Native and Non-native Secondary Structure and Dynamics in the pH 4 Intermediate of Apomyoglobin†

David Eliezer, John Chung, H. Jane Dyson, and Peter E. Wright*

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

Received November 4, 1999; Revised Manuscript Received January 7, 2000

ABSTRACT: The partly folded state of apomyoglobin at pH 4 represents an excellent model for an obligatory kinetic folding intermediate. The structure and dynamics of this intermediate state have been extensively examined using NMR spectroscopy. Secondary chemical shifts, $^1H-^1H$ NOEs, and amide proton temperature coefficients have been used to probe residual structure in the intermediate state, and NMR relaxation parameters T_1 and T_2 and $\{^1H\}-^{15}N$ NOE have been analyzed using spectral densities to correlate motion of the polypeptide chain with these structural observations. A significant amount of helical structure remains in the pH 4 state, indicated by the secondary chemical shifts of the $^{13}C^{\alpha}$, ^{13}CO, $^1H^{\alpha}$, and $^{13}C^{\beta}$ nuclei, and the boundaries of this helical structure are confirmed by the locations of $^1H-^1H$ NOEs. Hydrogen bonding in the structured regions is predominantly nativelike according to the amide proton chemical shifts and their temperature dependence. The locations of the A, G, and H helix segments and the C-terminal part of the B helix are similar to those in native apomyoglobin, consistent with the early, complete protection of the amides of residues in these helices in quench-flow experiments. These results confirm the similarity of the equilibrium form of apoMb at pH 4 and the kinetic intermediate observed at short times in the quench-flow experiment. Flexibility in this structured core is severely curtailed compared with the remainder of the protein, as indicated by the analysis of the NMR relaxation parameters. Regions with relatively high values of $J(0)$ and low values of $J(750)$ correspond well with the A, B, G, and H helices, an indication that nanosecond time scale backbone fluctuations in these regions of the sequence are restricted. Other parts of the protein show much greater flexibility and much reduced secondary chemical shifts. Nevertheless, several regions show evidence of the beginnings of helical structure, including stretches encompassing the C helix−CD loop, the boundary of the D and E helices, and the C-terminal half of the E helix. These regions are clearly not well-structured in the pH 4 state, unlike the A, B, G, and H helices, which form a nativelike structured core. However, the proximity of this structured core most likely influences the region between the B and F helices, inducing at least transient helical structure.

Many proteins populate partially folded states at equilibrium under suitable conditions (*1*). For several small globular proteins, there is strong evidence for a close structural correspondence between such equilibrium states and kinetic intermediates observed during the folding process (*2*−*5*). According to fundamental thermodynamics, equilibrium states need not resemble kinetic intermediates, but a relationship between them is not precluded. Important undiscovered principles governing the behavior of proteins may be responsible for the observed similarities. Until such time as we understand the detailed behavior of protein free energy in phase space (i.e., the protein energy landscape), no general conclusions can be drawn on the relevance of equilibrium states to folding kinetics. In those specific cases where a close correspondence has been empirically demonstrated, however, the utility of detailed characterization of stable equilibrium states as an approximation to properties of the corresponding kinetic intermediates is indisputable.

Irrespective of any relation to kinetic intermediates, partially folded equilibrium states are of considerable interest in an in vivo context. A growing number of proteins are found to be at least partially unfolded in the absence of their binding partners (see, for example, *6*−*8*), and residual structure in partially folded states may also play an important role in function. Partially folded proteins are very likely involved in membrane interaction and translocation events (*9*), and partially folded proteins are implicated as precursors of in vivo protein aggregates of varying morphologies (*10*), including pathogenic amyloid fibrils and diffuse plaques associated with a number of neurodegenerative conditions such as Alzheimer's disease (*11*). To understand the behavior of partially folded states in all of the above contexts, it is necessary to obtain detailed information about local structure and dynamics in such systems.

Apomyoglobin at pH 4 in the absence of salt populates a partially folded state (*12*) bearing a strong resemblance to a kinetic intermediate formed early in the apomyoglobin folding process. This intermediate is obligatory, as shown

† This work was supported by Grants DK34909 and GM57374 from the National Institutes of Health.

* Correspondence should be addressed to this author at the Department of Molecular Biology, The Scripps Research Institute, MB-2, 10550 N. Torrey Pines Rd., La Jolla, CA 92037. Tel: (858) 784-9721, Fax: (858) 784-9822, Email: wright@scripps.edu.

10.1021/bi992545f CCC: $19.00
Published on Web 02/26/2000

by quench-flow mass spectrometry experiments (*13*). The similarity of this equilibrium state to the kinetic intermediate has been established through measurements of the radius of gyration (*14*), helical content (*2*), and pattern of protection of backbone amide protons from exchange with solvent protons (*2*), all of which are essentially identical for both kinetic and equilibrium intermediate states. Backbone amides in the A, G, and H helices and the C-terminal end of the B helix were found to be protected from exchange both at equilibrium and 5 ms into the folding process. These observations, together with mutagenesis data (*15*), led to the suggestion that the A, G, and H helices are fully formed in the intermediate and packed into a nativelike AGH hydrophobic core (*12*). This model accounted for the approximately 35% helical content inferred from CD[1] data. The small radius of gyration relative to the unfolded state, together with the cooperativity of the folding/unfolding transition, led to the conclusion that the intermediate is globular and compact.

We have used heteronuclear multidimensional NMR spectroscopy to characterize structure and dynamics in the sperm whale apomyoglobin pH 4 state at the level of individual residues. A small part of this work has appeared previously in the context of the delineation of the folding pathway of apomyoglobin (*16*). The current results provide a much more comprehensive view of the pH 4 state by greatly extending the chemical shift data to include all four nuclei which show strong correlations with secondary structure, by providing short-range NOE data and amide proton temperature coefficients to complement the chemical shift analysis, and by providing a complete set of dynamics parameters, allowing for a spectral density analysis of backbone motions. Using this extensive set of data, we have been able to identify specific boundaries for the elements of secondary structure in the intermediate, both for the well-formed AGH core helices and for the more transient helical regions outside the core.

MATERIALS AND METHODS

Sample Preparation. Apomyoglobin corresponding to the sperm whale sequence was expressed in *E. coli* and purified from inclusion bodies using an HPLC protocol as previously described (*17*). Samples of the pH 4 intermediate were prepared by dissolving ~3 mg of lyophilized protein in 0.4 mL of sodium acetate buffer (pH 6.1, 10 mM) in 90% ^1H_2O/10% ^2H_2O. Samples were passed through a Sephadex fine G-25 column equilibrated in the same buffer, and titrated to pH 4.0 using microliter volumes of 2.9 M deuterated acetic acid. Then 40 μL of deuterated ethanol was added to stabilize the protein against aggregation during the acquisition of NMR data (*18*); the final pH was 4.1. Samples were transferred to Shigemi-type NMR tubes to keep the air volume above the sample to a minimum, thus preventing distillation of the ethanol at the probe temperature of 50 °C. Final protein concentrations were around 300 μM.

NMR Spectroscopy. Experiments were performed on a Bruker DMX 750 MHz spectrometer at 50 °C. Pulse sequences employed gradient coherence selection. An external DSS sample was used to determine reference frequencies for all nuclei (*19*). Assignments were made primarily from a series of triple resonance experiments (see below), a number of which were run using partially deuterated (~80% by electrospray mass spectrometry) ^{13}C,^{15}N-labeled samples. Typical spectra were recorded with 512 complex points in the direct (^1H) dimension, 64 or 80 complex points in the first indirect (^{13}C) dimension, and 32 complex points in the second indirect (^{15}N) dimension. Typical spectral widths were 4496 Hz (6 ppm) for ^1H, 1825 Hz (24 ppm) for ^{15}N, 4000 Hz (21.2 ppm) for $^{13}C^{\alpha}$, 12 500 Hz (66.3 ppm) for $^{13}C^{\alpha\beta}$, and 1515 Hz (8 ppm) for ^{13}CO. NOEs were measured in an HSQC−NOESY−HSQC experiment (*20*) using partially deuterated protein and a mixing time of 300 ms. For R1 measurements, data were recorded with relaxation delays at 10, 20, 40, 80, 160, 320, 640, 1280, and 2000 ms, with repeat measurements at 10, 80, and 640 ms for the purpose of error estimates. For R2 measurements, time points were taken at 6, 10, 18, 34, 66, 130, 258, 514, and 682 ms, with repetitions at 6, 34, and 514 ms. CPMG ^{15}N refocusing pulses were at 1 ms intervals, with ^1H refocusing pulses interspersed after every second ^{15}N pulse to eliminate cross-correlation effects (*21*). Recycle delays of 2.5 s were used for both measurements. For steady-state NOE measurements, ^1H saturation was accomplished using 120° pulses spaced 18 ms apart for 3.6 s during the 5.2 s recycle delay and a bandwidth of 11 363 Hz. For the $^1H-^{15}N$ double resonance experiments, 128 complex points were typically collected in the indirect (^{15}N) dimension.

Data Analysis. NMR data were processed using the program Felix (Molecular Simulations Inc.). Data were typically apodized using a 60° shifted $\sin^2$ function before zero-filling and Fourier transformation. In general, base line corrections were not necessary. Maximum peak heights were used in processing the R1, R2, and steady-state NOE data. Decays were fit to monoexponential functions using in-house least-squares fitting software. Spectral densities were calculated as described by Peng and Wagner (*22, 23*) with in-house software using Numerical Recipes (*24*) routines. Random coil chemical shifts were taken from Wishart et al. (*25*), with corrections for residues preceding prolines. H^N random coil chemical shifts were extrapolated to 50 °C using amino acid-specific temperature coefficients from Merutka et al. (*26*). Temperature coefficients were calculated as the difference between the chemical shifts at the start and end points of the temperature range. For a handful of residues, this result was compared with that obtained from a linear regression to all the temperature points. The data were highly linear, and no significant differences were found between the values obtained by the two methods.

Resonance Assignments. The standard methods of establishing resonance assignments through nearest-neighbor connectivities were hampered by the high degree of spectral overlap, together with the broadness of some of the resonance lines in the pH 4 intermediate. To overcome these problems, it was necessary to collect triple resonance data utilizing the resonance dispersion of several heteronuclei as well as that of the NH (*27*). Connectivities based on the backbone ^{13}CO, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{15}N nuclei were obtained from the following pairs of experiments: HNCO (*28*) and H(CA)CO(CA)NH (*29*), HNCA and HN(CO)CA (*28*), HN(CA)CB (*30*) and CBCA(CO)NH (*31*), and the $^1H-^{15}N$ HSQC−NOESY−HSQC experiment (*20*). In addition, a series of

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; CD circular dichroism.

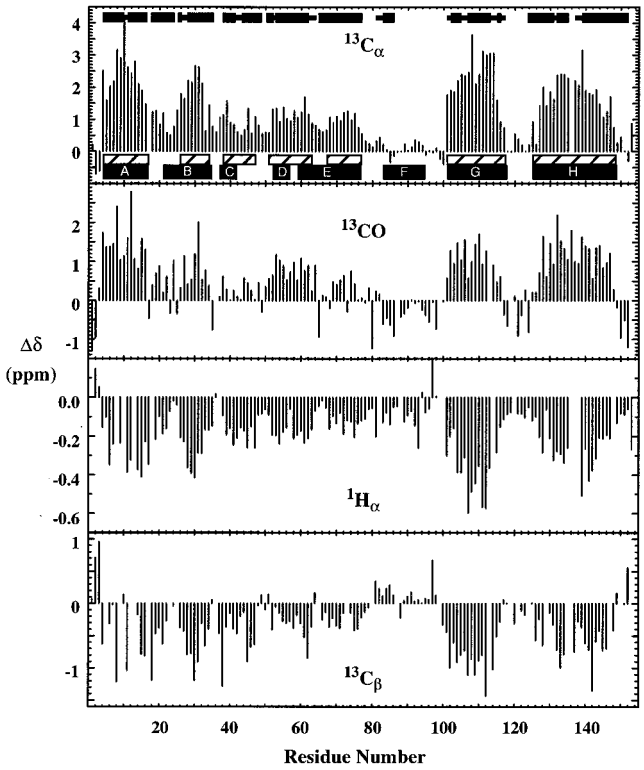

FIGURE 1: Secondary chemical shifts for $^{13}C^{\alpha}$, ^{13}CO, $^{1}H^{\alpha}$, and $^{13}C^{\beta}$ as a function of residue number in apomyoglobin at pH 4.1. Bars at the top of the figure indicate the presence of $H^N_i-H^N_{i+1}$ NOEs; the smaller bars indicate that the NOE was ambiguous due to resonance overlap. Black rectangles at the base of the top panel indicate the locations of helices in the native holomyoglobin structure (*51*). Hatched rectangles indicate putative boundaries for helical regions in the pH 4 intermediate, based on the chemical shift and NOE data.

2D $^{1}H-^{15}N$ HSQC experiments were collected at several pH values between 2 and 4. By tracking peaks, it was possible either to directly transfer resonance assignments previously obtained for the pH 2.3 acid-unfolded state or to significantly constrain the possible identity of resonances at pH 4.1. This information greatly facilitated finding the correct connectivities in the above-mentioned data sets, and led to nearly complete backbone resonance assignments. Resonance assignments for the pH 4 form of apomyoglobin have been deposited in the BioMagResBank (accession number 4568).

RESULTS

Secondary Structure. The specific resonance frequencies (chemical shifts) of NMR signals from proteins are highly sensitive to the structure surrounding the corresponding nuclei. Although efforts to quantitate this dependence in an exact fashion are in progress (*32*), the most common methods of interpreting chemical shift data currently rely on empirically observed correlations. A particularly strong correlation has been found between the deviations (termed secondary shifts) of $^{13}C^{\alpha}$ and H^{α} chemical shifts from their random coil values, and the involvement of the corresponding residues in secondary structure (*33, 34*). A similar correlation is found for $^{13}C^{\beta}$ and ^{13}CO chemical shifts. Figure 1 shows the secondary shifts for $^{13}C^{\alpha}$, ^{13}CO, H^{α}, and $^{13}C^{\beta}$ nuclei in apomyoglobin at pH 4.1. For $^{13}C^{\alpha}$ and ^{13}CO, positive secondary shifts indicate a preference for ϕ,ψ angles characteristic of helical conformations whereas negative values indicate a propensity toward extended conformations with ϕ,ψ angles in the β-sheet region. For H^{α} and $^{13}C^{\beta}$ nuclei, the trend is reversed, with negative values indicating helical conformations and positive values indicating extended conformations. It is immediately evident from the data that there is a strong propensity for helical structure throughout the pH 4 intermediate, with the exception of a region which corresponds to the F helix in the holoprotein structure. Well demarcated stretches of high helical propensity are evident in regions corresponding to the A, G, and H helices of the holoprotein structure, as well as the C-terminal end of the B helix. The amplitude of the secondary shifts in the A, G, and H regions is indicative of highly populated stable helical structure, although it is not as large as that observed in the native state. A decrease in the secondary shifts toward the ends of the helices, especially evident in the H helix region, suggests that the helices are frayed. In addition to these highly helical regions, a significant propensity toward helical conformations is evident in the region between the B and F helices.

Helical secondary structure is associated with a specific pattern of NOEs between protons, including $H^N_i-H^N_{i+1}$, $H^N_i-H^N_{i+2}$, $H^{\alpha}_i-H^N_{i+2}$, $H^{\alpha}_i-H^N_{i+3}$, $H^{\alpha}_i-H^N_{i+4}$, and $H^{\alpha}_i-H^{\beta}_{i+3}$ NOEs (*35*). Of these, the $H^N_i-H^N_{i+1}$ NOE signals are particularly strong in helical structure. While these NOEs are also present in β-sheet and turn structures, they are either much weaker (in β-sheets) or not contiguous (in turns), and can therefore be useful in discriminating between helical and nonhelical conformational preferences in unfolded or partly folded proteins and peptides (*36*). The location of significant $H^N_i-H^N_{i+1}$ NOEs observed in apomyoglobin at pH 4.1 is shown by the bars at the top of Figure 1. A good correlation is observed between the presence of $H^N_i-H^N_{i+1}$ NOEs and helical secondary shifts.

The detection and assignment of the $H^N_i-H^N_{i+1}$ NOEs was made possible by their high intensity and the relatively good dispersion of backbone ^{15}N chemical shifts. The necessity for using very low sample concentrations for NMR measurements of apomyoglobin at pH 4 (*18*) has impeded the detection of the medium-range NOEs typical of helical structure, which are less intense than the $H^N_i-H^N_{i+1}$ NOEs and most of which involve aliphatic or side chain protons with poorer chemical shift dispersion than the H^N.

Hydrogen Bonding. A measure of the participation of amide protons in intramolecular hydrogen bonds can be obtained from the temperature dependence of the H^N chemical shift (temperature coefficient) (*37, 38*). In the absence of structural changes, the chemical shift of an amide proton that is protected from exchange with solvent protons by intraprotein hydrogen bonds is less sensitive to changes in temperature than that of a proton which can exchange with the bulk solvent (and which therefore reflects the sensitivity of the bulk H_2O chemical shift to temperature). Random coil temperature coefficients in unstructured peptide models are around 8 ppb/K (*26*). Figure 2 shows the deviation from random coil values (*26*) of H^N temperature coefficients measured throughout apomyoglobin at pH 4.1 over a 10 °C range from 40 to 50 °C. Over this range, the optical properties of the intermediate undergo only small changes, suggesting that structural alterations are minimal. The A, G, and H helix regions clearly display depressed temperature coefficients, although there is an interruption

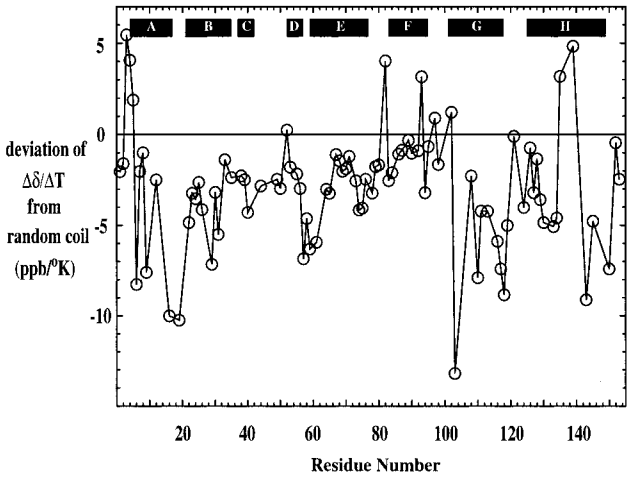

FIGURE 2: Deviation of the temperature coefficients from their random coil values (*26*) for backbone H^N chemical shifts in apomyoglobin at pH 4.1.

in the center of the H helix (This may be due to solvent-exposed amides on the outside of the helix. Similar periodicity is frequently observed for the helices in native proteins.) Lowered temperature coefficients are also apparent at the C-terminal end of the B helix and in a region between the D and E helices.

Hydrogen bonds can also exert a significant effect on the H^N chemical shift; indeed, the H^N temperature coefficient shows a correlation with the H^N chemical shift (*39*, *40*). One reflection of the partially structured nature of the apomyoglobin intermediate is the presence of several H^N resonances well outside the envelope defined by random coil chemical shifts, suggesting the formation of hydrogen bonds. Secondary structure and ring-currents arising from the proximity of aromatic rings also affect the H^N chemical shift, and it is difficult to separate the effects. The average magnitude of the H^N shifts induced by helical structure is 0.2 ppm (*41*). In the present case, therefore, shifts greater than 0.2 ppm are suggestive of hydrogen bond formation, although it is not possible to exclude effects from secondary structure or other sources. Figure 3 shows a $^1H-^{15}N$ HSQC spectrum of the apomyoglobin pH 4 intermediate. Peaks for which the H^N secondary chemical shift is greater than 0.2 ppm are indicated, with circles indicating a shift in the same direction observed in the native state, and squares indicating a shift in the opposite direction. The predominance of circled peaks suggests that many of the hydrogen bonds present in the intermediate may be nativelike. In general, the identity of the indicated peaks correlates quite well with the location of depressed temperature coefficients, although there are some exceptions. For example, Ser 3, Glu 4, Gly 5, and Trp 7 show significantly shifted amides, probably due to ring-current effects arising from the proximity of the tryptophan ring rather than to hydrogen bonding.

Dynamics. The relaxation rates of NMR signals are sensitive to the motions of the corresponding nuclei, and can be used to probe internal protein dynamics. The relaxation rates of longitudinal and transverse magnetization of backbone ^{15}N nuclei (R1 and R2), as well as the $^1H-^{15}N$ steady-state heteronuclear NOE, are sensitive to motions on a picosecond to nanosecond time scale. The $^1H-^{15}N$ NOE is typically most sensitive to higher-frequency motions of the backbone (*42*), with values near 1.0 indicating a lack of such

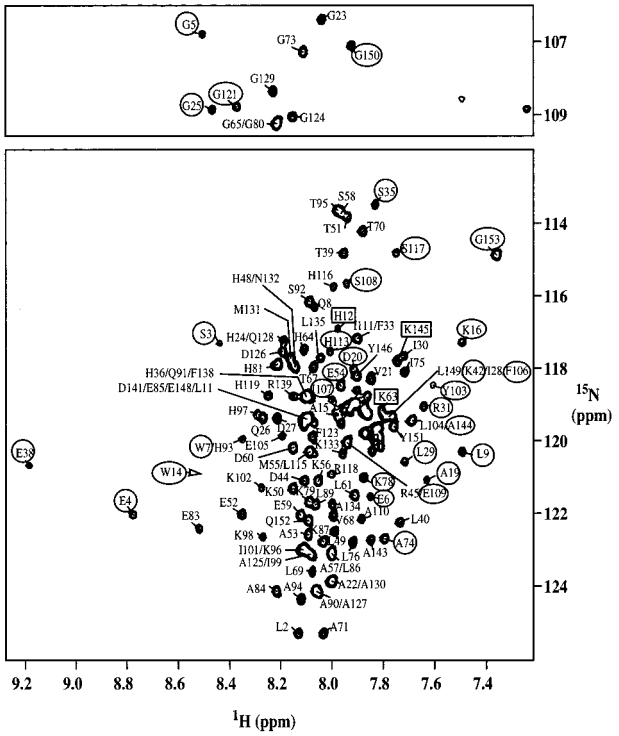

FIGURE 3: $^1H-^{15}N$ HSQC spectrum of apomyoglobin at pH 4.1. Residues with circled labels have H^N secondary shifts greater than 0.2 ppm, and are shifted in the same direction observed in the native state. Residues with boxed labels also have H^N secondary shifts greater than 0.2 ppm, but the shifts are in the opposite direction to that observed in the native state.

motions, and lower values indicating increased local flexibility of the polypeptide. The R1 and R2 relaxation rates and heteronuclear NOEs are plotted in Figure 4 for each backbone amide in the apomyoglobin intermediate. High values of the heteronuclear NOE are observed for the regions corresponding to the A, G, and H helices and the C-terminal end of the B helix. The average value in these regions, about 0.7, approaches the values of 0.8−0.9 found in the native apo- and holoproteins (*16*) and indicates that backbone flexibility on very fast time scales is limited. This observation is consistent with the strong helical character inferred from the secondary chemical shifts in these regions, and with the hypothesis that these segments comprise a well-packed hydrophobic core (hereafter referred to as the AGH[B] core), possibly in a nativelike topology (*43*). In contrast, the F helix region shows negative NOE values, suggesting an unstructured free-flight chain. Between the B and F helices, there is a steady decrease in the NOE values, suggesting an increasing degree of flexibility in moving from the relatively well-structured A helix toward the unstructured F helix.

Although all three relaxation parameters are influenced by motions over a range of time scales, the transverse relaxation rate, R2, is more sensitive to lower frequency (nanosecond) motions, and also reflects contributions from slower millisecond to microsecond exchange processes which may cause line broadening in the NMR spectrum. In the apomyoglobin intermediate, the high R2 values in the AGH-[B] core indicate the dominance of slower time-scale motions, perhaps rigid-body-like tumbling of this relatively well-packed region as an autonomous unit. The relaxation mechanism may also include contributions from slower exchange processes arising from larger scale motions of

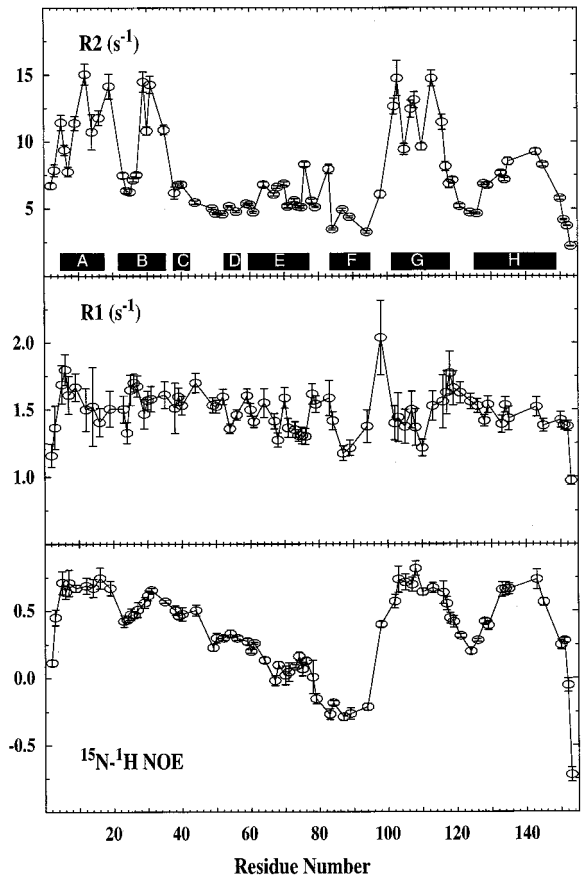

FIGURE 4: Backbone relaxation parameters measured throughout apomyoglobin at pH 4.1. Heteronuclear NOE values are most sensitive to the fastest time scale (picosecond to subnanosecond) motions, whereas R2 values are most sensitive to slower (nanosecond) motions or very slow (microsecond to millisecond) conformational exchange processes.

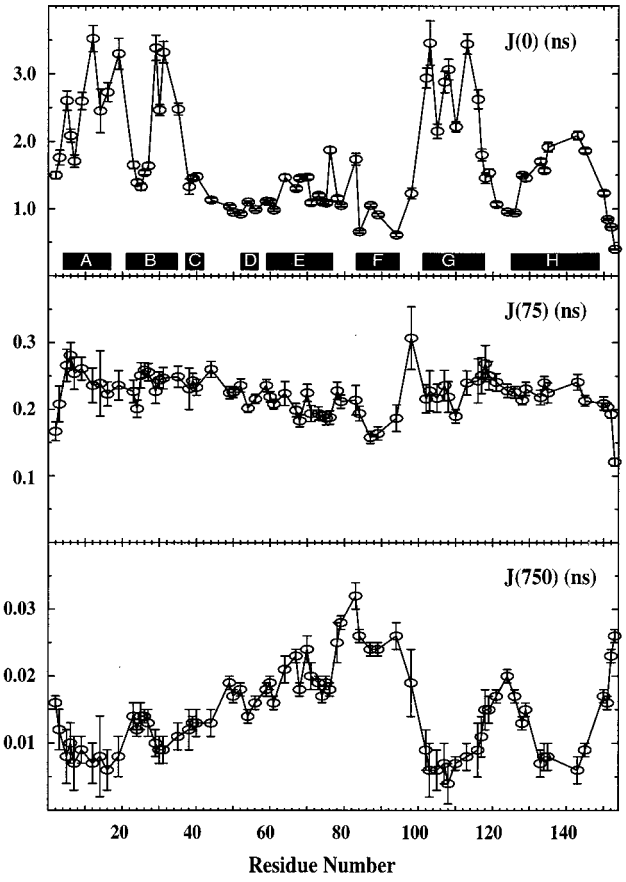

FIGURE 5: Spectral densities of backbone motions in the pH 4.1 apomyoglobin intermediate at three different frequencies.

elements within this core. Lower R2 values outside the core reflect the importance of faster internal motions in the relatively unstructured remainder of the protein. The R1 relaxation rate by itself does not discriminate effectively between faster and slower motions within the picosecond to nanosecond time scale.

The R1, R2, and NOE relaxation parameters are related by a set of simple linear equations to the spectral density function, $J(\omega)$, of the protein backbone at three specific frequencies: $\omega = 0$, ω_N, and $0.86\omega_H$ (*22*, *23*, *44*). The spectral density function at each of these three values is plotted as a function of residue number in Figure 5. $J(0.86\omega_H)$ reflects the highest frequency motions, and is largely determined by the heteronuclear NOE. $J(\omega_N)$ is most sensitive to R1, and like R1 does not discriminate well between faster and slower motions. $J(0)$ is most sensitive to R2, reflecting slower (nanosecond) motions, with possible contributions from exchange processes on the millisecond to microsecond time scale. Like the corresponding relaxation parameters, the per-residue spectral density functions indicate that the backbone AGH[B] core is less mobile on very fast time scales, and behaves more like a single rigid domain, with the possibility, however, of slower conformational exchange processes at discrete sites with unusually high $J(0)$s. The F helix region has a high degree of mobility on fast (subnanosecond) time scales, indicating a lack of structure. Between the B and F helices, there are varying degrees of restriction on faster motions.

Analysis of the dynamics of folded proteins is commonly done using the "model-free" method (*45*, *46*), utilizing the computer program ModelFree (*47*). These methods do not provide meaningful analysis of the data for partially folded proteins, due to the wide variation in local correlation times between residues in the more structured regions and those in the less-structured and unstructured regions. Efforts are currently under way to account for these differences in model-free calculations.

DISCUSSION

The AGH[B] Core. In the pH 4 apomyoglobin intermediate, the secondary chemical shifts of four different nuclei which can be used as indicators of secondary structure all agree in predicting significant helical structure in the regions corresponding to the A, G, and H helices and the C-terminal end of the B helix of the holoprotein. Consistent indications of highly helical secondary structure in the secondary chemical shifts are supported by $H^N_i-H^N_{i+1}$ NOEs and indications of hydrogen bonding in the same regions of the sequence. Combined with dynamics measurements of the backbone amides that approach those of the native protein, these observations provide strong support for the hypothesis that the A, G, and H helices and the C-terminal part of the B helix form a compact, rigid core in the equilibrium intermediate. The additional hypothesis that this core is arranged in a nativelike topology is not addressed by the current results, but is supported by other studies (*43*).

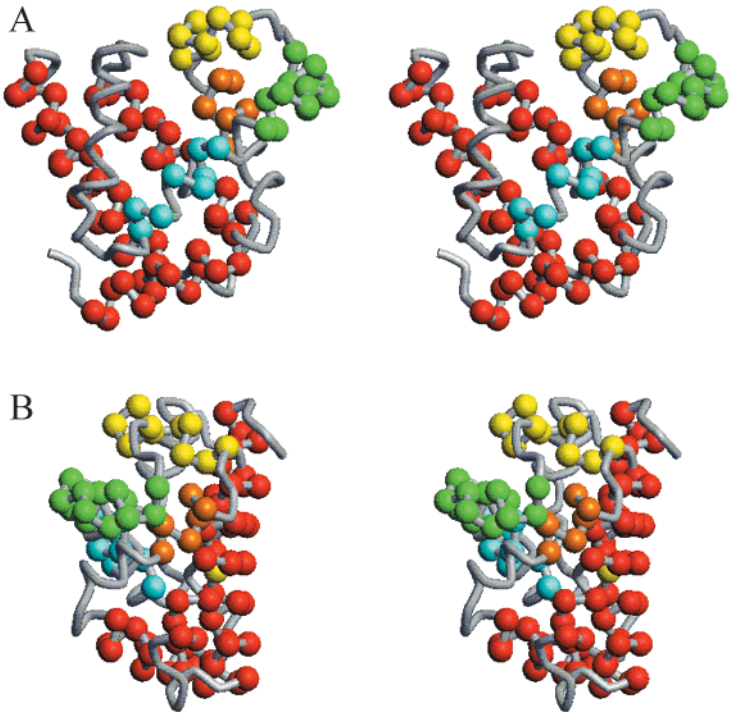

FIGURE 6: (A and B) Stereoviews from different viewpoints of the backbone trace of the crystal structure of MbCO (*51*) showing the locations of helical structure in the pH 4 state, deduced from Figure 1. Red spheres indicate the Cα positions of residues in helices A, G, and H. Orange spheres show the locations of residues in the B helix that show helical propensities in the pH 4 state, and yellow, green, and blue spheres show the locations of residues with helical propensities at pH 4 in the C−CD region, in the D−E region, and at the end of the E helix, respectively.

The F Helix. In holomyoglobin, the F helix is separated from the remainder of the protein by the heme moiety. This region of the apoprotein appears to be highly disordered at pH 4, with little indication of secondary structure from secondary chemical shifts or NOEs, no detectable intramolecular hydrogen bonding, and a high degree of backbone flexibility. In fact, the measurements presented here are very similar to those obtained for the same region in the acid unfolded state of the protein at pH 2.3 (*16*). In the folded apoprotein at pH 6.0, NMR backbone resonances corresponding to the F helix are rendered unobservable by a slow conformational exchange process (*48*). Apparently, the extensive contacts made with the heme group are necessary to stabilize the final native conformation of this region of the protein.

The Other Helices. Prior to this study, little information has been available about the state of the apomyoglobin intermediate outside the AGH[B] core. Infrared studies have suggested some "loose, solvated" helical structure which was ascribed to regions outside the core (*49*). A recent report suggests that the B through E helices (comprising all remaining regions which are fully structured in native apomyoglobin, but not at pH 4) may be considered as a second folding unit in apomyoglobin (*50*). This region clearly forms after the AGH[B] core during kinetic folding (*2*), but it is speculated (*50*) that it may constitute an independently folding nucleus (although the possibility that its formation may depend on an AGH core is also noted). The secondary shifts measured here indicate that at pH 4 this region of the protein has a significant propensity for helical conformations and, in contrast with the F helix region, is not as disordered as an unfolded polypeptide. Also evident in this region is a decrease in backbone flexibility on fast time scales, and a corresponding shift in spectral density from very fast time scales to somewhat slower time scales. Although these results do not rule out the possibility that the transient structure and restricted motions in this region originate from the nascent formation of a second independent nativelike core, a more likely explanation is that the weak intrinsic helical propensities of the C, D, and E helices and the N terminus of the B helix are stabilized in the intermediate by the nearby hydrophobic surface of the AGH[B] core. This leads to an increased population of helical conformations in these regions, which may transiently pack against the core, leading to the observed decrease in backbone flexibility. This hypothesis is supported by the observation that the indications of structure outside the core decrease with increasing distance from the core.

Helix Boundaries in the Intermediate. An interesting question regarding the folding of proteins is the extent to which secondary structure elements formed in early steps of the process resemble the final elements of the native state. The data presented here provide some of the first insights into the boundaries of secondary structure elements formed early in folding. The $^{13}C^{\alpha}$ secondary shifts alone, which have been previously reported (*16*), do not allow for unambiguous identification of helix boundaries. In contrast, the combination of four different secondary chemical shifts and the $H^{N}_{i}-H^{N}_{i+1}$ NOEs, measured on a per-residue basis, provides a sound basis for assigning putative boundaries to the observed helical segments. To pinpoint the locations of native and nonnative secondary structure, the helix boundaries for the pH 4 state, derived from Figure 1, are shown in Figure 6 mapped onto the crystal structure of MbCO (*51*). Well-populated helices A, G, and H are indicated by red spheres in Figure 6. Both the N- and C-termini of the G and H helices are

clearly evident in the secondary shift plots (Figure 1), and are within one or two residues of the boundaries observed in the native state. The N-terminus of the A helix is also clearly demarcated and coincides with that of the native protein, but the C-terminus of the A helix is not clearly defined. The ^{13}CO secondary shift of Val 17 is negative, suggesting a termination of the A helix at this residue, as in the native state. However, the ^{15}N, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ resonances of this residue were not reliably observed due to broadness of the resonance lines, preventing confirmation of this boundary from NOEs or other chemical shifts. Therefore, the possibility remains that at pH 4, the A helix extends beyond the native boundary, perhaps as far as Gly 23, where helix termination is strongly suggested in the chemical shift and NOE data. The B helix (orange spheres in Figure 6) appears to begin at His 26, in contrast with the native state starting position of Val 21, but the C-terminus of the B helix is Ser 35, the same as in the native state. Helical secondary shifts for all four nuclei suggest Glu 38 as the starting point for the (transiently populated) C helix (yellow spheres in Figure 6), as in the native state, and the termination point is most likely His 48. This is a considerably longer helical stretch than is observed in the native structure, where Lys 42 is the terminal residue.

The (transiently populated) D helix (green spheres in Figure 6) presents perhaps the most interesting case. It begins at Thr 51, as in the native state, but the first clear indication of an interruption is at His 64 or Gly 65, where ^{13}CO, H^{α}, and $^{13}C^{\beta}$ shifts show a transition, and where a $H^N_i-H^N_{i+1}$ NOE is missing. In the native protein, the D helix ends at Ala 57, and the E helix begins at Glu 59. Thus, in the intermediate, the D helix runs past the native D/E boundary and four residues into the native E helix, a conclusion which is supported by the low H^N temperature coefficients (Figure 2), which suggest that hydrogen bonding occurs for residues 57 through 65. This non-native helical segment is also populated in the acid-unfolded state of the protein at pH 2.3, as well as in synthetic peptides of the same sequence (J. Yao, D. Eliezer, J. Chung, H. J. Dyson, and P. E. Wright, unpublished data). These observations reflect the intrinsic helical propensity of this sequence, which is also predicted by the program AGADIR (*52*). It is likely that this element of non-native secondary structure has a relatively high population early in the folding process, persists through the formation of the AGH[B] intermediate, and is only later broken in the middle to form the D and E helices of the native structure. The transiently populated E helix of the pH 4 intermediate (blue spheres in Figure 6) begins around Thr 67 and extends to Leu 76 where the transition to the unfolded F helix is marked by the disappearance of $H^N_i-H^N_{i+1}$ NOEs and nonhelical secondary shifts, as well as a lower tendency for hydrogen bonding, as indicated by the temperature coefficients.

Although the helical regions outside the AGH[B] core are clearly not as highly populated as those within it, they nevertheless appear to have relatively well-defined boundaries, which in some cases are different from the boundaries observed for the corresponding helices in the final native structure, as illustrated by the hatched and solid boxes in Figure 1. It is clear that the pH 4 state contains largely nativelike helical propensities in the A, G, and H helices and part of the B helix; Figure 1 indicates that although these helices are frayed at the ends, the centers approach nativelike populations of helix. Figure 6B shows that these regions (red and orange spheres) form a domain-like half of the molecule which would be expected to be most nativelike. The remainder of the molecule (yellow, blue, and green) forms a second domain that is clearly less well-folded in the pH 4 intermediate. The lowered population of helical structure that is present in this region does not follow closely the helical boundaries found in the native folded structure. These results suggest that, while the nativelike helical structure is present and relatively stable in the A, G, and H helices as well as half of B, the other parts of the molecule are considerably more flexible. Nevertheless, helical structure, albeit rather unstable and not well-localized in native locations, is beginning to form in the intermediate: rearrangement and packing of the correct helical structure of the remainder of the molecule could then occur on a time scale of milliseconds, as observed in kinetic experiments (*2*).

CONCLUSIONS

Detailed structural information about partially folded states is essential to extending our understanding of proteins beyond the horizon of well-structured native states, into the poorly charted waters of on- and off-pathway folding intermediates, misfolded precursors to aggregation, intrinsically unstructured proteins, membrane-associated proteins, and other biologically relevant but poorly folded states. This study has demonstrated the efficacy of high-resolution solution NMR as a tool for probing such systems. The particular system studied, the pH 4 intermediate state of apomyoglobin, is known to closely resemble a major, on-pathway, early kinetic intermediate of this protein. The results provide a clear picture of the location and extent of secondary structure and of the backbone flexibility of this intermediate. This information will help to provide much needed constraints on both existing and future theoretical and computational models of protein folding.

ACKNOWLEDGMENT

We thank Linda Tennant for expert technical assistance and Jian Yao for helpful discussions.

REFERENCES

1. Ptitsyn, O. B. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 243−300, W. H. Freeman and Company, New York.
2. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
3. Colón, W., and Roder, H. (1996) *Nat. Struct. Biol. 3*, 1019−1025.
4. Raschke, T. M., and Marqusee, S. (1997) *Nat. Struct. Biol. 4*, 298−304.
5. Hosszu, L. L., Craven, C. J., Parker, M. J., Lorch, M., Spencer, J., Clarke, A. R., and Waltho, J. P. (1997) *Nat. Struct. Biol. 4*, 801−804.
6. Hyre, D. E., and Klevit, R. E. (1998) *J. Mol. Biol. 279*, 929−943.
7. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504−11509.
8. Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) *Cell 91*, 741−752.
9. Watts, A. (1995) *Biochem. Soc. Trans. 23*, 959−965.
10. Fink, A. L. (1998) *Folding Des. 3*, R9−R23.

11. Kelly, J. W. (1996) *Curr. Opin. Struct. Biol. 6*, 11−17.
12. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) *Science 249*, 1544−1548.
13. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) *Protein Sci. 8*, 45−49.
14. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O., and Tsuruta, H. (1995) *Science 270*, 487−488.
15. Kay, M. S., and Baldwin, R. L. (1996) *Nat. Struct. Biol. 3*, 439−445.
16. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
17. Jennings, P. A., Stone, M. J., and Wright, P. E. (1995) *J. Biomol. NMR 6*, 271−276.
18. Eliezer, D., Jennings, P. A., Dyson, H. J., and Wright, P. E. (1997) *FEBS Lett. 417*, 92−96.
19. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
20. Zhang, O., Forman-Kay, J. D., Shortle, D., and Kay, L. E. (1997) *J. Biomol. NMR 9*, 181−200.
21. Palmer, A. G., Skelton, N. J., Chazin, W. J., Wright, P. E., and Rance, M. (1992) *Mol. Phys. 75*, 699−711.
22. Peng, J. W., and Wagner, G. (1992) *J. Magn. Reson. 98*, 308−332.
23. Lefèvre, J. F., Dayie, K. T., Peng, J. W., and Wagner, G. (1996) *Biochemistry 35*, 2674−2686.
24. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) in *Numerical Recipes in C*, Cambridge University Press, Cambridge U.K.
25. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
26. Merutka, G., Dyson, H. J., and Wright, P. E. (1995) *J. Biomol. NMR 5*, 14−24.
27. Yao, J., Dyson, H. J., and Wright, P. E. (1997) *FEBS Lett. 419*, 285−289.
28. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
29. Löhr, F., and Rüterjans, H. (1995) *J. Biomol. NMR 6*, 189−197.
30. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson. 101*, 201−205.
31. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
32. de Dios, A. C. (1996) *Prog. NMR Spectrosc. 29*, 229−278.
33. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
34. Spera, S., and Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490−5492.
35. Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
36. Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 499−503.
37. Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) *Adv. Protein Chem. 37*, 1−106.
38. Dyson, H. J., and Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
39. Reymond, M. T., Huo, S. Q., Duggan, B., Wright, P. E., and Dyson, H. J. (1997) *Biochemistry 36*, 5234−5244.
40. Andersen, N. H., Neidigh, J. W., Harris, S. M., Lee, G. M., Liu, Z., and Hong, H. (1997) *J. Am. Chem. Soc. 119*, 8547−8561.
41. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
42. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.
43. Rischel, C., Thyberg, P., Rigler, R., and Poulsen, F. M. (1996) *J. Mol. Biol. 257*, 877−885.
44. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) *J. Biomol. NMR 6*, 153−162.
45. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
46. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
47. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) *J. Mol. Biol. 285*, 2133-2146.
48. Eliezer, D., and Wright, P. (1996) *J. Mol. Biol. 263*, 531−538.
49. Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1997) *Biochemistry 36*, 15006−15012.
50. Gilmanshin, R., Dyer, R. B., and Callender, R. H. (1997) *Protein Sci. 6*, 2134−2142.
51. Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G.A. (1986) *J. Mol. Biol. 192*, 133−154.
52. Muñoz, V., and Serrano, L. (1997) *Biopolymers 41*, 495−509.

BI992545F