *N* at longer refolding times is accompanied by a proportionate decrease in *I*, the implication being that *I* is converted to *N* as the protein folds (Fig. 3). The peak corresponding to *N* is centered at 17,584 Da, with 22 more protons protected from exchange than in the intermediate species. This is in general agreement with the results of Jennings and Wright (1993), who found 17 protected sites in the native apoprotein that were exchanged in the intermediate. The discrepancy may be due to the difference in the elapsed time between the preparation of samples for mass spectrometry and NMR and the measurements in each case. In practical terms, the shorter time between sample preparation and measurement for the mass spectrometry measurement means that there is less tendency for exchange-out of weakly hydrogen-bonded amide protons to occur from the holoprotein present in the final measurement sample. Finally, at a refolding time of 4 s, native apomyoglobin is the dominant state and there are only small amounts of intermediate. The decay of the intermediate state and the appearance of the native state both fit to a single exponential with a time constant of 1.38 ± 0.07 s (Fig. 3).

Species containing different fractions of deuterium in place of the amide protons can be distinguished by their masses in the electrospray mass spectrum. According to the hypothesis that the majority of the protein molecules follow a folding pathway that is represented by

$$U \rightarrow I \rightarrow N \quad (1)$$

where *U* represents the unfolded state in urea, *I* represents the intermediate, and *N* represents the fully-folded state, there are three main species to be expected in the mass spectrometry experiments. As stated previously, the mass spectrometry measurements by themselves give no direct evidence for the on-pathway nature of the intermediate *I*.

Due to the complexity of the operations involved in producing samples for the mass spectrometry experiment, the process has been summarized schematically in Figure 4. In this scheme, four different populations of exchangeable protons have been identified: those that exchange rapidly with solvent in the holoprotein

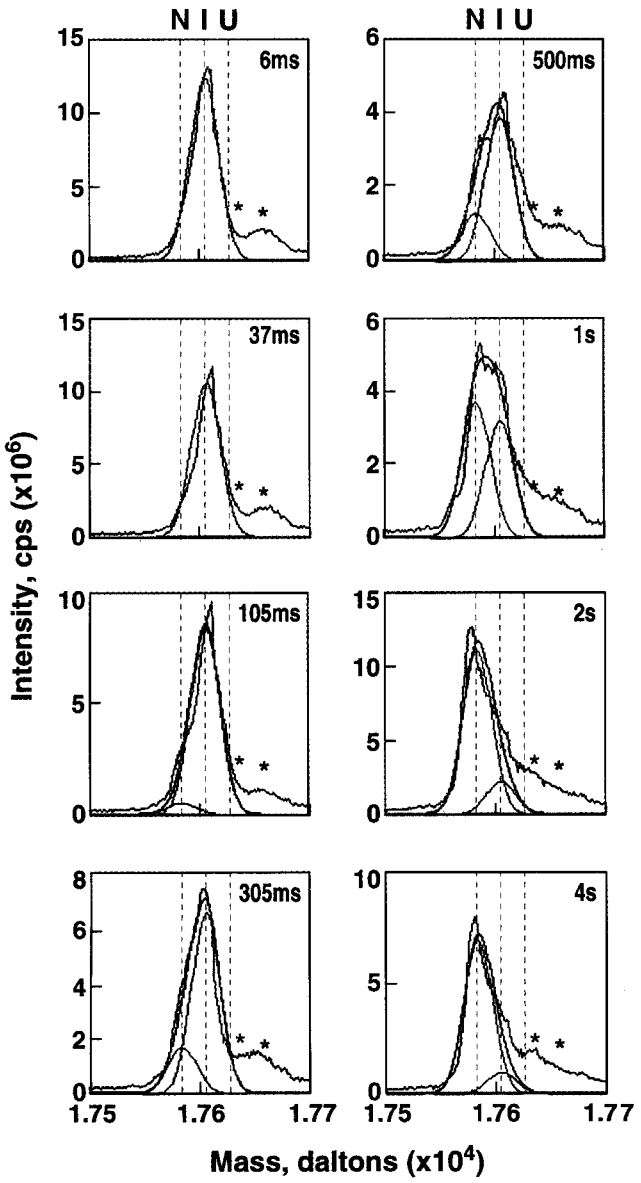

Fig. 2. Deconvoluted electrospray ionization mass spectra of apomyoglobin in the +7 charge state exchanged with deuterium at eight different refolding times: 6 ms, 37 ms, 105 ms, 305 ms, 500 ms, 1 s, 2 s, and 4 s. Black, experimental data; blue, overall fitted; red, fitted intermediate peak; green, fitted native peak. *I* stands for intermediate; *N* stands for native. The peaks corresponding to the *N* and *I* states were represented as Gaussians and the centers and widths were kept constant at each of the time points. The width at half-height was 24 Da. The expected positions for the masses of the folded native apomyoglobin (*N*) at 17,589 Da, the intermediate (*I*) at 17,604 Da, and the unfolded protein at 17,629 Da were calculated from values obtained by NMR (see text). The small peaks to the right of the main mass peaks do not correspond to any of the expected masses for apomyoglobin and most likely represent adducts of the protein with 2× Na^+ (additional mass 46). On the basis of the position of the artifact in the 6 ms panel relative to the mass of *I*, the expected position of the artifact has been indicated in all of the panels by asterisks, red for an artifact related to *I*, and green for an artifact related to *N*.

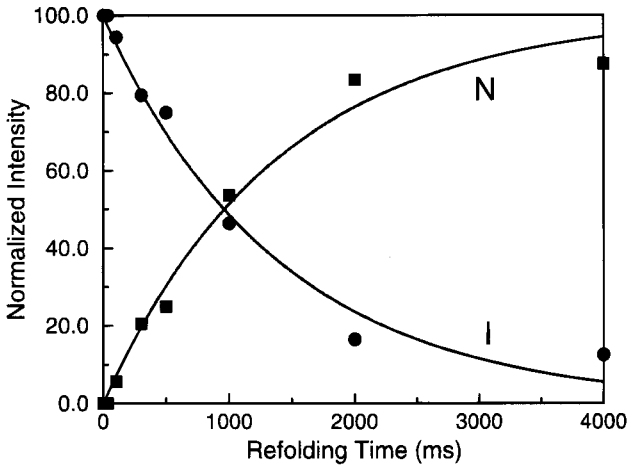

Fig. 3. Normalized intensities of intermediate and native states as a function of refolding time. The sum of the heights of the *N* and *I* peaks in each mass spectrum was equated to 1 and the proportion of *N* and *I* calculated to give the normalized values.

(termed group *W* protons), those that are protected within the dead time of the quench flow apparatus (termed *X*), those that are not protected in the dead time but that are protected in the folded apoprotein (termed *Y*), and those that are protected in the holoprotein but not in the folded apoprotein (termed *Z*). Figure 4 explains the populations of protons and deuterons to be expected at zero folding time and at two folding times, 6.4 ms and 4 s. The populations *W*–*Z* were estimated from NMR results (Hughson et al., 1990; Jennings & Wright, 1993). The number of amides in group *X* was estimated in three ways: (1) by counting the number of amides protected at the earliest time point (Jennings & Wright, 1993), which gave a value of 20, (2) by counting the number of amides with $P > 1$ in the equilibrium intermediate (Hughson et al., 1990), which gave a value of 23, and (3) by summing the intensities of residual cross peaks in a ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) spectrum obtained after the shortest possible refolding time. The intensities in the latter method were normalized by dividing the summed intensity by that of a reference cross peak (either V10 or V17), which gave a value of 25 (S. Cavagnero, S.S. Narula, Y. Thériault, H.J. Dyson, and P.E. Wright, unpubl. results). A value of 25 was used for *X* in Figure 4, since it is most likely to account both for fully and partially-deuterated sites. The value of *Y* was obtained similarly in three ways. The published data give values of 18 (Jennings & Wright, 1993) and 16 (Hughson et al., 1990) protons that are protected in the native apoprotein but not in the intermediate. There are 40 protons protected in the native apoprotein according to the ^{15}N-^{1}H HSQC experiment described above, but with a refolding time of 4 s (S. Cavagnero, S.S. Narula, Y. Thériault, H.J. Dyson, and P.E. Wright, unpubl. results). Subtraction of the value of *X* from this total gives 15 protons for *Y*; this value was used in the calculations in Figure 4. The value of *Z* can be obtained from the difference between the number of amides protected in the native apoprotein and in the holoprotein. This number was obtained from hydrogen exchange measurements and is estimated to be 40 protons (S. Cavagnero, S.S. Narula, Y. Thériault, H.J. Dyson, and P.E. Wright, unpubl. results).

It can be seen from Figure 4 that the masses expected on the basis of the NMR results and the masses obtained in the mass spectrometry experiments are quite consistent, indicating that the folding of all of the individual apomyoglobin molecules proceeds via the intermediate. While these mass spectrometry experiments

No Folding Time

Species	**U**	*0ms and* →	**I**	*complete* →	**N**	*add* →	**H**	*wash* →	**H**
Buffer	H_2O	*label with D_2O*	D_2O	*folding in D_2O*	D_2O	*heme in D_2O*	D_2O	*H_2O for MS*	H_2O
Protonated	WXYZ		-		-		-		W
Deuterated	-		WXYZ		WXYZ		WXYZ		XYZ

Mass (calculated) = 17549 + X + Y + Z = 17629

Earliest Time Point

Species	**U**	*6ms and* →	**I**	*complete* →	**N**	*add* →	**H**	*wash* →	**H**
Buffer	H_2O	*label with D_2O*	D_2O	*folding in D_2O*	D_2O	*heme in D_2O*	D_2O	*H_2O for MS*	H_2O
Protonated	WXYZ		X		X		X		WX
Deuterated	-		WYZ		WYZ		WYZ		YZ

Mass (calculated) = 17549 + Y + Z = 17604

Mass (measured) = 17606

Last Time Point

Species	**U**	*4s and* →	**I**	*complete* →	**N**	*add* →	**H**	*wash* →	**H**
Buffer	H_2O	*label with D_2O*	D_2O	*folding in D_2O*	D_2O	*heme in D_2O*	D_2O	*H_2O for MS*	H_2O
Protonated	WXYZ		XY		XY		XY		WXY
Deuterated	-		WZ		WZ		WZ		Z

Mass (calculated) = 17549 + Z = 17589

Mass (measured) = 17584

Fig. 4. Schematic representation of the masses to be expected during folding of apomyoglobin, according to Equation 1. The various forms of myoglobin encountered in the experiments are abbreviated *U* (unfolded form), *I* (kinetic intermediate), *N* (fully-folded apoprotein), and *H* (holoprotein after addition of heme). Exchange of a proton for a deuteron results in the addition of 1 Da to the mass of the protein. Four populations of exchangeable protons are distinguished: *W* (backbone and side-chain protons that exchange rapidly in the native holoprotein), *X* (amides that are protected at the earliest folding time), *Y* (amides that are protected in the native apoprotein but not at the earliest folding time), and *Z* (amides that are not protected in the apoprotein but that exchange slowly in the holoprotein). Estimates of the values of *W*, *X*, *Y*, and *Z* were made from published and unpublished NMR hydrogen exchange data (Hughson et al., 1990; Jennings & Wright, 1993; S. Cavagnero, S.S. Narula, Y. Theriault, H.J. Dyson, and P.E. Wright, unpubl. results) (see text). For the calculation of the expected masses, values of 25, 15, and 40 protons were assumed for populations *X*, *Y*, and *Z*, respectively.

show that the intermediate is obligatory in the folding of apomyoglobin, they do not provide conclusive proof that the intermediate is on-pathway. Such proof requires other experiments (Jennings & Wright, 1993; Jamin & Baldwin, 1996, 1998; Cavagnero et al., 1998).

Conclusions

Using quench-flow hydrogen-exchange pulse labeling followed by electrospray ionization mass spectroscopy, we have shown that apomyoglobin folds by a single pathway that includes an obligatory intermediate. All molecules of apomyoglobin form the intermediate within 6 ms after initiation of refolding, then slowly form the native state during the subsequent 4 s. The studies of Jennings and Wright (1993) show that the A, G, and H helices, together with a part of B, are formed within the dead time of the quench-flow apparatus (6 ms). The remainder of the B helix is completely formed within 1 s, followed by the formation of the C helix, C-D loop, and E helix within 2.5 s. The most important conclusion from the present work is that the native state is indeed absent at the beginning of refolding, and that it appears concomitantly with the disappearance of the intermediate species. This demonstration that apomyoglobin folds by a single dominant pathway validates the results of many kinetic studies of apomyoglobin and its mutants. It

further indicates that, while certain proteins such as hen lysozyme exhibit multiple-pathway folding behavior, other proteins such as apomyoglobin follow a single highly-preferred folding pathway on the energy landscape.

Materials and methods

Recombinant sperm whale apomyoglobin was expressed in *Escherichia coli* and purified using previously described methods (Mabbutt & Wright, 1985; Jennings et al., 1995). The samples used in the mass spectrometry experiments were uniformly labeled with ^{15}N, since in many cases a portion of the same samples was used for NMR experiments to confirm that the results obtained at each time point were the same using both techniques. All myoglobin samples contained an additional initiator methionine at the N-terminus. Quench-flow hydrogen-exchange pulse labeling of apoMb was performed at 5 °C using standard methods (Roder et al., 1988; Udgaonkar & Baldwin, 1988) with a Biologic model QFM-5 rapid-mixing quench flow apparatus. The fully-protonated protein was unfolded in 10 mM sodium acetate buffer and 6 M urea at pH 6.5. Refolding was initiated by a rapid 7.5-fold dilution into 10 mM sodium acetate buffer at pH* 5.6 in D_2O for variable time periods (6 ms, 37 ms, 105 ms, 305 ms, 500 ms, 1 s, 2 s, 4 s). This was immediately followed by a 20 ms labeling pulse consisting of a 1:1 fold dilution into a 100 mM CAPS buffer, pH* 10.4 in D_2O, to give a pH* of 10.0. Only protons involved in hydrogen-bonded structure in folded regions are protected from deuterium exchange under these conditions. Following the labeling pulse, the pH* was rapidly decreased to 5.9 by dilution into a quench buffer (200 mM MOPS pH* 2.8, in D_2O). The solution was collected into 1.2 mL of 50 mM phosphate buffer in D_2O, 100 mM KCN pH 8 at 4 °C, containing 1.2 molar excess of bovine heme (Sigma, St. Louis, Missouri) over the protein. Finally, the pH* of the solution containing the holoprotein was decreased by addition of 200 μL of quench buffer (200 mM MOPS, pH* 2.8, in D_2O), to give a final pH* of 6.2.

The samples of reconstituted wild-type apoMb obtained at different refolding times were concentrated to ~1 mL with an Amicon concentrator (YM10 membrane) and exchanged with 50 mM phosphate buffer in D_2O (pH* 6.0) without KCN. Under these conditions, most of the free heme precipitated onto the concentrator membrane. To remove excess buffer and salt, which interfere with the mass analysis, each sample was concentrated twice and diluted each time with 10 mM ammonium acetate, pH 6.0, in water.

All electrospray ionization mass spectrometry experiments were performed using an API I00 Perkin Elmer SCIEX single quadrupole mass spectrometer. Samples were typically introduced into the mass analyzer at a rate of 4.0 μL/min. The positive ions, generated by charged droplet evaporation, enter the analyzer through an interface plate and a 100 μm orifice. The declustering potential was maintained between 50 and 200 V and the emitter voltage was typically maintained at 5,000 V. Samples were injected into the electrospray immediately following a 1:2 dilution in HPLC Buffer B (60% acetonitrile in water, 0.1% TFA). These solution conditions were found to allow vaporization of the protein without causing appreciable unfolding and hence loss of the deuterium label. All the spectra are shown at charge +7 expressed on a mass scale. The scan range included other charge states to confirm that the results are independent of the particular charge state selected. Samples were prepared several times and each mass spectrum was acquired two or three times.

Peaks in the mass spectra were fitted to Gaussian curves using the program MATLAB. Because of salt effects that broaden the tails of peaks, the width of the Gaussian curves was kept constant at a width corresponding to that of the original 6 ms peak (17,606 Da). The centers of the peaks were fixed in all of the simulations. Relative intensities were calculated by normalizing the sum of intensities for all species in each spectrum and comparing the normalized intensities of the different species. Exponentials were fit using the program XMGR and plots were made using MATLAB.

Acknowledgments

We thank Stefan Prytulla, Jian Yao, David Eliezer, and Pat Jennings for helpful discussions during the progress of this work. Expert technical assistance was provided by Linda Tennant. This work was supported by grants GM 34909 (PEW) and GM 57374 (HJD) from the National Institutes of Health, and in part by the Wills Foundation (SC).

References

Bryngelson JD, Wolynes PG. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA 84*:7524–7528.

Cavagnero S, Dyson HJ, Wright PE. 1998. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J Mol Biol.* Forthcoming.

Eliezer D, Yao J, Dyson HJ, Wright PE. 1998. Structural and dynamic characterization of partially folded states of myoglobin and implications for protein folding. *Nature Struct Biol 5*:148–155.

Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science 249*:1544–1548.

Jamin M, Baldwin RL. 1996. Refolding and unfolding kinetics of the equilibrium folding intermediate of apomyoglobin. *Nature Struct Biol 3*:613–618.

Jamin M, Baldwin RL. 1998. Two forms of the pH 4 folding intermediate of apomyoglobin. *J Mol Biol 276*:491–504.

Jennings PA, Stone MJ, Wright PE. 1995. Overexpression of myoglobin and assignment of the amide, Cα and Cβ resonances. *J Biomol NMR 6*:271–276.

Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science 262*:892–896.

Katta V, Chait BT. 1991. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun Mass Spectrom 5*:214–217.

Kuriyan J, Wilz S, Karplus M, Petsko GA. 1986. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. *J Mol Biol 192*:133–154.

Mabbutt BC, Wright PE. 1985. Assignment of heme and distal amino acid resonances in the ^{1}H-NMR spectra of the carbon monoxide and oxygen complexes of sperm whale myoglobin. *Biochim Biophys Acta 832*:175–185.

Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM. 1993. Detection of transient protein folding populations by mass spectrometry. *Science 262*:896–900.

Roder H. 1989. Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol 176*:446–473.

Roder H, Elöve GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature 335*:700–704.

Sali A, Shakhnovich E, Karplus M. 1994a. How does a protein fold? *Nature 369*:248–251.

Sali A, Shakhnovich E, Karplus M. 1994b. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J Mol Biol 235*:1614–1636.

Shakhnovich EI, Gutin AM. 1993. Engineering of stable and fast-folding sequences of model proteins. *Proc Natl Acad Sci USA 90*:7195–7199.

Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature 335*:694–699.