Protein Science (2000), *9*:2457–2469. Cambridge University Press. Printed in the USA.

The design of a hyperstable mutant of the Abp1p SH3 domain by sequence alignment analysis

ARIANNA RATH[1] AND ALAN R. DAVIDSON[1,2]

[1]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

(RECEIVED June 6, 2000; FINAL REVISION September 12, 2000; ACCEPTED September 18, 2000)

Abstract

We have characterized the thermodynamic stability of the SH3 domain from the *Saccharomyces cerevisiae* Abp1p protein and found it to be relatively low compared to most other SH3 domains, with a T_m of 60 °C and a ΔG_u of 3.08 kcal/mol. Analysis of a large alignment of SH3 domains led to the identification of atypical residues at eight positions in the wild-type Abp1p SH3 domain sequence that were subsequently replaced by the residue seen most frequently at that position in the alignment. Three of the eight mutants constructed in this way displayed increases in T_m ranging from 8 to 15 °C with concomitant increases in ΔG_u of up to 1.4 kcal/mol. The effects of these substitutions on folding thermodynamics and kinetics were entirely additive, and a mutant containing all three was dramatically stabilized with a T_m greater than 90 °C and a ΔG_u more than double that of the wild-type domain. The folding rate of this hyperstable mutant was 10-fold faster than wild-type, while its unfolding rate was fivefold slower. All of the stabilized mutants were still able to bind a target peptide with wild-type affinity. We have analyzed the stabilizing amino acid substitutions isolated in this study and several other similar sequence alignment based studies. In approximately 25% of cases, increased stability can be explained by enhanced propensity of the substituted residue for the local backbone conformation at the mutagenized site.

Keywords: protein engineering; protein folding kinetics; protein stability; sequence alignment; SH3 domain; structural propensity

The design of mutations to increase the stability of proteins is an important goal of protein engineering (Russell & Taylor, 1995; Querol et al., 1996; Vieille & Zeikus, 1996). To this end, a variety of approaches have been used to stabilize proteins by mutagenesis. For example, significant protein stabilization has been achieved by engineering disulfide bonds into proteins (Sauer et al., 1986; Matsumura et al., 1989), substituting Gly residues in helices with Ala (Hecht et al., 1986), replacing residues to reduce the entropy of the unfolded state (Matthews et al., 1987), and removing unsatisfied polar groups (Hendsch et al., 1996). In addition, stabilizing substitutions in a variety of proteins have been selected by comparison with thermophilic homologues (Takagi et al., 1990; Kirino et al., 1994; Kawamura et al., 1998). Despite many successes in engineering thermostability, there are still no general principles guiding changes that increase protein stability.

Reprint requests to: Alan R. Davidson, Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada; e-mail: alan.davidson@utoronto.ca.

Abbreviations: Abp1p, *Saccharomyces cerevisiae* actin-binding protein 1; SH3, Src-homology 3; GuHCl, guanidine hydrochloride; CD, circular dichroism; WT, wild-type; T_m, temperature at the midpoint of the unfolding transition; ΔG_u, free energy of unfolding in water at 25 °C.

The design of stabilizing mutations through sequence alignment analysis provides a powerful alternative to other methods. Sequence alignment based methods generally involve comparing the sequence of a protein of interest to an alignment of homologous sequences. Positions in the protein of interest that are occupied by amino acids rarely seen in the related sequences are substituted with the amino acid that is seen most often. Studies performed on immunoglobulin variable domains, the DNA binding domain of p53, and a GroEL minichaperone have shown that both small and large sequence alignments can be used in this way to design stabilizing mutations (Steipe et al., 1994; Nikolova et al., 1998; Wang et al., 1999; Wirtz & Steipe, 1999). Many single amino acid substitutions that increased the free energy of unfolding (ΔG_u) of these proteins by more than 1 kcal/mol were isolated in these studies and between 25 and 50% of the mutants constructed provided some degree of stabilization. In addition, the effects of single stabilizing substitutions were generally fully additive, so that highly stabilized multiple mutants could be constructed. The advantages of sequence alignment methods are that they are simple to carry out, they do not require knowledge of the three-dimensional structure of the protein under investigation, and the resulting mutants generally maintain their biological function. In the work described here, we use a sequence alignment based approach to design mu-

tations that stabilize the SH3 domain from the yeast actin binding protein 1 (Abp1p).

Due to its small size (~60 residues) and its amenability to in vitro analysis, the SH3 domain provides an excellent model system for protein folding studies (Chen et al., 1996; Grantcharova & Baker, 1997; Plaxco et al., 1998; Filimonov et al., 1999). SH3 domains have been found in more than 350 proteins in a variety of organisms and function as mediators of specific protein–protein interactions by binding to PXXP-containing sequence motifs in target proteins (Pawson, 1995; Dalgarno et al., 1997). The SH3 domain fold is composed of five β-strands arranged into two sheets packed orthogonally (Fig. 1). The first sheet is formed by β-strands *a*, *e*, and the first half of *b*, while the second is formed by β-strands *c*, *d*, and the second half of *b*. A kink in β-strand *b* allows it to participate in both sheets. β-Strands *a* and *b* are separated by the long RT-Src loop, which possesses an irregular antiparallel structure. The shorter N-Src and distal loops are found between β-strands *b* and *c*, and *c* and *d*, respectively. The four residues N-terminal to β-strand *b* form a type II β-turn, and the three residues separating β-strands *d* and *e* are generally found in a 3_{10}-helical conformation. Peptide binding by the SH3 domain is mediated by a surface region rich in aromatic residues (Fig. 1) and by various polar residues located in the RT-Src and N-Src loops.

Abp1p, a 592 residue protein composed of an N-terminal actin-binding domain, a central Pro-rich region and a C-terminal SH3 domain, plays important roles in regulating the actin cytoskeleton in yeast (Drubin et al., 1990; Lila & Drubin, 1997). The Abp1p SH3 domain interacts with a PXXP containing site in Srv2p, an adenylate cyclase and actin-binding protein, and is required for Srv2p localization to cortical actin patches (Freeman et al., 1995). The Abp1p SH3 domain was chosen as an object of study because it possesses a number of amino acid residues in its sequence that are rarely found at their respective positions in other SH3 domains. In addition, the functional effects of mutations in this domain can ultimately be investigated through relatively simple in vivo studies in yeast.

The goal of the experiments described here was to determine whether data derived from an extensive SH3 domain sequence alignment analysis performed in our laboratory (S.M. Larson & A.R. Davidson, unpubl. obs.) could be used to predict stabilizing amino acid substitutions in the Abp1p SH3 domain. To this end, eight positions within the Abp1p SH3 domain that are occupied by residues not typically seen in SH3 domains were replaced with the amino acid most commonly observed at each position. Three of the eight substitutions were found to be significantly stabilizing, and combining these substitutions resulted in additive increases in stability. The folding thermodynamic and kinetic properties of these stabilized mutants have been fully characterized, as have their in vitro peptide binding activities. In an attempt to elucidate some general principles concerning the type of substitutions that lead to protein stabilization, we have also carefully examined the stabilizing substitutions made in other sequence alignment based mutagenesis studies.

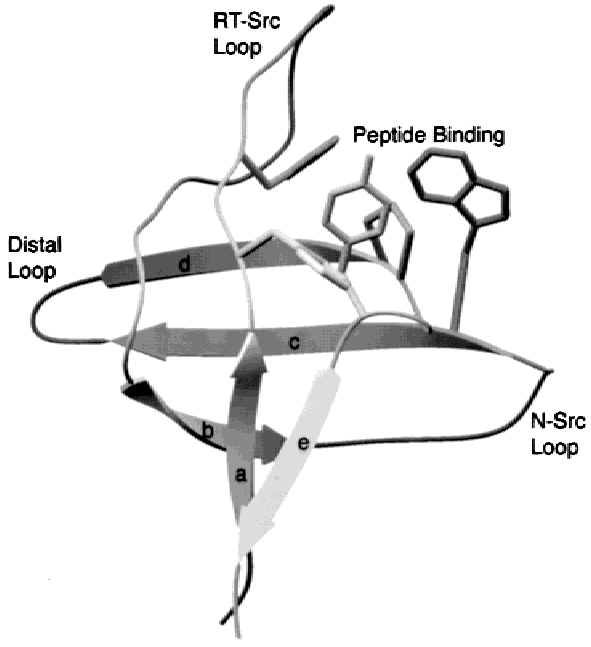

Fig. 1. Structure of the SH3 domain. Ribbon drawing of the Sem-5 SH3 domain based on the X-ray crystal structure (PDB identifier 1SEM; Lim et al., 1994b). The strands comprising the two β-sheets are labeled a–e. The side chains of conserved residues involved in peptide binding are shown. Graphics were produced using the program SETOR (Evans, 1993).

Results

Thermodynamic characterization of the Abp1p SH3 domain

The sequence of the Abp1p SH3 domain is shown in Figure 2 with the residue numbering system used in this work. The wild-type (WT) protein and all mutants described here possessed six histidine residues at their C-termini, which were added to allow purification by Ni-affinity chromatography. Figure 3 shows the circular dichroism (CD) spectrum of the WT Abp1p SH3 domain, which is similar to that of other SH3 domains with a maximum at about 220 nm (Lim et al., 1994a; Maxwell & Davidson, 1998). There is a marked difference in the CD spectra of the folded and unfolded Abp1p SH3 domain, with a large decrease in ellipticity at 220 nm. Unfolding of the Abp1p SH3 domain was monitored by the change in CD signal at 220 nm (Fig. 4A,B). Resulting thermal and chemical denaturation curves were cooperative and reversible and were fit by a two-state model. Because calculation of the free energy of unfolding (ΔG_u) of the stabilized mutants was somewhat complicated by the sloped baselines observed in the guanidine hydrochloride (GuHCl) denaturant melts monitored by CD, ΔG_u values were confirmed by also using Trp fluorescence to monitor GuHCl induced denaturation (data not shown). The ΔG_u values determined by both methods were in good agreement with one another, and the average of the two was reported (Table 1). The ΔG_u of the WT domain and V21K mutant were also determined by denaturation in urea, which produced more easily analyzed curves with less sloped baselines (data not shown). However, the more stable mutants described below could not be fully melted by this denaturant. The WT Abp1p SH3 domain was found to possess a midpoint transition temperature (T_m) of 60.3 °C and a ΔG_u of 3.08 kcal/mol at room temperature (Table 1). The seemingly low ΔG_u of this domain when compared to its T_m is likely a result of the small size and resulting low ΔC_p of the domain (Alexander et al., 1992).

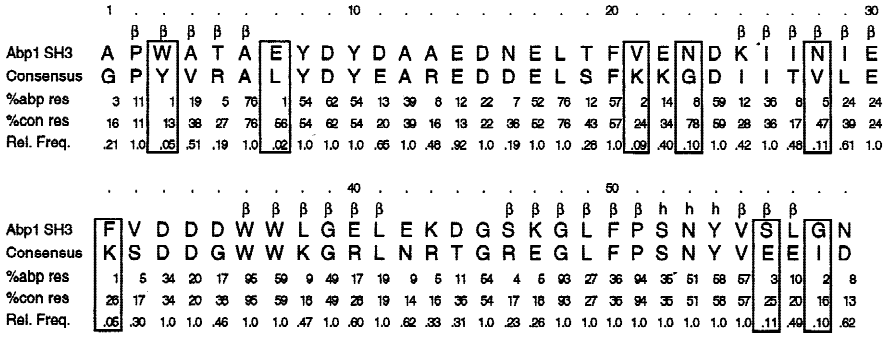

Fig. 2. Alignment of the Abp1p SH3 domain sequence with the SH3 domain consensus sequence. Amino acid residues are assigned numbers according to their position in our alignment. Based on this alignment, the most frequently observed amino acid residue at each position in the SH3 domain was determined and is given as the consensus sequence. The percent frequency of occurrence observed for each residue in the Abp1p SH3 domain and in the SH3 domain consensus sequence is indicated by %abp res and %con res, respectively. The relative frequency of occurrence was obtained by dividing the Abp1p SH3 domain residue frequency with that of the consensus residue. A structural alignment was used to determine the secondary structure common to solved SH3 domains. β indicates residues observed to be in β-sheet conformation, and h indicates those found in 3_{10} helical conformation.

Design of Abp1p SH3 domain mutations

Because the Abp1p SH3 domain was observed to be less stable than most other characterized SH3 domains (Lim et al., 1994a; Chen et al., 1996; Grantcharova & Baker, 1997; Knapp et al., 1998; Maxwell & Davidson, 1998), we hypothesized that this domain might contain residues at some positions that are not typically seen in SH3 domains, and that these unusual residues might cause the observed lowered stability. The sequence of the Abp1p SH3 domain was compared to a consensus SH3 sequence derived from an alignment of 266 nonredundant (no two sequences were more than 90% identical) SH3 domain sequences constructed and analyzed in our laboratory (S.M. Larson & A.R. Davidson, unpubl. obs.). At each position in the alignment, the frequency of occurrence of the residue found in the Abp1p SH3 domain was compared to the frequency of occurrence of the most commonly observed residue at that position. For example, at position 7 in the alignment, the most commonly observed residue, Leu, is found in 56% of SH3 domain sequences. Glu, the residue at this position in the Abp1p SH3 domain, is found in only 1% of sequences, indicating that it is a highly unusual residue at this position in SH3 domains. By dividing these two frequencies, a "relative frequency" of 0.02 can be assigned to this position in the Abp1p SH3 domain. The

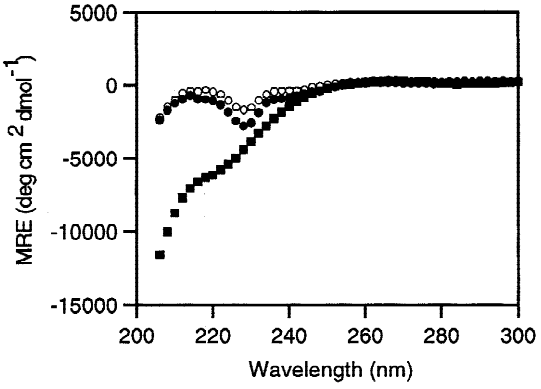

Fig. 3. CD spectra of WT and triple mutant Abp1p SH3 domains. The WT spectra were obtained from a 30 μM sample in a 0.1 cm cuvette. Folded (●) and unfolded samples (■) were at 25 and 90 °C, respectively. The triple mutant spectrum (○) was obtained at 25 °C from a 35 μM sample in a 0.1 cm cuvette. Mean residue ellipticity (MRE) is shown on the *y*-axis.

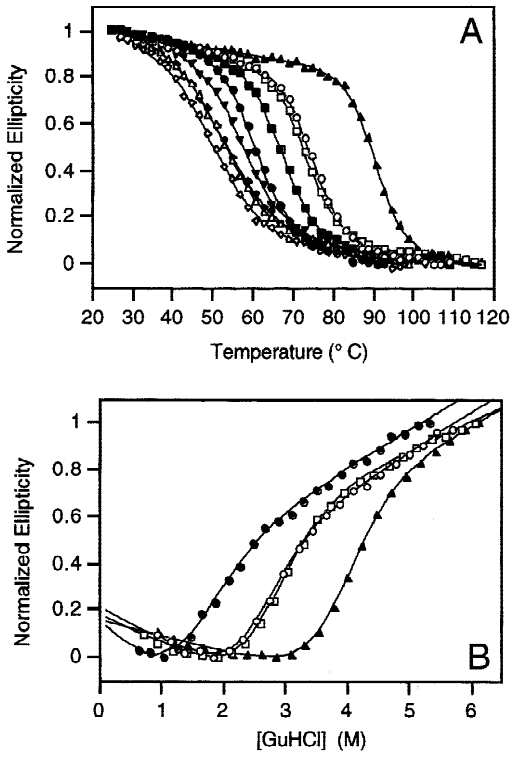

Fig. 4. Unfolding of WT and mutant Abp1p SH3 domains. **A:** Thermal denaturation experiments performed on WT (●), triple (▲), E7L (○), N23G (□), V21K (■), G58I (▼), N28V (△), W3Y (♦), and S56E (◇) mutant Abp1p SH3 domains. **B:** GuHCl denaturation of WT (●), and E7L (○), N23G (□), and triple (▲) mutant Abp1p SH3 domains monitored by CD. Ellipticity values are expressed as a fraction of the total change. The lines joining the points in each graph are theoretical fits to the data based on the thermodynamic parameters shown in Tables 1 and 2.

Table 1. *Protein folding thermodynamics and kinetics of stabilized mutants*

	Equilibrium experiments		Kinetic experiments				
Protein	T_m (°C)[a]	$\Delta\Delta G_u$[b] (kcal mol^{-1})	k_f[c] (s^{-1})	k_u (s^{-1}) × 10^3	$\Delta\Delta G_{ts\rightarrow u}$[d] (kcal mol^{-1})	$\Delta\Delta G_{n\rightarrow ts}$ (kcal mol^{-1})	$\Delta\Delta G_u$[e] (kcal mol^{-1})
WT	60.3		10.1 ± 1.1	27.4 ± 9.5			
E7L	75.5	1.23	17.3 ± 1.6	4.8 ± 1.6	0.31	0.70	1.01
V21K	68.1	0.38	22.5 ± 0.6	44.8 ± 4.3	0.47	−0.05	0.42
N23G	73.4	1.41	34.5 ± 2.1	10.8 ± 2.9	0.77	0.32	1.09
TRIPLE	92.2	3.35	100.2 ± 11.8	4.7 ± 3.6	1.44	1.09	2.53
Expected[f]	*95.7*	*3.02*			*1.55*	*0.97*	*2.52*

[a] T_m values were determined by temperature-induced unfolding experiments monitored by CD. Values were averaged from at least two experiments and the average errors in these values were ±2%.

[b] Equilibrium ΔG_u values (in H_2O) were determined by analysis of urea and/or GuHCl denaturation experiments at 25 °C. All ΔG_u values were averaged from denaturant melts monitored by CD and by tryptophan fluorescence emission at 340 nm except V21K, which was monitored by fluorescence only. Each value is based on at least two experiments, and the average errors in these measurements were ±5%. The WT Abp1p SH3 domain possesses a ΔG_u of 3.08 kcal/mol at 25 °C.

[c] Folding and unfolding rates shown are the rates in H_2O derived by extrapolation from the fitting of chevron plots. The errors associated with fitting the experimental data are shown.

[d] The changes in the free energy between the transition state (ts) and the native or unfolded states were calculated using folding rates determined at 0.3 M GuHCl and unfolding rates determined at 4 M GuHCl as follows: $\Delta\Delta G_{ts\rightarrow u} = -RT \ln(k_f \text{ at 0.3 M GuHCl (WT)}/k_f \text{ at 0.3 M GuHCl (mutant)})$; $\Delta\Delta G_{n\rightarrow ts} = -RT \ln(k_u \text{ at 4 M GuHCl (mutant)}/k_u \text{ at 4 M GuHCl (WT)})$. This method avoids inaccuracy arising from the extrapolations used to derive the folding and unfolding rates in H_2O.

[e] $\Delta\Delta G_u = \Delta\Delta G_{ts\rightarrow u} + \Delta\Delta G_{n\rightarrow ts}$.

[f] Expected values are those expected for the triple mutant if the effects of the individual stabilizing substitutions are completely additive.

average relative frequency for residues in the Abp1p SH3 domain is 0.66 with a standard deviation of 0.37. Most positions in the Abp1p SH3 domain have residues with relative frequencies above 0.29, the mean relative frequency minus one standard deviation. However, eight positions stand out as being atypical, possessing residues with relative frequencies less than or equal to 0.11, the mean relative frequency minus 1.5 times the standard deviation (boxed positions in Fig. 2). To assess the role of these significantly atypical residues in the stability and function of this SH3 domain, each was mutated individually to the residue seen in the consensus SH3 domain sequence.

Thermodynamic characterization of the Abp1p SH3 domain mutants

The effects of each of the eight single amino acid replacements constructed were assessed by thermal denaturation studies (Fig. 4A). Consistent with our hypothesis that atypical residues in the Abp1p SH3 domain cause its relatively low stability, the E7L, V21K, and N23G substitutions were found to be significantly stabilizing, increasing the T_m by 15.2, 7.8, and 13.1 °C, respectively (Table 1). Combining these three substitutions in one mutant produced a hyperstable SH3 domain with a T_m of 92.2 °C, which is more than 10 °C higher than that observed for any other WT SH3 domain. It can also be seen in Table 1 that the observed T_m of the triple mutant is only 3.5 °C less than what would be expected if the stabilizing effects of the three mutants were completely additive. The stabilizing effects observed by thermal denaturation were paralleled in chemical denaturation experiments (Fig. 4B). The E7L, V21K, and N23G point mutants increased the free energy of unfolding of the domain by 1.23, 0.38, and 1.41 kcal/mol, respectively (Table 1). The increase in ΔG_u observed for the triple mutant was very close to the sum of the increases observed for the individual mutants, again demonstrating the additivity of these effects. The very high correlation between the observed ΔT_m and $\Delta\Delta G_u$ values of these mutants (Fig. 5A) supports the accuracy of our data. Despite the large increase in stability observed for the triple mutant, its CD spectrum is almost identical to that of the WT domain (Fig. 3), indicating that a major alteration in the structure of this mutant is unlikely.

The five other substitutions tested were destabilizing to varying degrees (Table 2). The F31K mutant could not be produced in *Escherichia coli* at a level high enough to purify, indicating that this mutant is highly destabilized. Unstable mutants are generally rapidly degraded by cellular proteases (Parsell & Sauer, 1989; Inoue & Rechsteiner, 1994), and we have observed that the stability of SH3 domain mutants and their level of accumulation in *E. coli* are generally well correlated (A.A. DiNardo, A. Rath, & A.R. Davidson, unpubl. results). Because position 31 lies in a solvent exposed loop in all solved SH3 domain structures, the implied low stability of the F31K substitution is difficult to explain. Another

Table 2. *Thermostability of destabilized mutants*

Protein	T_m (°C)
WT	60.3
W3Y	51.5
N28V	51.5
F31K	n.d.[a]
S56E	49.8
G58I	57.0

[a] Not determined due to low expression of mutant.

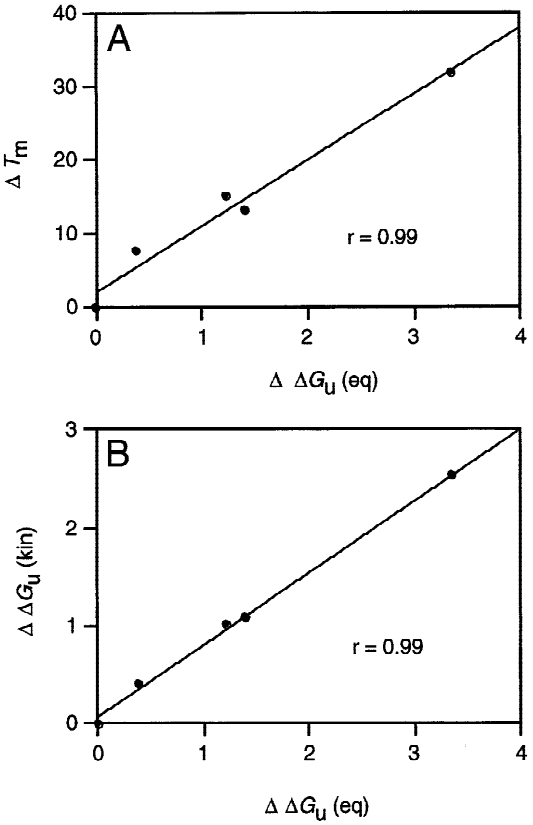

Fig. 5. Plots of the ΔT_m of each stabilized Abp1p SH3 domain mutant vs. its $\Delta\Delta G_u$ as measured by GuHCl denaturation (**A**), and the $\Delta\Delta G_u$ of each stabilized mutant derived by kinetic experiments versus its $\Delta\Delta G_u$ as measured by GuHCl denaturation (**B**).

surprising result was the lowered stability of the N28V mutant. This position lies in the hydrophobic core in most SH3 domain structures; thus, substitution of a polar residue with a hydrophobic one at this position was expected to be stabilizing. These results suggest that the Abp1p SH3 domain possesses structural features in the region comprising residues 28–31 that are distinct from most other SH3 domains. This region of the Abp1p SH3 domain may resemble the Nebulin or Phospholipase-C SH3 domains, which also possess Asn residues at position 28 (Kohda et al., 1993; Politou et al., 1998). The five destabilized Abp1p SH3 domain mutants were not characterized further in this work.

Folding kinetics of the stabilized Abp1p SH3 domain mutants

To gain insight into the mechanism by which the E7L, V21K, and N23G substitutions stabilize the Abp1p SH3 domain, the unfolding and refolding kinetics of the WT and mutant Abp1p SH3 domains were determined by stopped-flow fluorescence. The "chevron" plots of the unfolding and refolding rates of the WT and mutant domains in various concentrations of GuHCl are shown in Figure 6. The folding rates (k_f) and unfolding rates (k_u) extrapolated to 0 M GuHCl for each mutant are shown in Table 1. All three substitutions result in significant (two- to threefold) increases in the k_f with N23G having the largest effect. The E7L substitution had the largest effect on the unfolding rate, slowing it by fivefold, while the

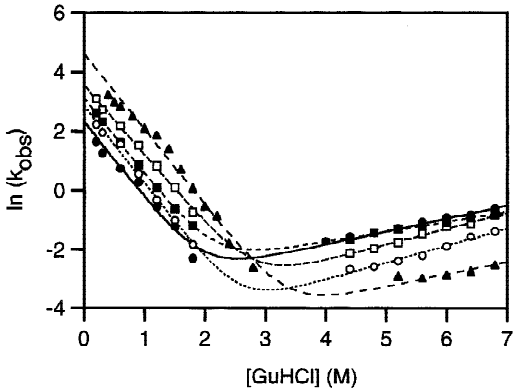

Fig. 6. Dependence of the observed unfolding and refolding rate constants of WT and stabilized mutant Abp1p SH3 domains on GuHCl concentration. Folding kinetics experiments were performed on WT (●), E7L (○), V21K (■), N23G (□), and triple (▲) mutant Abp1p SH3 domains. The lines joining the points in each graph are theoretical fits to the data.

N23G mutant displayed an approximately twofold reduction in its unfolding rate. Unexpectedly, the V21K substitution appears to significantly increase the unfolding rate.

To assess the additivity of the mutant effects on the folding and unfolding rates of the domain, the changes in free energy between the folding transition state and the unfolded state ($\Delta\Delta G_{ts\rightarrow u}$) and between the native state and the folding transition state ($\Delta\Delta G_{n\rightarrow ts}$) were calculated for each mutant compared to the WT protein (Table 1). These calculations were performed by comparing the WT and mutant folding rates measured at 0.3 M GuHCl and the unfolding rates measured at 4 M GuHCl. This method avoids inaccuracy that can be caused by extrapolation of data to 0 M GuHCl (Grantcharova et al., 1998). As expected from the observed increases in k_f, each mutant shows a significant positive $\Delta\Delta G_{ts\rightarrow u}$. The E7L and N23G mutants also show positive $\Delta\Delta G_{n\rightarrow ts}$ values, a result of their slowed unfolding rates. The V21K mutant, however, shows a $\Delta\Delta G_{n\rightarrow ts}$ of close to 0. The increase in k_u in H_2O observed for the V21K mutant may simply be an artifact of extrapolation. The $\Delta\Delta G_{ts\rightarrow u}$ and $\Delta\Delta G_{n\rightarrow ts}$ values observed for the triple mutant are very close to the sum of these values seen for the individual mutants. Thus, the kinetic effects of the three stabilizing mutants are completely additive, as were the equilibrium effects discussed above. Although the $\Delta\Delta G_u$ values calculated from the kinetic data (Table 1) are generally slightly lower than those determined by equilibrium denaturant melts, the correlation between these values is excellent (Fig. 5B).

The dependence of the natural logarithms of the folding and unfolding rate constants on the concentration of GuHCl (m_f and m_u, respectively) for the WT and mutant Abp1p SH3 domain are shown in Table 3. The α-value given in Table 3 is calculated from m_f and m_u, and is thought to reflect the degree of solvent exposure of the folding transition state ensemble relative to the native state (Jackson & Fersht, 1991). The similarity of these values for the WT and mutant Abp1p SH3 domains suggests that the folding pathway of the mutants has not been significantly perturbed from that of the WT domain. Interestingly, the average α-value observed here (0.84) is significantly higher than the values of ~0.7 observed for other SH3 domains (Plaxco et al., 1998), indicating that the Abp1p SH3 domain possesses a more compact transition state structure than other SH3 domains. It can also be seen in Table 3

Table 3. *m* values derived from equilibrium and kinetic experiments

Protein	Equilibrium *m* value	Kinetic experiments			
		m_f[a]	m_u	*m* value[b]	α[c]
WT	1.64 0.78[d]	−2.57 ± 0.13	0.44 ± 0.06	1.78	0.85
E7L	1.68	−2.61 ± 0.09	0.58 ± 0.06	1.88	0.82
V21K	0.64[d]	−2.57 ± 0.03	0.34 ± 0.02	1.72	0.88
N23G	1.54	−2.34 ± 0.06	0.54 ± 0.05	1.70	0.81
TRIPLE	1.64	−2.30 ± 0.07	0.42 ± 0.12	1.75	0.85

[a] m_f and m_u are the GuHCl dependence of k_f and k_u, respectively, as derived from the fitting of chevron plots. The errors associated with these fits are shown.
[b] $m = -RT(m_f - m_u)$.
[c] $\alpha = m_f/(m_f - m_u)$.
[d] Values determined from urea induced denaturation. All other experiments used GuHCl as the denaturant. *m* values are expressed in units of kcal mol^{-1} M^{-1}.

that the overall *m* values, which represent the dependence of ΔG_u on the concentration of GuHCl, calculated from the kinetic and equilibrium experiments are quite similar for WT and the mutants. The correspondence between the thermodynamic values calculated from the equilibrium and kinetic experiments supports our assumption in analyzing these data that the two-state model accurately represents the folding transitions of both the WT Abp1p SH3 domain and the stabilized mutants.

Peptide binding activity of the stabilized Abp1p SH3 domain mutants

Because each of the E7L, V21K, and N23G substitutions had a significant effect both on the stability and folding kinetics of the Abp1p SH3 domain, we wondered if the peptide binding activity of the mutants would be affected. Accordingly, the in vitro affinity of the WT and mutant SH3 domains for a short sequence derived from the known in vivo Abp1p SH3 domain binding partner, Srv2p (Lila & Drubin, 1997), was determined. Binding was monitored by Trp fluorescence as has been previously described (Lim et al., 1994a; Maxwell & Davidson, 1998). The WT domain was calculated to bind with a K_d of 4.5 μM, and all the stabilized mutants displayed a similar affinity (data not shown). Figure 7 shows that the titration curves of the WT and triple mutant domains are nearly indistinguishable.

Native molecular weight determination

To ensure that the stabilizing effect of the E7L/V21K/N23G replacements did not result from protein oligomerization, analytical ultracentrifugation was performed. Experiments were performed on WT and triple mutant Abp1p SH3 domain solutions at protein concentrations between 20 and 140 μM, which are similar to and higher than the concentrations used in the thermodynamic, kinetic, and peptide binding assays. Representative data are shown in Figure 8. The WT molecular weight averaged from all experiments was 7,853 ± 331, very close to the expected molecular weight of the monomeric Abp1p SH3 domain calculated from amino acid composition (7,922). Similarly, the average molecular weight determined for the triple mutant, 7,193 ± 507, is close to the expected weight of the monomer (7,878). Figure 7 also shows that the predicted mass distribution for an Abp1p SH3 domain dimer differs substantially from that observed experimentally. Further evidence that the WT and triple mutant domains were monomeric was provided by fast protein liquid chromatography (FPLC) gel filtration experiments. At loading concentrations as high as 2.5 mM, both proteins eluted in a single peak at positions expected for monomeric protein (data not shown).

Discussion

Stabilizing amino acid substitutions are reliably predicted by sequence alignment analysis

In this study, we have successfully used sequence alignment analysis to design a hyperstable triple mutant of the Abp1p SH3 domain. The 50% increase in T_m and 100% increase in ΔG_u displayed by this mutant over the WT domain is a strikingly large relative in-

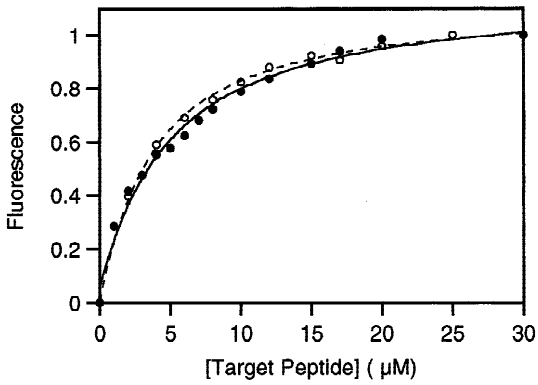

Fig. 7. Peptide binding activity of WT (●) and triple mutant (○) Abp1p SH3 domains. The peptide binding activity was monitored by tryptophan fluorescence emission at 330 nm. Fluorescence values are expressed as a fraction of the total change. The lines joining the points in each graph are theoretical fits to the data.

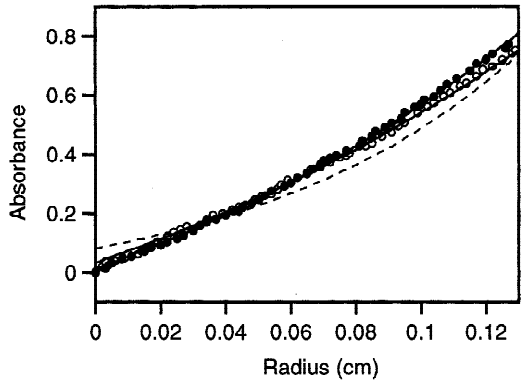

Fig. 8. Analytical ultracentrifugation of WT (●) and triple mutant (○) Abp1p SH3 domains. Sedimentation equilibrium data were obtained at 25 °C from a 28,000 rpm overnight centrifugation of a 100 μM solution of WT Abp1p SH3 and a 50 μM solution of the triple mutant. Equilibrium distribution of protein was determined by absorbance at 280 nm. Absorbance is expressed as a fraction of the maximum absorbance value. The radius given is the distance relative to the edge of the ultracentrifugation cell. The lines joining the points in each graph are theoretical fits to the data. The dashed line indicates the equilibrium protein distribution expected for an Abp1p SH3 domain dimer.

crease in thermodynamic stability. These results confirm previous findings demonstrating that stabilizing mutations can be identified by comparing the sequences of homologous proteins, even when none are thermophilic (Steipe et al., 1994; Nikolova et al., 1998; Wang et al., 1999; Wirtz & Steipe, 1999). The magnitudes of the increases in stability displayed by the single mutants characterized here are quite similar to those observed in other sequence alignment based studies (Table 4). For example, the three most stabilizing substitutions in the study on an immunoglobulin variable domain (IgL) increased the ΔG_u by 1.36, 1.05, and 0.38 kcal/mol (Steipe et al., 1994), and the stability increases observed in the study on the p53 DNA-binding domain ranged between 0.11–1.49 kcal/mol (Nikolova et al., 1998). Overall, between 25 and 50% of the mutants constructed in our study and these others proved to be stabilizing. The congruence of the results in our study and these others occurred despite the different types of sequence alignments used in each. The p53 study used a small alignment (22 sequences), the IgL study used a large alignment of closely related sequences, and we used a large and diverse alignment.

Interestingly, the stabilizing effects of substitutions designed by sequence analysis are generally additive (Serrano et al., 1993; Nikolova et al., 1998; Wang et al., 1999). In this study, we showed that the three stabilizing substitutions in the Abp1p SH3 domain are additive in their effects on both folding thermodynamics and kinetics. The additivity of the effects of these mutants implies that the stability conferred by these substitutions does not result from interaction between the mutated side chains (i.e., the effect of each substitution is independent of whether the WT or mutant residue is present at the other two positions).

The mechanism of stabilization by the Abp1p SH3 domain substitutions

Although the structure of the Abp1p SH3 domain has not yet been solved, examination of other SH3 domain structures provides some insight into the mechanisms of stabilization by the substitutions described here. In our laboratory, we have completed an extensive analysis of the structures of 18 different SH3 domains (S.M. Larson & A.R. Davidson, unpubl. obs.) and information derived from this analysis, and our sequence alignment analysis is used in the discussion below.

The position occupied by Asn23 in the Abp1p SH3 domain is found at the third position of a type II β-turn in almost every SH3 domain structure. Gly is highly favored at this position in type II β-turns because it must adopt an α_L backbone conformation, with a positive ϕ angle (Hutchinson & Thornton, 1994). This conformational requirement likely explains the stabilizing effect of the N23G substitution. Additionally, the large increase in folding rate displayed by this mutant (Table 1) is consistent with previous results indicating that position 23 lies in a region of the SH3 domain that is structured in the folding transition state (Grantcharova et al., 1998; Martinez et al., 1998). The replacement of various residues with glycine to reduce conformational strain and

Table 4. *Stabilizing mutants isolated in other sequence alignment-based studies*

Mutant	Protein[a]	$\Delta\Delta G_u$	Acc[b]	Conformation in WT structure
M233L	GroEL	1.8	0	α-helix
A223T	GroEL	1.6	0	β-strand, last residue preceding bend
A212E	GroEL	1.4	10	β-strand, at start following β-turn
A15L	IgL	1.4	57	Type II β-turn, second residue
N326T	GroEL	1.4	52	Type I β-turn, first residue
T294R	GroEL	1.3	23	α-helix, three residues before Gly at C-cap
N90Q	IgL	1.1	1	β-strand, last residue preceding turn
A223V	GroEL	1.0	0	β-strand, last residue preceding bend
N268D	p53	0.9	13	β-strand
I305L	GroEL	0.7	66	Type I β-turn, third residue
V203A	p53	0.5	0.5	Type I β-turn, fourth residue
E308S	GroEL	0.5	62	Loop, precedes short helix
K242Q	GroEL	0.5	74	α-helix, two residues before Gly at C-cap
F32Y	IgL	0.4	15	β-strand, at point of bend
L106I	IgL	0.4	21	β-strand
L201P	p53	0.4	84	Type I β-turn, second residue
E308K	GroEL	0.3	62	Loop, precedes short helix
V271L	GroEL	0.3	6	β-strand, two residues after Gly of C-cap
M133L	p53	0.3	0	β-strand
T63S	IgL	0.3	58	β-strand
K207N	GroEL	0.3	56	Loop, one residue before β-turn
Y236F	p53	0.3	0	β-strand, one residue before bend
T210K	GroEL	0.3	67	Type I β-turn, second residue

[a]The proteins referred to here are the GroEL minichaperone (Wang et al., 1999), an immunoglobulin light chain variable domain (Steipe et al., 1994), and the p53 DNA-binding domain (Nikolova et al., 1998).

[b]Solvent accessibility of side chains was calculated using the program ENVIRONMENTS (Bowie et al., 1991) using PDB files 1KID (GroEL), 1TSR (p53), and 2IMM (IgL).

thus increase stability has been observed in other proteins (Kimura et al., 1992; Kawamura et al., 1996; Martinez et al., 1998).

The stabilizing role of Leu at position 7 is difficult to discern from examination of SH3 domain structures, even though Leu appears at this position in more than half of SH3 domain sequences. Because the E7L mutant unfolds fivefold more slowly than WT (Table 1), this substitution is clearly acting primarily by stabilizing the folded state of the domain. However, the Leu side chain is 42% exposed on average in SH3 domain structures and does not appear to make conserved stabilizing interactions with surrounding residues. Position 7 invariably adopts helical ϕ/ψ angles and lies at the point where the extended chain of the first β-strand of the domain changes direction to begin the RT-Src loop. Leu may be the most favorable residue in this particular local backbone conformation.

The 21 position in SH3 domains is highly exposed (74% on average) and lies in the first position of the type II β-turn mentioned above. However, there is generally no distinct preference for Lys over Val at this turn position (Hutchinson & Thornton, 1994). Because the stabilization mediated by the V21K substitution is due solely to an increase in the folding rate of the domain (Table 1) and not to stabilization of the folded state, examination of folded SH3 domain structures might not be expected to shed light on the mechanism of stabilization by this mutant. Nevertheless, insight into the role of this position has been provided by examination of a library of frequently occurring sequence patterns, called "I-sites," that strongly correlate with local protein structural features (Bystroff & Baker, 1998). These I-site motifs define specific amino acid sequences that possess high intrinsic propensities to form particular backbone conformations independent of the tertiary structure of the protein in which they appear. Position 21 in all solved SH3 domain structures lies in a novel I-site motif called the "diverging type II β-turn" (Fig. 9A) (Yi et al., 1998). Strikingly, Lys is strongly favored at the diverging type II β-turn position corresponding to position 21 and Val is strongly disfavored. Thus, the V21K substitution increases the local propensity of this region to adopt the diverging type II β-turn structure. Polar residues have been speculated to be favored at certain positions in the diverging type II β-turn because they prevent the formation of potential alternative incorrect structures (i.e., negative design) (Yi et al., 1998). The increase in folding rate caused by the V21K substitution is consistent with this hypothesis in that a Val residue at position 21 could increase the likelihood of this region to fold into a nonnative structure. For productive folding to proceed, this nonnative structure would have to be unfolded then refolded into the correct structure, which would slow the folding process. The stabilizing effect of the N23G substitution is also consistent with the diverging type II β-turn because Gly is the most favored amino acid at the corresponding position in this I-site (Table 5).

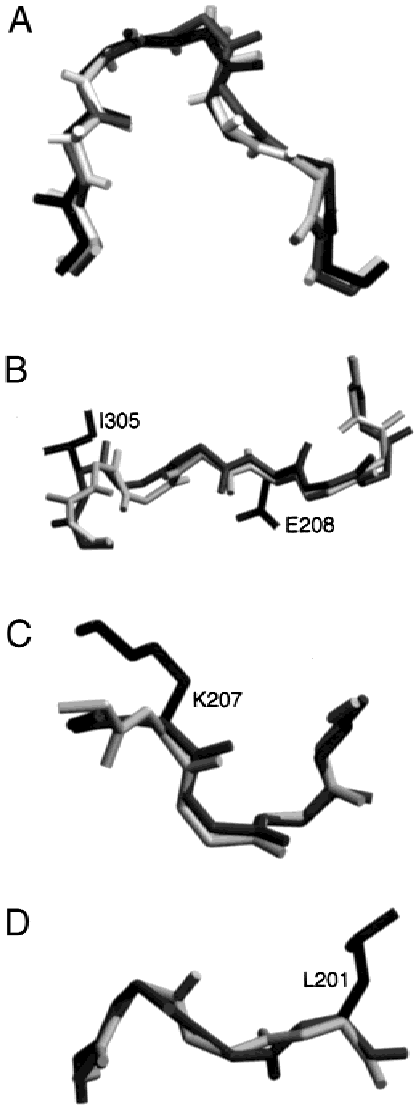

Fig. 9. A: The structures of the residues corresponding to Abp1p SH3 domain residues 20–26 (the Sem5 SH3 domain structure, 1SEM, is shown, black) and residues 13–19 of an immunoglobulin light-chain variable domain (2IMM, dark gray) are overlaid with a paradigm structure of the "diverging type II β-turn" I-site motif (light gray). **B:** The structure of residues 304–311 of the GroEL minichaperone (1KID, dark gray) is overlaid with a paradigm structure of the "α–α corner, Type I" I-site motif (light gray). **C:** The structure of residues 206–210 of the GroEL minichaperone (dark gray) is overlaid with a paradigm structure of the "DP-α N-Cap" I-site motif. **D:** The structure residues 198–201 of p53 DNA-binding domain (1TSR, dark gray) is overlaid with the paradigm structure a miscellaneous I-site motif (light gray). The paradigm structures for I-site motifs are found at the I-sites Library home page (http://isites.bio.rpi.edu/bystrc/Isites/index.html).

Analysis of stabilizing substitutions characterized in other sequence alignment based mutagenesis studies

Because several studies have now used sequence alignment based methods similar to that described here to successfully design stabilizing substitutions in various unrelated proteins (Steipe et al., 1994; Nikolova et al., 1998; Wang et al., 1999; Wirtz & Steipe, 1999), we wondered if there might be any features in common among these stabilized mutants. A summary of 23 different substitutions that were found to be stabilizing by more than 0.25 kcal/mol in three different studies are shown in Table 4 (in one of the studies cited above $\Delta\Delta G_u$ values could not be measured, so those mutants are not considered here). On examination of these data, there appears to be no clear trend to the types of substitutions observed except that almost all involve either polar to polar or nonpolar to nonpolar replacements. The solvent accessibility of the side chains of the replaced residues ranges from 0 to 84%, with 10 positions being more than 50% exposed. In addition, the locations of the substituted positions in terms of secondary structure are

Table 5. *Stabilizing substitutions located in I-sites*

Protein	Substitution	Change in I-site propensity of substitution[a]	I-Site designation		Alignment[b]	RMSD[c] (Å)
			Name	Number		
GroEL	I305L E308S E308K	Disfavored to favored Disfavored to favored Disfavored to neutral	α-α corner, Type I	10063	E<u>I</u>GM<u>E</u>LEK ELGMSPEQ	0.81
GroEL	K207N T210K	Disfavored to favored Favored to favored	DP-α N-Cap	5236	N<u>K</u>PE<u>T</u> WNPET/K	0.65
p53	L201P	Neutral to favored	Miscellaneous	4103	EGN<u>L</u> HGKP	0.53
IgL	A15L	Favored to disfavored	Diverging Type II β-turn	11007	VS<u>A</u>GERV VKAGERV	0.38
Abp1p	V21K N23G	Disfavored to favored Neutral to favored	Diverging Type II β-turn	11007	F<u>V</u>E<u>N</u>DKI FKEGDRI	0.42

[a]The residue preferences and other information pertaining to I-sites are taken from the I-sites Library home page (http://isites.bio.rpi.edu/bystrc/Isites/index.html). Favored residues are seen at their given positions at frequencies at least threefold greater than expected by chance and disfavored residues are seen at least threefold less frequently than expected by chance.

[b]Alignments of the protein sequence (top) and the most closely corresponding I-site consensus sequence (bottom) are shown. Underlined residues are those that were mutated to increase stability.

[c]RMSD is the root-mean-square deviation between the backbone atoms of the protein structure and the paradigm structure for the indicated I-site (see Fig. 9). The value shown for Abp1p is from comparison with the corresponding region of the Sem5 SH3 domain (1SEM).

variable with 11 substitutions in β-strand regions, nine in turns or loops, and three in helices. When the frequency of occurrence of stabilizing mutations within various secondary structure elements was compared with the fraction of β-strand, turn, loop, and helix present in each protein, no statistically significant trend was observed. Stabilizing substitutions did not seem to preferentially occur in any particular type of secondary structure. Examination of the three-dimensional structures of the WT proteins in these studies was not found to provide a great deal of insight into the mechanisms by which these substitutions cause increased stability (Steipe et al., 1994; Nikolova et al., 1998; Wang et al., 1999), Indeed, in the IgL study of Steipe et al. (1994), the mechanisms of stabilization remained obscure even after the structures of seven of the stabilized mutants were solved.

One mechanism that could explain the effects of some stabilizing substitutions is the improvement of local structural propensity. As described above, we found that two out of the three stabilizing substitutions in the Abp1p SH3 domain lie in the "diverging type II β-turn" I-site motif, and that both substitutions are predicted to greatly improve the local propensity for this structure. Encouraged by this finding, we investigated whether the stabilizing effects of any of the substitutions listed in Table 4 could be explained by improvement of local structural propensity, as assessed by the detection of I-sites. As shown in Table 5, seven stabilizing substitutions from the list in Table 4 clearly lie within defined I-site conformations, and five of these seven are predicted to be stabilizing by the consensus sequence for the corresponding I-site. For example, the Ile305 and Glu308 residues of GroEL lie in a conformation known as the "α–α corner, Type I." The paradigm structure for this motif overlays the relevant part of the GroEL structure with a root-mean-square deviation (RMSD) of only 0.81 Å (Fig. 9B). Examination of the sequence profile for this motif (Bystroff & Baker, 1998) shows that at the position corresponding to Ile305 in GroEL, Ile is strongly disfavored, and Leu is strongly favored; thus, the large increase in stability resulting from the I305L substitution is expected. Similarly, Glu at the position corresponding to Glu308 in GroEL is strongly disfavored in this local conformation, while Ser is strongly favored, and Lys is neither favored nor disfavored. Again, the effects of stabilizing substitutions are rationalized. In a similar manner, the location of mutated positions within I-sites can explain the stabilizing effects of the K207N substitution in GroEL and the L201P substitution in p53 (Table 5; Fig. 9C,D). In the case of the K207N substitution, the validity of the I-site match is supported by two destabilizing substitutions that were also characterized in this region (N206T and P208S), which both change relatively favorable residues in the I-site (P at 208 is actually the most commonly seen residue) to strongly disfavored ones. The stabilizing T210K substitution replaces one favorable residue with another.

The A15L substitution of IgL provides a single counterexample to the correlation of stabilization with improvement of I-site propensity. The structure of the region in which the Ala 15 position lies matches that of the diverging type II β-turn motif very closely (Table 5; Fig. 9A), but the A15L substitution actually changes this position from a relatively favorable residue to a strongly disfavored one. Interestingly, a stabilizing substitution at a different position in the homologous region of an immunoglobulin heavy chain variable domain has been isolated (Wirtz & Steipe, 1999), and this change also substitutes a favorable residue for the I-site conformation with an unfavorable one. The substituted residues within the diverging type II β-turn motif in the immunoglobulin variable domains are probably forming stabilizing nonlocal interactions.

As seen in Table 4, 13 of the 23 stabilizing substitutions lie at positions that are less than 25% solvent exposed in their respective structures. The observed increases in stability in these cases are probably caused by improvements in the interactions of these buried side chains with other side chains lying nearby in the tertiary

structure of the protein, many of which may not be close in the primary sequence. At these positions, the propensity of the local sequence to form a particular structure (as would be reflected in the I-site consensus sequences) may have less influence on the stability of the protein than do the tertiary interactions resulting from residue burial. Consistent with this notion, the five stabilizing substitutions that can be rationalized by their location within I-sites are all observed to lie in highly exposed positions within turn or loop conformations. Indeed, when only the exposed sites in Table 4 are considered, 50% of the stabilizing substitutions can be rationalized by their occurrence within I-sites. These data, although still preliminary due to the small number of mutations examined, suggest that detection of I-sites and mutation to residues matching I-site consensus sequences could provide a useful means to select stabilizing substitutions, especially for exposed residues in loop or turn regions of proteins.

Biological implications of the stabilizing Abp1p SH3 domain substitutions

The dramatic stabilization of the Abp1p SH3 domain observed when only three residues are substituted raises the possibility that the relatively low stability of this SH3 domain may actually be required for its in vivo function. One mechanism by which altered stability could affect in vivo function is by changing the rate of proteolysis. Even though both the WT and stabilized Abp1p SH3 domains have T_m values far above the normal growth temperature of yeast, it can be calculated from the ΔG_u values that ~1% of the WT Abp1p SH3 domain is unfolded at 30 °C at any given time, while only 0.003% of the stabilized triple mutant would be unfolded under the same conditions. In addition, large changes in the folding kinetics (Table 1) of the stabilized mutant means that it would stay unfolded for a much shorter time than the WT domain. Because proteases digest proteins at a much higher rate when they are in an unfolded state, this change in stability could have a significant effect on the in vivo degradation rate of this domain. The importance of proteolysis in regulating crucial cellular processes, such as the cell cycle, signal transduction, and gene regulation has been well documented (Hakak & Martin, 1999; Koepp et al., 1999; Salvat et al., 1999).

An alternate explanation for the presence of destabilizing residues in the Abp1p SH3 domain is that these residues are directly involved in a functionally important interaction. Although all the stabilized mutants bind to the proline-rich binding site of Srv2p with WT affinity (Table 1), the Glu7, Val21, and Asn23 residues could be crucial for some other biologically important interaction. The clustering of these three residues and three of the other atypical Abp1p residues on one surface of the SH3 domain structure, which is distinct from the PXXP-peptide binding surface (Fig. 10), supports this possibility. Protein–protein interactions mediated by SH3 domain surfaces other than the PXXP-peptide binding surface have been previously reported (Gorina & Pavletich, 1996; Vidal et al., 1998; Sondhi & Cole, 1999). Studies on the in vivo functional properties of the mutants described in this work will elucidate the role of these residues and may uncover new mechanisms of SH3 domain interactions.

What can be learned from sequence alignment based mutagenesis studies?

This study and others using similar procedures clearly demonstrate that sequence alignment methods provide a useful means to design

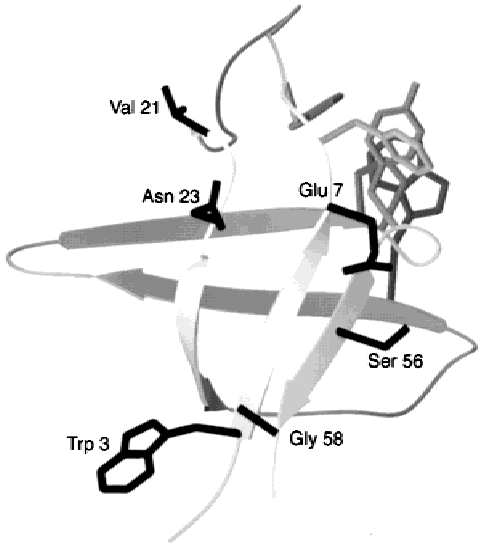

Fig. 10. Position of the unusual Abp1p residues on the SH3 domain. Ribbon drawing of the Sem-5 SH3 domain with the WT Abp1p SH3 domain amino acid residues superimposed on the structure at their respective positions. The side chains of the Abp1p SH3 domain residues Trp3, Glu7, Val21, Asn23, and Ser56 are shown. Because Gly58 lacks a side chain, one of its alpha hydrogens has been shown to indicate its position. The side chains of residues involved in polyproline peptide binding are also shown.

stabilizing amino acid substitutions. An advantage of these methods is that success does not require knowledge of the three-dimensional structure of the protein under study, or understanding of the mechanism by which stabilization is achieved. In addition, the resulting mutants generally maintain WT activity. On the other hand, the sequence alignment based approach may ultimately teach us little about the determinants of protein stability because the mechanisms by which the identified substitutions cause increased stability are so difficult to elucidate. In this study, by using the I-sites library, we have been able to find a rationalization for some of the stabilizing mutants isolated in this and other studies. Our findings suggest that the sequence alignment based approach can provide a means to identify regions of proteins where local structural propensity plays an important role in determining stability. In this regard, we predict that because detailed specific tertiary interactions (e.g., between highly buried side chains) are less likely to be conserved among homologues with relatively low sequence identity, studies utilizing alignments containing diverse sequences will be more likely to identify substitutions affecting local interactions. It is interesting to note that all but one of the stabilizing mutations occurring in I-sites identified here occurred in the SH3 domain or GroEL, and that relatively diverse sequence alignments (compared to those used in the other studies) were used to predict the stabilizing mutations in these proteins.

Continued use of simple sequence alignment based methodologies to successfully identify stabilizing amino acid substitutions in proteins could ultimately provide us with a large catalog of single amino acid substitutions that can stabilize proteins. Study of the nature of these substitutions could lead to the identification of a variety of general stabilization mechanisms and may provide further insights into the factors that determine the propensities of specific amino acid sequences for certain local structural environments.

Materials and methods

Plasmid construction and mutagenesis

Plasmid pDD3 (Lila & Drubin, 1997), containing DNA encoding Abp1p was obtained from the laboratory of Dr. Brenda Andrews. The Abp1p SH3 domain (residues 535–592) was amplified by polymerase chain reactions (PCR) using the oligonucleotides 5′GG TAGCCATGGCTCCTTGGGCCACAGCAG 3′ and 5′CGGACC TCGAGGTTGCCCAAAGACACATATTAG3′ and cloned into the pET21d(+) (Novagen, Madison, Wisconsin) expression vector between the *NcoI* and *XhoI* sites. The DNA encoding the Abp1p SH3 domain was cloned such that two residues (MA) were added to the N-terminus of the expressed protein. Eight residues (LEHHH HHH) were added to the C-terminus encoding the hexahistidine tag used for protein purification. Mutations were introduced into the SH3 gene by two-step site-directed PCR mutagenesis (Kuipers et al., 1991) using Vent® DNA polymerase (New England Biolabs, Beverly, Massachusetts). DNA sequences of the wild-type gene and all mutants were verified using the Sequenase® 2.0 kit (U.S. Biochemical, Cleveland, Ohio).

Protein expression and purification

All protein expression was performed in the *E. coli* strain KM1502, a Δ*slyD* variant of *E. coli* GJ1158, with the genotype *ompT hsdS gal dcm* Δ*malAp510 malP*::(*proUp*-T7 RNAP) *malQ*::*LacZ*hyb11 Δ(*zhf-900*::Tn*10*dTet). This strain contains the gene for T7 RNA polymerase under the control of the NaCl-inducible promoter *proU* (Bhandari & Gowrishankar, 1997). The *slyD* deletion prevents the expression of a 21 kD histidine-rich *E. coli* protein that binds to the Ni-affinity resin used for protein purification. After induction with NaCl and 3 h of growth at 37 °C, cells were harvested and lysed in 6 M GuHCl, 100 mM NaH_2PO_4, 10 mM Tris-HCl, 10 mM imidazole, pH 8.0. Purification was performed by a batch procedure in the same buffer using Ni-NTA agarose resin (Qiagen, Hilden, Germany). A low-pH buffer (6 M GuHCl, 0.2 M acetic acid) was used to elute purified proteins, which were refolded by dialysis into 10 mM Tris-HCl, pH 8.0, 0.2 mM EDTA, 250 mM KCl. All subsequent studies were performed in this buffer. Protein purity was verified by SDS-PAGE electrophoresis of samples followed by Coomassie staining. Protein concentrations were determined by ultraviolet absorbance at 280 nm using a molar absorption coefficient of 20,835 $M^{-1}\,cm^{-1}$ determined by the method of Pace (Pace et al., 1995).

Determination of thermodynamic parameters by CD and fluorescence spectroscopy

CD measurements were performed in an Aviv 62A DS circular dichroism spectrometer. Thermal denaturation was monitored by following the change in ellipticity at 220 nm and chemical denaturation was monitored at 230 nm. Proteins were heated from 25–109 °C in 2 °C increments, with a 1 min equilibration time and a 15 s averaging time for the CD measurement. More stable proteins ($T_m > 80$ °C) were monitored to a final temperature of 120 °C. Temperature melts were performed in a 0.1 cm pathlength cuvette on samples with protein concentrations ranging from 30 to 50 μM. Unfolding was not affected by protein concentration and was completely reversible in all cases. GuHCl denaturation experiments were performed at 25 °C using a Microlab 500 series automated titrator by stepwise mixing of a 20 μM protein solution with 0 M GuHCl into a 20 μM protein solution with high GuHCl (5 to 6.5 M). At each step, the sample was stirred in a 1.0 cm pathlength cuvette for 1 min, then a CD measurement averaged for 30 s was recorded.

Fluorescence measurements were performed in a Spex 0.22m Spectrometer using DM 3000 software. Chemical denaturation was monitored by fluorescence emission at 340 nm after excitation at 295 nm with a 2 nm slit width. Both GuHCl and urea denaturation experiments were performed manually at 25 °C by stepwise mixing of 1 μM protein solution in buffer into a 1 μM protein solution with high urea (8 M) or high GuHCl (6.5 M) as described above. Thermal and chemical denaturation data were fit as described previously (Maxwell & Davidson, 1998). Due to the large slope of the folded baselines of GuHCl melts monitored by fluorescence, they were fit with a fixed *m* value of 1.62. This value was derived by averaging the *m* values obtained for wild-type and all point mutants from the GuHCl melts monitored by CD. The ΔG_u of the WT domain was also confirmed by urea-induced unfolding, which resulted in a more extensive folded baseline and, for the same reason, the V21K mutant was only tested by urea melts.

Stopped-flow kinetics

Kinetic measurements were performed at room temperature in a BioLogic SFM-4 stopped-flow spectrometer. The pathlength of the observation chamber was 2 mm, with a dead time of mixing of 9 ms. Folding kinetics were followed by the change in fluorescence greater than 309 nm after excitation at 295 nm, with a 2 nm slitwidth. Unfolding of 20 μM Abp1p SH3 domain solutions in buffer was initiated by 10-fold dilution with GuHCl solutions of varying concentrations to give final GuHCl concentrations between 4 and 8 M. To initiate refolding, 20 μM unfolded SH3 domain in high GuHCl (3 to 6 M, depending on the SH3 domain stability) was diluted 10-fold with buffer or GuHCl solutions of varying concentrations to give final concentrations of 0.2–4 M GuHCl. Kinetics were measured at least three times under identical conditions, averaged, and analyzed as monoexponential functions using BioKine software provided by BioLogic. Unfolding and refolding rate constants were fit by nonlinear least-squares regression (using the program Kaleidagraph) to Equation 1:

$$\ln(k_{obs}) = \ln(k_f) - m_f[\mathrm{GuHCl}] + \ln(k_u) + m_u[\mathrm{GuHCl}] \qquad (1)$$

where k_{obs} is the observed folding or unfolding rate constant at a given GuHCl concentration, k_f and k_u are the folding and unfolding rate constants at zero denaturant, and m_f and m_u are the dependence of the natural logarithms of the folding and unfolding rate constants, respectively, on the concentration of denaturant.

Peptide binding assays

Peptide binding titrations were performed on pre-mixed solutions containing 1 μM SH3 domain and 0–30 μM target protein. The degree of binding of the target protein to the SH3 domain was monitored by tryptophan fluorescence emission at 330 nm after excitation at 295 nm with a 2 nm slitwidth. Measurements were made in a 0.5 cm cuvette in a Spex 0.22m Spectrometer using DM 3000 software. The target peptide sequence used corresponds to amino acids 337 to 376 of the protein Srv2p, an in vivo target of

the Abp1p SH3 domain (Lila & Drubin, 1997). This proline-rich sequence contains the binding site for the Abp1p SH3 domain. For ease of purification, the Srv2p sequence was expressed in *E. coli* with the N-terminal domain of λ repressor (residues 1–99) fused to its N-terminus and a FLAG epitope and a hexahistidine tag at its C-terminus, as previously described (Maxwell & Davidson, 1998). The target peptide was not cleaved from the fusion protein to perform binding assays, as the region of λ repressor used contains no tryptophan residues. Binding data were fit as described previously (Maxwell & Davidson, 1998).

Measurements of native molecular weight

Analytical ultracentrifugation was performed at 25°C at three protein concentrations of the WT (20, 50, 100 μM) and E7L/V21K/N23G triple mutant (25, 50, 140 μM) SH3 domains. Overnight spins were performed at 20, 24, 28, 32, 36, and 40,000 rpm. Equilibrium distribution of protein was determined by absorption at 280 nm. To determine the molecular weight of the protein, all 18 resulting data sets for each protein were fit were fit by nonlinear least-squares regression (using the program Kaleidagraph) to Equation 2:

$$A = A(0)\exp[\omega^2 M_r(1 - \nu\rho)(r^2 - r(0)^2)/2RT] \qquad (2)$$

where A is the absorbance at 280 nm at radius r, $A(0)$ is the absorbance at 280 nm at reference radius $r(0)$, R is the gas constant, T is the temperature, ρ is the buffer density, ν is the partial specific volume of the protein calculated from the amino acid composition, ω is the angular velocity, and M_r is the molecular weight of the protein. The values reported were averaged from 18 experiments for each protein.

Gel filtration chromatography was performed at room temperature by FPLC on a 25 mL Superdex-75 column (Pharmacia, Uppsala, Sweden) calibrated with BSA (67 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), myoglobin (17 kDa), lysozyme (14 kDa), and insulin (3.5 kDa). The column was washed with 2 volumes of buffer at a 1 mL/min flow rate. 2.5 mM (20 mg/mL) samples of WT and triple mutant Abp1p SH3 domain were loaded on the column and eluted at 1 mL/min, using the absorbance at 280 nm to detect protein.

Identification of I-sites

I-sites were identified using the I-sites server (http://honduras.bio.rpi.edu/~isites/ISLserver.html). Protein Data Bank (PDB) files for each structure of interest were submitted to the server and a list of I-sites that were probable from the sequence of the protein and were present in the structure of the protein was returned. I-sites corresponding to sites of stabilizing substitutions were identified through manual examination of this list.

Acknowledgments

We thank Karen Colwill and Brenda Andrews for providing a plasmid containing the gene encoding WT Abp1p, Reinhart Reithmeier for use of the spectrofluorometer, Avi Chakrabartty for use of the analytical ultracentrifuge, Julie Forman Kay for use of the stopped-flow spectrofluorometer, and Karen Maxwell for the gift of *E. coli* strain KM1502 and for critical reading of the manuscript. We also thank Chris Bystroff for a large amount of help in identifying I-sites. This work was supported by an operating grant from the Medical Research Council of Canada and a scholarship to AR from the National Sciences and Engineering Research Council of Canada.

References

Alexander P, Fahnestock S, Lee T, Orban J, Bryan P. 1992. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: Why small proteins tend to have high denaturation temperatures. *Biochemistry 31*:3597–3603.

Bhandari P, Gowrishankar J. 1997. An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. *J Bacteriol 179*:4403–4406.

Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*:164–170.

Bystroff C, Baker D. 1998. Prediction of local structure in proteins using a library of sequence-structure motifs. *J Mol Biol 281*:565–577.

Chen YJ, Lin SC, Tzeng SR, Patel HV, Lyu PC, Cheng JW. 1996. Stability and folding of the SH3 domain of Bruton's tyrosine kinase. *Proteins 26*:465–471.

Dalgarno DC, Botfield MC, Rickles RJ. 1997. SH3 domains and drug design: Ligands, structure, and biological function. *Biopolymers 43*:383–400.

Drubin DG, Mulholland J, Zhu ZM, Botstein D. 1990. Homology of a yeast actin-binding protein to signal transduction proteins and myosin-I. *Nature 343*:288–290.

Evans SV. 1993. SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules. *J Mol Graph 11*:134–138, 127–128.

Filimonov VV, Azuaga AI, Viguera AR, Serrano L, Mateo PL. 1999. A thermodynamic analysis of a family of small globular proteins: SH3 domains. *Biophys Chem 77*:195–208.

Freeman NL, Chen Z, Horenstein J, Weber A, Field J. 1995. An actin monomer binding activity localizes to the carboxyl-terminal half of the *Saccharomyces cerevisiae* cyclase-associated protein. *J Biol Chem 270*:5680–5685.

Gorina S, Pavletich NP. 1996. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science 274*:1001–1005.

Grantcharova VP, Baker D. 1997. Folding dynamics of the src SH3 domain. *Biochemistry 36*:15685–15692.

Grantcharova VP, 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.

Hakak Y, Martin GS. 1999. Ubiquitin-dependent degradation of active Src. *Curr Biol 9*:1039–1042.

Hecht MH, Sturtevant JM, Sauer RT. 1986. Stabilization of lambda repressor against thermal denaturation by site-directed Gly–Ala changes in alpha-helix 3. *Proteins 1*:43–46.

Hendsch ZS, Jonsson T, Sauer RT, Tidor B. 1996. Protein stabilization by removal of unsatisfied polar groups: Computational approaches and experimental tests. *Biochemistry 35*:7621–7625.

Hutchinson EG, Thornton JM. 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci 3*:2207–2216.

Inoue I, Rechsteiner M. 1994. On the relationship between the metabolic and thermodynamic stabilities of T4 lysozymes. Measurements in *Escherichia coli*. *J Biol Chem 269*:29241–29246.

Jackson SE, Fersht AR. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry 30*:10428–10435.

Kawamura S, Abe Y, Ueda T, Masumoto K, Imoto T, Yamasaki N, Kimura M. 1998. Investigation of the structural basis for thermostability of DNA-binding protein HU from *Bacillus stearothermophilus*. *J Biol Chem 273*:19982–19987.

Kawamura S, Kakuta Y, Tanaka I, Hikichi K, Kuhara S, Yamasaki N, Kimura M. 1996. Glycine-15 in the bend between two alpha-helices can explain the thermostability of DNA binding protein HU from *Bacillus stearothermophilus*. *Biochemistry 35*:1195–1200.

Kimura S, Kanaya S, Nakamura H. 1992. Thermostabilization of *Escherichia coli* ribonuclease HI by replacing left-handed helical Lys95 with Gly or Asn. *J Biol Chem 267*:22014–22017.

Kirino H, Aoki M, Aoshima M, Hayashi Y, Ohba M, Yamagishi A, Wakagi T, Oshima T. 1994. Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus*. *Eur J Biochem 220*:275–281.

Knapp S, Mattson PT, Christova P, Berndt KD, Karshikoff A, Vihinen M, Smith CI, Ladenstein R. 1998. Thermal unfolding of small proteins with SH3 domain folding pattern. *Proteins 31*:309–319.

Koepp DM, Harper JW, Elledge SJ. 1999. How the cyclin became a cyclin: Regulated proteolysis in the cell cycle. *Cell 97*:431–434.

Kohda D, Hatanaka H, Odaka M, Mandiyan V, Ullrich A, Schlessinger J, Inagaki F. 1993. Solution structure of the SH3 domain of phospholipase C-gamma. *Cell 72*:953–960.

Kuipers OP, Boot HJ, de Vos WM. 1991. Improved site-directed mutagenesis method using PCR. *Nucleic Acids Res 19*:4558.

Lila T, Drubin DG. 1997. Evidence for physical and functional interactions among two *Saccharomyces cerevisiae* SH3 domain proteins, an adenylyl cyclase-associated protein and the actin cytoskeleton. *Mol Biol Cell 8*:367–385.

Lim WA, Fox RO, Richards FM. 1994a. Stability and peptide binding affinity of an SH3 domain from the *Caenorhabditis elegans* signaling protein Sem-5. *Protein Sci 3*:1261–1266.

Lim WA, Richards FM, Fox RO. 1994b. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature 372*:375–379.

Martinez JC, Pisabarro MT, Serrano L. 1998. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol 5*:721–729.

Matsumura M, Signor G, Matthews BW. 1989. Substantial increase of protein stability by multiple disulphide bonds. *Nature 342*:291–293.

Matthews BW, Nicholson H, Becktel WJ. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci USA 84*:6663–6667.

Maxwell KL, Davidson AR. 1998. Mutagenesis of a buried polar interaction in an SH3 domain: Sequence conservation provides the best prediction of stability effects. *Biochemistry 37*:16172–16182.

Nikolova PV, Henckel J, Lane DP, Fersht AR. 1998. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc Natl Acad Sci USA 95*:14675–14680.

Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci 4*:2411–2423.

Parsell DA, Sauer RT. 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *J Biol Chem 264*:7590–7595.

Pawson T. 1995. Protein modules and signalling networks. *Nature 373*:573–580.

Plaxco KW, Guijarro JI, Morton CJ, Pitkeathly M, Campbell ID, Dobson CM. 1998. The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry 37*:2529–2537.

Politou AS, Millevoi S, Gautel M, Kolmerer B, Pastore A. 1998. SH3 in muscles: Solution structure of the SH3 domain from nebulin. *J Mol Biol 276*:189–202.

Querol E, Perez-Pons JA, Mozo-Villarias A. 1996. Analysis of protein conformational characteristics related to thermostability. *Protein Eng 9*:265–271.

Russell RJ, Taylor GL. 1995. Engineering thermostability: Lessons from thermophilic proteins. *Curr Opin Biotechnol 6*:370–374.

Salvat C, Aquaviva C, Jariel-Encontre I, Ferrara P, Pariat M, Steff AM, Carillo S, Piechaczyk M. 1999. Are there multiple proteolytic pathways contributing to c-Fos, c-Jun and p53 protein degradation in vivo? *Mol Biol Rep 26*:45–51.

Sauer RT, Hehir K, Stearman RS, Weiss MA, Jeitler-Nilsson A, Suchanek EG, Pabo CO. 1986. An engineered intersubunit disulfide enhances the stability and DNA binding of the N-terminal domain of lambda repressor. *Biochemistry 25*:5992–5998.

Serrano L, Day AG, Fersht AR. 1993. Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J Mol Biol 233*:305–312.

Sondhi D, Cole PA. 1999. Domain interactions in protein tyrosine kinase Csk. *Biochemistry 38*:11147–11155.

Steipe B, Schiller B, Pluckthun A, Steinbacher S. 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol 240*:188–192.

Takagi H, Takahashi T, Momose H, Inouye M, Maeda Y, Matsuzawa H, Ohta T. 1990. Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease. *J Biol Chem 265*:6874–6878.

Vidal M, Montiel JL, Cussac D, Cornille F, Duchesne M, Parker F, Tocque B, Roques BP, Garbay C. 1998. Differential interactions of the growth factor receptor-bound protein 2 N-SH3 domain with son of sevenless and dynamin. Potential role in the Ras-dependent signaling pathway. *J Biol Chem 273*:5343–5358.

Vieille C, Zeikus JG. 1996. Thermozymes: Identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol 14*: 183–191.

Wang Q, Buckle AM, Foster NW, Johnson CM, Fersht AR. 1999. Design of highly stable functional GroEL minichaperones. *Protein Sci 8*:2186–2193.

Wirtz P, Steipe B. 1999. Intrabody construction and expression III: Engineering hyperstable V(H) domains. *Protein Sci 8*:2245–2250.

Yi Q, Bystroff C, Rajagopal P, Klevit RE, Baker D. 1998. Prediction and structural characterization of an independently folding substructure in the src SH3 domain. *J Mol Biol 283*:293–300.