

Design of a Novel Globular Protein Fold with Atomic-Level Accuracy

Brian Kuhlman,^{1*}† Gautam Dantas,^{1*} Gregory C. Ireton,⁴ Gabriele Varani,^{1,2} Barry L. Stoddard,⁴ David Baker^{1,3,†}

A major challenge of computational protein design is the creation of novel proteins with arbitrarily chosen three-dimensional structures. Here, we used a general computational strategy that iterates between sequence design and structure prediction to design a 93-residue α/β protein called Top7 with a novel sequence and topology. Top7 was found experimentally to be folded and extremely stable, and the x-ray crystal structure of Top7 is similar (root mean square deviation equals 1.2 angstroms) to the design model. The ability to design a new protein fold makes possible the exploration of the large regions of the protein universe not yet observed in nature.

There are a large but finite number of protein folds observed thus far in nature, and it is not clear whether the structures not yet observed are physically unrealizable or have simply not yet been sampled by the evolutionary process or characterized by a structural biologist. Methods for de novo design of novel protein structures provide a route to resolving this question and, perhaps more importantly, a possible route to novel protein machines and therapeutics.

There has been considerable progress in the development of computational methods for identifying amino acid sequences compatible with a target structure (1–3), notably the pioneering complete redesign of a zinc finger protein by Mayo and co-workers (1). In general, these methods have not been used to create new protein structures but rather to redesign naturally occurring proteins so that they have enhanced stability or new functionality (4–6). Because of the strong steric restrictions in the cores of globular proteins, the packing of side chains in redesigned proteins is often quite similar to that in the original native protein (1, 7), and hence high-resolution protein backbone coordinates contain some memory of the original native sequence (8–12). When creating a new protein from scratch,

there is no such sequence memory to aid the process, and it is not even known whether the target backbone is designable. Thus, the computational design of novel protein structures is a more rigorous test of current force fields and optimization methodology than the redesign of naturally occurring proteins.

Because it is unlikely that any arbitrarily chosen protein backbone will be designable, it is essential that the design procedure include a search of nearby conformational space in addition to sequence space. With the exception of the method used by Desjarlais and Handel (2) to redesign the hydrophobic core of a small naturally occurring protein, most previous approaches have either opti-

mized the amino acid sequence for a large number of fixed backbone conformations (4, 12–14) or, as in the landmark design by Harbury and colleagues of coiled coil oligomers with a right-handed superhelical twist (15), refined the backbone conformation for a large number of fixed amino acid sequences (15, 16). The range of sequence-structure pairs that can be searched with the use of these approaches is restricted by the need to specify, in advance, a limited number of backbone conformations or amino acid sequences.

We have developed a general procedure for identifying very low free energy sequence-structure pairs that iterates between sequence optimization and structure prediction and can be applied to the design of any desired target structure. The same energy function is used to guide the search at all stages, and at each stage only the lowest energy sequence or structure identified in the previous iteration is optimized, thereby avoiding the large-scale and computationally expensive enumeration of alternative backbones or alternative sequences. Unlike the genetic algorithm of Desjarlais and Handel (2) in which randomly selected torsion angles and residue identities were simultaneously perturbed, our procedure iterates between full-scale optimization of sequence for a fixed backbone conformation and gradient-based optimization of the backbone coordinates for a fixed sequence. We used this approach to create a 93-residue α/β protein with a topology not present in the Protein Structure Database (PDB).

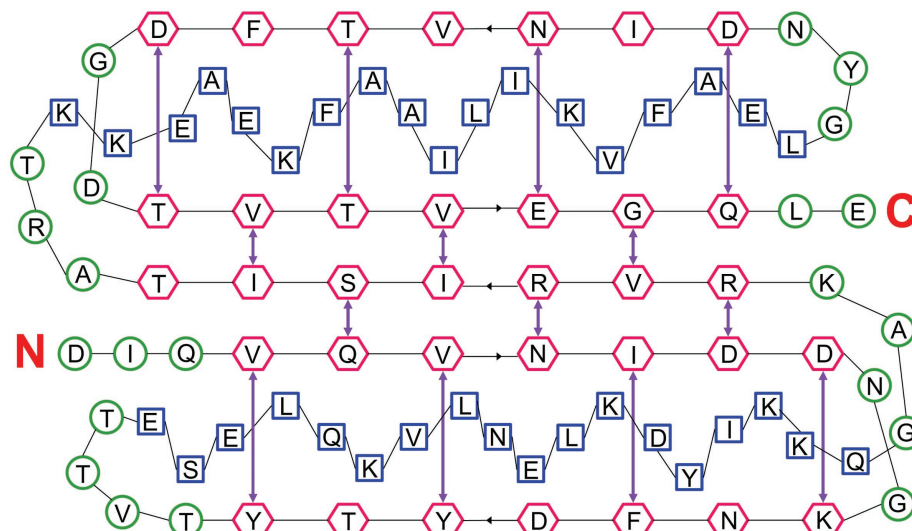


Fig. 1. A two-dimensional schematic of the target fold (hexagon, strand; square, helix; circle, other). Hydrogen bond partners are shown as purple arrows. The amino acids shown are those in the final designed (Top7) sequence.

¹Department of Biochemistry, ²Department of Chemistry, ³Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. ⁴Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA.

*These authors contributed equally to this work.

†Present address: Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599, USA.

‡To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu

Generation of starting models. The target structure for the de novo design process can range from a detailed backbone model to a back-of-the-envelope sketch. Because we aimed to create a novel protein fold, we selected a topology not present in the PDB according to the Topology of Protein Structure (TOPS) server (17). A rough two-dimensional diagram was created of the target structure (Fig. 1), and constraints were identified that define the topology (Fig. 1, arrows). Three-dimensional models satisfying the constraints were then generated by assembling three- and nine-residue fragments from the PDB with secondary structures consistent with the diagram using the Rosetta de novo structure prediction methodology (18), leading to 172 backbone-only models that had the desired topology and secondary structure content and had root mean square deviations (RMSDs) from each other of 2 to 3 Å.

Generation of starting sequences. A sequence was designed for each model with the use of the RosettaDesign (9) Monte Carlo search protocol and energy function, which is dominated by a 12-6 Lennard-Jones potential, an orientation-dependent hydrogen bonding term (19), and an implicit solvation model (20, 21). All amino acids except for cysteine were allowed at 71 of the 93 positions [~ 110 rotamers from Dunbrack's library (22) per position]; the remaining 22 surface β -sheet positions were restricted to polar amino acids (~ 75 rotamers per position). The search through the $110^{71} \times 75^{22}$ ($>10^{186}$) rotamer combinations took ~ 10 min for each model on a Pentium III (Intel) processor.

Because the starting backbone conformations were generated without regard to side-chain packing, it was anticipated that sequences with very low free energies might

not exist (i.e., the structures would not be designable). Indeed, the lowest energy sequences selected for the starting structures had energies considerably higher than those of native proteins of roughly the same size. In particular, the Lennard-Jones interaction energies for core residues were on average $0.8 \text{ kcal mol}^{-1}$ less favorable than the interaction energies for the same residues in native protein cores. The finding that low-energy sequences do not exist for protein backbones generated without regard to side-chain packing emphasizes the need to couple sequence design with backbone flexibility for general protein design.

Simultaneous optimization of sequence and structure. The critical feature of the design protocol is the cycling between sequence design, as described above, and backbone optimization. The goal of the backbone optimization step, to identify the lowest free energy backbone conformation for a fixed amino acid sequence, is formally analogous to the high-resolution structure prediction problem, and we used the Rosetta program (23), which we developed for structure prediction. The backbone torsion angles were optimized with the use of a Monte Carlo minimization protocol (24) in which each move has the following parts. (i) An initial perturbation, consisting of either a small random change in the torsion angles of one to five randomly selected residues or a substitution of the backbone torsion angles of one to three consecutive residues with torsion angles from a structure in the PDB. In the latter case, the torsion angles of neighboring residues were varied to minimize the displacement of the downstream portion of the chain. (ii) A rapid optimization of side-chain conformation for all residue positions that had a higher energy after step 1 by cycling through each rotamer at each position in turn and replacing the current side-chain conformation with the lowest energy rotamer conformation. (iii) Optimization of the backbone torsion angles in a 10-residue window surrounding the site of insertion by energy minimization using a quasi-Newton method (25). Moves were accepted or rejected on the basis of the energy difference between the final minimized structure and the starting structure according to the Metropolis criterion. The same energy function was used for backbone optimization and sequence design. Each round of backbone relaxation consisted of several thousand such Monte Carlo minimization moves; a full combinatorial optimization of side-chain rotamer conformations was carried out with the use of a Monte Carlo procedure every 20 moves.

For each starting structure, five independent simulations, each with 15 cycles of sequence design and backbone optimization, were used to obtain low-energy structure sequence pairs. Final energies were comparable to those observed for naturally occurring proteins. Proteins de-

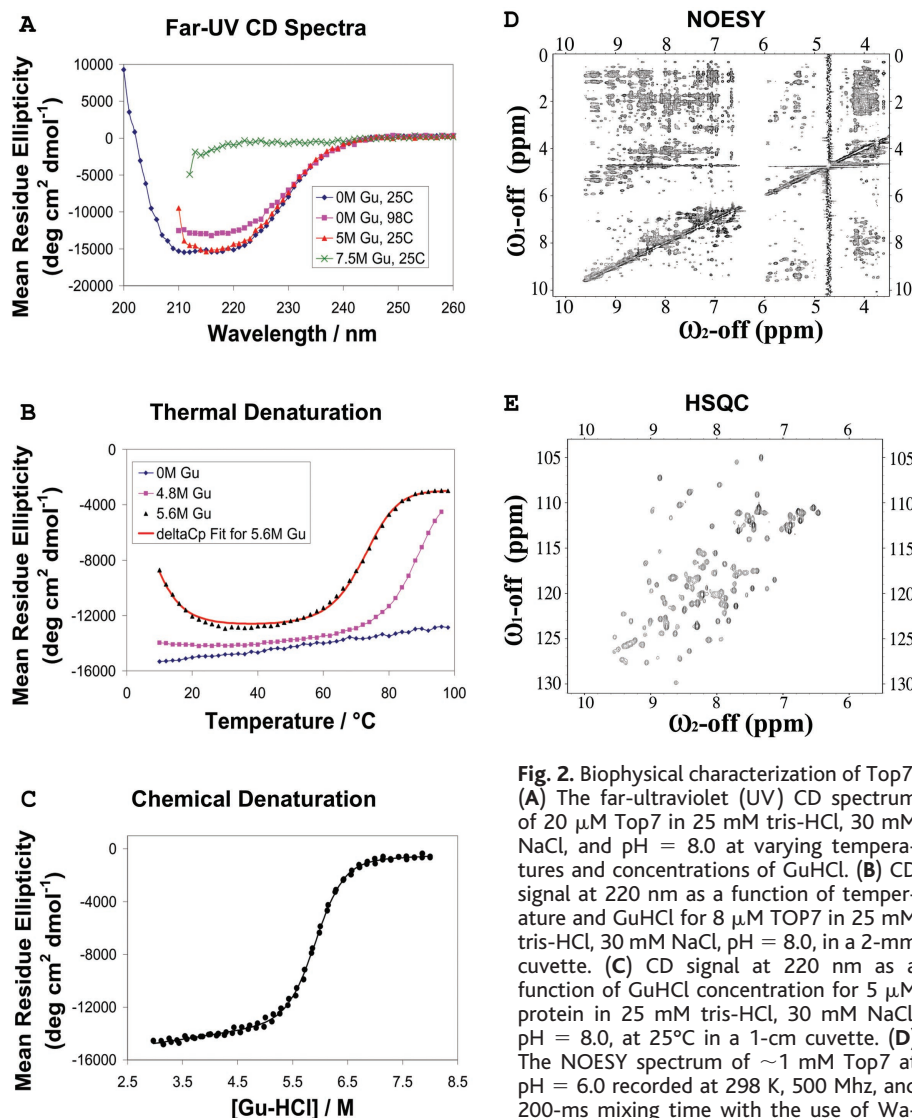


Fig. 2. Biophysical characterization of Top7. (A) The far-ultraviolet (UV) CD spectrum of $20 \mu\text{M}$ Top7 in 25 mM tris-HCl, 30 mM NaCl, and $\text{pH} = 8.0$ at varying temperatures and concentrations of GuHCl. (B) CD signal at 220 nm as a function of temperature and GuHCl for $8 \mu\text{M}$ TOP7 in 25 mM tris-HCl, 30 mM NaCl, $\text{pH} = 8.0$, in a 2-mm cuvette. (C) CD signal at 220 nm as a function of GuHCl concentration for $5 \mu\text{M}$ protein in 25 mM tris-HCl, 30 mM NaCl, $\text{pH} = 8.0$, at 25°C in a 1-cm cuvette. (D) The NOESY spectrum of $\sim 1 \text{ mM}$ Top7 at $\text{pH} = 6.0$ recorded at 298 K , 500 MHz , and 200-ms mixing time with the use of Watergate suppression. ω , frequency. (E) The

HSQC spectrum of $\sim 1 \text{ mM}$ ^{15}N -Top7 at $\text{pH} = 6.0$ recorded at 298 K and 500 MHz with the use of the fast HSQC scheme of Mori *et al.* (43).

RESEARCH ARTICLES

signed with the use of an initial version of the protocol with a damped Lennard-Jones repulsive term and a Monte Carlo optimization without the minimization step were observed experimentally to be quite stable but appeared to have somewhat molten cores (26). To optimize steric packing, the atomic radii were reparameterized on the basis of the distances of closest approach of atom pairs in high-resolution protein structures, explicit protons were included on all atoms, the penalty for atom-atom overlaps was greatly steepened, and the full Monte Carlo minimization protocol was used for varying the backbone, resulting in the generation of much lower energy sequence-structure pairs (20% of the final 860 models had more favorable Lennard-Jones energies than an average protein in the PDB). With these improvements, the protocol was used to design a protein sequence called Top7 (27).

The average Lennard-Jones energies for the buried residues in Top7 become favorable during the relaxation process (table S1), and, although the structural changes during the iterative refinement process are modest (the final protein backbone model has an RMSD of 1.1 Å from the starting model), they bring about dramatic changes in the designed sequence: Only 31% of the Top7 residues are identical to those in the starting sequence. Neither the Top7 sequence nor the sequence before the iterative sequence-structure refinement process have significant similarity to any naturally occurring protein sequence; the closest match to the Top7 sequence found with the use of PSI-BLAST (28) in the Non-Redundant protein sequence database is weaker than would be expected by random chance (E value = 1.6).

Biophysical and structural characterization of Top7. The folding, stability, and structure of the Top7 protein (29) were analyzed with the use of a variety of biophysical methods. The Top7 protein is highly soluble (at 25 to 60 mg ml⁻¹) and is monomeric as determined by gel filtration chromatography. The circular dichroism (CD) spectrum of Top7 is characteristic of α/β proteins (Fig. 2A), and the protein is remarkably thermally stable: The CD spectrum at 98°C is nearly indistinguishable from that at 25°C. At intermediate concentrations (~5 M) of guanidine hydrochloride (GuHCl), Top7 unfolds cooperatively with an increase in temperature and exhibits cold denaturation (Fig. 2B). Fitting these data to the Gibbs-Helmholtz equation gave a change in heat capacity at constant pressure (ΔC_p) per residue associated with unfolding of about 10 cal deg⁻¹ mol⁻¹, a typical value for well-folded proteins of this size (30). The GuHCl-induced chemical denaturation of Top7 is cooperative, and the steep transition is characteristic of the two-state unfolding expected for a small, monomeric, single-domain protein (Fig. 2C). Fitting the chemical denaturation data to a two-state unfolding model yields a free energy of unfolding of 13.2 kcal mol⁻¹ at 25°C, indicating that Top7 is more stable than most

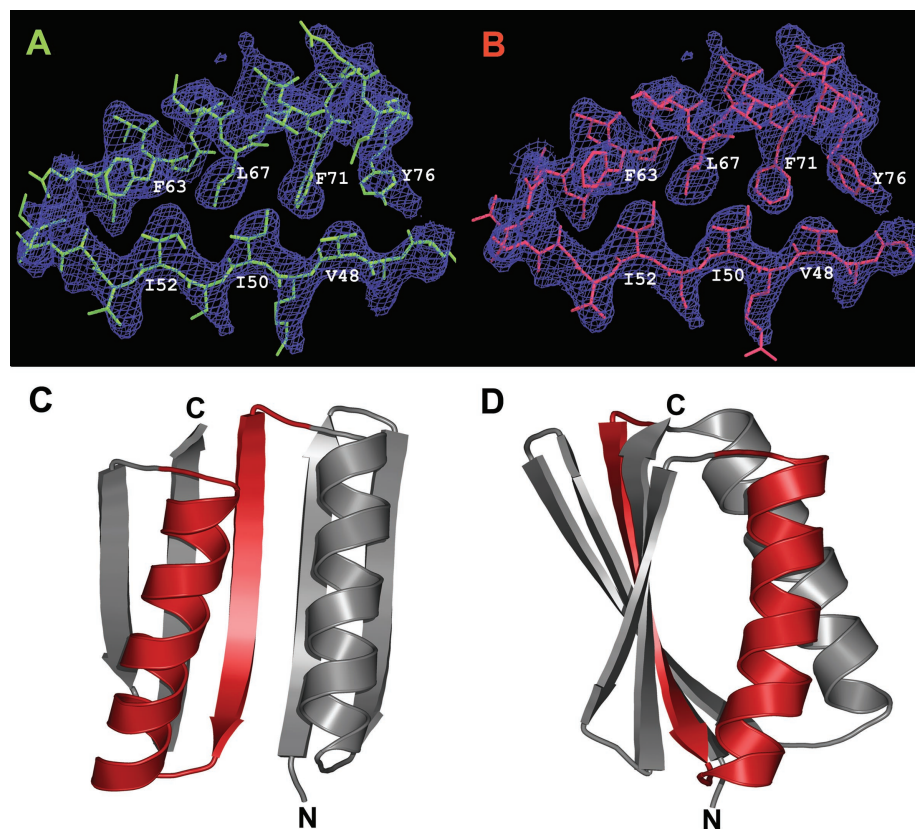


Fig. 3. Schematic representation of Top7 in unbiased SAD density. (A and B) Stick representations of residues 46 to 76 from the computationally designed Top7 (left, green) and from the 2.5 Å x-ray structure (right, red) are shown in unbiased density (blue). The map was generated from SAD phasing from a single SeMet-substituted variant of Top7, followed by density modification. (C and D) Ribbon diagrams of Top7 with residues 46 to 76 highlighted in red. The two diagrams are related by a 90° rotation around the vertical axis.

proteins of similar size (31). The nuclear Overhauser effect spectroscopy (NOESY) and heteronuclear single-quantum coherence (HSQC) spectra of Top7 (Fig. 2, D and E) exhibit features characteristic of a folded protein with substantial β -sheet content. The HSQC spectrum contains the expected number of cross peaks, and the dispersion is comparable to that of α/β proteins of similar size. Strong backbone NH-H α cross peaks and the observation of H α resonances downfield of the water signal (to 6 parts per million) indicate the presence of a β sheet, whereas NH-NH peaks are consistent with a partial helical character for the protein.

Crystallization trials with Top7 yielded crystals that diffracted to 2.5 Å. Remarkably, a strong molecular replacement (MR) solution to the phase problem was found with the use of the design model. This suggested immediately that the design model was quite close to the true structure: Even the small deviations of nuclear magnetic resonance (NMR) solution structures from x-ray crystal structures can make molecular replacement searches fail. To obtain unbiased phase information, we produced and crystallized a selenomethionyl (SeMet)-substituted variant of Top7 with a surface lysine at position 37 mutated to methionine, and we solved the

x-ray crystal structure to 2.5 Å by direct rebuilding into an unbiased single-wavelength anomalous difference (SAD) electron density map (Fig. 3B) and residual difference Fourier maps (32). The final R_{work} and R_{free} were 0.268 and 0.293, respectively (table S2).

The high-resolution crystal structure reveals that the Top7 protein adopts the designed topology (Fig. 4A). Indeed, the structure is strikingly similar to the design model at atomic resolution (1.17 Å RMSD over all backbone atoms). The overall protein structure is very well ordered, with the exception of two turns (comprising residues 11 to 15 and 24 to 31), each of which exhibit elevated B-factors and poor quality electron density. The first of these two turns and the immediately adjoining residues from its neighboring strand deviate the most from the computational model. However, even in this region, the all-atom RMSD between the two models does not exceed 2.8 Å. In contrast, the C-terminal half of the x-ray structure is well ordered and very similar to the computational model; for example, the region from Asp⁷⁸ to Gly⁸⁵ has an all-atom RMSD of 0.79 Å (Fig. 4B). Many side chains in the core of the solved structure are effectively superposable with those of the designed Top7 (Fig. 4C).

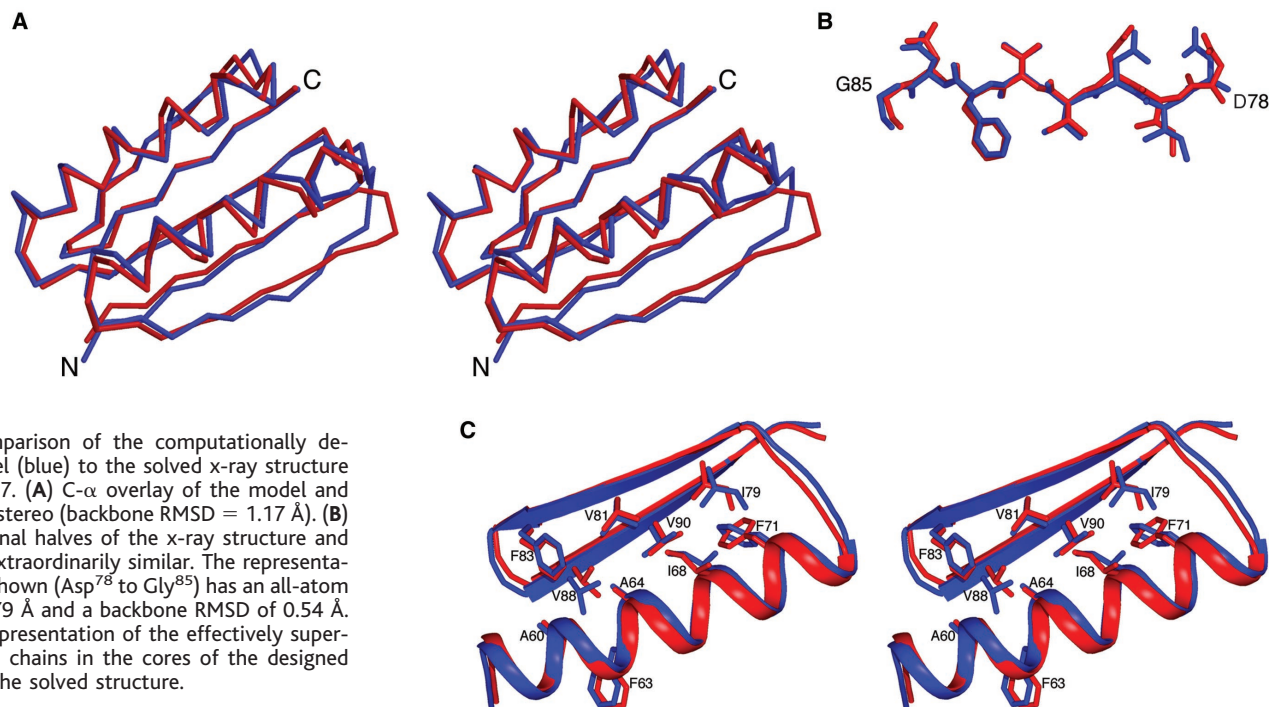


Fig. 4. Comparison of the computationally designed model (blue) to the solved x-ray structure (red) of Top7. **(A)** C- α overlay of the model and structure in stereo (backbone RMSD = 1.17 Å). **(B)** The C-terminal halves of the x-ray structure and model are extraordinarily similar. The representative region shown (Asp⁷⁸ to Gly⁸⁵) has an all-atom RMSD of 0.79 Å and a backbone RMSD of 0.54 Å. **(C)** Stereorepresentation of the effectively superposable side chains in the cores of the designed model and the solved structure.

Like the design model, the Top7 crystal structure is judged to be a novel topology by the TOPS server. The strongest structural similarity found in a Dali search of the PDB (33) is to a discontinuous portion of the 668-residue protein S-adenosylmethionine decarboxylase, which has a large 68-residue insertion between strands 1 and 2, and the third and fourth strands are connected by an unrefined loop instead of a helix. According to A. Murzin, the curator of the Structural Classification of Proteins (SCOP) database, the Top7 fold is not present in SCOP (34, 35).

Implications. The 1.17-Å backbone atom RMSD between the Top7 design model and the crystal structure implies that deep minima in the free energy function used in design correspond to deep minima in the actual free-energy landscape and hence are an important validation of the accuracy of current potential functions. This atomic-level accuracy contrasts sharply with the low accuracy of ab initio structure predictions for naturally occurring sequences: The most accurate structure predictions in the Critical Assessment of Structure Prediction experiments for 90- to 100-residue proteins have RMSDs greater than 4 Å from the experimentally determined structure. Why does the simultaneous optimization of sequence and structure identify the global free energy minimum, whereas the optimization of structure for fixed sequence does not? The answer may involve both of the challenges facing ab initio structure prediction, the vast size and ruggedness of the conformational space to be searched and the limited accuracy of current potential functions. The capability to alter the sequence and hence reconfigure the landscape may greatly facilitate

the search for low-free-energy protein structures as compared to standard ab initio prediction, where the sequence is fixed. In addition, Top7 lacks functional constraints, which can lead to locally suboptimal regions in native structures that are particularly challenging for structure prediction, and the more extensive optimization of the folding free energy may partially compensate for inaccuracies in the potential functions. Finally, it should be noted that the design process focused entirely on minimizing the free energy of the folded monomeric structure; attaining a highly stable new structure did not require extensive negative design against possible alternative conformations (36, 37) nor consideration of the kinetic process of protein folding (38).

The design of Top7 shows that globular protein folds not yet observed in nature not only are physically possible but can be extremely stable. This extends the earlier observation that helical coiled coil geometries not found in nature can be generated by computational protein design (15). The protein therapeutics and molecular machines of the future should thus not be limited to the structures sampled by the biological evolutionary process. The methods used to design Top7 are, in principle, applicable to any globular protein structure and open the door to the exploration and use of a vast new world of protein structures and architectures.

References and Notes

1. B. I. Dahiya, S. L. Mayo, *Science* **278**, 82 (1997).
2. J. R. Desjarlais, T. M. Handel, *Protein Sci.* **4**, 2006 (1995).
3. J. W. Ponder, F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
4. J. Reina *et al.*, *Nature Struct. Biol.* **9**, 621 (2002).

5. L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, *Nature* **423**, 185 (2003).
6. S. M. Malakauskas, S. L. Mayo, *Nature Struct. Biol.* **5**, 470 (1998).
7. E. C. Johnson, G. A. Lazar, J. R. Desjarlais, T. M. Handel, *Structure Fold. Des.* **7**, 967 (1999).
8. P. Koehl, M. Levitt, *J. Mol. Biol.* **293**, 1161 (1999).
9. B. Kuhlman, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10383 (2000).
10. A. Jaramillo, L. Wernisch, S. Hery, S. J. Wodak, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13554 (2002).
11. K. Raha, A. M. Wollacott, M. J. Italia, J. R. Desjarlais, *Protein Sci.* **9**, 1106 (2000).
12. A. Su, S. L. Mayo, *Protein Sci.* **6**, 1701 (1997).
13. S. M. Larson, J. L. England, J. R. Desjarlais, V. S. Pande, *Protein Sci.* **11**, 2804 (2002).
14. B. Kuhlman, J. W. O'Neill, D. E. Kim, K. Y. Zhang, D. Baker, *J. Mol. Biol.* **315**, 471 (2002).
15. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462 (1998).
16. A. E. Keating, V. N. Malashkevich, B. Tidor, P. S. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14825 (2001).
17. The TOPS database is available at www.tops.leeds.ac.uk/.
18. P. M. Bowers, C. E. Strauss, D. Baker, *J. Biomol. NMR* **18**, 311 (2000).
19. T. Kortemme, A. V. Morozov, D. Baker, *J. Mol. Biol.* **326**, 1239 (2003).
20. T. Lazaridis, M. Karplus, *Proteins Struct. Func. Genet.* **35**, 132 (1999).
21. Materials and methods are available as supporting material on Science Online.
22. R. L. Dunbrack, F. E. Cohen, *Protein Sci.* **6**, 1661 (1997).
23. R. Bonneau *et al.*, *Proteins* (suppl. 5), 119 (2001).
24. C. Rohl, C. E. M. Straus, K. M. S. Misura, D. Baker, *Methods Enzymol.*, in press.
25. We used the Davidson-Fletcher-Powell algorithm as described by W. H. Press *et al.*, in *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, ed. 2, 1992), pp. 428–429.
26. B. Kuhlman *et al.*, data not shown.
27. The sequence of Top7 is mgDIQVQVNIIDNGKNFDYTYTVTTESELQKVLNKLKDYIKKQGAKRVRISITARTKKEAEKFAAILIKVFAELGYNDINVTFDGDTVTVEQGLegglehhhhhh; the computationally designed sequence is in uppercase and residues added to allow expression and purification are in lowercase. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H,

- His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
28. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
 29. A synthetic gene with the Top7 sequence under the control of the T7 promoter, a C-terminal 6X His tag, and codon usage optimized for *Escherichia coli* was obtained from BlueHeron Biotechnologies (Bothell, WA). After expression in *E. coli*, the protein was readily purified to >95% homogeneity with the use of nickel affinity chromatography followed by anion exchange chromatography.
 30. J. K. Myers, C. N. Pace, J. M. Scholtz, *Protein Sci.* **4**, 2138 (1995).
 31. K. W. Plaxco, K. T. Simons, D. Baker, *J. Mol. Biol.* **277**, 985 (1998).
 32. The structure of Top7_K35M was solved by molecular replacement with the program EPMR (39) and by direct rebuilding into an unbiased SAD electron density map and residual difference Fourier maps. For molecular replacement, 19 large surface residues such as Lys, Arg, Gln, and Glu were truncated to Ala in the search model. The correlation coefficient for the initial MR search, using data to 4.0 Å resolution, was 0.52, compared with background of 0.36. For SAD phasing, the position of SeMet 37 was determined from an anomalous difference Patterson map. The initial phasing power and figure of merit for SAD phasing were 1.99 and 0.24, respectively, before density modification. An interpretable electron density map was obtained after density modification with solvent flipping with a solvent content of 43% with the use of CNS (40). An initial model was built with the use of XtalView (41) and O (42). The model was refined with CNS with the use of the mlhl target (maximum likelihood, Hendrickson-Lattman coefficients) with 5% of the data excluded for the calculation of the cross-validating free R. Of all the built residues, 88% are in the most favorable regions of Ramachandran space, and 12% are in the allowed regions. Statistics from phasing and refinement are shown in table S2.
 33. L. Holm, C. Sander, *Trends Biochem. Sci.* **20**, 478 (1995).
 34. T. J. Hubbard, A. G. Murzin, S. E. Brenner, C. Chothia, *Nucleic Acids Res.* **25**, 236 (1997).
 35. A. Murzin, personal communication.
 36. J. J. Havranek, P. B. Harbury, *Nature Struct. Biol.* **10**, 45 (2003).
 37. W. Jin, O. Kambara, H. Sasakawa, A. Tamura, S. Takada, *Structure (Cambridge)* **11**, 581 (2003).
 38. L. Mirny, E. Shakhnovich, *J. Mol. Biol.* **308**, 123 (2001).
 39. EPMR: A program for crystallographic molecular replacement by evolutionary search (C. R. Kissinger, D. K. Gehlhaar, Agouron Pharmaceuticals, La Jolla, CA).
 40. A. T. Brünger *et al.*, *Acta Crystallogr. D. Biol. Crystallogr.* **54**, 905 (1998).
 41. D. E. McRee, *J. Struct. Biol.* **125**, 156 (1999).
 42. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr. A* **47**, 110 (1991).
 43. S. Mori, C. Abeygunawardana, M. O. Johnson, P. C. van Zijl, *J. Magn. Reson. B.* **108**, 94 (1995).
 44. We acknowledge the expert assistance of B. Shen in crystallographic phasing, modeling, and refinement of the TOP7 structure, C. Rohl for aiding in the incorporation of RosettaDesign into Rosetta, C. Strauss for helping to generate the initial models of Top7, T. Leeper for help with two-dimensional NMR studies, and R. Kleivt and the Kleivt laboratory for help with preliminary NMR characterization of Top7. Academic users can obtain licensing information for RosettaDesign at www.unc.edu/kuhlmanpg/rosettadesign.htm. The coordinates and structure factors for the Top7 x-ray crystal structure have been deposited in the PDB with accession code 1QYS. B.K. was supported by a fellowship from the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation. This work was also supported by NIH grants to G.V., B.L.S., and D.B.

Supporting Online Material

www.sciencemag.org/cgi/content/full/302/5649/1364/DC1

Materials and Methods
Tables S1 to S6

21 July 2003; accepted 25 September 2003

Targeted Protein Degradation and Synapse Remodeling by an Inducible Protein Kinase

Daniel T. S. Pak*† and Morgan Sheng†

Synaptic plasticity involves the reorganization of synapses at the protein and the morphological levels. Here, we report activity-dependent remodeling of synapses by serum-inducible kinase (SNK). SNK was induced in hippocampal neurons by synaptic activity and was targeted to dendritic spines. SNK bound to and phosphorylated spine-associated Rap guanosine triphosphatase activating protein (SPAR), a postsynaptic actin regulatory protein, leading to degradation of SPAR. Induction of SNK in hippocampal neurons eliminated SPAR protein, depleted postsynaptic density-95 and Bassoon clusters, and caused loss of mature dendritic spines. These results implicate SNK as a mediator of activity-dependent change in the molecular composition and morphology of synapses.

Synaptic activity can induce a variety of changes within postsynaptic neurons, ranging from transient posttranslational modifications to altered programs of gene expression. Long-lasting forms of synaptic plasticity require new gene expression and protein synthesis (1–3). Some activity-inducible genes may mediate the conversion of short-term responses to long-term changes by altering synaptic structure (4, 5).

Numerous activity-inducible genes have been identified (6, 7). Notably, few protein

kinases are known to be inducible by synaptic activity at the mRNA level; the best-characterized induction profiles are those of SNK and FGF-inducible kinase (FNK) (8). However, the roles of these polo family kinases in plasticity are unknown.

One reasonable expectation is that synaptic remodeling will involve the dismantling and/or reorganization of key cytoskeletal and scaffolding protein complexes. In the postsynaptic density (PSD) of mammalian excitatory synapses, actin is the major cytoskeletal element, and scaffold proteins of the PSD-95 family are important for assembling glutamate receptors with their signaling-cytoskeletal complexes (9, 10). One PSD-95-interacting partner, SPAR [spine-associated Rap guanosine triphosphatase (GTPase) activating protein (GAP)], is a multidomain postsynaptic protein that controls dendritic spine shape by regulating

actin arrangement as well as signaling by the small GTPase Rap (11). Actin dynamics and Rap activity are both regulated by synaptic activity and involved in synaptic plasticity (12–16). With its postsynaptic location in the N-methyl-D-aspartate (NMDA) receptor-PSD-95 complex, SPAR is an attractive candidate for mediating activity-dependent remodeling of synapses.

Interaction of SPAR and SNK. We screened for SPAR-binding proteins with the use of individual domains of SPAR (Fig. 1A) as bait in the yeast two-hybrid system (17). SPAR contains two actin regulatory domains, termed Act1 and Act2, a GAP domain specific for Rap, a PDZ domain of unknown function, and a C-terminal region (termed GKBD) that binds specifically to the guanylate kinase domain of PSD-95 (11). When the Act2 domain was used to screen a brain cDNA library, one of the positive clones (clone 19) isolated was SNK, initially identified in fibroblasts as an mRNA transcript induced by mitogenic stimulation (18). Clone 19 encoded roughly the C-terminal half of SNK protein (amino acids 395 to 682; hereon termed SNKc) (Fig. 1A), including a motif characteristic of the polo family of kinases (19, 20). Full-length SNK also bound to Act2 in the yeast two-hybrid assay. Neither full-length SNK nor clone 19 interacted with the GKBD region of SPAR (Fig. 1A).

The interaction between SNK and SPAR was confirmed with the use of an in vitro precipitation assay in which GST fused to SNKc precipitated full-length SPAR expressed in COS-7 cells but not a SPAR construct with an internal deletion of the Act2 domain (SPAR Δ Act2) (Fig. 1B). The isolated Act2 domain also bound GST-SNKc with greater efficiency than full-length SPAR. GST alone failed to bind any of these SPAR constructs.

Picower Center for Learning and Memory, RIKEN Massachusetts Institute of Technology (MIT) Neuroscience Research Center, Howard Hughes Medical Institute, MIT, Cambridge, MA 02139, USA.

*Present address: Department of Pharmacology, Georgetown University, Washington, DC 20057, USA.
†To whom correspondence should be addressed. E-mail: dtp6@georgetown.edu (D.T.S.P.); msheng@mit.edu (M.S.)