Populating Partially Unfolded Forms by Hydrogen Exchange-Directed Protein Engineering

Jiro Takei,‡ Wuhong Pei,‡ Diep Vu, and Yawen Bai*

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 6114E, Bethesda, Maryland 20892

Received July 22, 2002; Revised Manuscript Received August 12, 2002

ABSTRACT: The native-state hydrogen exchange of a redesigned apocytochrome b_{562} suggests that at least two partially unfolded forms (PUFs) exist for this four-helix bundle protein under native conditions. The more stable PUF has the N-terminal helix unfolded. To verify the conclusion further and obtain more detailed structural information about this PUF, five hydrophobic core residues in the N-terminal helix were mutated to Gly and Asp to destabilize the native state selectively and populate the PUF for structural studies. The secondary structure and the backbone dynamics of this mutant were characterized using multidimensional NMR. Consistent with the prediction, the N-terminal region of the mutant was found to be unfolded while other parts of the proteins remained folded. These results suggest that native-state hydrogen exchange-directed protein engineering can be a useful approach to populating partially unfolded forms for detailed structural studies.

To understand the folding mechanism and function of a protein, it is important to characterize the structures of the protein at various thermodynamic states under native conditions, which includes unfolded, partially unfolded, and fully folded states. In favorable cases, native-state hydrogen exchange can identify partially unfolded forms (PUFs)[1] that are more stable than the unfolded state but are less stable than the native state. From such experiments, the structure of the PUFs can be inferred in terms of nativelike hydrogen bonding (*1*−*3*). To understand how PUFs are stabilized, however, it is necessary to determine the structure of PUFs at the atomic level. So far, the only method of obtaining a protein's structure at an atomic resolution in solution is by NMR. Nevertheless, this technique cannot be directly applied to determine the structure of the PUFs, since it requires that PUFs be the dominant species and soluble at a concentration of ∼1 mM. Under native conditions, PUFs only exist as a rarely populated species. Thus, new methods are needed to populate PUFs selectively as the dominant and soluble species. Figure 1 illustrates a native-state hydrogen exchange-directed protein engineering approach to achieving such a goal.

Previously, native-state hydrogen exchange on a redesigned four-helix bundle protein (Rd-apocyt b_{562}) based on apocytochrome b_{562} revealed that at least two PUFs exist under native conditions (*4*). The more stable PUF, PUF2 (see Figure 1), has the N-terminal helix (helix I) unfolded and other helices folded. Since mutations in the unfolded region of a PUF would have little effect on the stability of the PUFs (relative to the unfolded state) but could have a large effect on the stability of the native state, PUF2 may be selectively populated by mutations in the unfolded region of the PUF to destabilize the native state.

MATERIALS AND METHODS

Sample Preparation. The mutant named 4GD7 (W7D/L10G/L14G/V16G/I17G) was made on the basis of Rd-apocyt b_{562} (*4*) using the QuickChange kit (Stratagene). It was made consecutively with five single-site mutations. ^{15}N-labeled proteins and ^{13}C- and ^{15}N-labeled proteins were prepared as described previously (*4*). Samples for NMR studies were prepared by dissolving the lyophilized protein in 30 mM NaAc buffer (5% D_2O and 95% H_2O) and passing it through a 3 mL bed volume spin column loaded with G25 (fine grade) gel filtration resin and equilibrated with the same buffer. The final protein concentration is ∼1 mM for resonance assignment and dynamics studies.

Thermal Melting by Circular Dichroism. Thermal melting of 4GD7 was performed using a Jasco CD (J-720) instrument and monitored at 222 nm. The temperature increment was 1 °C/min. The protein concentration was ∼50 μM in the 30 mM NaAc buffer solution.

NMR Spectroscopy. Triple-resonance spectra for backbone resonance assignment of the 4GD7 mutant using the ^{15}N- and ^{13}C-labeled sample were recorded at pH 5.0 and 25 °C on a Bruker DRX 500 MHz NMR spectrometer. Three-dimensional (3D) CBCACONH (*5*), HNCACB (*6*), HNCO (*7*), and HN(CA)CO (*8*) experiments were carried out for backbone assignment. ^{1}H and ^{13}C chemical shifts were referenced to an internal standard (DSS), and ^{15}N chemical shifts were referenced indirectly to DSS using a ratio of 0.10132912 (*9*). In backbone dynamics studies (*10*) on Rd-apocyt b_{562}, T_1 inversion−recovery delays were set to 2, 5000, 5, 2000, 10, 1000, 20, 500, 50, 200, and 100 ms, and for T_2 CPMG delays, they were 1, 5000, 2, 2000, 5, 500, 10, 200, 20, 100, and 50 ms with a duty cycle of 0.45 ms. For the mutant 4GD7, T_1 inversion−recovery delays were

* To whom correspondence should be addressed. Telephone: (301) 594-2375. Fax: (301) 402-3095. E-mail: yawen@helix.nih.gov.

‡ These authors contributed equally.

[1] Abbreviation: PUF, partially unfolded form.

10.1021/bi026491c This article not subject to U.S. Copyright. Published 2002 by the American Chemical Society
Published on Web 09/20/2002

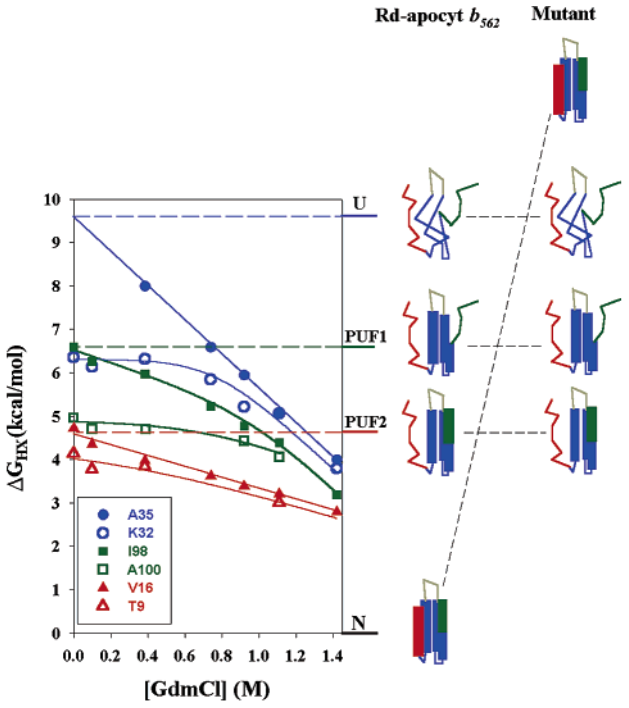

FIGURE 1: Illustration of the principle of the native-state hydrogen exchange-directed protein engineering for populating partially unfolded forms. The plot of ΔG_{HX} vs [GdmCl] for different species is adopted from Chu et al. (*4*). Mutations in helix I that increase the free energy of the native state relative to that of the unfolded state will have little effect on the free energy of the partially unfolded forms since the mutation sites are unfolded in the partially folded intermediates.

set to 2, 5000, 100, 3000, 5, 1000, 50, 10, 30, 40, 70, 3000, 15, and 90 ms. For T_2 CPMG delays, they were 1000, 2, 100, 10, 2, 50, 200, 80, 150, 30, 10, and 60 ms. In both T_1 and T_2 measurements, the initial relaxation delays were set to 3 s. Steady-state NOE values were determined by recording in the presence (NOE) and absence (NONOE) of ^{1}H saturation for 3 s. NOE and NONOE experiments were interleaved in one experiment to minimize the effect of the uncertainty of the instrument. Duplicate measurements of T_1, T_2, and NOE were performed to estimate the error.

Data Processing and Analysis. NMR spectra were processed using NMRPipe (*11*) and analyzed using NMRView (*12*). Relaxation time constants (T_1 and T_2) for each amide nitrogen were calculated by fitting HSQC peak intensities versus time to the simple exponential decay function $I(t) = I_0 \exp(-t/T)$, where $I(t)$ is the peak intensity, I_0 is the peak intensity at time zero, and T is the time constant of relaxation. NOE values were calculated by dividing the peak intensities in the NOE experiment by those in the NONOE experiment. The standard deviations for T_1, T_2, and NOE are estimated using the two independently measured data sets.

The overall tumbling correlation time, τ_m, was obtained from the T_1/T_2 ratio according to the method of Kay et al. (*13*). Residues with a more than 5% error in T_1, T_2, or NOE and those with an NOE of <0.6 were excluded in the calculation of τ_m. T_1, T_2, and NOE of an amide ^{15}N nucleus were calculated using the following equations (*14*):

$$1/T_1 = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad (1)$$

$$1/T_2 = 0.5d^2[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + {}^1/_6 c^2[3J(\omega_N) + 4J(0)] \quad (2)$$

$$\mathrm{NOE} = 1 + d^2(\gamma_H/\gamma_N)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]T_1 \quad (3)$$

The constants d^2 and c^2 are equal to $[(\mu_o h/8\pi^2)\gamma_H\gamma_N\langle 1/r_{NH}^3\rangle]^2$ and $[\omega_N(\sigma_{||} - \sigma_\perp)/3]^2$, respectively, where γ is the gyromagnetic ratio, h is Plank's constant, μ_o is the permeability of a vacuum, r_{NH} is the N−H bond length, taken to be 1.02 Å, and $\sigma_{||} - \sigma_\perp$ is the axially symmetric chemical shift tensor for backbone ^{15}N nuclei, taken to be −160 ppm.

Order parameters (S^2) were determined using the model free approach (*15*) with the spectral density function presented in eq 4.

$$J(\omega) = (S^2\tau_m)/(1 + \omega^2\tau_m^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2) \quad (4)$$

where $1/\tau = 1/\tau_m + 1/\tau_e$. The order parameters were obtained by minimizing the χ^2 value described in eq 5.

$$\chi^2 = [(T_{1c} - T_{1e})/\sigma_{T_1}]^2 + [(T_{2c} - T_{2e})/\sigma_{T_2}]^2 + [(\mathrm{NOE_c} - \mathrm{NOE_e})/\sigma_{NOE}]^2 \quad (5)$$

where subscripts c and e denote the calculated and experimentally obtained values, respectively, and σ_{T_1}, σ_{T_2}, and σ_{NOE} are the standard deviations of T_1, T_2, and NOE, respectively.

RESULTS

Thermal Melting and Secondary Structure Mapping by NMR. To test the idea illustrated in Figure 1, a 5-fold mutation, W7D/L10G/L14G/V16G/I17G, was made to replace the hydrophobic residues in helix I with Asp and Gly. These mutation sites are part of the core of Rd-apocyt b_{562} (see Figure 2). It was designed to destabilize the native state selectively and to minimize the possibility of aggregation. The 5-fold mutant termed 4GD7 was expressed in *Escherichia coli* and purified using reverse phase HPLC. To see whether the 4GD7 mutant is thermodynamically stable, a thermal melting experiment was performed using circular dichroism by monitoring the signal at 222 nm. The CD spectrum and the melting curve are shown in Figure 3. Despite the heavy mutations, the 4GD7 mutant is thermodynamically stable. The middle point of the melting curve is ~57 °C at pH 5.0. This result further supports the conclusion that PUF2 exists in Rd-apocyt b_{562} since otherwise such heavy mutations would have strongly destabilized the native state and completely unfolded the protein.

To characterize the structure of 4GD7 by NMR, the mutant was uniformly labeled with ^{15}N and/or ^{13}C. The backbone resonance was assigned at 25 °C and pH 5.0. The ^{1}H−^{15}N correlation spectrum is shown in Figure 4. The spectrum has a very good dispersion with few overlapped peaks. All of the peaks for residues 2−23 exist in the region between 8.3 and 8.5 ppm with higher intensity, a characteristic for an unfolded polypeptide. Figure 5 illustrates the deviation of the $^{13}C_\alpha$ chemical shifts from the random coil values (secondary shifts), a sensitive indicator of secondary structure (*16*). It clearly shows that the secondary $^{13}C_\alpha$ chemical shifts of the residues in helices II−IV are significantly larger than

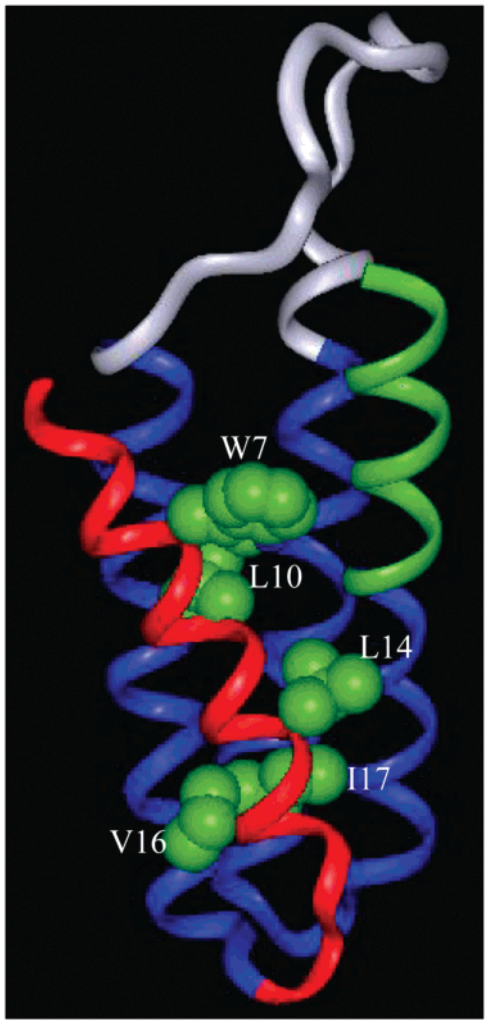

FIGURE 2: Structural illustration of the residues mutated in Rd-apocyt b_{562}.

the random coil values and similar to those of Rd-apocyt b_{562}. For the residues 1−23, however, they are close to the random coil values, indicating that this region is completely unfolded.

Backbone Dynamic Behavior of 4GD7. To characterize the mutant 4GD7 further, the dynamic behavior of the backbone was studied. In the dynamic studies, we measured the longitudinal (T_1) and transverse (T_2) relaxation times of the amide ^{15}N and the nuclear Overhauser effect (NOE) between the amide nitrogen and amide proton for 4GD7. For the purpose of comparison, the same dynamic parameters were also measured for Rd-apocyt b_{562}. The results are shown in Figure 6. The T_1, T_2, and NOE values are essentially uniformly distributed throughout the sequence of Rd-apocyt b_{562} except in the N- and C-termini, indicating that the molecule is well-folded. Similar results were observed for the residues 24−106 in 4GD7. However, significantly longer T_2 and negative NOE values for the NH group of residues 1−23 were observed in 4GD7. In addition, the $1/T_1$ and T_2 values for the residues in the folded region of 4GD7 are slightly larger than those in Rd-apocyt b_{562}, suggesting that the effective size of 4GD7 is slightly larger than that of Rd-apocyt b_{562}. These results support the conclusion that the region of residues 1−23 is unfolded in 4GD7.

Order Parameters. To illustrate the dynamic behavior quantitatively, the dynamic data were analyzed using the

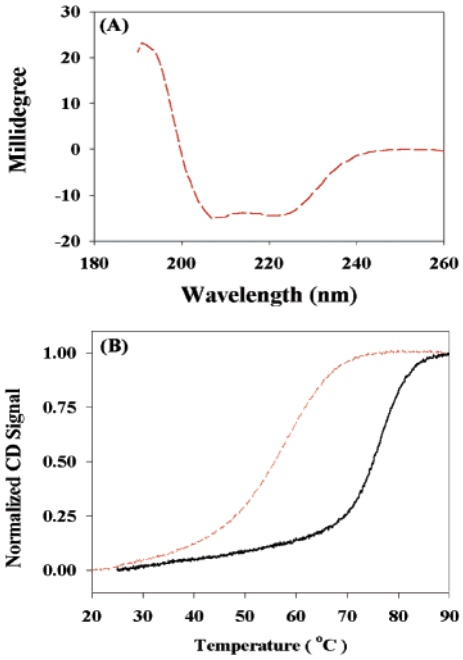

FIGURE 3: (A) CD spectrum of the 4GD7 mutant. (B) Normalized melting curve of 4GD7 (− − −) and Rd-apocyt b_{562} (—) monitored at 222 nm.

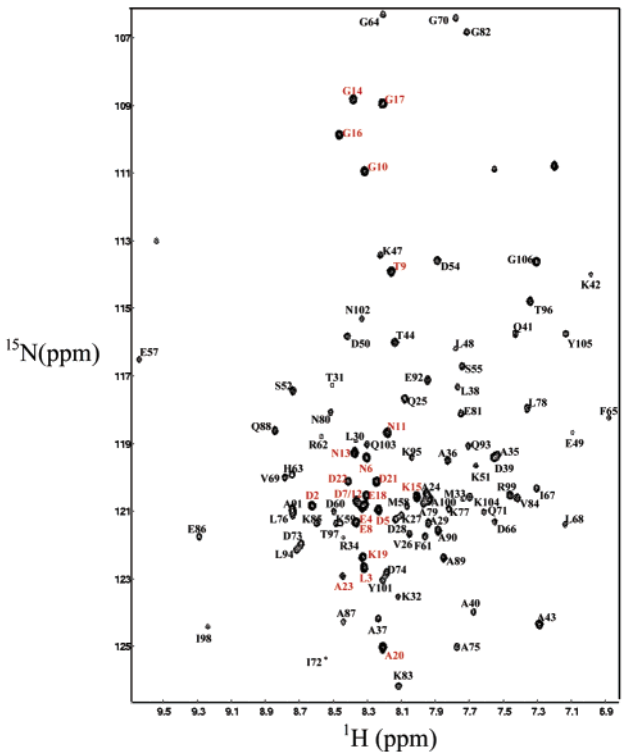

FIGURE 4: ^{1}H−^{15}N HSQC spectrum of 4GD7. The residues labeled in red are in helix I.

model free approach (*15*). In this approach, the dynamic behavior of a protein is described in terms of rotational correlation time (τ_m), order parameter (S^2), and effective correlation time τ_e. Here τ_m describes the tumbling rate for the whole protein. The order parameter describes the amplitude of the motions for each N−H vector, which can range between 0 and 1, representing complete disorder and complete order, respectively. τ_e was used to describe the effective internal motion of each N−H vector. The values

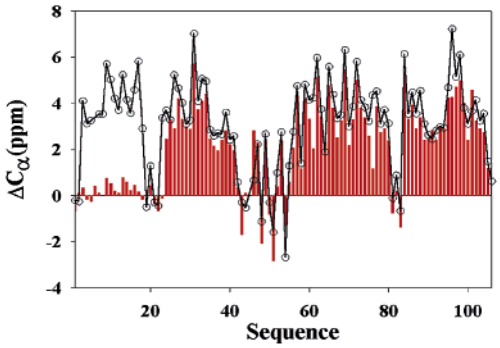

FIGURE 5: C_α chemical shift deviation from random coil values in Rd-apocyt b_{562} (○) and 4GD7 (|).

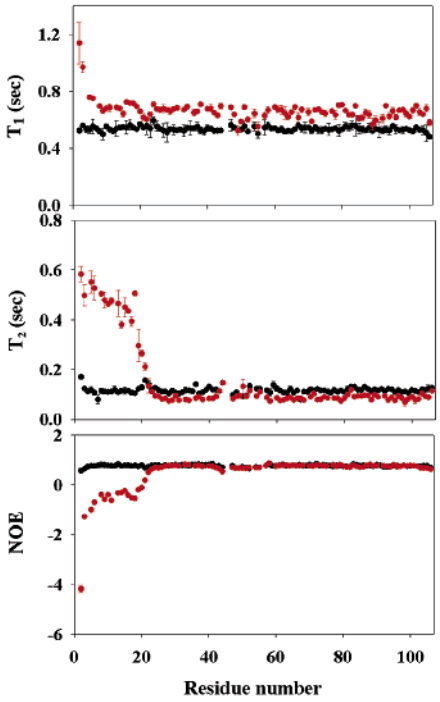

FIGURE 6: T_1, T_2, and NOE values for Rd-apocyt b_{562} (black) and 4GD7 (red).

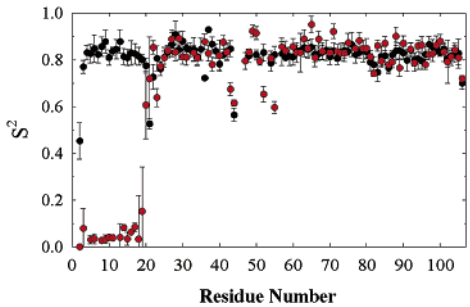

FIGURE 7: Order parameters for Rd-apocyt b_{562} (black) and 4GD7 (red).

of τ_m for 4GD7 and Rd-apocyt b_{562} are 9.8 and 7.2 ns, respectively. The larger τ_m for 4GD7 is consistent with the conclusion that helix I is unfolded in 4GD7. The order parameters for the two proteins were obtained by minimizing the χ^2 value described in eq 5 (see Materials and Methods). They were plotted in Figure 7. In the case of Rd-apocyt b_{562}, the order parameters for the residues in the helical structures are near or above 0.8. Only in the turn and terminal regions were smaller order parameters observed, suggesting that the helical regions are well-ordered in the structure of Rd-apocyt b_{562}. This is also the case for residues 24−106 in 4GD7. In contrast, the order parameters decrease dramatically to near zero in the region of residues 1−20, indicating helix I is completely unfolded.

DISCUSSION

Population of Partially Unfolded Forms for Structural Studies. So far, NMR is the only method of obtaining structural information of proteins at the atomic level in solution. However, it is not directly applicable to the study of the folding intermediates since it takes days to collect the data necessary for structural determination and needs a sample with a concentration of ~1 mM. Folding intermediates, on the other hand, populate only transiently during kinetic folding. Under native equilibrium conditions, they exist as only minor species. Therefore, to study the intermediates by NMR, approaches that can selectively populate folding intermediates or their mimics are needed. The first approach used to study the structure of folding intermediates was to make peptide mimics of the folded region of the intermediate. It has been successfully applied to study the folding intermediates involving formation of disulfide bonds in BPTI (*17*). However, previous attempts to use this approach to study partially unfolded states of other proteins in the absence of disulfide bonds were unsuccessful due to aggregations at concentrations suitable for structural characterization by NMR [see reviews by Peng and Wu (*18*)]. The second approach used for populating folding intermediates is the denaturation of proteins under mild denaturing conditions such as low pH. A typical example is the population of the intermediate of apomyoglobin at pH 4 (*19*, *20*). In the case of Rd-apocyt b_{562}, however, the protein remains folded under acidic conditions (unpublished results). The native-state hydrogen exchange-directed protein engineering method presented here provides another way of populating PUFs. A similar approach has been used to obtain an isolated ribonuclease H core fragment (*21*).

Gly Mutations as a General Method for Probing PUFs. The Gly mutations in helix I of Rd-apocyt b_{562} lead to the population of the PUF identified by the native-state hydrogen exchange. Gly mutations may be used as a general approach to investigating PUFs even in the absence of the information from native-state hydrogen exchange experiments in a way similar to denaturation under acidic conditions. If a protein can exist as a PUF with Gly mutations, one may infer that the region folded in the PUF has the ability to fold alone in the wild-type protein and is a potential folding intermediate. In fact, the approach of using Gly mutations in principle is more rational than the approach of using low pH to denature proteins. At low pH, all Asp and Glu residues in the protein were mutated to their respective neutral form regardless of whether they were in the folded or unfolded region.

Obviously, Gly mutations and denaturation at low pH alone cannot lead to a definitive conclusion about whether populated partially unfolded intermediates exist on the kinetic folding pathways of the proteins. Further kinetic testing is needed. In addition, mutations may have changed the folding pathways of the proteins. Nevertheless, the population of a thermodynamically stable partially unfolded state by Gly mutants can suggest the possible existence of a stable

intermediate and substantiate the conclusion that proteins may fold hierarchically in terms of structure when intermediates cannot be detected using other methods (*22*). It should also be noted that observation of kinetic intermediates in folding experiments alone does not prove that the intermediates are on the folding pathway either. A specific demonstration of whether the observed kinetic intermediate is on-pathway or off-pathway is necessary.

PUFs with Defined Structures under Native Conditions. One of the properties of 4GD7 is that it has a well-resolved NMR spectrum with sharp peaks even though the molecule is partially folded. The T_2 values for the residues in the folded region of the molecule are essentially uniform and smaller than those in the unfolded region. This characteristic indicates that there is little conformational heterogeneity in this PUF. This result is different from the partially unfolded states found in many other proteins under acidic conditions. In those cases, conformational changes on the millisecond time scale occurred, which are commonly described as molten globules (*23*).

Implications for the Studies of Amyloid Formation. The process of amyloid fibril formation involves conformational changes of a normally soluble protein and self-assembly into an insoluble fibrillar state. It is suggested that formation of the amyloid is the cause of many diseases such as Alzheimer's disease (*24*). Partially folded intermediates have been inferred as the precursors for formation of the amyloid in several proteins (*25*−*27*). To understand the process of amyloid formation, it is important to obtain detailed structural information about these partially folded intermediates. So far, the conformational properties of these species have been determined only for limited cases at a residue-specific level using amide hydrogen exchange and equilibrium denaturation coupled with NMR (*27*). No detailed information at the atomic level was available. In addition, these intermediates were populated at low pHs which are not physiological conditions. It is possible that the conformational properties of the intermediates at low pH may not be the same as those under physiological conditions. Thus, it seems necessary to study the intermediates formed under neutral-pH conditions, and the approach presented here may provide an alternative way for studying the amyloid precursor at the atomic level under native conditions.

CONCLUSIONS

To understand the mechanism of protein folding, it is necessary to characterize the structure of partially unfolded intermediates in detail. This is a difficult task since partially unfolded intermediates populate only transiently in kinetic folding. Under native and equilibrium conditions, they are much less populated than the native state. Therefore, the structures of the partially folded intermediates cannot be directly studied at the atomic level using NMR. In this paper, we tested a native-state hydrogen exchange-directed protein engineering method for populating the partially unfolded intermediates, which allows the structure of partially unfolded forms to be studied by NMR at atomic resolutions. It is shown that the partially unfolded state of the four-helix bundle protein can be selectively populated through substituting the large hydrophobic core residues with Gly and the charged residue Asp in the unfolded region of the PUF.

ACKNOWLEDGMENT

We thank Dr. Rieko Ishima for providing us with the program and help for the model free analysis and Dr. S. Walter Englander for comments on the manuscript.

REFERENCES

1. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
2. Raschke, T. M., and Marqusee, S. (1998) *Curr. Opin. Biotechnol. 9*, 80−86.
3. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 105−112.
4. Chu, R. A., Pei, W. H., Takei, J., and Bai, Y. (2002) *Biochemistry 41*, 7998−8003.
5. Grzesiek, S., Ikura, M., Gronenborn, A. M., Clore, G. M., and Bax, A. (1992) *J. Magn. Reson. 96*, 215−221.
6. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson., Ser. B 101*, 201−205.
7. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
8. Clubb, R. T., Thanabal, V., and Wagner, G. (1992) *J. Magn. Reson. 97*, 213−217.
9. 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.
10. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Key, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
11. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
12. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR 4*, 603−614.
13. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972.
14. Abragam, A. (1961) *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, U.K.
15. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
16. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
17. Oas, T. G., and Kim, P. S. (1988) *Nature 336*, 42−48.
18. Peng, Z. Y., and Wu, L. C. (2000) *Adv. Protein Chem. 53*, 1−47.
19. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
20. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 2894−2901.
21. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M., and Marqusee, A. S. (1999) *Protein Sci. 8*, 2251−2257.
22. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 26−33.
23. Arai, M., and Kuwajima, K. (2000) *Adv. Protein Chem. 53*, 209−282.
24. Dobson, C. M. (1999) *Trends Biochem. Sci. 24*, 329−332.
25. Jiang, X., Smith, C. S., Petrussi, H. M., Hammarstrom, P., White, J. T., Sacchettini, J. C., and Kelley, J. W. (2001) *Biochemistry 40*, 11422−11452.
26. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V., and Dobson, C. M. (2002) *Nat. Struct. Biol. 9*, 308−315.
27. McPharland, V. J., Kalverda, A. P., Homans, S. W., and Radford, S. E. (2002) *Nat. Struct. Biol. 9*, 326−331.

BI026491C