Protein Science (1999), *8*:2734–2741. Cambridge University Press. Printed in the USA.

Robustness of protein folding kinetics to surface hydrophobic substitutions

H. GU, N. DOSHI, D.E. KIM, K.T. SIMONS, J.V. SANTIAGO, S. NAULI, AND DAVID BAKER
Department of Biochemistry, University of Washington, Seattle, Washington 98195

(RECEIVED April 16, 1999; ACCEPTED September 6, 1999)

Abstract

We use both combinatorial and site-directed mutagenesis to explore the consequences of surface hydrophobic substitutions for the folding of two small single domain proteins, the src SH3 domain, and the IgG binding domain of Peptostreptococcal protein L. We find that in almost every case, destabilizing surface hydrophobic substitutions have much larger effects on the rate of unfolding than on the rate of folding, suggesting that nonnative hydrophobic interactions do not significantly interfere with the rate of core assembly.

Keywords: binary pattern; phage display; protein folding; protein L; SH3 domain

The pattern of hydrophobic and hydrophilic amino acids in a protein sequence is thought to be an important determinant of the structure of the native state (West & Hecht, 1995) and the assembly of the hydrophobic core has been proposed to be important and perhaps rate-limiting steps in protein folding (Agashe et al., 1995; Dill et al., 1995; Sosnick et al., 1996). A combinatorial search through alternative hydrophobic packing arrangements may be required to locate the native core. As the number of possible core packing configurations grows exponentially with the number of hydrophobic residues in the sequence, unless other factors favor the native arrangement of hydrophobic core residues, the substitution of polar surface residues for hydrophobic residues should slow the rate of folding. Thus, the contribution of large scale searching through alternative core packing arrangements to the rate limiting step in folding can potentially be assessed by determining the effect of surface hydrophobic substitutions on the rate of folding.

We have chosen two small proteins that have been studied in our laboratory as models to probe the consequences of surface hydrophobic substitutions on the rate and robustness of protein folding. The kinetics of folding of the src SH3 domain (Grantcharova & Baker, 1997) and the IgG binding domain of protein L (Scalley et al., 1997) have been extensively characterized. Phage display selection methods have been developed that provide a means to retrieve rare correctly folded sequences from large combinatorial libraries for both protein L (Gu et al., 1995) and the SH3 domain (Riddle et al., 1997). Here, we generate large combinatorial libraries for both proteins in which the hydrophobic-polar pattern for a portion of the sequence is largely disrupted. By analyzing the sequences of folded variants obtained using the phage display selection, we determine which surface positions tolerate surface hydrophobic substitutions and which do not. We then use site-directed mutagenesis to probe the consequences of introducing hydrophobic residues at specific sites on the folding and unfolding rates.

Reprint requests to: David Baker, Department of Biochemistry, University of Washington, Seattle, Washington 98195; e-mail: dabaker@u.washington.edu.

Results

src SH3 domain

To assess the importance of the native hydrophobic pattern to the folding of the src SH3 domain, we generated a large library of $\sim 10^7$ SH3 variants in which each of the surface polar residues in the first third of the src SH3 domain was biased toward a hydrophobic residue (Fig. 1A, shaded region). The library was designed such that at each of the 11 sites, the frequency of the naturally occurring (polar) residue was 0.2, and the frequency of valine or isoleucine, 0.8. The choice of valine or isoleucine was dictated by the genetic code and the specification that only the naturally occurring residue and one of the two hydrophobic residues were to be allowed at any one position.

To check on the quality of the library construction, a number of variants were sequenced prior to the phage selection step. As is evident in Table 1 ("Unselected"), the distribution of residues at each of the randomized surface positions was roughly that expected, with a preponderance of hydrophobic residues at each position.

Folded variants were recovered from the library by biopanning using magnetic beads coated with a proline rich peptide ligand for the src SH3 domain and a subsequent colony lift screen (Riddle et al., 1997). The sequences of these folded variants (Table 1, "Selected") contrasted sharply with those of the unselected clones (Table 1, "Unselected"). Isoleucine and valine residues were recovered frequently at 5 of the 11 surface positions. However, hydrophobic residues were

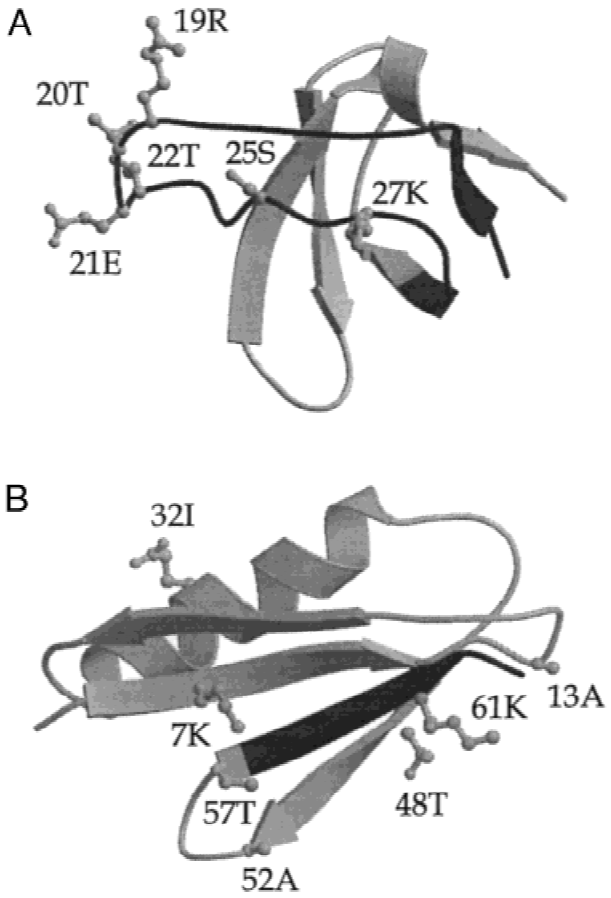

Fig. 1. Ribbon diagram illustrating the overall protein fold of (**A**) the src SH3 domain and (**B**) the protein L IgG binding domain. Mutagenized residues are darkly shaded.

excluded from the remaining six surface positions, which lie in the second half of the RT loop. The dramatic exclusion of hydrophobic residues from these positions could reflect a large reduction in stability, a disruption of ligand binding activity, or a pronounced increase in susceptibility to aggregation.

To probe the origins of the sensitivity to surface hydrophobic substitutions in the second portion of the RT loop, we used site-directed mutagenesis to introduce isoleucine residues at these positions one at a time. As indicated in Table 2, single hydrophobic substitutions at these positions did not substantially destabilize the protein, therefore the selection against hydrophobic residues at these positions is probably not due to a loss in stability. The binding affinities of the mutants for the proline rich peptide used in the phage selection were within a factor of five of that of the wild-type protein (data not shown), suggesting that the substitutions did not substantially alter protein function. The absence of strong effects on binding affinities in vitro [fivefold decreases in affinity do not significantly reduce phage recovery in the biopanning selection (Riddle et al., 1997)] taken together with the exclusion of hydrophobic surface residues in the selected variants suggest that surface hydrophobic substitutions at these positions may impair the assembly of the fusion protein onto the phage coat, or that there may be an insufficient accumulation of protein in the periplasm to be detectable in the colony lift screen. An increased tendency toward aggregation could contribute to either or both of these possible defects.

To explore the kinetic consequences of surface hydrophobic substitutions, we measured the kinetics of folding of the variants. Interestingly, replacement of polar residues by hydrophobic residues at most of the sites had relatively little effect on the rate of folding (Table 2); the largest change in the folding rate was only twofold. Most mutants also had only a modest affect on the unfolding rate, with the exception of S25I, which increased the unfolding rate 11-fold.

Table 1. *Sequences recovered from the SH3 domain surface hydrophobic library*[a]

	WT	M	T	F	V	A	L	Y	D	Y	E	S	R	T	E	T	D	L	S	F	K	K	G	E	R	L
Selected	NL1										V		I	I												I	
	NL2			I									I	I												I	
	NL3			I									I	I													
	NL4			I									I														
	NL5										V		I	I												I	
	NL6												I													I	
	NL7										V		I	I													
Unselected	NLu1			I							V		I	I		I			I		I	I	V			I	
	NLu2										V		I	I	V	I			I		I	I	V			I	
	NLu3			I									I	I	V	I			I		I	I		G		I	
	NLu4			I									I		V	I			I			I	V				
	NLu5			I							V		I	I	V				I			I	V			I	
	NLu6			I		C							I	I	V	I			I		I	I	V			I	
Residues allowed				T							E		R	T	E	T			S		K	K	G			R	
				I							V		I	I	V	I			I		I	I	V			I	

[a]At each of the 11 mutated positions, the expected frequency of the wild-type residue and either isoleucine or valine was 0.2 and 0.8, respectively. The sequence differences between the wild-type sequence (top row) and the recovered sequences are indicated. The residues allowed at each of the varied positions are listed in the last two rows.

Table 2. *Kinetic and thermodynamic parameters for SH3 domain point mutants*

Protein	ΔG (kcal mol^{-1})	k_f (s^{-1})	k_u 4M (s^{-1})
wt	4.1	57	3.0
E21I	4.7	56	2.6
K27I	3.9	51	2.8
S25I	2.6	57	35
T22I	2.3	36	5.2
R19I/T20I	3.6[a]	53	3.8

[a]Calculated from kinetic data.

Protein L

The structure of protein L consists of an α-helix packed against a four-stranded β-sheet; the order of secondary structure elements is $\beta\beta\alpha\beta\beta$ (Fig. 1B). Except for A52 and A63, all of the exposed positions on strands 3 and 4 are occupied by polar amino acids, and the buried positions in the interior strand are occupied by large hydrophobic residues (Fig. 2). To probe the dependence of folding on the hydrophobic-polar pattern in this part of the protein, a phagemid library was constructed in which six residues (L58–A63) on the fourth β-strand of protein L were replaced by synthetic oligonucleotide cassettes encoding 12 amino acids at each position (Fig. 1B, shaded region). Hydrophobic side chains of different sizes and shapes (Phe, Leu, Ile, Val, Ala), as well as positively charged (Lys), negatively charged (Glu, Asp), and polar (Ser, Thr, Tyr, Asn) side chains were allowed in the randomization. Cysteines and prolines were omitted due to complications arising from disulfide bond formation and peptide bond isomerization, respectively; methionines were omitted because they complicate CNBr cleavage of the N-terminal His-tag. The number of independent transformants in the phagemid library (5×10^6) was slightly greater than the number of different sequences allowed by design (3×10^6). The nucleotide frequencies in the unselected phage population were checked by sequencing individual clones and were close to the expected frequencies (data not shown). The recovery rate of functional variants from the library in biopanning experiments using IgG coated magnetic beads (Gu et al., 1995) was 0.01%, and in a secondary colony lift screen (Gu et al., 1995), approximately 20% of the selected strand variants showed IgG binding activity.

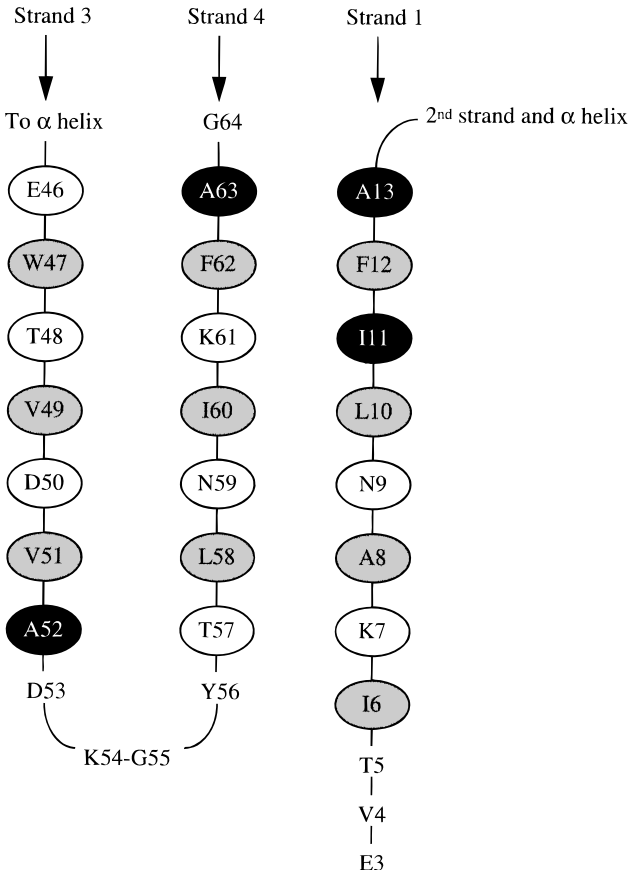

Fig. 2. Schematic illustration of strands 1, 3, and 4 of protein L. Circles, strand residues; grey, buried positions; black, surface hydrophobic residues.

The sequences of variants that exhibited IgG binding activity in both the phage display selection and the colony lift screen were determined (Table 3; some of these sequences were reported in Kim et al. (1998a) and are shown again here for convenient reference). The buried positions were almost always substituted with large hydrophobic amino acids (the methionine in the last sequence may have been due to a spontaneous mutation or oligonucleotide synthesis error). Furthermore, most of the repacked cores had a larger calculated volume than that of wild-type: the average combined burial volume of positions 58, 60, and 62 was 571 Å^3 compared to 540 Å^3 in the wild-type (Table 3). As noted in studies of the central helix of protein L (Kim et al., 1998b), underpacking the core appears to be more deleterious to protein L than overpacking.

In contrast, the surface exposed positions showed few sequence constraints and the nearly perfect alternating polar/nonpolar pattern of the wild-type protein was not preserved in the combinatorial mutants (Table 3). All three exposed positions tolerated nonpolar residues; however, the hydrophobic substitutions were somewhat disfavored (Table 4). Interestingly, although the majority of combinatorial mutants had one nonpolar residue on the surface and almost one-quarter of the mutants had two surface nonpolar residues, there were no mutants with all three surface positions substituted by nonpolar amino acids, despite the fact that 13% of the sequences contain three surface hydrophobic residues in the unselected population. Negative selection was particularly dramatic for residues 59 and 61, which were never simultaneously substituted with hydrophobic amino acids in the recovered mutants (Table 4, last column).

The above observations suggested that at least one polar residue on the exposed side of strand 4 is required for the proper folding of protein L. To further explore the effect of surface hydrophobic residues on folding, we generated a series of mutants with phenylalanine substitutions at exposed positions. A potential problem with studying polar to hydrophobic substitutions on the surface of a protein is that they can destabilize proteins considerably (Pakula & Sauer, 1990; Predki et al., 1996). To probe dramatic changes in the hydrophobic pattern without completely destabilizing the protein, we took advantage of the fact that phenylalanine residues in adjacent β-strands can form energetically favorable interactions (Smith & Regan, 1995). To systematically remove the binary pat-

Table 3. *Functional amino acid sequences recovered from the phage selection and screen*[a]

58	59	60	61	62	63	Volume buried (Å^3)
Y	T	F	T	Y	A	611
Y	T	F	Y	Y	V	
Y	T	F	T	Y	D	
F	N	F	S	Y	A	610
I	Y	F	T	Y	A	576
L	T	Y	V	Y	T	
L	N	Y	Y	Y	N	
L	N	Y	I	Y	K	
L	T	Y	S	Y	A	
I	Y	F	T	Y	V	576
I	T	F	T	Y	E	
L	F	F	N	F	I	575
L	Y	F	N	F	Y	
L	S	F	S	F	L	
L	T	F	A	F	E	
L	F	F	S	Y	L	575
L	S	F	A	Y	S	
L	N	F	V	Y	I	
L	V	F	N	Y	I	
L	S	F	N	Y	V	
L	F	F	N	Y	T	
L	V	F	T	Y	T	
L	F	F	S	Y	I	
L	F	F	S	Y	L	
L	T	Y	Y	F	I	575
L	V	Y	T	F	A	
I	Y	F	I	F	N	575
I	F	F	N	F	N	
V	T	F	Y	F	E	548
V	S	F	Y	F	L	
V	N	F	S	F	L	
L	Y	I	K	Y	V	541
L	F	I	T	Y	T	
L	N	I	K	F	A	Wt (540)
L	N	I	I	F	E	540
L	Y	I	T	F	T	
L	T	I	V	F	T	
L	Y	L	L	F	A	
I	N	V	I	Y	A	515
L	L	M	Y	L	E	507

[a]The residue numbers are indicated at the top of the table. Mutants with the same residues at the buried positions (58, 60, 62) are grouped together; the total buried volume of these three residues (Creighton, 1993) is listed on the right. Buried positions (58, 60, 62) and nonpolar amino acids appearing at exposed positions are highlighted in grey and black, respectively.

tern, we constructed mutants 2Phe and 3Phe, which have two and three phenylalanine substitutions on the fourth β-strand, and mutants 4Phe and 6Phe, which have paired phenylalanine substitutions on the third and fourth strand (Fig. 3). The fourth strand in 3Phe and 6Phe consists entirely of hydrophobic residues; the original amphilicity of the strand is completely lost.

Biophysical characterization of the Phe mutants confirmed that at least one polar residue on the surface of the fourth strand is required for proper folding of protein L. Circular dichroism (CD) spectra of 2Phe and 4Phe resembled that of protein L, whereas CD spectra of 3Phe and 6Phe were significantly different (Fig. 4A). The change in the fluorescence emission spectra of 2Phe and 4Phe upon folding were almost identical to that of protein L, whereas the emission spectra of 3Phe and 6Phe show relatively little change during folding (Fig. 4B). In addition, the equilibrium GuHCl denaturation profile of 3Phe as monitored by CD was much less cooperative than that observed for protein L, 2Phe, and 4Phe. Gel-filtration chromatography showed that 3Phe but not 2Phe or 4Phe forms large oligomers in the absence of denaturant (data not shown). Taken together, these results indicate that 2Phe and 4Phe, but not 3Phe and 6Phe, possess a protein L like overall fold.

In summary, both the random library and point mutant studies suggest that at least one polar residue is required on the surface of the fourth beta strand for proper folding: no viable protein se-

Table 4. *Expected and observed frequencies of nonpolar residues at surface positions in selected protein L variants* [a]

	#59	#61	#63	0-NP	1-NP	2-NP	3-NP	NP(#59,#61)
Expected from design	0.50	0.50	0.50	0.12	0.37	0.38	0.13	0.25
Observed	0.28	0.26	0.26	0.15	0.62	0.23	0	0
Expected from obs. ind. freq.				0.26	0.46	0.24	0.04	0.07

[a] The first three columns are the expected and observed frequencies of nonpolar residues at the surface positions 59, 61, and 63. The next four columns are the expected and observed frequencies of sequences with 0, 1, 2, or 3 nonpolar amino acids at these surface positions. The final column is the expected and observed frequency of sequences with nonpolar amino acids at both positions 59 and 61. The first row is the frequency expected from the library design, and the second row, the observed frequency in the selected sequences. The third row is the expected frequency of multiple substitutions given the observed frequencies at the individual positions.

quences with all the surface residues hydrophobic were recovered from the selection, and the designed mutants (3Phe and 6Phe) with all surface residues hydrophobic did not fold properly.

As described in the introduction, the extent to which the folding rate is limited by a search through possible hydrophobic core packing arrangements can potentially be assessed by determining the effects of surface hydrophobic substitutions on the folding rate. Despite the dramatic increase in exposed hydrophobic surface area, the refolding rates of 2Phe (81 s^{-1}) and 4Phe (75 s^{-1}) were slightly faster than that of wild-type protein L (60 s^{-1}). The unfolding rates of both proteins were considerably faster than that of the wild-type protein (Table 5).

To probe the consequences of surface hydrophobic substitutions at sites distributed throughout the protein L structure, the folding kinetics of polar to hydrophobic substitutions in the first strand (K7F), the first β-turn (A13V, G15V), and the α-helix (E32I) were also determined. Like the 2Phe and 4Phe mutants, K7F and E32I increased the unfolding rate and had virtually no effect on the folding rate. The K7F mutation also reduced the denaturant dependence of the unfolding rate (Fig. 5), suggesting that the introduced phenylalanine residue is more buried in the folding transition state than it is in the native state. A13V and G15V were the only mutations examined in this study, which had a significant effect on the folding rate. However, the first β-turn, in which these two residues lie, was previously found to be largely structured in the folding transition state: the G15A and N14A mutations have large effects on the folding rate but very little effect on the unfolding rate (Table 5, Gu et al., 1997). It is possible that the reduction in the folding rate in the A13V and G15V mutants is due to steric interference with turn formation rather than the changes in hydrophobicity per se. This is supported by consideration of the two substitutions at position 15: while the G15A mutation produces a 10-fold decrease in the folding rate, the subsequent mutation to valine produces little further decrease in the folding rate but does increase the unfolding rate.

Discussion

Binary patterning

What range of hydrophobic/hydrophilic patterns are compatible with protein L and SH3 domain folding? The random sequence results and the biophysical properties of the Phe mutants demonstrate that at least one polar amino acid must be present on the

	α-helix	47	48	49	50	51	52	β-turn II				57	58	59	60	61	62	63	64
WT		W	T	V	D	V	A	D	K	G	Y	T	L	N	I	K	F	A	G
2PHE		W	T	V	D	V	A					F	L	N	I	F	F	A	G
3PHE		W	T	V	D	V	A					F	L	F	I	F	F	A	G
4PHE		W	F	V	D	V	F					F	L	N	I	F	F	A	G
6PHE		W	F	V	F	V	F					F	L	F	I	F	F	A	G

Fig. 3. Amino acid sequences of the second β-hairpin for protein L and the Phe mutants. Grey, buried positions; black, hydrophobic residues.

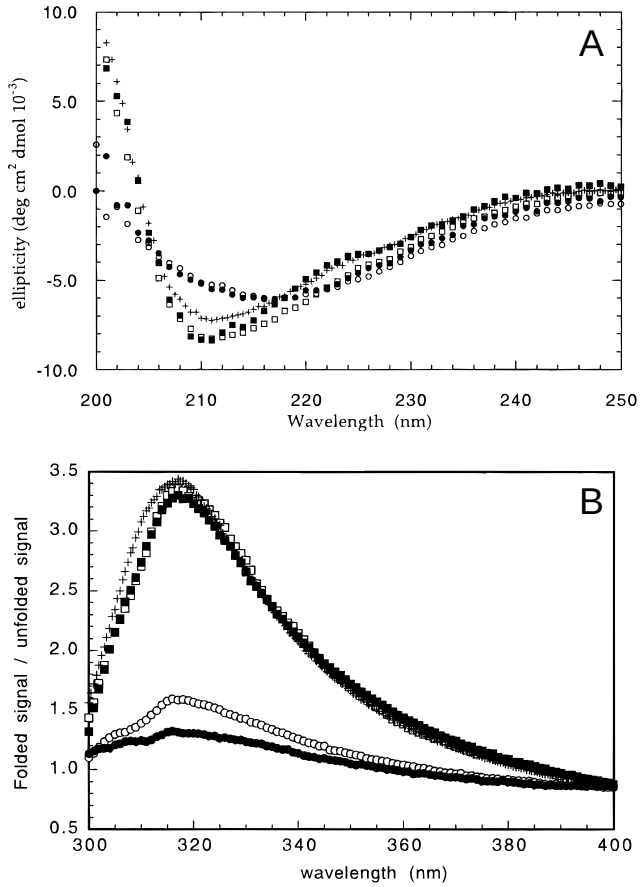

Fig. 4. A: CD spectra of protein L (+), 2Phe (□), 3Phe (○), 4Phe (■), and 6Phe (●). Each spectrum is the average of six scans. Protein concentrations were 20 μM. **B:** The ratio of fluorescence emission intensities of folded (0 M GuHCl) to unfolded (5.4 M GuHCl) protein L (+), 2Phe (□), 3Phe (○), 4Phe (■), and 6Phe (●). Protein concentrations were 12 μM. The change in the fluorescence spectra of 2Phe and 4Phe upon folding is very similar to protein L, while there is relatively little change in the spectra of 3Phe and 6Phe.

surface of the fourth strand in order for protein L to fold properly. As in naturally occurring proteins (Eisenberg et al., 1984; West & Hecht, 1995; Bystroff & Baker, 1998), an alternating polar/nonpolar pattern is not required in surface β-strands. It appears that the primary requirement on the surface of the fourth β-strand in protein L is some minimal amount of polarity. In the absence of strand polarity, there may be little to distinguish the two sides of the β-strand during folding; either side may be able to participate in a hydrophobic core. In the case of the SH3 domain, surface hydrophobic substitutions in the second half of the RT loop appear to compromise the assembly of the SH3-gene VIII fusion protein into the phage coat and/or the accumulation of soluble fusion protein in the periplasm. No surface hydrophobic substitutions were present in this region of SH3 in any of the selected variants, and single hydrophobic surface substitutions in this region did not significantly destabilize the protein.

In a very complementary recent study, the tolerance of multiple surface hydrophobic substitutions in alpha helices was studied using P22 Arc repressor as a model system (Cordes & Sauer, 1999). The two different helices studied had quite different tolerances to surface hydrophobic substitutions: in one helix up to five polar residues could be mutated to hydrophobic residues without loss of activity, while in the other, more than two such substitutions abolished activity. The effect of the substitutions was also found to depend on the broader sequence context: a triple hydrophobic substitution was tolerated in the wild-type protein but not in a mutant having a nearly identical structure. Taken together, the results on Arc repressor, the SH3 domain, and protein L suggest that the effect of surface hydrophobic substitutions is strongly dependent on the broader sequence and structural context of the protein.

Effect of surface exposed nonpolar residues on folding kinetics

The 11 surface hydrophobic substitutions in protein L and the SH3 domain examined in this study generally had little effect on the folding rate. These results suggest that an unrestricted combinatorial search through alternative core packing arrangements is not rate limiting; such a search should be slowed by an increase in the number of hydrophobic residues (particularly in the case of the 4PHE mutant with the simultaneous introduction of four surface phenylalanine residues). Instead, native-like local interactions may bias the polypeptide chain toward the formation of the correct native core packing arrangement. In this scenario, disrupting such local interactions would have a large affect on the folding rate. Mutations made in both protein L and SH3 that destabilize a specific β-hairpin do, in fact, reduce the folding rate considerably (Gu et al., 1997; Grantcharova et al., 1998). Formation of the β-hairpin in both proteins may be important in guiding formation of the native core by properly orienting critical hydrophobic residues. Additionally, the folding transition state ensemble may be so overdetermined by the global hydrophobic–hydrophilic pattern that it is not significantly perturbed by surface hydrophobic substitutions.

Overall, the lack of significant effects of surface hydrophobic substitutions suggests protein folding rates are determined more by the balance between favorable native interactions and configurational entropy loss than by a competition between native and nonnative interactions or the roughness of the free energy landscape. The primacy of native interactions in determining folding rates is consistent with experimental findings that protein folding mechanisms are largely determined by the topology and the distribution of contacts in the native state, and the encouraging success models based on native state topology have had in predicting features of folding reactions (Alm & Baker, 1999a, 1999b). The robustness of proteins to surface hydrophobic substitutions was presumably important for the evolution of protein binding sites that involve surface hydrophobic patches.

Materials and methods

Materials

All media, chemicals, and enzymes used in molecular biological and biophysical experiments were as described previously (Gu et al., 1995; Scalley et al., 1997).

Methods

Phagemid library construction and selection

The strand phagemid libraries were constructed using previously described procedures (Gu et al., 1995; Riddle et al., 1997)

Table 5. *Kinetic and thermodynamic parameters for protein L point mutants*

Protein	ΔG (kcal mol^{-1})	m (kcal mol^{-1} M^{-1})	k_f (s^{-1})	m_f (kcal mol^{-1} M^{-1})	k_u 4M (s^{-1})	m_u (kcal mol^{-1} M^{-1})
wt	4.6	1.9	61	1.5	0.6	0.5
2Phe	3.7	2.0	81	1.5	5.9	0.5
4Phe	3.6[a]	1.6[a]	75	1.1	4.3	0.5
K7F	5.3	1.9	73	1.5	1.7	0.4
A13V	4.4	2.2	20	1.7	2.0	0.6
G15V	3.1	2.6	4.3	1.8	3.2	0.6
E32I[b]	3.9	2.0	55	1.5	4.2	0.6
G15A[c]	3.6	2.1	7.1	1.8	1.2	0.6
N14A[c]	2.9	1.9	7.4	1.7	0.9	0.5

[a] Calculated from kinetic data.
[b] From Kim et al. (1998b).
[c] From Gu et al. (1997).

using oligonucleotides from Ransom Hill Bioscience, Inc. (Ramona, California) with sequences 5′-gat aaa ggt tat act DHW DHW DHW DHW DHW DHW gga tag atg cac-3′, and 5′-a cgc gtt tcc tcc gtg cat cta tcc-3′ for the protein L library, and 5′-g ggg gcc atg ggg aat tca cat atg AET ttc gtg gct ctc tat gac tat-3′, and 5′-gg ggg ctg cag TFT ttc TFC TIT TIT gaa GFT caa gtc TJT TIC AJT TFT gga CIC ata gtc ata gag agc-3′ for the SH3 library, where E = 75% T and 25% C, F = 75% A and 25% C, I = 75% A and 25% T, and J = 75% A and 25% G. The preparation of the affinity matrix and the biopanning and colony lift screen of the phagemid libraries, was performed as described previously (Gu et al., 1995; Riddle et al., 1997).

Construction, expression, and characterization of Phe and point mutants

The Phe mutants were constructed using PCR mutagenesis with primers from Ransom Hills Bioscience, Inc. The point mutations were made with the QuickChange site-directed mutagenesis kit (Strategene, La Jolla, California). All sequences were verified by DNA sequencing. Expression and purification was carried out as described previously (Gu et al., 1995). The identities of mutant proteins were verified by mass spectrometry. Equilibrium CD and fluorescence measurements and stopped flow experiments were performed as previously described (Scalley et al., 1997).

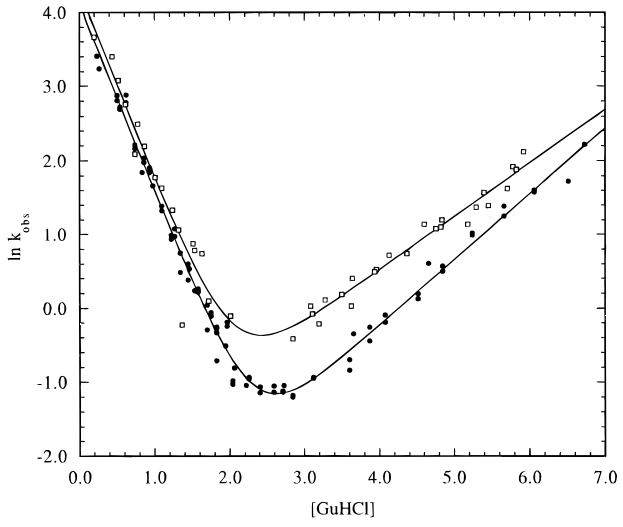

Fig. 5. The denaturant dependence of the observed folding and unfolding rates of K7F (□) and wild-type protein L (●).

Acknowledgments

We thank Brian Kuhlman and Qian Yi for helpful comments on the manuscript. This work was supported by a grant from the NIH and by young investigator awards to D.B. from the NSF and the Packard Foundation.

References

Agashe VR, Shastry MC, Udgaonkar JB. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature 377*:754–757.
Alm E, Baker D. 1999a. Matching theory and experiment in protein folding. *Curr Opin Struct Biol 9*:189–196.
Alm E, Baker D. 1999b. Prediction of protein folding mechanisms from free energy landscapes derived from native structures. *Proc Natl Acad Sci USA 96*:11305–11310.
Bystroff C, Baker D. 1998. Prediction of local structure in proteins using a library of sequence-structure motifs. *J Mol Biol 281*:565–577.
Cordes MH, Sauer RT. 1999. Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Sci 8*:318–325.
Creighton TE. 1993. *Proteins—Structures and molecular properties, 2nd ed*. New York: WH Freeman and Company. p 4.
Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*:561–602.
Eisenberg D, Weiss RM, Terwilliger TC. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA 81*: 140–144.
Grantcharova V, Baker D. 1997. Folding dynamics of the Src SH3 domain. *Biochemistry 36*:15685–15692.
Grantcharova V, Riddle DS, Santiago JV, Baker D. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol 5*:714–720.
Gu H, Kim DE, Baker D. 1997. Contrasting roles for symmetrically disposed beta turns in the folding of a small protein. *J Mol Biol 274*:588–596.
Gu H, Yi Q, Bray ST, Riddle DS, Shiau AK, Baker D. 1995. A phage display system for studying the sequence determinants of protein folding. *Protein Sci 4*:1108–1117.
Kim DE, Gu H, Baker D. 1998a. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc Natl Acad Sci 95*:4982–4986.

Kim DE, Yi Q, Gladwin ST, Baker D. 1998b. The single helix in Protein L is disrupted at the rate limiting step in folding. *J Mol Biol 284*:807–815.
Pakula AA, Sauer RT. 1990. Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface. *Nature 344*:363–364.
Predki PF, Aagrawal V, Trunger AT, Regan L. 1996. Amino-acid substitutions in a surface turn modulate protein stability. *Nat Struct Biol 3*:54–58.
Riddle D, Santiago J, Grantcharova V, Baker D. 1997. Functional rapidly folding proteins from simplified amino acid sequences. *Nat Struct Biol 4*:805–809.
Scalley ML, Yi Q, Gu H, McCormack A, Yates JR III, Baker D. 1997. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry 36*:3373–3382.
Smith CK, Regan L. 1995. Guidelines for protein design: The energetics of beta sheet side chain interactions. *Science 270*:980–982.
Sosnick TR, Mayne L, Englander SW. 1996. Molecular collapse: The rate-limiting step in two-state cytochrome c folding. *Proteins 24*:413–426.
West MW, Hecht MH. 1995. Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Sci 4*:2032–2039.