FEBS Letters 417 (1997) 92–96 FEBS 19404

Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR

David Eliezer, Patricia A. Jennings[1], H. Jane Dyson, Peter E. Wright*

Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Received 11 September 1997

Abstract Conditions have been determined under which the equilibrium molten globule state of apomyoglobin is stable and remains monomeric for periods of time sufficient for the application of three-dimensional heteronuclear NMR experiments. The quality of initial two-dimensional NMR spectra suggests that sequence-specific assignments can be made for a majority of the protein resonances under these conditions. A pH titration of the protein followed using two-dimensional ^{1}H-^{15}N correlation experiments indicates that the equilibrium intermediate undergoes fast exchange on the chemical shift time scale with the unfolded state and intermediate time scale exchange with the native state, and suggests a strategy to assist with backbone resonance assignments. The conditions and techniques described may be applicable to the characterization of other equilibrium folding intermediates.

Key words: Apomyoglobin; Molten globule; Protein folding; NMR assignment

1. Introduction

Apomyoglobin has been observed to fold through a kinetic intermediate which shows a high degree of similarity to a partially unfolded equilibrium state populated under mildly acidic conditions [1–3]. Because of this similarity, it is expected that structural properties of the equilibrium 'molten globule' state reflect structure in the kinetic folding intermediate. While the kinetic intermediate is too short-lived to allow for a direct detailed structural characterization, the equilibrium state provides a unique opportunity for obtaining insights into structure formation during the folding process.

NMR is the technique best suited for detailed structural characterization of proteins in non-native conformations. Previous studies of the apomyoglobin equilibrium intermediate, however, have indicated that this state is prone to aggregation [3]. The conditions (low temperature and low concentration) which have been used previously to characterize the structure of this intermediate by amide proton exchange trapping [3] are not suitable for high-resolution NMR studies. Here we describe conditions under which apomyoglobin populates the equilibrium intermediate state, remains largely monomeric for the extended periods of time required to execute modern three-dimensional NMR pulse sequences, and yields high quality spectra. The assignment of resonances in the intermediate state remains difficult because of severe spectral overlap and line-widths which are broad compared to those in spectra of fully unfolded proteins. We demonstrate that in addition to the use of triple-resonance spectra, a series of two-dimensional heteronuclear NMR spectra collected throughout the equilibrium acid-induced unfolding of apomyoglobin can be used to transfer assignments from the acid-unfolded state (which is amenable to direct assignment strategies) to the equilibrium intermediate. Using two-dimensional NMR to monitor unfolding also provides information about the transition rates between the equilibrium intermediate and the native and acid-unfolded states.

2. Materials and methods

2.1. Sample preparation

For NMR experiments uniformly and selectively labeled apomyoglobin was expressed in *Escherichia coli* and purified from inclusion bodies as previously described [4,5]. Sample purity was greater than 95% as judged by analytical HPLC and electrospray mass spectrometry. Purified lyophilized protein was dissolved and unfolded in a solution containing 6 M urea and 10 mM sodium acetate at pH 6.1 and allowed to equilibrate for 30 min. This solution was then diluted 7-fold using 10 mM sodium acetate, pH 6.1 and the protein was allowed to refold (to the native state) and equilibrate for 60 min at 4°C. This solution was dialyzed extensively against 10 mM sodium acetate, pH 6.1, to remove the remaining urea and any remaining HPLC solvents. The native protein solution was then concentrated using Amicon gas-driven concentrators to approximately 300 μM final apomyoglobin concentration, as judged by UV absorption at 280 nm. 10% by volume D_2O was added during the concentration process. An alternative protocol in which lyophilized protein was dissolved directly in 10 mM sodium acetate, 10% D_2O, pH 6.1 and subsequently passed through a Sephadex G-25 column (equilibrated with the same buffer) was used in some of the more recent experiments. To populate the equilibrium intermediate state, native protein samples were titrated to pH 4.0 using a small volume of 3.6 M acetic acid. 10% by volume of ethanol was added to many of the samples to retard the aggregation of the intermediate state (as discussed below). Freshly prepared samples were used for all measurements. HCl was used to titrate samples below pH 3.3. All pH values were determined at room temperature.

For small angle X-ray scattering experiments, apomyoglobin was prepared starting with holomyoglobin from natural sources, using the method of Teale [6], as a 2 mM stock solution in 10 mM acetate buffer at pH 6.1. Scattering data from native apoprotein samples prepared at various concentrations using this stock solution (not shown) show no indication of oligomeric species. The stock solution was diluted two-, four-, eight-, and 16-fold to yield samples of 1 mM, 500 μM, 250 μM and 125 μM concentration. The intermediate state was then prepared from these samples as described above. Scattering data were collected using a portion of each of the freshly prepared samples of the intermediate, as described below. The remaining portion of each sample was sealed and stored in a bath at 50°C for a period of 52 h. At the end of this time, another set of scattering data were collected for the purpose of comparison with the data from the freshly prepared samples.

Samples for fluorescence and circular dichroism measurements were prepared in the same way as NMR samples, except that the final protein concentration used was about 2 μM. No ethanol was used

*Corresponding author. Fax: (1) (619) 784-9822.

[1]Present address: Department of Chemistry and Biochemistry, University of California, San Diego, CA 92037, USA.

PII S0014-5793(97)01256-8

for the fluorescence measurements. Circular dichroism data were collected both prior to and following the addition of 10% by volume ethanol.

2.2. Nuclear magnetic resonance

NMR spectra were collected on a Bruker 750 MHz DMX spectrometer using an actively shielded gradient triple-resonance probe. HSQC spectra [7] were collected using recently published pulse sequences [8]. These two-dimensional spectra exhibit a single correlation peak for each amide proton in the protein molecule. Each peak is labeled by the resonance frequencies of the proton itself and of the attached nitrogen atom.

2.3. Small angle X-ray scattering

Small angle X-ray scattering (SAXS) data were collected at the Stanford Synchrotron Radiation Laboratory using the Biotechnology Small Angle X-ray Scattering camera situated on beamline 4-2. The X-ray photon energy was calibrated to 8980 eV using a high-flux multi-layer X-ray monochromator. The sample-to-detector distance was 2.3 m. The sample was contained in a ceramic sample cell with thin (~20 μm) quartz windows and a 1 mm sample pathlength. The cell was kept in thermal contact with an aluminum block which in turn was maintained at 50°C using a circulating water-bath. Samples were exposed for 30 s each. Four to six such exposures were collected at each protein concentration. No evidence of radiation damage was observed.

2.4. Fluorescence and circular dichroism

Fluorescence data were collected on an SLM Instruments SPF-500C spectrofluorometer using a 1 cm pathlength cell with excitation at 278 nm (2 nm bandwidth) and detection at 320 nm (4 nm bandwidth). Circular dichroism (CD) data were collected on an Aviv model 61DS spectropolarimeter using a 1 cm pathlength cell. The sample cell holders were maintained at 50°C using a circulating bath. The signal was monitored until thermal equilibrium was reached before data were recorded.

3. Results and discussion

Direct NMR studies of partially folded equilibrium intermediates are made difficult by poor dispersion and broad lines (see for example [9,10]). It is therefore necessary to search for the 'best' conditions possible when attempting such studies. NMR spectra of proteins are temperature dependent, with spectra collected at higher temperatures typically exhibiting narrower lines. Thus, one possible approach to improving spectra from equilibrium intermediates is to work at higher temperatures when possible. The original description of the

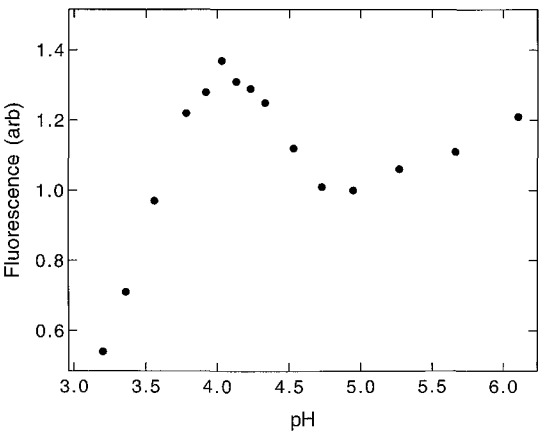

Fig. 1. Fluorescence from apomyoglobin in 10 mM acetate at 50°C as a function of pH. Excitation was at 278 nm and emission was monitored at 320 nm. The peak around pH 4 is characteristic of the 'molten globule' equilibrium intermediate state.

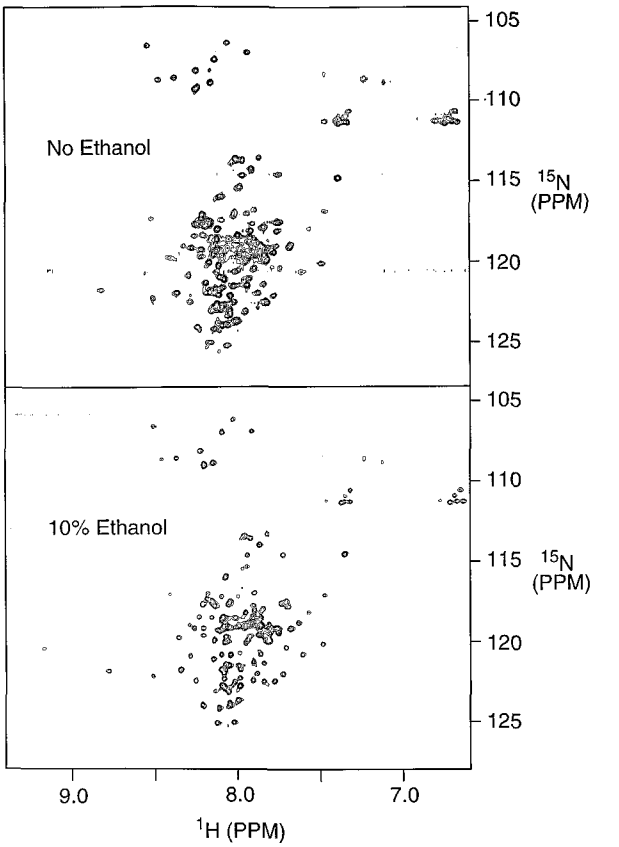

Fig. 2. ^{1}H-^{15}N HSQC spectra of apomyoglobin in 10 mM acetate, 50°C, pH 4.0 in the absence (top) and presence (bottom) of 10% by volume ethanol.

apomyoglobin equilibrium intermediate state [11] indicates that this state is populated at pH 4 over a range of temperatures between −6°C and 50°C. This is borne out in the NMR data presented below. Although most published studies have been performed at low (4°C) or room (25°C) temperatures, we decided to take advantage of the fact that this state is stable at temperatures as high as 50°C for the purpose of NMR characterization. A fluorescence monitored titration of apomyoglobin between pH 6.1 and pH 3.3 at 50°C (Fig. 1) exhibits the peak in fluorescence intensity which is characteristic of the equilibrium intermediate, and confirms the earlier observation by Griko et al. [11]. Work by Damaschun and colleagues [12] confirms that other properties characteristic of the intermediate state persist at high temperatures. While the helicity at pH 4 (as monitored by CD) is slightly lower at 50°C than at 4°C, the transitions between the pH 4 state and the acid-unfolded and native states are cooperative at both temperatures [13], implying that the intermediate is essentially fully populated.

^{1}H-^{15}N HSQC spectra of sub-millimolar samples of freshly prepared apomyoglobin at pH 4 in 10 mM acetate at 50°C (Fig. 2) appear significantly more tractable than those collected at room temperature or lower (data not shown). Nevertheless it is clear that in order to obtain resonance assignments, sophisticated heteronuclear triple-resonance experiments are required. Unfortunately, the quality of the spectra degrades severely within hours of sample preparation, making it impracticable to run the necessary triple-resonance experiments, which typically require several days of data collection. Equilibrium intermediates are in general highly prone

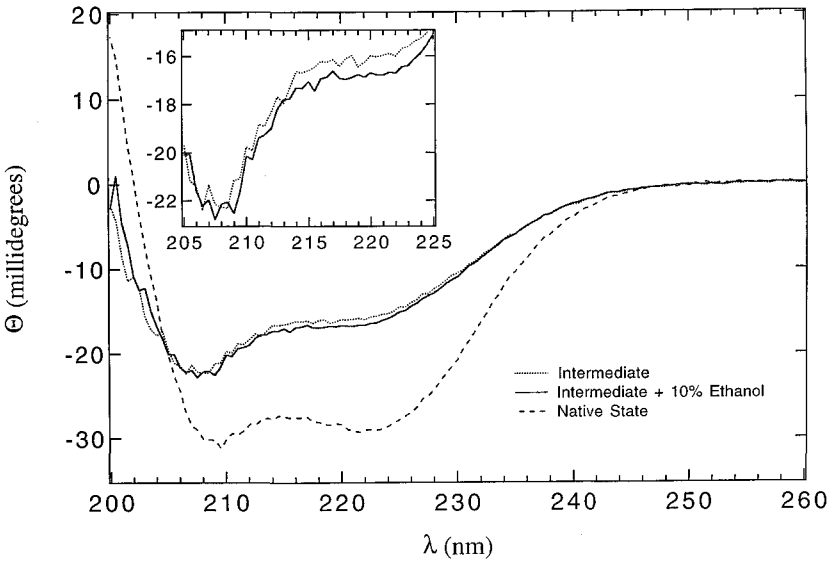

Fig. 3. Circular dichroism spectra of apomyoglobin in 10 mM acetate, 50°C in the native state (pH 6.3) and in the equilibrium intermediate state (pH 4.0) in the absence and presence of ethanol. The inset shows the change induced in θ_{222} by the presence of ethanol.

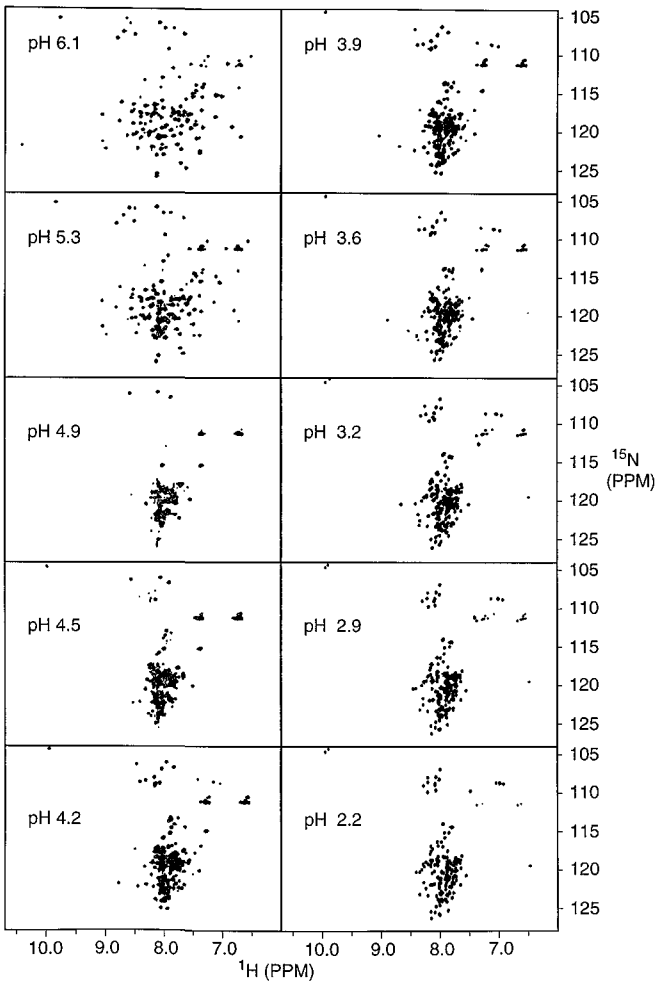

Fig. 4. ^{1}H-^{15}N HSQC spectra of apomyoglobin in 10 mM acetate, 50°C at different pH values between 2.3 and 6.1. The spectral degradation between pH 4 and 6 is thought to reflect transitions occurring between the native and intermediate states on an intermediate chemical shift time scale, which is estimated to be in the order of milliseconds.

to aggregation, which in turn severely degrades the quality of NMR spectra, and it is therefore likely that sample aggregation is responsible for the observed time-dependent degradation of the NMR spectra. Because aggregation is believed to be largely mediated by hydrophobic interactions, and because alcohols generally weaken hydrophobic interactions (see for example [14]), we attempted to eliminate or retard the aggregation of our samples by the addition of small volumes of ethanol. We found that by adding 10% by volume of ethanol to the samples, the quality of HSQC spectra remained acceptable for a period exceeding 3 days after sample preparation, providing sufficient time for the execution of sophisticated triple-resonance experiments. Although some loss in resonance intensity was observed after 3 days, the position of the resonances remained constant, and no new resonances were observed. After longer periods of time (a week) spectra were usually noticeably degraded, with split and missing resonances.

In order to evaluate the effects of adding 10% ethanol to our samples, we compared two-dimensional HSQC spectra of samples with and without ethanol (Fig. 2). Although slight shifts are observed in the positions of some resonances, in general the spectra are very similar in the presence and absence of ethanol. Fig. 3 shows circular dichroism data at 50°C from the native state and the intermediate with and without ethanol. A slight increase in θ_{222} is apparent in the presence of ethanol, but the corresponding increase in the total estimated helicity is only around 2%. Based on these observations we conclude that the addition of ethanol does not significantly perturb the structure of the apomyoglobin equilibrium intermediate state.

In order to ascertain more fully the extent of aggregation in our samples, we carried out a series of SAXS measurements (which are highly sensitive to sample aggregation) on samples of different protein concentrations. SAXS data (not shown) were collected from samples freshly prepared with 10% by volume ethanol at four different protein concentrations (1.0, 0.5, 0.25, and 0.125 mM) and from the same samples after storage at 50°C for 52 h. The fresh sample data at all four concentrations do not show any evidence of aggregation. The data at 1 mM protein concentration indicate that significant aggregation has occurred after 52 h. At 0.5 and 0.25 mM

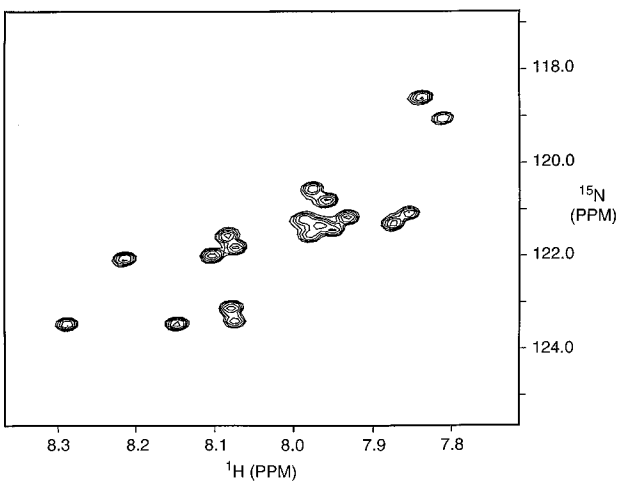

Fig. 5. ^{1}H-^{15}N HSQC spectra at pH 2.3, 50°C of apomyoglobin selectively ^{15}N-labeled at lysine backbone sites.

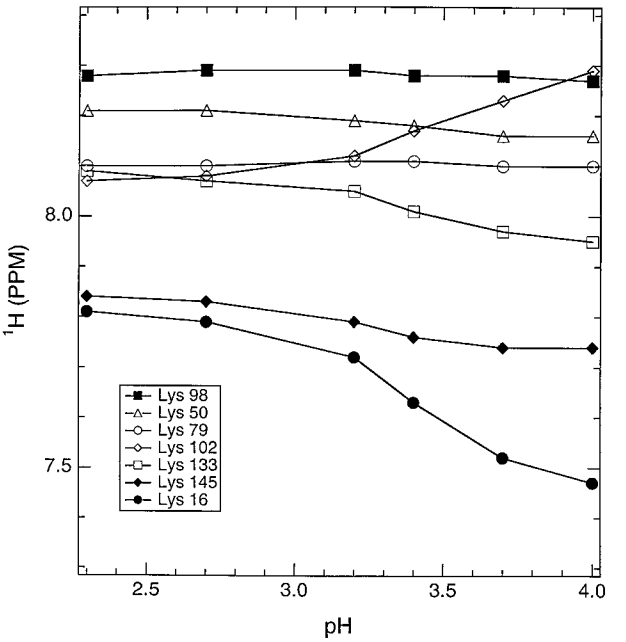

Fig. 6. ^{1}H chemical shift as a function of pH for several resonances in Fig. 5. Tentative assignments (based on assignments of the acid-unfolded state) are indicated.

protein concentration a slight increase in the small angle scattering suggests that after 52 h a small amount (less than 10%) of the protein sample may oligomerize. The data at 125 μM show essentially no change after 52 h, from which we conclude that no significant oligomerization or aggregation has occurred under these conditions. These results are consistent with the observations made using NMR, and indicate that at sufficiently low protein concentrations in the presence of 10% ethanol, the apomyoglobin equilibrium intermediate remains largely monomeric for a period of several days.

A variety of triple resonance NMR experiments have been acquired for the apomyoglobin intermediate under the above conditions. The resulting spectra are of high quality, but resonance assignment is made difficult by the rather poor resonance dispersion. In order to facilitate the assignment process, a strategy was devised to transfer resonance assignments from native apomyoglobin at pH 6.1 [5] to the pH 4 intermediate. By collecting two-dimensional HSQC spectra at incremental pH values between 6 and 4 (Fig. 4), we hoped to connect each resonance in the intermediate state with its corresponding assigned resonance in the native state. Unfortunately, there is a severe degradation in the quality of the spectra midway between the native and intermediate states, which most likely indicates that the transition between the native and intermediate states occurs at a rate which is intermediate on the chemical shift time scale. Using the known field strength and estimating the chemical shift changes involved, this time scale is calculated to be in the order of milliseconds. This behavior precludes the possibility of transferring resonance assignments from the native state to the pH 4 intermediate. However, spectra between pH 4 and pH 2.3 (where the protein is maximally unfolded by acid) remain of high quality, with resolved resonances throughout. Careful examination suggests that in each spectrum, each resonance appears only once, at a chemical shift value intermediate between that which it assumes at

the pH 4 and pH 2.3 endpoints. This indicates that the transitions between the pH 4 intermediate and the acid unfolded state occur at a rate which is fast on the chemical shift time scale (100 μs or less). Thus, although resonance assignments cannot be easily transferred from the native state to the pH 4 intermediate, assignments from the acid-unfolded state can be transferred to the intermediate state by simply tracking the position of each peak as the pH is raised incrementally.

The observation that the transitions between the intermediate and native state are slow (leading to exchange broadening of NMR resonances even at 50°C) while those between the acid-unfolded and intermediate states are fast is entirely consistent with previous kinetic data on the folding of apomyoglobin [1], which show that the intermediate state forms very rapidly, whereas the subsequent transition to the native state is orders of magnitude slower. The quantitative estimates above are also consistent with other published values. Jamin and Baldwin [15] report a time scale of 20 ms for the urea unfolding of the apomyoglobin pH 4 intermediate at 4.5°C. An increase in this rate to values such as we estimate above is reasonable given the difference in the temperatures (4.5°C versus 50°C).

The titration-based assignment strategy can be extended by using selectively labeled protein samples. Fig. 5 shows a ^{1}H-^{15}N HSQC spectrum of lysine-labeled apomyoglobin (prepared as part of the characterization of the native apoprotein [5]) at pH 2.3. The spectrum shows all 19 Lys resonances and is considerably simplified when compared to those in Fig. 4. Using such selectively labeled samples in a pH titration greatly facilitates the tracking of individual resonances. Fig. 6 shows the ^{1}H chemical shift of some of the better resolved Lys NH cross peaks in Fig. 5 as a function of pH between 2.3 and 4. The resonances which exhibit the largest changes appear to be in regions corresponding to the A, G, or H helices of the native protein. The large changes that occur with decreasing pH clearly reflect unfolding of the A-G-H helical core [3] of the intermediate state. An apparent pK of approximately 3.4 for the unfolding of the A-G-H intermediate can be estimated from Fig. 6. This value is in good agreement with fluorescence and CD data on unfolding of the intermediate state, and could be associated with the protonation of Asp or Glu residues, as is implicated in the unfolding transition (with an apparent pK of 3.6) of peptides corresponding to the B helix [16]. Resonances which fail to show significant chemical shift changes during the titration, e.g. Lys-79 and Lys-98, originate from residues which are believed to be in largely unstructured regions of the intermediate state [3]. Thus, chemical shift changes upon pH titration can be used to identify folded versus unfolded regions of the intermediate, and to determine whether its unfolding is a two-state or more complex process.

Using recently obtained resonance assignments of apomyoglobin at pH 2.3 [17] in conjunction with HSQC titration data and the triple-resonance data discussed above, the assignment of backbone resonances of the apomyoglobin pH 4 intermediate state has been nearly completed. The complete assignments will be published elsewhere.

4. Conclusions

We have successfully prepared long-lived samples of the apomyoglobin equilibrium intermediate under conditions suitable for the collection of high quality NMR data. In order to assign resonances in the highly overlapped spectra of the intermediate, we have devised a strategy which combines modern triple-resonance NMR experiments with the transfer of assignments from the acid-unfolded state (using selective labeling when necessary). The assignment transfer process has also yielded information about the transition rates between the native and intermediate states and between the intermediate and acid-unfolded states. Titration curves for individual resonances between pH 2.3 and 4 suggest that the largest structural changes between these two conditions occur in the A, G, and H helix regions. Extensive analysis of pH titration curves may help to evaluate the cooperativity of this transition, using probes located at residues throughout the polypeptide. The complete set of resonance assignments will lead to a wealth of information about the intermediate state through analysis of the chemical shift, relaxation, and NOE data and should provide a picture of both structural and dynamical propensities in a folding intermediate at an unprecedented residue-specific level of detail. We anticipate that the ideas and techniques described in this work may prove effective in NMR studies of partially folded states of other proteins as well.

Acknowledgements: This work was supported by Grants DK34909 (P.E.W.) and GM38794 (H.J.D.) from the National Institutes of Health and through an NIH National Research Service Award (D.E.). Small angle X-ray scattering data were collected at the Stanford Synchrotron Radiation Laboratory, which is supported by the U.S. Department of Energy and the NIH. We thank John Chung for assistance with NMR experiments, Jian Yao, Li Shi, Stefan Prytulla and Shohei Koide for helpful discussions, and Linda Tennant and Martine Reymond for expert technical assistance.

References

[1] Jennings, P.A. and Wright, P.E. (1993) Science 262, 892–896.
[2] Eliezer, D., Jennings, P.A., Wright, P.E., Doniach, S., Hodgson, K.O. and Tsuruta, H. (1995) Science 270, 487–488.
[3] Hughson, F.M., Wright, P.E. and Baldwin, R.L. (1990) Science 249, 1544–1548.
[4] Jennings, P.A., Stone, M.J. and Wright, P.E. (1995) J. Biomol. NMR 6, 271–276.
[5] Eliezer, D. and Wright, P.E. (1996) J. Mol. Biol. 263, 531–538.
[6] Teale, F.W.J. (1959) Biochim. Biophys. Acta 35, 543.
[7] Bodenhausen, G. and Ruben, D.J. (1980) Chem. Phys. Lett. 69, 185–189.
[8] Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) J. Biomol. NMR 4, 845–858.
[9] Baum, J., Dobson, C.M., Evans, P.A. and Hanley, C. (1989) Biochemistry 28, 7–13.
[10] Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. and Dobson, C.M. (1993) Biochemistry 32, 1707–1718.
[11] Griko, Y.V., Privalov, P.L., Venyaminov, S.Y. and Kutyshenko, V.P. (1988) J. Mol. Biol. 202, 127–138.
[12] Gast, K., Damaschun, H., Misselwitz, R., Muller-Frohne, M., Zirwer, D. and Damaschun, G. (1994) Eur. Biophys. J. 23, 297.
[13] Griko, Y.V. and Privalov, P.L. (1994) J. Mol. Biol. 235, 1318–1325.
[14] Thomas, P.D. and Dill, K.A. (1993) Protein Sci. 2, 2050–2065.
[15] Jamin, M. and Baldwin, R.L. (1996) Nature Struct. Biol. 3, 613–618.
[16] Kiefhaber, T. and Baldwin, R.L. (1995) J. Mol. Biol. 252, 122–132.
[17] Yao, J., Dyson, H.J. and Wright, P.E. (1997) in preparation.