doi:10.1016/j.jmb.2011.02.002 J. Mol. Biol. (2011) 407, 744–763

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

A Polypeptide "Building Block" for the β-Trefoil Fold Identified by "Top-Down Symmetric Deconstruction"

Jihun Lee[1], Sachiko I. Blaber[1], Vikash K. Dubey[2] and Michael Blaber[1]*

[1]*Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL 32306-4300, USA*
[2]*Department of Biotechnology, Indian Institute of Technology, Guwahati 781039, India*

Received 22 December 2010; received in revised form 31 January 2011; accepted 2 February 2011
Available online 16 February 2011

Edited by I. Wilson

Keywords:
protein design;
protein symmetry;
protein evolution;
top-down symmetric deconstruction;
β-trefoil

Fibroblast growth factor-1, a member of the 3-fold symmetric β-trefoil fold, was subjected to a series of symmetric constraint mutations in a process termed "top-down symmetric deconstruction." The mutations enforced a cumulative exact 3-fold symmetry upon symmetrically equivalent positions within the protein and were combined with a stability screen. This process culminated in a β-trefoil protein with exact 3-fold primary-structure symmetry that exhibited excellent folding and stability properties. Subsequent fragmentation of the repeating primary-structure motif yielded a 42-residue polypeptide capable of spontaneous assembly as a homotrimer, producing a thermostable β-trefoil architecture. The results show that despite pronounced reduction in sequence complexity, pure symmetry in the design of a foldable, thermostable β-trefoil fold is possible. The top-down symmetric deconstruction approach provides a novel alternative means to successfully identify a useful polypeptide "building block" for subsequent "bottom-up" *de novo* design of target protein architecture.

Introduction

A long-standing goal in *de novo* protein design has been the exploitation of a hierarchical design strategy, utilizing appropriately designed peptide "building blocks," to spontaneously assemble (via oligomerization or concatenation) the desired target architecture.[1–13] Such hierarchical design strategies are "bottom-up" in that polypeptides are designed from first principles to have folding and thermodynamic properties that promote correct assembly of the target structure. Symmetric or periodic protein architecture is favored for such *de novo* design and offers a number of potential advantages. A symmetric design constraint substantially reduces the conformational search in design algorithms and can simplify folding simulations, thereby substantially accelerating the design calculations.[7,12,14] Elements of symmetry that result in efficient structural compaction during folding likely contribute to an efficient funneled energy landscape of folding.[15] Structural symmetry can also result in multiple folding nuclei with an associated redundancy within the folding pathway.[16] However, there are also significant unresolved questions regarding the practical limitations of symmetric protein design. For example, exact primary-structure symmetry within a symmetric architecture involves a substantial reduction in sequence complexity—one hallmark of natively unstructured proteins.[17,18] Exact primary-structure symmetry within repeated domains provides opportunities for domain mismatches producing misfolded forms with near-native Gibbs energy, and low sequence identities could have a crucial and general role in safeguarding proteins against misfolding and aggregation;[19] furthermore,

*Corresponding author. E-mail address: michael.blaber@med.fsu.edu.

Present address: J. Lee, 12441 Parklawn Drive, Room 108, NIAID, NIH, Rockville, MD 20852, USA.

Abbreviations used: FGF-1, fibroblast growth factor-1; SD, symmetric deconstruction; SPR, surface plasmon resonance; FGFR-1c, FGF receptor-1c; FGFR, FGF receptor; GuHCl, guanidine HCl; DSC, differential scanning calorimetry; ITC, isothermal titration calorimetry; $(His)_6$, hexahistidine.

primary-structure symmetry is one feature of amyloid-type aggregates.[20] Thus, while symmetric protein architecture offers attractive advantages for *de novo* design, there is a need for novel approaches to successfully identify foldable peptide building blocks from those that might otherwise misfold or aggregate.

We have previously described the structural properties of subdomain fragments of a symmetric polypeptide derived from human fibroblast growth factor-1 [FGF-1, a protein belonging to the 3-fold symmetric β-trefoil fold; Protein Data Bank (PDB) code 2AFG] via a novel "top-down symmetric deconstruction (SD)" approach.[21] This prior work provided experimental support for a "conserved architecture" model of evolution for the β-trefoil fold via gene duplication and fusion processes; however, it did not provide a detailed description of the deconstruction process. Here, we describe in detail the top-down SD method that produced the symmetric β-trefoil polypeptide, detailing the function, folding, and stability properties of intermediary forms between the FGF-1 starting protein and the final symmetric polypeptide, as well as structural details of an important intermediary mutant. The results describe a loss of FGF-1-specific function relatively early in the deconstruction process, while "function-competent" properties of folding and stability are maintained (or enhanced). The end result of the top-down SD is a simple 42-amino-acid peptide building block that can subsequently be utilized in hierarchical bottom-up *de novo* protein design to

Protein	Sequence	ΔΔG (kJ/mol)	*m*-value (kJ/molM)
FGF-1	PKLLYCSNGG---HFLRILPDGTVDG--TRDRSDQHIQLQLSAESVG EVYIKSTETG---QYLAMDTDGLLYG---SQTPNEECLFLERLEENH YNTYISKKHAEKNWFVGLKKNGSCKRGPRTHYGQKAILFLPLPVSSD	-	18.8
SYM2	PKLLYCSNGG---HFLRILPDGTVDG--TRDRSDQHIQLQLSAESVG EVYIKSTETG---QYLAMDTDGLVYG---SQTPNEECLFLERLEENH YNTYISKKHAEKNWFLGLKKNGSCKRGPRTHYGQKAILFLPLPVSSD	5.3	20.6
SYM3	PKLLYCSNGG---HFLRILPDGTVDG--TRDRSDQHIQFQLSAESVG EVYIKSTETG---QYLAMDTDGLVYG---SQTPNEECLFLERLEENH YNTYISKKHAEKNWFLGLKKNGSCKRGPRTHYGQKAILFLPLPVSSD	3.2	19.1
SYM4	PKLLYCSNGG---HFLRILPDGTVDG--TRDRSDQHIQFQLSAESVG EVYIKSTETG---QYLAMDTDGLVYG---SQTPNEECLFLERLEENH YNTYISKKHAEKNWFLGLKKNGSVKRGPRTHYGQKAILFLPLPVSSD	1.9	19.2
SYM5	PKLLYCSNGG---HFLRILPDGTVDG--TRDRSDQHIQFQLSAESVG EVYIKSTETG---QYLAMDTDGLVYG---SQTPNEECLFLERLEENH YNTYISKKHAEKNWFLGIKKNGSVKRGPRTHYGQKAILFLPLPVSSD	0.8	19.4
SYM6	PKLLYCSNGG---HFLRILPDGTVDG--TRDRSDQHIQFQLSAESVG EVYIKSTETG---QYLAIDTDGLVYG---SQTPNEECLFLERLEENH YNTYISKKHAEKNWFLGIKKNGSVKRGPRTHYGQKAILFLPLPVSSD	10.9	19.4
SYM6ΔΔ	PKLLYCSNGGHFLRILPDGTVDGTRDRSDQHIQFQLSAESVG EVYIKSTETGQYLAIDTDGLVYGS-QTPNEECLFLERLEENH YNTYISKKHGWFLGIKKNGSVKQT-HYGQKAILFLPLPVSSD	-16.1	17.7
SYM7ΔΔ	PKLLYCSNGGHYLRILPDGTVDGTRDRSDQHIQFQLSAESVG EVYIKSTETGQYLAIDTDGLVYGS-QTPNEECLFLERLEENH YNTYISKKHGWYLGIKKNGSVKQT-HYGQKAILFLPLPVSSD	-16.8	ND
SYM8ΔΔ	PKLLYCSNGGHYLRILPDGTVDGTRDRSDQHCQFQLSAESVG EVYIKSTETGQYLAIDTDGLVYGS-QTPNEECLFLERLEENH YNTYISKKHGWYLGIKKNGSVKQT-HYGQKACLFLPLPVSSD	-17.1	ND

Fig. 1. SD transform #1: introduction of a symmetric constraint on the hydrophobic core region of FGF-1. Gray shading indicates positions related by 3-fold structural symmetry where either two or all three residues are identical. Boxed positions indicate the specific site(s) of mutation in construction of that particular mutant. $\Delta\Delta G$ values are a measure of the stability effect (referencing FGF-1 and determined in ADA buffer), where a negative value indicates an increase in overall thermostability. Thermodynamic data for FGF-1 are from Blaber *et al.*,[22] data for SYM2–SYM6 are from Brych *et al.*,[23,24] data for SYM6ΔΔ are from Brych *et al.*,[25] and $\Delta\Delta G$ data (determined in crystallization buffer) for SYM7ΔΔ and SYM8ΔΔ are from Dubey *et al.*[26]

generate a foldable, thermostable β-trefoil fold (either as a threefold repeat sequence within a single polypeptide or as a noncovalent homotrimer assembly). Top-down SD therefore provides a novel alternative strategy in the successful development of peptide building blocks for subsequent bottom-up hierarchical protein design of symmetric protein architecture. In the case of the β-trefoil fold, a purely symmetric architecture is shown to be a satisfactory solution to the design of a soluble, foldable, thermostable protein architecture.

Results

Transform #1: Top-down SD of core positions

Attempts to introduce a symmetric constraint on the hydrophobic core-packing group, without also making alterations to the asymmetric FGF-1 tertiary structure, met with limited success. An alternative core-packing group involving five residue positions (SYM5 mutant, Fig. 1) was identified that was essentially equivalent to the FGF-1 thermostability and folding cooperativity.[23,24] Folding and unfolding kinetic data showed that the SYM5 mutant exhibited folding and unfolding rates similar to that of the FGF-1 protein including the characteristic biphasic folding behavior.[24] Attempts at introducing a further symmetric constraint (Ile) at symmetry-related positions 25, 67, and 111 (SYM6 mutant) essentially failed due to substantial (10.9 kJ/mol) destabilization. However, subsequent deletion mutations in the third trefoil-fold subdomain, which increased the tertiary-structure symmetry, enabled the symmetric Ile solution at positions 25, 67, and 111 (SYM6ΔΔ mutant, Fig. 2) with a substantial (16.1 kJ/mol) increase in stability in comparison to FGF-1.[24] This increase in stability was due exclusively to an increase in the folding kinetics; additionally, the SYM6ΔΔ mutant exhibited single-exponential folding behavior (unlike FGF-1 and SYM2–SYM6 mutants).[25] Construction of the SYM6ΔΔ mutant also coincided with essential loss of affinity for heparin–Sepharose and an order of magnitude increase in K_D for sucrose octasulfate (a disaccharide mimic of heparin).[25]

Subsequent symmetric mutations SYM7ΔΔ and SYM8ΔΔ addressed additional buried positions (separate from the central hydrophobic core). Symmetry-related positions Ile42, Cys83, and Ile130 were initially constrained to Ile residues (by mutation Cys83Ile); however, this failed due to substantial (9.9 kJ/mol) destabilization.[26] Thus, this symmetric set of buried residues was constrained to Cys mutations (SYM8ΔΔ), which were tolerated with essentially neutral effect on stability (in comparison to SYM6ΔΔ; Fig. 1). In FGF-1, the symmetry-related residues Leu14, Ile56, and Tyr97 form a core–solvent interface at the "top" of the central β-barrel. Attempts to enforce a symmetric constraint as a final step of transform #1 by evaluating Ile56Leu, Tyr97Ile, or Leu14Ile point mutations failed due to substantial destabilizing and precipitation; thus, these positions remained wild type (i.e., asymmetric) at the end of transform #1.

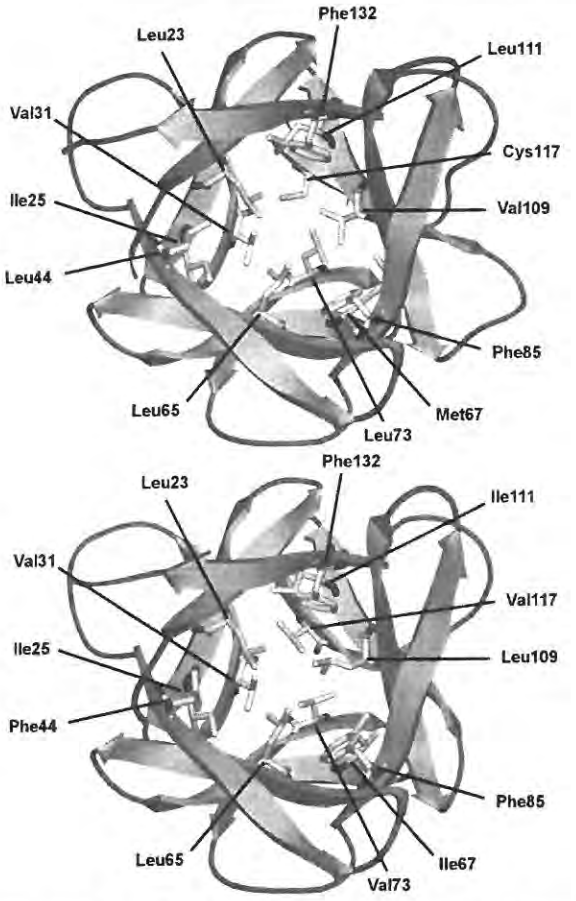

Fig. 2. Top: ribbon representation of the FGF-1 structure (2AFG) showing the residues that make up the central hydrophobic core. Bottom: similar ribbon representation of the SYM6ΔΔ(F108Y) mutant structure showing the equivalent hydrophobic core residues and illustrating the essential completion of transform #1 in the SD.

Overall, the application of transform #1 resulted in an increase in 3-fold primary-structure symmetry from 2% (i.e., 1 of 42 positions) in FGF-1 to 21% (i.e., 9 of 42 positions) in SYM8ΔΔ. Furthermore, deletions in the third trefoil-fold subdomain resulted in its length equaling that of the second trefoil-fold subdomain, thereby increasing the tertiary-structure symmetry. Substantial nonadditive and cooperative effects were observed in the application of transform #1. The SYM3 mutant was 5.0 kJ/mol more stable than the simple sum of

the constituent point mutations;[23] the SYM5 mutant was 12.0 kJ/mol more stable than the sum of the constituent point mutations;[24] and the Met67Ile point mutation was 9.4 kJ/mol destabilizing in the SYM5 background, but was completely neutral as regards stability when constructed in the same protein with the above-described tertiary-structure deletions in the third trefoil-fold subdomain.[25]

Surface plasmon resonance (SPR) data for the binding constants of the SYM6ΔΔ and SYM7ΔΔ mutants with the extracellular domain of the FGF receptor-1c (FGFR-1c) protein demonstrated an approximately 2 orders of magnitude increase in K_D (i.e., reduction in receptor affinity) for SYM6ΔΔ and another order of magnitude increase in K_D for SYM7ΔΔ (Table 1). Conversely, the 3T3 fibroblast mitogenic assay showed an order of magnitude increase in mitogenic potency for the SYM6ΔΔ mutant in comparison to FGF-1 (Table 2). Crystal structures were reported for the SYM2,[23] SYM3,[23] SYM4,[24] and SYM5[24] mutants. Although the other core mutants did not crystallize, a Phe108Tyr mutation within the SYM6ΔΔ protein crystallized and yielded a 1.60 Å resolution data set and is reported here (Table 3).

Transform #2: SD of reverse turns

The buried free cysteine residues introduced into the SYM8ΔΔ mutant at positions 42 and 130 proved problematic for folding and stability studies due to the need for reducing agents in the buffer; thus, these residue positions were reverted to Ile as part of the initial application of transform #2 with the design of the SYM9ΔΔ mutant (Fig. 3). The introduction of a (partial) symmetric constraint on β-turn regions, ending with the SYM11ΔΔ mutant, was achieved with a further substantial increase in stability and slight reduction in folding cooperativity m value. The SYM11ΔΔ mutant (the product of transform #2) was 23.2 kJ/mol more stable than the SYM8ΔΔ mutant and 41.6 kJ/mol more stable than the FGF-1 starting protein. None of the mutants in this transform (i.e., SYM9ΔΔ, SYM10 ΔΔ, and SYM11 ΔΔ) exhibited any detectable binding to the FGFR-1c protein in SPR studies (Table 1); furthermore, the SYM10ΔΔ mutant was devoid of any detectable mitogenic activity (Table 2). After the application of transform #2, the 3-fold symmetry of the primary structure increased from 21% (9 of 42 positions in SYM7ΔΔ) to 26% (11 of 42 positions in SYM11ΔΔ) with no change in the tertiary structure (i.e., length) of the individual trefoil-fold subdomains (Fig. 3).

Table 1. SPR binding constants for FGF-1 and mutant proteins with soluble FGFR-1c

Protein	K_D (M)	R_{max} (RU)	χ^2 (RU^2)
FGF-1	$(6.96 \pm 0.04) \times 10^{-9}$	24.9	0.185
SYM6ΔΔ	$(2.09 \pm 0.02) \times 10^{-7}$	27.1	0.016
SYM7ΔΔ	$(1.76 \pm 0.81) \times 10^{-6}$	61.8	0.020
SYM9ΔΔ	N.D.		
SYM10ΔΔ	N.D.		
SYM11ΔΔ	N.D.		
SYM12ΔΔ	N.D.		
SYM13ΔΔ	N.D.		
Symfoil-1	N.D.		
SYM6ΔΔ/K12V/P134V	N.D.		
SYM7ΔΔ/K12V/P134V	N.D.		

N.D., no binding detected.

Table 2. Summary of the mitogenic activity of FGF-1 and mutant proteins with 3T3 fibroblasts

Protein	EC_{50} (ng/ml)
FGF-1	58.4 ± 25.4
SYM6ΔΔ[25]	0.84 ± 0.43
K12V/P134V[27]	1.80 ± 0.90
SYM6ΔΔ/K12V/P134V	741 ± 302
SYM10ΔΔ	N.D.

N.D., no activity detected.

Table 3. Crystallographic data collection and refinement statistics for SYM6ΔΔ (F108Y mutant)

Data collection	
Space group	*P*1
Cell dimensions	
a, *b*, *c* (Å)	46.9, 56.9, 61.9
α, β, γ (°)	64.9, 89.7, 71.2
Resolution (Å)	50.00–1.60 (1.66–1.60)[a]
Mosaicity (°)	0.92
Matthews coefficient ($Å^3$/Da)	2.12
Redundancy	3.7 (2.6)
Total reflections	256,934
Unique reflections	68,593
$I/\sigma I$	30.8 (2.9)
Completeness (%)	96.3 (88.4)
R_{merge}	11.1 (34.3)
Refinement	
Resolution (Å)	27.71–1.60
R_{work}/R_{free}	17.0/20.3
No. of atoms	
Protein	3977
Ligand/ion	12
Water	578
B-factor	
Protein	23.8
Ligand/ion	23.5
Water	33.7
RMSDs	
Bond length (Å)	0.006
Bond angle (°)	1.01
Ramachandran plot:	
Most favored (%)	91.2
Additionally allowed (%)	8.6
Generously allowed (%)	0.2
Disallowed region (%)	0.0
PDB code	3O3Q

[a] Values in parentheses are for the highest-resolution shell.

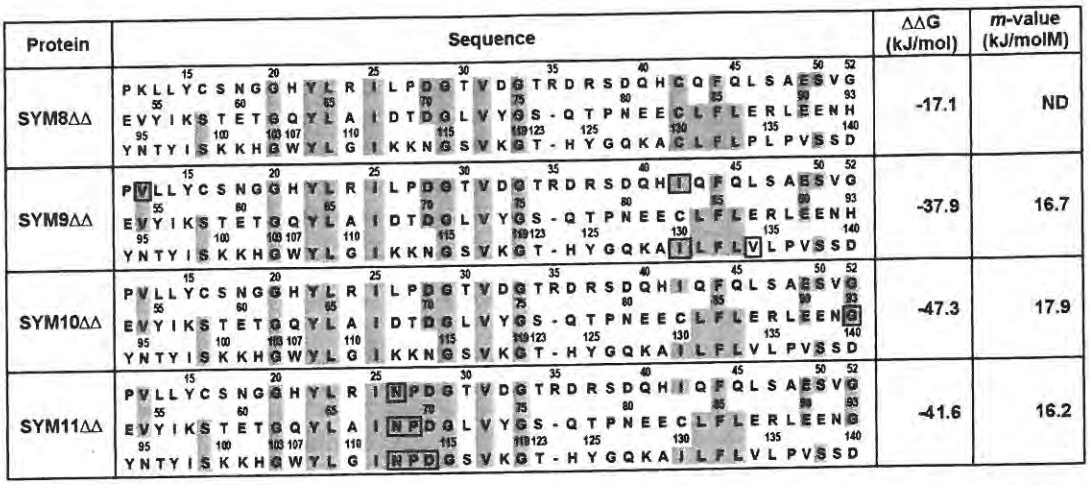

Fig. 3. SD transform #2: introduction of a symmetric constraint on reverse turn regions. Gray shading indicates positions related by 3-fold structural symmetry where either two or all three residues are identical. Boxed positions indicate the specific site(s) of mutation in construction of that particular mutant. $\Delta\Delta G$ values are a measure of the stability effect (referencing FGF-1), where a negative value indicates an increase in overall thermostability. Thermodynamic data are from Lee and Blaber.[21]

Transform #3: SD of β-strands

The introduction of a (partial) symmetric constraint on β-strand secondary structure, starting with the SYM11ΔΔ mutant and ending with the SYM13ΔΔ mutant, was achieved with a further gain in stability (Fig. 4). The SYM13ΔΔ mutant was 7.8 kJ/mol more stable than the SYM11ΔΔ mutant and 47.4 kJ/mol more stable than the FGF-1 starting protein; however, there was a gradual decrease observed in the folding cooperativity *m* value from 18.9 kJ/mol M in FGF-1 to 16.2 kJ/mol M in SYM11ΔΔ and 13.9 kJ/mol M in SYM13ΔΔ. SPR studies confirmed that none of the mutations in this transform (i.e., SYM11ΔΔ, SYM12ΔΔ, and SYM13ΔΔ) possess any detectable affinity for the FGFR-1c protein (Table 1). The 3-fold symmetry of the primary structure increased from 26% (11 of 42 positions in SYM11ΔΔ) to 33% (14 of 42 positions in SYM13ΔΔ) (Fig. 4).

At this point, in an attempt to speed the SD process, we constructed a combined SYM13ΔΔ/Cys16Ser/Thr59Ser/Lys100Ser/Asp140Gly/

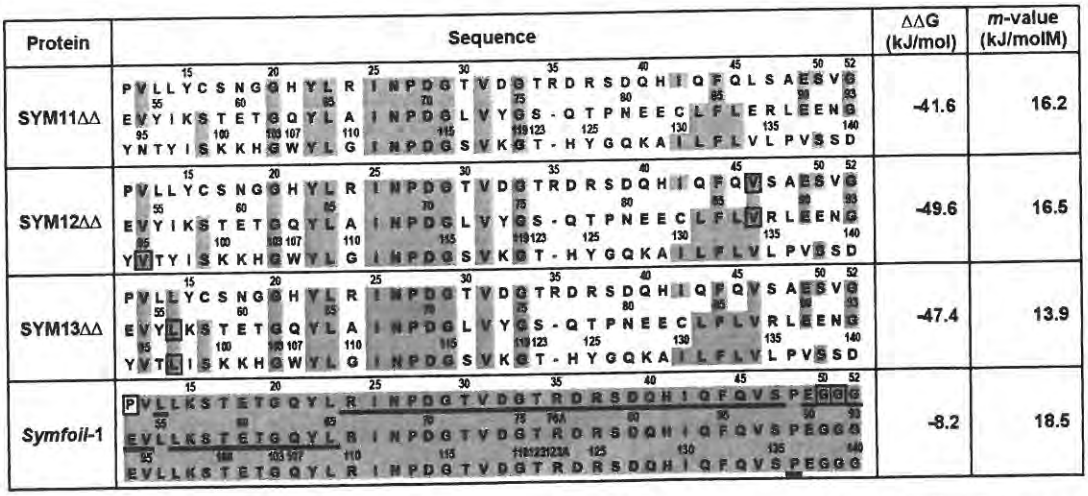

Fig. 4. SD transform #3: introduction of a symmetric constraint on β-strand regions. Gray shading indicates positions related by 3-fold structural symmetry where either two or all three residues are identical. Boxed positions indicate the specific site(s) of mutation in construction of that particular mutant. Underlined positions indicate the regions of the SYM13ΔΔ mutant utilized in the chimera construct of the Symfoil-1 mutant. $\Delta\Delta G$ values are a measure of the stability effect (referencing FGF-1), where a negative value indicates an increase in overall thermostability. Thermodynamic data are from Lee and Blaber.[21]

His41Ala/Ile42Thr/Phe44Trp/Glu82Ala/Cys83Thr/Phe85Trp/Ile130Thr/Phe132Trp mutant in the SYM13ΔΔ mutant with the combined goal of eliminating the buried free thiols at positions Cys16 and Cys83, increasing the symmetric constraint on β-strand secondary structure, and simultaneously introducing buried Trp residues as a useful fluorescence probe of protein folding. Although this mutant was folded and soluble, it exhibited poor stability ($\Delta G = 17.0$ kJ/mol) and notably poor folding cooperativity m value (5.3 kJ/mol M). Thus, this mutant was abandoned, and, instead, a chimera strategy was pursued to eliminate the buried free thiols and increase the 3-fold symmetric constraint. The initial chimera design focused on invoking a symmetric constraint utilizing residue positions that comprise the van der Waals contacts of Ser58 and Ile42 as a means to eliminate reactive buried thiols at symmetry-related positions Cys16 and Cys83. However, since this substitution of primary structure involved a substantial portion of the remaining asymmetry within the SYM13ΔΔ mutant, the chimera strategy was expanded to comprise regions that defined the entire molecule and thereby enforce a complete primary- and tertiary-structure symmetric constraint in a single mutagenesis step (producing the Symfoil-1 protein; Fig. 4). In comparison to SYM13ΔΔ, the Symfoil-1 protein exhibited a substantial (39.7 kJ/mol) loss of stability; however, the folding cooperativity m value increased from 13.9 to 18.5 kJ/mol M, essentially recovering the folding cooperativity of the FGF-1 starting protein. Furthermore, while the Symfoil-1 protein was destabilized in comparison to SYM13ΔΔ, it was nonetheless significantly (8.3 kJ/mol) more stable than the FGF-1 starting protein. Notably, in contrast to FGF-1 [which aggregates upon thermal denaturation in the absence of low concentrations of added guanidine HCl (GuHCl) denaturant[22]], differential scanning calorimetry (DSC) analysis of the Symfoil-1 protein demonstrated reversible, two-state thermal denaturation (i.e., $\Delta H_{\text{van't Hoff}}/\Delta H_{\text{cal}} \sim 1.0$), with excellent recovery of enthalpy upon cooling, under all buffer conditions tested (Table 4).

A comparison of the primary structure of FGF-1 with Symfoil-1 shows that the SD was associated with a total of 76 mutated positions and 4 overall deleted positions. The first trefoil-fold subdomain of Symfoil-1 is 67% identical (28 of 42 positions) with no insertions or deletions relative to the same subdomain in FGF-1; the second trefoil-fold subdomain is 39% identical (16 of 41 positions) with one additional residue in the Symfoil-1 subdomain; and the third trefoil-fold subdomain is 11% identical (5 of 47 positions) with five residues deleted in the Symfoil-1 subdomain. Notably, while the FGF-1 sequence contains instances of all 20 common amino acids, the Symfoil-1 sequence contains only 16 unique amino acids and is devoid of Ala, Cys, Trp, and Met residues. While the elimination of reactive thiols (Cys residues) was part of the design strategy, loss of the other residues was an unanticipated consequence of the SD. A crystal structure of the Symfoil-1 protein confirmed the precise 3-fold

Table 4. DSC data of FGF-1 and mutant proteins in ADA buffer

Protein	ΔH (kJ/mol)	ΔC_p (kJ/mol K)	T_m (K)	H_{vH}/H_{cal}	$\Delta\Delta G$ (kJ/mol)[a]
0.7 M GuHCl					
FGF-1[22]	257±3	9.33±0.53	312.6±0.10	1.08±0.10	—
Symfoil-1	311±6	6.06±0.63	325.4±0.12	1.02±0.12	−10.4
Symfoil-2	394±6	6.97±1.11	335.3±0.11	1.02±0.03	−21.3
Symfoil-3	417±4	8.64±0.56	338.4±0.05	1.18±0.01	−24.5
Symfoil-4T	426±3	9.24±0.70	340.2±0.09	1.16±0.03	−26.5
Symfoil-4V	472±4	8.68±0.19	343.5±0.02	1.14±0.01	−31.6
Symfoil-4P	539±5	4.66±0.84	354.0±0.03	1.11±0.01	−43.2
0 M GuHCl					
Symfoil-1	400±2	6.36±0.49	333.9±0.05	1.04±0.02	
Symfoil-2	460±8	6.16±0.94	341.7±0.08	1.13±0.06	
Symfoil-3	494±5	8.24±0.65	344.5±0.03	1.10±0.02	
Symfoil-4T	501±9	8.67±0.67	346.6±0.03	1.12±0.03	
Symfoil-4V	557±4	5.77±2.61	348.9±0.09	1.04±0.04	
Symfoil-4P	599±10	5.67±1.01	358.1±0.04	1.05±0.07	
	ΔH^0 (kJ/mol)[b]	ΔC_p (kJ/molK)[b]	T_m (K)[b]	H_{vH}/H_{cal}[b]	K_D (M^2)
Monofoil-4P (20 μM)	434±3	8.61±0.33	333.6±0.05	0.84±0.10	$0.93\times10^{-8}\pm0.40$
Monofoil-4P (40 μM)	399±5	7.61±0.07	334.0±0.10	0.93±0.07	$2.07\times10^{-8}\pm0.70$
Monofoil-4P (80 μM)	419±2	7.45±0.07	335.7±0.10	0.94±0.06	$5.72\times10^{-8}\pm1.80$
Difoil-4P (20 μM)	701±5	8.70±0.10	345.1±0.05	0.97±0.04	$3.58\times10^{-11}\pm2.5$
Difoil-4P (40 μM)	733±5	8.86±0.08	346.2±0.05	0.97±0.04	$9.56\times10^{-11}\pm4.0$

[a] Determined at 332 K in reference to FGF-1.
[b] Two-state trimer dissociation model.

symmetric structural properties, which extends to identical side-chain rotamers and regions of solvent structure.[21] Additionally, the Symfoil series of mutants contained Tris in the crystallization mother liquor, and this C_3 symmetric molecule was observed precisely bound on the 3-fold axis of structural symmetry at the "bottom" of the β-barrel (Fig. 5). Isothermal titration calorimetry (ITC) studies of Tris binding (using the Symfoil-2 protein) demonstrated that Tris binding is an enthalpy-driven process (Supplemental Fig. 1).

Transform #4: Stability optimization/SD fragmentation

The stability optimization of the Symfoil-1 protein, progressing through the Symfoil-2, Symfoil-3, Symfoil-4T, Symfoil-4V, and culminating in the Symfoil-4P protein, resulted in a significant increase in stability with modest reduction in the folding cooperativity m value (Fig. 6). The Symfoil-4P protein was 35.9 kJ/mol more stable than the Symfoil-1 protein and 42.6 kJ/mol more stable than the initial FGF-1 protein. The progressive increase in stability for this set of mutants was associated not only with an increase in folding rate constant but an even more substantial decreasing of the unfolding rate constant (Table 5). The slowest kinetic data that could be practically obtained were on the order of $1 \times 10^{-4}\ s^{-1}$ ($\ln k_{obs}$ approximately −9); thus, data for the "unfolding arm" of the Symfoil-4P mutant were effectively truncated at this region (Fig. 7). There was excellent agreement between the midpoint of denaturation as determined from the isothermal equilibrium data and folding kinetics, further supporting the two-state folding assumption. There was similar agreement for the derived $\Delta\Delta G$ stability values, although these values deviated somewhat with increasing stability (and the associated longer incubation times required for equilibrium).

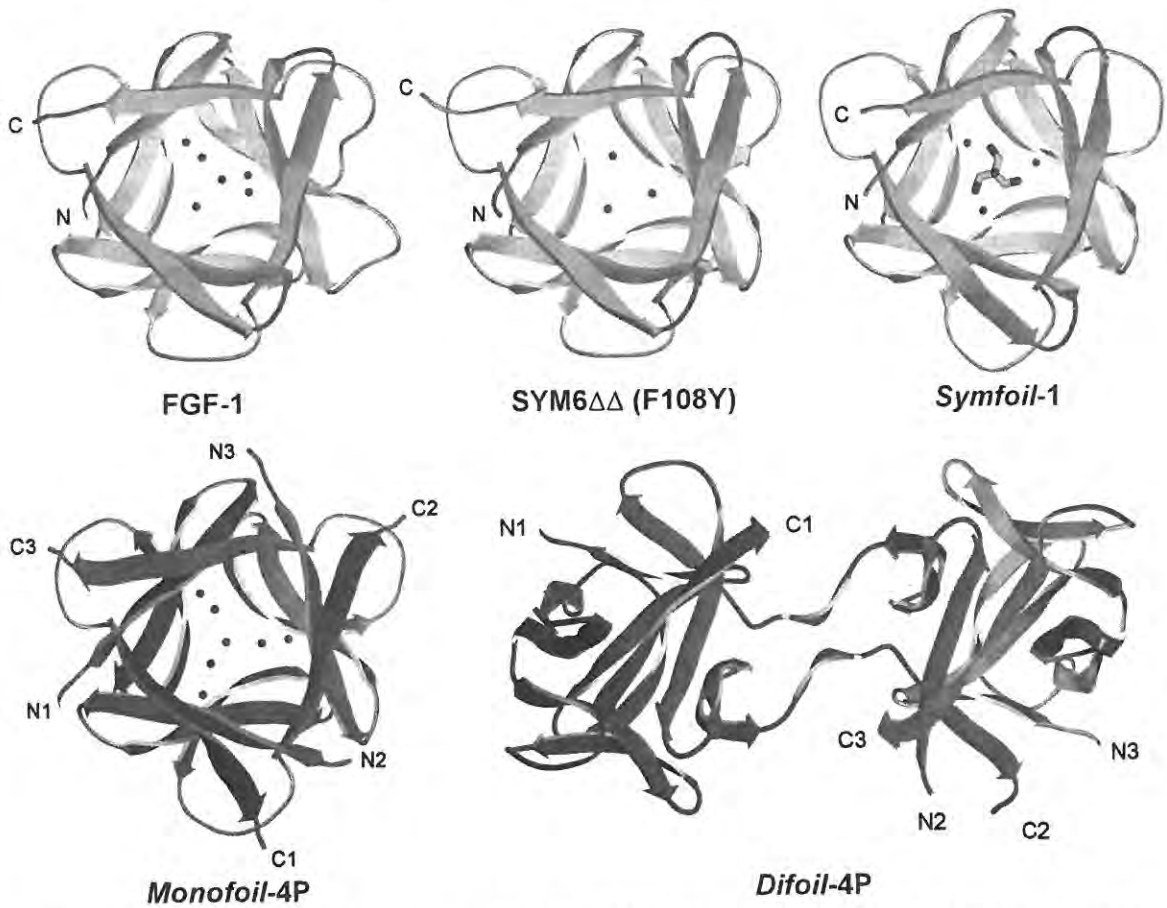

Fig. 5. Ribbon diagram of the X-ray structures of FGF-1,[28] SYM6ΔΔ(Phe108Tyr), Symfoil-1, Monofoil-4P, and Difoil-4P proteins.[21] FGF-1, SYM6ΔΔ(Phe108Tyr) and Symfoil-1 structures are colored according to their secondary structure (β-strands are yellow, turns/coils are gray). The Monofoil-4P and Difoil-4P structures are colored according to the individual polypeptide chains (chain A, red; chain B, green; chain C, blue). In each case except Difoil-4P, the view is down the 3-fold axis of structural symmetry. Select solvent structure at the bottom of the β-barrel in the vicinity of the 3-fold axis is shown; additionally, the Symfoil series of proteins crystallized in the presence of Tris buffer that bound at this 3-fold axis (shown in stick figure in the Symfoil-1 structure).

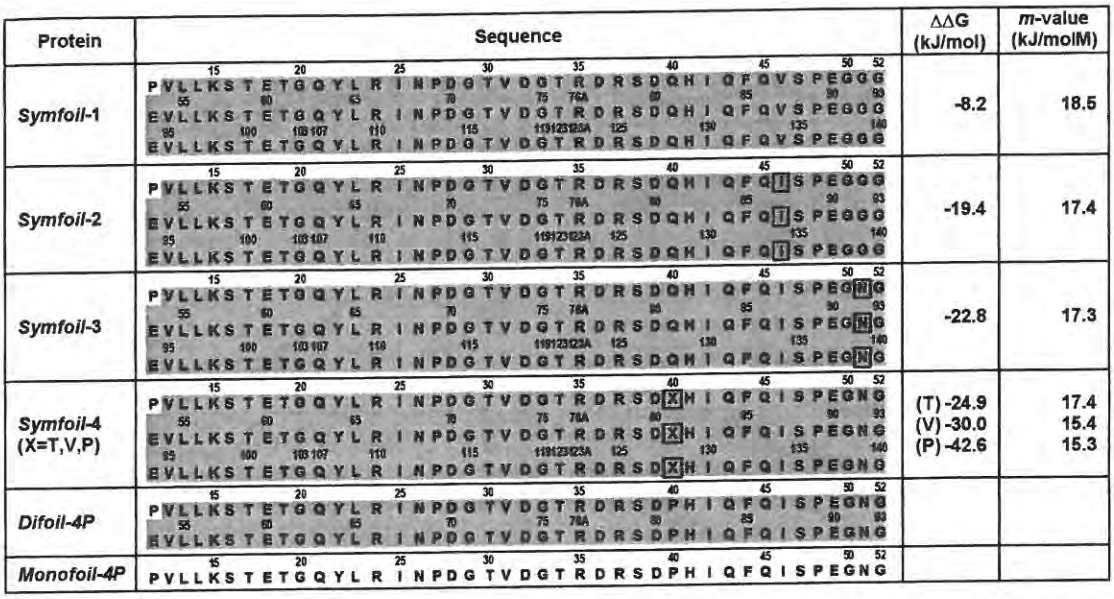

Fig. 6. SD transform #4: stability optimization and symmetric fragmentation. Gray shading indicates positions related by 3-fold structural symmetry where either two or all three residues are identical. Boxed positions indicate the specific site(s) of mutation in construction of that particular mutant. $\Delta\Delta G$ values are a measure of the stability effect (referencing FGF-1), where a negative value indicates an increase in overall thermostability. Thermodynamic data are from Lee and Blaber.[21]

The protease-resistance properties for this set of mutants correlated with their stability, and the most stable Symfoil-4P mutant was 3–4 orders of magnitude more resistant to proteolysis than the FGF-1 starting protein (Fig. 8).

Expressed Monofoil-4P and Difoil-4P proteins, representing monomer and dimer trefoil-fold subdomains of the Symfoil-4P protein, respectively, were soluble during purification. The Monofoil-4P and Difoil-4P proteins resolved as single peaks on calibrated size-exclusion chromatography, with apparent masses, indicating a homogenous oligomeric form (with $n \sim 3$) and with no detectable monomer form present with either protein.[21] In analytical ultracentrifugation studies, Monofoil-4P and Difoil-4P proteins sedimented as homogeneous forms with $S_{20,w}$ values of 2.31 ± 0.036 and 3.18 ± 0.021 S, respectively, and with corresponding molecular masses of 18.8 and 29.8 kDa, respectively.[21] The molecular mass of the Monofoil-4P (6.4 kDa) and Difoil-4P (11.2 kDa) polypeptides therefore indicated a homotrimer in solution for both polypeptides, with no evidence of monomeric or higher-order oligomeric forms.

DSC analyses showed that the molar calorimetric enthalpy for the Monofoil-4P trimer was approximately 80% that of the Symfoil-4P protein, while the molar calorimetric enthalpy for the Difoil-4P trimer was approximately 23% greater than that of the Symfoil-4P protein (Fig. 9). The $H_{van't\ Hoff}/H_{cal}$ ratio was near unity for both the Monofoil-4P and Difoil-4P polypeptides, obtained with a trimer dissociation model (Table 4). There were only modest concentration-dependent effects on the thermodynamic parameters over the 20–80 μM range evaluated, and the calorimetrically derived K_D value for the Monofoil-4P protein was 10–50 nM, while the Difoil-4P exhibited a subnanomolar (0.04–0.10 nM) K_D

Table 5. Folding and unfolding kinetic parameters for FGF-1 and *Symfoil* mutant proteins in ADA buffer with GuHCl as the denaturant

Protein	k_f (s^{-1})	m_f (kJ/mol M)	k_u (1×10^{-6} s^{-1})	m_u (kJ/mol M)	m value (kJ/mol M)	C_m (M)	$\Delta\Delta G$ (kJ/mol)	$\Delta\Delta G_u$ (kJ/mol)	$\Delta\Delta G_f$ (kJ/mol)
FGF-1	3.74	−16.4	808	1.1	17.5	1.19	—	—	—
Symfoil-1	5.30	−14.1	119	3.2	17.3	1.53	−5.9	2.3	−3.6
Symfoil-2	19.6	−13.8	12.0	2.7	16.5	2.14	−15.7	8.6	−7.2
Symfoil-3	51.1	−13.6	4.27	3.2	16.7	2.42	−20.4	10.6	−9.8
Symfoil-4T	73.6	−12.9	11.5	2.6	15.5	2.51	−20.4	8.8	−11.6
Symfoil-4V	115	−12.5	1.69	2.9	15.3	2.92	−26.4	13.2	−13.2
Symfoil-4P	134	−10.5	1.07	2.0	12.5	3.70	−31.3	15.4	−15.9

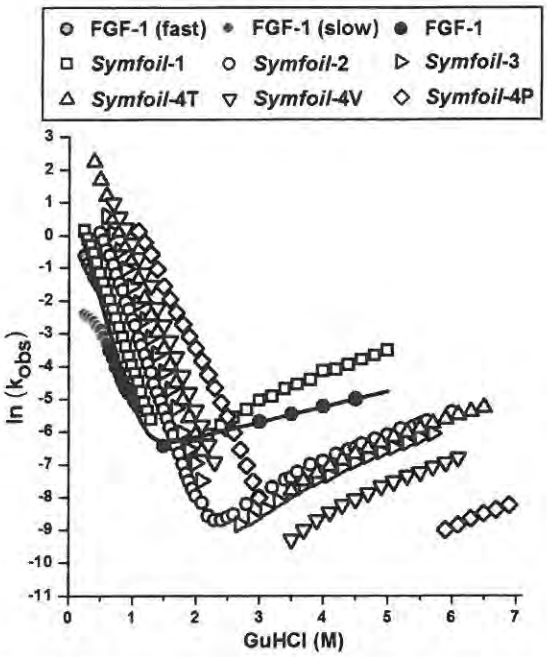

Fig. 7. Folding and unfolding kinetics for FGF-1 and Symfoil mutant proteins. The left-hand side of the figure shows the "folding arms" and the right-hand side shows the "unfolding arms." To facilitate identification of the FGF-1 folding profile, a line representing the global fitted function (i.e., "chevron plot") is shown. FGF-1 exhibits both a fast and a slow folding phase, whereas none of the Symfoil proteins exhibit this property. Stability optimization of the Symfoil series of proteins was associated with an increase in the folding kinetics but an even more pronounced decrease in the unfolding kinetics.

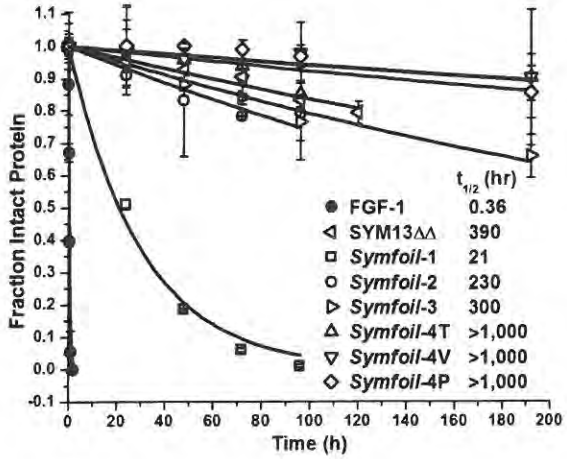

Fig. 8. Trypsin proteolysis of FGF-1, SYM13ΔΔ, and Symfoil mutant proteins. Resistance to proteolysis generally correlates with thermostability. The proteolytic degradation half-life of the optimized Symfoil protein (i.e., Symfoil-4P) is 3–4 orders of magnitude greater than that of the FGF-1 protein.

value. The Monofoil-4P trimer was thus less stable than the Difoil-4P trimer, which in turn was less stable than the Symfoil-4P protein; however, each of these oligomers had a higher T_m and enthalpy of unfolding than FGF-1. Crystal structures of the Monofoil-4P and Difoil-4P proteins confirmed homotrimer assemblies that reconstituted single and dual β-trefoil protein folds with essential structural identity to the Symfoil-4P protein[21] (Fig. 5).

Discussion

Evidence of structural symmetry is widespread in naturally evolved protein folds,[29–31] presumably reflecting gene duplication and fusion events responsible for their emergence and suggesting a strategy to simplify bottom-up hierarchical *de novo* design (i.e., via symmetric self-assembly of an appropriately designed polypeptide building block).[3,4,32] Despite the attractiveness of a symmetric design principle, bottom-up design efforts have not achieved widespread success. Here, we have asked whether an alternative approach (SD) can be used to successfully identify a polypeptide building block for a symmetric target architecture. This novel approach offers potential advantages for *de novo* design but also involves risks related to poorly understood aspects of protein folding. SD involves a significant reduction in sequence complexity; however, reduced

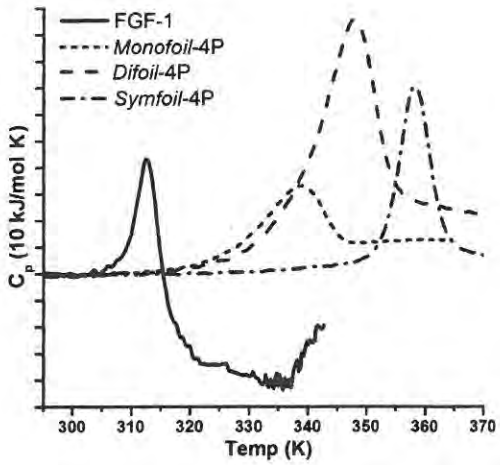

Fig. 9. DSC endotherms for FGF-1, Monofoil-4P, Difoil-4P, and Symfoil-4P polypeptides in ADA buffer. Endotherms for Monofoil-4P and Difoil-4P proteins are normalized per molar concentration of trimer assembly. FGF-1 has mesophilic stability and irreversibly aggregates during thermal denaturation, giving rise to the noisy exothermic signal post denaturation; Monofoil-4P, Difoil-4P, and Symfoil-4P proteins do not exhibit this property. The Symfoil-4P protein exhibits hyperthermophilic stability (i.e., T_m > 85 °C).

sequence complexity and symmetric primary structure are features associated with natively unstructured proteins and protein aggregation.[17–20] Thus, it is not apparent, *a priori*, that SD can produce a foldable polypeptide. Additionally, the final stage of SD involves fragmentation to yield the repeating polypeptide motif; with the expectation that this peptide will spontaneously assemble (as a homomultimer) and yield the correct target architecture. Such fragmentation potentially exposes substantial hydrophobic surface, and multimer assembly is achieved with a significant entropic penalty; additionally, in the case of the β-trefoil fold, no known example exists of its assembly via a homotrimer. Thus, it is not obvious that the solubility, folding, and thermodynamic properties of a polypeptide building block, necessary to spontaneously assemble as a homotrimer forming a β-trefoil fold, are realistically achievable. However, these types of issues are inherent in all *de novo* protein design efforts that attempt to exploit bottom-up hierarchical design of symmetric protein architecture. A primary advantage of SD is that the design process starts with a naturally evolved protein that is soluble, thermostable, and foldable; as long as mutations to enforce a symmetric constraint remain within foldable, thermostable "sequence space," a design solution is possible.

The 3-fold symmetric β-trefoil fold was selected as the target architecture for SD. Structural symmetries higher than 2-fold provide opportunities for subdomain mismatches during folding that can contain a significant fraction of native-like contacts but are nonetheless misfolded; thus, the β-trefoil architecture addresses folding issues inherent to higher-symmetry architecture. Additionally, unlike certain other symmetric protein architectures [e.g., the β-propeller or $(\beta/\alpha)_8$-barrel], the β-trefoil architecture contains a substantial and cooperatively packed central hydrophobic core (comprising ~14%, or 18 of 126, of the total residues); thus, the β-trefoil fold is representative of a single-domain globular protein. The β-trefoil protein selected for SD is FGF-1, which exhibits a substantial lack of symmetry at both the primary- and the tertiary-structure level as well as poor mesophilic thermostability; thus, attempting SD with FGF-1 is nontrivial. An initial size-exclusion chromatography study of peptide fragments of FGF-1 (residues 10–52, 53–93, and 94–140; representing the three independent trefoil-fold subdomains), either alone or in various combinations, yielded no evidence of self-assembly into a multimeric structure (and, in fact, these peptides were largely insoluble). Overall, therefore, the selection of the β-trefoil as the target architecture for SD (and the selection of FGF-1 as the proxy) is highly appropriate.

The strategy utilized in pursuing SD involved the sequential application of symmetric transforms starting with the core and then proceeding to add reverse turn and β-strand structural elements, respectively. Interplay between primary- and tertiary-structure symmetry became apparent when SD of the core region was attempted. Symmetric mutations that were poorly tolerated in the FGF-1 background were neutral or favorable when constructed within a mutant background that increased the tertiary-structure symmetry (involving the deletion of two extended surface loop regions within the third trefoil-fold subdomain). Thus, symmetric transforms of the primary and tertiary structure were mutually supportive in terms of stability and folding.

The completion of transform #1 (symmetric core) did not completely abolish FGF-1 mitogenic activity, although the symmetric tertiary-structure deletions essentially abolished heparin-binding functionality. Curiously, although the SYM6ΔΔ mutant had substantially reduced heparin affinity, its mitogenic activity in 3T3 fibroblast cell assays increased by almost 2 orders of magnitude. A ternary complex of FGF-1, FGFR, and heparin is reportedly essential for FGF-1 signal transduction;[33–35] additionally, the heparin-binding affinity of FGF-1 promotes sequestration onto heparin and heparan glycans present in the extracellular matrix and cell surface.[36–38] Thus, while the SYM6ΔΔ mutant retains sufficient heparin affinity for signal transduction, it may not readily be sequestered on cell-surface heparan proteoglycans (and a greater effective concentration may be available in solution for receptor binding). Studies of additional stability mutants at the end of transform #1 provide evidence of a loss of essential dynamics necessary for FGFR-binding function. The SYM6ΔΔ mutant is highly active in the 3T3 fibroblast mitogenic assay (Table 2). A Lys12Val/Pro134V double mutant in FGF-1 that is also highly active in the mitogenic assay was previously described[27] (Table 2). Both mutations are substantially stabilizing; the SYM6ΔΔ mutant provides 16.1 kJ/mol of additional thermostability, and the Lys12Val/Pro134V double mutant provides 19.1 kJ/mol of additional thermostability. The combination of these mutations produces a mutant that is, surprisingly, essentially inactive in the mitogenic assay (Table 2). We conclude that the substantial increase in thermostability (i.e., $\Delta\Delta G$, approximately −35 kJ/mol) eliminates structural dynamics essential for signal transduction function. Thus, multiple aspects of FGF-1 functionality were being eliminated early in the SD process; certainly by the development of the SYM10ΔΔ mutant (in transform #2), key properties of heparin-binding, FGFR affinity, dynamics essential for FGF function, and associated mitogenic activity were lost. Nonetheless, the protein remained within foldable, thermostable sequence space.

When a stability screen for mutant selection was utilized, the thermostability increased continuously during the application of transforms #1–3 (Fig. 10).

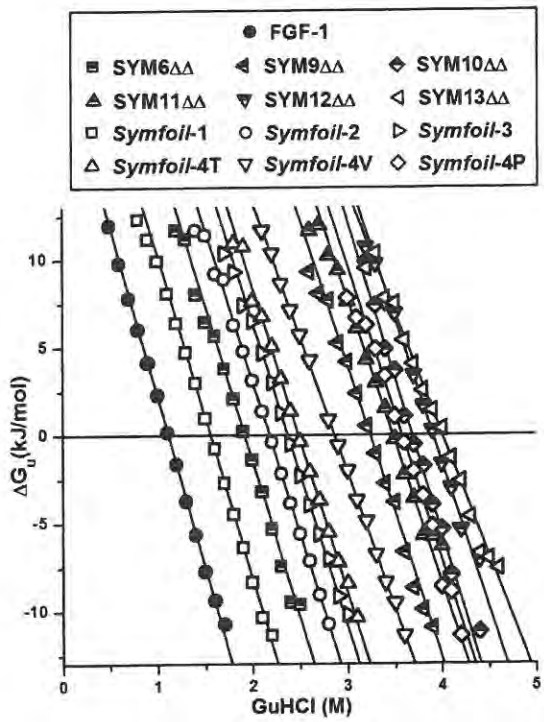

Fig. 10. ΔG_u *versus* GuHCl denaturant concentration, as determined from isothermal equilibrium denaturation, for select mutants in the SD. Data shown are overlaid with the associated fitted function from a two-state model (see Materials and Methods for model definition).

Thus, the imposition of a symmetric constraint had a net effect of a loss of function, but a greater than twofold gain in ΔG_u, compared to FGF-1. Although FGF-1 exhibits poor mesophilic stability, the many mutant examples with increased thermostability suggest that the β-trefoil architecture is intrinsically capable of substantial thermostability, and there is no reason to believe that the various mutant forms described in the present study represent the optimum possible thermostability. The "stability/function trade-off" hypothesis posits that functional residues are often accommodated at the expense of overall stability,[39–44] and the results are consistent with this hypothesis. The present results prompt us to propose a "symmetry/function trade-off" hypothesis in that aspects of structural asymmetry are associated with function (at the expense of stability). We note that development of a symmetric β-trefoil fold resulted in the emergence of an unanticipated novel "functionality," namely, binding specificity for a C_3 symmetric molecule (Tris).

While the stability of the original mesophilic FGF-1 protein could be moved into the hyperthermophile range (i.e., T_m >85°C with the SYM10ΔΔ, SYM12ΔΔ, SYM13ΔΔ, and Symfoil-4P mutant proteins), the folding cooperativity m value exhibited a modest decline. Auton and Bolen have reported that the summation of transfer free energies from water to osmolyte for a comparison of native- and denatured-state models can accurately predict protein folding m values,[45] with the greatest contribution to the magnitude of the folding m value provided by the peptide backbone (and with side chains having a comparatively minor influence). This suggests that a specific tertiary structure has an intrinsic folding m value with comparatively minor modulation in response to amino acid substitution; thus, conservation of tertiary structure is consistent with conservation of folding m value. We note, however, that all mutants in the SD less stable than FGF-1 exhibit a higher folding m value; conversely, all mutants more stable than FGF-1 exhibit a lower folding m value. This effect is relatively subtle but suggests an inverse correlation between thermostability and folding m value, with FGF-1 residing on this continuum (Fig. 11). The origin of this effect may be residual structure in the osmolyte-induced unfolded form as the thermostability increases; this correlation suggests that this may be largely unavoidable. Thus, when a stability screen for SD is utilized, enhanced thermostability is an attainable goal, but maintenance of folding m value may be difficult to achieve. The folding and unfolding kinetic data (Fig. 7) show that stability enhancement for the Symfoil series of proteins was realized through a combination of increased folding as well as decreased unfolding kinetic constants. These results are achieved through native-state stabilization (in comparison to the

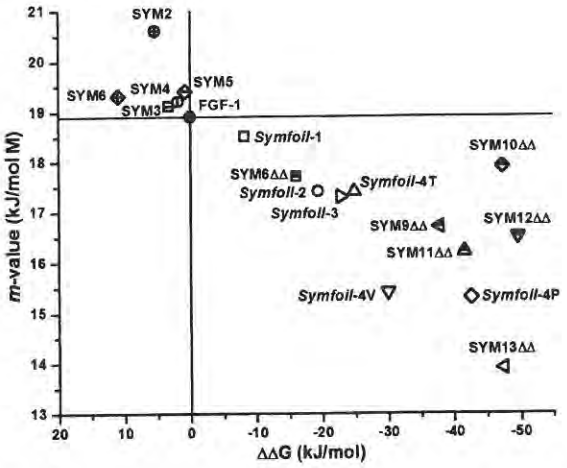

Fig. 11. Effects on change in protein stability ($\Delta\Delta G$) plotted *versus* folding cooperativity m value for mutants in the SD (from isothermal equilibrium denaturation data). A negative value of $\Delta\Delta G$ indicates increased stability; an increase in m value indicates increased folding cooperativity. Increasing thermostability appears to correlate with a decline in folding cooperativity m value.

folding transition state; i.e., $\Delta\Delta G_u > 0$) as well as folding transition-state stabilization (in comparison to the denatured state; i.e., $\Delta\Delta G_f < 0$; Table 5). Thus, the symmetric primary structure is compatible with both an efficient native structure and folding transition state in terms of thermodynamics and folding kinetics.

The use of a chimera construct to achieve a final symmetric polypeptide (i.e., the Symfoil-1 protein) proved to be a successful and rapid method to complete the SD (while maintaining a foldable, stable polypeptide). Immediately prior to the chimera design, the SD transforms increased the primary-structure symmetry from 2% in FGF-1 to 33% in the SYM13ΔΔ mutant. A useful heuristic in undertaking a chimera approach to achieve a symmetric primary-structure constraint would be to increase the primary-structure symmetry (e.g., by point mutation) to a critical threshold prior to attempting chimera design; in this regard, 33% primary-structure symmetry may be a useful minimum target.

The Symfoil-1 protein contains only 16 of the 20 common amino acids. Notably absent in the primary structure is Ala, known for high α-helical propensity.[46–48] The Symfoil β-trefoil architecture is devoid of α-helix and contains only β-strand and reverse-turn secondary structure. Conversely, the Symfoil sequence is comparatively rich in β-branched residues Ile, Thr, and Val that are known to have a high β-strand-forming propensity.[49–51] The Symfoil-1 protein, involving both a 3-fold symmetric constraint and elimination of a significant subset of amino acids, represents a substantial reduction in sequence complexity in relationship to the FGF-1 protein. While reduced sequence complexity and symmetric primary-structure features are associated with natively unstructured proteins and aggregation,[17–20] Symfoil-1 exhibits enhanced stability and folding properties in comparison to FGF-1 (i.e., no thermal-induced aggregation, reversible two-state denaturation with no biphasic folding, and enhanced thermostability). Therefore, the results show that symmetric sequence design can be compatible with efficient folding and stability, thus providing support for purely symmetric design principles in *de novo* protein design. The derived Monofoil-4P polypeptide is 66% identical to FGF-1 trefoil-fold 1, 39% identical to trefoil-fold 2, and 11% identical to trefoil-fold 3, but is devoid of known FGF-1 functionality. It is therefore plausible that in FGF-1 the first two trefoil folds provide the majority of the essential structural determinants for folding and stability; correspondingly, elements of heparin- and receptor-binding functionality are known to residue within the third trefoil fold.[28,33]

The Monofoil-4P 42-mer polypeptide spontaneously folded as a stable homotrimer recapitulating the β-trefoil architecture. The Difoil-4P 82-mer polypeptide, representing a gene duplication/fusion of the Monofoil-4P peptide, also produced the β-trefoil target architecture (as two complete β-trefoil folds within a homotrimer complex).[21] No evidence of soluble monomeric forms or insoluble aggregates was observed during the purification of either polypeptide; furthermore, no refolding step was utilized in the purification of either polypeptide, and thermal denaturation was essentially reversible. Further work is needed to understand the process of trimeric assembly and whether the Monofoil-4P peptide has any preformed structure prior to oligomerization. Additionally, the oligomeric assemblies formed by both Monofoil-4P and Difoil-4P peptides recapitulate the same degree of symmetry as the parental FGF-1 protein (with no higher-order forms). In contrast to the β-trefoil, the β-propeller fold is one where different copy numbers of each repeating subdomain (e.g., four to eight "blades") are known to exist as an integral whole.[52] A major structural difference is that while the β-trefoil contains a central cooperative hydrophobic core-packing group, the β-propeller architecture can be described as a circularly closed linear-repeat protein where the primary hydrophobic packing is between adjacent repeating elements (and no centralized packing group exists). Thus, whereas an additional repeat might be accommodated with β-propeller-type architecture, accommodation of an additional repeat within the β-trefoil fold would require a substantial alteration to the design of the central cooperative core-packing group. Since all known β-trefoil proteins are single polypeptide chains, the Monofoil-4P and Difoil-4P multimeric structures are unique, and despite the obvious entropic penalty, they spontaneously assemble to generate stable β-trefoil architecture. In this regard, the development of the hyperthermostable Symfoil-4P polypeptide may have been critical to the success of the fragmentation step in the SD.

The Monofoil-4P polypeptide is a useful building block for traditional bottom-up hierarchical *de novo* design; it can assemble as a homotrimer or as a concatenated threefold repeat within a single polypeptide to spontaneously generate a soluble, foldable, thermostable, precisely defined β-trefoil fold. The resulting β-trefoil is functionally benign (i.e., no known FGF-1 function has carried over into the Monofoil-4P polypeptide) yet is thermostable and therefore function competent. Introduction of novel function can proceed via directed or random mutagenesis coupled with a functional selection or screen. In this regard, mutating the Monofoil-4P polypeptide could be utilized to search for 3-fold symmetric compatible function, whereas mutation of the Symfoil-4P polypeptide could explore function enabled by the introduction of asymmetry within the symmetric architecture. Preliminary studies (not shown) with mix and match of the Monofoil-4P

and Difoil-4P polypeptides have shown that heterodimers with 1:1 stoichiometry readily form, recapitulating an intact β-trefoil fold. In this regard, the Difoil-4P polypeptide could serve as a scaffold into which mutant Monofoil-4P type peptides could assemble to create intact soluble β-trefoil proteins with novel functionality.

The top-down SD describes a logical approach that is broadly applicable and is not architecture dependent. What is described in this report is the sequential targeting of core (transform #1), turn (transform #2), and, finally, secondary-structure (β-strand; transform #3) positions. This logic is, in principle, applicable in the SD of any aqueously soluble symmetric globular protein. Although the β-trefoil contains β-strand secondary structure and no α-helix, it is a straightforward modification of transform #3 to focus instead on α-helix. Each transform was developed by utilizing statistical preferences of amino acids, either conserved residues for the particular architecture (i.e., β-trefoil; transform #1) or conserved residues for the secondary structure (i.e., Asx-Pro-Asx-Gly in β-turns in transform #2 or preferred β-strand residues in transform #3). As with the conceptually related method of retrosynthetic analysis (RA) in the design of organic synthesis strategies (pioneered by Corey and Cheng[53]), SD of protein architecture follows a specific logic even though the end point is not known *a priori*; this is the strength of the method, which can provide a useful alternative to bottom-up approaches to *de novo* protein design.

Materials and Methods

Symmetric deconstruction

Top-down SD is loosely based on principles of RA formalized by Corey and Cheng[53] for the purpose of efficiently designing synthetic pathways of complex organic molecules. In RA, the target molecule (i.e., desired synthesis product) is considered first, and is conceptually "deconstructed" into increasingly simpler molecules using knowledge of efficient bond synthesis; each deconstruction step is termed a "transform." This process continues until useful precursors for synthesis are identified, at which point an efficient synthetic pathway has been generated. More important, RA begins with no preconceived ideas regarding the best precursor(s) for the synthesis; the method identifies such molecules. In contrast to RA applied to organic synthesis (a conceptual process), SD developed herein is experimental. SD is not concerned with issues of peptide bond synthesis; instead, SD reduces sequence complexity via the progressive application of a symmetric primary-structure constraint, while maintaining foldability, for the purpose of identifying a useful peptide building block for the target architecture. A brief description of the application of SD in support of a specific evolutionary mechanism of the β-trefoil fold has previously been reported;[21] here, we provide a detailed biophysical characterization of this process. SD starts with a naturally evolved (i.e., foldable and thermostable) protein that is a member of the target architecture. The peptide motif generated by SD is useful as a building block for subsequent bottom-up hierarchical *de novo* design of the target architecture.

The target architecture selected for SD was the β-trefoil, a common protein fold that in conceptually idealized form has internal 3-fold rotational symmetry[54,55] (although naturally evolved β-trefoil proteins exhibit substantial asymmetry in their primary and tertiary structures). The β-trefoil "proxy" selected for SD was human FGF-1 (see Figs. 1, 2 and 5). In choosing FGF-1, maximum advantage was taken of a comparatively large amount of published mutational data on protein stability and folding. Sequence alignment of the three trefoil-fold subdomains within FGF-1 identifies only a single position sharing identical residues (i.e., Gly29, Gly79, and Gly115). Additionally, each of the three trefoil-fold subdomains exhibit different lengths (due to relative insertions/deletions); thus, SD of FGF-1 requires substantial redesign of both the primary and tertiary structure of the protein. Furthermore, FGF-1 is a weak mesophile as regards thermostability (ΔG=21.0 kJ/mol);[22] thus, mutational change must be generally compatible with folding and stability. The transforms applied in the SD of FGF-1 followed a logic of imposing a cumulative symmetric constraint on core, reverse-turn, and β-strand secondary structure, respectively, by targeted mutagenesis. Symmetric mutations were retained if they were either neutral or favorable to protein stability, folding, and solubility. Such properties are essential for protein function;[56] thus, SD includes a screen for function-competent biophysical properties (but no specific functionality). Whenever possible, prior mutational data were utilized in identifying potential symmetric mutations. Sequence comparison between the three trefoil-fold subdomains of FGF-1 was utilized to identify candidate symmetric mutations; for example, if the same residue was present at two out of three symmetry-related positions, mutation to that residue at the third position was typically evaluated. Potential symmetric mutations were analyzed by molecular modeling using FGF-1[28] or mutant X-ray structures. A description of the individual transforms in SD of FGF-1 follows.

Transform #1: SD of core positions

Previously published work has detailed the development of a series of symmetric core-packing mutations (SYM2-5[23,24]; Fig. 1) that systematically introduced a symmetric primary-structure constraint on the central hydrophobic core of FGF-1. Without changes to the FGF-1 tertiary structure, only limited symmetry within the core region was possible without also incurring substantial destabilization (compare SYM5 and SYM6);[24] however, deletion mutations that increased the tertiary-structure symmetry enabled the accommodation of a symmetric and thermostable hydrophobic core (SYM6ΔΔ).[25] Thus, the SD method included both primary- and tertiary-structure symmetric constraint mutations. Production of the SYM8ΔΔ mutant represents essential completion of transform #1, and this mutant was utilized as the starting point for transform #2.

Transform #2: SD of reverse turns

The SYM8ΔΔ mutant was initially modified by the inclusion of two previously described stabilizing point mutations (Lys12Val and Pro134Val)[27] and reversion of previously introduced Cys mutations at positions 42 and 130 (due to undesirable thiol reactivity) to produce the SYM9ΔΔ mutant (Fig. 3). The Lys12Val and Pro134Val mutations are located in the adjacent first and last β-strands of the β-barrel of the architecture, respectively. The N- and C-terminus β-strands represent a "discontinuous" β-turn, as the 3-fold symmetry-related positions describe β-turns #4 and #8 (see Figs. 1 and 5). The break between the N- and C-terminus β-strands describes a region of known structural weakness,[57] and the Lys12Val and Pro134Val mutations stabilize these adjacent β-strand interactions.[27] A His93Gly mutation, located within β-turn #8 and reported to stabilize FGF-1,[58,59] was introduced into the SYM9ΔΔ mutant to produce the SYM10ΔΔ mutant. This mutation increased the primary-structure symmetry between symmetry-related β-turns #4 and #8. Prior studies of the role of the statistically preferred β-turn motif Asx-Pro-Asx-Gly at individual β-turns #2, 6, and 10 in FGF-1 indicated a consistently favorable effect on thermostability with the sequence Asn-Xxx-Asp-Gly (where Xxx represents the retained FGF-1 residue at this position).[60] The canonical Asn-Pro-Asp-Gly sequence was therefore substituted into the SYM10ΔΔ mutant at each of these symmetry-related β-turns to produce the SYM11ΔΔ mutant (Fig. 3). The SD of β-turn regions was not taken to completion at this stage, and the SYM11ΔΔ mutant was considered a successful "proof of concept" and was utilized as the starting point for transform #3.

Transform #3: SD of β-strands

Taking advantage of the Lys12Val mutation (described above), we introduced an Asn95Val mutation into SYM11ΔΔ resulting in a symmetric (Val) deconstruction of positions 12, 54, and 95 within β-strands 1, 5, and 9, respectively. Similarly, Leu46Val and Glu87Val mutations were introduced, resulting in a further symmetric (Val) constraint within symmetry-related β-strands 4, 8, and 12, respectively. The combination of these mutations resulted in the SYM12ΔΔ mutant (Fig. 4). Ile56Leu and Tyr97Leu mutations (identified as structurally compatible symmetric mutations by manual modeling) were subsequently introduced into the SYM12ΔΔ mutant, increasing the symmetric constraint within β-strands 1, 5 and 9, respectively, and producing the SYM13ΔΔ mutant. The SD of β-strand regions was not taken to completion at this stage, and the SYM13ΔΔ mutant was considered a successful proof of concept.

Completion of transforms #2 and #3 was accomplished simultaneously via a chimera design strategy. The SYM13ΔΔ mutant contained two buried free Cys residues at positions 16 and 83, which were known to substantially limit the functional half-life of FGF-1 via thiol reactivity while in the unfolded and exposed state.[61,62] The local packing environment around each of these residues is highly optimized for Cys and intolerant to substitution without significant destabilization.[63] Subsequently, a chimera design was chosen so as to substitute the packing environment surrounding positions Cys16 and Cys83 with the packing environment surrounding the symmetry-related non-Cys residues Ser58 and Ile42, respectively. Threading of this chimera sequence onto the SYM6ΔΔ/Phe108Tyr X-ray structure suggested no bad contacts, either within individual trefoil-fold subdomains or, critically, at the subdomain interface(s), with one exception involving residues within symmetry-related β-turns 4, 8 and 12. These turns were therefore designed using a simple $(Gly)_3$ linker sequence. A single Pro residue was retained at position 10 (the initial residue of the first trefoil-fold subdomain) to promote solvent accessibility of the preceding N-terminal hexahistidine [$(His)_6$]-FGF-1(1–10) residue sequence. The resulting chimera sequence was termed "Symfoil-1" (for symmetric β-trefoil protein #1; Fig. 4). The residue numbering of all synthetic mutants was chosen so as to correspond as closely as possible with the numbering of the initial FGF-1 protein.

Transform #4: Stability optimization/symmetric fragmentation

Since the intermediate asymmetric "SYMΔΔ" mutant proteins demonstrated higher stability than Symfoil-1, optimization of Symfoil-1 stability was pursued prior to attempting fragmentation of the repeating peptide motif. Manual inspection of the X-ray structure of Symfoil-1 indicated that symmetry-related residue positions Val46, 87, and 134 (located at the solvent interface at the top of the β-barrel) each contain a "crevice" adjacent to Val $C^{\gamma 2}$, suggesting that an Ile mutation might be accommodated with enhanced van der Waals interactions. Subsequently, Ile mutations were introduced at each of these positions to create the Symfoil-2 variant (Fig. 6). Symmetry-related β-turns 4 and 8 were initially designed to be simple $(Gly)_3$ linkers, although these were considered to be a nonoptimal β-turn solution. "Asn scanning" of positions i (residue 49) through $i+3$ (residue 52) was performed on β-turn 4 in the Symfoil-2 protein and demonstrated a stability improvement with Asn mutation at the $i+2$ position. Asn mutations were therefore introduced at symmetry-related positions 51, 92, and 139 of the Symfoil-2 protein to produce Symfoil-3. A Gln40Pro mutation was reported to stabilize the FGF-1 protein.[64] Since Gln40 was conserved in Symfoil-3, Pro mutations were introduced at symmetry-related positions 40, 81, and 128 of Symfoil-3 to produce Symfoil-4; alternative constructs utilizing either Val or Thr mutations were also introduced at these positions, and these different forms are identified as Symfoil-4T, V, or P mutants, respectively. The Symfoil-4P mutant (the most stable Symfoil variant) was selected as the starting point for the symmetric fragmentation of the 42-residue repeating polypeptide motif. Dimer and monomer trefoil-fold subdomains were constructed by mutagenesis of Symfoil-4P; the resulting monomeric and dimeric trefoil-fold polypeptides are referred to as Monofoil-4P and Difoil-4P, respectively (for monomeric trefoil fold and dimeric trefoil fold, respectively).

Mutagenesis and protein purification

Construction of all the SYMΔΔ mutants utilized a synthetic gene for the 140-amino-acid form of human FGF-1[28,65–67] containing an additional amino-terminal $(His)_6$ tag as previously described.[23] The QuikChange™ site-directed mutagenesis protocol (Agilent Technologies,

Santa Clara, CA) was used to introduce all mutations, which were confirmed by nucleic acid sequence analysis (Biomolecular Analysis Synthesis and Sequencing Laboratory, Florida State University). Construction of the Symfoil-1 protein involved complete gene synthesis utilizing unique codons at symmetry-related positions such that these could be individually targeted during subsequent mutagenesis. The Monofoil-4P and Difoil-4P genes were generated by mutation of the Symfoil-4P gene to introduce a stop codon at position Glu53 or Glu94, respectively. Additionally, for crystallization trials of Monofoil-4P and Difoil-4P polypeptides, the codons for N-terminus residues 1–10 were deleted (i.e., the N-terminus $(His)_6$-tag was followed immediately by residue position 11 of the FGF-1 numbering scheme). Expression and purification of recombinant proteins followed previously published procedures[23] and utilized Ni-NTA chelation and Superdex 75 size-exclusion chromatography (GE Healthcare, Piscataway, NJ). Purified protein was exchanged into crystallization buffer [50 mM sodium phosphate, 0.1 M NaCl, 10 mM $(NH_4)_2SO_4$, 2 mM dithiothreitol (DTT), pH 7.5] for crystallization studies (with DTT omitted for all Symfoil protein variants) or ADA buffer [20 mM *N*-(2-acetamido)iminodiacetic acid (ADA), 0.1 M NaCl, 2 mM DTT, pH 6.6] for biophysical studies (with DTT omitted for all Symfoil protein variants). An extinction coefficient of E_{280} nm (0.1%, 1 cm) = 1.26[68,69] was used to determine protein concentration for FGF-1. Due to the variation in number of Trp, Tyr, and Cys residues, the extinction coefficient for all mutant forms was determined by the method of Gill and von Hippel.[70]

Isothermal equilibrium denaturation

Isothermal equilibrium denaturation by GuHCl was performed as previously described[22] with either fluorescence or circular dichroism (CD) as the spectroscopic probe. FGF-1 unfolding monitored by CD spectroscopy exhibits excellent agreement with results obtained by fluorescence spectroscopy and is a useful alternative spectroscopic probe in cases where fluorescence cannot be utilized.[59] FGF-1 contains a single buried Trp residue at position 107, and although this exhibits atypical native-state quenching, it is nonetheless useful as a spectroscopic probe of unfolding. This native-state quenching appears due to an adjacent Pro121 residue that is deleted in all the SYMΔΔ mutants; thus, for these mutants there is little discrimination between the native- and denatured-state fluorescence, and the unfolding was followed by CD signal. All Symfoil mutants delete Trp107, and their fluorescence signal is largely contributed by Tyr residues at positions 22, 64, and 108. Unfolding was followed by both fluorescence and CD, which were found to be in excellent agreement (similarly for Monofoil-4P and Difoil-4P polypeptides); subsequently, all reported values for Symfoil mutants are fluorescence data. Fluorescence data were collected on a Cary Eclipse fluorescence spectrophotometer (Varian Inc., Palo Alto, CA) equipped with a Pelletier controlled-temperature regulator at 298 K and using a 1.0 -cm path-length cuvette. Protein samples (5.0 μM) were equilibrated in ADA buffer at 298 K in 0.1 M increments of GuHCl. For unfolding analysis monitored by CD, 25 μM protein samples were equilibrated in ADA buffer at 298 K in 0.1 M increments of GuHCl. CD data were collected on a Jasco model 815 CD spectrophotometer (Jasco Inc, Easton, MD) equipped with a Pelletier controlled-temperature regulator at 298 K and using a 1-mm path-length cuvette. Both fluorescence and CD data were analyzed with the general-purpose nonlinear least-squares fitting program DataFit (Oakdale Engineering, Oakdale, PA) implementing a six-parameter, two-state model[71] as previously described.[22] The effect of a given mutation on the stability of the protein ($\Delta\Delta G$) was calculated by taking the difference between the midpoint of denaturation (C_m value) for reference and mutant proteins and multiplying by the average of the m values, as described by Pace and Scholtz,[72] and where a negative value indicates the mutation is stabilizing in relationship to the reference protein. In the case of Monofoil-4P and Difoil-4P proteins, the data were analyzed with a trimer-to-monomer isothermal equilibrium denaturation model,[73,74] where subunit dissociation occurs simultaneously with two-state unfolding:

$$N_3 \Leftrightarrow 3D \tag{1}$$

and the equilibrium dissociation constant K_D is expressed as:

$$K_D = \frac{27C_t^2F_D^3}{1 - F_D} \tag{2}$$

where C_t is the molar concentration of trimer form and F_D is the fraction of denatured monomer. See Backmann *et al.*[73] for details of the derivation of the function describing the experimental spectroscopic data in terms of the fitted thermodynamic parameters.

Differential scanning calorimetry

All DSC data were collected on a VP-DSC microcalorimeter (GE Healthcare) as previously described.[22] Briefly, 40 μM protein samples were equilibrated at 298 K in ADA buffer without DTT and in the presence of varying concentrations of GuHCl. A scan rate of 15 K/h was used for all proteins to meet the equilibrium assumption of the thermodynamic model. Molar heat capacity data were analyzed with the DSCfit software package.[75] The Monofoil-4P and Difoil-4P mutants were analyzed with a trimer-to-monomer thermal denaturation model[73,74] (described above) and implemented with the DataFit nonlinear least-squares fit software package (Oakdale Engineering) (see Backmann *et al.*[73] for details of the derivation of the function describing the experimental molar heat capacity data in terms of the fitted thermodynamic parameters).

Folding/unfolding kinetic analysis

Folding and unfolding kinetic data followed previously described methods.[59] Briefly, denatured protein samples for folding kinetics measurements were prepared by overnight dialysis against ADA buffer containing 2.5–5.0 M GuHCl. All folding kinetic data were collected with an Applied Photophysics SX20 stopped-flow system (Applied Photophysics Ltd., Surrey, UK) at 298 K. Folding was initiated by a 1:10 dilution of 20 μM denatured protein into ADA buffer with denaturant concentrations varying

in increments of 0.05 or 0.1 M up to the midpoint of denaturation, as determined by isothermal equilibrium denaturation measurements. The data collection strategy was designed to span approximately five half-lives or >97% of the expected fluorescence signal change between the fully denatured and native states. For unfolding kinetics measurements due to the comparatively slower kinetics, unfolding kinetics measurements were performed by manual mixing. Protein samples (~20 μM) were dialyzed against ADA buffer overnight at 298 K. Unfolding was initiated by a 1:10 dilution into ADA buffer with a final GuHCl concentration of 1.5–7.5 M in 0.2 M increments. All unfolding data were collected with a Varian Eclipse fluorescence spectrophotometer equipped with a Pelletier controlled-temperature unit at 298 K. Data collection times for each protein were designed so as to quantify the fluorescence signal over three to four half-lives or >93% of the total expected amplitude. The kinetic rates and amplitudes *versus* denaturant concentration were calculated from the time-dependent change in fluorescence with a single-exponential model. Folding and unfolding rate constant data were fit to a global function describing the contribution of both rate constants to the observed kinetics as a function of denaturant as described by Fersht.[76]

X-ray crystallization studies

Purified mutant protein in crystallization buffer was concentrated to 9–15 mg/ml, and crystal screening was performed with either the hanging-drop or sitting-drop vapor diffusion method at room temperature. Diffraction-quality crystals of a Phe108Tyr mutation of the SYM6ΔΔ protein grew in 1 week from vapor diffusion against 0.2 M magnesium formate. Crystals were mounted with Hampton Research nylon mounted cryoturns and cryocooled in a stream of gaseous nitrogen at 100 K. Diffraction data were collected at the Southeast Regional Collaborative Access Team (SET-CAT) 22-BM beam line ($\lambda = 1.00$ Å) at the Advanced Photon Source, Argonne National Laboratory, using a MarCCD 300 detector (Mar USA, Evanston, IL). A single-crystal diffraction data set was collected, and diffraction data were indexed, integrated, and scaled with the DENZO or HKL2000 software package.[77,78] Molecular replacement and refinement utilized the PHENIX software package,[79] with 5% of the data in the reflection files set aside for R_{free} calculations.[80] His-tagged FGF-1 (PDB code 1JQZ) was used as the search model in molecular replacement for the SYM6ΔΔ/Phe108Tyr mutant. Model building and visualization utilized the COOT molecular graphics software.[81]

SPR studies of FGFR-1c binding

FGF-1 and mutant proteins were immobilized as the "ligand" on the SPR sensor chip, and the extracellular domain of human FGFR-1c was used as the soluble "analyte" in all studies. Recombinant FGFR-1c protein was expressed and purified from an insect cell–Baculovirus system as previously described.[82] The FGFR-1c recombinant protein contains the ligand-binding immunoglobulin-like domains D2, D3, and interconnecting linker (residues 131 to 365) and includes an N-terminal His tag for purification purposes. Prior to SPR analysis, the purified FGFR was dialyzed against HBS-EP+ buffer [10 mM Hepes, 150 mM NaCl, 3 mM EDTA (ethylenediaminetetraacetic acid), 0.005% (v/v) surfactant P20, pH 7.4], and 1:2 serial dilutions spanning a concentration range of 256 –1.0 nM were made in 1× HBS-EP+ buffer; additionally, the FGF-1 and mutant proteins were passed over a HiLoad Superdex 75 26/60 size-exclusion column (GE Healthcare) to ensure a monodisperse sample and then diluted to 1–10 μg/ml in 10 mM Mes buffer (pH 6.0). The sensor chip was prepared by immobilizing FGF-1 or mutant protein on a Series S Sensor Chip CM5 (GE Healthcare, Uppsala, Sweden) by covalent amine coupling following the manufacturer's suggested protocol. Chip sample surfaces were prepared with a target surface density of ~50 RU. The reference cell surface was prepared under identical conditions with only buffer injections. A surface stability test and mass transfer control experiment were performed to optimize the conditions for interaction kinetics analyses. SPR data were collected at 25 °C on a Biacore T-100 instrument (GE Healthcare). The association/dissociation phase was measured by flowing 0–256 nM FGFR-1c analyte over the FGF-1/mutant protein CM5 Sensor Chip at a flow rate of 75 μl/min for 280 s. At the end of each sample injection, HBS-EP+ buffer was passed over the sensor chip at the same flow rate for 400 s to monitor the dissociation phase. NaCl (2.5 M) was injected at 50 μl/min for 120 s to fully regenerate the sensor surface. The control flow cell response was subtracted from the ligand sample cell for each receptor injection; the 0 M concentration sensorgram values were subsequently subtracted from the analyte runs, and the resulting sensorgrams were analyzed with the Biacore T100 Evaluation v2.0 software package (GE Healthcare). Association (k_a) and dissociation (k_d) rate constants were determined by fitting with the "bivalent analyte" model, based on the known X-ray structure of human FGF-1 in complex with FGFR-1c (PDB code 1EVT).[83]

3T3 fibroblast mitogenic assays

Purified protein was equilibrated in TBS buffer [0.14 M NaCl, 5.1 mM KCl, 0.7 mM Na_2HPO_4, and 24.8 mM Tris base (pH 7.4)], and assay of the mitogenic activity toward 3T3 fibroblasts was performed as previously described.[25] Briefly, NIH 3T3 fibroblasts were plated in Dulbecco's modified Eagle's medium (Invitrogen, Carlsbad, CA) supplemented with 0.5% (v/v) newborn calf serum (Sigma-Aldrich Corp., St. Louis, MO) for 48 h at 37 °C with 5% (v/v) CO_2. The quiescent serum-starved cells were stimulated with fresh medium supplemented with FGF-1 or mutant protein (0–10 μg/ml) and incubated for an additional 48 h. After this incubation period, the cells were counted with a hemacytometer (Hausser Scientific, Horsham, PA). The protein concentration yielding one-half maximal cell density (EC_{50} value) was used for quantitative comparison of functional mitogenic activity.

ITC Tris-binding studies

All ITC data were collected on a VP-ITC microcalorimeter (MicroCal LLC, Northampton, MA). Titrations were performed at 298 K, and all samples were equilibrated in

50 mM sodium phosphate and 0.1 M NaCl (pH 8.0) buffer. All samples were filtered and degassed for 10 min prior to loading. Symfoil-2 (200 μM) was titrated with 20 mM Tris; each injection was performed over 12 s with a postinjection equilibration period of 240 s. The samples were titrated against 40 injections at 6 μl per injection of Tris. The titration curves were fit with the manufacturer's software (MicroCal Origin) employing a model with a single ligand-binding site. The Symfoil-2 protein was selected for titration studies due to quantity of protein available and the shared Tris-binding property observed among the set of Symfoil mutant proteins. Fitting of the ITC data, while allowing the stoichiometric parameter n to float, did not result in convergence; however, since the X-ray data were unambiguous in showing a 1:1 protein–Tris complex, the stoichiometric n parameter was fixed to 1.0, thus enabling fitting convergence.

Protease-resistance studies

FGF-1 and mutant proteins (0.025 mM) were incubated with trypsin (Promega, Madison, WI) (0.25 μM; for 100:1 molar ratio, respectively) in TBS buffer at 37 °C to evaluate resistance to proteolysis. Time points were taken at various intervals (spanning minutes to days, depending on the particular mutant) and added to SDS sample buffer and immediately incubated at 95 °C for 5 min to halt the digestion reaction. Samples were resolved on 16.5% Tricine SDS-PAGE visualized with Coomassie Brilliant Blue staining. The stained gels were scanned and the amount of intact protein was quantified with UN-SCAN-IT densitometry software (Silk Scientific, Orem, UT).

PDB accession numbers

Model coordinates for the refined SYM6ΔΔ/Phe108Tyr structure have been deposited in the PDB with code 3O3Q. Crystal structures of the Symfoil-1, Symfoil-2, Symfoil-4T, -4V, -4P, Monofoil-4P and Difoil-4P proteins have previously been reported.[21]

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.02.002

Acknowledgements

This work was funded by grant 0655133B from the American Heart Association and by additional research support provided by the Florida State University College of Medicine. We thank Dr. Claudius Mundoma (Physical Biochemistry Facility) and Dr. T. Somasundaram (X-ray Crystallography Facility, Kasha Laboratory Institute of Molecular Biophysics) for valuable suggestions and technical assistance. We also thank Ms. Pushparani Dhanarajan (Molecular Cloning Facility, Department of Biological Sciences, Florida State University) for helpful comments. Analytical ultracentrifugation calculations were performed on the Ultra Scan LIMS cluster at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio and the Lonestar cluster at the Texas Advanced Computing Center, supported by NSF Teragrid grant MCB070038 (to Borries Demeler). We acknowledge the instrumentation facilities at Biomedical Proteomics Laboratory, College of Medicine at Florida State University. Use of the "mail-in crystallography" facility of SER-CAT for diffraction data collection is acknowledged. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. W-31-109-Eng-38. The SYM6ΔΔ/Phe108Tyr mutant X-ray structure has been deposited in the PDB (accession 3O3Q).

References

1. DeGrado, W. F., Regan, L. & Ho, S. P (1987). The design of a four-helix bundle protein. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 521–526.
2. Regan, L. & DeGrado, W. F. (1988). Characterization of a helical protein designed from first principles. *Science*, **241**, 976–978.
3. Richardson, J. & Richardson, D. C. (1989). The *de novo* design of protein structures. *Trends Biochem. Sci.* **14**, 304–309.
4. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. (1990). De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science*, **249**, 884–891.
5. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S. & Richardson, D. C. (1994). Betadoublet: *de novo* design, synthesis, and characterization of a β-sandwich protein. *Proc. Natl Acad. Sci. USA*, **91**, 8747–8751.
6. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T. & DeGrado, W. F. (1995). Protein design: a hierarchic approach. *Science*, **270**, 935–941.
7. Fu, X., Kono, H. & Saven, J. G. (2003). Probabilistic approach to the design of symmetric protein quaternary structures. *Protein Eng.* **16**, 971–977.
8. Offredi, F., Dubail, F., Kischel, P., Sarinski, K., Stern, A. S., Van de Weerdt, C. *et al.* (2003). De novo backbone and sequence design of an idealized α/β-barrel protein: evidence of stable tertiary structure. *J. Mol. Biol.* **325**, 163–174.
9. Tsai, H. H., Tsai, C. J., Ma, B. & Nussinov, R. (2004). In silico protein design by combinatorial assembly of protein building blocks. *Protein Sci.* **13**, 2753–2765.
10. Heinemann, M. & Panke, S. (2006). Synthetic biology—putting engineering into biology. *Bioinformatics*, **22**, 2790–2799.
11. Haspel, N., Wainreb, G., Inbar, Y., Tsai, H. H., Tsai, C. J., Wolfson, H. J. & Nussinov, R. (2007). A hierarchical protein folding scheme based on the building block folding model. *Methods Mol. Biol.* **350**, 189–204.

12. He, Y., Zhou, R., Huang, Y. & Xiao, Y (2009). Foldable subunits of helix protein. *Comput. Biol. Chem.* **33**, 325–328.
13. Armstrong, C. T., Boyle, A. L., Bromley, E. H., Mahmoud, Z. N., Smith, L., Thomson, A. R. & Woolfson, D. N. (2009). Rational design of peptide-based building blocks for nanoscience and synthetic biology. *Faraday Discuss.* **143**, 305–317.
14. Andre, I., Bradley, P., Wang, C. & Baker, D. (2007). Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. USA*, **104**, 17656–17661.
15. Wolynes, P. G. (1996). Symmetry and the energy landscapes of biomolecules. *Proc. Natl Acad. Sci. USA*, **93**, 14249–14255.
16. Lowe, A. R. & Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc. Natl Acad. Sci. USA*, **104**, 2679–2684.
17. Wootton, J. C. & Federhen, S. (1993). Statistics of local complexity in amino acid sequences and sequence databases. *Computers Chem.* **17**, 149–163.
18. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered proteins. *Proteins*, **42**, 38–48.
19. Wright, C. F., Teichmann, S. A., Clarke, J. & Dobson, C. M. (2005). The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*, **438**, 878–881.
20. Hoang, T. X., Marsella, L., Trovato, A., Seno, F., Banavar, J. R. & Maritan, A. (2006). Common attributes of native-state structures of proteins, disordered proteins, and amyloid. *Proc. Natl Acad. Sci. USA*, **103**, 6883–6888.
21. Lee, J. & Blaber, M. (2011). Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc. Natl Acad. Sci. USA*, **108**, 126–130.
22. Blaber, S. I., Culajay, J. F., Khurana, A. & Blaber, M. (1999). Reversible thermal denaturation of human FGF-1 induced by low concentrations of guanidine hydrochloride. *Biophys. J.* **77**, 470–477.
23. Brych, S. R., Blaber, S. I., Logan, T. M. & Blaber, M. (2001). Structure and stability effects of mutations designed to increase the primary sequence symmetry within the core region of a β-trefoil. *Protein Sci.* **10**, 2587–2599.
24. Brych, S. R., Kim, J., Logan, T. M. & Blaber, M. (2003). Accommodation of a highly symmetric core within a symmetric protein superfold. *Protein Sci.* **12**, 2704–2718.
25. Brych, S. R., Dubey, V. K., Bienkiewicz, E., Lee, J., Logan, T. M. & Blaber, M. (2004). Symmetric primary and tertiary structure mutations within a symmetric superfold: a solution, not a constraint, to achieve a foldable polypeptide. *J. Mol. Biol.* **344**, 769–780.
26. Dubey, V. K., Lee, J. & Blaber, M. (2005). Redesigning symmetry-related "mini-core" regions of FGF-1 to increase primary structure symmetry: thermodynamic and functional consequences of structural symmetry. *Protein Sci.* **14**, 2315–2323.
27. Dubey, V. K., Lee, J., Somasundaram, T., Blaber, S. & Blaber, M. (2007). Spackling the crack: stabilizing human fibroblast growth factor-1 by targeting the N and C terminus β-strand interactions. *J. Mol. Biol.* **371**, 256–268.
28. Blaber, M., DiSalvo, J. & Thomas, K. A. (1996). X-ray crystal structure of human acidic fibroblast growth factor. *Biochemistry*, **35**, 2086–2094.
29. Orengo, C. A., Jones, D. T. & Thornton, J. M. (1994). Protein superfamilies and domain superfolds. *Nature*, **372**, 631–634.
30. Thornton, J. M., Orengo, C. A., Todd, A. E. & Pearl, F. M. (1999). Protein folds, functions and evolution. *J. Mol. Biol.* **293**, 333–342.
31. Kim, C., Basner, J. & Lee, B. (2010). Detecting internally symmetric protein structures. *BMC Bioinf.* **11**, 303–318.
32. Ghirlanda, G., Lear, J. D., Ogihara, N. L., Eisenberg, D. & DeGrado, W. F. (2002). A hierarchic approach to the design of hexameric helical barrels. *J. Mol. Biol.* **319**, 243–253.
33. Pantoliano, M. W., Horlick, R. A., Springer, B. A., Van Dyk, D. E., Tobery, T., Wetmore, D. R. *et al.* (1994). Multivalent ligand-receptor binding interactions in the fibroblast growth factor system produce a cooperative growth factor and heparin mechanism for receptor dimerization. *Biochemistry*, **33**, 10229–10248.
34. Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature*, **407**, 1029–1034.
35. Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A. *et al.* (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol. Cell. Biol.* **6**, 743–750.
36. Bashkin, P., Doctrow, S., Klagsbrun, M., Svahn, C. M., Folkman, J. & Vlodavsky, I. (1989). Basic fibroblast growth factor binds to subendothelial extracellular matrix and is released by heparitinase and heparin-like molecules. *Biochemistry*, **28**, 1737–1743.
37. Weiner, H. L. & Swain, J. L. (1989). Acidic fibroblast growth factor mRNA is expressed by cardiac myocytes in culture and the protein is localized to the extracellular matrix. *Proc. Natl Acad. Sci. USA*, **86**, 2683–2687.
38. Chintala, S. K., Miller, R. R. & McDevitt, C. A. (1994). Basic fibroblast growth factor binds to heparan sulfate in the extracellular matrix of rat growth plate chondrocytes. *Arch. Biochem. Biophys.* **310**, 180–186.
39. Schreiber, G., Buckle, A. M. & Fersht, A. R. (1994). Stability and function: two constraints in the evolution of barstar and other proteins. *Structure*, **2**, 945–951.
40. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA*, **92**, 452–456.
41. Beadle, B. M. & Shoichet, B. K. (2002). Structural basis of stability–function tradeoffs in enzymes. *J. Mol. Biol.* **321**, 285–296.
42. Bloom, J. D., Wilke, C. O., Arnold, F. H. & Adami, C. (2004). Stability and the evolvability of function in a model protein. *Biophys. J.* **86**, 2758–2764.
43. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA*, **103**, 5869–5874.
44. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002.

45. Auton, M. & Bolen, D. W. (2005). Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl Acad. Sci. USA*, **102**, 15065–15068.
46. Argos, P. & Palau, J. (1982). Amino acid distribution in protein secondary structures. *Int. J. Pept. Protein Res.* **19**, 380–393.
47. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. (1990). Relative helix-forming tendencies of nonpolar amino acids. *Nature*, **344**, 268–270.
48. Blaber, M., Zhang, X. J. & Matthews, B. W. (1993). Structural basis of amino acid α helix propensity. *Science*, **260**, 1637–1640.
49. Kim, C. A. & Berg, J. M. (1993). Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. *Nature*, **362**, 267–270.
50. Minor, D. L., Jr & Kim, P. S. (1994). Measurement of the β-sheet-forming propensities of amino acids. *Nature*, **367**, 660–663.
51. Smith, C. K., Withka, J. M. & Regan, L. (1994). A thermodynamic scale for the β-sheet forming tendencies of the amino acids. *Biochemistry*, **33**, 5510–5517.
52. Chaudhuri, I., Soding, J. & Lupas, A. N. (2008). Evolution of the β-propeller fold. *Proteins*, **71**, 795–803.
53. Corey, E. J. & Cheng, X. M. (1989). *The Logic of Chemical Synthesis*. John Wiley & Sons, Inc., New York.
54. McLachlan, A. D. (1979). Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). *J. Mol. Biol.* **133**, 557–563.
55. Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). β-Trefoil fold patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *J. Mol. Biol.* **223**, 531–543.
56. Soskine, M. & Tawfik, D. S. (2010). Mutational effects and the evolution of new protein functions. *Nat. Rev.* **11**, 572–582.
57. Bernett, M. J., Somasundaram, T. & Blaber, M. (2004). An atomic resolution structure for human fibroblast growth factor 1. *Proteins*, **57**, 626–634.
58. Arakawa, T., Horan, T. P., Narhi, L. O., Rees, D. C., Schiffer, S. G., Holst, P. L. *et al.* (1993). Production and characterization of an analog of acidic fibroblast growth factor with enhanced stability and biological activity. *Protein Eng.* **6**, 541–546.
59. Kim, J., Brych, S. R., Lee, J., Logan, T. M. & Blaber, M. (2003). Identification of a key structural element for protein folding within β-hairpin turns. *J. Mol. Biol.* **328**, 951–961.
60. Lee, J., Dubey, V. K., Longo, L. M. & Blaber, M. (2008). A logical OR redundancy with the Asx-Pro-Asx-Gly type I β-turn motif. *J. Mol. Biol.* **377**, 1251–1264.
61. Culajay, J. F., Blaber, S. I., Khurana, A. & Blaber, M. (2000). Thermodynamic characterization of mutants of human fibroblast growth factor 1 with an increased physiological half-life. *Biochemistry*, **39**, 7153–7158.
62. Lee, J. & Blaber, M. (2009). The interaction between thermostability and buried free cysteines in regulating the functional half-life of fibroblast growth factor-1. *J. Mol. Biol.* **393**, 113–127.
63. Lee, J. & Blaber, M. (2009). Structural basis for conserved cysteine in the fibroblast growth factor family: evidence for a vestigial half-cystine. *J. Mol. Biol.* **393**, 128–139.
64. Zakrzewska, M., Krowarsch, D., Wiedlocha, A., Olsnes, S. & Otlewski, J. (2005). Highly stable mutants of human fibroblast growth factor-1 exhibit prolonged biological action. *J. Mol. Biol.* **352**, 860–875.
65. Gimenez-Gallego, G., Conn, G., Hatcher, V. B. & Thomas, K. A. (1986). The complete amino acid sequence of human brain-derived acidic fibroblast growth factor. *Biochem. Biophys. Res. Commun.* **128**, 611–617.
66. Linemeyer, D. L., Menke, J. G., Kelly, L. J., Disalvo, J., Soderman, D., Schaeffer, M. T. *et al.* (1990). Disulfide bonds are neither required, present, nor compatible with full activity of human recombinant acidic fibroblast growth factor. *Growth Factors*, **3**, 287–298.
67. Ortega, S., Schaeffer, M. T., Soderman, D., DiSalvo, J., Linemeyer, D. L., Gimenez-Gallego, G. & Thomas, K. A. (1991). Conversion of cysteine to serine residues alters the activity, stability, and heparin dependence of acidic fibroblast growth factor. *J. Biol. Chem.* **266**, 5842–5846.
68. Zazo, M., Lozano, R. M., Ortega, S., Varela, J., Diaz-Orejas, R., Ramirez, J. M. & Gimenez-Gallego, G. (1992). High-level synthesis in *Escherichia coli* of a shortened and full-length human acidic fibroblast growth factor and purification in a form stable in aqueous solutions. *Gene*, **113**, 231–238.
69. Tsai, P. K., Volkin, D. B., Dabora, J. M., Thompson, K. C., Bruner, M. W., Gress, J. O. *et al.* (1993). Formulation design of acidic fibroblast growth factor. *Pharm. Res.* **10**, 649–659.
70. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
71. Eftink, M. R. (1994). The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophys. J.* **66**, 482–501.
72. Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: a Practical Approach* (Creighton, T. E., ed.), pp. 299–321, Oxford University Press, Oxford.
73. Backmann, J., Schafer, G., Wyns, L. & Bonisch, H. (1998). Thermodynamics and kinetics of unfolding of the thermostable trimeric adenylate kinase from the archaeon *Sulfolobus acidocaldarius*. *J. Mol. Biol.* **284**, 817–833.
74. Jelesarov, I. & Lu, M. (2001). Thermodynamics of trimer-of-hairpins formation by the SIV gp41 envelope protein. *J. Mol. Biol.* **307**, 637–656.
75. Grek, S. B., Davis, J. K. & Blaber, M. (2001). An efficient, flexible-model program for the analysis of differential scanning calorimetry protein denaturation data. *Protein Pept. Lett.* **8**, 429–436.
76. Fersht, A. R. (1999). *Kinetics of Protein Folding. Structure and Mechanism in Protein Science*. W. H. Freeman and Co., New York.
77. Otwinowski, Z. (1993). Proceedings of the CCP4 Study Weekend: Data Collection and Processing. SERC Daresbury Laboratory, England.
78. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
79. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Loerger, T. R., McCoy, A. J. *et al.* (2008).

Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435.
80. Brunger, A. T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
81. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
82. Lee, J., Blaber, S. I., Irsigler, A., Aspinwall, E. & Blaber, M. (2009). X-ray structure and biophysical properties of rabbit fibroblast growth factor 1. *Acta Crystallogr., Sect. F.* **65**, 1097–1104.
83. Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. & Mohammadi, M. (2000). Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell*, **101**, 413–424.