# Conformational and Dynamic Characterization of the Molten Globule State of an Apomyoglobin Mutant with an Altered Folding Pathway†

Silvia Cavagnero,‡ Chiaki Nishimura, Stephan Schwarzinger,§ H. Jane Dyson,* and Peter E. Wright*,§

*Department of Molecular Biology MB-2 and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037*

*Received July 19, 2001; Revised Manuscript Received October 5, 2001*

ABSTRACT: Kinetic and equilibrium studies of apomyoglobin folding pathways and intermediates have provided important insights into the mechanism of protein folding. To investigate the role of intrinsic helical propensities in the apomyoglobin folding process, a mutant has been prepared in which Asn132 and Glu136 have been substituted with glycine to destabilize the H helix. The structure and dynamics of the equilibrium molten globule state formed at pH 4.1 have been examined using NMR spectroscopy. Deviations of backbone $^{13}C^{\alpha}$ and $^{13}CO$ chemical shifts from random coil values reveal high populations of helical structure in the A and G helix regions and in part of the B helix. However, the H helix is significantly destabilized compared to the wild-type molten globule. Heteronuclear $\{^1H\}-^{15}N$ NOEs show that, although the polypeptide backbone in the H helix region is more flexible than in the wild-type protein, its motions are restricted by transient hydrophobic interactions with the molten globule core. Quench flow hydrogen exchange measurements reveal stable helical structure in the A and G helices and part of the B helix in the burst phase kinetic intermediate and confirm that the H helix is largely unstructured. Stabilization of structure in the H helix occurs during the slow folding phases, in synchrony with the C and E helices and the CD region. The kinetic and equilibrium molten globule intermediates formed by N132G/E136G are similar in structure. Although both the wild-type apomyoglobin and the mutant fold via compact helical intermediates, the structures of the intermediates and consequently the detailed folding pathways differ. Apomyoglobin is therefore capable of compensating for mutations by using alternative folding pathways within a common basic framework. Tertiary hydrophobic interactions appear to play an important role in the formation and stabilization of secondary structure in the H helix of the N132G/E136G mutant. These studies provide important insights into the interplay between secondary and tertiary structure formation in protein folding.

A mechanistic understanding of how a polypeptide chain achieves its native folded conformation is one of the greatest challenges in structural biology. Nuclear magnetic resonance is a particularly powerful technique for studies of protein folding landscapes, since it offers a unique opportunity for mapping of the structural and dynamic features of unfolded, partially folded, and native states at the level of individual amino acid residues (reviewed in ref *1*). We have focused on the detailed elucidation of the folding pathways of apomyoglobin (apoMb),[1] which is amenable to both kinetic and equilibrium analysis of the folding process (*2*−*12*). High-resolution experimental characterization of folding events and intermediates is essential for a detailed understanding of the interplay between hydrophobic compaction and secondary structural propensities in directing the folding process. Wild-type apoMb folds via an obligatory kinetic intermediate in which helical structure is stabilized in the A, G, and H helices and in part of the B helix (*3*, *8*, *13*). ApoMb also forms an equilibrium molten globule intermediate which is maximally populated at pH 4 and which is very similar in structure to the kinetic intermediate (*2*, *4*, *7*, *10*). Detailed NMR studies of the equilibrium pH 4 intermediate of apoMb are of great importance for the insights they can provide into the structure and dynamics of molten globules. Key questions that remain to be answered include the role of secondary structure in formation and stabilization of molten globule intermediates and the detailed relationship between equilibrium and kinetic molten globule states. To obtain insights into these issues, we have initiated parallel kinetic and equilibrium NMR studies of the folding of apomyoglobin mutants.

In this paper, we describe the secondary structure and dynamic features of the equilibrium molten globule state of a mutant of apoMb in which Asn132 and Glu136 have been replaced with glycine (N132G/E136G). This mutant was designed to have a significantly decreased helical propensity in the H helix (*9*). Previous kinetic studies have indicated

† This work was supported by Grant DK34909 from the National Institutes of Health. S.C. received support from the Wills Foundation, and S.S. acknowledges an Erwin Schrödinger Fellowship from the Austrian Science Fund (FWF, Project CHE-1736).

* To whom correspondence should be addressed. Phone: (858) 784-9721. Fax: (858) 784-9822. E-mail: wright@scripps.edu or dyson@scripps.edu.

‡ Present address: Department of Chemistry, University of Wisconsin−Madison, 1101 University Ave., Madison, WI 53706.

§ Present address: Universität Bayreuth-Lehrstuhl Biopolymere, Universitätsstrasse 30/BGI, D-95440 Bayreuth, Germany.

[1] Abbreviations: apoMb, apomyoglobin; N132G/E136G, apoMb mutant where residues in positions 132 and 136 have been replaced by glycine; NMR, nuclear magnetic resonance.

10.1021/bi011500n CCC: $20.00 
Published on Web 11/07/2001

that the mutant protein forms both equilibrium and kinetic intermediates; however, unlike that in the wild-type protein, the H helix is not stabilized in the kinetic intermediate (*9*). Here, we report on detailed quench flow hydrogen exchange pulse labeling experiments for probing the folding pathway of the N132G/E136G mutant, together with direct NMR characterization of the structure and dynamics of the equilibrium molten globule state formed by this protein. These studies confirm that the mutated H helix no longer becomes stabilized and folded within the burst phase intermediate, and show that it folds during the slower phase, on a time scale comparable to that of the other slowly folding helices in the wild-type protein (*3*). The equilibrium NMR experiments reveal that the structure of the equilibrium molten globule is similar to that of the kinetic intermediate. Thus, the introduction of destabilizing mutations into the H helix affects the structure of both the kinetic and equilibrium intermediates in similar ways, providing strong evidence that these states are closely related. Finally, this work provides important information about the relationship between secondary structure formation and long-range hydrophobic interactions in determining the structure of apomyoglobin folding intermediates.

## MATERIALS AND METHODS

*Sample Preparation.* Samples of $^{15}$N-labeled and doubly $^{13}$C- and $^{15}$N-labeled N132G/E136G apoMb were prepared as previously described (*9*). The unlabeled protein, used in the fluorescence equilibrium titrations, was overexpressed in standard Luria broth and purified using the same procedure that was used for the isotopically labeled samples. NMR samples of the molten globule intermediate of N132G/E136G were prepared by dissolving the lyophilized protein in 10 mM sodium acetate buffer (pH 4.10). The sample was passed through a 3 mL bed volume spin column loaded with G25 (fine grade) gel filtration resin equilibrated with the same buffer. Ethanol was added to a total concentration of 10% (v/v) to prevent aggregation during the multidimensional NMR experiments (*7*, *10*), and the pH was readjusted to 4.10 by adding small aliquots of 0.1 M acetic acid or sodium acetate as needed. The final protein concentration was 0.3 mM.

A peptide with an acetyl-GADAQGAM**G**KAL**G**LFRK-DIAAKYKELG-$NH_2$ sequence, corresponding to the apoMb H helix with mutations N132**G** and E136**G**, was synthesized using an Applied Biosystems peptide synthesizer. The peptide was purified by HPLC, and its composition was checked by mass spectrometry.

*Fluorescence and Circular Dichroism-Detected pH Titrations.* The pH titrations of the N132G/E136G apoMb mutant were carried out with a circular dichroism (CD) spectropolarimeter (Aviv, model 202) equipped with a total fluorescence module, a quantum counter, an automated titration system, a Peltier element for temperature control, and an Orion SureFlow-ROSS pH electrode. The step size in the titrations was set to 0.05 pH unit, and the protein concentration was 5 $\mu$M as determined by absorption spectroscopy at 280 nm (extinction coefficient of 15 200 $M^{-1}$ $cm^{-1}$) and 288 nm (extinction coefficient of 10 800 $M^{-1}$ $cm^{-1}$) (*14*). CD was measured at 222 nm for 5 s. A quartz cell with a 1 cm path length was employed. The excitation wavelength was set to 278 nm, and the total fluorescence was detected for 5 s using a 305 nm high-pass filter. The titrations were carried out in a constant volume mode starting from a low pH (solution A consisting of 5 $\mu$M protein and 10 mM acetic acid, containing 5 mM HCl, which gives a pH of 2.3) to a higher pH (solution B consisting of 5 $\mu$M protein and 10 mM acetic acid, with the pH adjusted to 7 with NaOH). Titrations were carried out at both 25 and 50 °C. The effect of 10% ethanol on the molten globule state was only investigated at 25 °C due to solvent evaporation at 50 °C. To minimize the experiment time, data were acquired over a smaller pH range at 50 °C.

*CD Characterization of the N132G/E136G H Helix Peptide.* The secondary structure content of the mutant N132G/E136G H helix peptide was determined from the molar ellipticity. All samples were 15 $\mu$M in 1 mM phosphate buffer at pH 5.0, and spectra were recorded at 5 and 25 °C.

*NMR Spectroscopy of the Equilibrium Molten Globule State.* Triple-resonance spectra for assignment of backbone resonances of the molten globule state of the doubly $^{13}$C- and $^{15}$N-labeled N132G/E136G were recorded at pH 4.1 and 50 °C on a Bruker DMX 750 MHz spectrometer. The probe temperature was calibrated using methanol (*15*). Spectra were referenced to the water resonance at 4.5 ppm, and indirectly in the $^{13}$C and $^{15}$N dimensions (*16*). NMR spectra were processed using NMRPipe (*17*) and analyzed using NMRView (*18*).

Due to the partially unfolded nature of this species, many resonances were severely overlapped and a number of peaks were very broad which made resonance assignment difficult. Assignments were made on the basis of connectivities observed between backbone $^{13}$CO, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{15}$N nuclei in three-dimensional HNCA, HN(CO)CA, HNCO, (HCA)CO(CA)NH, CBCA(CO)NH, and HNCACB spectra (*19*−*23*). Due to their low inherent sensitivity and, in part, to peak broadness, the HNCACB and (HCA)CO(CA)NH spectra were of only limited utility; nevertheless, these spectra were useful for confirming some of the connectivities identified in other experiments. All assignments for N132G/E136G were made independently from the triple-resonance spectra but, in regions distant in sequence from the site of mutation, were checked for consistency with the assignments for the wild-type molten globule (*10*). Assignments of the $^{15}$N, NH, and $^{13}$CO resonances could be made for 138 residues in the molten globule state of the mutant protein, while $^{13}C^{\alpha}$ assignments were obtained for 149 residues. The resonance assignments have been deposited in the BioMagResBank (accession number 5158).

*$^{1}$H−$^{15}$N Heteronuclear NOE Measurements.* Heteronuclear NOEs were measured on a $^{15}$N-labeled sample using previously published methods (*24*) at 50 °C and 750 MHz. Heteronuclear NOE values were determined from resonance peak heights as described by Stone et al. (*25*, *26*) using in-house programs and/or programs developed by A. G. Palmer (*27*). Uncertainties in heteronuclear NOE values were calculated as described by Nicholson et al. (*28*).

*Quench Flow Hydrogen Exchange Measurements.* Hydrogen exchange pulse labeling experiments were performed at 5 °C using a Biologic QFM5 Quench Flow instrument. Lyophilized $^{15}$N-labeled N132G/E136G apomyoglobin was dissolved in 10 mM sodium acetate buffer (pH 5.9) in $H_2O$ containing 6 M urea. The unfolded protein solution (2.2 mg/

mL) was rapidly mixed with a 7.5-fold volume of 10 mM sodium acetate buffer (pH* 5.9) in $^2H_2O$, to a final urea concentration of 0.7 M. The N132G/E136G apomyoglobin is fully folded at this urea concentration (*9*). After the various refolding times (6.4 ms, 37 ms, 74 ms, 104 ms, 200 ms, 300 ms, 500 ms, 1 s, 2 s, 4 s, and 8 s), the pH* of the solution was increased to 10.0 for 20 ms by the addition of 100 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) buffer in $^2H_2O$. At the end of the deuterium labeling pulse, the pH was rapidly dropped to 6.1 with 300 mM 3-(*N*-morpholino)-propanesulfonic acid (MOPS) buffer in $^2H_2O$. To stabilize the protein for NMR analysis and minimize the amide proton exchange rate, a 3-fold molar excess of bovine heme (Sigma) in 50 mM phosphate buffer (pH* 9.0) containing 100 mM KCN in $^2H_2O$ was added to the solution and the holoprotein was reconstituted. The pH* of the solution was then decreased to 5.7 by the addition of 300 mM MOPS buffer in $^2H_2O$. The solutions were concentrated using a Centriprep 10 (Amicon) cartridge and exchanged into 50 mM phosphate buffer in $^2H_2O$ (pH* 5.6). Sodium dithionite (final concentration of 1.3 mg/mL) was added to the solution under an atmosphere of CO, and the excess dithionite was removed by using a Sephadex G15 column equilibrated with CO-saturated phosphate buffer in $^2H_2O$ (50 mM, pH* 5.6).

*NMR Spectroscopy of Quench Flow Samples.* The $^1H-^{15}N$ HSQC spectra of the pulse-labeled N132G/E136G myoglobin samples, in the form of the CO complex, were recorded at 35 °C on a Bruker AMX500 spectrometer equipped with pulsed field gradients. Changes in the intensity of the amide cross-peaks were monitored as a function of refolding time. The data were collected in the indirect dimension with quadrature detection using the TPPI−States method. One-dimensional $^1H$ spectra were also recorded, and the signal intensity of amides was calibrated on the basis of the volumes of the methyl proton resonances of V68, L29, and V17 observed between 0.0 and −2.5 ppm in the one-dimensional proton spectra. The signal intensity at each time point was also normalized using the intensity of the same signal at the final point (8 s).

*Mass Spectrometry of Quench Flow Samples.* Electrospray ionization (ESI) mass spectra were recorded with a Perkin-Elmer API 100 double-quadrupole mass spectrometer. The holomyoglobin solutions obtained at different refolding times were exchanged into 10 mM ammonium acetate buffer in $^2H_2O$ (pH* 5.6) and then into pure $H_2O$ using a Centricon 10 (Amicon) cartridge. The samples were mixed with a 2-fold volume of a 60% acetonitrile/0.1% TFA (v/v) mixture prior to measurements. The declustering potential was maintained between 50 and 200 V, whereas the emitter voltage was maintained at 4200 V. The spectra were analyzed using the +7 charge state peak with a resolution of 0.1 *m*/*z*.

## RESULTS

*Fluorescence and Circular Dichroism.* The conditions under which the equilibrium molten globule state of the N132G/E136G apomyoglobin is maximally populated were determined previously using CD and fluorescence spectroscopy (*9*). At room temperature, both wild-type and N132G/E136G apoMb form an equilibrium molten globule state at pH 4.10. NMR spectra of the wild-type molten globule were recorded at pH 4.1 and 50 °C in the presence of 10% ethanol to stabilize the protein against aggregation during the several

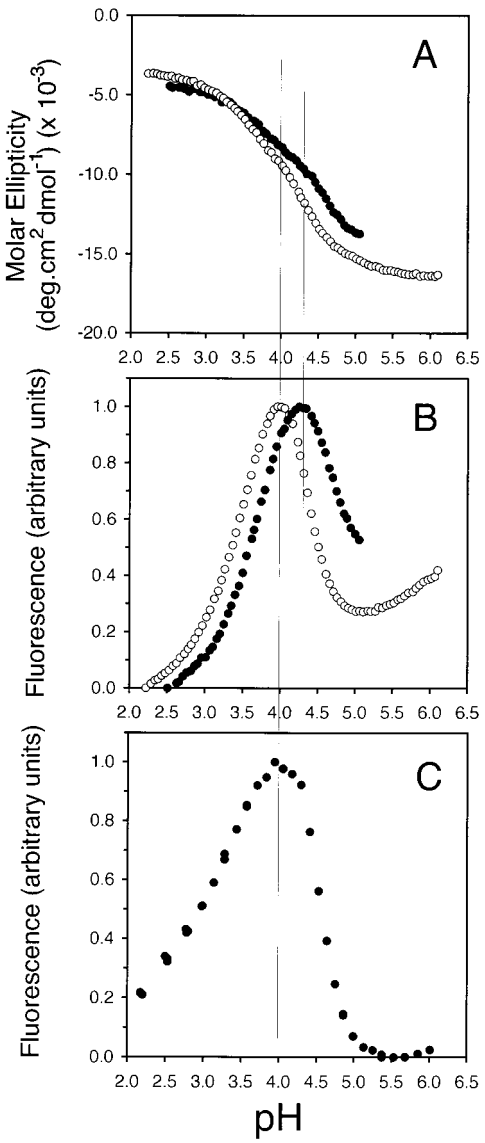


FIGURE 1: pH titrations of the N132G/E136G apoMb mutant detected by (A) circular dichroism and (B) fluorescence at 25 (○) and 50 °C (●). (C) Trp fluorescence emission-detected titration at 25 °C in the presence of 10% ethanol. For easier comparison, each of the fluorescence traces was normalized to be 0 at the first data point and to be 1 at its maximum. Vertical lines show the position of the fluorescence maximum and the CD inflection point, each of which indicates the maximal population of the molten globule state.

days required for three-dimensional NMR data acquisition (*6*, *7*, *10*). We have shown previously that the wild-type molten globule is fully formed at pH 4.1 at this temperature and have established that this small amount of ethanol does not significantly perturb its structure. To ensure that the equilibrium molten globule of N132G/E136G is highly populated under the same conditions, circular dichroism and fluorescence-detected pH titrations of N132G/E136G apoMb were performed at 25 and 50 °C (Figure 1). Both techniques show that the molten globule state is fully populated at 50 °C, although the fluorescence maximum shifts ~0.2 unit to a higher pH at this temperature. Due to the destabilization of the H helix, the plateau corresponding to the maximal population of the molten globule is less pronounced in the CD spectrum of the mutant protein than it is for the wild type. Addition of 10% ethanol has virtually no effect on the position of the fluorescence maximum of N132G/E136G at

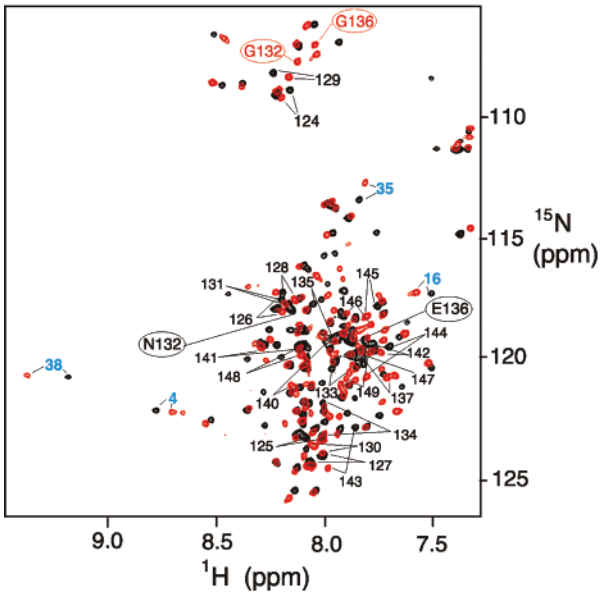


FIGURE 2: HSQC spectra (750 MHz) of the molten globule states of the N132G/E136G mutant (red) and wild-type apomyoglobin (black) at 50 °C and pH 4.10. Assignments for H helix residues are indicated. Well-resolved backbone peaks whose proton resonances have been significantly shifted as a result of the N132G/E136G mutation are labeled in blue. Mutated residues are circled.

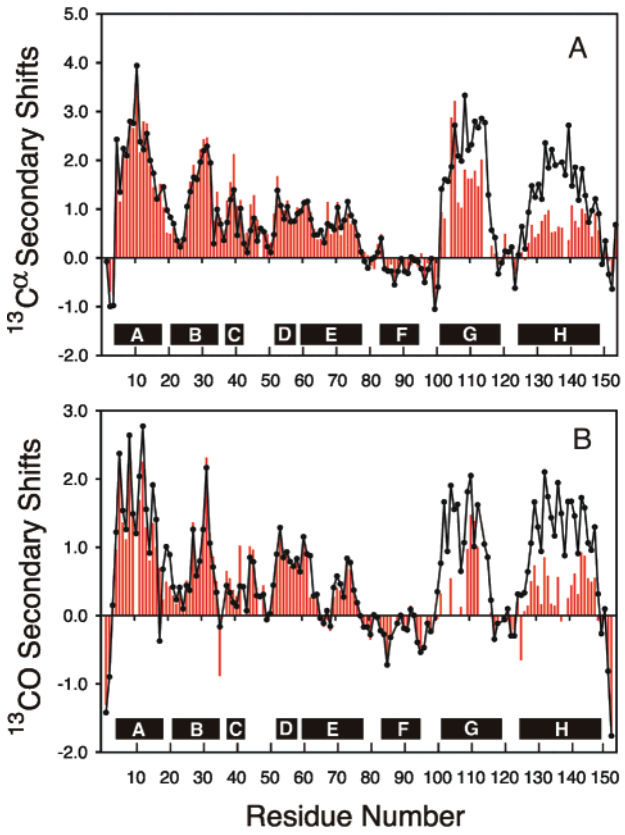


FIGURE 3: (A) $^{13}C^{\alpha}$ and (B) $^{13}CO$ secondary shifts (expressed as deviations from random coil values) for the pH 4.10 state of the N132G/E136G apoMb mutant (histogram) and wild-type apoMb (data points) (*10*). The reference random coil shifts were corrected for the local amino acid sequence as described in the text. The black bars at the bottom of each figure indicate the location of helical structure in the native folded myoglobin.

25 °C (Figure 1C). Because of solvent evaporation, fluorescence pH titrations in the presence of 10% ethanol were not performed at 50 °C; evaporation is minimized in NMR experiments by sealing the samples tightly in a Shigemi tube (*10*). The titration data show clearly that N132G/E136G forms a highly populated molten globule under the same experimental conditions that were used for previous NMR studies of the wild-type protein (*7*).

*Backbone NMR Resonance Assignments of N132G/E136G.* Resonance assignments of fully or partially unfolded states of proteins are made difficult by the poor chemical shift dispersion. Double- and triple-resonance-based methods performed on isotopically labeled samples exploit the greater chemical shift dispersion of the $^{13}CO$ and $^{15}N$ resonances to perform backbone assignments (*29*). In addition to the routine HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB experiments (*30*), the (HCA)CO(CA)NH experiment (*20*) exploits the significant chemical shift dispersion of $^{13}CO$. This experiment provides inter-residue connectivities for backbone assignments of protein-unfolded states. Partially folded species, including the apoMb equilibrium folding intermediate, present additional challenges, since they frequently display intermediate time scale motions which broaden the (HCA)CO(CA)NH peaks. Backbone assignments of the N132G/E136G apoMb were made at pH 4.1 using the listed triple-resonance experiments and were compared with known assignments for the wild-type protein as an additional check. Several resonances for residues in the G helix have very low signal-to-noise ratios, and some are broadened beyond detection. Continuous sequential connectivities could not be determined in this region of the sequence.

The $^{1}H-^{15}N$ HSQC spectrum of the equilibrium molten globule intermediate of N132G/E136G apoMb is shown in Figure 2, superimposed on the HSQC spectrum of the wild-type molten globule. The two additional glycine cross-peaks (G132 and G136) can readily be observed as two additional peaks at the top of the figure. The overall chemical shift dispersion in both dimensions is very similar for the wild-type and mutant molten globules, and many resonances are only slightly shifted as a result of the mutation. However, several resonances belonging to residues distant from the site of mutation experience significant changes in $^{1}H^{N}$ or $^{15}N$ shifts, indicating long-range environmental perturbations; well-resolved $^{1}H$ and $^{15}N$ resonances that experience significant shifts are indicated by blue labels in Figure 2.

*Secondary Structure Mapping by NMR.* The deviations of the $^{13}C^{\alpha}$ and $^{13}CO$ chemical shifts from random coil values (secondary shifts) for the pH 4 state of the N132G/E136G mutant apoMb are shown in Figure 3, together with those of the wild-type protein (*10*). The random coil reference values were those determined for apoMb unfolded in urea at pH 2.3 (*31*), with sequence-dependent corrections as described by Schwarzinger et al. (*32*). To compensate for pH-dependent shifts for Asp and Glu, the random coil values for the GG(D/E)AGG peptides (*33*) were used for these residues, corrected for the sequence contributions of the neighboring alanine residue as described elsewhere (*34*). The importance of correcting the random coil shifts for local amino acid sequence, especially for $^{13}CO$ shifts, has been discussed previously (*12*, *32*, *34*). Regions of the polypeptide with positive secondary shifts for $^{13}C^{\alpha}$ and $^{13}CO$ show a preference for $\phi$ and $\psi$ angles in the α-helical region (*35*, *36*). Figure 3 shows that the distribution of helical structure

Table 1: Estimation of the Percentage of Helical Structure in the pH 4 State of Mutant and Wild-Type ApoMb[a]

| helix (residues) | % helix wild type $^{13}CO$ | % helix wild type $^{13}C^{\alpha}$ | avg % helix wild type | % helix N132G/E136G $^{13}CO$ | % helix N132G/E136G $^{13}C^{\alpha}$ | avg % helix N132G/E136G |
|---|---|---|---|---|---|---|
| A (4−17) | 74.7 | 81.6 | 78 | 67.1 | 77.0 | 72 |
| B (21−35) | 31.7 | 42.2 | 37 | 32.2 | 47.0 | 40 |
| C (37−42) | 15.6 | 30.3 | 23 | 27.2 | 45.1 | 36 |
| D (52−57) | 43.5 | 34.5 | 39 | 42.5 | 39.0 | 41 |
| E (59−77) | 19.6 | 25.5 | 23 | 15.6 | 25.4 | 21 |
| F (83−95) | −11.4 | −5.9 | −9 | −11.0 | −4.3 | −8 |
| G (101−118) | 51.4 | 67.8 | 60 | nd[b] | 48.7 | 49 |
| H (125−149) | 56.0 | 49.4 | 53 | 16.5 | 21.0 | 19 |

[a] The secondary shifts corresponding to a fully formed helix were 2.8 ppm for $^{13}C^{\alpha}$ atoms and 2.1 ppm for $^{13}CO$ atoms [values averaged from those given by Wishart and Sykes (*47*)]. [b] Not determined due to missing assignments.

in the molten globule state of N132G/E136G is very similar to that of the wild-type molten globule, with the exception of the H helix, which is considerably less populated for the mutant protein. The population of helical structure in the G helix region also appears to be somewhat diminished in the mutant.

An estimate of the population of helical structure in the molten globules formed by the mutant and wild-type protein was made from the deviations of chemical shifts from random coil values, as described elsewhere (*12*). Briefly, the secondary chemical shifts for $^{13}C^{\alpha}$ and $^{13}CO$ were averaged over the residues that form the secondary structure elements in native myoglobin. The helical populations are summarized in Table 1. Within experimental uncertainties, there are no significant differences in the population of helical structure in the A, B, D, or E helix regions between the wild-type and mutant proteins. Neither protein shows significant helical structure in the F helix region in the molten globule state. The most notable effects of the N132G/E136G mutation are the dramatic destabilization of structure in the H helix (population decreases from 53 to 19% in the mutant protein) and the smaller decrease in the population of helical structure in the G helix region. The data also suggest a slight increase in the level of helical structure in the C helix region upon mutation, but given the large uncertainties in determining the population of this short helix, the significance of this observation is not clear. We do note, however, that significant $^{15}N$ or $^{1}H^{N}$ shift changes are observed for residues 35 and 38, near the start of the C helix.

*$^{1}H-^{15}N$ Steady State NOE Experiments.* The $\{^{1}H\}-^{15}N$ heteronuclear NOE data for the molten globule state of both mutant and wild-type apoMb are shown in Figure 4. The NOE values are remarkably similar for residues 2−96, encompassing helices A−F of the folded protein, although very small differences can be discerned in parts of the A helix. However, substantially decreased heteronuclear NOEs for residues in the H helix and, to a lesser extent, the G helix indicate that these segments of the polypeptide have become more dynamic as a result of the N132G/E136G mutation. The flexibility of the H helix in the pH 4 state of the mutant is comparable to that of the E helix region in both mutant and wild-type proteins. Although the H helix region has clearly become much more flexible as a result of the mutation, its motions are more restricted than those of the F helix segment which behaves as a free flight random coil polypeptide in both the wild-type (*7*, *10*) and mutant molten globules. This important observation bears on the structure

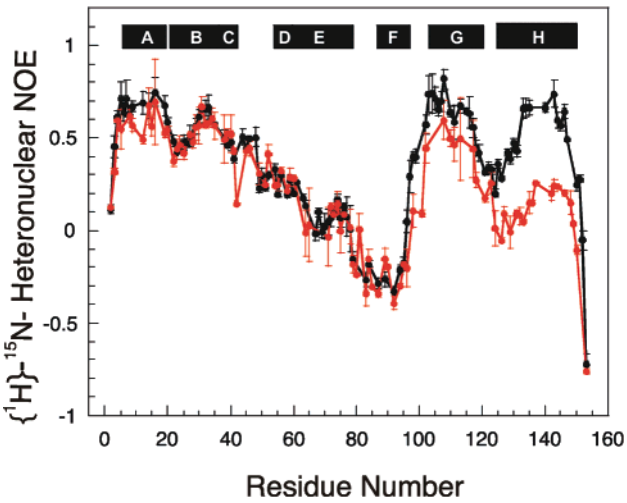


FIGURE 4: Plot of $\{^{1}H\}-^{15}N$ heteronuclear NOEs vs residue number for N132G/E136G (red) and wild-type apoMb (black). Data were collected at 50 °C under conditions identical to those used for the backbone resonance assignments.

of the molten globule in the mutant and wild-type proteins (see the Discussion).

*N132G/E136G H Helix Peptide.* A 27-residue peptide consisting of the H helix sequence of the N132G/E136G mutant protein (Ac-GADAQGAMGKALGLFRKKDIAAK-YKELG-$NH_2$) was prepared by solid phase peptide synthesis, and the helix content was measured by far-UV CD (data not shown) at 25 and 4 °C. In contrast to a peptide with the wild-type sequence (*37*), the mutant peptide displays a CD spectrum typical of a random coil, confirming that the H helix region of N132G/E136G apoMb has a negligible intrinsic helicity.

*Quench Flow Hydrogen Exchange Experiments.* Preliminary quench flow data for the N132G/E136G mutant, at a single refolding time (6.4 ms), have been reported previously (*9*). The quench flow refolding experiments have now been extended to additional refolding times to sample the entire folding process (Figure 5). As for the wild-type protein (*3*), amide protons in the A and G helix regions are almost fully protected from exchange in the burst phase intermediate, as are residues 28−30 in the B helix. However, amides in the H helix region are largely unprotected in the burst phase; formation of stable helical structure in this region occurs only during the slow folding phase, at a rate similar to that for folding of the remainder of the B helix, the C and E helices, and the CD loop. This is in marked contrast to wild-type apomyoglobin, in which the H helix is folded and stabilized

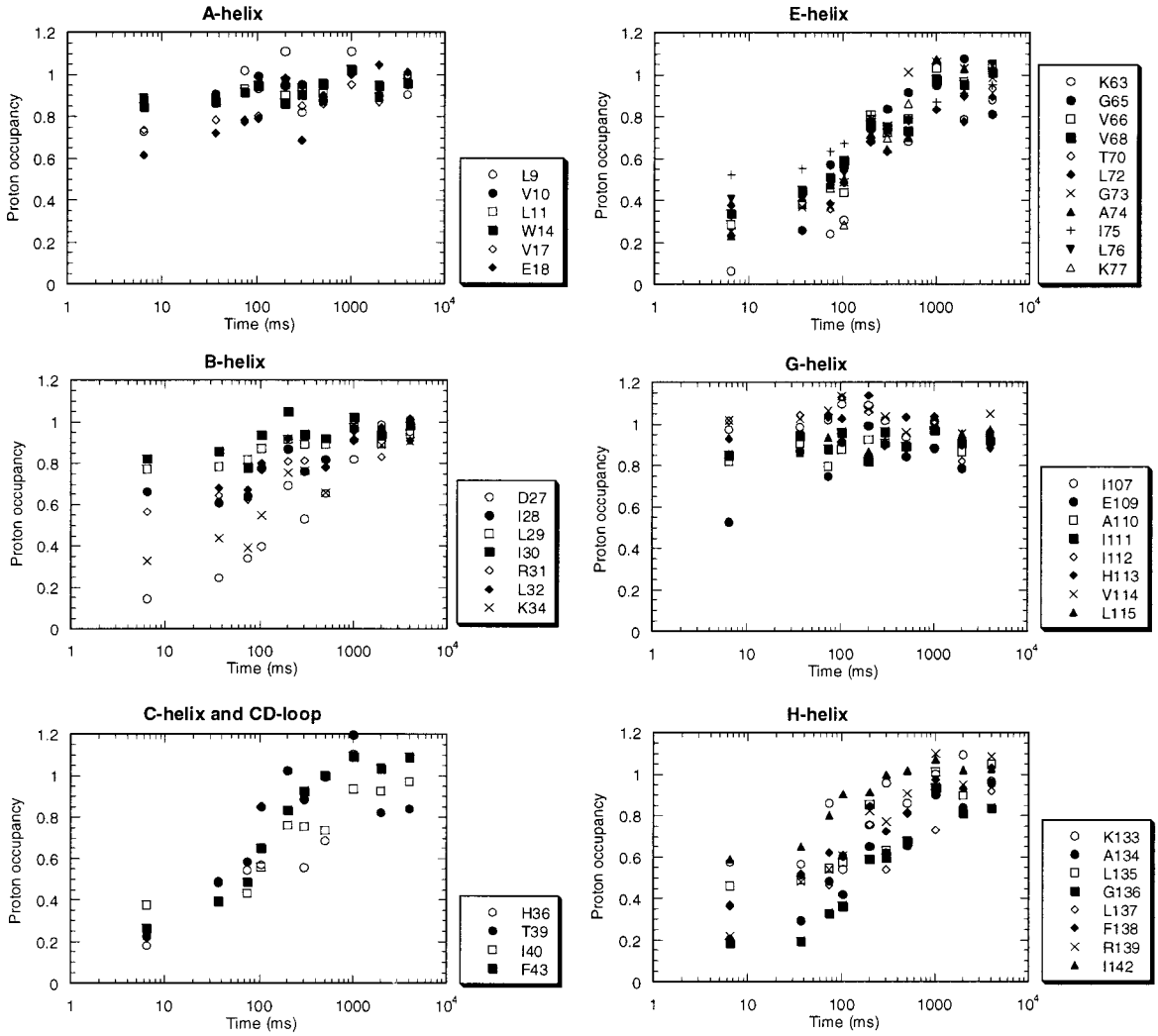


FIGURE 5: Quench flow experiments with the N132G/E136G mutant. Following the pH pulse labeling, the apoproteins were reconstituted with heme and the HSQC spectra of the holoproteins were collected. On the basis of the backbone signal assignments the mutant (*9*), the proton occupancies were determined for each residue as a function of refolding time.

against amide proton exchange during the burst phase. The same quench flow samples were also subjected to analysis by mass spectrometry (data not shown). The data show that the folding pathway of the N132G/E136G mutant is a simple sequential one, as for the wild-type protein (*8*), with initial formation of the intermediate or I state, followed by decay of this state over a time period of ~1 s concomitant with formation of the fully folded native apoprotein. Interestingly, the rate of this slower folding step is the same for the mutant and wild-type proteins (*9*), indicating that the lack of helical structure in the H helix neither facilitates nor hinders the normal folding of the remainder of the molecule.

## DISCUSSION

*Structure of the Mutant ApoMb Molten Globule.* Although the CD spectra of the folded (pH 6) forms of the wild-type and N132G/E136G apoMb are indistinguishable (*9*), the equilibrium (pH 4) and kinetic molten globule states differ significantly in structure. Whereas in the wild-type protein the A, G, and H helices adopt stable helical structure in the burst phase kinetic intermediate, the H helix folds only during the slow refolding steps for the mutant apomyoglobin. This is shown both by the quench flow hydrogen exchange data (Figure 5) and by the reduced burst phase ellipticity, indicating lower helical content of the kinetic intermediate, in stopped flow CD experiments (*9*). The CD spectrum of the equilibrium molten globule state shows ~25% less helix in the mutant than in the wild-type protein (*9*), consistent with the loss of helical structure in the H helix region indicated by the chemical shift data (Figure 3 and Table 1). The work presented here thus provides direct evidence that the equilibrium and kinetic molten globules of the N132G/E136G mutant are extremely similar in structure, as they are for the wild-type protein. This is an important result because it establishes that mutations that modify the structure of the equilibrium molten globule cause parallel changes in the structure of the burst phase kinetic intermediate, providing compelling evidence that these two states are, to a first approximation, one and the same. It follows that detailed structural analysis of the equilibrium molten globule is of direct relevance to understanding the properties of the kinetic folding intermediate.

The NMR chemical shifts for the pH 4 state of the N132G/E136G mutant indicate a significant reduction of helical content in the H helix. In addition, it appears that the destabilization of the H helix has consequences for the stability of the G helix, which shows slightly reduced helicity (Figure 3). If it is assumed that G and H helices are packed

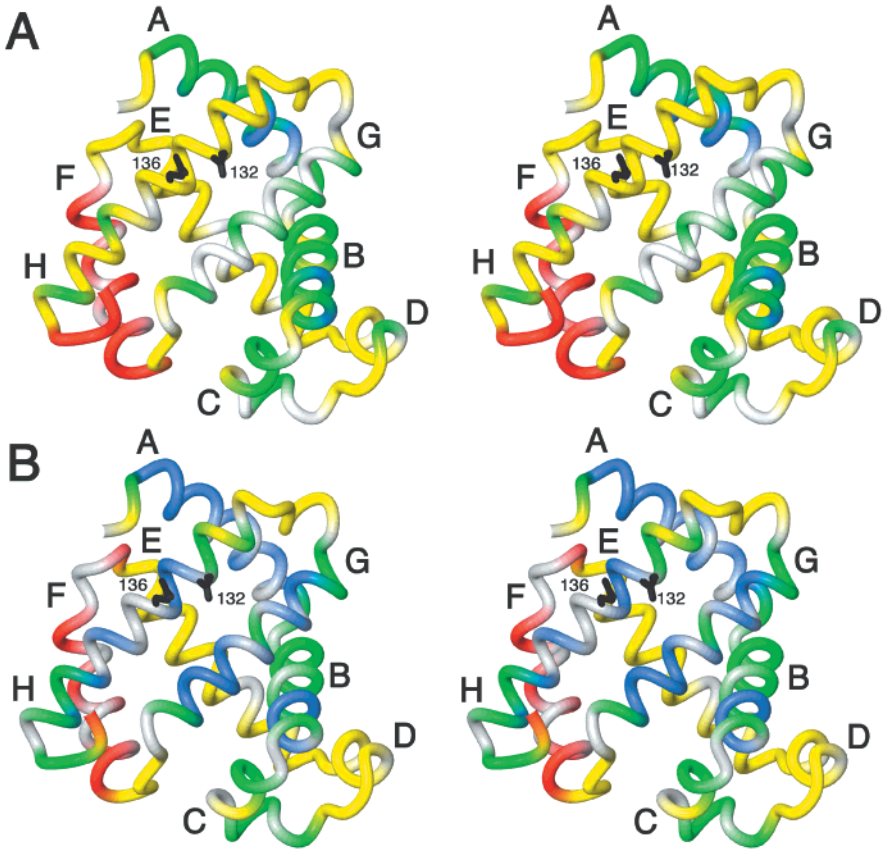


FIGURE 6: {$^1$H}−$^{15}$N heteronuclear NOE data for the pH 4.1 state of (A) the N132G/E136G mutant and (B) the wild-type protein, mapped onto the crystal structure of wild-type myoglobin (*48*). Data are from Figure 5: red, −1 < {NOE} < −0.1; yellow, −0.1 < {NOE} < 0.4; green, 0.4 < {NOE} < 0.6; and blue, {NOE} > 0.6. The figure was generated with the program MOLMOL (*49*).

together in the molten globule state of the wild-type protein, then the observed destabilization of the G helix in N132G/E136G probably reflects disruption of nativelike packing interactions with the H helix region.

*Changes in Dynamics Resulting from the Mutations.* Although the population of helical structure in the H helix region of the N132G/E136G molten globule is greatly diminished relative to that of the wild type (∼20 vs ∼50%, Table 1), the H helix segment does not behave as a free-flight random coil chain. Rather, the {$^1$H}−$^{15}$N heteronuclear NOEs are weakly positive, indicating a degree of motional restriction similar to that observed for the E helix (Figures 4 and 6). Motional restriction probably arises from transient, presumably hydrophobic, contacts between the H helix region and the compact core of the molten globule. Since the isolated, mutant H helix peptide shows no measurable propensity for spontaneous helix formation, the small (∼20%) residual helicity observed for the H helix in the molten globule state probably reflects formation of transient helical structure stabilized by these fluctuating contacts with the molten globule core. Further evidence that the H helix region remains partially associated with the core of molten globule comes from the similar fluorescence emission intensity for the pH 4 state of the mutant and wild-type proteins (Figure 1).

The destabilization of helical structure in the G helix region inferred from the chemical shift data is also reflected in slightly increased flexibility [decreased heteronuclear {$^1$H}−$^{15}$N NOEs (Figure 4)]. Furthermore, the flexibility of two residues near the middle of the A helix (Leu9 and His12) is increased in the N132G/E136G molten globule. These two residues are located precisely at the site where the A and H helices cross in the holomyoglobin structure (Figure 6). In contrast, the heteronuclear NOEs for the remainder of the protein, including the structured C-terminus of the B helix, are the same as for the wild type within experimental error. Thus, despite the absence of well-formed helical structure in the H helix, the backbone flexibility of the polypeptide chain remains the same as in the wild-type molten globule, with the exception of the H helix itself and of local regions that would contact the H helix in a native folding topology.

*Evidence for Nativelike Topology in the ApoMb Molten Globule.* The observed changes in backbone flexibility in localized regions distant from the site of mutation provide compelling evidence for nativelike packing of the polypeptide chain in the core of the molten globule. Thus, the loss of helical structure and increased flexibility in the G helix region of the N132G/E136G mutant would be expected if there are nativelike packing interactions between the G and H helices in the wild-type molten globule. Similarly, the slightly increased flexibility of residues 9 and 12 implies a population of molecules with nativelike contacts between the A and H helix regions, which are at least partially disrupted by the H helix mutations. The lack of motional perturbations in other regions of the polypeptide chain, from helices B−F, indicates that these regions make no contacts with the H helix in the molten globule state.

*Development of Secondary Structure during Kinetic Refolding.* As for wild-type apomyoglobin, the A and G helices and part of the B helix in the mutant protein adopt

stable helical structure in the burst phase intermediate formed within 6.4 ms of the initiation of refolding. However, in contrast to that of the wild-type protein, the mutated H helix does not fold during the burst phase; instead, this region of the molecule forms stable helical structure only during the slow phases of folding. Since the mutant H helix peptide has a negligible propensity to form helical structure spontaneously, formation and stabilization of secondary structure in the mutated H helix during refolding must be mediated by tertiary interactions. That is, the correct folding of the H helix sequence occurs as a result of interactions with the compact molten globule core, not as a result of the docking of a preformed helix.

The diffusion−collision model of protein folding (*38*, *39*) postulates that folding results from the productive collisions of preformed secondary structure elements. This model is consistent with the experimental findings for a number of single-domain helical proteins that fold rapidly by a two-state mechanism (*40*−*44*). The model has also been applied to the folding of apomyoglobin (*45*). Our results for the folding of the N132G/E136G mutant are partly consistent with the diffusion−collision model in that helix H folds more slowly when it is destabilized (*9*). They also, however, highlight some of its shortcomings. For instance, destabilization of the H helix has no effect on the overall rate of folding of apomyoglobin to its native state. In addition, studies of the equilibrium molten globule reveal that the H helix region does not diffuse freely prior to its incorporation into the molten globule core; the polypeptide makes at least transient hydrophobic contacts with the core that significantly restrict its backbone motions and lead to formation of helical structure. Folding data for the H64F mutant provide further evidence that the intrinsic helicity of local areas of the polypeptide is not the primary determinant of apomyoglobin folding rates or pathways. The H64F substitution causes no change in intrinsic helicity yet results in stabilization of the E helix during the burst phase due to improved hydrophobic packing (*11*). Despite the success of the diffusion−collision model in predicting folding rates for small helical proteins, our studies suggest that it is an oversimplification in the case of apomyoglobin, and point toward the importance of tertiary interactions in formation and stabilization of helical structure during protein folding.

Analysis of the urea dependence of the refolding and unfolding rates of N132G/E136G apomyoglobin suggests that the H helix, while largely unfolded in the burst phase intermediate, folds into stable helical structure before formation of the transition state ensemble (*9*). Since the quench flow data show that the H helix folds in apparent synchrony with the C and E helices, the slow folding regions of the B helix, and the CD loop (Figure 5), we conclude that these structural elements are also stabilized prior to the transition state. Thus, most of the helical structure of apomyoglobin appears to be already formed in the transition state ensemble. Since the intrinsic propensity of the C, E, and mutant H regions to spontaneously form helical structure is extremely small (ref *12* and this work), these results imply that formation and stabilization of helical structure are driven by tertiary interactions, and that the ability of these interactions to organize the secondary structure improves along the folding pathway.

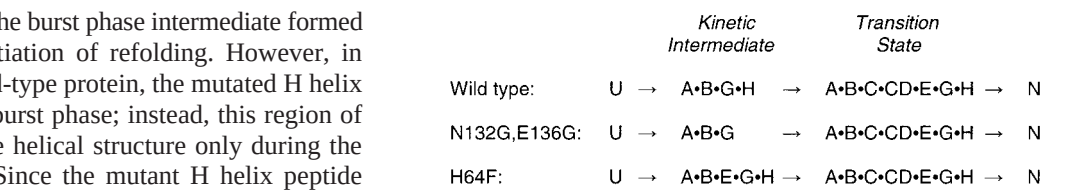


FIGURE 7: Summary of kinetic folding pathways for wild-type and mutant apomyoglobins. The information for H64F is from ref *11*.

The apomyoglobin folding process therefore proceeds by way of rapid collapse to form a compact hydrophobic core in which certain elements of secondary structure are stabilized. The precise nature of this kinetic intermediate, in terms of its helical content, is influenced by intrinsic helical propensities (this work) and by local hydrophobic packing interactions (*11*). The helical structure of the kinetic intermediate, for both the wild-type and mutant proteins, is very similar to that of the equilibrium molten globule intermediate formed at acidic pH. Apomyoglobin is rather tolerant of mutations in terms of its ability to fold. However, its folding pathway is not conserved but is strongly influenced by subtle changes in amino acid sequence (Figure 7). Thus, whereas the wild-type protein folds by way of an A−B−G−H helical intermediate, mutations that affect intrinsic helical stability (N132G/E136G) or local hydrophobic packing interactions (H64F) lead to changes in the structure of the kinetic intermediate. Given the mutability of the apomyoglobin folding intermediate, it is not at all surprising that the folding pathway of a distant member of the globin family, apoleghemoglobin, is not conserved (*46*). Nevertheless, a common feature of the folding pathway for all four proteins is the rapid formation (within a few milliseconds) of a compact, helical burst phase intermediate.

## ACKNOWLEDGMENT

We gratefully acknowledge helpful advice and technical assistance provided by Linda Tennant, Ted Foss, Gerard Kroon, John Chung, and Brendan Duggan.

## REFERENCES

1. Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 499−503.
2. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) *Science 249*, 1544−1548.
3. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
4. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O., and Tsuruta, H. (1995) *Science 270*, 487−488.
5. Eliezer, D., and Wright, P. (1996) *J. Mol. Biol. 263*, 531−538.
6. Eliezer, D., Jennings, P. A., Dyson, H. J., and Wright, P. E. (1997) *FEBS Lett. 417*, 92−96.
7. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
8. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) *Protein Sci. 8*, 45−49.
9. Cavagnero, S., Dyson, H. J., and Wright, P. E. (1999) *J. Mol. Biol. 285*, 269−282.
10. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 2894−2901.
11. Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 11227−11237.
12. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) *Biochemistry 40*, 3561−3571.

13. Jamin, M., and Baldwin, R. L. (1998) *J. Mol. Biol. 276*, 491−504.
14. Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
15. Van Geet, A. L. (1970) *Anal. Chem. 42*, 679−680.
16. 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.
17. Delaglio, F., Grzesiek, S., Vuister, G. W., Guang, Z., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
18. Johnson, B. A., and Blevins, R. A. (1994) *J. Chem. Phys. 29*, 1012−1014.
19. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
20. Löhr, F., and Rüterjans, H. (1995) *J. Biomol. NMR 6*, 189−197.
21. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
22. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 99*, 201−207.
23. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson. 101*, 201−205.
24. 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.
25. Stone, M. J., Chandrasekhar, K., Holmgren, A., Wright, P. E., and Dyson, H. J. (1993) *Biochemistry 32*, 426−435.
26. Stone, M. J., Fairbrother, W. J., Palmer, A. G., III, Reizer, J., Saier, M. H., Jr., and Wright, P. E. (1992) *Biochemistry 31*, 4394−4406.
27. Palmer, A. G., III (1996) http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html.
28. Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A., and Torchia, D. A. (1992) *Biochemistry 31*, 5253−5263.
29. Yao, J., Dyson, H. J., and Wright, P. E. (1997) *FEBS Lett. 419*, 285−289.
30. Bax, A. (1994) *Curr. Opin. Struct. Biol. 4*, 738−744.
31. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) *J. Biomol. NMR 18*, 43−48.
32. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) *J. Am. Chem. Soc. 123*, 2970−2978.
33. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
34. Bai, Y., Chung, J., Dyson, H. J., and Wright, P. E. (2001) *Protein Sci. 10*, 1056−1066.
35. Spera, S., and Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490−5492.
36. Wishart, D. S., and Nip, A. M. (1998) *Biochem. Cell Biol. 76*, 153−163.
37. Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., and Wright, P. E. (1993) *Biochemistry 32*, 6337−6347.
38. Karplus, M., and Weaver, D. L. (1979) *Biopolymers 18*, 1421−1437.
39. Karplus, M., and Weaver, D. L. (1994) *Protein Sci. 3*, 650−668.
40. Dobson, C. M., Sali, A., and Karplus, M. (1998) *Angew. Chem., Int. Ed. Engl. 37*, 868−893.
41. Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869−10873.
42. Huang, G. S., and Oas, T. G. (1995) *Biochemistry 34*, 3884−3892.
43. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) *Biochemistry 37*, 5337−5343.
44. Myers, J. K., and Oas, T. G. (2001) *Nat. Struct. Biol. 8*, 552−558.
45. Pappu, R. V., and Weaver, D. L. (1998) *Protein Sci. 7*, 480−490.
46. Nishimura, C., Prytulla, S., Dyson, H. J., and Wright, P. E. (2000) *Nat. Struct. Biol. 7*, 679−686.
47. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
48. Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133−154.
49. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.

BI011500N