1. Mirny, L.A., Abkevich, V.I. & Shakhnovich, E.I. *Proc. Natl. Acad. Sci. USA* **95**, 4976–4981 (1998).
2. Michnick, S.W. & Shakhnovich, E. *Folding & Design* **3**, 239–251 (1998).
3. Plaxco, K.W., Simons, K.T. & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).
4. Shakhnovich, E. *Folding & Design* **3**, R108–R111 (1998).
5. Thirumalai, D. & Klimov, D.K. *Folding & Design* **3**, R112–R118. (1998).
6. Martinez, J.C., Pisabarro, M.T. & Serrano, L. *Nature Struct. Biol.* **5**, 721–729 (1998).
7. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. *Nature* **359**, 851–855 (1992).
8. Blanco, F.J., Ortiz, A.R. & Serrano, L. *J. Biomol. NMR* **9**, 347–357 (1997).
9. Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L. & Serrano, L. *Biochemistry* **33**, 2142–2150 (1994).
10. Viguera, A.R., Blanco, F.J. & Serrano, L. *J. Mol. Biol.* **247**, 670–681 (1995).
11. Viguera, A.R., Serrano, L. & Wilmanns, M. *Nature Struct. Biol.* **3**, 874–880 (1996).
12. Viguera, A.R. & Serrano, L. *Nature Struct. Biol.* **4**, 939–946 (1997).
13. Grantcharova, V.P., Riddle, D.S., Santiago, J.V. & Baker, D. *Nature Struct. Biol.* **5**, 714–720 (1998).
14. Fersht, A.R. *Curr. Opin. Struct. Biol.* **5**, 79–84 (1995).
15. Fersht, A.R., Itzhaki, L.S., elMasry, N.F., Matthews, J.M. & Otzen, D.E. *Proc. Natl. Acad. Sci. USA* **91**, 10426–10429 (1994).
16. Riddle D.S. *et al. Nature Struct. Biol.* **6**, 1016–1024 (1999)
17. Chiti, F., *et al*. *Nature Struct. Biol.* **6**, 1005–1009 (1999).
18. Villegas, V., Martinez, J.C., Avilés, F.X. & Serrano, L. *J. Mol. Biol.* **283,** 1027–36 (1998).
19. Kunkel, T.A.. *Proc. Natl. Acad. Sci. USA* **82**, 488–492 (1985).
20. Gill, S.C. & Hippel, P.H. *Anal. Biochem.* **182**, 319–326 (1989).
21. Prieto, J., Wilmans, M., Jimenez, M.A., Rico, M. & Serrano, L. *J. Mol. Biol.* **268**, 760–778 (1997).
22. Johnson, C.M. & Fersht, A.R. *Biochemistry* **34**, 6795–6804 (1995).

Experiment and theory highlight role of native state topology in SH3 folding

David S. Riddle[1–3], Viara P. Grantcharova[1,2], Jed V. Santiago[2], Eric Alm[2], Ingo Ruczinski[2] and David Baker[2]

[1]These authors contributed equally to this work. [2]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [3]Present address: Department of Immunology, Mayo Clinic, Rochester, Minnesota 55904, USA.

We use a combination of experiments, computer simulations and simple model calculations to characterize, first, the folding transition state ensemble of the src SH3 domain, and second, the features of the protein that determine its folding mechanism. Kinetic analysis of mutations at 52 of the 57 residues in the src SH3 domain revealed that the transition state ensemble is even more polarized than suspected earlier: no single alanine substitution in the N-terminal 15 residues or the C-terminal 9 residues has more than a two-fold effect on the folding rate, while such substitutions at 15 sites in the central three-stranded β-sheet cause significant decreases in the folding rate. Molecular dynamics (MD) unfolding simulations and *ab initio* folding simulations on the src SH3 domain exhibit a hierarchy of folding similar to that observed in the experiments. The similarity in folding mechanism of different SH3 domains and the similar hierarchy of structure formation observed in the experiments and the simulations can be largely accounted for by a simple native state topology-based model of protein folding energy landscapes.

Three independent lines of investigation suggest that protein folding rates and mechanisms are largely determined by native state topology[1]. First, dramatic changes in amino acid sequence, produced either in the laboratory[2,3] or by the evolutionary process[4], that do not alter the overall topology of a protein usually have relatively little effect on protein folding rates. Second, comparison of the consequences of mutations on folding kinetics in distantly related homologs suggests that folding transition state structure is conserved despite differences in amino acid sequence and stability[5,6]. Third, the folding rates of small proteins are strongly correlated with a property of the native state topology: the average sequence separation between residues that make contacts in the three-dimensional structure (the contact order)[7]. The influence of native state topology on protein folding rates and mechanisms is a consequence of the relatively large entropic cost of forming nonlocal interactions early in folding: simple topologies with mostly local interactions are more readily formed than those with many nonlocal interactions, and for a given topology, local interactions are more likely to be formed early in folding than nonlocal interactions.

SH3 domains are an ideal system to investigate how topology determines folding mechanisms. Over 400 different naturally occurring SH3 domain sequences have been identified, more than 10 high-resolution structures have been determined, and the stability and folding kinetics of a number of these proteins have been characterized[6,8–14]. We had found that many of the residues conserved in a phage display selection for simplified src SH3 domain variants[2] played an important role in determining the folding mechanism[5]. Kinetic analysis of mutations in 20 of the 57 positions in the protein suggested that the distribution of structure in the transition state ensemble was localized to one portion of the molecule, and that the folding transition state of the src SH3 domain resembled that of the α-spectrin SH3 domain, which has an almost identical topology but only 36% sequence identity[6]. Though suggestive, with less than half of the residues accounted for, these results did not thoroughly characterize the transition state ensemble of the protein, and did not provide an explanation for the similarity in the src and spectrin SH3 transition states.

In this paper we present a combination of experiments, computer simulations, and simple model calculations aimed at detailed characterization of the src SH3 transition state and its structural origins. The experiments fully map out the transition state ensemble by probing the kinetic consequences of mutations of every residue that makes appreciable interactions in the native state. The computer simulation studies assess the robustness of the hierarchy of structure formation to the numerous approximations and likely inaccuracies in computational models of folding. Finally, the simple model calculations probe the topological features that determine the way SH3 domains fold. Our results provide perhaps the most comprehensive picture of the rate-limiting step in folding of an all β-sheet protein available to date.

Experimental studies

The SH3 domain is a 57-residue globular protein that consists of two antiparallel β-sheets orthogonally packed to form a single hydrophobic core (Fig. 1). Here we describe the effects of mutations of all residues more than 10% buried in the native

structure (52 of 57 residues in the protein) on the rates of folding and unfolding, and the picture of the folding transition state that emerges from these data.

The method we employ was pioneered by Fersht and coworkers[15] and has emerged as the predominant experimental procedure for the detailed characterization of folding transition states[16–20]. The extent to which a residue's interactions are formed in the transition state is summarized by the Φ_f value ($\Delta\Delta G_{u-\ddagger}/\Delta\Delta G_{u-f}$) , which is the change in the free energy of the transition state brought about by mutation of the residue normalized by the change in overall stability[15]. A Φ_f value of 1 indicates that all of a residue's interactions are formed in the transition state, whereas a Φ_f value of 0 means that the residue does not make stabilizing interactions in the transition state. Intermediate Φ_f values indicate partially formed interactions or interactions formed in a fraction of the transition state ensemble; the relationship between the actual Φ_f value and the extent of structure formation is not necessarily linear. As emphasized by Fersht and coworkers[16], the most straightforward class of mutations to interpret are those that remove a small number of methyl groups, such as isoleucine to valine, alanine to glycine, and valine to alanine, as these are least likely to change the folding mechanism and the structure of the folded and unfolded states. In this study, we have also mutated polar residues to alanine to examine the role of polar interactions and hydrogen bonds in the transition state, and have substituted glycine residues with alanine to probe turn formation in the transition state. To guard against possible artifacts due to changes in denatured state structure and/or folding mechanism, we draw conclusions only from results that are consistent among a number of neighboring residues.

To facilitate presentation of our results, we have divided the src SH3 domain into five structural regions and discuss them in order of increasing importance in the folding transition state.

N- and C-terminal strands (strands 1 and 5) and 3_{10}-helix. The N- and C-termini of the SH3 domain come together to form an antiparallel β-sheet stabilized by nonlocal side chain–side chain interactions (Fig. 1 and Table 1). A short 3_{10}-helix (PSNY, residues 57–60) precedes strand 5 and is responsible for the 90° transition from one sheet to the other. It is remarkable that almost all mutations in this region (12 out of 14) either exclusively affect the unfolding rate or do not change protein stability (Fig. 2*a*,*b*).

The extremely low Φ_f values (Fig. 1, Table 2) suggest that the N- and C-termini are largely unstructured in the transition state ensemble.

RT Loop. Residues 14–25 (YDYESRTETDLS) form the large, relatively disordered RT loop (Fig. 1), which is functionally important for binding proline-rich peptides. The crystal structure of the src SH3 domain reveals a small stretch of regular β-sheet pairing within the loop, as well as quite a few intraloop hydrogen bonds involving the side chains of D15, S18, D23 and S25 (Table 1). Hydrogen/deuterium (HD) exchange experiments[8] indicate, however, that this part of the molecule is flexible. As with the N- and C-termini, almost all mutations (eight out of nine) in the RT loop have Φ_f values close to 0 (Fig. 2*c*, Table 2). L24A is the only mutation in the RT loop that lowers k_f, but its predominant effect is still on k_u. The RT loop and the N- and C-termini are clearly the parts of the SH3 domain that are least structured in the transition state (Fig. 1).

Diverging type II β-turn. The transition from the RT loop to the central three-stranded sheet formed by the n-src loop and the distal loop β-hairpin is made by the diverging turn (FKKGERLQ, residues 26–33) (Fig. 1). It is stabilized by hydrophobic contacts between the central core residues, F26 and L32, and a hydrogen bond between the side chain carboxyl of E30 and the backbone amine of K27.

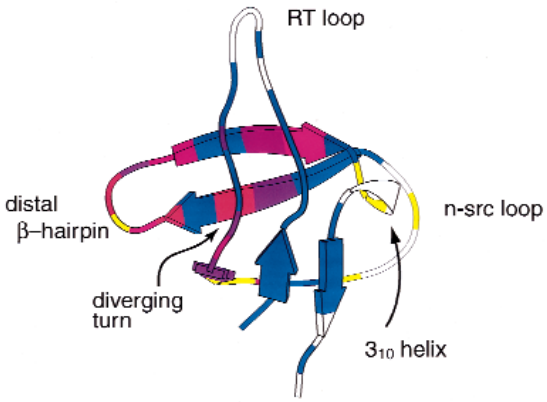

Fig. 1 Structure[35] of the src SH3 domain colored by Φ_f value from red (1) to blue (0). Residues colored in white were either not mutated or the mutation did not affect ΔG significantly. Residues colored in yellow increased or decreased both k_f and k_u, suggesting that these mutations affect the transition state more than the native state. Φ_f values were calculated as described in the Methods. The image was created using Molscript[36].

All of the structurally important residues in this region have intermediate Φ_f values (Fig. 2*e*, Table 2). NMR studies[21] of an isolated peptide with the sequence FKKGERL suggest that the diverging turn conformation is partially populated in the denatured state. Thus, the interactions made by the diverging turn residues in the transition state may be greater than indicated by the Φ_f values, since the reference state (the denatured state) is already partially ordered. Recent double-mutant experiments (V.G. and D.B., unpublished results) suggest that the additional interactions made by the diverging turn in the transition state include a nonlocal hydrogen bond network involving E30 in the diverging turn and S47 and T50 in the distal β-hairpin. The partial Φ_f values of the core residues F26 and L32 suggest that these residues also make some of their interactions with hydrophobic residues in the distal loop β-hairpin in the transition state.

N-src loop. The n-src loop (IVNNTEGDWW, residues 34–43) (Fig. 1) has an unusual shape: the two end residues, I34 and W43, are part of the hydrophobic core whereas the intervening sequence forms a large, almost rectangular turn around W43. W42 is only peripherally associated with the core and together with W43 lines the peptide binding site. There is limited local hydrogen bonding within the n-src loop, and two nonlocal hydrogen bonds connect it to the 3_{10}-helix (Table 1).

The large number of mutations with unusual kinetic consequences suggests that this region may adopt nonnative conformations in the transition state (Fig. 2*d*, Table 2). I34 is a central hydrophobic core residue with many neighbors (Table 1), yet neither I34A nor I34V affect stability significantly (ΔΔG -0.33 and -0.09 kcal mol^{-1}, respectively). Kinetic analysis shows that the two I34 mutants slow both the folding and unfolding rates simultaneously, suggesting that the mutations destabilize the transition state more than the native or denatured states. I34 appears to be critical for core formation during folding, but strained in the native state because of slight overpacking of

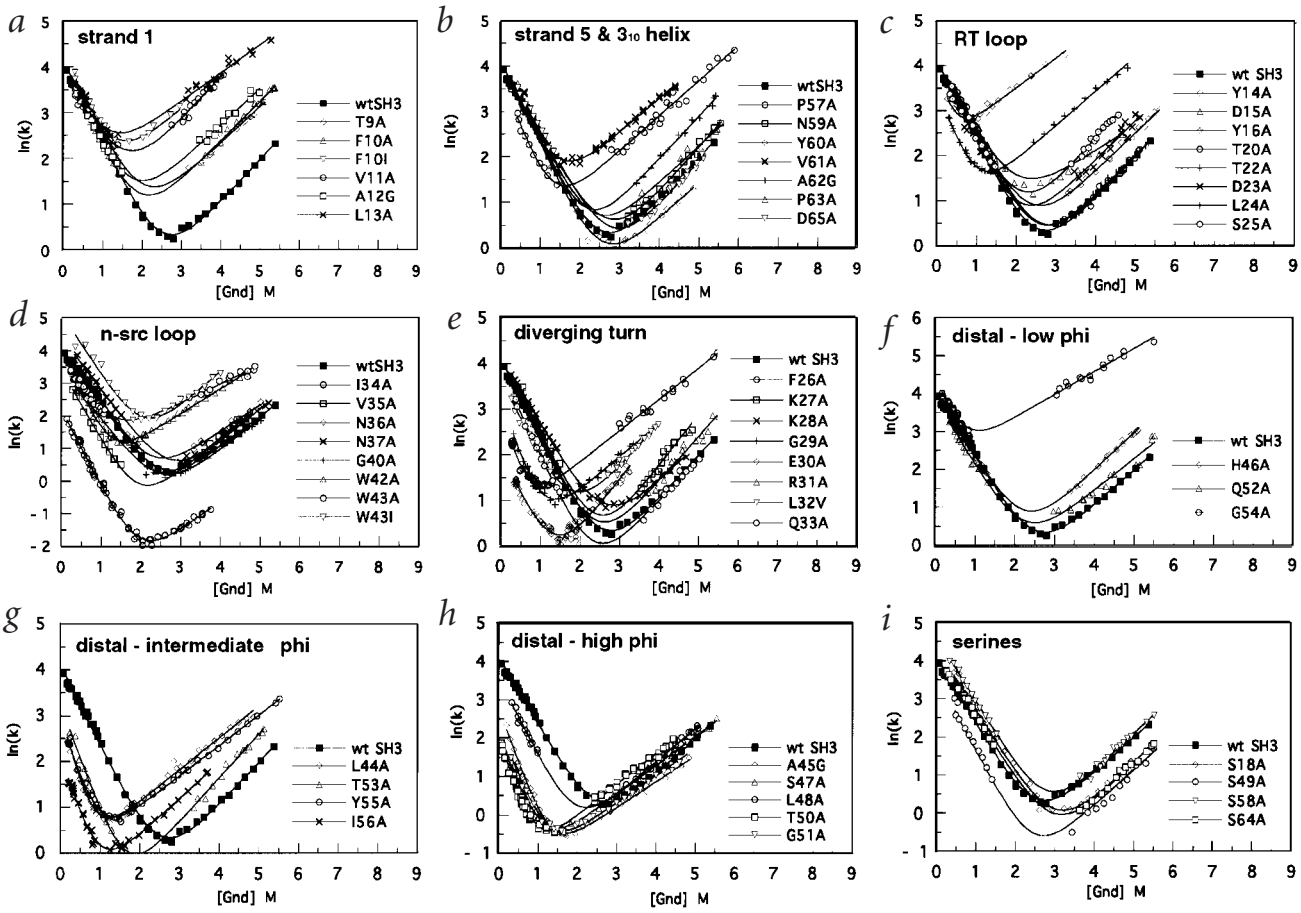

Fig. 2 Dependence of the rate of folding and unfolding on the denaturant concentration for all the mutants grouped into structural regions (***a–h***) as shown in Fig. 1. ***i***, Serine to alanine substitutions with unusual behavior. The data for the wild type (wt) protein (■) is shown in all panels for comparison. The solid lines represent the fits to the experimental data.

the hydrophobic core in the native src SH3 domain. Overpacking is most likely due to burial of the bulky W43 in the native state, but not in the transition state ($\Phi_f = 0.15$ for the W43A mutation). On the solvent-exposed side, V35A is the only mutation for which an unambiguous Φ_f value can be calculated (0.77); its interaction with L44 in the distal β-hairpin appears to be partially formed in the transition state. The N37 side chain appears to make unfavorable interactions in the transition state as the N37A mutation speeds both folding and unfolding. Chain reversal at the tip of the n-src loop appears to be important at the transition state as the G40A mutation, which stiffens the chain, slows both k_f and k_u. As mutations in I34, N37 and G40 appear to selectively stabilize or destabilize the transition state relative to the native and unfolded states, it seems likely that the residues in the n-src loop are ordered at the transition state, but in a nonnative conformation (perhaps a tight hairpin, rather than the distorted loop present in the native state).

Distal β-hairpin. Strands 3 (LAHS, residues 44–47) and 4 (RTGYI, residues 52–56) form the distal β-hairpin, the most regular element of secondary structure in the SH3 domain (Fig. 1). They are connected by a tight type I β-turn and stabilized by numerous backbone and side chain hydrogen bonds, including an extensive network of hydrogen bonds among the turn residues S47 and T50 and the peptide backbone (Table 1).

Mutations throughout the distal β-hairpin can be grouped into three categories based on their effects on kinetics (Table 2). Mutations with Φ_f values of 0 (Fig. 2*f*) include H46A and Q52A (both exposed polar residues) and G54A. Among the residues with intermediate Φ_f values (Fig. 2*g*, Table 2) L44, T53 and Y55 interact at the solvent-exposed side of the hairpin and appear to be only partially associated at the transition state. I56, on the other hand, is an integral part of the hydrophobic core and intimately involved in the transition state as judged by the large decrease in k_f upon mutation to alanine; it does, however, make additional interactions after the transition state as well ($\Phi_f = 0.71$). Mutations with Φ_f values of 1 are clustered around the turn (S47A, L48A, T50A and G51A) or are part of the hydrophobic core (A45G) (Fig. 2*h*), suggesting that the β-turn is fully formed in the transition state and the center of the β-hairpin is associated with the hydrophobic core. As mentioned earlier, residues S47 and T50 also make nonlocal interactions with the diverging turn. S49 (Fig. 2*i*) might also take part in this hydrogen bond network at the transition state as the S49A mutation decreases k_f; the decrease in k_u brought about by the mutation may be due to partial burial of the -OH group in the native state without a suitable hydrogen bonding partner.

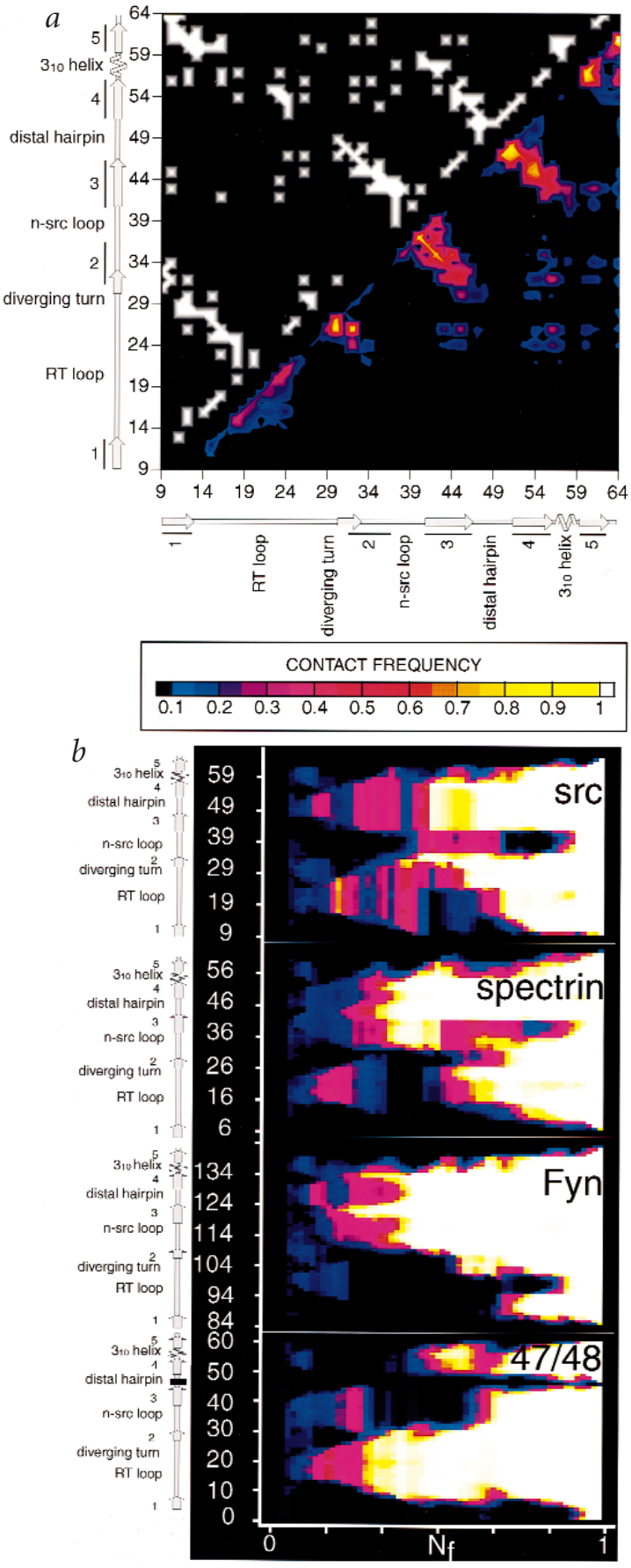

Fig. 3 Theoretical analysis of SH3 folding. ***a***, *Ab initio* simulation of src SH3 folding using ROSETTA. The folding of the src SH3 domain was simulated using ROSETTA as described in Methods. All SH3 domain structures were removed from the data base of short fragments used for building up conformations. A total of 500 independent simulations were carried out, and all conformations from the 20 trajectories that produced structures within 4.5 Å r.m.s.d. of the native structure were combined to calculate the frequency of side chain–side chain contacts for each pair of residues in the protein (lower right triangle; color scheme is shown below the figure). For comparison, the contact distribution in the native structure is shown in the upper left. ***b***, Hierarchy of SH3 domain folding in model calculations based on native state topology. Calculations were performed on the src, spectrin and fyn SH3 domains and the 47–48 circular permutant of the spectrin SH3 domain in which the distal hairpin has been cut. The reaction coordinate, N_f, is the fraction of ordered residues ($N_f = 0$ is the fully unfolded state and $N_f = 1$ is the fully folded state). The y-axis indicates position along the sequence. All configurations of the system were enumerated, and the Boltzmann averaged frequency of ordering of each residue, as a function of N_f, is indicated by the color (black-blue, 0–0.25; blue-magenta, 0.25–0.50; magenta-red, 0.50–0.75; red-yellow, 0.75–0.88; and yellow-white, 0.88–1.00). The top panel was shown in Alm and Baker[31]. It is important to note that segments of the protein not contiguous along the sequence still interact in the model if contacting in the three-dimensional structure, for example in the top panel, the high population of the diverging turn/strand 2 and the distal loop β-hairpin at $N_f = 0.6$ indicates that more surface area is buried within and between these structural elements than within any other substructure with the same number of residues ordered in the protein.

Putting the pieces together, the following picture of the transition state of the src SH3 domain emerges (Fig. 1). The distal β-hairpin is the most ordered structural element in the transition state. The diverging turn and strand 2 are partially ordered and interact with residues in the distal β-hairpin, and this effectively constrains the n–src loop and specifies the three-stranded topology of the central β-sheet in the protein. The clustering of mutations that selectively stabilize or destabilize the transition state in the vicinity of the n-src loop (Fig. 1, yellow) suggests that the loop may have a nonnative configuration in the transition state. In contrast, the two terminal strands, the RT loop and the 3_{10}-helix are mostly unstructured and contribute few stabilizing interactions in the transition state.

Because of the complexities associated with interpreting any one mutation, consistency within a large set of mutations is critical for constructing a plausible picture of structure in the transition state. The segment of sequence between residues 26 and 58 contains only 28 of the 43 positions for which mutations significantly affected the rate of folding and/or unfolding, but this segment contains 25 of the 27 positions with either (i) Φ_f values greater than 0.15 or (ii) mutations that selectively stabilize or destabilize the transition state (Table 2). The probability of such a partitioning if the observed Φ_f values were randomly distributed in the sequence is 1 in 758,000.

Computer simulations

To further elucidate the hierarchy of structure formation in the SH3 domain, we compare our experimental findings with the results from two complementary computational methods: recently published molecular dynamics (MD) simulations of src SH3 domain

Table 1 Interactions in the native state of src SH3 domain[1]

	Residue	Burial %	Hydrophobic interactions[2]	sm H bonds[3]	ss H bonds[4]
β1	T9	13	11, 31, 48, 64		33
	F10	76	32, 34, 37, 43, 61, 63		
	V11	64	9, 13, 31, 64		
	A12	100	26, 30, 32, 61		
	L13	54	11, 14, 62		
RT loop	Y14	44	13, 28, 60		
	D15	48	17, 28	28	
	Y16	85	19, 23, 26, 42, 56, 57, 60		23
	E17	25	15		
	S18	87	23	25	
	R19	8	16, 20, 42	23	
	T20	39	19, 23	22, 23	
	E21	7	22		
	T22	26	21, 55	20	
	D23	79	16, 18, 20, 42, 55	19, 20	16
	L24	69	25, 26, 32, 45, 47, 52, 56		
	S25	51	24	18	
	F26	90	12, 16, 24, 30, 32, 56, 57, 60, 61		
diverging turn	K27	34	30, 50	30	
	K28	30	14, 15	15	
	G29	22			
	E30	53	12, 26, 27, 32, 47, 49, 50	27	47
β2	R31	21	9, 11, 64		
	L32	94	10, 12, 24, 26, 30, 34, 45, 47, 56, 61		
	Q33	46	35, 46, 48		9
n–src loop	I34	77	10, 32, 37, 43, 45, 56, 61		
	V35	51	33, 36		
	N36	28	35, 38		
	N37	53	10, 34, 43		
	T38	7	36, 43	43	
	E39	21			
	G40	33		58	
	D41	43	42		
β3	W42	50	16, 19, 23, 41, 55, 57		
	W43	92	10, 34, 37, 38, 56, 58, 61	38	
	L44	83	53, 55		
	A45	99	24, 32, 34, 56		
	H46	68	33, 48, 53		
	S47	79	24, 30, 32, 52	49, 50	30
distal hairpin	L48	36	9, 33, 46		
	S49	12	30, 50	47	
	T50	21	27, 30, 49, 52	47	
	G51	54			
	Q52	44	24, 47, 50		
β4	T53	53	44, 46		
	G54	94			
	Y55	62	22, 23, 42, 44		
	I56	99	16, 24, 26, 32, 34, 43, 45, 57, 61		
3_{10} helix	P57	96	16, 26, 42, 56, 60		
	S58	87	43	40	
	N59	16	60		
	Y60	63	14, 16, 26, 57, 59, 61		
β5	V61	95	10, 12, 26, 32, 34, 43, 56, 60		
	A62	77	13, 63		
	P63	34	10, 62		
	S64	0	9, 11, 31, 65		
	D65	0	64		

[1]The crystal structure of the src tyrosine kinase[35] was used as a model for the native state of the SH3 domain (1fmk).
[2]Hydrophobic interactions were defined using the Voronoi procedure[22].
[3]Main chain–side chain hydrogen bonding.
[4]Side chain–side chain hydrogen bonding.

unfolding[22], and folding simulations of the src SH3 domain using our *ab initio* folding method, ROSETTA, which recently showed considerable promise in structure prediction in the CASP3 experiment[23] (folding of the SH3 domain starting with an unfolded polypeptide is computationally prohibitive using MD; ROSETTA achieves the vast speed-up necessary by simplifying both the conformational search strategy and the potential function). The chain representations, potentials and conformational sampling methods used by the two approaches are radically different; any common features observed in the two simulations are thus likely to reflect properties of the overall fold rather than specific residue–residue interactions.

ROSETTA uses local structural information from the protein data base and a simplified potential function to fold amino acid sequences to compact protein-like structures[23–25] (see Methods). Even for a small protein such as the SH3 domain, only a fraction of ROSETTA trajectories pass through native-like conformations. Inspection of individual successful trajectories suggested that the order of events in folding was quite similar to that observed experimentally. To get a more quantitative picture of the conformations sampled, we identified the substructures populated most frequently in 20 successful folding trajectories that produced structures within 4.5 Å r.m.s.d. (on Cα) of the native state. The occupancy of all side chain–side chain contacts (both native and nonnative) was averaged over all conformations in the 20 trajectories (Fig. 3*a*). Interactions formed early in the trajectory and persisting throughout have high occupancy in Fig. 3*a*, whereas contacts formed late have low occupancy. Thus, while this analysis does not single out the transition state ensemble (this could potentially be done using the p_{fold} method[26], but would be extremely computationally expensive), it provides information about the overall hierarchy to folding in the simulations. As is evident in Fig. 3*a*, the distal β-hairpin, the n-src loop and the diverging turn are highly populated during the simulations, while the RT loop and the sheet formed by the N- and C-terminal strands are very rarely populated, suggesting that they are the last elements to be structured in the protein (the contact map for the native protein is shown above the diagonal for comparison).

Table 2 Kinetic parameters of src SH3 folding[1]

	Mutant	ln(k_f)	ln(k_u)	m_f	m_u	$\Delta\Delta G_u$	Φ_f
	WT	3.55 ± 0.03	1.07 ± 0.04	1.02 ± 0.019	0.54 ± 0.013	NA	NA
	T9A	3.67 ± 0.03	2.35 ± 0.04	1.01 ± 0.037	0.46 ± 0.017	-0.64 ± 0.08	-0.11 ± 0.04
β1	F10A	3.41 ± 0.03	2.39 ± 0.03	1.06 ± 0.029	0.51 ± 0.012	-0.84 ± 0.07	0.10 ± 0.03
	F10I	3.67 ± 0.04	4.06 ± 0.03	1.08 ± 0.11	0.44 ± 0.091	-1.65 ± 0.17	-0.05 ± 0.02
	V11A	3.45 ± 0.04	3.79 ± 0.03	0.85 ± 0.057	0.55 ± 0.033	-1.64 ± 0.12	0.03 ± 0.02
	A12G	3.46 ± 0.03	2.73 ± 0.05	1.08 ± 0.049	0.47 ± 0.026	-1.00 ± 0.09	0.05 ± 0.02
	L13A	3.62 ± 0.02	3.87 ± 0.06	1.20 ± 0.098	0.36 ± 0.02	-1.49 ± 0.13	-0.03 ± 0.01
	Y14A	3.59 ± 0.03	1.68 ± 0.03	0.96 ± 0.042	0.50 ± 0.011	-0.31 ± 0.06	-0.08 ± 0.09
	D15A	3.70 ± 0.03	2.02 ± 0.08	0.98 ± 0.051	0.41 ± 0.03	-0.43 ± 0.13	-0.22 ± 0.10
	Y16A	3.40 ± 0.07	4.94 ± 0.06	–[2]	0.39 ± 0.018	-2.27 ± 0.26	0.03 ± 0.03
RT loop	Y16F	3.54 ± 0.04	1.37 ± 0.02	0.97 ± 0.036	0.65 ± 0.093	-0.18 ± 0.10	–[3]
	S18A	3.80 ± 0.04	0.47 ± 0.06	0.96 ± 0.028	0.55 ± 0.051	0.52 ± 0.10	0.28 ± 0.06
	R19A	3.64 ± 0.04	1.22 ± 0.07	1.08 ± 0.034	0.61 ± 0.017	-0.07 ± 0.08	–[3]
	T20A	3.69 ± 0.03	1.10 ± 0.04	1.01 ± 0.033	0.52 ± 0.017	0.06 ± 0.07	–[3]
	T22A	3.60 ± 0.03	1.14 ± 0.03	1.00 ± 0.03	0.52 ± 0.016	-0.01 ± 0.07	–[3]
	D23A	3.43 ± 0.05	1.91 ± 0.07	1.06 ± 0.083	0.51 ± 0.043	-0.56 ± 0.13	0.13 ± 0.07
	L24A	2.76 ± 0.03	3.44 ± 0.03	1.45 ± 0.06	0.45 ± 0.0089	-1.79 ± 0.09	0.26 ± 0.01
	S25A	3.49 ± 0.04	2.40 ± 0.03	0.95 ± 0.041	0.58 ± 0.04	-0.82 ± 0.08	0.03 ± 0.04
	F26A	2.17 ± 0.03	3.21 ± 0.04	–[2]	0.40 ± 0.0096	-1.97 ± 0.10	0.40 ± 0.01
	K27A	3.60 ± 0.04	1.79 ± 0.05	1.01 ± 0.051	0.65 ± 0.045	-0.44 ± 0.11	-0.06 ± 0.09
diverging turn	K28A	3.73 ± 0.03	1.45 ± 0.04	0.88 ± 0.017	0.49 ± 0.022	-0.09 ± 0.07	–[3]
	G29A	2.30 ± 0.04	2.74 ± 0.03	1.54 ± 0.088	0.43 ± 0.017	-1.66 ± 0.12	0.44 ± 0.02
	E30A	1.50 ± 0.03	2.34 ± 0.03	1.09 ± 0.054	0.65 ± 0.027	-1.94 ± 0.13	0.62 ± 0.02
	R31A	3.43 ± 0.03	1.45 ± 0.10	1.00 ± 0.06	0.57 ± 0.035	-0.32 ± 0.12	0.23 ± 0.08
β2	L32A	1.40 ± 0.08	2.61 ± 0.11	–[2]	–[2]	-2.26 ± 0.37	0.55 ± 0.05
	L32V	3.10 ± 0.04	2.68 ± 0.03	1.03 ± 0.05	0.56 ± 0.031	-1.21 ± 0.11	0.22 ± 0.02
	Q33A	3.17 ± 0.04	1.00 ± 0.04	1.11 ± 0.033	0.55 ± 0.027	-0.21 ± 0.09	–[4]
	I34A	1.40 ± 0.04	-0.63 ± 0.03	1.39 ± 0.027	0.51 ± 0.03	-0.32 ± 0.12	–[4]
	I34V	3.09 ± 0.05	0.75 ± 0.09	1.16 ± 0.093	0.50 ± 0.019	-0.09 ± 0.12	–[4]
n–src loop	V35A	2.6 ± 0.06	1.33 ± 0.05	1.27 ± 0.046	0.59 ± 0.041	-0.77 ± 0.12	0.77 ± 0.05
	N36A	3.64 ± 0.05	1.50 ± 0.05	1.12 ± 0.035	0.47 ± 0.017	-0.20 ± 0.09	–[3]
	N37A	3.91 ± 0.03	1.30 ± 0.04	0.94 ± 0.032	0.59 ± 0.023	0.07 ± 0.06	–[4]
	G40A	2.99 ± 0.05	1.02 ± 0.06	0.92 ± 0.037	0.55 ± 0.036	-0.28 ± 0.08	–[4]
	W42A	3.03 ± 0.05	2.83 ± 0.02	1.62 ± 0.048	0.45 ± 0.0084	-1.29 ± 0.10	0.25 ± 0.03
	W43A	3.24 ± 0.03	2.97 ± 0.06	1.16 ± 0.069	0.35 ± 0.025	-1.20 ± 0.11	0.15 ± 0.02
	W43I	4.45 ± 0.05	3.30 ± 0.03	1.10 ± 0.081	0.62 ± 0.073	-0.77 ± 0.13	–[4]
β3	L44A	2.10 ± 0.05	2.52 ± 0.05	1.88 ± 0.065	0.41 ± 0.0095	-1.64 ± 0.15	0.54 ± 0.03
	A45G	1.71 ± 0.07	0.86 ± 0.06	1.70 ± 0.044	0.39 ± 0.011	-0.92 ± 0.15	1.20 ± 0.08
	H46A	3.47 ± 0.03	2.00 ± 0.02	0.99 ± 0.029	0.58 ± 0.016	-0.62 ± 0.06	0.08 ± 0.04
	S47A	1.23 ± 0.04	1.29 ± 0.03	1.50 ± 0.035	0.44 ± 0.0074	-1.46 ± 0.09	0.95 ± 0.03
	L48A	2.82 ± 0.04	1.38 ± 0.03	1.20 ± 0.059	0.51 ± 0.019	-0.61 ± 0.08	0.72 ± 0.04
distal hairpin	S49A	2.98 ± 0.06	0.07 ± 0.06	1.20 ± 0.061	0.61 ± 0.026	0.18 ± 0.11	–[4]
	T50A	0.99 ± 0.05	1.59 ± 0.02	1.84 ± 0.054	0.47 ± 0.012	-1.79 ± 0.10	0.86 ± 0.02
	G51A	1.39 ± 0.04	1.02 ± 0.06	1.51 ± 0.11	0.41 ± 0.019	-1.21 ± 0.14	1.06 ± 0.06
	Q52A	3.29 ± 0.03	1.41 ± 0.08	1.03 ± 0.059	0.49 ± 0.021	-0.35 ± 0.12	0.45 ± 0.09
	T53A	2.33 ± 0.06	1.65 ± 0.04	1.38 ± 0.048	0.56 ± 0.021	-1.11 ± 0.11	0.68 ± 0.03
β4	G54A	3.79 ± 0.02	4.59 ± 0.05	1.60 ± 0.17	0.36 ± 0.02	-1.81 ± 0.12	-0.08 ± 0.01
	Y55A	2.10 ± 0.04	2.34 ± 0.04	1.64 ± 0.063	0.39 ± 0.0062	-1.52 ± 0.10	0.56 ± 0.02
	I56A	1.34 ± 0.03	2.02 ± 0.02	1.64 ± 0.067	0.46 ± 0.016	-1.84 ± 0.10	0.71 ± 0.02
	P57A	2.98 ± 0.04	2.89 ± 0.04	1.36 ± 0.098	0.45 ± 0.014	-1.36 ± 0.11	0.24 ± 0.02
	S58A	4.06 ± 0.03	1.14 ± 0.05	0.99 ± 0.034	0.58 ± 0.023	0.24 ± 0.08	–[4]
3_{10} helix	N59A	3.55 ± 0.03	1.28 ± 0.04	0.87 ± 0.041	0.58 ± 0.016	-0.14 ± 0.07	–[3]
	Y60A	3.48 ± 0.05	0.65 ± 0.04	1.06 ± 0.034	0.47 ± 0.034	0.23 ± 0.09	–[3]
	V61A	3.67 ± 0.04	3.29 ± 0.02	1.15 ± 0.044	0.44 ± 0.013	-1.18 ± 0.09	-0.06 ± 0.03
β5	A62G	3.59 ± 0.04	1.97 ± 0.03	1.12 ± 0.045	0.55 ± 0.019	-0.53 ± 0.08	-0.02 ± 0.07
	P63A	3.64 ± 0.04	1.44 ± 0.08	1.04 ± 0.062	0.48± 0.031	-0.14 ± 0.11	–[3]
	S64A	3.66 ± 0.03	0.40 ± 0.04	0.95 ± 0.026	0.59 ± 0.018	0.44 ± 0.06	0.14 ± 0.05
	D65A	3.69 ± 0.03	0.98 ± 0.05	0.92 ± 0.032	0.57 ± 0.02	0.13 ± 0.07	–[3]

[1]All experiments were done at pH 6 and 295 K. Kinetic measurements were done by stopped-flow fluorescence. Rate of folding (k_f) is reported at 0.3 M Gnd; rate of unfolding (k_u) is reported at 4 M Gnd to avoid extrapolation. $\Delta\Delta G_u$, Φ_f and standard errors were calculated as described in the Methods section.
[2]Parameters could not be reliably measured.
[3]Mutation has no (or very small) effect on stability, that is, $\Delta\Delta G_u \leq 0.20$ kcal mol^{-1}.
[4]Mutation either increases or decreases both k_f and k_u.

High-temperature MD unfolding simulations have provided insights into the folding of a number of small proteins[27–30]. Tsai *et al.*[22] carried out 30 independent simulations of src SH3 domain unfolding, and analyzed the order in which the structural elements are disrupted in the unfolding process. Overall, the hierarchy of unfolding was consistent with, but less pronounced than, the hierarchy observed in the experiments and in the *ab initio* simulations: the interactions between the N- and C-terminal strands were lost earlier than those within the distal loop β-hairpin, the n-src loop, and between the diverging turn and the distal loop β-hairpin.

Although the overall features of the simulations were consistent with the experimental results, there also were some inconsistencies. While residues in the three-stranded sheet made extensive contacts in the *ab initio* simulations and have high Φ_f values in the experiments, the C-terminus also made numerous contacts in the simulations (Fig. 3*a*), but contains mainly low Φ_f values. In the MD simulations, the RT loop remained ordered until quite late in the unfolding process. These discrepancies notwithstanding, the overall concordance between the simulations and the experiments is quite intriguing given that the MD simulations were carried out at 498 K and the *ab initio* simulations do not explicitly model side chains, and sugests that the hierarchy to folding of the SH3 domain is determined by fairly coarse-grained features of the structure.

Native state topology-based model calculations

To isolate those features of the native topology responsible for determining folding mechanisms, we recently developed a simple native state topology-based model of the folding free energy landscape and folding process[1,31]. In this model, the folding landscape is approximated by considering only conformations in which each residue is either ordered as in the native structure or completely disordered, and all ordered residues occur in one or two contiguous stretches of the protein sequence. The free energy of each of these conformations is determined by the balance between attractive native interactions, taken to be proportional to the surface area buried within the ordered region in the native structure, and the entropic cost of chain ordering, a function of the number of residues ordered and the loop length between the ordered segments.

We use this simple approach to model the folding free energy landscape of the three SH3 domains whose folding mechanisms have been probed by mutation: the src SH3 domain, the fyn SH3 domain (A. Davidson, pers. comm.), and the α-spectrin SH3 domain[6,12]. As a control, the same calculations were carried out on the α-spectrin SH3 permutant found by Serrano and coworkers[12] to have a significantly changed folding transition state. A natural reaction coordinate in this model is simply the fraction of residues ordered as in the native state, N_f. To determine the order in which the different parts of the protein fold as N_f increases, we enumerated all configurations allowed by the model with a particular value of N_f, determined their free energies, and computed the Boltzmann weighted frequency of ordering each residue[31] (Fig. 3*b*). Close to the unfolded state ($N_f = 0$) most residues have low frequency of ordering (black color), while close to the native state ($N_f = 1$) almost all residues are ordered (white color). There are interesting similarities in the hierarchies of structure formation in the three native SH3 domains obtained with this model (Fig. 3*b*). The first regions of the proteins to become ordered are the three hairpin loops (the distal loop β-hairpin, the RT loop and the n-src loop). By $N_f \sim 0.5$, the predominant region ordered in all three proteins is the three-stranded sheet formed by the distal loop β-hairpin and the n-src loop. The decrease in the relative population of the RT loop occurs because ordering additional residues increases the entropic cost of structure formation without significant increases in the attractive native interactions; in contrast, the ordering of the residues in the three-stranded sheet formed by the n-src loop and the distal loop β-hairpin produces significant gains in attractive interactions (the three-stranded sheet has a much higher density of stabilizing interactions than other portions of the protein of similar length). For all three proteins, the first and last strands become ordered very late in the folding process, consistent with the fact that they are stabilized primarily by nonlocal interactions. The similarities of the plots are consequences of the similarities in the topology of the three proteins. Notably, the hierarchy of folding is significantly altered in the SH3 domain circular permutant (Fig. 3*b*).

Overall, the hierarchy of structure formation observed in this simple model is consistent with the experimental results (Fig. 1): the residues with high Φ_f values in the src SH3 domain lie in the central three-stranded sheet, and for α-spectrin[6] and fyn (A. Davidson, pers. comm.), mutations in the distal loop β-hairpin have high Φ_f values. The lack of treatment of local sequence structure propensities may account for the roughly equal tendencies of the n-src loop and the distal loop β-hairpin to form in the model; the higher Φ_f values in the distal loop β-hairpin in src and α-spectrin[6] may reflect more complete ordering of this structure in the transition state due to stabilizing local interactions such as hydrogen bonds not included in the model. It should be emphasized that the two-segment model does not simply identify the longest contiguous stretch of interacting residues; for example, in barnase the N-terminal helix is correctly predicted to associate with the C-terminal sheet in the transition state[31].

In summary, the similarity in the hierarchy of folding observed experimentally in the src and α-spectrin SH3 domains, in the *ab initio* and MD simulations, and in the simple model calculations suggests that the folding mechanism of SH3 domains is largely determined by the topology of the native protein. The success of the simple model in reproducing the hierarchy observed both experimentally and in the simulations suggests that the folding process of this protein is largely determined by the balance between the entropic cost of chain ordering and the formation of attractive native interactions; nonnative interactions and conformations (that is, kinetic traps) appear to play a relatively minor role in shaping the folding process. The structural polarization of the SH3 domain folding transition state can be viewed as a consequence of the low free energy cost of ordering the low contact order[7] central three-stranded sheet, relative to the much higher contact order sheet formed by the N- and C-termini together with the RT loop. The importance of the computational work described in this paper in supporting this hypothesis may be seen by considering the alternative hypothesis that structural polarization in the transition state ensemble is a consequence of inhomogeneities in inter-residue interaction strengths: the strongest interactions are the last to break during unfolding and the most likely to nucleate the refolding process. Since the distal loop β-hairpin has the most extensive intraloop hydrogen bonding, if only the experimental data were available it could equally well be argued that the origin of structural polarization of the SH3 transition state was the greater stabilization of the distal loop β-hairpin relative to the

other structural elements in the protein. The *ab initio* folding simulations, however, have no prior knowledge that the interactions within the distal loop β-hairpin are stronger than in the other loops, and the simple free energy landscape model does not consider hydrogen bonding at all. Thus, the fact that a similar hierarchy to structure formation is observed in the calculations and experiments helps to distinguish between two hypotheses that are equally consistent with the experimental data.

The accompanying papers from the Dobson[32] and Serrano[33] groups strongly support the idea that native state topology is a dominant determinant of protein folding mechanisms. Martinez and Serrano[33] show that the folding transition state of the α-spectrin SH3 domain is similar to that of the src SH3 domain and is not significantly altered by changes in pH that produce large changes in stability. Chiti *et al.*[32] show that folding transition state structure is conserved in a second pair of proteins with similar native structures but with only 13% sequence identity: acylphosphatase and the activation domain of procarboxypeptidase 2. Chiti *et al.*[32] also show that the correlation between folding rates and contact order observed among two-state folding proteins generally also holds within a set of five nonhomologous proteins that exhibit the AcP topology.

The combination of experiment, simulation and theory employed in this paper, together with comparisons of the folding of structurally related proteins such as those in the accompanying papers, has the potential to distinguish the robust features of the folding process from those dependent on high-resolution detail, and to trace the origins of these robust features to basic physical principles. We believe that this integration of complementary approaches will be critical for obtaining a complete understanding of the folding process.

Methods

Mutagenesis. 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, protein solutions were made in 50 mM sodium phosphate, pH 6, and the temperature was held constant at 295 K. The stability of the point mutants was assessed by guanidine (Gnd) denaturation using either circular dichroism (CD) or fluorescence as reported[5]. The kinetics of folding and unfolding were followed by fluorescence on a Bio-Logic SFM-4 stopped-flow instrument. The unfolding reaction for the wild type protein was determined to behave as a two-state process[8], and the kinetic and equilibrium data for the mutants were fit to a two-state model. Equilibrium data (not shown) were generally in agreement with the kinetic estimates of stability for the less destabilized mutants.

Φ value analysis. Φ_f values were calculated only for mutants that were destabilized by more than 0.2 kcal mol^{-1} relative to the wild type protein. In order to avoid extrapolations, we compared folding rates at 0.3 M Gnd and unfolding rates at 4 M Gnd. In calculating ΔΔG for each mutant we assumed that it is independent of the denaturant concentration, which is warranted since the m values for the mutants are not very different from wild type. ΔΔG and Φ_f were computed using $\Delta\Delta G = -RT(\ln(k_{0.3M}^{wt}/k_{0.3M}^{mut}) + \ln(k_{4M}^{mut}/k_{4M}^{wt}))$ and $\Phi_f = -RT\ln(k_{0.3M}^{wt}/k_{0.3M}^{mut})/\Delta\Delta G$. For error analysis, we decided on a procedure that makes use of the many independent measurements in the linear portions of the V curves shown in Fig. 2. The estimates and confidence regions for $\ln(k_{0.3M}^{wt}/k_{0.3M}^{mut})$ and $\ln(k_{4M}^{mut}/k_{4M}^{wt})$ were obtained by simultaneously fitting the linear portions of the mutant and wild type V curves to $\ln k_f(Gnd)^{wt} = \ln k_f(Gnd)^{mut} + \delta_f$ and $\ln k_u(Gnd)^{wt} = \ln k_u(GND)^{mut} + \delta_u$. The error estimates for the Φ_f values presented in Table 2 represent 95% confidence intervals (roughly twice the standard deviation) for the Φ_f value, generated by repeatedly (10,000 times) sampling from the δ_f and δ_u distributions and recomputing the Φ_f values using $\Phi_f = \delta_f / (\delta_f + \delta_u)$.

***Ab initio* folding simulations.** The *ab initio* folding method, ROSETTA, utilizes a backbone plus side chain centroid-based representation of the chain; local interactions are satisfied by building structures up from short (three- and nine-residue) segments of known structures with sequences similar to those of the sequence being folded[24], while the nonlocal interactions that stabilize proteins are treated using a low-resolution scoring function with terms representing hydrophobic burial, strand pairing and specific pair interactions such as charge pairing and disulfide bonding[25]. A Monte Carlo simulated annealing strategy is used to sample conformational space; a move consists of a substitution of a three- or nine-residue segment of the chain by a randomly chosen fragment from a known structure with a similar local sequence. The protocol used to simulate the SH3 domain folding here was the same as that used in our CASP3 structure predictions[23]. All SH3 domain structures were removed from the data base of short fragments used for building up conformations. Because of the very large size of the conformational space, the trajectories and final structures for different runs can vary considerably. A total of 500 independent simulations were carried out, and all conformations from the 20 trajectories that produced structures within 4.5 Å r.m.s.d. of the native structure were combined. The frequency of each contact (defined as a pair of side chain centroids within 8 Å) in the pooled set of conformations was then computed.

Simple model calculations. The folding free energy landscape of the SH3 domain was approximated using the two-segment model described[31]. The free energy landscapes of the src, spectrin and fyn SH3 domains, and the 47–48 circular permutant of the spectrin SH3 domain were approximated by considering only configurations in which (i) each residue is fully ordered as in the native state or fully disordered, and (ii) the ordered residues occur in one or two contiguous stretches of the sequence. The free energy of each configuration was computed from the equation $F = -\gamma\Delta ASA + \alpha RTN + \beta RT\ln(\Delta L)$; all parameters were taken from the literature or from simple off-lattice calculations. In the first term, which represents the favorable interactions made in the partially ordered configuration, ΔASA is the difference in exposed surface area between the partially ordered configuration and the unfolded state (estimated from the sum of the native tripeptide surface areas) and $\gamma = 16$ cal mol^{-1} Å^{-2}. In the second term, which represents the entropic cost of ordering each residue in the ordered segments, N is the number of residues ordered and $\alpha = 2.9$. In the third term, which represents the entropic cost of closing the loop between the two ordered segments[34], ΔL is the length of the loop and $\beta = 1.8$.

Acknowledgments

We thank members of the Baker group for useful comments on the manuscript, and L. Serrano, C. Dobson and their coworkers for sharing their manuscripts before publication. This work was supported by grants from the NIH and the ONR and Young Investigator awards to D.B. from the NSF and the Packard Foundation and by the Molecular Biophysics Training Grant from the NIH to E.A.

Correspondence should be addressed to D.B.
email: dabaker@u.washington.edu

Received 30 April, 1999; accepted 1 September, 1999.

1. Alm, E. & Baker, D. *Curr. Opin. Struct. Biol.* **9**, 189–196 (1999).
2. Riddle, D.S. *et al. Nature Struct. Biol.* **4**, 805–809 (1997).
3. Kim, D.E., Gu, H. & Baker, D. *Proc. Natl. Acad. Sci USA* **95**, 4982–4986 (1998).
4. Perl, D. *et al*. *Nature Struct. Biol.* **5**, 229–235 (1998).
5. Grantcharova, V.P., Riddle, D.S., Santiago, J.V. & Baker, D. *Nature Struct. Biol.* **5**, 714–720 (1998).
6. Martinez, J.C., Pisabarro, M.T. & Serrano, L. *Nature Struct. Biol.* **5**, 721–726 (1998).
7. Plaxco, K.W., Simons, K. & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).
8. Grantcharova, V.P. & Baker, D. *Biochemistry* **36**, 15685–15692 (1998).
9. Plaxco, K.W. *et al. Biochemistry* **37**, 2529–2537 (1998).
10. Guijarro, J.I., Morton, C., Plaxco, K.W., Campbell, I.D. & Dobson, C.M. *J. Mol. Biol.* **276**, 657–667 (1998).
11. Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L. & Serrano, L. *Biochemistry* **33**, 2142–2150 (1994).
12. Viguera, A.R., Serrano, L. & Wilmanns, M. *Nature Struct. Biol.* **3**, 874–879 (1996).
13. Maxwell, K.L. & Davidson, A.R. *Biochemistry* **37**, 16172–16182 (1998).
14. Lim, W.A., Fox, R.O. & Richards, F.M. *Protein Sci.* **3**, 1261–1266 (1994).
15. Fersht, A.R. *Curr. Opin. Struct. Biol.* **5**, 79–84 (1994).
16. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. *J. Mol. Biol.* **254**, 260–288 (1995).
17. Milla, M.E., Brown, B.M., Waldburger, C.D. & Sauer, R.T. *Biochemistry* **34**, 13914–13919 (1995).
18. Serrano, L., Matouschek, A. & Fersht, A.R. *J. Mol. Biol.* **224**, 805–818 (1992).
19. Lopez-Hernandez, E. & Serrano, L. *Folding & Design* **1**, 43–55 (1995).
20. Kragelund, B.B. *et al*. *Nature Struct. Biol.* **6**, 594–601 (1999).
21. Yi, Q., Bystroff, C., Rajagopal, P., Klevit, R.E. & Baker, D. *J. Mol. Biol.* **283**, 293–300 (1998).
22. Tsai, J., Levitt, M. & Baker, D. *J. Mol. Biol.* **291**, 215–225 (1999).
23. Simons, K.T., Bonneau, R., Ruczinski, I. & Baker, D. *Proteins Struct. Funct. Genet.*, **in the press** (1999).
24. Simons, K.T., Kooperberg, C., Huang, E. & Baker, D. *J. Mol. Biol.* **268**, 209–25 (1997).
25. Simons, K.T. et al. Proteins Struct. Funct. Genet. **34**, 82–95 (1999).
26. Du, R., Pande, V.S., Grosberg, A.Y., Tanaka, T. & Shaknovich, E.I. *J. Chem. Phys.* **108**, 334–350 (1998).
27. Li, A. & Daggett, V. *J. Mol. Biol.* **257**, 412–429 (1996).
28. Bond, C.J., Wong, K.B., Clarke, J. Fersht, A.R. & Daggett, V. *Proc. Natl. Acad. Sci. USA* **94**, 13409–13413 (1997).
29. Alonso, D.O. & Daggett, V. *Protein Sci.* **7**, 860–874 (1998).
30. Lazaridis, T. & Karplus, M. *Science* **278**, 1928–1931 (1997).
31. Alm, E. & Baker, D. *Proc. Natl. Acad. Sci. USA*, **96**, 11305–11310 (1999).
32. Chiti, F. *et al.* Nature Struct. Biol. **6**, 1005–1009 (1999).
33. Martínez, J.C. & Serrano, L. *Nature Struct. Biol.* **6**, 1010–1016 (1999).
34. Jacobson, H. & Stockmayer, W.H. *J. Chem. Phys.* **18**, 1600–1606 (1950).
35. Xu, W., Harrison, S.C., & Eck, M.J. *Nature* **385**, 595–602 (1997).
36. Kraulis, P.J. *J. Appl. Crystallogr.* **24**, 946–950 (1991).