Introduction to Self-Assembling DNA Nanostructures for Computation and Nanofabrication.

Thomas H. LaBean
Department of Computer Science
Duke University

[PREPRINT. NOT for DISTRIBUTION. To be published 2002 by World Scientific. CBGI 2001 Proc.]

Introduction.

DNA, well-known as the predominant chemical for duplication and storage of genetic information in biology, has also recently been shown to be highly useful as an engineering material for construction of special purpose computers and micron-scale objects with nanometer-scale feature resolution. Properly designed synthetic DNA can be thought of as a programmable glue which, via specific hybridization of complementary sequences, will reliably self-organize to form desired structures and superstructures. Such engineered structures are inherently information-rich and are suitable for use directly as computers or as templates for imposing specific patterns on various other materials. In theory, DNA can be used to create any desired pattern in two or three dimensions and simultaneously to guide the assembly of a wide variety of other materials into any desired patterned structure. Given diverse mechanical, chemical, catalytic, and electronic properties of these specifically patterned materials, DNA self-assembly techniques hold great promise for bottom-up nanofabrication in a large number of potential applications in wide ranging fields of technology. Starting with background for understanding why the physical, chemical, and biological properties of DNA make it extremely useful as a "smart" material for nanoengineering projects, this chapter traces the historic development of DNA-based nanofabrication, outlines its major successes, and presents some possible future applications in fields as diverse as electronics, combinatorial chemistry, nano-robotics, and gene therapy,.

DNA-based nanoengineering as a field is related to computational biology, bioinformatics, and genome informatics rather tangentially; it is more closely allied with biomolecular computation (BMC) – the engineering of biological macromolecules for production of artificial information processing systems. Rather than using binary, electronic computers for analyzing information extracted from biological systems, BMC seeks to utilize biomolecules directly as active parts of engineered computers. The concluding section of this chapter contains some speculation into the possibility of coming full circle and applying BMC and DNA-based nanoengineering principles and systems to the extraction and processing of information directly from biological DNA, that is, the possible use of natural DNA molecules as inputs for artificial DNA-based machines.

Background.

Chemistry and Biology of DNA. DNA (deoxyribonucleic acid) is a linear polymer whose monomeric residues are made up of one sugar group (deoxyribose), one phosphate group, and one nitrogenous base (either adenine, cytosine, guanine, or thymine; designated A, C, G, and T, respectively). Details of the structure are available in any biochemistry or molecular biology textbook, but a few pertinent points will be mentioned here. First, neighboring residues are joined by a chemical bond between the nth phosphate and the $(n+1)^{th}$ sugar group such that a polymeric backbone is formed of alternating sugar and phosphate groups. The backbone has chemical directionality due to asymmetry in the placement of phosphate groups on the sugar, with each sugar having one phosphate bound to its 5' carbon and one phosphate bound to its 3' carbon. This asymmetry gives the entire polynucleotide chain two distinct ends – the 5' and the 3', as shown in Figure 1. Two DNA strands hybridize (form hydrogen bonds) to one another in anti-parallel fashion, thus the 5' end of one strand points toward the 3' end of its complementary strand in the famous Watson-Crick double-stranded form (or double helix).

The second pertinent point regarding the chemical nature of DNA is that the nitrogenous bases (or simply bases) form hydrogen bonded pairs in tongue-and-groove fashion providing specificity of annealing. The base groups decorate the sugar-phosphate backbone with regular spacings and provide the physico-chemical energy which zips the DNA together in its predictable helical structure. In double-

helical DNA (or double-strand DNA, abbreviated to dsDNA), G bases pair specifically with C residues and A bases pair with T bases. G and C are said to be *complementary*, as are A and T. DNA strands of exact Watson-Crick complementarity will form stable hydrogen-bonded structures under standard temperature and solution conditions (see Figures 1 and 2). Some alternative base pairings have been found to form fairly stable hydrogen bonding [see for example Peyret *et al*, 1999], however, careful design of the sequences, as well as very slow annealing protocols, can successfully avoid alternative pairings and ensure that perfectly complementary strand matchings are highly favored.

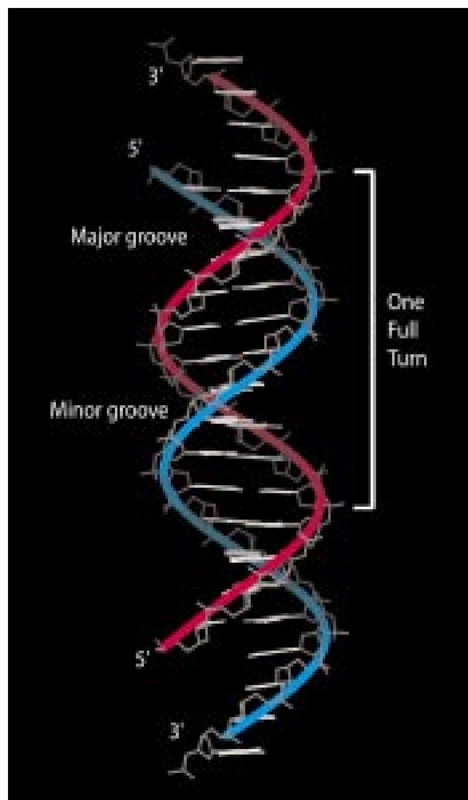


Figure 1. *Double-stranded DNA shown in the standard, right-handed, B-form double helix with four base ssDNA sticky-ends appended to the 3' ends of both strands. Strand backbones are highlighted with colored ribbons; bases (light gray) are viewed edgewise and can be seen to point toward their hydrogen bonding partner on the opposite strand. One full turn of DNA has a length of 3.4 nm along the vertical helix axis and contains on average 10.5 bases; the helix diameter is approximately 2 nm. The concave faces of the helix are known as the major and minor grooves; they are geometrically distinct and can be used to identify strand polarity -- for example, when looking into the minor groove, the strand on the bottom (in this orientation) always has its 3' end pointing down (toward the bottom of the page). Understanding the geometric constraints of DNA structure is essential to successful design of DNA-based objects and materials.*

The third important point stems directly from the exceptional stability and specificity of dsDNA. If a short segment of single-strand (ssDNA) is appended to a longer strand which participates in a double-helical domain, the ssDNA will act as a "smart glue", binding specifically to a complementary ssDNA segment located on another ds-domain. These ssDNA segments are known as *sticky-ends*. Complementary sticky-end pairs therefore act as address labels and can be used to specify which dsDNA domains are allowed to anneal to one another.

Finally, the "folding rules" which dictate the three-dimensional (3D) structure of DNA in solution are simple compared to other biological macromolecules, making DNA a more salutary engineering material than proteins, for example, whose folding rules have yet to be completely understood. Given proper pH and cation concentration, dsDNA will reliably adopt standard B-form helical structure with predictable dimensions as shown in Figure 1. In summary, important points of DNA chemistry include: anti-parallel alignment of backbones in hybridized strands, base-pairing specificity for high-fidelity annealing of sequences to their complements, and annealing by heating and slow cooling for double helix formation.

The task of engineering specific physical structures from DNA benefits from the tools evolved during the eons of biological evolution on Earth and especially from those now thoroughly researched and commercialized during the more recent biotechnological revolution. Enzymes can be purchased which perform highly specific chemical reactions upon DNA molecules. For example, *phosphatases* and *kinases* remove, add, and exchange phosphate groups from DNA backbones; *ligases* stitch together breaks in the backbone to form a single chemical strand from two or more shorter strands; and *restriction endonucleases* cleave the backbone at specific sites dictated by local base sequence. In addition, chemical synthesis methods for the production of DNA have advanced to the point where DNA strands of

any desired sequence can be ordered on-line from commercial production companies and shipped the next day for less than a dollar per residue.

DNA as a Structural Material. Since the publication of the 3D structure of dsDNA half a century ago [Watson and Crick, 1953], the vast majority of research on DNA structure has centered around DNA as it relates to known biological systems. However, twenty years ago Nadrian Seeman recognized the inherent potential of DNA as an engineering material and proposed visionary new uses for the polymer [Seeman, 1982]. Seeman's pioneering work originally focused on the creation of regular 3D lattices of DNA which could be used as scaffolding for the rapid, orderly binding of proteins to speed the formation of suitable crystals for 3D protein structure elucidation in x-ray diffraction studies.

Seeman noted that linear dsDNA can interact with only two other double-helices since it can display at most two sticky-ends, i.e. its maximum valence is two. Construction materials with valence = 2 are only really useful for making linear superstructures like railroad cars connected in a long train. A larger variety of substructures and an ability to interact with a greater number of neighboring components is required in order to advance even modest fabrication goals. Seeman pointed out that DNA in biological systems can exhibit structures with increased valence including replication forks (valence = 3) and Holliday junctions found in genetic recombination (valence = 4) as shown in Figure 2. One problem with these natural multivalent structures is that they involve repeated base sequences, so base-pairing partners are not perfectly specified and the junctions are mobile. The junctions, or strand crossover points, between the dsDNA domains are free to migrate up or down the helices by swapping one perfect sequence match for another perfect sequence match (see the right-hand drawing in Figure 2, if the top helix is pulled up while the bottom helix is pulled down, the left and right helices will become shorter as the top and bottom helices become longer). Seeman worked out a sequence symmetry minimization strategy in order to form, for the first time, immobile junctions - - branch points in the dsDNA which are unable to migrate up and down the helix. Note that the oligonucleotides are still normal, linear DNA polymers; the branch junctions occur in the arrangement of strand exchange crossovers between the double helical arms. Seeman has pioneered the use of branched DNA structures for the construction of geometric objects, knots, and Borromean rings [Chen *et al*, 1989; Chen & Seeman, 1991; Zhang & Seeman, 1994; Du & Seeman, 1994]. These early construction projects yielded many important technical developments including the use of oligonucleotide assemblies bound to insoluble resin beads for control of construction.

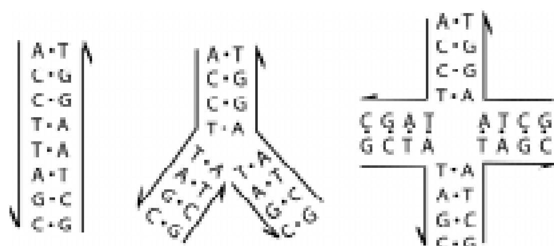


Figure 2. *Representations of unbranched, 3-branched, and 4-branched DNA. Fully base-paired, anti-parallel DNA can take on various forms depending upon the lengths and connectivities of the annealed strands. Normal double-helical DNA (left) involves two strands in a single helical domain. A 3-branched junction (center) involves four strands and three helical domains; it is a structural analog of a replication fork observed in biology. The 4-branched junction (right), involving four strands and four helical domains, is a structural analog of the Holliday junction used by biological systems in genetic recombination. Note that, in the branched structures shown, alternative base-pairings are available due to sequence symmetry around the branch point which will allow the junction to migrate up and down the helices. Properly designed sequences avoid such migration. 4-branch junctions have been used most extensively in engineered tile structures. Their four helical domains tend to stack into two domains in which two strands exchange between helices (as explained further in the next figure).*

One problem with many early DNA constructs was that the structural flexibility of the branched DNA complexes allowed undesired circular products to be formed during assembly of large superstructures from stable substructures. Again, innovation from Seeman's lab solved the problem by producing double-crossover (DX) complexes which act as rigid structural components for assembly of larger superstructures. The concept has now been extended further to produce more complex structures including triple-crossovers (TX) as shown in Figure 3. This class of DNA objects, often referred to as

'tiles', contain multiple oligonucleotide strands (ssDNA) which base-pair along parallel, coplanar helix axes. The helices are connected by exchange of two strands at each crossover point (crossovers are structural analogs of Holliday junctions). Rigid and thermally stable, these multi-helix tiles carry multiple, programmable sticky-ends for encoding neighbor relations to dictate tile-to-tile interactions used in specific assembly of patterned superstructures. DNA tiles are formed by heating an equimolar solution of linear oligonucleotides above 90° C to melt out base-paired structures, then slowly cooling the solution to allow specific annealing to form the desired structure. Tiles are stabilized in solution by the presence of magnesium counter ions (Mg^{++}) which allow close helix packing by shielding the negative charges on the DNA backbones from one another.

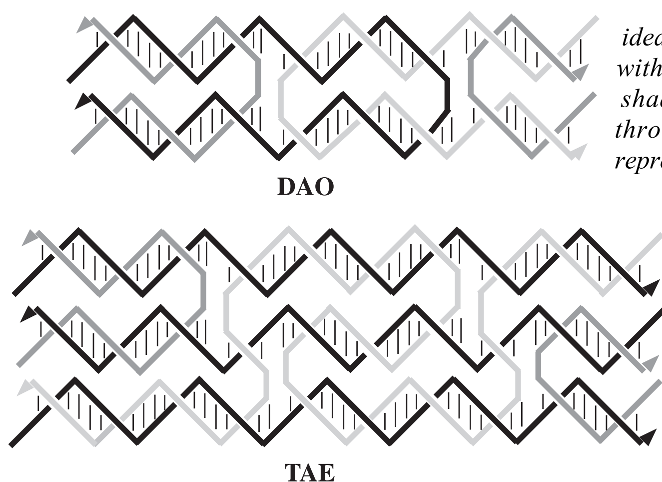


Figure 3. *Example DX and TX tiles drawn as an idealized projection of 3D helices onto the plane of the page with helix axes lying horizontal on the page. Strands are shaded for ease of tracing individual oligonucleotides through the complexes. Each straight strand segment represents a half-turn around the helix. Vertical segments of strands indicate strand exchange (junction) sites where strands cross over from one helix to another. Note that two strands are exchanged at each crossover point. Arrowheads indicate 3' ends of strands. Thin vertical hashes indicate base-pairing between strands. Unpaired segments on 5' ends represent sticky-ends. The top complex is a DAO double-crossover, so called because of its Double (two) ds-helices, Anti-parallel strand exchange points, and Odd number of helical half-turns between junctions. The bottom complex is a TAE (Triple, Anti-parallel, Even number of half-turns between crossovers). Anti-parallel crossovers cause strands to reverse their direction of propagation through the complex upon exchanging helices. For example, the lightest gray strand in the DAO begins in the right-hand side of the top helix; it propagates left until it crosses over to the bottom helix, then it continues back to the right until it reaches the right-hand end of the tile. The effect of spacing between crossover points can be seen by comparing the strand trace of the DAO with that of the TAO. The TAO contains three strands (black) which span the entire width of the tile; they are the non-exchanging strands at each of the crossover points. With an odd number of half-turns between crossovers (see DAO), no strands span the width of the tile. Many other strand topologies are possible; these shown and several others have been experimentally tested. Note that the figure also shows how the minor groove of one helix is designed to pack into the major groove of neighboring helices.*

Design of DNA tiles and superstructures requires two separate phases: first, geometric design and second, chemical or sequence design. The geometric design phase involves modeling and examination of strand topology (paths of the oligonucleotides through the tiles), spacing of crossover points to ensure proper orientation of neighboring helical domains (for example, to ensure flatness of 2D lattices), lengths of sticky-ends, and overall internal compatibility of components with each other and the superstructure design. Once the geometric constraints of the target structure are established, specific base sequences can be designed which guarantee formation of the desired structure.

Design of Base Sequences for DNA Nanoconstruction. To properly design base sequences of DNA for nanoassemblies, one must consider positive as well as negative design constraints: a sequence must not only match its desired hybridization site, but it must also hold no significant complementarity to any other DNA segment, thus avoiding formation of undesired alternative structures. Many approaches and strategies for sequence design have been pursued [see for example, Seeman, 1990; Baum, 1996; Deaton *et al*, 1996; Marathe *et al*, 2000; Reif *et al*, 2001]. Primary among design constraints is Hamming distance: no sequence can be included which contains more than some threshold number of exact matches with any other sequence or the complement of any sequence already contained in the set. Thresholds are chosen based on the lengths of sequences required and known limitations from hybridization experiments.

An example constraint might require at least three mismatches between every pair of subsequences of length eight. For longer strands, a sliding window is used to tabulate all subsequences of a given sequence. Such search and design problems require the use of electronic computers to keep track of the huge number of possibilities; therefore, custom software has been developed by several research groups to find good solutions to combinatorial optimization of sequence design. Besides Hamming distance, other design criteria include exclusion of certain undesired subsequences for example, palindromes which may form undesired hairpins, long stretches of G and C which, due to stronger base stacking interactions may distort the structure away from standard B-form double helix. Often, homogenization of base composition within and between strands is desirable in order to increase the likelihood of isothermal annealing. If individual regions of the structure have similar base composition they will have similar melting temperatures and formation of all parts of a tile will occur nearly simultaneously during the cooling process. Careful sequence design is critical for successful assembly of complex objects from synthetic DNA oligonucleotides since base-pair formation is the driving force of the self-organization process.

Experiments and Applications.

DNA-based Computation. The first experimental proof of the feasibility of DNA-based computing came from Adleman, when he used DNA to encode and solve a simple instance of a hard combinatorial search problem [Adleman, 1994]. He demonstrated the use of artificial DNA to generate all possible solutions to a Hamiltonian path problem (given a set of nodes connected by a set of one-way edges, answer the question of whether or not there exists a path which goes through each node once and only once). For large graphs, the problem can be very difficult for an electronic computer to solve since there are an astronomical number of possible paths and there is no known algorithm for finding the correct answer. Adleman's approach was to assign a 20-base DNA sequence to each node in an example graph, then to synthesize edge strands containing the complement to the 3' half of a starting node fused with the complement to the 5' half of the ending node for each valid edge in the graph. The sets of oligonucleotides encoding nodes and edges were annealed and ligated, thereby generating long DNA strands representing all possible paths through the graph. Non-Hamiltonian paths were then discarded from the DNA pool, first by size separation of the path DNA (discard strands greater than or less than the length of a Hamiltonian path, which is equal to the product of the number of nodes times the length of the node sequence), and second by a series of sequence-based separation steps involving DNA probes complementary to each node sequence (discard path sequences if they failed to contain any one of the required nodes). By this experimental protocol, Adleman was able to recover DNA strands encoding the Hamiltonian path through the example graph.

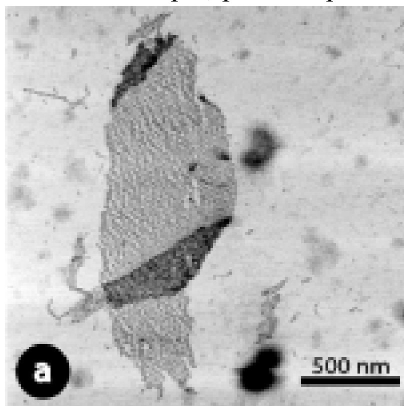
The primary contributions of Adleman's seminal paper were the revolutionary concepts that synthetic DNA could be made to carry information in non-biological ways and that the inherent massive parallelism of molecular biology operations could be harnessed to solve computationally hard problems. His experiment showed that DNA could be used as an integral part of a functioning computer. Since that time, some limits have been noted on the size of combinatorial search problems which can be implemented in DNA because of the exponential growth of search spaces and the volume constraints on wet computing techniques [Reif, 1998]. In addition to volume constraints, Adleman's original algorithm involved rather inefficient and tedious laboratory steps, the total number of which increased at least linearly with problem size. These concerns have been sidestepped by more recent theoretical and experimental advances including the development of computation by self-assembly.

Algorithmic Self-Assembly. Another fundamental insight which has shaped understanding of DNA-based computing and nanoengineering was made by Winfree when he realized that DNA annealing by itself and, specifically, annealings between DNA complexes being developed by Seeman were capable of carrying out computation [Winfree *et al*, 1998; Winfree, 1998]. This line of reasoning, developed theoretically and experimentally by Winfree in collaboration with Seeman and others, follows a theoretical model of computing known as Wang tiling [Wang, 1961]. In the Wang tiling model, unit

square tiles are labeled with symbols on each edge such that tiles are allowed to associate only if their edge symbols match. Tiling models have been designed which successfully simulate single-tape Turing Machines and are therefore capable of universal computation [Berger, 1966; Robinson, 1971; Wang, 1975]. The recognition that DNA tiles, exemplified by DX and TX complexes (see Figure 3), could represent Wang tiles in a physical system, where edge symbols are incarnated as sticky-ends, led to proofs that DNA tilings are capable of universal computation.

Computation by self-assembly of DNA tiles is a significant advance over earlier DNA-based computing schemes because self-assembly involves only a single-step in which the computation occurs during the annealing of carefully designed oligonucleotides. Contrast this with Adleman's experiment in which the annealing step generated all possible solutions and where a long series of laboratory steps was required to winnow the set by discarding incorrect answers. Self-assembly without errors will theoretically only allow formation of valid solutions during the annealing step, thereby eliminating the laborious phase involving a large number of laboratory steps. The first report of a successful computation by DNA self-assembly demonstrated example XOR calculations [Mao et al, 2000]. XOR, an addition operation without the carry-bit, was performed using tiles carrying binary values (1 or 0) to specifically assemble an input layer which then acted as a foundation upon which output tiles assembled based on the values encoded on the input tiles. The prototypes also demonstrated the use of readout from a reporter strand which was formed by ligation of strands carrying single bit-values from each tile in the superstructure. The scheme is currently being extended to harness the massively parallel nature of the annealing reaction by allowing random assembly of the input layers, followed by specific assembly of the output layers in order to simultaneously compute the entire lookup table for pairwise XOR (and eventually addition) up to some modest input length (perhaps 20 bits) [details described in LaBean *et al*, 2000a].

Patterned DNA Nanostructures. Programmed self-assembly of DNA objects promises further advances not only in biomolecular computation but also in nanofabrication as a means of creating complex, patterned structures for use as templates or scaffolds for imposing desired structures on other materials. Simple, periodic patterns have been successfully implemented and observed on superstructures formed from a variety of different DNA tiles including DX tiles [Winfree *et al*, 1998b], TX tiles [LaBean *et al*, 2000], triangular tiles [Yang *et al*, 1998], and rhombus-like tiles [Mao *et al*, 1999]. Figure 4 shows 2D lattice constructed from two types of TX tiles, A and B*, where the B* tiles display two extra dsDNA stem-loops (hairpins) protruding out of the tile plane, one each on the top and bottom faces of the tile. Sticky-ends on the four corners of each tile program neighbor relations such that A tiles only bind to B* tiles and vice versa resulting in the observed stripe pattern. Large lattice superstructures formed by such systems have been observed (at least 10 microns by 3 or 4 microns and containing tens of thousands of tiles).



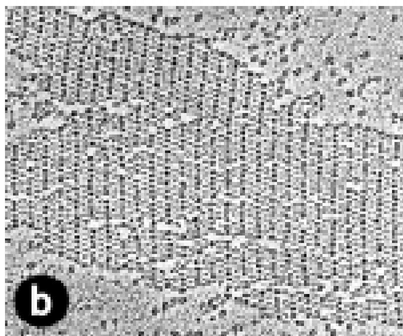
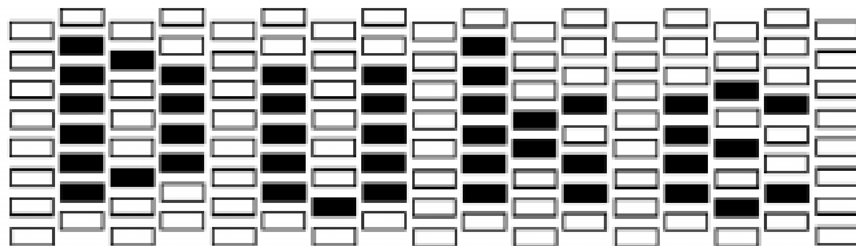


Figure 4. *TX tile lattices formed by annealing eight strands and visualized by atomic force microscopy (AFM -- panel a) and transmission electron microscopy (TEM -- panel b). Lattice displaying periodic patterns (stripes in this case) was designed using two types of TX tile, A and B*. The B* tiles contained an extra hairpin of DNA projected out of the lattice plane on each side of the tile. A tiles bind only to B* tiles and vice versa by virtue of properly coded sticky-ends at the four corners of each tile. The hairpins impart distinct features which can be microscopically observed. The TEM sample (panel b) was prepared by platinum rotary shadowing resulting in the B* tiles' extra hairpin causing them to take on a darker color than the A tiles.*

Larger tiles sets with more complicated association rules are currently being developed for the assembly of aperiodic patterns which will be used in the fabrication of patterned objects useful for nanotechnology applications (examples are given in Figure 5). 2D tile arrays can be thought of as molecular fabric or tapestry which contain a large number of addressable pixels. Individual tiles can carry one or more pixels depending upon the placement of observable features or binding sites. Overall connectivity can be programmed either with unique sticky-ends defined for each tile in the array or by assembly of crossover junctions which specifically stitch together distant segments of a single long scaffold strand as shown in Figure 6.



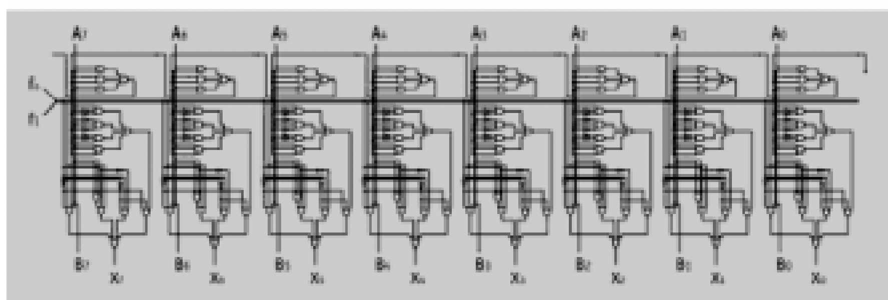


Figure 5. *Examples of simple and complex aperiodic structures as possible fabrication targets for DNA-based self-assembly. A relatively simple aperiodic structure such as writing a word in addressable pixels on a DNA tile array (top) would help improve methods for eventual assembly of very complex structures such as entire circuit layouts (bottom).*

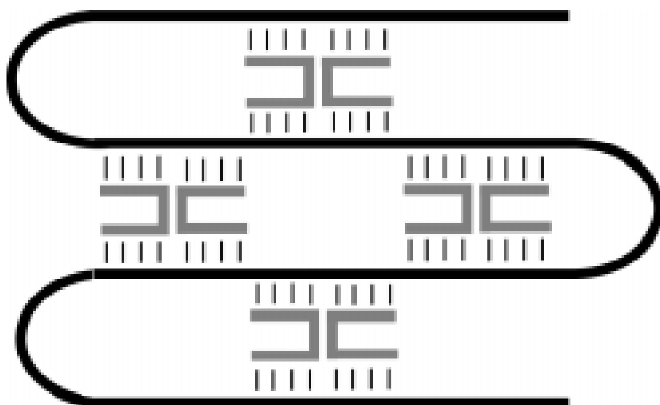


Figure 6. *Schematic of a grid structure formed by annealing specific short oligonucleotides (gray) onto a preexisting long ssDNA (black). 2D arrays might be assembled not only from short synthetic strands but also making use of the larger-scale connectivity information available in long strands of ssDNA. Pixels on such a lattice would be individually addressable by virtue of their specific ordering along the large scaffold strand. Assembly of multi-tile superstructures around input scaffold strands of moderate size has been demonstrated [LaBean et al, 1999]. The possibility of using very long ssDNA from biological sources is currently being investigated.*

Computer simulations and theoretical analysis of self-assembly processes have pointed to some potential difficulties including the possibility of assembly errors leading to trapping of incorrectly formed structures [Winfree, 1998; Rothemund, 2000; Reif, 1998]. An experimentally observed error rate of 2-5%, encompassing annealing and ligation errors, was noted in the XOR computational complex [Mao *et al*, 2000]. Several approaches exist to address such issues including more complicated annealing schedules, variable length sticky-ends for non-isothermal tile associations, and stepwise assembly controlled by time-stepped addition of critical oligonucleotide components. Readout methods which sample an ensemble of reporter strands as well as error-tolerant designs for the overall system are also being developed.

Patterned Immobilization of Other Materials on DNA Arrays. Implicit in the preceding discussion of DNA self-assemblies as templates for specific patterning of other materials is the need for attachment chemistries capable of immobilizing these materials onto DNA arrays. Materials of interest might include metal nanoparticles, peptides, proteins, other nucleic acids, and carbon nanotubes among others. A variety of strategies and chemistries are being developed including thiols (-SH), activated amino groups, biotin-avidin association, and annealing of pre-bound, complementary DNA. Oligonucleotides, chemically labeled with a thiol group on either the 5' or the 3' end readily bind to gold and have already been used via simple complementary DNA annealing to impart 3D ordering on gold nanospheres [Alivisatos *et al*, 1996; Mirkin *et al*, 1996; Mucic *et al*,1998] and gold nanorods [Mbindyo *et al*, 2001]. In those studies, gold was labeled with multiple copies of a single DNA sequence, then linear dsDNA was formed between complementary strands attached to adjacent gold particles. More specific chemistries are available including nanogold reagents which make use of 1.4 nm diameter gold clusters, each functionalized with a single chemical moiety for specific reaction with a thiol or a free amino group (Nanoprobes, Inc., Yaphank, NY). These reagents have been used to target the binding of single gold nanoparticles to specific locations on DNA nanoassemblies. Figure 7 shows preliminary results of targeted binding of nanogold to a filamentous DNA tile superstructure, followed by deposition of silver onto bound gold for the fabrication of nanometer scale (~50 nm diameter) metallic wires. A similar technique has been reported for construction of a conducting silver wire on a length of ssDNA [Braun *et al*, 1998]. Ongoing studies focus on formation of smaller (~10 nm diameter) metal wires laid out in specific patterns on 2D tile lattices. The long-term goal of these metalization studies is the self-assembly of electronic components and circuits at length scales below those available by lithography techniques.

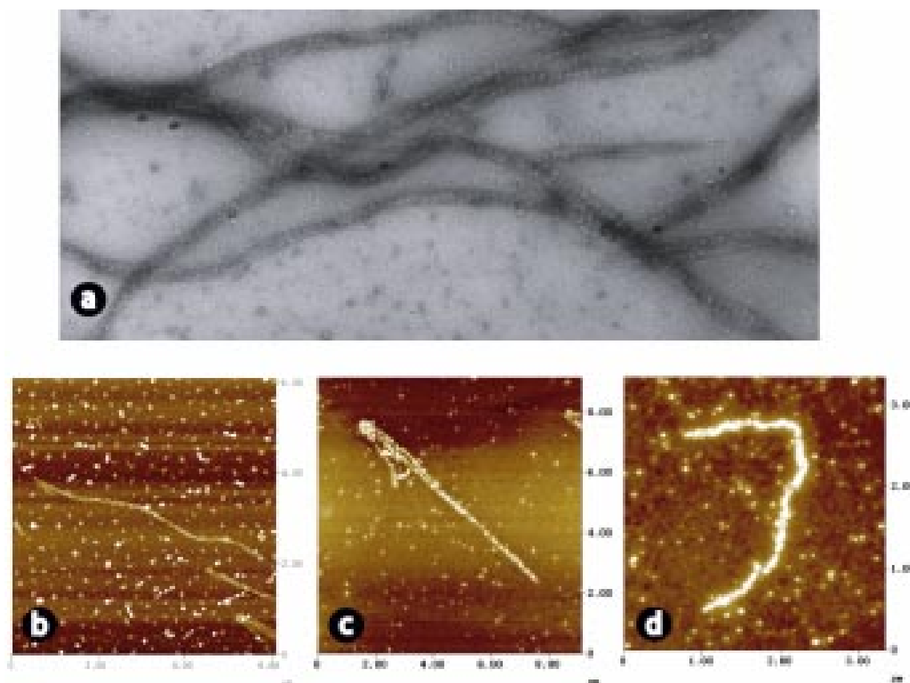


Figure 7. *Targeted metallization of a complex DNA superstructure. a). Filaments of DNA lattice constructed from AB* tiles as in Figure 4 but with the addition of two thiol (-SH) groups to the dsDNA stem protruding from one side of the B* tile and one amino group on the end of the dsDNA stem protruding from the other side of the tile. It appears that the thiol sulfurs associate with one another causing the lattice to curve and form tubes of quite uniform diameter. Experiments are ongoing to further clarify details of the structure. b). Same as a) with addition of 1.4 nm nanogold targeted to the amino groups on the protruding DNA stem. c). Same as b) with addition of 2 minute development with a silver enhancement procedure which deposits silver upon existing bound gold particles. d). Same as b) but with 5 minute silver enhancement. Progressive build-up of metal atoms is observed, with perhaps a few more minutes of silver binding required to form a complete, conductive wire. Note that these DNA filaments still have sticky-ends available at both ends which can be used for orienting the entire filament prior to metal binding.*

A novel approach to targeted binding which has yet to be experimentally tested is the display of "aptamer" domains, which have been artificially evolved for specific binding of antibodies (immunoglobulin proteins) to DNA or RNA [Tsai *et al*, 1992]. Techniques have been developed for *in vitro* selection of specific nucleic acid/antibody pairs. The antibody can be utilized as an adapter molecule, binding not only to its DNA epitope displayed on a 2D lattice but also to another protein of interest (this scheme will be further developed below). The well-known association between biotin and avidin has also been shown to be useful for targeted binding of the streptavidin protein to DNA lattice carrying an oligonucleotide labeled with the small biotin molecule [Winfree *et al*, 1998b]. The development of these and other attachment strategies has just begun. Many advances and new insights can be expected.

Summary and Future Directions.

The field of self-assembling DNA nanofabrication has already yielded successes on several fronts including binary computation, periodic tilings in two dimensions, and targeted immobilization of metallic nanoparticles. DNA has been shown to be well-suited for programmed construction of micron-scale objects with nanometer-scale feature resolution. Eventually, DNA-based self-assemblies may serve a critical role in the pattern formation step of electronic circuit fabrication, outperforming lithography by creating thousands or millions of copies of a desired structure simultaneously and at length-scales unavailable with current production techniques. Known hurdles which must be overcome include reduction of the error rate of strand hybridization, positioning of DNA objects relative to macroscopic contacts, and successful construction of complex, aperiodic patterns by algorithmic assembly.

Alternative chemistries should also be explored for backbone and bases. For example, inosine, which is able to pair with any of the four standard bases, might be useful for promiscuous annealing in some applications. Also, artificial bases with specific pairing partners could be incorporated in order to increase the information density of the polymers [Tae *et al*, 2001]. More stable backbone variants might be investigated such as PNA, which contains the normal DNA bases linked via peptide backbone chemistry and readily forms DNA-like double helix [see for example Hanvey *et al*, 1992]. DNA has been exploited primarily because techniques for very specific manipulations of its base sequence and backbone connectivity have been perfected for use in recombinant molecular genetics for biotechnology applications. However, other polymers with programmable interactions might be more suitable in the long-run for some nanofabrication applications.

Possible Future Applications. Some possible fields of application for future DNA nanotechnologies might include electronic circuit lay-out, organization of materials for batteries or flat panel displays, macromolecular patterned catalysts for chemical assembly lines, combinatorial chemistry, sensorless sorting of nanometer-scale objects, DNA sequence comparison, and perhaps gene therapy.

• Electronics and Chemistry. DNA self-assemblies may find uses not only in templating nanometer scale electronic circuits alluded to in preceding sections but also in preparation of patterned catalyst arrays. For example, nanoparticulate metals used to catalyze the formation of single-walled carbon nanotubes have previously been used when randomly distributed in aerogels [Su *et al*, 2000]. If attachment chemistries can be adapted for the binding of such nanoparticles to DNA tile lattices, then coordinated synthesis of ordered arrays of carbon nanotubes might be possible. Such ordered nanotube arrays might be useful in advanced electrical storage batteries, flat panel displays with ultra-fine pixel density, or very strong, multi-tube fibers. This approach is especially attractive because current synthesis methods generally yield tangled masses of nanotubes which have been difficult to sort and organize. Other target catalysts include protein enzymes or surface catalysts which, when ordered in series, could act as macromolecular chemical assembly-lines. Patterned stripes of catalysts could act sequentially to carry out a sequence of specific reactions or even repeated cycles of reactions on a stream of substrate flowing past.

• Combinatorial Chemistry. Brenner and Lerner proposed the use of DNA for tagging chemical compounds with specific labels [Brenner and Lerner, 1992] for use in combinatorial chemistry. They suggested that DNA labels could be decoded to reveal the identity of active molecules drawn by a

screening assay from a vast pool of candidate chemicals. It is possible that DNA tile structures could be used further to hold chemical reactants close together in space, thereby facilitating their reaction. The product of the reaction would remain bound to the tile, decoding of each strand of the tile would reveal the identity of each reactant used in the formation of active compounds. Encoding labels for reactants rather than final compounds would decrease the number of specific labels required.

• Sensorless Sorting. DNA tile lattices specifically decorated with protein rotary motors or environmentally responsive peptides might prove useful for sensorless sorting of poorly soluble nano-scale objects such as "buckyballs" or fragmented carbon nanotubes. Sensorless sorting involves an array of effectors capable of repetitive motion which act to organize objects into specific orientations and move them along a path comparable to a conveyor belt. Carbon nanotubes might be an interesting target object for sorting because they are poorly soluble in aqueous solution and they are difficult to purify and sort yet they are objects of intense study due to their unique structural and electronic properties. A possible scenario might involve a DNA array acting to organize a set of protein rotary motors which then provide a sweeping motion to coax nanotubes into alignment and feed them down a channel. Such an elaborate system could prove useful for simultaneously orienting large numbers of carbon nanotube into position for use as wires in a circuit, for example.

• General Nanofabrication. Self-assembling DNA-based structures also hold great potential in "seeding" for the autonomous growth of complex structures by bottom-up nano-fabrication. A molecular machine built of and fueled by DNA has been demonstrated experimentally [Yurke *et al*, 2000]. The technique introduces the possibility of setting up a cascade of annealing reactions which, once begun, run sequentially without further intervention, and result in formation of a complex structure inaccessible by simple annealing procedures.

• Gene Sequence Comparisons. DNA is also the perfect molecule for comparison of a set of related DNA sequences. If a family of genes (e.g. analogous genes from different organisms) are annealed together with synthetic strands designed to bridge between related sequences, then the existence or the morphology of the resulting superstructure might convey information about the extent of sequence similarity in the gene set.

• Gene Therapy. It is difficult to imagine any better material for the construction of a therapeutic agent targeted toward DNA than DNA itself. Target sequence specificity is readily programmable, complex structures which bring together fairly distant regions of a long strand may be possible, stability at physiologic-like temperature and solution conditions, and the ability to organize non-DNA materials may contribute to the usefulness of DNA tiles as therapeutics. As is the case with conventional gene therapy agents, delivery may be the key limiting factor. Experiments are planned which will test the encoding of complete DNA tiles on a single cloning vector. This will not only increase the yield and decrease the cost of tiles, but it may mitigate the problem of delivery of multiple strands to a target location. A self-assembling DNA tile structure for gene therapy could make use of the fact that participation in crossover complexes increases resistance to nuclease enzymes over that of standard dsDNA. A properly delivered complex which specifically hybridizes with a target site on cellular DNA or mRNA may act to sequester the bound nucleic acids and turn off an undesired cellular response. Alternatively, if distant regions of the cellular nucleic acid were held close together within a DNA crossover complex it might be possible to activate a cellular repair mechanism and cause the excision of some portion of a faulty gene or perhaps the delivery of a corrected copy. It also might be possible to design DNA assemblies which act as diagnostics to probe for multiple mutations or multiple, specific alleles simultaneously.

References.

Adleman, L.M. (1994) "Molecular computation of solutions to combinatorial problems", *Science* **266**, 1021-1023.

Alivisatos, A.P., K.P. Johnsson, X. Peng, T.E. Wilson, C.J. Loweth, M.P. Bruchez Jr., and P.G. Schultz (1996) Organization of 'nanocrystal molecules' using DNA. *Nature*, 382, 609-611.

Baum, E. B. (1996) “DNA Sequences useful for computation”, Proceedings of the 2nd Annual DIMACS Meeting on DNA Based Computers, Princeton University, June 1996.

Braun, E., Y. Eichen, U. Sivan, and Ben-Yoseph, G. (1998) "DNA-templated assembly and electrode attachment of a conducting silver wire", *Nature* **391**, 775-778.

Brenner, S., and Lerner, R.A. (1992) “Encoded combinatorial chemistry”, *Proc. Nat. Acad. Sci. USA* **89**, 5381-5383.

Chen, J-H., Kallenbach, N.R., and N.C. Seeman (1989) "A specific quadrilateral synthesized from DNA branched junctions", *J. Am. Chem. Soc.* **111**, 6402-6407.

Chen, J-H. and N.C. Seeman (1991) “The synthesis from DNA of a molecule with the connectivity of a cube”, *Nature* **350**, 631-633).

Deaton, R., R.C. Murphy, M. Garzon, D.R. Franceschetti, and S.E. Stevens, Jr. (1996) “Good encodings for DNA-based solutions to combinatorial problems”, Proceedings of the 2nd Annual DIMACS Meeting on DNA Based Computers, June 1996.

Du, S.M., and Seeman, N.C. (1994) "The construction of a trefoil knot from a DNA branched junction motif", *Biopolymers* **34**, 31-37.

Hanvey, J.C., Peffer, N.J., Bisi, J.E., Thomson, S.A., Cadilla, R., Josey, J.A., Ricca, D.J., Hassman, C.F., Bonham, M.A., and Au, K.G., (1992) "Antisense and antigene properties of peptide nucleic acids", *Science* **258**, 1481-1485.

LaBean, T.H., Winfree, E., and Reif, J.H. (2000a) “Experimental progress in computation by self-assembly of DNA tilings”, 5th DIMACS Workshop on DNA Based Computers, MIT, June, 1999. DNA Based Computers, V, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, (ed. E. Winfree), American Mathematical Society, 2000.

LaBean, T. H., Yan, H., Kopatsch, J., Liu, F., Winfree, E., Reif, J.H. & Seeman, N.C. (2000b) “The construction, analysis, ligation and self-assembly of DNA triple crossover complexes”, *J. Am. Chem. Soc.* **122**, 1848-1860.

Liu, F., Sha, R., and Seeman, N.C. “Modifying the surface features of two-dimensional DNA crystals.” *J. Am. Chem. Soc.* **121**, 917-922 (1999).

Mao, C., Sun, W.,. Shen, Z., and. Seeman, N.C. (1999) “A DNA nanomechanical device based on the B-Z transition”, *Nature* **397**, 144-146).

Mao, C. Sun, W., and Seeman, N.C. (1999) “Designed two-dinesional DNA Holliday junction arrays visualized by atomic force microscopy”, *J. Am. Chem. Soc.* **121**, 5437-5443.

Mao, C., T.H. LaBean, J. H. Reif, and N.C. Seeman, N.C. (2000) “Logical computation using algorithmic self-assembly of DNA triple-crossover molecules”, *Nature*, **407** 493-496.

Marathe, A., Condon, A.E., and Corn, R.M. (2000) “On combinatorial DNA word design”, *DNA Based Computers V*, DIMACS Workshop on DNA Based Computers (5th, MIT, June 14-15, 1999) editors E. Winfree & D.K. Gifford, DIMACS Series in Discreet Mathematics and Theoretical Computer Science, **54**, 75-90.

Mbindyo, J.K.N., Reiss, B.D., Martin, B.R., Keating, C.D., Natan, M.J., and Mallouk, T.E. (2001) *Advanced Materials* **13**, 249-254.

Mirkin, C.A., Letsinger, R.L., Mucic, R.C., and Storhoff, J.J. (1996) "A DNA-Based Method For Rationally Assembling Nanoparticles Into Macroscopic Materials," *Nature* **382**, 607-609.

Mucic, R. C.; Storhoff, J. J.; Mirkin, C. A.; Letsinger, R. L., (1998) "DNA-directed synthesis of binary nanoparticle network materials", *J. Am. Chem. Soc.* **120**, 12674-12675.

Reif, J.H. (1998) "Paradigms for biomolecular computation", First International Conference on Unconventional Models of Computation, Auckland, New Zealand, January 1998. Published in *Unconventional Models of Computation*, edited by C.S. Calude, J. Casti, and M.J. Dinneen, Springer Publishers, pp 72-93.

Reif, J.H., LaBean, T.H., Pirrung, M., Rana, V.S., Guo, B., Kingsford,C., and Wickham, G.S. (2001) "Experimental construction of very large scale DNA databases with associative search capability", 7th InternationalMeeting on DNA Based Computers (University of South Florida, June 10-13, 2001) editors N. Jonoska and N.C. Seeman.

Rothemund, P.W.K. (2000) “Using lateral capillary forces to compute by self-assembly”, *Proc. Nat. Acad. Sci.* **97**, 984-989

Peyret, N., Seneviratne, P. A., Allawi, H.T., SantaLucia, J., Jr. (1999) “Nearest-neighbor thermodynamics and NMR of DNA sequences with A-A, C-C, G-G, and T-T mismatches”, *Biochemistry* **38**, 3468.

Seeman, N.C. (1982) “Nucleic acid junctions and lattices”, *J. Theor. Biol*. **99**, 237-247.

Seeman, N.C. (1990) “De novo design of sequences for nucleic acid structural engineering.” *Journal of Biomolecular Structure and Dynamics* **8**: 573-581.

Su, M., Zheng,,B., and Liu, J. (2000) "A scalable CVD method for the synthesis of single walled carbon nanotubes with high catalyst productivity," *Chem. Phys. Letts* **322**, 321-326.

Tae, E.L., Wu, Y., Xia, G., Schultz, P.G., and Romesberg, F.E., (2001) "Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs", *J. Am. Chem. Soc.* **123**,7439-40.

Tsai, D.E., Kenan, D.J. and Keene, J.D. (1992) "In vitro selection of an RNA epitope immunologically cross-reactive with a peptide", *Proc. Nat. Acad. Sci., U.S.A.* **89**, 8864-8868.

Watson, J.D., and Crick, F.H.C. (1953) “Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid”, *Nature* **171**, 737-738.

Winfree, E. (1998) Algorithmic self-assembly of DNA. Ph.D. Thesis, Caltech.

Winfree, E., Yang, X., & Seeman, N.C., (1998a) “Universal computation via self-assembly of DNA: Some theory and experiments”, In Landweber, L.F. & Baum, E.B. (eds.), *DNA Based Computers II: DIMACS Workshop, June 10-12, 1996,* American Mathematical Society, Providence, RI.

Winfree, E., Liu, F., Wenzler, L.A., and Seeman, N.C. (1998b) “Design and Self-Assembly of Two-Dimensional DNA Crystals, *Nature* **394**, 539-544.

Yang, X., Wenzler, L.A. Qi, J., Li, X., and Seeman, N.C. (1998) “Ligation of DNA triangle containing double crossover molecules”, *J. Am. Chem. Soc.* **120**, 9779-9786.

Yurke, B., Turberfield, A.J., Mills, A.P., Simmel, F.C. and Neumann, J.E. (2000) "A DNA-fuelled molecular machine made of DNA", *Nature* **406**, 605-608.

Zhang, Y., and Seeman, N.C. (1994) "The construction of a DNA truncated octahedron", *J. Am. Chem. Soc.* **116**, 1661-1669.

Acknowledgements. Funding for some of the research described here was provided by DARPA and NSF in grants to John Reif who has been invaluable to the progress of this research. Thanks also go to Wolfgang Frey who performed the AFM shown in Figure 4, David Anderson, who performed the rotary shadowing for the sample shown in Figure 4., and Dage Liu who did the experiments summarized in Figure 7. Thanks also to Hao Yan and Lizbeth Videau for critical reading of the manuscript.

Index terms: Biomolecular computation, DNA, DNA-based nanotechnology, macromolecular engineering.