ENN Encyclopedia of Nanoscience and Nanotechnology
www.aspbs.com/enn

Nucleoprotein Assemblies

Steven S. Smith

Kaplan Clinical Research Laboratory, City of Hope, National Medical, Duarte, California, USA

CONTENTS

1. OVERVIEW

Bionanotechnology is an emerging field with great promise in molecular science. The field is so new that it has yet to be formally defined. However, it can be characterized as a primitive technology that takes advantage of the properties of highly evolved natural products like nucleic acids and proteins by attempting to harness them to achieve new and useful functionalities on the nanoscale.

It is appropriate to compare bionanotechnology to the technology of the paleolithic in which natural products like wood and stone were shaped into implements and devices. Much like the paleolithic technologists who used knowledge of their surroundings to identify components and processes used in tool-making, modern bionanotechnologists use knowledge of chemistry, biochemistry, and molecular biology to identify components and processes for the construction of self-assembling materials and devices. Paleolithic technology was limited by the basic properties of natural products like wood and stone. Wood and stone could be shaped and modified, but they could not be forced into the arbitrary *de novo* designs that can be attained with our modern understanding of wood composites and ceramics. In a like fashion, bionanotechnology is limited by the basic properties of biomolecules. The field will certainly mature in the near future, but this maturation will require, among other things, advances in our understanding of the nature of nucleic acid and protein folding so that arbitrary *de novo* design of such molecules can be contemplated.

Like other forms of nanotechnology, bionanotechnology seeks to define approaches to the fabrication of useful materials and devices. However, the construction principles utilized in the field often originate in biology and the goals are often biomemetic (e.g., the construction of biosensors [1]) or aimed at the solution of long-standing research problems (e.g., protein crystallization [2]. Nonbiological problems have also been approached in attempts at the construction of electronic circuitry using biomolecules [3] and the construction of a fuelled nanomechanical oscillator [4] and a nanomechanical switch [5]. At the heart of these approaches is the concept of self-assembly. Self-assembly of ordered elements is a defining property of living things. Moreover, the progressive increase in the complexity of the processes used by living things in self-assembly is a defining property of evolution. During the roughly 4.2 billion years of prebiotic and Darwinian evolution that have taken place on earth, an almost incomprehensible variety of molecular structures, functionalities, and associations have appeared. This evolutionary experience is stored in modern living systems. For this reason, modern living systems comprise a wealth of addressable macromolecular components. Of these, the nucleic acids and proteins are the most easily manipulated and the best understood. Thus, it is not surprising that they have been the first to be exploited to produce nucleoprotein-based addressing systems in nanobiotechnology [6–11] that attempt to exploit self-assembly and ordered proximity.

1.1. Creation of Biomolecular Machines by Directed Evolution

A machine is usually defined anthropomorphically as a device having a purpose [12]. A more general definition is as an assemblage of parts that transmit forces, motion, and energy one to another in a predetermined manner [13]. Such definitions are meant to encompass the spectrum of machines from the simple lever and wedge to the most complex networks of automata. In biology this definition encompasses an even broader spectrum of entities from the simplest biomolecular catalysts to the planetary biosphere. The central thesis in modern biology has been that the planetary biosphere arose spontaneously. Consequently an enormous amount of scientific research has been focused on

ISBN: 1-58883-001-2/$35.00

understanding how the earth underwent a transformation from an apparently lifeless state to the biosphere that it has now become.

A significant body of knowledge in support of this spontaneous transformation has now been accumulated in the field of prebiotic chemistry. This body of knowledge suggests pathways for the spontaneous production of each of the components of nucleic acids and proteins from conditions likely to have been present on the primitive earth [14, 15]. Most of this work has been aimed at constructing plausible routes to self-assembling chemical systems that would become subject to Darwinian natural selection. Working backward from apparent molecular fossils, the stated goal of these studies is to plausibly order the events that occurred on the prebiotic earth that gave rise to Darwinian evolution [16, 17].

The discovery of unifying principles in this area has been stimulated enormously by the discovery that nucleic acids themselves can serve as catalysts and must therefore be listed with peptides and proteins as the raw materials of simple biomachines. The current hypothesis is that the early stages of prebiotic evolution generated a series of supramolecular aggregates composed largely of RNA. These aggregates, sometimes termed metabolosomes [18], are postulated to have attained a high degree of complexity permitting them to carry out a complex series of chemical transformations. The level of complexity at which evolution by natural selection could take hold has been called the Darwinian threshold [17]. It is thought that the entity that crossed this threshold was a supramolecular aggregate that possessed the capacity for self-assembly, self-replication, and perhaps translation (i.e., the capacity to convert information stored in ribonucleic acid sequences into protein sequences [17]).

In efforts designed to recreate the basic designs for the postulated primitive nucleic acid machines, simple laboratory procedures for their construction have been developed using the principles of directed evolution *in vitro* [19–24]. Generally these techniques involve the production of a pool or library of nucleic acid sequences that are subjected to a selection process involving binding or catalysis. Those members of the library that can perform the binding or catalytic task dictated by the selection criterion are retained by the selective process. Initially, this represents only a tiny fraction of the initial library. This fraction is copied so as to increase the total number of copies of the selected species. After repeating this process many times, the resulting pool is thus reduced in complexity to a few representative molecules having the desired properties.

1.1.1. Aptamers

Aptamers represent perhaps the simplest class of molecular machine that has been produced by using bionanotechnology. The word aptamer is derived from the Greek word "aptus" meaning to fit, in this case, a molecule adapted to fit into another molecule. Under the first definition above [12], aptamers represent a class of molecular machine that can be considered to be devices having a specified purpose: binding to a ligand. Ligands that have been bound include organic dyes [23], proteins [25], other small molecules [26], and other nucleic acids [27]. Aptamers have been produced from both DNA [25] and RNA [23]. Although most aptamers are essentially static machines, in the sense that a wedge is a static machine, more recently aptamers have been produced that exhibit behavior that can be regulated by a small molecule [28]. This suggests that the ligand either modifies the structure of the aptamer so that it fits into its target site properly or that the ligand forms part of a complex complementary to the targeted binding site. In short certain aptamers may be dynamic machines since they appear to move in order to bind.

1.1.2. Ribozymes

It is a short step from binding to catalysis because biological catalysts are selected so as to bind to and thereby stabilize transition states in a given reaction pathway. This property of catalytic biomolecules allows them to drive many organic reactions at ambient temperatures with reasonable efficiencies. RNAs with these properties are readily isolated by directed *in vitro* evolution techniques. *In vitro* selected ribozymes exhibit rate enhancements ranging from 10^3 to 10^5 over the uncatalyzed reactions. RNAs that have been isolated by these methods include those that catalyze carbon–carbon bond formation in a Diels–Alder reaction [29], phosphate bond cleavage in cleaving single-stranded DNA [22], phosphate bond formation during RNA polymerization [30], RNA ligation [31], carbon–nitrogen bond formation during self-alkylation [32], and carboxyl attack on phosphorous during amino acid activation.

1.1.3. Deoxyribozymes

Most secondary structure in nucleic acids involves hydrogen bonding. Since DNA lacks the 2′ hydroxyl group on ribose that is present in RNA, it possesses a reduced potential for hydrogen bond formation and was initially thought to be less suitable for *in vitro* selection of aptamers and catalytic nucleic acids because it cannot adopt as many secondary structures as RNA (i.e., its conformation space [33] is restricted relative to RNA). This initial concern notwithstanding, catalytic DNAs of many types have also been selected by directed *in vitro* evolution. Porphyrin metallation [34], peroxidase activity [35], phosphoesterase activity [36], and DNA degrading activity [37] have all been documented, suggesting that the collection of accessible deoxyribozymes may be nearly as extensive as the collection of ribozymes.

1.1.4. Enzymes

Biologically occurring deoxyribozymes have not yet been observed. Biologically occurring ribozymes are rare, but several are known to exist. Catalytic proteins (enzymes), on the other hand, are the overwhelmingly predominant catalytic molecules in living things, with many thousands of different forms present in the average mammal, for example. Thus, by any measure, the collection of biologically occurring catalytic proteins (enzymes) is currently much larger than the collection of known catalytic nucleic acids since it encompasses the vast majority of the aggregate number of gene products contained in all species now present on earth. Even so, directed evolution techniques have been applied to the

development of new enzymes. Here, the methods are generally based on the selection of whole bacteria. In this process the bacteria develop previously unknown catalytic functions by recruiting proteins that are normally expressed in the bacteria for other purposes through a process of mutation and selection [38, 39]. For example, bacteria that are normally unable to use a certain sugar for growth are exposed to a DNA damaging agent that introduces random changes (mutations) in the genetic code of the organism. Most of these changes are deleterious and either cause the bacteria to die or grow more slowly than it would in its natural growth medium. However, if the mutated bacteria are forced to grow on the sugar that they are normally unable to use, a small fraction of them (perhaps 1 in 10^8) will be able to use the sugar in question for slow growth. Subsequent rounds of this mutagenesis and selection process yield new enzyme functionalities associated with absorbing and breaking down the sugar in question. Such *de novo* protein development completes the spectrum of currently known techniques for the *de novo* development of nanoscale components available for nanobioscience and nanobiotechnology.

Each of these *de novo* methods takes advantage of the almost incomprehensible conformation spaces available to proteins, RNAs and DNAs [33] as noted above. Given the hydrogen bonding capacities of each polymer and the number of commonly observed monomeric building blocks (4 each for nucleic acids and 20 for proteins), the number of available conformations is canonically ordered as follows: protein $\gg$ RNA $>$ DNA when each oligomer contains the same number of monomers. It is worth noting that even the conformations available to the most restricted system (the DNAs) is 4^n (where n = sequence length in nucleotides) assuming only one conformation per sequence. It is clear that the vast collection of shapes available to oligmers of modest length ($n > 20$) offers an almost inexhaustible wealth of potential devices available for nanoscale assembly.

1.2. Modification of Existing Biomolecular Machines

Directed evolution has provided a productive route to the *de novo* selection of new nanoscale machines; however, with proteins, it has often been simpler to modify existing molecules for purposes related to their current function. This makes recombinant DNA technology among the most powerful tools in bionanotechnology. Recombinant DNA technology is generally used to place desired proteins under the control of specific DNA sequences, called promoters, that permit the production of large quantities of the desired material. This is generally coupled with the second tool, site-directed mutagenesis, which allows the modification of the protein itself. In site-directed mutagenesis, recombinant DNA techniques from molecular biology are used to alter the genetic code so as to modify the amino acid sequence of the protein. The resulting proteins have generally been redesigned so as to suit a particular application. In general these alterations improve stability or functionality of the protein in a given application.

These techniques have been used to optimize [40, 41] the capacity of the light harvesting protein bacteriorhodopsin to cycle between two stable photochemical states so as to permit the construction of a three-dimensional memory device. In this device the modified protein is first immobilized in a matrix. A paging laser at 570–630 nm is used to cycle the engineered bacteriorodopsin from an all-*trans* state through a series of 13-*cis* retinal states to a light-adapted protonated all-*trans* retinal state. A full power write laser at 680 nm acting orthogonally through an active matrix liquid crystal light modulator is then used to convert volume elements of the protein-containing matrix from the light adapted all-*trans* retinal isomer to the 9-*cis* isomer to set the state of engineered protein in the matrix, thus creating a three-dimensional (3D) (volumetric) information storage pattern. The 570–630 nm paging laser is coupled with the 680 nm laser at low power to read the information by projecting the stored data onto a charge-coupled device. Bacteriorhodopsin is stable in this state and data stored in this manner are stable for decades in the absence of blue light or high temperatures. Erasure is accomplished with a 410–430 nm diode-pumped laser that converts the 9-*cis* to one of the all-*trans* states.

This working system integrates bionanotechnology (protein engineering and 3D immobilization techniques) with laser based input–output (IO) technology to achieve a storage device that is resistant to the high levels of radiation and shock required for satellite communications, while achieving a 1000-fold improvement in memory storage capacity over currently operational devices. The ultimate goal of bionanotechnology is to create devices based on biomolecules that achieve the even greater degree of complexity and miniaturization exhibited by living things.

2. SUPRAMOLECULAR ASSEMBLIES

A key step in the origin of life appears to have been the origin of translation [17, 18]. Translation is the process by which RNA sequence information is copied into peptide or protein sequence information. In general, this is assumed to have occurred at a time in prebiotic evolution when catalytic RNAs provided the only mechanism by which metabolic interconversion of chemical compounds could be accomplished. The appearance of translation is thought to have moved evolution to a new plane in which supramolecular aggregates (SMAs) appeared that had crossed the Darwinian threshold (i.e., the point at which the complexity of the aggregate permitted replication, self-assembly, and natural selection). This transition is thought to have been achieved by incorporating information storage and feedback to the storage mechanisms [17].

The modern ribosome is generally considered to be a molecular fossil representing a highly refined version of one of the key components of these supramolecular aggregates or SMAs. The more primitive SMAs that gave rise to them are generally thought of as metabolosomes in which nucleic acid scaffolds of RNA called organizing centers held catalytic RNAs in place by base-pairing between a short region of the nucleic acid scaffold and a short region of the catalytic RNA [18]. In this model, the alignment of sequestered catalytic RNAs was set by the sequence of the complementary RNA in the organizing center. This alignment and proximity is thought to have facilitated sequential reactions in the

metabolosome. The linear arrangement along the organizing center is thought to have been translated into the linear arrangement of amino acids, thus promoting the formation of peptides by ordering condensation reactions through appropriate juxtaposition of the amino acid adapters composed of RNAs. It is thought that these adapters ultimately evolved into modern transfer RNAs or tRNAs and that they continue to serve this purpose in modern living things. Bionanotechnological approaches to the construction of supramolecular aggregates have adapted many of these biological principles.

2.1. Creation of SMAs Using Base-Pairing in Complementary Nucleic Acids

Several nucleic acid scaffolds have been constructed using DNA complementarity. These scaffolds are assembled by annealing complementary DNAs followed by enzymatic ligation, or splicing, of the subsections of a larger entity. The scaffolds are defined by their connectivity. A cage having the edge connectivity of a cube [42] has been characterized, as well as a system with the connectivity of Borromean rings [43], the connectivity of a truncated octahedron [44], and the connectivity of a 2D lattice [45]. Thus, the topology available for the construction of supramolecular aggregates is quite elaborate. This topology has been used in the construction of a two-state switch that is actuated by changes in salt concentration [5] and in the construction of a DNA-fuelled oscillator [4] that is activated by the addition of oligodeoxynucleotides.

The two-state switch is based in the well-known transition of DNA from the right-handed B-form double helix seen at low salt to the left-handed Z-form double helix seen at high salt. In Z-DNA the phosphate backbone zigzags around the stacked base pairs instead of winding smoothly around them in a helical pattern as it does in B-form DNA [5]. This conformational change twists the DNA from the right- to the left-handed form, and it is favored by sequences that are rich in guanine–cytosine base pairs. Interestingly, it was found that a double-crossover molecule could be made to stabilize two duplex strands of DNA into a side by side arrangement that appears to remain in the B-form at high salt. Double-crossover is a term from genetics used to explain genetic exchanges that often occur during the DNA-strand exchanges that characterize recombination in biology. Physically a system of this type comprises four strands of DNA intertwined so that two of the strands cross over from one of the double helices to the other and back again so as to constrain them to lie side by side. This rigid system resists conformational change as salt concentration is altered. When two double-crossover systems are connected by a short double-stranded region that is rich in guanine–cytosine base pairs, the two double-crossover regions at either end of the double-stranded region can be made to lie on the same side of the double-stranded region at low salt and on opposite sides of the region after the B to Z transition induced by high salt. Resonant intramolecular energy transfer from chromophores placed at the ends of the double-crossover strands that lie near the linking region can be used to detect the change in proximity induced by the increase in salt concentration. In effect, the system becomes a biosensor that can detect changes in salt concentration [5].

The DNA-fuelled oscillator likewise induces nanomechanical motion in a supramolecular assembly formed from DNA [4]. Here again, resonant intramolecular energy transfer is used to detect the close approach of two chromophores during the formation of a duplex waste product from the sequential addition of two complementary fuel strands. The system essentially makes use of strand displacement in DNA. Here a chromophore and a fluorescence quencher are placed at each end of a short DNA strand during synthesis. This short strand is hybridized (annealed) to two strands that are complementary to the ends of the chromophore-containing strand, extend beyond the ends by several nucleotides, but do not cover the central nucleotides of the chromophore-bearing strand. In this state the duplex DNA region is extended by the stiff stacking interactions along the Z-axes of the two duplex regions, thus maintaining the ends of the strand containing the chromophore and the fluorescence quencher at a distance adequate to prevent quenching of the fluorescent signal. Once this state is achieved, a closing fuel strand is added that is complementary to the unpaired ends of each of the previously added strands but is longer than one of them so as to leave another unpaired region. Once annealed, this strand forces the chromophore-containing strand into a hairpinlike conformation that juxtaposes the chromophore and the quencher, extinguishing the fluorescence signal by resonant intramolecular energy transfer. The system remains in this conformation until an opening fuel strand is added that is completely complementary to the closing fuel strand added previously. Annealing between the two fuel strands begins at the short unpaired region and continues by strand displacement until the opening fuel strand that maintains the chromophore containing strand in the hairpin conformation is stripped away to relax the system and permit fluorescence to reappear [4].

2.2. Creation of SMAs Using Complementary Protein Domains

Multisubunit proteins can exhibit a form of complementarity between subunits that results in self-assembly of defined multimolecular protein aggregates. Protein–protein interaction is not easily predicted, as can be anticipated from the increased complexity intermolecular interactions possible between the 20 common amino acids compared to that of the 5 common bases that comprise the nucleic acids. However, as the number of known 3D protein structures has increased it has become possible to catalog numerous interactions between internal protein domains that can be reproduced with synthetic peptides. For example, the SH2 and SH3 domains of the Src tyrosine kinases are known to interact with phosphotyrosine and proline rich motifs to form internal associations that influence kinase activity. SH3 and SH2 are well characterized examples of protein domains whose natural folds are spontaneously adopted by recombinant peptides containing the native peptide sequence [46, 47]. Each isolated domain retains its affinity for phophotryosine (SH2) or proline rich peptides (SH3)

that interact with these domains to form complex associations involving the Src kinases.

In general, domain–domain interactions also govern the spontaneous association of multisubunit proteins. Dimeric proteins often associated with DNA-binding proteins [48] represent the simplest examples of these aggregates. In addition, the multimeric assemblies that form viral capsids [49] provide examples of the complexity that is achieved in biology with simple interactions of this type. The symmetry rules that govern the self-assembly of these subunits into complex multimers are well known [49, 50].

These advances in the understanding of protein domain structure have allowed the construction of fusion proteins (i.e., chimeric proteins constructed by recombinant DNA techniques that fuse disparate functional domains in to a single unit) with predictable subunit–subunit complementarity that have been used in the formation of a closed tetrahedron protein cage and an extended protein filament [51]. The closed protein cage was designed by searching a protein database for dimeric and trimeric protein structures that begin or end in an alpha helix. Molecular modeling was than used in computer-aided design of a fusion between domains thus identified. Once models were constructed, they were examined to determine whether or not they possessed appropriate symmetry for self-assembly into a closed geometric figure. One fusion between the trimeric bromoperoxidase of *Streptomyces aureofaciens* and the dimeric M1 matrix protein of the influenza virus was shown to be capable of the formation of a tetrahedral cage by molecular modeling. Molecular cloning techniques were used to express the designed fusion protein in *E. coli*. The purified fusion protein was shown to form the expected tetrahedral cage based on sedimentation velocity, light scattering, and electron microscopy experiments. Similar methods permitted the construction of an extended filament [51].

Biomolecular motors are also potential components of a variety of protein containing devices that have obvious potential in structuring or redirecting flow and mixing in nanofluidic systems. The F1-adenosine triphosphate synthase (F1-ATPase) provides a well-studied example of a molecular motor. It is also a naturally occurring supramolecular protein aggregate, comprising three alpha, three beta, and one gamma protein subunits. It is effectively a three-point stepper motor that generates ATP from ADP and inorganic phosphate by using rapid sequential conformational changes driven by a proton gradient. The conformational changes associated with the stepwise rotation of the protein rotor (i.e., the gamma subunit) between the surrounding sets of alpha and beta subunits cause the enzyme to catalyze the formation of ATP. This mechanism of action was deduced through studies of the biochemical kinetics of the system [52]. It was later supported by 3D structure determination [53] and unequivocally demonstrated by chemical means [54] and by direct real-time observation [55, 56] of the reverse reaction in which ATP is hydrolyzed. These advances in understanding of the system have permitted it to be modified using site-directed mutagenesis so that it could be mounted on an engineered substrate in order for it to be able to produce controlled rotation of synthetic arms. Rotation can be initiated by the addition of one compound (2 mM adenosine triphosphate) and halted by the addition of another compound (sodium azide). With further protein modification, rotation could be halted by the addition of the metal ion Zn^{2+} and reinitiated by the addition of the metal ion chelator EDTA. In short, the rotation of this modified machine can now be controlled by changing its chemical environment [57–59].

2.3. Addressable Supramolecular Assemblies Based on Nucleic Acid Complementarity

As noted above the key feature postulated for metabolosomes is the ability of the nucleic acid scaffolds that formed the organizing centers to sequester and align nucleic acid adapters. Base-pairing between a short region of the nucleic acid scaffold and a short region of the adapter is thought to have generated the appropriate alignment of the adapters [18]. This concept has also been employed in several applications in bionanotechnology. See Figure 1.

For example gold nanoparticles can be linked to DNA by attaching alkanethiol groups to the ends of single-stranded DNA molecules [10, 11]. When the modified DNAs are

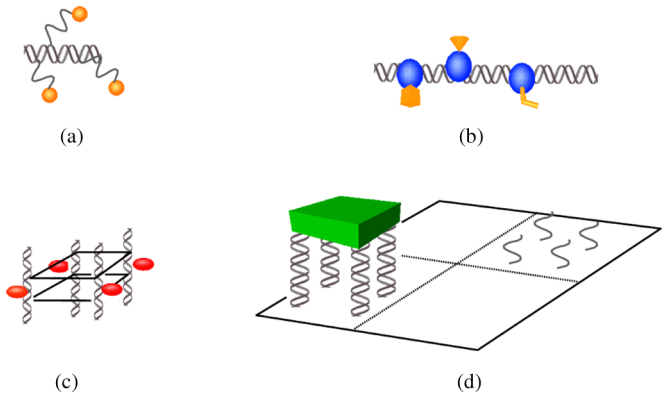

Figure 1. Addressable nucleoprotein assemblies. (a) Molecular addressing using nucleic acid complementarity. Here the central duplex is formed either by DNA • DNA interaction or by DNA • RNA interaction along a DNA or RNA organizing molecule. Short tethers of DNA are used to link thermostable moieties to which other molecules can be attached (e.g., gold or streptavadin molecules) indicated by the spheres [6, 11]. (b) Molecular addressing using protein–nucleic acid interaction. Here the central duplex contains 5FdC residues that form covalent complexes with DNA (cytosine-5) methyltransferase fusion proteins as they catalyze methylation of the DNA target. The fusion proteins are ordered along the DNA sequence based on the biospecificities of the cytosine methyltransferses of which more than 40 are known [8]. Each fusion protein can carry a different functionality indicated by the irregular shapes attached to each sphere. (c) Molecular addressing to a lattice. In this proposal [70] a lattice of DNA is ligated together so as to serve as a host for guest molecules (ovoid shaped objects). The purpose of this proposal is to drive crystallization so as to facilitate 3D structure determination. (d) Molecular addressing of silicon objects to a patterned DNA surface. In this proposal, DNA is patterned onto a surface (depicted here as a checkerboard). Silicon objects bearing complementary DNA strands on one of their surfaces are moved into the vicinity of their complementary strands on the surface pattern where they should attach biospecifically [9]. The attached object is linked to the DNA strands on the front left of the surface. Noncomplementary DNA states at the right rear of the surface are ignored.

exposed to a gold surface they adsorb by forming covalent bonds between the gold and sulfur atoms. Like other metal sulfur bonds the Au–S bond is very strong and can easily link a long DNA molecule tightly to the surface of the gold particle. The sequences of the attached DNA molecules can be chosen so as to be complementary to a linker strand. This permits the nanoparticles to spontaneously self-assemble into aggregates when mixed with the linker strand [10]. If they are complementary to a linear DNA strand, then they can be tethered to duplex DNA in an ordered fashion [11] around the screw axis of the B-DNA molecule (36°/base-pair). This approach can also be used to order the thermostable-protein streptavadin along the screw axis of a DNA helix [6]. Here biotinylated DNA is allowed to interact with streptavadin prior to nucleic acid hybridization. The steptavadin protein binds to the end of the DNA due to the innate capacity of streptavadin to bind extremely tightly to biotin. A DNA or an RNA organizing molecule is then used to sequester and align the tethered streptavadin molecules [6] much as originally envisioned by Gibson and Lamond [18] for the metabolosome. When DNA is used as the organizing molecule the streptavadin moieties should be arrayed in accordance with the screw axis of the B-DNA molecule (36°/base-pair). When RNA is used this orientation is expected to be in accordance with the screw axis of the RNA • DNA hybrid (30°/base-pair).

Both the gold and the streptavadin aggregates provide the opportunity for secondary assembly via the gold or streptavadin moiety. For example, the streptavidin molecule binds biotin with a stoichiometry of four biotin residues per protein moiety, leaving three unused sites on the DNA–streptavadin conjugate. These unused sites have been used to attach biotinylated antibodies to DNA-tethered streptavadin [6]. This permits the derivatization of microstructured surface arrays of DNA using the base-pairing complementarity of the array so as to convert it to a protein array of different antigen specificities [6]. Further, by adding biotin to both of the 5′ ends of complementary oligdeoxynucleotides it has been possible to form ring structures or networks of DNA linked by two, three, or four biotin contacts per streptavadin residue [60].

Biotinylated proteins can be linked to streptavadin and then targeted to DNA or RNA organizing molecules. Although high-fidelity annealing is generally only achieved at temperatures above 50 °C, annealing with bound proteins can be carried out at room temperature [61]. Alternatively, biotinylated proteins can be sequestered without regard to order along an organizing DNA or RNA molecule by incubating them at low temperature with streptavadin molecules tethered to DNA or RNA [62].

Using these methods, the effects of protein proximity have also been studied with this system. Here, the NAD(P)H/FMN oxidoreductase was placed adjacent to luciferase in order to improve the properties of light emission by the system that is normally coupled by these enzymes in their soluble form. Chemical biotinylation of proteins is not regioselective and can damage proteins. To avoid this problem, fusion proteins were prepared that contained a biotin acceptor region from the *E. coli* biotin carboxy carrier protein. This served as a biospecific attachment site for *E. coli* biotin ligase for the regiospecific attachment of biotin to the luciferase and to the NAD(P)H/FMN oxidoreductase. Streptavadin-containing adapters were attached to the fusion proteins and annealed to a DNA-linker strand that linked the two enzymes either to the same or to different strands. The SMAs thus created were then deposited on a microplate surface again through a streptavadin–biotin linkage to the end of the carrier strand. A 2.0- to 2.5-fold rate enhancement was observed when the proteins were adjacent on a linear DNA compared to randomly placed on a surface [61].

Each of these systems results in a free (often soluble) nanoscale assembly. On a larger scale DNA sequences can be manipulated by electrophoresis in ways that immobilize them on the micro- and submicroscale [63]. Such arrays become templates for addressing silicon and gold components [9] as well as biotin linked antibodies and proteins [6, 61, 62].

2.4. Addressable Supramolecular Assembly Based on Protein-DNA Specificity

The feasibility of an approach to addressable self-assembly that enables the construction of ordered assemblies and devices has also been described [7, 8, 64–66]. The principal advance brought about by this work is the ability to place fusion proteins in preselected positions on a DNA scaffold under conditions that do not denature sensitive protein components. The protein–nucleic acid structures self-assemble at 5FdC [7, 8, 64] or dU substituted [66] recognition sites. Order is specified during the synthesis of the DNA by placing DNA methyltransferase recognition sites along the DNA. Two- and three-address assemblies are obtained in good yield [65], and fusion proteins have been constructed and targeted to preselected sites on linear [8] or branched targets [67].

One advantage of fusion protein-targeting over streptavadin protein-tethering is that a covalent linkage is established between the cytosine methyltransferase and the DNA that is heat stable. Thus, an array of ordered fusion proteins that might find application in a thermal cycling system could be constructed by this method without difficulty, whereas the streptavadin system would break down under repeated thermocycling not only at the level of nucleic acid but also at the streptavadin–biotin linkage. With the tethering system, one could imagine ligating the tethers in place after assembly had taken place at lower temperatures. However, streric problems would reduce yields in many cases and ligation would not solve the problems associated with the biotin–streptavadin linkage in this application.

An advantage of tethered assemblies like those ordered by streptavadin conjugation is that the length of the tether can be varied in order to accommodate proteins or other elements of unknown dimensions by trial and error. On the other hand, the use of methyltransferase fusion proteins in ordering elements on DNA or elements that contain DNA requires considerable effort in computer-aided design. This is because the structural imperatives that are imposed by such assemblies must be anticipated, insofar as is possible. As noted above for the construction of SMAs from fusion proteins [51] extensive use of the 3D structures of proteins

available in protein databases is made [7, 8, 64, 66] in developing potential fusions and SMAs based on this method. Once it is determined that the proposed design can assemble spontaneously, it is important to test these predictions in several ways. Several interesting phenomena have surfaced in producing these assemblies.

For example, as DNA is added to either side of the four base-pair binding site in a linear DNA molecule, it has been shown that the rate of the coupling reaction between methyltransferases and the targeted 5-fluorocytosine residues increases in an apparently sigmoid fashion [8]. The initial inflection point in this reaction rate occurs at about 25 to 30 bp of total length for two unrelated bacterial cytosine methyltransferases (M•*Hha*I and M•*Msp*I). This was interpreted as the kinetic footprint of the enzymes (i.e., the minimum space required for unhindered approach by a methyltransferase to its targeted recognition sequence). However, the 3D structure of M•*Hha*I bound to DNA shows that it only physically covers about 13 bp of DNA [68], suggesting that it could be made to decorate DNA recognition sites that are closer then the 25 bp inferred from the kinetic experiments. This inference from the 3D structure turns out to have been correct, since three-point assembly along a linear molecule has been achieved with center to center placements as close as 17 bp [65].

3. DESIGNS

Although a considerable number of technical and structural feats have been achieved in bionanotechnology, an even larger number of designs for devices and assemblies based on addressable assembly have been proposed. See Table 1. Of these, many continue to be biomemetic with antecedents in prebiotic evolution. Proposals for macromolecular carcerands [69] and host–guest systems [70, 71] containing DNA form an important set of examples, since the protective nature of these assemblies touches upon one of the long-standing problems in prebiotic evolution: understanding how DNA ultimately became intertwined with the postulated RNA world. Given the rules that have emerged defining the attributes of the entity that first crossed the Darwinian threshold, it would appear that DNA was not required for the initial evolution of primitive supramolecular aggregates or metabolosomes [17, 18]. Nevertheless, the best evidence suggests that DNA was recruited early on in the evolution of these SMAs [16].

Recent discoveries on the structure of the supramolecular aggregates present in the macronucleus in protozoans and the nucleolus in primitive eukaryotes suggest that these structures may make use of the cohesive ends of DNA produced by telomerase. Telomerase is a molecular fossil with a structure resembling that of a ribosome, in that it comprises several proteins assembled on a functional RNA. In this case, however, one of the protein moieties serves to complement the RNA moiety in such a way that the supramolecular assembly becomes a primitive reverse transcriptase capable of generating short (often hexameric) DNA repeats that are capable of forming strong associations based primarily on hoogsteen pairing between guanine residues. Many of these associations can survive even in boiling water and thus can provide considerable integrity to a supramolecular structure. Moreover, in an RNA world, one can safely assume that competition between evolving supramolecular aggregates would involve the production of RNA degrading enzymes or ribozymes selected for their destructive power against competing metabolosomes. Thus, the evolution of a primitive ribonucleoprotein assemblage capable of generating a protective DNA cage or matrix that was impervious to RNA-degrading activities would offer an important selective advantage for such a system. In short, telomerase-like reverse transcriptases may have been both enveloped and protected by their DNA products in the distant past. This possibility is consistent with the association of the modern telomerases (as molecular fossils) with macronuclei [72, 73] and with the membraneless aggregate of nucleic acid and protein called the nucleolus [74, 75]. Of these the nucleoli of the lower eukaryotes (e.g., *D. Discodium* and *P. polycephalum*) [76, 77] may represent living molecular fossils that evoke the freestanding supramolecular aggregate postulated as the antecedent of modern living systems [16, 17]. The DNA contained in these nucleoli is a short linear element that ends in cohesive telomeric DNA and encodes the other important molecular fossil from the RNA world: the ribosomal RNA [77, 78]. As with other nucleoli these are membraneless structures that are enclosed in an apparently freestanding fashion by the nucleus. That DNA came to be associated with the RNA world in this fashion is also consistent with the findings of Ohno who pointed out that all modern genes appear to have evolved from iterated repetitive elements of the type produced by modern telomerases [79].

These systems form biomemetic models for nanoscale cages, carcerands, and extended 3D assemblies that can be constructed with the bionanotechnological tools for supramolecular assembly and addressing described above. The utility of such assemblies lies in their potential as devices that may some day acquire the complexity attributed to nanorobots that can be directed to perform complex tasks in a nanoscale environment [80]. Targeted drug delivery, selective cell destruction, selective cell remodeling, telemetered detection of cellular abnormalities [80], and the creation of new life forms [81] are often suggested as long-range goals for these technologies.

More immediate goals for these technologies revolve around their potential for contributing to understanding

Table 1.

Concept	Proposed	Achieved
DNA array addressing		[63]
DNA-adapter-linked protein ordering on linear DNA		[6]
Fusion protein ordering on linear DNA		[8]
Protein assembly on branched DNA		[67]
DNA-adapter-linked gold ordering on linear DNA		[11]
DNA directed gold particle aggregation to 2D lattice		[10]
DNA directed protein assembly on 3D lattice	[70]	
DNA directed silicon component addressing aggregation to 2D array	[9]	

biological molecules and their interactions. A key area here has been the desire to control solid state 3D assembly of macromolecules so as to permit structure determination by X-ray diffraction. To this end, designs for extended 3D crystals based on DNA cages with oriented guests [70, 71] have also been proposed in the hope that macromolecules that interact with DNA scaffolds could be forced to adopt regular crystalline arrays that would diffract to high resolution. These nucleic acid-based proposals are supplemented by proposals for the assembly of 3D crystals based on protein cages and extended structures [51]. As with the DNA technology described above, designs for cages and shells, double-layer rings, two-dimensional layers, and helical filaments have all been proposed based on the ordered assembly of protein domain fusions [51].

Designs for nucleoprotein cages have also been proposed [69]. Here closed structures take advantage of ordered placement of methyltransferase targeted protein donors (e.g., proline rich peptides) and methyltransferase targeted protein acceptors (SH3 domains). When donors are ordered along one set of branched or linear DNA scaffolds and acceptors are ordered along another set, self-assembly can be initiated by mixing the two sets of biostructures. Using this method, it should be possible to achieve assembly of relatively large DNA cages and extended assemblies akin to those that characterize the macronucleus and the nucleolus as noted above.

Ordered assembly using biomolecules on an even larger scale has also been proposed [9]. Here, one envisions a two-step process utilizing complementary oligodeoxynucleotides. In the first step a surface is decorated with single-stranded oligodeoxynucleotides in a predetermined pattern using well known techniques [63]. In the second step, etching techniques are applied to a separate silicon-on-insulator wafer. The wafer is etched with potassium hydroxide around elements protected on their surface by a mask of Au/Cr. These islands are then released from the silicon insulator layer with hydrofluoric acid and linked to thiols at the ends of synthetic oligodeoxynucleotides through interactions with the gold surface. The silicon elements can then to be moved in an electric field until they contact the complementary oligodeoxynucleotide in the previously prepared surface causing them to become fixed in a predetermined pattern [9]. In principle such a system would permit macroscale addressing for assembly of electronic devices.

Designs for ordering functional proteins so as to produce vectored chemical processes on the nanoscale have also been proposed [6–8, 82]. In this case, the implication is that ordered proximity for a set of enzymes or ribozymes performing a sequence biochemical reactions would produce advances in understanding the nature of vectored biological catalysis and its potential applications in signal amplification.

4. CONCLUSION

Bionanotechnology adapts not only the results (functional proteins and nucleic acids) but also the processes (e.g., directed evolution, supramolecular aggregate, and metabolosome construction) of molecular evolution. Molecular addressing systems, based on DNA complementarity and DNA–protein interaction selectivity, are now available for the ordered assembly of a variety of functional elements from biology. Molecular motors, DNA-based switches, DNA-based oscillators, enzymes, ribozymes, deoxyribozymes, gold particles, chromophores, fluorescence quenching agents, antibodies, aptamers, and nucleic acid binding proteins can all be ordered along nucleic acid scaffolds. The potential for construction of useful devices utilizing the extraordinary wealth of functionality made possible by ordering these elements is quite broad. Given the success of this paradigm it is reasonable to assume that clues from biology will continue to be applied to the construction of useful bionanotechnological devices. It is likely that the successful construction of these devices will shed new light on both the origin and nature of living things. Moreover, one can anticipate the possibility that one of these devices might itself cross the Darwinian threshold at some time in the future.

GLOSSARY

Aptamer A short nucleic acid composed of DNA or RNA that has been adapted by selection using directed evolution to fit into a molecular surface.

Biosensor A device designed to detect a specific biological molecule, system of biological molecules, or biologically produced signal.

Conformation space A three-dimensional vector space occupied by representations of the 3D shapes that can be assumed by a molecule.

Darwinian threshold The point at which a system can begin to evolve by self-replication variation and natural selection.

Directed evolution The process by which molecules are selected from a replicable library. Multiple rounds of selection and replication generate a desired product.

DNA Deoxyribonucleic acid. A naturally occurring polymer composed of deoxy ribonucleotides. Although commonly found as a double-stranded, right-handed helix it can adopt one-, three-, and four-stranded forms having a variety of shapes.

F1 ATPase A naturally occurring multisubunit complex of proteins capable of synthesizing ATP in living systems.

Hydrogen bond A weak polar bond formed between two electron-rich atoms, one of which is covalently bonded to a hydrogen atom.

Macronucleus A large inclusion found in the cells of certain living things that contains multiple copies of genes that are in constant use by the organism.

Membraneless structure A substructure often present inside a living cell that is not bounded by a lipid bilayer membrane.

Metabolosome A nucleoprotein assembly that has been designed or selected for functionally carrying out a series of chemical reactions.

Molecular fossil A molecule that has been preserved by evolution over eons of geologic time stretching to or nearly to the dawn of life.

Molecular machine A machine comprising an assembly of a small number of molecules.

Molecular modeling A computer-aided design process in which computational chemistry software programs are used to create models of macromolecules based on electronic structure, and molecular mechanics calculations coupled with 3D experimental data.

Mutation A change in the information storage mechanism of an evolving system that alters the information it contains but may or not alter the form or function of the system.

Nucleolus A substructure within the cell nucleus that is composed largely of the genes responsible for the production of ribosomal RNA.

RNA Ribonucleic acid. A naturally occurring polymer composed of ribonucleotides. Although commonly found in its single-stranded form, it can adopt two-, three-, and four-stranded forms having a variety of shapes.

Selection of whole bacteria Directed evolution in which bacteria and not macromolecules are used as the substratum for growth.

Transition state An intermediate and generally unstable state in a chemical reaction in which reactants are transformed into products resembling both states but identical to neither.

REFERENCES

1. C. S. Yun, G. A. Khitrov, D. E. Vergona, N. O. Reich, and G. F. Strouse, *J. Am. Chem. Soc.* 124, 7644 (2002).
2. N. C. Seeman, *Nanotechnology* 2, 149 (1991).
3. A. Gil, P. J. de Pablo, J. Colchero, J. Gómez-Herrero, and A. M. Baró, *Nanotechnology* 13, 309 (2002).
4. B. Yurke, A. J. Turberfield, A. P. Mills, Jr., F. C. Simmel, and J. L. Neumann, *Nature* 406, 605 (2000).
5. C. Mao, W. Sun, Z. Shen, and N. C. Seeman, *Nature* 397, 144 (1999).
6. C. M. Niemeyer, T. Sano, C. L. Smith, and C. R. Cantor, *Nucleic Acids Res.* 22, 5530 (1994).
7. S. S. Smith, Biological and Biomedical Science and Technology Division, ONR 34196-3, p. 161, 1995.
8. S. S. Smith, L. Niu, D. J. Baker, J. A. Wendel, S. E. Kane, and D. S. Joy, *Proc. Natl. Acad. Sci. USA* 94, 2162 (1997).
9. R. Bashir, *Superlattices Microstruct*. 29, 1 (2001).
10. C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, *Nature* 382, 607 (1996).
11. A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez, Jr., and P. G. Schultz, *Nature* 382, 609 (1996).
12. "Encyclopedia Britannica." William Benton, Chicago, 1960.
13. "Webster's Third New International Dictionary." Merriam-Webster, Springfield, 1986.
14. D. H. Kenyon and G. Steinman, "Biochemical Predestination." McGraw–Hill, New York, 1969.
15. G. Zubay, "Origins of Life: On Earth and in the Cosmos." Academic Press, San Diego, 2000.
16. S. A. Benner, A. D. Ellington, and A. Tauer, *Proc. Natl. Acad. Sci. USA* 86, 7054 (1989).
17. C. R. Woese, *Proc. Natl. Acad. Sci. USA* 99, 8742 (2002).
18. T. J. Gibson and A. I. Lamond, *J. Mol. Evol.* 30, 7 (1990).
19. D. R. Mills, R. L. Peterson, and S. Spiegelman, *Proc. Natl. Acad. Sci. USA* 58, 217 (1967).
20. R. Levisohn and S. Spiegelman, *Proc. Natl. Acad. Sci. USA* 63, 805 (1969).
21. C. Tuerk and L. Gold, *Science* 249, 505 (1990).
22. D. L. Robertson and G. F. Joyce, *Nature* 344, 467 (1990).
23. A. D. Ellington and J. W. Szostak, *Nature* 346, 818 (1990).
24. L. Gold, *Harvey Lect.* 91, 47 (1995).
25. L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas, and J. J. Toole, *Nature* 355, 564 (1992).
26. T. Kato, K. Yano, K. Ikebukuro, and I. Karube, *Nucleic Acids Res.* 28, 1963 (2000).
27. D. G. Attardi and G. P. Tocchini-Valentini, *RNA* 8, 904 (2002).
28. M. Vuyisich and P. Beal, *Chem. Biol.* 9, 907 (2002).
29. T. M. Tarasow, S. L. Tarasow, and B. E. Eaton, *Nature* 389, 54 (1997).
30. E. H. Ekland and D. P. Bartel, *Nature* 382, 373 (1996).
31. E. H. Ekland, J. W. Szostak, and D. P. Bartel, *Science* 269, 364 (1995).
32. P. A. Lohse and J. W. Szostak, *Nature* 381, 442 (1996).
33. S. S. Smith and L. Crocitto, *Molecular Carcinogenesis* 26, 1 (1999).
34. Y. Li and D. Sen, *Nat. Struct. Biol.* 3, 743 (1996).
35. P. Travascio, Y. Li, and D. Sen, *Chem. Biol.* 5, 505 (1998).
36. R. R. Breaker and G. F. Joyce, *Chem. Biol.* 2, 655 (1995).
37. N. Carmi and R. R. Breaker, *Bioorg. Med. Chem.* 9, 2589 (2001).
38. J. H. Campbell, J. A. Lengyel, and J. Langridge, *Proc. Natl. Acad. Sci. USA* 70, 1841 (1973).
39. B. G. Hall, *Genetics* 101, 335 (1982).
40. R. Birge, N. B. Gillespie, E. W. Izaguirre, A. Kusnetzow, A. F. Lawrence, D. Singh, Q. W. Song, E. Schmidt, J. Stuart, S. Seetharaman, and K. J. Wise, *J. Phys. Chem. B* 103, 10746 (1999).
41. K. Wise, N. Gillespie, J. Stuart, M. Krebs, and R. Birge, *Trends Biotechnol.* 20, 387 (2002).
42. J. H. Chen and N. C. Seeman, *Nature* 350, 631 (1991).
43. C. Mao, W. Sun, and N. C. Seeman, *Nature* 386, 137 (1997).
44. Y. Zhang and N. Seeman, *J. Am. Chem. Soc.* 116, 1661 (1994).
45. E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, *Nature* 394, 539 (1998).
46. V. P. Grantcharova and D. Baker, *Biochemistry* 36, 15685 (1997).
47. G. Siegal, B. Davis, S. M. Kristensen, A. Sankar, J. Linacre, R. C. Stein, G. Panayotou, M. D. Waterfield, and P. C. Driscoll, *J. Mol. Biol.* 276, 461 (1998).
48. A. C. Glasgow, M. F. Bruist, and M. I. Simon, *J. Biol. Chem.* 264, 10072 (1989).
49. D. L. D. Caspar and A. Klug, *Cold Spring Harb. Symp. Quant. Biol.* 27, 1 (1962).
50. I. Klotz, D. W. Darnall, and N. R. Langerman, in "The Proteins" (H. Neurath and R. L. Hill, Eds.), Vol. I, p. 293. Academic Press, Seattle, 1975.
51. J. E. Padilla, C. Colovos, and T. O. Yeates, *Proc. Natl. Acad. Sci. USA* 98, 2217 (2001).
52. P. D. Boyer, *Biochim. Biophys. Acta* 1140, 215 (1993).
53. J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker, *Nature* 370, 621 (1994).
54. T. M. Duncan, V. V. Bulygin, Y. Zhou, M. L. Hutcheon, and R. L. Cross, *Proc. Natl. Acad. Sci. USA* 92, 10964 (1995).
55. D. Sabbert, S. Engelbrecht, and W. Junge, *Nature* 381, 623 (1996).
56. H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita, Jr., *Nature* 386, 299 (1997).
57. R. K. Soong, H. P. Neves, and J. J. Schmidt, *Biomed. Micro.* 3, 71 (2001).
58. R. K. Soong, G. D. Bachand, H. P. Neves, A. G. Olkhovets, H. G. Craighead, and C. D. Montemagno, *Science* 290, 1555 (2000).
59. C. D. Montemagno, *Nanotechnology* 10 (1999).
60. C. M. Niemeyer, M. Adler, S. Gao, and L. Chi, *Angew Chem. Int. Ed. Engl.* 39, 3055 (2000).
61. C. M. Niemeyer, J. Koehler, and C. Wuerdemann, *Chembiochem.* 3, 242 (2002).
62. C. M. Niemeyer, L. Boldt, B. Ceyhan, and D. Blohm, *Anal. Biochem.* 268, 54 (1999).
63. M. J. Heller, *Annu. Rev. Biomed. Eng.* 4, 129 (2002).
64. S. S. Smith, L. Niu, D. J. Baker, J. A. Wendel, S. E. Kane, and D. S. Joy, *Miami Biotechnol. Short Rep.* 8, 13 (1997).
65. S. S. Smith, *Nano Letters* 1, 51 (2001).
66. J. A. Wendel and S. S. Smith, *Nanotechnology* 9, 297 (1998).

67. A. Laayoun and S. S. Smith, *Nucleic Acids Res.* 23, 1584 (1995).
68. S. Klimasauskas, S. Kumar, R. J. Roberts, and X. Cheng, *Cell* 76, 357 (1994).
69. S. S. Smith, *Nanotechnology* 13, 413 (2002).
70. N. C. Seeman, *DNA Cell Biol.* 10, 475 (1991).
71. N. C. Seeman, *Annu. Rev. Biophys. Biomol. Struct.* 27, 225 (1998).
72. T. M. Bryan, J. M. Sperger, K. B. Chapman, and T. R. Cech, *Proc. Natl. Acad. Sci. USA* 95, 8479 (1998).
73. K. G. Murti and D. M. Prescott, *Chromosome Res.* 10, 165 (2002).
74. K. T. Etheridge, S. S. Banik, B. N. Armbruster, Y. Zhu, R. M. Terns, M. P. Terns, and C. M. Counter, *J. Biol. Chem.* 277, 24764 (2002).
75. J. S. Andersen, C. E. Lyon, A. H. Fox, A. K. Leung, Y. W. Lam, H. Steen, M. Mann, and A. I. Lamond, *Curr. Biol.* 12, 1 (2002).
76. A. F. Cockburn, W. C. Taylor, and R. A. Firtel, *Chromosoma* 70, 19 (1978).
77. V. M. Vogt and R. Braun, *J. Mol. Biol.* 106, 567 (1976).
78. S. Johansen, T. Johansen, and F. Haugli, *Curr. Genet.* 22, 305 (1992).
79. S. Ohno, *Semin. Cell Dev. Biol.* 10, 517 (1999).
80. R. J. Feitas, "Nanomedicine: Basic Capabilities." Landes Bioscience, Georgetown, 1999.
81. S. J. Sowerby, N. G. Holm, and G. B. Petersen, *Biosystems* 61, 69 (2001).
82. C. M. Niemeyer, B. Ceyhan, S. Gao, L. Chi, S. Peschel, and U. Simon, *Colloid Polym. Sci.* 279, 68 (2001).