Accelerated Publications

Structural Examination of Φ-Value Analysis in Protein Folding

Hanqiao Feng,‡ Ngoc-Diep Vu,‡ Zheng Zhou, and Yawen Bai*

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

Received August 31, 2004; Revised Manuscript Received October 11, 2004

ABSTRACT: Protein folding intermediates and transition states are commonly characterized using a protein engineering procedure (Φ-value analysis) based on several assumptions, including (1) intermediates and transition states have native-like conformations and (2) single mutations from larger hydrophobic residues to smaller ones do not perturb their structures. Although Φ-value analysis has been widely used, these assumptions have not been tested to date because of the lack of high-resolution structures of intermediates and transition states. We recently have determined the structure of a folding intermediate for a four-helix bundle protein (Rd-apocytochrome b_{562}) using NMR. The intermediate has the N-terminal helix unfolded. The other three helices fold in a native-like topology with extensive non-native hydrophobic interactions. Here, we have determined the Φ values for 14 hydrophobic core residues, including those with significant non-native interactions. All of the Φ values are in the normal range from 0 to 1, indicating that these non-native interactions cannot be identified by the common Φ-value analysis, and therefore, the first assumption is not valid for this intermediate. We also determined the structure of a mutant (F65A) of the intermediate and found that the structure of the intermediate is not perturbed by the mutation, supporting the second assumption. Together, these results suggest that Φ-value analysis may be valid for characterizing the energetics of the interactions between the mutated residue and others, but not for determining the detailed structures of intermediates and transition states because non-native interactions may exist and may not be identifiable by the common Φ-value analysis.

To understand the mechanism of protein folding, it is necessary to characterize the folding energy landscape in detail, which includes the structures of intermediates and transition states. Knowledge of these structures is critical for establishing the correct model for describing the general folding behavior of proteins. For the past decade, a larger number of transition states (*1*) and three partially unfolded intermediates (*2*−*4*) have been characterized by using protein engineering and Φ-value analysis (*5*, *6*) aimed at obtaining structural information about the partially unfolded states. In the Φ-value analysis, a parameter, Φ, is measured as the ratio of the free energy change in the intermediate ($\Delta\Delta G_{IU}$) or the transition state ($\Delta\Delta G^{\dagger}_{TS\_U}$) to that in the native state ($\Delta\Delta G_{NU}$) upon a single mutation: $\Phi_I = \Delta\Delta G_{IU}/\Delta\Delta G_{NU}$ or $\Phi_{TS} = \Delta\Delta G^{\dagger}_{TS\_U}/\Delta\Delta G_{NU}$. A Φ value of zero appears to have a clear structural meaning: it indicates that the mutated residues are unfolded or do not form more interactions with other residues in the intermediates and transition states than those in the unfolded state. To derive structural information from other Φ values, two major assumptions are commonly made (*6*): (1) the folded region of the intermediate or transition state has a native-like structure, and (2) mutations from large hydrophobic residues to smaller ones do not perturb the structure of the intermediates. With these assumptions, if a mutation yields a Φ value of 1, it is interpreted to mean that the mutated residue forms the full contacts with its neighbor residues in the intermediate or transition state, as it does in the native state. Residues with fractional Φ values are considered to have partially formed native interactions. For example, Φ values have been used as native-like constraints to derive the structures of partially unfolded states in a way similar to the determination of protein structures by nuclear magnetic resonance (NMR)[1] with NOEs as constraints (*7*−*9*).

Using a native state hydrogen exchange method (*10*), we recently have identified a partially unfolded intermediate for a four-helix bundle protein, Rd-apocytochrome b_{562} (Rd-apocyt b_{562}) (*11*). This intermediate has the N-terminal helix unfolded while the other three helices remain folded. We have populated this intermediate by substituting the hydrophobic residues in the N-terminal helix with glycines without affecting the folded region of the intermediate, and determined its structure using multidimensional NMR (see panels A and B of Figure 1) (*12*). Although the structure of the intermediate has native-like topology for the folded region,

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

‡ These authors contributed equally to this work.

[1] Abbreviations: HX, hydrogen exchange; CD, circular dichroism; NMR, nuclear magnetic resonance.

10.1021/bi048126m
Published on Web 10/21/2004

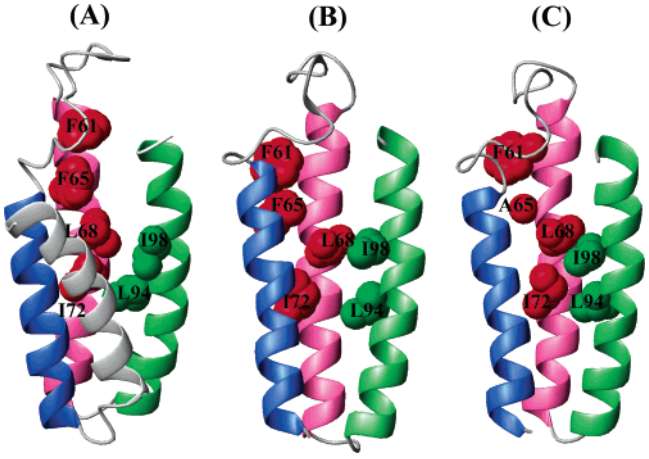

FIGURE 1: Illustration of the non-native side chain interactions in the intermediates: (A) native structure, (B) non-native intermediate with the N-terminal helix unfolded, and (C) F65A mutant of the non-native intermediate. Representative hydrophobic residues are shown in CPK models. For clarity, the N-terminal helix has been omitted in panels B and C.

it has extensive non-native local conformations, including distortions of the helical backbones and broad non-native hydrophobic interactions. In this paper, this glycine mutant will be called a "non-native intermediate".

To see how the non-native interactions in the intermediate of Rd-apocyt b_{562} are reflected in Φ values, we measured the Φ values at 14 positions that are occupied by the large hydrophobic core residues in the intermediate. All of the mutations yielded normal Φ values, from 0 to 1, although some of the residues participate in significant non-native interactions in the intermediate. To test whether mutations from larger hydrophobic residues to smaller ones may change the structure of the intermediate, we determined the structure of a mutant of the intermediate with a Phe mutated to Ala (F65A). No significant perturbation of the structure was observed.

MATERIALS AND METHODS

Protein Sample Preparation. Mutations were made using the quick-change kit (Stratagene) based on the plasmid that encodes the gene for the non-native intermediate (*13*). Protein expression and purification were carried out as described previously (*11*). The buffer is 20 mM NaAc at pH 5.0. Ultrapure urea was purchased from Gibco BRL. The concentration of the urea was measured using a refractometer (*14*). Isotopically enriched proteins for determination of the structure of the F65A mutant were grown on M9 minimal medium containing 1 g/L [U-^{15}N]$^{15}NH_4Cl$ and/or 4 g/L [U-^{13}C]glucose (Isotec) for doubly (^{13}C and ^{15}N) and/or singly (^{15}N or ^{13}C) labeled proteins as the sole sources of nitrogen and carbon. NMR samples included ^{13}C/^{15}N-, ^{15}N-, and ^{13}C-labeled proteins and the nonlabeled protein at concentrations of ~2 mM (95% H_2O and 5% D_2O) at pH 5.0 with 20 mM NaAc-d_4 as the buffer.

Thermal Melting Experiments. Equilibrium melting monitored by circular dichroism (CD) at a wavelength of 222 nm was used to measure the stability of the intermediate and its mutants. The experiments were carried out at pH 5.0 (20 mM NaAc). The melting curves were fitted with eq 1 using a fixed ΔC_p of 0.45 kcal mol^{-1} K^{-1} (*15*):

$$\theta = \theta_u + (\theta_f - \theta_u)/(1 + \exp\{(-\Delta H_m/R)(1/T - 1/T_m) + (\Delta C_p/R)[(T_m/T - 1) + \ln(T/T_m)]\}) \quad (1)$$

Here, θ is the measured CD signal, and θ_f and θ_u are linear functions of temperature, representing the folded and unfolded baselines, respectively (*15*). T_m is the melting temperature, and ΔH_m is the enthalpy of unfolding at T_m. Equation 2 is used to calculate the unfolding free energy after T_m and ΔH_m are obtained from the curve fitting.

$$\Delta G_{IU} = \Delta H_m(1 - T/T_m) - \Delta C_p[T - T_m - T \times \ln(T/T_m)] \quad (2)$$

Measurement of Hydrogen Exchange Rates. Hydrogen exchange experiments for the wild-type intermediate and F65A mutant were carried out by recording a series of ^{1}H−^{15}N HSQC spectra as a function of time. The amide protons in protein samples were first exchanged with deuterium in D_2O at pD 7 and 50 °C for the wild-type protein and at 25 °C for F65A. The hydrogen exchange rates were measured by changing D into H in H_2O with 5% D_2O (50 mM NaAc) at 25 °C. The NMR spectra were processed using NMRPipe (*16*) and analyzed using NMRView (*17*). The exchange rates were measured by fitting the peak intensity as a function of time using NMRView.

Equilibrium Unfolding by Urea. Equilibrium unfolding was performed using urea as a denaturant. The content of the secondary structure was monitored using CD at 222 nm (JASCO), at different concentrations of urea. The stock protein concentration in the buffer (pH 5.0, 50 mM NaAc) is ~0.3 mg/mL. The sample was prepared by a 10-fold dilution of the stock solution in various urea concentrations. The experimental data were fitted using eq 3 to obtain the global unfolding free energy (*18*):

$$\mathrm{CD}([\mathrm{urea}]) = [a_N[\mathrm{urea}] + b_N + (K_{NU}^{H_2O} \times 10^{m_{NU}[\mathrm{urea}]}) (a_U[\mathrm{urea}] + b_U)]/(1 + K_{NU}^{H_2O} \times 10^{m_{NU}[\mathrm{urea}]}) \quad (3)$$

where CD is the measured ellipticity at 222 nm, $K_{NU}^{H_2O}$ is the global unfolding equilibrium constant in water, a_N, b_N, a_U, and b_U are the parameters of the baselines for the native and unfolded states, respectively, and m_{NU} is the coefficient for the denaturant which is dependent on the logarithm of K_{NU}. The unfolding free energy, ΔG_{IU}, was then calculated from $-RT \ln K_{IU}$. These experiments were carried out at pH 5.0 and 25 °C.

NMR Spectroscopy for Structure Determination. NMR spectra were collected at 25 °C on Bruker (Bellerica, MA) DRX 500 MHz spectrometers equipped with a 5 mm *x,y,z*-shielded pulse field gradient triple-resonance probe. A series of three-dimensional spectra [CBCA(CO)NH, HNCACB, HCCH-TOCSY, ^{15}N-edited TOCSY, and ^{15}N/^{13}C-edited NOESY] in addition to two-dimensional ^{1}H NOESY spectra were collected for complete assignments and NOE measurement. An HNHA experiment was used to determine $^3J_{\alpha N}$ coupling constants (*19*). NMR data were processed using NMRPipe (*16*) and analyzed using NMRView (*17*).

Calculation of the Structure. Structural calculations were carried out using NIH X-PLOR (*20*). An extended polypep-

tide chain of reasonable geometry was used as the initial template. Backbone dihedral angles were then randomized before each cycle of the simulated annealing (SA) protocol. Each SA structure was optimized by restrained refinement. Ten structures with low energies and few NOE and dihedral angle violations were selected from 50 SA structures. For each such starting structure, 10 additional structures (i.e., 100 structures) were calculated. The ones with low energies and no NOE and dihedral angle violations from each set were used for a second round of refinement to calculate another 10 structures. This refinement procedure was repeated once more to obtain the final 10 structures with the lowest energies and no violation of restraints. These structures were further checked by PROCHECK-NMR (version 3.5.4) (*21*).

RESULTS AND DISCUSSION

Non-Native Interactions in the Intermediate. Panels A and B of Figure 1 illustrate the structures of the native state of Rd-apocyt b_{562} and the folded regions of the non-native intermediate, respectively. The typical non-native hydrophobic side chains are shown in CPK models. Using a probe with a radius of 1.4 Å and 10 NMR structures determined for each protein, we found that the non-native intermediate structure buries 258 ± 80 Å^2 more hydrophobic surface area than the putative native-like intermediate that has the N-terminal helix removed from the native structure (*12*, *22*), whereas the accessible surface area for the polar atoms is increased by 58 ± 85 Å^2 in the non-native intermediate, indicating that the non-native intermediate is more compact. Since hydrophobic interactions are stabilizing forces, the loss of solvent-accessible hydrophobic surfaces in the non-native intermediate suggests that partially unfolded intermediates tend to relax to a more stable conformation rather than staying at the otherwise unstable native-like conformation. Indeed, this non-native intermediate has extensive local structure reorganization throughout the protein structure, including main chain distortion and side chain repacking. For example, F61 and F65 face the inside of the protein and pack with other hydrophobic residues in the native structure; on the other hand, they face the outside and lie between the two middle helices. L68 and I98 are far apart in the native structure, but they packed with each other in the intermediate. In contrast, I72 and L94 interact closely in the native state, but they are separated within the intermediate. Table 1 lists the residue pairs that have their heavy atoms changed from <5 Å in the native-like intermediate to >7 Å in the non-native-like intermediate, and *vice versa*. These non-native interactions are also illustrated in Figure 2. It should be noted that these are large conformational changes that are well beyond the rmsds (0.8 Å) for all heavy atoms of these structures. None of the above differences in distance was found among the 10 NMR structures of either the native state or the non-native intermediate alone.

Selection of Mutations for Φ-Value Analysis. To investigate whether non-native interactions in the intermediate of Rd-apocyt b_{562} could be detected by the Φ-value analysis, we selected 14 hydrophobic core residues for mutational studies to measure the Φ values. These 14 residues were chosen because they were mutated in the earlier studies in an effort to characterize the rate-limiting transition state, and they include the residues with significant non-native interactions in the wild-type intermediate (*11*). In addition, the free energy changes in the native state upon mutations from these residues to Ala or Gly are known to be larger than 1.7 kcal/mol, a value that has been suggested to be critical in obtaining reliable Φ values (*23*). All of the mutations are from big hydrophobic side chains to smaller ones (Phe → Ala, Met → Ala, Leu → Ala, Ile → Ala, Val → Ala, and Ala → Gly).

Measurement of the Free Energy Change of the Intermediate and Its Mutants. To measure the Φ values of these hydrophobic core residues, the unfolding free energy for the non-native intermediate and its mutants was measured using thermal melting by monitoring the signal of circular dichroism (CD) at 222 nm. All melting curves monitored by CD could be fitted perfectly with a two-state model (eq 1) using a ΔC_p ranging from 0.2 to 0.7 kcal mol^{-1} K^{-1} (see Figure 3), indicating that T_m and ΔH_m can be measured accurately. Using the T_m and ΔH_m values obtained from the curve fitting

Table 1: Thermodynamic Parameters and Φ Values of the Intermediate

	T_m (K)	ΔH_m (kcal/mol)	ΔG_{IU}	$\Delta\Delta G_{IU}$	$\Delta\Delta G_{NU}$	Φ_I	non-native interactions[a]
4GD7	337.6	55.77	5.44				
M33A	327.0	48.46	3.69	1.75	3.50	0.50	
A36G	332.8	56.65	5.06	0.38	2.71	0.14	
L48A	333.4	37.42	3.08	2.36	4.23	0.56	
M58A	335.0	47.23	4.25	1.19	3.61	0.33	
F61A	330.1	43.36	3.48	1.96	5.45	0.36	101
F65A	324.0	41.18	2.82	2.62	2.50	1.04	66, 69, 101
L68A	321.0	47.51	3.02	2.42	4.03	0.60	72, 97, 98, 101
I72A	322.3	50.18	3.35	2.09	4.67	0.45	76, 94
L76A	332.8	41.32	3.46	1.98	3.43	0.58	80
V84A	330.2	58.77	4.99	0.45	3.85	0.12	
A87G	322.3	55.27	3.74	1.70	3.03	0.56	
A90G	325.3	62.86	4.74	0.70	2.87	0.24	
L94A	328.2	57.22	4.61	0.83	1.73	0.48	98
I98A	327.6	51.76	4.04	1.40	3.83	0.37	102

[a] The numbers represent the residues that have heavy atoms within 5 Å of those in the corresponding residues in the first column in the native-like intermediate but beyond 7 Å in the wild-type intermediate or *vice versa*.

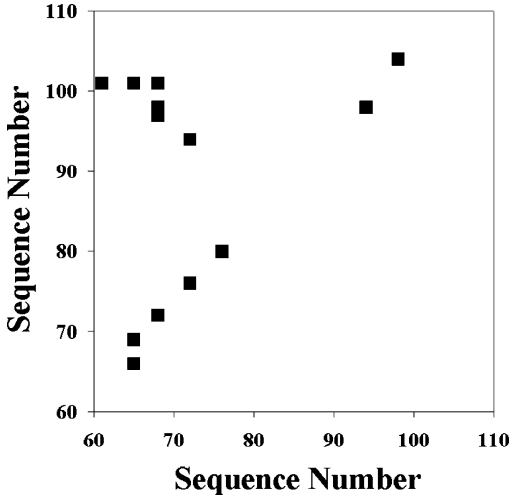

FIGURE 2: Contact map of the significant non-native interactions in the non-native intermediate. Filled squares represent non-native contacts (see Table 1). A pair of residues is considered to have a non-native contact in the non-native intermediate if any two heavy atoms of two residues are within 5 Å in the putative native-like intermediate but beyond 7 Å in the non-native intermediate or *vice versa*. On the basis of this criterion, no such contact is found among the 10 different NMR structures of either the native-like intermediate or the non-native intermediate.

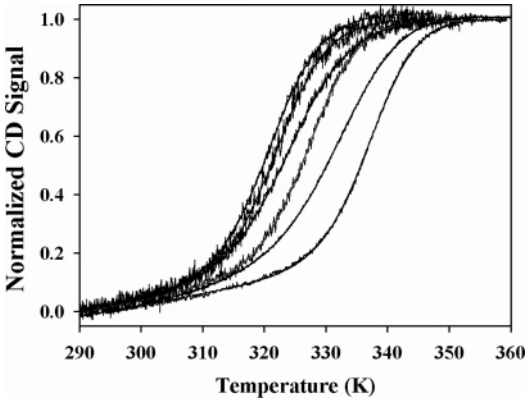

FIGURE 3: Normalized melting curves for the representative mutants of the non-native intermediate: L68A, I72A, F65A, I98A, L76A, and wild-type intermediate (from left to right).

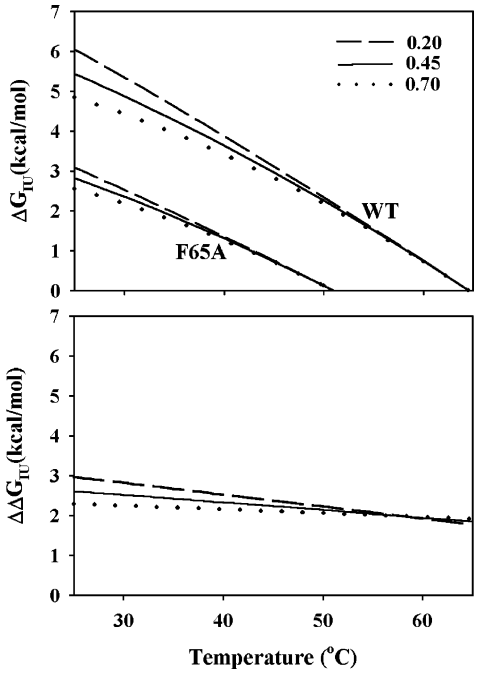

FIGURE 4: Effects of different ΔC_p values [0.2 (− − −), 0.45 (—), and 0.7 (···)] on the ΔG_{IU} values (A) and $\Delta\Delta G_{IU}$ values (B) of the non-native intermediate and the F65A mutant.

described above, the unfolding free energies were calculated using eq 2 with a ΔC_p of 0.45 kcal mol^{-1} M^{-1} for all proteins. We choose 0.45 kcal mol^{-1} K^{-1} for the non-native intermediate because apocytochrome b_{562}, which involves the unfolding of the major part of the C-terminal helix, has a ΔC_p of ∼0.5 (*24*). Simulation studies show that the effect of ΔC_p on the unfolding free energy of the intermediate ΔG_{IU} is small (less than ±0.6 kcal/mol) at 25 °C, if the ΔC_p is in the range of 0.2−0.7 kcal mol^{-1} K^{-1} (see Figure 4A). The effect of ΔC_p on $\Delta\Delta G_{IU}$ is even smaller (±0.3 kcal/mol) for the same range of ΔC_p values (see Figure 4B), indicating that reliable $\Delta\Delta G_{IU}$ values for the intermediates can be measured with a ΔC_p of 0.45 kcal mol^{-1} K^{-1}. The measured ΔG_{IU} and $\Delta\Delta G_{IU}$ values for the wild-type intermediate and its mutants are listed in Table 1.

Independent Evaluation of the Measurement of $\Delta\Delta G_{IU}$. To test the accuracy of the free energy change measured using the thermal melting method, hydrogen exchange rates, k_{ex}, of the amide protons of the wild-type intermediate and the F65A mutant were measured. The ΔG_{HX} values, plotted

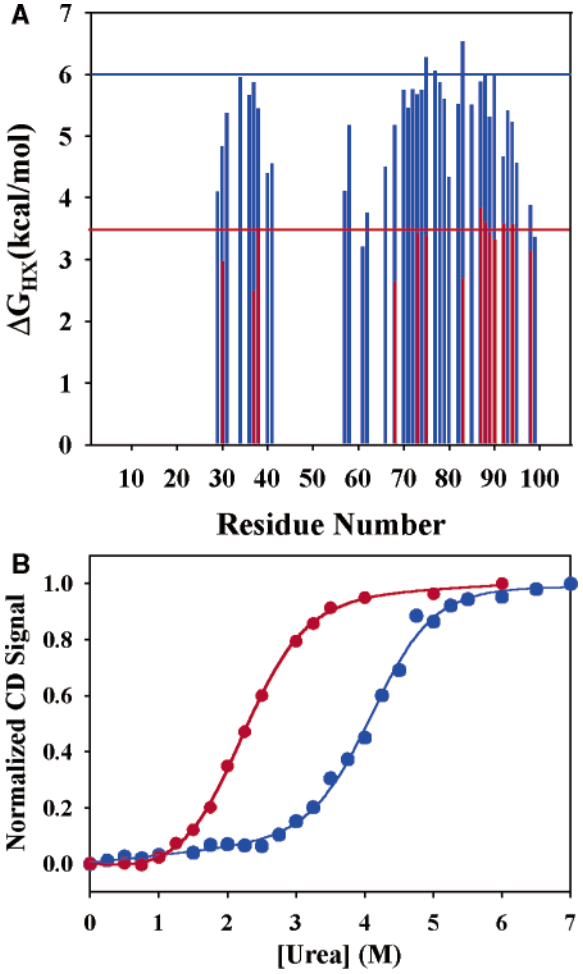

FIGURE 5: Measurements of unfolding free energy by hydrogen exchange and urea denaturation. (A) ΔG_{HX} values for the non-native intermediate (blue) and F65A mutant (red). The lines represent the estimated global unfolding free energy. (B) Urea melting data of the non-native intermediate (blue) and the F65A mutant (red). The solid lines are the fitting curves using eq 2.

in Figure 5A, were calculated for measurable slowly exchanging amide protons from $-RT \ln(k_{ex}/k_{int})$, where k_{int} is the intrinsic exchange rate constant based on peptide models (*25*). It has been shown that the ΔG_{HX} from very slowly exchanging amide protons can be used to measure the global unfolding free energy (*25*, *26*). The averaged values for the very slowly exchanging amide protons are ∼6.0 and ∼3.5 kcal/mol for the wild type and the F65A intermediate, respectively. These values are ∼0.5 kcal/mol larger than those measured using the thermal melting method. Nevertheless, the difference in the free energy change ($\Delta\Delta G_{IU}$) between the wild-type intermediate and the F65A mutant measured by hydrogen exchange (2.5 kcal/mol) is very similar to that measured by thermal denaturation (2.6 kcal/mol). The fact that all three helices have amide protons that exchange through global unfolding indicates that both the wild type and the F65A intermediates could not populate partially unfolded intermediates, and supports the two-state analysis. Moreover, equilibrium denaturation experiments were performed using urea as the denaturant and monitored by CD. The ellipticities of the CD signal at a wavelength of 222 nm were measured and plotted against the urea concentrations, as shown in Figure 5B. The experimental data were fitted with a two-state model using eq 3. The fitting

Table 2: Unfolding Free Energy of the Wild Type and F65A Mutant Measured by Different Methods at 25 °C and pH 5.0 in Units of Kilocalories per Mole

	ΔG_{IU}(WT)	ΔG_{IU}(F65A)	ΔG_{IU}(WT) − ΔG_{IU}(F65A)
thermal melting	5.5	2.8	2.7
hydrogen exchange	6.0	3.5	2.5
urea melting	5.2	2.8	2.4

leads to values of 5.2 and 2.8 kcal/mol for the unfolding free energy of the non-native intermediate and the F65A mutant, respectively. These values are very close to the values of 5.5 and 2.8 kcal/mol, respectively, measured from the thermal melting experiments, supporting the use of the thermal melting method in measuring the unfolding free energy of the intermediates. These unfolding free energy values are listed in Table 2.

Normal Φ Values with Non-Native Interactions. Using the $\Delta\Delta G_{IU}$ values given above and those of $\Delta\Delta G_{NU}$ measured earlier for the native state (*11*), we calculated Φ values for the 14 residues, as shown in Table 1. All of the residues have normal Φ values, ranging from 0 to 1, suggesting that if the structure of the intermediate were not known, the Φ values would have been interpreted in terms of native-like interactions. In particular, since the Φ value of F65 is close to 1, it would be concluded that F65 had a full native environment. In contrast, F65 has the most dramatic change in side chain conformation. Its aromatic side chain rotates nearly 180° away from the hydrophobic core and packs between the two middle helices (see panels A and B of Figure 1). In this case, the large Φ value of F65A is likely due to the relatively small free energy change (2.8 kcal/mol; see Table 1) in the native state. This is because a mutation from a buried Phe to Ala often causes a much larger (~5 kcal/mol) destabilization (*27*). Moreover, we found that the mutation from Phe to Leu at this position slightly increased the stability of the native state, although the number of carbon atoms was decreased (unpublished result). Since mutations from Phe to Leu normally decrease the stability of the protein (*27*, *28*), this increase in stability suggests that the side chain of F65 is overpacked in the native structure. In addition, F61 is nearly fully buried in the native-like intermediate, whereas it is significantly exposed in the non-native intermediate. Thus, the small Φ value (0.36) of F61 is more consistent with the structure of the non-native intermediate than with the structure of the putative native-like intermediate. All other fractional Φ values in Table 1 can be explained by the partial exposure of these residues in the non-native intermediate, which is due to the unfolding of the N-terminal helix of Rd-apocyt b_{562}. These results show that normal fractional Φ values can arise from non-native interactions and partial exposure of surface areas after partial unfolding.

A Structural Test for a Nondisruptive Mutation. In Φ-value analysis, mutations from large hydrophobic residues to smaller ones are commonly made, including Ile, Val, Phe, Tyr, and Met to Ala or Ala to Gly. It is assumed that such mutations will not affect the conformations of other residues in the partially unfolded intermediates and transition states. To test whether this would be the case for the non-native intermediate, we have determined the high-resolution structure of the F65A intermediate using multidimensional NMR. The rmsd value for all heavy atoms is 0.82 Å. The parameters describing the structure are listed in Table 3. Figure 6 shows the superposition of the 10 NMR structures based on all heavy atoms. The structure is very well defined. Panels B and C of Figure 1 compare the structures of the F65A mutant and non-native intermediate with the same typical hydrophobic side chains in CPK models. The two structures are very similar, with only minor differences, suggesting that a mutation from a residue as large as Phe to Ala with a change

Table 3: Parameters for the Folded Region in the F65A Mutant of the Intermediate

	rmsd from ideal geometry	
	⟨dgsa⟩	average
bonds (Å)	0.00471 ± 0.00006	0.00471
angles (deg)	0.579 ± 0.010	0.579
impropers (deg)	0.482 ± 0.012	0.482
NOE (all)	0.0558 ± 0.007	0.0558

	all residues/helical region
backbone atoms	0.50/0.39
all heavy atoms	0.82/0.70

Experimental Restraints	
NOEs	
intraresidue	819
sequential NOE ($\|i-j\|=1$)	1039
long-range NOE ($\|i-j\|>5$)	193
H-bonds	104
dihedral angles	166
C_α and C_β chemical shifts	103

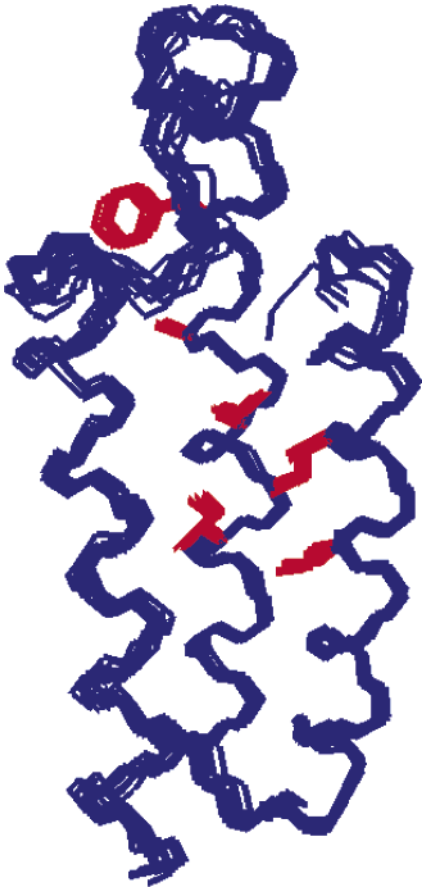

FIGURE 6: Illustration of the quality of the NMR structure for the F65A intermediate. Ten structures are superimposed on the basis of all heavy atoms. For the backbone, only C_α atoms are shown (blue). For side chains, only the typical residues (F61, F65, L68, I72, L94, and I98) are shown (red).

of free energy of 2.8 kcal/mol has little effect on the initial structure. This conclusion is further supported by the similar *m* values of the non-native intermediate (0.97 kcal mol^{-1} M^{-1}) and the F65A mutant (0.94 kcal mol^{-1} M^{-1}) from the urea melting experiments (see Figure 5). Therefore, these results suggest that mutations from large hydrophobic residues to smaller ones can be "nondisruptive".

Implications for Transition States. Since transition states are partially unfolded structures, we speculate that they also tend to reorganize and have non-native interactions. Earlier protein engineering studies on a number of proteins, including CI2, the α-spectrin SH3 domain, src SH3, CheY, and AcP (see ref *31* for detailed discussions), have also suggested that non-native interactions occur in the transition states in these proteins. In these cases, non-native interactions were identified on the basis of Φ values that were smaller than 0 or larger than 1. Such interactions are further illustrated by a computer simulation study (*29*). In a more recent study on the transition state of ubiquitin using a bi-His metal-binding method, ψ values significantly larger than 1 were found at several positions (*30*). Similarly, several residues in the WW domain have Φ values that are significantly larger than 1 (*31*). Again, these results suggest the existence of non-native interactions in the transition states.

Our finding that partially unfolded intermediates with a broad distribution of non-native interactions can still yield normal Φ values suggests that the transition states with normal Φ values for many proteins could also have non-native interactions. This is a crucial issue because most of the experimental and theoretical studies on the transition states have assumed that Φ values in the range from 0 to 1 represent native-like interactions. In addition, both the non-native intermediate and the F65A mutant have well-defined structures, suggesting that all protein molecules may go through this specific structure during folding before they reach the native state. This result would be contrary to the popular view that protein molecules fold to the native state through numerous different pathways, and fractional Φ values are the evidence for the existence of numerous different structures in the intermediates and transition states (*32*, *33*).

CONCLUSIONS

Protein engineering is a very useful method for identifying the unfolded regions in intermediates and transition states, and Φ-value analysis can provide the energetic information for the interactions between the mutated residue and others in the partially unfolded states. To obtain more detailed structural information about folding intermediates and transition states, it is commonly assumed that intermediates and transition states have native-like conformations and single mutations from larger hydrophobic residues to smaller ones do not perturb their structures. Although the structure determined for the F65A mutant supports the latter assumption, the normal Φ values measured for residues with significant non-native interactions in the folding intermediates of Rd-apocyt b_{562} suggest that Φ values may not be simply interpreted as native-like interactions and used to derive detailed structures of intermediates and transition states. Finally, our findings suggest that non-native interactions may exist more commonly than previously thought in partially unfolded states. They could be the norm rather than exception, and therefore need to be taken into account in future computer simulation studies on protein folding.

REFERENCES

1. Goldenberg, D. P. (1999) Finding the right fold, *Nat. Struct. Biol. 6*, 987−990.
2. Matouschek, A., Serrano, L., and Fersht, A. R. (1992) The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analyzed by a protein engineering procedure, *J. Mol. Biol. 224*, 819−835.
3. Raschke, T. M., Kho, J., and Marqusee, S. (1999) Confirmation of the hierarchical folding of RNase H: a protein engineering study, *Nat. Struct. Biol. 6*, 825−831.
4. Bulaj, G., and Goldenberg, D. P. (2001) Φ-values for BPTI folding intermediates and implications for transition state analysis, *Nat. Struct. Biol. 8*, 326−330.
5. Matthews, C. R., and Hurle, M. R. (1987) Mutant sequences as probes of protein folding mechanisms, *BioEssays 6*, 254−257.
6. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, *J. Mol. Biol. 224*, 771−782.
7. Li, A., and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10430−10434.
8. Vendruscolo, M., Paci, E., Dobson, C. M., and Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state, *Nature 409*, 641−645.
9. Hubner, I. A., Shimada, J., and Shakhnovich, E. E. (2004) Commitment and nucleation in the protein G transition state, *J. Mol. Biol. 336*, 745−761.
10. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange, *Science 269*, 192−197.
11. Chu, R. A., Pei, W. H., Takei, J., and Bai, Y. (2002) Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein, *Biochemistry 41*, 7998−8003.
12. Feng, H., Takei, J., Lipsitz, R., Tjandra, N., and Bai, Y. (2003) Specific non-native hydrophobic interactions in a hidden intermediate: implications for protien folding, *Biochemistry 42*, 12461−12465.
13. Takei, J., Pei, W., Vu, D., and Bai, Y. (2002) Populating partially unfolded forms by hydrogen exchange-directed protein engineering, *Biochemistry 41*, 12308−12312.
14. Pace, C. N. (1986) Determination and analysis of urea and guanidinium hydrochloride denaturation curves, *Methods Enzymol. 131*, 266−280.
15. Minor, D. L., Jr., and Kim, P. S. (1994) Measurement of the β-sheet-forming propensities of amino acids, *Nature 367*, 660−663.
16. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX Pipes, *J. Biomol. NMR 6*, 277−293.
17. Johnson, B. A., and Blevins, R. A. (1994) NMRview: a computer program for the visualization and analysis of NMR data, *J. Biomol. NMR 4*, 603−614.
18. Santoro, M. M., and Bolen, D. M. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range, *Biochemistry 31*, 4901−4907.
19. Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: A new approach for measuring homonuclear three-bond $J_{HNH\alpha}$ coupling constants in ^{15}N-enriched proteins, *J. Am. Chem. Soc. 115*, 7772−7777.
20. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR Molecular Structure Determination Package, *J. Magn. Reson. 160*, 65−73.
21. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
22. Feng, H., and Bai, Y. (2004) Repacking of hydrophobic residues in a stable mutant of apocytochrome b_{562} selected by phage-display and proteolysis, *Proteins 56*, 426−429.
23. Sanchez, I., and Kiefhaber, T. (2003) Origin of unusual Φ-values in protein folding, *J. Mol. Biol. 334*, 1077−1085.

24. Fuentes, E. J., and Wand, A. J. (1998) Local stability and dynamics of apocytochrome b_{562} examined by the dependence of hydrogen exchange on hydrostatic pressure, *Biochemistry 37*, 9877−9883.
25. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) Protein stability parameters measured by hydrogen exchange, *Proteins 20*, 4−14.
26. Huyghues-Despointes, B. M., Scholtz, J. M., and Pace, C. N. (1999) Protein conformational stabilities can be determined from hydrogen exchange rates, *Nat. Struct. Biol. 6*, 910−912.
27. Kim, D. E., Fisher, C., and Baker, D. (2000) A breakdown of symmetry in the folding transition state of protein L, *J. Mol. Biol. 298*, 971−984.
28. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding, *J. Mol. Biol. 254*, 260−288.
29. Li, L., Mirny, L. A., and Shakhnovich, E. I. (2000) Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus, *Nat. Struct. Biol. 7*, 336−342.
30. Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004) Discerning the structure and energy of multiple transition states in protein folding using Φ-analysis, *J. Mol. Biol. 337*, 463−475.
31. Crane, J. C., Koepf, E. K., Kelly, J. W., and Gruebele, M. (2000) Mapping the transition state of the WW domain β-sheet, *J. Mol. Biol. 298*, 283−292.
32. Ozkan, S. B., Bahar, I., and Dill, K. A. (2001) Transition states and the meaning of Φ-values in the protein folding kinetics, *Nat. Struct. Biol. 8*, 765−769.
33. Onuchic, J., and Wolynes, P. G. (2004) Theory of protein folding, *Curr. Opin. Struct. Biol. 14*, 70−75.

BI048126M