REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE July 17, 2002	3. REPORT TYPE AND DATES COVERED Final: May 1998 – May 2002 (Includes 1 year no-cost extension)

4. TITLE AND SUBTITLE
Design of De Novo Beta-Sheet Proteins

5. FUNDING NUMBERS
DAAG55-98-1-0084

6. AUTHOR(S)
Michael H. Hecht, Associate Prof. of Chemistry

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Princeton University, Princeton, NJ 08544

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
36604.3-LS

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

20030519 061

13. ABSTRACT (Maximum 200 words)

The goal of the proposed research was to design combinatorial libraries of *de novo* β-sheet proteins. The design of these libraries was based on a 'binary code' strategy, which specifies the sequence locations of polar and nonpolar amino acids, but allows the exact identities of the amino acid side chains to be varied combinatorially. The binary code strategy is made possible by the organization of the genetic code, which uses NAN codons to encode polar amino acids and NTN codons to encode nonpolar amino acids (N denotes a mixture of the DNA bases A, G, C & T.)

The initial research proposal focused on designing libraries of soluble monomeric β-sheet proteins. In the intervening years we have achieved this goal. We have also achieved several additional goals not outlined in the original proposal. We have constructed several collections of *de novo* β-sheet proteins. Among these we have proteins that

(1) fold intramolecularly as monomers in aqueous solution
(2) self-assemble into amyloid-like fibrils
(3) assemble into monolayers at an air/water interface
(4) form ordered structures templated by inorganic surfaces.

Projects 2 and 3 were published during the earlier years of this project. [West et al. (1999) *PNAS* 96, 11211; Broome & Hecht (2001) *J. Mol. Biol.* 296,961; and Xu et al. (2001) *PNAS* 98, 3652.] Project 1 was published recently [Wang & Hecht (2002) *PNAS* 99, 2760]. Project 4 was also published recently [Brown, Aksay, Saville & Hecht (2002) *J. Am. Chem. Soc.* 124, 6846.]

14. SUBJECT TERMS	15. NUMBER OF PAGES
Protein design, β-sheet proteins, Binary code, Combinatorial Libraries, Polar/nonpolar patterning, Self-assembly, Amyloid fibrils, Biomaterials	6
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

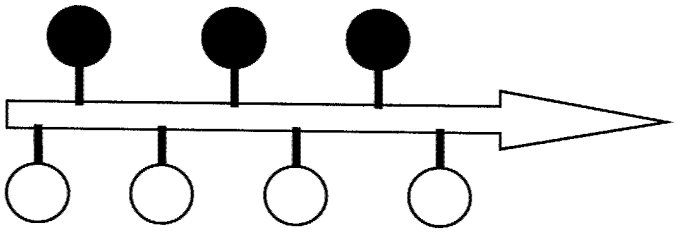
NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

Design of De Novo Beta-Sheet Proteins (Final Technical Report)

We have designed and constructed several combinatorial libraries of *de novo* β-sheet proteins. Design of the amino acid sequences of these libraries is based on the binary patterning of polar and non-polar amino acids. As shown in figure 1, alternating sequences of polar and nonpolar amino acids are consistent with the formation of amphiphilic beta strands. Figures 2 and 3 show a schematic of the designed template that we used to construct our libraries of *de novo* β-sheet proteins. All sequences in our libraries were designed to share the identical pattern of polar and nonpolar residues. However, the precise identities of these side chains were not constrained and were varied combinatorially.

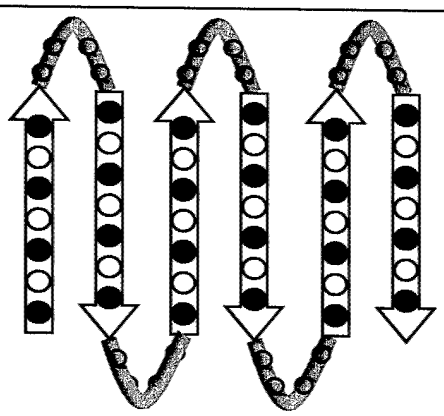
Figure 1. Alternating pattern of polar and nonpolar amino acids specifies an amphiphilic β strand.



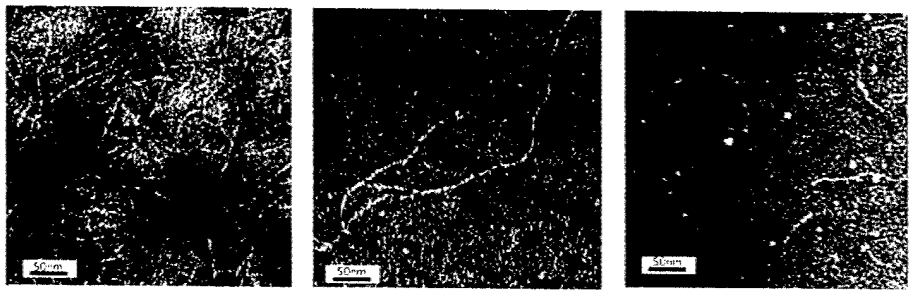
O●O●O●O-turn-O●O●O●O-turn-O●O●O●O-turn-O●O●O●O-turn-O●O●O●O-turn-O●O●O●O

Figure 2: Pattern of polar (O) and nonpolar (●) amino acid residues in a library of protein sequences designed to contain 6 β-strands, punctuated by 5 turns. Combinatorial diversity was generated using the degenerate DNA codon NTN to encode the nonpolar amino acids Met, Leu, Ile, Val or Phe; and NAN to encode the polar amino acids Lys, His, Glu, Gln, Asp or Asn. (N is a mixture of the 4 DNA bases).

Figure 3. Designed binary patterning for a combinatorial library of *de novo* β-sheet 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.



We have expressed and purified several proteins from our designed libraries. Biophysical characterization of these proteins demonstrated that in aqueous solution they self-assemble into fibrils resembling the amyloid fibrils found in several neurodegenerative diseases such as Alzheimer's and the Prion diseases. These fibrils are visible by electron microscopy (shown in figure 4). Circular dichroism spectroscopy demonstrates that the *de novo* fibrils, like natural amyloid fibrils, are composed of β-sheet secondary structure. Moreover, they bind the diagnostic dye, Congo Red. Thus, binary patterning of polar and nonpolar residues arranged with the appropriate periodicity can direct protein sequences to form fibrils resembling amyloid. The model amyloid fibrils assemble and disassemble reversibly, providing a tractable system both for basic studies into the mechanisms of fibril assembly and for the development of molecular therapies that interfere with this assembly. This work is summarized in a publication (West et al. (1999) De Novo Amyloid Proteins From Designed Combinatorial Libraries. *Proc. Nat. Acad. Sci* 96, 11211-11216).



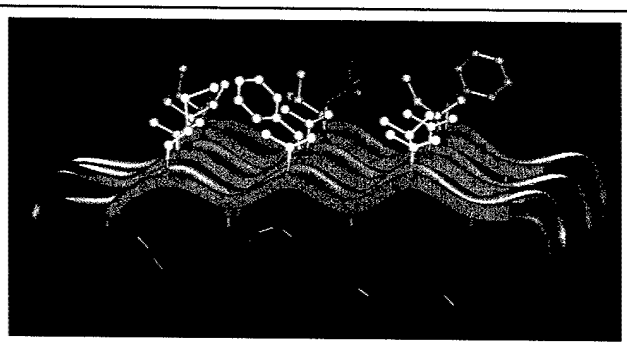
protein #4 protein #17 protein #23

Figure 4: Electron microscope image of amyloid-like fibrils formed by de novo designed beta sheet proteins.

The finding that our designed sequences assembled into amyloid-like fibrils, prompted us to probe the distribution of binary (polar/nonpolar) alternating patterns in the sequences of natural proteins. Analysis of a database of natural protein sequences for all possible patterns of polar and nonpolar amino acids revealed that alternating patterns (e.g. PNPNPNP) occur significantly less often than other patterns with similar compositions. The under-representation of alternating binary patterns in natural protein sequences, coupled with our observation that such patterns promote amyloid-like structures in *de novo* proteins, suggests that sequences of alternating polar and nonpolar amino acids are inherently amyloidogenic and consequently are disfavored by evolutionary selection. This work is summarized in a publication (Broome & Hecht (2000) Nature Disfavors Sequences of Alternating Polar and Nonpolar Amino Acids: Implications for Amyloidogenesis. *J. Mol. Biol.* 296, 961-968).

The formation of fibrils can be considered 1-dimensional self-assembly. In an effort to explore the potentialities of our libraries of *de novo* proteins for the fabrication of novel biomaterials we next probed the ability of our de novo proteins to self-assemble into 2-dimensioal arrays at an air/water interface. 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 square Angstroms) 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, since the designed binary pattern is compatible with a wide variety of different sequences, it may be possible to fabricate β-sheet monolayers using combinations of side chains that are explicitly designed to favor particular applications of novel biomaterials. A model of one of these proteins assembled into an amphiphilic monolayer is shown in figure 5. This work is summarized in a publication (Xu, Wang, Groves, Hecht (2001) Self-Assembled Monolayers from a Designed Combinatorial Library of *De Novo* β-sheet Proteins – ***Proc. Natl Acad. Sci.*** 98, 3652-3657).

Figure 5: Molecular model of a de novo β-sheet protein at an air/water interface: The model shows six anti-parallel 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 (towards air) and a hydrophilic face (towards water), thereby facilitating formation of a β-sheet monolayer at the air/water interface.



Recently, we have used our libraries of *de novo* beta-sheet proteins to construct ordered structures for the fabrication of multi-layered biomaterials analogous to those found in marine shells (e.g. abalone). To induce ordered structures, we have layered the proteins onto an ordered surface (e.g. graphite). The goal of this project is to use the inherent molecular order of the graphite surface to template assembly of our proteins into organized structures that are aligned in directions specified by the underlying graphite surface. The resulting structures were analyzed by atomic force microscopy (AFM). As shown below in figures 6 and 7, the graphite surface templates the proteins to organize into structures that maintain order at the scale of several microns. Given that the individual proteins are only ~30 Angstroms long, this suggest that the surface can template the assembly of structures comprising several million protein molecules.

This work is summarized in a recent publication [Brown CL, Aksay IA, Saville DA, Hecht MH (2002) Template-Directed Assembly of a *De Novo* Designed Protein. - ***J. Am. Chem. Soc.*** 124, 6846].

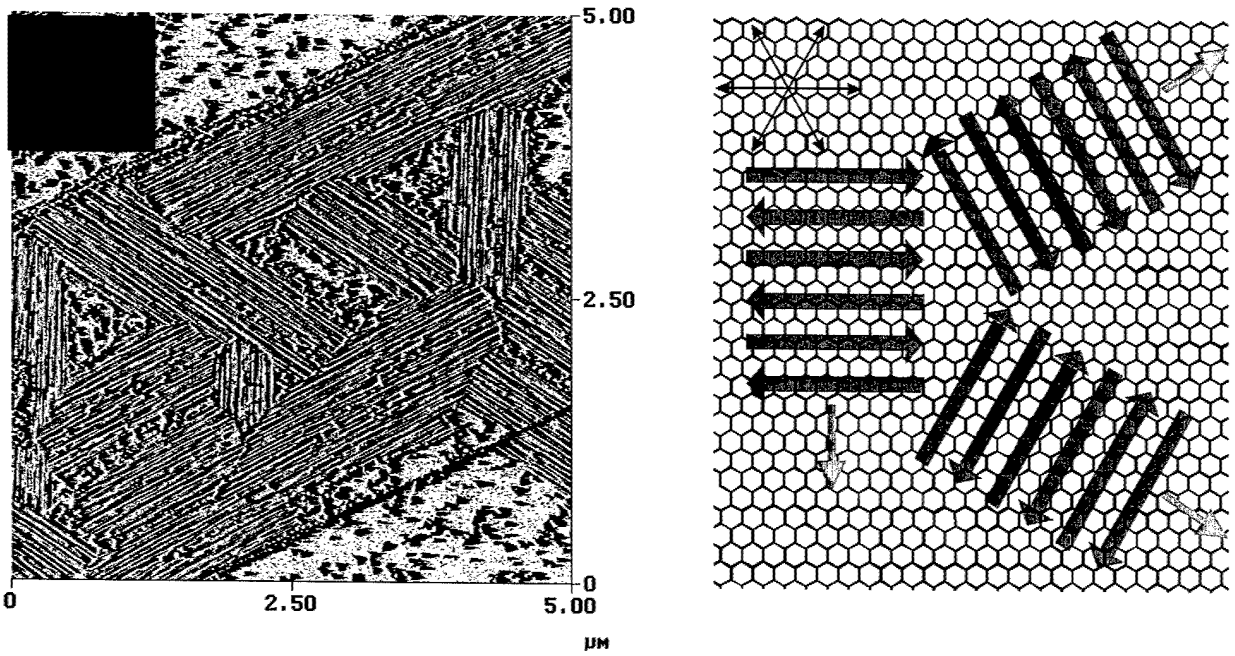


Figure 6: **(A)** AFM image of protein 17-6 deposited on highly ordered pyrolytic graphite (HOPG). Inset shows a Fourier transform of this image. The 3-fold symmetry is apparent both in the AFM image and in its Fourier transform. *Methods:* Protein was dissolved at pH 11 to break apart any existing aggregates. The sample was then lyophilized and re-dissolved in pure water, in which it persisted in the monomeric form, as determined by SEC (*not shown*). 10μL of protein at ~300μg/mL in pure water were deposited onto freshly cleaved grade ZYH pyrolytic graphite and allowed to dry slowly in a humidified environment. The adsorbed protein was imaged under ambient conditions 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 non-ordered 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 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 using contact mode AFM.

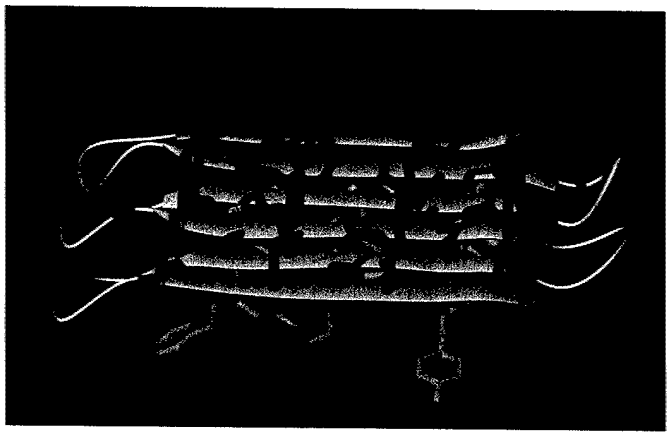


Figure 7: The sequence of protein 17-6 modeled as a flat 6-stranded β-sheet. The sheet is amphiphilic with polar residues (red) projecting up, and nonpolar residues (green) projecting down. The blue β-strands in this figure correspond to the blue arrows in figure 6B.

Returning to the original goal of the proposed research, we have recently produced monomeric β-sheet proteins. Our initial libraries (see figure 2) were designed to encode proteins containing six amphiphilic beta strands separated by reverse turns. Each beta strand was designed to be seven residues long, with polar (O) and nonpolar(●) amino acids arranged with an alternating periodicity (O●O●O●O). The initial design specified the identical polar/nonpolar pattern for all of the beta strands; no strand was explicitly designated to form the edges of the resulting β-sheets. With all β-strands preferring to occupy interior (as opposed to edge) locations, intermolecular oligomerization was favored, and the proteins assembled into amyloid-like fibrils. To assess whether explicit design of edge-favoring strands might tip the balance in favor of monomeric β-sheet proteins, we redesigned the first and/or last β-strands of several sequences from the original library. In the redesigned β-strands, the binary pattern is changed from O●O●O●O to O●OKO●O (K denotes lysine). The presence of a lysine on the nonpolar face of a β-strand should disfavor fibrillar structures because such structures would bury an uncompensated charge. The nonpolar→lysine mutations, therefore, would be expected to favor monomeric structures in which the O●OKO●O sequences form edge with the charged lysine side chain accessible to solvent. To test this hypothesis, we constructed several 2nd generation sequences in which the central nonpolar residue of either the N-terminal β-strands, the C-terminal β-strand (or both) is changed to lysine. The strategy is shown schematically in figure 8. Characterization of the redesigned proteins shows that they indeed form monomeric β-sheet proteins. (Wang W. & Hecht MH (2002) Rationally designed mutations convert de novo amyloid-like fibrils into soluble monomeric b-sheet proteins. ***Proc. Natl Acad. Sci.(USA)*** 99, 2760-2765)

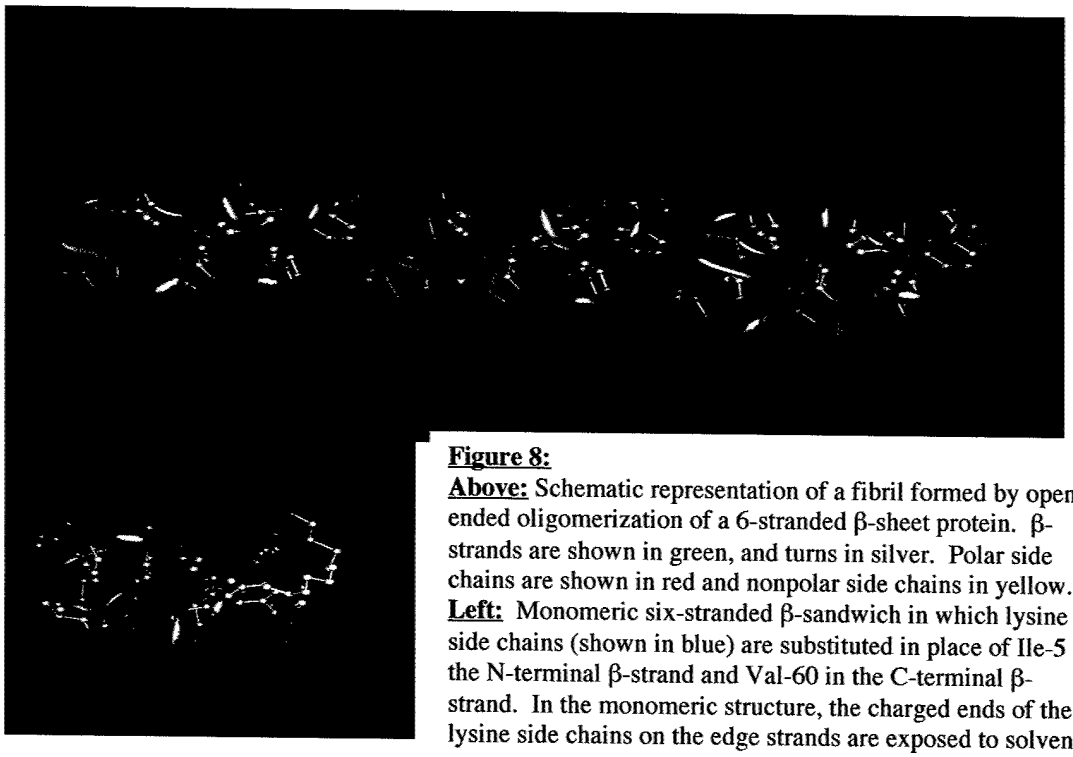


Figure 8:
Above: Schematic representation of a fibril formed by open-ended oligomerization of a 6-stranded β-sheet protein. β-strands are shown in green, and turns in silver. Polar side chains are shown in red and nonpolar side chains in yellow.
Left: Monomeric six-stranded β-sandwich in which lysine side chains (shown in blue) are substituted in place of Ile-5 in the N-terminal β-strand and Val-60 in the C-terminal β-strand. In the monomeric structure, the charged ends of the lysine side chains on the edge strands are exposed to solvent.

List of Publications

West MW, Wang W, Patterson J, Mancias JD, Beasley JR & Hecht MH (1999) De Novo Amyloid Proteins From Designed Combinatorial Libraries. *Proc.Natl Acad. Sci.* 96, 11211-11216.

Broome BM & Hecht MH (2000) Nature Disfavors Sequences of Alternating Polar and Nonpolar Amino Acids: Implications for Amyloidogenesis. *J. Molecular Biology* 296 (4), 961-968.

Xu, G, Wang W, Groves JT, Hecht MH (2001) Self-Assembled Monolayers from a Designed Combinatorial Library of *De Novo* β-sheet Proteins – *Proc. Natl Acad. Sci.(* 98, 3652-3657.

Hecht MH, West MW, Patterson J, Mancias JD, Beasley JR, Broome BM & Wang W. (2001) Designed Combinatorial Libraries of Novel Amyloid-like Proteins. Pages 127-138 in ***Self-assembling Peptide Systems in Biology, Medicine and Engineering,*** (edited by A. Aggeli, N. Boden, S Zhang) Kluwer Academic Publishers, The Netherlands.

Wang W, Hecht MH (2002) Rationally designed mutations convert de novo amyloid-like fibrils into soluble monomeric β-sheet proteins. ***Proc. Natl Acad. Sci.(USA)*** 99, 2760-2765.

Brown CL, Aksay IA, Saville DA, Hecht MH (2002) Template-Directed Assembly of a *De Novo* Designed Protein. - ***J. Am. Chem. Soc.*** 124, 6846-6846.

Scientific Personnel supported by this project:

Michael H. Hecht - principal Investigator: Partial summer salary

Weixun Wang - Graduate student
Dave Moffet - Graduate student
Yinan Wei - Graduate student

Bede Broome – Undergraduate Student
Christina Brown - Undergraduate Student