© 1997 Nature Publishing Group http://www.nature.com/nsmb

Functional rapidly folding proteins from simplified amino acid sequences

David S. Riddle, Jed V. Santiago, Susan T. Bray-Hall, Nikunj Doshi, Viara P. Grantcharova, Qian Yi and David Baker

Early protein synthesis is thought to have involved a reduced amino acid alphabet. What is the minimum number of amino acids that would have been needed to encode complex protein folds similar to those found in nature today? Here we show that a small β-sheet protein, the SH3 domain, can be largely encoded by a five letter amino acid alphabet but not by a three letter alphabet. Furthermore, despite the dramatic changes in sequence, the folding rates of the reduced alphabet proteins are very close to that of the naturally occurring SH3 domain. This finding suggests that despite the vast size of the search space, the rapid folding of biological sequences to their native states is not the result of extensive evolutionary optimization. Instead, the results support the idea that the interactions which stabilize the native state induce a funnel shape to the free energy landscape sufficient to guide the folding polypeptide chain to the proper structure.

Previous studies have shown that helical portions of proteins can be replaced with a subset of the naturally occurring amino acids without drastically affecting structure and function[1,2]. In complementary studies, entire helical bundle architectures have been built from a reduced amino acid alphabet[3,4], although for the most part they do not appear to have the ordered packing characteristic of biological proteins. While very suggestive, these studies leave open the question: can well-ordered, β-sheet containing protein architectures be built from sequences simplified through their entire length?

To address this question, we use a phage display selection strategy[5,6] to simplify the sequence of the 57 residue src SH3 domain. The SH3 domain has a complex β-barrel-like structure wherein residues spread throughout the sequence come together to create the binding site for a proline-rich peptide[7]. Because peptide binding requires proper folding of the SH3 domain, a selection for binding activity necessarily selects for the SH3 fold. Combinatorial libraries of SH3 variants displayed on the surface of M13 phage were constructed in which all residues not involved in binding were biased towards a small set of amino acids. Phage displaying properly folded, simplified SH3 variants were isolated by biopanning with proline-rich peptide coated paramagnetic beads in conjunction with a colony lift screen (see Methods).

Which amino acids should be included in a reduced alphabet capable of building complex protein architectures? Since globular proteins have nonpolar interiors and polar exteriors, any viable reduced alphabet must contain both nonpolar and polar residues. Although cooperatively folded helical structures were obtained from random sequences predominantly composed of one nonpolar and two polar amino acids[8], we found that such an alphabet was not sufficient for building the SH3 domain: in experiments with a reduced alphabet consisting of the nonpolar residue isoleucine (I) and the polar residues lysine (K) and glutamate (E), the majority of the alanines (A) and glycines (G) in the wild type sequence could not be replaced. Therefore, further experiments utilized the five letter amino acid alphabet I,K,E,A, and G. At positions where structural and phylogenetic data (Table 1, columns 4 and 5) suggested that one of the reduced alphabet residues might not be tolerated, additional residues were included; the residues allowed at each position are listed in Table 1, column 6.

Because of technical limitations on the total number of sequences which can be surveyed using phage display, each third of the SH3 domain was initially simplified independently. Sequences of the functional variants from these three libraries revealed that considerable simplification had been achieved (Fig. 1; Table 1, column 7). For example, in the first third of the protein (Fig. 1*a*) 16 of the 19 residues not involved in binding were converted to I, K, E, A, or G in the most simplified sequences. Simplification was successful at 38 of the 40 positions that were varied in these libraries. Table 1 provides some insight into the failure of the three letter alphabet: 9 of the 38 simplified positions tolerated A or G but not I (buried residues) or K or E (exposed positions). The requirement for the two additional residues appears to reflect space constraints at core positions not tolerating the large isoleucine side chain and a requirement for the increased conformational flexibility of glycine in turns. The conservation patterns evident in Fig. 1 provide important clues for understanding the sequence dependence of SH3 domain structure and function, and are being investigated through study of the corresponding point mutants.

Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA

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

© 1997 Nature Publishing Group http://www.nature.com/nsmb

To verify that the selection and screen yielded folded proteins, the biophysical properties of a number of the functional variants were assessed. Fluorescence, circular dichroism and one dimensional NMR studies indicated that each of the variants was folded and stable under the conditions used for the selection (data not shown). Thermodynamic parameters and peptide binding constants for simplified variants from each of the three libraries are summarized in Table 2.

SH3 variants simplified throughout their entire length were generated by randomly splicing together the simplified segments of a number of the partially simplified variants. In the splicing procedure, simplification of an additional five residues in restriction sites at the boundaries of the individual segments was also attempted. Two variants came through both the biopanning selection and the colony lift screen (Fig. 2*a*); their sequences are shown in Fig. 2*b*. Both variants contain large numbers of the reduced alphabet residues. In the more simplified variant, FP2, 40 of the 45 residues at which simplification was attempted are I, K, E, A, or G. Fig. 2*c* illustrates the positions at which simplification was achieved (red) for FP2 in the context of the structure of wild type SH3. Three of the five positions that resisted simplification (black) are at or near the binding site; the protein scaffold that supports the binding site is 95% I, K, E, A, and G. Since there are exposed large hydrophobic residues at the binding site that may compromise stability, further simplification could likely be achieved if the requirement for binding were relaxed. For example, of the 12 positions held fixed in this study, half were shown to tolerate alanine substitutions in the Sem5 SH3 domain; judging from the effects on expression levels, only one of these mutants appeared to have significantly decreased stability[9].

The structures adopted by the fully simplified sequences were probed using a range of biophysical methods. Gel filtration chromatography showed FP2 was monomeric with a Stokes radius very similar to wild type (Fig. 3*a*). The CD spectra of wild type and FP2 (Fig. 3*b*), while not identical, are no more different from each other than those of naturally occurring SH3 domains with nearly identical backbone conformations[10]. The differences are probably due to subtle contributions of aromatic residues at the binding site since the spectra of FP1 and wild type are very similar in the presence of peptide ligand (Fig. 3*c*). The CD and tryptophan fluorescence emission spectra of FP1 and FP2 undergo large shifts upon denaturation similar to those seen with wild type (Fig. 3*b*,*d*). Chemical denaturation data for both simplified proteins were fit well by a two state model (Fig. 3*e*); the free energy of folding of FP1 is only 0.7 kcal mol^{-1} less than wild type (Table 2). Both simplified SH3 domains bound the proline-rich peptide ligand used in the selection, albeit with somewhat lower affinity than the wild type protein (Table 2). Finally, one- and two-dimensional NMR spectra of FP1 and FP2 show the wide dispersion characteristic of native proteins (Fig. 3*f*). A well ordered backbone is indicated by a dispersion in the amide proton region (Fig. 3*f*, left panel) close to that of the wild type protein, and NOESY-HSQC experiments (data not shown) showed that the dispersion in the alpha proton region was ~2 p.p.m., considerably greater than the ~0.5 p.p.m. characteristic of disordered chains. The wide dispersion in the aliphatic region (Fig. 3*f*, right panel) suggests that the side chains in the core have well defined environments.

It is broadly accepted that a major stumbling block in protein design is the generation of well packed, ordered cores; many design efforts have yielded 'molten globule' structures with relatively fluid cores[1,3]. The ordered core packing suggested by the NMR results for FP1, FP2 (Fig. 3*f*) and a number of the partially

Table 1 Simplification strategy[1]

Residue number	Residue name	Secondary structure	Percent buried	Phylogenetic variation	Residues allowed in design	Residues observed in functional variants
9	T	C	13	E,I,L,Q,R,T,V	K,E	K,E
10	F	S	76	A,F,V,Y	I,v,l,f	V,F
11	V	S	64	I,K,Q,R,V	I,K	I,K
12	A	S	100	A	I,V,A,T	A
13	L	S	54	I,K,L	I,K	I,K
• 14	Y	C	44	F,H,Y	Y	Y
§ 15	D	C	48	D,P	D	D
• 16	Y	C	85	F,Y	Y	Y
17	E	C	25	E,K,Q,V	K,E	K,E
18	S	C	87	A,K,P,S	K,E,T,A	A
19	R	C	8	E,K,Q,R,S	K,E	K,E
20	T	C	39	E,H,R,S,T	K,E,T,A	K,E,T,A
21	E	C	7	D,E,G,P,S	K,E	K,E
22	T	C	26	D,E,G,R,S,T	K,E	K,E
• 23	D	C	79	D,E	D	D
24	L	C	69	I,L,V	I,v,l,f	L
25	S	C	51	A,D,G,S,T	K,E,T,A	E,T,A
26	F	C	90	F,I,L,M	I,v,l,f	I,F
27	K	C	34	E,H,I,K,T	K,E	K,E
28	K	C	30	E,K,L,R	K,E	K,E
29	G	C	22	G	K,E,R,G	G
30	E	C	53	A,D,E	K,E	E
31	R	S	21	I,K,Q,R,V	I,K	I,K
§ 32	*L*	S	94	F,I,L,M,V	*I,L*	*I,L*
§ 33	*Q*	S	46	H,I,K,Q,R,T,V	*K,E*	*E*
34	I	S	77	I,L,N,V	I,v,l,f	I,V,L,F
35	V	S	51	L,M,N,V	I,v,l,f	I,V,L,F
36	N	C	28	D,E,K,N,S	K,E	E,D
37	N	C	53	E,G,K,N,S	K,E,D,N	K,E,D,Q
38	T	C	7	D,E,S,T	K,E,T,A	K,E,T,A,*
39	E	C	21	D,E,N,P	K,E	K,E
40	G	C	33	D,G,I,K,P,Q	K,E,R,G	G,R
• 41	D	C	43	D,E,G,N	D	D,G
• 42	W	S	50	W	W	W
§ 43	W	S	92	L,M,W,Y	W	W
44	L	S	83	E,K,L,N,R	K,I,E,V	I,K,V
45	A	S	99	A,C,G,V	I,V,A,T	A
46	H	S	68	D,E,H,Q,R,Y	K,E	K,E
47	S	S	79	L,N,S,Y	I,V,A,T	A,T,S,N
§ 48	*L*	C	36	E,G,I,K,L,N	*I,L*	*I*
§ 49	*S*	C	12	A,N,R,S,T	*A,G,S,T*	*A,G*
§ 50	*T*	C	21	K,N,R,T	*A,G,S,T*	*S,T*
51	G	C	54	D,G,R	K,E,R,G	G
52	R	S	44	E,K,Q,R	K,E,R,G	K,E,R,G
53	T	S	53	E,Q,R,T,V	K,I,E,V,T,A	I,E,V,T
54	G	S	94	G	K,E,R,G	G,R
• 55	Y	S	62	D,F,I,L,W,Y	I,F,Y,N	Y,F
56	I	S	99	F,I,V	I,v,l,f	I,V
• 57	P	S	96	P	P	P
§ 58	*S*	H	87	A,G,K,S	*A,S*	*A,S*
• 59	N	H	16	N,T	N	N
• 60	Y	H	63	F,Y	Y	Y
61	V	S	95	I,V	I,V	I,V
62	A	S	77	A,E,R,T	K,E,A,T	A,T
63	P	S	34	D,E,K,M,P,R,Y	K,E,A,T,P,Q	E,A,P,Q
64	S	C	0	A,I,K,L,M,S,V	K,E,A,T,S	E,A
65	D	-	-	D,G,N	K,E,D,N	E,D,N

[1]Residue burial in the src SH3 structure (1SRL) was determined using the NACCESS program[24]; secondary structure assignments were taken directly from 1SRL (C, coil; S, strand; H, helix). Residues which appear in a multiple sequence alignment of the SH3 family[25] more than once are indicated in column 5. All amino acid mixtures were equimolar except for I,v,l,f, which was 70% I and 10% each V, L, and F. •, residues contacting the peptide ligand; §, residues at a restriction site or oligonucleotide annealing region. Italicized positions are at restriction sites which were kept constant in the three initial libraries but were allowed to change when the libraries were combined. A deletion at or near T38 was observed three times. The numbering system is that of ref. 7.

© 1997 Nature Publishing Group http://www.nature.com/nsmb

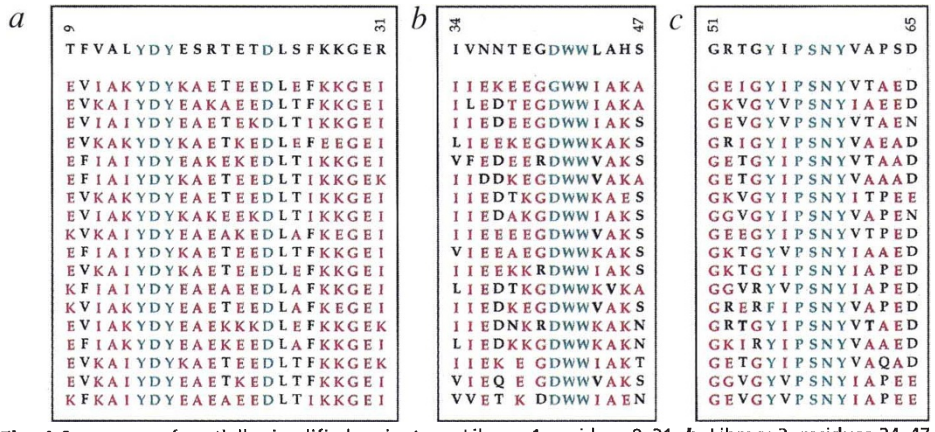

Fig. 1 Sequences of partially simplified variants. ***a***, Library 1, residues 9–31. ***b***, Library 2, residues 34–47. ***c***, Library 3, residues 51–65. For (*a*–*c*), wild type is at top. Red, I, K, E, A, G; black, residues which did not tolerate simplification; light blue, residues where simplification was not attempted.

simplified mutants (data not shown) is thus quite striking. Why were molten globule structures not obtained in the selection? One possibility is that the selection for function eliminates molten globule structures. Alternatively, the stability of a small β-barrel such as the SH3 domain may require tight core packing. The free energy cost associated with the transition from a tightly packed to a fluid core may be greater for β-barrels than for helical bundles. The loss in favourable packing interactions in such a transition is likely to be offset by a greater gain in entropy for helical bundles because main-chain hydrogen bonds and burial of nonpolar residues in the core are maintained over a larger ensemble of conformations.

The simplified sequences provide an opportunity to investigate basic issues concerning the evolution of the rapid and cooperative folding of small protein domains. Protein function requires that the native states of proteins be both stable and kinetically accessible. Natural selection clearly selects for stability, but it is not clear whether kinetic accessibility is also under selective pressure[11]. There are several reasons to suspect that folding rates may be under natural selection: first, the apparent difficulty of locating the native state amidst the vast number of nonnative conformations[12], second, the biological convenience of observed time-scales for folding of small proteins (milliseconds to seconds), and third, the existence of proteins for which the native state, although very stable, is not accessible on a biological time-scale[13]. The influence of natural selection can be assessed by comparing the rates of folding of sequences not generated by natural selection to those of biological proteins.

The simplified SH3 domain sequences make possible such a test. The folding rates of the simplified proteins were compared to that of the wild type protein (Fig. 4, Table 2). Strikingly, FP2 refolds at almost exactly the same rate as WT, while FP1 refolds even faster. Thus, the dramatic changes in the amino acid sequence and charge distribution do not significantly perturb the refolding of the protein. This suggests that the sequence of the src SH3 domain has not been highly optimized by natural selection for rapid folding. The independence of folding kinetics on the details of the sequence is a welcome confirmation of current theoretical work on protein folding which holds that the kinetics of folding are determined more by general features of the free energy landscape[14,15] than by a sequence of specific interactions. Rapid folding appears to be a byproduct of selection for stability; the interactions which stabilize the native structure are also likely to stabilize partially native structures, producing a funnel leading into the native minimum[14,15].

In conclusion, the results presented here show that a β-sheet containing protein can be built largely from a five letter amino acid alphabet. This extends similar findings on simple helical bundles[3,4,16] and α-helical portions of larger proteins[1,2] to a considerably more complex protein architecture. It has been proposed that the origin and early evolution of protein synthesis involved a reduced amino acid alphabet[17–19]; our results suggest that the heterogeneity of interactions and sequence complexity

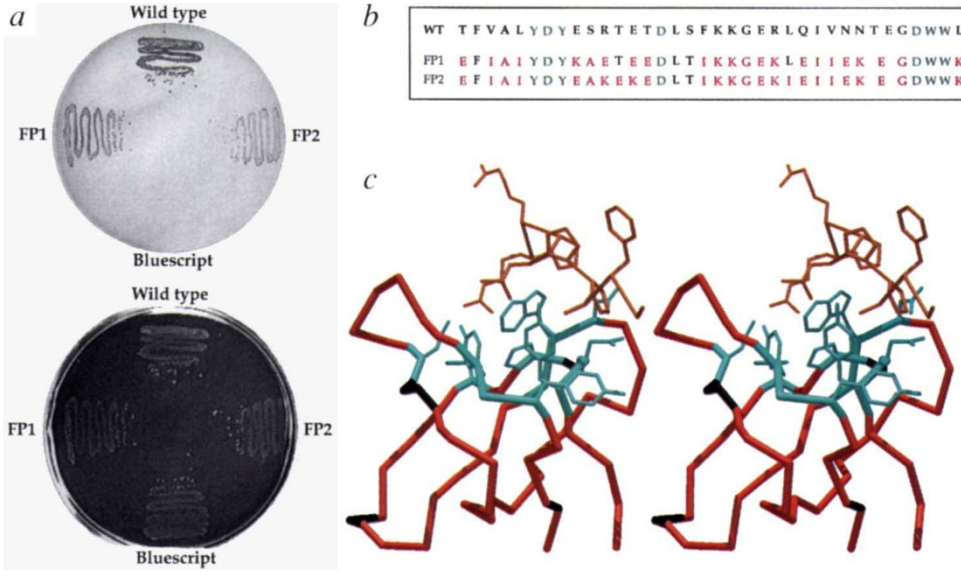

Fig. 2 Fully simplified variants. ***a***, Filter binding assay (top) for streaks of *E. coli* cells (bottom) bearing phagemids encoding wild type SH3, the two fully simplified variants, and a negative control (pBluescript). ***b***, Sequences of fully simplified variants. ***c***, Stereo diagram showing the positions of simplified residues for FP2 in the wild type SH3 structure (1SRL). Side chains of SH3 residues involved in ligand binding are displayed; the peptide ligand is shown in orange. The colour scheme is as in Fig. 1. Two of the five residues not simplified in FP2 (black) were simplified in other variants. (*c*) prepared with MOLSCRIPT[26] and Raster3D[27,28].

© 1997 Nature Publishing Group http://www.nature.com/nsmb

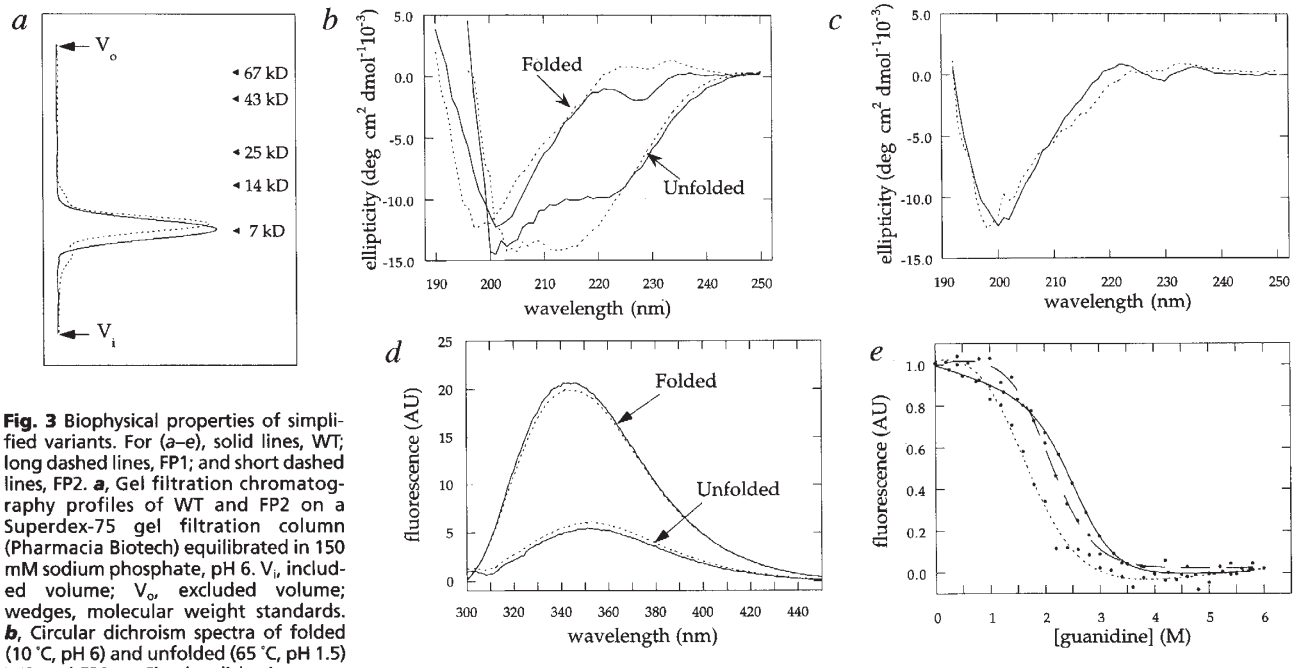

Fig. 3 Biophysical properties of simplified variants. For (*a*–e), solid lines, WT; long dashed lines, FP1; and short dashed lines, FP2. ***a***, Gel filtration chromatography profiles of WT and FP2 on a Superdex-75 gel filtration column (Pharmacia Biotech) equilibrated in 150 mM sodium phosphate, pH 6. V_i, included volume; V_o, excluded volume; wedges, molecular weight standards. ***b***, Circular dichroism spectra of folded (10 °C, pH 6) and unfolded (65 °C, pH 1.5) WT and FP2. ***c***, Circular dichroism spectra of folded WT and FP2 in the presence of 50 μM peptide ligand (EGREARALPPLPRY). ***d***, Fluorescence emission spectra of folded (50 mM sodium phosphate, pH 6) and chemically unfolded (50 mM sodium phosphate, 5 M guanidine, pH 6) WT and FP2. ***e***, Equilibrium guanidine denaturation of WT, FP1 and FP2 monitored through tryptophan fluorescence. ***f***, One-dimensional proton NMR spectra of WT, FP1 and FP2 in 50mM sodium phosphate, 90% H_2O/10% D_2O, pH 6, 22 °C. The left panel is the amide proton region, the right panel is the upfield portion of the aliphatic region. The H^1 chemical shift is referenced to H_2O at 4.70 p.p.m.

required to encode single unique structures are sufficiently small for this to have been possible. The lack of change of the folding kinetics upon drastic change in sequence suggests that the sequences of biological proteins are not heavily optimized by natural selection for rapid folding.

Methods

The src SH3 domain was cloned into a previously described phage display vector[6] between *Nco*I and *Mlu*I sites. A *Sal*I site was introduced at residues 48–50; this resulted in the conservative substitution T49S. This new site, along with an already extant *Pst*I site, divided the protein roughly into thirds. Each third was independently replaced by an oligonucleotide cassette encoding the simplified library described in Table 1. The cassettes were prepared from degenerate oligonucleotides (Ransom Hill Biochemical, Ramona, CA) using standard procedures[20]; following Klenow extension, the cassettes were digested with the appropriate pair of restriction enzymes and ligated into similarly digested SH3 phage display vector. Phage libraries were prepared following electroporation of *E. coli* using standard procedures[6]; the complexity of the libraries averaged 5 X 10^7. The library of fully simplified sequences was generated by splicing together variants with a simplified first third (Fig. 1*a*), a simplified second third (Fig. 1*b*) and an oligonucleotide cassette encoding simplified variants of the third segment. In the PCR splicing process, the residues spanning the *Sal*I and *Pst*I restriction sites were biased toward the reduced alphabet.

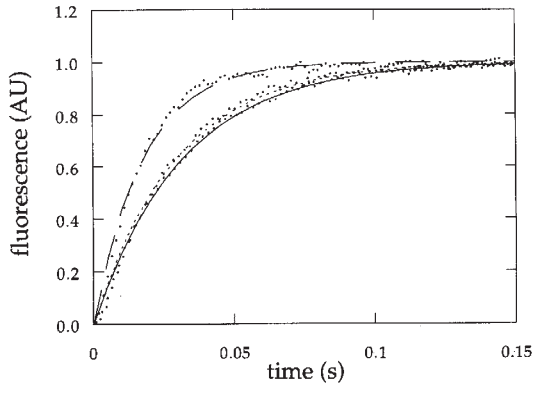

Fig. 4 Kinetics of folding. WT, FP1, and FP2 were denatured in guanidine and refolding was initiated by rapid removal of the denaturant by dilution to 0.4 M guanidine, 50 mM sodium phosphate pH 6. Folding was monitored through change in fluorescence signal (see Fig. 3*d*) using a stopped-flow instrument[22]. The lines are simple exponential fits of the data and are coded as in Fig. 3.

© 1997 Nature Publishing Group http://www.nature.com/nsmb

To prepare the affinity matrix, an oligonucleotide encoding the src SH3 ligand RALPPLPRY[21] was inserted downstream of a peptide which is endogenously biotinylated in *E. coli* (PINPOINT vector, Promega, Madison, WI). Extracts of cells expressing the biotinylated fusion protein were incubated with streptavidin-coated paramagnetic beads (Dynabeads M-280, Dynal, Inc., Lake Success, NY) for 30 min at room temperature; the beads were then washed five times with 1 ml 0.5% Tween, TBS, pH 7.0 (TTBS). For biopanning, ~10^9 library phage were incubated with the peptide coated beads for 1 h at room temperature, the beads were washed with TTBS, and phage displaying functional variants were eluted with pH 2 buffer. For the secondary colony screen, nitrocellulose filters (0.05 mm pore size; Schleicher and Schuell, Keene, NH) were overlaid on LB plates containing freshly grown colonies transfected with the eluted phage. The colonies were washed from the filters with deionized water, the filters were blocked for 10 min in TTBS, and incubated with the biotinylated peptide complexed to streptavidin-conjugated alkaline phosphatase. The subsequent washing and colour reaction were carried out as described[6]. For biophysical analysis, variants were cloned into pET15b (Novagen, Madison, WI) and overexpressed and purified as described[6]. All expression constructs contained an N-terminal 6 X His tag to facilitate purification; the first SH3 domain residue in the constructs corresponds to threonine 9 in the numbering system of ref. 7.

Circular dichroism, NMR and equilibrium and stopped-flow fluorescence measurements were carried out and analyzed as described[22]. The peptide H-EGREARALPPLPRY (Research Genetics, Huntsville, AL) was used for K_d measurements[23]; the first five residues are present in the fusion protein used in the selection and screen.

Table 2 Thermodynamic and kinetic parameters for wild type SH3 and simplified SH3 variants

Protein	$\Delta G_{u\text{-}f}^{H_2O}$ (kcal mol^{-1})	$k_f^{H_2O}$ (s^{-1})	$k_u^{H_2O}$ (s^{-1})	K_d (μM)
wt SH3	3.7	57	0.1	7.5
S1	1.8	42	1.8	5.8
S2	3.4	52	0.2	60
S3	3.7	59	0.1	37
FP1	3.0	93	0.5	150
FP2	1.7	57	3.0	38

S1, S2, and S3 are the first SH3 variants listed in Fig. 1*a*, 1*b*, and 1c respectively. The overall free energy of folding ($\Delta G_{u\text{-}f}^{H_2O}$), the rates of folding ($k_f^{H_2O}$) and unfolding ($k_u^{H_2O}$), and the affinity for the src peptide ligand EGREARALPPLPRY were determined as described previously[22,23].

Acknowledgments

We thank K. Walsh for use of his FPLC system, R. Klevit for use of her NMR instrument, and D. Agard, M. Hecht, J. Onuchic, D. Shortle, D. Teller, and members of the Baker group for comments on the manuscript. This work was supported by a grant from the Office of Naval Research and by Young Investigator awards to D.B. from the NSF and the Beckman foundation.

Received 8 May, 1997; accepted 1 August, 1997.

1. Shang, Z. *et al.* Design of a "minimAl" homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *Proc. Natl. Acad. Sci. USA* **91**, 8373–8377 (1994).
2. Heinz, D.W., Baase, W.A. & Matthews, B.W. Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrates the redundancy of information in an amino acid sequence. *Proc. Natl. Acad. Sci. USA* **89**, 3751–3755 (1992).
3. Regan, L. & Degrado, W.F. Charcterization of a helical protein designed from first principles. *Science* **241**, 976–978 (1988).
4. Kamteker, S., Schiffer, J.M., Xiong, H., Babik, J.M. & Hecht, M.H. Protein design by binary patterning of polar and nonpolar amino acid sequences. *Science* **262**, 1680–1685 (1993).
5. Scott, J.K. & Smith, G.P. Searching for peptide ligands with an epitope library. *Science* **249**, 386–390 (1990).
6. Gu, H. *et al.* A phage display system for studying the sequence determinants of protein folding. *Prot. Sci.* **4**, 1108–1117 (1995).
7. Feng, S., Chen, J.K., Yu, H., Simon, J.A. & Schreiber, S.L. Two binding orientations for peptides to the src SH3 domain: development of a general model for SH3 ligand interactions. *Science* **266**, 1241–1247 (1994).
8. Davidson, A.R., Lumb, K.J. & Sauer, R.T. Cooperatively folded proteins in random libraries. *Nature Struct. Biol.* **2**, 856–863 (1995).
9. Lim, W.A. & Richards, F.M. Critical residues in an SH3 domain from Sem-5 suggest a mechanism for proline-rich peptide recognition. *Nature Struct. Biol.* **1**, 221–225 (1994).
10. Viguera, A.R., Arrondo, J.L.R., Musacchio, A., Saraste, M. & Serrano, L. Characterization of the interaction of natural proline-rich peptides with five different SH3 domains. *Biochemistry* **33**, 10925–10933 (1994).
11. Baker, D. & Agard, D.A. Kinetics versus thermodynamics in protein folding. *Biochemistry* **33**, 7505–7509 (1994).
12. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).
13. Baker, D., Sohl, J.L. & Agard, D.A. A protein-folding reaction under kinetic control. *Nature* **356**, 263–265 (1992).
14. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Prot. Struc. Funct. Genet.* **21**, 167–195 (1995).
15. Dill, K.A. & Chan, H.S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19 (1997).
16. Handel, T.M., Williams, S.A. & DeGrado, W.F. Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **261**, 879–885 (1993).
17. Woese, C. *The Genetic Code*, Harper & Row, New York (1967).
18. Crick, F.H.C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
19. Wong, J.T. A co-evolution theory of the genetic code. *Proc. Nat. Acad. Sci. USA* **72**, 1909–1912 (1975).
20. Reidhaar-Olsen, J.F. *et al.* Random mutagenesis of protein sequences using oligonucleotide cassettes. *Methods Enzymol.* **208**, 564–586 (1991).
21. Yu, H. *et al.* Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945 (1994).
22. Scalley, M.L. *et al.* Kinetics of folding of the IgG binding domain of peptostreptoccocal protein L. *Biochemistry* **36**, 3373–3382 (1997).
23. Chen, J.K., Lane, W.S., Brauer, A., Tanaka, A. & Schreiber, S.L. Biased combinatorial libraries: novel ligands for the SH3 domain of phosphatidylinositol 3-kinase. *J. Am. Chem. Soc.* **115**, 12591–12592 (1993).
24. Hubbard, S.J. & Thornton, J.M. "NACCESS" computer program, Department of Biochemistry, University College London (1993).
25. Koyama, S. *et al.* Structure of the PI3K SH3 domain and analysis of the SH3 family. *Cell* **72**, 945–952 (1993).
26. Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
27. Bacon, D. & Anderson, W.F. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* **6**, 219–220 (1988).
28. Merritt, E.A. & Murphy, M.E. Raster 3D Version 2.0 A program for photorealistic molecular graphics. *Acta Crystallogr.* **D50**, 869–873 (1994).