

The Highly Cooperative Folding of Small Naturally Occurring Proteins Is Likely the Result of Natural Selection

Alexander L. Watters,^{1,6} Pritilekha Deka,² Colin Corrent,³ David Callender,³ Gabriele Varani,^{2,3} Tobin Sosnick,⁴ and David Baker^{3,5,*}

¹Molecular and Cellular Biology Program

²Department of Chemistry

³Department of Biochemistry

University of Washington, Seattle, WA 98195, USA

⁴Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, USA

⁵Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

⁶Present address: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94143-2240, USA.

*Correspondence: dabaker@u.washington.edu

DOI 10.1016/j.cell.2006.12.042

SUMMARY

To illuminate the evolutionary pressure acting on the folding free energy landscapes of naturally occurring proteins, we have systematically characterized the folding free energy landscape of Top7, a computationally designed protein lacking an evolutionary history. Stopped-flow kinetics, circular dichroism, and NMR experiments reveal that there are at least three distinct phases in the folding of Top7, that a nonnative conformation is stable at equilibrium, and that multiple fragments of Top7 are stable in isolation. These results indicate that the folding of Top7 is significantly less cooperative than the folding of similarly sized naturally occurring proteins, suggesting that the cooperative folding and smooth free energy landscapes observed for small naturally occurring proteins are not general properties of polypeptide chains that fold to unique stable structures but are instead a product of natural selection.

INTRODUCTION

A critical consideration in the accurate modeling and engineering of biological systems is the contribution of evolutionary pressures; are the physical properties of its components sufficient to predict the behavior of a system, or is it also necessary to understand the evolutionary pressures operating on the system? A route to identifying such pressures is to compare the properties of systems lacking evolutionary history to those of naturally occurring systems. Due to recent advances in de novo protein

design, this is now possible for protein folding (Kuhlman and Baker, 2004).

Protein structures are stabilized primarily by the sequestration of hydrophobic residues away from solvent in the protein core (Dill, 1990). Because hydrophobic interactions are quite nonspecific, alternative hydrophobic core packing arrangements could exist that, while less stable than the native packing arrangement, could be more stable than the unfolded state. This would produce a rugged folding free energy landscape in which the folding of a polypeptide chain could be slowed by trapping in nonnative minima. In agreement with this picture, many theoretical models predict that protein folding should be a complex, slow process characterized by the trapping of the polypeptide chain in nonnative minima (Chan, 2000; Fukunishi, 1998; Go, 1984; Kaya and Chan, 2003; Sali et al., 1994). However, a large body of work carried out over the past 15 years suggests that rugged folding free energy landscapes are quite rare in small, naturally occurring proteins; these proteins generally have highly cooperative folding transitions in which nonnative states are not significantly populated (Jackson, 1998; Jackson and Fersht, 1991; Krantz et al., 2002; Krantz and Sosnick, 2000). Is this discrepancy caused by fundamental differences between these simple models and real polypeptide chains, or is it due to evolutionary pressures that have operated on naturally occurring proteins (Jewett et al., 2003)? If the cooperative folding and smooth free energy landscapes of small, naturally occurring proteins are a general property of any polypeptide folding to a unique stable structure, then a protein capable of folding to a stable native state but lacking an evolutionary history would be expected to possess a smooth free energy landscape.

Previous attempts to explore evolutionary pressures on protein folding landscapes have involved characterization of novel sequences that adopt naturally occurring protein

structures. For example, novel folded variants of protein L and the Src SH3 domain have been identified using phage-display technology (Gu et al., 1995; Riddle et al., 1997), while computational protein design has allowed for the complete redesign of sequences capable of folding to naturally occurring structures (Dahiyat and Mayo, 1997; Dantas et al., 2003; Dobson et al., 2006; Gillespie et al., 2003). The proteins generated by these two methods show two-state cooperative folding; however, these methods were unable to generate proteins completely lacking an evolutionary history—with a single exception (Gillespie et al., 2003), every protein generated by these methods can be readily identified as sequence homologs of naturally occurring proteins (Dahiyat and Mayo, 1997; Dantas et al., 2003; Gu et al., 1995; Riddle et al., 1997).

The complete separation of protein structure from evolutionary history was recently achieved through the design of a sequence that folds into a novel structure (Kuhlman et al., 2003). Because the sequence and structure of the Top7 protein differ from those of known proteins, it is an ideal candidate to study the free energy landscape of a protein lacking an evolutionary history. Like naturally occurring proteins, Top7 folds to a unique stable native state, but preliminary kinetic studies suggested the folding of Top7 may not be highly cooperative (Scalley-Kim and Baker, 2004).

In order to explore the differences between Top7 and naturally occurring proteins, we have systematically characterized the free energy landscape by searching for stable nonnative states of Top7 using NMR, stopped-flow kinetics, and mutational and fragment studies. Our results demonstrate that the folding free energy landscape of Top7 is much more complex than those of small, naturally occurring proteins. We discuss the implications of these findings for our understanding of the evolutionary pressures on protein folding and future endeavors in protein engineering.

RESULTS

In protein folding kinetics studies, unfolded protein in denaturant is diluted into buffer lacking denaturant, and the formation of secondary and tertiary structure during folding is followed by monitoring the change in circular dichroism (CD) signal of the polypeptide chain and/or intrinsic fluorescence of tryptophan residues. The folding processes of most small naturally occurring proteins can be fit with a single exponential (Ae^{-kt}), consistent with a single cooperative transition between the unfolded and native states, and the logarithm of the rate constant k is typically a linear function of the denaturant concentration. Deviations from single-exponential kinetics and/or significant nonlinearity in the denaturant dependence of folding kinetics suggest the population of metastable nonnative states during folding (Baldwin, 1996).

Folding of Top7 Involves at Least Three Distinct Kinetic Phases

Folding and unfolding experiments monitoring tertiary structure formation using the intrinsic fluorescence of the tryptophan at position 83 demonstrate that for most guanidine concentrations a triple exponential ($A_{\text{fast}}e^{-k_{\text{fast}}t} + A_{\text{middle}}e^{-k_{\text{middle}}t} + A_{\text{slow}}e^{-k_{\text{slow}}t}$) is needed to fit the folding reaction of Top7, whereas the unfolding reaction can be fit to a single exponential. Figure 1A shows a kinetic trace of Top7 folding fit using a triple-exponential model (one exponential for the fast increase in fluorescence and two for the slower decrease in fluorescence). Figure 1B shows the guanidine dependence of the three rate constants (k_{fast} , k_{middle} , k_{slow}) from the triple-exponential fit. The guanidine dependence of protein folding rates reflects the amount of surface area buried during the corresponding transition (Myers et al., 1995), and the strong guanidine dependence of the fast phase, k_{fast} (m value = 1.1 kcal mol⁻¹ M⁻¹, see Experimental Procedures), indicates that a significant fraction of the surface area is buried during this transition. Below 4 M guanidine, the middle and slow phases, k_{middle} and k_{slow} , are guanidine independent, suggesting that these phases correspond to transitions in which no additional surface area is buried, most likely structural rearrangements between collapsed states. Because Top7 does not possess any prolines, neither of these guanidine-independent phases is due to *cis-trans* proline isomerizations, which cause multiphasic folding in some naturally occurring proteins (Calloni et al., 2003; Grantcharova and Baker, 1997).

Figure 1C shows the guanidine dependence of the amplitudes of the three phases (A_{fast} , A_{middle} , A_{slow}). At higher guanidine concentrations the amplitudes of the fast and slow phases approach zero, suggesting that the nonnative structures associated with the two phases are disappearing. The disappearance of these phases suggests that at high guanidine concentrations Top7 displays the single-exponential folding and unfolding kinetics, with rate constants dependent on the guanidine concentration (Figure 1B), that are characteristic of cooperatively folding proteins.

Previous folding experiments have shown that apparent complexities in the folding of small, naturally occurring proteins can be ascribed to transient aggregation (Silow and Oliveberg, 1997; Silow et al., 1999). However, the rate constants and associated amplitudes of the three phases are not dependent on protein concentration (Figure 1D), suggesting the complex folding kinetics of Top7 are not due to intermolecular interactions.

Secondary Structure Is Formed during the Rapid-Burst Phase

Continuous-flow circular dichroism experiments were carried out to monitor the formation of secondary structure during folding. At low concentrations of guanidine, 80% of the circular dichroism signal of the native protein is recovered during the first 2 ms of folding (Figure 1E). The kinetics and guanidine dependence of the rapid formation

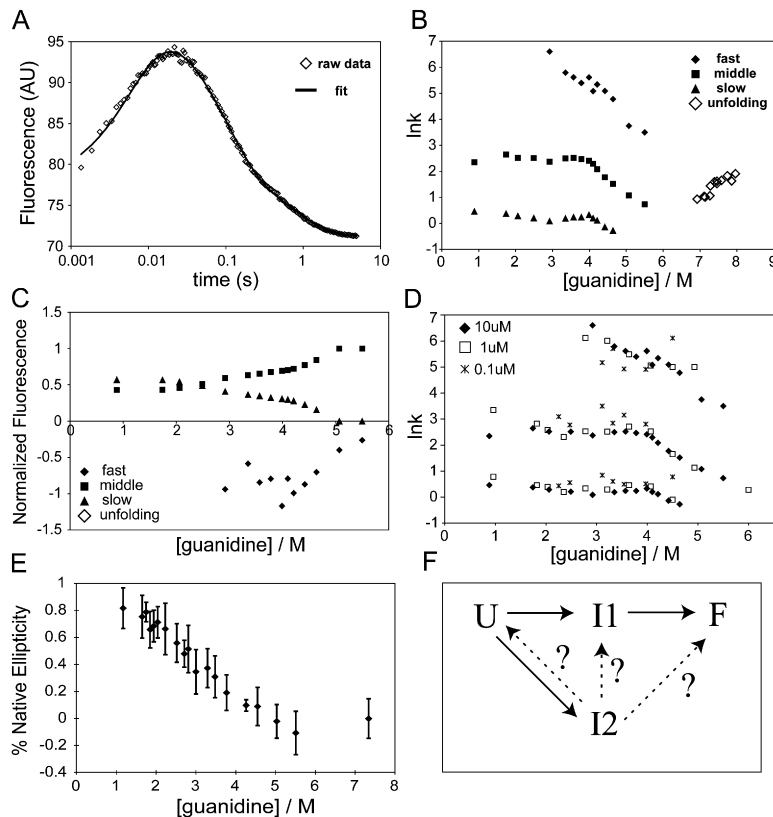


Figure 1. Kinetic Experiments Reveal Multiple Nonnative States Populated during the Folding of Top7

The folding and unfolding of Top7 was monitored by stopped-flow fluorescence and continuous-flow circular dichroism. (A) Kinetic trace from stopped-flow fluorescence experiment in which Top7 was rapidly diluted from 7.0 M guanidine to 4.1 M guanidine, overlaid with the best triple-exponential fit ($A_{\text{fast}}e^{-k_{\text{fast}}t} + A_{\text{middle}}e^{-k_{\text{middle}}t} + A_{\text{slow}}e^{-k_{\text{slow}}t}$) to the data. (B) The guanidine dependence of the three observable rate constants needed to fit the folding reactions— k_{fast} , k_{middle} , and k_{slow} (filled symbols) and the single rate constant needed for the unfolding reactions (open symbol)—at low guanidine concentrations, the fast phase is completed during the dead time of the instrument (2 ms). (C) The guanidine dependence of the associated amplitudes (below 4 M guanidine it is difficult to estimate the amplitude of the fast phase, and therefore the data are very noisy). The amplitudes have been normalized against the sum of the amplitudes for the middle and slow phases ($A_{\text{middle}} + A_{\text{slow}}$). (D) Concentration dependence of the rate constants. 10 μM (filled diamonds), 1 μM (open squares), and 0.1 μM (crosses) Top7. (E) Continuous-flow circular dichroism experiments reveal that after 2 ms (dead time of the instrument), over 80% of native CD signal is recovered at guanidine concentrations less than 2 M. Y axis: circular dichroism signal at 222 nm 2 ms after initiation of refolding expressed as a percentage of

native CD signal at 222 nm. Each data point represents the average of 8 to 15 independent 40 ms continuous-flow experiments that were sampled using 0.5 ms bins. The error bars represent the standard error of the average from these runs. (F) The triple exponential folding kinetics of Top7 suggests that, at low guanidine concentrations, the folding of Top7 proceeds through two parallel pathways with two or more intermediates (see text).

of secondary structure parallel those of the fast phase observed in the fluorescence experiments. Thus, at low guanidine concentrations unfolded Top7 rapidly collapses to nonnative conformations containing significant amounts of secondary structure.

Four-State Model

At high guanidine concentrations, the folding kinetics of Top7 are similar to those of small naturally occurring proteins. The kinetics near the equilibrium unfolding transition of Top7 (~ 6 M guanidine) can be fit quite well with a single exponential, and the rate is linearly dependent on the denaturant concentration (Figures 1C and 1D and Scalley-Kim and Baker, 2004). Under these conditions, folding appears to be a cooperative transition between the open unfolded state and the compact folded state, as observed for most small naturally occurring proteins.

As the guanidine concentration is decreased, the free energy landscape of Top7 becomes much more complex, possessing multiple stable minima. At low guanidine concentrations (< 4 M), the folding of Top7 is a triple-exponential reaction, indicating that at least four states are populated during the folding reaction (U, I1, I2, and F) (Ikai

and Tanford, 1973). Given the large number of parameters in a four-state system and the multiple possible connections between the states, we do not attempt to fit a specific quantitative model to the data. Instead, we take a more qualitative approach and propose the simplified model outlined in Figure 1F and described in the following paragraphs.

The middle phase is likely to represent a transition from a collapsed species to the native state. This interpretation is based on the continuity in the values of the rate constants associated with the middle phase and the values of the rate constants associated with the single-exponential folding transition at high guanidine concentrations, which clearly represents a transition to the native structure. At lower guanidine concentrations, the middle phase is guanidine independent, suggesting that it represents a transition to the native state from a collapsed species (I1). The switch from the guanidine dependence of the folding rate at high guanidine concentration to the guanidine independence at lower concentrations is seldom observed for small, naturally occurring proteins (Jackson, 1998)—in the rare cases where this is observed, the decreases in the guanidine dependence of the folding rates

Table 1. Location and Contribution to Stability of Top7 Mutants Examined in the Kinetics Studies

Protein	Location	m Value ^a	ΔG (3 M) ^a	$\Delta\Delta G$ (3 M)
Wild-type ^b	NA	-2.22	6.54	NA
K41E/K42E/K57E ^b	1st and 2nd helix	-2.22	6.68	0.14
F17Q/Y19L ^c	2nd strand	-2.26	6.81	0.27
G14A ^c	1st hairpin	-2.29	5.98	-0.56
Y21L ^c	2nd strand	-2.14	6.33	-0.21
L29A ^c	1st helix	-1.83	3.77	-2.77
N34G ^c	1st helix	-1.84	5.40	-1.14
V48A ^c	3rd strand	-2.13	4.87	-1.67
F63A ^c	2nd helix	-2.05	4.47	-2.07
A64G/A65G ^c	2nd helix	-2.17	5.28	-1.26
L67A ^c	2nd helix	-2.13	5.84	-0.70
V81A ^c	4th strand	-2.01	5.01	-1.53
G85A ^c	2nd hairpin	-2.12	5.35	-1.16
V90A ^c	5th strand	-2.09	5.45	-1.10

^a Values were determined from CD-monitored guanidine denaturation curves (see Experimental Procedures).

^b F83W mutation.

^c Mutations are in the context of K41E, K42E, K57E, and F83W.

occur only at very low denaturant concentrations, and the rates do not become completely denaturant independent (Ferguson et al., 1999; Jemth et al., 2004).

The stopped-flow fluorescence and circular dichroism experiments suggest that the fast phase reflects the rapid collapse of the chain into conformations having significant secondary structure (I1 and I2). Because the signal change associated with the fast phase is in the opposite direction of the fluorescence change associated with the final formation of the native state (Figure 1A), the fast phase cannot represent a direct transition from the unfolded to folded state. Such a rapid burst phase is generally not observed for proteins fewer than 100 amino acids (Grantcharova and Baker, 1997; Jackson, 1998; Jacob et al., 2004; Krantz et al., 2002; Krantz and Sosnick, 2000; Plaxco et al., 1999). In the case of Top7, the rapid collapse may be due to the extremely hydrophobic nature of the protein core (the free energy decrease associated with hydrophobic association may outweigh the entropic cost of chain collapse).

The very slow phase could represent a parallel kinetic channel or an additional obligatory step in a sequential process. The former is more plausible, because if the middle phase represents a transition to the native state, it cannot be preceded by a much slower obligatory step, as it would no longer be observable (Bai, 2003). Therefore, the slow phase suggests that an alternative collapsed state (I2) is formed in parallel with I1 during the rapid collapse phase; because of the slow rate of escape from I2, it may be viewed as a kinetic trap (Figure 1F). The denaturant independence of the slow phase suggests that transitions out of this state do not involve significant changes

in surface area. The marked similarity in the denaturant dependencies of the middle phase and the slow phase, notably the sharp kink at 4 M guanidine, suggests that transitions from both states may involve a similar structural rearrangement.

Top7 Mutants Further Highlight the Folding Complexity

To probe the structure and sequence features critical for the three kinetic phases, and to attempt to simplify the complexity of the kinetics, we examined the folding kinetics of 11 point mutants of Top7 (Table 1 and Figure 2L). The results further highlight the complexity of Top7 folding. Remarkably, no mutation simplified the kinetics (all three phases are present for all the proteins examined), presumably because the mutations do not significantly destabilize the nonnative states—in fact, some mutations appear to make the folding process even more complex (see the Supplemental Data available with this article online). The mutations do change the rates associated with the different phases, but to further complicate the analysis, these changes differ for nearly all of the mutants.

We can roughly classify the effects of the mutations as follows. L67A (Figure 2H), A64G/A65G (Figure 2G), and V81A (Figure 2I) slow down the fast phase; L29A (Figure 2C), V48A (Figure 2E), and F63A (Figure 2F) speed up one or both of the two slower phases; L67A, G85A (Figure 2J), and to a lesser extent Y21L (Figure 2B), A64G/A65G, V81A, and V90A (Figure 2K) slow down one or both of the two slower phases; and G14A and N34G do not affect any phase. To interpret these results, we assume

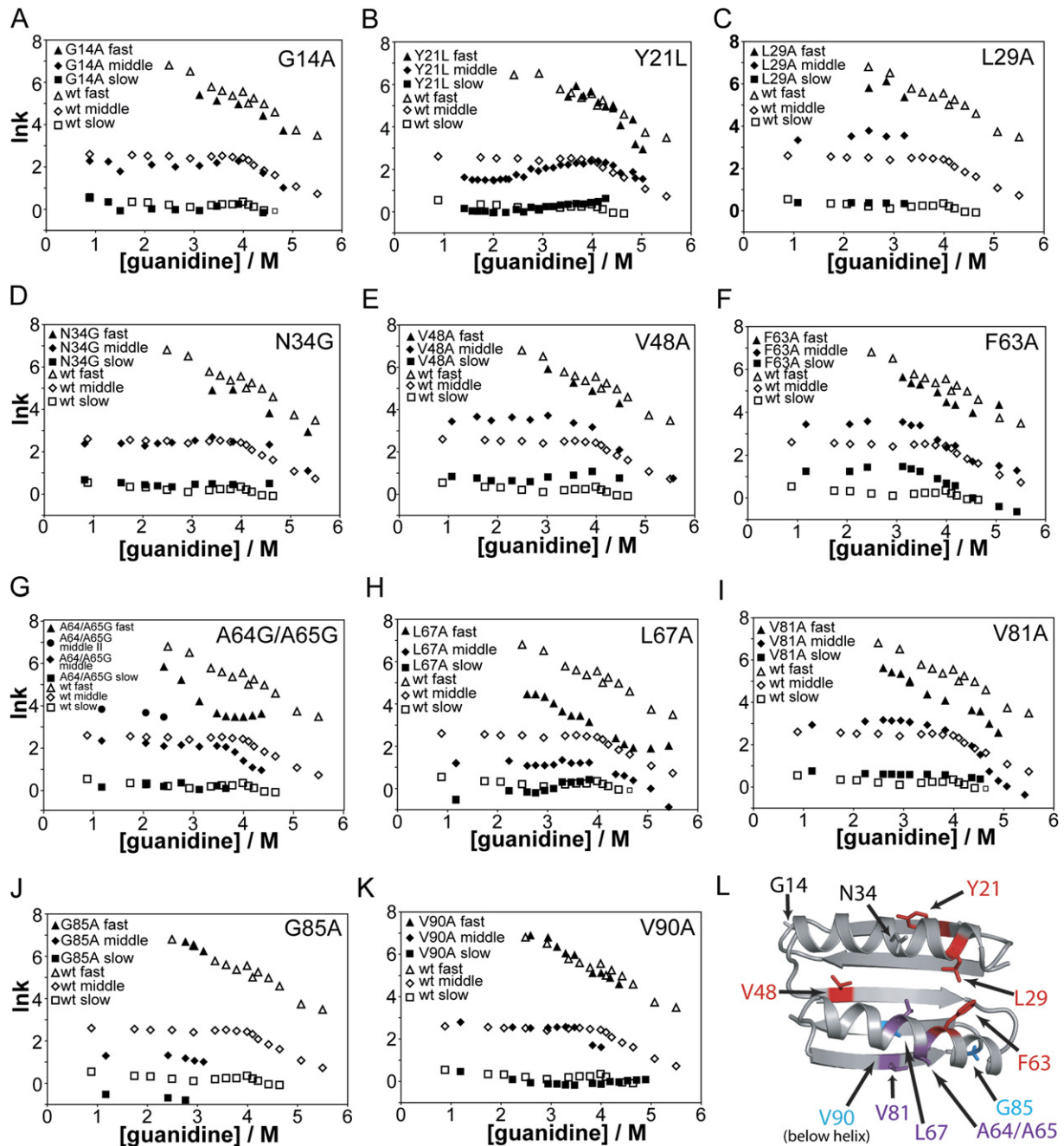


Figure 2. Guanine Dependence of the Folding Kinetics of Top7 Mutants

The logarithms of the rate constants for wild-type Top7 folding (open symbols) and (closed symbols) (A) G14A, (B) Y21L, (C) L29A, (D) N34G, (E) V48A, (F) F63A, (G) A64G/A65G double mutant, (H) L67A, (I) V81A, (J) G85A, and (K) V90A are plotted versus guanidine concentration. (L) Structure of Top7 showing the location of the residues that are critical in the formation of the intermediates (red residues), the native state (blue), both (purple), or neither (gray). All mutants contained three additional surface mutations (K41E, K42E, and K57E) to reduce perceived protein degradation. These mutations had no apparent effect on the equilibrium (Table 1) and kinetic properties (data not shown) of Top7. At low guanidine concentrations, four phases are needed to fit A64G/A65G (see Supplemental Data).

that because all the mutations involve side-chain truncations they destabilize intermediates and/or transition states (TSs) relative to the unfolded state (Matouschek

and Fersht, 1991) and that the interactions in transition states are subsets of the interactions in the stable structures they lead to. Therefore, in the model outlined above

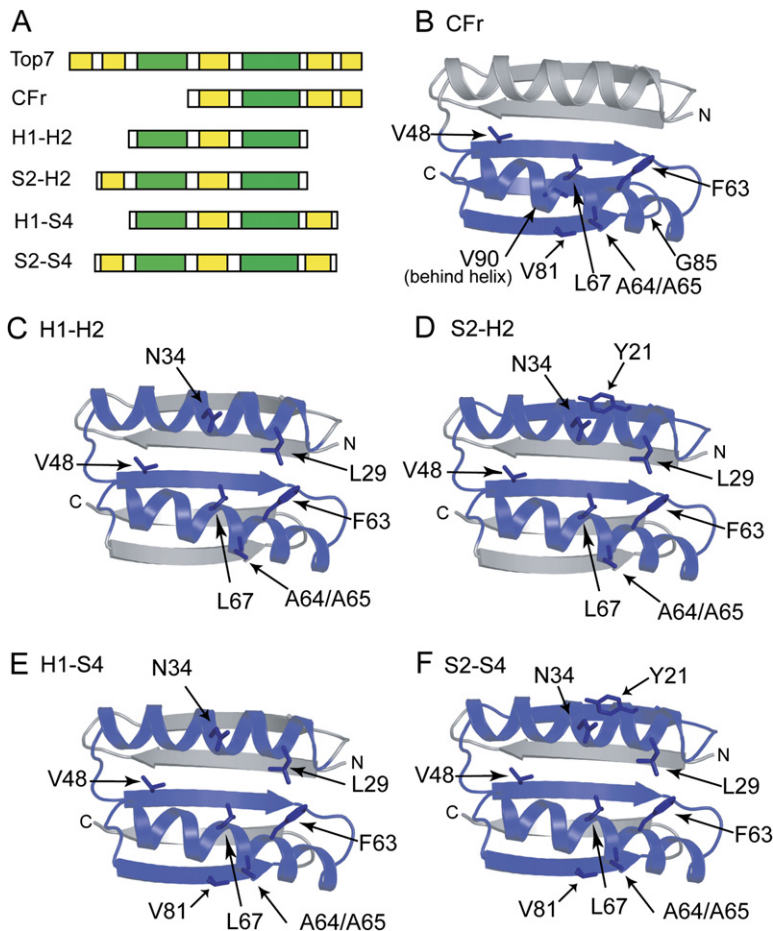


Figure 3. Schematic Diagram of Top7 Fragments

(A) Linear representation of the full-length Top7 and the five fragments studied showing elements of secondary structure—strand, yellow; helix, green. Structures of Top7 highlighting the regions contained in the five subfragments studied (portions in blue). (B) CFr (residues R47 to L92), (C) H1-H2 (E26 to Y76), (D) S2-H2 (K15 to Y76), (E) H1-S4 (E26 to D84), and (F) S2-S4 (K15 to D84). For comparison purposes, the residues analyzed in the mutational studies are indicated in the corresponding fragments.

(Figure 1F), mutations that destabilize one or both of the nonnative structures should speed up one or both of the slower phases, and if they also destabilize the transition states leading into these states, they should slow down the fast phase. In contrast, mutations that destabilize the transition states out of the nonnative states toward the native state should slow down the middle and/or the slow phase.

Given the complexity, detailed interpretation of the mutational results must proceed with caution, but in light of the considerations above, the residues in the middle portion of Top7 (L29, V48, F63, A64, A65, L67, and V81) appear important for the formation and/or stability of one or more of the nonnative states, and residues in the C-terminal portion (A64, A65, L67, V81, G85, and V90) appear important for the transition to the native state. Y21L does not affect the stability of the native state, but does slow the middle phase, perhaps by stabilizing the nonnative states.

Fragments of Top7 Are Stable in Isolation

To identify structural elements of Top7 that may be responsible for the formation of the folding intermediates, five fragments of Top7 were purified and analyzed. The C-terminal fragment (CFr) consists of residues 45–94 (Figure 3B)

and contains residues suggested by the mutational studies to be critical for the formation of the native state. The other fragments contain residues that appeared to be critical for the stability of I1 and/or I2. The smallest fragment, H1-H2 (residues 28–76), contains both helices and strand 3. The remaining fragments contain the H1-H2 core in addition to strand 2 and/or strand 4—S2-H2, H1-S4, and S2-S4 (Figures 3C–3F). CD analysis demonstrated that all five fragments form stable structures (Figure 4A) and, with the exception of H1-H2, possess significant amounts of secondary structure (Figure 4B). In the context of the full-length protein, the region corresponding to CFr forms a compact substructure, while the β strands of the other four fragments do not pair with each other (Figure 3), suggesting that the strands of these fragments in isolation are stabilized by nonnative contacts. In contrast, fragments of small naturally occurring proteins are seldom stable in isolation (Camarero et al., 2001; Flanagan et al., 1992; Ladurner et al., 1997).

To gain further insight into the complex kinetics observed in Top7 folding, we examined the folding kinetics of CFr. Interestingly, the folding of CFr can be fit to a single exponential (Figure 4C), suggesting that the complex kinetics observed for Top7 may reflect inhibition of proper

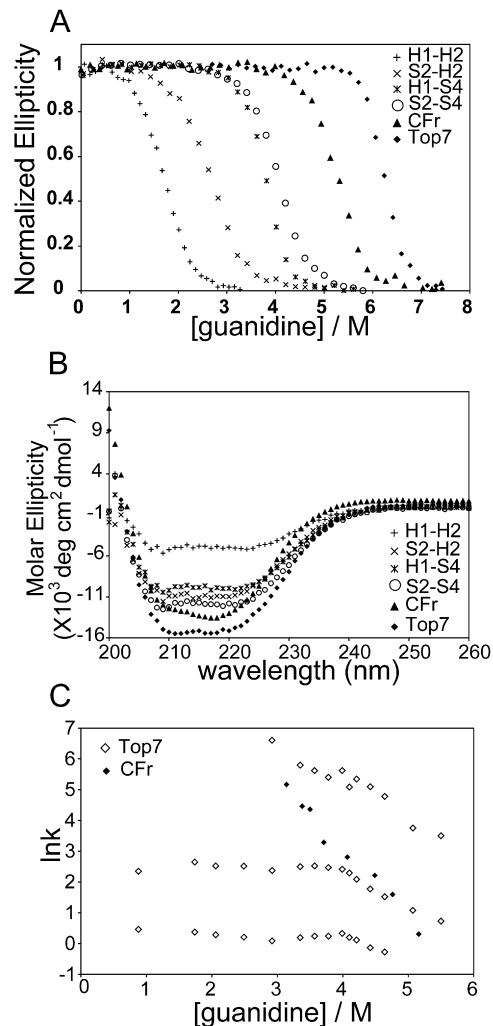


Figure 4. Biophysical Characterization of Top7 Fragments

(A) Guanidine melts monitored at 220 nm by CD and (B) CD wavelength scans of full-length Top7 and the five fragments. (C) The guanidine dependence of the rate constant for the single-exponential fits of the folding of CFr (closed symbols) and the three observed rate constants for the folding of Top7 (open symbols).

folding of the C-terminal region by nonnative interactions with the N-terminal region. This hypothesis is supported by the observation that mutations in the C-terminal region (A64, A65, L67, V81, G85, and V90) have the greatest effect on the formation of the native state, while residues in both the N- and C-terminal regions are critical for stabilizing the nonnative structures (Y21, L29, V48, F63, A64, A65, L67, V81), and fragments that contain these residues appear to be forming nonnative structures in isolation. Unfortunately, kinetic and thermodynamic parameters for the five fragments cannot be used to build a quantitative model of Top7's free energy landscape, because size-exclusion chromatography (data not shown) and NMR (Dantas et al., 2006) suggest these fragments form homodimers at equilibrium.

Equilibrium NMR Experiments Suggest the Presence of a Nonnative Conformation at Higher Guanidine Concentrations

The multiexponential folding kinetics suggests that at low guanidine concentrations there are multiple distinct states of Top7 that are stable relative to the unfolded state. NMR experiments were used to directly probe for these states and investigate possible structural changes that were not observed by the denaturation melts monitored by CD and fluorescence (Kuhlman et al., 2003; Scalley-Kim and Baker, 2004; Srimathi et al., 2002; van Mierlo et al., 2000). Triple resonance experiments on samples of singly (^{15}N) and doubly (^{15}N , ^{13}C) labeled protein were used to assign the backbone resonances of a Top7 variant with equilibrium folding properties indistinguishable from Top7 (Figure S2).

While at low guanidine concentrations each observable peak can be assigned to a specific residue in the native state, at intermediate and higher guanidine concentrations (4–6.5 M) additional, nonnative peaks appear, suggesting that a nonnative state is populated (Figure 5). To determine the effects of denaturant on the relative population of the native and nonnative structures, we identified and monitored a subset of native and nonnative (NN) peaks that could be followed across a broad range of guanidine concentrations (see Supplemental Data for a list of peaks used in this analysis). From this analysis, it is clear that the loss of secondary structure (as monitored by CD) is not coincident with the loss of specific tertiary structure (as monitored by NMR) (Figure 6A). This contrasts with most small, naturally occurring proteins for which the loss of tertiary structure is coincidental with the loss of secondary structure (Kuhlman et al., 1998).

The nonnative HSQC peaks can be divided into two groups based on their guanidine dependencies (Figure 6B). The first group consists of peaks whose growth in intensity coincides with the decrease in native peak intensity (NN-5, -6, and -9) and presumably corresponds to the alternate, nonnative structure. The second group consists of peaks (NN-18, -22, -23) whose appearance and growth parallel the loss of CD signal and hence likely correspond to the unfolded state. Analysis of these nonnative peaks demonstrates that a nonnative structure is populated around 4.5 M guanidine, well below the guanidine concentrations where secondary structure is lost and the native NMR resonances completely disappear. While possessing native-like CD and fluorescent signals, the alternate structure formed around 4.5 M guanidine does not appear to be tightly packed, since most of the nonnative peaks are in the poorly dispersed region of the proton spectrum. These characteristics suggest that this nonnative state is similar to polypeptide chains that form partially folded and molten globule structures (Chamberlain and Marqusee, 1998).

DISCUSSION

The major result of this work is that the folding free energy landscape of Top7, a protein not generated by the natural

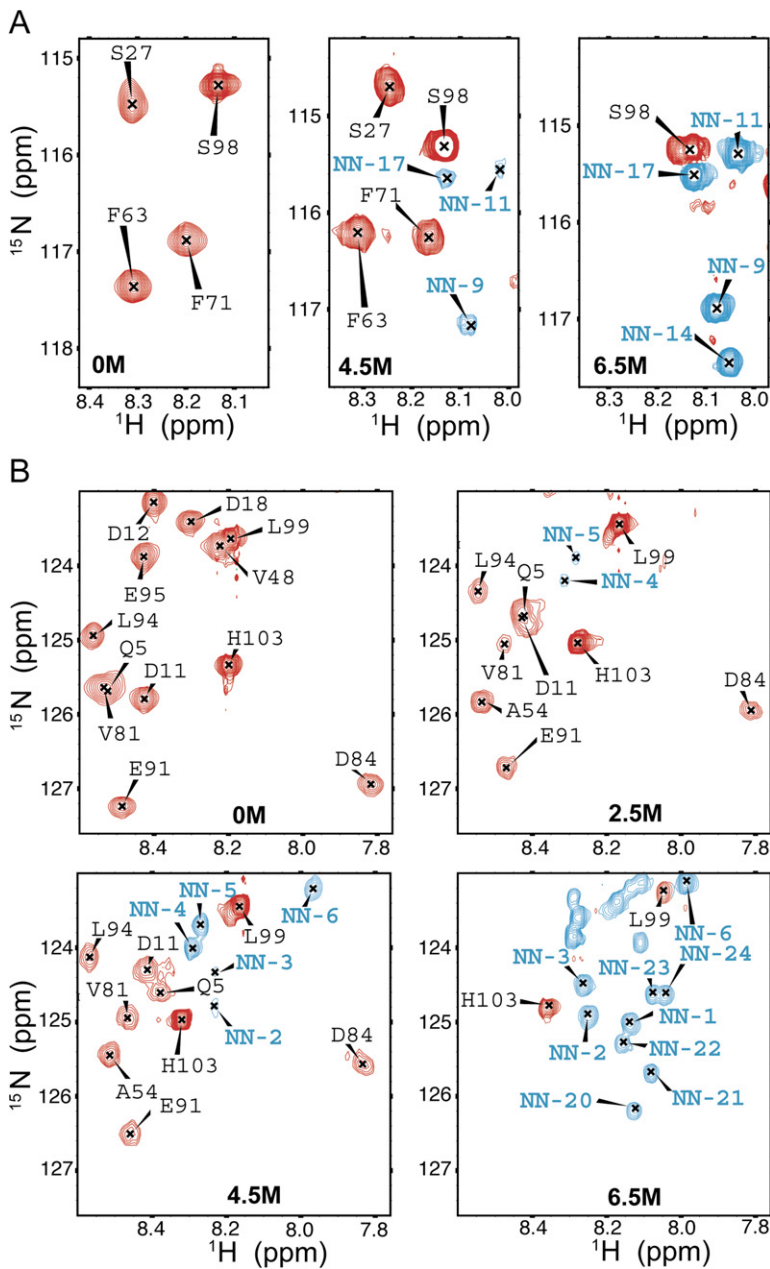


Figure 5. NMR Spectra of Top7 Reveal a Partially Structured Alternative State at Intermediate Concentrations of Denaturant

HSQC spectra of Top7 for the regions around (A) serine 98 (S98) and (B) histidine 103 (H103) at multiple guanidine concentrations show the disappearance of native peaks (red peaks, labeled by residue type and number) and the appearance of nonnative peaks (blue peaks, labeled with NN and an arbitrary numbering system). The disordered residues S98, L99, and H103 are part of the affinity tag, and their peaks do not change significantly during the guanidine melt. Each HSQC is referenced to the S98 peak.

evolutionary process, differs significantly from those of similarly sized naturally occurring proteins. First, stopped-flow kinetic experiments reveal that the folding of Top7 is multiphasic, with at least two nonnative structured states populated during the folding process. Second, NMR equilibrium studies demonstrate that a nonnative structure is populated at guanidine concentrations well below the equilibrium unfolding transition monitored by fluorescence and circular dichroism. Third, multiple fragments of Top7 appear stable in isolation. These results stand in sharp contrast to studies performed on small naturally occurring proteins, where folding reactions can generally be fit to a single exponential, only one stable state is observed in

equilibrium experiments, and fragments are not stable in isolation (Jackson, 1998). Exceptions among small, naturally occurring proteins only show small deviations from this classical behavior compared to the differences we observed for Top7. For example, a recent equilibrium NMR study suggested that the folding of the small protein BBL may not be as cooperative as other small proteins (Sadqi et al., 2006), yet BBL exhibits single-exponential folding kinetics in contrast to the at least triple-exponential folding kinetics of Top7 (Ferguson et al., 2004, 2005). Taken together, these results demonstrate that the free energy landscape of Top7 has an unusually high degree of complexity—a characteristic consistent with the hypothesis

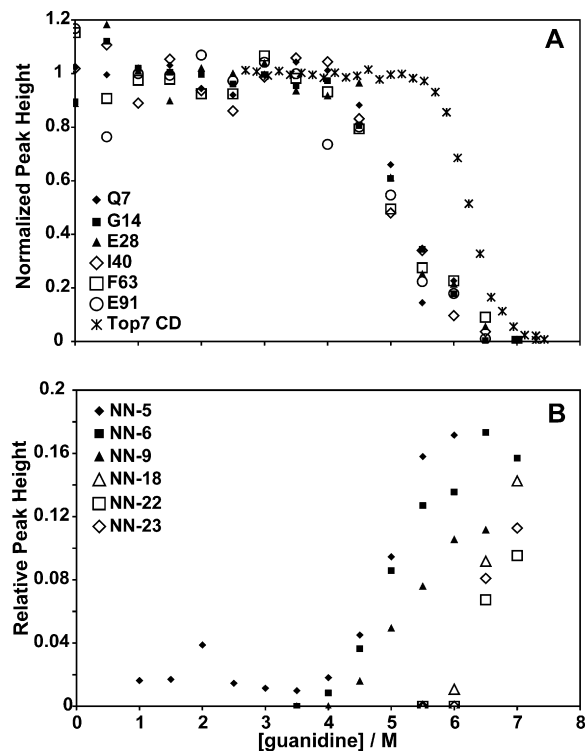


Figure 6. Equilibrium Denaturation of Top7 Monitored by 2D NMR

^{15}N HSQC spectra were collected on Top7 over a range of guanidine concentrations. The structural changes occurring in Top7 as the guanidine concentration was increased were monitored by following the height and volume of both native backbone N-H peaks identified in the absence of denaturant and nonnative backbone N-H peaks arising in the presence of denaturant. To account for the loss of signal quality at higher guanidine concentrations due to reduced probe performance in high salt concentrations, peak heights were normalized against the sum of all examined peak heights at each guanidine concentration. (A) HSQC denaturation profiles for specific native peaks compared to the denaturation profile for global unfolding obtained in the circular dichroism experiments. (B) Denaturation profiles for nonnative peaks that do not overlap heavily with other native or nonnative peaks. As discussed in the text, these profiles fall into two distinct groups. Similar profiles are obtained using peak volumes rather than peak heights (data not shown).

that the relatively smooth free energy landscapes of small, naturally occurring proteins are not a property common to all polypeptides that fold to a well-behaved, stable structure, but rather a product of natural selection. Further studies involving more computationally designed proteins will be needed to confirm this hypothesis.

Recent experimental studies have suggested an evolutionary disadvantage to proteins that possess stable, nonnative states. Partially stable, nonnative structures play roles in nucleating amyloid formation (Dobson, 2003; Horwich, 2002) and there appears to be evolutionary pressure at the sequence level against amyloid formation (Dobson, 1999; Stefani and Dobson, 2003; Wright et al., 2005). Both the accumulation of collapsed nonnative species during

folding and the stability of subfragments might exacerbate the propensity to form amyloid fibrils: the former could aggregate via exposed hydrophobic regions, and the latter could accumulate as byproducts of mistranslation or partial proteolysis and aggregate as well. Thus, selection against aggregation could lead to the cooperative folding landscapes seen for naturally occurring proteins.

Protein chaperones, which protect misfolded proteins from aggregation, appear to have evolved, in part, as global suppressors of the deleterious consequences that arise from the accumulation of misfolded or partially folded proteins (Rutherford and Lindquist, 1998; Yahara, 1999). With such a global defense against protein misfolding, it would appear unnecessary for evolution to select for specific polypeptides possessing smooth free energy landscapes. However, in contrast to the folding of large, multi-domain proteins, which show multiphasic folding and often require chaperones to avoid aggregation during the folding process, almost all small naturally occurring proteins fold in a chaperone-independent, cooperative manner (Jackson, 1998; Krantz et al., 2002; Maxwell et al., 2005). While one explanation for the preponderance of these smooth free energy landscapes is that such landscapes are a fundamental property of small proteins, the rugged free energy landscape of Top7 demonstrates that this is not the case. A more parsimonious argument is that the generally smooth free energy landscapes of small, naturally occurring proteins are the product of natural selection acting on the level of individual protein sequences, perhaps because the deleterious effects of stable partially functional conformations cannot be completely overcome by the chaperone machinery.

We have identified three general characteristics of Top7 that differ from naturally occurring proteins. These findings illuminate mechanisms by which small, naturally occurring proteins may achieve smooth free energy landscapes and should aid in future engineering of native-like proteins for *in vivo* activity

First, Top7 is significantly more stable than most naturally occurring proteins of similar size (Kuhlman et al., 2003). Because of the approximate additivity of the interactions stabilizing proteins, an exceptionally stable protein such as Top7 is likely to have relatively stable substructures that could contribute to relatively stable nonnative collapsed species. Yet, many small naturally occurring proteins have been extensively computationally redesigned and stabilized, and these stabilized versions retain cooperative folding (Dantas et al., 2003; Gillespie et al., 2003; Scalley-Kim and Baker, 2004). Conversely, while many of the mutants presented in this work destabilize Top7, they do not appear to simplify the folding process (Table 1, Figure 2, and Figure S1). Thus, extensive stabilization of the native state does not in itself give rise to complex folding.

The second mechanism by which proteins may achieve smooth free energy landscapes is through the incorporation of specific interactions to destabilize nonnative conformations. Naturally occurring proteins contain specific,

often destabilizing interactions to overcome the rather promiscuous interactions that stabilize tertiary (hydrophobic interactions) and secondary structure (backbone hydrogen bonding between β strands and α helices). Unlike Top7, many naturally occurring proteins contain buried polar interactions in their hydrophobic core that are generally destabilizing but appear to be critical for specifying the native state (Bolon and Mayo, 2001). In addition, edge strands of naturally occurring proteins may be selected to inhibit nonspecific *intermolecular* interactions across β sheet interfaces (Richardson and Richardson, 2002) and this may also be true for *intramolecular* interactions involving β strands. In the case of Top7, while the canonical nature of strands 1, 3, 4, and 5 (they are flat and have a high percentage of isoleucine and valine) leads to an extremely stable native state, their sequence and structural similarity may allow stable nonnative strand pairings.

The third mechanism by which naturally occurring proteins may have evolved smooth free energy landscapes is through the favoring of topologies with a significant fraction of contacts between residues distant in the sequence. In Top7, the majority of the stabilizing contacts are between residues nearby in the linear amino acid sequence. Proteins with a significant fraction of nonlocal contacts are expected to fold more cooperatively, as these contacts are unlikely to be formed in partially folded structures (Abkevich et al., 1995; Go and Taketomi, 1978; Govindarajan and Goldstein, 1995). The most N- and C-terminal elements of secondary structure in small naturally occurring proteins pack against each other more often than would be expected by chance, and this has been found to be correlated with cooperative folding (Krishna and Englander, 2005); the N- and C-terminal strands of Top7 do not pair with each other. The preponderance of short-range interactions may allow the N- and C-terminal regions of Top7 to remain stable yet fluctuate relative to each other and allow the exploration of a wide range of nonnative structures with at least partially native substructures. In small naturally occurring proteins, such fluctuations could disrupt function; native states possessing locally unfavorable substructures that are stabilized by long-range favorable interactions (Baker et al., 1992) would be more rigid and hence more suitable for carrying out specific functions. An evolutionary preference for relatively rigid structures stabilized by nonlocal interactions may explain why some topologies appear to be more common than others (Murzin et al., 1995).

The complex folding free energy landscape of Top7 provides an ideal testing ground for simulations of protein folding, and conversely, fully understanding this complexity will require computational modeling. Because of the high cooperativity in the folding of small naturally occurring proteins, the opportunities for comparison between computer simulations and experiment are limited (for example, there are no folding intermediates to monitor). In contrast, the much less cooperative folding of Top7 provides a rich context in which to compare simulations with experimental data; for example, the multiple states

observed in Top7 folding should be populated in folding simulations.

Insights into the determinants of the ruggedness of protein folding free energy landscapes from studies of Top7 and other designed proteins has direct bearing on the design of new proteins for *in vivo* applications. Since misfolded states and stable subfragments can be quite deleterious, proteins designed for therapeutic purposes should have highly cooperative folding transitions analogous to those of naturally occurring proteins. Based on the above discussion, this goal is likely to be facilitated by the design of proteins possessing nonlocal topologies, specific polar interactions, and secondary structural elements deviating from ideal geometries.

EXPERIMENTAL PROCEDURES

Top7 Expression and Purification

All experiments use a previously described tryptophan mutant of Top7, F83W (Scalley-Kim and Baker, 2004). All proteins were expressed using pET29b (Novagen), purified using nickel affinity and ion-exchange chromatography (Dantas et al., 2006; Kuhlman et al., 2003; Scalley-Kim and Baker, 2004), and stored in 50 mM sodium phosphate, pH 7.0. The fragments H1-H2, S2-H2, H1-S4, and S2-S4 were purified under denaturing conditions (3 M guanidine) and stored in 1 M guanidine.

Kinetic Experiments

Stopped-flow kinetic data were acquired using a BioLogic SFM4/QFM4, with a 0.8 mm cuvette, and the results were fit using the BioLogic analysis software (version 3.30) and KaleidaGraph 3.6. Folding experiments, buffered in standard conditions agreed upon by the protein-folding community, 50 mM sodium phosphate, pH 7.0 (Maxwell et al., 2005), were performed at final protein concentrations of 0.1 μ M (Top7), 1 μ M (Top7), and 10 μ M (Top7, CFr, Top7 mutants). Stopped-flow fluorescence of Top7 and CFr were carried out at 298K, with an excitation of 280 nm and emission cutoff of 309 nm (Top7 folding) or 385 nm (Top7 unfolding and CFr folding and unfolding). The CD stopped-flow experiments were performed at 283K. Other aspects of the stopped-flow experiments were carried out as previously described (Krantz et al., 2002; Krantz and Sosnick, 2000; Scalley et al., 1997). A linear fit to the k_{fast} data in Figure 1C was used to estimate the guanidine dependence of the fast phase (m value).

Equilibrium Experiments

Guanidine melts were performed at 298K in a 1 cm cuvette, at 2 μ M protein concentration and 50 mM sodium phosphate, pH 7.0, using an Aviv CD spectrometer 16A DS and a Hamilton syringe titrating device (VWR). To estimate the free energy of folding, guanidine melts for Top7 mutants were fit to a two-state model (Scalley-Kim and Baker, 2004). The denaturation curves for Top7 fragments were normalized using a two-state model with sloping baselines (Scalley et al., 1997).

NMR

Backbone amide ^1H , ^{15}N , $\text{C}\alpha$, and $\text{C}\beta$ resonances were assigned using a standard set of triple-resonance experiments (HNCA, HNCO, HNCACB, ^{15}N , and ^{13}C NOESY) (Sattler et al., 1999). HSQC spectra for samples of 1 mM ^{15}N -labeled Top7 in the appropriate guanidine concentration were recorded at 298K at 500 MHz using pulse sequences designed to minimize the guanidinium signal (Zhang et al., 1994). Thirty-five native peaks and twenty-four nonnative peaks were used to analyze the guanidine dependence of the native and nonnative peak heights and volumes. The denaturation curves for individual folded peaks were normalized using a two-state model, assuming a flat unfolded baseline (Scalley et al., 1997).

Supplemental Data

The Supplemental Data for this article, including Supplemental Experimental Procedures, Results, Tables, and Figures, can be found online at <http://www.cell.com/cgi/content/full/128/3/613/DC1/>.

ACKNOWLEDGMENTS

We would like to acknowledge Steve Reichow for his help in NMR data collection and analysis; Jens Lipfert and Ziad Eletr for their help in characterizing the folded state of Top7; Devin and Jennifer Strickland, Adarsh Pandit, Ali Shandiz, Michael Baxa, and the rest of the Sosnick lab for their help and hospitality during our collaboration; and Charlotte Berkes, Tanja Kortemme, Gautam Dantas, Tom Fazzio, Brian Yeh, and Wendell Lim for their suggestions during the writing of this manuscript. Figures displaying Top7's 3D structure were generated using Pymol (DeLano, 2002). This work was supported in part by grants from NIH-NIGMS (G.V.), NIBIB (G.V.), NIH (T.S. and D.B.), and HHMI (D.B.).

Received: August 17, 2006

Revised: November 17, 2006

Accepted: December 28, 2006

Published: February 8, 2007

REFERENCES

- Abkevich, V.I., Gutin, A.M., and Shakhnovich, E.I. (1995). Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *J. Mol. Biol.* *252*, 460–471.
- Bai, Y. (2003). Hidden intermediates and Levinthal paradox in the folding of small proteins. *Biochem. Biophys. Res. Commun.* *305*, 785–788.
- Baker, D., Sohl, J.L., and Agard, D.A. (1992). A protein-folding reaction under kinetic control. *Nature* *356*, 263–265.
- Baldwin, R.L. (1996). On-pathway versus off-pathway folding intermediates. *Fold. Des.* *1*, R1–R8.
- Bolon, D.N., and Mayo, S.L. (2001). Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. *Biochemistry* *40*, 10047–10053.
- Calloni, G., Taddei, N., Plaxco, K.W., Ramponi, G., Stefani, M., and Chiti, F. (2003). Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J. Mol. Biol.* *330*, 577–591.
- Camarero, J.A., Fushman, D., Sato, S., Girit, I., Cowburn, D., Raleigh, D.P., and Muir, T.W. (2001). Rescuing a destabilized protein fold through backbone cyclization. *J. Mol. Biol.* *308*, 1045–1062.
- Chamberlain, A.K., and Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry* *37*, 1736–1742.
- Chan, H.S. (2000). Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins* *40*, 543–571.
- Dahiyat, B.I., and Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. *Science* *278*, 82–87.
- Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003). A large scale test of computational protein design: Folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* *332*, 449–460.
- Dantas, G., Watters, A.L., Lunde, B.M., Eletr, Z.M., Isern, N.G., Roseman, T., Lipfert, J., Doniach, S., Tompa, M., Kuhlman, B., et al. (2006). Mis-translation of a computationally designed protein yields an exceptionally stable homodimer: implications for protein engineering and evolution. *J. Mol. Biol.* *362*, 1004–1024.
- DeLano, W.L. (2002). The Pymol Molecular Graphics System (<http://www.pymol.org>) (DeLano Scientific).
- Dill, K.A. (1990). Dominant Forces In Protein Folding. *Biochemistry* *29*, 7133–7155.
- Dobson, C.M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* *24*, 329–332.
- Dobson, C.M. (2003). Protein folding and misfolding. *Nature* *426*, 884–890.
- Dobson, N., Dantas, G., Baker, D., and Varani, G. (2006). High-resolution structural validation of the computational redesign of human U1A protein. *Structure* *14*, 847–856.
- Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C., and Radford, S.E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* *286*, 1597–1608.
- Ferguson, N., Schartau, P.J., Sharpe, T.D., Sato, S., and Fersht, A.R. (2004). One-state downhill versus conventional protein folding. *J. Mol. Biol.* *344*, 295–301.
- Ferguson, N., Sharpe, T.D., Schartau, P.J., Sato, S., Allen, M.D., Johnson, C.M., Rutherford, T.J., and Fersht, A.R. (2005). Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *J. Mol. Biol.* *353*, 427–446.
- Flanagan, J.M., Kataoka, M., Shortle, D., and Engelman, D.M. (1992). Truncated staphylococcal nuclease is compact but disordered. *Proc. Natl. Acad. Sci. USA* *89*, 748–752.
- Fukunishi, Y. (1998). Folding-unfolding energy change of a simple sphere model protein and an energy landscape of the folding process. *Proteins* *33*, 408–416.
- Gillespie, B., Vu, D.M., Shah, P.S., Marshall, S.A., Dyer, R.B., Mayo, S.L., and Plaxco, K.W. (2003). NMR and temperature-jump measurements of de novo designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics. *J. Mol. Biol.* *330*, 813–819.
- Go, N. (1984). The consistency principle in protein structure and pathways of folding. *Adv. Biophys.* *18*, 149–164.
- Go, N., and Taketomi, H. (1978). Respective roles of short- and long-range interactions in protein folding. *Proc. Natl. Acad. Sci. USA* *75*, 559–563.
- Govindarajan, S., and Goldstein, R.A. (1995). Optimal local propensities for model proteins. *Proteins* *22*, 413–418.
- Grantcharova, V.P., and Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry* *36*, 15685–15692.
- Gu, H., Yi, Q., Bray, S.T., Riddle, D.S., Shiau, A.K., and Baker, D. (1995). A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* *4*, 1108–1117.
- Horwich, A. (2002). Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J. Clin. Invest.* *110*, 1221–1232.
- Ikai, A., and Tanford, C. (1973). Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. *J. Mol. Biol.* *73*, 145–163.
- Jackson, S.E. (1998). How do small single-domain proteins fold? *Fold. Des.* *3*, R81–R91.
- Jackson, S.E., and Fersht, A.R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* *30*, 10428–10435.
- Jacob, J., Krantz, B., Dothager, R.S., Thiagarajan, P., and Sosnick, T.R. (2004). Early collapse is not an obligate step in protein folding. *J. Mol. Biol.* *338*, 369–382.
- Jemth, P., Gianni, S., Day, R., Li, B., Johnson, C.M., Daggett, V., and Fersht, A.R. (2004). Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc. Natl. Acad. Sci. USA* *101*, 6450–6455.
- Jewett, A.I., Pande, V.S., and Plaxco, K.W. (2003). Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J. Mol. Biol.* *326*, 247–253.

- Kaya, H., and Chan, H.S. (2003). Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* **326**, 911–931.
- Krantz, B.A., and Sosnick, T.R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696–11701.
- Krantz, B.A., Mayne, L., Rumbley, J., Englander, S.W., and Sosnick, T.R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* **324**, 359–371.
- Krishna, M.M.G., and Englander, S.W. (2005). The N-terminal to C-terminal motif in protein folding and function. *Proc. Natl. Acad. Sci. USA* **102**, 1053–1058.
- Kuhlman, B., and Baker, D. (2004). Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struct. Biol.* **14**, 89–95.
- Kuhlman, B., Boice, J.A., Fairman, R., and Raleigh, D.P. (1998). Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding. *Biochemistry* **37**, 1025–1032.
- Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368.
- Ladurner, A.G., Itzhaki, L.S., de Prat Gay, G., and Fersht, A.R. (1997). Complementation of peptide fragments of the single domain protein chymotrypsin inhibitor 2. *J. Mol. Biol.* **273**, 317–329.
- Matouschek, A., and Fersht, A.R. (1991). Protein engineering in analysis of protein folding pathways and stability. *Methods Enzymol.* **202**, 82–112.
- Maxwell, K.L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M.A., Brown, A.G., Friel, C.T., Hedberg, L., Horng, J.C., Bona, D., Miller, E.J., et al. (2005). Protein folding: defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). Scop - A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
- Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
- Plaxco, K.W., Millett, I.S., Segel, D.J., Doniach, S., and Baker, D. (1999). Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat. Struct. Biol.* **6**, 554–556.
- Richardson, J.S., and Richardson, D.C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* **99**, 2754–2759.
- Riddle, D.S., Santiago, J.V., BrayHall, S.T., Doshi, N., Grantcharova, V.P., Yi, Q., and Baker, D. (1997). Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **4**, 805–809.
- Rutherford, S.L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342.
- Sadqi, M., Fushman, D., and Munoz, V. (2006). Atom-by-atom analysis of global downhill protein folding. *Nature* **442**, 317–321.
- Sali, A., Shakhnovich, E., and Karplus, M. (1994). How does a protein fold? *Nature* **369**, 248–251.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucleic Mag. Res. Sp.* **34**, 93–158.
- Scalley, M.L., Yi, Q., Gu, H., McCormack, A., Yates, J.R., and Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry* **36**, 3373–3382.
- Scalley-Kim, M., and Baker, D. (2004). Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J. Mol. Biol.* **338**, 573–583.
- Silow, M., and Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. USA* **94**, 6084–6086.
- Silow, M., Tan, Y.J., Fersht, A.R., and Oliveberg, M. (1999). Formation of short-lived protein aggregates directly from the coil in two-state folding. *Biochemistry* **38**, 13006–13012.
- Srimathi, T., Kumar, T.K., Chi, Y.H., Chiu, I.M., and Yu, C. (2002). Characterization of the structure and dynamics of a near-native equilibrium intermediate in the unfolding pathway of an all beta-barrel protein. *J. Biol. Chem.* **277**, 47507–47516.
- Stefani, M., and Dobson, C.M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678–699.
- van Mierlo, C.P., van den Oever, J.M., and Steensma, E. (2000). Apo-flavodoxin (un)folding followed at the residue level by NMR. *Protein Sci.* **9**, 145–157.
- Wright, C.F., Teichmann, S.A., Clarke, J., and Dobson, C.M. (2005). The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**, 878–881.
- Yahara, I. (1999). The role of HSP90 in evolution. *Genes Cells* **4**, 375–379.
- Zhang, O., Kay, L.E., Olivier, J.P., and Forman-Kay, J.D. (1994). Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* **4**, 845–858.