doi:10.1016/j.jmb.2006.07.080 J. Mol. Biol. (2006) 363, 262–278

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Folding Energy Landscape of the Dimerization Domain of *Escherichia coli* Trp Repressor: A Joint Experimental and Theoretical Investigation

B. Robert Simler[1], Yaakov Levy[2], José N. Onuchic[2] and C. Robert Matthews[1]*

[1]*Department of Biochemistry and Molecular Pharmacology University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605 USA*

[2]*Center for Theoretical Biological Physics, University of California at San Diego, La Jolla CA 92093, USA*

**Corresponding author*

Enhanced structural insights into the folding energy landscape of the N-terminal dimerization domain of *Escherichia coli* tryptophan repressor, $[2\text{-}66]_2$ TR, were obtained from a combined experimental and theoretical analysis of its equilibrium folding reaction. Previous studies have shown that the three intertwined helices in $[2\text{-}66]_2$ TR are sufficient to drive the formation of a stable dimer for the full-length protein, $[2\text{-}107]_2$ TR. The monomeric and dimeric folding intermediates that appear during the folding reactions of $[2\text{-}66]_2$ TR have counterparts in the folding mechanism of the full-length protein. The equilibrium unfolding energy surface on which the folding and dimerization reactions occur for $[2\text{-}66]_2$ TR was examined with a combination of native-state hydrogen exchange analysis, pepsin digestion and matrix-assisted laser/desorption mass spectrometry performed at several concentrations of protein and denaturant. Peptides corresponding to all three helices in $[2\text{-}66]_2$ TR show multi-layered protection patterns consistent with the relative stabilities of the dimeric and monomeric folding intermediates. The observation of protection exceeding that offered by the dimeric intermediate in segments from all three helices implies that a segment-swapping mechanism may be operative in the monomeric intermediate. Protection greater than that expected from the global stability for a single amide hydrogen in a peptide from the C-helix possibly and another from the A-helix may reflect non-random structure, possibly a precursor for segment swapping, in the urea-denatured state. Native topology-based model simulations that correspond to a funnel energy landscape capture both the monomeric and dimeric intermediates suggested by the HX MS data and provide a rationale for the progressive acquisition of secondary structure in their conformational ensembles.

Keywords: protein folding; hydrogen exchange; segment swapping; mass spectrometry; Gō model

Introduction

Complementary experimental and theoretical studies of protein folding reactions can provide insights into the energy surfaces that control the decoding of information contained in the one-dimensional amino acid sequence into the three-dimensional structure. The link between the energy surface and the sequence defining it, i.e. the solution of the protein folding problem, requires detailed knowledge of the structures represented by the transient intermediates and transition states

Present addresses: B. R. Simler, Genzyme Corporation, Framingham, MA 01701, USA; Y. Levy, Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel, 76100.

Abbreviations used: $[2\text{-}66]_2$ TR, dimeric *Trp* repressor fragment containing residues 2-66; CHCA, α-cyano-4-hydroxycinnaminic acid; HX, hydrogen/deuterium exchange; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; TFA, trifluoroacetic acid; TR, *Trp* repressor.

E-mail address of the corresponding author: c.robert.matthews@umassmed.edu

that intervene between the native and unfolded forms.

From one theoretical perspective, the funneled energy landscape of evolutionarily selected proteins governs their robust ability to efficiently organize the polypeptide chain into a specific structure.[1,2] Because energetic frustration is sufficiently small, models that exploit topological information contained in the X-ray crystal structures are able to capture the dominant structural and thermodynamic properties of the folding mechanisms of both small proteins[3–5] and more complex proteins such as homooligomers.[6–10] The topological features of the native protein are often sufficient to predict the general features of the structures and thermodynamic properties of the states populated between distinct unfolded and folded protein conformational ensembles.

From an experimental perspective, mutational analysis has proven to be an invaluable tool for identifying the side-chains that participate in very short-lived transition states and folding intermediates.[11–13] A variety of biophysical tools are also available to monitor various higher-order structural properties of transient intermediates.[14–18] Hydrogen exchange (HX) methods in combination with NMR have proven to be a particularly useful source of information on the site-specific development of secondary structure in kinetic intermediates that appear during folding reactions.[19–23] The exchange of backbone amide hydrogen atoms with solvent can be retarded by several orders of magnitude by the formation of stable hydrogen bonds with backbone carbonyl oxygen atoms, i.e. by the formation of secondary structure, during folding. Recently, mass spectrometry has become a viable alternative to NMR for the analysis of HX experiments on transient folding intermediates.[24–27] Although pulse-quench HX mass spectrometry (MS) analysis coupled with proteolytic digestion lacks the site-specific resolution provided by NMR, MS-based methods offer rapid identification of labeled peptides from large (>30 kDa) proteins and access to micromolar protein concentrations that can minimize solubility problems sometimes encountered in the NMR-based analysis.

The application of a complementary approach, the native-state HX method,[28–30] can provide insights into the secondary structures of rare, high-energy partially-folded conformations in equilibrium with the native conformation. Although the dynamic folding reaction beginning from the unfolded state may not sample the entire equilibrium energy surface defined by native-favoring conditions, the equilibrium surface provides a framework for testing the role of its local minima in guiding the flow of material from the unfolded to the native state of a protein.

The dimeric Trp repressor (TR) from *Escherichia coli* provides a particularly intriguing target for native-state HX MS studies of its equilibrium folding energy surface. Its 107 residues per chain form a highly helical, intertwined protein that can be described as a domain-swapped or segment-swapped dimer.[31,32] Helices A, B, and C comprise the intertwined dimerization domain in which helices A and B from one chain are docked on the C-helix from the partner chain. Helices D, E and the intervening turn form the DNA reading heads, and helix F buttresses the reading heads against the dimerization domain (Figure 1). The full-length wild-type protein folds *via* a complex mechanism featuring three parallel channels that appear to arise from the random association of two folded monomeric intermediates with differing, non-native structures.[33–37] The three dimeric intermediates then convert independently to the

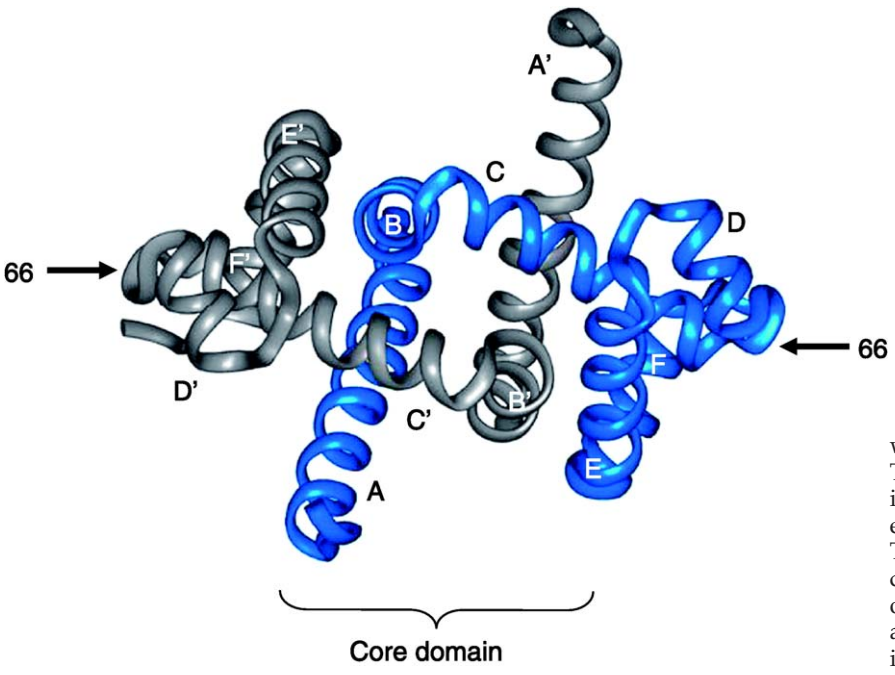

Figure 1. Crystal structure of wild-type *E. coli Trp* repressor.[44] The two monomers are displayed in blue and gray. The six helices in each monomer are denoted A–F. The core region, $[2\text{-}66]_2$ TR, which is comprised of the first three helices of each monomer A, B and C, ends at residue 66, the position of which is indicated.

native dimer *via* rate-limiting rearrangement reactions. Fortunately, the folding mechanism can be greatly simplified by the truncation of TR to produce its N-terminal dimerization domain (denoted $[2\text{-}66]_2$TR). The removal of the DNA reading heads, i.e. the D and E helices, and the C-terminal F helix yields a stable dimeric form whose single-channel folding mechanism closely resembles one of the channels for the full-length protein.[38,39]

The proposed kinetic mechanism for the folding of the dimerization domain of TR involves both monomeric and dimeric intermediates,

$$2U \leftrightarrows 2I \leftrightarrows I_2 \leftrightarrows N_2 \qquad \text{Scheme 1}$$

where the unfolded state, U, proceeds to a partially-folded monomeric intermediate, I, in the sub-millisecond time-scale. After a near diffusion-limited association reaction produces the I_2 species, this latter intermediate undergoes a slow rearrangement step to the native dimer, N_2. The ellipticity of I at 222 nm is ~50% of the native form, as inferred from a well-folded monomeric version of $[2\text{-}66]_2$ TR obtained by mutating leucine 39 to glutamic acid (unpublished results). The ellipticity at 222 nm of I_2 is ~75% of that of the native form, based upon the amplitude of the CD signal after the 5 ms dead-time of mixing when I_2 is the dominant form in solution. Unfortunately, the sub-millisecond folding and association of the subunits at micromolar concentrations preclude conventional quenched-flow HX experiments to examine the protection offered in both of the transient intermediates.

The relative simplicity of the single-channel folding reaction for $[2\text{-}66]_2$ TR, the preservation of the transient monomeric and dimeric intermediates observed in each of the channels for the full-length protein, and the progressive enhancement of secondary structure in these intermediates make $[2\text{-}66]_2$ TR an ideal candidate for a comparative native-state HX MS/native topology-based simulation of its equilibrium folding energy surface. The results, which are in excellent agreement, demonstrate the presence of a stable sub-core of structure in the dimer and imply a segment-swapping mechanism for dimer formation from a pair of folded monomers.

Results

Experimental analysis

Proteolytic digestion of [2-66]$_2$ TR

Insights into the differential protection against exchange experienced by various segments of $[2\text{-}66]_2$ TR can be obtained by pepsin digestion of partially exchanged samples under acidic conditions. The substantial retardation of exchange at pH 2.5 minimizes the loss of label to back-exchange during proteolysis and, thereby, enables MALDI MS analysis of the protection patterns. The hydrogen exchange behavior of the eight peptides reported in this study (Figure 2) provide nearly complete coverage of $[2\text{-}66]_2$ TR (only leucine 20 is not included in this set of peptides), and the termini of several of the peptides correlate well with boundaries of interest, i.e. the borders of the helices. The three remaining peptides, 10-21, 11-21 and 42-60, largely overlap these eight peptides and, therefore, do not provide significant additional information on the hydrogen exchange properties of $[2\text{-}66]_2$ TR.

Amide hydrogen exchange behavior in [2-66]$_2$TR peptides

Native-state exchange experiments on $[2\text{-}66]_2$ TR were conducted by transferring the protein from H_2O to ^2H_2O and allowing the exchange to occur for various lengths of time at pH 7.2 and 25 °C before quenching, digestion, and mass analysis. A representative set of mass spectra *versus* time for peptide 26-35 exhibits a continuous shift of the isotopic envelope to higher mass as the incubation time is increased (Figure 3). This response can be understood in terms of the classical kinetic model for the exchange of an amide hydrogen with the solvent:

$$F(H) \underset{k_1}{\overset{k_{-1}}{\rightleftarrows}} U(H) \xrightarrow{k_2} U(D) \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} F(D) \qquad (1)$$

where F and U represent the folded, i.e. protected, and unfolded, i.e. unprotected, forms of a given amide in a protein. H and D represent the protonated and deuterated forms of this amide, and k_{-1}, and k_2 represent the rate constants for the closing reaction and for the amide hydrogen exchange reaction, respectively.[40] If the local unfolding/refolding reaction occurs many times before the amide hydrogen is

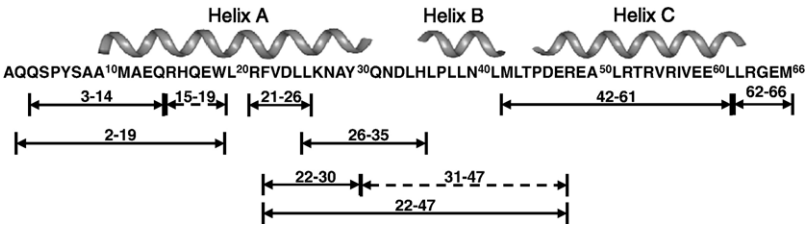

Figure 2. Peptide coverage map of $[2\text{-}66]_2$ TR. The eight peptides whose HX data were processed are shown as double-headed arrows underneath the amino acid sequence for $[2\text{-}66]_2$ TR. The location of the A, B and C helices are shown above the amino acid sequence. Broken line arrows indicate segments that were analyzed by subtracting the masses of two peptides from each other.

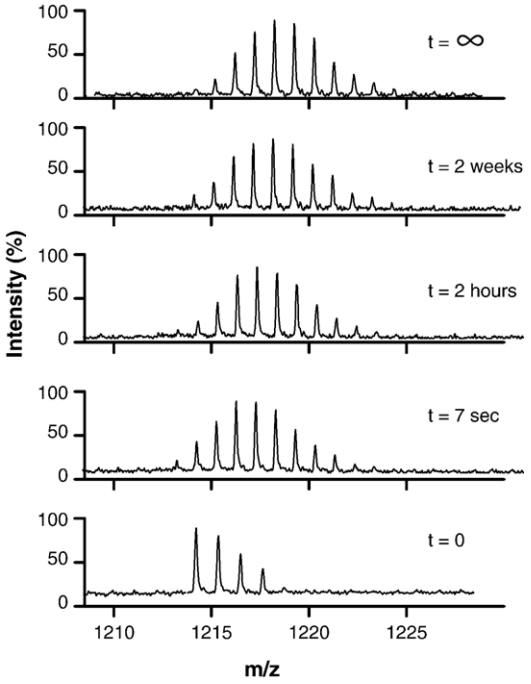

Figure 3. Time-dependent shift of the isotopic envelope for peptide 26-35 in 10 mM phosphate (pH 7.2) at 25 °C in ^2H_2O. The progressive shift of the isotopic envelope to higher mass/charge (m/z) ratio for the +1 ion demonstrates EX2 exchange kinetics.

exchanged, i.e. the EX2 limit where $k_{-1} \gg k_2$, the phenomenological rate constant, k_{ex}, for isotopic exchange becomes;

$$k_{ex} = Kk_2 \qquad (2)$$

where K is the equilibrium constant between the folded and unfolded states ($K = k_1/k_{-1}$). For multiple amide hydrogen atoms in a peptide, EX2 kinetics are expected to produce a progressive shift in the isotopic envelope, as seen in Figure 3. For EX1 kinetics, where $k_2 \gg k_{-1}$, the disappearance of the folded peak coincides with the appearance of the unfolded peak and implies the simultaneous exchange of all amide hydrogen atoms. All of the peptides examined in this study exhibited EX2 kinetics.

Four classes of amide hydrogen exchange behavior were observed in the eight peptic peptides derived from $[2\text{-}66]_2$ TR incubated in ^2H_2O for up to 24 h at pH 7.2 and at 25 °C. A graphical representation illustrating the data analysis leading to the identification of these four classes is shown in Figure 4 (for a more detailed description of the analysis, see Materials and Methods). The fastest exchanging amide hydrogen atoms, i.e. those in side-chains or in backbone positions exposed to solvent, were assumed to exchange at a rate, k_f, equal to that calculated for unprotected amide hydrogen atoms.[41] These amide hydrogen atoms are expected to be >99% exchanged by the time the first datum point was taken after a 7 s exposure to ^2H_2O. Depending upon the peptide, up to two classes of amide hydrogen atoms exchanged *via* processes whose apparent rate constants could be measured by fitting the exchange curves to a sum of exponentials, as described in Materials and Methods. The apparent rate constants obtained from these fits, k_{i1} and k_{i2}, are ~10^2-fold (Int1) and ~10^4-fold (Int2) slower than the rate for a solvent-exposed amide hydrogen. The resistance to exchange for more than 24 h implies that the apparent rate constant of the slowly exchanging hydrogen atoms, k_s, is $\geq$ ~10^6-fold slower than that for a solvent-exposed amide hydrogen under these conditions. Because the modulation of the apparent rate constants by varying protein or urea concentration over limited ranges is less than the two order-of-magnitude difference between the apparent rate constants (Table 1), this empirical classification scheme is sufficiently robust to allow useful comparisons among the peptides.

Dependence of amide hydrogen exchange rates on the concentrations of urea and protein

To obtain information on the partially unfolded forms of $[2\text{-}66]_2$ TR whose populations are enhanced in the presence of a chemical denaturant, the effect of low concentrations of denaturant, 0.5 M and 1.0 M urea, on amide hydrogen exchange behavior was investigated. Although the native dimeric state is the dominant population under these conditions, the proposed folding mechanism (Scheme 1) includes dimeric and monomeric intermediates through which exchange is likely to be enhanced at increasing concentrations of urea. The denaturant insensitivity of local fluctuations[42] and the urea-dependence of the free energy differences, i.e. the m-values, for the subglobal transition from N_2 to I_2, 0.78 kcal mol^{-1} M^{-1},

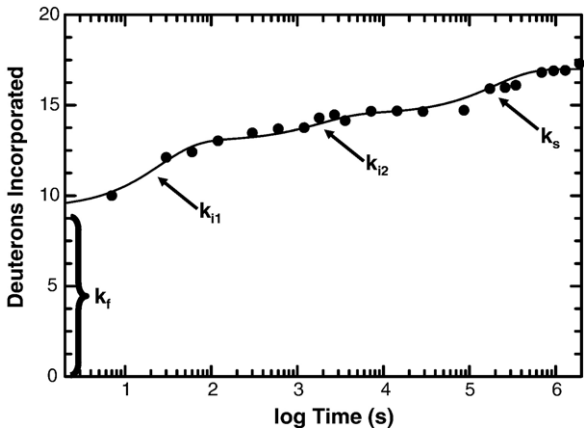

Figure 4. Graphical interpretation of the data-fitting routine for peptide 42-61 obtained by incubating 20 μM $[2\text{-}66]_2$ TR in 10 mM phosphate (pH 7.2) at 25 °C in the absence of denaturant for up to four weeks in ^2H_2O. Amide hydrogen atoms exchanging *via* the fast reaction, k_f, are quantified by extrapolating the exponential fits back to 1 s. The other three exponentials, k_{i1}, k_{i2}, and k_s, are indicated.

Table 1. H/D-exchange parameters for $[2\text{-}66]_2$ TR peptides

Peptide	Condition	N_{fast}	k_f[a] (s^{-1})	N_{int1}	k_{i1} (s^{-1})	N_{int2}	k_{i2} (s^{-1})	N_{slow}
3-14 (10)[c]	10 μM	8.3 (0.2)	3.1	1.6 (0.2)	$1.0\ (0.3)\times10^{-2}$	nd[b]	nd	0.1 (0.2)
	20 μM	8.1 (0.1)	3.1	1.6 (0.3)	$3.1\ (1.9)\times10^{-2}$	nd	nd	0.1 (0.3)
	20 μM+0.5 M urea	9.8 (0.2)	3.1	nd	nd	nd	nd	0.2 (0.2)
	20 μM+1.0 M urea	9.9 (0.3)	3.1	nd	nd	nd	nd	0.1 (0.2)
	50 μM	7.3 (0.3)	3.1	2.2 (0.4)	$4.0\ (2.1)\times10^{-2}$	nd	nd	0.5 (0.4)
15-19 (4)	10 μM	0.2 (0.2)	3.6	0.1 (0.1)	$8.0\ (4.0)\times10^{-3}$	nd	nd	3.7 (0.3)
	20 μM	0.3 (0.3)	3.6	nd	nd	nd	nd	3.7 (0.3)
	20 μM+0.5 M urea	0.2 (0.2)	3.6	0.3 (0.2)	$1.0\ (0.5)\times10^{-2}$	nd	nd	3.5 (0.3)
	20 μM+1.0 M urea	0.2 (0.2)	3.6	0.3 (0.2)	$4.0\ (3.0)\times10^{-3}$	nd	nd	3.5 (0.3)
	50 μM	0.2 (0.1)	3.6	0.3 (0.2)	$4.0\ (2.0)\times10^{-2}$	nd	nd	3.5 (0.2)
21-26 (5)	10 μM	2.0 (0.3)	1.2	1.2 (0.2)	$3.5\ (1.4)\times10^{-2}$	0.5 (0.2)	$9.4\ (6.8)\times10^{-4}$	1.3 (0.3)
	20 μM	1.8 (0.3)	1.2	1.2 (0.2)	$3.1\ (2.2)\times10^{-2}$	0.7 (0.1)	$1.4\ (1.0)\times10^{-4}$	1.3 (0.3)
	20 μM+0.5 M urea	2.3 (0.4)	1.2	1.0 (0.2)	$8.3\ (5.7)\times10^{-2}$	0.4 (0.1)	$7.9\ (4.3)\times10^{-4}$	1.3 (0.2)
	20 μM+1.0 M urea	2.3 (0.3)	1.2	1.3 (0.2)	$8.0\ (6.0)\times10^{-2}$	nd	nd	1.4 (0.2)
	50 μM	1.3 (0.2)	1.2	1.5 (0.2)	$1.5\ (0.5)\times10^{-2}$	0.8 (0.1)	$4.7\ (1.3)\times10^{-4}$	1.5 (0.2)
26-35 (9)	10 μM	7.5 (0.3)	3.3	0.2 (0.1)	$2.0\ (1.7)\times10^{-2}$	0.1 (0.1)	$7.5\ (4.8)\times10^{-4}$	1.2 (0.3)
	20 μM	5.9 (0.4)	3.3	0.8 (0.2)	$2.3\ (1.1)\times10^{-2}$	0.6 (0.2)	$3.8\ (2.5)\times10^{-4}$	1.7 (0.4)
	20 μM+0.5 M urea	6.6 (0.3)	3.3	1.5 (0.3)	$1.6\ (0.7)\times10^{-2}$	nd	nd	0.9 (0.3)
	20 μM+1.0 M urea	6.6 (0.3)	3.3	1.8 (0.3)	$4.4\ (1.7)\times10^{-2}$	nd	nd	0.6 (0.3)
	50 μM	5.2 (0.3)	3.3	0.8 (0.2)	$5.0\ (3.4)\times10^{-2}$	0.8 (0.1)	$4.7\ (1.3)\times10^{-4}$	2.2 (0.3)
31-47 (14)	10 μM	5.4 (0.3)	2.6	4.6 (0.3)	$3.0\ (1.2)\times10^{-2}$	2.0 (0.3)	$7.5\ (4.0)\times10^{-4}$	2.0 (0.2)
	20 μM	4.2 (0.3)	2.6	4.7 (0.3)	$1.8\ (1.1)\times10^{-2}$	2.5 (0.2)	$6.4\ (3.8)\times10^{-4}$	2.6 (0.3)
	20 μM+0.5 M urea	10.7 (0.5)	2.6	1.6 (0.3)	$7.0\ (3.2)\times10^{-2}$	nd	nd	1.7 (0.4)
	20 μM+1.0 M urea	10.2 (0.6)	2.6	1.8 (0.4)	$9.0\ (6.2)\times10^{-2}$	nd	nd	2.0 (0.4)
	50 μM	2.9 (0.2)	2.6	5.3 (0.3)	$7.1\ (4.1)\times10^{-2}$	2.1 (0.2)	$6.8\ (4.1)\times10^{-4}$	3.7 (0.3)
42-61 (18)	10 μM	9.9 (0.7)	3.2	4.5 (0.5)	$1.0\ (0.4)\times10^{-2}$	1.5 (0.3)	$5.7\ (2.9)\times10^{-4}$	2.1 (0.5)
	20 μM	9.4 (0.5)	3.2	3.7 (0.4)	$2.3\ (1.1)\times10^{-2}$	1.5 (0.2)	$9.4\ (6.8)\times10^{-4}$	3.4 (0.3)
	20 μM+0.5 M urea	10.3 (0.5)	3.2	4.7 (0.5)	$4.5\ (2.3)\times10^{-2}$	0.8 (0.2)	$8.3\ (4.4)\times10^{-3}$	2.2 (0.5)
	20 μM+1.0 M urea	11.0 (0.6)	3.2	4.2 (0.4)	$5.5\ (3.1)\times10^{-3}$	0.6 (0.2)	$2.5\ (1.7)\times10^{-3}$	2.2 (0.5)
	50 μM	7.4 (0.4)	3.2	4.5 (0.3)	$4.1\ (2.2)\times10^{-2}$	1.8 (0.2)	$5.0\ (2.1)\times10^{-4}$	4.3 (0.3)
62-66 (4)	10 μM	2.8 (0.2)	4.2	0.9 (0.2)	$8.0\ (0.5)\times10^{-3}$	nd	nd	0.3 (0.2)
	20 μM	2.7 (0.3)	4.2	0.8 (0.3)	$1.2\ (0.5)\times10^{-2}$	nd	nd	0.4 (0.3)
	20 μM+0.5 M urea	2.8 (0.3)	4.2	0.7 (0.3)	$9.0\ (5.1)\times10^{-3}$	nd	nd	0.5 (0.3)
	20 μM+1.0 M urea	3.0 (0.2)	4.2	0.6 (0.2)	$7.0\ (4.5)\times10^{-3}$	nd	nd	0.4 (0.2)
	50 μM	2.6 (0.3)	4.2	0.9 (0.2)	$1.3\ (0.6)\times10^{-2}$	nd	nd	0.5 (0.3)

Conditions for all experiments were 10 mM phosphate, pH 7.2, 25 °C. Errors are listed in parentheses.

[a] The values for the k_f exchange rates were determined by averaging the exchange rates for the individual amino acids in the peptide as determined by the method of Bai and Englander[41] for a free, unstructured amino acid.

[b] Not detectable.

[c] The number of exchangeable amide hydrogen atoms in each peptide, accounting for the proline residues at positions 6, 37 and 45, are shown in parentheses.

and for the global transition from N_2 to 2U, 2.00 kcal mol^{-1} M^{-1},[43] have direct implications for the urea-dependence of the exchange rate constants for individual amide hydrogen atoms. Exchange through local fluctuations is expected to be urea-independent, exchange through the I_2 state is expected to increase by a factor of ~4-fold at 1.0 M urea, and exchange through the U state is expected to increase by a factor ~30 at 1.0 M urea. Because the *m*-value for the stable monomeric form of $[2\text{-}66]_2$ TR, i.e. L39E [2-66], is 0.64(±0.04) kcal mol^{-1} M^{-1} (unpublished results and see Supplementary Data Figure S1), the total *m*-value for the unfolding of a pair of folded monomers, i.e. 2I, is estimated to be 1.28 kcal mol^{-1} M^{-1}. These data imply that the *m*-value for the N_2 to 2I transition is 0.72 kcal mol^{-1} M^{-1} and within error of that for the I_2 state relative to the N_2 state, 0.78 kcal mol^{-1} M^{-1}. The near identity of the *m*-values for I_2 and 2I implies that the protection in both states would be expected to diminish almost equally at increasing concentrations of urea. The near equality also implies that the net change in buried surface area does not change significantly when I self-associates to form I_2.

An example of the change in HX behavior as a function of the concentration of denaturant for peptide 42-61 is shown in Figure 5(a), and the complete results for all of the peptides and segments examined are shown Table 1. In general, as the concentration of urea was increased from 0 to 1.0 M, the number of amide hydrogen atoms exchanging through the fast process, N_f, increased and the number through the slow process, N_s, decreased. Although the variations in the number of hydrogen atoms exchanging through the two intermediary processes with increasing concentration of urea did not show a consistent change, the errors in the estimated amplitudes precludes an unambiguous assessment of their significance. The observed increase of the apparent rate constant k_{i1} by up to fourfold accompanying an increase in the concentration of urea from 0 to 1.0 M for peptides 21-26 and 26-35, segment 31-47, and peptide 42-61 (Table 1) is in agreement with expectations for exchange through the I_2 species (Scheme 1). The k_{i2} apparent rate constant, which can be monitored at 1 M urea only for the 42-61 peptide, also increased by a factor of ~4. However, as for the amplitudes, the errors in

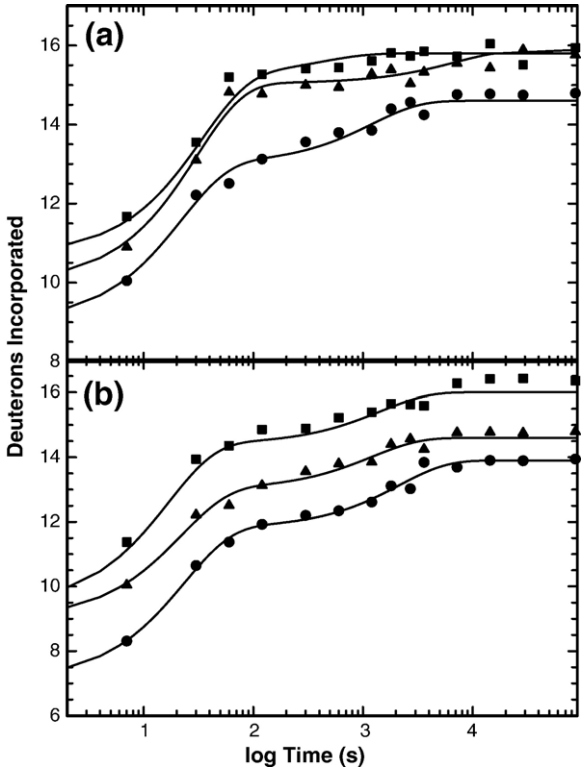

Figure 5. (a) Urea-dependence of the HX properties of peptide 42-61. Urea concentrations for exchange were 0 M (circles), 0.50 M (triangles) and 1.00 M (squares). (b) Protein concentration-dependence of the HX properties of peptide 42-61. Protein concentrations for exchange were 10 μM (squares), 20 μM (triangles) and 50 μM (circles). The log time coordinate enables the display of the full data set out to 24 h. The inflections highlight the various exchange regimes. Exchange conditions were 10 mM phosphate, pH 7.2, 25 °C in ^2H_2O.

the apparent rate constants are sufficiently large to make these conclusions tenuous. The amplitudes of the k_{i2} apparent rate constants for peptides 21-26 and 26-35 and segment 31-47 were below the limits of detection at 1 M urea. Because segment 15-19 exchanges only through the slow, k_s, process, it was not possible to monitor the effect of the concentration of urea on k_{i1} and k_{i2}. The dominance of the fast exchange process for peptides 3-14 and 62-66 also precluded an assessment of the urea dependence of k_{i1} and k_{i2}.

The protein concentration dependence of the amide hydrogen exchange behavior was examined in an effort to gain structural insight into the role of monomeric and dimeric intermediates in exchange. An example of the change in HX behavior as a function of protein concentration is shown in Figure 5(b). As the concentration of protein was increased, the apparent number of amide hydrogen atoms exchanging through the fast, k_f, process decreased and the number exchanging through the slow, k_s, process increased for peptide 26-35, segment 31-47, and peptide 42-61. No reliable trend was observed for the k_{i1} and k_{i2} processes for these peptides or segments. The small number of amide hydrogen atoms in peptide 21-26, the dominance of the fast exchange process for peptides 3-14 and 62-66, and the dominance of the slow exchange process for segment 15-19 made it impossible to monitor the effect of protein concentration on exchange through these various processes.

Although the quality of the data extracted from the fits of the urea and protein concentration-dependence of the exchange profiles for the eight peptides from $[2\text{-}66]_2$ TR was not sufficient to distinguish the contributions from monomeric and dimeric species, it is apparent that an individual peptide can display fast, intermediate and slow exchange behavior. When combined with previous thermodynamic information about the folding free energy surface,[43] this information makes it possible to extract useful insights into the structures of partially folded forms that contribute to the exchange properties of the peptides.

The N and C termini of $[2\text{-}66]_2$TR

Two of the peptic peptides correspond to the termini of $[2\text{-}66]_2$ TR: 3-14, which covers nearly the entire N terminus including the first five residues of the A helix (residues 9–31) and 62-66, which includes the final residues of the C helix (residues 45–63; Figure 2). Although most of the amide hydrogen atoms in both peptides exchange at the fast rate, a small, but discernible fraction exchange *via* the k_{i1} rate (Table 1). This observation is not surprising for peptide 3-14, because residues 2–4 in this segment cannot be visualized in the crystal structure of the full-length protein (Figure 1).[44] The $\sim 10^2$-fold reduction in the rate of exchange for 1–2 amide hydrogen atoms and the vectorial nature of hydrogen bonds in an α-helix, i.e. the NH points towards the N terminus, implies that the N terminus of the A helix is protected in the native state but not in the partially folded state corresponding to the k_{i1} process. The observed fast exchange for all but one of the amide hydrogen atoms in peptide 62-66 (Table 1) shows that this amide hydrogen also exchanges through the k_{i1} process.

The A helix of $[2\text{-}66]_2$TR

The six peptides that encompass various parts of the A helix (residues 9–31) provide a rather detailed view of its behavior. The amide hydrogen exchange properties of the three non-overlapping segments that can be generated by subtracting appropriate pairs of peptic peptides are summarized in Table 1. The exchange properties of segment 15-19 can be highlighted by subtraction of the data for peptide 3-14 from that for peptide 2-19. Although the N terminus of these two peptides differs by one residue, the probable absence of structure at this position implies that the amide hydrogen at alanine 2 should experience fast exchange. Therefore, after the MS data from peptide 3-14 were subtracted from the data for peptide 2-19, an additional deuteron was subtracted from every datum point to reflect the

expected >99% exchange at alanine 2 by the first time-point, 7 s. As shown in Table 1, segment 15-19, which lies completely within the A helix, is very resistant to exchange. The exchange rates of at least three of the four amide hydrogen atoms are retarded more than ~10^4-fold, regardless of the concentrations of protein and urea examined. Because the number of amide hydrogen atoms is estimated from the amplitudes of the fits of the exchange data to a sum of two exponentials, this value should be considered as an equivalent number of hydrogen atoms and not necessarily as a precise measure of the number of amide hydrogen atoms exchanged in a peptide with multiple exchange sites.

Peptide 21-26, which lies completely within the A helix, displays more complex exchange behavior. In the absence of urea, the equivalent of two amide hydrogen atoms exchange rapidly, a third *via* the k_{i1} process, a fourth via the k_{i2} process and a fifth *via* the slow process. Peptide 26-35, which encompasses the six C-terminal residues of the A helix as well as four residues from the linker with helix B, also displays complex exchange properties. Six or seven amide hydrogen equivalents exchange *via* the fast process, one or two *via* the k_{i1} process, perhaps one *via* the k_{i2} process, and one or two *via* the slow process.

Thus, the A helix displays a heterogeneous amide hydrogen exchange behavior, with strong protection for a short, continuous segment, residues 15–19, under a variety of conditions and multi-layered protection patterns for the other residues.

The B and C helices of [2-66]$_2$ TR

Only one rather long peptide, 22-47, could be generated that includes the short B helix (residues 35–42). However, this peptide also encompasses the last nine residues of the A helix, the linkers between the three helices, and the first three residues of the C helix. Fortunately, the production of peptide 22-30 enabled a more refined view of the behavior of the B helix by subtraction of its exchange properties from those of the 22-47 peptide (Table 1). The exchange properties of most of the C helix, residues 45–63, could be examined in the 42-61 peptide (Table 1).

Segment 31–47 displayed a complex exchange pattern involving fast, intermediate and slow exchange behavior: at 10–20 μM protein, the equivalent of amide hydrogen atoms exchange *via* the fast process, about five *via* the k_{i1} process, two *via* the k_{i2} process, and two *via* the slow process. Interestingly, as the concentration of protein was raised to 50 μM, the number exchanging through the fast process decreased to about three amide hydrogen equivalents and the number exchanging through the slow process increased correspondingly to four equivalents; the number exchanging through the k_{i1} and k_{i2} processes were unaffected. Peptide 42-61 also displayed a complex protection pattern reflecting all three time regimes: about nine hydrogen equivalents exchanged through the fast process, four or five through the k_{i1} process, one or two through the k_{i2} process, and three or four through the slow process.

Exchange on long time-scales

To gain more insight on the amide hydrogen atoms of [2-66]$_2$ TR that were resistant to exchange after 24 h of incubation in ^2H_2O, the exchange reaction was performed for four weeks at 20 μM protein under the same conditions, pH 7.2 and 25 °C. As shown in Table 2, the exchange data for all but one of the peptides could be fit to a third exponential, whose rate constant indicates ~10^6-fold slower exchange than for a solvent-exposed amide hydrogen. Segment 15–19 contains two or three strongly protected amide hydrogen equivalents, peptide 21-26 contains one, peptide 26-35 contains one, segment 31-47 contains two, and peptide 42-61 contains one or two strongly protected amide hydrogen equivalents. Surprisingly, peptide 42-61, and possibly segment 15-19, each contain a single protected amide hydrogen equivalent after four weeks of incubation in ^2H_2O. As will be described below, the estimated protection factor for these two amide hydrogen atoms exceeds that estimated from the global stability, potentially implying super-protection.[45–47] Although the relative error in estimating the number of super-protected amides is high (Table 2), the reproducibility of these results enhances the likelihood that at least one, if not both, of these amide hydrogen atoms may be super-protected.

Table 2. H/D-exchange parameters for [2-66]$_2$ TR on long time scales

Peptide	N_{fast}	k_f[a] (s^{-1})	N_{int1}	k_{i1} (s^{-1})	N_{int2}	k_{i2} (s^{-1})	N_{slow}	k_s (s^{-1})	$N_{superprotected}$
3-14	8.4 (0.4)	3.1	1.3 (0.2)	$2.0\ (1.0)\times10^{-2}$	nd[b]	nd	nd	nd	0.3 (0.2)
15-19	0.3 (0.2)	3.6	nd	nd	nd	nd	3.2 (0.3)	$6.1\ (3.3)\times10^{-6}$	0.5 (0.3)
21-26	1.5 (0.3)	1.2	1.4 (0.2)	$2.0\ (1.3)\times10^{-2}$	0.8 (0.1)	$2.1\ (1.0)\times10^{-4}$	1.3 (0.2)	$2.2\ (1.3)\times10^{-6}$	nd
26-35	6.2 (0.4)	3.3	0.6 (0.2)	$1.5\ (0.8)\times10^{-2}$	0.5 (0.2)	$5.9\ (3.2)\times10^{-4}$	1.4 (0.3)	$2.4\ (0.8)\times10^{-6}$	0.3 (0.3)
31-47	4.5 (0.4)	2.6	4.5 (0.3)	$2.1\ (0.7)\times10^{-2}$	2.5 (0.2)	$5.9\ (3.2)\times10^{-4}$	2.4 (0.3)	$6.1\ (3.3)\times10^{-6}$	0.1 (0.2)
42-61	9.3 (0.5)	3.2	3.7 (0.4)	$1.8\ (1.2)\times10^{-2}$	1.5 (0.2)	$5.6\ (2.4)\times10^{-4}$	2.5 (0.4)	$5.1\ (2.6)\times10^{-6}$	1.0 (0.5)
62-66	2.8 (0.3)	4.2	0.8 (0.3)	$2.0\ (0.8)\times10^{-2}$	nd	nd	0.3 (0.1)	$8.8\ (5.0)\times10^{-7}$	0.1 (0.3)

Conditions for all experiments were 20 μM [2-66]$_2$ TR, 10 mM phosphate, pH 7.2, 25 °C. Errors are listed in parentheses.

[a] The values for the k_f exchange rates were determined by averaging the exchange rates for the individual amino acids in the peptide as determined by the method of Bai and Englander[41] for a free, unstructured amino acid.

[b] Not detectable.

Theoretical Analysis

Native topology-based simulations of the folding of $[8\text{-}66]_2$ TR

The intertwined structure of the Trp repressor (Figure 1) offers an opportunity for self-association *via* a segment-swapping mechanism.[31,32] In this mechanism, interactions similar to those that define the dimer interface can be formed in an intramolecular fashion to stabilize a monomeric form. Accordingly, the native interactions were symmetrized in the folding simulation to allow unswapping of helices A and A′. These interactions should not be treated as purely non-native interactions (although they do not stabilize the dimeric TR), because they follow the same energetic and topological arguments whether they are formed in an intra- or an intermolecular fashion.

The native topology-based simulations on $[8\text{-}66]_2$ TR detected a dimeric intermediate and, when unswapping of helices A and A′ was allowed, a monomeric intermediate as well. A trajectory that monitors the total number of contacts during the simulation (i.e. monomeric contacts, Q_A and Q_B, as well as interfacial contacts, $Q_{Interface}$), illustrates the existence of an on-pathway intermediate. While the dimeric intermediate is seen in a time-series analysis of the trajectories (Figure 6(a)), it is not observed clearly in the free energy profile when plotted along a single-order parameter such as Q_{Total}, the radius of gyration, or the distance between the two monomers (data not shown). This first-order analysis probably fails to detect these intermediates because the intermediates have moderate stability and because a single reaction coordinate is not sufficient to describe the complex dynamics of the bimolecular folding reaction. The complexity reflects an overlap and interference between the coordinates of the folding reaction of the individual subunits, their binding reaction, and the various transition state ensembles.

Figure 6(b) and (c) show four-dimensional plots of the free energy landscape of the folding/association of the $[8\text{-}66]_2$ TR projected onto three reaction coordinates. One reaction coordinate indicates the degree of folding of a single monomer, another reaction coordinate indicates the degree of binding (i.e. formation of the interface), and the third measures the separation distance between the two monomers. Both free energy landscapes obtained from the Gō (Figure 6(b)) and symmetrized-Gō (Figure 6(c)) simulations indicate the existence of the dimeric intermediate, I_2. In the dimeric intermediate, the two monomers are partially folded and the interface between them is partially formed. Plotting the number of interfacial contacts that do not involve helices C and C′ (i.e. interfacial contacts between helices A, A′, B, and B′) *versus* those that

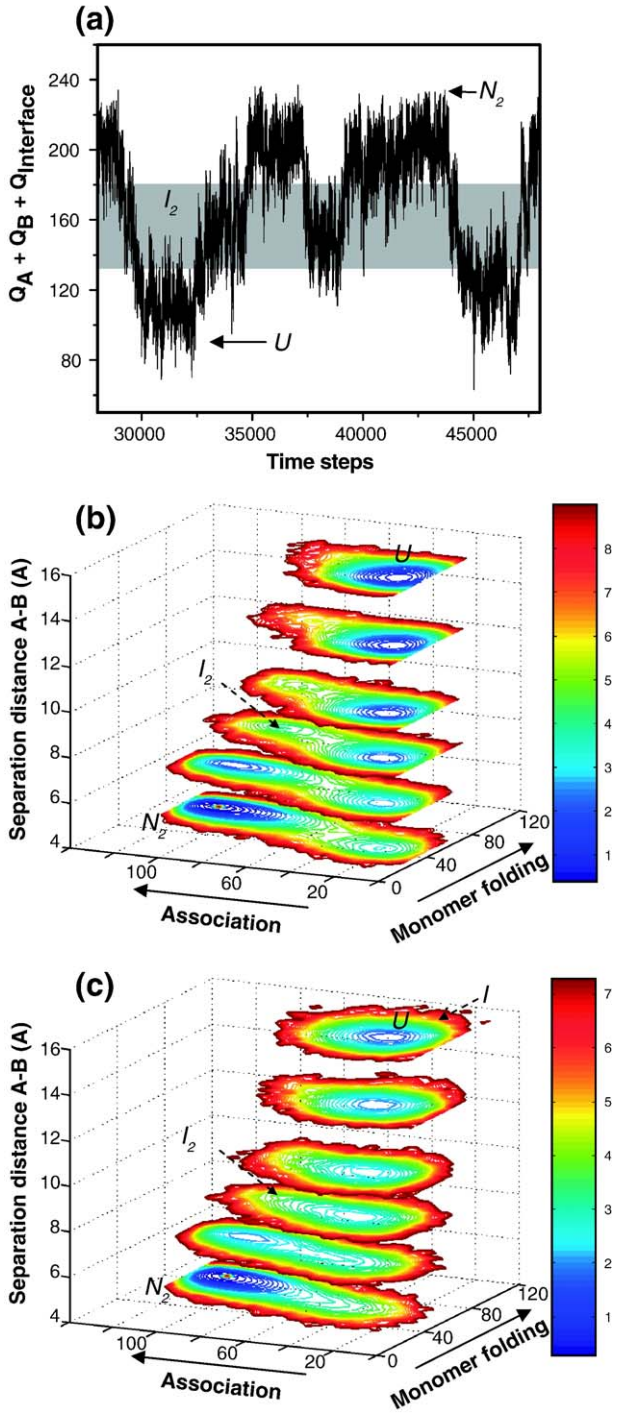

Figure 6. The folding/association mechanism of the $[8\text{-}66]_2$ Trp repressor from Gō and symmetrized-Gō models. (a) A typical trajectory of the folding of $[8\text{-}66]_2$ TR simulated with the symmetrized-Gō model. The trajectory illustrates the unfolded and folded states (*U* and N_2, respectively) and the on-pathway dimeric intermediate (I_2). Similar trajectories are observed when simulating using the Gō model. The monomeric intermediate (I) is not seen in this trajectory, as it is relatively unstable around the folding temperature. (b) and (c) The four-dimensional free energy surfaces for folding and association of $[8\text{-}66]_2$ TR at the folding temperature using Gō and symmetrized-Gō models, respectively. The reaction coordinates in this free energy plot are the number of contacts of a single subunit (i.e. monomer folding), the number of contacts in the dimer interface (i.e. association), and the separation distance between the two subunits.

involve only interfacial contacts between the residues in helices C and C′ and any other residues may indicate which part of the interface is formed first and will reflect the existence of coupling between the packing of the interface regions (Figure 7(a)). This plot shows that the interface is formed initially by packing between helices A, A′, B, and B′. The contact map (Figure 7(b)) of the dimeric intermediate illustrates that helices A and B from the two monomers interact while the intra- and intermolecular interactions with helices C and C′ are very transient (Figure 7(b)). Similar properties of the dimeric intermediate are found when simulating with the model that includes native contacts only.

A monomeric intermediate, I, is detected only when unswapping of helices A and A′ is taken into account. In the monomeric intermediate, helix A interacts with helix C by using similar interactions that exist in the interface of the dimer when helix A interacts with helix C′ (Figure 8). Intramolecular interactions between helices A and C introduce strain and, consequently, helix B and the N-terminal of helix C are less structured than in the dimeric structure. To illustrate the monomeric intermediate (Figure 8(b)), the monomeric contacts that are defined in the symmetrized-Gō (but are not part of the simple Gō ,) are plotted *versus* the monomeric contacts that are defined in the simple Gō model. It is clear that an almost fully folded monomer can be found with a significant number of unswapped contacts. The participation of many unswapped contacts at this state suggests a folded monomer that has almost all the helices formed but packed in an intramolecular fashion. The monomeric intermediate is populated only partially under the conditions of these simulations but is expected to be more highly populated at lower concentrations of protein by the law of mass action. To better characterize the monomeric intermediate, the folding of a monomeric $[8\text{-}66]_2$ TR was simulated with the use of the monomeric contacts that are defined by the symmetrized-Gō model. In these simulations, which may represent a very dilute solution, the unswapped monomer was stabilized by interactions between helices A and C (Figure 8(c)).

A detailed comparison of the theoretical and experimental results is provided in Discussion.

Discussion

The folding energy landscape of $[2\text{-}66]_2$ TR

Kinetic and equilibrium studies of $[2\text{-}66]_2$ TR[38,39,43] have shown that the native dimer, N_2, is in equilibrium with a dimeric intermediate, I_2, a monomeric intermediate, I, and the unfolded monomer. Relative to the U state, the standard state free energy of the native dimeric state is –13.3 kcal mol^{-1}.[43] Chevron analysis of the kinetic folding data revealed that the dimeric intermediate is 2.5(±0.5) kcal mol^{-1} less stable than N_2 in the absence of denaturant.[38] At 20 μM $[2\text{-}66]_2$ TR, where most of these studies were conducted, the apparent stabilities of N_2 and I_2 *versus* U are predicted to be 7.0 kcal mol^{-1} and 4.5 kcal mol^{-1} ($\Delta\Delta G = -RT$ ln[protein] = 6.3 kcal mol^{-1}). A monomeric version of $[2\text{-}66]_2$ TR, constructed by replacing an interfacial leucine 39 with glutamic acid, L39E, was determined to be 1.1 kcal mol^{-1} more stable than the U state (unpublished results; and see Supplementary Data Figure S1). For the reaction shown in Scheme 1 (implying a dimeric reference state), the free energy of 2I is 2.2 kcal mol^{-1} lower than 2U. This information can be used to construct the reaction coordinate diagram shown in Figure 9. The features of the energy landscape described by this diagram

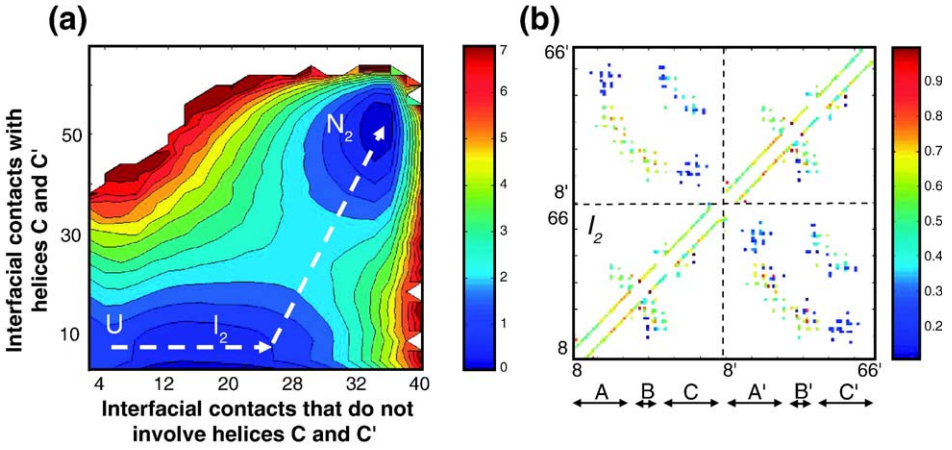

Figure 7. The dimeric intermediate in the folding/association of $[8\text{-}66]_2$ TR obtained from the native topology-based models. (a) Free energy surface of the folding of $[8\text{-}66]_2$ TR projected on two reaction coordinates. The first reaction coordinate measures the interfacial contacts between helices A, B, A′, and B′, and the second measures all the interfacial contacts with helices C and C′. It indicates that helix C and C′ are not interacting intermolecularly in the dimeric intermediate. (b) The contact probability map for the dimeric intermediate. The upper and lower half show the contact probabilities of the dimeric intermediate from the Gō and symmetrized-Gō models, respectively. The flexibility of helices C and C′ is indicated by the lower probabilities for the interfacial contacts that involved the residues of helices C and C′.

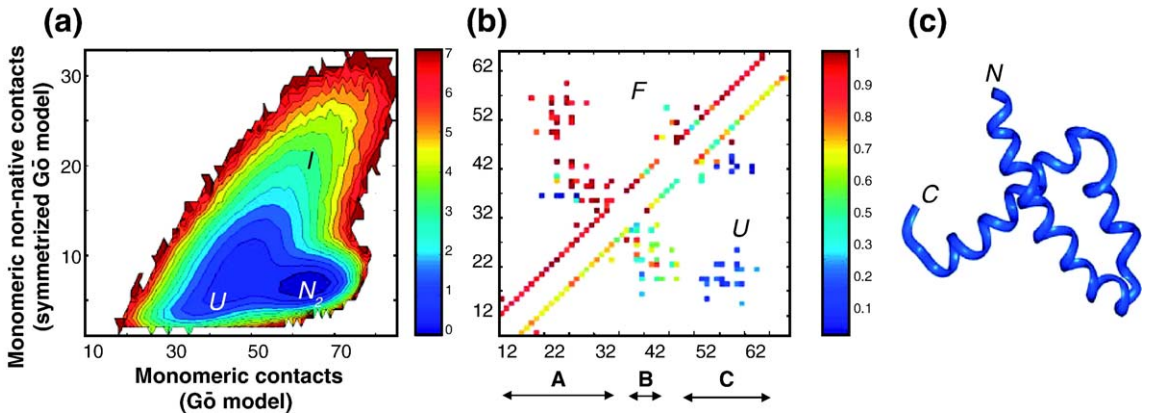

Figure 8. The monomeric intermediate in the folding/association of $[8\text{-}66]_2$ TR obtained from the native topology-based models. (a) Free energy surface of the folding of $[8\text{-}66]_2$ TR projected on the native contact of the monomer (as defined in the Gō model) and the monomeric non-native contacts (as defined in the symmetrized-Gō model). This plot shows that an almost fully folded monomer can be found with a large amount of non-native contacts that result in packing of the α-helices. (b) The contact probability at the folded (F) and unfolded (U) state of monomeric TR (residues 8–66) from simulations of an isolated monomer using the symmetrized-Gō potential. (c) A typical structure of folded monomeric [8-66] TR.

serve as a basis for the interpretation of the apparent protection factors derived from the HX data shown in Tables 1 and 2.

Structural insights into partially folded states in $[2\text{-}66]_2$ TR

The Int1 process

The ~50-fold protection against exchange for amide hydrogen atoms that exchange through the Int1 process at 20 μM protein (Table 1) implies that the partially folded state(s) through which exchange occurs is ~2.3 kcal mol^{-1} less stable than the native state. Although a retardation of this magnitude in other proteins has been attributed to local fluctuations in the structure,[42,47,48] the existence of a dimeric intermediate state for $[2\text{-}66]_2$ TR, I_2, with nearly the same estimated stability relative to the native dimeric state, N_2, +2.5(±0.5) kcal mol^{-1},[38] offers another vehicle for exchange for at least a fraction of the hydrogen atoms that exchange through the Int1 process. Thus, it is likely that a subset of the amide hydrogen atoms in peptides 21-26 and 26-35, segment 31–47, and peptide 42-61 are protected in the native state but not in the I_2 folding intermediate for $[2\text{-}66]_2$ TR. Radford and her colleagues have drawn similar conclusions about the presence of partially folded states in their HX studies of small, monomeric bacterial immunity proteins.[49] The conclusion that at least some of these Int1 process amide hydrogen atoms stabilize the A, B, and C helices in N_2 but not in I_2 is supported by the previous observation that the I_2 species has ~25% less far-UV CD signal than the N_2 state.[38]

The native topology-based model simulations suggest that helices A and B from different subunits in I_2 interact while helix C is more flexible. This dimeric intermediate was detected also when unswapping was modeled to account for intra- and intermolecular interactions between helices A and C. These structural features are seen in the protection factor calculated from the simulations (Figures 10 and 11). Helices A and B have a larger fraction of protected residues than helix C, and the number of protected residues is decreased in all three peptides compared to the native dimeric state.

The Int2 process

The $\sim 2 \times 10^3$ protection against exchange observed for the amide hydrogen atoms that exchange through the Int2 process at 20 μM protein (Table 1) implies an effective free energy of +4.3 kcal mol^{-1} relative to the N_2 state. This estimated free energy is significantly more positive than that afforded by the

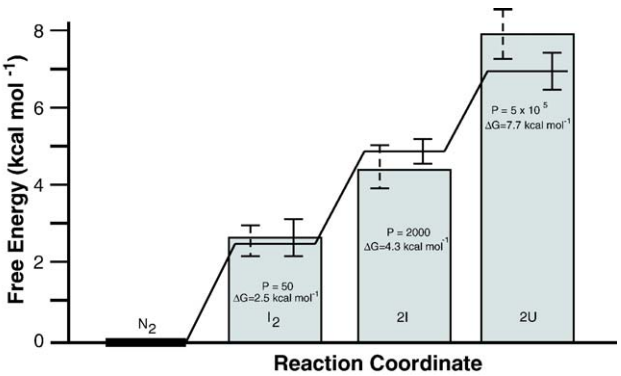

Figure 9. A free energy diagram for the folding of $[2\text{-}66]_2$ TR. Four states are proposed to exist; the native dimer N_2, a dimeric intermediate I_2, a monomeric intermediate I, and the unfolded monomer U. Continuous lines indicate the relative free energies at equilibrium and at 20 μM $[2\text{-}66]_2$ TR as determined previously.[38,43] Gray bars indicate the average protection factors and implied free energies obtained for each amide hydrogen exchange rate averaged over all peptides under exchange conditions of 20 μM protein.

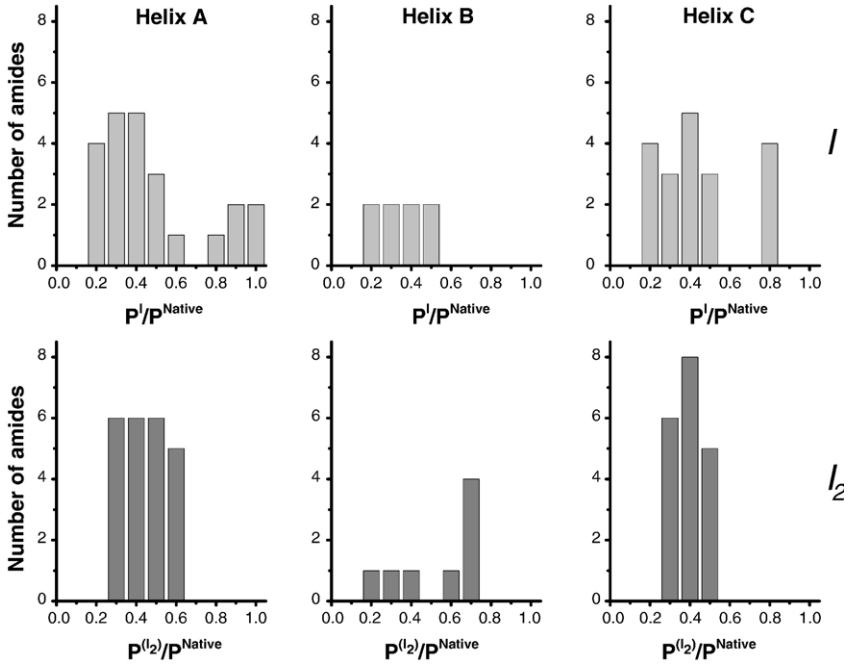

Figure 10. Histograms of the relative protection factors of the residues comprising helices A, B, and C in the monomeric and dimeric intermediates, I and I_2. The protection factors are estimated from the native topology-based model simulations supplemented by swapped/unswapped interhelical interactions. The protection factor was assumed to be proportional to the number of interactions that any residue forms in any state (equation (5)).

I_2 state, but less than that expected for the U state of $[2\text{-}66]_2$ TR at 20 μM, 7.0 kcal mol^{-1}. These amide hydrogen atoms could exchange through a combination of global unfolding and local fluctuations,[49] or through a partially folded state whose free energy is bounded by the I_2 and U states. Given the observation of another slow-exchange reaction, k_s, whose implied stability is very similar to that for the global unfolding reaction (see below), the latter explanation is more likely. This putative additional state could reflect another, as yet undetected, dimeric species or a folded monomeric state that is competent to form the I_2 state *via* a bimolecular docking reaction.

The aforementioned L39E monomeric version of $[2\text{-}66]_2$ TR is a candidate for such a higher-energy species. The estimated free energy of a pair of monomers is 2.2 kcal mol^{-1} lower than the U state or, at 20 μM protein, 4.8 kcal mol^{-1} higher than the N_2 state. This value is in remarkable agreement with that estimated for the additional protection revealed by the Int2 process, 4.3 kcal mol^{-1}. The retention of

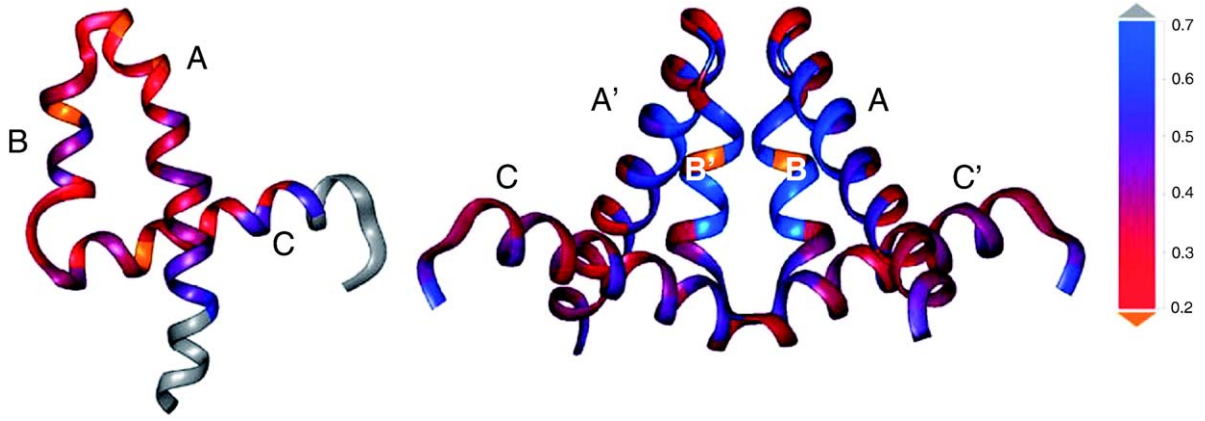

Figure 11. The relative protection factors in the monomeric I (left), and dimeric I_2 (right) intermediates of $[8\text{-}66]_2$ TR as calculated from the native topology-based simulations. Relative protection factors <0.2 or >0.7 are colored orange and gray, respectively. The dimeric intermediate displays much more protection than the monomeric intermediate (the flexible termini in I have high relative protection factors but their absolute protection factors are low in comparison to those of the helices). The relative protection factors illustrate the interactions between helices A and C and the strain in helix B in state I as well as the stabilization of the dimeric interface between helices A, A′, B, and B′ in state I_2.

~50% of the native far-UV ellipticity by L39E [2-66] TR is consistent with the loss of amide hydrogen equivalents through the k_{i2} process and the eventual exchange through the k_s process (see below).

The intertwined nature of the dimeric TR core implies that a structural rearrangement must occur in the monomer to provide the observed protection in all three helices *via* the subsequent k_s process. The helix-swapping required to achieve this rearrangement would be analogous to domain and segment swapping reported for other multimeric proteins.[31] The simulations of $[8\text{-}66]_2$ TR that allow interactions between helices A and C within the same or between different subunits support the existence of a monomeric species. In this high-energy intermediate, helices A and C interact using the interactions that stabilize the native state when these interaction are formed intermolecularly after helix swapping. This structural feature is seen in the simulated protection factors for state I. In this state, helices A and C have more residues with higher protection than helix B. This pattern of protection factors for the intermediate I is different from those calculated for the intermediate I_2 (Figure 10). Helix B is, in particular, much less protected in I than in I_2. The latter observation may reflect the need to unravel the B helix to some extent to allow segment swapping in the monomer. In addition, some of the low relative protection factors observed for several residues in the B as well as in the A and C helices in the I state stem from the fact that these residues gain additional interactions from the interface with the other subunit in the I_2 state.

The k_s process

The strong protection, $\sim 5 \times 10^5$ retardation relative to solvent-exposed amide hydrogen atoms, displayed by a sub-set of hydrogen atoms from segment 15–19, peptides 21-26 and 26-35, segment 31–47, and peptide 42-61 (Table 1), implies that these hydrogen atoms exchange *via* a global unfolding mechanism through the U state. The stability predicted on the basis of the HX MS experiment, 7.7 kcal mol^{-1}, and that measured by urea denaturation, 7.0 kcal mol^{-1}, are remarkable similar and within the estimated errors.

The ultra-slow process

Surprisingly, a single amide hydrogen equivalent in peptide 42-61 and, possibly, another in segment 15–19, remain protected after four weeks of incubation at pH 7.2 and 25 °C (Table 2). The absence of protection after the same period of time in segment 31–47 means that the amide hydrogen(s) displaying this ultra-slow exchange must reside in residues 48–61, i.e. entirely within the C helix. The putative protected amide hydrogen in segment 15–19 is contained in the A helix. The stability estimated to account for this level of retardation, >8.0 kcal mol^{-1}, exceeds that of the global unfolding reaction. This situation implies that one or two amide hydrogen atoms are protected against exchange in an otherwise unfolded state of $[2\text{-}66]_2$ TR. Such "super-protection" has been observed in other proteins,[45–47] and has been interpreted to reflect the presence of non-random, perhaps residual, structure in chemically denatured proteins.

Although the possibility of super-protection in these two distal segments of the protein may be due to local elements of structure that ultimately form two of the three helices in $[2\text{-}66]_2$ TR, it is intriguing to consider that both segments might experience a mutual stabilization in a long-range hydrophobic cluster. Schwalbe, Dobson, and their colleagues have demonstrated long-range, cooperative interactions between distal segments of lysozyme.[50] Either local or non-local residual structure could play an important role in directing the earliest folding events of $[2\text{-}66]_2$ TR.

Dependence of the distribution of amide hydrogen atoms exchanging through the fast, intermediate and slow processes on the concentrations of denaturant and protein

With the exception of the amide hydrogens that exchange through the fast process, all of the remaining amide hydrogen atoms in peptides or segments representing all three helices in $[2\text{-}66]_2$ TR experience protection factors that are comparable to or exceed the stabilities of previously identified dimeric or monomeric intermediates. Thus, the generally observed increase of the number of amide hydrogen atoms exchanging through the fast process at increasing concentrations of urea is consistent with the enhancement of the populations of partially folded states. Conversely, the decrease in the number of fast-exchanging amide hydrogen atoms at increasing concentrations of protein is consistent with the predicted shift from the monomeric to dimeric forms by the law of mass action. Unfortunately, the small apparent number of amide hydrogen atoms exchanging through the Int1 and Int2 processes for most of the peptides and the magnitude of their estimated errors obscures the expected changes in rate constants in response to changes in the concentration of urea or protein. Thus, it was not possible to determine whether the species responsible are dimeric or monomeric.

The heterogeneous exchange behavior of the A, B, and C helices

The observation that the three peptides or segments derived from the A helix show distinctly different exchange properties demonstrates that this helix does not behave as a uniform unfolding module.[51] Segment 15–19 exchanges only through slow or ultra-slow, k_{us}, processes, while peptides 21-26 and 26-35 exchange through fast, intermediate and slow processes. The B and C helices, similar to the A helix, exchange through fast, intermediate, and slow processes; one amide hydrogen in the C helix exchanges through the ultra-slow process.

Given the relationship between the Int1 process and exchange through the I_2 intermediate discussed above, it is reasonable to conclude that amide hydrogen atoms exchanging through the slower processes, i.e. Int2, k_s, and k_{us}, reflect partial helical structure (or at least non-random structure) retained for all three helices in the I_2 and, possibly, other higher-energy states. The variations in the protection factors for the three helices are observed in the native topology-based model simulations, where the protection factors of each residue in a given state, i.e. the folded state and the monomeric and dimeric intermediates, is estimated by the degree of formation of its non-bonded contacts relative to the unfolded state (see Materials and Methods). The histograms of the calculated protection factors (Figures 10 and 11) support the picture obtained from the HX MS data that the helices do not act as a single unit.

The native-state HX behavior of full-length TR was examined by Jardetzky and colleagues using HX NMR methodology.[52] These studies indicated that helices A to C as well as the F helix contain residues whose amide hydrogen atoms exchange on slow time-scales. Although the limited exchange times employed did not allow a full exploration of the energy landscape for the more stable, full-length protein, this study showed that the helices do not exchange as a single unit. As for the HX MS studies on $[2\text{-}66]_2$ TR, the exchange rates of the amide hydrogen atoms in these helices can vary by comparable orders of magnitude. Thus, the "foldon hypothesis" proposed for cytochrome *c* by Englander and his colleagues[23] does not hold for full-length TR or for $[2\text{-}66]_2$ TR.

Structural implications of the HX data for the folding mechanism of $[2\text{-}66]_2$ TR

The equilibrium folding surfaces of $[2\text{-}66]_2$ TR determined from simulations and HX MS measurements are in good agreement. Both theoretical and experimental approaches are consistent with the existence of monomeric and dimeric intermediates that reflect a progressive development of structure during folding. The agreement supports the presumption, validated previously for other monomeric and dimeric proteins,[3,5–7,9,10] that the native contacts are sufficient to capture essential features of complex folding reactions. However, the simplicity of the model limits the accuracy of the calculated stabilities of the intermediates and the associated protection factors.

Given the caveat that the partially folded forms detected in the hydrogen exchange experiments under equilibrium conditions are not necessarily those that appear during the actual folding reaction, a conceptual picture of the folding reaction emerges. The reaction begins from an unfolded state that contains non-random, perhaps residual, native-like structure in two key segments of a single TR chain. These segments include short stretches of the A and C helices, which could be the helices that dock on each other in the partner subunit in the native dimeric form, namely, 15–19 and 48–54. These unfolded conformers, containing either local or non-local residual structure, collapse within a few milliseconds to monomeric intermediates with substantial helical structure. The mean residue ellipticity of a stable model for the monomeric intermediate is ~55% of that for the native dimeric form.[36] The observation that segments containing all three helices display the Int2 exchange process implies that segments of all three helices may be present in this burst-phase species. The partially folded monomers then associate to form the I_2 intermediate and, in doing so, increase the helical structure to about 75% of that found in the native dimeric state. Finally, the I_2 intermediate undergoes the rate-limiting transformation to the native state and acquires the full complement of helical structure. The observation of the Int1 process for segments containing all three helices suggests that all of the helical regions increase in size when the dimeric precursor matures to the native state.

The progressive increase in the helical structure for each of the segments containing the three helices in native $[2\text{-}66]_2$ TR during the folding reaction is not consistent with the predictions of the framework model of folding.[53] In this model, the elements of secondary structure are postulated to form independently and early in the reaction, and guide their complex assembly into the native conformation. The results are, however, consistent with a hydrophobic collapse model,[53,54] in which small, possibly non-native, elements of structure could guide the concerted condensation of the chains and the formation of helices in non-native monomeric species. Collisions between these non-native forms results in the formation of a dimeric intermediate, mediated by a segment-swapping reaction that increases the length of each helix. The maturation of the dimeric intermediate to the native conformation, which allows all three helices to reach their ultimate length, represents the rate-limiting step in the folding of $[2\text{-}66]_2$TR.

Materials and Methods

Materials

The expression and purification of $[2\text{-}66]_2$ TR has been described.[43] Ultrapure urea was purchased from ICN Biomedicals Inc. All other reagents were purchased from Sigma-Aldrich and were reagent grade.

Mass spectrometry

Matrix-assisted laser desorption ionization (MALDI) time-of-flight mass spectra were obtained on a Micromass M@LDI-LR instrument in the reflectron mode. A sampling rate of 0.5 ns was used with a source voltage of 15 kV, a pulse voltage of 2175.0 V and a micro-channel plate voltage of 1850 V. The matrix suppression delay was set

to 400 mass units and delayed extraction was used with a nominal pulse delay of 500 ns. Between 50 and a 100 spectra were obtained and averaged for every sample. The matrix was formed by drying 10 mg ml^{-1} α-cyano-4-hydroxycinnaminic acid dissolved in 70% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid (TFA), pH 2.5, on a MALDI target plate, and the target plates were chilled on ice before use. Lock mass calibration was conducted for each sample using angiotensin I (1296.6853 Da), angiotensin II (1060.5692 Da), and Sub-P (1347.7200 Da). All analyzed masses were monoisotopic MH^+ masses.

Native-state HX exchange and digestion conditions

Native-state exchange experiments were initiated by a 20-fold dilution of $[2\text{-}66]_2$ TR from H_2O buffer (10 mM phosphate, pH 7.2) into ^2H_2O buffer (10 mM deuterated phosphate, pH 7.2) at 25 °C. In experiments done in the presence of urea, the ^2H_2O buffer contained undeuterated urea while the ^2H_2O buffer contained deuterated, d_4-urea. After various exchange times, a 30 μl sample was removed and quenched by a 1:1 (v/v) dilution into an H_2O quench solution containing 1% TFA, pH 2.5. After quenching, the samples were immediately flash-frozen in liquid nitrogen and stored at –70 °C until digestion and analysis by MALDI MS.

Pepsin digestion was done by mixing pepsin and $[2\text{-}66]_2$ TR in a 2:1 and 1:1 (w/w) ratio of pepsin to protein immediately after $[2\text{-}66]_2$ TR had been thawed. Pepsin was dissolved in H_2O with 0.2% TFA, pH 2.5 and chilled on ice before use. Although the digestions resulting from the two different pepsin to protein ratios yielded 11 identical peptides, the 2:1 pepsin to protein ratio yielded better centroid fits to isotopic envelopes for smaller peptides (<2200 Da); at the 1:1 pepsin to protein ratio, the analysis of the larger peptides was improved. Immediately after digestion was initiated, 0.5 μl of the mixture was spotted on the chilled MALDI target plate pre-loaded with 0.5 μl of CHCA matrix. Digestion occurred as the mixture dried on the plate, lasting ~5 min. Immediately after drying, the target plate was transferred to the mass spectrometer for analysis.

Data analysis

Micromass MassLynx Software was used to measure the extent of deuterium incorporation by determining the centroid of the isotopic envelope of the peptides from the digested $[2\text{-}66]_2$ TR. The centroid was then compared to the centroids of fully undeuterated and fully deuterated $[2\text{-}66]_2$ TR to correct for deuterium loss during analysis. A fully deuterated sample was prepared by incubating $[2\text{-}66]_2$ TR for 24 h in ^2H_2O and 8 M d_4-urea. This sample was then subjected to two rounds of lyophilization and incubation in a d_4-urea, ^2H_2O solution for 24 h. After being redissolved in ^2H_2O for a third time, the urea was removed by dialysis against ^2H_2O containing 10 mM deuterated phosphate, pH 7.2. Deuterium incorporation, D, was then calculated by:

$$D = \frac{m - m_{0\%}}{m_{100\%} - m_{0\%}} \times N \qquad (3)$$

where N is the total number of exchangeable amide hydrogen atoms in the peptide, m is the mass of the analyzed sample, $m_{0\%}$ is the mass of undeuterated sample and $m_{100\%}$ is the mass of the fully deuterated sample. The extent of back-exchange, based on the number of amide hydrogen atoms that remain in the fully deuterated sample after sample workup (digestion and spotting onto a MALDI plate)[27,40] was fairly constant over all peptides analyzed in the fully deuterated sample, approximately 30(±5)%.

Exchangeable amide hydrogen atoms were classified into three groups; fast, intermediate, and slow, based on their apparent exchange rates. Operationally, the fast-exchanging hydrogen atoms were those replaced by deuterium within the dead-time of the manual mixing experiments, 7 s. The apparent exchange rate constants, k_i, and the apparent number of amide hydrogen atomss in a given peptide that exchanged by these rates, N_i, were determined by fitting the exchange data as a function of time (from 7 s to four weeks) to the equation:

$$N_t = N_0 = \Sigma N_i(1 - e^{k_i t}) \qquad (4)$$

where N_t is the total number of amide hydrogen atoms exchanged at time t, and N_0 is the number exchanged at time zero. Data were analyzed using the in-house software package Savuka 5.2.

In almost all of the experiments, two exponentials were adequate to describe the exchange data in this intermediate regime for samples assayed up to 24 h. The fitted rate constants were ~10^2-fold and ~10^4-fold smaller than that expected for an unprotected amide hydrogen. Fast-exchanging amide hydrogen atoms were assumed to exchange at a rate, k_f, equal to that determined by the method of Englander and his colleagues for the intrinsic exchange rate for unprotected amides.[41] Under the conditions used in these experiments, pH 7.2 and 25 °C, the average rate constant was assumed to be 3.1 s^{-1}. The number of amide hydrogen atoms exchanging in the fast process was calculated from equation (2) using the fitted rate constants and extrapolation to 1 s, where >95% of the typical fast-exchanging amide hydrogen atoms would have already exchanged. The number of amide hydrogen atoms that exchange *via* the slow rate, k_s, was calculated by subtracting the number of amide hydrogen atoms exchanging *via* the fast and intermediate rates from the total number of exchangeable amide hydrogen atoms in a given peptide.

As the concentration of urea in which the samples were incubated was increased, the exchange data for some of the peptides could not be fit to an exponential in the intermediate time regime because all the amide hydrogen atoms exchanged only *via* the fast and slow exchange rates. In these instances, the data for the first few hours were fit to a linear function (reflecting the very slow rate of exchange) and the number of fast-exchanging amide hydrogen atoms determined from the extrapolated value at 1 s. The number of slowly exchanging amide hydrogen atoms was found by subtracting the number of fast-exchanging amides from the total number of exchangeable amides in the given peptide.

To examine in more detail the kinetic properties of the slow-exchanging amide hydrogen atoms, samples containing $[2\text{-}66]_2$ TR at several concentrations of protein were incubated for four weeks under the above conditions. The exchange data for peptides containing slow-exchanging hydrogen atoms could be fit to three apparent rate constants that are ~10^2-fold, ~10^4-fold, and ~10^6-fold smaller than that expected for an unprotected amide hydrogen. A representation of the data fitting method is shown in Figure 4.

Simulation models

The folding mechanism of the dimerization domain of *E. coli* Trp repressor, $[8\text{-}66]_2$ TR, was studied by simulation using simple models. The first model is based on the native topology of the protein where only native interactions are included (Gō model).[5] The native topology-based model has been applied recently to study protein–protein interactions in oligomeric proteins.[6,7] In these studies, the association mechanisms and the structure of the binding transition state ensemble successfully reproduced the binding behavior found in the laboratory. The native topology-based model for $[8\text{-}66]_2$ TR was constructed on the basis of the crystal structure of Trp repressor (PDB entry 2wrp, which includes residues 5–108). In these simulations residues 5–7 were excluded because they are flexible and do not participate in helical packing. In the second model, unswapping of helices A and A′ is allowed in addition to the native interactions that define helix-swapping between the two subunits. This model, which is called the symmetrized-Gō model, has been used recently to study protein association *via* domain-swapping by allowing each monomeric contact to be formed also in an intermolecular fashion.[55–57] The symmetrized-Gō model is relevant to any homo-oligomeric protein as intra- and intermolecular interactions can, in principle, interconvert.

The lack of a structure for the monomeric Trp repressor necessitated the construction of the symmetrized-Gō model for the dimeric structure. Accordingly, each interfacial contact between the two monomers is copied and allowed to be formed intramolecularly within each monomer. In addition, any non-helical contact in the monomers is allowed to interchange. The Gō-model includes 238 native contacts, of which 70 and 98 are intra- and intersubunit interactions, respectively. The symmetrized-Gō model includes 356 contacts where 220 are formed intramolecularly and 136 are interfacial contacts. The Hamiltonian used in the Gō model simulations is similar to those used in the study of the oligomerization mechanism of other proteins.[7,8] The details of the symmetrized-Gō Hamiltonian can be found elsewhere.[56,57] In the symmetrized-Gō model, the dihedral angles in the non-helical regions are fourfold less stringent than any other dihedral angle in the protein. This additional flexibility is necessary to allow a conformational change that is required for the domain-swapping process. To ensure efficient simulation of the bimolecular folding reaction, the $[8\text{-}66]_2$ TR system was simulated in a sphere that corresponds to a concentration of about 500 times larger than that used in the experiments.

The equilibrium native topology-based model trajectories are used to estimate the protection factor in the EX2 limit for each residue at any conformational ensemble.[58,59] The protection factor P_i (the ratio between the intrinsic rate of HX in the unstructured protein over the HX rate at a given conformational ensemble) was assumed to be proportional to the number of native interactions that any residue forms with all other residues. In other words, the free energy difference between the exchange-competent and exchange-incompetent states is described in terms of formation of native contacts:

$$\ln P_i \approx \sum_j p_{ij}^{\mathrm{X}} \varepsilon_{ij} - \sum_j p_{ij}^{\mathrm{U}} \varepsilon_{ij} \qquad (5)$$

where U is the unfolded state, X stands for all other states (folded and intermediate states), p_{ij} is the probability of forming a native contact, and ε_{ij} is the the strength of the contacts between residues i and j (set to 1 for all contacts). The protection factor of each residue in a given state is estimated by the degree of formation of its native contacts. When residue j forms most of its native contacts, i.e. most of its p_{ij} are close to 1, it is exchange-incompetent and its protection factor is high. Conversely, when residue j in state X has a number of contacts similar to that in the unfolded state, its protection factor is low.

Acknowledgements

This work was supported by N.I.H. grant GM 54836 to C.R.M. and by N.S.F. grants PHY-0216576 and 0225630 with additional support from MCB-0543906 to J.N.O. We acknowledge the assistance of James Evans and Karen Green in the operation of the mass spectrometry instruments. We acknowledge Dr Osman Bilsel for his help in the data analysis and Dr Jill Zitzewitz for her valuable discussions and help with the manuscript.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2006.07.080

References

1. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl Acad. Sci. USA*, **89**, 8721–8725.
2. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
3. Chavez, L. L., Onuchic, J. N. & Clementi, C. (2004). Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* **126**, 8426–8432.
4. Koga, N. & Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* **313**, 171–180.
5. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953.
6. Levy, Y. & Onuchic, J. N. (2006). Mechanisms of protein assembly: lessons from minimalist models. *Accts Chem. Res.* **39**, 135–142.
7. Levy, Y., Cho, S. S., Onuchic, J. N. & Wolynes, P. G. (2005). A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* **346**, 1121–1145.
8. Levy, Y., Wolynes, P. G. & Onuchic, J. N. (2004). Protein topology determines binding mechanism. *Proc. Natl Acad. Sci. USA*, **101**, 511–516.
9. Finke, J. M. & Onuchic, J. N. (2005). Equilibrium and kinetic folding pathways of a TIM barrel with a funneled energy landscape. *Biophys. J.* **89**, 488–505.

10. Gosavi, S., Chavez, L. L., Jennings, P. A. & Onuchic, J. N. (2006). Topological frustration and the folding of interleukin-1beta. *J. Mol. Biol.* **357**, 986–996.
11. Matthews, C. R. (1987). Effect of point mutations on the folding of globular proteins. *Methods Enzymol.* **154**, 498–511.
12. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.
13. Khan, F., Chuang, J. I., Gianni, S. & Fersht, A. R. (2003). The kinetic pathway of folding of barnase. *J. Mol. Biol.* **333**, 169–186.
14. Matthews, C. R. (1993). Pathways of protein folding. *Annu. Rev. Biochem.* **62**, 653–683.
15. Wallace, L. A. & Matthews, C. R. (2002). Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys. Chem.* **101-102**, 113–131.
16. Kataoka, M. & Goto, Y. (1996). X-ray solution scattering studies of protein folding. *Fold. Des.* **1**, R107–R114.
17. Utiyama, H. & Baldwin, R. L. (1986). Kinetic mechanisms of protein folding. *Methods Enzymol.* **131**, 51–70.
18. Eftink, M. R. (1995). Use of multiple spectroscopic methods to monitor equilibrium unfolding of proteins. *Methods Enzymol.* **259**, 487–512.
19. Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature*, **335**, 700–704.
20. Udgaonkar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature*, **335**, 694–699.
21. Parker, M. J. & Marqusee, S. (2001). A kinetic folding intermediate probed by native state hydrogen exchange. *J. Mol. Biol.* **305**, 593–602.
22. Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997). Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Sci.* **6**, 1101–1109.
23. Maity, H., Maity, M., Krishna, M. M., Mayne, L. & Englander, S. W. (2005). Protein folding: the stepwise assembly of foldon units. *Proc. Natl Acad. Sci. USA*, **102**, 4741–4746.
24. Deng, Y. & Smith, D. L. (1999). Rate and equilibrium constants for protein unfolding and refolding determined by hydrogen exchange-mass spectrometry. *Anal. Biochem.* **276**, 150–160.
25. Smith, D. L., Deng, Y. & Zhang, Z. (1997). Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J. Mass. Spectrom.* **32**, 135–146.
26. Wintrode, P. L., Rojsajjakul, T., Vadrevu, R., Matthews, C. R. & Smith, D. L. (2005). An obligatory intermediate controls the folding of the alpha-subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* **347**, 911–919.
27. Zhang, Z. & Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522–531.
28. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
29. Milne, J. S., Xu, Y., Mayne, L. C. & Englander, S. W. (1999). Experimental study of the protein folding landscape: unfolding reactions in cytochrome c. *J. Mol. Biol.* **290**, 811–822.
30. Xu, Y., Mayne, L. & Englander, S. W. (1998). Evidence for an unfolding and refolding pathway in cytochrome c. *Nature Struct. Biol.* **5**, 774–778.
31. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455–2468.
32. Xu, D., Tsai, C. J. & Nussinov, R. (1998). Mechanism and evolution of protein dimerization. *Protein Sci.* **7**, 533–544.
33. Gittelman, M. S. & Matthews, C. R. (1990). Folding and stability of trp aporepressor from Escherichia coli. *Biochemistry*, **29**, 7011–7020.
34. Mann, C. J., Royer, C. A. & Matthews, C. R. (1993). Tryptophan replacements in the trp aporepressor from Escherichia coli: probing the equilibrium and kinetic folding models. *Protein Sci.* **2**, 1853–1861.
35. Mann, C. J. & Matthews, C. R. (1993). Structure and stability of an early folding intermediate of Escherichia coli trp aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. *Biochemistry*, **32**, 5282–5290.
36. Shao, X. & Matthews, C. R. (1998). Single-tryptophan mutants of monomeric tryptophan repressor: optical spectroscopy reveals nonnative structure in a model for an early folding intermediate. *Biochemistry*, **37**, 7850–7858.
37. Gloss, L. M., Simler, B. R. & Matthews, C. R. (2001). Rough energy landscapes in protein folding: dimeric E. coli Trp repressor folds through three parallel channels. *J. Mol. Biol.* **312**, 1121–1134.
38. Gloss, L. M. & Matthews, C. R. (1998). Mechanism of folding of the dimeric core domain of Escherichia coli trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway dimeric intermediate. *Biochemistry*, **37**, 15990–15999.
39. Gloss, L. M. & Matthews, C. R. (1998). The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of Escherichia coli Trp repressor are dominated by enthalpic contributions. *Biochemistry*, **37**, 16000–16010.
40. Deng, Y. & Smith, D. L. (1998). Identification of unfolding domains in large proteins by their unfolding rates. *Biochemistry*, **37**, 6256–6262.
41. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75–86.
42. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. (2003). Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci.* **12**, 153–160.
43. Gloss, L. M. & Matthews, C. R. (1997). Urea and thermal equilibrium denaturation studies on the dimerization domain of *Escherichia coli* Trp repressor. *Biochemistry*, **36**, 5612–5623.
44. Zhang, R. G., Joachimiak, A., Lawson, C. L., Schevitz, R. W., Otwinowski, Z. & Sigler, P. B. (1987). The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature*, **327**, 591–597.
45. Wildes, D., Anderson, L. M., Sabogal, A. & Marqusee, S. (2006). Native state energetics of the Src SH2 domain: evidence for a partially structured state in the denatured ensemble. *Protein Sci.* **15**, 1769–1779.
46. Nicholson, E. M., Mo, H., Prusiner, S. B., Cohen, F. E. & Marqusee, S. (2002). Differences between the prion protein and its homolog Doppel: a partially structured state with implications for scrapie formation. *J. Mol. Biol.* **316**, 807–815.

47. Llinas, M., Gillespie, B., Dahlquist, F. W. & Marqusee, S. (1999). The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nature Struct. Biol.* **6**, 1072–1078.
48. Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature Struct. Biol.* **3**, 782–787.
49. Gorski, S. A., Le Duff, C. S., Capaldi, A. P., Kalverda, A. P., Beddard, G. S., Moore, G. R. & Radford, S. E. (2004). Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J. Mol. Biol.* **337**, 183–193.
50. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T. *et al.* (2002). Long-range interactions within a nonnative protein. *Science*, **295**, 1719–1722.
51. Hoang, L., Bedard, S., Krishna, M. M., Lin, Y. & Englander, S. W. (2002). Cytochrome c folding pathway: kinetic native-state hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **99**, 12173–12178.
52. Finucane, M. D. & Jardetzky, O. (1996). The pH dependence of hydrogen-deuterium exchange in trp repressor: the exchange rate of amide protons in proteins reflects tertiary interactions, not only secondary structure. *Protein Sci.* **5**, 653–662.
53. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 77–83.
54. Ptitsyn, O. B. (1998). Protein folding: nucleation and compact intermediates. *Biochemistry (Mosc)*, **63**, 367–373.
55. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002). Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *J. Mol. Biol.* **324**, 851–857.
56. Yang, S., Cho, S. S., Levy, Y., Cheung, M. S., Levine, H., Wolynes, P. G. & Onuchic, J. N. (2004). Domain swapping is a consequence of minimal frustration. *Proc. Natl Acad. Sci. USA*, **101**, 13786–13791.
57. Cho, S. S., Levy, Y., Onuchic, J. N. & Wolynes, P. G. (2005). Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation. *Phys. Biol.* **2**, S44–S55.
58. Best, R. B. & Vendruscolo, M. (2006). Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. *Structure*, **14**, 97–106.
59. Dixon, R. D., Chen, Y., Ding, F., Khare, S. D., Prutzman, K. C., Schaller, M. D. *et al.* (2004). New insights into FAK signaling and localization based on detection of a FAT domain folding intermediate. *Structure*, **12**, 2161–2171.

Edited by P. Wright

(Received 22 May 2006; received in revised form 27 July 2006; accepted 28 July 2006)
Available online 2 August 2006