



## Coding and geometrical shapes in nanostructures: A fractal DNA-assembly

ALESSANDRA CARBONE<sup>1</sup> and NADRIAN C. SEEMAN<sup>2</sup>

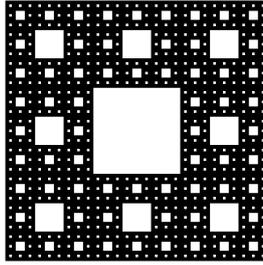
<sup>1</sup>*Institut des Hautes Études Scientifiques, 35, Route de Chartres, F-91440 Bures-sur-Yvette, France (E-mail: carbone@ihes.fr);* <sup>2</sup>*Department of Chemistry, New York University, New York, NY 10003, USA (E-mail: ned.seeman@nyu.edu)*

**Abstract.** Fractal patterns represent an important class of aperiodic arrangements. Generating fractal structures by self-assembly is a major challenge for nanotechnology. The specificity of DNA sticky-ended interactions and the well-behaved structural nature of DNA parallelogram motifs has previously led to a protocol that appears likely to be capable of producing fractal constructions [A. Carbone and N.C. Seeman, A route to fractal DNA assembly, *Natural Computing* **1**, 469–480, 2002]. That protocol depends on gluing the set of tiles with special ‘glue tiles’ to produce the fractal structure. It is possible to develop a fractal-assembly protocol that does not require the participation of gluing components. When designed with similar DNA parallelogram motifs, the protocol involves sixteen specific tiles, sixteen closely related tiles, and a series of protecting groups that are designed to be removed by the introduction of specific strands into the solution. One novel aspect of the construction on the theoretical level is the interplay of both geometry and coding in tile design. A second feature, related to the implementation, is the notion of generalized protecting groups.

**Key words:** coding of tiles, DNA assembly, DNA molecules, DNA nanotechnology, geometry of tiles, Sierpinski fractals, tiling

### 1. Introduction

In previous work (Carbone and Seeman, 2002), we have described a route to fractal assembly using tiles based on DNA parallelograms. The target was to build a Sierpinski square (see Figure 1) through an iterative assembly process. Given a basic parallelogram-shaped DNA tile (Mao et al., 1999) with an appropriate coding of the sides, the protocol was as follows: at step  $i = 0$ , eight basic DNA tiles were used as building blocks and were assembled with the help of glue tiles (also constructed from DNA parallelograms) to form a “square with a hole in the middle” arrangement. The basis of the binding between tiles is the intermolecular Watson-Crick complementarity of DNA cohesive ends, used in biotechnology (Cohen et al., 1973), DNA nanotechnology (Seeman, 1982) and DNA-based computation (Winfree, 1996). The result was a DNA tile of larger dimension and with an appropriate coding of



*Figure 1. Sierpinski square fractal in  $\mathbb{R}^2$ .* Its construction goes as follows. Let  $U$  denote the unit square  $[0,1] \times [0,1]$  in  $\mathbb{R}^2$ . We can decompose  $U$  into 9 (closed) squares  $U_i$  in the obvious manner, so that  $U_i$ 's have disjoint interiors and side-length  $1/3$ . Take  $U$  and replace it with the union of all the  $U_i$ 's except the one in the center of  $U$  to produce a set  $K_1$  which is the union of 8 squares. Apply this process to each of the squares and repeat the procedure indefinitely to get a sequence of compact sets  $K_i$  in the plane, where each  $K_i$  consists of  $8^i$  squares of length  $3^{-i}$ , and where these squares have disjoint interiors. The set  $K = \cup_i K_i$  is a compact subset of  $\mathbb{R}^2$  called Sierpinski square fractal or Sierpinski carpet. In the paper we consider the finite approximations  $K_i$  of  $K$ . In the picture we see  $K_4$ . The index  $i$  is called *dimension* of the fractal.

its sides. Step  $i + 1$  assembled eight DNA tiles obtained at step  $i$  with the help of glue tiles and produced tiles of increased size.

Here, we have sought to design a synthetic protocol that would avoid the use of gluing elements. We describe how to assemble a Sierpinski square-like fractal starting from a set of 32 DNA parallelogram tiles, without using glues. The tiles are illustrated in Figures 2 and 3. The physical nature of the tiles is that they consist of four DNA double helices; those with parallel helix axes (the red pair or the blue pair) lie in the same plane, but the other pair lie in a second plane, about 2 nm from the first plane. The tiles are drawn as rhomboids, but are not restricted to equal distances between their vertices on both layers. The vertices correspond to the crossover points of 4-arm immobile DNA junctions (Seeman, 1982); each of the four junctions consists of a red helix joined to a blue helix by strand connectivity (Figure 4). A feature of the self-assembly is a coding on each edge of the parallelogram that involves the lengths of the helices, one long and one short. The tiles naturally assort themselves into two groups, shown in Figures 2 and 3. Above each group is a prototype tile, labeled **T** (Figure 2) or **T** (Figure 3). The difference between the two groups is that the positions of the long red helix and the short red helix of the **T** group have been reversed in the **T** group, although the positions of the blue helices remain unchanged. The tiles are all drawn in the same orientation, with the red helices vertical. The acute angle in the lower left-hand corner has been labeled  $\alpha$  and the obtuse angle in the upper left-hand corner has been labeled  $\beta$ , as an aid to follow the orientation of

the tiles. It is evident from Figures 2 and 3 that eight different extensions protrude from each tile. There are two possible extensions per side and each member of the pair has a different length, one being *long* and the other *short*. In each tile, the long extensions are both on the same helix, as are the short extensions. The two lengths are fixed for all the tiles.

The parallelogram-like geometry of the tiles themselves is well-defined; indeed, periodic arrays of these tiles have been used to quantitate the angles at their branchpoints (Mao et al., 1999; Sha et al., 2000, 2002). They share the structural integrity of their shape with DNA double crossover molecules (Li et al., 1996), a system proposed by Winfree for use in algorithmic DNA assembly, including “Pascal triangles, mod 2” (Winfree, 1996). In addition to their well-defined geometries, our tiles are characterized by a specific code for each side. There are two basic coding labels that label the sides of the tiles, **a** and **a\*** (see Figures 2 and 3), where **a\*** is designated to be the complement of **a** in the Watson-Crick sense.

An important property of the well-structured nature of the tiles is that we can use both traditional Watson-Crick affinity *and* tile geometry to define their assembly properties. We have pointed out previously (Li et al., 1996) that structural integrity is necessary for periodic constructions based on DNA sticky ends. The same feature is even more important for algorithmic assemblies (Winfree, 1996, 2000), such as the one described here. Indeed, these tiles are *essentially* Wang tiles: besides the coding, they carry geometrical information. Two tiles *match* if their sides are coded with opposite labels (i.e., one with **a** and the other with **a\***) *and* if their geometrical boundaries allow the fit for both helices at a point where they bond.

Tiles have been designated in groups labeled **1–6** in Figure 2. The tiles labeled **b**, **r** and **t** contain two circled combining sites, both labeled **a**; these are the positions where they are designated to combine with other tiles in the assemblies. Tile **1** can combine in any of its adjacent pairs of **a** sites, so they have not been labeled in Figure 2. The **b** and **r** tiles contain one **a** and one **a\*** site that are designated not to combine; the **a\*** site is in a vertical position (between two red helices) in the **b** tiles and in a horizontal position (between two blue helices) in the **r** tiles. The **t** tiles again have two **a** sites in their circled combining sites, but have two adjacent **a\*** sites in both their non-combining sites. The **1**, **2**, **3**, and **4** labels of **b**, **r** and **t** are the four circular permutations of the combining sites. In the following, we refer to tiles **1**, **2**, **3**, and **4** as being of class **1**, because of the adjacent nature of their **a**-containing combining sites. In all there are 13 possible configurations of these tiles. Tiles **5** and **6** have opposite combining edges that are labeled **a\***, and two others labeled **a**. Tile **56** combines these features, with four **a\*** sites that are not circled (as in tile **1**), and any of its adjacent pairs can combine. In total, the number of tiles

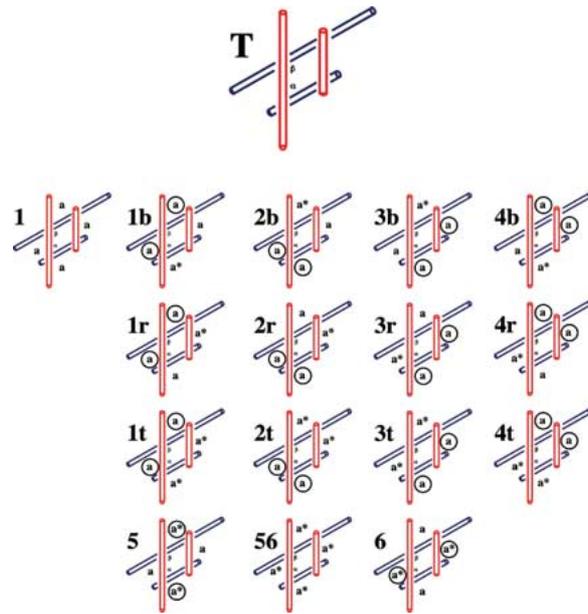


Figure 2.

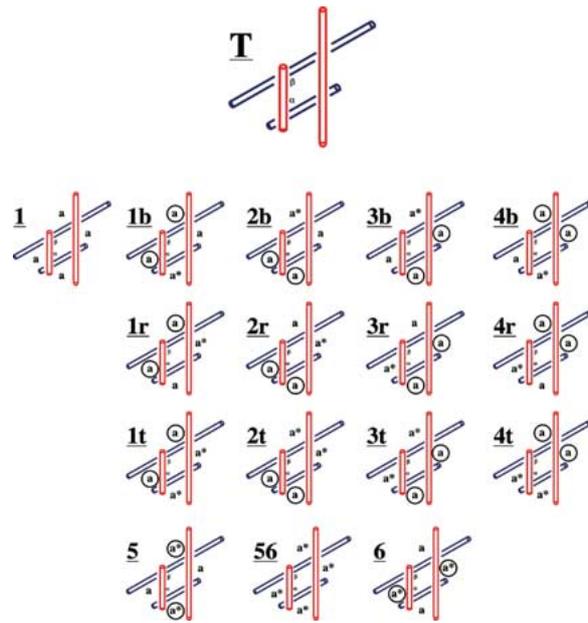


Figure 3.

is 16. As mentioned above, there are two versions of each of the 16 tiles. The other version is shown in Figure 3; these tiles are labeled as T. The tiles in both Figures 2 and 3 are drawn with the red helices vertical. It is clear that the tiles in Figure 3 are identical to those in Figure 2, both in the locations of their combining sites and in the labeling of their edges. The difference between the two tile types is that the positions of the long and short red helices, have been reversed in Figure 3: The long helices are on the left in Figure 2 and the short helices are on the right; the long helices are on the right in Figure 3 and the short helices are on the left.

## 2. Fractal layers

A *fractal layer* is a parallelogram fractal of dimension  $k + 1$  that is constructed in one step out of parallelogram fractals of dimension  $k$  used as building blocks. The tiles in Figures 2 and 3 have dimension 0. In the first stage of the construction, we combine the tiles of Figures 2 and 3 in such a way that physical boundaries as well as coding complementarity are satisfied. As seen in Figures 5 and 6, we assemble these tiles in a specified manner:

---

←

*Figure 2. First set of 16 tiles used in the fractal assembly.* A prototype tile rhombus-shaped tile labeled 'T' is shown at the top of the drawing. It contains two red DNA helices above two blue DNA helices. The upper blue helix and the left red helix contain long extensions going beyond the borders of the rhombus. The lower blue helix and the right red helix contain short extensions beyond the border of the helix. The upper left corner of the rhombus contains a  $\beta$  in the obtuse angle and the lower left corner of the rhombus contains an  $\alpha$  in the acute angle to aid in orientation. There are three basic families of tiles, called **1**, **5** and **6**, and their variations. Circles indicate the binding sites where the tiles are supposed to assemble. Tile **1** has four possible preferred pairs of sides (corresponding to the four corners), so none are listed. Tiles named **2**, **3** and **4** are variations of **1**, but where the pair of preferred sides changed. Tiles **1**, **2**, **3** and **4** all pair through adjacent edges labeled **a**. The tiles labeled **b** contain a vertical edge labeled **a\***, those labeled **r** contain a horizontal edge labeled **a\***, and those labeled **t** contain both. Tiles **5** and **6** are used as intermediary tiles during the assembly, because they contain opposite **a\*** edges. Their variation, **56** is also shown. Notice that the sides coded **a\*** in **5** lie between red extensions; in **6** they lie between blue extensions.

*Figure 3. The second set of 16 tiles used in the fractal assembly.* These tiles are labeled analogously to those in Figure 2, except that their names are underlined; thus, the prototype tile is labeled 'T'. The same conventions apply as in Figure 2. The difference between the tiles in this figure and those in Figure 2 is that the positions of the red helices have been reversed; whereas the long helices are on the left and the short are on the right in that figure, in this drawing the long helices are on the right and the short helices are on the left. The mirror image relationship between the red helices of the two tile types allows them to be combined, so that one type of tile is on the corners and the other type is on the edge, perhaps following a  $180^\circ$  rotation.

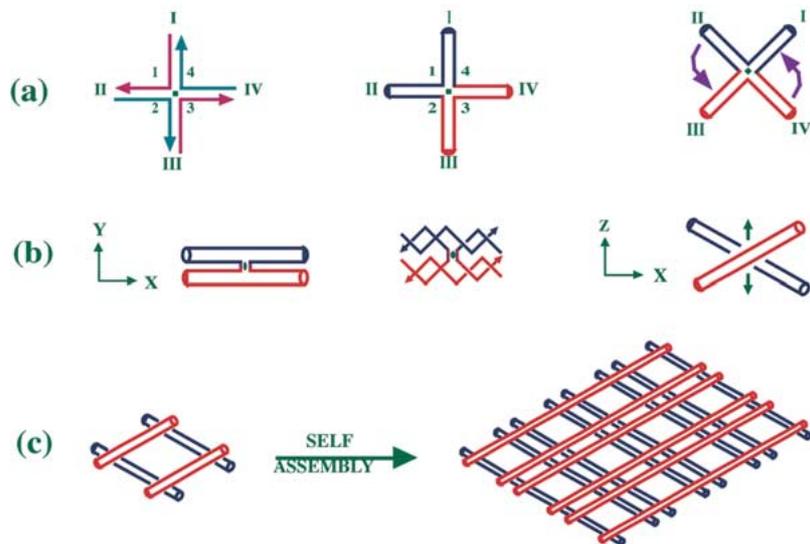


Figure 4.

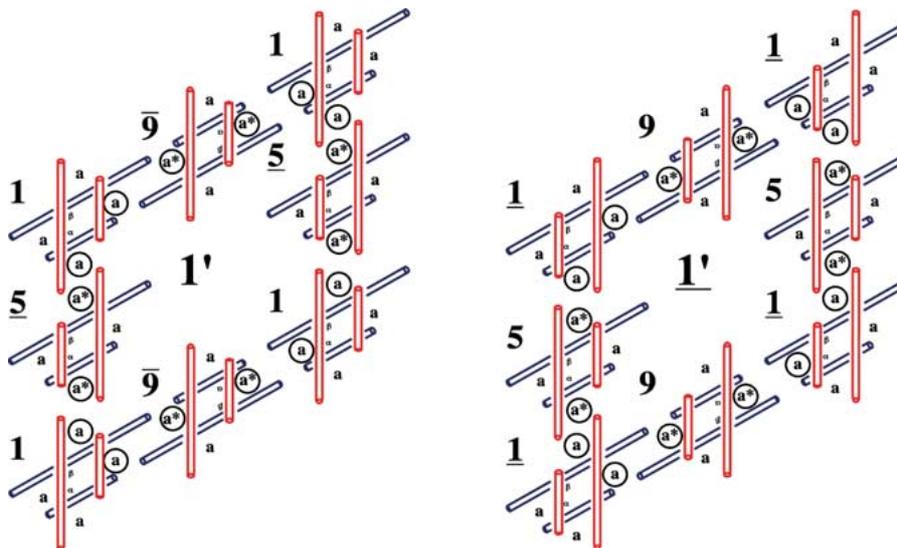


Figure 5.

we have chosen to illustrate explicitly the assembly of a few fractal layers of dimension 1, in all there are 32 of them, and the others are put together similarly. The assemblies that we generate “look like” the tiles of the set itself, both in coding and in geometrical shape; hence, there is a fractal relationship between the tile-like units of successive dimensions. Take, for instance, the assembly of Figure 5. It illustrates a fractal layer representing tile **1** at dimension 1 (designated **1'**). Tile **1** is constituted by four sides coded **a**, and by a *long-short* sequence of helices for each one of its sides. The codes along the sides of **1'** are **aaa**, **aaa**, **aaa**, **aaa**. The geometrical shape of **1'** respects the geometrical properties of tile **1**. That is, the upper side is a sequence of *long-short-long-short-long-short* helices going from left to right, the right vertical side is also a *long-short-long-short-long-short* sequence of helices reading from top to bottom, the bottom side is a *long-short-long-short-long-*

---

←

*Figure 4. Schematic representations of the Holliday junction, its use as a parallelogram component and their assembly into 2D Arrays. (a) Representations of the Holliday junction.* The left panel shows the Holliday junction as four strands, in two homologous pairs, as represented by the colors. The middle panel represents the same structure, but emphasizing its helical structure; here, adjacent helices are colored the same. The right panel rotates the molecule and indicates with curved arrows the folding of the helices into a stacked structure. *(b) The stacked structure.* The left panel shows a view down the dyad axis of the Holliday junction. The dyad axis is indicated by the small lens-shaped figure. The upper helical domain is rotated 30° about the vertical so that its right end penetrates the page, and the lower helical domain is rotated 30° about the vertical so that its left end penetrates the page. The X and Y axes of a right-handed coordinate system are shown to help orient the reader. The central panel shows a representation of the strand structure of this molecule. The right panel shows another cylindrical view, but now with the dyad axis vertical. The molecule has been rotated 90° about the X axis. The dyad axis is indicated by the double arrows. *(c) Schematic representation of the parallelogram and its self-assembly into a 2D array.* In the left panel, four molecules, in the orientation of the right panel of (b) are combined to form a parallelogram. Typically, there are six turns of DNA in each helix, and four turns between crossover points, leading to 1-turn overhangs on the ends. The red helices are about 2 nm closer to the reader than the blue helices. The right panel shows the 2-dimensional self-assembly product of the motif; the long separations between helices contain four helical turns, and the short separations contain two helical turns. Note that the latticework array contains two separate layers, an upper layer oriented from lower left to upper right, and a lower layer oriented from lower right to upper left.

*Figure 5. First fractal layer of tiles **1** and **1'**.* The left panel contains tile **1'** and the right panel contains tile **1**. The boundaries of the **1'** tile are coded by the same letters (i.e., triplets **aaa**) and the geometrical shape of the sides of the fractal tile consists of a repeat of the basic motif “*long-short* helices”. Note that the corners are **1** tiles, but the edge tiles are **5** and **6** tiles. The **6** tile has been rotated 180°. The right panel contains the **1'** tile. The corners and edges have been replaced with the corresponding tile from the other group, with **1** tiles on the corners, and edges comprised of **5** and inverted **6** tiles.

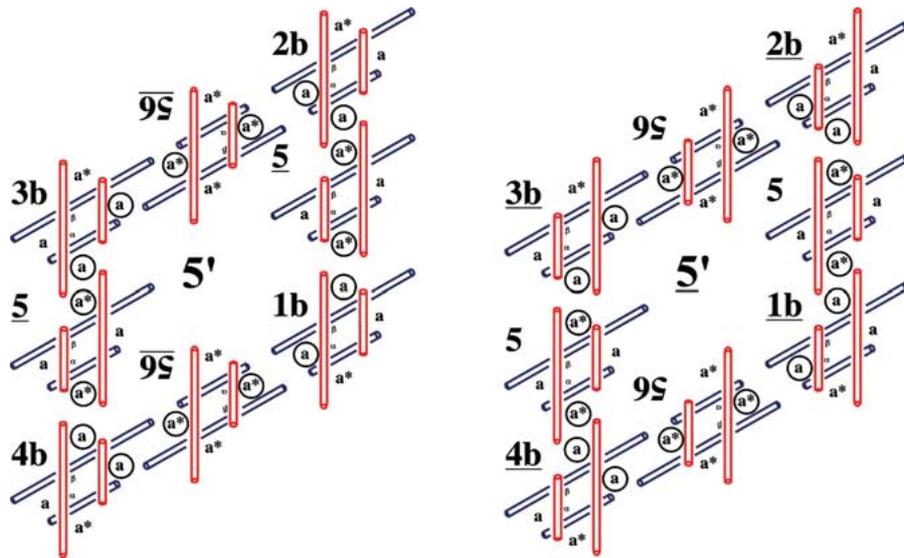


Figure 6. First fractal layers of basic tiles  $5$  and  $\bar{5}$ . Both the coding and the geometrical shape of the sides correspond to the characteristics of the  $5$  and  $\bar{5}$  tiles.

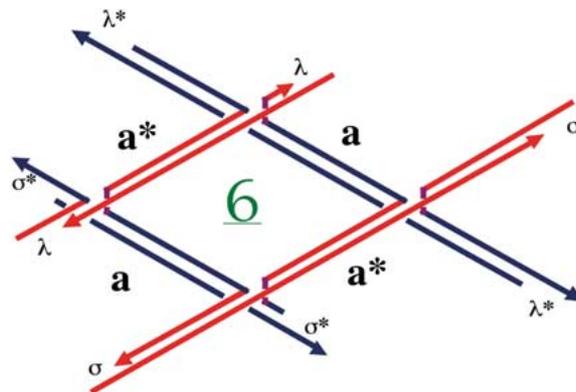


Figure 7. The strand structure of the  $6$  tile. This drawing shows how DNA strands hybridize in a stable parallelogram configuration. Arrowheads represent 3' ends. Note that the sticky ends on the red helices contain 5' overhangs, and the sticky ends on the blue helices contain 3' overhangs. The red helices are about 2 nm closer to the reader than the blue helices. Sides coded  $a$  and  $a^*$  correspond to *short* and *long* extensions coded  $\sigma$ ,  $\lambda$  and  $\lambda^*$ ,  $\sigma^*$  respectively, where  $\lambda$ ,  $\lambda^*$  and  $\sigma$ ,  $\sigma^*$  are complementary sequences. Note that some strands pass between layers. Although not an issue here, the parallel-strand representation for DNA used can lead to visual artifacts, such as apparent 5'-5' juxtapositions of sticky ends. The way to imagine the actual strands is to picture them forming a zigzag path between the parallel strands.

*short* sequence of helices reading from left to right and the left vertical side is a *long-short-long-short-long-short* sequence of helices reading from top to bottom. The consistency of the coding and the geometry of fractal layers  $\mathbf{5}'$  and  $\underline{\mathbf{5}'}$  (Figure 6) can be checked easily.

If a zero-layer tile is among the  $\mathbf{T}$  tiles in Figure 2, corner tiles of a first layer tile modeled on it will also be  $\mathbf{T}$  tiles. They will be joined together by  $\underline{\mathbf{T}}$  tiles from Figure 3. Likewise, if the zero-layer model tile is among the  $\underline{\mathbf{T}}$  tiles, its corner tiles will be  $\underline{\mathbf{T}}$  tiles, and they will be joined by  $\mathbf{T}$  tiles. We have organized the labeling of the tiles so that any given  $\underline{\mathbf{T}}$  tile of a given layer  $k$ , can be constructed easily by knowing the components of the  $\mathbf{T}$  tile: Switching  $\mathbf{T}$  tiles (of dimension  $k-1$ ) to  $\underline{\mathbf{T}}$  tiles and the  $\underline{\mathbf{T}}$  tiles to  $\mathbf{T}$  tiles will generate the  $\underline{\mathbf{T}}$  tiles of dimension  $k$ . For example, the left panel of Figure 5 shows tile  $\mathbf{1}'$ , with  $\mathbf{1}$  tiles at the corners,  $\mathbf{5}$  tiles on the left and right sides and inverted  $\mathbf{6}$  tiles on the top and bottom edges. Likewise, the right panel of Figure 5 contains the  $\underline{\mathbf{1}}$  tile, with  $\underline{\mathbf{1}}$  tiles at the corners,  $\mathbf{5}$  tiles on the left and right sides and inverted  $\mathbf{6}$  tiles on the top and bottom edges. The same point is evident in Figure 6, showing the  $\mathbf{5}'$  and  $\underline{\mathbf{5}'}$  tiles.

One applies the same approach to generate fractal layers of dimension  $k > 1$ . For example, tile  $\mathbf{5}''$  is obtained by assembling tiles  $\mathbf{1b}'$ ,  $\mathbf{2b}'$ ,  $\mathbf{3b}'$ ,  $\mathbf{4b}'$ ,  $\underline{\mathbf{5}'}$ , and  $\mathbf{56}'$  for  $k = 2$  (compare with Figure 6). In general, fractal layers have sides coded either  $(\mathbf{a})^{3^k}$  or  $(\mathbf{a}^*)^{3^k}$ , and geometrical shapes determined by  $3^k$  repeats of *long-short* helical motifs. In fact, iteration  $k$  defines an assembly of  $3 \times 3$  tiles of dimension  $k-1$  (with a gap in the center), and the sides of a tile of dimension  $k$  are coded either by a sequence  $(\mathbf{a})^{3^{k-1}} \cdot (\mathbf{a})^{3^{k-1}} \cdot (\mathbf{a})^{3^{k-1}} = (\mathbf{a})^{3^k}$  or by a sequence  $(\mathbf{a}^*)^{3^{k-1}} \cdot (\mathbf{a}^*)^{3^{k-1}} \cdot (\mathbf{a}^*)^{3^{k-1}} = (\mathbf{a}^*)^{3^k}$ , where  $(\mathbf{a})^{3^k}$  or  $(\mathbf{a}^*)^{3^k}$  is a sequence made of  $3^k$  repeated  $\mathbf{a}$ 's or  $\mathbf{a}^*$ 's. The parallelogram shapes are respected, for all  $k > 1$ .

### 3. DNA tiles suitable for a fractal assembly

The strand structure of DNA parallelogram tiles suitable for implementing the appropriate coding and geometry of the tiles described in the introduction is shown in Figure 7. The 8 helical extensions beyond the edges take different lengths according to the abstract description of the tile in question (see Figures 2 and 3). Helices are either *long* or *short*. The *long* helix of a side coded  $\mathbf{a}$ , is constituted by a sticky-end with sequence  $\lambda$ , and the *short* helix of the same side has a sticky-end with sequence  $\sigma$  (Figure 7). The complementary side  $\mathbf{a}^*$  is formed by a *short* helix with code  $\lambda^*$  and by a *long* helix with code  $\sigma^*$ . We set the sequence  $\lambda$  to be complementary to  $\lambda^*$  and the sequence  $\sigma$  to be complementary to  $\sigma^*$ . See Figure 7. The distance

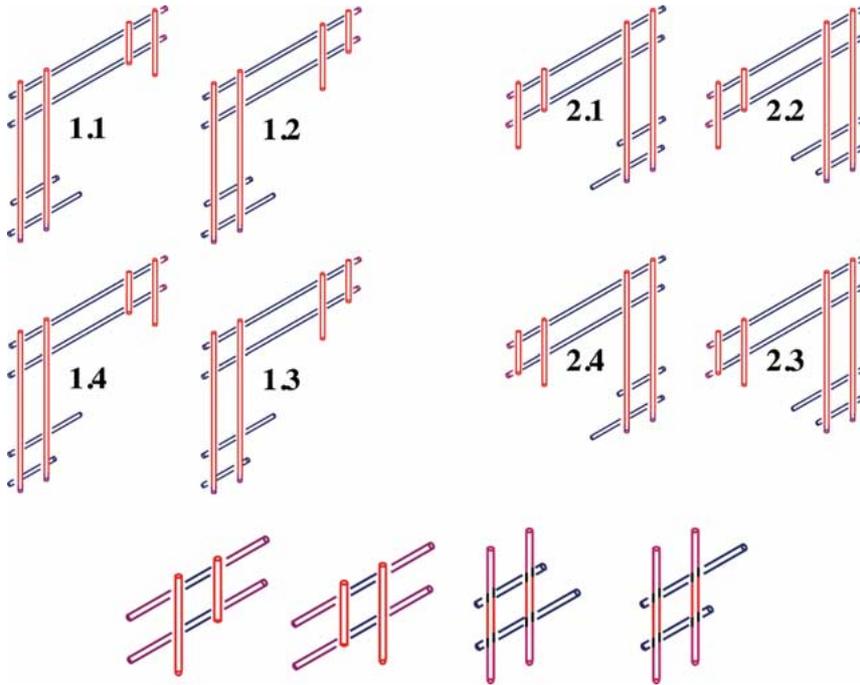


Figure 8.

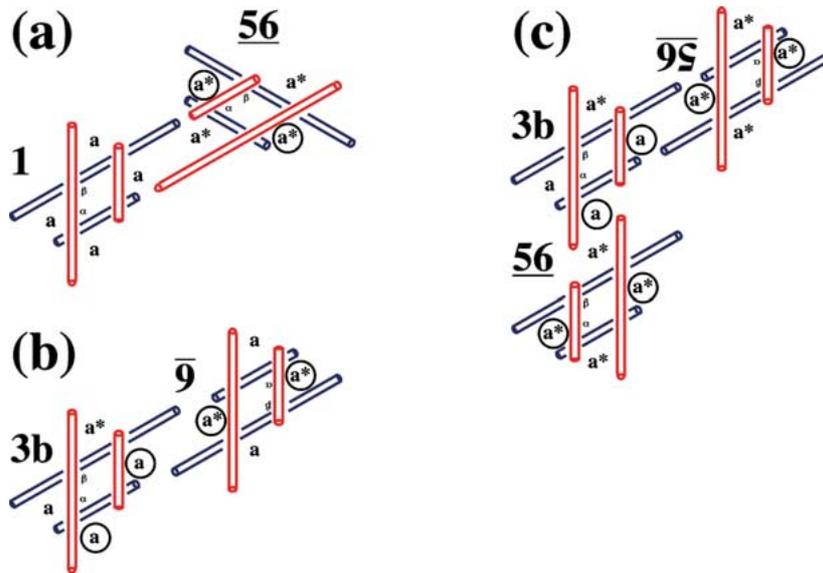


Figure 9.

that is the sum of  $\lambda$  and  $\sigma$  may be an integral multiple of DNA half-turns. We will arbitrarily restrict this number to even multiples of half-turns, i.e., full turns of DNA. This restriction lends another level of control over the construction, because tiles that turn upside down and bind (as, e.g., the top and bottom edges of tile 1b), would be misphased by a half-turn in their other helix. Thus, the helical nature of DNA has a component of rotational, as well as length control in it.

It is important to realize that our scheme is predicated on the notion that *both* sides of an edge will pair with another edge; a single interaction between a  $\lambda$  sticky end and a single  $\sigma^*$  sticky end could result in an erroneous structure. This is not an unfamiliar notion in nucleic acid hydrogen bonding, which avails itself of cooperativity in almost all of its aspects. The stabilization that results from a second favorable interaction that can occur within a context where one interaction is already established clearly benefits from entropic gain, relative to a system that must join more molecules to achieve interactions characterized by the same individual free energies. The bases themselves pair with two or three hydrogen bonds, and single hydrogen bonding interactions, unless otherwise supported by other structural features, are not regarded as major stabilizing elements. The successful prototype XOR calculation performed recently (Mao et al., 2000) also requires that

---

←

*Figure 8. Protecting groups.* Two sets of protecting groups are shown schematically in the upper panel. Those that protect obtuse angles are labeled 1.1, 1.2, 1.3 and 1.4; those that protect acute angles are labeled 2.1, 2.2, 2.3 and 2.4. Within each group, the differences have to do with the long/short features of the corner that is protected. Note that the tips of the protecting groups are shown to be purple. The purple extensions are used to construct long protections in successive fractal layers. To do this one uses extra tiles that are illustrated in the lower figure. The coding sequence associated with all long extensions is the same. The same holds for short extensions. In practical terms, a tile binds to a protecting group through protecting strands that will recognize the coding sequences of the protecting group and of the coding sequence of the tile, as illustrated in Figure 11. As a consequence, the geometrical shape of the protection is the only characteristic that discriminates tiles binding to the protecting groups, and this allows having 8 protecting groups instead of 32 (coming from all possible combinations of  $\mathbf{a}/\mathbf{a}^*$  coding and shapes of two consecutive sites). Protecting strands will be also employed for the 4 tiles used to elongate protecting groups on fractal layers of non-trivial dimension.

*Figure 9. Possible types of mispairing.* (a) *An impossible mispair.* The different sticky ends on the two layers prevent this type of mispair, despite the appropriate geometrical fits of the extensions on the **1** and **56** tiles. (b) *A feasible assembly of tiles 3b and 6.* The geometrical shape, coding and planes of the helices match accordingly. However, this is an undesired pairing, because the top row combines  $\mathbf{a}$  and  $\mathbf{a}^*$  extensions. (c) *Tile 56 is assembled with tile 3b in two different ways.* The one on the right is appropriate, but the lower interaction mixes extension types.

two separate double helical pairings occur successfully, in a context where a single one could result in an incorrect assembly. Nevertheless, as in all algorithmic assemblies, it is key that kinetic intermediates be eliminated by ensuring that equilibrium is reached for any particular reaction where competing alternatives exist. It is likely that this system will be most effective if assemblies are formed at temperatures where double interactions compete effectively against single interactions.

As noted above, in the physical realization of a parallelogram tile, pairs of helices lie in two distinct parallel planes. Our approach requires that there be no cross-reactivity between the two planes. We must exclude mispairing between similar edges, such as the top of **1b** and the right side of **2r**, or the top of **2t** and the right side of **4t**. The exclusion of pairing between the layers is accomplished readily by using 5' overhangs for the sticky ends on the helices in the upper plane, and 3' overhangs for the sticky ends on the helices in the lower plane; in Figure 7, the strand structure indicates that the upper plane contains the red helices, and the lower plane contains the blue helices; if acceptable to the target design, different lengths for the blue extensions and red extensions could also be used.

#### 4. Parallelogram protection

It is a common practice in chemical synthesis (e.g., Caruthers, 1985), and even in DNA nanotechnology (Seeman, 1992; Zhang and Seeman, 1992) to direct assembly in the desired direction by using protecting groups. The use of protecting groups prevents the occurrence of interactions of similar free energy to the target interactions; once the desired bonding has occurred, the protecting groups are removed. To avoid the construction of potentially competitive, undesirable shapes during assembly in this system, it is necessary to protect the corners of the tiles during assembly. The purpose is to block the access to binding pairs of *consecutive* edges of a tile. We use a combination of several parallelograms forming corner-covers to protect corners. At the first stage of the assembly, protecting groups consist of 3 parallelograms, two of which bind to the two consecutive sites of a tile. A third parallelogram is used to join these two protecting parallelograms together. Figure 8 illustrates the four kinds of protecting groups for a tile: they correspond to all combinations of *long-short* helices that can lie on a corner of either shape, either an acute angle or an obtuse angle. In successive stages of the assembly, the length of the protecting groups might vary. Ideally one should ask for protecting groups that completely cover the two sides of the tile. In practice, it might be enough that only a short part of the sides is covered, to ensure

that undesired binding is unstable. Proper control of the temperature during assembly might make short sides feasible.

Groups **1**, **5** and **6** contain tiles whose corners, i.e., pairs of successive sides, have geometrically different shapes (this is a crucial feature of our set of tiles) and because of this we can design protecting groups that are corner specific for each tile. In Figure 8 we illustrate a complete set of different corner sides with specific shapes.

Protecting groups are to be put in solution along with tiles of a specific kind (that is, either **1** or **5** or **6**) and they select, with no ambiguity, the appropriate binding site because of the geometry of the tile (see below and legend to Figure 8 for how the use of protecting strands makes selection depend on *geometry* only, avoiding *coding* discrimination between **a/a\*** sites). The outcome assembly is then used for the assembly of tiles at a larger scale. Figure 10 illustrates the desired assembly of tiles **1** and **5**, after the two incorrect options illustrated in Figure 9.

A consideration of the detailed nature of the protecting groups is key if the construction is not to get bogged down in complex syntheses. Figure 11 illustrates the strand structure of a protecting tile, such as one of those shown in Figure 10. The protecting group ties up the sticky ends of the tile on the bottom, just as those in Figure 10 tie up those of the tile on the lower right. The three parallelograms labeled '**P**' form the corners of the protecting group, similar to 1.3 of Figure 8. This is an example of an obtuse protecting group, but acute protecting groups are also necessary, as seen in Figure 8. The protected tile is labeled '**T**'. It seems possible to take a modular approach to protection by having the actual protection effected by strands that bind to the geometry-enforcing large protecting group (ensuring that only the proper edges are protected) and to the tile to be protected. These are shown in Figure 11 as four doubly colored single strands. One part forms a bridge between the protecting group and the sticky end on the '**T**' tile, by forming base pairs with both. The other part, colored differently, corresponds to a non-pairing extension of this strand, similar to the extensions used by Yurke et al. (2000) to remove strands from a DNA nanomechanical device. Removal is achieved by adding the complete complement (optionally containing biotin) to the doubly colored strand, both the extension and the part paired to the larger units. The system works by the complement binding to the extension and then invading the duplex formed by the other strand *via* a process known as branch migration. Once the new duplex is formed, it can be removed by magnetic streptavidin beads that bind biotin (Yan et al., 2002). This mechanism should allow the use of **different** extensions attached to the **same** parts of the strands that form duplex; this procedure would permit selective deprotection of complexes held together by the same sticky ends. Such an approach

would rely on careful control of the kinetics of the system, because it is predicated on kinetics, rather than equilibrium thermodynamics. To produce protecting groups with the greatest generality, we have introduced a series of strands in Figure 11 that are colored orange and dark green. These strands (whose sequences are all different), can bind to the generic protecting group, lending it the specificity to bind the bridge strands. These strands are also removed by adding full complements.

Protection at higher dimensions using DNA suffers from the problem that more-and-more complex protecting groups would need to be manufactured. As noted above, it might be possible to protect only the corners of large fractal tiles without protecting the entire edge of the tile. A way to enable this approach would be to use peptide nucleic acids (PNA) for the apparatus of the protecting group (Nielsen et al., 1991). PNA binds more tightly to DNA than DNA, because its backbone is uncharged. Thus, it should out-compete other molecules available to displace it. Insofar as is known, PNA behaves exactly the same as DNA, except for a slightly different twist, which needs to be considered, but is not an insurmountable problem on the single-stranded regions of the protecting groups.

## 5. Fractal assembly in solution

Two tiles assemble along a side if three conditions are satisfied simultaneously: [1] The geometry of their sides fits (i.e., *long* extensions are opposite to *short* ones); [2] the coding of opposite extensions is complementary; and [3] the helices corresponding to opposite extensions lie on the same layer. Keeping this in mind, we will explain the assembly process, by considering the specific example of fractal layer **1'**. All other assemblies follow an analogous schema. To assemble **1'**, we take four vessels containing multiple copies of tile **1** and add into each one of them a protecting group for a specific corner of the tile. Because of the geometry of the sides of the tiles, with no ambiguity, a specific corner is chosen by the protecting group to assemble on the tile. The tiles with their corners protected are then mixed in new vessels with tiles **5** and **6**, one type after the other, by respecting the arrangement illustrated in Figure 5, say going clockwise from the top-right tile **1** back to it. There is no ambiguity in the way that tiles **5** and **6** assemble to each tile **1** because of the geometry of their sides *and* of the coding. Both these conditions play a role here. It is important to point out that there will be no interactions between two different copies of the same tile. The non-pairing between layers, the different coding within the same layer, and the inability (noted above) of tiles to bind when upside down are the basis for this key condition.

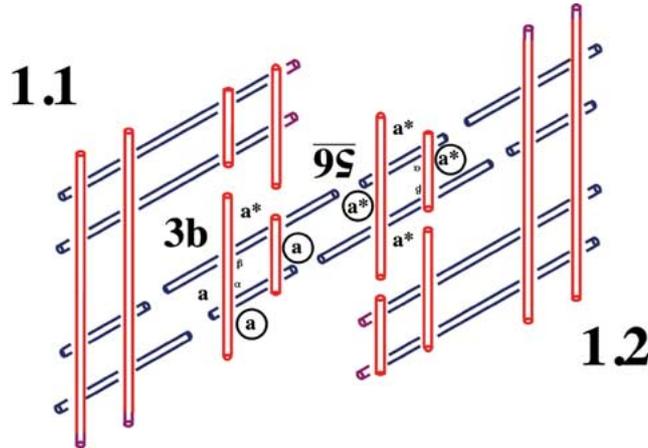


Figure 10. An example of protection. This example shows how the inappropriate interaction shown in Figure 9c can be prevented by the presence of protecting group 1.2. Note, however, that the desired interaction is still feasible.

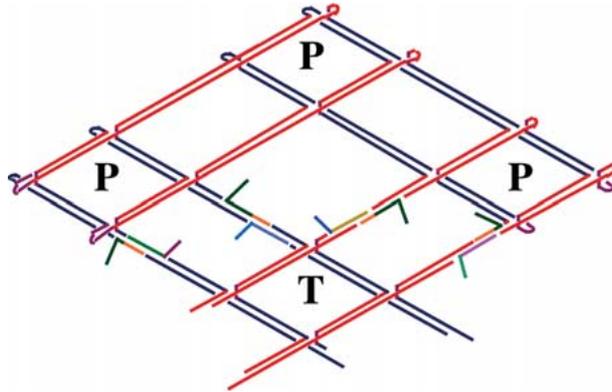


Figure 11. Generalized protection. The triple tile unit whose corner tiles are labeled 'P' forms the protecting group for the tile labeled 'T'. It covers the sticky ends that should not react if the T tile is to act as a corner tile without other pairings that might adulterate the structure. Note that the actual protection is accomplished by four short colored strands that have differently colored unpaired extensions. The colored strand is complementary both to a target region on the protecting strand, and to a sticky end on the T tile. These protecting strands are to be removed by Yurke et al.-type (2000) full complements that bind to the extensions and then remove the rest of the strand by branch migration. Different extensions would allow for a standard protecting group that could be combined with the same sequence in different loci, and would enable the selective removal of particular protecting strands. The orange-green strands increase the generality of the construction. They bind to the protecting group on one side (the green side) and to the colored strands on the other side (the orange side). Thus, generic protecting groups can be used in all cases. The orange-green strands may be removed by the same techniques as the other strands.

Whenever the assembly asks for the use of intermediary tiles of type **56** or **56** that have four sites coded  $\mathbf{a}^*$ , then we need to be a bit more careful in assembling the fractal layer. This is because some ambiguity might appear in their assembly with tiles of type **1**. Figure 9c illustrates an example that occurs in the assembly of fractal layer **5'**: To avoid the ambiguity coming from two different sites of **56**, one needs to use protecting groups for tile **56** to avoid the undesired site. The intermediate tile is then put in a tube with tile **3b** equipped by its protection, and the assembly is formed. See Figure 10. The protection on the intermediate tile **56** is then taken away to allow the completion of the fractal layer: Tile **2b** hybridizes to tile **56** and so on. Here, the removal of the protection on **56** needs to be specific and one can achieve this as discussed at the end of section “Parallelogram protection”.

To conclude, let us notice that there are variations of tile **1** which have opposite sides made of helices lying on the same plane and coded  $\mathbf{a}$  and  $\mathbf{a}^*$ . It might seem that such sides can glue to each other, but in fact their geometry does not allow this to happen, being one side the mirror image of the other. Variation 56 of tiles **5** and **6** has opposite helices with identical code.

## 6. Discussion

In a fractal assembly, the tiles produced at larger scales should satisfy the properties of the basic tile boundaries. This means that the properties of the coding and of the geometrical shape of the boundaries (if any) should be passed on at larger scales. Given the Sierpinski square fractal, one could ask what is the minimum number of square tiles that can assemble uniquely into a desired shape when the only boundary conditions concern the coding. The design of such individual tiles with “differently-labeled” sides is not at all obvious and it would be a challenge to verify that such a task is possible at all. We show that by combining geometrical shapes *and* coding we can build a fractal assembly with the 16 basic tiles in Figure 2 and their mates in Figure 3. The design of such a set of tiles and of suitable protections requires satisfying adequate *closure* conditions: a set of parallelogram tiles  $N \subset m$  is *closed* with respect to a property  $P$  if all tiles in  $N$  satisfy  $P$  and all tiles in  $M$  satisfying  $P$  are in  $N$ . For us,  $N$  is one of the sets **1**, **5** or **6** and the property  $P$  concerns the *geometry* of the boundaries (i.e., corners of tiles have geometrically different shapes) and the *coding* of their sides. The set of variations of tile **1**, for instance, is constituted by tiles with a fixed boundary shape, and where the coding is such that a pair of consecutive edges are labeled  $\mathbf{a}$ , and the remaining two edges are labeled either  $\mathbf{a}$  or  $\mathbf{a}^*$ . In a similar spirit, one defines protecting groups by requiring closure on the *geometrical shape* of pairs of consecutive edges. Using the notion of closure, the reader can formally show

that fractals  $X^k$  of dimension  $k > 1$  sharing the characteristics (i.e., coding and geometrical shape) of a tile  $X$  in the set of basic tiles, can always be constructed.

We have noticed above that the lengths of the extensions can vary between the two layers of helices, to create differently shaped fractal layers, not just rhomboidal constructs. As we have illustrated in a variety of assemblies, the only rotations required of the individual tiles are  $180^\circ$  rotations. Consequently, the lengths of the two directions within the parallelogram are also independent of each other, so that the parallelograms themselves are not restricted to be rhomboids. They may have any ratio of edges between their vertices compatible with an integral number of half-turns of DNA.

The way that we ensure there is no interaction between the two layers utilizes a chemical difference between the two  $5'$  end and the  $3'$  end. Although the sequences are the same, the presence of material  $5'$  or  $3'$  to the overhangs prevents interaction between the two types of sticky ends. It is important to realize that the  $3'$  and  $5'$  sticky ends are used in different directions, either horizontal or vertical but never mixed; this happens because the only rotations we use are  $180^\circ$  rotations perpendicular to the plane of the parallelograms. If one were to regard these ends as being different, then we would not have a single label,  $\mathbf{a}$  and its complement,  $\mathbf{a}^*$ , but in fact two different labels,  $\mathbf{a}$  and  $\mathbf{b}$  and their complements  $\mathbf{a}^*$  and  $\mathbf{b}^*$ . Initial experimental attempts to implement this system are likely to use only the  $3'$  and  $5'$  differences, without resorting to two different types of sequences; the extra level of control possibly available from a second coding word does not appear necessary. However, if this system were to be extended to 3D, the differences between  $5'$  and  $3'$  sticky ends would not be enough. A second (and perhaps for generality a third) type of coding would be required. Related 3D fractal systems will be discussed elsewhere.

The interdomain angle found in Holliday junctions has been measured by AFM to be  $\sim 63^\circ$  (Mao et al., 1999). However, changing the sequence that flanks the site can decrease the angle by up to  $20^\circ$  (Sha et al., 2002). The Bowtie junction (Sha et al., 1998, 2000) is closely related to the Holliday junction, but its two crossover strands contain  $5', 5'$  and  $3', 3'$  linkages, instead of the conventional  $5', 3'$  linkages found in conventional DNA. The Bowtie junction has an interdomain angle around  $-67^\circ$  ( $113^\circ$ ), which approximately changes the  $\alpha$  and  $\beta$  angles to their supplements, interconverting obtuse and acute angles. Thus, a closely related shape is available from the scheme outlined here.

Recently, parallelogram tiles have been produced with helix axes lying in the same plane (Yan and Seeman, 2002). Their implementation in this context appears technically more complicated, because only a single double helical

extension in each direction would exist. As a result, it would not be possible to use the double-cohesion of two helices to enforce the fidelity of sticky ends. The advantage of using “planar” parallelograms is that the complex combinatorial aspects of the solution decrease: The number of tiles used is 16, the number of protecting corners is four and only two tile types are needed to act as protecting groups.

Fractal constructions can be thought of as a way to generate fixed geometrical shapes of controlled size. Winfree proposed a schema to build periodic arrays of size  $n \times m$  by generating repeatedly the Boolean truth table for  $n$  entries until  $m$  rows of the table have been filled (Winfree, 1996, 2000). One would like to have a way to do the same with other shapes other than rectangles, such as triangles, hexagons, etc. Fractal assembly allows constructing fixed sizes that are powers of some value: for instance, for the Sierpinski fractal, the size of the squares is  $3^k$ , where  $k$  is the dimension. We will discuss elsewhere the construction of fractal arrangements from other shapes.

A variety of periodic and aperiodic constructions have been proposed (Winfree, 1996, 2000) and executed from the components of DNA nanotechnology. Those reduced to practice include periodic arrays built from double crossover (DX) molecules (Winfree et al., 1998; Liu *et al.*, 1999), triple crossover molecules (TX) (LaBean et al., 2000), conventional (Mao et al., 1999; Sha et al., 2002) and Bowtie parallelograms (Sha et al., 2000), and an aperiodic assembly that performs a cumulative XOR calculation (Mao et al., 2000). All of these constructions use uniformly shaped tiles that rely exclusively on the specificity of sticky ended cohesion for their assembly. The fractal assembly proposed here relies both on cohesion and on molecular shape (long and short extensions) to establish complementarity. Although related to the geometrical constraints used in the cumulative XOR calculation, these geometrical determinants are fundamentally new, and are likely to present a new set of principles and problems in DNA nanotechnology.

### Acknowledgments

This research has been supported by grants GM-29554 from the National Institute of General Medical Sciences, N00014-98-1-0093 from the Office of Naval Research, grants DMI-0210844, EIA-0086015, DMR-01138790, and CTS-0103002 from the National Science Foundation, and F30602-01-2-0561 from DARPA/AFSOR.

## References

- Carbone A and Seeman NC (2002) A route to fractal DNA assembly. *Natural Computing* 1: 469–480
- Caruthers MH (1985) Gene synthesis machines: DNA chemistry and its uses. *Science* 230: 281–285
- Churchill MEA, Tullius TD, Kallenbach NR and Seeman NC (1988) A Holliday recombination intermediate is twofold symmetric. *Proc. Nat. Acad. Sci. (USA)* 85: 4653–4656
- Cohen SN, Chang ACY, Boyer HW and Helling RB (1973) Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Nat. Acad. Sci. (USA)* 70: 3240–3244
- LaBean T, Yan H, Kopatsch J, Liu F, Winfree E, Reif JH and Seeman NC (2000) The construction, analysis, ligation and self-assembly of DNA triple crossover complexes. *J. Am. Chem. Soc.* 122: 1848–1860
- Li X, Yang X, Qi J and Seeman NC (1996) Antiparallel DNA double crossover molecules as components for nanoconstruction. *J. Am. Chem. Soc.* 118: 6131–6140
- Liu F, Sha R and Seeman NC (1999) Modifying the surface features of two-dimensional DNA crystals. *J. Am. Chem. Soc.* 121: 917–922
- Mao C, Sun W and Seeman NC (1999) Two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *J. Am. Chem. Soc.* 121: 5437–5443
- Mao C, LaBean T, Reif JH and Seeman NC (2000) Logical computation using algorithmic self-assembly of DNA triple crossover molecules. *Nature* 407: 493–496
- Nielsen PE, Egholm M, Berg RH and Buchardt O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254: 1497–1500
- Seeman NC (1982) Nucleic acid junctions and lattices. *J. Theor. Biol.* 99: 237–247
- Seeman NC (1992) The design of single-stranded nucleic acid knots. *Molec. Eng.* 2: 297–307
- Sha R, Liu F, Bruist MF and Seeman NC (1998) Parallel helical domains in DNA branched junctions containing 5', 5' and 3', 3' linkages. *Biochemistry* 38: 2832–2841
- Sha R, Liu F, Millar DM and Seeman NC (2000) Atomic force microscopy of parallel DNA branched junction arrays. *Chem. & Biol.* 7: 743–751
- Sha R, Liu F and Seeman NC (2002) Atomic force measurement of the interdomain angle in symmetric Holliday junctions. *Biochem.* 41: 5950–5955
- Winfree E (1996) On the computational power of DNA annealing and ligation. In: Richard Lipton J and Baum EB (eds) *DNA Based Computers: Proceedings of a DIMACS Workshop, April 4, 1995, Princeton University (Volume 27 in DIMACS)*, pp. 187–198. American Mathematical Society
- Winfree E (2000) Algorithmic self-assembly of DNA: Theoretical motivations and 2D assembly experiments. *J. Biol. Mol. Struct. & Dyns. Conversation* 11(2): 263–270
- Winfree E, Liu F, Wenzler LA and Seeman NC (1998) Design and self-assembly of two-dimensional DNA crystals. *Nature* 394: 539–544
- Yan H, Zhang X, Shen Z and Seeman NC (2002) A robust DNA mechanical device controlled by hybridization topology. *Nature* 415: 62–65
- Yan H and Seeman NC (2002) Edge-sharing motifs in DNA nanotechnology. *J. Supramol. Chem.*, in press
- Yurke B, Turberfield AJ, Mills AP, Jr., Simmel FC and Neumann JL (2000) A DNA-fuelled molecular machine made of DNA. *Nature* 406: 605–608
- Zhang Y and Seeman NC (1992) A solid-support methodology for the construction of geometrical objects from DNA. *J. Am. Chem. Soc.* 114: 2656–2663

