THE PROTEIN SOCIETY

Introduction of a polar core into the de novo designed protein Top7

Benjamin Basanta,[1,2,3] Kui K. Chan,[4] Patrick Barth,[5,6,7] Tiffany King,[8] Tobin R. Sosnick,[8,9] James R. Hinshaw,[10] Gaohua Liu,[11,12] John K. Everett,[11,12] Rong Xiao,[11,12] Gaetano T. Montelione,[11,12,13] and David Baker[1,2,14]*

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195

[2]Institute for Protein Design, University of Washington, Seattle, Washington 98195

[3]Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, Washington 98195, USA

[4]Enzyme Engineering, EnzymeWorks, California 92121

[5]Structural and Computational Biology and Molecular Biophysics Graduate Program, Baylor College of Medicine, Houston, Texas 77030

[6]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030

[7]Department of Pharmacology Baylor College of Medicine, Houston, Texas 77030

[8]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637

[9]Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637

[10]Department of Chemistry, University of Chicago, Chicago, Illinois 60637

[11]Department of Molecular Biology and Biochemistry, Center of Advanced Biotechnology and Medicine, The State University of New Jersey, Piscataway, New Jersey 08854

[12]Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854

[13]Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854

[14]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195

Received 2 October 2015; Revised 4 February 2016; Accepted 8 February 2016
DOI: 10.1002/pro.2899
Published online 13 February 2016 proteinscience.org

Abstract: Design of polar interactions is a current challenge for protein design. The de novo designed protein Top7, like almost all designed proteins, has an entirely nonpolar core. Here we describe the replacing of a sizable fraction (5 residues) of this core with a designed polar hydrogen bond network. The polar core design is expressed at high levels in *E. coli*, has a folding free energy of 10 kcal/mol, and retains the multiphasic folding kinetics of the original Top7. The NMR structure of the design shows that conformations of three of the five residues, and the designed hydrogen bonds between them, are very close to those in the design model. The remaining two residues, which are more solvent exposed, sample a wide range of conformations in the NMR ensemble. These results show that hydrogen bond networks can be designed in protein cores, but also highlight challenges that need to be overcome when there is competition with solvent.

Additional Supporting Information may be found in the online version of this article.

Broader Audience Statement: Natural proteins have primarily hydrophobic cores and polar exteriors. With the long-term goal of making "inside out" proteins with polar cores, we explore the properties of a designed protein with a polar hydrogen bond network in its core

Benjamin Basanta and Kui K. Chan contributed equally to this work

Grant sponsor: US National Institutes of Health; Grant number: U54GM094597 (to G.T.M.); Grant sponsor: NIH; Grant number: GM055694 (to T.R.S); Grant sponsor: Howard Hughes Medical Institute.

*Correspondence to: David Baker, University of Washington, Molecular Engineering and Sciences, Box 351655, Seattle, WA 98195-1655. E-mail: dabaker@u.washington.edu

Keywords: protein design; hydrogen bonds; protein folding; protein NMR; protein core polar interactions

Introduction

De novo protein design has advanced considerably in recent years. New protein structures, functions, and interactions have been successfully designed.[1–5] Both the new structures and new interactions are stabilized primarily by nonpolar interactions; there has been less success with hydrogen bonding and polar interaction networks.

The de novo designed Top7 protein has an entirely hydrophobic core and is extremely stable.[6] Characterization of the folding kinetics of Top7 showed the process was considerably more complex than that of most native proteins with the same size, with multiple distinct kinetic phases. Analysis of fragments of Top7 showed that many had appreciable structure in isolation. The high stability and low folding cooperativity of Top7 suggests that there may have been more selection for folding cooperativity than stability during evolution.[7]

With the long-range goal of designing an "inside out" protein—with a polar core and a nonpolar exterior—that could be stable in organic solvent, we explored the replacement of a portion of the Top7 hydrophobic core with a polar hydrogen bond network. The polar-core Top 7 (Top7_PC) is folded and moderately stable, and structural characterization shows that the majority of the hydrogen bonds in the designed network are intact.

Results

With the goal of designing an "inside out" protein, we adapted the RosettaMembrane all-atom energy function which utilizes an implicit model of the lipid surrounding the transmembrane portions of membrane proteins.[8] We modified the energy function such that the protein environment was treated as a uniform implicit apolar solvent instead of a heterogeneous lipid membrane. We first entirely redesigned Top7 using the apolar potential, keeping the backbone conformation fixed, and allowing all 20 amino acids at all 94 positions to assess whether networks of polar residues with low energy configurations could be designed in its core. The calculations converged on very similar low energy designed structures, with most core positions of the original Top7 recapitulated except for five positions substituted with polar residues forming a connected hydrogen bond network with each residue stabilized by two hydrogen bonds [Fig. 1(A,B)]. The polar network bridges several secondary structural elements in the Top7 core including the first and third beta strands and the C-terminal helix. The backbone and side-chain dihedral angles of the designs were optimized by minimization in the Rosetta forcefield; the lowest energy of the refined designs is shown in Figure 1(C).

In a first round of experiments, we attempted to express the two lowest energy designed "inside-out" proteins in *Escherichia coli* with the hope that they would form inclusion bodies that could then be solubilized in organic solvent. However, there was little or no expression of these rather unusual proteins. To evaluate the designed polar core independent of such a drastic surface redesign, we replaced the surface residues on the two lowest energy polar-core Top7 designs with those of the original water-soluble design. As we anticipated that the polar core designs would be considerably destabilized, we decided to follow the example of crambin, a small protein with a quite polar core,[9] and incorporate a disulfide bond into the design. We searched for pairs of positions at which substituted cysteines could form a disulfide bond by introducing cysteines at every residue and using the Rosetta all atom energy function to identify residue pairs with low energy disulfide bonds. We identified three pairs of positions where disulfide bonds with good geometry could be introduced [Fig. 1(D), only the model for the best behaved protein is shown].

We tested the designed disulfides initially in the context of the original nonpolar core Top7. The three disulfides were introduced by mutagenesis, and the proteins expressed and purified. One of the three introduced disulfides—between residues 12 and 45—was formed quantitatively, as assessed using Ellman's reagent. Guanidine hydrochloride (GuHCl) denaturation experiments tracking far-UV circular dichroism (CD) showed the introduced disulfide shifted the denaturation midpoint, C_m, of Top7, an already very stable protein, from $\sim$6.5M to $\sim$8M GuHCl (Top7_CC, Supporting Information Fig. 1). Because little or no unfolded baseline could be measured, we could not determine the increase in folding free energy by fitting the full unfolding transition; instead we estimate from the shift in midpoint and the m-value of Top7 that the disulfide increases stability by $\sim$3 kcal/mol.

We next introduced the 12 to 45 disulfide bond into the protein having the designed core hydrogen bond network, keeping the original Top7 surface residues at all other positions [Fig. 1(E)]. The resulting protein "Top7_PC" expressed at high levels in *E. coli* and could be readily purified. The CD spectrum of Top7_PC was very similar to that of the wild-type protein [Supporting Information Figs. 4 and 2(A) of Ref. 6]. Rotational correlation time (τ_c) estimates

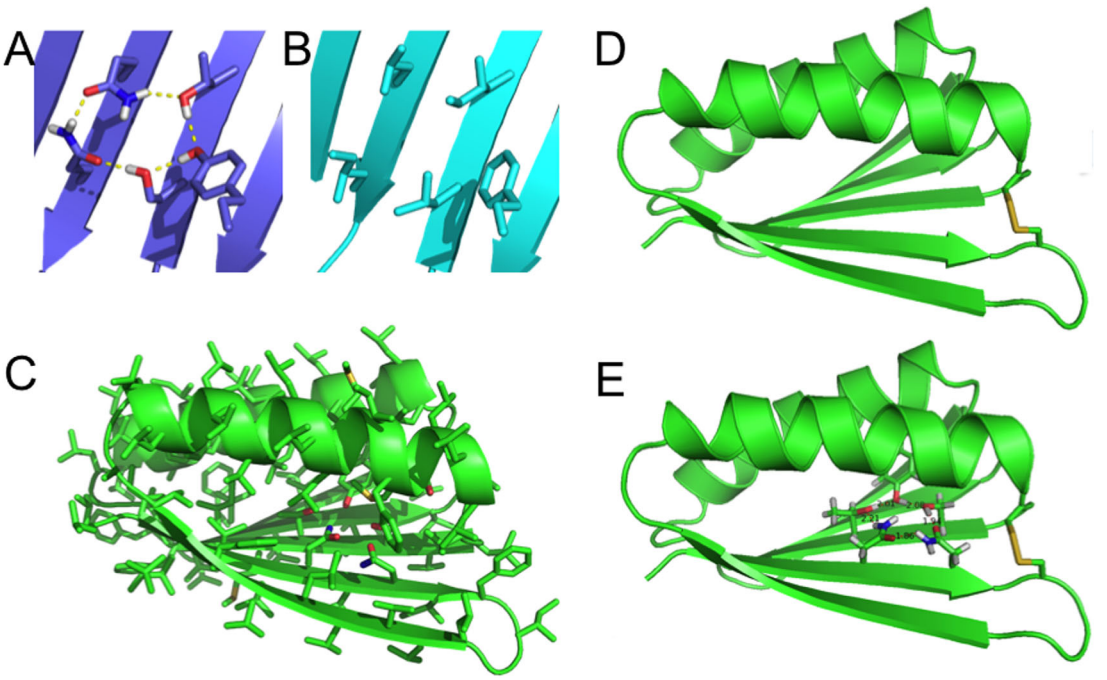

Figure 1. Top7_PC and Top7 models. Close up view comparing the core region of Top7 before (B) and after (A) hydrogen-bond network incorporation. (C) Hydrogen-bond network in the context of the whole structure of one of the initial "inverted" Top7 models. (D) Model of the disulfide-bonded variant of Top7. (E) Model of disulfide-bonded Top7 with core hydrogen-bond network.

based on ^{15}N NMR nuclear relaxation measurements (T_1/T_2) indicate that Top7_PC is predominantly monomeric in solution (Supporting Information Fig. 3). The GuHCl unfolding midpoint was lower than that of either Top7 and Top7_CC but the C_m is still high for a monomeric protein (~5*M*; Supporting Information Fig. 1). GuHCl melting experiments in absence and presence of reducing agent (*tris*(2-carboxyethyl)phosphine, TCEP) suggest that the disulfide bond increases stability in this context by ~1.4 kcal/mol (from a ΔG of unfolding of 10.3 kcal/mol in oxidizing conditions—Fig. 2 and Supporting Information Fig. 3—to 8.9 kcal/mol in reducing conditions—Fig. 3 and Table I).

To compare the thermodynamics of folding of Top7 and Top7_PC, we conducted temperature denaturation measurements monitoring CD at 227 nm at GuHCl concentrations ranging from 4 to 7*M* in conditions likely to maintain the sample in reduced conditions (see Methods). The comparison between the folding kinetics of Top7 and Top7_PC is easier to interpret in the absence of the disulfide bond, i.e., in reducing conditions, since in this case the only difference between the two is the introduction of the hydrogen bond network. The thermodynamic parameters ΔH_m, ΔS_m, ΔC_p, T_m and denaturant *m*-value were calculated from global fits to the data (Table I, Fig. 3, Supporting Information Figs. 1 and 2).

Top7 is more stable than Top7_PC by ~5 kcal mol^{-1} at all measured temperatures. Otherwise, the thermodynamic properties for the two proteins are similar. Both proteins are stabilized entropically at low temperature but enthalpically at high temperature. The *m*-value for Top7 is 15% higher than Top7_PC, whereas the ΔC_p value is 9% lower. The lower ΔC_p value for Top7 is unexpected as this protein has more hydrophobic burial, a property that heavily contributes to ΔC_p.[10] Possibly, the partial exposure of polar surface in the *native* state of Top7_PC due to an incompletely packed core in reducing conditions could increase the ΔC_p (e.g., residues Asn 18 and Ser 56; see structural studies below). This increased polar exposure in the native

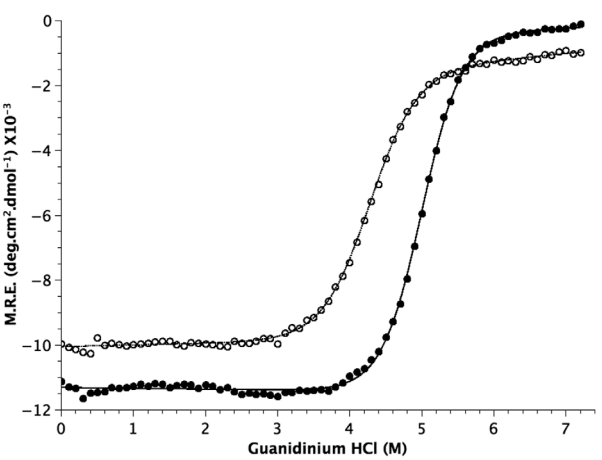

Figure 2. Chemical denaturation of Top7_PC in the presence (open circles) and absence (filled circles) of reducing agent. Lines are two-state unfolding model fits.

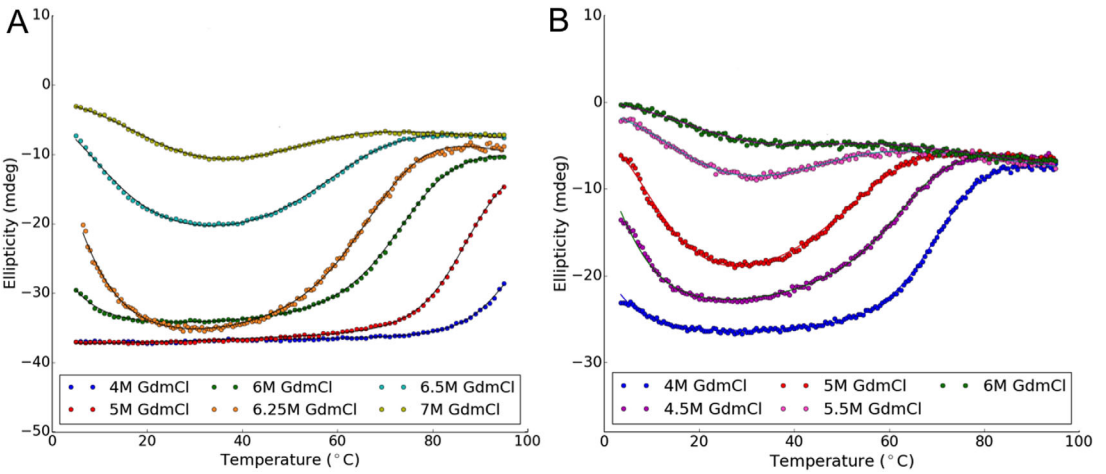

Figure 3. Thermodynamic properties of Top7 (A) and Top7_PC (B). CD-monitored thermal denaturation at multiple GuHCl concentrations with fits to the data using Eq. (3).

state of Top7_PC is consistent with this variant having a smaller *m*-value, a quantity that is sensitive to backbone exposure.[11]

We next investigated the folding kinetics of Top7_PC monitoring fluorescence of either the sole tryptophan (Trp 83) and 8-anilinonaphthalene-1-sulfonic acid (ANS), a compound that strongly fluoresces upon association with hydrophobic regions found in some folding intermediates, but typically not in either the unfolded and folded states[12] (Fig. 4). The original Top7 protein exhibits complex multiphasic folding kinetics with a weak dependence of the folding rate constants on denaturant concentration for the two slower phases, and a very fast burst phase with strong denaturant dependence [see Fig. 4(B)]. It was hypothesized that the rate limiting step in Top7 folding consists in structural rearrangements of collapsed states.[7] Therefore, the reduction in hydrophobic content in Top7_PC was anticipated to reduce or eliminate the slow folding phases by reducing the stability of any such intermediates. Reduced Top7_PC's folding behavior shared some of the complexity found in Top7. Refolding tracking tryptophan fluorescence was nearly single exponential with a rate constant between 1 and 4 s^{-1} that had only a mild denaturant dependence from 1 to 5*M* GdmCl. Fitting with an additional phase of ~10 s^{-1}, also of decreasing fluorescence intensity, slightly reduced the residuals. ANS-monitored refolding identified additional fast phases with largely denaturant independent rate constants. The ANS fluorescence signal decreased with time indicating that hydrophobic surface was being buried in these phases. This decrease indicates that a partially folded species (when stable) was formed within the 2 ms instrumental dead-time. At 2.2*M* Gdm and below, the fastest phase had a rate constant of 105 to 137 s^{-1} while two minor phases had rate constants of 3 to 11 s^{-1} and <1 s^{-1}, respectively. At 3.6*M* Gdm, the faster phases were no longer observed, and the observed rate constant was 3 s^{-1}, similar to the rate constant of the major phase monitored by tryptophan fluorescence. As with Top7, the reduced form Top7_PC likely folds via multiple intermediate species with a rate-limiting step that involves reorganization of partially formed structure.

We next carried out NMR structural studies to determine the three-dimensional structure of Top7_PC, focusing on the extent of formation of the designed hydrogen bond networks. NMR structure quality statistics are presented in Table II. The overall topology of disulfide-oxidized Top7_PC is very close to the design model [and the original Top7 structure, Fig. 5(A)]. Analysis of the NOEs involving the polar core residues suggests that three of the five residues in the core of the network are positioned as in the design model, and making the designed hydrogen bonds [Fig. 5(C)]. The remaining two residues, Asn 18 and Ser 56, are at the periphery of the network and more solvent exposed, and do not make the designed hydrogen bonds [NOE Fig. 5(B), overall structure Fig. 5(A)].

To determine if disulfide reduction significantly affects the 3D structure of TOP7_PC, we compared [^{15}N-^{1}H]-HSQC spectra of Top7_PC in both disulfide-

Table I. *Thermodynamic Parameters for Top7 and Reduced Top7_PC*

Protein	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	ΔC_p (kcal mol^{-1} K^{-1})	m (kcal mol^{-1} M^{-1})	ΔG (kcal mol^{-1})
Top7	0.9 ± 0.4	−13.0 ± 0.3	0.91 ± 0.01	2.17 ± 0.05	13.9 ± 0.4
Top7_PC	−3.0 ± 0.2	−11.9 ± 0.3	1.00 ± 0.02	1.85 ± 0.03	8.9 ± 0.3

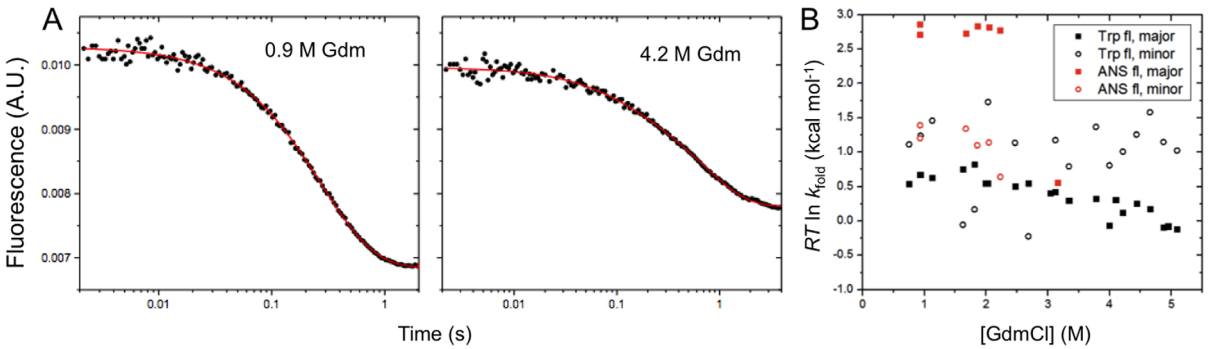

Figure 4. Folding kinetics for Top7_PC. (A) Double exponential fits to refolding signals at 0.9 and 4.2*M* GuHCl, (Complete buffer composition in Methods). (B) Relaxation rate constants for the major and minor folding phases as a function of GuHCl concentration followed by tryptophan (Trp fl.) and ANS (ANS fl.) fluorescence.

oxidized and reduced (10 m*M* DTT) conditions. These data (Supporting Information Fig. 5) demonstrate the backbone structure is essentially identical under both conditions. However, several resonances are significantly broader and have weaker intensities in the disulfide-reduced protein, including residues Asn16 (backbone and sidechain amides), Asn17 (backbone and sidechain amides), Asn18 (backbone), Thr19, Ser56, Arg57, Thr58, and Tyr79, some of which contribute to the hydrogen-bonded network of the polar core [Fig. 5(C)]. We attribute this to conformational exchange broadening on the NMR timescale in the polar core, which while observed for these same residues also in the disulfide-oxidized form, appears to be much more pronounced in the disulfide-reduced form of Top7_PC.

Encouraged by the stability of Top7_PC and the accuracy of recapitulation of the hydrogen bond network, we returned to the overall challenge of designing an inside-out protein. We redesigned the surface of each of the two helices and the beta sheet individually to be entirely apolar, but again were unable to express the apolar surface proteins in *E. coli* even in inclusion bodies.

Discussion

An extended hydrogen bond network was introduced into the hydrophobic core of Top7 using computational protein design. This substitution of nonpolar core residues with a polar network is the opposite of a classic experiment by Waldburger *et al.*,[17] who replaced a buried hydrogen bond network with a cluster of nonpolar residues. Not surprisingly, given that formation of a buried polar network requires stripping away of attractive interactions with water in the unfolded state, in both experiments the more nonpolar arrangement produces a less stable protein.

Why do Asn 18 and Ser 56 not form the designed hydrogen bonds? There are at least three possible contributing factors. First, as noted above, these residues are more solvent exposed in the three-dimensional structure of TOP7_PC, and hence can make hydrogen bonds with water instead of the designed intra-protein hydrogen bonds. In contrast, the residues involved in the part of the network that was recapitulated—Tyr 79, 16, 58—are more buried and hence had fewer other options for hydrogen-bond satisfaction. A key feature in the design of hydrogen bond networks may be to eliminate as much as possible competing interactions with other amino acids and with solvent. Second, the hydrogen bond geometry between Asn 18 and Asn 16 is somewhat strained (the angle at the hydrogen bond is 130° instead of 180°). Without this hydrogen bond, neither Asn 18 nor Ser 56 would be held in the designed configuration. Third, the region surrounding the S1 -> S2 turn has higher B factors in the original Top7 structure,[6] and hence may be more sensitive to substitution with the network forming residues than the remainder of the protein.

We were unsuccessful with further steps towards the long-range goal of making "inside out" proteins with a polar core that could be stable in organic solvent since *E. coli* did not produce variants of the polar core design with nonpolar surface residues. The *E. coli* quality control machinery may degrade such proteins soon after synthesis. Further progress towards the goal of proteins with polar cores stable in organic solvent will likely require chemical synthesis or in vitro translation.

Methods

Computational design

Rosetta design calculations were carried out as described in the text. Standard Rosetta Monte Carlo sequence design calculations were carried out starting from the Top7 backbone using a uniform apolar solvation model. Each Monte Carlo move consists of substitution of a randomly selected rotamer of a randomly selected amino acid at a randomly selected position; a typical design trajectory involves 1

Table II. *NMR Structure Quality Statistics for Top7_PC*

NMR distance and dihedral restraints	
Distance restraints	
Total NOE	1855
Intra-residue	436
Inter-residue	
Sequential ($\|i-j\|=1$)	445
Medium-range ($\|i-j\|\leq 4$)	374
Long-range ($\|i-j\|\geq 5$)	600
Hydrogen bonds	48
Total dihedral angle restraints	134
Φ	67
Ψ	67
Total number of restricting restraints	2037
Total number of restricting restraints per residue	22.6
Restricting long-range restraints per residue	6.9
Structure statistics	
Violations	
RMS of distance violation/restraint[a] (Å)	0.01
RMS of dihedral angle violation/restraint (°)	0.91
Max distance restraint violation (Å)	0.39
Max dihedral angle violation (°)	7.60
Average r.m.s.d. to representative conformer[b] (Å)	
Backbone atoms	0.69 ± 0.06
Heavy atoms	1.48 ± 0.14
RPF scores[c]	
Recall	0.94
Precision	0.951
F-measure	0.945
DP-score	0.805
Structure quality factors (raw/*Z*-score[d])	
Procheck *G*-factor (Φ/Ψ only)[b]	0.06/0.55
Procheck *G*-factor (all dihedral angles)[b]	0.01/0.06
Verify3D	0.42/−0.64
ProsaII	0.89/0.99
MolProbity clashscore	11.90/−0.52
Ramachandran plot summary from Richardson's lab	
Most favored regions (%)	98.6
Allowed regions (%)	1.4
Disallowed regions (%)	0

[a] Calculated by using sum over r^{-6}.

[b] Calculated among 20 refined structures for ordered residues that have sum of Φ and Ψ order parameters[13] $S(\Phi)+S(\Psi) > 1.8$.[14] Residue range: 11–19, 28–50, 55–82, 87–91, 95–99. RMSD values were calculated by PDBStat,[15] the mediod structure is the seventh conformer.

[c] Analyzed for the 20 lowest energy refined NMR structures, by using PDBSTAT PSVS 1.4, and RPF/DP software.[14,16]

[d] With respect to mean and standard deviation for a set of 252 X-ray structures with sequence lengths <500, resolution ≤1.80 Å, *R*-factor ≤0.25, and *R*-free ≤0.28; a positive value indicates a "better" score.

million such moves Ref. 41 and takes ~2 min (wall time) on a single core of an Intel® Xeon® CPU E7-2850 2.00GHz (24576 KB cache). Surface residues grafted from the original Top7 to give the soluble, highly expressed Top7_PC design are those with SASA higher than 5 Å^2, as calculated using the LeGrand method[18] with a probe radius of 2.2Å (Using the Top7 1QYS crystal structure numbering: D3, Q5, Q7, N9, D11, D12, N13, G14, K15, 16N, 17F, 18D, 19Y, 20T, 21Y, 22T, 24T, 25T, 26E, 27S, 28E, 30Q, 31K, 32V, 33L, 34N, 35E, 37M, 38D, 39Y, 41K, 42K, 43Q, 44G, 46K, 47R, 49R, 51S, 53T, 55R, 56T, 57K, 58K, 59E, 61E, 62K, 65A, 66I, 68I, 69K, 70V, 72A, 73E, 75G, 77N, 78D, 79I, 80N, 81V, 82T, 83F, 84D, 85G, 86D, 87T, 89T, 91E, 93Q, 94L), except for M37(K), N16(T) and F83(W).

Refolding measurements

Refolding measurements were conducted with a Biologic SFM-4000 instrument integrated with a PTI mercury lamp and a Hanamatsu 10722-110 photomultiplier. Measurements were conducted in 50 m*M* sodium phosphate, 150 m*M* NaCl, 20n*M* TCEP, pH 7 to 7.5. Tryptophan and ANS (Acros Organics) fluorescence, respectively, were measured using an excitation at $\lambda = 283$ and 370 nm, and measuring total emission above 310 and 410 nm, respectively, using glass filters. Protein concentration varied from 3 to 20 μ*M*. The final concentration of ANS and dimethyl sulfoxide (used to solubilize the ANS) was between 340 and 570 μ*M* and 0.3 to 0.6% by volume, respectively.

Circular dichroism

We conducted temperature denaturation measurements monitoring the CD signal at 227 ± 5 nm (pathlength = 1 cm) from 3 to 95°C at GuHCl concentrations ranging from 4*M* to 7*M* using a Jasco 715 spectrometer. The samples were in 20 n*M* TCEP, 5m*M* sodium phosphate, pH 7.0 at a concentration of 2 μ*M*. A global fit to the data (Table I, Fig. 2, Supporting Information Figs. 1 and 2) was conducted to calculate the thermodynamic parameters ΔH_m, ΔS_m, ΔC_p, T_m and denaturant *m*-value using the Eqs. (1) to (3).

$$\Delta H(T) = \Delta H_{293} + \Delta C_p * (T - 293) \tag{1}$$

$$\Delta S(T) = \Delta S_{293} + \Delta C_p \ln\left(\frac{T}{293}\right) \tag{2}$$

$$\Delta G(T, [\mathrm{GdmCl}]) = \Delta H_{293} - T\Delta S_{293} + \Delta C_p\left[(T\text{-}293) - T\ln\left(\frac{T}{293}\right)\right] - m[\mathrm{GdmCl}] \tag{3}$$

We conducted denaturant titration experiments shown in Figure 2 and Supporting Information Figure 1 monitoring the CD signal at 227 ± 5 nm (pathlength = 1 cm) from 0 to 8*M* GuHCl using an AVIV 420 CD spectrometer (Biomedical, Inc.). The samples were in 1 m*M* TCEP (only in reducing conditions), 5 m*M* sodium phosphate, pH 7.0 at a concentration of 4 μ*M*.

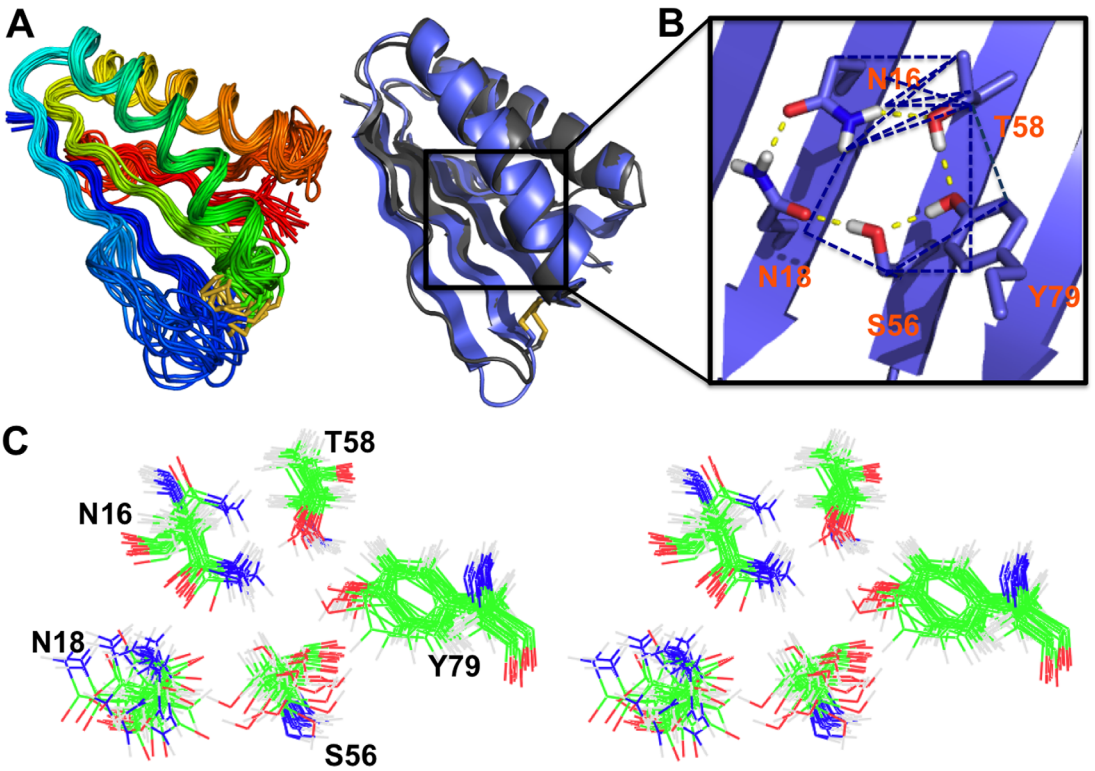

Figure 5. NMR structure of Top7_PC. (A) Left, NMR structure ensemble (disulfide shown as sticks). Right, Comparison of NMR average structure (gray) to TOP7_PC model (blue). The cysteines forming the disulfide bonds are shown as sticks. (B) Hydrogen bond network zoom-in; blue dashed lines indicate observed NOEs, and yellow dashed lines indicate hydrogen bonds. (C) Stereo view of hydrogen-bond network residues in the 20 NMR structure conformers.

Protein sample preparation for structural studies

Synthetic genes coding for the modified TOP7 protein designs were expressed and purified by following the standard protocols.[19] [U-^{15}N, 5%-^{13}C]- or [U-^{15}N, U-^{13}C]-enriched proteins were expressed using MJ9 minimal media.[20] The [U-^{15}N, 5%-^{13}C]-labeled proteins were generated for stereo-specific assignments of isopropyl methyl groups of valines and leucines[21] and for residual dipolar coupling (RDC) measurements.[22] The final purified proteins each contained an N-terminal tag with sequence MHHHHHHR. Samples were validated to have correct molecular weight by MALDI-TOF mass spectrometry, and were >98% homogeneous based on analytical SDS polyacrylamide gel electrophoresis (SDS-PAGE). For structural studies, the purified protein concentration was 0.37 m*M* in a low salt buffer (100 m*M* NaCl, 10 m*M* Tris-HCl, pH 7.5, 0.02% NaN_3).

NMR studies for structure determination

All NMR spectra were recorded at 25°C using cryogenic NMR probes. Triple resonance NMR data, simultaneous 3D $^{15}N/^{13}C_{aliphatic}/^{13}C_{aromatic}$-edited NOESY[23] (mixing time: 100 ms) and 3D ^{13}C-edited aromatic NOESY (mixing time: 100 ms) spectra were acquired on a Bruker AVANCE 800 MHz spectrometer. All NMR data were processed using the program NMRPipe[24] and analyzed using the program XEASY.[25] Spectra were referenced to external DSS.

Rotational correlation time measurements

Rotational correlation times (τ_c) were computed from 1D ^{15}N T_1 and T_2 relaxation data at 25°C as described elsewhere.[26] The rotational correlation tie τ_c was then calculated from the ^{15}N T_1/T_2 ratio using the following approximation:[27,28]

$$\tau \approx \frac{\sqrt{\frac{6T_1}{T_2}-7}}{4\pi\nu_N}$$

where ν_N is the resonance frequency of ^{15}N in Hz. Finally, values of τ_c for were plotted against protein molecular weight and compared with data for known monomeric proteins. For full length Top7_PC, the rotational correlation time $\tau_c = 5.8$ ns, corresponding to a molecular weight of ~9.6 kDa consistent with the expected monomeric molecular weight of 12.1 kDa.

NMR resonance assignments

Sequence-specific backbone resonance assignments were determined in a largely automated fashion with the program AUTOASSIGN[29] as described

previously.[30] The full resonance assignments were obtained by using the standard NESG high-throughput protocol.[30] The obtained backbone assignments, together with random coil side-chain chemical shift values, were then used to simulate peak lists that facilitated manual analysis of side-chain resonance assignments. Simultaneous 3D $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-NOESY and CCH-TOCSY were then analyzed manually to obtain nearly complete side-chain assignments. Chemical shift data were deposited in the Biological Magnetic Resonance Bank. (BMRB accession code: 25666).

NMR structure determination

NMR structure calculations followed standard protocols.[30] Helical secondary structures were initially identified from chemical shift data using TALOS_N.[31] Initial NOESY peak lists of expected intra-residue, sequential, and α-helical medium-range NOE peaks were then generated from the resonance assignments. These were next manually edited by visual inspection of the NOESY spectra. Subsequent manual peak picking was then used to identify remaining, primarily long-range NOESY cross peaks.[30] Backbone dihedral angle restraints were derived from the chemical shifts using the program TALOS_N[31] for residues located in well-defined secondary structure elements. The program CYANA[32,33] was then used to automatically assign NOEs and to calculate structures. The automatic NOESY analysis program AUTOSTRUCTURE[34,35] was used in parallel to guide iterative cycles of noise/artefact peak removal, peak picking and NOE assignments. The 20 conformers with the lowest target function value were then refined in explicit water[36] using the program CNS.[37] The structural statistics and global structure quality factors included analyses by Verify3D,[14] ProsaII,[38] PROCHECK,[39] and MolProbity.[40] Raw and statistical *Z*-scores were computed using the Protein Structure Validation Server (PSVS ver. 1.4).[14] The global goodness-of-fit of the final structure ensembles with the NOESY peak list data (RPF and DP scores) was assessed using the RPF analysis program.[16] Coordinates and restraints were deposited in the Protein Data Bank (PDB ID: 2N4E).

Acknowledgments

Authors thank the many scientists of the Northeast Structural Genomics Consortium for their support. Authors also thank Prof. James Prestegard, University of Georgia, for providing ^{15}N-^{1}H residual dipolar coupling data.

References

1. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D (2012) Principles for designing ideal protein structures. Nature 491:222–227.
2. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D (2013) Computational design of ligand-binding proteins with high affinity and selectivity. Nature 501:212–216.
3. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D (2014) Accurate design of co-assembling multi-component protein nanomaterials. Nature 510:103–108.
4. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble alpha-helical barrels. Science 346:485–488.
5. Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G, DeGrado WF (2014) De novo design of a transmembrane Zn2+-transporting four-helix bundle. Science 346:1520–1524.
6. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302:1364–1368.
7. Watters AL, Deka P, Corrent C, Callender D, Varani G, Sosnick T, Baker D (2007) The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. Cell 128:613–624.
8. Barth P, Schonbrun J, Baker D (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. Proc Natl Acad Sci USA 104: 15682–15687.
9. Teeter MM, Roe SM, Heo NH (1993) Atomic resolution (0.83 A) crystal structure of the hydrophobic protein crambin at 130 K. J Mol Biol 230:292–311.
10. Makhatadze GI, Privalov PL (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. J Mol Biol 213:375–384.
11. Guinn EJ, Kontur WS, Tsodikov OV, Shkel I, Record MT (2013) Probing the protein-folding mechanism using denaturant and temperature effects on rate constants. Proc Natl Acad Sci USA 110:16784–16789.
12. Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers 31:119–128.
13. Hyberts SG, Goldberg MS, Havel TF, Wagner G (1992) The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. Protein Sci 1:736–751.
14. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66:778–795.
15. Tejero R, Snyder D, Mao B, Aramini JM, Montelione GT (2013) PDBStat: a universal restraint converter and restraint analysis software package for protein NMR. J Biomol NMR 56:337–351.
16. Huang YJ, Powers R, Montelione GT (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 127:1665–1674.
17. Waldburger CD, Schildbach JF, Sauer RT (1995) Are buried salt bridges important for protein stability and conformational specificity? Nat Struct Biol 2:122–128.
18. Le Grand S, Merz K (1993) Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. J Comput Chem 14:349–352.
19. Acton TB, Xiao R, Anderson S, Aramini J, Buchwald WA, Ciccosanti C, Conover K, Everett J, Hamilton K, Huang YJ, Janjua H, Kornhaber G, Lau J, Lee DY, Patel D, Rossi P, Sahdev S, Shastry R, Swapna GVT,

Tang Y, Tong S, Wang D, Wang H, Zhao L, Montelione GT (2011) Preparation of protein samples for NMR structure, function, and small-molecule screening studies. Methods Enzymol 493:21–60.
20. Jansson M, Li Y-C, Jendeberg L, Anderson S, Montelione G, Nilsson B (1996) High-level production of uniformly 15N-and 13C-enriched fusion proteins in *Escherichia coli*. J Biomol NMR 7:131–141.
21. Neri D, Szyperski T, Otting G, Senn H, Wuethrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. Biochemistry 28:7510–7516.
22. Tjandra N, Grzesiek S, Bax A (1996) Magnetic field dependence of nitrogen−proton J splittings in 15N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. J Am Chem Soc 118:6264–6272.
23. Shen Y, Atreya HS, Liu G, Szyperski T (2005) G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. J Am Chem Soc 127:9085–9099.
24. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293.
25. Bartels C, Xia TH, Billeter M, Güntert P, Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 6:1–10.
26. Aramini JM, Ma L-C, Zhou L, Schauder CM, Hamilton K, Amer BR, Mack TR, Lee H-W, Ciccosanti CT, Zhao L, Xiao R, Krug RM, Montelione GT (2011) Dimer interface of the effector domain of non-structural protein 1 from influenza A virus: an interface with multiple functions. J Biol Chem 286:26050–26060.
27. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28:8972–8979.
28. Fushman D, Weisemann R, Thüring H, Rüterjans H (1994) Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. J Biomol NMR 4:61–78.
29. Moseley HN, Monleon D, Montelione GT (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol 339:91–108.
30. Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. Proc Natl Acad Sci USA 102:10487–10492.
31. Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR 56: 227–241.
32. Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298.
33. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227.
34. Huang YJ, Tejero R, Powers R, Montelione GT (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins 62:587–603.
35. Huang YJ, Moseley HNB, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperski T, Montelione GT (2005) An integrated platform for automated analysis of protein NMR structures. Methods Enzymol 394:111–141.
36. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506.
37. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr 54:905–921. D
38. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17:355–362.
39. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291.
40. Word JM, Bateman RC, Presley BK, Lovell SC, Richardson DC (2000) Exploring steric constraints on protein mutations using MAGE/PROBE. Protein Sci 9:2251–2259.
41. Rohl CA, Strauss CEM, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383:66–93.