Supporting Information

M. Longo et al. 10.1073/pnas.1219530110

SI Materials and Methods

Protein Expression and Purification. Primitive version (PV)1 and PV2 mutants were constructed in the background of the de novo designed simplified β-trefoil protein (*Symfoil*-4P) synthetic gene (1). The constructs contained an additional amino-terminal $(His)_6$ tag but deleted residues 1–10 of the *Symfoil*-4P protein. These deleted residues are not part of the fundamental β-trefoil architecture but comprise an unstructured N-terminal extension and were deleted to promote crystallization and reduce the biotic amino acid composition. The numbering scheme of the FGF-1 protein is retained in the PV1 and PV2 mutants for purposes of comparison (Fig. 1). The QuikChange site-directed mutagenesis protocol (Agilent Technologies) was used to introduce all mutations, which were confirmed by nucleic acid sequence analysis (Biomolecular Analysis Synthesis and Sequencing Laboratory, Florida State University). Expression and purification of recombinant proteins followed previously published procedures (2) and used Ni-NTA chelation and Superdex 75 size-exclusion chromatography (GE Healthcare). Purified protein was exchanged into 50 mM sodium phosphate, 0.1 M NaCl, 10 mM $(NH_4)_2SO_4$, pH 7.5 ("crystallization buffer") for crystallization studies or 20 mM *N*-(2-acetamido)iminodiacetic acid (ADA), 0.1 M or 2.0 M NaCl, pH 6.6 ("ADA buffer") for all biophysical studies. The extinction coefficients for FGF-1 and *Symfoil*-4P mutant form were determined by the method of Gill and von Hippel (3); concentration of PV1 and PV2 were determined by bicinchoninic acid assay in reference to a known *Symfoil*-4P (1, 4) concentration standard.

X-Ray Crystallography. Purified mutant protein in crystallization buffer was concentrated to 10–15 mg/mL and crystal screening was performed using either the hanging-drop or sitting-drop vapor diffusion method at room temperature. Two different orthorhombic crystal forms of the PV1 mutant grew from 1.5 M $(NH_4)_2SO_4$, 0.10 M Li_2SO_4, and either 0.1 M Hepes pH 7.4 (form 1) or 0.1 M Tris pH 7.0 (form 2). An orthorhombic crystal form of the PV2 mutant grew from 1.5 M $(NH_4)_2SO_4$, 0.11 M Li_2SO_4, 0.1 M Tris pH 7.0. Crystals were mounted using Hampton Research nylon mounted cryo-turns and cryo-cooled in a stream of gaseous nitrogen at 100 K. Diffraction data for both PV1 crystals were collected at the ×25 beam line of the National Synchrotron Light Source at Brookhaven National Laboratory using an ADSC Q315 CCD detector. Diffraction data for PV2 were collected in-house, using a Rigaku RU-H2R rotating anode X-ray source equipped with an Osmic confocal mirrors (MarUSA) and a MarCCD165 detector. A single-crystal diffraction dataset was collected in each case and diffraction data were indexed, integrated, and scaled using the DENZO or HKL2000 software package (5, 6). Molecular replacement and refinement used the PHENIX software package (7), with 5% of the data in the reflection files set aside for R_{free} calculations (8). The structure was solved by molecular replacement, in which the *Symfoil*-4P de novo designed protein [Protein Data Bank (PDB) IDcode 3O4D] was used as the search model for PV1; subsequently, PV1 was used as the search model for PV2. Model building and visualization used the COOT molecular graphics software (9).

1. 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(1): 126–130.
2. Brych SR, Blaber SI, Logan TM, 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(12):2587–2599.
3. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182(2):319–326.
4. Lee J, Blaber SI, Dubey VK, Blaber M (2011) A polypeptide "building block" for the β-trefoil fold identified by "top-down symmetric deconstruction." *J Mol Biol* 407(5):744–763.
5. Otwinowski Z (1993) Oscillation data reduction program. *Proceedings of the CCP4 Study Weekend: "Data Collection and Processing,"* eds Sawyer L, Isaacs N, Bailey S (SERC Daresbury Laboratory, Daresbury, England), pp 56–62.
6. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276(part A):307–326.
7. Zwart PH, et al. (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419–435.
8. Brunger AT (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472–475.
9. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.

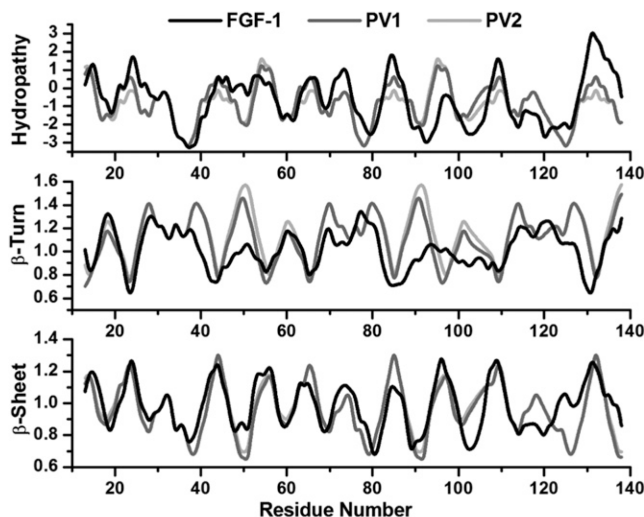

Fig. S1. Hydropathy, β-turn propensity, and β-strand propensity plots for FGF-1, PV1, and PV2 mutant proteins.

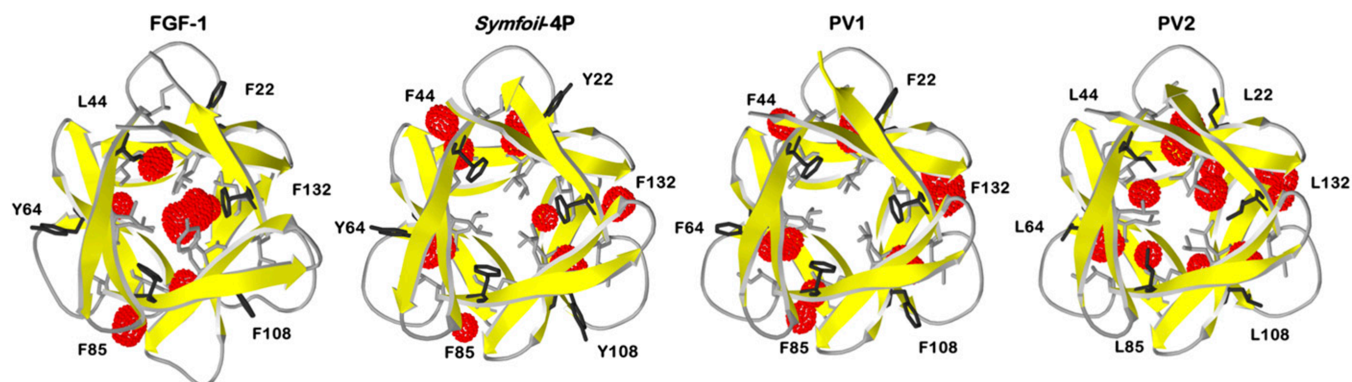

Fig. S2. Ribbon diagrams of FGF-1 (2AFG), *Symfoil*-4P (3O4D), PV1 (3QYX), and PV2 (4D8H) proteins from their respective X-ray structures. The view in each case is parallel to the threefold axis of internal rotational symmetry characteristic of the β-trefoil fold. The set of 21 hydrophobic packing groups is indicated by (light gray) stick representation, and six positions associated with aromatic side chains substituted by Leu in PV2 are indicated (dark gray). Also shown are the solvent excluded cavities (red) within each structure identified using a 1.2 Å probe radius (1).

1. Connolly ML (1993) The molecular surface package. *Journal of Molecular Graphics* 11:139–141.

Core Position	FGF-1	*Symfoil*-4P	PV1	PV2
14	L	L	L	L
22	F	Y	F	L
23	L	L	L	L
25	I	I	I	I
31	V	V	V	V
42	I	I	I	I
44	L	F	F	L
56	I	L	L	L
64	Y	Y	F	L
65	L	L	L	L
67	M	I	I	I
73	L	V	V	V
83	C	I	I	I
85	F	F	F	L
97	Y	L	L	L
108	F	Y	F	L
109	V	L	L	L
111	L	I	I	I
117	C	V	V	V
130	I	I	I	I
132	F	F	F	L
# Carbons	96	99	99	81
Alphabet size	7	5	4	3
% Pre-biotic	67	83	83	100

Fig. S3. The 21 solvent-excluded core-packing positions in FGF-1, *Symfoil*-4P, PV1, and PV2 mutant proteins. Also shown are number of carbons, alphabet size and percent prebiotic amino acids in these sets. The shaded positions identify amino acids belonging to the prebiotic set.

Table S1. Crystallographic data collection and refinement statistics

	PV1*	PV1[†]	PV2[‡]
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell constants (Å)			
	a = 38.5	*a* = 46.7	*a* = 46.6
	b = 46.6	*b* = 48.7	*b* = 48.7
	c = 63.9	*c* = 67.6	*c* = 64.9
Maximum resolution (Å)	1.50	1.40	1.90
Mosaicity (°)	0.43	0.40	0.49
Redundancy	6.4	12.8	12.3
Mol/ASU	1	1	1
Matthew coefficient ($Å^3$/Da)	1.98	2.65	2.30
Total reflections	120,892	398,211	247,513
Unique reflections	18,903	31,023	12,167
I/σ (overall)	43.4	65.9	38.2
I/σ (highest shell)	6.2	8.0	3.8
Completion overall (%)	99.6	99.9	97.2
Completion highest shell (%)	99.7	99.9	76.2
R_{merge} overall (%)	8.0	7.4	9.5
R_{merge} highest shell (%)	34.9	30.1	34.2
Nonhydrogen protein atoms	1,009	1,031	974
Solvent molecules/ion	154/10	217/13	177/1
R_{cryst} (%)	16.7	18.2	16.5
R_{free} (%)	21.1	20.5	21.4
rmsd bond length (Å)	0.006	0.006	0.006
rmsd bond angle (°)	1.08	1.10	1.07
Ramachandran plot			
Most favored (%)	93.5	95.7	100.0
Additional allowed (%)	6.5	4.3	0.0
Generously allowed (%)	0.0	0.0	0.0
Disallowed region (%)	0.0	0.0	0.0
PDB ID code	3Q7W	3Q7X	4D8H

Mol/ASU, molecules per asymmetric unit; I/Σ, intensity/standard deviation of intensity (signal-to-noise); R_{merge}, agreement among multiple measurements of the same reflection; R_{cryst}, crystallographic R index; R_{free}, crystallographic R index of the test set.

*1.5 M $(NH_4)_2SO_4$, 0.1 M Hepes pH7.4, 0.10 M Li_2SO_4.

[†]1.5 M $(NH_4)_2SO_4$, 0.1 M Tris pH 7.0, 0.10 M Li_2SO_4.

[‡]1.5 M $(NH_4)_2SO_4$, 0.1 M Tris pH 7.0, 0.11 M Li_2SO_4.

Table S2. DSC data for the thermal denaturation of FGF-1, *Symfoil*-4P, PV1, and PV2 mutant proteins in 0.1 M and 2.0 M NaCl

Protein	$\Delta H(T_m)$ (kJ·mol^{-1})	T_m (°C)	$\Delta H_{van't\ Hoff}/\Delta H_{cal}$	ΔT_m 2.0–0.1 M NaCl (°C)
0.1 M NaCl				
FGF-1	PPT			
Symfoil-4P*	599 ± 10	85.0 ± 0.1	1.05 ± 0.07	
PV1	490 ± 3	70.7 ± 0.1	0.96 ± 0.08	
PV2	157 ± 5	34.2 ± 0.2	0.87 ± 0.19	
2.0 M NaCl				
FGF-1	PPT			16.3[†]
Symfoil-4P	726 ± 2	100.4 ± 0.1	1.13 ± 0.01	15.4
PV1	620 ± 9	88.3 ± 0.2	0.88 ± 0.07	17.6
PV2	357 ± 2	64.5 ± 0.1	0.84 ± 0.04	30.3

H, enthalpy; T_m, melting temperature; $H_{van't\ Hoff}$, van't Hoff enthalpy; H_{cal}, calorimetric enthalpy; PPT, precipitation.

*From ref. (4).

[†]Apparent ΔT_m determined from endotherm peak.