

Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments

Jennifer E. Padilla^{*†‡}, Christos Colovos^{*†‡}, and Todd O. Yeates^{*†§}

^{*}Department of Chemistry and Biochemistry, and [†]Department of Energy Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, CA 90095-1569

Communicated by Paul D. Boyer, University of California, Los Angeles, CA, December 22, 2000 (received for review November 9, 2000)

A general strategy is described for designing proteins that self assemble into large symmetrical nanomaterials, including molecular cages, filaments, layers, and porous materials. In this strategy, one molecule of protein A, which naturally forms a self-assembling oligomer, A_n , is fused rigidly to one molecule of protein B, which forms another self-assembling oligomer, B_m . The result is a fusion protein, A-B, which self assembles with other identical copies of itself into a designed nanohedral particle or material, $(A-B)_p$. The strategy is demonstrated through the design, production, and characterization of two fusion proteins: a 49-kDa protein designed to assemble into a cage approximately 15 nm across, and a 44-kDa protein designed to assemble into long filaments approximately 4 nm wide. The strategy opens a way to create a wide variety of potentially useful protein-based materials, some of which share similar features with natural biological assemblies.

A central goal of nanotechnology is to design and fabricate novel materials with sizes in the nanometer range (1). These materials fall into different architectural classes, such as compact clusters (2), hollow shells (3–5) and tubes (6–8), two-dimensional layers (9, 10), and nanoporous materials (11–13). In recent years, a wide variety of chemical building blocks and synthetic strategies have been investigated. Specific methods have produced interesting new materials, but a single general approach for fabricating materials having many different architectures and symmetries has not emerged. In the present work, we move a step closer to this fundamental goal with a method for engineering self-assembling nanomaterials by combining naturally symmetric protein components.

Symmetry provides a powerful tool for building large regular objects. Our general strategy for using symmetry to construct protein nanomaterials is illustrated in Fig. 1. Many natural proteins have evolved on their own to self associate through noncovalent interactions. Those that associate two at a time (dimers) or three at a time (trimers) are relatively common (Fig. 1*a*). These natural proteins make up our raw building materials. We refer to them here as oligomerization domains. Following specific geometric rules, two oligomerization domains are connected by genetic manipulation into a single larger molecule called a fusion protein. A fusion protein therefore carries two parts, each with a strong tendency to associate with other copies of itself. As a consequence of this design, many identical copies of the fusion protein self assemble into a highly symmetric object or extended material we call a protein nanohedron or a protein nanohedral material.

A rich variety of nanohedral structures can be designed that take forms such as cages and shells or unbounded layers, filaments, or three-dimensional crystals. We refer to these as separate architectural classes. Within each architectural class, several different symmetries may be possible. Which symmetries and architectures can be designed with this strategy, and what geometric rules of construction must be followed to achieve them? Those that can be built by fusing together two dimeric or trimeric oligomerization domains are listed in Table 1. A fusion protein carries two virtual symmetry axes, one from each oligomerization domain (Fig. 1*b*). To achieve a given symmetry and

architecture, there is a construction rule (Table 1) that describes the fixed geometric relationship the two symmetry axes must have with each other. For example, for a nanohedral protein layer with hexagonal ($p6$) symmetry, the construction rule is that the symmetry axis from the dimeric oligomerization domain and that from the trimeric oligomerization domain must be nonintersecting and parallel (Fig. 1*d*).

To satisfy the construction rule for a particular design, the two oligomerization domains in a fusion protein must be held together in a relatively rigid fashion. If the oligomerization domains are instead fused in arbitrary or flexible ways, the fusion proteins might assemble in an irregular fashion to give a material whose structure and properties cannot be anticipated. The method we used in this work to enforce the correct disposition of the separate oligomerization domains was to choose them from the set of oligomeric proteins whose three-dimensional structures are known and which begin or end in α -helices. The two oligomerization domains may then be fused by genetically engineering a connecting segment of amino acids that strongly prefers the α -helical conformation. The continuous α -helix extending from one oligomerization domain into the other provides the rigidity and directionality required to produce a fusion protein composed of two predictably connected oligomerization domains (Fig. 1*c*) with symmetry elements conforming to prescribed geometric rules (Table 1). This strategy has been applied successfully to produce two new protein materials.

Experimental Procedures

A Designed Cage. To design a symmetric cage, we identified those dimeric and trimeric protein structures in the Protein Data Bank (14) that begin or end in an α -helix. By using a computer program, the dimers and trimers were considered in all pairwise combinations as potential oligomerization domains. The two oligomerization domains were connected computationally in three dimensions by a short intervening α -helical linker with a variable number of amino acid residues (Fig. 1*c*). Each hypothetical fusion protein model was examined to see whether the component oligomerization domains were arranged in a way that nearly satisfied the construction rules for a cage (Table 1). After checking for steric clashes in fully assembled models, several promising designs were obtained. We chose one with tetrahedral symmetry, intended to self assemble from 12 copies of a 49-kDa fusion protein.

The oligomerization domains of the fusion protein were the trimeric bromoperoxidase (ref. 15; DNA kindly provided by Karl-Heinz van Pée, Technische Universität, Dresden, Germany) and the dimeric M1 matrix protein of influenza virus (ref. 16; DNA kindly provided by Ming Luo, University of Alabama, Birmingham) connected by a nine-residue helical linker. According to the design, the tetrahedral cage was expected to be

[†]J.E.P. and C.C. contributed equally to this work.

[§]To whom reprint requests should be addressed at: UCLA Department of Chemistry and Biochemistry, P.O. Box 951569, Los Angeles, CA 90095-1569. E-mail: yeates@mbi.ucla.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

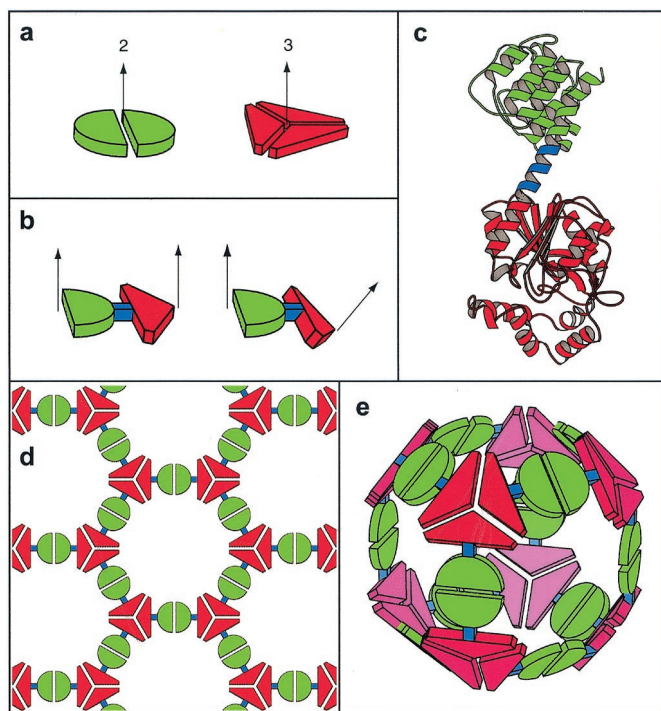


Fig. 1. A general strategy for designing fusion proteins that assemble into symmetric nanostructures. (a) The green semicircle represents a natural dimeric protein (i.e., a protein that associates with one other copy of itself), whereas the red shape represents a trimeric protein. The symmetry axes of the natural oligomers are shown. (b) The two natural proteins are combined by genetic methods into a single fusion protein. Each of the original natural proteins serves as an “oligomerization domain” in the designed fusion protein. Two different hypothetical fusion proteins are shown to illustrate that the oligomerization domains can be joined rigidly in different geometries. (c) A ribbon diagram of a fusion protein showing one method for joining two oligomerization domains (red and green) in a relatively rigid fashion. One of the natural oligomerization domains must end in an α -helical conformation, and the other must begin in an α -helical conformation. The two are then linked by a short stretch of amino acids (blue) that have a strong tendency to adopt an α -helical conformation. Thus, the two oligomerization domains are joined physically in a predictable orientation. (d) A designed fusion protein self assembles into a particular kind of nanostructure that depends on the geometry of the symmetry axes belonging to its component oligomerization domains (Table 1). A molecular layer arises from an arrangement like that in *b* (Left). (e) A cubic cage arises from an arrangement like the one in *b* (Right).

approximately 15 nm on an edge and about 9 nm from the center to the perimeter.

PCR primers were designed to amplify a DNA fragment containing the bromoperoxidase gene [Protein Data Bank (PDB) ID code 1bro] and add to it the nine-residue linker KALEAQKQK [an amino acid sequence chosen from a portion of a long helix in ribosomal protein L9 (PDB ID code 1div)] and part of influenza virus matrix protein M1 (PDB ID code 1aa7). The final product consisted of residues 1–276 of bromoperoxidase, followed by the amino acid sequence KALEAQKQK, followed by residues 3–164 of influenza virus matrix protein M1. After ligation into a pET21b vector and bacterial transformation, the protein was produced in *Escherichia coli* BL21(DE3) cells growing at 37°C by inducing expression at an optical density at 600 nm of 0.6 with 1 mM isopropyl-thio- β -D-galactoside and harvesting 3 h later. Cells were lysed in a French press in buffer A [100 mM NaCl/10 mM Hepes/0.02% azide (pH 7.2)] and a mixture containing EDTA, PMSF, benzamidine, lysozyme, DNase I, and RNase A. After centrifugation, the soluble cell lysate was precipitated with ammonium sulfate. The fusion

protein was abundant in the 35–40% ammonium sulfate fraction. The pellet was resuspended first in Buffer A, then run over a Q-Sepharose anion exchange column and eluted with a 0.1 M to 1 M NaCl gradient. After buffer exchange back to Buffer A by using a centrifuge concentrator, the protein was run over a nickel chelating column and eluted with buffer A plus 500 mM imidazole. Although the protein does not have a polyhistidine tail, it bound to the column with useful affinity. The protein was further purified on a Sephacryl S200h size exclusion column (Amersham Pharmacia) in Buffer A. Approximately 1 mg of pure protein was obtained per liter of cell culture.

For electron microscopy, carbon support films mounted on grids were made hydrophilic immediately before use by high-voltage, alternating current-glow discharge. Samples at 0.02 mg/ml were applied directly onto grids and allowed to adhere for 2 min. Grids were rinsed with three drops of distilled water and negatively stained with 1% (vol/vol) uranyl acetate for 1 min. Specimens were examined in a Hitachi (Tokyo) H-7000 electron microscope at an accelerating voltage of 75 kV.

For analysis by equilibrium sedimentation, a sample at 0.7 mg/ml in buffer A was spun at 4,000 rpm in a Beckman Optima XL-A-70 analytical ultracentrifuge. A curve was fitted to the data by using Beckman ORIGIN-based software (Version 3.01) to determine the molecular mass.

Dynamic light scattering was measured from a filtered sample in Buffer A at 0.3 mg/ml by using a Protein Solutions, Inc. (Charlottesville, VA) Dyna Pro Molecular Sizing instrument and DYNAMICS 4.0 software.

Table 1. Construction rules for designing nano-hedral protein materials from dimeric and trimeric oligomerization domains

Symmetry*	Construction	Geometry of symmetry elements [†]	
Cages and shells			
T [‡]	Dimer-trimer	54.7°	I
O [§]	Dimer-trimer	35.3°	I
I	Dimer-trimer	20.9°	I
Double-layer rings			
D _n	Dimer-dimer	180°/n	I
Two-dimensional layers			
p6	Dimer-trimer	0°	N
p321	Dimer-trimer	90°	N
p3	Trimer-trimer	0°	N
Three-dimensional crystals			
I2 ₁ 3	Dimer-trimer	54.7°	N
P4 ₁ 32 or P4 ₃ 32	Dimer-trimer	35.3°	N
P23	Trimer-trimer	70.5°	N
Helical filaments			
Helical	Dimer-dimer	Any angle,	N

This table gives only those symmetries that can be constructed by combining two oligomerization domains of the dimeric or trimeric type. Other kinds of oligomerization domains, such as tetramers, would give additional possibilities not listed here.

*T, O, I, and D_n refer to tetrahedral, octahedral, icosahedral, and dihedral symmetry, respectively. The remaining symbols are denoted by their Hermann–Mauguin symbols.

[†]The angle formed between the two symmetry elements is given, followed by I or N to denote intersecting or nonintersecting axes.

[‡]See Fig. 2c.

[§]See Fig. 1e.

^{||}See Fig. 1d.

^{||}See Fig. 3b.

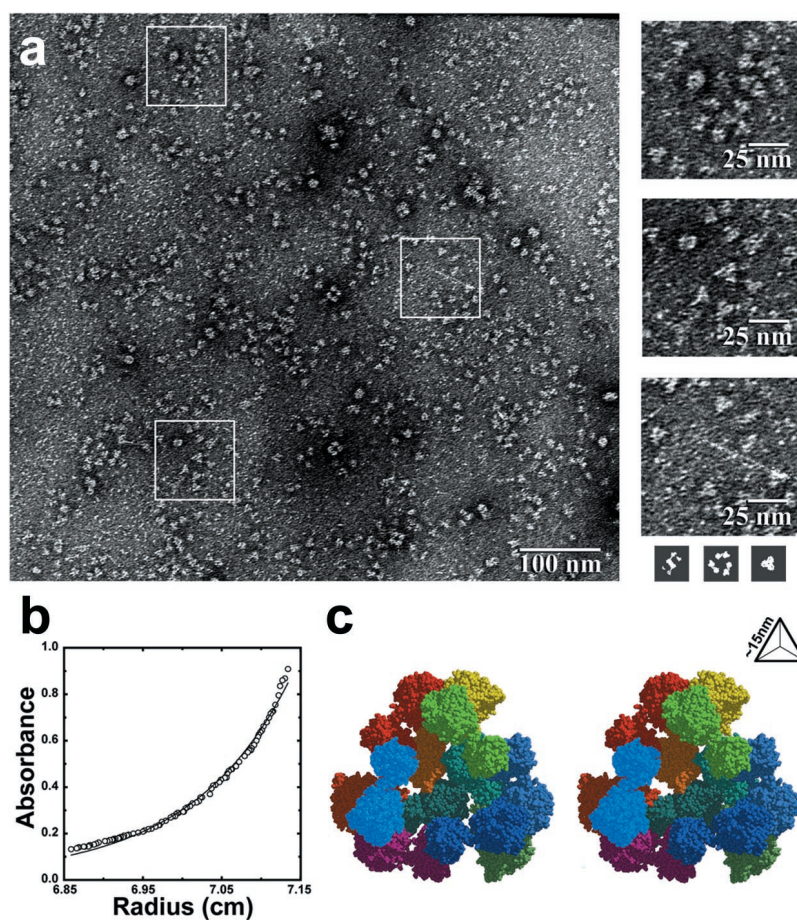


Fig. 2. Characterization of a designed tetrahedral protein cage. (a) Negatively stained electron micrographs show images of discrete particles. The images left in the heavy-atom stain are consistent with the sizes of the largest faces of the cage. For size comparison (shown to scale, *Bottom Right*), three simulated images were calculated from the atomic coordinates of the cage in three orientations where it would make the most extensive contacts with the surface of the electromagnetic support grid. As a rough approximation, it was assumed that the complex would leave a footprint in a layer of heavy-atom stain 15 Å thick. (b) Equilibrium sedimentation shows that the major component has a molecular mass of approximately 550 kDa, corresponding roughly to 11.3 subunits (close to the anticipated value of 12). A small degree of polymorphism is evident from the residual difference between the experimental and theoretical curves. (c) A stereo model of the tetrahedral protein cage as it was intended to assemble from 12 copies of the 49-kDa engineered fusion protein (shown in Fig. 1c). The view is through one of the four large openings in the cage. The particle radius is approximately 9 nm, and the edge length is approximately 15 nm. The separate protein subunits are colored differently.

A Designed Filament. A protein filament was chosen as the second design target to demonstrate the construction of an unbound nanomaterial. According to the construction rules (Table 1), the filament architecture can be built from a fusion protein with two dimeric oligomerization domains, connected in such a way that their symmetry axes do not intersect. By using the same computational strategy described above, we identified pairs that could be fused with satisfactory geometry from among a set of potential oligomerization domains (Fig. 3). The chosen design was a fusion of influenza virus matrix protein M1 (16) and carboxylesterase (ref. 17; DNA kindly provided by Ook Joon Yoo and Se Won Suh, Seoul National University, Seoul, Korea) connected by a 5-aa α -helical linker (Fig. 3). This 44-kDa fusion protein was engineered and purified from *E. coli*.

The carboxylesterase gene (Protein Data Bank ID code 1auo) was amplified from a plasmid by using PCR, and the DNA was extended by using primers to include the five-residue linker KDTDS (an amino acid sequence chosen from a helix connecting two domains of calmodulin) and a portion of influenza virus matrix protein M1. This DNA was digested and ligated into the pET21b vector containing influenza virus matrix protein M1. The fusion protein contained residues 1–216 of carboxylesterase, followed by the amino acid sequence KDTDS, followed by

residues 3–164 of influenza virus matrix protein M1. The resulting DNA construct was then digested and ligated into the pET15b vector to add an N-terminal histidine tag.

The protein was expressed in *E. coli* BL21(DE3) cells at 37°C, induced with 1 mM isopropyl-thio- β -D-galactoside, and harvested 3 h later. As expected, the solubility was very low, so the protein in inclusion bodies was solubilized in 8 M urea/100 mM Tris (pH 8.0), then purified. The unfolded protein was run over a nickel chelating column and eluted with the urea buffer plus 500 mM imidazole. The purified protein was refolded by dialyzing against a refolding buffer [100 mM Tris (pH 7.8)/1.3 M urea/100 mM glycine/0.1 mM GSSG (oxidized glutathione)/1 mM reduced glutathione] and then dialyzing against PBS. Electron microscopy was performed as before, on a suspension where the concentration was approximately 1 mg/ml.

Results

Designed Cage. After purification, the 49-kDa fusion protein, designed to form a tetrahedral cage from 12 subunits, remained soluble at concentrations up to 20 mg/ml. Various experimental methods were used to reveal the mode of assembly of the designed protein (Fig. 2). The most definitive results came from equilibrium sedimentation, which gave a shape-independent

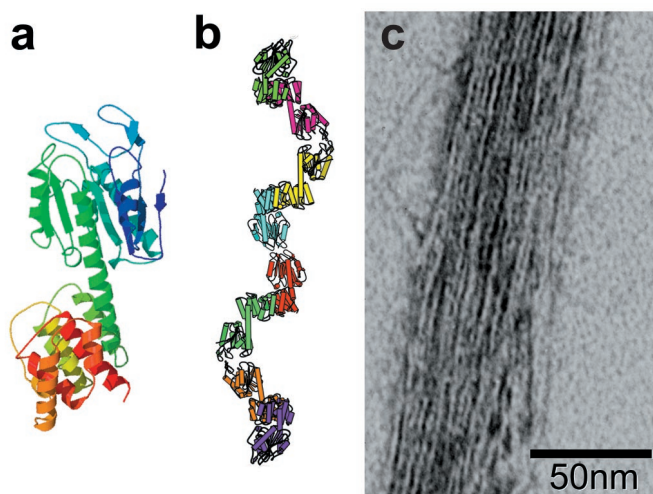


Fig. 3. Electron microscopy and model of a designed protein filament. (a) A ribbon model of a single molecule of the designed fusion protein. (b) A ribbon model of the protein filament as it was intended to assemble, with separate protein molecules colored differently. (c) Negatively stained electron micrograph of a bundle of filaments formed by the designed fusion protein. The bundle is 15–20 filaments across and reveals details indicative of the individual dimeric oligomerization domains that make up the fusion protein. In addition to bundles, networks of filaments were also observed in other micrographs.

molecular mass of 550 kDa (Fig. 2*b*), corresponding very nearly to the expected 590 kDa for 12 subunits. Light-scattering and electron microscopy (Fig. 2*a*) further supported the finding that the designed protein assembled essentially as intended. Discrete particles were seen in the negative stain electron micrographs at sizes consistent with the model of the tetrahedral complex. The molecular friction coefficient was measured by dynamic light scattering. Although the frictional coefficient can be converted to a valid hydrodynamic radius only for a solid, roughly spherical particle, the observed values were consistent with the design. The frictional coefficient equated to a spherical particle with a radius between 6 and 7.5 nm, which is as close to the model value of 9 nm for the designed assembly as could be expected for a cage-like structure.

In all experiments, minor components could also be detected, some smaller and some larger than 12 subunits, reflecting association–dissociation equilibrium and possible flexibility or polymorphism in the assembled particles. Although the symmetry of the protein cage could not be verified by these experiments, the correct architectural class in the expected size range was demonstrated.

Designed Protein Filament. The 44-kDa fusion protein which had been designed to self assemble into filaments was expressed in *E. coli* and extracted from insoluble inclusion bodies. The purified insoluble material was sonicated and examined by electron microscopy and found to contain essentially linear filaments, each approximately 4 nm wide. As is relatively common for natural protein filaments, these designed filaments are found arranged in networks or bundles under electron microscopy conditions (Fig. 3*c*). The width of the individual filaments and the visible details are consistent with the features of the designed filament.

Discussion and Conclusions. A general strategy for designing many types of protein assemblies was tested in the laboratory on two designs, a cage and a filament. These successful experiments demonstrate two of the possible architectural classes and thereby validate the protein-design strategy. Now, an endless variety of nanohedral protein materials can be investigated. Useful applications may arise for many of the different architectural classes.

Cage-like structures can be built to follow any of the cubic point symmetries (tetrahedral, octahedral, and icosahedral) assembling respectively from 12, 24, and 60 copies of a fusion protein (Fig. 1*e*). Depending on the specific shapes of the oligomerization domains, some cage structures may be relatively open, whereas others may be more compact. Hollow structures may be useful for delivering drugs or genes or for stabilizing, shielding, or sequestering other molecules in their interior volumes (18, 19). More compact structures might be useful for presenting multiple copies of viral or bacterial antigens or other chemical groups on their surfaces.

Although their construction has not been demonstrated yet, the design strategy also provides for ordered protein arrays that extend indefinitely in two or three dimensions (Fig. 1*d*). By using different approaches, some success has been reported already with nucleic acids (10) and with a carbohydrate-binding protein designed to assemble into an ordered array (20). Well ordered molecular layers may have applications as biosensors or detectors (21). Layers and solids with large pores might also be useful as molecular sieves. Porous materials have been fabricated from silicates and more recently from metal sulfides and metal phosphates, but it has been difficult to exceed a pore diameter of roughly 1.4 nm (11, 13). Two- and three-dimensional nanohedral protein materials could have precisely defined pore sizes in the 5-nm to 20-nm range.

Many recent efforts in protein engineering have focused on redesigning the internal structures of individual protein domains. Various studies have succeeded in increasing thermal stability (22, 23), incorporating binding sites for metals or ligands (24, 25), and controlling relatively simple oligomerization (26, 27). These tend to be challenging problems, mainly because success depends on a detailed understanding of protein energetics, binding, or catalysis.

The nanohedral protein-design strategy developed here differs not only in its goal to create large symmetric assemblies, but also in the nature of the protein redesign. Here, no attempt is made to alter the internal structures of individual protein domains. Instead, it is sufficient to fuse naturally occurring proteins without making any internal modifications. Another attractive feature of the method is the combinatorial power that comes from connecting multiple protein components. In fact, this combinatorial feature is critical to satisfy the specific geometric rules of construction (Table 1).

The creation of two protein materials with very different architectures shows the potential power of this design strategy. The experiments also identify some limitations and obstacles to be addressed in future work. The filament results confirm that bacterial expression and purification may be problematic for some unbound nanohedral materials. The tetrahedral cage results show that flexibility or association–dissociation equilibria may lead to some heterogeneity in the nanostructures. Nonetheless, these first experiments point to a very promising avenue for creating a vast array of new biological materials.

Beyond their potential uses as novel materials, designed nanohedral structures could improve our understanding of many natural biological assemblies. The design and study of synthetic structures might shed light on the evolution of such diverse natural assemblies as viral capsids (28), clathrin cages (29), bacterial S-layers (21), and on the formation of protein fibers implicated in Alzheimer's disease and other amyloid pathologies (30).

We thank Mari Gingery for electron microscopy, Martin Phillips of the Department of Energy Instrumentation Facility, University of California Los Angeles for ultracentrifugation; Daniel Anderson for technical assistance; Florence Lee for expression and purification of the filament design protein; and David Eisenberg for helpful discussions. We are indebted to Ming Luo, Karl-Heinz van Pée, Ook Joon Yoo, and Se Won Suh for contributing the clones used to demonstrate the design strategy. This work was supported by the Department of Energy Grant DE-FC03-87ER60615 and the National Institutes of Health Grant GM31299.

1. Whitesides, G. M., Mathias, J. P. & Seto, C. T. (1991) *Science* **254**, 1312–1319.
2. Herron, N. & Thorn, D. L. (1998) *Adv. Mater.* **10**, 1173–1184.
3. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. (1985) *Nature (London)* **318**, 162–163.
4. Zhang, Y. & Seeman, N. C. (1994) *J. Am. Chem. Soc.* **116**, 1661–1669.
5. Huang, H. Y., Remsen, E. E., Kowalewski, T. & Wooley, K. L. (1999) *J. Am. Chem. Soc.* **121**, 3805–3806.
6. Iijima, S. (1991) *Nature (London)* **354**, 56–58.
7. Ghadiri, M. R., Granja, J. R., Milligan, R. A., McRee, D. E. & Khazanovich, N. (1993) *Nature (London)* **366**, 324–327.
8. Ajayan, P. M. & Ebbesen, T. W. (1997) *Rep. Prog. Phys.* **60**, 1025–1062.
9. Russell, V. A., Evans, C. C., Li, W. J. & Ward, M. D. (1997) *Science* **276**, 575–579.
10. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. (1998) *Nature (London)* **394**, 539–544.
11. Bu, X., Feng, P. & Stucky, G. D. (1997) *Science* **278**, 2080–2085.
12. Chui, S. S. Y., Lo, S. M. F., Charmant, J. P. H., Orpen, A. G. & Williams, I. D. (1999) *Science* **283**, 1148–1150.
13. Li, H. L., Laine, A., O'Keefe, M. & Yaghi, O. M. (1999) *Science* **283**, 1145–1147.
14. Abola, E. E., Sussman, J. L., Prilusky, J. & Manning, N. O. (1997) *Methods Enzymol.* **277**, 556–571.
15. Hecht, H. J., Sobek, H., Haag, T., Pfeifer, O. & van Pée, K. H. (1994) *Nat. Struct. Biol.* **1**, 532–537.
16. Sha, B. D. & Luo, M. (1997) *Nat. Struct. Biol.* **4**, 239–244.
17. Kim, K. K., Song, H. K., Shin, D. H., Hwang, K. Y., Choe, S., Yoo, O. J. & Suh, S. W. (1997) *Structure (London)* **5**, 1571–1584.
18. Rebek, J. (1996) *Chem. Soc. Rev.* **25**, 255–264.
19. Douglas, T. & Young, M. (1998) *Nature (London)* **393**, 152–155.
20. Dotan, N., Arad, D., Frolow, F. & Freeman, A. (1999) *Angew. Chem. Int. Ed. Engl.* **38**, 2363–2366.
21. Sára, M. & Sleytr, U. B. (1996) *Micron* **27**, 141–156.
22. Zhang, X. J., Baase, W. A., Shoichet, B. K., Wilson, K. P. & Matthews, B. W. (1995) *Protein Eng.* **8**, 1017–1022.
23. Malakauskas, S. M. & Mayo, S. L. (1998) *Nat. Struct. Biol.* **5**, 470–475.
24. Wilcox, S. K., Putnam, C. D., Sastry, M., Blankenship, J., Chazin, W. J., McRee, D. E. & Goodin, D. B. (1998) *Biochemistry* **37**, 16853–16862.
25. Wisz, M. S., Garrett, C. Z. & Hellinga, H. W. (1998) *Biochemistry* **37**, 8269–8277.
26. Ogihara, N. L., Weiss, M. S., Degrado, W. F. & Eisenberg, D. (1997) *Protein Sci.* **6**, 80–88.
27. Nautiyal, S., Woolfson, D. N., King, D. S. & Alber, T. (1995) *Biochemistry* **34**, 11645–11651.
28. Caspar, D. L. D. & Klug, A. (1962) *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.
29. Smith, C. J. & Pearse, B. M. F. (1999) *Trends Cell Biol.* **9**, 335–338.
30. Kelly, J. W. (1996) *Curr. Opin. Struct. Biol.* **6**, 11–17.