Published in final edited form as:

J Phys Chem B. 2018 December 13; 122(49): 11228–11239. doi:10.1021/acs.jpcb.8b06895.

Complex Folding Landscape of Apomyoglobin at Acidic pH Revealed by Ultrafast Kinetic Analysis of Core Mutants

Takuya Mizukami[†], Ming Xu[†], Ruzaliya Fazlieva[†], Valentina E. Bychkova[‡], and Heinrich Roder^{†,*}

[†] Molecular Therapeutics Program, Fox Chase Cancer Center, Philadelphia, PA 19111

[‡] Laboratory of Protein Physics, Institute of Protein Science RAS. Pushchino, Moscow Region, Russia

Abstract

Under mildly acidic conditions (pH 4–4.5) apomyoglobin (apoMb) adopts a partially structured equilibrium state (*M*-state) that structurally resembles a kinetic intermediate encountered at a late stage of folding to the native structure at neutral pH. We have previously reported that the *M*-state is formed rapidly (< 1 ms) via a multi-state process, and thus offers a unique opportunity for exploring early stages of folding by both experimental and computational techniques. In order to gain structural insight into intermediates and barriers at the residue level, we studied the folding/unfolding kinetics of 12 apoMb mutants at pH 4.2 using fluorescence-detected ultrafast mixing techniques. Global analysis of the sub-millisecond folding/unfolding kinetics vs. urea concentration for each variant, based on a sequential four-state mechanism ($U \rightleftharpoons I \rightleftharpoons L \rightleftharpoons M$), allowed us to determine elementary rate constants and their dependence on urea concentration for most transitions. Comparison of the free energy diagrams constructed from the kinetic data of the mutants with that of wild-type apoMb yielded quantitative information on the effects of mutations on the free energy (ΔG) of both intermediates and the first two kinetic barriers encountered during folding. Truncation of conserved aliphatic side chains on helices A, G and H give rise to a stepwise increase in ΔG as the protein advances from *U* towards *M*, consistent with progressive stabilization of native-like contacts within the primary core of apoMb. Helix-helix contacts in the primary core contribute little to the first folding barrier ($U \rightleftharpoons I$), and thus are not required for folding initiation, but are critical for the stability of the late intermediate, *L*, and the *M*-state. Alanine substitution of hydrophobic residues at more peripheral helix-helix contact sites of the native structure, which are still absent or unstable in the *M*-state, show both positive (destabilizing) as well as negative (stabilizing) ΔG , indicating that non-native contacts are formed initially and weakened or lost as a result of subsequent structural rearrangement steps.

The following graphic will be used for the TOC:

*Corresponding author. roder@fccc.edu.

SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b06895.

- Three figures with additional data on equilibrium and kinetic analysis of folding of apomyoglobin mutants (PDF)
- PDF jp8b06895_si_001.pdf (1.53 MB)

INTRODUCTION

There has been major progress in recent years in our understanding of the folding mechanism of small, single-domain proteins, which often fold cooperatively on the microsecond to millisecond time scale in an apparent two-state mechanism.^{1–5} Much of this progress has been driven by advances in experimental techniques for rapid initiation and detection of folding reactions.^{6–14} Meanwhile, the development of massive parallel computers and distributed computing networks has made it possible to simulate the process of folding for small protein modules with atomic resolution.^{3–4, 15–16} However, it remains unclear at this time to what degree the lessons learned from these “ultrafast two-state folders” can be extrapolated to larger and structurally more complex proteins, which generally exhibit slower (>10 ms), more complex folding reactions with transient intermediates giving rise to multiple kinetic phases.^{17–21}

By monitoring folding and unfolding reactions over a wide time range ($\sim 10^{-5}$ to 10^2 s), using improved turbulent^{8, 22} and laminar⁹ mixing schemes in combination with conventional stopped-flow measurements, we and others have obtained unambiguous evidence for accumulation of transient compact states prior to the rate-limiting step in folding for proteins of various sizes and structural types.^{12, 23–27} These include several proteins with less than 100 residues that were previously thought to be two-state folders.^{24, 28–30} The observation of exponential fluorescence-detected folding phases on the sub-millisecond time scale for a number of proteins has been attributed to a barrier-limited transition to a compact-state ensemble that can be described as a thermodynamic state (macro-state) different from the unfolded state,^{23, 30–31} whereas others³² suggested that these states represent a shift in the ensemble of unfolded states from expanded to more compact conformers with decreasing denaturant concentration. Using ultrafast H/D exchange measurements, we have recently shown that helices in the 60s, 70s and 90s region of cytochrome c form and coalesce within 140 μ s of initiating refolding while the N-terminal helix remains unstructured, indicating that the initial collapse can involve specific structural events.³³ However, fundamental questions remain unanswered concerning the origin of kinetic barriers and the properties and kinetic roles of intermediate states. Are folding intermediates (i) productive “stepping stones” in finding the native fold, (ii) non-obligatory states populated along parallel pathways, or (iii) misfolded dead-end states? Although it has only rarely been possible to obtain unambiguous kinetic evidence that early (< 1 ms) intermediates accumulate on a direct path between the unfolded ensemble (*U*) and native state (*N*)²⁹, protein folding is often modeled as a linear sequence of states leading from *U* to *N*. In contrast, a simple funnel model of protein folding in the energy landscape theory

attributes intermediates to kinetic traps or topological frustration^{34–35}, and large-scale atomistic simulations predict complex folding mechanisms with a large number of states connected via a dense network of pathways.^{3–4, 16, 36}

A major objective in protein folding research is to understand the sequence determinants for initiating and directing folding. It has been proposed that proteins contain a critical set of residues that form a “folding nucleus” necessary for crossing the rate-limiting barrier (transition-state ensemble) for folding.^{37–44} In contrast to kinetically populated intermediate states, the transition states in protein folding cannot be investigated directly, but only via the influence of mutations on the rate of folding or unfolding.^{45–46} While the folding transition states revealed by this so-called mutational Φ -value method are consistent with the notion of a folding nucleus in some proteins, others display a more diffuse transition state⁴⁴ consistent with the heterogeneous ensemble of states observed in simulations.^{3–4, 36}

Apo-myoglobin (apoMb) is a classic example of a multi-state folder, i.e., a protein populating a series of partially structured states during the process of folding or unfolding,^{27, 47–54} and even under certain equilibrium conditions, such as mildly acidic pH.^{55–56} Although high-resolution X-ray and NMR structures have been solved only for holomyoglobin, a heme protein consisting of a single 153-residue domain with eight α -helices (Figure 1), the available solution NMR and H/D exchange data for apoMb indicate that removal of the non-covalently bound heme group results in limited structural perturbations, including unfolding of the F-helix and increased disorder at the N-terminus of the G-helix and adjacent loops, but the remaining helices are still intact and arranged in a native-like topology.^{47, 57–59}

ApoMb adopts three structurally distinct equilibrium states as a function of pH, including the native state at pH 6 and above, an acid-unfolded state at pH 2, and a partially folded equilibrium state at moderately acidic pH (4.0–4.5).⁵⁶ This “pH 4 intermediate” (termed “*M*-state” in this paper) has structural properties resembling a kinetic intermediate populated within a few ms of initiating folding at pH 6, including stable secondary structure in helices A, G, H and parts of B, which are thought to form native-like helix-helix contacts.^{47, 52, 60} In a sequence-based analysis of the globin superfamily, Ptitsyn *et al.*⁴² identified an evolutionarily conserved cluster of hydrophobic residues not directly involved in cofactor or ligand binding, including Val10 and Trp14 on helix A, Ile111 and Leu115 on helix G, and Met131 and Leu135 on helix H in sperm whale myoglobin. These hydrophobic side chains occupy central positions at helix-helix contact sites between helices A, G and H, and are thought to play a key role during the early stage of the apoMb folding process,^{42, 60–61} but previous kinetic studies have lacked the time resolution to test this hypothesis.

Using NMR-detected pulsed H/D exchange methods,^{62–63} Jennings *et al.* showed that amide protons in helices A, G and H and the N-terminus of B become protected from solvent exchange within the first 6 ms of refolding under native conditions,⁴⁷ indicating the existence of a partially structured state in which a stable core of interacting helices forms early in the folding of apoMb. By combining H/D exchange labeling with mutagenesis, focusing on the bulky hydrophobic residues at helix-helix contact sites on the A, B, G, H and E helices, Nishimura *et al.*^{58, 64–65} showed that the early kinetic intermediate contains both native-like secondary structure in helices A, G, H and parts of helix B as well as non-native packing interactions between helices G and H.^{65–66} Equilibrium H/D exchange and

mutagenesis studies showed that the A, G and H helices contain stable H-bonded structure and engage in mutual hydrophobic interactions in the pH 4 intermediate.^{55, 60, 67} By working at pH 4.2, we can thus focus on the folding events that occur prior to the rate-limiting step in forming the *M*-state.

Using continuous-flow mixing coupled with CD and SAXS detection, Uzawa *et al.* observed a major chain compaction coupled with a large increase in helix content within 350 μ s of the refolding of horse apoMb.⁵³ Their data are consistent with a sequential four-state mechanism containing two on-pathway folding intermediates. In a kinetic study of the role of hydrophobic side chains on the folding of apoMb, Samatova, et al. showed that packing interactions involving the conserved core residues on helices A, G and H were established within milliseconds of refolding while less conserved hydrophobic residues on helices B, D and E did not participate in stabilizing early intermediates, but became important during the late stages of acquiring the native state at pH 6.2.⁵⁴ In most previous studies the very early events in apoMb folding were only manifest as a “burst phase” in measurements using conventional stopped-flow or quenched-flow techniques. In particular, the time-course of forming the “A-G-H intermediate” is yet to be fully resolved. Recent advances in rapid mixing techniques make it possible to detect protein folding dynamics with microsecond time resolution. For example, we developed a fluorescence-detected turbulent-mixing continuous-flow apparatus capable of monitoring protein folding process with an instrumental dead time as short as 30- μ s time resolution,⁸ making it feasible to elucidate complex folding reactions on the sub-millisecond time scale involving multiple kinetic intermediates.²⁷

In this study, we combined ultrafast mixing with site-directed mutagenesis to explore how the conserved residues in the A-G-H helix core and the less conserved residues at peripheral helix-helix contact sites contribute to the formation and unfolding of the “pH 4 intermediate” (*M*-state) of apoMb. The *M*-state is significantly destabilized by mutations in the A-G-H helix core, confirming that this three-helix cluster is intact in this partially unfolded state. Global least-squares fitting of a four-state mechanism to the kinetics of folding and unfolding on the sub-millisecond time scale allowed us to construct free energy diagrams for 12 apoMb variants. Comparison with the free energy landscape of wild-type apoMb²⁷ provides quantitative information on the effects of mutations on free energies of all major states and kinetic barriers encountered during the early stages of apoMb folding and yields new structural insight, including evidence for formation of both native-like and non-native interactions in early folding intermediates followed by structural rearrangement steps.

METHODS

Materials.

Plasmids encoding most of the apoMb variants used in this study were prepared at the Institute of Protein Research in Pushchino⁵⁴ while others were a generous gift from Drs. Jane Dyson and Peter Wright (The Scripps Research Institute). Ultra-pure urea was from MP Biochemicals, Inc. (Solon, OH). Trifluoroacetic acid was from Sigma-Aldrich (St. Louis, MO). Sodium acetate, acetonitrile, HPLC grade methanol and water, and other

chemicals were from Fisher Scientific (Pittsburg, PA). Purchased reagents were used as received.

Protein Expression and Purification.

ApoMb mutants were prepared using a QuickChange® II site-directed mutagenesis kit (Stratagene) following manufacturer's protocol. The mutations were confirmed by sequencing the pET17b plasmid encoding the mutants. Proteins were expressed using the pET17b/BL-21 (DE3) expression system and purified using reverse-phase HPLC as described.²⁷ The purity was checked by SDS-PAGE.

Urea-induced Equilibrium Unfolding of apoMb Mutants.

The urea-induced unfolding of apoMb mutants at pH 4.2 (20 mM sodium acetate; NaAc) was monitored by Trp fluorescence and far-UV CD spectroscopies as described elsewhere.²⁷ A PTI QM-2000 spectrofluorometer (Photon Technology International, Inc., South Brunswick, NJ) was used for measuring Trp fluorescence. Far UV-CD was measured on an AVIV 62DS spectropolarimeter (AVIV Biomedical Inc., Lakewood, NJ). All experiments were done at 11 °C. The normalized unfolding data were fitted to a three-state equilibrium model ($M \rightleftharpoons L_{eq} \rightleftharpoons U$) following a global fitting procedure,^{27, 68} in which the thermodynamic parameters, $C_{m,ML}$, $C_{m,LU}$, (the mid-point concentrations of urea-induced transition between M and L_{eq} and between L_{eq} and U , respectively) and equilibrium m values, m_{ML} and m_{LU} , (linear dependence of free energy dependence on urea concentration M and L_{eq} and between L_{eq} and U , respectively), were treated as global parameters, while the signal of each state and each wavelength were treated as local parameters.

Microsecond Folding/Unfolding Kinetics of apoMb Mutants.

A home-build ultrafast continuous-flow turbulent mixing instrument⁸ and SX-20 stopped-flow instrument (Applied Photophysics, Ltd., Leatherhead, UK) were used to measure the kinetics of folding/unfolding of apoMb mutants on the submillisecond and longer time scales, respectively. The changes in tryptophan fluorescence (>305 nm) upon excitation at 290 nm were measured as a function of urea concentration in 20 mM NaAc, pH 4.2 at 11 °C as described previously.²⁷ Typically, initial protein concentration was 50–60 μ M for a 1:5 dilution experiment. The folding reaction was initiated by diluting apoMb at pH 2.0 with refolding buffer (20 mM NaAc, pH 4.2 containing various concentrations of urea). The unfolding kinetics was measure by diluting the protein solution in 20 mM NaAc to urea solution in 20 mM NaAc, pH 4.2 to reach the desired final urea concentrations. The kinetic traces were fitted to multi-exponential functions using Igor Pro software (WaveMetrics Inc., Lake Oswego, OR).

Kinetic Modeling.

A general matrix-based method was used for kinetic modeling,^{27, 69} using the following four-state mechanism (Scheme 1),

where U is the unfolded state, I and L are kinetic intermediates, and M is an equilibrium intermediate populated at pH 4.2. A rate matrix was constructed based on the differential equations describing the proposed kinetic model.

$$\dot{X} = KX \quad 1$$

where X is a state vector and K is the rate matrix corresponding to Scheme 1. The microscopic rate constant for transitions from state i to state j , k_{ij} , is defined by

$$\log k_{ij} = \log k_{ij}^0 + m_{ij}C/RT \quad 2$$

where k_{ij}^0 is the microscopic rate in the absence of urea, m_{ij} is the kinetic m -value describing the urea-dependence, C is the concentration of urea, R is gas constant and T is the absolute temperature. The time, t , dependence of the fluorescence intensity, $f(t)$, is obtained by

$$f(t) = F_{\text{int}} P \exp(-\Lambda t) P^{-1} X_0 \quad 3$$

$$\Lambda = P^{-1} K P \quad 4$$

Where F_{int} is the vector of fluorescence intensities of each state, P is a matrix of four eigenvectors, X_0 is the state vector at the zero time, and Λ is a diagonal matrix with eigenvalues as the diagonal elements. The majority of kinetic parameters, k_{ij} and m_{ij} , as well as the intrinsic fluorescence intensity, F_{int} , were optimized by global fitting of the kinetic traces, chevron plots and equilibrium titration curves of each variant using a non-linear least squares fitting program build in Igor. The equilibrium titration curves were generated by integrating the equilibrium emission spectra between 305 nm and 450 nm at each urea concentration. The parameters m_{IU} , m_{LM} and m_{ML} were poorly defined for most mutants and were held constant at values of 0.2 kcal/mol, -0.27 kcal/mol and 0.27 kcal/mol, respectively, based on the WT apoMb data. In addition, k_{LM} was also held constant at $1.0 \times 10^5 \text{ s}^{-1}$ for all mutants with single-exponential unfolding kinetics (all except V10A, L69A and L76A). The relative weight of kinetic traces, chevron plots and equilibrium titration curves was set to 1:50:0.1 for the least squares fitting.

The free-energy diagram for each variant was estimated from the elementary rate constants and kinetic m -values, assuming Arrhenius behavior (eq 5):

$$\Delta G_{ij}^{\ddagger} = -RT \log \left(\frac{k_{ij}}{A_0} \right) - m_{ij}C \quad 5$$

where $\Delta G_{ij}^{\ddagger}$ is the height of free energy barrier between states i and j relative to state i , and a value of $5.0 \times 10^5 \text{ s}^{-1}$ was assumed for the Arrhenius factor A_0 .¹

RESULTS

Effects of Core Mutations on the Stability of apoMb.

The effect of mutations on the stability of apoMb was measured by monitoring changes in tryptophan fluorescence and far-UV CD associated with urea-induced unfolding of each variant (Supporting Information, Figure S1). The equilibrium unfolding data were fitted to a three-state model ($M \rightleftharpoons L_{eq} \rightleftharpoons U$) using a global fitting procedure.^{27, 68} Briefly, the urea-dependent tryptophan fluorescence spectra of each mutant were transposed into a set of urea-dependent transitions at individual wavelengths, which were then fitted to the three-state model using global least-squares fitting. The thermodynamic parameters, C_{mML} , m_{ML} , C_{mLU} and m_{LU} , were treated as global parameters as they are independent of wavelength while the intercepts of the spectroscopic signature for each state and their urea-dependency were treated as local parameters (see, e.g., Figure S3). In addition to Trp fluorescence, CD-monitored urea-unfolding data were included and the hybrid dataset was globally fitted to the three-state model,²⁷ which was sufficient to describe the equilibrium data of all variants studied. The thermodynamic parameters thus derived are listed in Table 1, and key parameters, including the free energies of the M - and L_{eq} -states relative to U , G_{MU} and G_{LU} , and $\beta_L = 1 - m_{ML}/m_{MU}$, which provides a measure of the burial of solvent-accessible surface area in the L_{eq} -state relative to that in M , are displayed in Figure 2.

The mutations in the G- and H-helices, I111A, L115A, M131A and L135A, result in substantial destabilization of both the M -state (red bars in Fig 2A) and the equilibrium intermediate L_{eq} (black bars in Figure 2A). The fact that the same set of variants also shows distinctly lower values for β_L (Figure 2B) indicates that these mutations in the center of the primary hydrophobic core result in a less compact equilibrium intermediate compared to WT apoMb as well as all other variants studied. In contrast, the changes in G_{MU} and β_L caused by Ala substitution of I28, M55, L69 and L76 are small or slightly positive, which is consistent with previous evidence, based on mutational analysis and hydrogen exchange data^{52, 67} that these more peripheral residues are not involved in tight, native-like, packing interactions.

Microsecond Folding/Unfolding Kinetics of apoMb Mutants.

For each of the twelve mutants depicted in Figure 1, we carried out continuous-flow fluorescence measurements of the kinetics of folding and unfolding at pH 4.2 vs. urea concentration (starting from the U -stat at pH 2 and the M -state at pH 4.2, respectively), using conditions matching those used in our previous study of WT apoMb.²⁷ The fluorescence signal we monitor is dominated by a pair of largely buried tryptophan residues

on helix A, Trp7 and Trp14, which report on the formation of the central hydrophobic core of the apoMb structure. The kinetic data for two representative variants, V10A and L76A, are shown in Figs. 3 and 4, and those of the remaining variants are included in Supplementary Materials (Figure S2). The folding traces of all mutants are biphasic and qualitatively resemble those of WT apoMb, indicating that an early kinetic intermediate, I , accumulates during folding of all variants studied. The unfolding traces for V10A (Figure 3), L69A (Figure S2) and L76A (Figure 4) can be fitted to double exponential functions consisting of a fast increasing phase and a slow decreasing phase, although the fast phase is less pronounced compared to the WT²⁷, indicating that the unfolding mechanism is conserved. For all other mutants the fast increasing phase is negligible or absent, and the unfolding traces can be fitted to a single exponential function. However, the urea dependence of the major phases in both folding and unfolding qualitatively resembles that of WT apoMb, indicating that all proteins share a common mechanism, but that the time constant of the fast phase is less than the dead time of the mixing apparatus ($\sim 40 \mu\text{s}$) for most variants.

A chevron plot for each mutant is constructed by plotting the apparent folding/unfolding rates as a function of urea concentration on a logarithmic scale (Figs. 3, 4 and S2). We previously reported that the chevron plot of WT apoMb consists of three phases; the fastest phase (λ_1) corresponds to the fast refolding phase; the middle phase (λ_2) corresponds to the fast unfolding phase; and the slowest phase (λ_3) corresponds to the slow (rate-limiting) folding and unfolding phases. Phase λ_3 of WT apoMb shows a pronounced downward curvature (roll-over) in the unfolding limb, indicating that a different barrier limits the rate of unfolding at intermediate (2 – 4 M urea) compared to high (> 4 M) urea concentrations. Although most mutations cause significant changes in the observable rate constants, the shape of $\log(\lambda_3)$ vs. urea concentration is qualitatively similar to that of the wild-type with a clear roll-over above 2 M urea, confirming that the folding/unfolding mechanism is not altered by the mutations.

Kinetic Modeling.

We used a general matrix method⁷⁰ for solving the differential equations describing a sequential four-state kinetic mechanism²⁷ coupled with a least-square fitting procedure for modeling the folding kinetics of apoMb mutants (see Methods). The parameters for WT apoMb were used as the initial guess for fitting the combined data of the mutants, including raw kinetic traces, chevron plots and equilibrium transitions. Two parameters, m_{IU} and m_{LM} , were not well constrained by the available data for some of the mutants and were held constant at the values of the wild type. The elementary rate constants and kinetic m values obtained from the kinetic modeling for the microsecond folding/unfolding kinetics of the twelve apoMb mutants studied are listed in Table 2 along with fitting errors for all of the uniquely determined rate constants. The relative errors for most rate constants are less than 15%, indicating that these parameters are statistically significant. In our global fitting procedure, we assigned relatively high weights to chevron plots; as a result, the urea-dependence of observable rate constants is accurately predicted by the model (solid colored lines in panel C of Figure 3, 4 and S2). We also used the final kinetic parameters, along with additional parameters describing the urea-dependent intrinsic fluorescence of the states, to

predict the time course of folding and unfolding (back lines in the upper panels of Figures 3, 4 and S2). Our ability to qualitatively reproduce the raw kinetic data using a single set of global fitting parameters was critical for discriminating among alternative kinetic mechanisms. For example, a sequential 4-state mechanism in which *L* is a high-energy state rather than a populated intermediate failed to exhibit the observed bell-shaped time course of unfolding (panel B in Figures 3 and 4).

The changes in elementary rate constants caused by mutations can in some cases directly identify key residues involved in a specific folding/unfolding event. For example, while most mutations in the primary hydrophobic core (V10A, I111A, M131A and L135A) have little or no effect on k_{UI} or result in a moderate increase (L115A), the V17A mutation gives rise to a significant decrease in k_{UL} , indicating that interactions of the V17 side chain are important for driving the initial folding step. In contrast, mutations in the secondary hydrophobic core (F33A, F46A and M55A) cause a moderate and those in the B- and E-helices (I28A, L69A and L76A) a more substantial increase in k_{UL} , suggesting that bulky side chains at these sites make it more difficult for the protein to traverse the initial barrier for folding.

Energetic Perturbations due to Mutations.

To quantify the energetic perturbations caused by truncating bulky hydrophobic side chains, we used Eq. 5 to calculate the relative free energies of adjacent ground- and transition-states from the elementary rate constants in the absence of denaturant (k_{ij}). This approach is based on the premise that our relatively conservative mutations (truncation of bulky side chains) do not alter the folding mechanism (i.e., the number and connectivity of the major species involved), but only perturb the relative free energies of different conformational states. Free energy diagrams for the mutant proteins were constructed as described by Xu et al.²⁷ and compared with that of the WT protein (e.g., Figs. 3D and 4D). Errors in each energy level were estimated from the fitting errors of the elementary rate constants using standard error propagation. In the panels on the left of Figure 5, we compare the free energy profiles of WT apoMb (black) with those of the mutants (color), and the panels on the right show the corresponding free energy differences, $\Delta G = G_{\text{mutant}} - G_{\text{WT}}$, for each of the observable states, i.e., *TS1*, *I*, *TS2*, *L* and *M* for all mutants and *TS3* for those with biphasic unfolding kinetics (V10A, L69A and L76A).

Most of the variants with mutations in the primary hydrophobic core (Figure 5A–D) show a gradual and largely monotonic increase in ΔG as the protein progresses from *U* to *M*, indicating that the A/G/H-helix interface is progressively stabilized via cumulative addition of native-like interactions. One exception to this trend is V10A, which shows a sharp decrease in ΔG to a value near 0 in TS3 before rising again to its final positive value at the *M*-state, suggesting that some of the contacts present in the *L*-state have to be disrupted again in order to traverse the final transition state. However, since V10A is the only mutation in the primary core with an observable TS3, it is not clear whether this is an exceptional or more general phenomenon. The fact that mutations on all three core helices (V17A, L115A, M131A and L135A) exhibit high ΔG values in the *L*-intermediate, comparable to those of the *M*-state (corresponding to a Φ -value near unity; see below), indicates that the primary

hydrophobic core is well developed in this late folding intermediate. In contrast, the *L*-state

ΔG values for V10A and I111A are much lower than those of the *M*-state, indicating that the N-termini of the A- and G-helices become fully structured only in the *M*-state.

In addition to the primary core, the native apoMb structure at neutral pH has a separate hydrophobic cluster surrounded by the B-, C- and D-helices at the periphery of the structure (Figure 1).⁷¹ Consistent with the lack of significant protection observed in prior H/D exchange experiments at pH 4.0,⁵² we found that Ala substitution of F33, F46 or M55 has little or no effect on the stability of the *M*-state estimated from the kinetic data (Figure 5G), indicating that these side chains are in a largely disordered region of the *M*-state. Nevertheless, these mutations give rise to clear kinetic changes (Figure S2) and concomitant free energy perturbations, including both stabilizing (negative ΔG) as well as destabilizing (positive ΔG) effects (Figure 5G). Ala substitution at all three sites results in a net stabilization of *TS1*, suggesting that the bulkier aromatic and aliphatic side chains form unfavorable interactions that raise this initial folding barrier. The significantly lower ground-state free energy of the *L*-state ($\Delta G_L = -0.9$ kcal/mol) caused by the F46A mutation (Figure 5G) indicates that the native Phe side chain engages in transient non-native contacts in the *L*-intermediate that are lost again in the *M*-state.

The I28A in the B-helix and the L69A and L76A mutations in the E-helix result in only modest (in some cases slightly stabilizing) effects on the free energy of the *M*-state (Figs. 2 and 5H), which is consistent with prior mutational and H/D exchange data^{52, 67} indicating that this region of the protein is largely unstructured in the pH 4 equilibrium state. Thus, it was quite unexpected that these mutations result in significant kinetic perturbations corresponding to a decrease in ΔG for all of the transition states and intermediates relative to *U* (Figure 5H). This effect is especially striking in the case of L69A and L76A, which show virtually identical profiles with ΔG values ranging from -0.6 kcal/mol (for *TS1*, *I*, and *TS2*) to -1 kcal/mol (for *L* and *TS3*). Thus, bulky side-chains hinder folding to the *M*-state by forming energetically unfavorable intramolecular or protein-solvent interactions. The fact that the E-helix mutations exhibit more negative ΔG in *L* and *TS3* compared to *M* suggests that the leucine side chains of WT apoMb engage in transient non-native interactions in *L* and *TS3* that are resolved in the *M*-state.

DISCUSSION

For proteins that exhibit cooperative (two-state) folding behavior, insight into mechanisms of folding can be obtained only indirectly by observing the effects of extrinsic parameters (i.e., temperature, pressure or solvent composition), mutations or ligand binding on the kinetics of folding and unfolding. For example, the mutant Φ -value strategy^{46, 72} can provide insight into the role of individual side chains in stabilizing transition-state ensembles. Proteins such as apoMb that undergo multi-state folding transitions offer additional opportunities for probing the structural characteristics of intermediate states by coupling rapid mixing and other techniques for fast initiation of folding with time-resolved biophysical techniques, including optical spectroscopy, pulsed H/D exchange and SAXS.^{20, 22, 47, 62–63, 73–82} In this study, we combined multi-state kinetic analysis with mutagenesis in order to gain further insight into transient intermediates and kinetic barriers encountered prior to the rate-limiting

folding step. Most previous applications of systematic mutational analysis to study early folding events in multi-state folders, except for a study of homeodomain folding using the temperature-jump method,⁸³ focused on three-state systems with kinetically unresolved early intermediates.^{54,84–86}

Non-cooperative Folding Equilibrium.

Some structural insight can already be gleaned from our equilibrium analysis of the apoMb variants, which revealed a well-populated equilibrium intermediate (Figure 2). The Tanford β -value defined as the fractional m -value of an intermediate (or transition state) relative to the total m -value for unfolding is often used as a reaction coordinate of protein folding as it reports on changes in solvent-accessible surface area. Mutations in the primary core, especially in the G- and H-helices, give rise to a pronounced decrease in the β -value of L_{eq} (β_L), indicating that L_{eq} of the mutant is more highly solvated compared to WT apoMb. This conclusion is based on the premise that the mutation does not alter the folding mechanism. Our finding that the folding/unfolding kinetics of mutant proteins are qualitatively similar to those of wild type protein (Figs. 3, 4 and S2) is a strong indication that the mutations do not alter the mechanism in terms of the number and connectivity of populated states. The observed mutational effects on β_L may be in part due to the fact that, because of the marginal stability of M , other intermediates, including L and I , can also contribute to the equilibrium population at pH 4.2.^{27, 87} According to Figure 5, the L -state is the second most stable state for the wild type and most mutants, but the mutations in the G- and H-helices preferentially destabilize the L -state so that both L and I accumulate. Because I is less structured than L , this may account for the apparent decrease in the equilibrium β -value.

Multi-state Mutational Φ -values.

To summarize the mutational effects on apoMb folding dynamics, Figure 6A compares the free energy perturbations for all mutants, using bars of different color to indicate the ground- and transition states of the underlying four-state mechanism. For each point i along this “reaction coordinate”, G_i reflects the average number and strength of interactions lost by truncating a specific side chain to alanine, and thus provides qualitative structural information for the various (transition) states encountered during folding. For mutations that cause a significant change in the free energy of the M -state relative to U ($-G_M$), which includes all residues in the primary core (Figure 6B), we can calculate mutational Φ -values as

$$\Phi_i = \Delta\Delta G_i / \Delta\Delta G_M \quad \text{EQ6}$$

where G_i represents the mutational free energy perturbations for each of the observable ground and transition states, i , and G_M is the equilibrium free energy change for the M -state derived from the kinetic data (see Methods). While Φ -values can provide residue-specific information on the fraction of native-like contacts formed at different stages of folding, because of the uncertainty in G_M the error associated with Φ_i is generally much larger than that of the corresponding G_i . This uncertainty is compounded by the lack of

information on the final barrier, *TS3*, for most mutants. Thus, we have to use caution in interpreting Φ -values.

Among the twelve residues targeted for Ala substitution in this study, six residues are located in the primary hydrophobic core of apoMb (Figure 1). Five of these (Val10 on helix A; Ile111 and Leu115 on helix G; and Met131 and Leu135 on helix H) are among the most highly conserved residues in the globin fold family (excluding residues involved in functionally important interactions with the heme).⁴² Instead of Trp 14, another strictly conserved core residue, we chose to mutate Val17 at the C-terminus of helix A as we are relying on Trp14 as an intrinsic fluorescence probe. In addition, several hydrophobic residues outside the A/G/H helix cluster were also subjected to alanine substitution, including Phe33, Phe46 and M55 forming a peripheral hydrophobic core in native structure of apoMb, as well as residues in helices B (Ile28) and E (Leu69 and Leu76), which contact the primary A-G-H helix core in the native structure, but appear to be still disordered in the *M*-state.^{52, 55, 88}

Progressive Stabilization of the Primary Hydrophobic Core.

Among the mutations targeting conserved residues on helices A, G and H (V10A, V17A, I111A, L115A, M131A and L135A), most exhibit a gradual, more or less monotonic, increase in Φ (Figure 5A–D) as the protein advances from *U* towards *L*, indicating that the hydrophobic core residues interact only weakly during the initial steps (especially *TS1* and *I*) and become progressively more tightly packed during the later steps (*TS2* and *L*). Our observation that the profiles of free energy perturbations are qualitatively similar for the core residues suggests that the A, G and H helices are involved in mutual contacts, although their relative arrangement is not necessarily the same as in the native structure. The V17A and L135A mutants also show monotonic increases in Φ -value with fractional values in *I* and *TS2* and values near unity in *L* consistent with stepwise accrual of native-like (or “*M*-like”) interactions. A characteristically different behavior is seen for M131A, which reaches Φ -value near 1 already at an earlier stage (*I* and *TS2*), and L115A, whose Φ -values remain low or slightly negative until *TS2*. However, in terms of free energy perturbations, these mutants are similar to the other primary core variants, suggesting that the apparent differences in Φ -values may be an artifact due to errors in G_M ; note that $G_M \sim 0.5$ kcal/mol for L115A and L135A while all other A/G/H-helix mutants of values near 1 kcal/mol, which is more consistent with the independently measured equilibrium values (Figure 2). Despite this uncertainty, a common feature of all primary core mutants is that the Φ -value at *TS1* is small (0 ± 0.3), indicating that the initial transition state encountered during folding is largely unstructured. This argues against the hypothesis that these highly conserved residues play a key role in folding initiation by forming a network of mutual interactions required for crossing the initial folding barrier.⁴² On the other hand, one or two residues on each of the three central helices reach Φ -values near one at the *L*-state, indicating that mutual contacts between the A, G and H-helices are critical for stabilizing this late intermediate. Given that *M*-state is analogous to a late intermediate in the formation of the *N*-state at higher pH, our results are fully consistent with the conclusion that interactions among these core residues are formed well before the rate-limiting barrier in folding to *N*,⁵⁴ based on stopped-flow experiments at pH 6.2.

Kinetic Perturbations due to Truncation of Hydrophobic Residues in Unstructured Regions of the *M*-state.

Samatova et al. showed that hydrophobic residues in peripheral locations, including I28 on the B-helix, F33, F46 and M55 in the secondary hydrophobic core and L69 and L76 on the E-helix, participate in the final step of folding to the native state at pH 6.2,⁵⁴ but contribute little to the stability of a (late) intermediate detected in their stopped-flow experiments. Since the stability of the equilibrium state at pH 4 is largely unaffected by Ala substitution of these residues (Figs 2 and 6), this observation further supports the notion that the *M*-state represents a stable analogue of a late kinetic intermediate encountered at higher pH during folding to the *N*-state.⁴⁷ Thus, it was quite unexpected that many of these mutations in poorly structured regions of the *M*-state result in significant kinetic changes in our sub-millisecond folding/unfolding experiments (Figure 4, Figure S2). This may in part reflect a difference between equilibrium and kinetic intermediates, as suggested by Nishimura et al.⁵², based on pulsed H/D exchange measurements at pH 6.2, which showed stronger protection in the B-helix and weaker protection in the D-helix in the kinetic intermediate compared to the equilibrium state at pH 4.2. However, this does not fully explain our observation that truncation of large hydrophobic side chains can result in significant stabilization of transient intermediates and transition-state ensembles during formation of the *M*-state, which is especially striking in the case of the E-helix residues L69 and L76, and to a lesser degree I28 and F46 (Figure 6). Since these side chains are likely to be in largely solvent-accessible locations both in the initial (*U*) and final (*M*) state, our observations imply that they form transient (non-native) intramolecular contacts or unfavorable solvent interactions that effectively slow down the kinetics in both the folding and unfolding direction. The largest stabilization due to Ala-replacement of L69 and L76 is seen for the late folding transition state, *TS3*, suggestive of a highly non-native conformation that is resolved via structural rearrangements in forming the *M*-state. This is reminiscent of observations, based on H/D-exchange and FRET data, that during folding at neutral pH the G- and H-helices initially form non-native packing interactions, which are resolved by subsequent rearrangements involving translation of the H-helix relative to the A/G-helix pair by one helical turn.^{65–66} Our findings indicate that misfolding and rearrangement events occur also during the very early stages of folding under conditions (pH 4.2) where the folding reaction ends at the partially folded *M*-state.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Drs. H. Jane Dyson and Peter Wright for providing expression plasmids for apomyoglobin variants, and Dr. Vincent Voelz for helpful discussions. The work was supported by NSF Grant MCB-1412378 and NIH grant GM116911 to H.R., and NIH Grant CA06927 to the Fox Chase Cancer Center. The Spectroscopy Support and DNA sequencing Facilities at the Fox Chase Cancer Center provided technical support.

REFERENCES

- (1.) Kubelka J; Hofrichter J; Eaton WA, The Protein Folding 'Speed Limit'. *Curr. Opin. Struct. Biol* 2004, 14, 76–88. [PubMed: 15102453]
- (2.) Schaeffer RD; Fersht A; Daggett V, Combining Experiment and Simulation in Protein Folding: Closing the Gap for Small Model Systems. *Curr. Opin. Struct. Biol* 2008, 18, 4–9. [PubMed: 18242977]
- (3.) Bowman GR; Voelz VA; Pande VS, Taming the Complexity of Protein Folding. *Curr. Opin. Struct. Biol* 2011, 21, 4–11. [PubMed: 21081274]
- (4.) Lindorff-Larsen K; Piana S; Dror RO; Shaw DE, How Fast-Folding Proteins Fold. *Science* 2011, 334, 517–20. [PubMed: 22034434]
- (5.) Cellmer T; Buscaglia M; Henry ER; Hofrichter J; Eaton WA, Making Connections between Ultrafast Protein Folding Kinetics and Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A* 2011, 108, 6103–8. [PubMed: 21441105]
- (6.) Dyer BR; Gai F; Woodruff WH; Gilmanshin R; Callender RH, Infrared Studies of Fast Events in Protein Folding. *Acc. Chem. Res* 1988, 31, 709–716.
- (7.) Eaton WA; Muñoz V; Thompson PA; Chan C-K; Hofrichter J, Submillisecond Kinetics of Protein Folding. *Curr. Opin. Struct. Biol* 1997, 7, 10–14. [PubMed: 9032067]
- (8.) Shastry MCR; Luck SD; Roder H, A Continuous-Flow Capillary Mixer to Monitor Reactions on the Microsecond Time Scale. *Biophys. J* 1998, 74, 2714–2721. [PubMed: 9591695]
- (9.) Knight JB; Vishwanath A; Brody JP; Austin RH, Hydrodynamic Focusing on a Silicon Chip: Mixing Nanoliters in Microseconds. *Phys. Rev. Lett* 1998, 80, 3863–3866.
- (10.) Lapidus LJ; Eaton WA; Hofrichter J, Measuring the Rate of Intramolecular Contact Formation in Polypeptides. *Proc. Natl. Acad. Sci. U. S. A* 2000, 97, 7220–5. [PubMed: 10860987]
- (11.) Pollack L; Tate MW; Finnefrock AC; Kalidas C; Trotter S; Darnton NC; Lurio L; Austin RH; Batt CA; Gruner SM, et al., Time Resolved Collapse of a Folding Protein Observed with Small Angle X-Ray Scattering. *Phys. Rev. Lett* 2001, 86, 4962–5. [PubMed: 11384392]
- (12.) Bilsel O; Kayatekin C; Wallace LA; Matthews CR, A Microchannel Solution Mixer for Studying Microsecond Protein Folding Reactions. *Rev. Sci. Instr* 2005, 76, 014302.
- (13.) Gai F; Du D; Xu Y, Infrared Temperature-Jump Study of the Folding Dynamics of Alpha-Helices and Beta-Hairpins. *Methods Mol Biol* 2007, 350, 1–20. [PubMed: 16957314]
- (14.) Schuler B; Eaton WA, Protein Folding Studied by Single-Molecule FRET. *Curr. Opin. Struct. Biol* 2008, 18, 16–26. [PubMed: 18221865]
- (15.) Snow CD; Sorin EJ; Rhee YM; Pande VS, How Well Can Simulation Predict Protein Folding Kinetics and Thermodynamics? *Annu Rev Biophys Biomol Struct* 2005, 34, 43–69. [PubMed: 15869383]
- (16.) Noe F; Schutte C; Vanden-Eijnden E; Reich L; Weikl TR, Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. U. S. A* 2009, 106, 19011–6. [PubMed: 19887634]
- (17.) Kuwajima K, The Molten Globule State as a Clue for Understanding the Folding and Cooperativity of Globular-Protein Structure. *Proteins: Struct., Funct., Genet* 1989, 6, 87–103. [PubMed: 2695928]
- (18.) Matthews CR, Pathways of Protein Folding. *Annu. Rev. Biochem* 1993, 62, 653–683. [PubMed: 8352599]
- (19.) Roder H; Colón W, Kinetic Role of Early Intermediates in Protein Folding. *Curr. Opin. Struct. Biol* 1997, 7, 15–28. [PubMed: 9032062]
- (20.) Roder H; Maki K; Cheng H, Early Events in Protein Folding Explored by Rapid Mixing Methods. *Chem Rev* 2006, 106, 1836–61. [PubMed: 16683757]
- (21.) Englander SW; Mayne L; Krishna MM, Protein Folding and Misfolding: Mechanism and Principles. *Q. Rev. Biophys* 2007, 40, 287–326. [PubMed: 18405419]
- (22.) Chan C-K; Hu Y; Takahashi S; Rousseau DL; Eaton WA; Hofrichter J, Submillisecond Protein Folding Kinetics Studied by Ultrarapid Mixing. *Proc. Natl. Acad. Sci. U. S. A* 1997, 94, 1779–1784. [PubMed: 9050855]

23.) Shastry MCR; Roder H, Evidence for Barrier-Limited Protein Folding Kinetics on the Microsecond Time Scale. *Nat. Struct. Biol* 1998, 5, 385–392. [PubMed: 9587001]
24.) Roder H; Maki K; Cheng H; Shastry MC, Rapid Mixing Methods for Exploring the Kinetics of Protein Folding. *Methods* 2004, 34, 15–27. [PubMed: 15283912]
25.) Garcia P; Bruix M; Rico M; Ciofi-Baffoni S; Banci L; Ramachandra Shastry MC; Roder H; de Lumley Woodyear T; Johnson CM; Fersht AR, et al., Effects of Heme on the Structure of the Denatured State and Folding Kinetics of Cytochrome B(562). *J. Mol. Biol* 2005, 346, 331–44. [PubMed: 15663948]
26.) Lapidus LJ; Yao S; McGarrity KS; Hertzog DE; Tubman E; Bakajin O, Protein Hydrophobic Collapse and Early Folding Steps Observed in a Microfluidic Mixer. *Biophys. J* 2007, 93, 218–24. [PubMed: 17416618]
27.) Xu M; Beresneva O; Rosario R; Roder H, Microsecond Folding Dynamics of Apomyoglobin at Acidic Ph. *J Phys Chem B* 2012, 116, 7014–25. [PubMed: 22475221]
28.) Park S-H; Shastry MCR; Roder H, Folding Dynamics of the B1 Domain of Protein G Explored by Ultrarapid Mixing. *Nat. Struct. Biol* 1999, 6, 943–947. [PubMed: 10504729]
29.) Capaldi AP; Shastry RMC; Kleanthous C; Roder H; Radford SE, Ultra-Rapid Mixing Experiments Reveal That Im7 Folds *Via* an on-Pathway Intermediate. *Nat. Struct. Biol* 2001, 8, 68–72. [PubMed: 11135674]
30.) Teilum K; Maki K; Kragelund BB; Poulsen FM; Roder H, Early Kinetic Intermediate in the Folding of Acyl-CoA Binding Protein Detected by Fluorescence Labeling and Ultrarapid Mixing. *Proc. Natl. Acad. Sci. U. S. A* 2002, 99, 9807–12. [PubMed: 12096190]
31.) Hagen SJ; Eaton WA, Two-State Expansion and Collapse of a Polypeptide. *J. Mol. Biol* 2000, 297, 781–9. [PubMed: 10731428]
32.) Krantz BA; Mayne L; Rumbley J; Englander SW; Sosnick TR, Fast and Slow Intermediate Accumulation and the Initial Barrier Mechanism in Protein Folding. *J. Mol. Biol* 2002, 324, 359–71. [PubMed: 12441113]
33.) Fazelinia H; Xu M; Cheng H; Roder H, Ultrafast Hydrogen Exchange Reveals Specific Structural Events During the Initial Stages of Folding of Cytochrome C. *J. Am. Chem. Soc* 2014, 136, 733–40. [PubMed: 24364692]
34.) Bryngelson JD; Onuchic JN; Socci ND; Wolynes PG, Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins: Struct., Funct., Genet* 1995, 21, 167–195. [PubMed: 7784423]
35.) Onuchic JN; Wolynes PG, Theory of Protein Folding. *Curr. Opin. Struct. Biol* 2004, 14, 70–5. [PubMed: 15102452]
36.) Voelz VA; Jäger M; Yao S; Chen Y; Zhu L; Waldauer SA; Bowman GR; Friedrichs M; Bakajin O; Lapidus LJ, et al., Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment. *J. Am. Chem. Soc* 2012, 134, 12565–12577. [PubMed: 22747188]
37.) Shakhnovich E; Abkevich V; Ptitsyn OB, Conserved Residues and the Mechanism of Protein Folding. *Nature* 1996, 379, 96–98. [PubMed: 8538750]
38.) Fersht AR, Nucleation Mechanisms in Protein Folding. *Curr. Opin. Struct. Biol* 1997, 7, 3–9. [PubMed: 9032066]
39.) Klimov DK; Thirumalai D, Lattice Models for Proteins Reveal Multiple Folding Nuclei for Nucleation-Collapse Mechanism. *J. Mol. Biol* 1998, 282, 471–92. [PubMed: 9735420]
40.) Michnick SW; Shakhnovich E, A Strategy for Detecting the Conservation of Folding-Nucleus Residues in Protein Superfamilies. *Fold Des* 1998, 3, 239–51. [PubMed: 9710570]
41.) Ptitsyn OB, Protein Folding and Protein Evolution: Common Folding Nucleus in Different Subfamilies of C-Type Cytochromes? *J. Mol. Biol* 1998, 278, 655–666. [PubMed: 9600846]
42.) Ptitsyn OB; Ting KL, Non-Functional Conserved Residues in Globins and Their Possible Role as a Folding Nucleus. *J. Mol. Biol* 1999, 291, 671–82. [PubMed: 10448045]
43.) Galzitskaya OV; Finkelstein AV, A Theoretical Search for Folding/Unfolding Nuclei in Three-Dimensional Protein Structures. *Proc. Natl. Acad. Sci. U. S. A* 1999, 96, 11299–304. [PubMed: 10500171]
44.) Daggett V; Fersht AR, Is There a Unifying Mechanism for Protein Folding? *Trends Biochem. Sci* 2003, 28, 18–25. [PubMed: 12517448]

45.) Matthews CR, Effect of Point Mutations on the Folding of Globular Proteins. *Methods Enzymol* 1987, 154, 498–511. [PubMed: 3431461]
46.) Matouschek A; Kellis JT Jr.; Serrano L; Fersht AR, Mapping the Transition State and Pathway of Protein Folding by Protein Engineering. *Nature* 1989, 340, 122–6. [PubMed: 2739734]
47.) Jennings PA; Wright PE, Formation of a Molten Globule Intermediate Early in the Kinetic Folding Pathway of Apomyoglobin. *Science* 1993, 262, 892–895. [PubMed: 8235610]
48.) Eliezer D; Jennings PA; Wright PE; Doniach S; Hodgson KO; Tsuruta H, The Radius of Gyration of an Apomyoglobin Folding Intermediate. *Science* 1995, 270, 487–488. [PubMed: 7570004]
49.) Loh SN; Kay MS; Baldwin RL, Structure and Stability of a 2nd Molten Globule Intermediate in the Apomyoglobin Folding Pathway. *Proc. Natl. Acad. Sci. U. S. A* 1995, 92, 5446–5450. [PubMed: 7777528]
50.) Jamin M; Baldwin RL, Refolding and Unfolding Kinetics of the Equilibrium Folding Intermediate of Apomyoglobin. *Nat. Struct. Biol* 1996, 3, 613–618. [PubMed: 8673605]
51.) Jamin M; Yeh SR; Rousseau DL; Baldwin RL, Submillisecond Unfolding Kinetics of Apomyoglobin and Its Ph 4 Intermediate. *J. Mol. Biol* 1999, 292, 731–40. [PubMed: 10497035]
52.) 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. U. S. A* 2005, 102, 4765–70. [PubMed: 15769860]
53.) Uzawa T; Akiyama S; Kimura T; Takahashi S; Ishimori K; Morishima I; Fujisawa T, Collapse and Search Dynamics of Apomyoglobin Folding Revealed by Submillisecond Observations of α -Helical Content and Compactness. *Proc. Natl. Acad. Sci. USA* 2004, 101, 1171–1176. [PubMed: 14711991]
54.) Samatova EN; Melnik BS; Balobanov VA; Katina NS; Dolgikh DA; Semisotnov GV; Finkelstein AV; Bychkova VE, Folding Intermediate and Folding Nucleus for I $\rightarrow$ N and U $\rightarrow$ I $\rightarrow$ N Transitions in Apomyoglobin: Contributions by Conserved and Nonconserved Residues. *Biophys. J* 2010, 98, 1694–702. [PubMed: 20409491]
55.) Hughson FM; Wright PE; Baldwin RL, Structural Characterization of a Partly Folded Apomyoglobin Intermediate. *Science* 1990, 249, 1544–1548. [PubMed: 2218495]
56.) Barrick D; Baldwin RL, Three-State Analysis of Spermin Whale Apomyoglobin Folding. *Biochemistry* 1993, 32, 3790–3796. [PubMed: 8466917]
57.) Cocco MJ; Lecompte JT, The Native State of Apomyoglobin Described by NMR Spectroscopy: Interactions with the Paramagnetic Probe Hytempo and the Fluorescent Dye Ans. *Protein Science* 1994, 3, 267–281. [PubMed: 8003963]
58.) Nishimura C; Dyson HJ; Wright PE, The Apomyoglobin Folding Pathway Revisited: Structural Heterogeneity in the Kinetic Burst Phase Intermediate. *J. Mol. Biol* 2002, 322, 483–9. [PubMed: 12225742]
59.) Nishimura C; Dyson HJ; Wright PE, Consequences of Stabilizing the Natively Disordered F Helix for the Folding Pathway of Apomyoglobin. *J. Mol. Biol* 2011, 411, 248–63. [PubMed: 21640124]
60.) Hughson FM; Baldwin RL, Use of Site-Directed Mutagenesis to Destabilize Native Apomyoglobin Relative to Folding Intermediates. *Biochemistry* 1989, 28, 4415–4422. [PubMed: 2765493]
61.) Kay MS; Baldwin RL, Packing Interactions in the Apomyoglobin Folding Intermediate. *Nat. Struct. Biol* 1996, 3, 439–445. [PubMed: 8612074]
62.) 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]
63.) 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]
64.) 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]
65.) Nishimura C; Dyson HJ; Wright PE, Identification of Native and Non-Native Structure in Kinetic Folding Intermediates of Apomyoglobin. *J. Mol. Biol* 2006, 355, 139–56. [PubMed: 16300787]

66.) Aoto PC; Nishimura C; Dyson HJ; Wright PE, Probing the Non-Native H Helix Translocation in Apomyoglobin Folding Intermediates. *Biochemistry* 2014, 53, 3767–80. [PubMed: 24857522]
67.) Hughson FM; Barrick D; Baldwin RL, Probing the Stability of a Partly Folded Apomyoglobin Intermediate by Site-Directed Mutagenesis. *Biochemistry* 1991, 30, 4113–4118. [PubMed: 2021603]
68.) Latypov RF; Cheng H; Roder NA; Zhang J; Roder H, Structural Characterization of an Equilibrium Unfolding Intermediate in Cytochrome C. *J. Mol. Biol* 2006, 357, 1009–25. [PubMed: 16473367]
69.) Walkenhorst WF; Green SM; Roder H, Kinetic Evidence for Folding and Unfolding Intermediates in Staphylococcal Nuclease. *Biochemistry* 1997, 36, 5795–5805.
70.) Berberan-Santos MN; Martinho JMG, The Integration of Kinetic Rate Equations by Matrix Methods. *J. Chem. Educ* 1990, 67, 375–379.
71.) Lecomte JT; Sukits SF; Bhattacharya S; Falzone CJ, Conformational Properties of Native Sperm Whale Apomyoglobin in Solution. *Protein Sci* 1999, 8, 1484–91. [PubMed: 10422837]
72.) Fersht AR; Matouschek A; Serrano L, The Folding of an Enzyme. I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol* 1992, 224, 771–782. [PubMed: 1569556]
73.) Roder H; Wüthrich K, Protein Folding Kinetics by Combined Use of Rapid Mixing Techniques and NMR Observation of Individual Amide Protons. *Proteins: Struct., Funct., Genet* 1986, 1, 34–42. [PubMed: 2835760]
74.) Roder H, Structural Characterization of Protein Folding Intermediates by Proton Magnetic Resonance and Hydrogen Exchange. *Methods Enzymol* 1989, 176, 446–473. [PubMed: 2811697]
75.) Radford SE; Dobson CM; Evans PA, The Folding of Hen Lysozyme Involves Partially Structured Intermediates and Multiple Pathways. *Nature* 1992, 358, 302–307. [PubMed: 1641003]
76.) Miranker A; Robinson CV; Radford SE; Aplin RT; Dobson CM, Detection of Transient Protein Folding Populations by Mass Spectrometry. *Science* 1993, 262, 896–899. [PubMed: 8235611]
77.) Jones CM; Henry ER; Hu Y; Chan C-K; Luck SD; Bhuyan A; Roder H; Hofrichter J; Eaton WA, Fast Events in Protein Folding Initiated by Nanosecond Laser Photolysis. *Proc. Natl. Acad. Sci. U. S. A* 1993, 90, 11860–11864. [PubMed: 8265638]
78.) Takahashi S; Yeh S-R; Das TK; Chan C-K; Gottfried DS; Rousseau DL, Folding of Cytochrome C Initiated by Submillisecond Mixing. *Nat. Struct. Biol* 1997, 4, 44–50. [PubMed: 8989323]
79.) Shastry MCR; Sauder JM; Roder H, Kinetic and Structural Analysis of Submillisecond Folding Events in Cytochrome C. *Acc. Chem. Res* 1998, 31, 717–725.
80.) Gruebele M; Sabelko J; Ballew R; Ervin J, Laser Temperature Jump Induced Protein Folding. *Acc. Chem. Res* 1998, 31, 699–707.
81.) Bilsel O; Matthews CR, Barriers in Protein Folding Reactions. *Adv. Protein Chem* 2000, 53, 153–207. [PubMed: 10751945]
82.) Kathuria SV; Guo L; Graceffa R; Barrea R; Nobrega RP; Matthews CR; Irving TC; Bilsel O, Minireview: Structural Insights into Early Folding Events Using Continuous-Flow Time-Resolved Small-Angle X-Ray Scattering. *Biopolymers* 2011, 95, 550–8. [PubMed: 21442608]
83.) Banachewicz W; Religa TL; Schaeffer RD; Daggett V; Fersht AR, Malleability of Folding Intermediates in the Homeodomain Superfamily. *Proc. Natl. Acad. Sci. U. S. A* 2011, 108, 5596–5601. [PubMed: 21422286]
84.) Capaldi AP; Kleanthous C; Radford SE, Im7 Folding Mechanism: Misfolding on a Path to the Native State. *Nat Struct Biol* 2002, 9, 209–16. [PubMed: 11875516]
85.) Matouschek A; Serrano L; Fersht AR, The Folding of an Enzyme. Iv. Structure of an Intermediate in the Refolding of Barnase Analysed by a Protein Engineering Procedure. *J. Mol. Biol* 1992, 224, 819–35. [PubMed: 1569559]
86.) Nölting B; Golbik R; Neira JL; Soler-Gonzalez AS; Schreiber G; Fersht AR, The Folding Pathway of a Protein at High Resolution from Microseconds to Seconds. *Proc. Natl. Acad. Sci. U. S. A* 1997, 94, 826–830. [PubMed: 9023341]
87.) Jamin M; Baldwin RL, Two Forms of the pH 4 Folding Intermediate of Apomyoglobin. *J. Mol. Biol* 1998, 276, 491–504. [PubMed: 9512718]

88.) Uzawa T; Nishimura C; Akiyama S; Ishimori K; Takahashi S; Dyson HJ; Wright PE, Hierarchical Folding Mechanism of Apomyoglobin Revealed by Ultra-Fast H/D Exchange Coupled with 2D NMR. Proc. Natl. Acad. Sci. U. S. A 2008, 105, 13859–64. [PubMed: 18779573]

Figure 1.
Structure of sperm whale myoglobin (PDB 2JHO) with the heme group removed. The side chains that were subjected to Ala-substitution in this study are shown in the ball-and-stick representation.

Figure 2.

Summary of mutational perturbations in stability (A) and β_L -value (B) obtained by equilibrium fluorescence and CD measurements vs. urea concentration. Red bars in panel A show the unfolding free energy of the *M*-state, and gray bars shown that of the *L*-state.

Figure 3.

The folding kinetics of V10A variant at pH 4.2. Panels A and B show the refolding and unfolding kinetic traces at various urea concentrations monitored by Trp fluorescence, respectively. The solid lines represent the time course of folding (A) or unfolding (B) predicted by global fitting of a four-state mechanism (Scheme 1) to the combined kinetic data. Panel C shows the corresponding chevron plots. Open circles were obtained by individual fitting of kinetic traces to single or double exponential functions. The rates assigned to λ_1 , λ_2 and λ_3 are shown in red, green and blue, respectively. Solid lines represent simulated chevron plots obtained by the global fitting of the combined kinetic and equilibrium data. The corresponding elementary rate constants are shown in the dashed lines. Black lines represent the fitted rate constants of WT apoMb. Panel D shows the free energy diagrams for V10A (red) and WT apoMb (black) derived from the elementary rate constants in the absence of urea.

Figure 4.

The folding kinetics of L76A variant at pH 4.2. As in Figure 3, panels A and B show the refolding and unfolding kinetic traces vs. urea concentration, and panels C and D show the corresponding chevron plot and free energy diagrams, respectively.

Figure 5.

Free energy diagrams (A-D) of apoMb mutants at pH 4.2 in the absence of urea and the free energy perturbations, $\Delta\Delta G$ of each mutant relative to WT apoMb (E-H). The activation free energy for each transition is calculated from the elementary rate constants shown in Table 2, using an Arrhenius pre-exponential factor of $5.0 \times 10^5 \text{ s}^{-1}$. Except for V10A, L69A, L76A and WT, the values for TS3 are shown in dotted lines because a fast unfolding phase (λ_2) that defines this barrier was not observed for these mutants.

Figure 6.

Summary of mutational perturbations in kinetic parameters, including $\Delta\Delta G$ values for all variants studied (A) and Φ -values of primary core mutants (B-D). In panel A $\Delta\Delta G$ values of *TS1* are shown in blue, *I*-state in cyan, *TS2* in green, *L*-state in orange, and *M*-state in red (see inset). For V10A, L69A and L76A, $\Delta\Delta G$ values of *TS3* are shown in gray, while for the other mutants, $\Delta\Delta G$ values of *TS3* are not shown.

Thermodynamic parameters of wild-type and mutant apoMb at pH 4.2. Urea-unfolding of the mutants monitored by Trp fluorescence and far-UV CD spectroscopy were fitted to a three-state equilibrium model following a global fitting procedure.

Table 1.

	m (kcal mol ⁻¹ M ⁻¹)		β_L (=1- m_{NL}/m_{MU})		G_{NL} (kcal mol ⁻¹)	G_{LU} (kcal mol ⁻¹)	G_{MU} (kcal mol ⁻¹)
	$M \rightleftharpoons L_{eq}$	$L_{eq} \rightleftharpoons U$	$M \rightleftharpoons U$				
WT1	0.72 ± 0.03	1.34 ± 0.01	2.06 ± 0.03	0.65 ± 0.02	0.25 ± 0.11	2.08 ± 0.01	2.33 ± 0.11
WT2	0.92 ± 0.05	1.35 ± 0.01	2.28 ± 0.05	0.59 ± 0.02	0.35 ± 0.13	2.16 ± 0.01	2.51 ± 0.13
V10A	0.76 ± 0.04	0.98 ± 0.01	1.74 ± 0.04	0.56 ± 0.02	0.82 ± 0.08	1.12 ± 0.04	1.94 ± 0.09
V17A	0.80 ± 0.04	1.22 ± 0.01	2.02 ± 0.04	0.60 ± 0.02	0.70 ± 0.07	1.44 ± 0.02	2.14 ± 0.07
I111A	1.16 ± 0.06	1.00 ± 0.02	2.16 ± 0.06	0.46 ± 0.03	0.99 ± 0.08	0.58 ± 0.06	1.57 ± 0.10
L115A	0.77 ± 0.04	0.89 ± 0.01	1.66 ± 0.04	0.54 ± 0.02	1.22 ± 0.21	0.24 ± 0.13	1.46 ± 0.25
M131A	0.89 ± 0.03	0.87 ± 0.01	1.76 ± 0.03	0.49 ± 0.02	1.28 ± 0.10	0.32 ± 0.09	1.60 ± 0.13
L135A	0.90 ± 0.03	0.83 ± 0.01	1.73 ± 0.03	0.48 ± 0.02	1.43 ± 0.14	0.06 ± 0.20	1.49 ± 0.25
I28A	0.71 ± 0.03	1.27 ± 0.01	1.98 ± 0.03	0.64 ± 0.01	0.53 ± 0.04	1.82 ± 0.01	2.35 ± 0.04
F33A	0.56 ± 0.05	1.27 ± 0.01	1.83 ± 0.05	0.69 ± 0.03	0.22 ± 0.16	1.46 ± 0.01	1.68 ± 0.16
F46A	0.92 ± 0.03	1.20 ± 0.01	2.12 ± 0.03	0.57 ± 0.02	1.00 ± 0.04	1.65 ± 0.01	2.64 ± 0.04
M55A	0.75 ± 0.03	1.27 ± 0.01	2.02 ± 0.03	0.63 ± 0.02	0.72 ± 0.05	1.95 ± 0.01	2.66 ± 0.05
L69A	0.82 ± 0.03	1.19 ± 0.01	2.02 ± 0.03	0.59 ± 0.01	0.93 ± 0.04	1.75 ± 0.01	2.68 ± 0.04
L76A	0.58 ± 0.03	1.13 ± 0.01	1.71 ± 0.03	0.66 ± 0.02	0.91 ± 0.16	1.40 ± 0.03	2.32 ± 0.17

Table 2.

Elementary rate constants in ms^{-1} (A) and kinetic m -values in cal/mol/M (B) for apoMb mutants at pH 4.2.

A.	V10A	V17A	I28A	F33A	F46A	M55A	L69A	L76A	I111A	L115A	M131A	L135A	WT
k_{UI}	23.1 ± 3.6	12.3 ± 1.5	38.0 ± 0.8	39.8 ^b	31.6 ^b	24.5 ± 0.9	43.0 ± 0.7	48.9 ± 0.8	18.2 ± 2.4	28.7 ± 1.8	17.2 ± 1.7	19.0 ± 1.4	18.0 ^a
k_{IU}	17.6	13.9	11.9	9.66	21.5	12.3	5.71	6.30	12.9	7.90	17.2	11.2	7.00 ^a
k_{UL}	± 1.8 8.46	± 0.3 8.33	± 0.2 11.6	± 0.45 4.54	± 0.8 7.64	± 0.3 6.56	± 0.48 8.54	± 0.53 8.25	± 0.8 10.2	± 1.12 5.30	± 0.6 8.54	± 1.1 5.43	8.50 ^a
	± 1.13	± 0.98	± 0.2	± 0.06	± 0.15	± 0.20	± 0.14	± 0.13	± 1.3	± 0.28	± 0.78	± 0.39	
k_{LI}	2.19	6.24	1.02	2.91	0.23	1.71	0.92	1.05	2.12	6.12	2.88	4.14	1.90 ^a
	± 0.09	± 0.31	± 0.05	± 0.17	± 0.01	± 0.12	± 0.03	± 0.03	± 0.06	± 0.54	± 0.18	± 0.22	
k_{LM}	23.1 ± 0.9	100 ^b ± 0.61	100 ^b ± 0.9	100 ^b ± 0.35	100 ^b ± 7	100 ^b ± 0.84	12.5 ± 0.05	12.2 ± 0.05	100 ^b ± 8.5	100 ^b ± 0.54	100 ^b ± 0.52	100 ^b ± 0.81	9.00 ^a
k_{ML}	5.48 ± 0.15	9.72 ± 0.61	10.9 ± 0.9	5.66 ± 0.35	125 ± 7	9.09 ± 0.84	3.20 ± 0.05	2.88 ± 0.05	60.5 ± 8.5	6.71 ± 0.54	7.31 ± 0.52	9.94 ± 0.81	1.01 ^a
B.	V10A	V17A	I28A	F33A	F46A	M55A	L69A	L76A	I111A	L115A	M131A	L135A	WT
m_{UI}	-349	-295 ^{ab}	-610	-295 ^{ab}	-295 ^{ab}	-295 ^{ab}	-413	-415	-295 ^{ab}	-295 ^{ab}	-295 ^{ab}	-403	-295 ^a
m_{IU}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^{ab}	205 ^a
m_{UL}	-349	-240	-308	-429	-429	-393	-541	-535	-452	-766	-417	-391	-495 ^a
m_{LI}	104	42	193	63	308	115	163	148	142	44	108	44	47 ^a
m_{LM}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^{ab}	-270 ^a
m_{ML}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^{ab}	270 ^a

^aValues are taken from ref.²⁷^bValues are fixed during fitting.