Cell PRESS
Structure
Article

Molecular Evolution of Cystine-Stabilized Miniproteins as Stable Proteinaceous Binders

Hung-Ju Chang,[1,2,3] Hung-Ju Hsu,[1,4] Chi-Fon Chang,[1] Hung-Pin Peng,[1,2,5] Yi-Kun Sun,[1] Hui-Ming Yu,[1] Hsi-Chang Shih,[1] Chun-Ying Song,[1] Yi-Ting Lin,[1] Chu-Chun Chen,[1] Chia-Hung Wang,[1,6] and An-Suei Yang[1,*]
[1]Genomics Research Center
[2]Taiwan International Graduate Program
Academia Sinica, Taipei 115, Taiwan
[3]Institute of Biochemical Science, National Taiwan University, Taipei 106, Taiwan
[4]Graduate Institute of Life Sciences, National Defence Medical Center, Taipei 114, Taiwan
[5]Institute of BioMedical Informatics
[6]Institute of Biochemistry and Molecular Biology
National Yang-Ming University, Taipei 112, Taiwan
*Correspondence: yangas@gate.sinica.edu.tw
DOI 10.1016/j.str.2009.01.011

SUMMARY

Small cystine-stabilized proteins are desirable scaffolds for therapeutics and diagnostics. Specific folding and binding properties of the proteinaceous binders can be engineered with combinatorial protein libraries in connection with artificial molecular evolution. The combinatorial protein libraries are composed of scaffold variants with random sequence variation, which inevitably produces a portion of the library sequences incompatible with the parent structure. Here, we used artificial molecular evolution to elucidate structure-determining residues in a smallest cystine-stabilized scaffold. The structural determinant information was then applied to designing cystine-stabilized miniproteins binding to human vascular endothelial growth factor. This work demonstrated a general methodology on engineering artificial cystine-stabilized proteins as antibody mimetics with simultaneously enhanced folding and binding properties.

INTRODUCTION

Small cystine-stabilized proteins as scaffolds for proteinaceous binders are advantageous (Craik et al., 2006; Smith et al., 1998; Silverman et al., 2005) in that the cystine cross-linked miniprotein structures are stable in extreme conditions and are relatively easy to produce (Smith et al., 1998; Silverman et al., 2005; Souriau et al., 2005). The underlying rationale for this direction was attained from a large number of naturally occurring cases. One set of examples are knottins, a large family (more than 1200 known sequence members to date) (Gracy et al., 2008) of small (<~50 residues) cystine-stabilized proteins for which the main biological function is molecular recognition on proteins, sugars, and lipids. These proteins provide a rich source of potentially useful scaffolds for ever expanding applications in molecular recognition.

Monoclonal antibodies and their recombinant fragments are the mainstream scaffolds in protein recognition engineering, but alternative nonantibody scaffolds are frequently desirable to overcome the limitations imposed by antibody production and storage (Skerra, 2007). To design antibody mimetics, we constructed repertoires of combinatorial protein-recognition surfaces on a small cystine-stabilized scaffold with a spatially contiguous patch composed of two discontinuous variable sequence segments. Such protein recognition surfaces mimic antibody variable domain paratope structures, which are composed of sequence-wise distant complementarity determining regions (CDRs) forming a spatially contiguous antigen recognition surface. In conjunction with powerful phage display-based artificial molecular evolution (Silverman et al., 2005; Skerra, 2007; Hsu et al., 2008, 2006; Ladner et al., 2004; Sidhu et al., 2003; Lowman, 1997), the large repertoires of combinatorial proteins are important tools enabling protein engineering processes mimicking in vivo antibody-antigen interaction selection and reinforcement.

Engineering protein binding needs insight into protein folding. The reason is that recognition surfaces are frequently preassembled from sequence-wise distant segments, arranged in close proximity in space through a stably folded structure (Wodak and Janin, 2002). The sequence design in this work was based on a 23-residue autonomous folding unit (dubbed Min-23) (Heitz et al., 1999) from EETI-II (ecballium elaterium trypsin inhibitor II, a member of the knottin family) with four cysteines forming 1-3/2-4 disulfide bonding configuration (Souriau et al., 2005). Structural determinant information is essential in determining invariable residues, which are indispensable in supporting the targeting surfaces on the scaffold as designed. The folding mechanisms of Min-23 (Heitz et al., 1999) and its parent protein EETI-II (Heitz et al., 1995; Wentzel et al., 1999; Le-Nguyen et al., 1993) have been well-characterized, but the structural determinants underlying the native disulfide bonding configuration are not clearly understood.

Structural determinant information has been difficult to attain. Although successes in consensus-based protein engineering approaches (Kajander et al., 2006) have generally validated the underlying hypothesis that the fold-specifying and structure-stabilizing residues occur with high frequencies in multiple

alignments of evolutionarily related sequences (Socolich et al., 2005; Davidson, 2006), conserved sequence features in a evolutionary sequence profile reflect mixed information beyond just structural determinants. Furthermore, the sequence analysis becomes progressively ineffective for proteins with decreasing coverage of related sequences. Hence, the bioinformatics methodology is inapplicable for engineering in silico designed proteins, for which related protein sequences do not exist in protein sequence databases. In this regard, we used artificial molecular evolution (Hsu et al., 2006 ; Sieber et al., 1998 ; Michnick and Sidhu, 2008) to select from random collections of combinatorial sequences for foldable sequences, for which the conserved sequence features revealed only fold-specifying and structure-stabilizing residues (Hsu et al., 2006). This process not only mimics the evolution of protein sequences in nature, but also the methodology was specific in elucidating key structural determinant residues under artificially controllable folding conditions regardless of the availability of naturally occurring related sequences.

The structural determinant information for Min-23 structure was applied to designing cystine-stabilized proteinaceous binders against human vascular endothelial growth factor (VEGF-A) (Gatto and Cavalli, 2006). VEGF-A is a major proangiogenesis protein secreted by tumor cells and binds to its receptor (VEGFR2) and to neuropilin on endothelial cells to stimulate the angiogenesis process (Folkman, 2007). Monoclonal antibody trapping VEGF has been developed (bevacizumab, or Avastin from Genentech/Roche) as an angiogenesis inhibitor, and has demonstrated efficacy against colon cancer, non-small-cell lung cancer, and breast cancer (Bergers and Hanahan, 2008). With the molecular evolution and recombinant phage display techniques, we demonstrated that cystine-stabilized miniproteins binding to VEGF-A could be readily obtained from the protein libraries designed based on the structural determinants for the scaffold structure.

RESULTS

Intact Fraction as a High-Throughput Protein Stability Measurement

One key technique in the molecular evolution methodology for stability engineering is the intact fraction measurement (Figure 1) that quantifies the stability of a phage-displayed cystine-stabilized protein on a high-throughput platform; the intact fraction is the percentage of the phage-displayed proteins remaining folded (intact) after the treatment of protein denaturants followed by bovine factor Xa (fXa) digestion. The stability measure required that the fXa cleavage site (-IEGR-) was embedded in the displayed protein sequences, in such way that the site was protected from fXa digestion in a cystine-stabilized structure but became accessible for fXa cleavage upon protein unfolding. In addition to the designated cutting site, no other fXa cleavable sites should exist in the displayed sequences. As depicted in Figure 1, the phage-displayed proteins were unfolded by guanidine hydrochloride (GdnHCl) in the presence of dithiothreitol (DTT) or Tris(2-carboxyethyl)phosphine to reduce disulfide bonds, followed by the N-ethylmaleimide (NEM) treatment to block cysteines from reforming cystine-stabilized structure. Only the GdnHCl/DTT-unfolded and NEM-blocked polypeptides

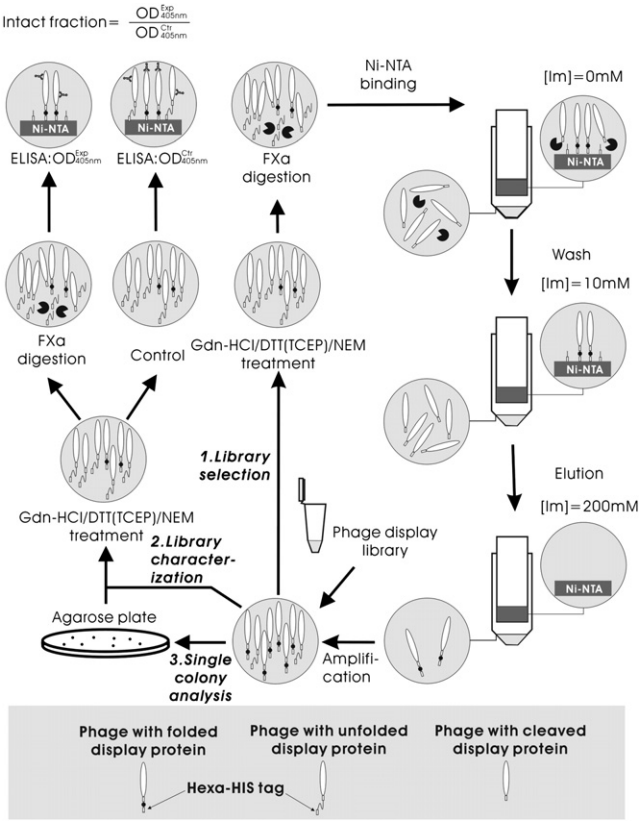

Figure 1. Flow Chart Depicting Procedures for Library Selection, Library Characterization, and Single-Colony Analysis

(1) Library selection: Phage-display library (depicted as the Eppendorf tube in the lower part of the figure) is treated with denaturants followed by fXa digestion. Phage particles with intact hexa-HIS tag are enriched with Ni-NTA beads. The selected phage particles are amplified and used as input phage display library for the next selection cycle.

(2) Library characterization: Phage display library (depicted as the Eppendorf tube in the lower part of the figure) is treated with denaturants. The treated phage solution is parted into two equal portions: one treated with fXa and the other one used as control. The intact fraction is calculated as the ratio of ELISA signal readouts from the fXa-treated portion over the control.

(3) Single-colony analysis: A single colony is picked from agarose plate and phage particles are rescued and cultured overnight. The intact fraction of the phage is determined as described in (2).

were cleaved at the exposed cutting site (-IEGR-) by fXa, and consequently the N-terminal hexa-His tag was removed from the phage particle. The intact fraction was then determined by a pair of enzyme-linked immunosorbent assays (ELISAs), which determined the ratio of phages capable of binding to Ni^{2+}-coated surface from phage solutions before and after the fXa digestion.

Two control phages were constructed to validate the intact fraction measurement. The pC-Xa phage (Figure 2) was the positive control, for which the displayed peptide mimics the exposed fXa cutting site in a fully unfolded polypeptide appending to the phage particle. As expected, the measured intact fractions were all below 10% at various denaturant conditions (Figure 3A). In contrast, the negative control pC-Min-23 phage (Figure 2) was insensitive to the denaturant conditions and fXa digestion (Figure 3A), indicating that no additional fXa cutting sites existed

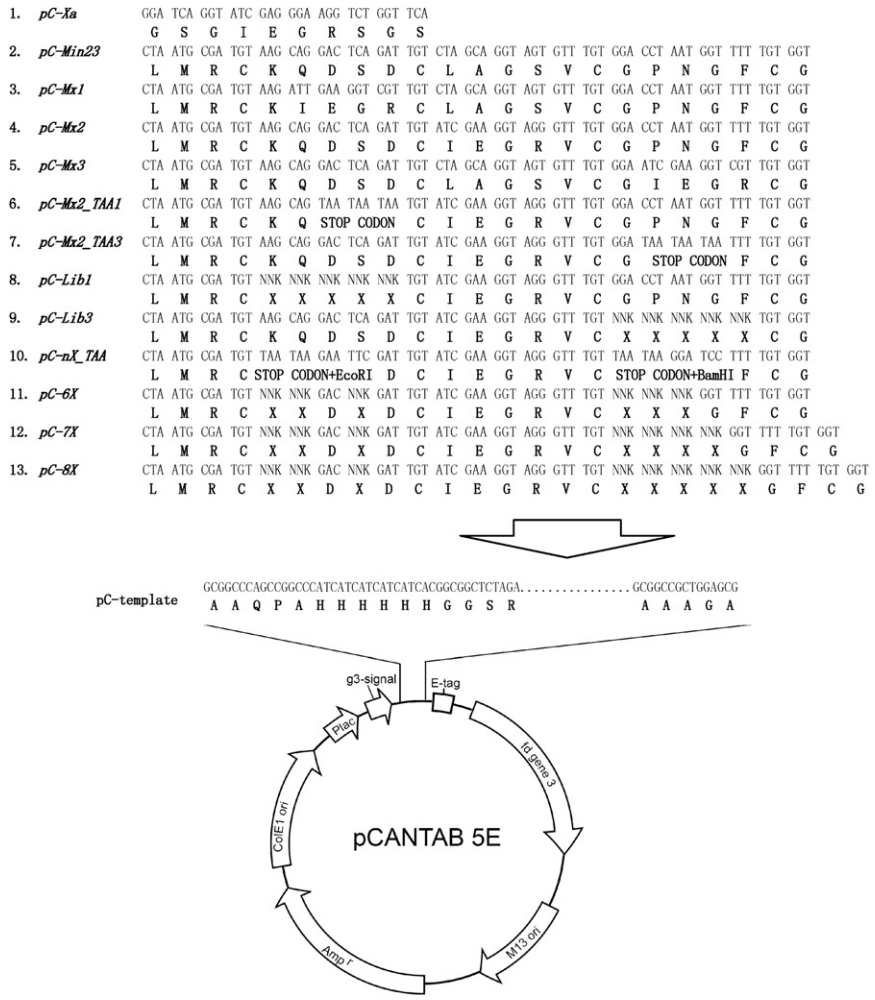

Figure 2. Phagemid Constructs for Phage-Displayed Proteins and Combinatorial Libraries

DNA fragments were cloned with the pCANTAB5E phagemid as shown. Restriction enzyme XbaI (restriction site TCTAGT) and NotI (restriction site GCGGCCGC) were used for the cloning procedure, except for pC-Xa where SpeI (restriction site ACTAGT) was used in place of XbaI. The DNA fragments were inserted after the g3 signal sequence, which is N-terminal to the E-tag followed by the gene 3 in the pCANTAB5E phagemid.

in the pIII fusion protein and that the pC-Xa intact fraction measurements reflected only the cleavage of the -IEGR- site between the pIII protein and the hexa-His tag.

Design of Cystine-Stabilized Scaffolds for Phage Display

Three Min-23 variants with IEGR replacement for the residues between cysteines (constructs pC-Mx1∼3, Figure 2) showed different stability under various denaturant conditions. The three variants were equally stable in the presence of only fXa, indicating that disulfide bonding either protected the fXa cleavage site or kept the cleaved peptide fragments connected through the disulfide linkage. In the presence of denaturants, pC-Mx2 remained stable (Figure 3A) whereas the sequences of pC-Mx1 and pC-Mx3 yielded less stable structures (data not shown), in agreement with the 1-3/2-4 disulfide bonding configuration of the Min-23 scaffold. This is because for the N-terminal HIS tag to be cleaved from pC-Mx2, both disulfide bonds need to be disrupted (assuming the 1-3/2-4 configuration); whereas for pC-Mx1 and pC-Mx3, only one disulfide bond needs to be disrupted for the N-terminal HIS tag cleavage. Thus, it was not unexpected that pC-Mx2 was more stable than the other two phages in the presence of denaturant and fXa. In comparison, the stability of Mx2 was indistinguishable from that of Min-23 (Figure 3A). The stability of the Mx2 structure was expected: the local sequence between the second and the third cysteine (-LAGS-) in Min-23 forms type II turn in the EETI-II X-ray structure (Kratzner et al., 2005) (Figure 4D). This local structure is highly compatible with the fXa-cleavable sequence (-IEGR-), as judged by the sequence preferences for the tight turn in known protein structures (Peng and Yang, 2007; Yang and Wang, 2003).

The structure of the refolded Mx2 polypeptide was determined with 2D nuclear magnetic resonance (NMR) (Figure 4). As shown in the superimposed structures (Figure 4D), the structure of Mx2 is similar to the corresponding domain in the wild-type EETI-II X-ray structure; the characteristic 1-3/2-4 disulfide bonding configuration in Min-23 was conserved in the Mx2 structure. To further confirm the disulfide bonding configuration, we did the structure calculation for all 3 possible connectivities, the target functions were 4.28×10^{-2}, 2.80×10^{-5}, and 9.78×10^{-3} for (9-15, 21-27), (9-21, 15-27), (9-27,15-21) in Mx2, respectively. Because the lower target function indicates lower violations, these results confirmed that (9-21, 15-27) was the most favorable configuration for the disulfide bonds. The floating disulfide computation also confirmed that the (9-21, 15-27) disulfide connectivity was the most stable species.

To the limit of phage tolerance of 2 M GdnHCl and 10 mM DTT, the intact fraction of the Mx2 phage was 78 ± 18% (Figure 3A). To confirm this intact fraction as the ratio of folded over total conformations, the refolded Mx2 polypeptide was treated with 2 M GdnHCl and 10 mM DTT for 30 min before measuring the fraction of the folded conformation with reverse-phase high-performance liquid chromatography (HPLC). The result showed that 48% of the Mx2 polypeptide remained folded. The discrepancy reflected the intrinsic differences between the two methods in measuring the fraction of the polypeptide remaining disulfide bonded after the denaturant/protease treatment. The HPLC method measured the fraction of fully disulfide-bonded polypeptide, whereas the intact fraction method required that, in addition to the reduction of the disulfide bonds, the IEGR site in the

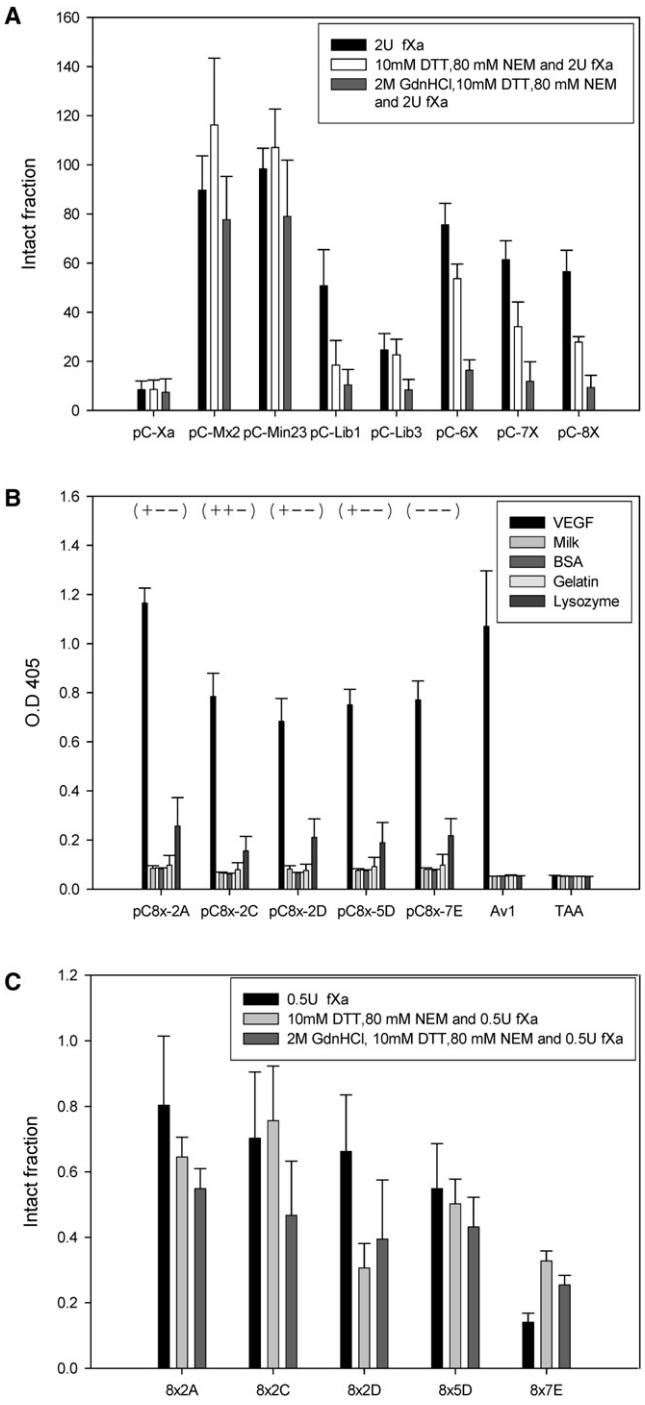

Figure 3. Selection and Screening for Cystine-Stabilized Binders against Human VEGF-A

(A) The histogram compares the intact fractions of controls and phage-displayed libraries with (pC-6∼8X) and without (pC-Lib1∼3) the structural determinant residues. The intact fractions were obtained after treatment of 2U of fXa for 2 hr (black bar); 10 mM DTT, 80 mM NEM, and 2U fXa for 2 hr (white bar); 2M GdnHCl, 10mM DTT, 80 mM NEM, and 2U fXa for 2 hr (gray bar). The standard deviations are shown as the error bars, calculated with 12 repeats of the measurements.

(B) Five pC-8X phage-displayed variants were selected from 48 randomly picked phage colonies after two rounds of biopanning against human VEGF-A. The *y* axis shows the ELISA OD at 405 nm developed after the binding of the phage variants to the wells coated with 1 μg VEGF-A, milk, bovine serum albumin (BSA), fish gelatin, and hen lysozyme, respectively. All wells were blocked with 3% fish gelatin before phage binding. The displayed sequences for these variants are shown in Table 3. The stability class—e.g., (++−), (+−−), and (−−−)—for each of the variants is labeled above the histogram based on the intact fraction measurements as shown in panel (C). The error bars indicated the standard deviations derived from four repeats of the measurements. In addition to the five pC-8X variants, positive control was carried out with a recombinant phage displaying anti-VEGF antibody fragment scFv (Av1), for which the phagemid was constructed based on the published sequence (Fuh et al., 2006). The negative control (TAA) was carried out with null recombinant phage containing only wild-type pIII capsid protein.

(C) The histogram shows intact fractions of the selected VEGF-binding variants from pC-8X library (see panel [B]). The detailed experimental conditions are described in Experimental Procedures. The error bars indicate the standard deviations of the measurements with three repeats.

reduced polypeptide displayed on the recombinant phage be fully accessible to fXa protease digestion—any steric hindrance against fXa to reach the fXa-cleavable sequence would result in incomplete digestion and thus higher measured intact fraction than the fraction measured by HPLC. Thus, the intact fraction measurements can only be treated as a semiquantitative measure for folding stability, suitable for high-throughput experimental platforms, and the stability comparison based on the intact fraction measurements needs to be carried out under strictly consistent experimental conditions (see below).

Elucidation of Structural Determinant with Molecular Evolution

With the Mx2 established as a stable template structure and the fXa cleavage site fitfully embedded in the cystine-stabilized scaffold, sequence requirements for the folding of the scaffold structure were elucidated with the selection/amplification cycles depicted in Figure 1. Two protein libraries were constructed based on the pC-Mx2 template (pC-Lib1 and pC-Lib3 shown in Figure 2). The libraries covered variants derived from degenerate DNA codon replacements for the residues other than the fXa cutting site and the cysteines with complexity greater than 10^8 each (Experimental Procedures), which was enough to cover all possible variants (20^5) in each of the libraries. Phage-displayed cystine-stabilized structures from the libraries were then enriched with the selection/amplification cycles (Figure 1 and Experimental Procedures). Figure 5 shows the intact fractions of the input libraries under the selection conditions of the following selection cycle. None of the denaturant conditions in Figure 5 affected the infectivity of the recombinant phage to its host *E. coli*. The progressive resistance to the increasingly stringent denaturant conditions (Figure 5) indicated that the selection procedure was effective in enriching cystine-stabilized structures displayed on the recombinant phage. Single colonies were randomly picked from the phage libraries after the 0, 4th, and 8th selection cycle. The intact fractions for each monoclonal displayed protein were determined under three denaturant conditions: (1) GdnHCl (0 M) / DTT (0 M) / NEM (0 M); (2) GdnHCl (0 M) / DTT (10 mM) / NEM (80 mM); (3) GdnHCl (2 M) / DTT (10 mM) / NEM (80 mM). Tables S1 and S2 (available online) show total 237 sequences from pC-Lib1 and pC-Lib3 and the intact fractions under the above denaturant conditions. We clustered the phage-displayed proteins into four classes with decreasing structural stability: class (+++), intact

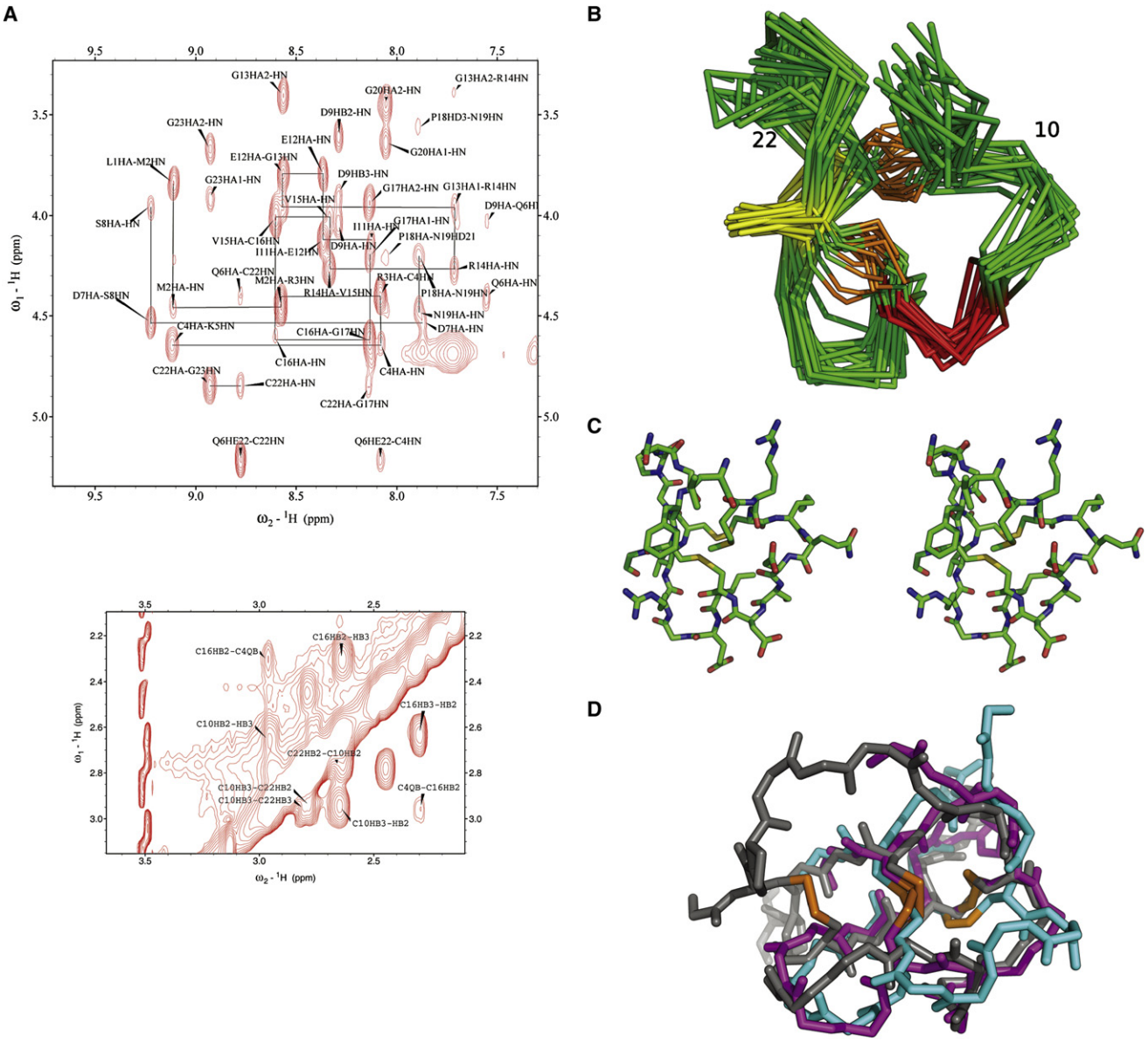

Figure 4. Structure Determination of Mx2 with 2D NMR

(A) The amino acid sequence for Mx2 is LMRCKQDSDCIEGRVCGPNGFCG. The NOESY fingerprint region (upper panel) showed the NOE connectivity between NH and $H_\alpha H_\beta$. The H_β region (lower panel) indicated the disulfide bonding between Cys4-Cys16 (equivalent to Cys 9-Cys21 in EETI-II numbering) and Cys10-Cys22 (equivalent to Cys 15-Cys27 in EETI-II numbering). More details of the NMR experimental conditions can be found in Experimental Procedures.

(B) The ensemble of 20 NMR structures of the refolded Mx2 polypeptide. Helices are colored in red, sheets in yellow, and turns in green. Disulfide bonds are represented in orange sticks. The structure refinement details are shown in Table 1.

(C) This panel shows the side-by-side stereo display of the Mx2 structure.

(D) Superimposition of the Mx2 and NL3 (see Figure 6) structures with the parent protein EETI-II (Protein Data Bank code 1h9h) (Kratzner et al., 2005). The RMSD of EETI-II (gray) with Mx2 (cyan) and NL3 (magenta) are 2.92 Å and 1.33 Å, respectively.

fractions were all above 70% for all three denaturant conditions; class (++−), intact fraction for denaturant condition (3) was below 70%, the other two denaturant conditions were above 70%; class (+−−), only the intact fraction for denaturant condition (1) was above 70%; class (−−−), none of the intact fractions were above 70% for all three denaturant conditions.

The intact fraction measurement uncertainty was determined with six sets of experiments (two control phages: pC-Xa and pC-Mx2 and three denaturant conditions as shown in Tables S1 and S2). Each set of the experiments repeatedly measured the intact fraction 20 times and calculated the standard deviation. The largest standard deviation was 22%, measured with pC-Mx2 phage in GdnHCl (2 M) / DTT (10 mM) / NEM (80 mM) (Tables S1 and S2). Based on this standard deviation, any displayed polypeptide sequence with the intact fraction greater than 70% is within the experimental error to have similar stability

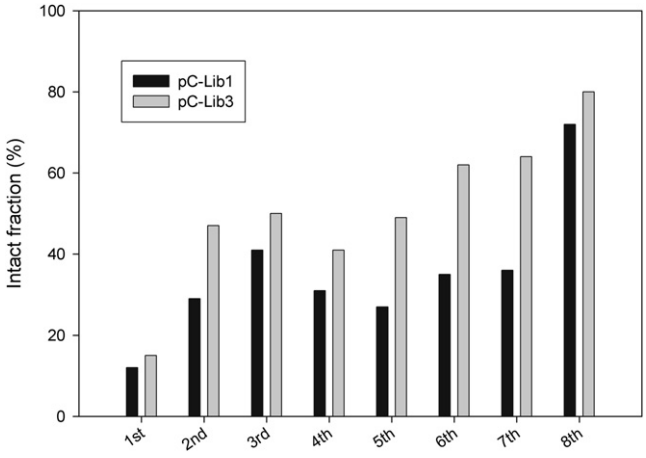

Figure 5. Progressive Resistance of the Phage-Displayed Libraries against Increasingly Stringent Denaturant Conditions at Each Selection/Amplification Cycle

The *y* axis shows the intact fraction of the input library under the selection condition of each selection cycle in the *x* axis: first round, DTT/NEM; 2nd round, GdnHCl (2 M) / DTT / NEM; 3rd round, GdnHCl (4 M) / TCEP gel / NEM; fourth through eighth rounds, GdnHCl (6 M) / TCEP gel / NEM, where DTT and NEM were kept at 10 mM and 80 mM, respectively. None of these denaturant conditions affected the infectivity of the recombinant phages within two orders of magnitude in colony forming units.

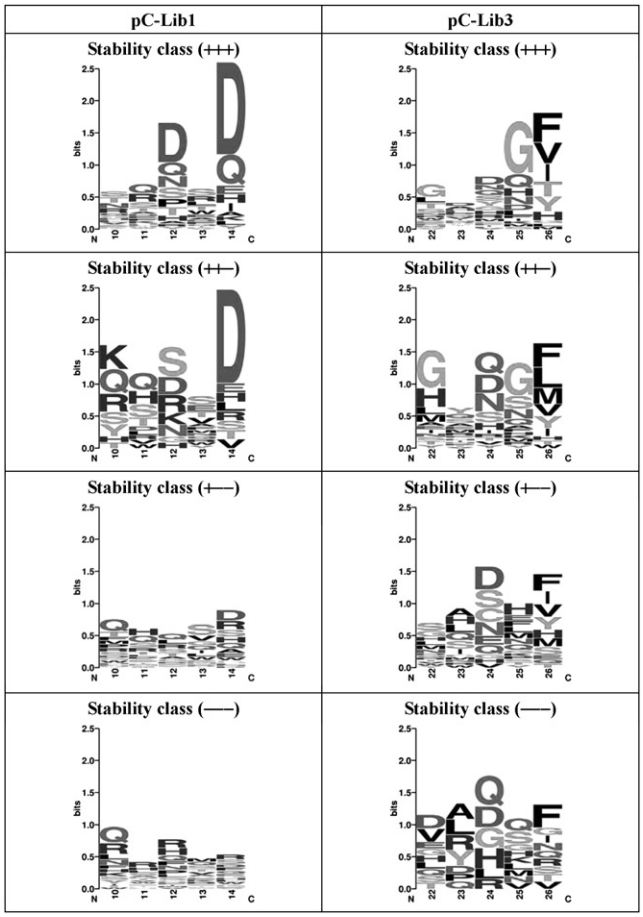

Figure 6. Sequence Profiles for the Sequences from pC-Lib1 and pC-Lib3 with Decreasing Structural Stability

The corresponding sequences and detailed stability measurements are listed in Table S1 (pC-Lib1) and S2 (pC-Lib3). For pC-Lib3 sequences, entries with more than two His in the sequence shown in Table S2 were not included as part of the sequence profile. The reason is that these sequences could be selected by Ni-NTA binding owing to the His-rich sequence rather than the hexa-His tag designed for the selection of the stable structures. This situation is not applicable to sequence members from pC-Lib1.The logo plots were generated by the WebLogo program (http://weblogo.berkeley.edu/) (Crooks et al., 2004) with the background amino acid occurrence frequencies in consistent with the degenerate codons used in the phage display system as shown in Figure 2. See Table S3 for the list of the amino acid occurrence frequencies.

as the Mx2 polypeptide. We used 70% as the threshold to classify the polypeptide sequences as described above.

The LOGO plots in Figure 6 show the sequence preferences for all the four stability classes from the pC-Lib1 and pC-Lib3 library respectively. The sequence preferences emerged in Figure 6 are similar with corresponding parts of the evolutionarily conserved sequence pattern (Figure S1): Among the pC-Lib1 (+++) members, the two most conserved residues (residues 12 and 14 in EETI-II numbering, see Figure 6) are in consensus with the evolutionary sequence pattern, where these two residues are also highly conserved (Figure S1). Similar conclusions applied to the sequence preferences among the pC-Lib3 (+++) members. The two most conserved amino acids (residues 25∼26) are consistent with the two most conserved residues in the homolog sequence-based sequence profile (Figure S1) in the sequence ranging from residue 22 to residue 26. The sequence preference features in the two (+++) classes became progressively diminished with decreasing stability as evident in the sequence profiles derived from the other three stability classes (Figure 6). This result indicated that the sequence features emerged in the (+++) classes are critical to the folding and/or stability of the cystine-stabilized Mx2 structure.

To answer the question as to whether the sequences in the (+++) class folded into cystine-stabilized structure with the 1-3/2-4 disulfide bond configuration, we determined the structure of refolded NL3 polypeptide with 2D NMR spectroscopy (Figure 7). The sequence of NL3 (LMR**C**KQDSD**CIEGR**V**C**VDH DY**C**G) is a (+++) class member from pC-Lib3 with very different sequence from the (+++) class consensus sequence shown in Figure 6. Even so, the NMR structures (Figure 7) are similar to the Mx2 structure and the parent structure EETI-II (Figure 4D) in both general folding topology and the disulfide bond configuration. To further confirm the disulfide bonding configuration, we did the structure calculation for all 3 possible connectivities, the target functions were 2.7×10^{-1},8.7×10^{-2}, and 1.8×10^{-1} for (9-15, 21-27), (9-21, 15-27), (9-27,15-21), respectively. These results confirmed that (9-21, 15-27) was the most favorable configuration for the disulfide bonds.

Cystine-Stabilized Phage Display Libraries for VEGF-A Binding

The sequence profiles shown in Figure 6 provided structural determinant information in cystine-stabilized phage-displayed library constructs. In particular, Asp12, Asp14, Gly25, and Phe26 are the structural determinant residues for the scaffold structure and should remain invariable in the library designs. Based on this insight, three libraries (pC-6X∼pC-8X, see Figure 2) were

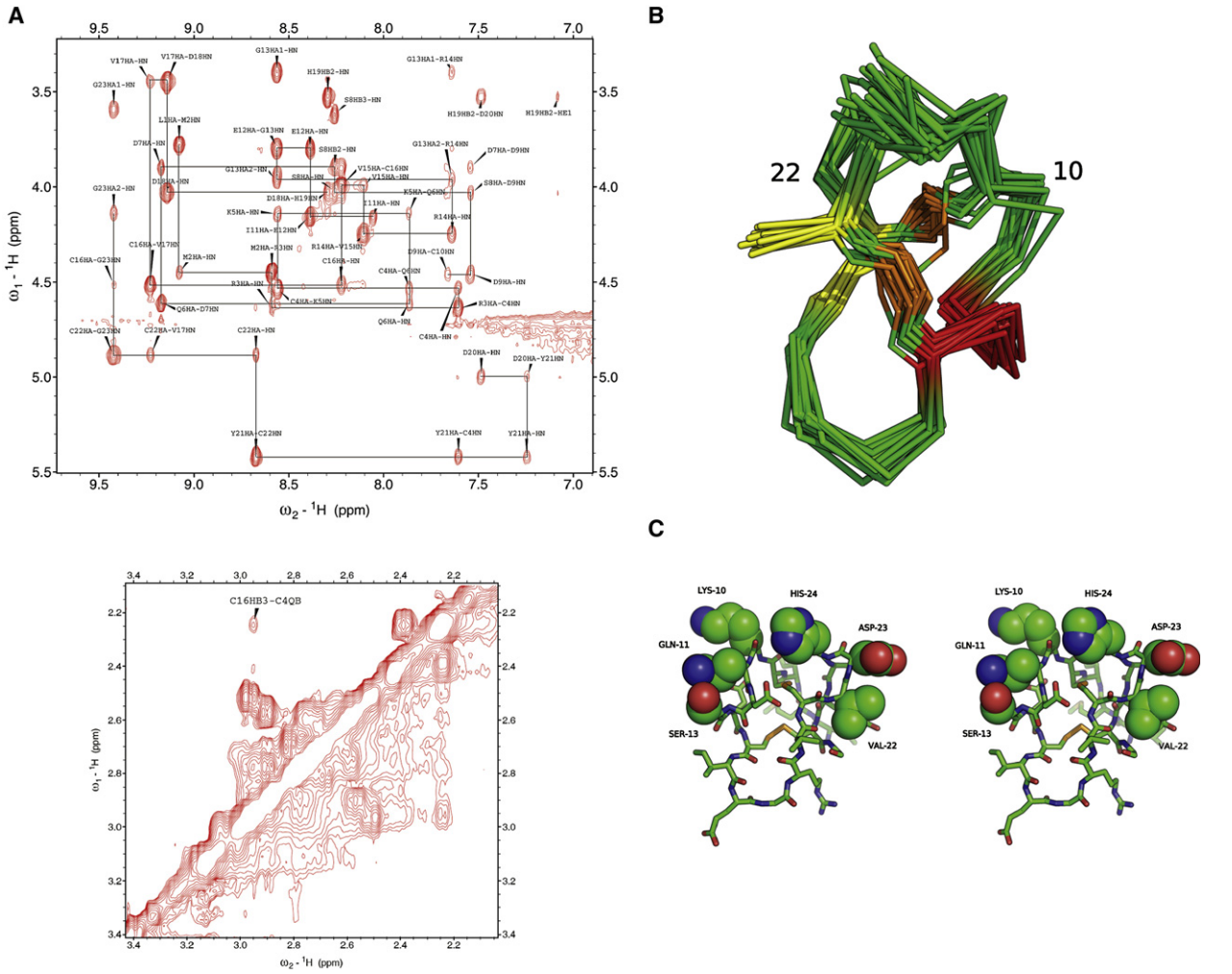

Figure 7. Structure Determination of NL3 with 2D NMR

(A) The amino acid sequence for NL3 is LMRCKQDSDCIEGRVCVDHDYCG. The fingerprint region (upper panel) showed the NOE connectivity between NH and $H_\alpha H_\beta$. The H_β region (lower panel) indicated the disulfide bonding between Cys4-Cys16 (equivalent to Cys 9-Cys21 in EETI-II numbering). The second disulfide link Cys10-Cys22 (equivalent to Cys15–Cys27 in EETI-II numbering) was not resolved in the NMR spectra but mass spectroscopy analysis indicated that two disulfide bonds were evident in NL3, suggesting that the Cys10-Cys22 disulfide bond was also formed. More details of the NMR experimental conditions can be found in Experimental Procedures.

(B) The ensemble of 20 NMR structures of the refolded NL3 polypeptide. Helices are colored in red, sheets in yellow, and turns in green. Disulfide bonds are represented in orange sticks. The structure refinement details are shown in Table 2.

(C) This panel shows the side-by-side stereo display of the NL3 structure. The side chains of the variable residues in the pC-6X library are shown in the space-filling model, and form a contiguous protein-recognition surface on the scaffold.

constructed. The intact fractions at various denaturant conditions were compared among the libraries (Figure 3A). As expected, the pC-6X∼pC-8X libraries indeed contained more cystine-stabilized phage-displayed proteins than the libraries constructed without considering the structural determinant residues (pC-Lib1 and pC-Lib3, Figure 3A). Moreover, the stabilities of the displayed proteins decreases with increasing loop size between the third and the fourth cysteine, suggesting that the elucidated structural determinant residues were more relevant to the original loop size in the Mx2 structure.

The binding properties of variants in pC-8X were further tested against human VEGF-A. Specific phage displayed binders were identified after only two selection cycles (Figure 3B). Even though pC-8X contained the least abundant cystine-stabilized variants among the three libraries, four (one in ++− and three in (+−−) stability class; Figures 3B and 3C) of the five randomly picked VEGF binders as shown in Figure 3B formed disulfide linkages to stabilize the proteinaceous binders. In comparison, the affinities toward VEGF-A were comparable between the selected pC-8X variants and the anti-VEGF antibody fragment scFv (data for Av1 in Figure 3B), but the anti-VEGF scFv (Av1) was relatively more specific in VEGF recognition than the pC-8X variants (Figure 3B).

DISCUSSION

The 23-residue autonomous folding unit (Min-23) from EETI-II is the smallest cystine-stabilized beta-sheet (CSB) motif known

Table 1. Summary of Structural Constraints and Structural Statistics Derived from 2D NMR for Mx2

NMR Distance and Dihedral Constraints	
Distance constraints	
Total NOE	106
Intraresidue	48
Interresidue	58
Sequential (|i-j| = 1)	43
Medium range (|i-j| < 4)	8
Long range (|i-j| > 5)	7
Hydrogen bonds	4
Total dihedral angle restraints	9
Phi	9
Structure Statistics from CYANA	
Target function (Å^2)	2.8E-05
Average upper distance limit (Å)	0.0001
Average dihedral angle constraints (°)	0.0016
Maximum dihedral angle violation (°)	0.05
Maximum upper distance limit (Å)	0.01
Average pairwise rmsd[a] (Å)	
Heavy	0.95 ± 0.16
Backbone	0.53 ± 0.15
Ramachandran analysis	
Most favorable region	72.1%
Additional allowed regions	27.9%
Generally allowed regions	0%
Nonfavorable region	0%

[a] Pairwise root-mean-square deviation was calculated among 20 refined structures for secondary structure region (7–10, 15 and 16, 22 and 23).

Table 2. Summary of Structural Constraints and Structural Statistics Derived from 2D NMR for NL3

NMR Distance and Dihedral Constraints	
Distance constraints	
Total NOE	183
Intraresidue	64
Interresidue	119
Sequential (|i − j| = 1)	66
Medium range (|i − j| < 4)	29
Long range (|i − j > 5)	24
Hydrogen bonds	2
Structure Statistics from CYANA	
Target function (Å^2)	8.71 E-02
Average upper distance limit (Å)	0.10
Maximum upper distance limit (Å)	0.13
Average pairwise rmsd[a] (Å)	
Heavy	0.69 ± 0.38
Backbone	0.29 ± 0.11
Ramachandran analysis	
Most favorable region	63.5%
Additional allowed regions	36.0%
Generally allowed regions	0.5%
Nonfavorable region	0%

[a] Pairwise rmsd was calculated among 20 refined structures for secondary structure region (7–10, 15 and 16, 22 and 23).

so far (Heitz et al., 1999), and is an ideal model system for scrutinizing sequence-structure relationships governing the folding of the motif. The small protein class in SCOP database (Murzin et al., 1995) is composed of 85 folds to date. More than half of these folds belong to disulfide-rich families. Nearly 50% of the disulfide-rich families contain a CSB motif (Souriau et al., 2005). The prevalence of the motif resulting from convergent evolution and the permissiveness in sequence requirements makes the CSB appealing as an ideal scaffold for protein engineering, where sequence tolerance and folding stability/specificity need to coexist. Still, the structural determinants for the CSB are largely unknown. The information is critical not only for understanding the structural determinants in the folding of the CSB, but also in designing combinatorial libraries with variants retaining the CSB structure.

Even with the small size of Min-23, the potential sequence variation of the noncysteine residues is still astronomically vast (20^{19} variants), and hence neither experimental nor computational methodologies provide effective means to comprehensively elucidate the sequence-structure relationships of the protein. In this study, we constructed combinatorial protein libraries to exhaustively explore the sequence requirements within two consecutive 5-residue sequence segments of Min-23. The complexities of the combinatorial libraries (10^8∼10^9) was enough to cover all the possible variants in these sequence segments (20^5). The first segment is the sequence between the first (Cys9) and the second cysteine (Cys15); the second segment is the sequence between the third (Cys21) and the fourth (Cys27). These two segments are important structural regions (3_{10} helix and type I beta-turn, respectively), and are known to involve in the folding initiation or specification of the CSB motif (Heitz et al., 1995, 1999; Wentzel et al., 1999). We are particularly interested in these two regions because our design of antibody mimetics based on the Min-23 scaffold required that the sequence in these two regions to be replaced by hypervariable loops mimicking the CDRs in the antibody—we needed to know which residues in these regions are structural determinants that are irreplaceable if the CSB fold is to be retained.

The limitation of the current experimental design is that the structure-wise stable variants with sequence variegation between Cys9 and Cys15 were evolved in the context of the wild-type sequence between Cys21 and Cys27. Similarly, the stabilizing

Table 3. Phage-Displayed Sequences of Randomly Picked Binders against Human VEGF-A from pC-8X Library

8x-2A	LMRC**RS**D**A**DCIEGRVC**SYFSQ**GFCG
8x-2C	LMRC**TS**D**L**DCIEGRVC**AAWLS**GFCG
8x-2D	LMRC**GS**D**S**DCIEGRVC**PPSRS**GFCG
8x-5D	LMRC**LS**D**P**DCIEGRVC**PSLSL**GFCG
8x-7E	LMRC**LF**D**L**DCIEGRVC**SFWPR**GFCG

The bold characters represent the variable regions in phage-displayed protein library design.

sequences between Cys21 and Cys27 were evolved in the context of the wild-type sequence between Cys9 and Cys15. Thus, the resultant stabilizing sequences have been expected to vary around the local sequence space optimum corresponding to the wild-type sequence of Min-23. This expectation was confirmed by comparing the sequence profiles shown in Figure 6 and Figure S1. The consistency of the two sets of sequence profiles have supported the validity of the artificial molecular evolution methodology presented in this work. Nevertheless, experimental designs with co-evolution of both sequence segments would provide more complete sequence-structure relationships outside the wild-type local optimum in the sequence space. But such experiments would need phage display libraries to cover at least 20^{10} complexity, which is about four orders of magnitude beyond current complexity limitation in phage-displayed protein libraries.

The combinatorial exploration for structural determinants in these two regions revealed that Asp12 and Asp14 between Cys9 and Cys15 were consensus residues in the cystine-stabilized structures. Gly25, and Phe26 to a lesser degree, were consensus residues between Cys21 and Cys27 in the cystine-stabilized structures. In the former region, the Asp12 side-chain carboxyl group is hydrogen bonded with the backbone NH group of Cys21 and the Asp14 side-chain carboxyl group is hydrogen bonded with the backbone NH group of Lys10 according to the X-ray structure of EETI-II (Kratzner et al., 2005). Moreover, the Asp14 has high propensity to adopt a type I turn as needed in the local structure. In the latter region, Gly25 (backbone torsion angle in the α_L region) is a critical residue for the local tight turn structure and Phe26 involves in tertiary interactions with Leu7 and Arg9. Other than these consensus residues, the rest of the residues in the 3_{10} helix structure in the former region and the type I tight turn of the latter region were permissive and could be because the torsion angle requirements for these structures can be accommodated by all amino acids. The structure of NL3 (Figure 7), however, indicated that sequences without the consensus sequence features could still form the native structures, although sequences matched with the sequence features above were preferred in folding into stable native structure.

The structure determinant information above was derived from large scale experiments on stability determination. Large scale experimental methods in elucidating structure-stabilizing and fold-specifying residues require a selection method to isolate folded structures from unfolded ones and a high-throughput measurement for the stability of each of the monoclonal proteins. Experimental methods based on enzyme digestion or binding specificity have been developed to select for stable structures from protein display systems (see reviews Kotz et al., 2004; Magliery and Regan, 2004). But high-throughput measurements for the stabilities of the displayed proteins have received less attention, perhaps because of experimental limitations in high-throughput protein expression and in high-throughput protein stability measurement with conventional experimental techniques (Aucamp et al., 2005; Bommarius et al., 2006). The intact fraction measurement described in this work provides a general strategy for high-throughput stability measurements of cystine-stabilized protein structures. The correlation of the sequence features and the structural stability indicated that the intact fraction measurements closely reflected the stabilities of the phage-displayed proteins.

This high-throughput methodology is subject to the concentration limit of the denaturants tolerated by the recombinant phage (2 M GdnHCl and 10mM DTT). Structural stability above this denaturant threshold cannot be discriminated with this method. Hence, the most effective application of the method is to discriminate disulfide-linked structures with well-buried disulfide bonds from protein structures with surface disulfide bonds found in scrambled cysteine-rich structures. This is particularly feasible in small CSB structures. In the Mx2-based combinatorial libraries, three two-disulfide bonding configurations are possible: (9-21,15-27), (9-15,21-27), and (9-27,15-21). With fXa cutting site situated between Cys15 and Cys21, the (9-15,21-27) species is cleavable by fXa without the disulfide bond reduction, so that these species should belong to the (−−−) class. The (9-27,15-21) species have both disulfide bonds exposed to solvent due to lacking the interlocking cystine structure and the small size of the miniprotein. This conclusion was derived from a molecular simulation of the model structures, for which the simulation details and results are shown in the Figure S2. The simulation indicated that the (9-27,15-21) conformation was less capable of sequestering the disulfide bonds from the water environment than the native NMR structure of Mx2. The (9-27,15-21) species is cleavable by fXa after the DTT reduction and NEM blocking, and should belong to the (+−−) class. Only the native configuration (9-21,15-27) is resistant to DTT/NEM/fXa treatment and can be unfolded in the presence of both GdnHCl and DTT. These species should belong to either (++−) or (+++) class, as demonstrated in Mx2 and NL3 molecules by the structure determination (Figures 4 and 7) and the intact fraction measurements at various denaturant conditions (Figure 3A and Table S2).

The selection/amplification cycles based on the same principle as in the intact fraction measurement (Figure 1) were effective in enriching variants that were resistant to increasingly stringent denaturant conditions. In combination with the high-throughput intact fraction measurements for monoclonal protein stability analysis, the artificial evolution procedure generated sequence profiles for the variants clustered according to the structural stability against denaturants and fXa digestion (Figure 6). Unlike sequence profiles derived from homologous proteins that have been evolved from single or multiple ancestors to accomplish structural stability in a natural environment, these sequence profiles revealed the emergence of the sequence patterns as the variant structures becoming more accommodating to artificial stability conditions. Comparing the emerged consensus sequence profiles from directed evolution (Figure 6, class [++−] and [+++]) with the sequence profile from natural homologous sequences (Figure S1), we found that the sequence patterns in the corresponding areas were strikingly similar. The consistency has confirmed the underlying hypothesis for the consensus-based protein engineering approach, where amino acid residues specifying the fold and stabilizing the structure occur in higher frequency in the evolution-derived sequence profiles. More importantly, the artificial evolution methodology can be applied to artificially designed proteins for which the homologous protein sequences are not available in nature. In this work, we initiated a methodology to design small cystine-stabilized proteins without the need to start with a known sequence in nature. With

the methodology, we will be able to first computationally design the basic requirements in the sequences for designated structural details and then artificially evolve the variants to find the sequence requirements that have or have not been attained in the initial in silico design of a stable structure.

The sequence requirements for a stable structure are prerequisites for designing protein binders with paratopes composed of nonsequential protein segments. This type of conformational paratope is known to confer the combinatorial power to antibody-antigen recognitions. In designing the two-loop variable surfaces (Figure 7C) on the Mx2 scaffold, the structural determinant residues discussed above remained invariable, and indeed the variants designed as such are benefited in structural stability from these key residues (Figure 3A). The structural stability in turn increases the binding affinity in tentative protein-recognition complexes, and thus we anticipated to identify binders with reasonable binding affinity/specificity (Figure 3B) and structure stability (Figure 3C).

VEGF-A antagonists have potential therapeutic applications. The stabilization of known peptide VEGF-A epitope has recently been accomplished by Gunasekera et al. (2008) through grafting VEGF-A binding peptide onto a plant-derived cyclotide scaffold with three conserved disulfide bonds, which confer the extraordinary stability of the cyclotide. We demonstrated in this work that alternative potential VEGF-A binders can be derived on the basis of a cystine-stabilized miniprotein scaffold from a phage-displayed protein library, where stabilizing disulfide bonding configuration has been specifically designed and selected to form a stable structure with required binding specificity.

In summary, with the fXa cutting site embedded in the sequences of the Mx2-based libraries, molecular evolution with both binding-selection against the target protein and stability-selection against fXa digestion after denaturant treatment enabled a protein engineering procedure simultaneously enhancing recognition specificity and structural stability for the cystine-stabilized binders. This methodology can be generalized to engineering artificial minisequences initially designed in silico.

EXPERIMENTAL PROCEDURES

Phage Display Library Construction

pCANTAB5E (GE-Amersham) phagemids for recombinant M13KE phage particles displaying various polypeptide sequences and libraries on the minor capsid pIII protein were constructed as shown in Figure 2. The phagemid templates with the TAA stop codons in the sites for amino acid randomization (i.e., pC-Mx2_TAA1, pC-Mx2_TAA3, pC-nX_TAA in Figure 2) were designed to ensure that only the library phagemids with designated degenerate codon replacement would produce pIII fusion proteins for phage surface display. The phage display libraries were constructed on the basis of these TAA-templates with the oligonucleotide-directed mutagenesis method initially proposed by Kunkel et al. (1987). In this work, we followed the protocol of Sidhu and Weiss (2004). Details of the methodology have been described in Hsu et al. (2008). Sequences for the DNA primers are listed in Supplemental Experimental Procedures. The complexities of the libraries were 4.4×10^8, 5.0×10^8, 2.4×10^9, 7.0×10^8, and 1.5×10^9 for pC-Lib1, pC-Lib3, pC-6X, pC-7X, and pC-8X, respectively.

Molecular Evolution of Cystine-Stabilized Proteins

The 10 ml rescued recombinant phage particles ($\sim 10^{11}$ cfu/ml) were precipitated with PEG/NaCl (20% PEG 8000 and 2.5 M NaCl) and then resuspended in 100 μl phosphate buffer (50 mM NaH_2PO_4, 300 mM NaCl [pH 8.0]) with concentration of GdnHCl ranging from 0 M to 2 M. Dithiothreitol was added to the phage solution to make final concentration of 10 mM for 20 min, followed by adding NEM to the solution to make final concentration of 80 mM for 30 min. Both reactions were carried out at room temperature. GdnHCl and DTT were used to unfold the polypeptide chain and to reduce disulfide bonds in the displayed proteins; reaction of excessive NEM with the thiol group in free cysteines blocked the reformation of disulfide bonds in the displayed proteins. After the reaction, the phage particles were precipitated with 20% PEG/NaCl and resuspended in 60 μl bovine fXa reaction buffer (20 mM Tris-HCl [pH 6.8], 50 mM NaCl, 1 mM $CaCl_2$) containing 2 units fXa (QIAGEN) for 1 hr at room temperature. More than 90% of the cleavable sites engineered in the displayed proteins that were unfolded by the GdnHCl/DTT/NEM treatment were cleaved and the hexa-His tags were removed from the phage particles. The well-folded proteins with protected disulfide bonds remained intact with the hexa-His tag and the intact recombinant phages were isolated through binding to Ni-NTA beads. This was done by adding the reaction mixture to 1 ml of Ni-NTA binding buffer (50 mM NaH_2PO_4, 300 mM NaCl [pH 8.0]) containing 50 μl Ni-NTA agarose beads (QIAGEN) for 1 hr. The Ni-NTA beads were transferred to a micro filter (Millipore Ultrafree MC, UFC30SV00) and washed 21 times, each with 500 μl washing buffer (50 mM NaH_2PO_4, 300 mM NaCl [pH 8.0], 10 mM imidazole). The Ni-NTA bound phage particles were then eluted from the Ni-NTA beads with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, [pH 8.0], 200 mM imidazole). The eluted phage particles were amplified overnight and were used for the next round of selection.

The above selection protocol is not suitable for GdnHCl concentration higher than 3 M because few phage particles remained infective after the GdnHCl/DTT/NEM treatment above the 3 M GdnHCl threshold concentration. We found that immobilized TCEP (Tris(2-carboxyethyl)phosphine on agarose bead, Pierce product #77712) reduced disulfide bonds to less extent than 10 mM DTT but had much less adverse effect on phage infectivity in the presence of a high concentration of GdnHCl up to 6 M. The GdnHCl/TCEP-gel/NEM-based selection was different from the GdnHCl/DTT/NEM-based selection described above in only one step: instead of adding DTT followed by NEM to the phage solution, we added 100 μl TCEP gel to 100 μl phage solution (100 mM Tris-acetate, 10 ml EDTA [pH 8]) in the presence of GdnHCl (3∼6 M) and 80 mM NEM. The reaction mixture was mildly rocked for 1 hr at room temperature before fXa digestion as described above.

Intact Fraction Measurement

Mid-log *E. coli* ER2738 infected with the selected phage from the selection procedure was spread on 2YT/ampicillin(100 μg/ml) agarose plate. Overnight colonies were picked with GENETIX Qpix II colony picker to 96-well deep well culture plate. Each well contained 1 ml 2YT, ampicillin 100 μg, and 10^8 cfu helper phage M13KO7. One hour after inoculation, 200 μl 2YT with 300 μg/ml kanamycin and 600 μg/ml ampicillin were added to each well. The culture plates were incubated in 37°C with vigorous shaking overnight. The overnight culture was centrifuged at 3000 g at 4°C for 30 min. Then 100 μl supernatant from each well was added to the corresponding well of a 96-well His-sorb plate (QIAGEN). After 1 hr of binding at room temperature, the His-sorb plates were washed with 3 × 250 μl PBST (phosphate-buffered saline [pH 7.5], Tween 20 0.05% washed three times, each time with 250 μl) and 1 × 250 μl PBS. Anti-M13-HRP (1:5000, GE-Amersham) in PBS + skim milk (2.5%) was added to each well for 1 hr. Clear the anti-M13-HRP from the His-sorb plates with 3 × 250 μl PBST and 1 × 250 μl PBS washing buffer. Then 100 μl ABTS substrate (BioFix) was added to each well for 10 min before adding 1% (final concentration) SDS to terminate the chromogenic reaction. The optical density (OD) of each well was measured at 405 nm. Phage species with OD_{405nm} greater than 0.5 were selected for measuring the extent of unfolding after the GdnHCl/DTT/NEM treatment (see below). If the OD_{405nm} was greater than 2, dilution was necessary so that the quantitative ELISA readings (see below) remained linearly related to the concentration of the phage particles with intact hexa-His tag.

For each of the selected phage solutions from the deep well culture plate, the phage solution was centrifuged at 8100 g in a microcentrifuge tube. Then 200 μl of the supernatant was transferred to a centrifugal filter (Millipore microcon YM-100, molecular weight cut-off 100 kDa) and spun at 6600 g for 5 min to remove phage culture medium. All recombinant phage particles were expected to remain in the filter. The filter was washed with 200 μl phosphate buffer (50 mM NaH_2PO_4, 300 mM NaCl [pH 8.0]). After removing the wash buffer by centrifuge (6600 g, 5 min), 100 μl fXa reaction buffer containing GdnHCl (0∼2M) and 10 mM

DTT was added to the filter for 20 min, followed by adding NEM (final concentration 80 mM) to the filter for 30 min. The reaction mixture was removed from phage particles by centrifuge and by washing the filter two times, each with 200 μl fXa reaction buffer. The phage particles were then transferred from the filter to a micro tube by adding 35 μl fXa reaction buffer to the filter and collecting the solution by centrifuging (2000 g, 2 min) the upside-down filter in the collection tube. Then 10 μl of the collected phage solution was added to a well of a nonabsorbing plate containing 90 μl fXa reaction buffer and 0.2 U fXa. In a second well of the same plate, 10 μl of the collected phage solution was added to 90 μl fXa reaction buffer without fXa. After 2 hr reaction at room temperature, 50 μl of 1.5 M GdnHCl was added to all wells to terminate the enzymatic reaction. Then 100 μl solution was transferred from each well to a corresponding well in a His-sorb plate and the ELISA procedure was performed as described above. The intact fraction was calculated with the OD_{405nm} from the sample with fXa over the OD_{405nm} from the sample without fXa.

The library characterization procedure is similar to the single-colony analysis: the amplified phage library was precipitated with PEG/NaCl and then treated with GdnHCl/DTT(or TCEP-gel)/NEM as described above. After the denaturation reaction, the phage particles were isolated from the reaction mixture with PEG/NaCl precipitation and were partitioned into two equal portions, one treated with fXa and the other added only buffer. The intact fraction of the phage particles after the GdnHCl/DTT (or TCEP-gel)/NEM/fXa treatment was evaluated with the ELISA procedure described above.

Biopanning of Phage Library to VEGF-Conjugated Bead

After PEG/NaCl precipitation, phages were dissolved in PBST milk ($\sim 10^{12}$ cfu/ml) and incubated at room temperature for 30 min. Nonspecificity binding phages were removed first by incubating the phages with ethanol-amine blocked Ultralink Biosupport bead (Pierce) at 4°C for 2 hr. The supernatant was transferred to the VEGF-A conjugated Ultralink Biosupport bead and incubated at 4°C for 2 hr. The expression and purification of human VEGF-A are described below. The beads were washed 10 × 1 ml PBST (0.05% Tween 20 in PBS buffer) and 4 × 1 ml PBS. The VEGF binding phages were then eluted with glycine buffer (0.2 M [pH 2.0]). The eluted phages were neutralized with Tris buffer (1 M [pH 9.1]) and then added into the fresh *E. coli* host ER2738 for reamplification.

Peptide Synthesis, Purification, and Refolding

Peptides were synthesized with automated solid-phase 9-fluorenymethoxycarbonyl peptide synthesizer. Lyophilized crude synthetic peptide was dissolved in 10 ml reducing buffer (100 mM Tris, 6 M GdnHCl, 100 mM DTT, 1 mM EDTA [pH 8.0]) at room temperature for 2 hr. The solution was titrated with acetic acid to pH 3 and injected into reverse-phase HPLC to purify the fully reduced polypeptide, which was identified with mass spectrometry. The acetonitrile in the collected fraction was removed by lyophilization, and the polypeptide was redissolved in refolding buffer (100 mM Tris, 1 mM EDTA, 0.5 M GdnHCl, 10 mM reduced glutathione, 1 mM oxidized glutathione [pH 8.5]) overnight at 4°C. The refolded polypeptide was isolated with reverse-phase HPLC. The fraction containing the fully refolded polypeptide with two pairs of disulfide bonds was identified by mass spectrometry. The actual disulfide bonding configuration and the 3D structures were determined with 2D NMR.

Synthetic Peptide Structure Determination with 2D NMR

The purified refolded polypeptide as described above was dissolved in 90% H_2O and 10% D_2O at pH 3 to final concentration of 1 mM. All NMR experiments were carried out at 285 K or 298 K on the Bruker Avance 600MHz spectrometer. Assignment of 1H chemical shifts were based on 2D DQF-correlated spectroscopy (COSY), total correlation spectroscopy (TOCSY), and nuclear Overhauser enhancement spectroscopy (NOESY) experiments where water suppression was achieved using 3-9-19 pulse sequence with gradients. TOCSY spin lock was 70 ms, and NOESY spectra were collected with mixing time of 150 ms and 300 ms. All data were acquired and processed using TOPSPIN (Bruker, Germany) and further analyzed using Sparky software package (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

Structure calculation was done using CYANA (Guntert, 2004) where simulated annealing protocol with torsion angle molecular dynamics was applied. Semiautomated NOE cross-peak assignments were performed using CANDID module in the CYANA software and manually checked for correctness in an iterative manner. The $^3J_{NH\alpha}$ coupling constants estimated from DQF-COSY spectra were converted to backbone ϕ angles. These ϕ angles were constrained in the range of $-110 \pm 30°$ for residues with $^3J_{NH\alpha}$ coupling constants larger than 9 Hz. Disulfide bond constraints were based on NOE observation between Hβ of Cys residues, and the distance upper limit between sulfur atoms of disulfide bridges was set to 2.1 Å. One hundred random conformers were annealed in 10,000 steps for seven cycles and 20 structures with the lowest target function after the final cycle were selected.

VEGF Expression and Purification

Human VEGF-A DNA encoding amino acid sequence 34–135 was synthesized in fragments and jointed with recursive polymerase chain reaction. The synthesized gene was constructed into pET15b vector and transformed to *E. coli* strain BL21. A single colony was picked and cultured in 3 ml Luria-Bertani Broth (LB) / ampicillin (100 ug/ml) medium at 37°C overnight. 1ml overnight culture was poured into 1 l LB/ampicillin medium and incubated at 37°C for 3–4 hr until the OD_{595} reached to 0.6∼0.8. The cells were inducted by 1 mM IPTG at 37°C for 3–4 hr and harvested by centrifugation at 4500 g for 20 min. The lysed cell was centrifuged and the pellet was dissolved in denaturing buffer (20 mM Tris, 8 M urea, 20 mM DTT [pH 7.5]). The protein solution was dialyzed against the buffer (20 mM Tris, 2 M urea, 1 mM cysteine [pH 8.]) for 6 hr, and then against the buffer (20 mM Tris, 400 mM NaCl [pH 8.0]) overnight. The refolded protein was purified with Ni-NTA resin (QIAGEN). The purified protein solution was dialyzed against PBS buffer for 6 hr and exposed to open air for disulfide mediated dimerization.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, Supplemental References, two figures, and three tables and can be found with this article online at http://www.cell.com/structure/supplemental/S0969-2126(09)00076-8.

ACKNOWLEDGMENTS

A.S.Y. would like to acknowledge the financial support from National Health Research Institutes grant NHRI-EX95-9525EI, National Science Council grant NSC 96-2311-B-001-030-MY3, and support from Genomics Research Center at Academia Sinica Taipei Taiwan.

Received: September 5, 2008
Revised: January 16, 2009
Accepted: January 16, 2009
Published: April 14, 2009

REFERENCES

Aucamp, J.P., Cosme, A.M., Lye, G.J., and Dalby, P.A. (2005). High-throughput measurement of protein stability in microtiter plates. Biotechnol. Bioeng. *89*, 599–607.

Bergers, G., and Hanahan, D. (2008). Modes of resistance to anti-angiogenic therapy. Nat. Rev. Cancer *8*, 592–603.

Bommarius, A.S., Broering, J.M., Chaparro-Riggers, J.F., and Polizzi, K.M. (2006). High-throughput screening for enhanced protein stability. Curr. Opin. Biotechnol. *17*, 606–610.

Craik, D.J., Cemazar, M., Wang, C.K., and Daly, N.L. (2006). The cyclotide family of circular miniproteins: nature's combinatorial peptide template. Biopolymers *84*, 250–266.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome Res. *14*, 1188–1190.

Davidson, A.R. (2006). Multiple sequence alignment as a guideline for protein engineering strategies. Methods Mol. Biol. *340*, 171–181.

Folkman, J. (2007). Angiogenesis: an organizing principle for drug discovery? Nat. Rev. Drug Discov. *6*, 273–286.

Fuh, G., Wu, P., Liang, W.C., Ultsch, M., Lee, C.V., Moffat, B., and Wiesmann, C. (2006). Structure-function studies of two synthetic anti-vascular endothelial

growth factor Fabs and comparison with the Avastin Fab. J. Biol. Chem. *281*, 6625–6631.

Gatto, B., and Cavalli, M. (2006). From proteins to nucleic acid-based drugs: the role of biotech in anti-VEGF therapy. Anticancer. Agents Med. Chem. *6*, 287–301.

Gracy, J., Le-Nguyen, D., Gelly, J.C., Kaas, Q., Heitz, A., and Chiche, L. (2008). KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucleic Acids Res. *36*, D314–D319.

Gunasekera, S., Foley, F.M., Clark, R.J., Sando, L., Fabri, L.J., Craik, D.J., and Daly, N.L. (2008). Engineering stabilized vascular endothelial growth factor-A antagonists: synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides. J. Med. Chem. *51*, 7697–7704.

Guntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol. Biol. *278*, 353–378.

Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1995). Folding of the squash trypsin inhibitor EETI II. Evidence of native and non-native local structural preferences in a linear analogue. Eur. J. Biochem. *233*, 837–846.

Heitz, A., Le-Nguyen, D., and Chiche, L. (1999). Min-21 and min-23, the smallest peptides that fold like a cystine-stabilized beta-sheet motif: design, solution structure, and thermal stability. Biochemistry *38*, 10615–10625.

Hsu, H.J., Chang, H.J., Peng, H.P., Huang, S.S., Lin, M.Y., and Yang, A.S. (2006). Assessing computational amino acid beta-turn propensities with a phage-displayed combinatorial library and directed evolution. Structure *14*, 1499–1510.

Hsu, H.J., Tsai, K.C., Sun, Y.K., Chang, H.J., Huang, Y.J., Yu, H.M., Lin, C.H., Mao, S.S., and Yang, A.S. (2008). Factor Xa active site substrate specificity with substrate phage display and computational molecular modeling. J. Biol. Chem. *283*, 12343–12353.

Kajander, T., Cortajarena, A.L., and Regan, L. (2006). Consensus design as a tool for engineering repeat proteins. Methods Mol. Biol. *340*, 151–170.

Kotz, J.D., Bond, C.J., and Cochran, A.G. (2004). Phage-display as a tool for quantifying protein stability determinants. Eur. J. Biochem. *271*, 1623–1629.

Kratzner, R., Debreczeni, J.E., Pape, T., Schneider, T.R., Wentzel, A., Kolmar, H., Sheldrick, G.M., and Uson, I. (2005). Structure of Ecballium elaterium trypsin inhibitor II (EETI-II): a rigid molecular scaffold. Acta Crystallogr. D Biol. Crystallogr. *61*, 1255–1262.

Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. *154*, 367–382.

Ladner, R.C., Sato, A.K., Gorzelany, J., and de Souza, M. (2004). Phage display-derived peptides as therapeutic alternatives to antibodies. Drug Discov. Today *9*, 525–529.

Le-Nguyen, D., Heitz, A., Chiche, L., el Hajji, M., and Castro, B. (1993). Characterization and 2D NMR study of the stable [9-21, 15-27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor EETI II. Protein Sci. *2*, 165–174.

Lowman, H.B. (1997). Bacteriophage display and discovery of peptide leads for drug development. Annu. Rev. Biophys. Biomol. Struct. *26*, 401–424.

Magliery, T.J., and Regan, L. (2004). Combinatorial approaches to protein stability and structure. Eur. J. Biochem. *271*, 1595–1608.

Michnick, S.W., and Sidhu, S.S. (2008). Submitting antibodies to binding arbitration. Nat. Chem. Biol. *4*, 326–329.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Peng, H.P., and Yang, A.S. (2007). Modeling protein loops with knowledge-based prediction of sequence-structure alignment. Bioinformatics *23*, 2836–2842.

Sidhu, S.S., and Weiss, G.A. (2004). Constructiong phage display libraries by oligonucleotide-directed mutagenesis. In Phage Display, First Edition, T. Clackson and H.B. Lowman, eds. (New York: Oxford University Press).

Sidhu, S.S., Fairbrother, W.J., and Deshayes, K. (2003). Exploring protein-protein interactions with phage display. ChemBioChem *4*, 14–25.

Sieber, V., Pluckthun, A., and Schmid, F.X. (1998). Selecting proteins with improved stability by a phage-based method. Nat. Biotechnol. *16*, 955–960.

Silverman, J., Liu, Q., Bakker, A., To, W., Duguay, A., Alba, B.M., Smith, R., Rivas, A., Li, P., Le, H., et al. (2005). Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. Nat. Biotechnol. *23*, 1556–1561.

Skerra, A. (2007). Alternative non-antibody scaffolds for molecular recognition. Curr. Opin. Biotechnol. *18*, 295–304.

Smith, G.P., Patel, S.U., Windass, J.D., Thornton, J.M., Winter, G., and Griffiths, A.D. (1998). Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. J. Mol. Biol. *277*, 317–332.

Socolich, M., Lockless, S.W., Russ, W.P., Lee, H., Gardner, K.H., and Ranganathan, R. (2005). Evolutionary information for specifying a protein fold. Nature *437*, 512–518.

Souriau, C., Chiche, L., Irving, R., and Hudson, P. (2005). New binding specificities derived from Min-23, a small cystine-stabilized peptidic scaffold. Biochemistry *44*, 7143–7155.

Wentzel, A., Christmann, A., Kratzner, R., and Kolmar, H. (1999). Sequence requirements of the GPNG beta-turn of the Ecballium elaterium trypsin inhibitor II explored by combinatorial library screening. J. Biol. Chem. *274*, 21037–21043.

Wodak, S.J., and Janin, J. (2002). Structural basis of macromolecular recognition. Adv. Protein Chem. *61*, 9–73.

Yang, A.S., and Wang, L. (2003). Local structure prediction with local structure-based sequence profiles. Bioinformatics *19*, 1267–1274.