Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain

Viara P. Grantcharova[1,2], David S. Riddle[1,2], Jed V. Santiago[1] and David Baker[1]

Experimental and theoretical studies on the folding of small proteins such as the chymotrypsin inhibitor 2 (CI-2) and the P22 Arc repressor suggest that the folding transition state is an expanded version of the native state with most interactions partially formed. Here we report that this picture does not hold generally: a hydrogen bond network involving two β-turns and an adjacent hydrophobic cluster appear to be formed in the folding transition state of the src SH3 domain, while the remainder of the polypeptide chain is largely unstructured. Comparison with data on other small proteins suggests that this structural polarization is a consequence of the topology of the SH3 domain fold. The non-uniform distribution of structure in the folding transition state provides a challenging test for computational models of the folding process.

A detailed understanding of how amino acid sequences determine protein three-dimensional structures requires the identification of residues and interactions that play critical roles in the folding process. In previous experiments directed at establishing the minimal sequence requirements for the folding of a small protein, the SH3 domain, a combinatorial library selection strategy was used to obtain functional SH3 domains composed primarily of I, K, E, A and G outside of the binding site[1]. The folding rates of these simplified SH3 variants were found to be very similar to that of the wild type protein, suggesting that the residues critical to folding kinetics must have been conserved in the selection. Here we investigate the roles of the conserved residues (Table 1) in the folding reaction by studying the consequences of alanine substitutions at these positions on the thermodynamics and kinetics of folding. Initially we focused on a non-local hydrogen bonding network involving Glu 30, Ser 47 and Thr 50 (Fig. 1) due to the strong selection pressure observed at these positions (Glu 30 is absolutely conserved and Ser 47 is frequently recovered even though it was not allowed in the mutagenesis strategy). The finding that these residues are important for the kinetics of folding prompted us to extend the analysis to residues throughout the src SH3 domain in order to obtain a more complete picture of the folding transition state.

Contributions to src SH3 stability

The structure of the src SH3 domain consists of two β-sheets orthoganally packed around a hydrophobic core[2,3]. Within the sheets, strands are joined by the RT, n-src and distal loops, while the crossovers between the two sheets occur at a diverging type II β-turn and a short 3_{10}-helix (Fig 1*a*). The residues conserved in the combinatorial library selection were located in the distal loop and diverging turn (Gly 29, Gly 51), the hydrogen bond network between them (Glu 30, Ser 47 and Thr 50; Fig. 1*b*), and the hydrophobic core (Phe 10, Leu 24, Ile 34, Ala 45, Ile 56). Substitutions in the hydrophobic core were most destabilizing, with mutations in the center of the core (Ala 45, Leu 32, Ile 56) having larger effects ($\Delta\Delta G_U$ from 1.8–3.6 kcal mol^{-1}) than mutations at the periphery (Leu 24, Phe 10, Val 61, Ile 34; $\Delta\Delta G_U$ from 0.6–1.8 kcal mol^{-1}). Mutations in the hydrogen bonding network also decreased ΔG_U significantly ($\Delta\Delta G_U$ from 1.8–2.5 kcal mol^{-1}), indicating the importance of these interactions in stabilizing the native state. Each of the mutated residues makes both local and non-local hydrogen bonds and thus the total energetic cost of the mutations is consistent with previous estimates of 1–2 kcal mol^{-1} per hydrogen bond[4]. Partially exposed aromatic residues lining the peptide binding pocket (Trp 42, Tyr 16, Tyr 55) can be viewed as extensions of the hydrophobic core and alanine substitutions at these positions decreased stability. On the other hand, mutation of the completely solvent exposed Tyr 60 to alanine increased ΔG_U, probably by destabilizing non-native conformations in which this residue is partially buried. The remaining mutations probe the integrity of the various loops. Glycine to alanine substitutions in the distal loop (G51) and the diverging turn (G29) decreased stability, while disrupting interactions in the n-src (G40, N36) and RT loops (D15, S18) either did not affect or slightly increased stability. The interpretation that the first two structural elements have more rigid structural requirements than the second two is consistent with our previous finding that amide protons in the distal loop hairpin and diverging turn are more protected from exchange than amide protons in the n-src and RT loops [5].

Kinetics of folding and unfolding

The kinetics of folding of the mutant proteins were characterized using stopped-flow fluorescence. The mutations were found to fall into three categories depending on whether the folding rate (k_f, Fig. 2*a*), the unfolding rate (k_u, Fig. 2*b*), or both were affected (Table 2). It is convenient to use simple transition state theory to interpret the kinetic data; computational models of folding for which the folding rate and the free energy difference between the unfolded state and the transition state ($\Delta G_{U-\ddagger}$) can be determined independently[6,7], suggest that the approximation rate $\cong$ D exp ($-\Delta G_{U-\ddagger}$/RT) provides an excellent estimate of the folding rate (D is the frequency of transitions between related

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [2]Both of these authors contributed equally to this work.

Correspondence should be addressed to D.B. *email: baker@ben.bchem.washington.edu*

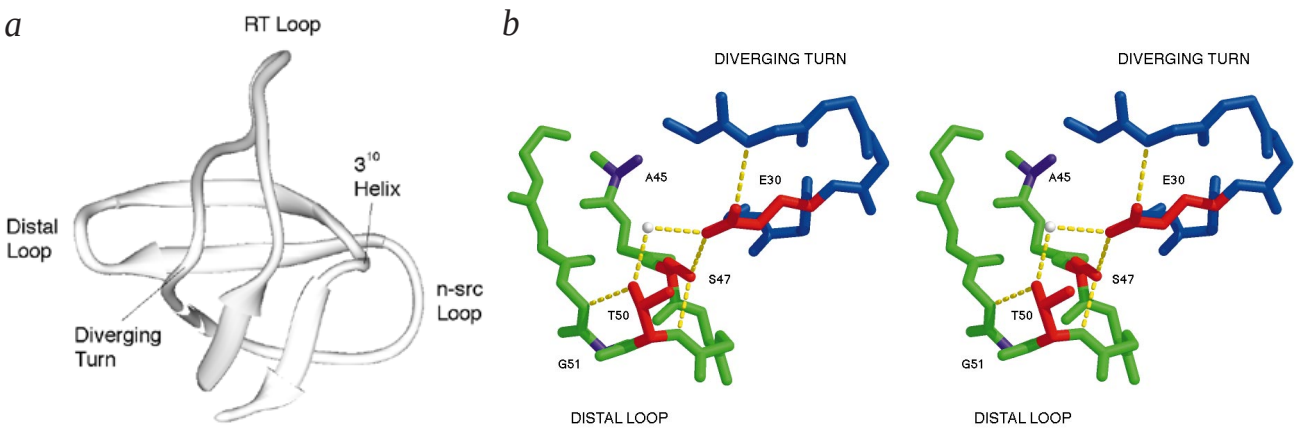

Fig. 1 ***a***, Ribbon diagram of the src SH3 domain crystal structure[2] with the loops and turns labeled. ***b***, Non-local hydrogen bond network connecting the distal loop (Ser 47 and Thr 50) and the diverging turn (Glu 30). Atomic coordinates were taken from the crystal structure of src SH3 domain within the context of the intact tyrosine kinase [2]. A coordinating water molecule may stabilize the interaction. Hydrogen bonding residues are shown in red; the two other residues in this region with high Φ_F values (Gly 51 and Ala 45) are shown in magenta. Both images were created with MidasPlus[38,39].

conformations). A simple but useful interpretation of the kinetic data based on this expression is that mutations which decrease k_f, but do not alter k_u disrupt interactions stabilizing both the transition and the native state while mutations which increase k_u, but do not change k_f disrupt interactions formed after the transition state[8]. A mutation will simultaneously decrease k_f and increase k_u in this model if some but not all of the interactions made by the residue in the native state are also made in the transition state. The structure of the folding transition state for the src SH3 domain deduced from the kinetic measurements using this model is presented below.

Distal loop hairpin. The distal loop hairpin consists of two β-strands connected by a tight β-turn (Fig.1*a*). Several mutations probe the integrity of the hairpin and reveal that it is relatively well structured in the transition state. Gly 51, located in the β-turn, has a positive ϕ angle which is disfavored for all amino acids except glycine; therefore, a substitution with alanine is likely to disrupt formation of the turn. The G51A mutation slowed the folding rate suggesting that the two strands joined by this turn are brought together at the transition state. Ser 47 and Thr 50 also contribute to the structure of the turn by forming local hydrogen bonds between their side chain hydroxyl oxygens and adjacent backbone amide protons (Fig.1*b*). Mutation of these residues significantly slowed the folding rate confirming the near-native structure of this part of the molecule in the transition state. Leu 44 and Tyr 55 interact with each other near the base of the hairpin on the more solvent exposed side; alanine substitutions at these positions slowed the folding rate and increased the unfolding rate, suggesting that the base of the hairpin is partially structured in the transition state.

Diverging type II β-turn. The diverging type II β-turn joining the RT and the n-src loops is stabilized by a hydrophobic interaction between two residues flanking the turn and a local hydrogen bond between the carboxylate of Glu 30 and a backbone amide proton (Fig.1). A local structure prediction program based on a library of recurrent sequence-structure motifs identified this region as the most likely portion of the protein to adopt structure in isolation, and a seven residue peptide corresponding to this turn has been found by NMR to be partially structured in solution[9]. Mutation of Glu 30 and Gly 29 to alanine affected both the folding and the unfolding rate suggesting that these residues have not fully formed all of their contacts in the transition state. However, the interpretation is somewhat complicated by the fact that these mutations are likely to disrupt residual structure in the denatured state.

Hydrophobic core. The hydrophobic residues we have characterized fall roughly into two classes. The first class consists of residues in a hydrophobic cluster formed by the base of the distal loop hairpin and the strand following the diverging turn. Mutations in all of these residues (Ala 45, Ile 34, Leu 32, Ile 56) significantly slowed the folding rate. The second class consists of residues outside of this hydrophobic cluster that are, for the most, partially solvent exposed. Mutations in these residues (Phe 10, Leu 24, Tyr 16, Val 61) had relatively small affects on the folding rate. Taken together, these results suggest that the hydrophobic interactions between the base of the distal hairpin and the strand following the diverging turn are at least partially formed in the folding transition state. The I34A mutation slows both the folding and the unfolding rate suggesting that it destabilizes the transition state more than the native state; the loss of interactions may be partially compensated by structural rearrangements or a relief of strain in the native state.

Hydrogen-bond network. Mutations in the hydrogen bond network residues (S47A, T50A and E30A) significantly reduce the folding rate. To our knowledge, this is the first example of the formation of a hydrogen bond cluster in a folding transition state. As all three residues are involved in both local and non-local hydrogen bonds in the native state it is hard to distinguish conclusively which interactions are important for stabilizing the transition state. However, the large effect of truncations of hydrophobic residues which pack between the distal loop hairpin and the diverging turn (L32A and A45G) on k_f suggests that the distal loop hairpin and diverging turn are not only well structured, but also closely opposed at the transition state. Therefore, it is likely that Glu 30 and Ser 47 are positioned in the proper geometry for the formation of a tertiary hydrogen bond.

Unstructured regions. The remainder of the src SH3 domain appears to be disordered in the transition state. Mutations in the RT loop (D15A, S18A), the n-src loop (N36A, G40A) and the cluster of surface aromatics either had no effect on kinetics

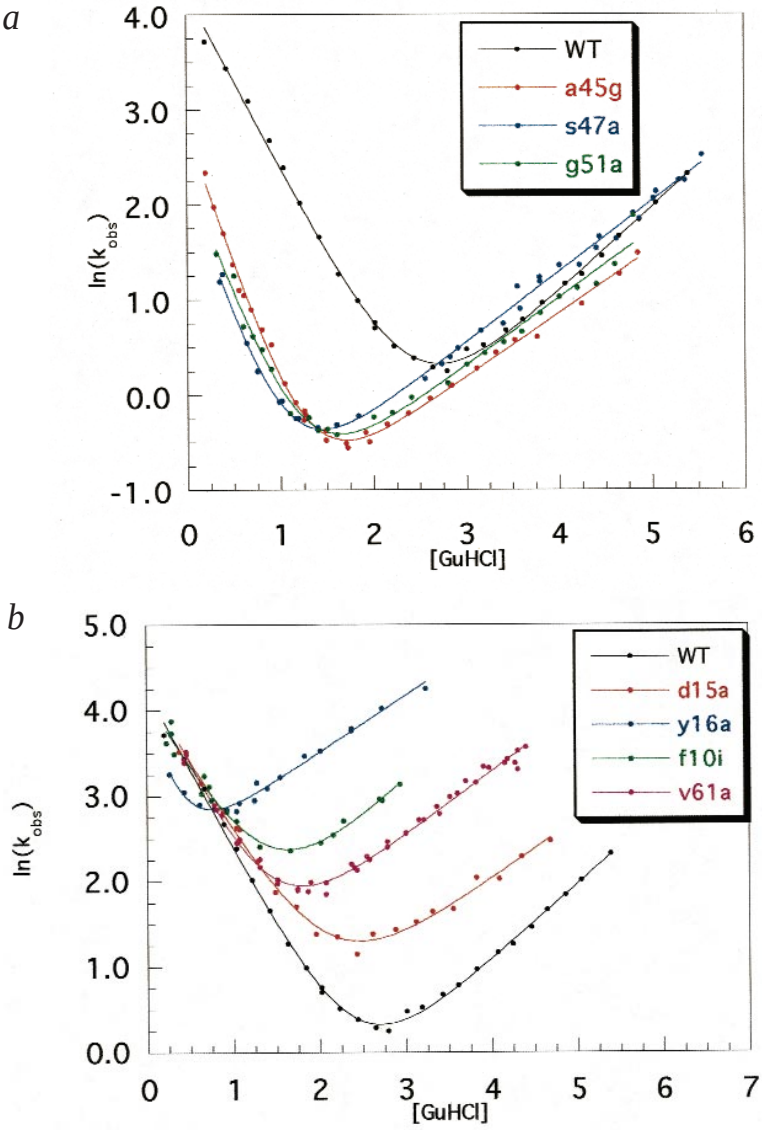

Fig. 2 Dependence of the rate of folding and unfolding on the denaturant concentration for ***a***, mutants which lower k_f and ***b***, mutants which increase k_u. The data for the wild type protein (black) is shown in both panels for comparison. For several of the mutants, the dependence of the folding rate on the guanidine concentration is greater than that of the wild type protein; these mutations may cause some expansion of the denatured state. The color scheme for the mutants is as follows: (*a*) A45G, red; S47A, blue; G51A, green: (*b*) Y16A, blue; F10I, green; V61A, magenta; D15A, red. The solid lines represent the fits to the experimental data.

or exclusively increased the unfolding rate, indicating that these regions do not play a role in guiding the protein towards its folded conformation. These structural elements are located on the opposite side of the molecule from the distal loop hairpin and the diverging turn and constitute the peptide binding site.

Structure of the folding SH3 transition state

To estimate the extent to which an interaction is formed in the transition state, it is convenient to normalize the effect of a mutation on the folding rate by its effect on overall stability (all other things being equal, it is expected that more drastic mutations will have larger effects). The ratio ($\Delta\Delta G_{U-\ddagger}/\Delta\Delta G_{U-F}$), termed the folding Φ_F value by Fersht and co-workers[8], conveniently describes the degree of structure formation around each residue in the transition state. Fig. 3*a* shows a schematic of the src SH3 structure[2] color-coded by Φ_F values on a continuous scale from 1 (red) to 0 (blue). Mutations with Φ_F values close to one (red) affect largely the folding, but not the unfolding rate, and are therefore likely to disrupt interactions stabilizing the transition state. Residues in the distal loop hairpin, diverging turn, and in the hydrophobic core between them have the highest Φ_F values and thus most native-like interactions in the transition state. Moving away from this high Φ_F (red) zone, the Φ_F values gradually decrease to 0 (blue) indicating that the remainder of the protein is largely unstructured in the transition state. The NMR study mentioned earlier showed that the diverging turn is at least partially formed in the denatured protein. The kinetic data reported here suggest that the rate-limiting step in folding involves formation of the distal loop β-hairpin and the docking of the hairpin onto the diverging turn and the strand following it. Once these elements are brought together the three-stranded β-sheet they form could serve as a nucleus around which the RT loop and the two terminal β-strands rapidly assume their native conformation.

NMR studies of the diverging β-turn peptide[9] and of the denatured state of the drk SH3 domain[10] suggest that both turns are partially sampled in the denatured state. If the rate limiting step in folding is a productive collision between these two structural elements, the length of the n-src loop, which connects the diverging turn and the distal loop, should impact the folding rate. The SH3 domain of PI3 kinase has an elongated n-src loop and folds significantly more slowly[11] than any of the other SH3 domains and several of the variants obtained in the SH3 simplification experiment[1] had deletions in the n-src loop. The strong complementarity of these structural elements is vividly illustrated by the recently published structure of the Eps8 SH3 domain dimer[12] in which the distal loop of one molecule is paired with the diverging turn of another.

The segments of the src SH3 domain found to be most structured in the transition state also include the residues most protected from hydrogen-deuterium (HD) exchange with the solvent[5]. There is not a residue-by-residue correspondence — mutagenesis probes side chain interactions, while HD exchange reports on hydrogen bonding and solvent accessibility of backbone amide protons. However, the distal hairpin and the diverging turn were indeed among the most highly protected parts of the protein. If the large fluctuations which lead to the unfolding transition state are amplified versions of the local fluctuations which contribute to HD exchange, the highest rates of exchange would be expected in parts of the protein most disrupted in the transition state[13]. A similar correlation has been observed for protein L[14], but not for CI-2[15].

Sequence conservation and Φ_F values

One of the goals of this study was to understand the sequence

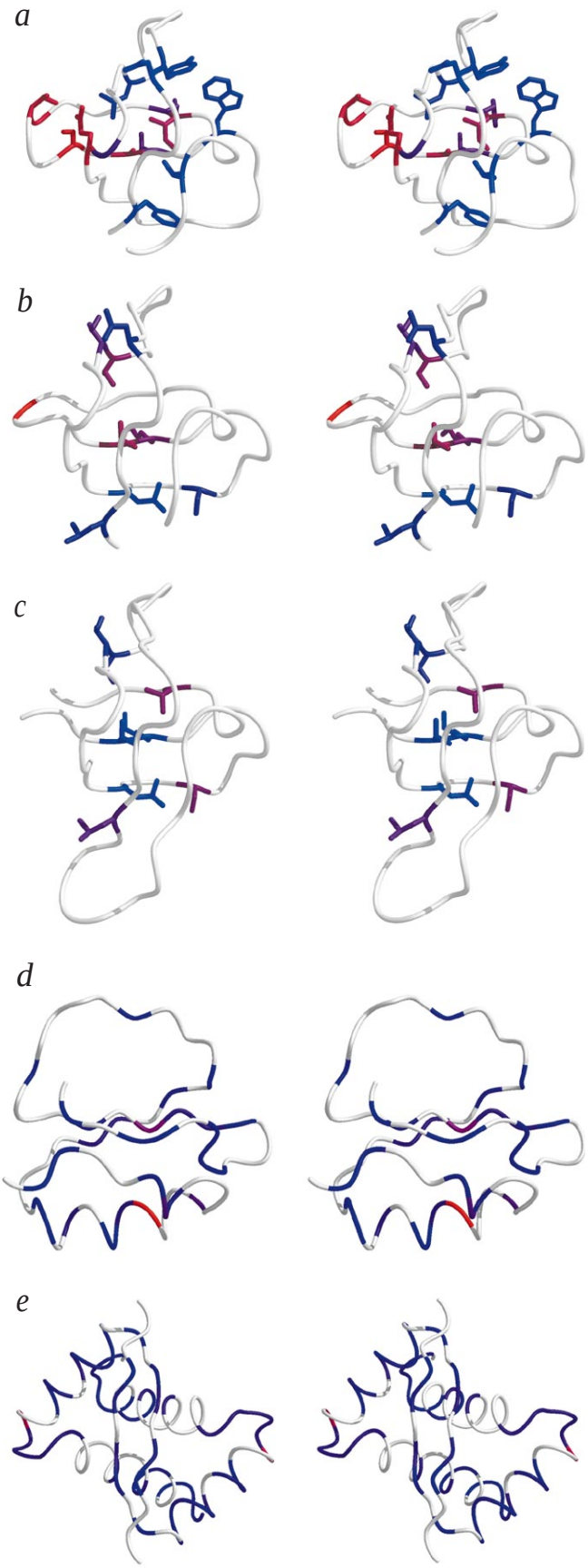

Fig. 3 ***a***, Structures of the src SH3 domain, ***b***, the wt α-spectrin SH3 domain and ***c***, the distal loop permutant, ***d***, CI2 and ***e***, the Arc repressor dimer colored by Φ_F value on a continuous scale from red (1) to blue (0). Experimentally determined values for CI2 were obtained from the original literature[22] and include only alanine substitutions. Φ_F values for Arc repressor were calculated using data in ref. 23 and 25. Two residues that probably make extensive interactions in the src SH3 folding transition state are not colored in (*a*) because of complications in data interpretation: Ile 34 appears to make stronger interactions in the transition state than in the native state (both the folding and unfolding rates are slowed), and Leu 32, which has a substantially decreased folding rate, is so destablized that an accurate Φ_F value could not be obtained. These residues are in the strand following the diverging turn. All images were created with MidasPlus[38,39].

constraints observed in previous phage selection experiments[1]. Two thirds of the conserved residues exhibited high Φ_F values. One exception was Leu 24 (Φ_F = 0.2), however, mutation of this residue to alanine was found to substantially reduce the protein's affinity for the peptide substrate used in the selection (data not shown). When many residues are changed simultaneously, those important for kinetics could be conserved because they play important roles in specifying protein structure. A similar conclusion was reached in a lattice simulation study[16].

It was suggested previously, based on a comparison of Φ_F values for CI-2 and phylogenetic sequence variation within the family of small protease inhibitors homologous to CI-2[16], that residues involved in the rate-limiting step of folding are conserved in evolution. However, for proteins such as CI-2 in which hydrophobic core residues have the highest Φ_F values, it is difficult to disentangle selection for stability from selection for kinetics as core residues generally are the most critical for stability. The SH3 domain is particularly well suited for addressing the relationship between evolutionary conservation and Φ_F values because several residues which make important interactions in the transition state lie outside of the hydrophobic core. We find no relationship between phylogenetic variation and Φ_F values for the SH3 mutants described in this paper (data not shown). In fact, the opposite was observed both with and without exclusion of the binding residues: residues with high Φ_F values were somewhat less conserved than residues with low Φ_F values. The considerably greater opportunity for compensatory mutations in evolution (the replacement of the hydrogen bonding Glu 30, Ser 47 pair in the src SH3 domain by two hydrophobic residues in the PLC γ SH3 domain, for example) compared to the phage selection may account for the differences in the correlation between Φ_F values and sequence conservation in evolution and in the phage selection.

Topology and transition state structure

A previous study on the α-spectrin SH3 domain assessed the importance of topology in determining the structure of the folding transition state[17]. Circular permutants had very similar native structures, but displayed substantially different distributions of Φ_F values (Fig. 3*b*,*c*). The distal loop permutant was particularly affected, consistent with our results on the involvement of this structural element in the transition state. Changes in chain connectivity thus alter transition state structure. Our data on the src SH3 domain provides the first opportunity to examine the effect of sequence divergence on the detailed structure of the transition state (the src and α-spectrin SH3 domains are only 34% identical in sequence). The placement of the transition state along the reaction coordinate

Table 1 Positions conserved in the combinatorial mutagenesis selection

Position	% Burial	Amino acids[1] (% observed/% allowed)[2]					
Phe 10	76	Ile (0/70)	Val (67/10)	Leu (0/10)	Phe (33/10)		
Leu 24	69	Ile (0/70)	Val (0/10)	**Leu** (100/10)	Phe (0/10)		
Gly 29	22	Lys (0/25)	Glu (0/25)	Arg (0/25)	**Gly** (100/25)		
Glu 30	53	Lys (0/50)	**Glu** (100/50)				
Ala 45	99	Ile (0/25)	Val (6/25)	**Ala** (94/25)	Thr (0/25)		
Ser 47	79	Ile (0/25)	Val (0/25)	Ala (22/25)	Thr (6/25)	Asn (16/0)	Ser (56/0)
Thr 50	21	Ala (0/25)	Gly (0/25)	Ser (50/25)	Thr (50/25)		
Gly 51	54	Lys (0/25)	Glu (0/25)	Arg (0/25)	**Gly** (100/25)		
Ile 56	99	Ile (67/70)	Val (33/10)	Leu (0/10)	Phe (0/10)		

[1]The residues conserved more than 90% in the selected variants are indicated in bold.
[2]The first number in the parentheses is the percent recovery of the residue in the phage selection experiments, the second number is the percent recovery expected in the absence of selective pressure given the design of the library.

is very similar for the fyn[18], src[5], and spectrin[19] SH3 domains (the m_f/m ratios for the three proteins are within experimental error), suggesting that their transition states may have similar structures. Comparison of our results with the effects of mutations in the α-spectrin SH3 domain[17,20] suggests that this is indeed the case. In earlier studies of point mutants in the spectrin SH3 domain, two mutants which reported on the proximity of the distal loop to the rest of the protein had the highest Φ_F values, and in the accompanying paper by Serrano and coworkers[21], an Asp to Gly substitution in the distal loop β-turn is shown to have a Φ_F value of 1. The similarity between the src and α-spectrin SH3 domain folding transition states (compare Fig. 3a and 3*b*) is remarkable given their significant divergence in sequence and suggests that the exact identities of the amino acids are not important for the structure of the transition state. In fact, the critical hydrogen bonding residue Ser 47 in the distal loop of src SH3 is replaced in α-spectrin by a valine which contributes to the hydrophobic core. Taken together, the observations that the structure of the transition state is (i) altered by changes in topology (circular permutants), but (ii) largely invariant to the large number of substitutions between the SH3 domains strongly support the idea that topology is a dominant determinant of the folding mechanism of this family of proteins.

Previous work on CI-2[22] and the P22 Arc repressor[23] suggested that the transition states of these proteins represent expanded forms of the native state with most interactions partially formed. The picture for src SH3 domain is quite different: its transition state appears to be quite polarized, with one portion of the molecule much more highly ordered than the rest. For both CI-2 and the Arc repressor, the extent to which a residue contributes to the stability of the transition state is roughly proportional to its contribution to native state stability, as evidenced by the linear relationship between $\Delta\Delta G_U$ and $\Delta\Delta G_{U-\ddagger}$ observed in Brønsted plots[22,23]. In contrast, such a plot for the src SH3 mutants (data not shown) is more similar to that of barnase[22,24], a larger protein that folds through an intermediate: the data are scattered between lines with slopes of 0 and 1 indicating non-uniform formation of structure in the transition state.

Comparison of the structures of CI-2 and Arc repressor colored by Φ_F value (Fig. 3*d*,*e* respectively) with that of the src SH3 domain (Fig. 3*a*) accentuates the difference between the transitions states of these proteins. Both CI-2 and Arc repressor show a relatively uniform distribution of low (blue) and intermediate (magenta) Φ_F values, while the src SH3 domain is split into a high Φ_F value region (red) and a low Φ_F value (blue) one. Although a greater number of mutants was generated for CI-2 and the Arc repressor, the absolute number of residues with Φ_F values greater than 0.5 is still larger for src SH3 than for the two other proteins: six for src SH3 domain *versus* one for Arc repressor[23,25] and two for CI2[22]. The unusual features of the src SH3 transition state may reflect the dominant effect of topology in specifying the transition state structure. Unlike the other proteins whose transition states have been characterized, the SH3 domain consists predominantly of β-sheets. The rate-limiting step in folding may involve docking of transiently formed structural elements; in β-sheet proteins such local structure is likely to occur in the vicinity of β-turns, and in helical proteins, near the middle of helices. Thus, the folding transition states of β-sheet proteins may be expected to be more polarized and not as centered on the hydrophobic core as those of helix containing proteins. More generally, the topology of β-sheet proteins is to some extent determined by the positions and changes in chain orientation in the β-turns (in the SH3 domain, the diverging turn is one of the two transitions between the sheets, and the distal loop β-turn sets up the hydrophobic contacts along the distal loop β-hairpin), and thus formation of a small number of critical β-turns could be coupled to the formation of sufficient favorable native interactions to overcome the entropic barrier to folding. The polarization may also reflect the importance of hydrogen bonds in stabilizing the transition state: hydrogen bonds have much stronger orientational constraints than hydrophobic interactions and are not likely to be stabilizing unless almost fully formed.

In the past several years there has been considerable discussion of the differences between the 'new' and the 'classical' views of protein folding[26–29]. For small, single domain proteins with kinetics and thermodynamics well described by a two-state model, the distinction primarily concerns the breadth of the transition state ensemble: in the classical view, the transition state consists of a relatively well defined set of conformations, whereas for the funnel shaped energy landscapes suggested by the new view, the set of conformations can be extremely diverse (in the limit of the models of ref. 6, the transition state ensemble includes all conformations with a particular degree of order). It is important to note that the transition state approximation is valid independent of the homogeneity (or lack thereof) of the transition state, and that both old and new views are consistent with the simple exponential kinetics observed for the folding of small proteins. In the sequence simplification experiments[1], the folding rate was relatively

Table 2 Kinetic and thermodynamic parameters for wt SH3 domain and mutants[1]

Name	ΔG^{H_2O} (kcal mol^{-1})	$k_f^{0.3}$ (s^{-1})	$k_u^{3.5}$ (s^{-1})	$\Phi_F^{H_2O}{}_{kin}$	$\Phi_F^{H_2O}{}_{eq}$	$\Phi_F^{0.3}{}_{Cm}$
WT	3.7	39	2.1	—	—	—
F10I	2.7	39	32	0.05	0.08	0.00
D15A	3.3	42	5.5	0.01	0.02	-0.09
Y16A	0.9	24	91	0.10	0.11	0.08
S18A	4.1	46	1.2	*[2]	*	*
L24A	1.9	17	21	0.21	0.25	0.20
G29A	2.1	12	10	0.33	0.39	0.29
E30A	1.8	4.6	7.4	0.79	0.67	0.67
L32A	0.1	4.8	9.6	^[3]	^	^
I34A	3.0	5.1	0.4	5.5	1.61	1.31
N36A	4.0	45	3.3	*	*	*
G40A	3.8	20	1.8	*	*	*
W42A	2.5	27	12	0.07	0.08	0.11
L44A	1.5	7.8	9.3	0.37	0.35	0.34
A45G	2.2	6.9	1.8	0.72	0.58	0.54
S47A	1.2	2.9	2.4	0.87	0.60	0.75
T50A	1.9	3.2	3.5	0.73	0.72	0.57
G51A	2.0	4.7	1.9	0.77	0.67	0.64
Y55A	1.9	8.8	7.5	0.39	0.43	0.35
I56A	1.4	3.9	4.7	0.58	0.49	0.49
Y60A	4.1	40	1.2	*	*	*
V61A	3.1	40	18	-0.01	-0.01	-0.01

[1]All experiments were performed at 295 K using 50 mM sodium phosphate as buffer. $k_f^{0.3}$ is the folding rate in 0.3 M guanidine, $k_u^{3.5}$, the unfolding rate in 3.5 M guanidine. Details on the three different methods for calculating Φ_F values are described in the Methods section.
[2]These mutants (*) were more stable than wt SH3.
[3]Due to the very low stability of this mutant (^) thermodynamic and kinetic parameters are only rough estimates.

unchanged by drastic changes in the sequence, consistent with a simple funnel picture in which interactions stabilizing the native state also stabilize partially folded conformations. The clustering of mutations which primarily affect the folding rate (have high Φ_F values) in the distal loop and diverging turn suggest that the transition state ensemble for the src SH3 domain is relatively well defined and thus that the folding funnel departs considerably from symmetry. Taken together, our data suggest that the folding free energy landscape of the src SH3 domain is somewhere between that envisioned in the classical view of folding and the extreme of a completely symmetrical funnel.

Protein folding landscapes could in principle deviate from symmetry either because of heterogeneities in inter-residue contact energies or because of asymmetries in the folded structure. The robustness of the SH3 domain transition state to large changes in sequence (and hence changes in the residue-residue interaction energies) indicated by the strong similarities between the src and spectrin SH3 transition states and the near wild type folding rates of the simplified SH3 variants, and the contrast with the more delocalized transition states of Arc repressor and CI-2 suggest that topological features of the SH3 domain fold rather than heterogeneities in the contact energies are likely to be responsible for the departure from symmetry. The importance of topology in determining folding mechanism is further highlighted by a comparison to λ repressor[30]; two relatively subtle Gly to Ala substitutions in one of the helices of this protein caused a much larger change in m_f/m (from 0.4 to 0.8) than the very large number of sequence changes between src SH3, fyn SH3, spectrin SH3, and the simplified SH3 variant FP1 (ref. 1) (m_f/m = 0.69, 0.68, 0.69 and 0.63, respectively). By analogy with these results, differences in topology may also underlie the differences in folding scenarios derived from studies of lattice models of proteins (the delocalized nuclei of ref. 31 versus the specific nucleus of ref. 32, for example; it was anticipated that the delocalized nuclei scenario may be more common for small helical proteins[31]). The recent finding that the folding rates and m_f/m ratios for small single domain proteins are correlated strongly with the average separation between contacting residues suggests that the relationship between topology and folding mechanism is quite general[33].

Our data present a challenge for methods that seek to predict folding transition state structure using molecular dynamics[34] and other computational approaches[31,35]. Agreement between theory and experiment has been claimed in a number of cases involving CI2, but for this protein it is only necessary to predict that most interactions are partially formed to achieve reasonable success. The highly polarized src SH3 transition state will provide a much more rigorous test of computational models as it requires the precise identification of crucial residues. To make at least part of the test blind, we are currently determining Φ_F values for the remaining residues in the structure and invite predictions of these as a test of computational models of protein folding.

Methods

Mutagenesis and purification. The SH3 gene was cloned into the NdeI and BamHI sites of the pET 15b expression vector (Novagen). Mutagenesis was accomplished using the Quick Change Site-Directed mutagenesis kit (Stratagene). Plasmids harboring the point mutations were transformed into BL21 cells, and protein was overexpressed and purified[5]. The His•Tag® was not removed for the purposes of this study. All mutants were sequenced to ensure that the mutagenesis was successful and the purified proteins were analyzed by mass spectrometry to confirm that each mutation was the expected one.

Biophysical analysis. In all experiments, proteins solutions were made in 50 mM sodium phosphate (JT Baker), pH 6, and the temperature was held constant at 295 K. The stability of the point mutants was assessed by guanidine denaturation using either CD or fluorescence as described[36]. The kinetics of folding and unfolding were followed by fluorescence on a Bio-Logic SFM-4 stopped-flow instrument[36]. The unfolding reaction for the wild type protein was well modeled as a two-state process [5], and the kinetic and equilibrium data for the mutants were fit to a two-state model.

Φ value analysis. There are several different ways of measuring the values of $\Delta\Delta G_{U\text{-}F}$ and $\Delta\Delta G_{U\text{-}‡}$ that determine Φ_F. Because of the possible errors introduced by extrapolation we report three estimates of the Φ_F value for all the mutants destabilized by more than 0.5 kcal mol^{-1}: (i) for $\Phi_F^{H_2O}{}_{kin}$, both $\Delta\Delta G_{U\text{-}F}$ and $\Delta\Delta G_{U\text{-}‡}$ were computed from kinetic data extrapolated to H_2O[15]; (ii) for $\Phi_F^{H_2O}{}_{eq}$, $\Delta\Delta G_{U\text{-}F}$ was computed from equilibrium data and $\Delta\Delta G_{U\text{-}‡}$ from kinetic data extrapolated to H_2O; (iii) $\Phi_F^{0.3}{}_{Cm}$, $\Delta\Delta G_{U\text{-}F}$ was computed using (ΔC_m m_{avg}) and $\Delta\Delta G_{U\text{-}‡}$ from the folding rate in 0.3 M

guanidine[37]. The values obtained by the three methods match very closely confirming the validity of our results. Small differences are seen only in the significantly destabilized mutants for which estimates of the equilibrium ΔG_U are not very accurate due to the lack of a folded baseline.

Acknowledgments

We thank M. Eck for providing us with the atomic coordinates of src tyrosine kinase prior to submitting them in the Brookhaven protein data bank, Q. Yi for mass spectrometry analysis of all the SH3 mutants, J. Onuchic and members of the Baker group for useful comments on the manuscript, and L. Serrano and coworkers for sharing their manuscript on the spectrin folding transition state prior to publication. This work was supported by a grant from the Office of Naval Research and Young Investigator awards to D. B. from the NSF and the Packard Foundation.

Received 11 March, 1998; accepted 25 June, 1998.

1 Riddle, D. S., Santiago, J. V., Bray, S. T., Doshi, N., Grantcharova, V. P. & Baker, D. Functional rapidly folding proteins from simplified amino acid sequences. *Nature Struct. Biol.* **4**, 805–809 (1997).
2. Xu, W., Harrison, S. C., & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).
3. Yu, H., Rosen, M. K., & Schreiber, S. L. ^{1}H and ^{15}N assignments and secondary structure of the Src SH3 domain. *FEBS Lett.* **324**, 87–92 (1993).
4. Pace, N. C., Shirley, B. A., Mcnutt, M. & Gajiwala, K. Forces contributing to the stability of proteins. *FASEB* **10**, 75–83 (1996).
5. Grantcharova, V. P. & Baker, D. Folding dynamics of the src SH3 domain. *Biochemistry* **36**, 15685–15692 (1998).
6. Doyle, R., Simons, K., Qian, H. & Baker, D. Local interactions and the optimization of protein folding. *Proteins Struct. Func. Gen.* **29**, 282–291 (1997).
7. Socci, N. D., Onuchic, J. N. & Wolynes, P. G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **104**, 5860–5868 (1996).
8. Fersht, A. R. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84 (1994).
9. Yi, Q., Bystroff, C. & Baker, D. Prediction and structure characterization of an independently folding substructure in the src SH3 domain. *J. Mol. Biol.*, **in the press** (1998).
10. Zhang, O. & Forman-Kay, J. D. NMR studies of unfolded states of an SH3 domain in acqueos solution and denaturing conditions. *Biochemsitry* **36**, 3959–3970 (1997).
11. Kishan, K. V. R., Scita, G., Wong, W. T., Di Fiore, P. P. & Newcomer, M. E. The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nature Struct. Biol.* **4**, 739–743 (1997).
12. Guijarro, J. I., Morton, C., Plaxco, K. W., Campbell, I. D. & Dobson, C. M. Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J. Mol. Biol.* **276**, 657–667 (1998).
13. Woodward, C. Is the slow-exchanging core the protein folding core? *TIBS* 18, 359–**360** (1993).
14. Gu, H., Kim, D. & Baker, D. Contrasting roles for the symmetrically disposed β-turns in the folding of a small protein. *J. Mol. Biol.* **274**, 588–596 (1997).
15. Neira, J. L., Itzhaki, L. S., Otzen, D. E., Davis, B. & Fersht, A. R. Hydrogen exchange in chymotrypsin inhibitor 2 probed by mutagenesis. *J. Mol. Biol.* **270**, 1–12 (1997).
16. Shakhnovich, E., Abkevich, V. & Ptitsyn, O. Conserved residues and the mechanism of protein folding. *Nature* **379**, 96–98 (1996).
17. Viguera, A.R., Serrano, L. & Wilmanns, M. Different folding transition states may result in the same native structure. *Nature Struct. Biol.* **3**, 874–879 (1996).
18. Plaxco, K.W., Guijarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D. & Dobson, C. M. The folding kinetics and thermodynamics if the fyn SH3 domain. *Biochemistry* **37**, 2529–2537 (1998).
19. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., & Serrano, L. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry* **33**, 2142–2150 (1994).
20. Prieto, J., Wilmans, M., Jimenez, M. A., Rico, M. & Serrano, L. Non-native interactions in protein folding and stability: introducing a helical tendency in the all β-sheet α-spectin SH3 domain. *J. Mol. Biol.* **268**, 760–778 (1997).
21. Martinez, J. C., Pisabarro, M. T. & Serrano, L. Obligatory steps in protein folding and conformational diversity of the transition state. *Nature Struct. Biol.* **5**, 721–729 (1998).
22. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288 (1995).
23. Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. P22 Arc Repressor: Transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* **34**, 13914–13919 (1995).
24. Serrano, L., Matouschek, A. & Fersht, A. The folding of an enzyme. 3. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805–818 (1992).
25. Milla, M. E., Brown, B. M. & Sauer, R. T. Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nature Struct. Biol.* **1**, 518–523 (1994).
26. Baldwin, R. L. Matching speed and stability. *Nature* **369**, 183–184 (1994).
27. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619–20 (1995).
28. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19 (1997).
29. Pande, V.S., Grosberg, A.Y., Tanaka, T. & Rokhsar, D. S. Pathways for protein folding:is a new view needed? *Curr. Opin. Struct. Biol.* **8**, 68–79 (1998).
30. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. the energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nature Struct. Biol.* **4**, 305–310 (1997).
31. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. Protein folding funnels: the nature of the transition state ensemble *Folding Design* **1**, 441–450 (1996).
32. Abkevich, V. I., Gutin, A.M. & Shakhnovich, E. I. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* **33**, 10026–10036 (1994).
33. Plaxco, K. W., Simons, K. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).
34. Li, A. & Daggett, V. Identification and characterization of the unfolding transition state of chymotypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* **257**, 412–429 (1996).
35. Schoemaker, B. A., Wang, J. & Wolynes, P. G. Structural correlations in protein folding *Proc. Natl. Acad. Sci. USA* **95**, 777–782 (1997).
36. Scalley, M. L. *et al.* Kinetics of folding of the IgG binding domain of peptostreptoccocal protein L. *Biochemistry* **36**, 3373–3382 (1996).
37. Jackson, S. E., Moracci, M., elMarsy, N., Johnson, C. & Fersht, A.R. Effect of cavity-creating mutations in the hdrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32**, 11259–11269 (1993).
38. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. The MIDAS display system. *J. Mol. Graph.* **6**, 13–27 (1988).
39. Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E. & Langridge, R. Conic: a fast renderer for space-filling molecules with shadows. *J. Mol. Graph.* **9**, 230–236 (1991).