NMR Structural and Dynamic Characterization of the Acid-Unfolded State of Apomyoglobin Provides Insights into the Early Events in Protein Folding[†,‡]

Jian Yao,[§] John Chung, David Eliezer,[||] Peter E. Wright,* and H. Jane Dyson*

Department of Molecular Biology MB2, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received December 6, 2000

ABSTRACT: Apomyoglobin forms a denatured state under low-salt conditions at pH 2.3. The conformational propensities and polypeptide backbone dynamics of this state have been characterized by NMR. Nearly complete backbone and some side chain resonance assignments have been obtained, using a triple-resonance assignment strategy tailored to low protein concentration (0.2 mM) and poor chemical shift dispersion. An estimate of the population and location of residual secondary structure has been made by examining deviations of $^{13}C^{\alpha}$, ^{13}CO, and $^{1}H^{\alpha}$ chemical shifts from random coil values, scalar $^{3}J_{HN,H\alpha}$ coupling constants and ^{1}H-^{1}H NOEs. Chemical shifts constitute a highly reliable indicator of secondary structural preferences, provided the appropriate random coil chemical shift references are used, but in the case of acid-unfolded apomyoglobin, $^{3}J_{HN,H\alpha}$ coupling constants are poor diagnostics of secondary structure formation. Substantial populations of helical structure, in dynamic equilibrium with unfolded states, are formed in regions corresponding to the A and H helices of the folded protein. In addition, the deviation of the chemical shifts from random coil values indicates the presence of helical structure encompassing the D helix and extending into the first turn of the E helix. The polypeptide backbone dynamics of acid-unfolded apomyoglobin have been investigated using reduced spectral density function analysis of ^{15}N relaxation data. The spectral density $J(\omega_N)$ is particularly sensitive to variations in backbone fluctuations on the picosecond to nanosecond time scale. The central region of the polypeptide spanning the C-terminal half of the E helix, the EF turn, and the F helix behaves as a free-flight random coil chain, but there is evidence from $J(\omega_N)$ of restricted motions on the picosecond to nanosecond time scale in the A and H helix regions where there is a propensity to populate helical secondary structure in the acid-unfolded state. Backbone fluctuations are also restricted in parts of the B and G helices due to formation of local hydrophobic clusters. Regions of restricted backbone flexibility are generally associated with large buried surface area. A significant increase in $J(0)$ is observed for the NH resonances of some residues located in the A and G helices of the folded protein and is associated with fluctuations on a microsecond to millisecond time scale that probably arise from transient contacts between these distant regions of the polypeptide chain. Our results indicate that the equilibrium unfolded state of apomyoglobin formed at pH 2.3 is an excellent model for the events that are expected to occur in the earliest stages of protein folding, providing insights into the regions of the polypeptide that spontaneously undergo local hydrophobic collapse and sample nativelike secondary structure.

There has been a great deal of recent interest in the unfolded states of proteins, both because of their relevance for understanding protein folding and because it is now evident that many proteins are intrinsically unstructured in their biologically functional states (*1*). Due to the disordered nature of the unfolded states of proteins, they are unsuitable for study by X-ray crystallography. However, high-resolution NMR experiments can provide detailed insights into the structure and dynamics of unfolded states, and studies of several proteins have been reported, including lysozyme (*2*−*4*), 434 repressor (*5*), FK506-binding protein (*6*), BPTI (*7*), protein G, B domain (*8*), barnase (*9*), an SH3 domain (*10*, *11*), and staphylococcal nuclease (*12*, *13*).

Experimental difficulties with NMR studies of unfolded proteins include the high degree of overlap of resonances, particularly those of protons, and the tendency of unfolded proteins to aggregate in solution. At low pH, there is also the possibility of chemical degradation of the protein through acid hydrolysis. The problem of resonance dispersion is approached by labeling of the protein with ^{15}N, ^{13}C, and ^{2}H and utilizing the dispersion of the backbone ^{15}N and ^{13}CO resonances, the chemical shifts of which are determined primarily on the local amino acid sequence (*14*). In the present work, we have utilized a triple-resonance strategy,

[†] This work was supported by Grant GM57374 from the National Institutes of Health.

[‡] The resonance assignments for the ^{1}H, ^{15}N, and ^{13}C chemical shifts for apomyoglobin at pH 2.3 have been deposited in the BioMagResBank (http://www.brmb.wisc.edu) under BMRB accession number 4676.

* Corresponding authors. Tel. (Wright): 858 784-9721 or (Dyson): 858 784-2223. Fax: 858 784-9822. Email: wright@scripps.edu or dyson@scripps.edu.

[§] Present address: Genencor International, 924 Page Mill Rd., Palo Alto, CA 94304.

[||] Present address: Department of Biochemistry, Weill Medical College of Cornell University, 1300 York Ave., New York, NY 10021.

10.1021/bi002776i CCC: $20.00
Published on Web 02/24/2001

which exploits the intrinsic resonance dispersion of the ^{15}N and ^{13}CO resonances, to obtain almost complete backbone resonance assignments for acid-unfolded apomyoglobin. Chemical shifts and ^{15}N relaxation rates are then used to characterize the structural propensities and backbone dynamics of this unfolded state.

Apomyoglobin has been the subject of extensive study in a number of laboratories, and is a paradigm for studies of protein folding pathways. The protein folds by way of an on-pathway kinetic intermediate (*15*, *16*). Equilibrium measurements by a variety of methods have shown formation of a molten globule intermediate at pH 4 (*17*, *18*). Circular dichroism measurements indicate that apomyoglobin, while largely unfolded at pH ~2, retains a small amount of residual helical structure that is abolished upon addition of denaturants (*17*). The intermediate state at pH 4 was shown by amide exchange trapping to contain helical structure in the A, G, and H helices (*19*), and quench-flow hydrogen exchange measurements together with stopped-flow CD gave evidence that the same intermediate is formed on the kinetic folding pathway (*20*). The structure and dynamics of the equilibrium intermediate have been probed by an extensive series of NMR experiments (*21*), and a brief report comparing the behavior of the unfolded, intermediate, and native states has been published (*22*). In this paper, we report detailed characterization of the structure and dynamics of the acid-unfolded state of apomyoglobin by NMR.

MATERIALS AND METHODS

Protein Production. Sperm whale apomyoglobin uniformly labeled with ^{15}N and ^{13}C was expressed and purified by a method similar to that described previously (*23*) except that $FeCl_3$, $ZnSO_4$, and thiamin were omitted from the growth medium since they have no influence on the yield of apoMb, and the post-induction temperature was maintained at 37 °C. Solubilization of inclusion bodies was accomplished by resuspension in a solvent mixture containing 85% CH_3CN, 0.1% TFA, and 10mM DTT. Solubilization was enhanced by sonication and repeated washing. After centrifugation to remove insoluble material, the supernatant was diluted with deionized water and applied directly to a preparative reverse-phase HPLC column. The fractions containing apomyoglobin were pooled and lyophilized.

Peptide Characterization. The peptide Ac-TEAEMKASED-LKK-NH_2 was synthesized by standard solid-phase techniques and purified by reverse-phase HPLC. CD spectra were recorded at 4 °C on an Aviv 61DS spectropolarimeter calibrated using a standard solution of *d*-camphorsulfonic acid. Peptide concentrations were determined by quantitative amino acid analysis, and spectra were recorded at pH 5.1 and pH 2.3. The ^1H NMR spectra of the peptide were completely assigned using standard 2D homonuclear NMR experiments.

NMR Sample Preparation. NMR samples were prepared by dissolving lyophilized protein directly in a solution of 10 mM acetic acid-d_4 in 95% ^1H_2O/5% ^2H_2O, for which the pH of the solution had been adjusted to 2.3 using HCl. The final HCl concentration was ~5 mM. To ensure that the protein solution contained no extraneous salt and was present at the correct pH, the solution was passed twice through a Sephadex G-25 spin column and subsequently diluted to the appropriate protein concentration (200 μM) with the acetate−HCl solution, pH 2.3. Protein concentration was determined by UV absorbance at 280 nm, using a molar extinction coefficient of $\epsilon_{280} = 15\,900$ (*24*).

NMR Spectroscopy. All NMR spectra were acquired on a Bruker DMX spectrometer operating at a proton frequency of 750 MHz. All heteronuclear NMR experiments were performed on uniformly ^{15}N-labeled or ^{15}N,^{13}C-doubly-labeled samples. All NMR experiments employed pulsed field gradients and gradient-selected sensitivity enhancement. The ^1H carrier was set on the water resonance and was shifted to 6.9 ppm during detection. The ^1H spectral width was usually set at 3894.08 Hz. The ^{15}N carrier was set at 117 ppm and that for $^{13}C^{\alpha}$ at 55 ppm. When both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ were to be excited, the carrier was centered at 42 ppm, and at 38 ppm for excitation of side chain ^{13}C resonances. For ^{13}CO, the carrier was centered at 176 ppm. These carrier positions were optimized for the unfolded protein and are slightly different from those normally used for folded proteins. When the ^{13}C channel was to be used for pulses other than at the carrier frequency, shifted laminar pulses (*25*) were used with an excitation null at the carrier frequency. Water suppression was usually achieved by water flip-back pulses (*26*) and gradient coherence selection (*27*). The proton chemical shifts were referenced with respect to internal dioxane at 3.75 ppm (*28*). The chemical shifts for nitrogen and carbon were indirectly referenced (*28*). Unless otherwise indicated, all experiments were performed at 25 °C.

2D ^1H-^{15}N HSQC spectra (*29*) were acquired with 512 (t_2) and 64 (t_1) complex points. Spectral widths were 3894 (^1H) and 2281 Hz (^{15}N). Coherence selection with gradients, water flip-back pulses, and sensitivity enhancements were employed. For the ^{13}C,^{15}N-doubly-labeled sample, ^{13}C decoupling was achieved by a π pulse on the carbonyl and a WALTZ-16 decoupling pulse on the α carbons. Pulsed field gradient constant time 3D HNCO data were acquired (*27*) with 512, 32, and 64 complex points in the t_3, t_2, and t_1 dimensions, respectively, with spectral widths of 3894 (^1H), 2281 (^{15}N), and 1132 Hz (^{13}CO). Proton decoupling was achieved with a WALTZ-16 composite pulse at a field strength of 5555 Hz. Coherence selection with gradients, water flip-back pulse, and sensitivity enhancement were employed. The constant time period during ^{15}N evolution and the delay during coherence transfer from ^{15}N to ^{13}CO were both set to 24.8 ms.

Pulsed field gradient 3D CBCA(CO)NH (*26*, *30*) and 3D HNCACB (*31*) spectra were acquired with spectral widths of 4194 (HN), 1900 (^{15}N), and 12 255 Hz ($C^{\alpha\beta}$). A total of 512 (t_3), 34 (t_2), and 54 (t_1) complex points were acquired for CBCA(CO)NH, and 512 (t_3), 42 (t_2), and 64 (t_1) complex points were acquired for HNCACB. Coherence selection with gradients, water flip-back pulse, and sensitivity enhancement were employed. Pulsed field gradient constant time 3D HNCA (*32*) spectra were acquired with 512 (HN), 45 (^{15}N), and 124 (C^{α}) complex points and spectral widths of 3894 (t_3), 2534 (t_2), and 4529 Hz (t_1). The constant time evolution period was set to 26.6 ms. Proton decoupling was achieved by a composite WALTZ-16 pulse at a field strength of 5814 Hz. The 3D(HCA)CO(CA)NH spectra were acquired as described (*33*), except that pulsed field gradients were added for coherence selection and sensitivity enhancement. The

spectral widths were 4195 (HN), 2281 (^{15}N), and 1250 Hz (^{13}CO) with 512, 34, and 30 complex points in the three dimensions, respectively. Acquisition times were 122.1, 14.9, and 24 ms, respectively. Proton decoupling was achieved by a WALTZ-16 scheme with an rf field of 5814 Hz, applied continuously without the spin lock pulses since water suppression was achieved by coherence selection with gradients.

The 3D C(CO)NH-TOCSY spectrum was acquired with pulsed field gradient coherence selection and sensitivity enhancement. The spectral widths were 3894 (HN), 1900.4 (^{15}N), and 12 255 Hz (^{13}C) with 512, 40, and 64 complex points in the respective dimensions and acquisition times of 131.5, 21, and 5.2 ms, respectively. ^{13}C isotropic mixing was achieved with two cycles of DIPSI-3 (*34*) and a mixing time of 13.04 ms at a field strength of 8.3 kHz. The 3D ^{15}N-edited TOCSY−HSQC (*29*) spectra were acquired with spectral widths of 4496.4, 2281, and 6751 Hz with 512, 38, and 128 complex points in the t_3 (HN), t_2 (^{15}N), and t_1 (^{1}H) dimensions, respectively. ^{1}H isotropic mixing was achieved with 14 cycles of DIPSI-2RC (*35*) with a total mixing time of 58 ms at a field strength of 10.4 kHz.

A 3D ^{15}N-edited PFG NOESY−HSQC spectrum (*29*) was acquired with spectral widths of 4496.4, 2281, and 6751 Hz with 512, 32, and 256 complex points in the t_3 (HN), t_2 (^{15}N), and t_1 (^{1}H) dimensions, respectively, and with a mixing time of 300 ms. The 3D HSQC−NOESY−HSQC spectra (*36*−*38*) were acquired with spectral widths of 3754, 2281, and 2281 Hz and 256, 64, and with 128 complex points in the t_3 (HN), t_2 (^{15}N), and t_1 (^{15}N) dimensions, respectively, with a mixing time of 245 ms. A 3D water flip-back HNHA experiment (*39*) was carried out with spectral widths of 3894, 2281, and 6751 Hz and with 512, 87, and 80 complex points in the t_3 (HN), t_2 (^{15}N), and t_1 (^{1}H) dimensions, respectively.

Spectral Analysis and Processing. Spectral processing was performed on Silicon Graphics Indigo2 or Indy workstations using Felix 95.0. Time domain data in the ^{15}N and ^{13}C dimensions were linear predicted to increase the number of data points by a third and zero-filled to double the original data size before Fourier transformation. Gaussian window functions were applied in all dimensions.

Temperature Coefficients and Hydrogen Exchange. Amide proton exchange rates were measured from a series of ^{1}H-^{15}N HSQC spectra recorded at 5 °C immediately following exchange of apomyoglobin into D_2O containing 10 mM acetate at pH* 2.3 using a spin column. A total of 54 HSQC spectra were recorded (acquisition time 12.5 min) over a period of 42 h. Amide proton temperature coefficients were determined from a series of ^{1}H-^{15}N HSQC spectra of apomyoglobin at pH 2.3 recorded over the temperature range 3−25 °C.

Measurements of Relaxation Parameters. All relaxation measurements were performed at 750 MHz at a temperature of 25 °C, with the probe temperature calibrated using methanol (*40*). The pulse sequences used to obtain T_1, T_2, and heteronuclear NOE spectra are those described by Farrow et al. (*41*), which employ gradient selection and sensitivity enhancement, as well as minimal water suppression. Typically, 10 time points were collected for the T_1 and T_2 experiments, and each set included at least 2 duplicate measurements to allow estimation of the uncertainty. {^{1}H}-^{15}N heteronuclear NOE spectra were recorded at 750 MHz; each spectrum was measured 4 times. In the direct dimension (^{1}H), the carrier frequency was set on the water resonance with a spectral width of 5.2 ppm; the indirect (^{15}N) dimension was centered at 116.6 ppm with a 30 ppm spectral width. A total of 512 complex data points with 128 complex increments were collected. Typical T_1 experiments were acquired with 16 scans and a repetition delay of 2.5 s, while T_2 experiments were acquired with 16 scans and a 1 ms delay between 180 pulses in the Carr−Purcell−Meiboom−Gill (CPMG) pulse train. Each {^{1}H}-^{15}N NOE experiment (both saturated and nonsaturated pairs) was acquired with 28 scans. The ^{1}H saturation was achieved by the application of 120° ^{1}H pulses separated by 18 ms, for a 3.6 s period. In addition, a 1.54 s recycle delay was employed.

RESULTS

Optimization of Sample Conditions Used for NMR Spectroscopy. Considerable effort was required to find reproducible solution conditions for 3D NMR studies of acid-unfolded apomyoglobin. Solution conditions had to be found to optimize viscosity, sensitivity, and sample stability. The quality of NMR spectra of acid-unfolded apomyoglobin depends critically on the sample concentration and experimental temperature. A series of HSQC spectra at different sample concentrations revealed the loss of a number of cross-peaks at concentrations above 500 μM at 25 °C. There is no precipitation or cloudiness at protein concentrations above 1 mM, but the sample becomes gel-like above a concentration of 2 mM. The present study was conducted using samples of acid-unfolded apomyoglobin at a concentration of 200 μM and 25 °C, which gave the maximum number of resolved peaks.

Another important factor to consider in determining optimum experimental conditions is the lifetime of the sample. At the temperature used in this study, the lifetime of acid-unfolded apomyoglobin is about 1 week. After a prolonged period of time at pH 2, apomyoglobin undergoes chemical degradation, which results in new peaks in the HSQC spectra. The degradation is more severe at higher temperatures. Control experiments, in which the samples were sealed in NMR tubes and protease inhibitors were added, indicate that the degradation is caused simply by the acidic conditions, rather than by oxidation or protease digestion. HPLC analysis of the degradation products indicates that they are of low molecular weight, probably short peptides or single amino acids.

Resonance Assignments. The ^{15}N-^{1}H HSQC spectrum of unfolded apomyoglobin at 25 °C and pH 2.3 is shown in Figure 1. The chemical shift dispersion of unfolded proteins is significantly smaller than that of folded proteins (*14*). For unfolded apomyoglobin, the chemical shifts of $^1H^N$ are within a range of only 1 ppm, and those of ^{15}N are within a range of 20 ppm. At 750 MHz, 135 backbone ^{1}H-^{15}N cross-peaks were distinguishable, out of a total of 149. The strategy used for assignment of the acid-unfolded form of apomyoglobin utilized the superior intrinsic chemical shift dispersion of the ^{13}CO resonances in unfolded proteins (*14*). More than half of the resonance assignments could be made directly through the sequential connectivities observed in the (HCA)-CO(CA)NH spectrum. Figure 2 shows representative strips and sequential connectivities. The HNCO spectrum was used

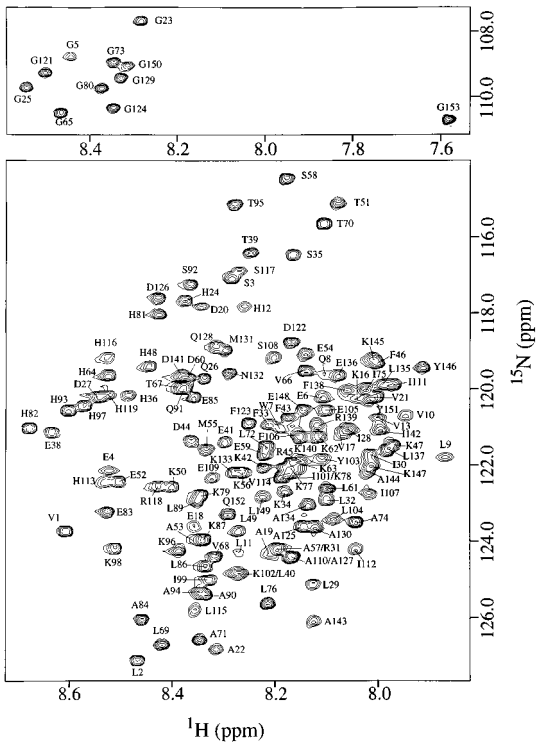

FIGURE 1: 750 MHz ^{1}H-^{15}N HSQC spectrum of 0.2 mM acid-unfolded apomyoglobin, uniformly labeled with ^{15}N, in 95% H_2O/5% $^{2}H_2O$, 10 mM acetic acid-d_6/HCl at pH 2.3 and 25 °C. All residues are assigned except Trp14 and Ala15, which are broadened beyond detection. The upper panel shows the region containing all the glycine residues. The lower panel shows the region of other backbone amide groups. The assignments are given with the one-letter code for amino acids and the residue number in the primary sequence. Residues assigned to an overlapped peak are separated by a slash (/).

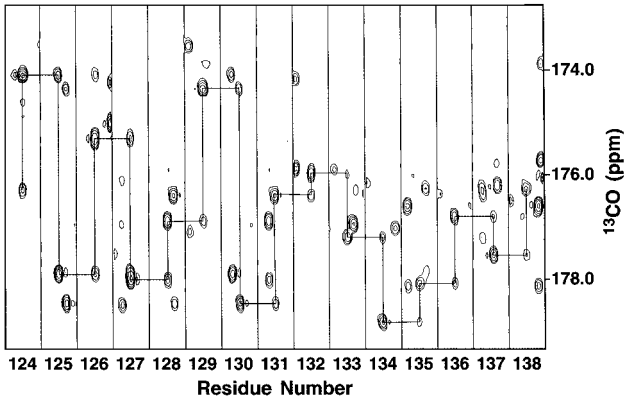

FIGURE 2: Strips of the 3D (HCA)CO(CA)NH spectrum of acid-unfolded apomyoglobin for residues 124 through 138, the first half of the H helix. Sequential connectivities are shown as connected lines. Experimental conditions are the same as in Figure 1. In each strip, the intraresidue peak is slightly stronger than the interresidue peak.

to distinguish interresidue peaks from intraresidue peaks in the (HCA)CO(CA)NH spectrum. TOCSY−HSQC, CBCA(CO)NH, and C(CO)NH-TOCSY spectra were then used to confirm the assignments and resolve ambiguities. The remainder of the sequential assignments were made using HNCACB and CBCA(CO)NH spectra. The assignments have been deposited in BioMagResBank.

The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances are shifted only slightly from their random coil values. Although this creates problems for

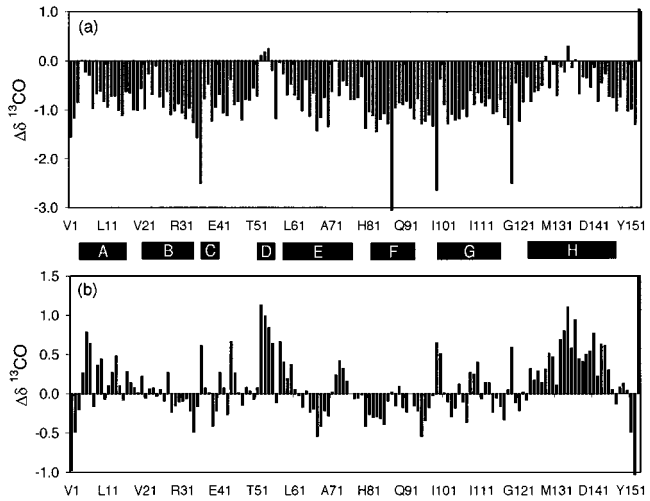

FIGURE 3: Chemical shift deviations from random coil values for ^{13}CO resonances of acid-unfolded apomyoglobin. (a) Secondary ^{13}CO shifts obtained by subtracting random coil shifts, uncorrected for sequence-dependent variations, from the experimental ^{13}CO chemical shifts. (b) Corrected ^{13}CO secondary shifts (see text). The positions of helices in the folded structure of myoglobin are indicated by black bars.

sequential assignments using standard C^{α}-based methods, it is an advantage in that residue type can be determined with reasonable accuracy simply by observing the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ peak positions. Further, the CBCA(CO)NH spectrum gives $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ assignments for the preceding residue, which can be used to identify dipeptide fragments within the protein sequence. The residue-specific assignments for lysine and histidine residues were confirmed by recording ^{15}N HSQC spectra of samples ^{15}N-labeled at either lysine or histidine residues. Complete assignments for the backbone nuclei ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, ^{13}CO) and $^{13}C^{\beta}$ were obtained except for two residues, Trp14 and Ala15, whose ^{1}H, ^{15}N resonances were broadened beyond detection. Sequence-specific assignments are indicated in Figure 1 for all assigned ^{1}H-^{15}N pairs.

Deviation of Chemical Shifts from Random Coil Values. The chemical shift deviations from random coil values (secondary shifts) (*42*) in acid-unfolded apomyoglobin are shown in Figures3 and 4. Secondary shifts were calculated by subtracting random coil values determined for Ac-GGXGG-NH_2 peptides in 8 M urea at pH 2.3 (*43*) from the measured chemical shifts for apoMb at pH 2.3. Since the chemical shifts of certain nuclei are influenced both by the local backbone structure and by the nature of the neighboring amino acids in the sequence, it is important to correct for local sequence-dependent contributions if the secondary shifts are to be used to probe backbone conformational propensities. This was accomplished using sequence-dependent correction factors determined for the set of Ac-GGXGG-NH_2 peptides (*59*); the random coil shifts were corrected for the influence of the two amino acids immediately preceding and the two amino acids immediately following each residue. For example, the random coil reference shift for W7 (in the sequence GEWQL) was corrected for the influence of the G5, E6, Q8, and L9 residues. The dataset of Schwarzinger et al. (*43*) does not provide residue-specific correction factors for X-Pro sequences (where X is any amino acid except Gly): a $^{13}C^{\alpha}$ correction factor of 2.0 ppm [the average correction factor determined by Wishart (*44*) for non-glycine residues] was therefore applied to residues immediately

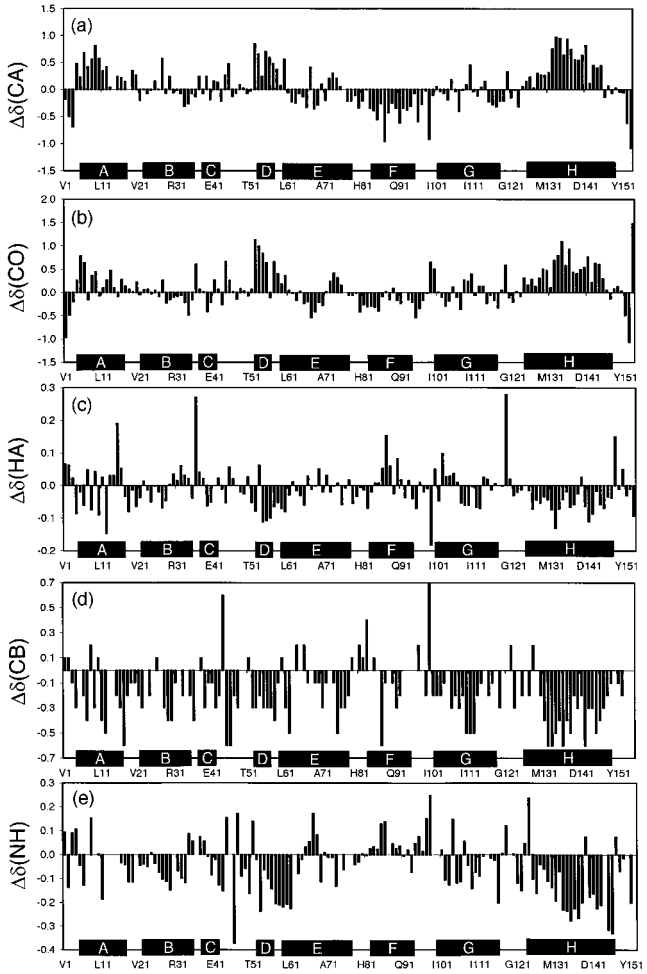

FIGURE 4: Secondary chemical shifts, corrected for sequence-dependent contributions, for (a) $^{13}C^{\alpha}$, (b) ^{13}CO, (c) $^{1}H^{\alpha}$, (d) $^{13}C^{\beta}$, and (e) $^{1}H^{N}$ resonances of acid-unfolded apomyoglobin. Regions corresponding to the helices of the native protein are marked with black bars. Sequence-dependent corrections were made as described in the text using the correction factors of Schwarzinger et al. (*59*). Reference random coil shifts were those determined for the peptide Ac-GGXGG-NH_2 in 8 M urea at pH 2.3 (*43*).

preceding proline. The effect of the sequence-dependent corrections on the secondary ^{13}CO shifts is illustrated in Figure 3. Without correction (Figure 3a), it is difficult to discern variations in backbone conformational propensities. However, in the sequence-corrected data (Figure 3b), regions that populate helical secondary structure are easily identified as having a consistent pattern of downfield secondary shifts (e.g., residues 124−145, which constitute the H helix region in holomyoglobin).

The sequence-corrected secondary chemical shifts of the $^{13}C^{\alpha}$, ^{13}CO, $^{1}H^{\alpha}$, $^{13}C^{\beta}$, and $^{1}H^{N}$ resonances of apomyoglobin are shown in Figure 4. The consistent pattern of downfield $^{13}C^{\alpha}$ shifts observed for residues corresponding to the A, H, and D/E helix regions of the folded protein (Figure 4a) clearly indicates that these segments of the polypeptide chain sample a significant population of helix even in the pH 2.3 unfolded state. Formation of helical structure in these regions is confirmed by the observation of matching patterns of downfield-shifted ^{13}CO resonances (Figure 4b). The secondary chemical shifts for $^{1}H^{\alpha}$ are much smaller and are therefore less sensitive to small populations of secondary structure in disordered polypeptides. In the A helix region, they are also

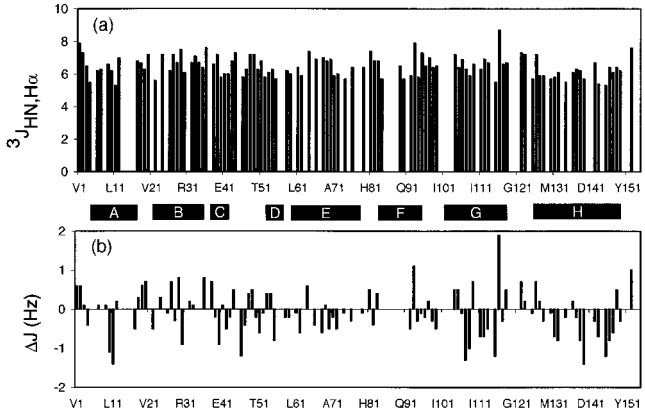

FIGURE 5: (a) Coupling constants $^{3}J_{HN,H\alpha}$ of acid-unfolded apomyoglobin measured from the 3D HNHA spectrum. (b) Deviations of $^{3}J_{HN,H\alpha}$ coupling constants from statistical random coil values, corrected for residue type and the nature of the preceding residue using the ALL random coil values (*47*, *48*). Regions corresponding to the helices of the native protein are marked with black bars.

influenced by the two tryptophan residues. Nevertheless, the $^{1}H^{\alpha}$ resonances are shifted slightly upfield in the H and D/E regions, reflecting formation of a small population of helical structure (Figure 4c). The $^{13}C^{\beta}$ secondary shifts (Figure 4d) are relatively poor indicators of backbone conformational propensities for apomyoglobin at pH 2.3, although they do show the expected pattern of upfield shifts in the H helix region. Unlike the significant shifts found in β-sheet structure, $^{13}C^{\beta}$ resonances are shifted only slightly from random coil positions in helical structure (*45*), which explains their poor diagnostic capabilities for the acid-unfolded state of apomyoglobin. The uncorrected ^{15}N and $^{1}H^{N}$ secondary chemical shifts were also relatively poor indicators of secondary structural propensities, presumably because of their sensitivity to differences in solvent and temperature between the apomyoglobin samples and the Ac-GGXGG-NH_2 reference peptides. Nevertheless, both nuclei show characteristic patterns of upfield-shifted resonances in the regions corresponding to the A and H helices and, for NH, the structured region of the D/E helix. After correction of the $^{1}H^{N}$ secondary shifts, using the measured temperature coefficients for amide proton resonances, for the 5 °C temperature difference between the random coil data and the data for apomyoglobin at pH 2.3, a much clearer indication of helical secondary structure in the H and D/E helix regions was obtained (Figure 4e). Overlap of resonances prevented measurement of temperature coefficients for many of the resonances in the A helix region so that temperature-compensated $^{1}H^{N}$ secondary shifts could not be determined for this region.

Backbone $^{3}J_{HN,H\alpha}$ Coupling Constants. The $^{3}J_{HN,H\alpha}$ coupling constants can in principle provide additional insights into backbone conformational preferences. In disordered proteins, however, conformational fluctuations average the coupling constants and make them insensitive to local structural preferences (*46*). In addition, corrections for residue type and for the nature of the preceding residue are necessary (*47*, *48*). Figure 5 shows the $^{3}J_{HN,H\alpha}$ scalar coupling constants measured from 3D HNHA spectra for the resolved peaks in the HSQC spectrum, together with the differences from statistical random coil coupling constants (*47*). These random coil *J* values were corrected for sequence context using the factors determined by Penkett et al. (*48*) for residues preceded

FIGURE 6: Summary of homonuclear NOEs observed in 0.2 mM acid-unfolded apomyoglobin. The thickness of bars indicates the relative intensities of the NOE peak observed in 3D NOESY−HSQC or 3D HSQC−NOESY−HSQC. Regions where no NOEs were observed are shown without bars. Hatched bars indicate overlapped or ambiguous cross-peaks.

by aromatic/β-branched side chains (class L) or preceded by other residue types (class S). While J values smaller than the expected random coil values are observed in some regions of the polypeptide, inspection of Figure 5 shows that the coupling constants are clearly a much poorer diagnostic of backbone secondary structural propensities than are secondary chemical shifts.

NOE Measurements. If the population of a conformation with defined secondary structure is above a certain threshold, characteristic NOEs can be observed. Sequential $d_{\alpha N}(i,i+1)$ NOEs are observed in virtually all unstructured states, since random coil proteins populate the β-region of ϕ,ψ space (*46*). Sequential $d_{NN}(i,i+1)$ NOEs are relatively strong in α-helical conformations and very weak in β-strands. Information on dihedral angle preferences in conformationally averaged states such as those found for small linear peptides can be derived from the relative intensities of these sequential NOE connectivities (*46*). Two NOE-based 3D experiments, 3D NOESY−HSQC and 3D HSQC−NOESY−HSQC, were performed. Figure 6 shows the observed NOEs in acid-unfolded apomyoglobin. Sequential $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ NOEs are observed throughout the sequence. Most of the NOE cross-peaks shown were obtained from the more sensitive NOESY−HSQC spectrum. Of these cross-peaks, only the most intense peaks were also observed in the less sensitive HSQC−NOESY−HSQC spectrum, which was used mainly to resolve $d_{NN}(i,i+1)$ NOEs where the NH resonances were overlapping. No medium- or long-range NOEs could be discerned above the noise level.

Polypeptide Chain Dynamics of Apomyoglobin at pH 2.3. Relaxation rates measured for the nuclei in acid-unfolded apomyoglobin can be used to obtain insights into the dynamics of the polypeptide chain. The relaxation rates R_1 ($=1/T_1$) and R_2 ($=1/T_2$) and the heteronuclear {^{1}H}-^{15}N NOE have been measured for all nonoverlapped cross-peaks in the ^{1}H-^{15}N HSQC spectrum, and are shown in Figure 7. The largest values of R_1 occur in the A helix and H helix regions, the smallest values at the N- and C-termini (Figure 7a). Below-average R_1 values are also observed in the center of the sequence in the region corresponding to the F helix. The value of R_2 is fairly uniform through much of the sequence, but significantly larger values are observed in the region corresponding to the A helix and, to a lesser extent, to the

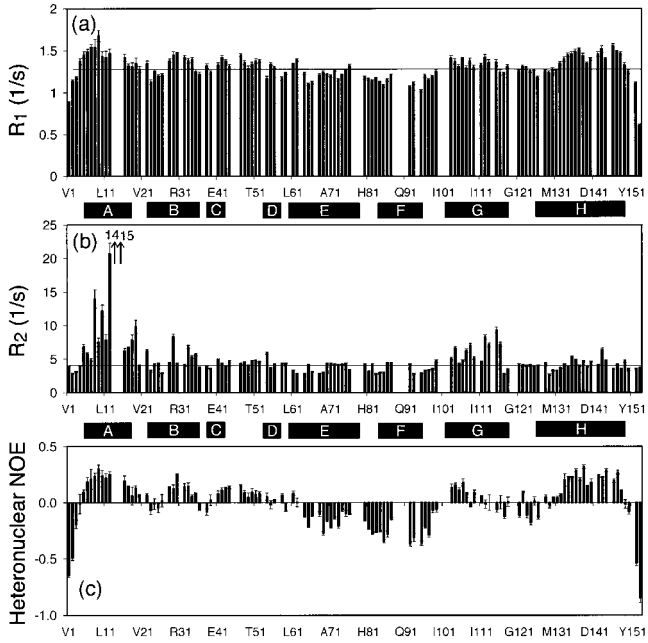

FIGURE 7: (a) R_1 relaxation rates, (b) R_2 relaxation rates, and (c) {^{1}H}-^{15}N heteronuclear NOEs for acid-unfolded apoMb at pH 2.3 and 25 °C. The horizontal lines in (a) and (b) indicate the mean R_1 (1.30 s^{-1}) and the 10% trimmed mean R_2 (4.5 s^{-1}). Black bars indicate the positions of the helices in the folded structure of myoglobin.

G helix (Figure 7b). The {^{1}H}-^{15}N NOE also shows pronounced sequence dependence (Figure 7c).

Analysis of the relaxation data using the model-free approach (*49*) is inappropriate for unfolded apomyoglobin, because the underlying assumption that the molecule tumbles isotropically with a single correlation time τ_m is unlikely to be valid. Thus, the relaxation data have been analyzed using the reduced spectral density method (*50*) to obtain insights into motions on various time scales. Values for the functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ are plotted on a per-residue basis in Figure 8.

Conformational Preferences of a D/E Peptide. In a previous study of isolated peptide fragments of myoglobin in aqueous solution, no significant helicity was observed for a peptide (Ac-TEAEMKA-NH_2) corresponding to the D helix (residues 51−57) of the folded protein (*51*). Consequently, it was somewhat surprising to observe helical structure in the D helix and start of the E helix in the acid-unfolded myoglobin. We therefore synthesized a longer peptide, spanning residues 51−63, of sequence Ac-TEAEMKASED-LKK-NH_2 and examined its conformational propensities by circular dichroism and by NMR (data not shown). This peptide spontaneously adopts helical structure in water solution, to a far greater extent than the shorter peptide representing the D helix alone. From the molar ellipticity ($\theta_{222} = -9050$ deg·cm^2·dmol^{-1} at 4 °C, pH 2.3), the average population of helix in the peptide is estimated to be 25%. NMR confirmed the presence of helix; the population estimated from the deviations of the $^1H^\alpha$ chemical shifts from random coil values was 40%.

Amide Proton Exchange Rates. Backbone amide proton exchange rates and temperature coefficients were measured for unfolded apomyoglobin at pH*2.3 and at 5 °C. Exchange rates were measurable for 83 out of the 149 amide protons; the cross-peaks of the remaining amide protons in the HSQC

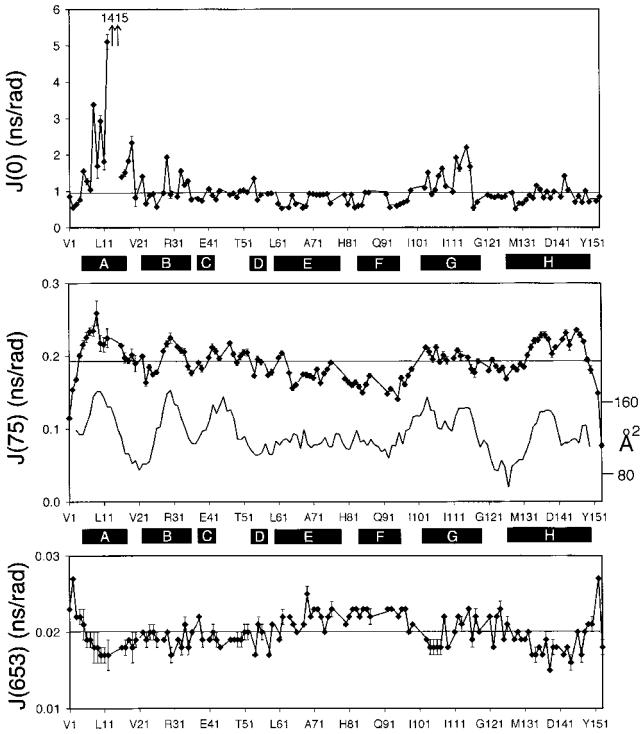

FIGURE 8: Calculated values of (a) $J(0)$, (b) $J(75)$, and (c) $J(653)$ for acid-unfolded apoMb, pH 2.3, 25 °C. The horizontal lines show the mean values of $J(75)$ (0.19 ns·rad^{-1}) and $J(653)$ (0.020 ns·rad^{-1}), and the 10% trimmed mean value of $J(0)$ (0.97 ns·rad^{-1}). (Inset) Average buried surface area, calculated using the values of Rose et al. (*53*) and averaged over a 7 residue window, is also shown in (b) (solid line, no data points, right-hand scale). Black bars indicate the positions of the helices in the folded structure of myoglobin.

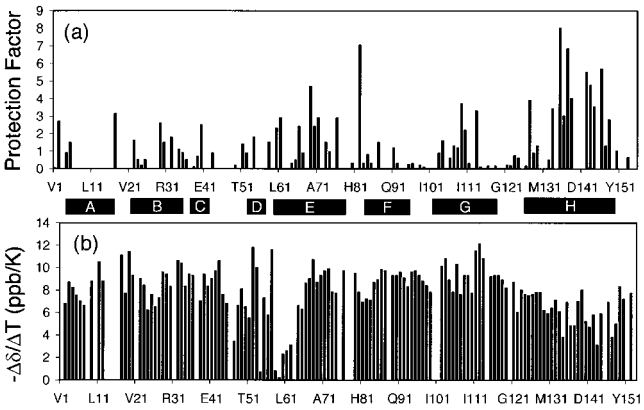

FIGURE 9: (a) Plot of the per-residue amide proton hydrogen exchange protection factors measured for apomyoglobin at pH 2.3, following buffer exchange into deuterated buffer. Protection factors were calculated using corrections for nearest-neighbor effects from the data of Bai et al. (*57*). (b) Temperature coefficients (ppb/K) measured for amide protons in apomyoglobin at pH 2.3. Black bars indicate the positions of the helices in the folded structure of myoglobin.

spectrum are either too weak or overlapped. Protection factors, defined as the ratio of the experimentally measured rates to the rates of the same amide proton in a random coil conformation, for these resonances are shown in Figure 9a. Although the protection factors are small, they are significantly and consistently larger in the H helix region than in other segments of the polypeptide. The highest protection factor is found for residue 135, in the middle of the H helix region. Amide proton temperature coefficients are also decreased significantly in a contiguous region that corresponds to the H helix in the folded protein (Figure 9b).

DISCUSSION

Residual Structure in Apomyoglobin at pH 2.3. The deviations of the $^{13}C^{\alpha}$, ^{13}CO, and, to a lesser extent, $^{1}H^{\alpha}$ chemical shifts from random coil values are highly sensitive to the conformational preferences of the polypeptide backbone. However, these secondary shifts are small in unfolded states of proteins. If backbone conformational propensities of unfolded proteins are to be determined reliably, it is essential to use appropriate random coil shift reference values and to make sequence-dependent corrections to the random coil shifts. Thus, in the case of acid-unfolded apomyoglobin, the appropriate chemical shift references are the random coil shifts determined for the peptides Ac-GGXGG-NH_2 at pH 2.3 in 8 M urea (*43*), systematically corrected for local sequence effects according to the method of Schwarzinger et al. (*59*). To give an example, uncorrected $^{13}C^{\alpha}$ secondary shifts, calculated from a different set of random coil values [derived from peptide GGXAGG at pH ~5 in 1 M urea (*44*)], are merely suggestive of a small population of helical structure in the A helix region (*22*). However, after application of sequence corrections and the use of a random coil data set at the correct pH, the $^{13}C^{\alpha}$ shifts provide convincing evidence for helical secondary structure in this region, with the corrected ^{13}CO shifts providing confirmation (Figure 4). Correction for the effects of the local amino acid sequence is especially important for the ^{13}CO shifts (Figure 3), and also for ^{15}N shifts if these are to be used to determine conformational propensities.

Strictly speaking, the observed deviations of the chemical shifts from random coil values indicate preferential population of the α-region of ϕ,ψ space, rather than formation of cooperatively folded helical secondary structure. However, as discussed elsewhere (*22*), several lines of evidence support formation of a small population of cooperatively folded helix in the H helix region of apomyoglobin at pH 2.3. These include the observation of motional restriction of the polypeptide backbone, decreased amide proton exchange rates (Figure 9a), and significantly lowered amide proton temperature coefficients (Figure 9b), indicative of a detectable population of hydrogen-bonded structures. In addition, the circular dichroism spectrum of apomyoglobin at pH 2.3 indicates the presence of a small population of ordered helical structure (*17*). In principle, the presence of ordered helical structure could be confirmed by the observation of medium-range ($i,i+3$) or ($i,i+4$) NOEs. Unfortunately, we were unable to observe any medium-range NOEs in either NOESY−HSQC or HSQC−NOESY−HSQC spectra; this is not surprising since the effective concentration of helical conformers is less than 40 μM, given the small population of helix (see below) and the low concentration at which the NMR experiments had to be performed to avoid aggregation.

The relative populations of α and β backbone dihedral angles can be assessed from the relative intensities of sequential NOEs (Figure 6). The ratio of the intensities of the $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOEs gives a measure of the population of helix in a particular amino acid sequence (*52*). The fact that both $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOEs were observed throughout the polypeptide provides clear evidence

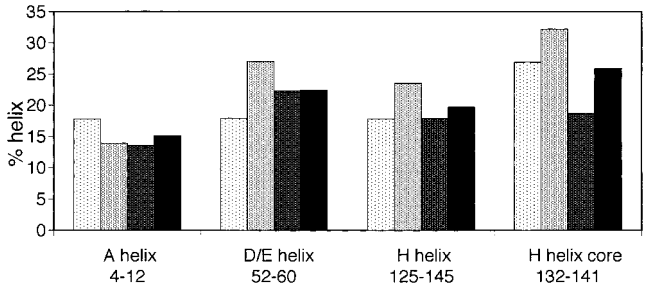

FIGURE 10: Populations of helix formed in acid-unfolded apomyoglobin, estimated from the deviations of $^{13}C^{\alpha}$, ^{13}CO, and $^{1}H^{\alpha}$ chemical shifts from random coil values, using the Ac-GGXGG-NH_2 peptides in 8 M urea as a reference (*43*) and correcting for local sequence effects (*59*). The percentage of helix was estimated from the secondary chemical shifts for each resonance, averaged over the regions indicated. The secondary shifts corresponding to fully formed helix were as follows: $^{13}C^{\alpha}$, 2.8 ppm; ^{13}CO, 2.1 ppm [values averaged from those given in (*42*)]; $^{1}H^{\alpha}$, −0.3 ppm (*58*). For each part of the apomyoglobin sequence (A helix, D/E helix, H helix, and H helix core), the 4 columns represent percentages of helix derived from the secondary $^{13}C^{\alpha}$ (lightest shading), ^{13}CO (light gray), and $^{1}H^{\alpha}$ (dark gray) chemical shifts, and the mean of these 3 values (black).

that the backbone fluctuates over both the α and β regions of ϕ,ψ space. The relative intensity of the $d_{NN}(i,i+1)$ NOEs is highest in the H helix and D/E helix regions, with the implication that these parts of the sequence have the highest population of dihedral angles in the α region of ϕ,ψ space in the acid-unfolded form of apomyoglobin. However, even in these regions, the helical structure is in rapid dynamic equilibrium with fully unfolded states. In general, there is a wide occurrence of $d_{NN}(i,i+1)$ NOEs, indicating that the acid-unfolded state samples α-region dihedral angles in most parts of the polypeptide chain. In contrast, few $d_{NN}(i,i+1)$ NOEs were observed in the unfolded state of plastocyanin, a β-sheet protein (Y. Bai, P. E. Wright, and H. J. Dyson, manuscript submitted).

The secondary chemical shifts reveal three regions of the polypeptide chain of apomyoglobin, corresponding to the A, D/E, and H helices in the native protein, which adopt significant populations of helical structure at pH 2.3 (Figure 4). The population of helix in the A, D/E, and H helices was estimated from the deviations of the $^{13}C^{\alpha}$, ^{13}CO, and $^{1}H^{\alpha}$ chemical shifts from random coil values (Figure 10). The populations derived from each of these resonances are in remarkable agreement. The H helix forms the largest region of contiguous helical structure. The consensus helix boundaries determined from the $^{13}C^{\alpha}$ and ^{13}CO secondary shifts are Ala125 and Lys145. The presence of helix in this region is confirmed by the $^{1}H^{\alpha}$ and temperature-compensated NH secondary shifts, and to a lesser extent by the $^{13}C^{\beta}$ shifts. This helix is clearly frayed at the ends, as judged by the decreased deviations of the $^{13}C^{\alpha}$ shifts from random coil values at the termini (Figure 4a). The average population for the H helix is 20% (and for residues 132−141, comprising the helical core, 26%), in excellent agreement with the population of helix (ca. 25%) in a monomeric synthetic peptide corresponding to the H helix of myoglobin, determined both from circular dichroism data and from NMR parameters (*52*).

The average populations of helix in the A and D/E helix regions at pH 2.3 are 15% and 22%, respectively (Figure 10). The presence of helical structure in the D helix sequence was unexpected, in view of its absence in the corresponding peptide (*51*). However, it is noticeable that the downfield-shifted C^{α} resonances in the pH 2.3 state (Figure 2) extend for several residues past the C-terminal end of the D helix (residue Ala57 in the folded protein) and into the E helix. That is, the residual helical structure seen in the unfolded state (residues Glu52−Asp60) is longer than the region corresponding to the D helix in the native structure. The previously studied D helix peptide mapped to the residues that form the D helix in the folded protein structure; we therefore synthesized and characterized a longer peptide (Thr51−Lys63) consisting of the D helix, the beginning of the E helix, and the intervening residues. Both CD and NMR spectra of this peptide indicate formation of a significant amount of ordered helical structure in water solution, consistent with the data for the pH 2.3 unfolded state of the protein. The failure to observe helical structure in the isolated D and E helix peptides (*51*) appears to have been because these peptides corresponded precisely to the helical boundaries in the native protein and failed to allow for formation of non-native structure spanning the D−E helix junction.

As noted previously, the presence of helical structure in the A helix region is only weakly indicated by uncorrected secondary chemical shifts. Once sequence-dependent corrections are made and the appropriate chemical shift references are used, the presence of a small population of helix in this region is evident from the $^{13}C^{\alpha}$, ^{13}CO, and, to a lesser extent, $^{1}H^{\alpha}$ chemical shifts. The maximum population of helix occurs between residues Glu4 and His12, but the secondary shifts suggest that the helix may extend more weakly as far as Val17, the terminus of helix A in the native protein.

The G helix region is one of the least helical regions in acid-unfolded apomyoglobin, although it forms stable helical structure in the molten globule intermediate (*21*). The F helix and FG loop regions, together with the B, C, and E helix regions, are essentially in random coil conformation, as indicated by the small secondary chemical shift deviations. Indeed, the $^{13}C^{\alpha}$ and ^{13}CO shifts even suggest that there may be a slight intrinsic propensity toward β-region backbone dihedral angles in the F helix region. Both the $^{13}C^{\alpha}$ and ^{13}CO secondary shifts suggest the possibility of a weakly populated helical turn between L72 and I75, toward the C-terminal end of the E helix region.

While the deviations of the chemical shifts from the appropriate random coil values provide a highly sensitive indication of the local secondary structural propensities, the $^{3}J_{HN,H\alpha}$ coupling constants are rather uninformative. This is probably because $^{3}J_{HN,H\alpha}$ is highly sensitive to the ϕ dihedral angle, especially in the α_R region, and differences in conformational averaging at local sites could have a large effect on the measured $^{3}J_{HN,H\alpha}$ values. In addition, more sophisticated sequence-dependent corrections for random coil coupling constants, beyond simple classification of the preceding residue as L or S (*48*), will probably be necessary if coupling constants are to play a useful role in identifying regions of transient secondary structure. As in the case of chemical shifts, it will probably be necessary to measure reference random coil coupling constants under solution conditions that are close to those of the unfolded protein being examined.

Dynamics of the Polypeptide Chain at pH 2.3. The ^{15}N relaxation parameters (Figure 7) and the reduced spectral density functions (Figure 8) indicate a highly flexible polypeptide chain. This is particularly evident in the generally small or negative values for the $\{^1H\}$-^{15}N NOE (Figure 7c). Indeed, as pointed out previously (*22*), the heteronuclear NOE provides an especially sensitive indicator of local variations in backbone motions on the picosecond to nanosecond time scale; residues with negative NOEs exhibit large-amplitude motions on a sub-nanosecond time scale, while residues with weakly positive NOEs are motionally restricted. While local variations in fast time scale motions are also evident in the R_1 and R_2 relaxation rates, the effects are less pronounced. Sequence-dependent variations in the spectral densities due to changes in fast time scale motions are most pronounced for $J(\omega_N)$ (Figure 8b). Anomalously large values of R_2 and $J(0)$ are observed in the A and G helix regions and reflect exchange contributions from slow time scale processes (see later discussion).

The spectral densities $J(\omega_N)$ and $J(0.87\omega_H)$ are sensitive only to fast (picosecond to nanosecond) time scale motions. Larger than average values of $J(\omega_N)$ are observed in several contiguous regions of the polypeptide chain, signifying restriction of sub-nanosecond time scale motions. These include residues Gly5−His12 in the A helix of native myoglobin, residues Ile28−Lys34 in the B helix, and residues Lys133−Glu148 in the H helix. Restriction of motions in these regions is confirmed by the observation of lower than average values for $J(0.87\omega_H)$. The region of restricted motion in the A helix corresponds closely to the segment shown by the secondary chemical shifts to form helical structure in acid-denatured apomyoglobin. Residues 133−148 lie in the central core and C-terminal part of the H helix structure that is populated in the pH 2.3 state. Interestingly, the residues which adopt helical structure in the D/E region in the pH 2.3 state exhibit only average $J(\omega_N)$ and $J(0.87\omega_H)$ values, suggesting that they retain significant flexibility. Although the backbone fluctuations of residues 28−34 are clearly restricted, there is no evidence from the chemical shift data that these amino acids populate nativelike helical structures in the pH 2.3 denatured state. This segment of the polypeptide is highly hydrophobic and buries a large surface area on folding (Figure 8b). It is therefore likely that the observed restriction of backbone motions arises through local hydrophobic collapse rather than transient secondary structure formation. Greater than average $J(\omega_N)$ values are also observed in the region between residues Lys42 and Lys47, which corresponds to the CD loop in the folded protein and contains two phenylalanine side chains, and between Tyr103 and Phe106 in the G helix of the folded protein. Neither of these segments displays measurable propensity to adopt ordered secondary structure at pH 2.3, but both are characterized by large buried surface area (Figure 8b). Thus, it is probable that the observed motional restrictions in these regions reflect local chain collapse and burial of hydrophobic side chains.

Three regions of the apomyoglobin backbone display significantly smaller than average values of $J(\omega_N)$ and above average $J(0.87\omega_H)$, indicating high flexibility on a sub-nanosecond time scale. Two of these regions consist of short stretches of polypeptide at the N- and C-termini of the protein. The other is a long stretch of contiguous residues from Leu69 to His97, encompassing the C-terminal half of the E helix, the EF loop, the entire F helix, and part of the FG loop of the folded protein. The chemical shifts indicate that the backbone conformations in this region are close to random coil, perhaps even with a slight preference toward the β-region of ϕ,ψ space. Thus, the chemical shifts and relaxation data together indicate that this ~30 amino acid segment near the center of the polypeptide behaves essentially as a free-flight random coil chain, separating two N- and C-terminal "domains" in which backbone motions are restricted due to a propensity for secondary structure formation and local hydrophobic collapse. We note that the $J(\omega_N)$ values for Ala74, Ile75, and Leu76 are increased slightly above those of neighboring residues, which may indicate some local motional restriction due to clustering of these hydrophobic side chains.

Finally, we point out that the principal features of the plot of $J(\omega_N)$ versus residue number correlate surprisingly well with the average buried surface area calculated using the parameters of Rose et al. (*53*). Interestingly, the correlation between $J(\omega_N)$ and hydrophobicity is much poorer; the Kyte−Doolittle hydrophobicity (*54*) is maximum near residue 70, in the middle of the E helix, yet the NMR parameters suggest little tendency for hydrophobic collapse and concomitant motional restriction in this region. Similar results are obtained using other hydrophobicity scales. Although the central region of the E helix is strongly hydrophobic, nearly half of the side chains are small (Gly, Ala, or Thr) and probably impart considerable flexibility to the polypeptide backbone. As a consequence, the buried surface area is small, and the entropic cost of restricting the backbone motions through local hydrophobic collapse is prohibitive. Burial of hydrophobic surface area appears to be of major importance in initiation of protein folding (*53*). As we have pointed out elsewhere, differences in the folding pathways of apomyoglobin and apoleghemoglobin can be explained by differences in buried surface area in the A and E helix regions (*55*). Thus, folding of apomyoglobin buries greater surface in the A helix than for apoleghemoglobin, whereas the converse is true for the E helix. Apomyoglobin folds via an A(B)GH helical intermediate, whereas apoleghemoglobin forms an EGH intermediate, with helix A only becoming stabilized late in the folding process.

Slow Motional Processes in ApoMb. The most noticeable features of the R_2 relaxation rate (Figure 7b) and the $J(0)$ spectral density (Figure 8a) are the greatly increased values in two localized segments of the polypeptide. The $J(0)$ values in these regions are much larger than expected on the basis of $J(\omega_N)$ and $J(0.87\omega_H)$ and provide evidence for slow motional processes, on the millisecond to microsecond time scale. Large $J(0)$ (and R_2) values are observed in two contiguous regions, formed by residues Gln8−Ala19 and residues Ile112−His116. We note that within the region of the A helix that exhibits the large R_2 values, the NH resonances of Trp14 and Ala15 are broadened beyond detection so that R_2 [and hence $J(0)$] could not be measured but must be very large. A plausible explanation for the broadening of resonances in these two highly localized regions (and nowhere else in the sequence) is that there is some transient (millisecond to microsecond time scale) contact in the acid-unfolded state between the A and G helix regions. This could be either intermolecular, through specific

interactions that lead to aggregation, or intramolecular, through transient compaction of the polypeptide chain leading to coalescence of the A and G regions. For two reasons, intermolecular association seems unlikely to be the origin of the enhanced R_2 relaxation. First, the anomalously large R_2 relaxation rates in the A and G regions persist in data sets acquired at concentrations of 200, 97, and 50 μM (data not shown). Second, data obtained at 250 μM concentration using nitroxide spin-labeled samples suggest negligible intermolecular association but unequivocally identify transient intramolecular interactions between the A and G helix regions (M. Lietzow, M. Jamin, H. J. Dyson, and P. E. Wright, manuscript in preparation). Although at this time we cannot completely rule out slow conformational fluctuations within locally compacted states as the origins of the enhanced R_2 relaxation, this seems unlikely given the nanosecond time scale flexibility of the polypeptide backbone in these regions. Further experiments are in progress to elucidate the detailed molecular events responsible for the anomalously large values of $J(0)$.

Implications for the Folding Pathway of Apomyoglobin. Among the three helices A, G, and H, which are stabilized in the molten globule state (*19*, *22*), only the A and H helices have appreciable helical content in the pH 2.3 acid-denatured form of apomyoglobin. The H helix shows about 20% population of helix, and the A helix less so (~15% between residues Glu4 and His12). In contrast, the G helix exhibits no significant secondary structural preferences. Since the three helices that are formed in the molten globule state at pH 4 are not the same ones that appear in the unfolded state at pH 2.3, we may definitively conclude that the pH 2.3 state is not a mixture of the molten globule state at pH 4 and some totally unfolded state. Restriction of backbone motions is evident in the $J(\omega_H)$ and $J(0.87\omega_H)$ spectral densities for the A and H helices as well as for regions of the B and G helices and the CD loop of the native protein structure. In the latter three instances, the motional restrictions appear to result from burial of side chains in local hydrophobic clusters, rather than from transient secondary structure formation. The current work therefore presents concrete experimental evidence for the formation of incipient local structure in the apomyoglobin polypeptide chain in a state far earlier in the folding process than has hitherto been accessible.

It is important to note that the non-native structure observed in the D−E helix sequence may also play a crucial role in the folding process by restricting the conformational space available to the polypeptide and by initiating formation of the D and E helices. This structure is non-native only to the extent that the helical structure formed transiently at pH 2.3 "runs through" the kink that separates the D and E helices in the native folded structure. Presumably long-range interactions formed at later stages of the folding process, possibly associated with stabilization of helical structure in the remainder of the E helix, lead to restructuring of this region of the polypeptide. Non-native structure has been invoked as a possible source of slowing of folding, due to the formation of off-pathway intermediates that must be unfolded before productive folding can proceed (*56*). This does not apply to the folding of apomyoglobin, since the helical propensity in the D−E helix sequence is small, and hence the structure can readily be melted out or modified with no effect on the folding pathway.

In summary, this work has identified several regions of the myoglobin sequence that spontaneously populate helical secondary structure or that collapse spontaneously into local hydrophobic clusters. Regions where local collapse occurs, as indicated by restriction of the backbone motions on the nanosecond time scale, tend to be those where there is a large change in the buried surface area on folding. These regions probably play an important role in initiation of protein folding and in directing the collapse so that the correct polypeptide chain topology is realized in the earliest compact molten globule states.

ACKNOWLEDGMENT

We thank Patricia Jennings, Martin Stone, Stefan Prytulla, Carlos Garcia, Silvia Cavagnero, Richard Kriwacki, Stephan Schwarzinger, and Yawen Bai for valuable discussions and input on NMR experiments.

REFERENCES

1. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
2. Evans, P. A., Topping, K. D., Woolfson, D. N., and Dobson, C. M. (1991) *Proteins: Struct., Funct., Genet. 9*, 248−266.
3. Buck, M., Radford, S. E., and Dobson, C. M. (1994) *J. Mol. Biol. 237*, 247−254.
4. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) *Biochemistry 36*, 8977−8991.
5. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) *Science 257*, 1559−1563.
6. Logan, T. M., Thériault, Y., and Fesik, S. W. (1994) *J. Mol. Biol. 236*, 637−648.
7. Lumb, K. J., and Kim, P. S. (1994) *J. Mol. Biol. 236*, 412−420.
8. Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) *Protein Sci. 4*, 2605−2615.
9. Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 305−321.
10. Zhang, O., and Forman-Kay, J. D. (1995) *Biochemistry 34*, 6784−6794.
11. Zhang, O., and Forman-Kay, J. D. (1997) *Biochemistry 36*, 3959−3970.
12. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol. 268*, 158−169.
13. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol. 268*, 170−184.
14. Yao, J., Dyson, H. J., and Wright, P. E. (1997) *FEBS Lett. 419*, 285−289.
15. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) *Protein Sci. 8*, 45−49.
16. Jamin, M., and Baldwin, R. L. (1998) *J. Mol. Biol. 276*, 491−504.
17. Griko, Y. V., Privalov, P. L., Venyaminov, S. Y., and Kutyshenko, V. P. (1988) *J. Mol. Biol. 202*, 127−138.
18. Griko, Y. V., and Privalov, P. L. (1994) *J. Mol. Biol. 235*, 1318−1325.
19. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) *Science 249*, 1544−1548.
20. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
21. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 2894−2901.
22. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
23. Jennings, P. A., Stone, M. J., and Wright, P. E. (1995) *J. Biomol. NMR 6*, 271−276.
24. Harrison, S. C., and Blout, E. R. (1965) *J. Biol. Chem. 240*, 299−303.
25. Patt, S. L. (1992) *J. Magn. Reson. 96*, 94−102.

26. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.
27. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 203−216.
28. 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.
29. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
30. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
31. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson. 101*, 201−205.
32. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
33. Löhr, F., and Rüterjans, H. (1995) *J. Biomol. NMR 6*, 189−197.
34. Shaka, A. J., Lee, C. J., and Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
35. Cavanagh, J., and Rance, M. (1992) *J. Magn. Reson. 96*, 670−678.
36. Ikura, M., Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 9020−9022.
37. Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., and Feeney, J. (1990) *J. Magn. Reson. 90*, 420−425.
38. Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D., and Bax, A. (1995) *J. Am. Chem. Soc. 117*, 9594−9595.
39. Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) *J. Biomol. NMR 4*, 871−878.
40. Van Geet, A. L. (1970) *Anal. Chem. 42*, 679−680.
41. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
42. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
43. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) *J. Biomol. NMR. 18*, 43−48.
44. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
45. Wishart, D. S., and Case, D. A. (2000) *Methods Enzymol.*
46. Dyson, H. J., and Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
47. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) *J. Mol. Biol. 255*, 494−506.
48. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M., and Smith, L. J. (1997) *J. Mol. Biol. 274*, 152−159.
49. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
50. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) *J. Biomol. NMR 6*, 153−162.
51. Reymond, M. T., Merutka, G., Dyson, H. J., and Wright, P. E. (1997) *Protein Sci. 6*, 706−716.
52. Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., and Wright, P. E. (1993) *Biochemistry 32*, 6337−6347.
53. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) *Science 229*, 834−838.
54. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105−132.
55. Nishimura, C., Prytulla, S., Dyson, H. J., and Wright, P. E. (2000) *Nat. Struct. Biol. 7*, 679−686.
56. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
57. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
58. Williamson, M. P. (1990) *Biopolymers 29*, 1423−1431.
59. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E. and Dyson, H. J. (2001) *J. Am. Chem. Soc.* (in press).

BI002776I