Design and Folding of a Multidomain Protein

Zheng Zhou,[‡,§] Hanqiao Feng,[‡,§] Hongyi Zhou,[||] Yaoqi Zhou,[||] and Yawen Bai*[,§]

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, and Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology and Biophysics, University at Buffalo, State University of New York, Buffalo, New York 14214

Received April 28, 2005; Revised Manuscript Received June 17, 2005

ABSTRACT: To test whether the folding process of a large protein can be understood on the basis of the folding behavior of the domains that constitute it, we coupled two well-studied small α-helical proteins, the B-domain of protein A (60 amino acids) and Rd-apocytochrome b_{562} (Rd-apocyt b_{562}, 106 amino acids), by fusing the C-terminal helix of the B-domain of protein A with the N-terminal helix of Rd-apocyt b_{562} without changing their hydrophobic core residues. The success of the design was confirmed by determining the structure of the engineered protein with multidimensional NMR methods. Kinetic studies showed that the logarithms of the folding/unfolding rate constants of the engineered protein are linearly dependent on concentrations of guanidinium chloride in the measurable range from 1.7 to 4 M. Their slopes (*m*-values) are close to those of Rd-apocyt b_{562}. In addition, the $^{1}H-^{15}N$ HSQC spectrum taken at 1.5 M guanidinium chloride reveals that only the Rd-apocyt b_{562} domain in the designed protein remained folded. These results suggest that the two domains have weak energetic coupling. Interestingly, the redesigned protein folds faster than Rd-apocyt b_{562}, suggesting that the fused helix stabilizes the rate-limiting transition state.

In recent studies on the mechanism of protein folding, significant effort has been devoted to small single-domain proteins. It has been shown that small proteins (<120 amino acids) generally fold without detectable folding intermediates in kinetic folding experiments (*1*, *2*). Their folding rates were found to correlate with various physical parameters of the structures of native proteins (*3*−*8*). Accordingly, a number of quantitative theoretical models have been developed to predict folding rates of proteins (*3*, *9*−*15*). These studies suggest that topology and the number of folded residues rather than detailed energetic interactions play a dominant role in determining the folding rate of small proteins (*2*, *3*, *16*).

Although folding intermediates are usually not detectable in the kinetic folding experiments for small proteins, hidden intermediates that exist after the rate-limiting transition states have been found for several small proteins, including cytochrome *c* (*16*−*18*), Rd-apocytochrome b_{562} (Rd-apocyt b_{562}) (*19*, *20*), barnase (*21*, *22*), and a construct of the third domain of PDZ (*23*). Moreover, the structures of the intermediates of Rd-apocyt b_{562} at atomic resolution have been obtained by protein engineering and multidimensional NMR (*24*−*26*). They have nativelike topology with broad distributions of non-native side-chain packing. Despite the existence of non-native interactions in these hidden intermediates, Rd-apocyt b_{562} folds very fast with a rate constant of $\sim 10^4$ s^{-1}, suggesting that repacking of the non-native side chains in the absence of topological misfolding has to be fast and cannot be the rate-limiting step, which is consistent with the earlier theoretical studies (*2*, *3*, *16*).

In contrast to the simple two-state folding behavior of small proteins, larger proteins (>120 amino acids) often have detectable early folding intermediates in kinetic folding experiments. Empirical correlations between the measured folding rates and the chain lengths or combination of chain length and topology of native proteins have also been found (*27*−*30*). A general physical model for describing the folding behavior of large proteins, however, is still not available. For example, the cause for the population of the early folding intermediates remains elusive. Several hypotheses have been proposed: (i) A molten-globule model suggests that fast formation of secondary structures and slow packing of side chains lead to the population of submillisecond folding intermediates (*31*). (ii) A misfolding−reorganization hypothesis suggests that an adventurous kinetic barrier causes the population of early folding intermediates that otherwise would exist after the rate-limiting transition state (*32*). (iii) The energy landscape theory suggests that the existence of early folding intermediates results from misfolding events on a rugged energy landscape (*33*, *34*). (iv) A multidomain hypothesis suggests that population of early folding intermediates could be due to the existence of coupled domains that fold with different rates (*24*, *35*). No definitive evidence has been found to prove or disprove any of the hypotheses as a general model for describing the folding behavior of larger proteins.

In principle, these hypotheses may be tested and distinguished by studying the structures and stability of the early folding intermediates. Practically, however, it is very challenging because partially unfolded intermediates are short-lived (less than seconds) during folding. They are very

* Corresponding author. E-mail: yawen@helix.nih.gov. Tel: 301-594-2375. Fax: 301-402-3095.

‡ These authors contributed equally to this work.

§ National Cancer Institute, National Institutes of Health.

|| University at Buffalo, State University of New York.

10.1021/bi050785r CCC: $30.25
Published on Web 08/19/2005

difficult to isolate for structural studies and prone for aggregation. So far, no high-resolution structures for any early folding intermediates have been determined.

Here, we took a reverse approach to investigate these issues. We coupled two small single-domain proteins and studied the folding behavior of the engineered multidomain protein. The goal was to test whether the folding behavior of a large protein with multidomains could be understood through the simple folding behavior of the domains that constitute it. Two well-studied small proteins, the protein A B-domain (PABD, 60 amino acids) and Rd-apocyt b_{562} (106 amino acids), were coupled together by fusing the C-terminal helix of the protein A B-domain with the N-terminal helix of Rd-apocyt b_{562} without changing their hydrophobic cores. The reengineered protein has 146 residues. We solved the structure of this redesigned protein by multidimensional NMR methods. We further characterized its folding pathway by stopped-flow fluorescence, hydrogen exchange, and NMR.

MATERIALS AND METHODS

Computer Modeling and Gene Cloning. To couple the protein A B-domain and Rd-apocyt b_{562} and obtain a six-helix structure, the C_α atoms of the protein A B-domain (PDB code 1BDC) from residues 42−60 was superimposed on those of Rd-apocyt b_{562} from residues 1−19 so that the rmsd is the smallest (rmsd = 2.61 Å). The coordinates of the protein A B-domain structure were then transformed according to the superposition and merged with the coordinates of Rd-apocyt b_{562}. Nineteen amino acid residues in the overlap region were chosen from either Rd-apocyt b_{562} or the protein A B-domain, depending on which residue in the isolated protein is less exposed to solvent. More specifically, the fraction of the solvent-accessible surface area in the total possible solvent-accessible surface area (ASA/ASA_0) is calculated for the 19 N-terminal residues in Rd-apocyt b_{562} and for the 19 C-terminal residues in the protein A B-domain. For each overlapped sequence position, the amino acid residue is chosen from the protein whose residue at the position has a smaller ASA/ASA_0. ASA of a residue is calculated by the Lee−Richards algorithm with a 1.4 Å probe (*36*). ASA_0 is the solvent-accessible area of the unfolded state (*37*).

On the basis of the design, the amino acid sequences of the protein A B-domain in bold (ADNKFNKEQQNAFYEIHPNLNEEQRNGFIQSLKDDPSQ**SANLLAEAKKLNDAQAPK**) and Rd-apocyt b_{562} in bold (**DLEDNWETLNDNLKVIEK**ADNAAQVKDALTKMRAAALDAQKATPPKLEDKSPDSPEMKDFRHGFDILVGQIDDALKLANEGKVKEAQAAAEQLKTTIRAYNQKYG) were substituted with a more hydrophobic, common amino acid sequence (**ALLLNWATLLNNLKAIEK**). It should be noted that the first residues and the last two residues of the protein A B-domain are not included in the design.

The PA_CYTB gene was generated using the overlap extension polymerase chain reaction 1. First, two fragments of PA were amplified in parallel. Forward primer PA1 (5′-CATATGGCGGATAACAAATTCAAC-3′) and reverse primer PA2 (5′-CTGGGTTCGGTTTCGCGAGAAGAAAATTTAACCCGATGGGAGGATTTCTTAAATTTTC-3′) were used to amplify the protein A part. Forward primer CYTB1 (5′-CTTCTTTTAAATTGGGCTACCCTCCTAAATAATTTAAAAGCTATCGAAAAAGCGGATAAC-3′) and reverse primer CYTB2 (5′-ATTAGTCTTCATACCTATTCTACCTAGG-3′) were used to amplify the Rd-apocyt b_{562} gene. Primers of PA2 and CYTB2 were partially complementary at their 5′ ends. Using primer PA1 and CYTB2 and the fragments obtained from the above PCR as a self-annealing template, a chimeric PA_CYTB gene was built up and amplified in the second round of PCR. This product contains the *Nde*I cutting site at the beginning and the *Bam*HI cutting site at the end of the PA_CYTB gene. It was cloned into pET17b(+) vector.

Protein Sample Preparation. Expression and purification of PA_CYTB and its mutants were done using the same procedure for obtaining Rd-apocyt b_{562}. Isotopically enriched proteins for NMR structure determination were grown in *Escherichia coli* [BL21(DE3)] on M9 minimal media containing 1 g/L [U-^{15}N]$^{15}NH_4Cl$ and/or 4 g/L [U-^{13}C]glucose (Isotec) for double (^{13}C,^{15}N) and/or single (^{15}N or ^{13}C) labeled proteins as the sole sources of nitrogen and carbon. NMR samples included $^{13}C/^{15}N$, ^{15}N, ^{13}C, and nonlabeled protein at concentrations of ~2 mM (95% H_2O and 5% D_2O) at pH 5.0 with 20 mM NaOAc-d_4 as buffer.

NMR Spectroscopy for Structure Determination. NMR spectra were collected on a Bruker (Bellerica, MA) DRX 500 MHz spectrometer equipped with a 5 mm *x*,*y*,*z*-shielded pulse field gradient triple resonance probe. A series of three-dimensional experiments [CBCA(CO)NH, HNCACB, HCCH-TOCSY, ^{15}N-edited TOCSY, $^{15}N/^{13}C$-edited NOSEY, 2D ^{1}H NOESY] were collected for complete assignments and NOE measurement. An HNHA experiment was used to determine $^3J_{\alpha N}$ coupling constants for ϕ-angles (*38*). NOE and T_2 of [^{15}N]amides were measured at 25 °C and pH 5.0. 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. 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. NOE was calculated from the intensities of cross-peaks by $NOE = I_{NOE}/I_{NONOE}$. NMR spectra were processed using NMRPipe (*39*) and analyzed using NMRView (*40*).

Calculation of Structures. Structural calculation was done using NIH X-PLOR (version 2.0.6). An extended polypeptide chain of reasonable geometry was used as the initial template. ϕ- and ψ-angles were then randomized before each cycle of the simulated annealing (SA) protocol. Each SA structure was optimized by restrained refinement. Ten structures with lower energy and fewer NOE and dihedral angle violations from 50 SA structures were selected 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 energy and no violation of restraints. These structures were further checked by PROCHECK-NMR (version 3.5.4) (*41*). The structure was determined using NMR.

Stopped-Flow Fluorescence. Stopped-flow fluorescence experiments were performed with a Biologic SFM4 machine (Grenoble, France). The fluorescence excitation wavelength was 280 nm, and emission was collected through a 320 nm cutoff filter. The kinetic folding experiment was initiated by diluting a solution of the protein in a high concentration of

denaturant to various final concentrations of denaturant. The folding and unfolding kinetic phases both show a single exponential phase. The following equations were used to fit the measured folding and unfolding rate constants as a function of GdmCl[1] concentrations:

$$k_{obs} = k_f + k_u \quad (1)$$

$$\log k_f = \log k_f^{H_2O} + m_f[\text{GdmCl}] \quad (2)$$

$$\log k_u = \log k_u^{H_2O} + m_u[\text{GdmCl}] \quad (3)$$

Here k_f and k_u are the folding and unfolding rate constants at corresponding denaturant concentrations. $k_f^{H_2O}$ and $k_u^{H_2O}$ are the folding and unfolding rate constants in water in the absence of denaturant. m_f and m_u are fitting parameters.

Hydrogen Exchange. The hydrogen exchange experiments were initiated by loading a 0.5 mL protein sample in H_2O at pH 5.0 (50 mM Mes) to the spin column (Sephadex 25, bed volume 3 mL) that was preequilibrated with D_2O at pD_{read} 5.0 (50 mM Mes). The column was spun for 1 min at 4000 rpm using a benchtop centrifuge. The eluted protein sample was immediately transferred into an NMR tube for collecting a series of $^1H-^{15}N$ HSQC spectra. The concentrations of the protein samples were ~1 mM. The NMR spectra were processed using NMRPipe (*42*) and analyzed using NMRView (*40*) to obtain the peak intensities as a function of exchanging time and the corresponding rate constants.

RESULTS AND DISCUSSION

Coupling of the B-Domain of Protein A and Rd-apocyt b_{562} by Protein Engineering. PABD is a three-helix bundle protein (Figure 1a). Early experiments demonstrated that PABD folds very rapidly (<1 ms) in the absence of detectable early folding intermediates (*43*). More recently, using NMR line-shape analysis (*44*) and continuous-flow methods (*45*), the folding rate (~10^5 s^{-1} at 25 °C) has been measured directly. In addition, a protein engineering study suggests that the rate-limiting transition state involves the formation of all three helices (*45*). Moreover, due to its small size and simple topology (see discussions in ref *45*), its folding pathway was investigated extensively by using molecular dynamics simulations.

Rd-apocyt b_{562} is a four-helix bundle protein (Figure 1b). Its folding pathway has been characterized in great detail as well. Kinetically, it folds in an apparently two-state manner with a rate constant of ~5×10^3 s^{-1} at 25 °C. Although no early folding intermediates are detectable in the kinetic folding experiments, two hidden partially unfolded intermediates that exist after the rate-limiting transition state have been identified by a native state hydrogen exchange method (*19*). One intermediate involves the unfolding of the N-terminal helix and a major part of the C-terminal helix. The other intermediate involves the unfolding of the N-terminal helix alone. In addition, the rate-limiting transition state has been characterized using the protein engineering method (*19*), which involves the formation of the two middle helices and a part of the C-terminal helix with the N-terminal helix unfolded. More recently, the high-resolution structures of the two intermediates have been solved by native state hydrogen exchange-directed protein engineering and multidimensional NMR methods (*24*, *26*).

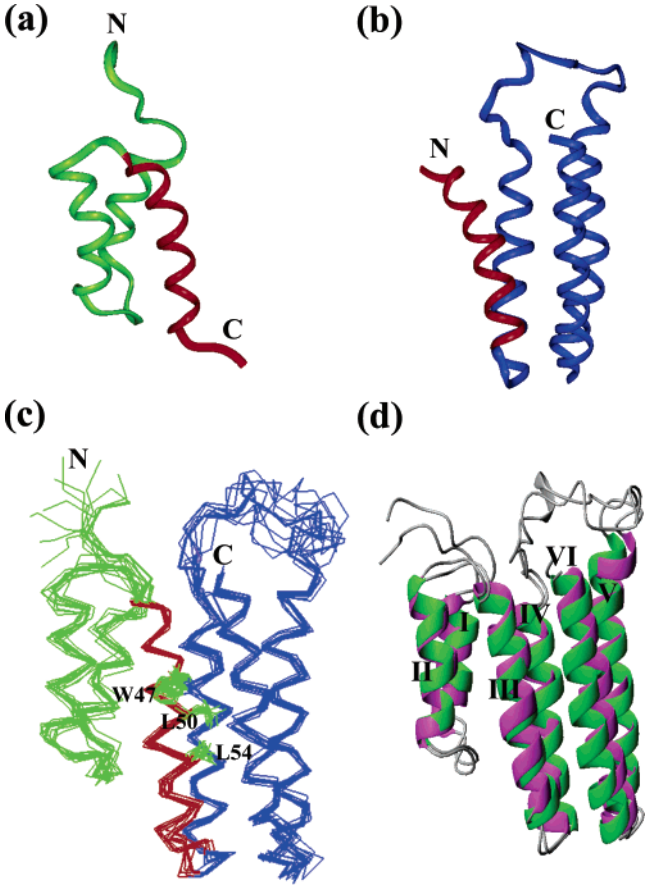

FIGURE 1: Design of a two-domain protein. (a) The ribbon structure of PABD (*48*). (b) The ribbon Rd-apocyt b_{562} (*49*). (c) NMR structures of PA_CYTB. Ten structures are superimposed on the C_α traces. The side chains of the three hydrophobic residues (W47, L50, and L54) are shown in lines. (d) Overlay of the computer-designed (green) and the experimentally determined (purple) structure. Helices in PA_CYTB are numbered I−VI from the N-terminus to the C-terminus.

To create a multidomain protein, the two proteins were coupled together by fusing the C-terminal helix of the B-domain of protein A (from residues 42−60) and the N-terminal helix of Rd-apocyt b_{562} (from residues 1−19) without changing their hydrophobic core residues. The packing of the side chains at the interface of the two domains was modeled using a computer program (see details in Materials and Methods). No amino acids in the first two helices of the B-domain of protein A or in the last three helices of Rd-apocyt b_{562} were changed. The reengineered protein is named as PA_CYTB, and the two domains in PA_CYTB are represented by PA and CYTB, respectively.

Structural Determination of PA_CYTB by NMR. To see whether the design is successful, we determined the structure of PA_CYTB by multidimensional NMR methods. Figure 1c shows the 10 NMR structures that are superimposed on the C_α atoms, along with the side chains of three hydrophobic side chains: W47, L50, and L54. The parameters describing the quality of the structure are listed in Table 1. Figure 1d shows the overlay of the computer-modeled and experimentally determined structures. The two structures have very similar topology; however, some deviation in the C-terminal

[1] Abbreviation: GdmCl, guanidinium chloride.

Table 1: Parameters for the NMR Solution Structure of PA_CYTC

	⟨dgsa⟩	average
rmsd from ideal geometry		
bonds (Å)	0.00435 + 0.00005	0.00435
angles (deg)	0.505 + 0.010	0.505
impropers (deg)	0.562 + 0.014	0.552
NOE (all)	0.0418 + 0.005	0.0418
all residues/helical region		
backbone atoms	1.29/0.59	
all heavy atoms	1.66/1.04	
experimental restraints		
intraresidue NOE	952	
sequential NOE ($\|i-j\| \leq 4$)	997	
long-range NOE ($\|i-j\| \geq 5$)	315	
H-bonds	154	
dihedral angles	138	

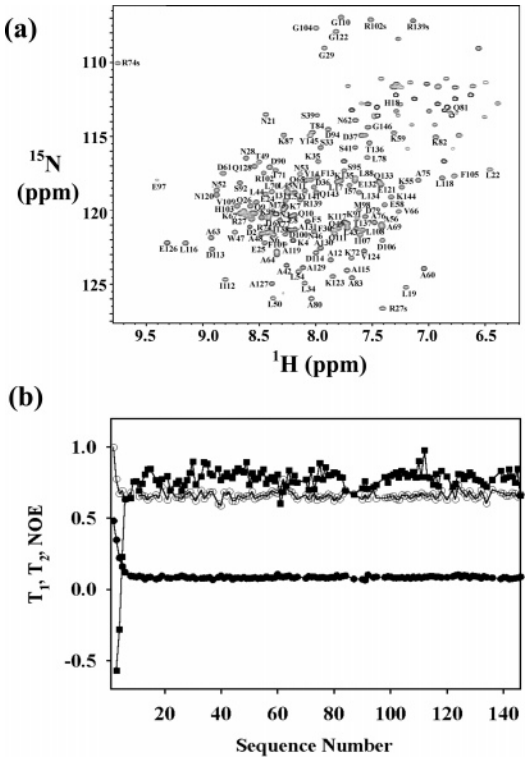

FIGURE 2: Dynamic properties of PA_CYTB. (a) ^{1}H−^{15}N HSQC spectrum of PA_CYTB. (b) T_1 (open circles), T_2 (filled circles), and NOEs (filled squares) of the backbone ^{15}N.

helix is clearly observable. The averaged rmsd of C_α atoms of the designed structure and 10 NMR structures is 2.4 Å.

We also studied the backbone ^{15}N dynamics of PA_CYTB. The ^{1}H−^{15}N cross-peaks in the HSQC spectrum are well dispersed (Figure 2a). T_1, T_2, and NOEs were measured. The T_2 value of ~80 ms for the amide ^{15}N in all six helices suggests that the protein is a monomer. NOE values are positive except for a few residues at the N-terminal end, indicating that the structure is well folded (Figure 2b).

Folding and Unfolding of PA_CYTB. To characterize the folding pathway of PA_CYTB, we used a stopped-flow apparatus (BioLogic) with a dead time of ~1 ms to measure the kinetic folding and unfolding rates by monitoring the fluorescence signal of the Trp residue that lies in the fused helix (see Figure 1c). Refolding was initiated by diluting the proteins in a high concentration of GdmCl to low concentrations. Unfolding was initiated by mixing the folded protein in buffer with high concentrations of GdmCl. Single

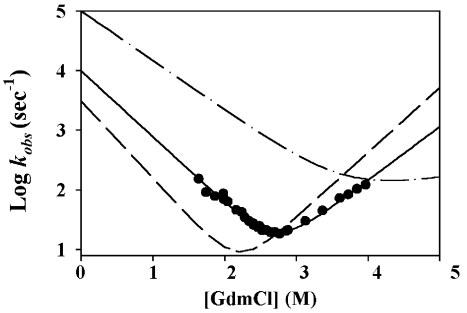

FIGURE 3: Kinetic folding and unfolding. Chevron plots of the protein A B-domain (dot-dashed line), PA_CYTB (solid circles), and Rd-apocyt b_{562} (dashed line).

kinetic folding traces were observed for both folding and unfolding. Figure 3 shows the logarithms of the folding and unfolding rate constants (chevron plot) for PA_CYTB, along with those of the B-domain of protein A and Rd-apocyt b_{562}. The slopes of the logarithms of the folding and unfolding rate constants (*m*-values) are very close to those of Rd-apocyt b_{562}, suggesting that the observed folding kinetics is likely to represent the folding of the CYTB domain. This is consistent with the observation that the sum of m_f and m_u (see Table 2) is much smaller than the value (~3.8 kcal mol^{-1} M^{-1}) expected for the size (145 amino acids) of the protein (*46*).

To test whether a partially unfolded intermediate with the CYTB domain alone folded can be detected under equilibrium conditions, we took a ^{1}H−^{15}N HSQC spectrum at 1.5 M GdmCl, which is close to the lowest concentration at which the kinetic folding rate can be measured (see Figure 3). The dispersed cross-peaks were assigned using triple resonance methods. The spectrum clearly indicates that the major change of the chemical shifts occurs to the residues in the first two helices in the PA domain (see Figure 4a). No cross-peaks for amides in these two helices were found outside the region from 8.0 to 8.5 ppm in the ^{1}H dimension, whereas cross-peaks for the amide protons in other helices are well dispersed, indicating that only the CYTB domain is folded at 1.5 M GdmCl. This conclusion is further supported by results from the amide hydrogen exchange experiment. The ΔG_{HX}s, defined as $-RT \ln(k_{ex}/k_{int})$, for amide protons are shown in Figure 4b. Here, k_{ex} is the measured exchange rate constant and k_{int} is the intrinsic exchange rate constant derived from peptide models (*46*), respectively. The amide protons that show small ΔG_{HX}s, i.e., less protection, are all in the first two helices of PA_CYTB.

It is interesting to note that PA_CYTB folds much faster than Rd-apocyt b_{562} (see Figure 3). Since the N-terminal helix of Rd-apocyt b_{562} was unfolded in the rate-limiting transition state (*19*), this result suggests that the reengineered helix made additional stabilizing interactions in the rate-limiting transition state. These additional stabilization interactions are likely due to the existence of the hydrophobic residues that substitute the hydrophilic residues in the N-terminal helix of Rd-apocyt b_{562} (see Figure 5).

It is also interesting to note that PA_CYTB is fully folded at zero denaturant. Although the kinetic folding of PA_CYTB is too fast to be studied by our stopped-flow machine under such conditions, it is reasonable to speculate that the PA domain may fold early and populate as a kinetic intermediate. This is because the protein A B-domain folds much faster

Table 2: Kinetic Parameters for the Protein and Its Mutants[a]

	k_f (s^{-1})	k_u (s^{-1})	m_f (kcal mol^{-1} M^{-1})	m_u (kcal mol^{-1} M^{-1})	K_{NU}	m_{NU} (kcal mol^{-1} M^{-1})
Rd-apocyt b_{562}	3.1×10^3	1.9×10^{-2}	1.30	1.09	6.13×10^{-6}	2.39
PA_CYTB	7.4×10^4	9.4×10^{-2}	1.50	0.79	1.27×10^{-6}	2.29

[a] The *m* values presented were obtained by multiplying 2.303*RT* (1.365) of those obtained by fitting the experimental data in Figure 3 using eqs 1−3.

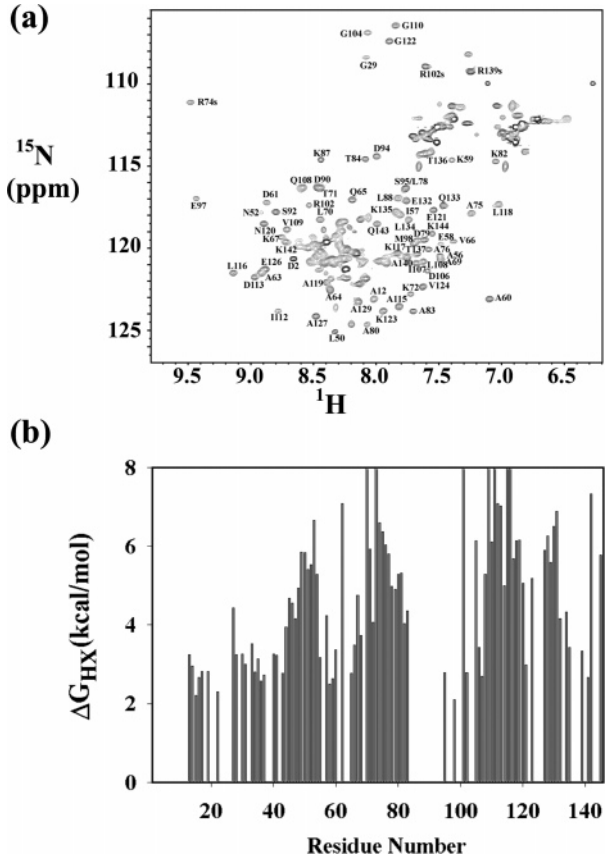

FIGURE 4: Population of the CYTB domain in equilibrium unfolding. (a) ^{1}H−^{15}N HSQC spectrum of PA_CYTB at 1.5 M GdmCl. No cross-peaks outside of the range from 8.0 to 8.5 ppm were found for helices I and II. (b) Hydrogen exchange results for PA_CYTB. The bars that reached to the top of the graph represent the amide protons that exchange too slowly to be studied. Their ΔG_{HX} values are larger than 8 kcal/mol.

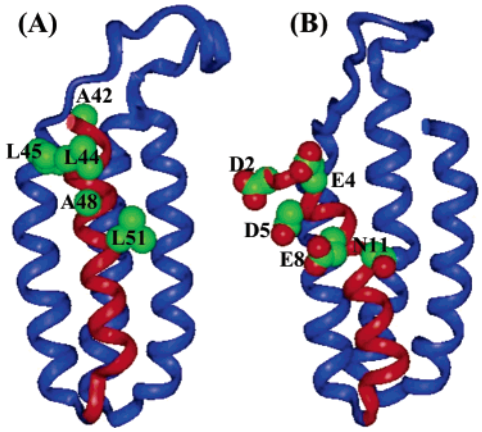

FIGURE 5: Structural illustration of the substitutions in the fused helix. (A) Reengineered side chains in the fused helix (red) of the CYTB domain in PA_CYTB in the absence of the PA domain. (B) The corresponding hydrophilic residues in Rd-apocyt b_{562}. The side chains of these residues are shown in CPK models.

than Rd-apocyt b_{562}. If so, the equilibrium intermediate with the CYTB domain folded could be off the folding pathway. Further investigation on submillisecond folding events will be pursued to test this hypothesis in the future.

The PA_CYTB versus the "Double CI2". Using a small apparent two-state protein CI2, Inaba et al. (*35*) previously have engineered a fusion protein, double CI2, in which one protein is inserted into the loop region of the other one. Their design is based on the rationale that the tertiary structure characteristic of CI2 is virtually conserved upon the cleavage of a loop and the N- and C-terminal ends of CI2 are close in space. Therefore, insertion of another CI2 to the loop region of CI2 is expected to keep both CI2s folded. Two conformers were found to coexist for both domains of CI2. Major phases of the folding and unfolding have double exponential kinetic traces (excluding proline isomerization). Of particular note, however, is that the double CI2 has nonlinear chevron plots in the folding region, which has led to the conclusion that a submillisecond early folding intermediate has populated during the folding of the double CI2. However, since CI2 alone folds slowly with a rate less than 100 s^{-1}, it is unclear why an intermediate can form in the submillisecond time scale for the double CI2. In contrast to double CI2, the two domains in PA_CYTB are not easily identifiable without knowing the history of the design. Therefore, it better represents many large naturally occurring proteins.

CONCLUSIONS

A multidomain protein has been created with the fusion of two small well-studied α-helical proteins. Although the redesigned protein has a well-defined structure at zero denaturant, the two domains are only weakly coupled energetically. At higher concentrations of denaturant, only the CYTB domain remains folded. The redesigned protein folds faster than Rd-apocyt b_{562}, suggesting that the fused helix in PA_CYTB forms additional interactions in the rate-limiting transition state. Further studies on the submillisecond folding events are needed to define the folding pathway of PA_CYTB at zero denaturant. In addition, reengineering of PA_CYTB to couple the two domains more strongly will be pursued.

REFERENCES

1. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Folding Des. 3*, R81−R91.
2. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W., and Sosnick, T. R. (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding, *J. Mol. Biol. 324*, 1−13.
3. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol. 277*, 985−994.

4. Gromiha, M. M., and Selvaraj, S. (2001) Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: Application of long-range order to folding rate prediction, *J. Mol. Biol. 310*, 27−32.
5. Mirny, L., and Shakhnovich, E. (2001) Protein folding theory: from lattice to all-atom models, *Annu. Rev. Biophys. Biomol. Struct. 30*, 361−396.
6. Zhou, H., and Zhou, Y. (2002) Folding rate prediction using total contact distance, *Biophys. J. 82*, 458−463.
7. Bai, Y., Zhou, H., and Zhou, Y. (2004) Critical nucleation size in the folding of small apparently two-state proteins, *Protein Sci. 13*, 1173−1181.
8. Gong, H., Isom, D. G., Srinivasan, R., and Rose, G. D. (2003) Local secondary structure content predicts folding rates for simple, two-state proteins, *J. Mol. Biol. 327*, 1149−1154.
9. Alm, E., and Baker, D. (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11305−11310.
10. Debe, D. A., and Goddard, W. A., III (1999) First principles prediction of protein folding rates, *J. Mol. Biol. 294*, 619−625.
11. Munoz, V., and Eaton, W. A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11311−11316.
12. Makarov, D., Keller, D. A., Plaxco, K. W., and Metiu, H. (2002) How the folding rate constant of simple, single-domain proteins depends on the number of native contacts, *Proc. Natl. Acad. Sci. U.S.A. 99*, 3535−3539.
13. Kaya, H., and Chan, H. S. (2003) Contact order dependent protein folding rates: kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences, *Protein Sci. 52*, 524−533.
14. Makarov, D., and Plaxco, K. W. (2003) The topomer search model: A simple, quantitative theory of two-state protein folding kinetics, *Protein Sci. 12*, 17−26.
15. Weikl, T., and Dill, K. A. (2003) Folding rates and low-entropy-loss routes of two-state proteins, *J. Mol. Biol. 329*, 585−598.
16. Sosnick, T. R., Mayne, L., and Englander, S. W. (1996) Molecular collapse: the rate-limiting step in two-state cytochrome *c* folding, *Proteins 24*, 413−426.
17. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange, *Science 269*, 192−197.
18. Bai, Y., and Englander, S. W. (1996) Future directions in folding: the multi-state nature of protein structure, *Proteins 24*, 145−151.
19. 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.
20. Takei, J., Pei, W., Vu, D., and Bai, Y. (2002) Populating partially unfolded forms by hydrogen exchange-directed protein engineering, *Biochemistry 41*, 12308−12312.
21. Vu, N. D., Feng, H., and Bai, Y. (2004) The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions, *Biochemistry 43*, 3346−3356.
22. Khan, F., Chuang, J. I., Gianni, S., and Fersht, A. R. (2003) The kinetic pathway of folding of barnase, *J. Mol. Biol. 333*, 169−186.
23. Feng, H., Vu, N. D., and Bai, Y. (2005) Detection of a hidden intermediate in the folding of the third domain of PDZ, *J. Mol. Biol. 343*, 1477−1485.
24. 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.
25. Feng, H., Vu, D., and Bai, Y. (2004) Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b_{562}: Native fold with non-native hydrophobic interactions, *J. Mol. Biol. 343*, 1477−1485.
26. Feng, H. Q., Zhou, Z., and Bai, Y. (2005) Protein folding pathway: Multiple folding intermediates at atomic resolution, *Proc. Natl. Acad. Sci. U.S.A. 102*, 5026−5031.
27. Galzitskaya, O. V., Garbuzynskiy, S. O., Ivankov, D. N., and Finkelstein, A. (2003) Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics, *Proteins 51*, 162−166.
28. Ivankov, D. N., Garbuzynskiy, S. O., Alm, E., Plaxco, K. W., Baker, D., and Finkel'shtein, A. V. (2003) Contact order revisited: influence of protein size on the folding rate, *Protein Sci. 12*, 2057−2062.
29. Naganathan, A. N., and Munoz, V. (2005) Scaling of folding times with protein size, *J. Am. Chem. Soc. 127*, 480−481.
30. Kamagata, K., Arai, M., and Kuwajima, K. (2004) Unification of the folding mechanism of non-two-state and two-state proteins, *J. Mol. Biol. 339*, 951−965.
31. Ptitsyn, O. B. (1995) How the molten globule became, *Trends Biochem. Sci. 20*, 376−379.
32. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) The barriers in protein folding, *Nat. Struct. Biol. 1*, 149−156.
33. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol. 4*, 10−19.
34. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis, *Proteins 21*, 167−195.
35. Inaba, K., Kobayashi, N., and Fersht, A. R. (2000) Conversion of two-state to multi-state folding kinetics on fusion of two protein foldons, *J. Mol. Biol. 302*, 219−233.
36. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol. 55*, 379−400.
37. Shrake, A., and Rupley, J. A. (1973) Environment and exposure to solvent of protein atoms: Lysozyme and insulin, *J. Mol. Biol. 79*, 351−371.
38. Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: A new approach for measuring homonuclear three-bond J(HNHa) coupling constants in ^{15}N-enriched proteins, *J. Am. Chem. Soc. 115*, 7772−7777.
39. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multi-dimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
40. 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.
41. 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.
42. 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.
43. Bai, Y., Karimi, A., Dyson, H. J., and Wright, P. E. (1997) Absence of a stable intermediate on the folding pathway of protein A, *Protein Sci. 6*, 1449−1457.
44. Myers, J. K., and Oas, T. G. (2001) Pre-organized secondary structure as an important determinant of fast protein folding, *Nat. Struct. Biol. 8*, 552−558.
45. Sato, S., and Fersht, A. R. (2004) Testing protein-folding simulations by experiment: B-domain of protein A, *Proc. Natl. Acad. Sci. U.S.A. 101*, 6952−6956.
46. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
47. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins 17*, 75−86.
48. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y., and Shimada, I. (1992) Three-dimensional solution structure of B domain of staphylococcal protein A: Comparisons of the solution and crystal structures, *Biochemistry 31*, 9665−9672.
49. 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.

BI050785R