Template-Directed Assembly of a *de Novo* Designed Protein

Christina L. Brown,[†,‡] Ilhan A. Aksay,[‡,§] Dudley A. Saville,[§] and Michael H. Hecht*[,†,‡]

*Department of Chemistry, Princeton Materials Institute, and Department of Chemical Engineering,
Princeton University, Princeton, New Jersey 08544*

Received March 7, 2002

A number of biological materials owe their unusual structural characteristics and mechanical properties to long-range order induced by the lamination of β-sheet proteins between layers of inorganic mineral.[1] In such composites, both the protein layer and the mineral layer adopt structures different from those they assume in isolation. Interactions between such layers and the ordered structures that result from these interactions enable nature to produce biomaterials that are simultaneously hard, strong, and tough.

With the long-term goal of constructing artificial biomaterials with laminated structures, we developed a biomimetic system using a highly ordered surface to template the assembly of a *de novo* designed β-sheet protein. The surface used is highly ordered pyrolytic graphite (HOPG), which forms individual crystallite domains extending over several microns.[2] The graphite lattice is hexagonal; therefore structures templated by HOPG are expected to show 3-fold symmetry.

The protein used in this study was chosen from a combinatorial library of amino acid sequences designed *de novo* to form β-sheet proteins.[3,4] The library design specified that each protein should contain 6 β-strands, and each β-strand would be 7 residues long, with polar and nonpolar amino acids arranged in an alternating pattern (Figure 1). This polar/nonpolar pattern (binary patterning[3–5]) is consistent with the formation of a β-sheet having all polar side chains projecting from one face, and all nonpolar side chains projecting from the opposing face (Figure 3). Formation of such facial amphiphiles would be favored at an interface between polar and nonpolar phases such as water and graphite.

Combinatorial diversity was incorporated into the library of *de novo* proteins by allowing polar residues to be His, Lys, Asn, Asp, Gln, or Glu; and nonpolar residues to be Leu, Ile, Val, or Phe.[3] These combinatorial sets of amino acids were encoded by libraries of synthetic genes with polar residues specified by the degenerate DNA codon VAS, and nonpolar residues by the degenerate DNA codon NTC (where V = A, G, or C; S = G or C; and N = A, G, C, or T). The proteins in this library have a total length (including six β-strands and intervening turns) of 63 residues. Their design and synthesis has been described.[3]

Previous studies of proteins from this library demonstrated that they self-assemble into (i) amyloid-like fibrils in a homogeneous aqueous environment[3] and (ii) β-sheet monolayers at an air/water interface.[4] Given the tendency of these *de novo* proteins to self-assemble in the absence of an external template, we sought to assess whether the presence of a highly ordered surface (template) would induce structures that recapitulate the order and symmetry of the template.

To assess the propensity of our *de novo* proteins to undergo template-directed assembly,[6] we focused on one protein, 17-6 (Figure 1), from our β-sheet library. When probing the ability of a surface to template protein assembly, it is important to present the template with a protein that is not preassembled. Previous studies demonstrated that protein 17-6 can exist in aqueous solution either as monomers or as self-assembled amyloid-like fibrils.[3,7] Physiological buffers favor assembly, whereas high pH and low salt favor monomers.[7] In the current study, the latter conditions were used to ensure that monomeric protein was presented to the HOPG surface.

A solution of protein 17-6 was deposited onto freshly cleaved HOPG, and the resulting assemblies were imaged with tapping mode atomic force microscopy (AFM). The AFM image shown in Figure 2A demonstrates that the protein assembles into parallel fibers on the HOPG surface. The fibers show three preferred orientations at 120° to each other. A Fourier transform of this image (Figure 2A, inset) confirms the 3-fold symmetry of the structure. This symmetry indicates that the hexagonal lattice of graphite (Figure 2B) directs nucleation of fibers on the HOPG surface. The straightness of the fibers and their persistence length of several microns suggest that the template also influences the addition of protein monomers onto the growing fiber.

To assess whether interactions between the HOPG surface and the protein are responsible for the ordered structure, we compare the current study to two different experiments in which this same protein was imaged on carbon surfaces:

(i) As described in previous work, in buffered saline solutions, protein 17-6 self-assembles into fibrillar structures resembling amyloid.[3,7] Electron microscopy of these preformed structures reveals individual fibrils (not groups of parallel rods) that are curved (not straight), and which occur in random orientations (without 3-fold symmetry). Thus, the fibrillar structures preformed in solution—in the absence of a template—do not form ordered arrays such as those in Figure 2A.

(ii) Deposition of monomeric protein 17-6 onto EM grids coated with amorphous carbon (rather than HOPG) does not yield ordered arrays *(not shown)*. Thus, the mere presence of assembly competent protein is not sufficient to induce structure; an ordered surface is necessary to template the structures shown in Figure 2A.

What is the structure of the protein assembled on the HOPG surface? The pattern of polar and nonpolar amino acids in the sequence of protein 17-6 (Figure 1) is consistent with formation of an amphiphilic β-sheet with the polar side chains projecting up toward water, and the nonpolar side chains pointing down toward the HOPG surface (Figure 3). To confirm that protein 17-6 at a polar/nonpolar interface indeed forms β-structure, we also deposited this protein onto ZnSe, which, like HOPG, is hydrophobic but, unlike HOPG, is transparent in the infrared. Transmission Fourier transform infrared spectroscopy (FTIR) of protein 17-6 on ZnSe

* Corresponding author. E-mail: hecht@princeton.edu.
† Department of Chemistry.
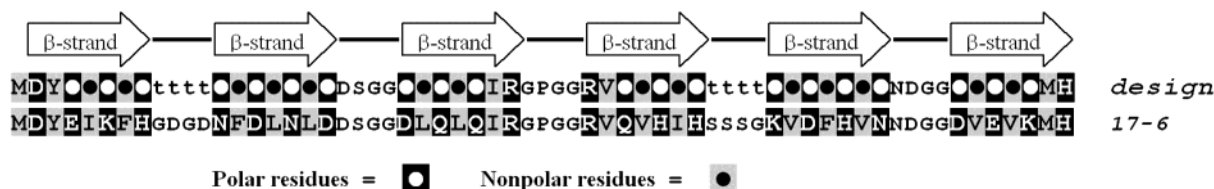‡ Princeton Materials Institute.
§ Department of Chemical Engineering.

Figure 1. Polar/nonpolar ("binary"') patterning of designed β-sheet proteins. β-Strands are shown as arrows. Polar residues are in white font on black background, and nonpolar residues are in black font on gray background. Combinatorial diversity was incorporated in the original library[3] at positions marked O, ●, and t (turn). The amino acid sequence of protein 17-6 is shown in the single letter code.
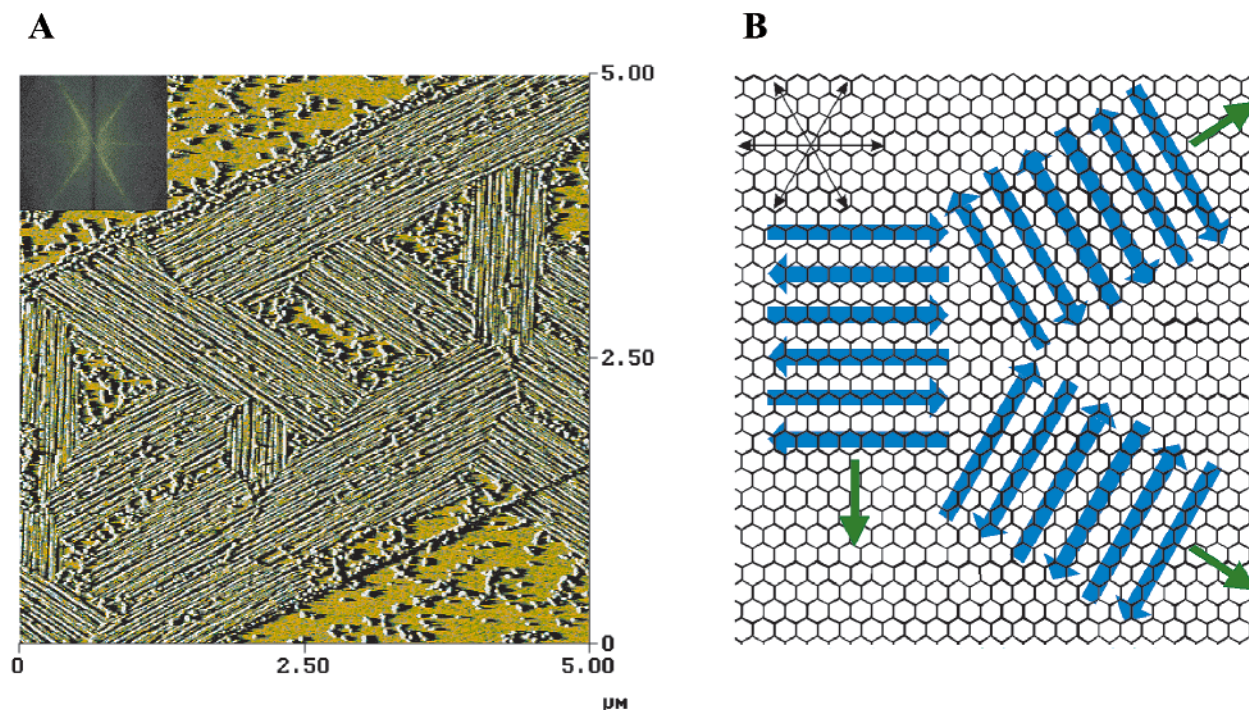

Figure 2. (A) AFM image of protein 17-6 deposited on highly ordered pyrolytic graphite (HOPG). The inset shows a Fourier transform of this image. The 3-fold symmetry is apparent both in the AFM image and in its Fourier transform. *Methods*: Protein was dissolved at pH 11 to break apart any existing aggregates. The sample was then lyophilized and redissolved in pure water, in which it persisted in the monomeric form, as determined by SEC (not shown). Ten microliters of protein at ~300 μg/mL in pure water was deposited onto freshly cleaved grade ZYH pyrolytic graphite monochromator (Advanced Ceramics Corp., Lakewood OH) and allowed to dry slowly in a humidified environment. The adsorbed protein was imaged under ambient conditions by using tapping mode AFM with a Nanoscope IIIa Scanning Probe Microscope from Digital Instruments, with Nanoscope IIIa software version 4.42r4, "TappingMode" Etched Silicon Probe tips, and a TappingMode cell. The globular deposits on the graphite likely consist of nonordered aggregates of the protein. The image shown here was collected in amplitude mode. Data collected in height mode showed the same features. (B) Schematic representation of a 6-stranded β-sheet protein assembled on a HOPG surface. β-Strands are shown as blue arrows. The 3-fold symmetry of the graphite template is recapitulated in the assembly of the protein. The long axis of the fibers (part A) is perpendicular to the β-strands and is indicated with green arrows. The relative orientation of the fibers to the graphite lattice was determined by imaging a sample of fibers and subsequently imaging the graphite lattice underneath the fibers by using contact mode AFM.

shows a maximum absorbance at 1621 cm^{-1} (Figure 4), indicating an antiparallel β-sheet structure.[8]

Assembly of β-structure into fibrillar or rodlike morphologies has been observed in several biological systems. Examples range from the silks of insects[9] to the amyloid fibrils of Alzheimer's disease.[10] In the "cross-β" structure[9b] the β-strands run perpendicular to the fiber axis, and the hydrogen bonds between β-strands are parallel to the fiber axis. These H-bonds can form either *intra*molecularly, between backbone N−H and C=O groups on successive strands of a β-hairpin, or *inter*molecularly, between N−H and C=O groups on the edge strands of neighboring β-sheets. Inter-molecular H-bonding would favor open-ended assembly of many molecules of β-sheet structure into long fibrils or rods (Figure 2).

The dimensions of a cross-β structure can be estimated from the known geometry of β-sheets. In the antiparallel β-structure, the distance per residue is 0.34 nm.[11] The 7-residue β-strands in Figure 3 would therefore have lengths of 2.4 nm. If the turns on either end contribute an additional ~0.5 nm each, then the total width

(along the β-strands) of the flat β-sheet in Figure 3 would be ~3.4 nm. This width can be compared with the dimensions of the fibers in Figure 2A. The apparent width of each fiber is ~20 nm. However, because AFM tips have a nonzero diameter, the true width of a structure is smaller than the apparent width in an AFM image. The actual width can be calculated from the observed width and an estimate of the tip diameter.[12,13] These calculations yield an estimated width of ~4 nm, consistent with the expected width (3.4 nm, see above) for a cross-β structure assembled from β-sheets similar to that shown in Figure 3.

The ordered structure in Figure 2A extends beyond the borders of the image. Thus the area of templated structure can be conservatively estimated as $>2 \times 10^7$ nm^2. By comparison, the area of a single protein molecule in the flat β-sheet shown in Figure 3 would be ~6.5 nm^2. [A Langmuir film of protein 17-6 at an air/water interface had an experimentally measured[4] molecular area of ~6 nm^2.] Thus, the ordered structures in Figure 2A comprise >3 million protein molecules. Since individual crystallites of HOPG
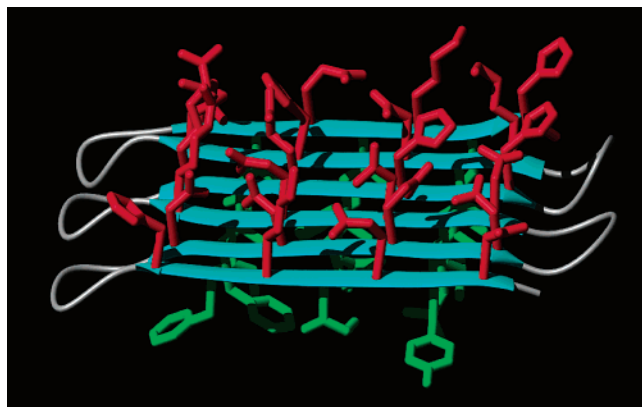
COMMUNICATIONS

Figure 3. The sequence of protein 17-6 modeled as a flat 6-stranded β-sheet. The sheet is amphiphilic with polar residues (red) projecting up, and nonpolar residues (green) projecting down. The figure was constructed with MOLMOL[14] and rendered with POV-ray (www.povray.org).
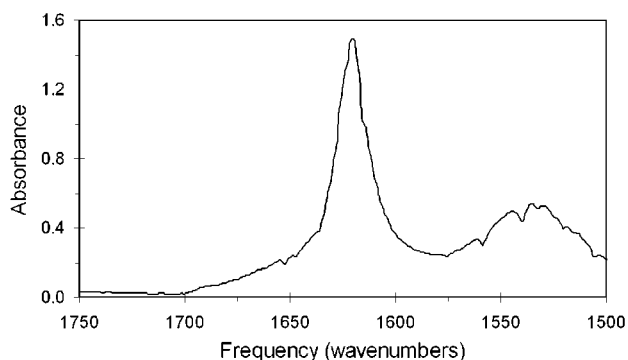

Figure 4. Fourier transform infrared spectrum of protein 17-6 on ZnSe. The peak at 1621 cm^{-1} indicates antiparallel β-structure. *Methods*: A solution of protein 17-6 at 3 mg/mL was heated to 70 °C for 45 min to ensure all the protein was in the monomeric form before deposition. Fifteen microliters was deposited onto preheated ZnSe at 70 °C (Harrick Sci Corp, Ossining NY). The sample was cooled to 50 °C (the approximate midpoint of the thermal denaturation[7]) and held at this temperature for 15 min. Samples were then cooled to room temperature and analyzed with a liquid nitrogen cooled FTIR spectrometer (Bruker Equinox 55).

are typically only several micrometers long,[2] the observed amount of template-induced assembly may in fact be limited by the degree of order in the HOPG itself. Indeed, an image recorded 50 μm away from that shown in Figure 2A showed similarly ordered 3-fold symmetric structures. However, as expected for a structure templated on a different crystallite, the orientations were rotated relative to those in Figure 2A.

The template-directed assembly described here was obtained by using a protein chosen arbitrarily from a combinatorial library of *de novo* sequences. The binary pattern of polar and nonpolar residues shown in Figure 1 can encode an enormous number of different sequences [6 hydrophilic residues at 24 polar sites, and 4 hydrophobic residues at 18 nonpolar sites would yield a theoretical diversity of $6^{24} \times 4^{18} = 3 \times 10^{29}$]. From amidst this diversity, sequence 17-6 was chosen arbitrarily. (It was chosen simply because earlier studies[3,4,7] had shown that it expressed at high levels and was easy to purify.) Neither genetic selections nor high-throughput screens were used to select this particular sequence for template-directed assembly. Therefore, our observation of template-directed assembly on HOPG suggests that binary patterning, if applied to an appropriately designed structural scaffold (e.g. Figure 1), can be used to generate large collections of assembly competent *de novo* proteins.

References

(1) Weiner, S.; Addadi, L. *J. Mater. Chem.* **1997**, *7*, 689−702.
(2) (a) Moore, A. In *Chemistry and Physics of Carbon*; Marcel Dekker Inc.: New York, 1973; Vol. 11, pp 69−187. (b) Ohler, M.; Baruchel, J.; Moore, A. W.; Galez, P.; Freud, A. *Nucl. Instrum. Methods Phys. Res., Sect. B* **1997**, *129*, 257−260
(3) (a) West, M. W.; Wang, W.; Patterson, J.; Mancias, J. D.; Beasley, J. R.; Hecht, M. H. *Proc. Natl Acad. Sci. U.S.A.* **1999**, *96*, 11211−11216. (b) West, M. W. Ph.D. Thesis, 1997, Princeton University
(4) Xu, G.; Wang, W.; Groves, J. T.; Hecht, M. H. *Proc. Natl Acad. Sci. U.S.A.* **2001**, *98*, 3652−3657
(5) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680−1685
(6) (a) Aksay, I. A.; Trau, M.; Manne, S.; Honma, I.; Yao, N.; Zhou, L.; Fenter, P.; Eisenberger, P. M.; Gruner, S. M. *Science* **1996**, *273*, 892−898. (b) McPherson, A.; Shlichta, P. *Science* **1988**, *239*, 385−387. (c) Reviakine, I.; Bergsma-Schutter, W.; Brisson, A. *J. Struct. Biol.* **1998**, *121*, 356−361. (d) Haggarty, L.; Watson, B. A.; Barteau, M. A.; Lenhoff, A. M. *J. Vac. Sci. Technol. B* **1991**, *9*, 1219−1222. (e) Kowalewski, T.; Holtzman, D. M. *Proc. Natl Acad. Sci. U.S.A.* **1999**, *96*, 3688−3693
(7) (a) Wang, W. Ph.D. Thesis, 2001, Princeton University (b) Patterson, J. BA Thesis, 1999, Princeton University (c) Mancias, J. D. BA Thesis, 2000, Princeton University (d) Brown, C. L. BA Thesis, 2001, Princeton University
(8) Wilson, D.; Valluzzi, R.; Kaplan, D. *Biophys. J.* **2000**, *78*, 2690−2701.
(9) (a) Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1−34. (b) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343−358.
(10) (a) Sunde, M.; Blake, C. *Adv. Protein Chem.* **1997**, *50*, 123−159. (b) Serpell, L. C. *Biochim. Biophys. Acta* **2000**, *1502*, 16−30.
(11) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; Freeman: New York, 1993.
(12) Engel, A.; Schoenenberger, C.-A.; Muller, D. J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 279−284.
(13) The radius of curvature of new AFM tips is 5−10 nm, however, during imaging tips are blunted and the radius is increased.
(14) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.

JA0261271