Self-assembled monolayers from a designed combinatorial library of *de novo* β-sheet proteins

Guofeng Xu*, Weixun Wang*, John T. Groves†, and Michael H. Hecht†

Department of Chemistry, Princeton University, Princeton, NJ 08544-1009

Edited by William F. DeGrado, University of Pennsylvania School of Medicine, Philadelphia, PA, and approved January 31, 2001 (received for review August 21, 2000)

A variety of naturally occurring biomaterials owe their unusual structural and mechanical properties to layers of β-sheet proteins laminated between layers of inorganic mineral. To explore the possibility of fabricating novel two-dimensional protein layers, we studied the self-assembly properties of *de novo* proteins from a designed combinatorial library. Each protein in the library has a distinct 63 amino acid sequence, yet they all share an identical binary pattern of polar and nonpolar residues, which was designed to favor the formation of six-stranded amphiphilic β-sheets. Characterization of proteins isolated from the library demonstrates that (*i*) they self assemble into monolayers at an air/water interface; (*ii*) the monolayers are dominated by β-sheet secondary structure, as shown by both circular dichroism and infrared spectroscopies; and (*iii*) the measured areas (500- 600 Å^2) of individual protein molecules in the monolayers match those expected for proteins folded into amphiphilic β-sheets. The finding that similar structures are formed by distinctly different protein sequences suggests that assembly into β-sheet monolayers can be encoded by binary patterning of polar and nonpolar amino acids. Moreover, because the designed binary pattern is compatible with a wide variety of different sequences, it may be possible to fabricate β-sheet monolayers by using combinations of side chains that are explicitly designed to favor particular applications of novel biomaterials.

Proteins play key roles in controlling the self-assembly of biological materials (1–5). In recent years, interest has grown in mimicking biological self assembly (2–11) with peptides or proteins designed *de novo*. This interest is motivated both by a desire to test our understanding of self assembly in natural systems and by the drive to lay foundations for the fabrication of novel protein-based biomaterials.

Self-assembled two-dimensional layers of proteins are of particular interest for the fabrication of biomaterials. Such layers play important roles in biological mineralization by controlling the size, orientation, and morphology of inorganic crystals at the surface of the protein layer (2, 5, 11–14). A well-studied example of protein/inorganic layering occurs in the nacre of mollusk shells ("mother of pearl") (4, 5, 11–17). These laminated structures are composed of alternating layers of a protein-rich matrix and aragonite (a crystal form of calcium carbonate). The proteins in the protein-rich layer are dominated by β-sheet secondary structure (1, 5, 11–14).

To explore the possibility of fabricating *de novo* β-sheet layers, we studied the self-assembly properties of a collection of *de novo* proteins derived from a combinatorial library of sequences designed to favor the formation of amphiphilic β-strands punctuated by reverse turns (18, 19). Each sequence in the library is distinct. However, all sequences are the same length (63 residues), and they all share the identical binary patterning of polar and nonpolar residues. Here we report that despite their different amino acid sequences, proteins derived from this library share the ability to assemble into monolayers at an air/water interface. The monolayers are dominated by β-sheet secondary structure, and their mean molecular areas match those expected for proteins folding into β-sheet structures with one polar and one nonpolar face.

Materials and Methods

Proteins. Construction of the combinatorial library of genes and the expression and purification of the proteins have been described (18, 19). The identity and purity of the purified proteins were verified by SDS/PAGE and electrospray mass spectroscopy. The purified proteins were dialyzed into water at pH 11.0 (adjusted with NaOH) and lyophilized overnight. The dry protein samples were dissolved in hexafluoroisopropanol (HFIP) at a concentration of ≈1 mg/ml. The exact concentration of the protein/HFIP solution was determined by quantitative amino acid analysis (Biotechnology Analytical Facility, Cornell University, Ithaca, NY). For Langmuir–Blodgett studies, the protein/HFIP solution was diluted with chloroform. The volume ratio of HFIP to chloroform was 1:8 for protein 17 and 1:6 for proteins 23 and 69.

Langmuir–Blodgett Studies. Langmuir films at the air/water interface were formed at 22°C by using a custom-built Teflon trough (110 mm × 95 mm) controlled by a commercial KSV 5000 (KSV Chemicals) Langmuir–Blodgett system. In a typical experiment, a solution of $CHCl_3$/HFIP containing the desired protein at a concentration of 0.1 mg/ml was added dropwise through a microsyringe onto the surface of deionized water that had been saturated with CO_2 and had a pH of 5.8. After layering of the film and evaporation of the organic solvents, the surface film was allowed to relax for 10–15 min. Isotherms were then recorded while the film was compressed at a constant rate of 20 mm/min, and surface pressure was measured by Wilhelmy plate. The mean molecular area of each protein was estimated by extrapolating the linear portion of each curve (between 15 mN/m and 25 mN/m) to a surface pressure equal to zero. Such extrapolations are routinely used to estimate the mean molecular area in films that are condensed but not under pressure (20).

Determination of Protein Secondary Structure. Secondary structure was analyzed both by circular dichroism (CD) and transmission infrared (IR) spectroscopy. Films were transferred from the air/water interface onto solid substrates (hydrophilic quartz slides for CD and CaF_2 for IR) by vertical deposition at a constant surface pressure of 25 mN/m.

Transmission Fourier transform IR spectra were obtained on

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HFIP, hexafluoroisopropanol; CD, circular dichroism; IR, transmission infrared; MREs, mean residue ellipticities.

*G.X. and W.W. contributed equally to this work.

†To whom reprint requests should be addressed. E-mail: hecht@princeton.edu or jtgroves@princeton.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.

www.pnas.org/cgi/doi/10.1073/pnas.071400098

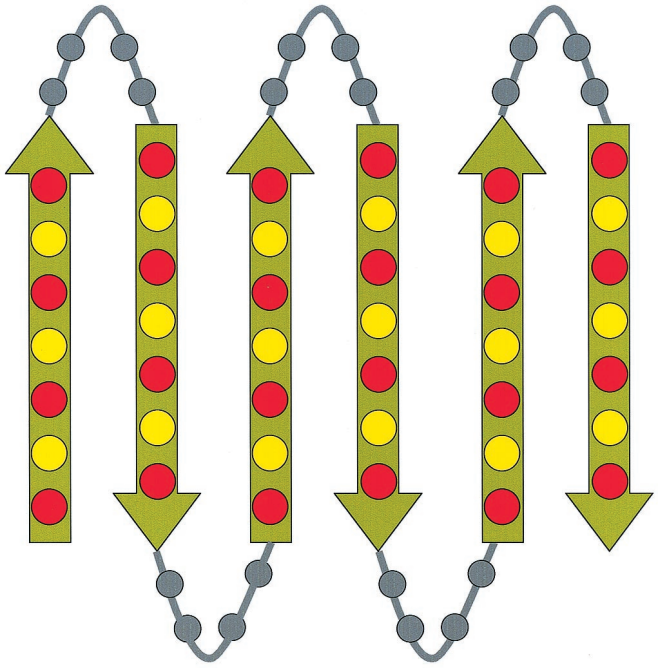

Fig. 1. The designed binary patterning for a combinatorial library of *de novo* proteins. Arrows designate β-strands, and gray loops designate reverse turns. Red circles represent positions that can be occupied by any of the polar amino acids His, Lys, Asn, Asp, Gln, or Glu. Yellow circles represent positions that can be occupied by any of the nonpolar amino acids Leu, Ile, Val, or Phe. Gray circles represent turn residues. (This figure shows 62 rather than 63 residues because the initiator methionine is not included in this schematic representation.)

a Nicolet-730 FT–IR instrument by scanning 512 times at a resolution of 4 cm^{-1} using blank CaF_2 as the background.

CD spectra were obtained for Langmuir films that had been transferred onto quartz slides [Chemglass (Vineland, NJ), 19 × 19 mm]. Before use, these slides were cleaned by immersion in "piranha" solution (30% H_2O_2, 70% H_2SO_4), rinsed in deionized water, and oven-dried. CD spectra were measured at 25°C by using an Aviv Associates (Lakewood, NJ) 62DS CD spectrometer. A blank quartz slide was used as background. The samples responsible for the observed CD spectra contain only two monolayers of protein (one on each side of the quartz slide). Therefore, samples were scanned repeatedly to improve the signal-to-noise ratio. The spectra shown in Fig. 4 represent 100 scans at 1-nm resolution. In some cases, we also attempted to improve the signal-to-noise ratio by dipping the quartz slide multiple times. An example of such an experiment is shown in Fig. 4 for protein 23.

The CD spectra are presented as raw data. Conversion to mean residue ellipticities requires an accurate assessment of the concentration of protein on the quartz slides. Although the concentration of protein in the original Langmuir monolayers was known from quantitative amino acid analysis, uncertainties about the efficiency of transfer to the quartz slides complicate assessment of protein concentration in the CD experiment. [For this reason, CD spectra of protein monolayers transferred onto quartz slides are often reported as raw data (e.g., refs. 21 and 22).] Nonetheless, by assuming that transfer occurs at 100% efficiency after a single dip of the quartz slide, we calculated mean residue ellipticities (MREs) at 217 nm for proteins 17 and 69. The calculated values are $\approx -9{,}000$ deg·cm^2/dmol. Because the actual transfer presumably occurs with less than perfect efficiency, the true MREs are likely to be somewhat more negative. These values demonstrate the monolayers are dominated by β-sheet secondary structure.

Model Building. The model shown in Fig. 6 was constructed using the INSIGHT/DISCOVER package of programs (Molecular Simulations, Waltham, MA). Initially, two seven-residue polyalanine β-strands were placed into the standard conformation for antiparallel β-structure (23). Constrained energy minimization brought the two strands to a separation distance consistent with hydrogen bonding between backbone amides and carbonyls on the adjacent strands. This pair of strands was then replicated twice to generate a total of six antiparallel β-strands. Successive β-strands were then linked by tetraglycine turns. Backbone coordinates for these turns were taken from tetrapeptide turns in natural proteins. For the first, third, and fifth turns, we used the mirror image of the backbone coordinates of residue 19–22 of prealbumin (2PAB). For the second and fourth turns, we used the backbone coordinates of residues 240–243 of penicillopepsin (3APP). After the turns had been joined to the β-strands, the side chains for the entire sequence of protein 17 were appended onto the backbone, and the structure was minimized.

Results

A Designed Combinatorial Library of *de Novo* β-Sheet Proteins. The proteins used in this study are derived from a library of amino acid sequences designed to form six amphiphilic β-strands punctuated by reverse turns (Fig. 1). Each of the six β-strands was designed to be seven residues long with polar (○) and nonpolar (●) amino acids arranged in an alternating pattern

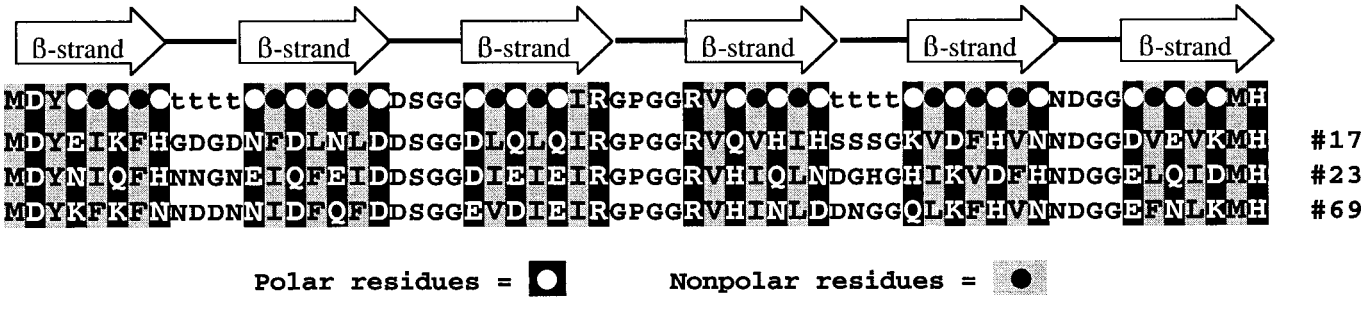

Fig. 2. (Top line) Binary sequence pattern of a designed combinatorial library of *de novo* proteins. The sequences are 63 residues long and are designed to form six β-strands (arrows) punctuated by turns. (Second line) Alternating pattern in the β-strands is indicated with polar residues (○) as white font on black background and nonpolar residues (●) as black font on gray background. Combinatorial diversity is incorporated at positions marked ○, ●, and t (turn). Fixed residues are incorporated at the termini and in some of the turns (18, 19). (Lines 3–5) Amino acid sequences (single-letter code) of three *de novo* proteins from the combinatorial library.

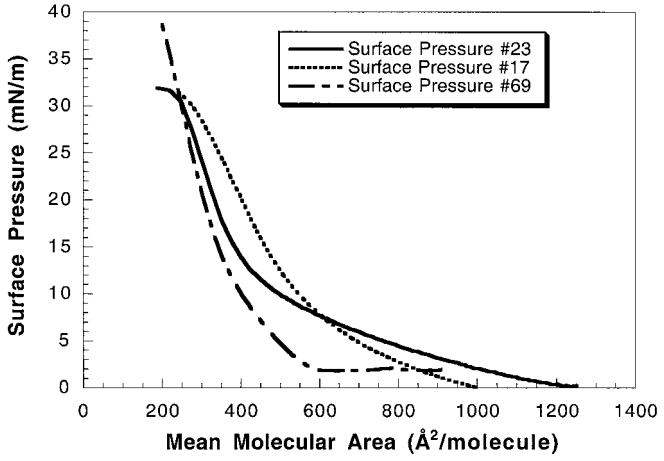

Fig. 3. Isotherms of Langmuir films at the air/water interface.

(○●○●○●○). This polar/nonpolar periodicity matches the structural repeat of an amphiphilic β-strand, with successive side chains pointing up-down-up-down, etc. (23–25). The β-strands were separated by reverse turns designed to be four residues long. By choosing turns with an even number of residues (e.g., ○●○●○●○-t-t-t-t-○●○●○●○), we designed an interruption in the alternating polar/nonpolar periodicity. This interruption would favor short β-strands separated by reverse turns, as shown in Fig. 1, and disfavor alternative structures containing long uninterrupted amphiphilic β-strands running the full length of the sequences. (In contrast, a sequence with an odd number of turn residues, such as ○●○●○●○-t-○●○●○●○, could maintain the alternating polar/nonpolar patterning by placing the putative turn residue on the hydrophobic face of a long uninterrupted amphiphilic β-strand; see *Discussion*.)

Within the constraints of the designed polar/nonpolar patterning, combinatorial diversity was incorporated into the library by allowing polar residues to be either His, Lys, Asn, Asp, Gln, or Glu and nonpolar residues to be either Leu, Ile, Val, or Phe. These combinatorial sets of amino acids were encoded by libraries of synthetic genes in which polar residues are encoded by the degenerate DNA codon NAN and nonpolar residues by the degenerate DNA codon NTN (where N represent a specified mixture of bases). The design and construction of this combinatorial library are described in greater detail in refs. 18 and 19.

A schematic diagram of the design and three representative amino acid sequences are shown in Fig. 2. The designed pattern of alternating polar and nonpolar residues is consistent with the formation of a β-sheet structure in which all of the hydrophilic side chains project from one face of the sheet and all of the hydrophobic side chains project from the opposing face. Formation of these facial amphiphiles would be favored at an interface between polar and nonpolar phases.

Monolayers at the Air/Water Interface. The three sequences shown in Fig. 2 were expressed in *Escherichia coli*, purified, and lyophilized. The lyophilized proteins were dissolved in a volatile organic solvent (hexafluoroisopropanol/chloroform) and then gently layered onto the aqueous surface in a Langmuir–Blodgett trough. All three proteins formed stable Langmuir films at the air/water interface. Plots of the surface pressure as a function of mean molecular area are shown in Fig. 3. The three isotherms are similar but not identical. They differ somewhat from one another, as expected for three distinctly different amino acid sequences. Yet they share several key properties: all three surface films are stable at surface pressures up to 30 mN/m. The films remained stable at surface pressures corresponding to their assembled state (20 ≈ 25 mN/m) for many hours without noticeable loss of protein into the aqueous subphase. From the isotherms in Fig. 3, the mean surface area occupied by each protein molecule was estimated by extrapolating the linear portion of each curve (between 15 and 25 mN/m) to a surface pressure equal to zero. Such extrapolations are routinely used to estimate the surface area per molecule in films that are condensed but not under pressure (20). These extrapolations yielded mean molecular surface areas of 600, 550, and 500 Å^2 for proteins 17, 23, and 69, respectively. The protein surface films were robust and were readily transferred onto solid substrates for further characterization.

β-Sheet Secondary Structure. The secondary structures of the surface films were determined both by CD and IR spectroscopy. To facilitate CD measurements, the films were transferred by vertical deposition onto quartz slides. The CD spectra display a single minimum at ≈217 nm (Fig. 4), demonstrating that the films are dominated by β-sheet secondary structure (26). Because of inherent uncertainties in the concentration of protein on the quartz slides, spectra are presented as raw data rather than MREs. However, by assuming that transfer from the air/water interface to the quartz slide occurred with 100%

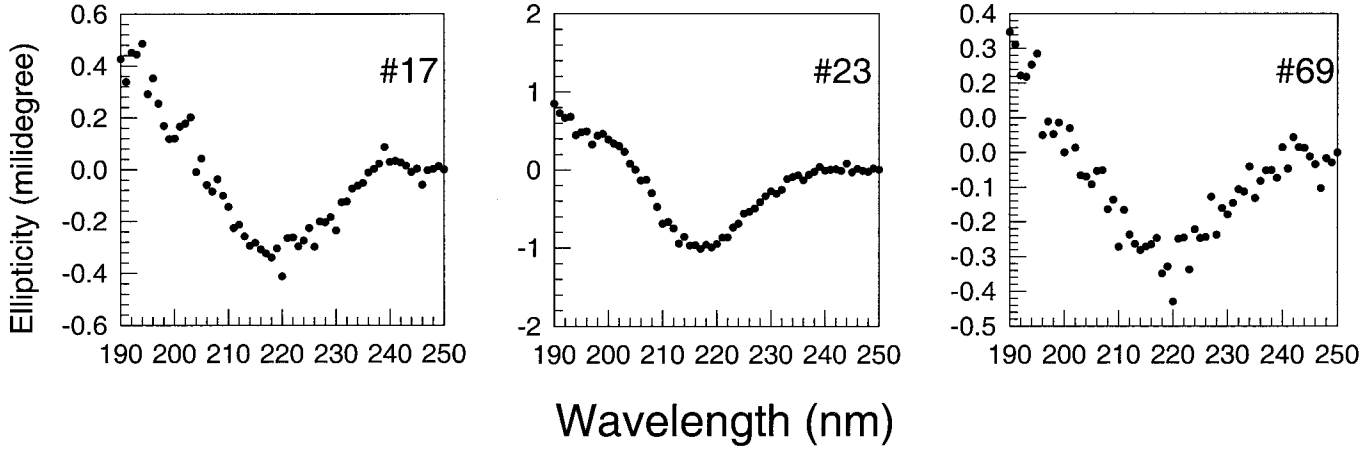

Fig. 4. CD spectra of the Langmuir–Blodgett films after transfer onto quartz slides. Each spectrum represents the average of 100 scans. The single minimum at 217 nm indicates β-sheet conformation. (Spectra for proteins nos. 17 and 69 result from a single dip of the quartz slide into the Langmuir–Blodgett film; the spectrum for protein no. 23 was measured after multiple dips; see text.)

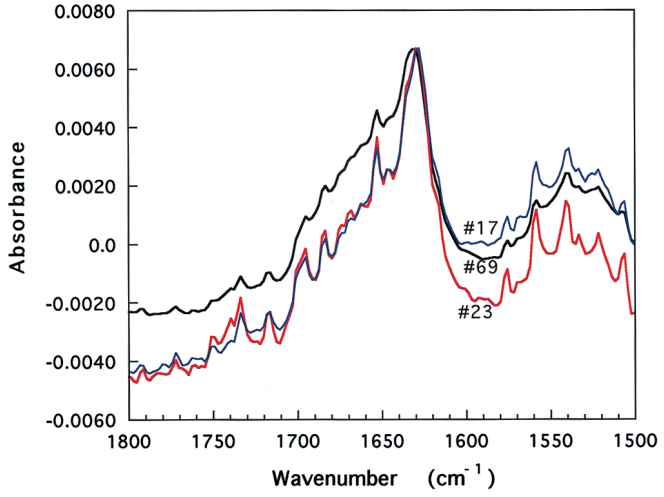

Fig. 5. Transmission Fourier transform-IR spectra of the Langmuir–Blodgett films after transfer onto CaF_2 slides. Each spectrum represents the average of 512 scans. The peak at 1,630 cm^{-1} indicates β-sheet conformation.

efficiency, we were able to estimate mean residue ellipticities at 217 nm. For proteins 17 and 69, these estimated values are $\approx -9{,}000$ deg·cm^2/dmol. (No attempt was made to estimate a MRE for no. 23; see *Materials and Methods*.) Because the actual transfer presumably occurs with less than perfect efficiency, the true MREs are likely to be somewhat more negative. These values demonstrate the monolayers are dominated by β-sheet secondary structure.

The secondary structures of the protein films were also analyzed by IR spectroscopy, as shown in Fig. 5. To facilitate the IR experiments, the films were first transferred by vertical deposition onto CaF_2 slides. For structural studies of proteins, the infrared absorption band in the region of $1{,}600 \approx 1{,}700$ cm^{-1} is assigned to the amide I stretching mode. The position of this peak is very sensitive to conformation and is commonly used as an indicator of protein secondary structure (27). α-Helical and random-coil conformations absorb at 1,650 cm^{-1} and 1,680 cm^{-1}, respectively, whereas the β-sheet conformation produces a strong characteristic absorption centered at 1,630 cm^{-1} (27). The films of all three proteins showed significant absorption peaks at 1,630 cm^{-1}, confirming that they are dominated by β-sheet secondary structure.

Self-Assembled β-Sheet Monolayers. The experimental results presented in Figs. 3–5 demonstrate that at an air/water interface, each of the three *de novo* proteins self assembles into a β-sheet film. To assess the likely structure of these films, we constructed a model of sequence 17 arranged as a flat β-sheet. The model was constructed by placing the seven residues in each of the six β-strands into the standard conformation for antiparallel β-structure. The distances between β-strands are dictated by the preferred hydrogen-bonding distance between the backbone amides and carbonyls of neighboring strands. Turns were built to connect the successive β-strands, and the energy was minimized. The resulting model is presented in Fig. 6. All nonpolar side chains (yellow) project from one face of the β-sheet; this face would orient toward the air phase. In contrast, all polar side chains (red) project from the opposing face of the sheet; this face, along with the polar turns (gray), would be stabilized by solvation in the aqueous subphase. Because all of the different sequences in the designed library share the same binary pattern of polar and nonpolar side chains, this model is generally applicable to all proteins in the collection.

The dimensions of the monolayer in this model are dictated by

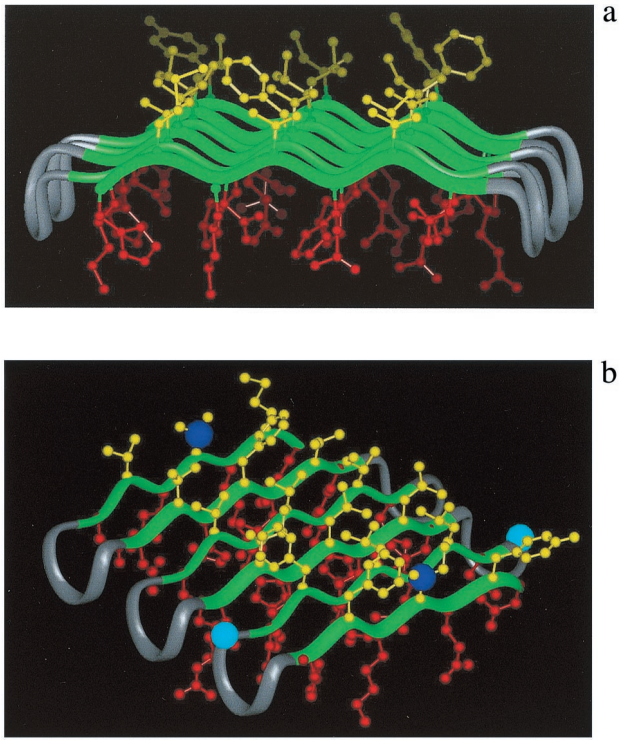

Fig. 6. (*a*) Molecular model of protein 17 at an air/water interface: The model shows six antiparallel β-strands. Each strand contains seven residues—four polar (red) and three nonpolar (yellow). The modeled conformation shows a facial amphiphile with a hydrophobic face (toward air) and a hydrophilic face (toward water), thereby facilitating formation of a β-sheet monolayer at the air/water interface. (*b*) Same model from a different angle. Measurements indicate the length and width of the β-sheet. The length of the β-strands is shown from the α-carbon of the residue (Asp-12) preceding the second strand to the α-carbon of the residue (Asp-20) after the second strand. These α-carbons are shown as turquoise balls. The distance between them is 27.2 Å. The width of the β-sheet is shown as the distance from the β carbon of the middle residue on the first strand (Ile-5) to the β carbon of the middle residue of the last strand (Val-60). These atoms are shown as blue balls. This distance is 23.6 Å. The area of the β-sheet is thus 27.2 Å × 23.6 Å = 642 $Å^2$. This area corresponds to the distance between the turquoise balls multiplied by the distance between the blue balls. This area does not include the polar turns, which are modeled as projecting down into aqueous solvent, nor does it include the area between proteins in the overall structure of the self-assembled monolayer.

the well-known dimensions of β-structure. In antiparallel β-structure, the length of a strand is 3.4 Å per residue (23). Hence, the length of a seven-residue β-strand measured from the α-carbon of the residue preceding the strand to the α-carbon of the residue following the strand (shown as turquoise balls in Fig. 6*B*) would be 8 × 3.4 Å = 27.2 Å. The distance across a β-sheet is dictated by hydrogen bonding between the backbone NH and CO groups of adjacent β-strands. In antiparallel β-structure, these H-bonds cause successive strands to be 4.72 Å apart (28); in the minimized model structure shown in Fig. 6, the adjacent β-strands are 4.72 Å apart. Consequently, the distance from the β-carbon of Ile-5 in the first β-strand to the β-carbon of Val-60 in the sixth β-strand (shown as blue balls in Fig. 6*B*) is 5 × 4.72 Å = 23.6 Å. The area of the sheet is thus 27.2 Å × 23.6 Å = 642 $Å^2$. (In Fig. 6*B*, this area corresponds to the distance between the turquoise balls multiplied by the distance between the blue balls.) Thus, the molecular area (642 $Å^2$) of the monolayer model presented in Fig. 6 is in reasonable agreement with the areas (500–600 $Å^2$) estimated from the Langmuir–Blodgett experiments in Fig. 3.

Discussion

For *de novo* proteins to play a significant role in the development of novel biomaterials, two goals must be achieved. First, methods must be developed for the large-scale production of many different *de novo* sequences. Second, these novel sequences must fold and self assemble into predictable structures with desired properties. The first of these goals—production of diverse sequences in large quantities—is readily achieved by constructing combinatorial libraries of synthetic genes and expressing them in bacteria. Achieving the second goal—assembly into predictable structures—is more subtle. Libraries of random sequences rarely yield proteins capable of self assembly into ordered structures. To produce collections of *de novo* proteins that are capable of self assembly, it is essential that combinatorial methodologies be guided by rational design. The design features that must be imposed on the library depend on which self-assembled structure is being targeted. If the desired structure is a facial amphiphile composed of flat β-sheets, then the sequences must be chosen from libraries explicitly designed to form amphiphilic β-strands.

The three sequences shown in Fig. 2 were chosen arbitrarily from a library designed to form amphiphilic β-strands (18, 19). All three folded into amphiphilic β-sheet structures at the air/water interface. The areas (500–600 Å^2) of the individual protein molecules in the Langmuir–Blodgett experiments (Fig. 3) are close to those expected for proteins folded into amphiphilic β-sheets (Fig. 6). Because each protein in the library has a different amino acid sequence, the results presented in this study suggest that the designed features shared by all members of the library—i.e., the binary patterning shown in Figs. 1 and 2—are sufficient to predispose these sequences to self assemble into β-sheet monolayers at an air/water interface.

The detailed structures of our monolayers cannot be known in the absence of high-resolution structural studies. Nonetheless, some structures, such as α-helical monolayers, are not consistent with the experimental data. Even among the various possible β structures, some are more likely than others. The structure modeled in Fig. 6 is consistent both with the sequence design (Figs. 1 and 2) and with the experimental results demonstrating β-sheet secondary structure (Figs. 4 and 5) and molecular areas in the range of 500–600 Å^2. In contrast, several alternative β-sheet topologies would be unlikely for the sequences in our collection. For example, a structure composed of two long amphiphilic β-strands separated by a single hairpin (rather than six short β-strands, as modeled in Fig. 6) was explicitly disfavored by incorporating features of "negative design" into our sequences (18, 19). Specifically, we chose interstrand linkers of four residues to ensure that the polar/nonpolar periodicity could not "run through" a desired turn sequence. This is diagrammed in Fig. 7, which shows that a long continuous amphiphilic β-strand is not possible for binary patterned β-strand sequences with an even number (e.g., four) of linker residues between successive β-strands. As shown in Fig. 7, if uninterrupted β-strands did form, their polar and nonpolar faces would be "flipped" by the four-residue linker sequence. Hence, long uninterrupted β-strands could not form facial amphiphiles at an air/water interface. The six-stranded structure shown in Fig. 6, however, can readily form a facial amphiphile and therefore represents a more reasonable model for our *de novo* β-sheet monolayers.

The assembly of the binary code proteins into amphiphilic monolayers is induced by the presence of a polar/nonpolar interface at the air/water boundary. In environments lacking a polar/nonpolar interface, these same protein sequences do not form monolayers but assemble into entirely different structures. For example, when dissolved in a homogeneous aqueous solution, these same proteins assemble into long fibrils (18), resembling the

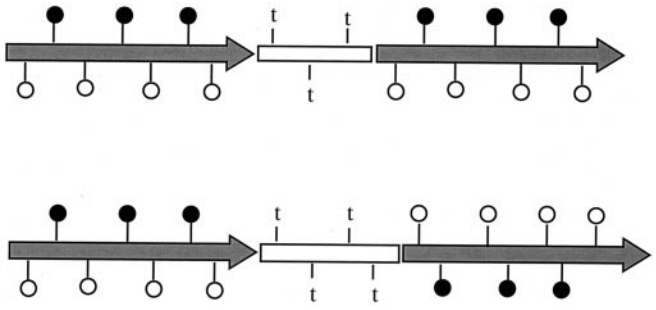

Fig. 7. Comparison of possible β-structures. Each β-strand is shown as an arrow and is composed of four polar (○) and three nonpolar (●) residues arranged in an alternating pattern (○●○●○●○). (*Upper*) If two successive β-strands were joined by an odd number of turn residues (three "t"s), then failure to form the desired hairpin could produce a continuous double-length amphiphilic β-strand with polar residues (○) pointing down toward the aqueous phase and nonpolar residues (●) pointing up toward the air. (*Lower*) Double-length amphiphilic β-strands are disfavored by "negative design." Incorporating an even number of turn residues (four "t"s) between successive β-strands disrupts the up–down nonpolar/polar patterning. Therefore, if the four turn residues failed to form the desired turn (as depicted in Figs. 1 and 6) but instead continued as β-structure, then the resulting polar/nonpolar patterning would be "flipped", and the double-length β-strand would *not* form a continuous amphiphile. Hence, for sequences in this designed binary code library (see Figs. 1 and 2), structures containing double-length β-strands would be disfavored at an air/water interface.

β-amyloid structures found in neurodegenerative diseases (29). In the polar aqueous environment, a nonpolar phase is not available, and the amphiphilic β-strands of the *de novo* proteins must bury their hydrophobic side chains in the interior of the fibril. Alternatively, when the *de novo* sequences are dissolved in HFIP, they form α-helical structures (data not shown). Fluorinated alcohols, such as triflouroenthanol and HFIP, are known to induce α-helical structure (23). In HFIP, which presents neither a polar/nonpolar interface nor a lipophobic environment, the drive to remove nonpolar side chains from solvent is not a dominant effect and is overcome by the solvent-induced α-helicity. When these same HFIP solutions are diluted with chloroform and then layered onto the air/water interface, the proteins undergo a structural transition and self assemble into amphiphilic β-sheet monolayers (Figs. 3–6). Thus the air/water interface clearly plays a vital role in dictating this structure.

Our finding that alternating polar/nonpolar sequences form β-sheet monolayers at the air/water interface is consistent with previous observations by using synthetic peptides. For example, DeGrado and Lear showed that a seven-residue peptide composed of alternating leucines and lysines formed Langmuir films with β-sheet structure at the air/water interface (30). A similar result was reported by Taylor for a 13-residue peptide composed of alternating valine and ornithine residues (21). More recently, Tirrell and coworkers showed that periodic peptides with the alternating sequence Pro-Glu-(Phe-Glu)$_n$-Pro form ordered β-sheet films at the air/water interface (31). Finally, Kelly and coworkers designed a peptidomimetic composed of two amphiphilic hexapeptide β-strands linked together by an artificial β-turn (22). This molecule also formed β-sheet monolayers at the air/water interface.

When reviewing these earlier studies, we found no consistent value for the mean molecular area per residue for peptides in β-sheet monolayers. For individual β-strand peptides (i.e., without hairpins or β-turns), the reported values range from 13 Å^2/residue for the 13-residue Val/Orn peptide described by Taylor (21) to 20 Å^2/residue for the 7-residue Leu/Lys peptide reported by DeGrado and Lear (30). For more complex peptides, in which a turn is designed to link successive β-strands,

there is also no consensus on the mean area expected per residue. Moreover, for linked peptides, it is not clear how much the turn between successive β-strands contributes to the observed area/residue. This may depend on whether the turn is predominantly hydrophobic [as in the peptidomimetic reported by Kelly and coworkers (22)], or hydrophilic, as in our proteins. Presumably hydrophobic turns would sit at the air/water interface, whereas hydrophilic turns might become solubilized by the aqueous phase, as shown in Fig. 6. The observed molecular areas of our binary code proteins ranged from 500 to 600 Å^2. If we assume that the polar turns contribute little to this area and consider only the 42 residues in the 6 β-strands, this would correspond to 12–14 Å^2 per residue. This value is smaller than that reported by Lear and DeGrado (30) but similar to the value reported by Taylor (21). Moreover, as described above, this experimental value is similar to that expected for the idealized β-sheet modeled in Fig. 6.

Our results with a combinatorial library of *de novo* 63 residue proteins demonstrate that the tendency of alternating patterns of polar and nonpolar residues to promote the formation of β-sheet monolayers is not limited to short peptides or repeating silk-like polymers. The three proteins used in this study (*i*) are comparable in length to domains of natural proteins; (*ii*) are composed of heterogeneous protein-like sequences, rather than polymer-like repeats; and (*iii*) differ significantly from one another.

The results presented herein demonstrate that members of a combinatorial library can have significantly different primary sequences (Fig. 2) and nonetheless self assemble into similar structures. Although we cannot be certain that all sequences patterned according to Figs. 1 and 2 will indeed form β-sheet monolayers, our findings demonstrate that this designed binary pattern is robust and can tolerate many possible combinations of polar and nonpolar side chains. The availability of this robust structural scaffold may facilitate design of novel β-sheet monolayers in which the polar and nonpolar faces might be designed explicitly to display combinations of side chains tailored for specific properties in the construction of novel biomaterials.

We gratefully acknowledge support from the National Science Foundation Materials Research Science and Engineering Center (DMR98–09483) and the Biological Sciences Directorate of the Army Research Office (M.H.H.).

1. Lowenstam, H. A. & Weiner, S. (1989) *On Biomineralization* (Oxford Univ. Press, Oxford, U.K.).
2. Weiner, S. & Addadi, L. (1997) *J. Mater. Chem.* **7,** 689–702.
3. Mann, S., Webb, J. & Williams, J. P. (1989) *Biomineralization: Chemical and Biological Perspectives* (VCH, New York).
4. Sarikaya, M. & Aksay, I. A. (1995) *Biomimetics: Design and Processing of Materials* (AIP Press, Woodbury, NY).
5. Addadi, L. & Weiner, S. (1992) *Angew. Chem. Int. Ed. Engl.* **31,** 153–169.
6. Krejchi, M. T., Atkins, E. D. T., Waddon, A. J., Fournier, M. J., Mason, T. L. & Tirrell, D. A. (1994) *Science* **265,** 1427–1432.
7. Aksay, I. A., Trau, M., Manne, S., Honma, I., Yao, N., Zhou, L., Fenter, P., Eisenberger, P. M. & Gruner, S. M. (1996) *Science* **273,** 892–898.
8. Xu, G., Yao, N., Aksay, I. A. & Groves, J. T. (1998) *J. Am. Chem. Soc.* **120,** 11977–11981.
9. Zhang, S., Yan, L., Altman, M., Lassle, M., Nugent, H., Frankel, F., Lauffenburger, D. A., Whitesides, G. M. & Rich, A. (1999) *Biomaterials* **20,** 1213–1220.
10. Cha, J. N., Stucky, G. D., Morse, D. E. & Deming, T. J. (2000) *Nature (London)* **403,** 289–292.
11. Heuer, A. H., Fink, D. J., Laraia, V. J., Arias, J. L., Calvert, P. D., Kendall, K., Messing, G. L., Blackwell, J., Rieke, P. C., Thompson, D. H., *et al.* (1992) *Science* **255,** 1098–1105.
12. Addadi, L., Moradian, J., Shay, E., Maroudas, N. G. & Weiner, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2732–2736.
13. Weiner, S. & Traub, W. (1980) *FEBS Lett.* **111,** 311–316.
14. Weiner, S. & Traub, W. (1984) *Philos. Trans. R. Soc. London B* **304,** 425–434.
15. Addadi, L. & Weiner, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4110–4114.
16. Falini, G., Albeck, S., Weiner, S. & Addadi, L. (1996) *Science* **271,** 67–69.
17. Belcher, A. M., Wu, X. H., Christensen, R. J., Hansma, P. K., Stucky, G. D. & Morse, D. E. (1996) *Nature (London)* **381,** 56–58.
18. West, M. W., Wang, W., Patterson, J., Mancias, J. D., Beasley, J. R. & Hecht, M. H. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11211–11216.
19. West, M. W. (1997) Ph.D. thesis (Princeton Univ., Princeton, NJ).
20. Maget-Dana, R. (1999) *Biochim. Biophys. Acta* **1462,** 109–140.
21. Taylor, J. W. (1990) *Biochemistry* **29,** 5364–5373.
22. Bekele, H., Fendler, J. H. & Kelly, J. W. (1999) *J. Am. Chem. Soc.* **121,** 7266–7267.
23. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties* (Freeman, New York), 2nd Ed.
24. Xiong, H., Buckwalter, B. L., Shieh, H. M. & Hecht, M. H. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6349–6353.
25. West, M. W. & Hecht, M. H. (1995) *Protein Sci.* **4,** 2032–2039.
26. Greenfield, N. & Fasman, G. D. (1969) *Biochemistry* **8,** 4108–4116.
27. Krimm, S & Bandikar, J. (1986) *Adv. Protein Chem.* **38,** 181–364.
28. Lotz, B. & Keith, H. D. (1971) *J. Mol. Biol.* **61,** 201–215.
29. Sunde, M. & Blake, C. (1997) *Adv. Protein Chem.* **50,** 123–159.
30. DeGrado, W. F. & Lear, J. D. (1985) *J. Am. Chem. Soc.* **107,** 7684–7689.
31. Rapaport, H., Kjaer, K., Jensen, T. R., Leiserowitz, L. & Tirrell, D. A. (2000) *J. Am. Chem. Soc*. **122,** 12523–12529.