doi:10.1016/S0022-2836(02)01100-2 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 324, 1031–1040

JMB

The equilibrium unfolding pathway of a $(\beta/\alpha)_8$ barrel

Joshua A. Silverman and Pehr B. Harbury*

Department of Biochemistry Stanford University, 279 Campus Drive West, Stanford CA 94305, USA

*Corresponding author

The $(\beta/\alpha)_8$ barrel is the most commonly occurring fold among enzymes. A key step towards rationally engineering $(\beta/\alpha)_8$ barrel proteins is to understand their underlying structural organization and folding energetics. Using misincorporation proton–alkyl exchange (MPAX), a new tool for solution structural studies of large proteins, we have performed a native-state exchange analysis of the prototypical $(\beta/\alpha)_8$ barrel triosephosphate isomerase. Three cooperatively unfolding subdomains within the structure are identified, as well as two partially unfolded forms of the protein. The C-terminal domain coincides with domains reported to exist in four other $(\beta/\alpha)_8$ barrels, but the two N-terminal domains have not been observed previously. These partially unfolded forms may represent sequential intermediates on the folding pathway of triosephosphate isomerase. The methods reported here should be applicable to a variety of other biological problems involving protein conformational changes.

Keywords: native-state exchange; triosephosphate isomerase; β/α barrel; MPAX; protein folding

Introduction

A number of distinct folding intermediates and pathways have been proposed for different $(\beta/\alpha)_8$ barrels,[1–6] but experimental data are limited. One experimentally tractable way to test the various proposals is to characterize barrel unfolding pathways. Native-state exchange experiments have proven useful for this purpose in studies of smaller proteins.[7,8] Under conditions that energetically favor the native state, a folded protein will occasionally populate partially unfolded conformations. Native-state exchange can be used to detect these partially unfolded forms by providing measurements of local stability at sites distributed throughout a protein structure. If the protein consists of a single cooperatively unfolding unit, the native and globally unfolded states will predominate, and the apparent stability at all sites will be identical. A more common observation is that different regions of the protein display different apparent stabilities, which occurs when portions of the structure locally unfold to produce intermediates more stable than the globally unfolded state. Regions of the protein structure that exhibit identical stability usually comprise cooperatively unfolding subdomains. In many cases, partially unfolded states generated by subdomain unfolding at equilibrium have been shown to correspond to intermediates on the kinetic folding pathway.[9]

Native-state exchange experiments utilize a reactive group protected by the protein structure as a probe of protein stability. Buried probes only react when exposed to solvent. The observed rate of reaction for each probe reflects the fraction of time that it is solvent-accessible, which is determined by the local stability of the protein. A buried probe can be exposed to solvent by at least two distinct mechanisms: large-scale protein unfolding and local structural fluctuations (LSFs).[10] The stability of cooperatively unfolding protein domains is linearly dependent on the denaturant concentration, with a slope (*m*-value) proportional to the amount of surface area exposed upon unfolding.[11] Cooperatively unfolding subdomains can therefore be distinguished by the clustering of *m*-values for sets of probes into distinct isotherms. In contrast, LSFs do not expose a significant amount of denaturant-binding surface area, and the extent of probe exposure by this mechanism is relatively independent of denaturant concentration. One mechanism for solvent exposure will dominate the rate of modification observed for a given probe.

Present address: J. A. Silverman, Maxygen Inc., 220 Penobscot Drive, Redwood City, CA 94063, USA.

Abbreviations used: GdmCl, guanidinium chloride; LSF, local structural fluctuation; MPAX, misincorporation proton–alkyl exchange; TIM, triosephosphate isomerase.

E-mail address of the corresponding author: harbury@cmgm.stanford.edu

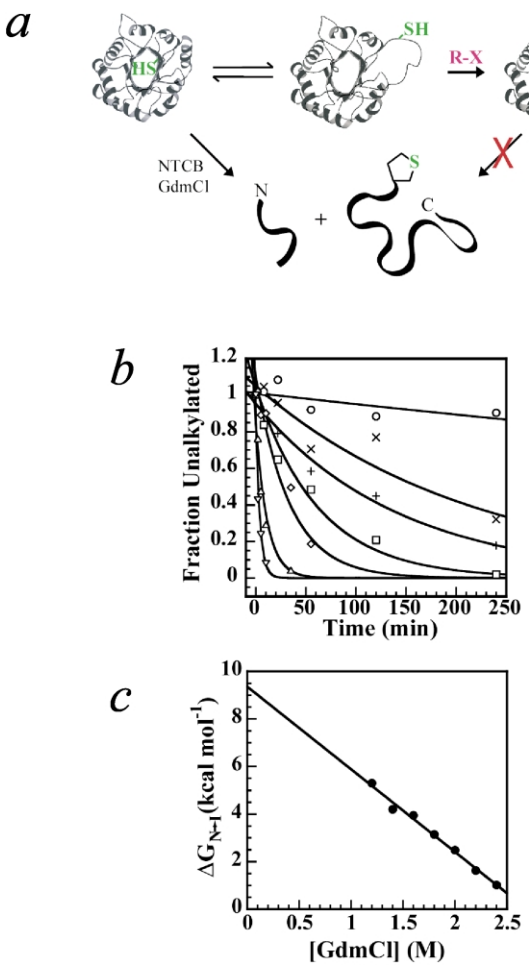

Figure 1. The native-state MPAX experiment. (a) A misincorporated cysteine residue buried in the protein structure is protected from solvent and is unreactive in the native state. The native state is in equilibrium with partially unfolded states that expose the misincorporated cysteine residue to solvent. When exposed to solvent, the cysteine sulfhydryl can react with an alkylating reagent (RX) to generate an irreversibly modified species. This species is resistant to cleavage of the protein backbone at cysteine residues by 2-nitro-5-thiocyanobenzoic acid (NTCB), while the unmodified protein is susceptible to cleavage. (b) Plot showing the rate of alkylation of C126 in 1.2 M (○), 1.4 M (×), 1.6 M (+), 1.8 M (□), 2.0 M (◇), 2.2 M (△), or 2.4 M (▽) guanidinium chloride (GdmCl). Continuous lines are fits to a single exponential process. (c) Plot of the local protein stability ($\Delta G_{N\rightarrow I}$) at C126 as a function of GdmCl concentration (calculated from the alkylation rates in (b) according to equation (1)). The continuous line is a fit to equation (2), extrapolated to zero denaturant concentration.

As the denaturant concentration increases, the rate of modification due to unfolding will be selectively accelerated and will eventually become faster than the rate of modification due to LSFs. Thus, the observed modification rate as a function of denaturant concentration can exhibit two exchange regimes: the first regime at low denaturant concentration representing LSFs, and the second regime at higher denaturant concentration representing protein unfolding.

The most common implementation of native-state exchange uses the exchange of hydrogen atoms for deuterium atoms at peptide bonds, which can be read out site-specifically by nuclear magnetic resonance (NMR). Unfortunately, high-resolution hydrogen exchange data have only been obtained for relatively small proteins due to the intrinsic limitations of NMR techniques. Mass-spectrometry approaches have been developed to detect hydrogen/deuterium exchange in large proteins, but they do not provide data at single amino acid residue resolution.[12] As demonstrated by Loh and co-workers, modification of cysteine residues represents an alternative approach for performing native-state exchange experiments.[13,14] The cysteine based native-state exchange scheme is shown in Figure 1. A cysteine probe buried in the protein structure reacts with a thiol-specific electrophile in solution only when it becomes exposed to solvent in an unfolded state. If refolding from the unfolded state is much faster than chemical modification of the cysteine probe (the EX2 limit), the observed rate of modification is directly proportional to the equilibrium constant between the native and unfolded protein states.[15] Thus, the equilibrium constant for local unfolding in the vicinity of each cysteine probe can be computed from the measured rate of alkylation (see Methods, equation (1)).

We have performed a comprehensive native-state exchange analysis of triosephosphate isomerase (TIM), a 247 amino acid $(\beta/\alpha)_8$ barrel protein that forms a dimer in solution. Our experiment is based on misincorporation proton–alkyl exchange (MPAX).[16] MPAX utilizes translational misincorporation of cysteine residues to generate an ensemble of proteins with single cysteine residue substitutions. The substitutions replace one amino acid residue type, and are distributed throughout the structure. The protein is incubated with the thiol-modifying reagent iodoacetamide in the presence of varying concentrations of denaturant, followed by cleavage of the protein backbone with 2-nitro-5-thiocyanobenzoic acid (NTCB) and separation of the cleavage products by gel electrophoresis. Alkylated cysteine residues are resistant

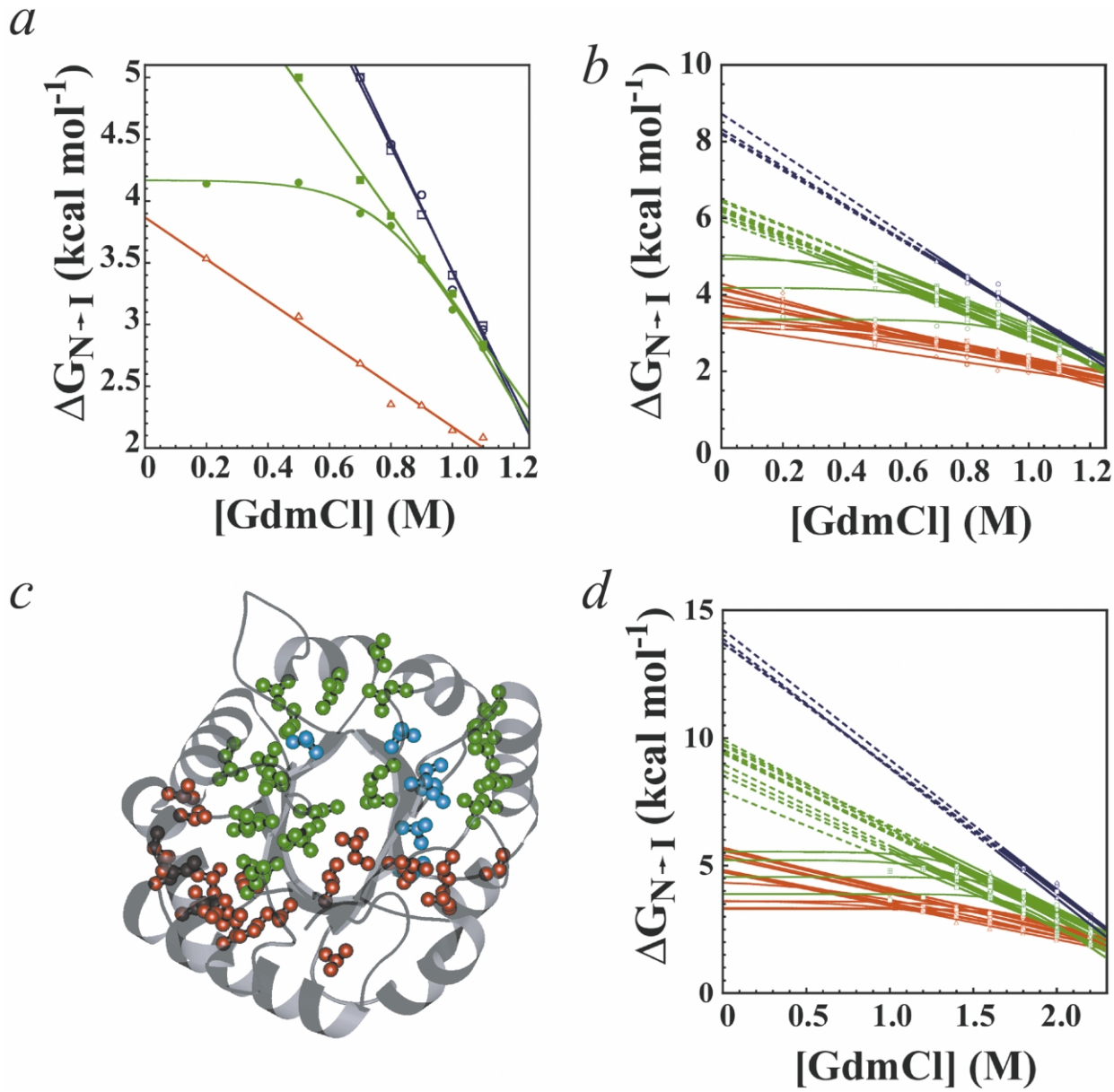

Figure 2. The local stabilities of probes in TIM fall into three classes. (a) Plot of representative data showing local protein stability ($\Delta G_{N\rightarrow I}$) as a function of GdmCl concentration for cysteine residues misincorporated at positions V38 (□), V61 (○), L117 (●), V80 (■), and V226 (△) in the C41V/C126A TIM protein. The continuous lines are fits of the data to equation (2) and are colored according to their *m*-value: red, $m < 2.5$; green, $2.5 < m < 4.5$; blue, $m > 4.5$. (b) Plot of $\Delta G_{N\rightarrow I}$ for cysteine residues misincorporated at isoleucine, leucine, and valine positions in the C41V/C126A TIM background. The continuous lines are fits of the data to equation (2), whereas broken lines are extrapolations. (c) Residues in the crystal structure of TIM are colored according to their *m*-value. (d) Plot of $\Delta G_{N\rightarrow I}$ for misincorporated cysteine residues in the C41V TIM background.

to cleavage by NTCB (Figure 1(a)). Thus, the loss of cleavage products over time reflects the rate of cysteine modification. Because the location of each misincorporated cysteine residue is known, data are obtained at single amino acid residue resolution. MPAX is not limited by protein size and can be applied to proteins of limited solubility, so it is ideal for the study of $(\beta/\alpha)_8$ barrels. Here we present a comprehensive analysis of the unfolding energy landscape of TIM by MPAX, identifying three cooperatively unfolding domains within the protein structure and two equilibrium unfolding intermediates.

Results

Native-state MPAX reveals three subdomains in TIM

To confirm that we could measure local protein stability using MPAX, we first measured the rate

Table 1. Thermodynamic parameters for misincorporated cysteine probes

			C41V/C126A			C41V		
Residue	Isotherm		$\Delta G^0_{N\rightarrow I}$ (kcal mol^{-1})	m (kcal mol^{-1} M^{-1})	ΔG_{LSF} (kcal mol^{-1})	$\Delta G^0_{N\rightarrow I}$ (kcal mol^{-1})	m (kcal mol^{-1} M^{-1})	ΔG_{LSF} (kcal mol^{-1})
V 7	R		3.93	1.55		5.33	1.47	
I 20	G		6.21	3.30		9.17	3.11	
I 23	G		6.21	3.30		9.17	3.11	
V 24	G		7.06	4.13	5.12	10.06	3.47	5.55
L 27	G		6.88	3.95		9.10	3.31	
V 36		B	8.69	5.27		13.76	4.88	
V 38		B	8.69	5.27		13.76	4.88	
I 40		B	8.34	4.84		13.94	5.12	
L 47	G		6.52	3.64		9.32	3.59	
L 53	G		6.52	3.64		9.32	3.59	
V 59		B	8.47	5.03		13.77	4.93	
V 61		B	8.47	5.03		13.77	4.93	
V 80	G		6.32	3.26		9.07	3.23	
I 83	G		6.50	3.29		9.12	3.28	
V 86	G		6.04	3.17	3.36	9.04	3.17	3.89
V 91		B	8.35	4.91		14.22	5.08	
I 92	G		6.78	3.37		9.26	3.25	
L 93	G		6.50	3.61		8.82	3.15	
I 109	G		6.93	3.46	4.95	10.04	3.55	5.21
L 117	G		6.59	3.67	4.18	9.25	3.35	4.56
V 121	G		6.10	3.05		n.d.	n.d.	
V 123	G		6.10	3.05		n.d.	n.d.	
I 124	G		5.99	3.18		n.d.	n.d.	
L 125	G		6.26	3.39		n.d.	n.d.	
C 126	G					9.92	3.47	
I 127	G		5.99	3.18		n.d.	n.d.	
V 142	R		4.13	1.84		5.58	1.49	
V 143	R		4.13	1.84		5.58	1.49	
L 147	R		4.14	1.86		5.64	1.63	4.43
V 150	R		3.71	1.47		5.65	1.55	
L 151	R		4.14	1.86		5.64	1.63	
V 154	R		3.73	1.46	3.30	5.31	1.42	3.19
V 160	G		6.82	3.69		9.78	3.28	
V 161	G		6.82	3.69		9.78	3.28	
V 162	G		6.82	3.69		9.78	3.28	
I 188	R		3.76	1.57	3.34	5.22	1.49	3.56
L 192	R		3.91	1.55		4.78	1.20	
L 196	R		3.91	1.55		4.78	1.20	
L 204	R		3.95	1.71		5.01	1.29	
I 206	R		3.33	1.20		4.76	1.27	
L 207	R		3.95	1.71		5.01	1.29	
V 226	R		3.60	1.41		5.69	1.55	3.61
L 230	R		3.05	1.50		4.18	1.03	
V 231	R		3.94	1.71		5.46	1.47	
V 241	R		3.89	1.68	3.65	5.49	1.48	4.09
I 243	R		3.02	1.02		4.83	1.29	
I 244	R		3.02	1.02		4.83	1.29	
Average	Red		3.7 ± 0.4	1.6 ± 0.3		5.2 ± 0.4	1.4 ± 0.2	
	Green		6.5 ± 0.3	3.5 ± 0.3		9.4 ± 0.4	3.3 ± 0.2	
		Blue	8.5 ± 0.2	5.1 ± 0.2		13.9 ± 0.2	5.0 ± 0.1	

Data are shown for probes in the C41V/C126A or C41V TIM variants. The isotherm to which each probe belongs is indicated in the second column: R, red; G, green; B, blue. Average values and standard deviations of $\Delta G^0_{N\rightarrow I}$ and *m*-value for probes in each isotherm are reported at the bottom of the table; n.d., not determined.

of iodoacetamide alkylation at C126, a naturally occurring cysteine residue in TIM, as a function of guanidinium chloride concentration ([GdmCl]) (Figure 1(b)). The plot of $\Delta G_{N\rightarrow I}$ *versus* [GdmCl] is linear. When extrapolated to zero denaturant concentration, the *y*-intercept corresponds to the free energy difference between the native state and the partially or fully unfolded state that exposes C126 to solvent ($\Delta G^0_{N\rightarrow I}$; Figure 1(c)). An exchange regime due to LSFs was not observed at this position, although data were only collected at GdmCl concentrations above 1 M.

To identify cooperatively unfolding domains within the TIM structure, the alkylation rate of 61 cysteine probes misincorporated at isoleucine, leucine, and valine residue positions was determined as a function of GdmCl concentration. Forty-six probes showed significant protection due to protein structure, in accordance with their level of solvent exposure in the crystal structure.[16] A

Table 2. Mutations in the helices of TIM establish a hierarchical relationship between I1, I2, and I3

Helix			I1	I2	I3
1	E22	$\Delta G^0_{N\rightarrow I}$ (Ala)	3.6	7.3	9.1
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.04	0.58	0.91
		$\Phi_{N\rightarrow I}$	0.04	0.64	1.01
2	V51	$\Delta G^0_{N\rightarrow I}$ (Ala)	4.2	7.0	9.3
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.07	0.06	1.00
		$\Phi_{N\rightarrow I}$	0.08	0.07	1.11
3	D81	$\Delta G^0_{N\rightarrow I}$ (Ala)	3.9	6.8	8.9
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.02	0.22	0.87
		$\Phi_{N\rightarrow I}$	0.02	0.24	0.97
4	F115	$\Delta G^0_{N\rightarrow I}$ (Ala)	3.4	4.7	6.6
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.08	0.93	0.84
		$\Phi_{N\rightarrow I}$	0.09	1.03	0.93
5	E153	$\Delta G^0_{N\rightarrow I}$ (Ala)	3.7	6.1	8.4
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.26	0.86	0.87
		$\Phi_{N\rightarrow I}$	0.29	0.96	0.97
6	A186	$\Delta G^0_{N\rightarrow I}$ (Ala)	3.8	6.5	8.5
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.79	0.88	0.93
		$\Phi_{N\rightarrow I}$	0.88	0.98	1.03
7	V218	$\Delta G^0_{N\rightarrow I}$ (Ala)	4.4	7.5	9.3
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.82	0.90	0.75
		$\Phi_{N\rightarrow I}$	0.91	1.00	0.83
8	D242	$\Delta G^0_{N\rightarrow I}$ (Ala)	4.6	7.6	9.2
		$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$	0.93	0.84	0.98
		$\Phi_{N\rightarrow I}$	1.03	0.93	1.09

Sites of Gly → Ala mutations in each helix are indicated. $\Phi_{N\rightarrow I}$ values report the fractional unfolding of the specified helix in I1–3. A $\Phi_{N\rightarrow I}$ value of zero indicates no unfolding, whereas a $\Phi_{N\rightarrow I}$ value of one indicates complete unfolding. $\Delta G^0_{N\rightarrow I}$ (Ala) reports the extrapolated stability of each subdomain in the alanine mutant. $\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$ and $\Phi_{N\rightarrow I}$ are calculated according to equations (3) and (4). Free energies are in units of kcal mol^{-1}.

two-regime exchange model incorporating both LSFs and subdomain unfolding was used to fit the observed alkylation rates of all cysteine probes (equation (2); Figure 2(a)). The *m*-values and extrapolated stabilities of the misincorporated cysteine residues cluster into three isotherms (Figure 2(b) and (d)). The values within each cluster are separated by more than two standard deviations from members of other clusters, confirming that the groups are distinct (Table 1). The isotherms are colored red, green, or blue in order of increasing $\Delta G^0_{N\rightarrow I}$ and *m*-value. The locations of misincorporated cysteine residues in each isotherm also cluster within the protein structure (Figure 2(c)). The three regions of structure are referred to as the red, green, or blue subdomains.

The presence of three isotherms indicates that there are at least three intermediates in the equilibrium unfolding pathway of TIM. We refer to the three states as I1, I2, and I3 in order of increasing $\Delta G^0_{N\rightarrow I}$. Since the native state of TIM is a dimer, a first step in characterizing the structures of these intermediates is to determine whether they are dimers or monomers. If an intermediate, I*, were a dimer, increasing protein concentration would alter the stability of the native and I* states equally, and no change in $\Delta G_{N\rightarrow I^*}$ would be observed. If I* were a monomer, however, the native state would be stabilized relative to I* by an increase in protein concentration, resulting in an apparent increase in $\Delta G_{N\rightarrow I^*}$. Alkylation rates were measured at TIM concentrations of 1 μM and 5 μM. $\Delta G_{N\rightarrow I}$ for all intermediates was found to increase by 0.8 ± 0.2 kcal mol^{-1} with a fivefold increase in protein concentration, indicating that all three intermediates are monomers.

A hierarchical relationship between I1, I2 and I3

The three intermediates might be arranged in a sequential pathway, three parallel pathways, or some combination of the two. In addition, one of the intermediates may or may not represent global protein unfolding. To distinguish between these possibilities, we determined the extent to which each alpha helix in TIM is unfolded in the I1, I2, and I3 intermediates. The stability of individual alpha helices was perturbed by mutating a solvent exposed residue to both glycine and alanine. The difference in helical propensity between glycine and alanine residues results in a 0.8–1.0 kcal mol^{-1} difference in helix stability.[17–20] If a helix remained folded in an intermediate I*, a glycine to alanine substitution would have similar effects on N and I*, and no change in $\Delta G_{N\rightarrow I^*}$ would be observed. Conversely, if the helix were unfolded in I*, the glycine to alanine substitution would stabilize N relative to I*, resulting in an increase of ~0.9 kcal mol^{-1} in $\Delta G_{N\rightarrow I^*}$ (the helix stability change expected for glycine to alanine mutations). Phi values quantifying the fractional unfolding of each helix in each of the intermediates can be computed as the ratio of $\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I^*}$ to 0.9 kcal mol^{-1} (equation (4), see Methods). Here, $\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I^*}$ is the observed difference in $\Delta G_{N\rightarrow I^*}$ between the alanine and glycine mutants. The phi value should equal zero if a helix is fully folded in an intermediate, and should equal one if the helix is completely unfolded. The experimentally measured phi values for mutations in each helix are shown in Table 2. $\Delta G_{N\rightarrow I}$ and *m*-values for all of the probes in each isotherm were perturbed identically by the helix substitutions, and average values are reported. Helices 6–8 are unfolded in I1, I2, and I3. In contrast, helices 1, 4, and 5 are folded in I1, but unfolded in both I2 and I3. Finally, helices 2 and 3 are folded in I1 and I2, and only unfold in I3. The analysis shows that when the blue domain unfolds, the red and green domains are already unfolded. Thus, I3 must represent global unfolding of the protein. Furthermore, the blue domain remains folded when the green domain unfolds, but the red domain has already unfolded. Finally, the data show that the red domain can unfold independently of the green and blue domains. The data support a model in which TIM unfolds along a sequential pathway: N → I1 → I2 → U.

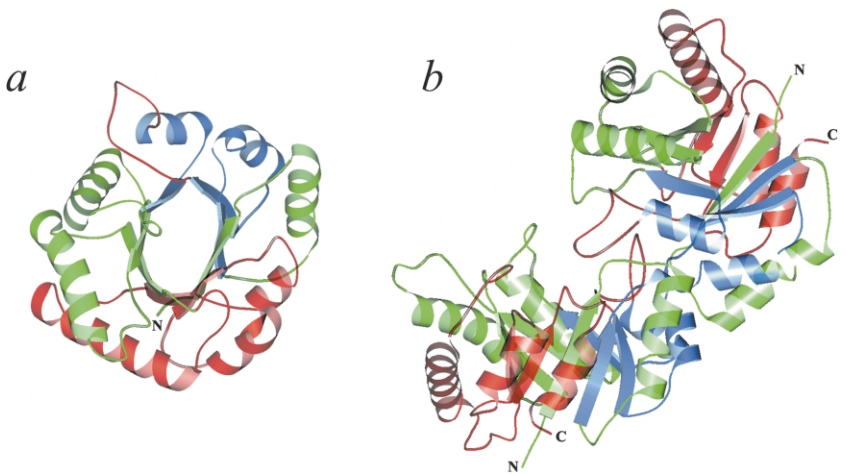

Figure 3. The subdomain structure of TIM. The three subdomains of TIM are colored red, green, and blue. (a) A single monomer, oriented with the active site facing the viewer. The red loop at the top forms the dimerization interface. (b) Side view of the TIM dimer. The active sites are located at the top of the lower monomer, and at the left side of the upper monomer. The N and C termini are indicated. Figures were generated using the 7tim pdb structure and MOLSCRIPT.[32]

Cooperatively unfolding subdomains

A remaining question is whether all of the probes in each isotherm become solvent exposed through one all-or-nothing subdomain unfolding event, or whether different probes in a given isotherm become exposed through distinct and non-cooperative unfolding processes. Three lines of evidence argue in favor of the all-or-nothing model. First, the exchange processes for all of the probes in a particular isotherm exhibit the same exchange free energy. Second, the exchange processes for all probes in each isotherm exhibit the same dependence on denaturant concentration (*m*-value). This fact implies that all of the unfolding events that lead to exchange of probes in a given isotherm must expose identical amounts of buried surface area. Finally, the exchange free energies for all of the probes in each isotherm are perturbed identically by Gly → Ala mutations in each of the eight helices of TIM. Thus all of the exchange processes for probes in a given isotherm must induce the same pattern of helix unfolding. These observations argue strongly that the three isotherms correspond to cooperative unfolding of subdomains within TIM, and that all of the probes in each isotherm exchange through the same partially unfolded state.

Discussion

We have used the MPAX method to map the unfolding energy landscape of triosephosphate isomerase. This is the first comprehensive analysis of $(\beta/\alpha)_8$ barrel unfolding at single amino acid resolution, and represents one of the largest proteins for which the unfolding energy landscape has been mapped. We have identified three distinct cooperatively unfolding units within the protein structure, and two intermediates presumably on the pathway between the folded and unfolded states.

One potential drawback of using the MPAX approach to measure unfolding energies is that mutation of a buried amino acid to cysteine could perturb the stability of the protein. We have previously presented evidence that this effect is small for substitution of isoleucine, leucine and valine positions in TIM.[16] We can now quantitate the energetic perturbation resulting from many single cysteine residue mutations. The alkylation rate of C126 provides a direct measure of the stability of I2 in the absence of misincorporated cysteine residues. The stability of I2 measured by probes in the green domain of the C41V protein differs from the wild-type value by less than 1.1 kcal mol^{-1} in all cases, with an average destabilization of 0.5 kcal mol^{-1}. The spread of the data within the other domains in the C126 and C126A backgrounds is consistent with these numbers, indicating that mutation of isoleucine, leucine, and valine residues to cysteine generally results in small energetic effects.

The apparent 5.4 kcal mol^{-1} destabilization of TIM caused by mutation of the native cysteine C126 to alanine differs from the small effects observed for mutation of hydrophobic amino acid to cysteine. Alanine, valine and serine substitutions at position 126 all resulted in a similar destabilization (data not shown), indicating that cysteine is highly favored at position 126. By contrast, mutation of the other conserved cysteine residue in TIM, C41, to valine destabilizes the protein by only 0.3 kcal mol^{-1}. While the general MPAX strategy of replacing hydrophobic residues with cysteine residues does not appear to perturb significantly the protein stability, the effect of removing conserved cysteine residues may be large in some cases.

One limitation of our experiment is that we only measure values of $\Delta G_{N\rightarrow I}$ that lie between 0 and 5 kcal mol^{-1}. Consequently, we have not characterized the behavior of most cysteine probes at low denaturant concentrations. Many of the probes would likely exhibit a second exchange regime due to LSFs under these conditions. The limitation on our $\Delta G_{N\rightarrow I}$ measurements arises from the fact that alkylation of cysteine thiol by iodoacetamide is slow: the pseudo-first order rate constant at

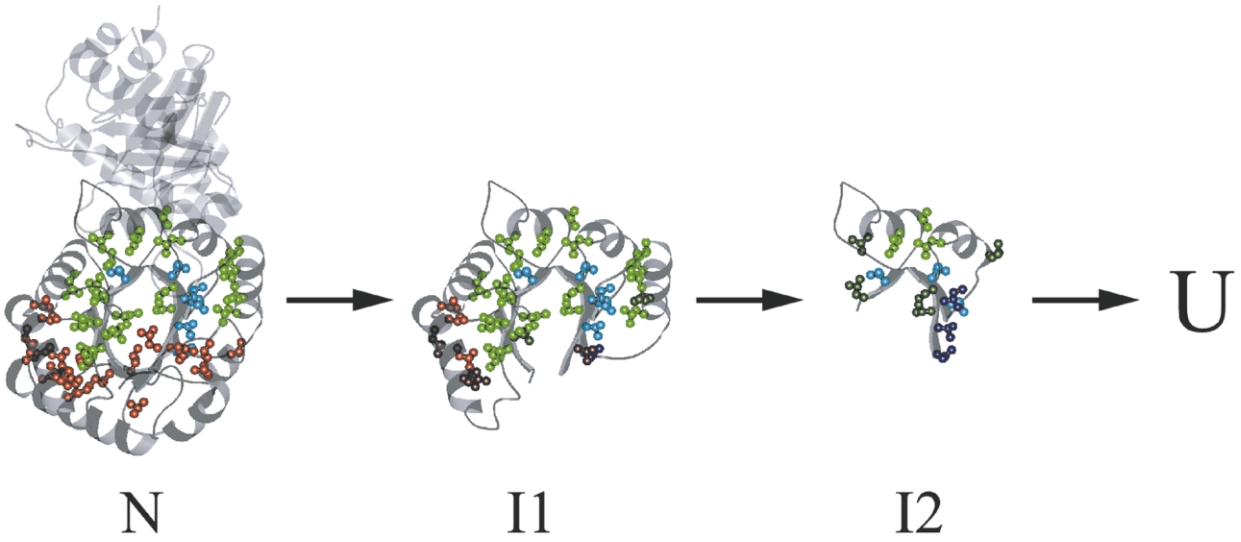

Figure 4. A sequential unfolding pathway for TIM. Models for the structures of I1 and I2 are shown. The native dimer (N) and the unfolded state (U) are also illustrated. Residue positions are colored according to their *m*-value: red, $m < 2.5$; green, $2.5 < m < 4.5$; blue, $m > 4.5$. Positions calculated to be more than 30% exposed to solvent in an intermediate structure are shaded a darker version of their original color.

10 mM iodoacetamide is 0.015 s^{-1}. In a four hour experiment, the largest protection factor that we can measure accurately is 4000, which corresponds to a $\Delta G_{N\rightarrow I}$ value of ~5.0 kcal mol^{-1}. Carrying out alkylation reactions for longer periods of time could expand the dynamic range for $\Delta G_{N\rightarrow I}$ measurements. However, a more expedient strategy would be to utilize a faster cysteine modification reaction. For example, Loh and co-workers have pioneered the use of methyl methanethiosulfonate (MMTS) for cysteine-based native-state exchange measurements.[14] The intrinsic rate for cysteine residue modification by MMTS is 10^5-fold faster than that for modification by iodoacetamide. If modification reactions by MMTS were to conform to the EX2 limit, then protection factors of 4×10^8 could be measured in a four hour experiment. A protection factor of that magnitude would correspond to an exchange free energy of ~12 kcal mol^{-1}.

What are the structures of the intermediates?

The fact that the stability of all intermediates relative to the native state is dependent on protein concentration indicates that all the intermediate states are monomeric, and suggests that the dimer dissociates prior to or concurrent with the unfolding of the red domain. As illustrated in Figure 3, the dimer interface is mainly formed by loop 3 (colored red), which packs into the active site of the opposite protomer (Figure 3(b)). Monomeric forms of TIM have been generated by introducing mutations into the tip of loop 3,[21] or by addition of denaturant.[22] These monomeric species are relatively compact, but show a significant loss of catalytic activity, stability, and circular dichroism signal relative to the native dimer. The characteristics of these species resemble those of the I1 state, suggesting that I1 may be more stable than the native dimer under these conditions. Alternatively, these species could represent the fully folded monomer, which may constitute an independent intermediate that was not observed in this study.

Although the three domains are clearly identified by the clustering of *m*-values, the boundaries between the domains are not well defined by this data alone. Scanning Gly → Ala mutagenesis represents a complementary approach, which quantifies the extent to which each helix is unfolded in the I1, I2, and I3 states. The data show that helices 6, 7, and 8 belong to the red domain, helices 1, 4, and 5 belong to the green domain, and helices 2 and 3 belong to the blue domain. The majority of phi values are close to either zero or unity, allowing an unambiguous assignment. However, fractional phi values are observed for helices 1, 3, and 5 (Table 2). Fractional values may indicate partial unfolding of the helix in an intermediate, or the presence of an ensemble of intermediate structures in which the helix is only folded in a subset of the structures.

Models of I1 and I2 were constructed by iteratively removing secondary structures in all possible combinations from the TIM crystal structure, followed by calculation of the solvent accessibility of all probes (Figure 4). A residue was designated as exposed if it was computed to be more than 30% solvent-accessible.[16] The best agreement with the accessibility data was obtained by assigning strands 7 and 8 to the red domain; strands 1, 5, and 6 to the green domain; and strands 2, 3, and 4 to the blue domain (Figure 3(a)). However, the agreement was not perfect. Three cysteine probes belonging to the red domain

are not exposed to solvent in the I1 model, and three probes belonging to the green domain are not exposed to solvent in the I2 model. Conversely, two probes belonging to the green domain and three probes belonging to the blue domain are improperly exposed to solvent in the I1 and I2 models, respectively. Unfolding of secondary structure units of the crystal structure cannot explain the observed protection patterns for these probes. One possible explanation is the formation of non-native structure in the intermediates. For example, helix 5 may rotate to cover the hydrophobic surface of the barrel interior exposed by unfolding of the red domain in I1. Beta strands 2–4, which make up the blue domain, have a very poor hydrogen-bonding geometry. When the rest of the barrel unfolds in I2, they may reorganize to optimize their interactions. Non-native structure in a folding intermediate was recently proposed to explain anomalous phi values in the Im7 protein[23] and has been observed in folding intermediates of several other proteins.[24–29]

Non-native structure is difficult to observe directly, and has not been reported in native-state hydrogen exchange experiments. One significant difference between MPAX and hydrogen exchange is that MPAX measures the solvent accessibility of side-chains while hydrogen exchange measures the breaking of backbone hydrogen bonds. Movements of secondary structure that expose only amino acid side-chains to solvent are invisible to hydrogen exchange. MPAX and hydrogen exchange can provide complementary data on the structure of proteins, to provide a more complete view of unfolding intermediates.

The equilibrium unfolding pathway of a $(\beta/\alpha)_8$ barrel

Analysis of PRAI[2] and tryptophan synthase[6,30] suggests that these $(\beta/\alpha)_8$ barrel proteins fold through two domains. The N-terminal subdomain consists of the first five helices and six strands, and the C-terminal subdomain comprises the last three helices and two strands. These studies led to a $6+2$ model for $(\beta/\alpha)_8$ barrel folding, in which the first six (β/α) units are proposed to fold first, followed by condensation of the last two (β/α) units. Recent studies on HisF, a remarkably symmetrical $(\beta/\alpha)_8$ barrel, have suggested a $4+4$ model, in which the N and C-terminal domains contain an equal number of secondary structure units.[3] Our results suggest that TIM unfolds according to a $3+3+2$ model. The red subdomain corresponds exactly to the C-terminal subdomain of PRAI and tryptophan synthase, but the N-terminal subdomain is divided into two structures. Because the blue domain represents the most stable region of the protein, one might speculate that it represents the site of initial structure formation, upon which the green and red domains fold sequentially. It is important to note, however, that knowledge of the equilibrium unfolding pathway of TIM cannot, in itself, determine the kinetic folding pathway.

In our hands, TIM does not unfold reversibly in spectrophotometrically monitored thermal or denaturant melts because it aggregates. Under the conditions employed for native-state MPAX, however, the protein transiently unfolds and refolds many times before alkylation occurs (demonstrated by the fact that the alkylation rates of all protected cysteine probes depend linearly on the concentration of alkylating reagent; see Methods). This apparent contradiction is reconciled by the fact that the folded protein is always favored by >2 kcal mol^{-1} in the MPAX experiment, so that $<3\%$ of the molecules are unfolded at any given time. Under these conditions, refolding of unstable intermediates is apparently faster than aggregation. It is interesting to note that the red and green isotherms merge with the blue isotherm before $\Delta G_{N\rightarrow I1}$ or $\Delta G_{N\rightarrow I2}$ reaches zero. This means that no intermediate would predominate at equilibrium in a denaturant melt at a protein concentration of 2 μM, and implies that TIM should exhibit two-state unfolding in the absence of aggregation or other kinetic traps.

This study paves the way for the detailed energetic analysis of other large and poorly behaved proteins. The MPAX technique allows the study of proteins at low (<1 μM) concentrations and is not limited by protein size. Thus, quantitative energetic frameworks previously limited to small proteins can now be applied to a variety of interesting biological problems.

Methods

Protein expression and purification

The yeast TIM gene in plasmid pH6_TIM_PKA[16] was used for all studies. TIM contains two native cysteine residues: C41 and C126. The presence of more than a single cysteine residue per molecule can lead to complications when MPAX data are used for protein stability measurements. Experiments were conducted in the C41V/C126A TIM background (except where noted) to eliminate this problem. The TIM protein construct contains a C-terminal protein kinase A site used for labeling with radioactive phosphate, and an N-terminal $6\times$His tag on a 30 amino-acid residue linker. TIM was co-expressed as described[16] with pMPAX plasmids that misincorporate cysteine residues in place of isoleucine, leucine, or valine residues. Glycine and alanine mutants were constructed by Kunkel mutagenesis[31] and co-expressed with the valine misincorporator pMPAX plasmid.

TIM protein was purified from inclusion bodies according to the following protocol. Inclusion bodies were isolated by two sonications in 50 mM $NaPO_4$ pH 7.5, 10 mM EDTA, 1% Triton X100. Protein was solubilized in 5 mM Tris pH 8, 5 M urea, 10 mM EDTA, 10 mM DTT and insoluble material was removed by centrifugation. The clarified supernatant was filtered through a 0.2 μm filter, loaded onto a HiTrap Q column (Pharmacia) equilibrated in 5 mM Tris pH 8, 5 M urea, and eluted

in a linear gradient of NaCl. TIM protein eluted at ~50 mM NaCl. TIM-containing fractions were pooled and an equal volume of Buffer A (250 mM sodium acetate, 125 mM Na_2HPO_4, 125 mM NaH_2PO_4, 5 M urea; unadjusted pH ~8) was added. The protein was then loaded on a Ni-NTA column (Pharmacia) equilibrated in Buffer A. TIM protein was eluted in a linear gradient of Buffer B (250 mM acetic acid, 250 mM H_3PO_4; unadjusted pH ~2). TIM-containing fractions were pooled, EDTA was added to 10 mM, DTT was added to 5 mM, and the pH was adjusted to 8 with 10 M NaOH. The protein was then dialyzed overnight into 5 mM Tris, pH 8, with several buffer changes. The dialysate was clarified by centrifugation and filtration, and loaded onto a HiTrap Q column equilibrated in 5 mM Tris, pH 8. TIM was eluted from the column in a linear gradient of NaCl, with the protein eluting at ~100 mM NaCl. Peak fractions were pooled and DTT was added to 0.1 mM. Proteins were stored at 4 °C for up to 1 month.

Native-state alkylation

The reagents most commonly used to denature proteins are urea and GdmCl. TIM unfolds with a half-life of 37 hours in 6 M urea, compared to a half-life of 15 seconds in 3 M GdmCl (data not shown). Therefore, GdmCl was chosen for MPAX studies to allow a rapid approach to equilibrium after dilution into denaturant.

A solution of 2 μM cysteine-misincorporated TIM was incubated with 10 mM iodoacetamide, 50 mM sodium bicine pH 8.6, in increasing concentrations of GdmCl. All protein concentrations are reported as the concentration of TIM monomers. At various timepoints, a 20 μL sample was taken and quenched with 50 μL of 20 mM β-mercaptoethanol, 1 mg/mL BSA. Samples were cleaved with NTCB and analyzed by gel electrophoresis as described.[16] The intensity of cleavage bands as a function of time was quantitated on a Phosphorimager (Molecular Dynamics) and fit to a single exponential.

The rate of alkylation of all cysteine probes was linearly dependent upon the concentration of iodoacetamide in up to 1.2 M GdmCl for the C41V/C126A protein and 2.2 M GdmCl for the C41V protein (data not shown), confirming that alkylation occurs by the EX2 mechanism under these conditions. In the EX2 regime, the free energy for unfolding of the native state to an intermediate is defined as:

$$\Delta G_{N\rightarrow I} = -RT \ln\left(\frac{k_{obs}}{k_{alk}}\right) \quad (1)$$

where k_{alk} is the intrinsic rate of alkylation of the misincorporated cysteine residue in the unfolded state, corrected for the effect of GdmCl on k_{alk}.[16] The apparent rate of alkylation at a probe can be described as the sum of the contributions from denaturant-independent LSFs that expose the cysteine residue to solvent and large-scale protein unfolding.[10] The observed free energy for formation of intermediates from the native state as a function of GdmCl concentration were fit to the following equation:

$$\Delta G_{N\rightarrow I}([\text{GdmCl}]) = -RT \ln(e^{-\Delta G_{LSF}/RT} + e^{(-\Delta G^0_{N\rightarrow I} + m[\text{GdmCl}])/RT}) \quad (2)$$

where ΔG_{LSF} is the free energy associated with LSFs, $\Delta G^0_{N\rightarrow I}$ is the free energy for unfolding of the protein subdomain to intermediate I in zero denaturant, and m is the denaturant dependence of the subdomain stability. Because the native state of TIM is a dimer, $\Delta G^0_{N\rightarrow I}$ for monomeric intermediates will be concentration dependent. The free energy values are reported per mole of monomer at a standard state concentration of 4 μM, and can be corrected to the standard state of 1 M by addition of 3.7 kcal mol^{-1}.

Gly → Ala phi analysis

For glycine/alanine mutant pairs, the difference in $\Delta G_{N\rightarrow I}$ between mutant proteins was defined as:

$$\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I} = -RT \ln\left(\frac{k^{Ala}_{obs}}{k^{Gly}_{obs}}\right) \quad (3)$$

where k_{obs} is the observed rate of alkylation of a cysteine probe in the glycine or alanine mutant background. $\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}$ at each misincorporated cysteine residue was measured in 1.7, 1.8, 1.9, and 2.0 M GdmCl (without extrapolation to zero denaturant concentration) and averaged for probes in the same subdomain. The fractional unfolding of a helix in each intermediate was calculated as:

$$\Phi_{N\rightarrow I} = \frac{\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}}{\Delta\Delta G^{Gly\rightarrow Ala}_{\text{Helix unfolding}}} = \frac{\Delta\Delta G^{Gly\rightarrow Ala}_{N\rightarrow I}}{0.9 \text{ kcal mol}^{-1}} \quad (4)$$

using the value of 0.9 kcal mol^{-1} from Myers *et al*.[18] for the energetic difference in the helical propensities of alanine and glycine.

Acknowledgements

We thank R. Baldwin and members of the Harbury lab for criticism and advice throughout the course of this work. J.S. is supported by the Paul and Mildred Berg Stanford graduate fellowship. This research was supported by a grant from the Chicago Community Trust to P.B.H., and a Terman fellowship to P.B.H.

References

1. Deng, Y. & Smith, D. L. (1998). Identification of unfolding domains in large proteins by their unfolding rates. *Biochemistry*, **37**, 6256–6262.
2. Eder, J. & Kirschner, K. (1992). Stable substructures of eightfold beta alpha-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry*, **31**, 3617–3625.
3. Hocker, B., Beismann-Driemeyer, S., 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.
4. 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.
5. Noland, B. W., Dangott, L. J. & Baldwin, T. O. (1999). Folding, stability, and physical properties of the alpha subunit of bacterial luciferase. *Biochemistry*, **38**, 16136–16145.

6. 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.
7. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238.
8. Raschke, T. M. & Marqusee, S. (1998). Hydrogen exchange studies of protein structure. *Curr. Opin. Biotechnol.* **9**, 80–86.
9. Chamberlain, A. K. & Marqusee, S. (2000). Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Advan. Protein Chem.* **53**, 283–328.
10. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
11. Myers, J. K., Pace, C. N. & 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.
12. Smith, D. L. (1998). Local structure and dynamics in proteins characterized by hydrogen exchange and mass spectrometry. *Biochemistry (Moscow)*, **63**, 285–293.
13. Feng, Z., Butler, M. C., Alam, S. L. & Loh, S. N. (2001). On the nature of conformational openings: native and unfolded-state hydrogen and thiol-disulfide exchange studies of ferric aquomyoglobin. *J. Mol. Biol.* **314**, 153–166.
14. Ha, J. H. & Loh, S. N. (1998). Changes in side-chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nature Struct. Biol.* **5**, 730–737.
15. Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Advan. Protein Chem.* **21**, 287–386.
16. Silverman, J. A. & Harbury, P. B. (2002). Rapid mapping of protein structure, interactions and ligand binding by misincorporation proton–alkyl exchange. *J. Biol. Chem.* **277**, 30968–30975.
17. O'Neil, K. T. & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, **250**, 646–651.
18. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1997). A direct comparison of helix propensity in proteins and peptides. *Proc. Natl Acad. Sci. USA*, **94**, 2833–2837.
19. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1997). Helix propensities are identical in proteins and peptides. *Biochemistry*, **36**, 10923–10929.
20. Huyghues-Despointes, B. M., Langhorst, U., Steyaert, J., Pace, C. N. & Scholtz, J. M. (1999). Hydrogen-exchange stabilities of RNase T1 and variants with buried and solvent-exposed Ala → Gly mutations in the helix. *Biochemistry*, **38**, 16481–16490.
21. Schliebs, W., Thanki, N., Jaenicke, R. & Wierenga, R. K. (1997). A double mutation at the tip of the dimer interface loop of triosephosphate isomerase generates active monomers with reduced stability. *Biochemistry*, **36**, 9655–9662.
22. Morgan, C. J., Wilkins, D. K., Smith, L. J., Kawata, Y. & Dobson, C. M. (2000). A compact monomeric intermediate identified by NMR in the denaturation of dimeric triose phosphate isomerase. *J. Mol. Biol.* **300**, 11–16.
23. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
24. Rothwarf, D. M. & Scheraga, H. A. (1996). Role of non-native aromatic and hydrophobic interactions in the folding of hen egg white lysozyme. *Biochemistry*, **35**, 13797–13807.
25. Balbach, J., Steegborn, C., Schindler, T. & Schmid, F. X. (1999). A protein folding intermediate of ribonuclease T1 characterized at high resolution by 1D and 2D real-time NMR spectroscopy. *J. Mol. Biol.* **285**, 829–842.
26. Neira, J. L. & Fersht, A. R. (1999). Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *J. Mol. Biol.* **285**, 1309–1333.
27. Eliezer, D., Chung, J., Dyson, H. J. & Wright, P. E. (2000). Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry*, **39**, 2894–2901.
28. Forge, V., Hoshino, M., Kuwata, K., Arai, M., Kuwajima, K., Batt, C. A. & Goto, Y. (2000). Is folding of beta-lactoglobulin non-hierarchic? Intermediate with native-like beta-sheet and non-native alpha-helix. *J. Mol. Biol.* **296**, 1039–1051.
29. Troullier, A., Reinstadler, D., Dupont, Y., Naumann, D. & Forge, V. (2000). Transient non-native secondary structures during the refolding of alpha-lactalbumin detected by infrared spectroscopy. *Nature Struct. Biol.* **7**, 78–86.
30. 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 beta/alpha barrel protein. *Protein Sci.* **8**, 1200–1209.
31. Kunkel, T. A., Bebenek, K. & McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125–139.
32. 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 8 July 2002; received in revised form 25 September 2002; accepted 25 September 2002)