Relationship between the Native-State Hydrogen Exchange and Folding Pathways of a Four-Helix Bundle Protein

Ruiai Chu,‡ Wuhong Pei,‡ Jiro Takei, and Yawen Bai*

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

Received March 26, 2002; Revised Manuscript Received May 3, 2002

ABSTRACT: The hydrogen exchange behavior of a four-helix bundle protein in low concentrations of denaturant reveals some partially unfolded forms that are significantly more stable than the fully unfolded state. Kinetic folding of the protein, however, is apparently two-state in the absence of the accumulation of early folding intermediates. The partially unfolded forms are either as folded as or more folded than the rate-limiting transition state and appear to represent the major intermediates in a folding and unfolding reaction. These results are consistent with the suggestion that partially unfolded intermediates may form after the rate-limiting step for small proteins with apparent two-state folding kinetics.

Many small proteins (ca. ≤120 amino acids) fold in a kinetically apparent two-state manner without the accumulation of partially folded intermediates (*1*, *2*). The folding studies on these proteins have focused on characterization of the rate-limiting transition state in an effort to understand the structural basis for the initiation of folding (*3*, *4*). Although it is commonly assumed that these proteins do not have partially unfolded intermediate states, theoretically, partially unfolded intermediates that are more stable than the unfolded state could still exist after the rate-limiting step of folding (*5*). Such intermediates hold important information for understanding the mechanism of protein folding (*6*) and perhaps for the function of proteins (*7*). Studies on these intermediates are lacking because conventional kinetic studies cannot detect folding events that occur after the rate-limiting step under native conditions.

Thermodynamic principles, however, require that protein molecules populate all possible states under equilibrium conditions and continually cycle through them according to their Boltzmann factors. A native-state hydrogen exchange approach can depict the structure and measure the free energy of the partially unfolded forms (PUFs)[1] through which proteins cycle under native and equilibrium conditions, even though they are much less populated than the native state (*8*, *9*). For example, three PUFs have been identified under native conditions for cytochrome *c* (*8*). Surprisingly, however, none of the three PUFs can be detected in the kinetic folding experiment when a misfold reorganization barrier is absent (*10*). These observations have led to the suggestion that the PUFs may exist after the initial rate-limiting step of folding (*5*, *6*, *11*). If this suggestion is correct, the native-state hydrogen exchange approach, coupled with traditional kinetic studies, provides a way to reveal the free energy landscape after the rate-limiting step of folding. A necessary condition for this hypothesis to be valid would be that the initial rate-limiting transition state not be more folded than the late intermediates. In this study, we investigated this issue by studying the relationship between the native-state hydrogen exchange and kinetic folding pathways of a redesigned four-helix bundle protein termed Rd-apocyt b_{562} based on apocytochrome b_{562}. We show that the folding behavior of the four-helix bundle protein satisfies such conditions and therefore further supports the late intermediate hypothesis.

MATERIALS AND METHODS

Proteins are expressed in *Escherichia coli* (BL21DE3) using the pET-17b vector and purified using ion exchange and reverse phase HPLC (water and acetonitrile). Mutations were introduced using a Quick-change mutagenesis kit (Stratagene, La Jolla, CA). Urea and GdmCl are ultrapure grade (Sigma). Their concentrations were measured using refractive indexes (*12*).

Stopped-Flow Kinetic Folding and Equilibrium Unfolding Experiments. Stopped-flow kinetic folding and unfolding were performed with a Biologic (Grenoble, France) SFM4 machine. Kinetic rate constants were measured by monitoring the change in the fluorescence signal. The excitation wavelength was 280 nm, and emission was collected at 320 nm. Folding was initiated by diluting proteins at high denaturant concentrations with folding buffer [50 mM NaAc (pH 5.0)]. Unfolding was initiated by mixing the protein solution in water with a high concentration of denaturant. The observed rate constants, k_{obs} ($=k_f + k_u$), versus denaturant concentrations were fitted with the following equations:

$$\log k_f([\text{urea}]) = \log k_f(H_2O) - m_f^{kin}[\text{urea}] \quad (1)$$

$$\log k_u([\text{urea}]) = \log k_u(H_2O) + m_u^{kin}[\text{urea}] \quad (2)$$

where k_f and k_u are the folding and unfolding rate constants, respectively.

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

‡ These authors contributed equally to this work.

[1] Abbreviations: RLTS, rate-limiting transition state; PUF, partially unfolded form; GdmCl, guanidinium chloride.

10.1021/bi025872n
Published on Web 06/01/2002

Native-State Hydrogen Exchange Experiments. The exchange reaction was initiated by dissolving protein samples with deuterated amide protons in an exchanging solution using a spin column filled with G25 Sephadex and pre-equilibrated with exchanging buffer (50 mM NaAc-*d*, 95% H_2O, and 5% D_2O). The deuterated Rd-apocyt b_{562} protein was prepared by dissolving proteins in D_2O (pD 7.0, 0.1 M Na_2PO_4) and heated at 60 °C for more than 30 min. $^{15}N-^{1}H$ HSQC spectra were recorded using a 500 MHz Avance instrument (Bruker). Protein concentrations are ~2 mM. The peak intensity as a function of time was measured and fitted to single-exponential kinetics to obtain the exchange rate constant.

RESULTS

A Stable Mutant of Apocytochrome b_{562}. Previously, it has been shown that PUFs exist for apocytochrome b_{562} under native conditions (*13*). Therefore, it is desirable to characterize the kinetic folding behavior of this protein and test whether PUFs exist after the rate-limiting transition state. Unfortunately, apocytochrome b_{562} has low thermodynamic stability and a fast folding rate that prevent the characterization of the transition state using the protein engineering approach (*14*), which requires the use of destabilizing mutants. In an attempt to test a protein design method using a phage display coupled with proteolysis, we selected a quintuple mutant (M7W/K98I/N99R/H102N/R106G) of apocytochrome b_{562}. This protein has a much higher thermodynamic stability. Therefore, Rd-apocyt b_{562} was chosen to test the late intermediate hypothesis.

Theory of Native-State Hydrogen Exchange. Exchangeable amide hydrogen (NH) involved in the hydrogen-bonded structure can exchange with solvent hydrogens only when they are transiently exposed to solvent in some kind of closed-to-open reaction, as indicated in eq 3 (*15*).

$$\mathrm{NH(closed)} \overset{K_{op}}{\rightleftharpoons} \mathrm{NH(open)} \xrightarrow{k_{int}} \mathrm{exchanged} \qquad (3)$$

where R is the gas constant, T is the absolute temperature, K_{op} is the equilibrium constant for the opening equation, and k_{int} is the intrinsic exchange rate constants from unfolded peptide models (*16*). In a case where the closing reaction is faster than the opening reaction (EX2 condition), the exchange rate of any hydrogen, k_{ex}, is determined by its chemical exchange rate in the open form, k_{int}, multiplied by the equilibrium opening constant, K_{op}.

$$k_{ex} = K_{op}k_{int} \qquad (4)$$

This leads to the free energy for the dominant opening reaction:

$$\Delta G_{HX} = -RT \ln K_{op} = -RT \ln(k_{ex}/k_{int}) \qquad (5)$$

Under the condition of the hydrogen exchange experiment for Rd-apocyt b_{562}, the folding rate constant is larger than the intrinsic hydrogen exchange rate constants. The free energy of hydrogen exchange, ΔG_{HX}, represents a combination of opening transitions from the EX2 mechanism, both structural unfolding and local fluctuations. Evaluating the rates of exchange as a function of denaturant GdmCl concentration allows us to distinguish and separate these two processes. The addition of GdmCl promotes the population of partially unfolded and fully unfolded forms, since these states have more surface exposed than the native state, but by definition have no effect on the local structure fluctuations (*17*, *18*). By examining the pattern of ΔG_{HX} versus GdmCl concentration of the amide proteins, one can obtain the structural and energetic information of the partially unfolded forms and fully unfolded state (*8*).

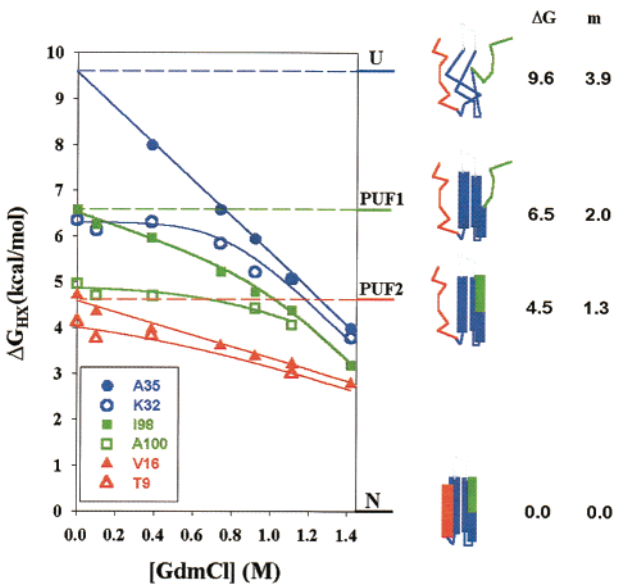

FIGURE 1: Summary of the native-state hydrogen exchange results of Rd-apocyt b_{562}. The free energies and m values of the PUFs are in units of kilocalories per mole and kilocalories per mole per molar (GdmCl), respectively. Different structural units are indicated with different colors on the structure of Rd-apocyt b_{562}. The region colored with gray cannot be defined unambiguously due to the lack of sufficient probes.

Identification of PUFs by Native-State Hydrogen Exchange. To see whether PUFs also exist for Rd-apocyt b_{562}, a native-state hydrogen exchange experiment was performed. The exchange behavior of amide protons could be classified into three different groups, based on their initial sensitivity to denaturant concentrations and how they converged at high denaturant concentrations, thus indicating the existence of PUFs. Figure 1 showed the three pairs of representative amide protons. The ΔG_{HX} of amide proton A35 shows a linear dependence on GdmCl concentration with a slope (m value) of 3.9 kcal mol^{-1} M^{-1} and a ΔG_{HX} of 9.6 kcal/mol at 0.0 M GdmCl. The m value is typical for the global unfolding of a protein with a size of ~100 amino acids, indicating that this amide proton exchanges in the fully unfolded state through a global unfolding process. Therefore, for this amide proton, the curve of ΔG_{HX} versus GdmCl concentration represents the global unfolding ΔG_u versus GdmCl concentration. Thus, the global unfolding free energy of Rd-apocyt b_{562} is 9.6 kcal/mol in water.

At low GdmCl concentrations, the exchange of amide proton K32 is dominated by local structural fluctuations, which expose little new GdmCl binding surface and therefore have an m value close to zero. The same hydrogen can also exchange through the global unfolding reaction. However, the global unfolded state is normally populated at a very low level and makes no significant contribution to the measured exchange. As the GdmCl concentration is increased, the global unfolded state is selectively promoted,

due to its high *m* value, and it dominates the exchange of this faster hydrogen. Eventually, the exchange curves merge to define a hydrogen exchange isotherm that reveals the global unfolding reaction (*18*).

The other two isotherms mirror the behavior seen for the global unfolding isotherm but exhibit serially decreasing ΔG and *m* values. Just as the highest-energy isotherm reflects the global unfolded state, the lower-lying isotherms reveal a sequence of partially unfolded states. The second isotherm is represented by amide hydrogen: Ile98 and A100 (Figure 1). The ΔG_{HX} of Ile98 shows significant GdmCl-dependent behavior at very low GdmCl concentrations. It has an *m* value of 2.0 kcal/mol, suggesting that the state (PUF1) in which the exchange occurs has a significantly exposed surface area. At a GdmCl concentration of zero, the ΔG_{HX} of Ile98 becomes the unfolding free energy of PUF1. The third isotherm is represented by V16 and T9 (Figure 1). It leads to PUF2 with an *m* value of 1.3 kcal mol^{-1} M^{-1} and an unfolding free energy of 4.5 kcal/mol.

Identification of Cooperative Structural Units. The amide protons that converge to the same isotherm form a cooperative structural unit. The assignment of many amide protons to each isotherm can be done simply on the basis of the following rule: hydrogen exchange curves that cross over each other belong to different isotherms. In other words, amide protons in the same isotherm should not cross over before they join the global unfolding isotherm. This criterion allows unambiguous assignment of K27, K32, A40, G70, Q71, A75, K77, A79, N80, K83, Q88, and A89 (Figure 2A−C) to the first isotherm and amide protons of E8, T9, L10, L14, V16, E18, F65, and D66 to the third isotherm (Figure 2E) but leaves uncertainties for residues V26, D28, E81, L94, K95, and those above residue L98, including R99, A100, Y101, and G106 (Figure 2D). These ambiguities can be further resolved with the assumption that amide protons constituting the same isotherm should be a cooperative structural unit in a continuous region of the native-state structure. Since amide protons of K27, A29, and K32 in helix II can be clearly assigned to the first isotherm and amide protons V26 and D28 are in the same region of helix II, they should also be in the same isotherm. Similarly, since A79, N80, K83, A87, Q88, and A89 are in the first isotherm, E81 should also be in the first isotherm. Amide protons after residue I98 cannot be in the first isotherm since such assignment will leave I98 as the only residue in the second isotherm with a large *m* value. Therefore, they should belong to the second isotherm along with I98. The somewhat nonlinear behavior of K95 and other amide protons in this second isotherm (Figure 2D) at low GdmCl concentrations suggests that K95 should be in the second isotherm and unfolding of the C-terminal end of helix IV is not highly cooperative. L94 can be either in the first isotherm or in the second isotherm.

The two amide protons, those of Phe65 and Asp66 (Figure 2E), at the N-terminal end of helix III are not part of helix I, although they fit in the third isotherm. On the basis of the structure of Rd-apocyt b_{562}, it seems reasonable to assume that the two residues along with the residues of the loop that connects the two middle helices form the fourth isotherm. In this paper, we assign them to the third isotherm. This ambiguity would not affect the conclusion made later in the paper.

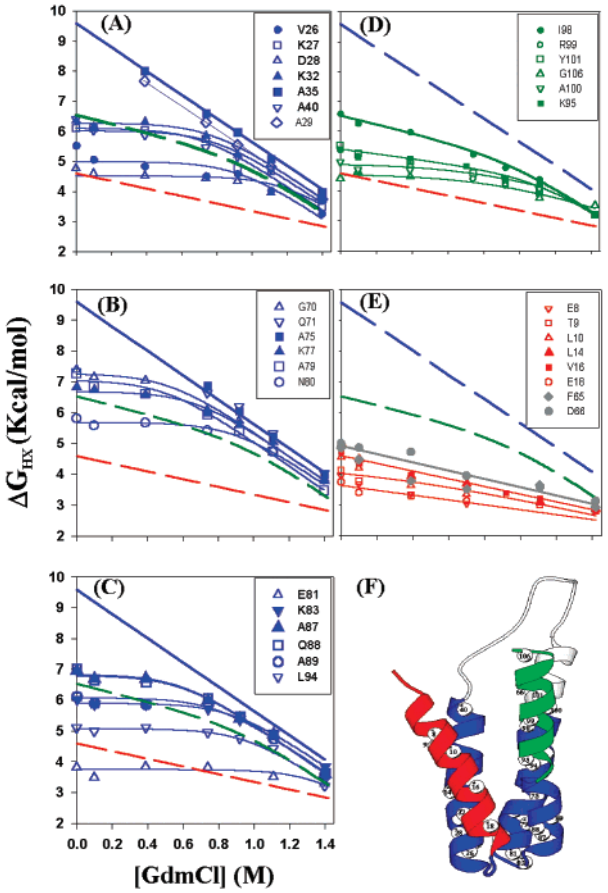

FIGURE 2: Native-state hydrogen exchange results of Rd-apocyt b_{562}. (A−C) The amide protons in the two middle helices and the N-terminal region of helix IV. (D) Amide protons in the C-terminal end of helix IV. (E) Amide protons in helix I and the loop between helix III and IV. (F) Positions of the observed amide protons in the structure of Rd-apocyt b_{562}. The structure is color coded to correspond to the hydrogen exchange isotherms. For fitting of the exchange curves, see refs *8*, *17*, and *18*.

Structures of the PUFs. The hydrogen exchange behavior of the amide protons in the second isotherm produced by PUF1 identifies the cooperative structural unit of the helical region at the C-terminal end of helix IV. Since helix I, which unfolds in the more populated PUF2, exchanges its hydrogens much more quickly than those in the second isotherm, we cannot determine if helix I is also unfolded in PUF1. Like the identification of the fully unfolded state from the amide protons in the first isotherm, the structural identity of PUF1 can be determined by considering the *m* values measured for the second and third isotherms. The *m* value associated with the PUF1 is 2.0 kcal mol^{-1} M^{-1}, indicating that ~50% of the total surface area exposed because of global unfolding is involved in producing PUF1 from the native state. On the other hand, unfolding of the larger cooperative unit, helix I (perhaps with the loop between the two middle helices), has an *m* value of 1.3 kcal mol^{-1} M^{-1}, indicating that helix I should also be unfolded in PUF1. These results are summarized in Figure 1.

Absence of Accumulation of Intermediates in Kinetic Folding. To see whether the PUFs can be detected in kinetic folding experiments, kinetic folding and unfolding of Rd-apocyt b_{562} were studied using a stopped-flow apparatus by monitoring the change in the fluorescence signal emitted from the Trp residue at position 7. Single-exponential kinetics

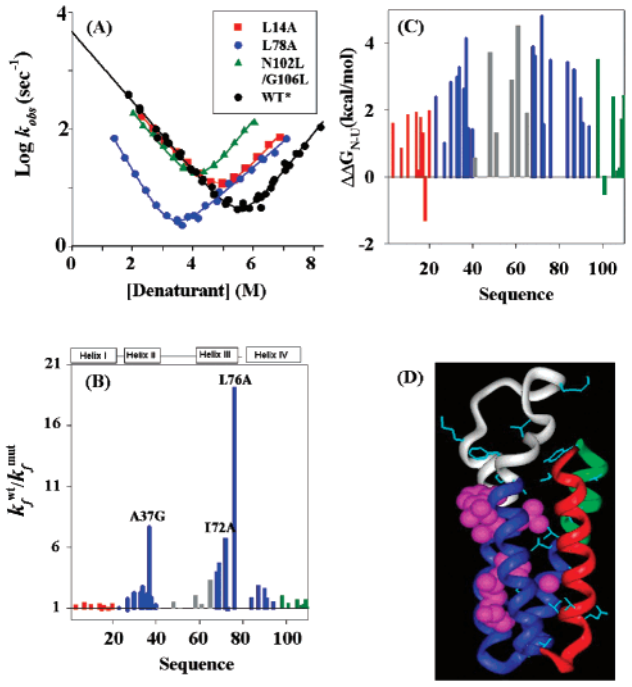

FIGURE 3: Mutational effects on the folding rate and distribution of Φ values in the structure of Rd-apocyt b_{562}. (A) Chevron curves of Rd-apocyt b_{562} and its mutants. (B) Mutational effects on folding rates. (C) Mutational effects on the stability of proteins. (D) Distribution of Φ values on the structure of Rd-apocyt b_{562}. The residues with Φ values larger than 0.2 are represented with the CPK models (purple). The residues with Φ values smaller than 0.2 are represented with lines.

was observed for folding and unfolding at pH 5.2 and 25 °C. The plot of the logarithm of the observed rate constant (k_{obs}) versus denaturant concentration is shown in Figure 3A. In this plot, both folding and unfolding limbs are linear functions of denaturant concentration. The kinetic rate constants at a denaturant concentration of zero are 4.3×10^3 s^{-1} for folding (k_f) and 5.6×10^{-4} s^{-1} for unfolding (k_u). The calculated free energy from $-RT \ln(k_u/k_f)$ is 9.3 kcal/mol. This value is in excellent agreement with the result measured from a hydrogen exchange experiment (9.6 kcal/mol), suggesting both kinetic folding and unfolding processes are apparently two-state without the accumulation of intermediates.

Structural Characterization of the Rate-Limiting Transition State. The simple folding kinetics of Rd-apocyt b_{562} allows structural characterization of the rate-limiting step by studying the effect of destabilizing mutations on the folding rate (*14*). The logarithm of the observed rate constants versus denaturant concentrations for the three typical mutants is shown in Figure 3A. The folding and unfolding rate constants of 43 mutants with mutation sites throughout the structure of Rd-apocyt b_{562} were measured. These mutations include all of the hydrophobic residues in the core and at the interfaces between the helices. The effects of these mutations on the folding rate and stability are shown in panels B and C of Figure 3, respectively. The kinetic and thermodynamic parameters are listed in Table 1.

A common way of describing a transition-state structure is to use Φ value analysis (*14*). Here, Φ is a parameter that describes the ratio of the free energy change in the transition state over that in the native state upon mutation. Theoretical Φ values are between 0 and 1. A value of 0 indicates the mutation site in the transition state is as unfolded as in the unfolded state; a value of 1 indicates the mutation site in the transition state is as folded as in the native state. A fractional value of Φ would indicate that the structure at the mutation site in the transition state is partially formed. The measured Φ values for these mutants are shown in Table 1 and depicted in Figure 3D. The Φ values of the residues in helix I and the C-terminal end of helix IV are uniformly small without exception (<0.16; see Table 1). Larger Φ values only occurred in the two middle helices and the N-terminal end of helix IV. Thus, the rate-limiting transition state of Rd-apocyt b_{562} has a partially unfolded structure involving interactions among residues in the two middle helices and the N-terminal end of helix IV, while other parts of the molecule unfolded. In addition, the few larger Φ values are significantly smaller than 1 (Table 1), which further supports this conclusion, since none of the side chains involves full interactions in the partially unfolded transition state as they did in the native state, based on the native structure of Rd-apocyt b_{562} (Figure 3D).

DISCUSSION

Are PUFs Late Folding Intermediates? The PUFs identified by the native-state hydrogen exchange experiment are equilibrium intermediates. Kinetically, they may be on-pathway (U ↔ PUFs ↔ N), off-pathway (PUFs ↔ U ↔ N), or isolated from the unfolded state (U ↔ N ↔ PUFs). Equilibrium hydrogen exchange data alone cannot tell which one is correct. It has been suggested that the PUF of T_4 lysozyme identified by native-state hydrogen exchange is kinetically isolated from the unfolded state as in the U ↔ N ↔ PUF equilibrium and therefore has nothing to do with the kinetic folding pathway (*19*). For Rd-apocyt b_{562}, we think this is unlikely because for the PUFs ↔ N ↔ U equilibrium to occur, it would mean that the PUFs have to unfold more slowly to the unfolded state than the native state does even though they are less stable than the native state by 4.5−6.5 kcal/mol. In contrast, we found that apocytochrome b_{562}, which has the C-terminal end of helix IV unfolded and is therefore a mimic of PUF1, unfolds to the unfolded state ~100 times faster than Rd-apocyt b_{562} does (unpublished results), indicating that the PUFs are not isolated from the unfolded state by the native state.

In addition, no early folding intermediates could be detected in the kinetic folding experiment, suggesting that the PUFs are late on-pathway folding intermediates. The structural characteristics of the PUFs and the rate-limiting transition state are consistent with this suggestion. The α [$=m_f^{kin}/(m_f^{kin} + m_u^{kin})$] value of the rate-limiting transition state, which represents the extent of burial of the solvent accessible surface upon folding, is 0.47 (see Table 2). This value is significantly smaller than the α value for the PUF2 (0.67) calculated using the relation $(m_u^{eq} - m_I)/m_u^{eq}$ (see Figure 1), suggesting that PUF2 is closer to the native state than the rate-limiting transition state. For PUF1, its structure is very similar to that of the rate-limiting transition state; i.e., they all involve the formation of the two middle helices and the N-terminal end of helix IV (Figures 2F and 3D) and have similar α values (Table 2). Since PUF1 is significantly more stable than the unfolded state, if it exists before the rate-limiting transition state of folding, it should accumulate

Table 1: Kinetic and Thermodynamic Parameters of Rd-apocyt b_{562} and Its Mutants[a]

	k_f (s^{-1})	k_u (s^{-1})	m_f^{kin} (kcal mol^{-1} M^{-1})	m_u^{kin} (kcal mol^{-1} M^{-1})	$-\Delta\Delta G_{-U}^{\ddagger}$ (kcal/mol)	$-\Delta\Delta G_{N-U}$ (kcal/mol)	Φ_f
wild type*	150.7	0.02	0.80	0.88	0.00	0.00	0.00
L3A	129.0	0.28	0.76	0.76	0.09	1.60	0.06
W7M	111.3	0.07	0.75	0.86	0.20	0.85	*b*
L10A	123.0	0.42	0.65	0.86	0.12	1.86	0.06
L14A	120.0	0.44	0.73	0.68	0.13	1.92	0.07
K15A	173.4	0.03	0.82	0.86	0.08	0.20	*b*
A16G	132.0	0.38	0.82	0.87	0.08	1.76	0.05
I17A	149.3	0.19	0.83	0.56	0.00	1.30	0.01
E18A	170.3	0.003	0.68	1.25	−0.07	−1.30	0.05
A20G	115.6	0.43	0.67	0.76	0.16	1.97	0.08
A23G	147.2	0.12	0.88	0.82	0.01	2.38	0.01
K27A	180.7	0.03	0.75	0.75	−0.11	0.17	*b*
K27G	86.0	0.06	0.75	0.84	0.33	1.00	0.33
L30A	67.8	1.20	1.22	0.53	0.47	2.83	0.17
M33A	66.8	1.34	0.71	0.76	0.48	2.97	0.16
R34A	92.9	0.56	0.75	0.64	0.29	2.25	0.13
R34G	55.2	1.84	1.00	0.64	0.59	3.26	0.18
A36G	71.3	0.79	0.80	0.84	0.44	2.61	0.17
A37G	19.7	3.10	0.92	0.71	1.20	4.15	0.29
L38G	82.9	0.13	0.64	0.61	0.35	1.44	0.24
A40G	123.7	0.18	0.82	0.80	0.11	1.41	0.08
A41G	158.8	0.06	0.83	0.82	−0.03	0.57	−0.05
L48A	99.2	0.40	0.83	0.65	0.25	3.73	0.07
K51G	161.2	0.20	0.72	0.73	−0.04	1.32	−0.03
M58A	74.9	0.38	0.68	0.72	0.41	2.91	0.14
F61A	112.6	3.46	0.79	0.52	0.17	4.52	0.15
F65A	45.7	1.80	0.78	0.67	0.71	1.92	0.37
L68A	38.8	0.40	0.80	0.82	0.80	3.88	0.21
V69G	32.4	2.10	0.73	0.76	0.91	3.61	0.25
I72A	22.7	10.70	1.02	0.71	1.12	4.79	0.23
D73A	172.4	0.33	0.83	0.67	−0.08	1.57	−0.05
L76A	7.9	0.41	0.98	0.72	1.74	3.48	0.50
V84A	83.6	0.40	0.82	0.69	0.35	3.42	0.11
A87G	54.2	1.70	1.21	0.54	0.60	3.20	0.19
A90G	59.6	0.42	0.82	0.65	0.55	2.34	0.23
A91G	86.8	0.19	0.99	0.73	0.33	1.60	0.20
L94A	103.0	0.19	0.87	0.72	0.23	1.50	0.15
I98A	78.5	3.78	0.82	0.72	0.39	3.48	0.11
Y101A	112.9	0.01	0.65	1.19	0.17	-0.52	*b*
Y105A	93.7	0.67	0.54	0.64	0.28	2.36	0.12
VAG	130.7	0.02	0.75	0.96	0.08	0.14	*b*
LDR	120.6	0.03	0.79	0.97	0.13	0.23	*b*
VGG	94.0	0.24	0.78	0.82	0.28	1.70	0.16
KIL	95.7	0.78	0.84	0.87	0.26	2.39	0.11

[a] Φ_f values were calculated using the equations $\Phi_f = \Delta\Delta G_{-U}^{\ddagger}/\Delta\Delta G_{N-U}$ where $\Delta\Delta G_{-U}^{\ddagger} = -RT \ln[k_f(\text{mut})/k_f(\text{wt})]$ and $\Delta\Delta G_{N-U} = -RT \ln[k_u(\text{mut})/k_f(\text{mut})] + RT \ln[k_u(\text{wt})/k_f(\text{wt})]$. For multiple mutants, VAG is I98V/N102A, LDR I98L/N102D/G106R, VGG I98V/N102G, and KIL I98K/N102I/G106L. These values are at 2.5 M urea with a standard error of ±5% for the folding rate constants and ±10% for *m* values. The errors in Φ values cannot be determined since k_u at 2.5 M urea relies on the extrapolation from the values at higher denaturant concentrations. These values are measured at 25 °C and pH 5.0 in 2.5 M urea. Here, the Rd-apocyt b_{562} is WT*. [b] $\Delta\Delta G_{N-U}$ is too small to obtain valuable Φ_f values.

Table 2: Solvent Shielding (α value) in the Rate-Limiting Transition State and the PUFs[a]

U	RLTS	PUF1	PUF2	N
0	0.47	0.49	0.67	1

[a] Here α is defined as $m_f^{kin}/(m_f^{kin} + m_u^{kin})$ for the rate-limiting transition state (see the *m* values in Table 1) and $(m_u^{eq} - m_{PUF})/m_u^{eq}$ for the PUFs (see the *m* values in Figure 1). Values of 0 and 1 are defined for U and N, respectively.

during folding and lead to a folding rate constant independent of denaturant concentration, which is not observed (Figure 3A). It should be noted that the similarity between PUF1 and the rate-limiting transition state should not be too surprising. For example, it has been shown that the transition state of cold shock protein B is very close to the native state with the unfolding rate constant being independent of denaturant concentration (*20*).

With the kinetic and structural data taken together, the folding pathway of Rd-apocyt b_{562} seems to be best described as a reaction scheme (eq 6) in which the initial step of folding is rate-limiting.

$$U \leftrightarrow RLTS \leftrightarrow \{PUF1 \leftrightarrow PUF2\} \leftrightarrow N \qquad (6)$$

However, it should be noted that these data do not provide absolute proof for the reaction scheme (eq 6). Here, RLTS represents the rate-limiting transition state, and the braces indicate that the order of formation of the PUFs is not defined.

Native-State Hydrogen Exchange and Protein Folding Pathways. To completely understand how proteins fold, it is important to study the relationship between the native-state hydrogen exchange results and kinetic folding pathways. A correct model for the folding pathway of a protein should provide consistent interpretations for both experimental

results. As shown in this paper, the native-state hydrogen exchange results and the kinetic folding pathway of Rd-apocyt b_{562} can be consistently interpreted using the late intermediate hypothesis. In addition, the very slowly exchanging amide protons are all in the two middle helices that form early in the kinetic folding (*21*, *22*). For the residues with slow exchange amide protons, however, large Φ values were not found (*21*). These results are similar to those observed in other proteins (*22*). It should be noted that the native-state hydrogen exchange results of barnase were initially thought to be inconsistent with the kinetic folding pathway (*23*). This problem, however, has been resolved recently (*24*, *25*). The revised kinetic folding pathway of barnase is now fully consistent with the native-state hydrogen exchange experiments. More interestingly, a late intermediate on the folding pathway of barnase has also been suggested, although the structure of the proposed late intermediate has not been characterized due to the change in the hydrogen exchange mechanism in the native-state hydrogen exchange experiment with barnase (*25*). Thus, the folding behavior of cytochrome *c*, Rd-apocyt b_{562}, and barnase can all be described using the late intermediate hypothesis. It remains to be seen whether other small proteins with apparent two-state folding behavior also have PUFs after the rate-limiting step of folding.

PUFs and Protein Engineering. To better understand the formation of the PUFs, the factors that stabilize the PUFs need to be studied. Therefore, more detailed structural information about PUFs would be extremely useful. On the basis of the diagram shown in Figure 1, it is possible to populate the PUFs by protein engineering and to study them using multidimensional NMR. For example, one can predict that destabilizing mutations in helix I would only reduce the stability of the native state while leaving the stability of PUFs unchanged since helix I is unfolded in the PUFs. Such destabilizing mutations would make PUF2 more stable than the native state. Therefore, PUF2 may be populated and studied directly using multidimensional NMR. Indeed, we have succeeded in populating PUF2 by substituting some of the hydrophobic residues in helix I with Gly and Glu to destabilize helix I. Structural characterization of the engineered product using multidimensional NMR indicates PUF2 is indeed populated, which further supports the conclusion shown in Figure 1 (unpublished results).

PUFs have other implications in protein engineering. For example, if one wants to engineer a protein to increase the stability of the native state while decreasing the relative population of PUFs (since PUFs may cause formation of amyloid or may be more sensitive to protease digestion), the mutation sites should be chosen in the unfolded region of PUFs. This is because stabilizing mutations in these regions would only increase the global stability of the native state, while it will have no effect on the stability of PUFs (see Figure 1). In contrast, if the stabilizing mutations are made in the folded region of the PUFs, the PUFs will also be stabilized and the relative population between the PUFs and the native state may not change significantly.

CONCLUSIONS

A more complete understanding of the energy landscape of proteins under native conditions is important to our understanding of the folding, design, and function of proteins. Native-state hydrogen exchange allows analysis of the structural energetics and suggests possible kinetic folding pathways. This paper presented kinetic and structural data that are fully consistent with the late intermediate hypothesis, in which partially unfolded intermediates are suggested to exist after the rate-limiting step in folding. It illustrates that, in favorable cases, native-state hydrogen exchange coupled with a protein engineering method could be a very useful way of studying the folding landscape that is not accessible for the conventional kinetic studies alone, and partially folded intermediates could be key elements of the folding pathways of small proteins with an apparent two-state folding kinetics. In addition, identification of such PUFs may have important implications for protein engineering and for understanding the formation of amyloid.

ACKNOWLEDGMENT

We thank Dr. S. Walter Englander for comments on the manuscript and Ms. Maritta Perry Grau for professional editing.

REFERENCES

1. Jackson, S. E. (1998) *Folding Des. 3*, R81−R91.
2. Plaxco, K. W., Simons, K. T., Ruczinski, I., and Baker, D. (2000) *Biochemistry 39*, 11177−11183.
3. Goldenberg, D. P. (1999) *Nat. Struct. Biol. 6*, 987−990.
4. Fersht, A. R. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 1525−1529.
5. Bai, Y., and Englander, S. W. (1996) *Proteins 24*, 145−151.
6. Englander, S. W. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 213−238.
7. Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) *Science 291*, 2429−2433.
8. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
9. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) *Nat. Struct. Biol. 3*, 782−787.
10. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
11. Sosnick, T. R., Mayne, L., and Englander, S. W. (1996) *Proteins 24*, 413−426.
12. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
13. Fuentes, E. J., and Wand, A. J. (1998) *Biochemistry 37*, 9877−9883.
14. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
15. Hvidt, A., and Nielsen, S. O. (1996) *Adv. Protein Chem. 21*, 287−386.
16. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins 17*, 75−86.
17. Qian, H., Mayo, S. L., and Morton, A. (1994) *Biochemistry 33*, 8167−8171.
18. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) *Proteins 20*, 4−14.
19. Schindler, T., Herrler, M., Marahiel, M. A., and Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.
20. Llinas, M., Gillespie, B., Dahlquist, F. W., and Marqusee, S. (1999) *Nat. Struct. Biol. 6*, 1072−1078.
21. Englander, S. W. (1998) *Trends Biochem. Sci. 23*, 378−381.
22. Li, R., and Woodward, C. (1999) *Protein Sci. 8*, 1571−1590.
23. Clarke, J., and Fersht, A. R. (1996) *Folding Des. 1*, 243−254.
24. Takei, J., Chu, R. A., and Bai, Y. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 10796−10801.
25. Fersht, A. R. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 14121−14126.

BI025872N