Chapter 1

DESIGN PRINCIPLES FOR SELF-ASSEMBLING DEVICES FROM MACROMOLECULES

F. Raymond Salemme
Imiplex Corp, Yardley PA

Abstract: Proteins are the natural building blocks for functional, self-assembling nanostructures. This short review cites the application of protein engineering technology to the formation of engineered functional polymers, as well as systems that self assemble as organized 2D molecular lattices on surfaces. Some consideration is given to approaches that could lead to self assembling structures of great complexity.
Key words: Bionanotechnology, protein structure, protein engineering, molecular lattices, biomaterials, streptavidin, molecular electronics

1. INTRODUCTION

One vision of nanotechnology anticipates the precision assembly of complex, large-scale systems incorporating individual devices engineered at the molecular scale. This chapter describes some initial studies and outlines some design principles that explore the assembly of macroscopic structures composed of proteins ordered at the molecular level. Biological systems form complex, self-assembling structures from amino acid polymers (proteins) encoded by deoxyribonucleic acid (DNA). Individual proteins containing thousands of atoms spontaneously fold to form unique three-dimensional (3D) structures replicated with atomic precision. Naturally occurring proteins exemplify many practical, structural, and signal transduction systems, including chemomechanical, electromechanical, optomechanical, and optoelectronic mechanisms, and frequently form cooperative hierarchical assemblies of great structural and functional complexity. The combination of recombinant DNA technology, enabling the

V. Renugopalakrishnan et al. (eds.), Bionanotechnology, 1-9.

synthesis of virtually any polypeptide sequence or functional domain fusion, together with lessons learned from sequentially assembled biological structures like virus particles and regular macromolecular assemblies like protein crystals, provide the basis for designing novel assemblies from engineered biological macromolecules. The successful development of these new material applications requires close integration of computer design, protein synthesis, biophysical measurements, and structural characterization using both X-ray diffraction and electron microscopy. Early results are exemplified by studies of engineered structural fibers based on architectural motifs found in viral spike structures. Examples of designs and assembly strategies for 2D ''semiconductor-like'' architectures based on biological macromolecular motifs are also presented.

2. DEVELOPMENT OF ORDERED POLYMERIC SYSTEMS BASED ON ADENOVIRUS TAILSPIKE PROTEIN ARCHITECTURE

During the early 1990s, an effort was launched through cooperation between the Dupont Central Research and Development Department and Dupont Polymer Products Department to investigate the use of proteins to form ordered polymeric systems.[1] The anticipated advantages of this approach included the ability, by using the methods of recombinant DNA technology, to produce high molecular weight polymers (MW$\sim$100,000 Da) with precisely defined amino acid sequences and composition. By creating polymers composed of repetitive blocks of amino acid sequences corresponding to known protein structural domains, it was envisioned that it would be possible to create polymeric structures that would spontaneously fold and assemble into 3D structures that were ordered at both the molecular and macroscopic levels. Initial objectives focused on producing materials that would emulate the unique mechanical properties of natural biological fibers like elastin and spider silk. At another level, it was believed that the molecular organization present in biological structures could be exploited to produce materials with, for example, novel nonlinear optical properties.

Efforts focused on the generation of repetitive consensus polymer sequences derived from the adenovirus spike protein. The adenovirus spike is a long stiff structure that emerges from the apex of the virus icosahedral capsid structure. Electron microscope and X-ray scattering studies of the adenovirus spike showed that it formed a long stiff shaft organized as a cross-β sheet structure. Sequence and composition analysis showed that the spike was composed of three identical polypeptide chains, each of which incorporated several approximate repeats of a homologous 15-residue

amino acid sequence. At Dupont, constructs were developed that expressed repetitive consensus sequences based on the natural sequence, in the hope that they would spontaneously assemble into trimeric, microfibrillar shafts whose intrinsic stiffness would facilitate spinning into macroscopic fibers in which the individual trimeric shaft axes were all aligned along the direction of the fiber axis. Three sequences were developed, which differed slightly with regard to repeating block sequences and extent of block polymerization, to produce polymer chains ranging from MW ~20,000 to ~100,000 Da.

The synthetic work was complemented by molecular modeling and energy minimization studies that envisioned that the shaft structure was organized as a twisted triangular beam in which each face of the beam was composed of a twisted antiparallel β-sheet with geometric and twist properties similar to that observed in many globular protein structures (Fig. 1).

In fact, the ''designed'' polypeptide chains were experimentally observed to form long fibrous shafts in electron microscope imaging and light scattering studies and to produce liquid crystalline mesophases. The liquid crystalline mesophases could be spun into macroscopic fibers. X-ray diffraction studies of spun fibers showed features characteristic of cross-β structure, although the

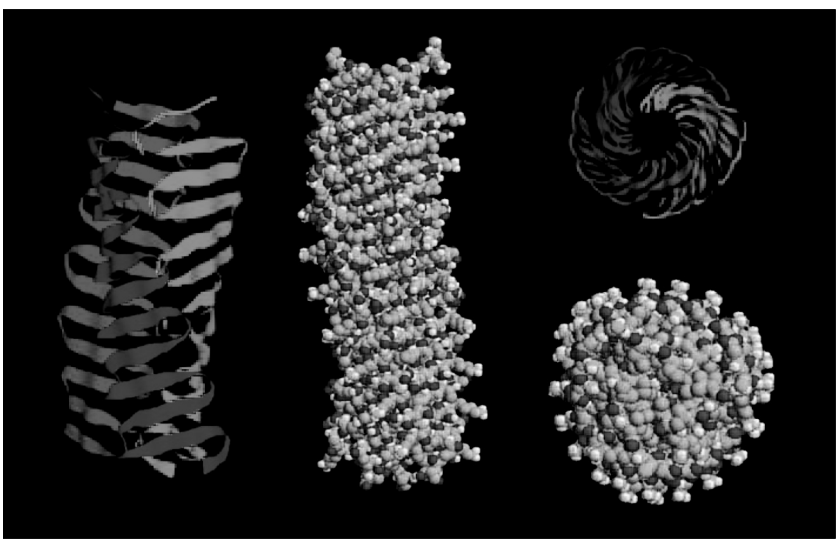

Figure 1. (A) A backbone model of a synthetic cross-β beam structure composed of three polypeptide chains in which each face of the beam is composed of a continuous, twisted, antiparallel β-sheet. (B) End views of the backbone and a space-filling model of the energy-minimized structure, illustrating the solid packing of the interior structure by hydrophobic amino acid residues of the repeating polypeptide sequence (Asparagine–Alanine–Leucine–Arginine–Isoleucine–Lysine–Glycine–Serine–Glycine–Leucine—Aspartic Acid–Phenyalanine —Aspartic Acid–Asparagine)$_n$.

features were not predominant along the fiber axis direction as expected if the shafts were closely aligned along the fiber axis. It was believed that additional experiments to improve spinning solvents and proper conditions could potentially produce more highly aligned polymeric materials, although this objective was not actively pursued at that time owing to changing priorities within Dupont Central Research.

Subsequently, the X-ray crystal structure of the adenovirus shaft was reported at near atomic resolution,[2] and it turned out to have major differences, but some similarities, with the structure that had been modeled. The model structure envisioned that each face of the trimeric shaft was composed of a single polypeptide chain, organized as repeating β-hairpins to form a contiguously hydrogen sheet whose H-bonds lay along the axis of the shaft. The actual virus structure is composed of polypeptide chains that form repetitive β-bends, which alternate "herringbone-fashion" between two adjacent faces of the trimeric beam structure. In addition, the hairpin bends in each chain twist so that the interactions between strands are primarily due to stacking between the sheets of the adjacent chains, rather than to interchain hydrogen-bonding (Fig. 2).

The true structural organization of the synthetic structures is further clouded by more recent studies[3] that investigated the propensity for both randomized and homopolymer amino-acid sequences to form amyloid struc-

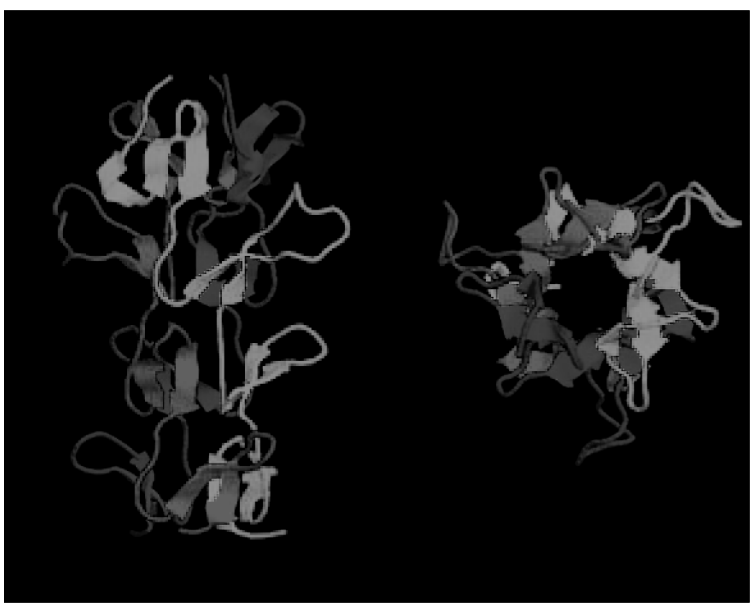

Figure 2. (A) View of a backbone model of a section of the adenovirus spike as determined from X-ray crystallography. (B) An end view.

tures. Both the electron microscope images of amyloid fibers and their diffraction patterns appear very similar to the corresponding data derived from the engineered synthetic adenovirus sequences. Obviously, much remains to be done to further characterize these structures, both as a means of manipulating them as new structural materials and potentially to suggest a means of ameliorating amyloid diseases.

3. TWO-DIMENSIONAL LATTICES ON SURFACES

Although many practical sensor applications are already based on the immobilization of proteins on semiconductor substrates, longer range objectives envision the self-assembly of complicated molecular circuitry on semiconductor substrates that could potentially emulate the function of present day metal-oxide-semiconductor field-effect transistor (MOSFET) devices but at substantially higher device densities. Key issues involved in the realization of such ideas involve the design and functional optimization of individual molecular components, followed by their controlled hierarchical organization into large-scale functional assemblies. The strategies for producing devices with molecular scale components combine aspects of the ''top–down'' approach used for conventional MOSFET device design with a ''bottom-up'' approach to self-assembly common to biological systems.

3.1 Two-Dimensional Lattices Based on Streptavidin

Early conceptual ideas[4] envisioned the formation of 2D molecular structures that could self-organize on self-assembling monolayer (SAM) surfaces. It is desirable in many biotechnology applications to link biomolecules irreversibly through a highly specific protein–ligand interaction. The high-affinity interaction between streptavidin and its small-molecule ligand biotin,[5] which can readily be linked covalently to a variety of other biomolecules, has been widely used for biotechnology applications. In an early hypothetical example of a self-assembling 2D lattice, it was envisioned that array structures could be formed by linking ''strut molecules'' constructed, for example, from DNA segments terminated with covalently linked biotin groups to ''nodes'' formed by the symmetric, tetrameric streptavidin molecule.[4] A remarkable paper by Ringler and Schulz[6] actually realized such 2D lattice structures on SAMs by linking streptavidin and an engineered fourfold symmetric protein to form molecular lattices with predetermined dimensions and symmetry. As shown in Fig. 3, the lattice is composed of two molecular components, streptavidin, a tetramer with D2 symmetry that incorporates four biotin binding sites, and

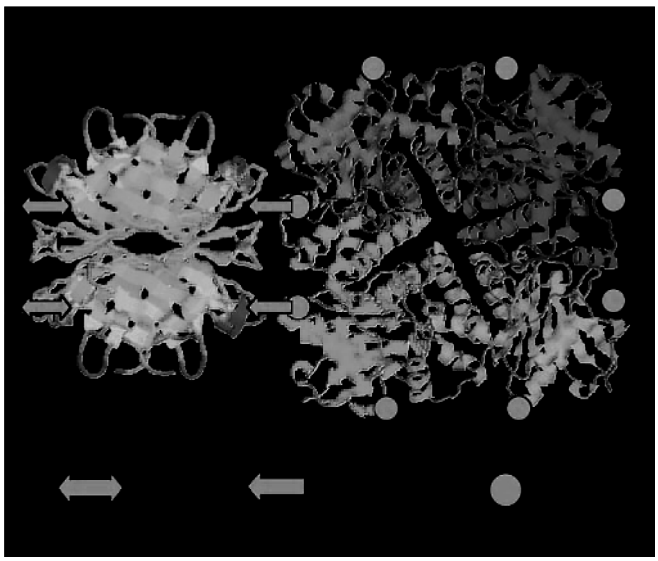

Figure 3. Molecular components of an engineered lattice formed of protein molecules arrayed on a 2D surface. The two molecular components are streptavidin, a tetramer with D2 symmetry that incorporates four biotin binding sites, and RhuA aldolase, a fourfold symmetric tetramer that has been engineered using recombinant DNA technology and then chemically modified to allow the covalent chemical attachment of biotin molecules. Circles represent sites of cysteine SH groups introduced into the tetrameric protein aldolase through site-directed mutagenesis, Arrows represents biotin groups that covalently react with the cystiene SH groups. Double arrows represent "di-biotin" crosslinking molecules.

RhuA aldolase, a fourfold symmetric tetramer that has been engineered using recombinant DNA technology to introduce surface cysteine sulfhydryl groups that were subsequently functionalized through the covalent attachment of biotin molecules.

The work of Ringler and Schulz demonstrates the potential for forming controlled 2D molecular assemblies of protein molecules on surfaces. High-density lattices composed of protein molecules with switchable magnetic or photosensitive prosthetic groups might for example be components of high-density memory arrays or sensors.

3.2 Self-Assembly of Hierarchical Organized Structures

The work of Ringler and Schulz[6] demonstrates the potential for self-assembly of regular 2D lattices with defined geometrical properties constructed of protein molecules on surfaces (see Fig. 4). However, substantially more complicated self-assembling structures can be envisioned that could

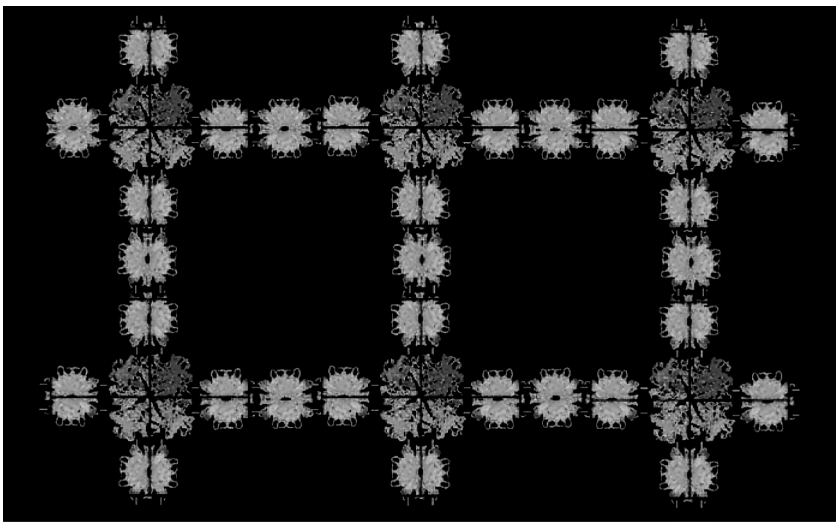

Figure 4. A schematic of the engineered 2D lattice formed by coordination of streptavidin and the engineered and modified form of RhuA aldolase, as described by Ringler and Schultz.

be constructed using a combination of molecular epitaxy and gated molecular assembly principles. There have been great advances in the use of atomic force microscopy to manipulate individual atoms on silicon or other semiconductor substrate surfaces.[7] Although it is probably unrealistic to imagine that devices of substantial complexity could actually be built using atom-by-atom construction principles, it is certainly feasible to produce a regular 2D lattice of individual atoms on a silicon surface. A regular lattice of gold atoms, for example, can be used as chemical immobilization sites for a 2D lattice of protein molecules linked through surface cysteine sulfhydryl groups. Additional different proteins can be added to immobilized proteins to eventually produce organized structures of potentially arbitrary complexity. In contrast to the 2D lattices that self assemble on SAM surfaces, whose formation is facilitated by the free diffusion of the assembling molecules on the SAM surface, the latter strategy would introduce fixed sites on the silicon surface to provide nucleation points for a sequence of hierarchical molecular assembly steps. The wealth of available X-ray crystal structures of proteins provides an extensive toolbox of molecular components that are able to undergo or perform chemomechanical, electromechanical, optomechanical, and optoelectronic processes. In addition, structural biology provides many examples of structures whose intermolecular interactions can be precisely controlled both in space and time.

While control of intermolecular interaction geometry is fundamental to any design process, lessons derived from the assembly of complex biological assemblies like viruses, suggest that it will be equally important to achieve temporal control over successive stages in the assembly of more complicated molecular devices. Again, structural biology provides numerous examples of how structural assembly processes can be "gated," usually as a consequence of either a precise biochemical modification of one of the molecular components or through the introduction of a small-molecule ligand that induces a conformational change in one of the molecular components so that it can interact stably with another.

The approach outlined in the preceding paragragh envisions a process in which the interactions of discrete molecular components are precisely controlled at every level and stage of device assembly. However, for many device applications, large-scale assemblies could potentially be spontaneously organized by other means. One interesting possibility involves systems based on finite automata.[4,8] For example, studies of interactions important for the habit development in protein crystals[9] and simulations of extended lattice structures formed by finite automata nucleated at a single point,[4,8] suggest that molecular systems incorporating components whose interactions are governed by simple interaction and symmetry rules, can spontaneously assemble to form highly complex, yet finite structures.

4. CONCLUSION

Although many outstanding issues will require resolution if useful functional devices are ultimately built from macromolecules, many of the required tools for simulation, modeling, protein engineering, structure determination, biophysics, and semiconductor technology are available to begin investigating the potential of such devices.

REFERENCES

1. O'Brien, J. P., Hoess, R. H., Gardner, K. H., Locke, R. L., Wasserman, Z. R., Weber, P. C., and Salemme, F. R., 1994, Design, synthesis, and fabrication of a novel, self-assembling fibrillar protein, in *Silk Polymers : Materials Science and Biotechnology* (D. Kaplan, W. W. Adams, B. Farmer, and C. Viney, eds.), ACS Symposium Series 55, pp. 104–117.
2. Van Raaij, M. J., Mitraki, A., Lavigne, G., and Cusack, S., 1992, A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein, *Nature* **401:**935–936.

3. Dobson, C. M., 2001, The structural basis of protein folding and its links with human disease, *Phil. Trans. Roy. Soc. Lond. B*. **356:**133–145.
4. Sligar, S. L., and Salemme, F. R., 1992, Protein engineering for molecular electronics, *Curr. Opin. Struct. Bio*. **2:**587–592.
5. Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., and Salemme, F. R., 1989, Structural origins of high affinity biotin binding to streptavidin, *Science* **243:**85–88.
6. Ringler, P., and Schulz, G., 2003, Self-assembly of proteins into designed networks, *Science* **302:**106–109.
7. Strocio, J. A., and Eigler, D. M., 1991, Atomic and molecular manipulation with the scanning tunnelling microscope, *Science* **254:**1319–1326.
8. Wolfram, S., 2002, *A New Kind of Science*, Wolfram Media.
9. Salemme, F. R., Genieser, L., Finzel, B. C., Hilmer, R. M., and Wendoloski, J. J., 1988, Molecular factors stabilizing protein crystals, *J. Cryst. Growth* **90:**273–282.