Identification and Characterization of Key Substructures Involved in the Early Folding Events of a (β/α)8-barrel Protein as Studied by Experimental and Computational Methods

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · DECEMBER 2005

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2005.08.070 · Source: PubMed

2 AUTHORS:

Satoshi Akanuma
Waseda University
45 PUBLICATIONS **384** CITATIONS
SEE PROFILE

Akihiko Yamagishi
Tokyo University of Pharmacy and Life Scie…
284 PUBLICATIONS **3,233** CITATIONS
SEE PROFILE

doi:10.1016/j.jmb.2005.08.070 J. Mol. Biol. (2005) 353, 1161–1170

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Identification and Characterization of Key Substructures Involved in the Early Folding Events of a $(\beta/\alpha)_8$-barrel Protein as Studied by Experimental and Computational Methods

Satoshi Akanuma[1,2] and Akihiko Yamagishi[1,2]*

[1]*Department of Molecular Biology, Tokyo University of Pharmacy and Life Science 1432-1 Horinouchi, Hachioji Tokyo 192-0392, Japan*

[2]*BIRD, JST (Japan Science and Technology Corporation) 5-3 Yonban-cho, Chiyoda Tokyo 102-0081, Japan*

**Corresponding author*

A number of studies have examined the structural properties of late folding intermediates of $(\beta/\alpha)_8$-barrel proteins involved in tryptophan biosynthesis, whereas there is little information available about the early folding events of these proteins. To identify the contiguous polypeptide segments important to the folding of the $(\beta/\alpha)_8$-barrel protein *Escherichia coli* *N*-(5′-phosphoribosyl)anthranilate isomerase, we structurally characterized fragments and circularly permuted forms of the protein. We also simulated thermal unfolding of the protein using molecular dynamics. Our fragmentation experiments demonstrate that the isolated $(\beta/\alpha)_{1\text{-}4}\beta_5$ fragment is almost as stable as the full-length protein. The far and near-UV CD spectra of this fragment are indicative of native-like secondary and tertiary structures. Structural analysis of the circularly permutated proteins shows that if the protein is cleaved within the two N-terminal $\beta\alpha$ modules, the amount of secondary structure is unaffected, whereas, when cleaved within the central $(\beta/\alpha)_{3\text{-}4}\beta_5$ segment, the protein simply cannot fold. An ensemble of the denatured structures produced by thermal unfolding simulations contains a persistent local structure comprised of β_3, β_4 and β_5. The presence of this three-stranded β-barrel suggests that it may be an important early-stage folding intermediate. Interactions found in $(\beta/\alpha)_{3\text{-}4}\beta_5$ may be essential for the early events of ePRAI folding if they provide a nucleation site that directs folding.

Keywords: $(\beta/\alpha)_8$-barrel protein; circular permutation; fragmentation; molecular dynamics simulation; protein folding

Introduction

Elucidation of protein folding mechanisms is fundamental to our understanding of how disordered polypeptide chains fold to their biologically active, native structures.[1,2] An increasing number of experimental and theoretical studies have shown that, for a given protein, kinetically accessible folding pathways are few in number.[3–6] Often, one or a few segments of a protein can fold independently and locally before the global native structure forms. In these cases, a transient intermediate ensemble appears in which certain segments of the polypeptide chain are partially folded, while the rest of the protein remains disordered.[7,8] One way of identifying such segments is to fragment the polypeptide chain and to then study the physical properties of the fragments.[9,10] If a fragment autonomously and properly folds, and is stable enough to maintain the fold, then it is likely that the fragment functions as an independent folding unit. A complementary approach, circular permutation of the polypeptide chain, can identify

Abbreviations used: PRAI, *N*-(5′-phosphoribosyl)-anthranilate isomerase; ePRAI, PRAI from *Escherichia coli*, this abbreviation is used to designate the full-length protein; IGPS, indoleglycerolphosphate synthase; αTS, α-subunit of tryptophan synthase; CD, circular dichroism; MD, molecular dynamics; RMSD, root-mean-square deviation; R_G, radius of gyration; IPTG, isopropyl-thio-β-D-galactopyranoside.

E-mail address of the corresponding author: yamagish@ls.toyaku.ac.jp

contiguous protein segments that are essential for folding and stability.[11–13] Permutation involves joining the original N and C termini through a peptide linker and also creating new termini by cleaving a peptide bond at a sequentially distant site. Unlike protein fragmentation, circular permutation perturbs the amino acid sequence without altering the amino acid composition.

N-(5′-Phosphoribosyl)anthranilate isomerase (PRAI), indoleglycerolphosphate synthase (IGPS), and the α-subunit of tryptophan synthase (αTS) catalyze successive reactions in tryptophan biosynthesis. These three enzymes belong to the same $(\beta/\alpha)_8$-barrel family, which has the most common protein fold known and is adopted by ~10% of all proteins whose three-dimensional structures have been solved.[14] The canonical $(\beta/\alpha)_8$-barrel protein has eight repeating units of a β-strand/α-helix module. The eight parallel β-strands form a central barrel that is surrounded by the eight α-helices. The crystal structure of PRAI from *Escherichia coli* (ePRAI) was solved as a part of an IGPS-PRAI bifunctional enzyme complex.[15] ePRAI is one of the smallest $(\beta/\alpha)_8$-barrel proteins known. It has 197 residues with no additional secondary structure elements either preceding the β_1 strand or following the α_8 helix. For ePRAI, the canonical α_5 helix is replaced by a loop in the structure. The truncated monofunctional ePRAI domain is as stable and as active as the bifunctional protein.[16]

Kinetic and equilibrium folding intermediates of ePRAI, as well as those of PRAI from *Saccharomyces cerevisiae* and those of αTS from *E. coli*, have been characterized extensively.[17–23] Structurally similar equilibrium folding intermediates have been found for the three enzymes. These intermediates have the first six βα units folded into native-like structures, while the last two βα units remain partially unfolded. An ensemble of unfolding intermediates, with structural properties similar to the equilibrium folding intermediates, appeared during our previous molecular dynamics (MD) simulation.[24] In that study, we computed MD trajectories for temperature-induced ePRAI unfolding so that its entire folding/unfolding pathway could be characterized. These intermediates probably appear relatively late in folding. Consequently, less is known about the early folding events of these $(\beta/\alpha)_8$-barrel proteins.

Here, we studied the structural characteristics of ePRAI fragments and circularly permutated variants. We identified a contiguous polypeptide segment of ePRAI that folds properly and is stable. We also computed MD trajectories for the temperature-induced unfolding of ePRAI, which produced an ensemble of denatured structures. Within the denatured ensemble, locally ordered regions were identified and characterized. These ordered regions may reflect the structures of early ePRAI folding intermediates.

Results

Fragmentation experiments

Figure 1 illustrates the ePRAI fragments and circularly permuted variants engineered for this study. The expressed fragments accumulated in inclusion bodies. The two N-terminal fragments, $(\beta/\alpha)_{1\text{-}6}$ and $(\beta/\alpha)_{1\text{-}4}\beta_5$, could be refolded as described in Materials and Methods. The extent of intermolecular association for ePRAI and its fragments, was investigated using analytical gel-filtration with two different initial protein concentrations (4 μM and 20 μM: Table 1 and Supplementary Figure 1(a)). The apparent molecular mass of ePRAI is slightly greater than its calculated

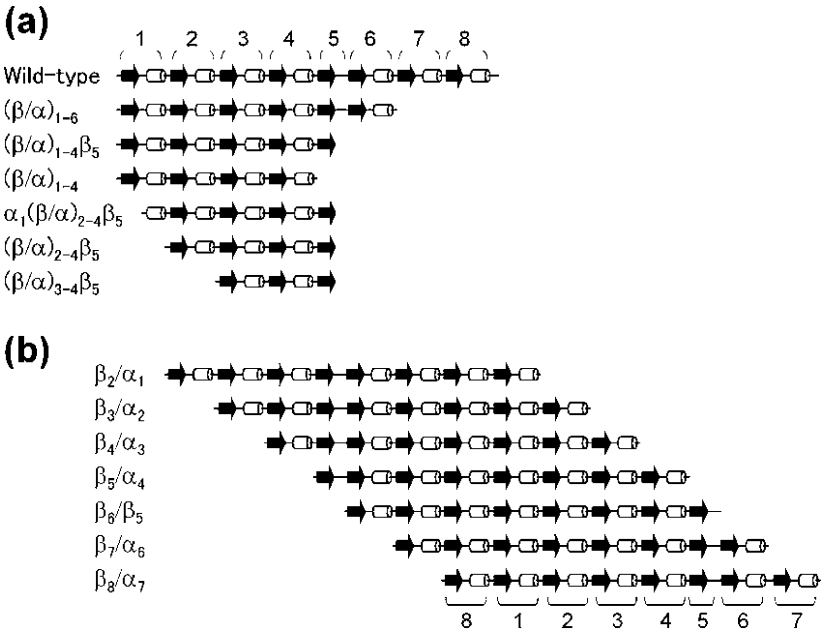

Figure 1. Linear secondary structure representations of ePRAI, (a) its truncated segments and (b) its circularly permuted variants constructed for this study. Arrows and cylinders indicate β-strands and α-helices, respectively. The β/α modules are numbered in sequential order from the original N terminus.

Table 1. Apparent molecular masses and deduced association states

Protein	Apparent molecular mass (kDa)[a]	Deduced association state
ePRAI	22.9	Monomer
$(\beta/\alpha)_{1\text{-}6}$	18.7	Monomer
$(\beta/\alpha)_{1\text{-}4}\beta_5$	14.5 + 26.1[b]	Monomer + dimer[b]
$(\beta/\alpha)_{1\text{-}4}$	Not refolded	
$\alpha_1(\beta/\alpha)_{2\text{-}4}\beta_5$	Not refolded	
$(\beta/\alpha)_{2\text{-}4}\beta_5$	Not refolded	
$(\beta/\alpha)_{3\text{-}4}\beta_5$	Not refolded	
β_2/α_1	27.2 + aggregate	Monomer + aggregate
β_3/α_2	Aggregate	Aggregate
β_4/α_3	Not refolded	
β_5/α_4	Not refolded	
β_6/β_5	27.5	Monomer
β_7/α_6	26.9 + 46.5 + 94.3[b]	Monomer + dimer + tetramer[b]
β_8/α_7	26.7	Monomer

[a] The calculated molecular mass of ePRAI is 21.1 kDa. The proteins were dissolved in 20 mM potassium phosphate buffer (pH 7.6), 300 mM NaCl so that initial protein concentrations were 4.0 and 20 μM.
[b] This peak was observed when the initial protein concentration was 20 μM.

molecular mass (21.1 kDa). The gel-filtration experiment indicates that $(\beta/\alpha)_{1\text{-}6}$ exists as monomer. Similarly, $(\beta/\alpha)_{1\text{-}4}\beta_5$ exists predominantly as a monomer, but a small elution peak corresponding to the dimeric state was also observed, especially when the initial protein concentration was 20 μM. The dimer might be a barrel-like structure that sequesters the otherwise exposed hydrophobic surfaces created by deletion of $(\beta/\alpha)_{6\text{-}8}$. A propensity to dimerize has also been reported for the imidazoleglycerol phosphate synthase $(\beta/\alpha)_{1\text{-}4}$ and $(\beta/\alpha)_{5\text{-}8}$ half-barrels,[25] as well as for $\alpha_0(\beta/\alpha)_{1\text{-}4}\beta_5$ of αTS.[9] The shorter fragments, $(\beta/\alpha)_{1\text{-}4}$, $\alpha_1(\beta/\alpha)_{2\text{-}4}\beta_5$, $(\beta/\alpha)_{2\text{-}4}\beta_5$, and $(\beta/\alpha)_{3\text{-}4}\beta_5$, precipitated during refolding and therefore could not be further purified or structurally characterized.

The far and near-UV circular dichroism (CD) spectra of ePRAI, $(\beta/\alpha)_{1\text{-}6}$, and $(\beta/\alpha)_{1\text{-}4}\beta_5$ are presented in Figure 2(a) and (b). The far-UV CD spectra monitor the secondary structure content of ePRAI and its fragments. The $(\beta/\alpha)_{1\text{-}6}$ spectrum is almost identical with that of ePRAI, indicating that $(\beta/\alpha)_{1\text{-}6}$ has a relative secondary structure content almost identical with that of ePRAI. The ellipticity of $(\beta/\alpha)_{1\text{-}4}\beta_5$ is greater than those of ePRAI and $(\beta/\alpha)_{1\text{-}6}$, which reflects the greater proportion of helical segments possible for the smaller fragment.[15]

Often, near-UV CD spectra are dominated by tryptophan ellipticity and thus can be used as a sensitive criterion for tertiary structure. Trp101 and Trp136 are found in β_5 and α_6, respectively. The shape and intensity of the near-UV CD spectra of both $(\beta/\alpha)_{1\text{-}6}$ and $(\beta/\alpha)_{1\text{-}4}\beta_5$ resemble those of the

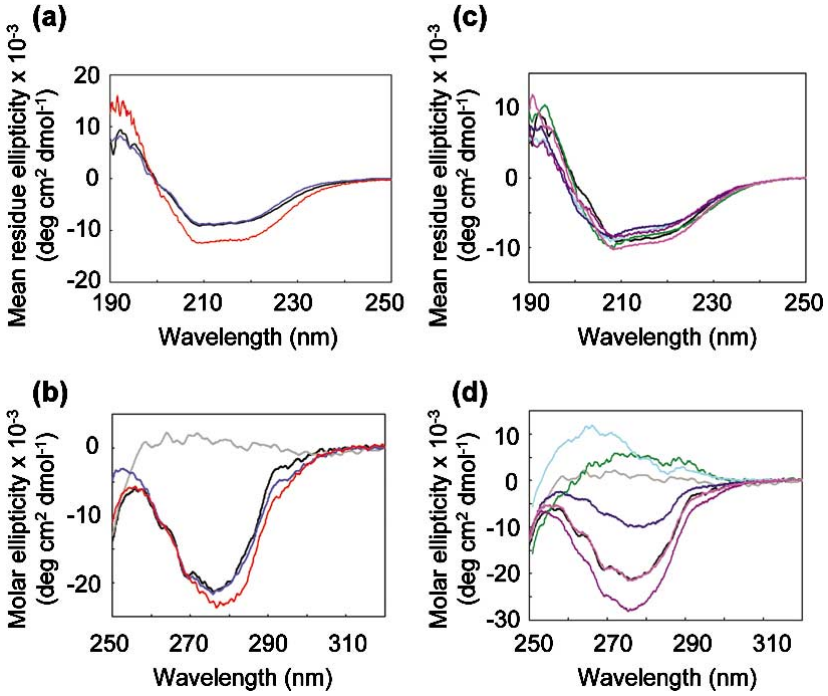

Figure 2. CD spectra of ePRAI and its fragments ((a) and (b)) and of the circularly permuted variants ((c) and (d)). (a) and (c) Far-UV CD spectra were acquired using a 0.1 cm path-length cell. Protein concentrations were 0.1–0.2 mg ml^{-1}. (b) and (d) Near-UV CD spectra were acquired using a 1.0 cm path-length cell. Protein concentrations were 1.0 mg ml^{-1}. The near-UV CD spectrum for ePRAI unfolded with 7 M urea is also shown. The color code is: black, wild-type ePRAI; gray, unfolded ePRAI with 7 M urea present; blue, $(\beta/\alpha)_{1\text{-}6}$; red, $(\beta/\alpha)_{1\text{-}4}\beta_5$; green, β_2/α_1; cyan, β_3/α_2; deep blue, β_6/β_5; purple, β_7/α_6; and magenta, β_8/α_7.

ePRAI spectrum, indicating that the two fragments have tertiary structures similar to that of ePRAI.

Figure 3(a) illustrates the disruption of secondary structure, monitored by the change in ellipticity at 220 nm as a function of urea concentration. The CD curve for ePRAI exhibits a cooperative unfolding transition between 1.0 M and 3.0 M urea, and a small transition between 3.0 M and 4.0 M urea. The break in the cooperative unfolding transition around 3 M urea suggests the presence of an equilibrium intermediate. In contrast, cooperative unfolding transitions were observed for the two fragments. These results indicate that $(\beta/\alpha)_{1\text{-}6}$ and $(\beta/\alpha)_{1\text{-}4}\beta_5$ have native-like, compactly folded tertiary structures.

In order to compare quantitatively the disruption of secondary structures in ePRAI, $(\beta/\alpha)_{1\text{-}6}$, and $(\beta/\alpha)_{1\text{-}4}\beta_5$, the equilibrium unfolding curves were also plotted as "apparent fraction of unfolded molecules" *versus* "urea concentration" (Figure 4). Quantitative analyses of the apparent stabilities for the full-length protein and the two fragments were performed by fitting the unfolding curves to an apparent three-state or a two-state transition model. The curve for ePRAI fits a three-state model. The first transition probably involves the disruption of $\alpha_6(\beta/\alpha)_{7\text{-}8}$ secondary structure, as observed for yeast PRAI unfolding.[22,26] However, the helical structures in $(\beta/\alpha)_{1\text{-}5}\beta_6$ may be disrupted to some extent, as judged by the magnitude of ellipticity around 3 M urea. The unfolding curves for $(\beta/\alpha)_{1\text{-}6}$ and $(\beta/\alpha)_{1\text{-}4}\beta_5$ best fit simple two-state models.

The resulting free-energy differences between the native and unfolded forms (ΔG°) of the fragments, between the native form and the intermediate (ΔG_1°) of ePRAI, and between the intermediate and the unfolded form (ΔG_2°) of ePRAI are listed in Table 2. The ΔG° value for $(\beta/\alpha)_{1\text{-}6}$ is similar to the ΔG_2° found for ePRAI, suggesting that the addition of $(\beta/\alpha)_{7\text{-}8}$ hardly influences $(\beta/\alpha)_{1\text{-}6}$ stability. However, $(\beta/\alpha)_{1\text{-}4}\beta_5$ has a remarkably decreased ΔG°. These results suggest that there are two structural boundaries for distinguishable folding modules; one is located after the fifth β strand and the other after the sixth one.

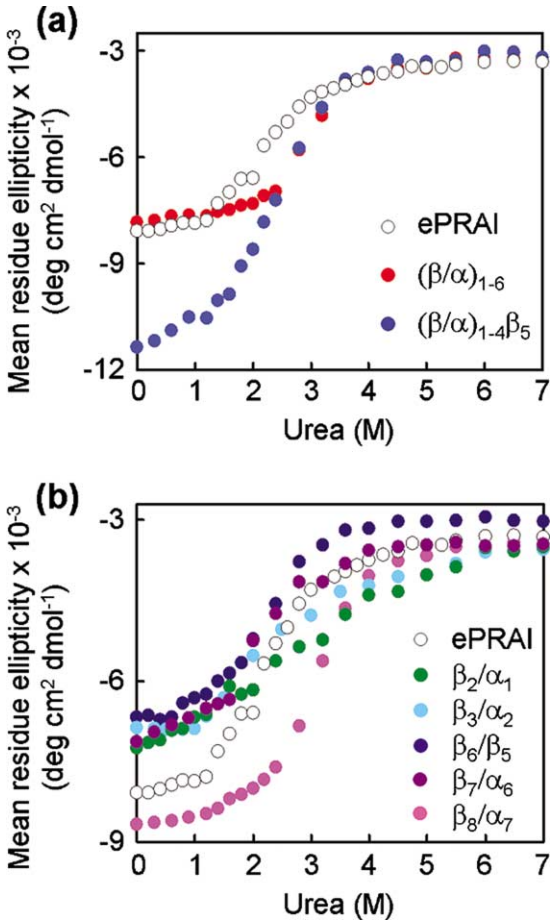

Figure 3. Equilibrium unfolding. Urea-induced unfolding of (a) ePRAI and its fragments and (b) of the circularly permuted variants, at 22 °C, was assessed using the 220 nm ellipticity. The unfolding of each protein was almost completely reversible as demonstrated by the recovery of ellipticity upon refolding. Protein concentrations were 0.1 mg ml^{-1}. Each datum point is the average of at least three independent measurements.

Circular permutation experiments

Seven circularly permuted variants (Figure 1(b)) were constructed in order to identify contiguous segments that are essential for ePRAI folding and stability. The natural termini were connected *via* a six-residue peptide linker (YDPSGG) and new termini were created at surface loops connecting the C termini of α-helices and the N termini of β-strands. The variants accumulated in inclusion bodies. Five of them could be refolded, yielding soluble proteins. The other two variants, β_4/α_3 and β_5/α_4, for which new termini were generated between α_3 and β_4 and between α_4 and β_5, respectively, could not be refolded to soluble

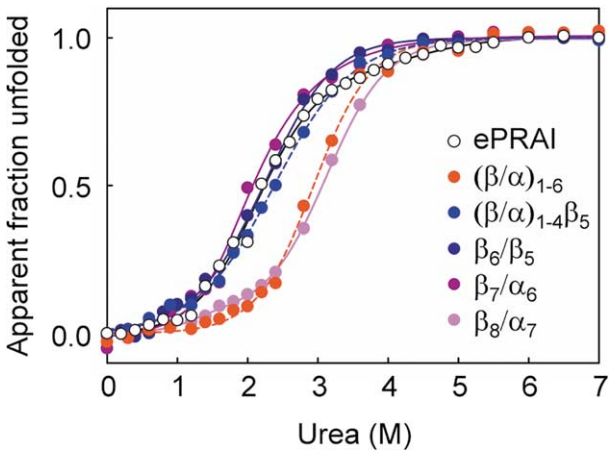

Figure 4. Normalized equilibrium unfolding plots for ePRAI, its fragments, and the circularly permuted variants. Continuous lines are the fits to apparent three-state equilibrium unfolding models. Broken lines are the fits to apparent two-state equilibrium unfolding models.

Table 2. Thermodynamic parameters obtained from equilibrium unfolding experiments

Protein	ΔG_1° (kcal mol^{-1})[a]	m_1 (kcal mol^{-1} M^{-1})[a]	ΔG_2° (kcal mol^{-1})[a]	m_2 (kcal mol^{-1} M^{-1})[a]	ΔG^o (kcal mol^{-1})[b]	m (kcal mol^{-1} M^{-1})[b]
ePRAI	2.79	1.32	4.44	1.02		
$(\beta/\alpha)_{1\text{-}6}$					4.21	1.43
$(\beta/\alpha)_{1\text{-}4}\beta_5$					2.52	1.06
β_6/β_5	2.27	1.27	3.27	1.35		
β_7/α_6	0.76	0.62	3.40	1.79		
β_8/α_7	3.25	2.69	4.37	1.40		

[a] Free energy differences, ΔG_1° and ΔG_2°, and cooperativity indices, m_1 and m_2, were obtained from curve fits using apparent three-state transition models.
[b] Free energy differences, ΔG°, and cooperativity indices, m, were obtained from curve fits using simple two-state models.

species. These observations suggest that the residues essential for folding are distributed over $(\beta/\alpha)_{3\text{-}4}\beta_5$ and that cleavage of $(\beta/\alpha)_{3\text{-}4}\beta_5$ causes ePRAI to completely lose the ability to fold.

The extents of intermolecular association for the circularly permuted variants are summarized in Table 1. Analytical gel-filtration analysis indicates that β_6/β_5 and β_8/α_7 are mainly monomeric. The β_7/α_6 elution profile shows that the variant exists predominantly as a monomer, but small elution peaks corresponding to dimeric and tetrameric states are also seen (Supplementary Figure 1(b)). The elution profile of β_2/α_1 indicates the existence of monomers, but also of oligomers and aggregates. β_3/α_2 tends to form aggregates while remaining soluble.

The structural properties of the purified circularly permutated variants were also investigated by CD spectroscopy. The far-UV CD spectra of β_6/β_5, β_7/α_6, and β_8/α_7 are almost identical with that of the wild-type protein (Figure 2(c)), which is also true of the aggregate-forming variants, β_2/α_1 and β_3/α_2. Therefore, the five purified variants and ePRAI have similar amounts and types of secondary structure. The shape of the β_8/α_7 near-UV CD spectrum is very similar to that of ePRAI (Figure 2(d)). Decreased and increased ellipticities were observed for the near-UV CD spectra of β_6/β_5 and β_7/α_6, respectively, although these spectra are still indicative of well-defined, native-like tertiary structures. The altered CD intensities are most likely caused by local changes in the environments of Trp101 in β_6/β_5 and Trp136 in β_7/α_6 caused by cleaving the loops connecting β_5 and β_6, and α_6 and β_7, respectively. For β_2/α_1 and β_3/α_2, as shown in Figure 2(d), the shapes of their near-UV CD spectra obviously differ from those of ePRAI and the other circularly permuted variants. The amplitudes are negative for the spectra of wild-type ePRAI, β_6/β_5, β_7/α_6 and β_8/α_7, whereas they are positive for β_2/α_1 and β_3/α_2. The spectral characteristics suggest non-native tertiary structures for β_2/α_1 and β_3/α_2.

The influence of the circular permutations on folding and stability was also investigated by urea-induced equilibrium folding/unfolding transition experiments. Figure 3(b) shows that β_6/β_5, β_7/α_6 and β_8/α_7 exhibit cooperative secondary structure unfolding transitions. Three-state fits best characterize the continuous lines in Figure 4(b). The associated free energy differences, ΔG_1° and ΔG_2°, and the cooperativity indices, m_1 and m_2, are listed in Table 2. A substantial shift in the unfolding transition to a higher urea concentration is observed for β_8/α_7. This variant exhibits an increased ΔG_1° and a similar ΔG_2° when compared to those of ePRAI. However, β_2/α_1 and β_3/α_2 may both be relatively flexible structures as reflected by unfolding transitions that are less cooperative than that found for ePRAI (Figure 3(b)).

Table 3. Structural properties of the denatured ensemble

	RMSD (Å)[a]		R_G (Å)[a]	
	Native states[b]	Denatured ensemble[c]	Native state[b]	Denatured ensemble[c]
Full-length	1.6 (0.2)	11.0 (0.6)	15.7 (0.1)	16.8 (0.7)
$(\beta/\alpha)_{1\text{-}2}$[d]	0.9 (0.1)	5.5 (2.8)	10.3 (0.1)	11.2 (0.7)
$(\beta/\alpha)_{3\text{-}4}\beta_5$[d]	0.8 (0.1)	4.4 (2.3)	10.1 (0.1)	10.9 (0.4)
$(\beta/\alpha)_6$[d]	2.1 (0.4)	10.7 (2.0)	11.7 (0.1)	13.7 (1.8)
$(\beta/\alpha)_{7\text{-}8}$[d]	1.4 (0.2)	11.3 (1.1)	10.5 (0.1)	13.6 (1.6)

[a] The RMSD and R_G values are derived from the C^α coordinates.
[b] The 10,000 native state structures were taken from 1 ps snapshots of the 300 K control simulation. The values are averaged over the 10 ns simulation. Standard deviations are given in parentheses.
[c] The values were calculated in the same manner as those of the native ensemble, except that the coordinates of 8000 structures were used.
[d] The segments $(\beta/\alpha)_{1\text{-}2}$, $(\beta/\alpha)_{3\text{-}4}\beta_5$, $(\beta/\alpha)_6$ and $(\beta/\alpha)_{7\text{-}8}$ correspond to residues 1–50, 51–105, 106–135 and 136–197, respectively. The deviations were calculated as follows. First, the segments were extracted from the protein structure at each time-step in the unfolding trajectory of the full-length protein. Then the extracted segment was optimally superimposed onto the corresponding segment of the initial structure and the deviations calculated.

TrpF activity can be assayed using a genetic complementation system that incorporates an *E. coli* strain lacking the chromosomal *trpF* gene.[27] The circularly permuted *trpF* genes coding for the folded variants β_6/β_5, β_7/α_6, and β_8/α_7 were cloned into the plasmid vector pUC119 and then the plasmids were introduced into *trpF*-defective *E. coli*. The resulting transformants grew on selective medium lacking tryptophan, indicating that β_6/β_5, β_7/α_6, and β_8/α_7 are catalytically active (data not shown).

Thermal unfolding MD simulations

Temperature-induced ePRAI unfolding was simulated at 500 K and 525 K. The first 7 ns of duplicate unfolding simulations at each temperature, starting from different initial structures, as well as a control simulation at 300 K, have been described.[24] Here, the unfolding simulations were continued to 20 ns and 12 ns at 500 K and 525 K, respectively, to produce an ensemble of unfolded structures. We sampled structures present between 18 ns and 20 ns at 500 K and structures present between 10 ns and 12 ns at 525 K. These structures represent the denatured state. From the four independent trajectories at two different temperatures, an ensemble of 8000 structures was assembled. The increases in the averaged C^α RMSD and in the average R_G that occurred by the end of the simulation ensure that the structures represent the unfolded state (Table 3). The average number of native hydrogen bonds found for species present in the control simulation is 81.5, whereas, the denatured structures have an average of 13.6 native hydrogen bonds (data not shown). The C^α RMSD values and the R_G values were also calculated for the four segments, residues 1–50, 51–105, 106–146 and 147–197 corresponding to $(\beta/\alpha)_{1\text{-}2}$, $(\beta/\alpha)_{3\text{-}4}\beta_5$, $(\beta/\alpha)_6$ and $(\beta/\alpha)_{7\text{-}8}$, respectively (Table 3). The RMSD and the R_G values for $(\beta/\alpha)_{1\text{-}2}$ and $(\beta/\alpha)_{3\text{-}4}\beta_5$ are significantly smaller than those of $(\beta/\alpha)_6$ and $(\beta/\alpha)_{7\text{-}8}$, indicating that the two N-terminal segments are relatively stable and retain compact structures when denatured. Table 4 lists the native hydrogen bonds present in at least 50% of the structures in the denatured ePRAI ensemble. These persistent hydrogen bonds are found in the central native β-barrel; one bond joins β_2 and β_3; three join β_3 and β_4; and four join β_4 and β_5. Therefore, the structures of the denatured ensemble are not fully disordered. Instead, the three-stranded β-sheet consisting of β_3, β_4 and β_5 remains significantly structured. It has been postulated that the residual structure in an unfolded protein might serve as a nucleation site.[28–30] Such a site might accelerate folding by limiting the accessible conformation space. The three-stranded β-barrel intermediate may be an essential scaffold upon which other segments of ePRAI fold.

Table 4. Persistent native hydrogen bonds present in at least 50% of the structures in the denatured ensemble

Atom 1	Atom 2	Frequency (%)
Gly24:O (β_2)	Val53:HN (β_3)	53
Gly54:HN (β_3)	Ala75:O (β_4)	76
Gly54:O (β_3)	Gln77:HN (β_4)	89
Phe56:HN (β_3)	Gln77:O (β_4)	76
Val76:HN (β_4)	Ala99:O (β_5)	51
Val76:O (β_4)	Trp101:HN (β_5)	63
Leu78:HN (β_4)	Trp101:O (β_5)	62
Leu78:O (β_4)	Ala103:HN (β_5)	54

The secondary structural element in which each residue is located in the native state is indicated in parentheses.

Discussion

Here, two experimental techniques, fragmentation and circular permutation, were used in conjunction with MD simulations to identify polypeptide segments important to the early events of ePRAI folding. It has been difficult to identify early folding intermediates and to characterize their structures by experimental methods, because such intermediates appear only transiently. MD protein unfolding simulations allow for the characterization of atomic-level resolution structures at each step of unfolding. However, one or more important features occurring during unfolding may be absent or difficult to identify given a limited number of MD trajectories. In contrast, experimental results reflect the average behavior of a large number of molecules. Therefore, experimental studies in combination with simulations, as shown herein, are the most powerful means available for studying early protein folding events.[31,32]

The $(\beta/\alpha)_8$-barrel proteins involved in tryptophan biosynthesis have been studied by methods that identify independent folding units.[9,21,22,26,33,34] Based on fragmentation and circular permutation experiments of yeast PRAI, Kirschner and coworkers[22,26] postulated a "6+2" folding model, for which the six N-terminal β/α modules fold first, providing a structural scaffold for the association and folding of the two C-terminal β/α modules. Equilibrium and kinetic folding/unfolding studies of ePRAI[21] and proteolytic fragmentation experiments of αTS[33,34] support the 6+2 folding mechanism. Zitzewitz *et al.*[9] constructed a series of amino-terminal fragments so that the structural boundaries of independent αTS folding units could be identified. In contrast to the 6+2 model, the urea-induced equilibrium unfolding study by Zitzewitz *et al.* demonstrates that $\alpha_0(\beta/\alpha)_{1\text{-}3}\beta_4$ is the smallest fragment able to fold autonomously. Moreover, the addition of $\alpha_4(\beta/\alpha)_5\beta_6$ to $\alpha_0(\beta/\alpha)_{1\text{-}3}\beta_4$ causes a remarkable increase in conformational stability and cooperativity that is not further improved upon by addition of the next two secondary structure elements, α_6 and β_7. Accordingly, β_4 and β_6 are structural boundaries for an autonomous αTS folding unit. Therefore, αTS folding is more likely to follow a "4+2+2" model. This alternative model is supported by

another αTS folding study that uses hydrogen exchange mass spectrometry as the probe.[17]

Here, we identified two ePRAI fragments that are capable of autonomous folding. The CD data indicate that the larger fragment $(\beta/\alpha)_{1\text{-}6}$ has a native-like secondary structure content, a compact tertiary structure, and unfolds cooperatively (Figures 2(a) and (b) and 3(a)). These structural properties are consistent with those of the corresponding fragment of yeast PRAI.[26] However, the far and near-UV CD spectra provide evidence that the smaller $(\beta/\alpha)_{1\text{-}4}\beta_5$ fragment also contains a significant amount of secondary structure and a native-like tertiary conformation (Figure 2(a) and (b)). The equilibrium unfolding curve shows the cooperative disruption of $(\beta/\alpha)_{1\text{-}4}\beta_5$ secondary structure by urea (Figures 3(a) and 4). These CD data indicate that the interaction of β_5 and β_6 is unrelated to the correct folding of $(\beta/\alpha)_{1\text{-}4}\beta_5$ although it causes the remarkable decrease in conformational stability. The results of our previous ePRAI unfolding simulations suggest that separation of β_6 and β_5 is unlikely to affect the subsequent unfolding of $(\beta/\alpha)_{1\text{-}4}\beta_5$.[24] Since removal of secondary structure elements from either the N or C terminus of $(\beta/\alpha)_{1\text{-}4}\beta_5$ seems to prevent folding altogether, $(\beta/\alpha)_{1\text{-}4}\beta_5$ is the smallest segment that is capable of autonomous folding. Covalent addition of $(\beta/\alpha)_6$ to $(\beta/\alpha)_{1\text{-}4}\beta_5$, yielding $(\beta/\alpha)_{1\text{-}6}$, causes a significant increase in conformational stability. These results suggest that there are two structural boundaries for ePRAI; one is located after the fifth β strand and the other after the sixth one. Consequently, the most plausible model for ePRAI folding is 5+1+2. It should be noted, however, that our fragmentation study does not provide sufficient data to exclude the 4+2+2 mechanism, as is found for αTS. The inability of $(\beta/\alpha)_{1\text{-}4}$ to fold may simply reflect that some interactions between β_4 and β_5 are crucial to the thermodynamic stability of $(\beta/\alpha)_{1\text{-}4}$ native structure. It is possible that $(\beta/\alpha)_{1\text{-}4}$ tends to form aggregates and to precipitate during refolding. An NMR experiment is one of the most powerful ways to elucidate the folding/unfolding mechanism of protein at atomic-level resolution. The NMR approach will provide us with additional information on ePRAI folding. Efforts are now in progress along this line.

Our fragmentation experiments suggest that structural characterization of $(\beta/\alpha)_{1\text{-}4}\beta_5$ is crucial if the ePRAI folding mechanism is to be understood. Additional insights into early ePRAI folding events were obtained by structurally characterizing circularly permuted variants and the denatured ensemble generated by unfolding simulations. Three permuted ePRAI variants β_6/β_5, β_7/α_6, and β_8/α_7 attained and preserved native-like features, and are monomeric. However, β_2/α_1 and β_3/α_2 did not fold to native-like tertiary structures (Figure 2(d)). These two variants partially aggregated during refolding (Table 1). Nevertheless, they contain elements of native-like secondary structure (Figure 2(c)). Based on αTS fragmentation studies,[9,18] Matthews and co-workers proposed that the kinetic pathway involves formation of βαβ super secondary structures during the early stages of folding, followed by assembly of these local folding modules. If this is also the case for ePRAI folding, cleavages at the N-terminal segment may prevent the assembly of the local N-terminal modules and lead to misfolding and aggregation.

Cleavages at the loops between α_3 and β_4, and between α_4 and β_5 caused the most striking effects on the folding of ePRAI. Expressed β_4/α_3 and β_5/α_4 circularly permuted variants could not be refolded, suggesting that destruction of an intact $(\beta/\alpha)_{3\text{-}4}\beta_5$ module completely prevents the protein from folding. Therefore, a natural arrangement of $(\beta/\alpha)_{3\text{-}4}\beta_5$ secondary structure elements is essential for proper ePRAI folding. Most likely, certain $(\beta/\alpha)_{3\text{-}4}\beta_5$ interactions nucleate ePRAI folding. However, as noted, $\alpha_1(\beta/\alpha)_{2\text{-}4}\beta_5$, $(\beta/\alpha)_{2\text{-}4}\beta_5$ and $(\beta/\alpha)_{3\text{-}4}\beta_5$ could not be refolded. Possibly, the native $(\beta/\alpha)_{3\text{-}4}\beta_5$ conformation is stabilized by tertiary interactions involving residues found in the first two βα units. In other words, the $(\beta/\alpha)_{3\text{-}4}\beta_5$ segment cannot retain its native structure without inter-(β/α) modular tertiary interactions. The urea-induced unfolding of $(\beta/\alpha)_{1\text{-}4}\beta_5$ is indeed a cooperative two-state transition. Cooperative unfolding involving several folding modules was observed also for barnase.[35]

A possible nucleation site is suggested by the results of our thermal unfolding simulations. There is evidence that unfolded proteins contain partially folded local structures that are potentially important for folding and stability.[28–30] These persistent structures are potential initiation sites and may play important roles during early stages of folding.[36] The structural properties of the members of the simulated denatured ensemble suggest that the three-stranded β-sheet-like structure consisting of β_3, β_4 and β_5 is highly stable and that the structural features of this region may endure in the unfolded state. This β-barrel intermediate persists the longest during unfolding simulations at temperatures of 550 K, 575 K and 600 K.[24] Accordingly, β_3, β_4, and β_5 may form unstable but distinct interactions during the early stages of folding, resulting in the formation of a folding nucleation site.

In conclusion, the ePRAI sequence can be divided into three portions with respect to folding (Figure 5). The C-terminal portion (shown in red) probably is not crucial to folding, since $(\beta/\alpha)_{1\text{-}4}\beta_5$ is stable enough to retain a native-like conformation and the far and near-UV CD spectra of the fragment are indicative of native secondary and tertiary structures. Our circular permutation experiments indicate that an intact N-terminal portion (shown in green) is required for the appearance of the global structure, but may not influence secondary structure formation. The middle portion (shown in blue) probably contains essential nucleation interactions, because when $(\beta/\alpha)_{3\text{-}4}\beta_5$ is cleaved, the protein cannot fold. Nevertheless, the possibility remains that the β_4/α_3 and β_5/α_4 circularly permuted variants do not fold simply because the cleaved

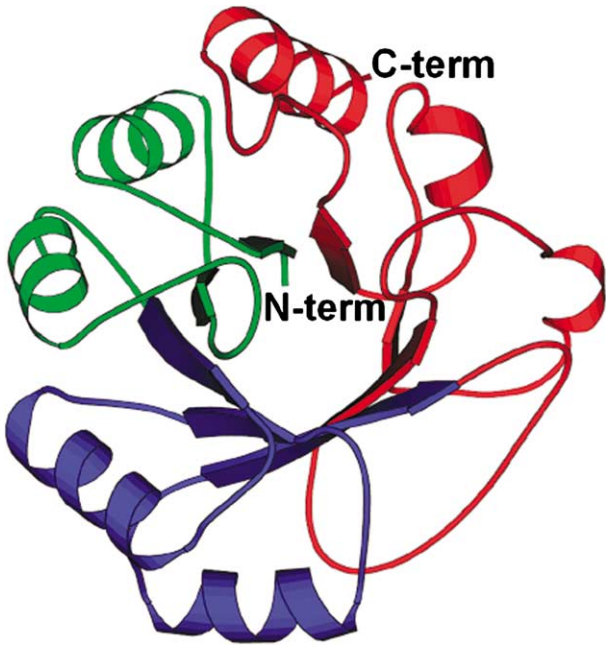

Figure 5. The ePRAI tertiary structure produced by MOLSCRIPT.[44] The segments corresponding to residues 1–50, 51–105 and 106–197 are shown in green, blue and red, respectively.

loops of these variants are crucial to the thermodynamic stability of the $(\beta/\alpha)_8$-barrel structure. However, the denatured ensemble produced by the thermal unfolding simulations contains persistent local structures involving β_3, β_4 and β_5 and this finding supports our preferred interpretation of the experiments. Thus, our *in vitro* and *in silico* studies indicate that $(\beta/\alpha)_{3\text{-}4}\beta_5$ contains interactions essential for the early events of ePRAI folding.

Materials and Methods

Construction of ePRAI fragments and circularly permuted variants

The *trpF* gene, encoding ePRAI, was cloned into pQE70 (Qiagen, Tokyo). The resulting plasmid was designated pQE70-*etrpF*. Partial sequences of the ePRAI gene were amplified using PCR, with pQE70-*etrpF* as the template. The circularly permuted ePRAI genes were also PCR amplified. All amplified DNA fragments were digested with SphI and HindIII and then cloned into the expression plasmid pQE70. The oligonucleotide primers used and the experimental details are given in Supplementary Table 1.

Protein preparation

For preparation of ePRAI, its fragments, and the circularly permuted variants, *E. coli* M15 (pREP4) (Qiagen, Tokyo) harboring the appropriate expression plasmid was cultivated in LB medium supplemented with ampicillin (150 μg ml^{-1}) and kanamycin (25 μg ml^{-1}). Overexpression was induced by adding IPTG to a final concentration of 0.25 mM when the culture reached an $A_{600\ nm}$ of 0.6–0.8. After cultivation for an additional 3 h at 37 °C, cells were harvested by centrifugation and disrupted by sonication. Since ePRAI and its derivatives collected in inclusion bodies, the soluble protein fraction was removed by centrifugation at 60,000*g* for 20 min. Insoluble protein was dissolved in 10 ml of 20 mM potassium phosphate buffer (pH 7.6) containing 6 M guanidine hydrochloride and the solution kept at room temperature for 30 min. An equal volume of 1 M guanidine hydrochloride, 20 mM potassium phosphate buffer (pH 7.6) was then added and the suspension remained at room temperature for 30 min. Then, the protein solution was centrifuged at 60,000*g* for 20 min and 30 ml of 2 M guanidine hydrochloride, 20 mM potassium phosphate buffer (pH 7.6) added to the supernatant. The solution was dialyzed twice against 5 l of 20 mM potassium phosphate buffer (pH 7.6). Precipitate was removed by centrifugation at 60,000*g* for 20 min and the soluble proteins purified using Resource Q (Amersham Biosciences, Piscataway) anion-exchange column chromatography. The purity of the proteins was >95% as judged by SDS-PAGE.

Analytical methods

Protein concentrations were determined by the method described by Gill and von Hippel[37] following the procedure described by Pace and colleagues.[38] For analytical gel-filtration, protein, in an initial volume of 0.2 ml, was eluted at a flow rate of 0.4 ml min^{-1} over Superdex 200 (1.0 cm×30 cm column dimensions; Amersham Biosciences, Piscataway) that was equilibrated with 20 mM potassium phosphate buffer (pH 7.6), 300 mM NaCl. Apparent molecular masses were determined using the protein elution volumes and a calibration curve produced using proteins of known molecular mass and elution volume. CD measurements were obtained using a J-720 spectropolarimeter (Jasco, Hachioji). The samples were dissolved in 20 mM potassium phosphate buffer (pH 7.6). All measurements were taken at 22 °C. Far-UV CD spectra were recorded from 185 nm to 250 nm. A 0.1 cm path-length cell was used. Protein concentrations were 0.1–0.2 mg ml^{-1}. Near-UV CD spectra were recorded from 250 nm to 320 nm. A 1.0 cm path-length cell was used. Protein concentrations were 1.0 mg ml^{-1}. To characterize their equilibrium unfolding properties, proteins were equilibrated with incremental amounts of urea in 20 mM potassium phosphate buffer (pH 7.6) at 22 °C. The samples were incubated at 22 °C for at least 18 h to ensure complete equilibration prior to data collection. Protein unfolding was estimated by monitoring the ellipticity at 220 nm and the values obtained using a 0.1 cm path-length cell, protein concentrations of 0.1 mg ml^{-1}, and a temperature of 22 °C. Equations for two and three-state equilibrium unfolding models have been described.[39,40] A non-linear least-squares fitting program found in SigmaPlot (Systat Software, Richmond) was used to estimate the free energy differences, ΔG°, ΔG°_1, and ΔG°_2, as well as the cooperativity indices, m, m_1, and m_2. Initial values for these parameters were chosen arbitrarily, and the regressions were reiterated until the R^2 values were minimized.

MD simulations

Thermal unfolding MD simulations were performed at 500 K or 525 K using the AMBER6.0 simulation package[41]

and the Parm96 force field[42] with the TIP3P water model.[43] The first 7 ns of the thermal unfolding simulations at 500 K and 525 K, as well as that of the control at 300 K, were calculated previously.[24] The simulation parameters were described therein. To generate an ensemble of unfolded structures, unfolding simulations were extended to 20 ns and 12 ns at 500 K and 525 K, respectively. The C^{α} root-mean-square deviation (RMSD) and the radius of gyration (R_G) were calculated as described.[24] To estimate the number of native hydrogen bonds in the simulation structures, hydrogen bonds in the native conformation were first identified using the Hbond module of the Insight II ver. 97 Molecular Modeling System (Accelrys, San Diego). A native hydrogen bond was defined as one that is present more than 70% of the time during the control simulation.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.08.070

References

1. Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, **426**, 884–890.
2. Dobson, C. M. (2000). The nature and significance of protein folding. In *Mechanisms of Protein Folding* (Pain, R. H., ed.), *Frontiers in Molecular Biology Series*, 2nd edit., pp. 1–33, Oxford University Press, Oxford.
3. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
4. Fersht, A. R. (2000). A kinetically significant intermediate in the folding of barnase. *Proc. Natl Acad. Sci. USA*, **97**, 14121–14126.
5. Dobson, C. M., Sali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37**, 868–893.
6. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
7. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. (2000). Understanding protein folding *via* free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339.
8. Fersht, A. R. (1999). *Structure and Mechanism in Protein Science, A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman, New York.
9. Zitzewitz, J. A., Gualfetti, P. J., Perkons, I. A., Wasta, S. A. & Matthews, C. R. (1999). Identifying the structural boundaries of independent folding domains in the alpha subunit of tryptophan synthase, a β/α barrel protein. *Protein Sci.* **8**, 1200–1209.
10. Ruiz Sanz, J., de Prat Gay, G., Otzen, D. E. & Fersht, A. R. (1995). Protein fragments as models for events in protein folding pathways: protein engineering analysis of the association of two complementary fragments of the barley chymotrypsin inhibitor 2 (CI-2). *Biochemistry*, **34**, 1695–1701.
11. Goldenberg, D. P. & Creighton, T. E. (1983). Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **165**, 407–413.
12. Hennecke, J., Sebbel, P. & Glockshuber, R. (1999). Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J. Mol. Biol.* **286**, 1197–1215.
13. Iwakura, M., Nakamura, T., Yamane, C. & Maki, K. (2000). Systematic circular permutation of an entire protein reveals essential folding elements. *Nature Struct. Biol.* **7**, 580–585.
14. Hegyi, H. & Gerstein, M. (1999). The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* **288**, 147–164.
15. Wilmanns, M., Priestle, J. P., Niermann, T. & Jansonius, J. N. (1992). Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *J. Mol. Biol.* **223**, 477–507.
16. Eberhard, M., Tsai Pflugfelder, M., Bolewska, K., Hommel, U. & Kirschner, K. (1995). Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry*, **34**, 5419–5428.
17. Rojsajjakul, T., Wintrode, P., Vadrevu, R., Matthews, C. R. & Smith, D. L. (2004). Multi-state unfolding of the alpha subunit of tryptophan synthase, a TIM barrel protein: insights into the secondary structure of the stable equilibrium intermediates by hydrogen exchange mass spectrometry. *J. Mol. Biol.* **341**, 241–253.
18. Zitzewitz, J. A. & Matthews, C. R. (1999). Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein. *Biochemistry*, **38**, 10205–10214.
19. Bilsel, O., Zitzewitz, J. A., Bowers, K. E. & Matthews, C. R. (1999). Folding mechanism of the α-subunit of tryptophan synthase, an α/β barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry*, **38**, 1018–1029.
20. Bilsel, O., Yang, L., Zitzewitz, J. A., Beechem, J. M. & Matthews, C. R. (1999). Time-resolved fluorescence anisotropy study of the refolding reaction of the α-subunit of tryptophan synthase reveals nonmonotonic behavior of the rotational correlation time. *Biochemistry*, **38**, 4177–4187.
21. Jasanoff, A., Davis, B. & Fersht, A. R. (1994). Detection of an intermediate in the folding of the $(\beta\alpha)_8$-barrel *N*-(5′-phosphoribosyl)anthranilate isomerase from *Escherichia coli*. *Biochemistry*, **33**, 6350–6355.
22. Luger, K., Hommel, U., Herold, M., Hofsteenge, J. & Kirschner, K. (1989). Correct folding of circularly permuted variants of a βα barrel enzyme *in vivo*. *Science*, **243**, 206–210.
23. Luger, K., Szadkowski, H. & Kirschner, K. (1990). An 8-fold βα barrel protein with redundant folding possibilities. *Protein Eng.* **3**, 249–258.
24. Akanuma, S., Miyagawa, H., Kitamura, K. & Yamagishi, A. (2004). A detailed unfolding pathway of a $(\beta/\alpha)_8$-barrel protein as studied by molecular dynamics simulations. *Proteins: Struct. Funct. Genet.* **58**, 538–546.
25. Höcker, B., Beismann, S. D., Hettwer, S., Lustig, A. & Sterner, R. (2001). Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nature Struct. Biol.* **8**, 32–36.
26. Eder, J. & Kirschner, K. (1992). Stable substructures of eightfold βα-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry*, **31**, 3617–3625.
27. Jurgens, C., Strom, A., Wegener, D., Hettwer, S.,

Wilmanns, M. & Sterner, R. (2000). Directed evolution of a $(\beta\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl Acad. Sci. USA*, **97**, 9925–9930.
28. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
29. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R. & Daggett, V. (2000). Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* **296**, 1257–1282.
30. Bond, C. J., Wong, K. B., Clarke, J., Fersht, A. R. & Daggett, V. (1997). Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc. Natl Acad. Sci. USA*, **94**, 13409–13413.
31. Fersht, A. R. & Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell*, **108**, 573–582.
32. Shea, J. E. & Brooks, C. L., 3rd (2001). From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499–535.
33. Miles, E. W., Yutani, K. & Ogasahara, K. (1982). Guanidine hydrochloride induced unfolding of the α subunit of tryptophan synthase and of the two α proteolytic fragments: evidence for stepwise unfolding of the two α domains. *Biochemistry*, **21**, 2586–2592.
34. Higgins, W., Fairwell, T. & Miles, E. W. (1979). An active proteolytic derivative of the alpha subunit of tryptophan synthase. Identification of the site of cleavage and characterization of the fragments. *Biochemistry*, **18**, 4827–4835.
35. Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3–9.
36. Dyson, H. J. & Wright, P. E. (2002). Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Advan. Protein Chem.* **62**, 311–340.
37. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
38. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423.
39. Sun, T. X., Akhtar, N. J. & Liang, J. J. (1999). Thermodynamic stability of human lens recombinant αA- and αB-crystallins. *J. Biol. Chem.* **274**, 34067–34071.
40. Motono, C., Yamagishi, A. & Oshima, T. (1999). Urea-induced unfolding and conformational stability of 3-isopropylmalate dehydrogenase from the thermophile *Thermus thermophilus* and its mesophilic counterpart from *Escherichia coli*. *Biochemistry*, **38**, 1332–1337.
41. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., Ross, W. S., Simmerling, C. L. *et al*. (1999). *AMBER 6.0*, University of California, San Francisco, CA.
42. Kollman, P., Dixon, R. W., Cornell, W. D., Chipot, C. & Pohorille, A. (1997). The development/application of "minimalist" organic/biochemical molecular mechanic force field using a combination of *ab initio* calculations and experimental data. In *Computer Simulations of Biological Systems* (van Gunsteren, W. F., ed.), vol. 3, pp. 83–86, Kluwer, Dordrecht.
43. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
44. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Edited by C. R. Matthews

(Received 25 March 2005; received in revised form 30 July 2005; accepted 29 August 2005)
Available online 22 September 2005