A "loop entropy reduction" phage-display selection for folded amino acid sequences

PHILIPPE MINARD,[1,3,4] MICHELLE SCALLEY-KIM,[2,3] ALEX WATTERS,[2,3] AND DAVID BAKER[1]

[1]Department of Biochemistry, Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

[2]Molecular and Cellular Biology Program, University of Washington, Seattle, Washington 98195, USA

(RECEIVED August 1, 2000; ACCEPTED October 27, 2000)

Abstract

As a step toward selecting folded proteins from libraries of randomized sequences, we have designed a 'loop entropy reduction'-based phage-display method. The basic premise is that insertion of a long disordered sequence into a loop of a host protein will substantially destabilize the host because of the entropic cost of closing a loop in a disordered chain. If the inserted sequence spontaneously folds into a stable structure with the N and C termini close in space, however, this entropic cost is diminished. The host protein function can, therefore, be used to select folded inserted sequences without relying on specific properties of the inserted sequence. This principle is tested using the IgG binding domain of protein L and the lck SH2 domain as host proteins. The results indicate that the loop entropy reduction screen is capable of discriminating folded from unfolded sequences when the proper host protein and insertion point are chosen.

Keywords: Phage-display; insertion; protein folding; protein evolution; chimeric proteins

What fraction of the vast number of possible polypeptide sequences are able to form a defined three-dimensional structure analogous to the folded states of natural proteins? An experimental answer to this fundamental question could, in principle, be obtained by an examination of the properties of a large number of randomly generated polypeptide sequences. We have developed a 'loop entropy reduction' screen to be used for selecting folded proteins from large collections of natural or artificial coding sequences.

Recent studies have shown that increasing the length of loop regions in protein structures typically results in a decrease in overall stability of the protein. The observed decrease in stability has been linked to the entropic cost of ordering the additional residues in the loop (Ladurner 1997; Nagi 1997). These data suggest that if the unfolded inserted sequence is of sufficient length, it may disrupt the folding of the host protein into which it is inserted. However, if the inserted sequence is capable of folding into an independent structure, the entropic cost to the host protein will be minimal and the host protein may retain its ability to fold (Betton et al. 1997; Collinet 2000). The functional integrity of the host protein is directly related to the conformational state of the inserted sequence, allowing the properties of the host protein to be used to select for folded inserted sequences.

Several questions must be addressed in the design of the loop entropy reduction screen: What protein should serve as the host protein? Into which loop should the sequences be inserted? How will the folded state of the host protein be evaluated? In this study, we have chosen to use the IgG-binding domain of protein L and the lck SH2 domain as host proteins. These proteins were selected because of the large amount of structural information available: High-resolution structures and thermodynamic stabilities have been deter-

Reprint requests to: David Baker, Department of Biochemistry, University of Washington, J 567 Health Sciences Building, Box 37-7350, Seattle, Washington 98195, USA; e-mail: dabaker@u.washington.edu; fax: 206-685-1792.

[3]These authors contributed equally to this work.

[4]Present address: Laboratoire de Modelisation et Ingenierie des Proteines, CNRS et Universite de Paris sud, 91405 CEDEX ORSAY, France.

Article and publication are at www.proteinscience.org/cgi/doi/10.1110/ps.32401

mined for both proteins (Wikstrom et al. 1993; Tong et al. 1996; Scalley et al. 1997). A phage-display format was chosen as a screening method because it has proven to be successful in selecting rare folded variants within a collection of highly randomized domains (Zhou et al. 1996; Riddle et al. 1997; Kim et al. 1998). Furthermore, the resistance of filamentous phage is compatible with the presence of destabilizing conditions such as high temperature or denaturants during the selection procedure, providing a method for adjusting the level of selection pressure (Kristensen 1998; Forrer et al. 1999; Jung et al. 1999).

Results

Initially, the IgG-binding domain of peptostreptococcal protein L was selected as a host protein. Two turns were chosen as points for insertion (Figure 1): the turn leading from the second β-strand into the α-helix, β_2–α, and the turn leading from the third β-strand into the fourth β-strand, β_3–β_4. To determine whether either or both insertion points could withstand the insertion of a folded sequence, preliminary experiments were performed using wild-type src SH3 as a model of a folded inserted sequence. The SH3 domain coding region was introduced into the protein L DNA sequence via two unique restriction sites (*Eco*RI/*Nde*I for β_2–α, and *Sal*I/*Kpn*I for β_3–β_4) in the protein L-gene VIII fusion construct (Gu et al. 1995). The amino acid sequences of the regions bordering the SH3 insertion are shown in Table 1. It is important to note that N and C termini of SH3 domains are close in space to one another, a property that is likely to be preferred by the screening method.

Phage displaying the chimeric proteins were made and screened for their ability to bind paramagnetic beads coated with IgG, which protein L binds to with high affinity (Kihlberg et al. 1996). The results show clearly that the turns cannot tolerate the insertion of SH3 equally (Table 2). Phage displaying SH3 inserted into the β_2–α turn were recovered at very low levels, whereas phage displaying SH3 inserted into the β_3–β_4 turn of protein L were recovered at levels similar to that of phage displaying wild-type protein L. Recently it was found that the residues packed at the β_2–α interface make important contributions to IgG binding, thus, the low recovery levels may reflect a disruption in IgG binding resulting from the proximity of the SH3 insertion (H. Svensson, pers. comm.). Because the β_3–β_4 turn tolerates insertion of a folded protein, this turn was used for all subsequent insertions into protein L.

A library of random sequences coding for 60–300 amino acids was constructed and inserted into the β_3–β_4 turn of protein L (see Materials and Methods). The library members that retained protein L function were isolated using IgG-coated beads. Several positives were identified by the screen, but initial examination of their sequences showed

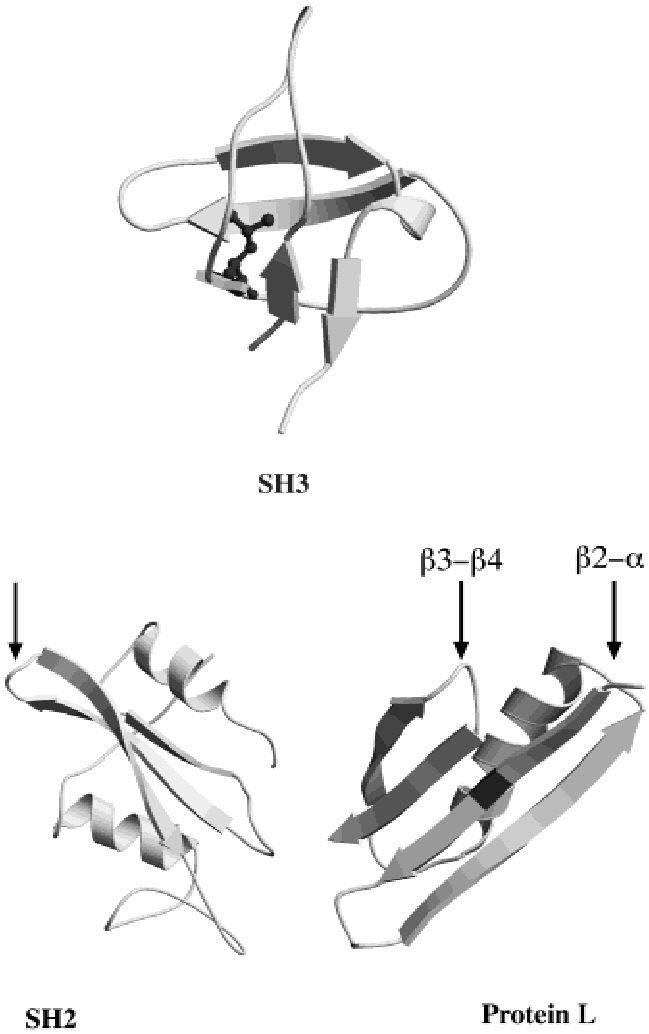

Fig. 1. Structure of the host and inserted domains. The surface loops in protein L and SH2 that are used for insertion are indicated by the arrows. The position of the A8G mutation in protein L is highlighted on the structure. Leu32 is mutated to a glutamate in the destabilized form of the SH3 domain. Images were made using Raster 3D (Bacon and Anderson 1988; Merrit and Murphy 1994) and Molscript (Kraulis 1991).

them to be highly polar and, thus, unlikely to fold into stable independent structures. One sequence identified by the screen, B11, was chosen for further analysis (see the legend to Table 1 for B11 sequence). Phage displaying the pL[B11] chimera were panned against IgG-coated beads to ensure the chimera exhibited binding activity (Table 2). The pL[B11] chimera was recovered at levels similar to that of wild-type protein L and the pL[SH3] chimera.

To examine the effect of folded and unfolded inserted sequences on the stability of protein L, the SH3 and B11 chimeric proteins, pL[SH3] and pL[B11], were overexpressed and purified. The stabilities of pL[SH3] and pL[B11] were measured using guanidine denaturation (Fig. 2; see Materials and Methods). Both pL[SH3] and pL[B11]

Table 1. *Phagemid constructs*

Host	Insert	Sequence
$pL_{\beta 2-\alpha}$	—	$T_{19}A_{20}[E_{21}F_{22}K_{23}G_{24}T_{25}F_{26}E_{27}K_{28}A_{29}T_{30}S_{31}E_{32}A_{33}Y_{34}A_{35}]Y_{36}A_{37}$
$pL_{\beta 2-\alpha}$	$SH3_{wt}$	$T_{19}A_{20}[E_{21}F_{22}K_{23}G_{24}T_{25}F_{26}$***G*****T$_9F_{10}A_{11}L_{12}$…(SH3)…A$_{62}P_{63}S_{64}D_{65}$***G***$E_{27}K_{28}A_{29}T_{30}S_{31}E_{32}A_{33}Y_{34}A_{35}]Y_{36}A_{37}$
$pL_{\beta 3-\beta 4}$	—	$T_{48}V_{49}[D_{50}V_{51}A_{52}D_{53}K_{54}G_{35}]T^*Y_{56}T_{57}L_{58}$
$pL_{\beta 3-\beta 4}$	$SH3_{wt}$	$T_{48}V_{49}[E^*_{50}V_{51}A_{52}D_{53}K_{54}$***SG*****T$_9F_{10}A_{11}L_{12}$…(SH3)…A$_{62}P_{63}S_{64}D_{65}$**$G_{55}]T^*Y_{56}T_{57}L_{58}$
$pL_{\beta 3-\beta 4}$	B11	$T_{48}V_{49}[E^*_{50}V_{51}A_{52}D_{53}K_{54}$***SG*****R$_1S_2P_3A_4$…(B11)…I$_{58}D_{59}T_{60}G_{61}S_{62}$**$G_{55}]T^*Y_{56}T_{57}L_{58}$
SH2	—	$R_{168}D_{169}V^*_{170}[D_{171}Q_{172}N_{173}Q_{174}G_{175}]T^*_{176}V_{177}V_{178}$
SH2	$SH3_{wt}$	$R_{168}D_{169}V^*_{170}[$***EVADKSG*****T$_9$F$_{10}$A$_{11}$L$_{12}$…(SH3)…A$_{62}$P$_{63}$S$_{64}$D$_{65}$**$G_{175}]T^*_{176}V_{177}V_{178}$
SH2	B11	$R_{168}D_{169}V^*_{170}[$***EVADKSG*****R$_1$S$_2$P$_3$A$_4$…(B11)…I$_{58}$D$_{59}$T$_{60}$G$_{61}$S$_{62}$**$G_{175}]T^*_{176}V_{177}V_{178}$
SH2	$SH3_{L32E}$	$R_{168}D_{169}V^*_{170}[$***EVADKSG*****T$_9$F$_{10}$A$_{11}$L$_{12}$…(SH3)…E*$_{32}$…A$_{62}$P$_{63}$S$_{64}$D$_{65}$**$G_{175}]T^*_{176}V_{177}V_{178}$

The sequence of the inserted protein and the linker region are typed in bold and italics, respectively. The position corresponding to the restriction site (EcoRI/NdeI) for $pL_{\beta 2-\alpha}$, SalI/KpnI for SH2 and $pL_{\beta 3-\beta 4\ a}$) are in brackets. Introduction of the KpnI site into protein L results in the insertion of a threonine residue between G_{55} and Y_{56}. Two point mutations, F170V and E176T, were introduced into SH2 to create the SalI and KpnI restriction sites and are indicated with *. Control experiments have shown that these mutations do not impact the binding activity of phage displaying protein L of SH2. The sequence of B11 follows: RSPAQVVDAQQNAVKDNEPSGSALGGRSAPGATRPSDQSGGSEDRSVPTEKPKEGPHID. Sequence numbering systems for SH2 and SH3 are described in Tong et al. (1996) and Riddle et al. (1997), respectively.

exhibited a cooperative and reversible folding transition with *m*-values (the denaturant dependence of the free energy of folding) similar to that of wild-type protein L (wild-type protein L: $m = 1.8$; pL[SH3]: $m = 1.7$; pL[B11]: $m = 1.8$). The free energy of unfolding for pL[SH3] was slightly reduced in comparison with wild-type protein L, whereas that of pL[B11] was drastically decreased (wild-type protein L: $\Delta G = 4.6$ kcal/mol; pL[SH3]: $\Delta G = 3.4$ kcal/mol; pL[B11]: $\Delta G = 1.0$ kcal/mol). These data confirm our assumption that the insertion of an unfolded sequence into protein L results in a large decrease in stability, whereas the insertion of a folded sequence results in minimal stability loss.

The equilibrium denaturation data suggest that if protein L were destabilized by 1–2 kcal/mol, the folding of the pL[B11] chimera, but not that of the pL[SH3] chimera, would be disrupted, thereby decreasing the permissiveness of protein L with respect to loop insertions. To destabilize protein L, panning experiments with phage displaying wild-type protein L, pL[SH3], and pL[B11] were performed in 1 M guanidine (Table 2). In comparison with panning in the absence of guanidine, an approximate 10-fold loss in recovery was observed for wild-type protein L and pL[SH3] phage whereas a 100-fold loss in recovery was observed for pL[B11] phage. However, the recovery of pL[B11] was still 10-fold above background recovery levels.

Table 2. *Panning recoveries for protein L experiments*

Host	Insert	% Recovery 0M gnd	% Recovery 1M gnd
pL	none	1.0×10^{-2}	2.4×10^{-3}
$pL_{\beta 2-\alpha}$	SH3	6.0×10^{-4}	—
$pL_{\beta 3-\beta 4}$	SH3	2.0×10^{-1}	1.6×10^{-2}
pL	B11	5.0×10^{-2}	3.6×10^{-4}
pL_{A8G}	none	9.0×10^{-3}	1.2×10^{-3}
pL_{A8G}	SH3	5.0×10^{-3}	8.4×10^{-5}
pL_{A8G}	B11	8.2×10^{-4}	0
Control	n/a	1.2×10^{-5}	1.2×10^{-5}

5×10^9 c.f.u. of freshly prepared phage were used as input. % recovery is calculated as 100 × (c.f.u. input/c.f.u. output). Control experiments correspond to phage displaying the SH2 domain rather than protein L.

We then destabilized protein L by mutating residue A8 to a glycine (Fig. 1). This mutation was chosen because it destabilizes the protein by 2.4 kcal/mol and preliminary studies suggest that it does not interfere with IgG binding (Kim et al. 2000). This strategy of destabilizing the host by mutagenesis was used successfully in phage-display experiments using protein G, a small single-domain protein with a topology identical to protein L: A wild-type protein G host tolerated 50% of randomized turn sequences whereas a destabilized protein G host tolerated only a small fraction of the randomized sequences (Zhou et al. 1996). Both B11 and SH3 were inserted into the β_3–β_4 turn of the A8G point mutant and subjected to phage panning experiments. The results were similar to those observed for the wild-type protein L chimera panned in the presence of guanidine; compared with wild-type protein L, pL_{A8G} and pL_{A8G}[SH3] phage experienced a 10-fold loss in recovery and pL_{A8G} [B11] phage experienced a 100-fold loss in recovery. Because the recovery levels of pL_{A8G} [B11] phage were still 10-fold above background levels, additional panning experiments were performed in the presence of 1 M guanidine (Table 2). Both pL_{A8G}[SH3] and pL_{A8G}[B11] were recovered at very low levels, however, indicating that these conditions are too stringent for the pL_{A8G} chimeric proteins.

In an effort to find a less permissive host protein, we turned to the lck SH2 domain. A surface loop in SH2 (Fig. 1) was chosen as the insertion point because of its central location in the molecule and its lack of involvement in ligand binding. To test the efficacy of SH2 as a host protein, wild-type SH3, B11, and, as an additional model for an unfolded sequence, a strongly destabilized SH3 point mu-

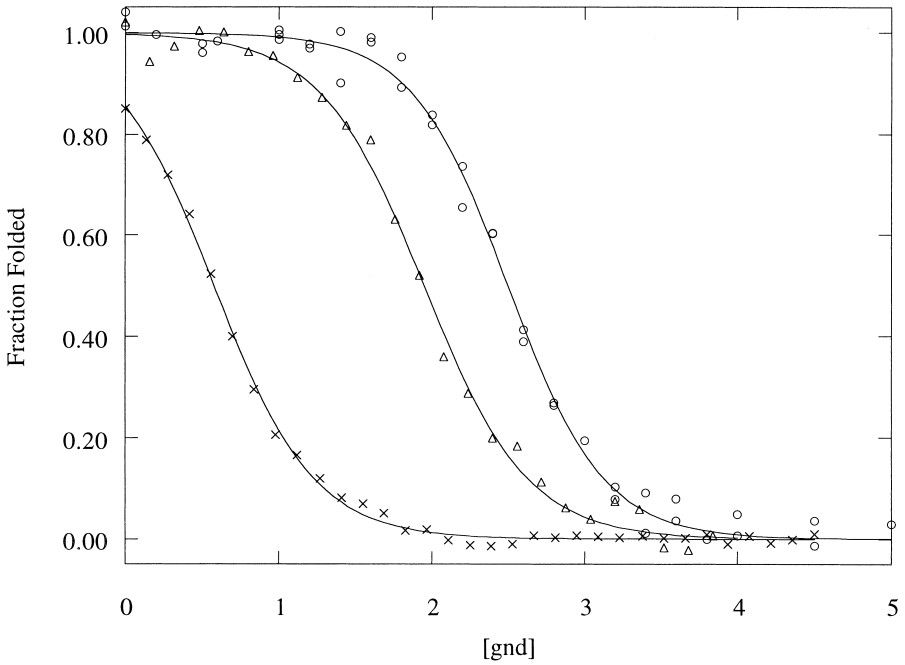

Fig. 2. Equilibrium denaturation melts of wild-type protein L, pL[SH3], and pL[B11]. Denaturation curves of wild-type protein L (open circles), pL[SH3] 0.(open triangles), and pL[B11] (crosses) were monitored by circular dichroism at 220 nm. The data were fit as described by Scalley et al. (1997). Protein concentrations were 10 ± 0.2 μM for wild-type protein L, 8 μM ± 0.2 μM for pL[B11], and 5 μM ± 0.2 μM for pL[SH3].

tant, $SH3_{L32E}$, were each inserted into the SH2 domain. Insertions were introduced into SH2 in a manner similar to protein L, using two unique restriction sites (*Sal*I/*Kpn*I) introduced into the SH2-gene VIII fusion construct. The amino acid sequences of the region bordering the insertions of wild-type SH3, $SH3_{L32E}$, and B11 into the SH2 domain are shown in Table 1.

To investigate the effects of the different insertions on SH2 function, phage displaying the SH2 insertion variants were panned using paramagnetic beads coated with SH2 ligand, a phosphotyrosyl peptide. The recovery of the different phage is reported in Table 3. The results show that the chimeric protein into which SH3 sequence was inserted is recovered with an efficiency on the same order of magnitude as the host protein without any insertion. The recovery is reduced to background levels when either $SH3_{L32E}$ or B11 is inserted into the host protein. These results indicate that the selection of phage displaying chimeric functional SH2 proteins can efficiently discriminate between folded and unfolded sequences without relying on any biological function of the inserted sequence.

Table 3. *Panning recoveries for SH2 experiments*

Host	Insert	% Recovery
SH2	none	1.0×10^{-2}
SH2	SH3	3.0×10^{-2}
SH2	SH_{L32E}	1.2×10^{-5}
SH2	B11	0
Control	n/a	2.4×10^{-5}

5×10^9 c.f.u. of freshly prepared phage were used as input. Control experiments correspond to phage displaying protein L rather than the SH2 domain.

Discussion

The results of this study show that a screen based on loop entropy reduction is capable of discriminating between folded and unfolded sequences. We have also shown that choosing the correct host protein and the position of insertion are critical to the success of the screen. For example, protein L will allow the insertion of a folded SH3 domain in the β_3–β_4 turn but not the β_3–α turn. However, the β_3–β_4 turn also accommodates the insertion of unfolded sequences, as seen by the retention of protein L function in the pL[B11] chimera.

By destabilizing protein L, we were able to reduce the recovery of the unfolded insert [B11] in comparison with the folded insert [SH3], although its recovery could not be reduced to background levels. The close proximity of the β_3–β_4 insertion site to the C terminus may contribute to the permissiveness of protein L because only a few residues are needed to complete the native structure, and a suitable replacement sequence may be found in the randomized insert.

In such a scenario, the remainder of the loop and the original C terminus would be extruded into the linker between protein L and gene VIII.

The underlying idea behind the loop entropy reduction screen is theoretically applicable to any host protein if a phenotypic screen related to the functional integrity of the host protein is available. A similar approach was employed using *Escherichia coli* RNase H1 as a host protein (Doi et al. 1997, 1998). In this study, random sequences were inserted into a surface loop of RNase H1, and chimeric proteins that retained RNase H1 function were selected using an in vivo assay. It was found that the inserted sequences that were folded maintained their structure on excision from the chimera. However, structural characterization of the chimeric proteins that came through the screen demonstrated that the inserted sequences were not always folded; three out of five of the chimeric proteins characterized were found to have unfolded inserted sequences. We observed a similar permissiveness with protein L as the host protein, providing further evidence that not all host proteins are equally good at discriminating between folded and unfolded inserts.

Other studies that have probed random sequence libraries for folded proteins have relied on expression of the random sequence as a screening method. Davidson and coworkers (Davidson and Sauer 1994; Davidson et al. 1995) constructed a library of random sequences consisting of three amino acids (Q, L, R). Interestingly, the fraction of sequences containing some degree of structure was large enough to allow detection of folded proteins using a screening method based on expression and solubility of protein from individual clones. In an extension of this work, Prijambada and coworkers (1996) constructed random sequence libraries containing all 20 amino acids. In that study, the investigators concluded that 8% of the random sequences were expressed and soluble, but no proteins with extensive secondary structure were found. To extend this work further, it is necessary to examine a larger number of sequences than is possible with expression-based screening methods.

In this study, we have employed a phage-display screening method that has several advantages over expression-based screening methods. First, phage-display techniques allows examination of many more sequences than is possible in an expression-based screen. Second, phage display has been proven successful in selecting rare folded variants within a collection of highly randomized domains (Zhou et al. 1996; Riddle et al. 1997; Kim et al. 1998). Additionally, the physical conditions of the selection step can be easily controlled and adapted to specific requirements. For instance, phage-panning experiments are compatible with the presence of reducing agents, denaturants, and proteases and these reagents effectively increase the selection pressure of the screen (Kristensen 1998; Jung et al. 1999). This feature may prove useful in eliminating false positives associated with inserted sequences that are marginally stable.

We are currently using the loop entropy reduction selection described in this paper to search for folded proteins in libraries of randomized synthetic sequences and libraries of shuffled genomic sequences. As a consequence of the design of the screen, it is very likely that only modules with their N and C termini in close proximity will be selected. Although this is certainly a limitation of the selection, it should be noted that the N and C termini are near one another in a disproportionately large number of globular proteins (Thornton 1983). It is attractive to speculate that this relatively high frequency of proximity between N and C termini is an evolutionary relic of a mechanism for generating complex, multidomain proteins from smaller folded units similar to our experimental selection strategy: the insertion of folded modules into loops of other folded modules. Thus, it is possible that the selection experiments may to some extent recapitulate the generation of the complex multidomain protein structures found in nature.

Materials and methods

Preparation and panning of phage

All phage were prepared as described in Gu et al. (1995). The preparation of IgG-coated magnetic beads and the subsequent protein L-panning experiments were performed as described in Gu et al. (1995), except for the guanidine-panning experiments where 1 M guanidine (USB, Ultrapure) was present in both the binding and washing steps. For the SH2-panning experiments, the streptavidin-coated magnetic beads (Dynabeads M-280) were coated with a biotinylated phosphotyrosine peptide (GGGGGGEPPQ[pY]EE IPIYL; synthesized by Sigma Genosys). A total of 20 μL of the streptavidin-coated beads (10 mg/mL) were incubated with 0.2 μg peptide, 0.5% TWEEN-TBS for 1 h, washed twice with 800 μL of 0.5% TWEEN-TBS, and resuspended in 20 μL of 0.5% TWEEN-TBS. The prepared beads (2 μL) were incubated with 5×10^9 phage particles for 1 h in 100 μL of a 4% milk, 0.1% TWEEN-TBS solution. The beads were washed 7 times with 800 μL of 0.5% TWEEN-TBS and resuspended in a final volume of 30 μL of 0.5% TWEEN-TBS. The phage bound to the beads were transfected into XLI Blue cells and plated onto LB agar with carbenicillin to quantitate the number of phage bound to the beads.

Random sequence library

A random sequence library was made by self-ligation of a highly degenerate cassette: $GGATCC(VNNNNB)_9GGATC$ where N = A,C,T,G; V = C,A,G; B = G,T,C; and GGATCC is the *Bam*HI cleavage site. The VNN, NNB polymer codes for polypeptides containing all amino acids in proportions similar to the sequence of typical soluble proteins. The cassette is symmetrical, allowing polymerization in both orientations. A Gly–Ser sequence corresponding to the *Bam*HI sequence occurs every 20 amino acids in the polymer. As only one stop codon occurs among 96 possibilities, a significant fraction of polymerized cassettes inserted in the host protein sequences are expected to be full length. The cloning of the random sequences into the host protein sequence

requires the proper restriction sites and flanking sequences at each of the ends. Two adaptor cassettes ("Start" and "Stop") containing a single *Bam*HI-cohesive extremity and one uncleaved (*Sal*I and *Kpn*I, respectively, for start and stop) restriction site at the other end were introduced in a low molar ratio (1/8) in the random-cassette ligation reaction. Under these conditions, a ladder of products between 100 bp and 1 kb was obtained. For ligation into the host protein, the polymerized cassettes were cleaved with *Sal*I and *Kpn*I. Fragments >200 bp were gel purified and ligated into the vector. The constructs were then electroporated into XLI Blue *E. coli* cells to give a library of 1.5×10^5 independent clones. One round of phage-display selection followed by a colony-lift assay was sufficient to identify several positive clones. Two sequences containing 60 and 80 amino acids were sequenced and both displayed highly polar sequences. One of these sequences, B11, was used in subsequent experiments.

Equilibrium denaturation

The methods described by Gu et al. (1995) were used for the overexpression and purification of the chimeric proteins. Circular dichroism equilibrium denaturation experiments were performed as described by Scalley et al. (1997).

Acknowledgments

We thank members of the Baker laboratory for helpful discussion and especially Karen Butner and David Kim for their help with design and experimental procedures. This work was supported in part by a NATO Fellowship to P.M., NIH training grants to M.S.-K. and A.W., and a NIH grant to D.B.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Bacon, D.J. and Anderson, W.F. 1988. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6:** 219–220.

Betton, J.M., Jacob, J.P., Hofnung, M., and Broome-Smith, J.K. 1997. Creating a bifunctional protein by insertion of beta-lactamase into the maltodextrin-binding protein. *Nat. Biotechnol.* **15:** 1276–1279.

Collinet, B., Herve, M., Pecorari, F., Minard. P., Eder, O., and Desmadril, M. 2000. Functionally accepted insertions of proteins within protein domains. *J. Biol. Chem.* **275:** 17428–17433.

Davidson, A.R and Sauer, R.T. 1994. Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl. Acad. Sci.* **91:** 2146–2150.

Davidson, A.R., Lumb, K.J., and Sauer, R.T. 1995. Cooperatively folded proteins in random sequence libraries. *Nat. Struct. Biol.* **2:** 856–864.

Doi, N., Itaya, M., Yomo, T., Tokura, S., and Yanagawa, H. 1997. Insertion of foreign random sequences of 120 amino acid residues into an active enzyme. *FEBS Lett.* **402:** 177–180.

Doi, N., Yomo, T., Itaya, M., and Yanagawa, H. 1998. Characterization of random-sequence proteins displayed on the surface of Escherichia coli RNase HI. *FEBS Lett.* **427:** 51–54.

Forrer, P., Jung, S., and Pluckthun, A. 1999. Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Curr. Opin. Struct. Biol.* **9:** 514–520.

Gu, H., Yi, Q., Bray, S.T., Riddle, D.S., Shiau, A.K., and Baker, D. 1995. A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* **4:** 1108–1117.

Jung, S., Honegger, A., and Pluckthun, A. 1999. Selection for improved protein stability by phage display. *J. Mol. Biol.* **294:** 163–180.

Kihlberg, B.M., Sjoholm, A.G., Bjorck, L., and Sjobring, U. 1996. Characterization of the binding properties of protein LG, an immunoglobulin-binding hybrid protein. *Eur. J. Biochem.* **240:** 556–563.

Kim, D.E., Gu, H., and Baker, D. 1998. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc. Natl. Acad. Sci.* **95:** 4982–4986.

Kim, D.E., Fisher, C., and Baker, D. 2000. A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298:** 971–984.

Kraulis, P.J. 1991. MOLSCRIPT- a program to produce both detailed and schematic plots of protei structures. *J. Appl. Crystallogr.* **24:** 946–950.

Kristensen, P.W. and Winter, G. 1998. Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des* **3:** 321–328.

Ladurner, A.G. and Fersht, A.R. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* **273:** 330–337.

Merrit, E.A. and Murphy, M.E.P. 1994. Raster 3D version 2.0. A program for photorealistic molecular graphics. *Acta Cyrstallogr.* sect. D 50, 869-873.

Nagi, A.D. and Regan, L. 1997. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* **2:** 67–75.

Prijambada, I.R., Yoma, T., Tanaka, F., Kawama, T., Yamamoto, K., Hasegawa, A., Shima, Y., Negora, S., and Urabe, I. 1996. Solubility of artificial proteins with random sequences. *FEBS Lett.* **382:** 21–25.

Riddle, D.S., Santiago, J.V., Bray-Hall, S.T., Doshi, N., Grantcharova, V.P., Yi, Q., and Baker, D. 1997. Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **4:** 805–809.

Scalley, M.L., Yi, Q., Gu, H., McCormack, A., Yates, J.R., and Baker, D. 1997. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry* **36:** 3373–3382.

Thornton, J. and Sibanda, B. 1983. Amino and carboxy-terminal regions in globular proteins. *J. Mol. Biol.* **167:** 443–460.

Tong, L., Warren, T.C., King, J., Betageri, R., Rose, J., and Jakes, S. 1996. Crystal structures of the human p56lck SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 A and 1.8 A resolution. *J. Mol. Biol.* **256:** 601–610.

Wikstrom, M., Sjobring, U., Kastern, W., Bjorck, L., Drakenberg, T., and Forsen, S. 1993. Proton nuclear magnetic resonance sequential assignments and secondary structure of an immunoglobulin lightchain-binding domain of protein L. *Biochemistry* **32:** 3381–3386.

Zhou, H.X., Hoess, R.H., and DeGrado, W.F. 1996. In vitro evolution of thermodynamically stable turns. *Nat. Struct. Biol.* **3:** 446–451.