Core Formation in Apomyoglobin: Probing the Upper Reaches of the Folding Energy Landscape†

Miriam Gulotta,‡ Rudolf Gilmanshin,‡,§ Thomas C. Buscher,|| Robert H. Callender,*,‡ and R. Brian Dyer*,||

Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, and Bioscience Division, Mail Stop J586, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received September 25, 2000; Revised Manuscript Received February 9, 2001

ABSTRACT: An acid-destabilized form of apomyoglobin, the so-called E state, consists of a set of heterogeneous structures that are all characterized by a stable hydrophobic core composed of 30−40 residues at the intersection of the A, G, and H helices of the protein, with little other secondary structure and no other tertiary structure. Relaxation kinetics studies were carried out to characterize the dynamics of core melting and formation in this protein. The unfolding and/or refolding response is induced by a laser-induced temperature jump between the folded and unfolded forms of E, and structural changes are monitored using the infrared amide I′ absorbance at 1648−1651 cm^{-1} that reports on the formation of solvent-protected, native-like helix in the core and by fluorescence emission changes from apomyoglobin's Trp14, a measure of burial of the indole group of this residue. The fluorescence kinetics data are monoexponential with a relaxation time of 14 μs. However, infrared kinetics data are best fit to a biexponential function with relaxation times of 14 and 59 μs. These relaxation times are very fast, close to the limits placed on folding reactions by diffusion. The 14 μs relaxation time is weakly temperature dependent and thus represents a pathway that is energetically downhill. The appearance of this relaxation time in both the fluorescence and infrared measurements indicates that this folding event proceeds by a concomitant formation of compact secondary and tertiary structures. The 59 μs relaxation time is much more strongly temperature dependent and has no fluorescence counterpart, indicating an activated process with a large energy barrier wherein nonspecific hydrophobic interactions between helix A and the G and H helices cause some helix burial but Trp14 remains solvent exposed. These results are best fit by a multiple-pathway kinetic model when U collapses to form the various folded core structures of E. Thus, the results suggest very robust dynamics for core formation involving multiple folding pathways and provide significant insight into the primary processes of protein folding.

A structural feature of many proteins is the existence of a hydrophobic core contained within its folded three-dimensional shape. Some hydrophobic cores are thermodynamically very stable. Myoglobin, a case in point, is a largely helical protein consisting of eight helical runs, called A−H [see Figure 1 in the preceding paper (*1*)]. Much of myoglobin's structure is left intact when the heme group is extracted to form apomyoglobin. ApoMb's[1] core consists of a cluster of largely hydrophobic residues of the A, G, and H helices, and the two ends of the protein are tied together upon formation of this core. This structural motif is clearly important in forming the folded protein. It forms very early in the folding pathway and represents the earliest formation of tertiary contacts (*2*−*4*). Subsequently, another condensation event takes place that involves the formation of more secondary structure and another set of apparently fixed tertiary contacts involving the B and E helices (*2*, *3*, *5*, *6*).

In this paper, we probe the kinetics of core melting and formation using an acid-destabilized (pH 3, low salt) form of horse heart apomyoglobin, a state that we have called E (*1*, *4*, *6*) which may correspond to the recently proposed form for sperm whale apoMb, called I_a (*7*). The E form of apoMb contains a relatively stable distinctive core, which is composed of no more than 30−40 residues of the 153 residues of apomyoglobin at the intersection of the A helix with the G and H helices and the G−H turn (*8*−*12*). The structure and thermodynamics of this state of apomyoglobin are the subject of our previous paper (*1*). This core is compact and melts cooperatively, yet it has significant structural heterogeneity (*1*). Hence, E consists of just the core with few other secondary and no other tertiary contacts. It is a relatively simple structure, and the dynamics of the folding of E may serve as a paradigm of core formation. There is no atomic resolution structure of E; a determination of this level of structure may not be possible given its heterogeneity. The structure of myoglobin is well-resolved, and Figure 1 is a representation of myoglobin based on the coordinates of Protein Data Bank entry 1AZI (*13*). Hydrophobic residues of the AGH intersection for myoglobin are quite numerous,

† This work was supported by the National Science Foundation (Grant MCB-9727439 to R.H.C.) and the Institute of General Medicine, National Institutes of Health Grants GM53640 (R.B.D.) and GM35183 (R.H.C.).

* To whom correspondence should be addressed. Phone: (718) 430-3024. Fax: (718) 430-8565. E-mail: call@aecom.yu.edu.

‡ Albert Einstein College of Medicine.

§ Current address: U.S. Genomics, Inc., 8 Saint Mary's St., Suite 922, Boston, MA 02215.

|| Los Alamos National Laboratory.

[1] Abbreviation: apoMb, myoglobin with the prosthetic heme group extracted.

10.1021/bi002256n CCC: $20.00
Published on Web 04/03/2001

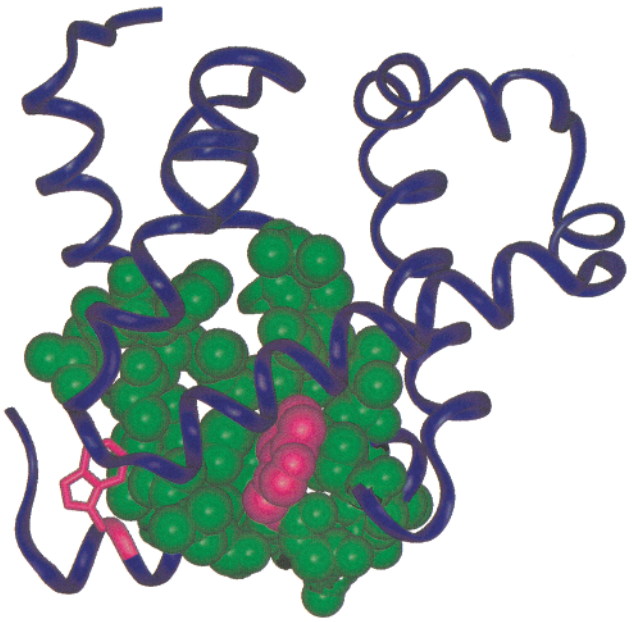

FIGURE 1: Diagram of myoglobin (heme not shown) based on Protein Data Bank entry 1AZI. The central part of the core is highlighted to illustrate that there is very little secondary structure and no tertiary structure apart from the core in the structure of E. van der Waals surfaces have been drawn for all of the hydrophobic residues in the core (shown in green). The two tryptophans, including their indol rings, are drawn in pink, as a stick diagram for Trp7 and as a van der Waals surface for Trp14. The rest of the protein is shown as a blue ribbon and is there to show how the rest of the structure forms around the core in the native protein. This figure was drawn using Insight II (Molecular Simulations, Inc.).

and Figure 1 emphasizes them by drawing the core's hydrophobic residues with van der Waals surfaces.

The formation time of E is faster than 100 μs, so the study of its dynamics requires special techniques that can characterize kinetics on such fast time scales. Our previous studies have shown that at high temperatures, E melts losing all tertiary contacts and nearly all secondary structure (*1*, *6*). Hence, E can be unfolded on fast time scales using a laser-induced temperature jump (*T*-jump) which rapidly (within 20 ns) induces a change in temperature of ~20 °C. The observed relaxation kinetics contain folding as well as unfolding rates so that the kinetics of both processes can be monitored (*14*). Using *T*-jump techniques, we have previously investigated the unfolding and folding kinetics of E by monitoring changes in the structurally sensitive amide I absorption band (*4*). Here, we report on an expanded set of measurements of the unfolding and folding kinetics of E using IR amide I′ absorbance and Trp fluorescence emission to follow secondary and tertiary structure structural changes, and over an extended temperature range. The relaxation kinetics following a *T*-jump are fast and nonexponential, and the observed rates depend on the probe technique that is employed. The results suggest very robust dynamics for core formation involving multiple folding pathways. The details of this process give significant new insight into the earliest events in the core formation of apomyoglobin.

MATERIALS AND METHODS

Materials. The preparation and purification of horse heart apoMb samples have been described in detail previously (*6*); see the preceding paper (*1*). The apoMb-E solution contained 20 mM NaCl at pH* 3.0. The NaCl concentration of the mixed solutions were calculated from the measured weights of the aliquots and their known densities. The pH* is the uncorrected (for D_2O) pH-meter reading at 20 °C. The pH* was controlled after the protein dissolution and always adjusted with minimal addition of DCl. Sample concentrations of 4−8 mg/mL were used for the IR measurements and 2 mg/mL for the fluorescence studies. Under these conditions, no aggregation was observed in the kinetics measurements (*4*, *6*).

Temperature-Jump Spectrometers. The essence of the kinetics spectrometer is the same for the IR and fluorescence temperature-jump instruments. A continuous laser beam, whose wavelength lies either in the IR within the structurally sensitive amide I′ protein band or at 275 nm, irradiates the sample. The change in signal induced by the *T*-jump (transmission for IR or backscattered fluorescence light) is detected in real time; hence, a very accurate baseline is automatically provided, and quite small changes can be determined. In both cases, the wavelength of the pump pulse corresponds to the peak of a weak D_2O near-IR absorption band ($\epsilon = 10.1$ cm^{-1} at 2 μm) and was chosen because 80% of the light is transmitted through the 100 μm path length cells used in the studies. The high transmission ensures a nearly uniform temperature profile in the approximately 9 nL (300 μm × 300 μm × 100 μm) laser interaction volume. The cells are identical for both the IR and fluorescence studies; they consist of sandwiched CaF_2 windows separated by a 100 μm spacer that is split into two separate compartments. The laser energy is absorbed by water (D_2O), and the temperature of the volume of water reaches its maximum value within ~20 ns (twice the fwhm of the pump pulse) since temperature thermalization and diffusion within water occur on sub-nanosecond time scales. The diffusion of heat out of the interaction volume takes about 20 ms in our cells.

The pump pulse and IR production used for the IR studies has been previously described (*15*, *16*). This spectrometer consists of a widely tunable (from 1600 to 1700 cm^{-1}) CW lead salt infrared diode laser (Laser Analytics) that functions as the probe for the IR measurements. An injection-seeded, Q-switched Nd:YAG laser (Spectra Physics GCR4) and Raman shifter (1 stokes shift in H_2 gas) produce the pump radiation at 2 μm (10 ns fwhm Gaussian pulse width) which is the source for the temperature jump. The combined total instrument response time is ~23 ns. The size of the *T*-jump was calibrated using the change in D_2O absorbance with temperature which acts as an internal thermometer in the range of 1632−1700 cm^{-1} (*15*). The *T*-jump was determined to within 2 °C

Figure 2 shows a schematic of the fluorescence spectrometer. The pump pulse used for the fluorescence studies employs a Spectra Physics GCR-4 Nd:YAG laser operating at 1.06 μm pumping a six foot Raman shifter filled with hydrogen gas (200 psi) to generate 20 mJ of 1.9 μm light. The 275 nm light is produced by a Spectra-Physics argon ion laser (model 2045). The UV probe laser beam was focused to about 25 μm in the sample with the IR pump pulse focused to 500 μm. This apparatus is capable of delivering enough light to create a 40−50 °C temperature jump; however, jumps were limited to <20 °C since randomly occurring sample "cavitations" were produced in the sample at higher-magnitude *T*-jumps. The UV laser can

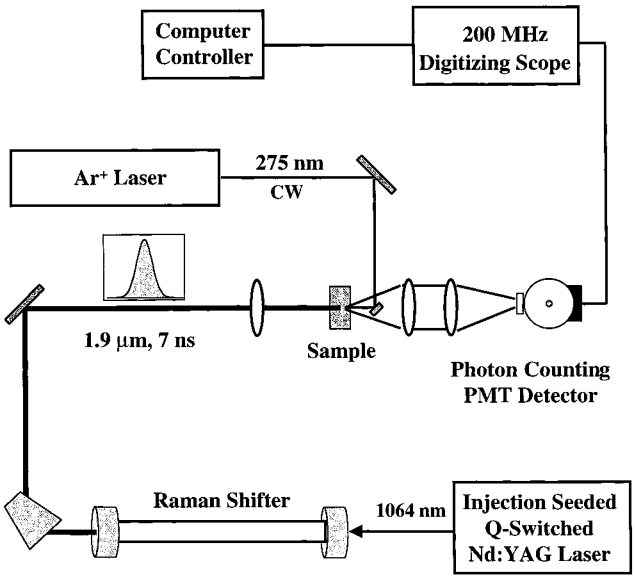

FIGURE 2: Schematic of the fluorescence *T*-jump spectrometer. See the text for details.

produce ~100 mW of 275 nm light; however, more than ~10 μW CW produced tryptophan photochemistry, yielding artifacts in the fluorescence signal. To increase the emission intensity (needed for short times), a shutter was placed in the path of the probe laser beam. This was opened for ~2 ms and timed so that the *T*-jump occurred ~0.5 ms after shutter opening. Hence, a probe intensity of ~1 mW within the open shutter condition could be used without inducing Trp photochemistry. The RC time constant in these experiments was 250 ns. A filter placed in front of the photomultiplier tube (RCA model C3134A photon counting tube) passed light over the entire emission profile of the Trp indole ring (300−400 nm at the 10% transmission points). A 200 μM solution of tryptophan was placed into the reference half of the split cell cuvette and translated into the laser interaction area before and after *T*-jump measurements of the sample were taken. The change in Trp fluorescence induced by the *T*-jump in the Trp reference solution was compared to static measurements, and this served to calibrate the size of the *T*-jump (accuracy, ±1 °C).

RESULTS

The equilibrium infrared and fluorescence E form melting curves obtained from the infrared amide I′ absorbance and the λ_{max} of the fluorescence, taken from the preceding paper (*1*), are shown in Figure 3. As discussed in the preceding paper (*1*), the amide I′ absorbance centered at 1648−1651 cm^{-1} reports on the formation of solvent-protected, native-like helix (*4*), and thus on the formation of a structured, hydrophobic core. Similarly, the fluorescence emission intensity and wavelength of the emission maximum are sensitive to the hydrophobic burial of the tryptophan indole ring and thus to core formation, albeit in a more localized sense than the IR measurements. The wavelength of the emission maximum shows the largest effects in static measurements, having a sharp red shift upon unfolding and exposure to water (*6*). The Trp residues are shown in pink in Figure 1 (Trp7 as a stick figure and Trp14 as a space-filling diagram). Notice that Trp7 is not included in the core. It is almost certain that Trp7 is solvated in E and in the

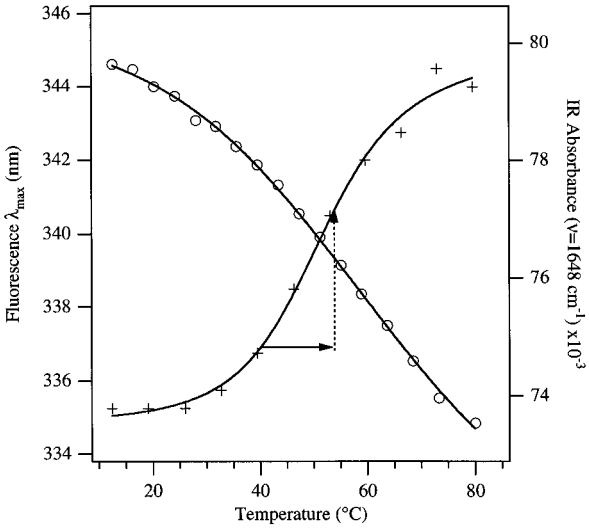

FIGURE 3: Equilibrium infrared (○) and fluorescence (+) melting curves for the E state of apoMb. The solid arrow shows the change in temperature on going from 42 to 54 °C, which represents the *T*-jump for the kinetic data shown in Figure 3, and the dashed arrow shows the relative magnitude of the change in λ_{max}. See ref *6* for the temperature dependence of the emission intensity through the melting transition. Both samples were in 20 mM NaCl at pH* 3.0 (D_2O). The concentration of apomyoglobin used in the fluorescence studies was 0.05 mg/mL. The concentration used for the IR studies was 4−8 mg/mL.

unfolded state so that its fluorescence emission does not change during the melting and, hence, does not affect the changes that are observed. In addition, at high temperatures the value of λ_{max} of the fluorescence emission of E is the same as that of L-tryptophan in solution; hence, the AGH core is completely unfolded at the high end of the melt [see the preceding paper (*1*)]. The melting curves generated with these two probes were obtained under identical conditions (except for a small difference in concentration). The infrared melting curve is broader than the fluorescence one, with a higher apparent T_m (fluorescence, $T_m = 49$ °C; IR, $T_m = 53$ °C) (data not shown). However, it is difficult to estimate the high-temperature baseline of the IR curve since the curve does not clearly "plateau" even at the highest temperatures. Moreover, there may be a small underlying linear contribution to the IR signal arising from changes in other parts of the protein which contain "solvated" helices. The infrared absorbance of solvated helices is broad and has a maximum near 1633 cm^{-1}, so part of this band overlaps the broad absorbance peak centered around 1648 cm^{-1}, the marker position for buried helix. Both of these issues make it difficult to determine thermodynamic parameters of the melt accurately from the IR data. Nevertheless, it is clear that the infrared and fluorescence probes reveal differences in melting behavior that depend on which structural feature is being probed.

The relaxation kinetics of the formation of U from E following a laser *T*-jump were probed by tryptophan fluorescence intensity and IR absorbance at the amide I′ frequency for the buried helix (1646 cm^{-1}) in a series of jumps to achieve various final temperatures. Figure 4 shows the results for a temperature jump from 42 to 54 °C probed by IR absorbance and by fluorescence emission. The final temperature was chosen to coincide closely to the midpoint of the equilibrium melt. The effect of the *T*-jump on the E

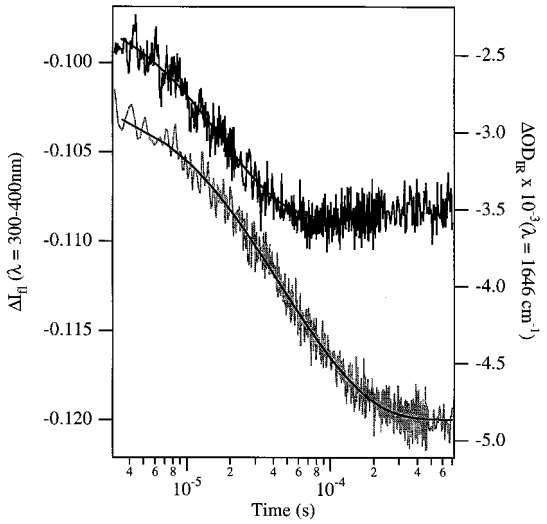

FIGURE 4: Fluorescence (top trace) and IR (bottom trace) kinetics traces for the E state of apoMb following a *T*-jump from 42 to 54 °C. Both samples were in 20 mM NaCl at pH* 3.0 (D_2O). The concentration of apomyoglobin used in the fluorescence studies was 2 mg/mL. The concentration used for the IR studies was 4−8 mg/mL.

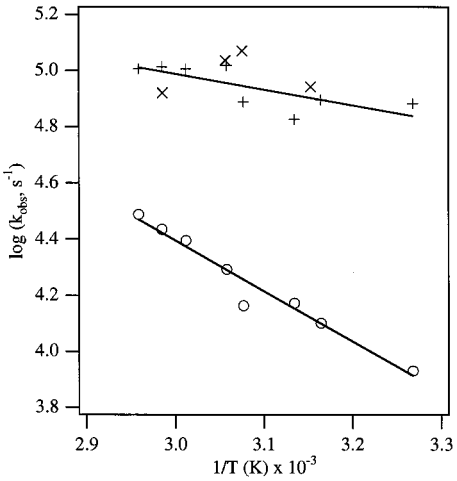

FIGURE 5: Temperature dependence of the two phases. The 1/*T* axis represents the temperatures reached during the *T*-jump. The data points show the rate dependence for the fluorescence data (+), the fast phase in the IR data (×), and the slow phase in the IR data (○). A linear least-squares regression algorithm using an Arrhenius function was used to determine the best straight line fit to the data (both the emission and IR data being used in the fit). The apparent activation energies obtained from the fit are 2 and 8 kcal/mol to the fast and slow phases, respectively. The errors in the rate constants, as judged from variations from run to run, are approximately ±2% for the fast phase and approximately ±1% for the slow phase. See the Figure 4 legend for sample details.

to U transition is shown schematically in Figure 3, with the horizontal line (solid) representing the *T*-jump and the vertical line (dashed) representing the relaxation. Single-exponential kinetics are observed in the fluorescence data, and the rate is unaffected by which salt is used, chloride or iodide (iodide transient not shown). Iodide is a strong quencher of tryptophan fluorescence; therefore, the signal sizes were much diminished relative to those of chloride, as expected. The emission results show a lifetime of 14 μs at this final temperature of 54 °C. In addition, the residual of transient minus the single-exponential fit shows only white noise that is less than 10% of the observed transient fluorescence signal. On the other hand, the IR relaxation is clearly not single-exponential, but is reasonably well fit with a biexponential function, with lifetimes of 14 (60%) and 59 μs (40%). The 14 μs component corresponds in time to the single phase observed in the fluorescence relaxation. In contrast, the slower IR phase (59 μs) is not observed under any conditions in the fluorescence relaxation and must be less than 10% of the observed transient based on the signal-to-noise ratio of the measurement.

Figure 5 shows the temperature dependence of the two relaxation phases. A fit of the data to an Arrhenius function is shown as solid lines in Figure 5. The fast phase is only weakly temperature dependent, while the show phase shows a moderate temperature dependence. The apparent activation energies for the two relaxation processes determined from the plot are 2 (fast phase) and 8 kcal/mol (slow phase). The actual activation energies for the constituent reactions that make up the relaxation processes depend on the specific kinetics model that best describes these reactions. For a two-state kinetics model (the preferred model as outlined below), where the unfolded state, U, is converted to E, the observed relaxation rate is proportional to the folding rate times 1/[U] (*14*). Since the concentration of U is temperature dependent, this affects the derived activation rate. Use of an Arrhenius function to fit a two-state model yields activation energies of 5 (fast phase) and 17 kcal/mol (slow phase).

DISCUSSION

Although myoglobin is a relatively simple protein, its folding dynamics involve the ordered formation of simpler structures. Specifically, the so-called AGH core, consisting of the overlapping portions of the A helix with the G−H turn and the G and H helices, forms very early in the folding process. Later events involve the formation of helices B−F and the final insertion of the heme group (*2*, *3*). This type of segmented folding behavior is observed in the thermodynamic behavior of the protein as well, since apomyoglobin contains at least three substates that are thermodynamically distinct (*1*, *6*, *16*, *17*). On the other hand, the acid-destabilized E state is a relatively simple structure (*1*). It is made up of the AGH core, ~30−40 residues in all, and a small amount of solvated helix. The core contains a number of tertiary contacts which result in cooperative-like melting. There are no tertiary contacts apart from the core in E as shown in the preceding paper (*1*). The results of the previous paper also show that the core of E consists of an ensemble of folded structures, which melt similarly but not identically. The small amount of solvated helix in E shows behavior thermodynamically (no cooperativity in melting) and kinetically [observed relaxation signals in response to a *T*-jump have a 40 ns lifetime (*4*)] distinctly different than that observed for its core region and so reside outside the core. The thermodynamics and folding dynamics of the two regions of E can be probed separately because (1) the Trp14 residue of apomyoglobin is located in the core and (*2*) the IR marker band due to desolvated amide linkages characteristic of the buried residues in the core of E (~1648 cm^{-1}) is well separated from the marker band for helical runs of protein which are solvated (~1632 cm^{-1}). Hence, core formation in the E state can be uniquely probed. The present relaxation experiments probe the emission changes of Trp14 [recall that

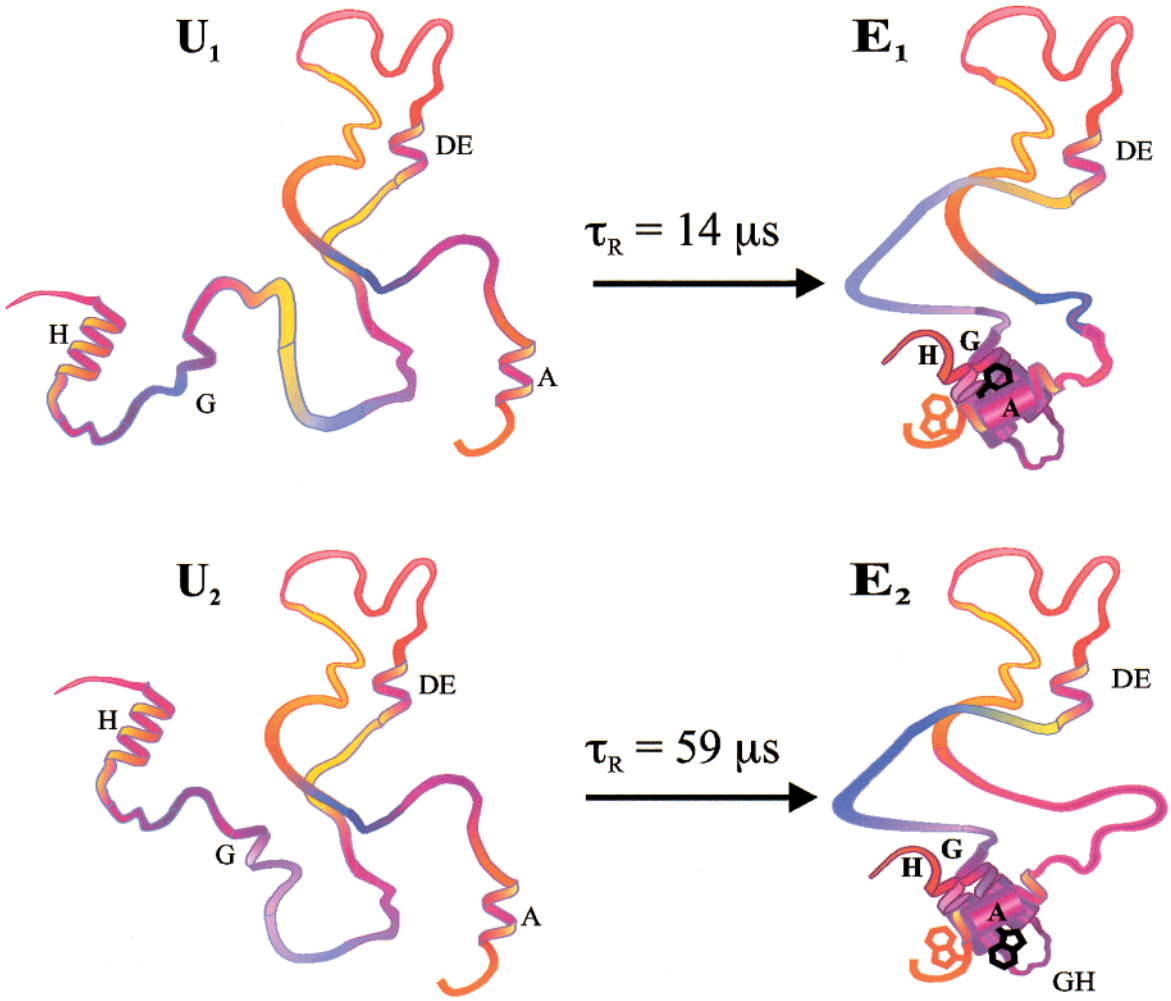

FIGURE 6: Schematic of the U to E transition based on the parallel folding pathway described in the text. τ_R is the relaxation time measured for each process. U_1 and U_2 are hypothetical subpopulations of the temperature-unfolded E state, and are in fast equilibrium with each other. Structure U_1 shows the G and H regions of the protein approximately parallel to one another, an alignment which is relatively close to the correct arrangement for the formation of tertiary structure of the collapsed E. Folding from U_1 results in the observed 14 μs process. Structure U_2, further away from the correct GH hairpin, leads to a "misformed" core conformation probably through nonspecific hydrophobic interactions and rearrangement. This interaction and rearrangement process takes 59 μs. In this structure, a core is formed but both the tryptophans are fully solvent exposed. The complete set of structural changes that take place from U to E are almost certainly more robust and complicated.

the emission of Trp7 does not appear to change when E melts (*1*)] and, hence, probe when core formation is initiated. The amide I′ absorption band at 1648 cm^{-1} is sensitive to the formation of buried helix or helical-like structures in the core of E so that the kinetics using IR absorption spectroscopy at this frequency give a picture of the kinetics of the formation of core secondary structure. The observed dynamics of core formation in E provides a picture of the earliest processes, on the upper reaches of the folding energy landscape of apomyoglobin, and may be representative of core formation in other proteins.

In our experiments, the pH- and salt-destabilized protein is in equilibrium between the high-temperature unfolded structure, U, and the more folded E state. Application of the temperature jump alters the equilibrium point, and the equilibrium shifts toward the unfolded state. The observed relaxation kinetics contain both folding and unfolding rates (see below) between the structures of E and U. From structural studies (*1*, *12*), the high-temperature unfolded form contains no tertiary structure and a small amount of secondary structure with about 5% of the protein forming helices or helix-like loops that show up in CD and IR measurements as helical signals. A recent NMR study of sperm whale apomyoglobin shows that the residues which make up the H helix and those that comprise a section of helices D and E retain some helical structure even at pH 2.3 (*12*). It is likely they also comprise the helical structure found in the heat-denatured U form at pH 3, which is the relevant unfolded species in our present study. In addition, the highly stable G−H turn may remain intact in the U state (*12*, *18*, *19*). The residues of the A and G helices do not form significant helical structure on their own in solution but have been shown to have a strong propensity for helix formation (i.e., "nascent" helix) (*20*). Schematic diagrams of the U and E states of apomyoglobin are shown in Figure 6.

A single fast phase is observed in the relaxation measurements of the Trp emission, while the IR kinetics data at the marker band position (1648 cm^{-1}) characteristic of core formation require at least two exponentials for a good fit. The relaxation observed for the buried Trp14 (fluorescence, 14 μs at 54 °C) correlates in time and temperature dependence with the relaxation observed for a significant fraction

of the buried helices (IR, 14 μs at 54 °C). Thus, the initial protection of Trp14 as part of the hydrophobic collapse involves the concomitant formation of compact secondary and tertiary structures. However, the slow IR phase has no counterpart in the fluorescence, implying that some folding pathways yield folded structures where both tryptophans are unprotected, but having buried helix.

There are two distinct classical minimal kinetic models capable of describing the data: a sequential folding pathway or parallel pathways. Since there are (at least) two observed relaxation events, either type of model must include at least two events and at least three species. We will assume the simplest case (that there are only two events) in describing these models. We also assume that a single core domain is formed in any one molecule, a reasonable assumption because of the small number of residues in the core (~36 residues), and the fact that the core melts completely and coorperatively makes it very unlikely that there is more than one core domain in a given molecule (*1*). Finally, both the IR and fluorescence measurements are sensitive to core formation, although fluorescence will only detect a core structure that buries one of the Trp indole rings. In the sequential scheme

Scheme 1

$$E_1 \leftrightarrows E_2 \leftrightarrows U$$

two conformations of E, E_1 and E_2, are in equilibrium and melting to U proceeds from E_1 through E_2. The relaxation kinetics equations describing these interconversions can be found in ref *7*. In our studies, the temperature jump induces melting so the relaxation takes place from left to right in Scheme 1. Hence, the fast phase (consisting substantially but certainly not exclusively of E_1 converting to E_2) must precede the slow phase (substantially the conversion between E_2 and U), or else the fast phase would not be observed. Such a kinetics model describes the data with difficulty, however, because the amplitudes of the two phases in the IR are almost equal while there is no trace of the slow phase in the Trp fluorescence kinetics (Figure 4). Using the observed relaxation rates and a wide range of reasonable equilibrium constants for the individual steps, we are unable to find a set of conditions that yields a single-exponential relaxation for the first step followed by a biexponential process with constituents of approximately equal amplitude.

This can be understood intuitively by considering a physically plausible scenario in which the structure protecting the Trp melts in the E_1 to E_2 transition (e.g., the A helix dissociates from the AGH core) followed by further changes in secondary structure in the E_2 to U transition (e.g., melting of the G and H helices). The fluorescence emission changes only occur from E_1 to E_2, while secondary structure changes occur about equally in both transitions. In this scenario, however, even if the further decrease in the level of buried helix in the E_2 to U transition does not perturb the tryptophan fluorescence, the E_2 to U phase kinetics would generally still be visible in the fluorescence data, because as the relatively slower E_2 to U transition proceeds it would perturb the E_1 to E_2 equilibrium with the slow relaxation time. Although awkward, there are situations where the sequential model would, however, fit the data. They rely on fortuitous situations and/or on situations where the relative cross sections of the IR and fluorescence signals do not match relative species concentrations. For example, if the Trp emission signals were equal in E_1 and U (unlikely since Trp14 is at least partially buried in E_1 and not in U) and smaller in E_2, fits to the data using the sequential model are possible.

The results are much more easily described by a kinetics model involving multiple pathways from the unfolded state to the heterogeneous E state:

Scheme 2

$$\begin{array}{c} E_1 \leftrightarrows U \\ \bullet \\ \bullet \\ \bullet \\ E_2 \leftrightarrows U \end{array}$$

where the three vertical dots represent the possibility of more than two folding routes. In this case, the E_1 to U represents the fast relaxation phase involving burial of the indole ring of Trp14 and the formation of secondary structure. The pathway involves a small enthalpic barrier based on the weak temperature dependence (Figure 5) and thus represents a pathway that is approximately energetically downhill under folding conditions (i.e., temperatures lower than the T_m). This pathway is very efficient. For comparison, applying the simple diffusion analysis of Eaton and co-workers (*21*) to a 110-residue polypeptide loop (approximately the size of apomyoglobin apart from the core in the E state), we estimated that it would take 70 μs for the two ends to come together. The observed rate ($k_{obs} = 10.1 \times 10^4$ s^{-1}) at 54 °C is close to the K_{eq} of 1 (that is, $k_1 \approx k_{-1}$) for the E_1 to U transition as measured by the Trp fluorescence melting thermodynamics (see Figure 3). In a two-state interconversion, the observed relaxation rate $k_{obs} = k_1 + k_{-1}$ so that the folding lifetime at this temperature is approximately τ_f ($1/k_{-1} = 18.6$ μs).

The E_2 to U pathway represents the observed slow kinetics relaxation phase. This pathway has a significant enthalpy barrier in the transition state of the folding pathway as indicated by its steep temperature dependence shown in Figure 5 and, hence, occurs more slowly. Furthermore, this pathway leads to a conformation of the protein in which Trp14 is not protected from solvent because no change in the fluorescence intensity is observed. This is consistent with the heterogeneity of the E state described in the preceding paper (*1*). Specifically, there are at least two conformers of E, one having considerably greater protection of Trp14 from solvent, as evidenced by the shift in the Trp fluorescence emission maximum upon titration with iodide. It is unlikely that Trp14 is completely protected in any native-like conformation of solely the AGH core, since it lacks the necessary E helix, which covers one side of Trp14 in the native structure (see Figure 1). It is easy to envision structures which only partially protect Trp14, or do not protect it at all.

The fast phase appears to be close to a downhill pathway to a native-like structure in which Trp14 is protected. In contrast, the slower phase appears to form a nonnative structure, with an unprotected Trp14, in an activated process. We propose the following structural model for explaining

the observed kinetics. The fast phase involves the diffusion-limited collapse of a subset of conformations of U that contains elements of secondary structure which resemble that found in the core and are close to a proper alignment. This probably includes a partially formed H helix with the G−H turn largely formed, a nascent G helix, and the rapidly fluctuating A helix. A similar collapse was suggested for the *T*-jump refolding of cold-denatured apomyoglobin in its native state (*22*). In contrast, the slow phase involves subsets of unfolded conformations which contain either less native-like secondary structure or secondary structure that is not well aligned. This collapse is then relatively less specific and is followed by a rearrangement over a barrier to a nonnative (i.e., unprotected Trp14) corelike structure. A schematic diagram of this scenario is shown in Figure 6.

There are several lines of evidence that support this model. First, as mentioned above, the U state is known to contain some fluctuating residual structure, which could certainly facilitate the downhill collapse. On the time scale of the formation of the AGH core, it is possible for portions of the isolated helices to quickly fluctuate in and out of helical conformations, since the time scale for helix formation is 50−200 ns (*15*, *16*, *23*). Thus, as the tertiary structure of the core is formed relatively slowly, the polypeptide can sample native-like secondary structures many times. On the other hand, the subpopulation of the unfolded state that does not have the G−H turn essentially preformed or does not form it within the diffusional encounter time of the two ends of apoMb can form hydrophobic interactions between the GH and A portions of the protein in a relatively "incorrect" orientation and is more likely to collapse in a nonspecific manner. Second, the rates and temperature dependencies are consistent with a relatively downhill process to form the AGH core versus an activated process to form a nonnative core. The fast rate is on the order of the diffusion time for the ends of the polypeptide chain to encounter one another, and the nucleation rate of a helix is at least 1 order of magnitude faster, which supports the possibility that the AGH core can be formed on this time scale by a direct downhill process. The apparent activation energy for this process (5 kcal/mol, uncorrected for the temperature-dependent change in viscosity) is also on the order of that expected for a diffusion-controlled reaction. In a sense, the slower process is pathological because it involves formation of nonnative structure. Nevertheless, this process provides insight into the time required for a nonspecific collapse followed by rearrangement over a barrier, which may be a significant rate-limiting process in many folding reactions. Finally, this picture is consistent with the heterogeneous nature of E; a particular conformation with its particular set of tertiary contacts condenses as determined by statistics. It clearly takes less time and is more efficient to condense into a compact corelike structure if there are multiple ways in which stable tertiary contacts can be made.

In summary, the formation of apoMb's core proceeds via multiple pathways, the fastest of which are close to energetically downhill, while others take longer because the collapsed structures that are initially formed require activation to reach tertiary contacts of a core structure. However, all pathways are still very fast (quite close to the physical limits placed by diffusion). Our results then strongly suggest that the specific sequence that encodes the A−G−H interaction is such that multiple conformers of corelike structures exist, all having high values of enthalpic stability (i.e., involving tertiary contacts between many residues). It is also appears that the amino acid sequence encodes few tertiary interactions in the absence of helices G and H because no other tertiary contacts appear to be made in apoMb without them. Both strategies appear to facilitate efficient (very fast) folding of E. This picture is quite different from a nonspecific hydrophobic collapse, however, because the cooperativity of the core clearly requires multiple, cooperative, tertiary interactions. Thus, the acid-destabilized forms of apoMb are very different from the molten globule of α-lactalbumin, for example, which appears to contain a noncooperative assembly of its constituent helices (*24*).

REFERENCES

1. Gilmanshin, R., Gulotta, M., Dyer, R., and Callender, R. (2001) *Biochemistry 40*, 5127−5136.
2. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
3. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) *Protein Sci. 8*, 45−49.
4. Gilmanshin, R., Callender, R. H., and Dyer, R. B. (1998) *Nat. Struct. Biol. 5*, 363−365.
5. Jamin, M., and Baldwin, R. L. (1996) *Nat. Struct. Biol. 3*, 613−618.
6. Gilmanshin, R., Dyer, R. B., and Callender, R. H. (1997) *Protein Sci. 6*, 2134−2142.
7. Jamin, M., and Baldwin, R. (1998) *J. Mol. Biol. 276*, 491−504.
8. Barrick, D., and Baldwin, R. L. (1993) *Protein Sci. 2*, 869−876.
9. Kay, M. S., and Baldwin, R. L. (1996) *Nat. Struct. Biol. 3*, 439−445.
10. Lecomte, J. T. J., Kao, Y. H., and Cocco, M. J. (1996) *Proteins: Struct., Funct., Genet. 25*, 267−285.
11. Luo, Y., Kay, M. S., and Baldwin, R. L. (1997) *Nat. Struct. Biol. 4*, 925−930.
12. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
13. Maurus, R., Bogumil, R., Nguyen, N., Mauk, A., and Brayer, G. (1998) *Biochem. J. 332*, 67.
14. Fersht, A. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Freeman and Co., New York.
15. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Woodruff, W. H., Callender, R. H., and Dyer, R. B. (1996) *Biochemistry 35*, 691−697.
16. Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W., and Dyer, R. B. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3709−3713.
17. Dyer, R. B., Gai, F., Woodruff, W., Gilmanshin, R., and Callender, R. H. (1998) *Acc. Chem. Res. 31*, 709−716.
18. Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E., and Dyson, H. J. (1993) *Biochemistry 32*, 6348−6355.
19. Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J., and Wright, P. E. (1993) *Biochemistry 32*, 6356−6364.
20. Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., and Wright, P. E. (1993) *Biochemistry 32*, 6337−6347.
21. Eaton, W. A., Munoz, V., Thompson, P. A., Chan, C. K., and Hofrichter, J. (1997) *Curr. Opin. Struct. Biol. 7*, 10−14.
22. Sabelko, J., Ervin, J., and Gruebele, M. (1998) *J. Phys. Chem. B 102*, 1806−1819.
23. Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997) *Biochemistry 36*, 9200−9210.
24. Schulman, B. A., and Kim, P. S. (1996) *Nat. Struct. Biol. 3*, 682−687.

BI002256N