Published in final edited form as:

J Mol Biol. 2011 August 5; 411(1): 248–263. doi:10.1016/j.jmb.2011.05.028.

Consequences of Stabilizing the Natively Disordered F helix for the Folding Pathway of Apomyoglobin

Chiaki Nishimura, H. Jane Dyson, and Peter E. Wright

Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA 92037

Abstract

The F helix region of sperm whale apomyoglobin (apoMb) is disordered, undergoing conformational fluctuations between a folded helical conformation and one or more locally unfolded states. In order to examine the effects of F helix stabilization on the folding pathway of apoMb, we have introduced mutations to augment intrinsic helical structure in the F helix of the kinetic folding intermediate and to increase its propensity to fold early in the pathway, using predictions based on plots of the average area buried upon folding (AABUF) derived from the primary sequence. Two mutant proteins were prepared, a double mutant P88K/S92K (F2) and a quadruple mutant, P88K/A90L/S92K/A94L (F4). Whereas the AABUF for F2 predicts that the F helix will not fold early in the pathway, the F helix in F4 shows a significantly increased AABUF and is therefore predicted to fold early. Protection of amide protons by formation of hydrogen-bonded helical structure during the early folding events has been analyzed by pH-pulse labeling. Consistent with the AABUF prediction, many of the F helix residues for F4 are significantly protected in the kinetic intermediate but are not protected in the F2 mutant. F4 folds via a kinetically trapped burst phase intermediate that contains stabilized secondary structure in the A, B, F, G, and H helix regions. Rapid folding of the F helix stabilizes the central core of the misfolded intermediate and inhibits translocation of the H helix back to its native position, thereby decreasing the overall folding rate.

Keywords

site-directed mutagenesis; NMR; protein folding; hydrogen-deuterium exchange; quench-flow pulse labeling

Introduction

Under conditions that favor the folded state of a globular protein, many polypeptides fold spontaneously into the native structure in the absence of extrinsic factors such as chaperones. This implies that the information required for the kinetic folding process, including the initiation and propagation steps and the final resolution of local non-native structures that may slow down the process, is fully encoded in the amino acid sequence. In an attempt to define the important attributes of local sequences in the various stages of

© 2011 Elsevier Ltd. All rights reserved.

Corresponding authors: wright@scripps.edu; dyson@scripps.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

folding of a protein, we have focused on the heme-free apoprotein of myoglobin (apoMb). This protein contains 153 amino acids, arranged in 8 helices (A – H) that together comprise the classic globin fold. All of the helices are well formed and stable in the heme-containing holoprotein, but in the absence of heme, the F helix undergoes exchange between folded and locally unfolded states on an intermediate time scale, which broadens its resonances in the NMR spectrum to the point where they are invisible.¹ This behavior probably reflects a functional requirement for the F helix to act as a gate for the insertion of the heme group. In order to explore the consequences of stabilization of this region to the early stages of apomyoglobin folding, we have engineered the sequence of sperm whale apomyoglobin to increase helix content in the F helix and, by adding long side chains in strategic positions, to increase the probability that the F helix will participate in the early steps of folding.

All of the variants of apomyoglobin so far studied, including the distantly related leghemoglobin from leguminous plants,^{2,3} fold via the formation of a burst phase intermediate that contains a subset of the helices that are present in the native folded structure.^{4,5} This intermediate is obligatory, that is, all members of the ensemble pass through this structure, and molecules that have reached this stage in the folding pathway do not subsequently unfold sufficiently to allow H/D exchange to occur for amides that are protected in the intermediate.⁶ The study of an extensive series of mutants of apomyoglobin⁷⁻¹¹ has allowed us to identify a contributing reason for the interruption of the folding process, resulting in formation of the intermediate: it appears that the H helix, which is present and stable in the burst phase intermediate, is translocated by approximately one helical turn relative to its position in the native folded protein. This translocation adversely affects docking of the E helix, and thereby impedes progress to the protein folding transition state; mutagenesis studies indicate that the E helix must fold and pack correctly prior to formation of the transition state.¹¹ No information was available from any of these experiments on the role of the F helix region in apoMb folding, since the amide resonances of F helix residues are broadened beyond detection in NMR spectra of the fully-folded apoprotein. However, it is highly likely that the F helix is completely unfolded in the kinetic intermediate, as it is in the equilibrium molten globule intermediate formed at pH ~4.^{12,13}

The role of individual amino acids in the initiation of the folding process in apomyoglobin has also been addressed by studying apoMb mutants under equilibrium conditions. Based on structural and dynamic characterization of the apoMb polypeptide chain in various unfolded and partly folded states (at pH ~2,¹⁴ pH ~4¹³ and 8 M urea at pH ~2¹⁵), we proposed that local clusters of side chains promote transient hydrophobic collapse and could play a key role in the initiation of folding. The location of these clusters can be predicted from the amino acid sequence, since they score highly in the quantity known as “average area buried upon folding” (AABUF)¹⁶. This index recognizes the potential for hydrophobic interactions inherent in long amino acid side chains, even if they also contain hydrophilic or charged groups. For example, a lysine residue, which is identified in classic hydrophobicity scales as “charged”, in fact contains an extensive aliphatic chain which is certainly capable in the context of an unfolded protein of making hydrophobic interactions.¹⁷ Regions of high AABUF appear to initiate folding by forming foci for local hydrophobic collapse and by attracting and consolidating transient long range interactions that are required for progression towards the folded state.¹⁸ To test these concepts, we made a set of mutations designed to alter the AABUF in the apoMb sequence while retaining the overall structure of the folded protein.¹⁰ The mutant proteins showed altered folding pathways, such that amide proton protection in the burst phase intermediate was observed in areas of the protein (the E helix) quite different from those of the wild-type protein (the A helix), and in precisely the positions to be expected if the AABUF hypothesis was correct.

The present paper extends this idea to the F helix, a region of the apoMb polypeptide where the structure is seriously compromised, not only in the burst phase and equilibrium intermediates, as is the E helix studied previously,¹⁰ but even in the fully folded apoprotein at pH ~6.¹ We asked firstly whether the unstable F helix could be engineered to form stable secondary structure in the apoprotein, and secondly whether it could be even further stabilized according to the AABUF hypothesis such that it would be preferentially folded in the burst phase intermediate formed in the kinetic folding process. We then asked what effect these modifications might have on the folding pathway of apomyoglobin, and whether stabilization of the F helix early in the folding pathway might cause changes elsewhere in the protein that result in faster or slower folding.

Results

Stabilization of the F helix by Mutation

Mutations were introduced into the F helix region of sperm whale apoMb to stabilize helical structure and to increase incrementally the local AABUF. The locations of the mutations in the double (P88K and S92K, termed F2) and quadruple (P88K/A90L/S92K/A94L, termed F4) mutants are shown in Figure 1A and their effects on the AABUF are shown in Figure 1B. The two substitutions introduced in the F2 mutant were sufficient for stabilization of helical structure to the point where new cross peaks could be observed in the HSQC spectrum in the native (pH ~6) state (Figure 2A). These new resonances arise from residues in the stabilized F helix and were readily assigned using standard triple-resonance techniques. The two mutations, which replace the helix-breaking Pro88 and Ser92 residues (located on the surface of the F helix in the holomyoglobin structure) with helix-promoting lysine side chains, were designed to stabilize the helix (the intrinsic helical propensity calculated for the F helix region using AGADIR¹⁹ increases from 0.2% to 12%) without significantly increasing the local AABUF.

In the F4 mutant, the mutations introduced in the F2 protein are retained, together with the replacement of two nearby alanine residues with leucine. The intrinsic helical propensity of the F4 sequence is further increased (22%, determined using AGADIR¹⁹) from that of the F2 mutant, and there is a very large increase in the AABUF (Figure 1B) due to replacement of two small alanine side chains with large hydrophobic leucines. The F4 mutant is therefore expected, like the F2 mutant, to show more stable structure in the folded apoprotein (reflected in the observation of additional cross peaks in the NMR spectrum at pH 6; Figure 2B), and in addition to have an increased tendency for population of the F helix early in the folding process, compared with either the WT protein or the F2 mutant.

The ¹H-¹⁵N HSQC spectra of the F2 and F4 mutant proteins at pH 6 are very similar to that of the WT apoMb under the same conditions, showing that the overall structures are conserved. The only differences relate to the appearance of resonances belonging to the F helix and FG turn, and slight shifts of cross peaks for residues near the C-terminus of the E helix, the N-terminus of the G helix, the CD turn, and the C-terminus of the H helix. A graph of the per-residue differences (WT – F2, WT – F4, F2 – F4) between the weighted average chemical shifts of ¹⁵N and ¹H is given in Figure S1. These residues are located in sites that immediately adjoin the F helix or are in regions where the F helix packs onto the myoglobin core in the X-ray structure of the holoprotein.²⁰ The newly-visible cross peaks in the ¹H-¹⁵N HSQC of the F2 and F4 mutants arise due to stabilization of helical structure in the F helix itself, as is clear from the ¹³Cα chemical shifts associated with the new resonances. Figure 3 compares per-residue plots of the Cα secondary shift (following subtraction of sequence-corrected random-coil chemical shifts) for wild-type, F2 and F4 mutant apoMbs. No data are available for the F helix residues of WT apoMb because their resonances are exchange broadened,¹ but both F2 and F4 show helix content in the F helix

comparable to those in the other helices. Additional residues near the C-terminus of the F helix are visible in the F4 mutant compared to the F2 mutant. In addition, in the F4 mutant, secondary structure is stabilized in the C-terminal region of the H helix (residues 136-148), which forms the docking site for the F helix in the holomyoglobin structure, and to a lesser extent at the C-terminus of the E helix (residues 74-77) and the N-terminus region of the G helix (residues 102-104). Thus, the mutations stabilize both the F helix structure and its packing interactions with neighboring regions of the polypeptide chain of the apoprotein.

Equilibrium and kinetic studies using CD and fluorescence

The effects of pH on the structure of apoMb are illustrated in Figure 4. The ellipticity at 222 nm in the CD spectrum (Figure 4A) shows a similar pattern of pH-dependent changes for the WT, F2 and F4 mutant proteins, but with slight differences that are consistent with the features that were designed into the mutants. All three proteins show a three-state transition from the acid-unfolded state at pH ~2, through an intermediate state at pH ~4, to a folded state at pH ~6. Both of the mutant proteins show increased helical structure in the pH ~6 state, consistent with the increased helical structure inferred for the F helix (Figure 3B, C); the increase in helical structure is similar between the two mutants, consistent with our design. Interestingly, the inflection point representing the pK_a of the transition $N \rightarrow I$ occurs at a lower pH for the F4 mutant than for the WT or F2, indicating increased acid-stability of the native state. The behavior of all three proteins is similar below pH ~4, although both mutants show a slight increase in the residual helical structure in the fully unfolded (pH ~2) state. The pH dependence of the tryptophan fluorescence (Figure 4B) shows a maximum value at pH ~4 for all three proteins, consistent with the formation of an equilibrium molten globule intermediate. The fluorescence data recapitulate the observations made from the CD data of Figure 4A, with the fluorescence maxima corresponding to the inflection points in the CD data. These data indicate that the intermediate of F4 is slightly destabilized at pHs lower than 4, compared with WT and F2.

The kinetics of refolding of the wild type and mutant proteins from their acid unfolded states were measured by pH-jump stopped flow experiments monitored by CD (Figure 5). Burst phase amplitudes and rate constants for the visible kinetic phase are summarized in Table 1. The burst phase amplitudes are all very similar, indicating that a similar amount of helical structure is formed in the dead time of the stopped-flow apparatus for all three proteins. The folding rate constant for F2 is decreased by approximately a factor of 2 compared to the value for WT, while that for F4 is decreased by nearly a factor of 6 from WT.

Unfolding of the proteins by urea denaturation was monitored by CD at pH 4 and pH 6, where the intermediate and native states, respectively are fully populated. Thermodynamic parameters calculated on the basis of these experiments are summarized in Table 1. At pH 6, the C_m value, representing the concentration of urea at which half of the molecules are unfolded, is significantly higher for F4 than for F2 or WT, indicating an increased stability towards urea in this protein that is reminiscent of the increased pH stability of its native state (Figure 4A). Consistent with this, ΔG of unfolding for F4 is significantly greater than that of WT. At pH 4, the mutant C_m values are somewhat lower than that of WT, reflecting a lowered stability of the intermediate towards denaturant, consistent with the results of the pH dependence of the fluorescence (Figure 4B). In addition, the m values (1.3 kcal/mol/M) of the mutants are higher than for WT (1.0 kcal/mol/M).

Effect of Mutations on the refolding pathway and kinetic intermediate

The location of helical structure in the burst phase intermediates of the F2 and F4 mutant proteins was elucidated in atomic detail by using the pH-pulse labeling technique.^{21,22} Experiments were performed with refolding times ranging from 9.2 ms to 6 s, and proton

occupancies following high pH pulse labeling were measured from the intensity of amide cross peaks in ^1H - ^{15}N HSQC spectra, recorded in DMSO solvent²³ to maximize the number of observable resonances. The changes in proton occupancy as a function of refolding time are illustrated in Figure 6 for the F2 and F4 mutant proteins. A full series of plots showing data for all resonances is given in Figures S2 and S3. For both proteins, amides from several residues in the A, G, and H helix regions are almost fully protected from exchange in the 9.2 ms burst phase intermediate. For F4, but not for F2 or wild type apoMb, the amides of several residues in the F helix and FG turn are also fully protected in the burst phase, showing that the F helix folds rapidly and forms part of the core of the F4 kinetic intermediate (Figures S2, S3). As the refolding time is increased from 9.2 ms to 6 s, additional amides in other regions of the protein become protected in a slow folding phase. The rate constants measured for the slow phase from individual residues are the same within experimental error. The average rate constants measured for the WT, F2 and F4 mutant proteins were 4.4 ± 1.0 , 3.3 ± 1.0 , and $1.4 \pm 0.5 \text{ s}^{-1}$, respectively.

A comparison of the relative stability of the burst phase intermediate in the mutant and wild type proteins can be made using the occupancy of amide sites during the high-pH labeling pulse of the quench flow experiment. Accurate values for the proton occupancy in the burst phase intermediates of F2 and F4 formed after a refolding time of 9.2 ms were obtained by fitting the exponential decay in cross peak intensity as the length of the pH 10.3 labeling pulse was increased from 9.1 to 70 ms. Values of A_0 , the proton occupancy at zero pulse label length, and k , the rate of amide proton exchange at pH 10.3, were determined and compared with values for the wild type burst phase intermediate under identical conditions. Values of A_0 for each residue are plotted in Figure 7A and B. The most striking observation is that several amides (of residues 86, 87, 93-96, 99) in the F helix and FG turn are protected from exchange in the burst phase intermediate formed during refolding of the F4 mutant, whereas amides in these regions remain unprotected for F2. Amides in the F helix and FG turn are also unprotected in both the burst phase and fully native state of WT apoMb.²³ These results imply that the design of the F4 mutant according to the AABUF hypothesis is correct: the F helix and FG turn fold rapidly and dock with the helix ABGH core during the burst phase, forming stabilized hydrogen bonded structure that protects amides from deuterium exchange. In contrast, the F helix remains largely unfolded in the burst phase intermediate formed by the F2 mutant.

A further notable difference from WT apoMb is in the GH turn (residues 119-123). In contrast to WT, in which GH turn amides are unprotected even in the native state, significant protection is observed in the burst phase for the F2 and F4 mutants. The A_0 values of residues of the GH turn for F4 (Figure 7B) are mostly greater than 0.5 and are significantly higher than those for F2 (Figure 7A), an indication that the additional mutations in F4 have further stabilized this region in the burst phase intermediate, relative to that formed by F2. This increased burst phase amide protection extends into the C-terminus of the G helix and the N-terminal region of the H helix; amides of residues 119, and 123-129 are protected in the burst phase intermediates of the F2 and F4 mutants, but not in that formed by WT, and substantially larger A_0 values are found for Ser117, Met131, Asn132, and Lys133 for the mutants. The differences between A_0 values for the WT and mutant proteins are most readily seen in the plot of Figure 7C. Increased amide protection is also evident at the N-terminus of helix G, where increased A_0 values are observed for residues 102 and 103 in the F4 mutant.

Protection factors for exchange of amides from the burst phase intermediate were calculated from the fitted exchange rate constants in the pH 10.3 labeling pulse using intrinsic exchange rates estimated from Bai et al²⁴. Overall, the protection factors are lower for F4 than for wild type apoMb, consistent with the slight decrease in stability of the equilibrium

intermediate (Table 1). It is notable, however, that the amides of residues 9-13, 17, 104, 108-116, 135, 137, and 138 have significantly higher protection factors for the F4 burst phase intermediate, suggesting stabilization of the central region of the AGH helical core by folding of the F helix (Figure 8, Table S1). For the F2 burst phase intermediate, protection factors in these regions of helices A, G, and H are mostly slightly smaller than for F4 and are closer to wild type values.

Additional long range effects of the F helix mutations are evident in decreased A0 values for residues 28-33 in the B helix region of the F4 mutant, which are located in the helix B – helix G interface in the native myoglobin structure (Figure 7B). This destabilization of helix B in the F4 burst phase intermediate is even more evident in amide exchange protection factors, which are smaller than those of the wild type protein for most residues in the B helix region (Figure 8). Thus, stabilization of helical structure and packing in the F helix region appears to have destabilized folding of the B helix in the burst phase intermediate. In both F2 and F4, A0 values for D helix residues are increased relative to WT (Figure 7A-C).

A direct comparison of the differences between the burst phase intermediates of F2 and F4 is shown in Figure 7D. This plot allows a comparison of a more extensive portion of the protein, since there are more amide probes in both F2 and F4 than in WT in several regions, including the N-terminal part of the B helix (D20 – Q26), the GH turn and the N-terminal part of the H helix (D122 – G129). The increased stabilization and enhanced packing of the F helix in F4 relative to F2 leads to protection of amides in the F helix region in the burst phase intermediate of F4 and increased protection (higher A0 values) in the G and H helices and GH turn. Several residues in the B helix (H24, I28, R31, L32, F33, and K34) appear to be better protected in the burst phase intermediate of F2 than in that of F4.

Equilibrium intermediate at pH4.1

The equilibrium intermediates of the F2 and F4 mutants have been characterized by NMR (Figure 9) under conditions (50°C and pH 4.1, 10 % ethanol) where the wild type protein has been successfully studied.¹³ The pH of 4.1 was chosen to give maximal formation of the intermediate on the basis of the fluorescence measurements (Figure 4B). Resonances were assigned for the pH 4.1 intermediate states for each of the mutants using 3D HNCO, HNCACO, and HNCA spectra of ¹³C-¹⁵N doubly labeled protein.

A comparison of the secondary C α and CO chemical shifts of the equilibrium intermediates of WT and the F2 and F4 mutants (Figure 10) reveals differences in the location and extent of helical structure. The F helix is clearly largely folded in the pH 4.1 intermediate state of F4 (blue bars) but not in the F2 (red bars) or WT (black bars) intermediates, consistent both with the AABUF plot (Figure 1B) and with the absence of exchange protected amide resonances in the F helix region of the F2 burst phase intermediate. The local effects caused by the presence of a folded F helix in the equilibrium pH 4 intermediate of F4 are observed most clearly in the F helix itself and on neighboring areas of the E and G helices (Figure 10C).

NMR relaxation studies on the equilibrium intermediates

Measurement of ¹⁵N relaxation rates gives information on local motion on a picosecond-nanosecond time scale. ¹⁵N R₂ and heteronuclear NOE values measured for the wild type and F2 and F4 mutant proteins at pH 4.1, 50°C, 10% ethanol are shown in Figure 11. The data for WT apoMb and the F2 mutant are very similar, with maxima in the R₂ and heteronuclear NOE profiles that indicate motional restrictions due to compaction and packing of helices A, B, G, and H.¹² The D, E, and F helix regions remain highly flexible in the WT and F2 molten globule states, as indicated by the uniformly low values for both R₂

and the heteronuclear NOE. The equilibrium intermediate of the F4 mutant also exhibits a high degree of motional restriction in the helix A, B, G, and H regions (Figure 11), but in addition shows greatly increased R_2 and heteronuclear NOE throughout the F helix. These changes are consistent with formation of helical structure in the F helix of the F4 equilibrium intermediate state, stabilized by packing against the ABGH hydrophobic core. Restrictions of backbone motions are also evident in the E helix, especially at its C-terminal end, in parts of the G and H helices, and in the GH turn (Figure 11B).

Discussion

In contrast to the heme-containing holoprotein, where the F helix is fully folded and provides a histidine ligand to bind the heme, the F helix in apomyoglobin is disordered and undergoes fluctuations on an intermediate time scale that causes broadening and loss of the resonances from the ^1H - ^{15}N correlated NMR spectrum.¹ The amino acid sequence of the F helix of sperm whale myoglobin (EAELKPLAQSHAI) contains a central strategically placed proline that is a good candidate for mutations aimed at stabilizing helical structure in this region. Published data on the P88A mutant²⁵ indicated that this protein contained significantly more helical structure than wild-type apomyoglobin. We prepared the P88A mutant protein, but found that it was unsuitable for NMR spectroscopy, due to a tendency to aggregate at higher concentrations. Modification of P88 and S92 by substitution with the long-chain helix-promoting amino acid lysine gave a protein that was highly soluble and stable in solution, the F2 mutant (Figure 2A). Additional substitutions (P88K/A90L/S92K/A94L) were introduced in F4 to promote helix formation and hydrophobic packing. The $^{13}\text{C}\alpha$ chemical shifts indicate that the F helix is folded in the native state for both F2 and F4 (Figure 3), and the CD data (Figure 4A) clearly show that both F2 and F4 contain more helical structure than WT in the native state at pH 6. Both mutants have significantly more probe residues (i.e. amides with significantly slowed rates of exchange) than the wild type protein in several regions, including the N-terminal parts of the B and H helices (D20-Q26, D126-G129) and the GH turn (D122 – G124). This is probably due to the overall stabilization of the native structure at pH 6 by the increased helical structure in the F helix.

Although stable hydrogen-bonded helical structure in the native state at pH 6 is necessary for the observation of burst-phase protection of amides in a given region, the presence of such structure does not guarantee that the region will be present in the burst-phase intermediate of folding – such a situation is seen, for example, for the E helix of WT apomyoglobin. The F helix region of the F2 mutant protein provides another example: the $\text{C}\alpha$ chemical shifts (Figure 3) and the CD spectrum (Figure 4A) provide evidence that the F helix of the F2 mutant is stable at pH 6, but the quench-flow NMR data (Figure 7A) show that this region does not participate in the burst phase folding intermediate. We have shown that a good correlate for early folding and burst-phase amide proton protection is the hydrophobicity of the immediately surrounding amino acid sequence. Specifically, the quantity “average area buried upon folding” (AABUF)¹⁶ has been used to design mutations to alter the folding pathway in predictable ways.¹⁰ According to this hypothesis, the mutations introduced in F4 (but not those of F2) should dramatically change the AABUF profile (Figure 1B) and promote early folding of the F helix. These predictions are supported by the pH pulse-labeling results (Figure 7B), which clearly show that the F helix is folded and its amides are exchange protected in the 9.2 ms burst phase intermediate of the F4 mutant.

The behavior of the equilibrium intermediates at pH 4 is also consistent with our design. The $^{13}\text{C}\alpha$ chemical shifts show a greater increase in helical structure in the pH 4 intermediate of F4 than for F2 (Figure 10), and the relaxation data obtained for the equilibrium intermediate (Figure 11) indicate substantial restriction of motions in the F helix

and its adjacent sequences, likely due to packing against the compact hydrophobic core formed by the A, G, and H helix regions in the pH 4 equilibrium molten globule^{12,13}. The close similarity between the chemical shifts of the equilibrium intermediates of WT, F2 and F4 (Figure 9, 10, S1) provides strong evidence that the structural ensemble is the same, except for the F helix and the C-terminal half of the F4 E helix. The F helix region (residues 83-95) has no measurable helical content in the WT equilibrium intermediate at pH 4.1, but helical structure in this region is slightly populated in the F2 mutant (9%) and highly populated in the F4 mutant (57%) (Table S2). The helical content of the remainder of the protein remains similar for all three proteins, with the exception of the E helix, which is significantly more helical in F4.

Differences in the proton occupancies (A0 values) in the kinetic burst phase intermediates formed during refolding between F4 apoMb and F2 and between F4 and WT are mapped to the structure of holomyoglobin in Figure 12. Firstly, the major effect on the burst phase intermediate is the stabilization of both the F helix and the G-H turn in the F4 mutant, which leads to exchange protection for several amides (green spheres); no F helix amides are protected in the burst phase for either WT apoMb or the F2 mutant, in accord with the predictions from the AABUF parameter (Figure 1B). However, it is noticeable that the effects on the rest of the F4 protein and on the F2 protein of incorporation of a new helix into the burst phase intermediate are rather small. This result contrasts with, for example, the marked specific effects throughout the molecule when the B helix is mutated according to the AABUF hypothesis.²⁶ Concomitant with stabilization of the F helix in the F4 mutant, increased A0 values (blue spheres) relative to F2 are observed for residues in both the N- and C-terminal regions of the G and H helices and in the GH turn; these effects can be readily explained by stabilization of the GH helical hairpin through packing with the stabilized F helix. Intriguingly, stabilization of the F, G, and H helices is accompanied by a decrease in burst phase proton occupancies and protection factors in the B helix, in regions which pack against the G helix in the native myoglobin structure (red spheres).

The most obvious effect of the F2 and F4 mutations is on the kinetics of folding: the folding rate of the slow phase is decreased (Figure 5, Table 1) and F2 has a smaller burst phase amplitude, which is an indication that there is less helical structure in the kinetic intermediate of F2. The slower folding rates likely arise from changes in the burst phase intermediates formed by F2 and F4 that raise the barrier to formation of the transition state which, for the wild type protein, is associated with docking of helix E.^{11,26} The small increase in proton occupancies and exchange protection factors in parts of the A, G, and H helices in F4 and to a lesser extent in F2 (Figures 7, 12) suggest that the central core of the burst phase intermediate has been stabilized relative to wild type protein by incorporation of the F helix to form an AFGH core. Since the burst phase intermediate is a kinetically trapped state, in which the H helix appears to be translocated by approximately one helical turn towards its N-terminus to maximize hydrophobic packing against helix G,¹¹ stabilization of this intermediate would be expected to impair progress towards the folded state. The degree of stabilization of the core is greater for F4 than for F2, consistent with the larger 6-fold decrease in folding rate associated with this mutation. Incorporation of helix F in the F4 burst phase intermediate appears to destabilize helix B, leading to smaller A0 values and/or decreased exchange protection factors for several B helix residues located in the BG interface in the native structure. We have observed previously that mutations which promote native-like structure that is not normally observed in the burst phase intermediate (such as stabilized helical structure in the E helix and native-like structure in the G-H turn region) appear to cause slowing of the overall folding rate.²⁶ We attribute this to the lowered tendency for such a stabilized intermediate to make the rearrangements necessary for progress to the protein folding transition state and thence to final folded state. Such a mechanism appears to be operative also in folding of the F4 mutant.

Methods

Protein preparation

Recombinant sperm-whale apoMb was expressed in *Escherichia coli* BL21-DE3 cells with a pET17b vector including the Mb gene.²⁷ Site-directed mutagenesis was performed using the mutated gene with QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). The ¹⁵N-labeled WT and mutant proteins for both pH-pulse labeling and the ¹⁵N,¹³C-double-labeled WT and mutant proteins for the resonance assignments were produced in M9 minimal medium containing ¹³C-glucose and ¹⁵N-ammonium sulfate as a carbon and nitrogen source. All proteins were purified by the same method as previously described for WT.¹¹ The same extinction coefficient (15,900 M⁻¹cm⁻¹ at 280 nm, pH6.1) was used for the determination of the protein concentration for the WT and mutants in solution, because the extinction coefficient should be not changed by the P88K/S92K and P88K/A90L/S92K/A94L mutations.²⁸

CD and fluorescence

CD spectra were measured on an Aviv 62DS spectrometer at 25 °C. The ellipticity of the proteins (10 μM) in 10 mM sodium acetate buffer was recorded at 222 nm as functions of the pH and concentration of urea.

Tryptophan fluorescence emission spectra were recorded using a Fluorolog-3 fluorescence spectrometer (Jobin Yvon-Horiba) at 25 °C. The spectra of proteins (2 μM) in 10 mM sodium acetate buffer were collected between 300 nm and 400 nm. The excitation wavelength was set at 288 nm with the bandwidth of both excitation and emission at 2 nm. The maximum intensities of the Trp fluorescence were collected as a function of pH. For both CD and fluorescence measurements, individual samples were prepared for each data point.

Stopped-flow CD

Helix formation after the initiation of protein folding was monitored using a DX-17MV Applied Photophysics instrument at 8 °C. The CD signal was measured at 225 nm after a pH-jump from pH 2.2 to pH 6.0 with a slit width of 2 nm. The final concentration of protein was 10 μM after a 1:5 dilution with 30 mM sodium acetate buffer. The dead time of the instrument was about 20 ms.

Quench-flow pH pulse-labeling

pH-pulse labeling H/D exchange was used to obtain residue-specific information on the rates of formation of helical structure during refolding of both the F2 and F4 mutants.⁵ Experiments were performed at 8 °C using a Biologic SFM-400 Quench Flow instrument (Grenoble, France). The protein samples (2.0 mg/ml) were unfolded by addition of HCl to bring the pH to 2.2 and refolding was initiated by dilution with 7.5 volumes of 20 mM sodium acetate buffer in D₂O to give a final pH* of 6.0. At end of the refolding period, pH-pulse labeling was performed at pH* 10.3, to exchange out all of the unprotected amides, by rapid mixing with a 100 mM 3-[cyclohexylamino]-1-propanesulfonic acid buffer in D₂O. At the end of the pulse labeling period, the pH was immediately decreased to pH* 6.0 by mixing with 300 mM 3-(N-morpholino) propanesulfonic acid buffer in D₂O. The pulse labeled protein was flushed into a tube containing 0.1 % (v/v) trifluoroacetic acid solution in D₂O (final pH* 2.5) and immediately frozen in liquid N₂ then lyophilized. The lyophilized protein was stored at -80°C until acquisition of NMR data. For the F2 and F4 mutants, pulse labeling experiments were performed over a range of refolding time between 9.2 ms and 6 s, with a high pH labeling pulse duration of 24 ms. To probe the extent of folding and stability of the burst phase intermediate, pulse labeling experiments were performed using a refolding

time of 9.2 ms and several pulse label durations (9.1, 12, 24, 46, and 70 ms). These experiments were performed for the F2 and F4 mutants and, to provide a proper control, data were recollected for the wild type apoMb.

Prior to NMR data collection, the lyophilized, pulse labeled protein samples were dissolved in a solution of 99.6 % (v/v) d_6 -DMSO (99.96 % deuterated, Cambridge Isotope Laboratories, Andover, MA) and 0.4 % (v/v) D_2O (99.8 % deuterated, CIL). NMR data acquisition for the pH labeled samples (about 0.2 mM) was initiated 8 min after the addition of the solvent. ^{15}N - 1H heteronuclear single quantum correlation (HSQC) spectra, with spectral widths of 9,615 Hz and 2,048 points in the 1H dimension and 1,538 Hz and 256 points in the ^{15}N dimension, were collected at 30 °C for each sample using a Bruker DRX600 spectrometer equipped with a cryoprobe. A small amount of artifactual H/D exchange occurs during data acquisition; the measured intensities of each cross peak in the series of HSQC spectra (14-16, recorded every 20 min) were fitted to an exponential decay to correct the proton occupancies for exchange-out in the DMSO.^{10,23} Amide cross peak intensities in the HSQC spectra were calibrated using the volume of the methyl proton resonances observed in the 1D proton spectra and normalized relative to the intensity of each cross peak at the longest refolding time (6 s). Burst phase proton occupancies measured with different pulse label lengths were fitted to the equation $P = (0.12 + A0 \cdot e^{-kt}) / 1.12$, where P is the experimentally measured proton occupancy, $A0$ is the proton occupancy at the beginning of the labeling pulse (at time zero), t is the length of the labeling pulse, and k is the rate constant for H/D exchange at the labeling pulse pH (10.3). The constant term (0.12 = 1/8.5) is the fraction of H_2O remaining after dilution with a 7.5-fold volume of refolding buffer in D_2O . Protection factors were calculated using intrinsic H/D exchange rate constants²⁴ estimated for the wild type and mutant sequences at pH 10.3.

NMR spectroscopy

Backbone resonance assignments of the mutant proteins in DMSO were made using 3D $HNCO$,²⁹ $HNCACO$,³⁰ $HNCA$,²⁹ $HNCOCA$,²⁹ $HNCACB$,³¹ and $CBCA(CO)NH$ ³² spectra recorded at 30 °C on a Bruker Avance750 spectrometer as described.²³ Backbone resonance assignments for the mutant proteins in the native state were made using the same set of NMR spectra acquired on a Bruker DRX600 spectrometer equipped with a cryoprobe. ^{15}N , ^{13}C -double-labeled samples of the two mutant proteins in DMSO were prepared in the same way as the samples obtained in the pH-pulse labeling experiments. Peaks were slightly shifted depending on the difference of the buffer components and salt concentration in the NMR buffer. For the native state spectra, the mutant ^{15}N , ^{13}C -double-labeled proteins were dissolved in 10 mM sodium acetate buffer at pH 6. For the spectra of the equilibrium molten globule states, the proteins were dissolved in 10 mM sodium acetate, pH 4.1, containing 10% v/v ethanol; spectra were acquired at 50 °C on a Bruker Avance750 spectrometer. All NMR data were collected in the indirect dimension with the TPPI-States method,³³ and processed using NMRPipe³⁴ and analyzed by NMRView.³⁵

Sequence-corrected secondary chemical shift values were calculated according to published methods.^{36,37} Per-residue and average percentage of helical structure were calculated according to published methods.³⁸

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Linda Tennant and Euvel Manlapaz for technical assistance, and members of the Wright and Dyson labs for valuable discussions. This work was supported by grant R01 DK34909 from the National Institutes of Health.

Reference List

1. Eliezer D, Wright PE. Is apomyoglobin a molten globule? Structural characterization by NMR. *J Mol Biol.* 1996; 263:531–538. [PubMed: 8918936]
2. Nishimura C, Prytulla S, Dyson HJ, Wright PE. Conservation of folding pathways in evolutionarily distant globin sequences. *Nat Struct Biol.* 2000; 7:679–686. [PubMed: 10932254]
3. Nishimura C, Dyson HJ, Wright PE. The kinetic and equilibrium molten globule intermediates of apoleghemoglobin differ in structure. *J Mol Biol.* 2008; 378:715–725. [PubMed: 18384808]
4. Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science.* 1993; 262:892–896. [PubMed: 8235610]
5. Nishimura C, Dyson HJ, Wright PE. The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J Mol Biol.* 2002; 322:483–489. [PubMed: 12225742]
6. Tsui V, Garcia C, Cavagnero S, Siuzdak G, Dyson HJ, Wright PE. Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* 1999; 8:45–49. [PubMed: 10210182]
7. Cavagnero S, Dyson HJ, Wright PE. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J Mol Biol.* 1999; 285:269–282. [PubMed: 9878405]
8. Garcia C, Nishimura C, Cavagnero S, Dyson HJ, Wright PE. Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. *Biochemistry.* 2000; 39:11227–11237. [PubMed: 10985768]
9. Nishimura C, Wright PE, Dyson HJ. Role of the B helix in early folding events in apomyoglobin: evidence from site-directed mutagenesis for native-like long range interactions. *J Mol Biol.* 2003; 334:293–307. [PubMed: 14607120]
10. Nishimura C, Lietzow MA, Dyson HJ, Wright PE. Sequence determinants of a protein folding pathway. *J Mol Biol.* 2005; 351:383–392. [PubMed: 16005892]
11. Nishimura C, Dyson HJ, Wright PE. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol.* 2006; 355:139–156. [PubMed: 16300787]
12. Eliezer D, Yao J, Dyson HJ, Wright PE. Structural and dynamic characterization of partially folded states of myoglobin and implications for protein folding. *Nature Struct Biol.* 1998; 5:148–155. [PubMed: 9461081]
13. Eliezer D, Chung J, Dyson HJ, Wright PE. Native and non-native structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry.* 2000; 39:2894–2901. [PubMed: 10715109]
14. Yao J, Chung J, Eliezer D, Wright PE, Dyson HJ. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry.* 2001; 40:3561–3571. [PubMed: 11297422]
15. Schwarzingier S, Wright PE, Dyson HJ. Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry.* 2002; 41:12681–12686. [PubMed: 12379110]
16. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. *Science.* 1985; 229:834–838. [PubMed: 4023714]
17. Dyson HJ, Wright PE, Scheraga HA. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc Natl Acad Sci USA.* 2006; 103:13057–13061. [PubMed: 16916929]
18. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA.* 2008; 105:6278–6283. [PubMed: 18434548]
19. Muñoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. *Structural Biology.* 1994; 1(no 6):399–409. [PubMed: 7664054]

20. Kuriyan J, Wilz S, Karplus M, Petsko GA. X-ray structure and refinement of carbon-monooxy (Fe II)-myoglobin at 1.5 Å resolution. *J Mol Biol.* 1986; 192:133–154. [PubMed: 3820301]
21. Udgaonkar JB, Baldwin RL. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature.* 1988; 335:694–699. [PubMed: 2845278]
22. Roder H, Elöve GA, Englander SW. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature.* 1988; 335:700–704. [PubMed: 2845279]
23. Nishimura C, Dyson HJ, Wright PE. Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc Natl Acad Sci USA.* 2005; 102:4765–4770. [PubMed: 15769860]
24. Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. *Proteins.* 1993; 17:75–86. [PubMed: 8234246]
25. Picotti P, Marabotti A, Negro A, Musi V, Spolaore B, Zamboni M, Fontana A. Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements. *Protein Sci.* 2004; 13:1572–1585. [PubMed: 15152090]
26. Nishimura C, Dyson HJ, Wright PE. Energetic frustration of apomyoglobin folding: role of the B helix. *J Mol Biol.* 2010; 396:1319–1328. [PubMed: 20043917]
27. Jennings PA, Stone MJ, Wright PE. Overexpression of myoglobin and assignment of the amide, C α and C β resonances. *J Biomol NMR.* 1995; 6:271–276. [PubMed: 8520219]
28. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 1995; 4:2411–2423. [PubMed: 8563639]
29. Grzesiek S, Bax A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson.* 1992; 96:432–440.
30. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson.* 1990; 89:496–514.
31. Wittekind M, Mueller L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α - and β -carbon resonances in proteins. *J Magn Reson.* 1993; 101:201–205.
32. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc.* 1992; 114:6291–6293.
33. Marion D, Ikura M, Tschudin R, Bax A. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J Magn Reson.* 1989; 85:393–399.
34. Delaglio F, Grzesiek S, Vuister GW, Guang Z, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–293. [PubMed: 8520220]
35. Johnson BA, Blevins RA. NMRView: A computer program for the visualization and analysis of NMR data. *J Biomol NMR.* 1994; 4:603–614.
36. Schwarzing S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence dependent correction of random coil NMR chemical shifts. *J Am Chem Soc.* 2001; 123:2970–2978. [PubMed: 11457007]
37. Schwarzing S, Kroon GJA, Foss TR, Wright PE, Dyson HJ. Random coil chemical shifts in acidic 8 M urea: implementation of random coil chemical shift data in NMRView. *J Biomol NMR.* 2000; 18:43–48. [PubMed: 11061227]
38. Yao J, Dyson HJ, Wright PE. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* 1997; 419:285–289. [PubMed: 9428652]
39. Koradi R, Billeter M, Wüthrich K. MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics.* 1996; 14:51–55.

Figure 1.

A. Amino acid sequence of sperm whale myoglobin showing the locations of the helices in the native folded protein, the A helix (red), B helix (orange), C helix (yellow), D helix (green), E helix (turquoise), F helix (blue), G helix (purple) and H helix (pink). Mutation sites on the F helix for the F2 mutant are indicated with red dots at P88 and S92. Mutation sites for the F4 mutant include both the red and blue dots at P88, A90, S92 and A94. B. Plot of AABUF for sperm whale myoglobin,¹⁰ calculated with a 9-residue moving average. Black: wild-type sequence; red: P88K/S92K mutant; blue: P88K/A90L/S92K/A94L mutant. Locations of helices in the structure of fully-folded myoglobin are shown at the top of the figure, color coded as in part A.

Figure 2. Portion of the ^1H - ^{15}N HSQC spectrum of apomyoglobin in 10 mM sodium acetate buffer at pH 6 and 30°C. Cross peaks for the wild type protein are shown in black. A. Comparison with the P88K/S92K mutant (red cross peaks). B. Comparison with the P88K/A90L/S92K/A94L mutant (blue cross peaks). The additional cross peaks observed for the mutant proteins are circled and labeled with large type, and selected assignments of cross peaks common with the wild type protein are labeled with small type. Labels and circles in red denote the cross peaks corresponding to the mutated residues.

Figure 3. Secondary α chemical shifts at pH 6.0 for A. wild type; B. P88K/S92K mutant; C. P88K/A90L/S92K/A94L mutant. Locations of helices in the structure of fully-folded myoglobin are shown, color coded as in Figure 1A.

Figure 4.

Effect of pH on the circular dichroism and fluorescence spectra of WT (black), P88K/S92K (red), and P88K/A90L/S92K/A94L (blue). A. Ellipticity at 222 nm in the CD spectrum. B. fluorescence emission intensity (excitation wavelength 288 nm, emission wavelengths 300-400 nm).

Figure 5.

Stopped-flow trace of folding after pH-jump from pH 2.2 to 6.0, monitored by CD ellipticity at 222 nm. Black, red, and blue lines correspond to wild type, P88K/S92K, and P88K/A90L/S92K/A94L, respectively.

Figure 6.

Time course of amide-protection of residues from selected regions of P88K/S92K (F2) and P88K/A90L/S92K/A94L (F4), together with the data for the F helix and FG loop of F4 (Complete data sets for each mutant are shown in Figures S2 and S3). Transition curves of proton occupancy, fitted using the method of least squares¹⁰ are shown for residues whose proton occupancy at 9.2 ms is below 0.7.

Figure 7.

Proton occupancy (A0) for the burst phase intermediate of A. wild type apomyoglobin (black) and P88K/S92K (red) and B. wild type apomyoglobin (black) and P88K/A90L/S92K/A94L (blue). Selected residues where A0 changes significantly between wild type and mutant are labeled. Areas where additional cross peaks are observed for the mutant proteins are circled and the residues labeled. Overall changes in the D helix (increased protection) and E helix (decreased protection) are indicated by arrows. C. Difference in proton occupancy A0 between the mutant and wild type proteins, P88K/S92K minus wild type (red) and P88K/A90L/S92K/A94L minus wild type (blue). Residues that show more significant differences in the mutants compared to wild type ($\Delta A0$ greater or less than 0.1)

are indicated. D. Difference of the proton occupancy between P88K/S92K and P88K/A90L/S92K/A94L. Residues that show more significant differences between the two mutants ($\delta A0$ greater or less than 0.1) are indicated. Locations of helices in the fully folded protein are indicated with horizontal bars colored as in Figure 1A.

Figure 8.

Plots of the amide proton protection factor calculated from the exchange rate constants at pH 10.3 (Table S1), plotted against residue number for WT (black), F2 (red) and F4 (blue) for A. the A and B helices and B. the G and H helices.

NIH-PA Author Manuscript

Figure 10.

Per-residue plots of secondary chemical shifts for the equilibrium intermediate states at pH 4.1 (50°C with 10% ethanol) for WT (black), F2 (red) and F4 (blue). A. Sequence-corrected $C\alpha$ secondary shifts for the D, E and F helices. B. Sequence-corrected CO secondary shifts for the D, E and H helices. C. Difference in the $C\alpha$ secondary shifts of F2 and F4. at pH 4.1. The horizontal lines at 2.8 in part A and 2.1 in part B show the estimated values of the sequence-corrected $C\alpha$ and CO secondary chemical shifts that correspond to 100% population of helix.³⁸

Figure 11.

A. R_2 relaxation and B. heteronuclear $[^1\text{H}]-^{15}\text{N}$ NOE data for the WT, F2 and F4 at pH 4.1, 50°C with 10% ethanol.

Figure 12.

Changes in the amide exchange protection and burst phase proton occupancy between WT, F2 and F4, mapped onto the structure of MbCO.²⁰ Backbone colors correspond to the locations of helices A-H, according to the legend of Figure 1A. Amides that show exchange protection in F4 but not in F2 (left figure) or WT (right figure) are shown as green spheres. Amides in the remainder of the protein that show increased burst phase proton occupancy in F4 relative to F2 or WT are shown as blue spheres, while amides that show decreased protection in F4 relative to F2 or WT are shown as red spheres. The figure was prepared using MolMol.³⁹

Table 1
Kinetic and Thermodynamic Data for Folding of F2 and F4

Protein	Burst phase Amplitude (% of total ellipticity change)	Rate constant (s ⁻¹)	pH 6			pH 4		
			ΔG (kcal mol ⁻¹)	m	Cm	ΔG (kcal mol ⁻¹)	m	Cm
WT	66	3.4 ± 0.1	-5.0 ± 0.2	1.57 ± 0.05	3.18	-1.5 ± 0.1	1.02 ± 0.02	1.43
F2	62	1.8 ± 0.1	-5.1 ± 0.1	1.48 ± 0.04	3.44	-1.3 ± 0.1	1.35 ± 0.07	0.94
F4	67	0.6 ± 0.1	-6.3 ± 0.2	1.47 ± 0.05	4.29	-1.3 ± 0.1	1.34 ± 0.06	0.97