Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR

Takanori Uzawa[†‡], Chiaki Nishimura[§], Shuji Akiyama[¶], Koichiro Ishimori[†‖], Satoshi Takahashi[†††‡], H. Jane Dyson[§], and Peter E. Wright[§‡‡]

[†]Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan; [§]Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037; [¶]PRESTO, Japan Science and Technology Agency, Saitama 332-0012, Japan; and [††]Institute for Protein Research, Osaka University, Suita 565-0871, Japan

Edited by Carl Frieden, Washington University School of Medicine, St. Louis, MO, and approved August 1, 2008 (received for review April 25, 2008)

The earliest steps in the folding of proteins are complete on an extremely rapid time scale that is difficult to access experimentally. We have used rapid-mixing quench-flow methods to extend the time resolution of folding studies on apomyoglobin and elucidate the structural and dynamic features of members of the ensemble of intermediate states that are populated on a submillisecond time scale during this process. The picture that emerges is of a continuum of rapidly interconverting states. Even after only 0.4 ms of refolding time a compact state is formed that contains major parts of the A, G, and H helices, which are sufficiently well folded to protect amides from exchange. The B, C, and E helix regions fold more slowly and fluctuate rapidly between open and closed states as they search docking sites on this core; the secondary structure in these regions becomes stabilized as the refolding time is increased from 0.4 to 6 ms. No further stabilization occurs in the A, G, H core at 6 ms of folding time. These studies begin to time-resolve a progression of compact states between the fully unfolded and native folded states and confirm the presence an ensemble of intermediates that interconvert in a hierarchical sequence as the protein searches conformational space on its folding trajectory.

protein folding | pulse labeling | rapid mixing

Most proteins fold rapidly from the highly heterogeneous conformational ensemble of the unfolded state into their well defined native conformations. For proteins with >100 residues, collapsed, partially folded intermediates are formed within hundreds of microseconds after the initiation of folding (1–4). Quench-flow pulse-labeling experiments have yielded considerable information about the development of secondary structure in such intermediates, on a time scale of milliseconds (5, 6). However, little is known about the processes that occur within the dead time (≈6 ms) of the conventional quench flow apparatus, nor about the dynamic behavior of the kinetic folding intermediates. To gain insights into the early folding processes for sperm whale apomyoglobin, we have performed pulsed hydrogen-deuterium (H/D) exchange experiments with submillisecond time resolution.

Apomyoglobin has been studied extensively by kinetic and equilibrium methods as a paradigm for understanding protein folding pathways and the structure of folding intermediates (7). The structure of apomyoglobin is similar to that of the holoprotein except that residues in the F helix and the C terminus of the H helix are disordered (8–10). During refolding, apomyoglobin forms an on pathway kinetic intermediate, in which major portions of the A, G, and H helices and part of the B helix are folded, within the 6-ms burst phase of conventional quench-flow H/D exchange experiments (11–14). These same regions adopt stable secondary structure in the equilibrium molten globule intermediate formed at pH 4.2 (10, 15–17). Recent H/D exchange experiments under a variety of conditions detected heterogeneity in the apomyoglobin kinetic intermediate, with the E helix partly folded in a subset of molecules in the conformational ensemble (12).

To obtain further information on the fluctuations and heterogeneity of the apomyoglobin intermediate, we adopted a rapid-mixing exchange protocol for determination of the elementary rate constants of H/D exchange (Scheme 1) (18–20):

$$\mathrm{NH_{closed}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{NH_{open}} \xrightarrow{k_{ex}} \mathrm{ND_{exchanged}} \qquad \textbf{[Scheme 1]}$$

where NH_{closed} is a folded state in which amide protons are protected from exchange, and NH_{open} is an open state from which exchange occurs. Data were obtained by using a three-solution mixer that can be applied for pulsed H/D exchange experiments in the submillisecond time domain (21, 22). In contrast to previous pulse-labeling experiments (11–13), in which amides were exchanged by a single high pH pulse in the EX1 regime ($k_{ex} \gg k_{cl}$), our protocol uses a wide range of labeling pulse strengths in both the EX1 and EX2 regimes to measure the pH dependence of the proton occupancy (20). Because the pH dependence of k_{ex} is known (23), the proton occupancy curves can be analyzed to give the values of k_{op} and k_{cl} without assuming EX1 or EX2 exchange mechanisms. The fluctuations of the residues in the kinetic intermediate can be assessed from the rate constants, whose ratio (k_{cl}/k_{op}) corresponds to the stability of the closed conformation. Furthermore, accurate values of [NH_{open}] at the moment of applying the pulse can be estimated. By estimating [NH_{open}] as a function of refolding time, the evolution of secondary structure along the apomyoglobin folding pathway can be deduced. Fractional values of [NH_{open}] indicate heterogeneity in the intermediate. The application of this method to refolding of sperm whale apomyoglobin shows that the initial collapse event leads to formation of a folding core comprised of parts of the A, G, and H helices, with

Author contributions: T.U., C.N., K.I., S.T., and P.E.W. designed research; T.U. and C.N. performed research; T.U., S.A., K.I., and S.T. contributed new reagents/analytic tools; T.U., C.N., S.T., H.J.D., and P.E.W. analyzed data; and T.U., S.T., H.J.D., and P.E.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[‡]Present address: Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106.

[‖]Present address: Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

[‡‡]To whom correspondence may be addressed. E-mail: st@protein.osaka-u.ac.jp or wright@scripps.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0804033105/DCSupplemental.

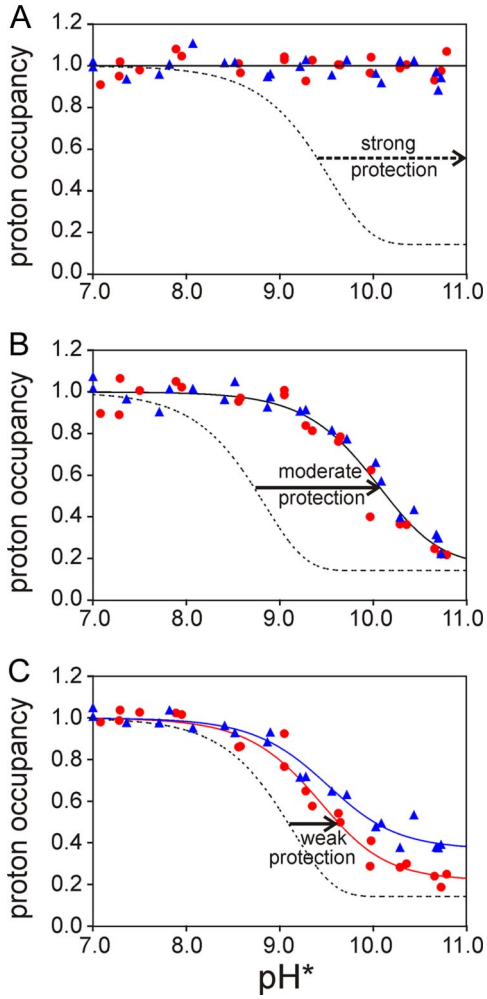

Fig. 1. Representative plots of amide proton occupancy in partially refolded apomyoglobin as the 3.6-ms labeling pulse pH is varied from 7.0 to 10.7. (*A*) V10 (A helix), representative of the highly protected residues (class S). (*B*) A143 (H helix), representative of moderately protected residues (class M). (*C*) V66 (E helix), representative of weakly protected residues (class W). Red circles and blue triangles represent the data obtained at t_f of 0.4 and 6 ms, respectively. The red lines and blue lines represent the fits to the 0.4- and 6-ms data, respectively, obtained by using Eqs. **2**-**5**. The red and blue lines are coincident in *A* and *B*. The dotted lines show the expected curves for unprotected amides at t_p of 3.6 ms.

fluctuating secondary structure in the B and E helix regions, which becomes stabilized on a millisecond time scale as these helices dock to the core domain.

Results

Time-Dependent Changes in the Proton Occupancy Profiles. To obtain information on fluctuations and heterogeneity in the folding intermediate of apomyoglobin, we measured proton occupancies (H_{obs}; Fig. 1) at refolding times of 6 and 0.4 ms and pulse strengths ranging from pH 7 to 10.7, all with duration of 3.6 ms. The resulting exchange profiles can be classified in three groups, exemplified by the data shown in Fig. 1 (a complete set of plots for all residues is included in supporting information (SI) Fig. S1). The amide protons of residues in the first group (termed group S) are strongly protected and cannot be exchanged within the 3.6-ms exchange period even by the strongest labeling pulse (Fig. 1*A*). Many residues in helices A and G belong to this group. Amide protons in group M are moderately protected from exchange; however, a significant decrease in H_{obs} is observed at high pH (Fig. 1*B*). Assuming the EX2 scheme, in which exchange occurs at the rate of k_{ex}·k_{cl}/k_{op}, the magnitude of the shift of the sigmoid H_{obs} curve to higher pH compared to the curve for the free amide (k_{ex}; Fig. 1*B*, dashed line) gives a measure of the stability (k_{cl}/k_{op}) of the amide. The amide protons of many residues located in helices B and H belong to this group. The third group (W) shows only a small shift of the sigmoid H_{obs} curve to alkaline pH (Fig. 1*C*), indicating weak protection. Even at the highest pulse strength, where the mechanism is EX1 and exchange is pH independent and limited by k_{op} (20), the proton occupancy after 6 ms of refolding time does not reach its limiting value of 0.14 (determined by the proportion of residual H_2O in the refolding buffer). The amide protons in the C and E helices belong to group W.

Analysis of HD Exchange Data. Amide protons in group S exhibit very similar exchange profiles at folding times of 6 and 0.4 ms, indicating that they are already fully protected within 0.4 ms of initiating refolding (Fig. 1*A*). With the sole exception of I30 (B helix), all of these amides belong to residues in the A and G helices. In contrast, the 0.4- and 6-ms refolding curves are not coincident for many of the amide protons belonging to groups M and W, an indication that the local folding process that results in the protection of these amides is continuing on a time scale of 0.4–6 ms. For each residue, the midpoint of the sigmoid transition is at nearly the same pH for the two refolding curves but there is a significant decrease in the plateau value of the proton occupancy at high pulse pH for the data at 0.4 ms relative to that at 6-ms refolding time (Fig. 1*C* and Fig. S1). This decrease can be attributed to an increased population of the NH_{open} state at 0.4-ms refolding time, rather than to differences in k_{cl} and k_{op} values, which would shift the pH midpoint of the sigmoid curve (20). In other words, many amide protons remain largely in the open conformation in the ensemble of compact states formed after 0.4 ms, and additional compaction or folding events reduce the fraction of NH_{open} after 6 ms of refolding. Amide protons located near the ends of the B helix and in the C and E helices exhibit this behavior, showing that these regions have not yet relaxed into stable hydrogen-bonded conformations in the compact states formed 0.4 ms after initiation of refolding.

For each residue, the pH-dependent proton occupancies at 0.4- and 6-ms refolding times were fitted simultaneously to the Hvidt equations (18,19) (Eqs. **2** and **3** and *Materials and Methods*), assuming that the values of k_{op} and k_{cl} are the same at both refolding times. These equations were derived without assuming either EX1 or EX2 regimes (18). Eq. **4** was used to describe the time-dependent changes in [NH_{open}], assuming that all amide protons are in the open form at the moment of initiation of the refolding reaction. Examples of fitted curves are shown in Fig. 1 as solid lines. All of the exchange profiles can be fitted by the above equations, with the exception of amides in group S, which show only limited dependence on pH. The values of the stability (k_{cl}/k_{op}) and [NH_{open}] obtained from the data fitting are shown as a function of residue number in Fig. 2 and are listed in Table S1. For amides in group S, only lower limits can be placed on values of k_{cl}/k_{op} (>100) and k_{cl} (more than ≈40,000 s^{-1}) (see *SI Text*). Moderately protected amides, with k_{cl}/k_{op} between 20 and 100, are found toward the C terminus of the B helix and at the N-termini of the G and H helices (Fig. 2*A*). Weakly protected amides were identified not only in the C and E helices and CD loop, but also in the terminal regions of the helices that were otherwise highly or moderately protected. The relative stability of each amide in the kinetic intermediate ensemble provides a clue to the nature of the core structure and its mechanism of formation.

The [NH_{open}] values in different regions of the protein (Fig. 2*B*) give additional insights that can be interpreted in a time-dependent way. Amide protons in the center of the A, G, and H

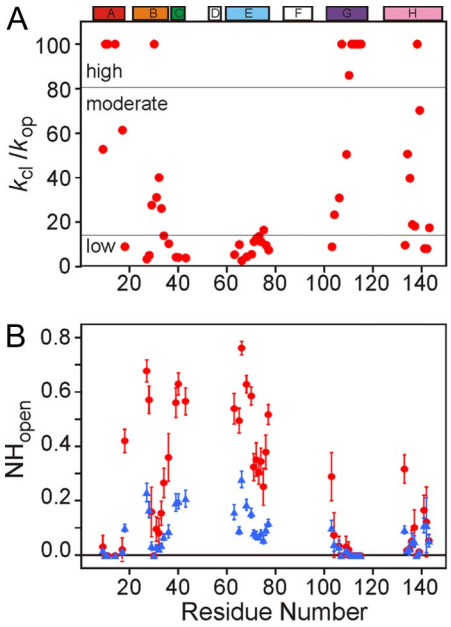

Fig. 2. Plots of stability and [NH_{open}] as a function of residue number. (*A*) The estimated stability (k_{cl}/k_{op}) for each amide. For highly protected amides with pH-independent proton occupancies, k_{cl}/k_{op} values are assigned a lower limit of 100. (*B*) [NH_{open}] values, estimated from k_{cl}, k_{op}, and t_f. The red circles and blue triangles represent [NH_{open}] at t_f of 0.4 and 6 ms, respectively. [NH_{open}] for highly protected amides is assigned a limiting value of 0. The location of the helices in the holomyoglobin structure is indicated at the top.

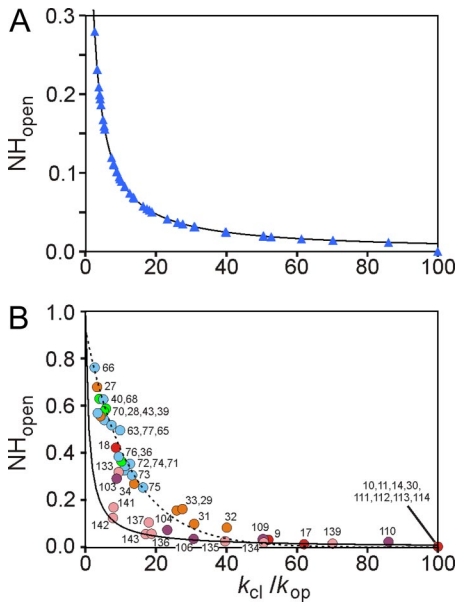

Fig. 3. Equilibration of amide proton exchange at 0.4 and 6 ms. (*A*) [NH_{open}] versus k_{cl}/k_{op} for a refolding time of 6 ms. Experimental data points are shown as blue triangles. The solid line depicts equilibrium values of [NH_{open}] at long refolding time, calculated by using the equation $[NH_{open}] = k_{op}/(k_{op} + k_{cl})$ derived from Eq. **4**. (*B*) Plot of [NH_{open}] versus k_{cl}/k_{op} for a refolding time of 0.4 ms. Data points are colored according to their location in the apomyoglobin helices, as depicted in Fig. 2. The highly protected amides are assigned values of $k_{cl}/k_{op} = 100$ and $[NH_{open}] = 0$. The solid line depicts the equilibrium values of [NH_{open}], and the broken line was calculated by using Eq. **4** with $k_{op} = 200$ s^{-1} and $t_f = 0.4$ ms.

helices and at I30 in the B helix are strongly protected and have extremely small [NH_{open}] values at both 0.4- and 6-ms refolding times; no decrease in [NH_{open}] is observed at the longer folding time. In contrast, most amides in the B, C, and E helices possess much larger values of [NH_{open}] at 0.4 ms, which decrease significantly after 6 ms of refolding time. The fractional [NH_{open}] values at 0.4 ms indicate that these amides are involved in fluctuating and heterogeneous conformations in the intermediate ensemble formed after 0.4 ms of refolding. Stabilization of secondary structure and reduction of heterogeneity occurs in these regions as folding progresses in the time range between 0.4 and 6 ms. The amides of several terminal residues of the helices of the AGH folding core (E18 in A helix, Y103 in G helix, and K133 in H helix) also display time-dependent [NH_{open}] values, reflecting end-fraying of the helices.

The difference in behavior of the various amides at 0.4- and 6-ms refolding time is emphasized by Fig. 3. For each amide, [NH_{open}], which is 1.0 in the unfolded state, starts to relax after the initiation of the refolding reaction to the equilibrium value defined by the respective k_{op} and k_{cl} rates. Limiting equilibrium values of [NH_{open}] can be calculated by using Eq. **4** for infinitely long refolding time, where the second term of the equation becomes zero and [NH_{open}] reduces to $k_{op}/(k_{op} + k_{cl})$. Values of [NH_{open}] obtained by fitting the experimental data at 6-ms refolding time (Fig. 3, blue triangles) coincide with the calculated equilibrium values of [NH_{open}] (Fig. 3, solid curve), showing that all amides have achieved equilibrium between the open and closed states at this time (Fig. 3*A*). In contrast, at 0.4-ms refolding time, only the amides located in the helix AGH folding core (plus I30) have reached their equilibrium values of [NH_{open}] (Fig. 3*B*). For amides in the B, C, and E helix regions, [NH_{open}] values are much larger than the limiting equilibrium values, again indicating that these regions of the protein remain in a fluctuating and heterogeneous state in the folding intermediate ensemble formed after 0.4 ms. For amides in the B, C, and E helices, the k_{op} values are all $\approx$200 s^{-1} (Table S1): their experimental [NH_{open}] values can be fitted by Eq. **4** by using $k_{op} = 200$ s^{-1} and $t_f = 0.4$ ms (broken line in Fig. 3*B*).

Discussion

Formation of the Folding Core Is Associated with Initial Hydrophobic Collapse. Conventional stopped-flow and quench-flow hydrogen exchange experiments on refolding of apomyoglobin at 5°C revealed formation of a helical intermediate within the 6-ms dead time of the instrument (11, 12). The present studies extend kinetic pulse-labeling experiments on apomyoglobin refolding into the submillisecond time regime and provide insights into the early folding events. In these experiments, most of the observable amide protons in the A, G, and H helix regions are significantly protected from exchange even at the shortest (0.4 ms) refolding time (Figs. 2*A* and 3*B*). Many amides in the A and G helix regions show little tendency to exchange even with the highest pH labeling pulse. Some amides in the middle of G (residues 106, 109, and 110) and those in the middle of H (residues 134–137 and 139) are more labile but still appear to have reached their maximal protection in the intermediate formed after 0.4 ms. These data show that the AGH core is fully formed after 0.4 ms of refolding time, although not all amides are equally well protected from exchange. With the exception of residue I30, and to a lesser extent R31 and L32, which have small values of [NH_{open}] (<0.1), amides in the B, C, and E helices are not strongly protected in the 0.4-ms intermediate, suggesting that stable helical structure has not yet formed.

Although stable helical structure is a prerequisite for the observation of amide proton protection during folding, it is not the major driving force in the folding process. Studies of isolated peptide fragments of apomyoglobin indicate that the H helix possesses the highest propensity for spontaneous formation of helical structure in aqueous solution, whereas the fragments corresponding to the G helix possess poor helix propensity in

isolation (24, 25). Both the G and A helices show extensive helix formation in longer peptides where interhelix contacts are possible (26). These data clearly show the importance of tertiary interactions for folding of the A and G helices. The A, B, G, and H helix regions are also the sites of transient long-range contacts formed in the acid-unfolded state of apomyoglobin at equilibrium, which have been directly related to the local hydrophobic nature of the sequence (27–29). Because hydrophobic collapse of the apomyoglobin polypeptide chain and formation of the first tertiary contacts occurs within 300 μs (1, 4, 30), our observation that the A, G, and H helices contain the most highly protected amides at 0.4 ms is consistent with a model in which collapse and coalescence of local hydrophobic clusters in these regions drives formation and stabilization of helical secondary structure. A strong propensity for spontaneous helix formation is not a prerequisite for efficient folding, as shown clearly by previous folding studies of H helix-destabilizing apomyoglobin mutations (31).

It is notable that, although amides in the middle of the H helix are well protected after 0.4 ms and show no decrease in [NH_{open}] values after 6 ms of refolding time, they are in general less well protected than amides in the A and G helices (Fig. 2). This finding contrasts with the native apoprotein and the pH 4 equilibrium molten globule intermediate, where the protection factors for exchange of amides in the H helix are comparable to those of the A and G helices (10).

Docking of B and E Helices to the Folding Core. Five amides in the B helix (residues 29, 30, 31, 32, and 33) have relatively small [NH_{open}] (<0.16) and are moderately protected at 0.4-ms refolding time. These residues contact the C/D region, the N-terminal region of the E helix, and the middle of the G helix in the native state. Of these regions, only the G helix contains amides that show high protection and low [NH_{open}], which implies that the primary contact for the B helix in the intermediate ensemble is to helix G. These observations are consistent with previous studies of the L32A mutant (32), in which the substitution caused a reduction in the observed proton occupancies at G-helix residues I107 and E109, which contact L32 in the folded state. We therefore suggest that, at 0.4-ms refolding time, the region between residues 29–33 of the B helix is sampling contacts with the fully structured G helix; the interactions with G and the helical structure of B become further stabilized as folding proceeds in the time domain from 0.4 to 6 ms, as evidenced by the decrease in [NH_{open}] observed for the B-helix residues over this time period.

The E helix can be divided into two regions that demonstrate different docking dynamics to the AGH core domain. At the C terminus of the E helix, residues 71–76 possess moderate NH_{open} at 0.4 ms and low protection (average $k_{cl}/k_{op} = 12 \pm 2$), whereas at the N terminus, residues 63–70 possess high NH_{open} and marginal protection (average $k_{cl}/k_{op} = 5 \pm 3$). The C-terminal half of the E helix is located at the junction of the A, G, and H helices in the folded protein (Fig. 4), providing a possible reason for its increased stability in the intermediate state. In addition, the marginal stability of the N-terminal half of the E helix, which contacts the B helix in the native structure, suggests inefficient packing between the E and B helices in the intermediate. These results imply that the docking of the B and E helices to the core occurs independently and that interactions between the B and E helices are not established during this time period. It follows that the formation of the correct contacts between the different helices that dock to the core might be an important kinetic bottleneck in the folding of apomyoglobin.

Cooperative Dynamics in the Intermediate Ensemble. Analysis of k_{op} and k_{cl} and the time-dependent changes in [NH_{open}] yields profound insights into the structural heterogeneity and dynamic cooperativity of the intermediate ensemble. Our model assumes

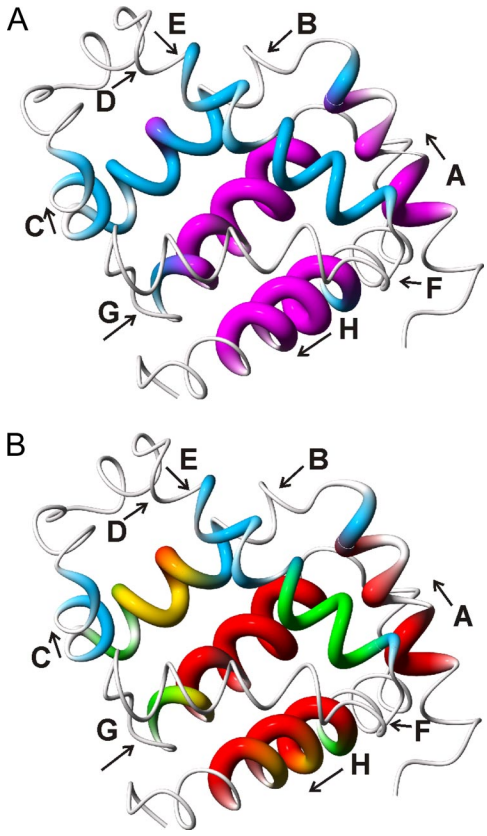

Fig. 4. Location of the most rapidly protected amides. (*A*) Distribution of fully and nonequilibrated amides in the 0.4-ms intermediate ensemble. The structure of holomyoglobin (38) is represented as a tube of varying radius. Residues whose amides are fully equilibrated after 0.4 ms of refolding, i.e., for which [NH_{open}] has reached the equilibrium value, are depicted in magenta with increased tube radius. Residues for which [NH_{open}] has not reached the equilibrium values and which are fluctuating between folded and unfolded states are depicted by a thinner blue tube. Very thin tubes represent regions that do not exhibit exchange protection and probably remain unstructured in the folding intermediate. (*B*) Mapping of k_{cl} values onto the holomyoglobin structure. Closing rates are depicted by variations in color and tube radius. Red, $k_{cl} > 9{,}500\ s^{-1}$; yellow, $5{,}000 < k_{cl} < 9{,}500\ s^{-1}$; green, $2{,}500 < k_{cl} < 5{,}000\ s^{-1}$; blue, $k_{cl} < 2{,}500\ s^{-1}$. This figure was prepared by using MolMol (39).

that the open-closed dynamics of different amides can be described by independent k_{op} and k_{cl} values. Exchange between open and closed states is fast enough ($k_{op} + k_{cl} > 1{,}000\ s^{-1}$) that the exchange process in the intermediate ensemble reaches equilibrium within the 6-ms refolding time, i.e., [NH_{open}] is equal to $k_{op}/(k_{op} + k_{cl})$ (Eq. **4**). That the system has equilibrated within this refolding time is confirmed by Fig. 3*A*, which shows that the fitted values of [NH_{open}] at 6 ms are coincident with the calculated equilibrium values, i.e., all protected amides in the intermediate ensemble have reached equilibrium after 6 ms.

In contrast to the data at 6 ms, the plot of [NH_{open}] versus k_{cl}/k_{op} for the data at 0.4 ms (Fig. 3*B*) shows that the amides fall into two groups. First, the amides in the AGH folding core are fully equilibrated even after 0.4 ms of refolding, and their [NH_{open}] values are identical to those of the equilibrated state at 6 ms. Amides in these regions are highly protected from exchange and k_{cl} values are large, i.e., folding is very rapid. Second, the amides in the B and E helices possess [NH_{open}] values that are significantly shifted from the equilibrated values; for these amides, $k_{op} + k_{cl}$ is not fast enough for equilibrium to be reached within the 0.4-ms refolding time. The B-helix amides cluster around k_{cl}/k_{op} of $\approx$20–40, and the amides in the N-terminal and

C-terminal halves of the E helix cluster around values of 5 and 10, respectively. The distribution of [NH_{open}] values for amides in these helices can be fitted reasonably well by Eq. **4** by using k_{op} of 200 s^{-1} and t_f of 0.4 ms (Fig. 3*B*). Amides in the C helix also belong to this group. The location of the equilibrated and nonequilibrated amides mapped onto the holomyoglobin structure is shown in Fig. 4*A*.

With the exception of the H helix, k_{op} values lie in a fairly narrow range (≈200–400 s^{-1}). The fact that k_{op} values are similar for most residues implies that H/D exchange occurs through a cooperative, global opening process that involves unfolding of the intermediate. Opening rates in the H helix, especially toward the C terminus, are significantly faster and suggest fraying of the helix. In contrast, considerable variation is seen in closing rates (k_{cl}), which provides important insights into the folding process. For residues in the G and H helices that display pH-dependent proton occupancies, folding is very fast and values of k_{cl} are mostly >10,000 s^{-1}. Although it is not possible to fit kinetic parameters for highly protected residues in helices A and G whose proton occupancies show no pH dependence, a lower limit of ≈40,000 s^{-1} can be placed on the k_{cl} rates. Closing rates near the ends of the A, G, and H helices are significantly slower and reflect the decreased stability of the helix terminals in the conformational ensemble of the intermediate.

The average value of k_{cl} in the central region of the B helix is ≈5,000 s^{-1}, suggesting that this region folds more slowly than the AGH core. The similar values of k_{cl} and k_{op} for these residues argue that both folding and unfolding of the central region of the B helix is cooperative. The N- and C-terminal halves of the E helix have the same k_{op} rates but display significantly different values of k_{cl}; for residues 63–70, the average k_{cl} is ≈1,400 s^{-1}, whereas the average k_{cl} for residues 71–76 is ≈3,000 s^{-1}. These results suggest that the E helix comprises two independent subdomains, and that folding and unfolding of each subdomain occurs cooperatively. It is also of note that k_{cl} values at the N terminus of the B helix (residues D27 and I28) are the same as the N-terminal residues of helix E, suggesting that these two regions, which are in contact in the native myoglobin structure, fold at the same rate. The k_{cl} values are mapped to the myoglobin structure in Fig. 4*B*.

Apomyoglobin Folding Mechanism. Based on the measured kinetic parameters, we propose a hierarchical model for folding of apomyoglobin. The A, G, and H helix regions undergo rapid collapse and fast formation of helical structure (red regions in Fig. 4*B*). The B, C, and E helices then dock onto the AGH core domain and fold in the following order: B helix (Fig. 4*B*, yellow), C-terminal subdomain of the E helix (Fig. 4*B*, green), and N-terminal subdomain of the E helix and C helix (Fig. 4*B*, blue). This interpretation is consistent with previous apomyoglobin folding schemes involving two sequential on pathway intermediates separated by an energy barrier (1, 33, 34). Jamin and Baldwin (33) identified two intermediates, designated I_a and I_b, that exist in equilibrium in the pH 4.2 molten globule state of sperm whale apomyoglobin and are populated transiently during kinetic refolding ($U \leftrightarrow I_a \leftrightarrow I_b \leftrightarrow N$). Fast kinetic refolding experiments with horse apomyoglobin revealed formation of an initial collapsed intermediate, I_1, within 0.3 ms, which converts to a second intermediate, I_2, with a time constant of 5 ms (1). It appears that I_1 and I_2 correspond to the I_a and I_b intermediates, respectively, of the sperm whale protein (1, 34). The two intermediates are equally compact (R_g, 23.7 Å) but differ in helical content (30% and 44% helix for I_1 and I_2, respectively). The initial intermediate, I_a/I_1, likely corresponds to the ensemble of heterogeneous conformations observed in the present work after 0.4 ms of refolding, containing helical structure in the A, G, and H regions. The second intermediate, I_b/I_2, would then correspond to the equilibrated state formed after 6 ms, in which additional helical structure has become stabilized in the B-, C-, and E-helix regions. The increase in helicity between I_1 and I_2 revealed by CD (1) is quite consistent with the present pulse-labeling experiments, which show increased exchange protection between 0.4 and 6 ms for regions that encompass 17–18 residues in the B, C, and E helices. Although the present data suggest that different domains of the B and E helices fold at different rates, these processes would be difficult to resolve by the low-resolution CD and fluorescence methods used in the previous kinetic measurements, which detected only the complete folding of both the B and E helices as a single kinetic event. Further studies on the opening and closing dynamics of the individual amides in helices B and H and on the heterogeneity of the intermediate ensemble are required to understand the kinetic cooperativity and heterogeneity in apomyoglobin folding.

Materials and Methods

Rapid Solution Mixer. For the pulse-labeling experiments, a series of three consecutive timed mixing steps is required: initiation of the refolding reaction, H/D exchange, and quench. A schematic drawing of the equipment, modified from the previous versions (2, 21, 22), is shown in Fig. S2. Usage and calibration of the mixer are described in *SI Text*.

H/D Exchange, pH Pulse Labeling for Apomyoglobin. ^{15}N-labeled sperm whale apomyoglobin was produced in *Escherichia coli* BL21-DE3 cells and purified as described (35). A solution containing the fully unfolded apomyoglobin (≈80 μM) at pH 2.2 in HCl/H_2O was placed in syringe P (Fig. S2). Refolding and H/D exchange was performed at room temperature. The refolding reaction was initiated by a 6-fold dilution of the acid-unfolded apomyoglobin with either 60 mM acetate buffer or 30 mM citrate buffer in ^2H_2O delivered from syringe Q, to a final pH* (measured pH, uncorrected for the deuterium isotope effect) of 5.8. After refolding times (t_f) of 0.4 or 6 ms, the H/D exchange reaction was initiated by a 1/6-fold dilution of the sample solution with ^2H_2O buffer containing 350-mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) delivered from syringe R at the desired pH* from 7 to 10.7 to enhance exchange of unprotected amide protons to deuterons. The H/D exchange labeling was allowed to proceed for 3.6 ms (t_p), and the pH* was then rapidly lowered to 5.7 by mixing with 250 mM citrate buffer in ^2H_2O, precooled on ice. To avoid further undesired H/D exchange, the samples were kept on ice or at 4°C after the final mixing.

Because holomyoglobin is more stable to exchange than apomyoglobin, heme was added to the exchange-labeled apomyoglobin and the CO complex was formed by a modification of previous procedures (12). The pulse-labeled samples were mixed with a 6-fold molar excess of hemin (Sigma) in a ^2H_2O buffer containing 250 mM Hepes and 0.1-mg/ml KCN adjusted to pH* 8.2. After a 5-s delay to allow heme insertion and isomerization, the pH* was decreased to 6.0 by addition of 700 mM citrate buffer in ^2H_2O. The holomyoglobin was concentrated by using a centrifugal filter (Amicon Ultra-15; Millipore) and the buffer was exchanged to 50 mM phosphate in ^2H_2O at pH* 5.6. The holomyoglobin was reduced by adding sodium dithionite to a final concentration of 1.3 mg/ml and purged with carbon monoxide to prepare the CO-bound complex. Excess dithionite was removed under a CO atmosphere by using a Sephadex G15 column equilibrated with 50 mM phosphate buffer in ^2H_2O at pH* 5.6. The eluate was collected directly in an NMR tube purged with CO.

NMR Spectroscopy. ^{15}N-^1H heteronuclear single quantum coherence (HSQC) spectra were recorded by using a Bruker DRX600 spectrometer at a probe temperature of 35°C. The spectral widths were adjusted to 9,615 Hz and 2,000 points in the ^1H dimension and 1,538 Hz and 128 points in the ^{15}N dimension. Proton occupancies were determined from HSQC cross-peak intensities (see *SI Text*).

Data Analysis. Amide hydrogen exchange was interpreted based on Scheme 1, in which a given amide can be in a protected state that is nonexchangeable (NH_{closed}), an open state from which exchange occurs (NH_{open}), or in the fully exchanged state ($ND_{exchanged}$) (18, 19, 36). k_{op} and k_{cl} are the local opening and local closing rates, and k_{ex} is the exchange rate of the unprotected amide. The latter depends on a variety of conditions and varies with pH as described in Eq. **1** (23, 37):

$$k_{ex} = k_{ex}^{pH7} \cdot 10^{(pH-7)}, \qquad [1]$$

BIOPHYSICS

where k_{ex}^{pH7} is the intrinsic exchange rate of the unprotected amide at pH 7. Because k_{ex} increases with increasing pH, the fraction of protonated amide (f_H) varies sigmoidally with labeling pulse pH, as expressed by Eq. **2** (20):

$$f_H = \left(\frac{k_{ex}[NH_{open}]_{t_f} - \lambda_2}{\lambda_1 - \lambda_2}\right) e^{-\lambda_1 t_p} + \left(\frac{\lambda_1 - k_{ex}[NH_{open}]_{t_f}}{\lambda_1 - \lambda_2}\right) e^{-\lambda_2 t_p} \quad [2]$$

where

$$\lambda_{1,2} = \frac{k_{op} + k_{cl} + k_{ex} \pm \sqrt{(k_{op} + k_{cl} + k_{ex})^2 - 4k_{op}k_{ex}}}{2} \quad [3]$$

$$[NH_{open}]_{t_f} = \frac{k_{op}}{k_{op} + k_{cl}} + \frac{k_{cl}}{k_{op} + k_{cl}} e^{-(k_{op}+k_{cl})t_f} \quad [4]$$

$[NH_{open}]_{t_f}$ is the population of the open state at refolding time t_f. In practice, we observe a proton occupancy (H_{obs}), which is the sum of f_H and the remaining fraction of H_2O after the second mixing (= 1/7 for the 7-fold dilution with refolding buffer and pulse labeling buffer in our experimental protocol).

$$H_{obs} = f_H + 1/7. \quad [5]$$

The H_{obs} values as a function of pH were fitted to the above equations by using the program IGOR Pro (Wavemetrics).

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant DK34909 (to P.E.W.). T.U. was supported by a fellowship from the Japan Society for the Promotion of Science to Young Scientists, and S.A. was supported by PRESTO, the Japan Science and Technology Agency.

1. Uzawa T, *et al.* (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. *Proc Natl Acad Sci USA* 101:1171–1176.
2. Uzawa T, *et al.* (2006) Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: Implication of the scaling relationship for the submillisecond intermediates of protein folding. *J Mol Biol* 357:997–1008.
3. Arai M, *et al.* (2007) Microsecond hydrophobic collapse in the folding of *Escherichia coli* dihydrofolate reductase, an α/β-type protein. *J Mol Biol* 368:219–229.
4. Lapidus LJ, *et al.* (2007) Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. *Biophys J* 93:218–224.
5. Englander SW (2000) Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* 29:213–238.
6. Bai Y (2006) Protein folding pathways studied by pulsed- and native-state hydrogen exchange. *Chem Rev* 106:1757–1768.
7. Wright PE, Baldwin RL (2000) in *Mechanisms of Protein Folding*, ed Pain RH (Oxford Univ Press, Oxford), pp 309–329.
8. Eliezer D, Wright PE (1996) Is apomyoglobin a molten globule? Structural characterization by NMR. *J Mol Biol* 263:531–538.
9. Lecomte JT, Sukits SF, Bhattacharjya S, Falzone CJ (1999) Conformational properties of native sperm whale apomyoglobin in solution. *Protein Sci* 8:1484–1491.
10. Nishimura C, Dyson HJ, Wright PE (2005) Enhanced picture of protein-folding intermediates by using organic solvents in H/D exchange and quench-flow experiments. *Proc Natl Acad Sci USA* 102:4765–4770.
11. Jennings PA, Wright PE (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262:892–896.
12. Nishimura C, Dyson HJ, Wright PE (2002) The apomyoglobin folding pathway revisited: Structural heterogeneity in the kinetic burst phase intermediate. *J Mol Biol* 322:483–489.
13. Nishimura C, Dyson HJ, Wright PE (2006) Identification of native and nonnative structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol* 355:139–156.
14. Tsui V, *et al.* (1999) Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci* 8:45–49.
15. Hughson FM, Wright PE, Baldwin RL (1990) Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–1548.
16. Eliezer D, Chung J, Dyson HJ, Wright PE (2000) Native and nonnative structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* 39:2894–2901.
17. Eliezer D, Yao J, Dyson HJ, Wright PE (1998) Structural and dynamic characterization of partially folded states of myoglobin and implications for protein folding. *Nat Struct Biol* 5:148–155.
18. Hvidt A (1964) A discussion of the pH dependence of the hydrogen-deuterium exchange of proteins. *C R Trav Lab Carlsberg* 34:299–317.
19. Hvidt A, Nielsen SO (1966) Hydrogen exchange in proteins. *Adv Protein Chem* 21:287–386.
20. Krishna MM, Hoang L, Lin Y, Englander SW (2004) Hydrogen exchange methods to study protein folding. *Methods* 34:51–64.
21. Akiyama S, Takahashi S, Ishimori K, Morishima I (2000) Stepwise formation of α-helices during cytochrome *c* folding. *Nat Struct Biol* 7:514–520.
22. Takahashi S, Yeh SR, Das TK, Gottfried DS, Rousseau DL (1997) Folding of cytochrome *c* initiated by submillisecond mixing. *Nat Struct Biol* 4:44–50.
23. Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17:75–86.
24. Shin H-C, *et al.* (1993) Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry* 32:6356–6364.
25. Waltho JP, Feher VA, Merutka G, Dyson HJ, Wright PE (1993) Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G and H helices of myoglobin. *Biochemistry* 32:6337–6347.
26. Reymond MT, Merutka G, Dyson HJ, Wright PE (1997) Folding propensities of peptide fragments of myoglobin. *Protein Sci* 6:706–716.
27. Lietzow MA, Jamin M, Dyson HJ, Wright PE (2002) Mapping long-range contacts in a highly unfolded protein. *J Mol Biol* 322:655–662.
28. Yao J, Chung J, Eliezer D, Wright PE, Dyson HJ (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* 40:3561–3571.
29. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA* 105:6278–6283.
30. Nishiguchi S, Goto Y, Takahashi S (2007) Solvation and desolvation dynamics in apomyoglobin folding monitored by time-resolved infrared spectroscopy. *J Mol Biol* 373:491–502.
31. Cavagnero S, Dyson HJ, Wright PE (1999) Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J Mol Biol* 285:269–282.
32. Nishimura C, Wright PE, Dyson HJ (2003) 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* 334:293–307.
33. Jamin M, Baldwin RL (1998) Two forms of the pH 4 folding intermediate of apomyoglobin. *J Mol Biol* 276:491–504.
34. Weisbuch S, *et al.* (2005) Cooperative sub-millisecond folding kinetics of apomyoglobin pH 4 intermediate. *Biochemistry* 44:7013–7023.
35. Jennings PA, Stone MJ, Wright PE (1995) Overexpression of myoglobin and assignment of the amide, Cα and Cβ resonances. *J Biomol NMR* 6:271–276.
36. Linderström-Lang K (1955) Deuterium exchange between peptides and water. *Chem Soc Spec Publ* 2:1–20.
37. Connelly GP, Bai Y, Jeng M-F, Englander SW (1993) Isotope effects in peptide group hydrogen exchange. *Proteins* 17:87–92.
38. Kuriyan J, Wilz S, Karplus M, Petsko GA (1986) X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5-Å resolution. *J Mol Biol* 192:133–154.
39. Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics* 14:51–55.