Molecular Tiling and DNA Self-assembly

Alessandra Carbone[1] and Nadrian C. Seeman[2]

[1] Institut des Hautes Études Scientifiques
35, route de Chartres, 91440 Bures-sur-Yvette, France
carbone@ihes.fr
[2] Department of Chemistry, New York University
New York, NY 10003, USA
ned.seeman@nyu.edu

Abstract. We examine hypotheses coming from the physical world and address new mathematical issues on tiling. We hope to bring to the attention of mathematicians the way that chemists use tiling in nanotechnology, where the aim is to propose building blocks and experimental protocols suitable for the construction of 1D, 2D and 3D macromolecular assembly. We shall especially concentrate on DNA nanotechnology, which has been demonstrated in recent years to be the most effective programmable self-assembly system. Here, the controlled construction of supramolecular assemblies containing components of fixed sizes and shapes is the principal objective. We shall spell out the algorithmic properties and combinatorial constraints of "physical protocols", to bring the working hypotheses of chemists closer to a mathematical formulation.

1 Introduction to Molecular Self-assembly

Molecular self-assembly is the spontaneous organisation of molecules under thermodynamic equilibrium conditions into a structurally well-defined and rather stable arrangement through a number of non-covalent interactions [5,26,52]. It should not be forgotten that periodic self-assemblies of molecules lead to crystals in one, two or three dimensions; we often do not understand the interactions between the constituents of a crystal, but their presence in our world was an existence-proof for 3D self-assembly long before the notion was voiced. By a *non-covalent* interaction, we mean the formation of several non-covalent weak chemical bonds between molecules, including hydrogen bonds, ionic bonds and van der Waals interactions. These interactions (of the order of 1-5 kcal/mol) can be considered *reversible* at normal temperatures, while covalent interactions (typically > 50 kcal/mol) are regarded as *irreversible*.

The self-association process leads the molecules to form stable hierarchical macroscopic structures. Even if the bonds themselves are rather weak, their collective interaction often results in very stable assemblies; think, for example, of an ice cube, held together by hydrogen bonds. Two important elements of molecular self-assembly are *complementarity* and *self-stability*, where both the size and the correct orientation of the molecules are crucial in order to have a complementary and compatible fitting.

The key engineering principle for molecular self-assembly is to design molecular building blocks that are able to undergo spontaneous stepwise interactions so that they self-assemble via weak bonding. This design is a type of "chemical programming", where the instructions are incorporated into the covalent structural framework of each molecular component, and where the running of the algorithm is based on the specific interaction patterns taking place among the molecules, their environment, and the intermediate stages of the assembly. The aim of the game is to induce and direct a controlled process.

Molecular self-assembly design is an art and to select from the vast virtual combinatorial library of alternatives is far from being an automatic task [19]. There are principles though, that could be mathematically analyzed and one of the purposes of this paper is to lead the reader towards such possibilities. We shall talk mainly about self-assembly from branched DNA-molecules, which in the last few years have produced considerable advances in the suggestion of potential biological materials for a wide range of applications [39]. Other directions using peptides and phospholipids have been also pursued successfully [57,35,4].

We shall start with an abstract overview of some of the principles governing self-assembly which have been investigated by chemists (for an introduction see also [27]), with a special emphasis on DNA self-assembly. With the desire to formalise in an appropriate mathematical language such principles and develop a combinatorial theory of self-assembly, we try to suggest mathematical structures that arise naturally from physical examples. All through the paper, we support our formalistic choices with experimental observations. A number of combinatorial and algorithmic problems are proposed. The word "tile" is used throughout the paper in a broad sense, as a synonym of "molecule" or of "combinatorial building block" leading to some assembly.

2   Examples of Molecular Self-assembly and Scales

Self-assembled entities may be either discrete constructions, or extended assemblies, potentially infinite, and in practice may reach very large sizes. These assemblies include such species as 1-dimensional polymolecular chains and fibers, or 2-dimensional layers and membranes, or 3-dimensional solids. Due to the exceptionally complicated cellular environment, the interplay of the different ligand affinities and the inherent complexity of the building blocks, it is not easy to predict, control and re-program *cellular* components. Proteins can in principle be engineered but to predict protein conformation is far from our grasp nowadays. At the other extreme lie chemical assemblies, such as organic or inorganic crystals, which are constituted by much simpler structural components that are not easily programmed. Within this spectrum of assembly possibilities, DNA self-assembly has revealed itself as the most tractable example of programmable molecular assembly, due to the high specificity of intermolecular Watson-Crick base-pairing, combined with the known structure formed by the components when they associate [31]. This has been demonstrated in recent years both theoretically and experimentally as we shall discuss later.

3   Molecular Self-assembly Processes

There are three basic steps that define a process of molecular self-assembly:

1. *molecular recognition*: elementary molecules selectively bind to others;
2. *growth*: elementary molecules or intermediate assemblies are the building blocks that bind to each other following a sequential or hierarchical assembly; cooperativity and non-linear behavior often characterize this process;
3. *termination*: a built-in halting feature is required to specify the completion of the assembly. Without it, assemblies can potentially grow infinitely; in practice, their growth is interrupted by physical and/or environmental constraints.

Molecular self-assembly is a *time-dependent process* and because of this, temporal information and kinetic control may play a role in the process, before thermodynamic stability is reached. For example, in a recent algorithmic self-assembly simulating a circuit constituted by a sequence of XOR gates [30], a template describing the input for the circuit, assembled first from DNA tiles as the temperature was lowered, because these tiles were programmed to have stronger interactions; the individual tiles that performed the gating functions, i.e. the actual computation of each XOR gate, assembled on the template later (at a lower temperature), because they interacted more weakly. If, as in this example, the kinetic product is an intermediate located on the pathway towards the final product, such a process is *sequential*. If not, then the process is said to *bifurcate*.

Molecular self-assembly is also a *highly parallel process*, where many copies of different molecules bind simultaneously to form intermediate complexes. One might be seeking to construct many copies of the same complex at the same time, as in the assembly of periodic 1D or 2D arrays; alternatively, one might wish to assemble in parallel different molecules, as in DNA-based computation, where different assemblies are sought to test out the combinatorics of the problem [1,22]. A sequential (or *deterministic*) process is defined as a sequence of highly parallel instruction steps.

Programming a system that induces *strictly sequential assembly* might be achieved, depending on the sensitivity of the program to perturbations. In a robust system, the instructions (that is the coding of the molecular interactions) are strong enough to ensure the stability of the process against interfering interactions or against the modification of parameters. Sensitivity to perturbations limits the operational range, but on the other hand, it ensures control on the assembly.

An example of *strong instructions* is the "perfect" pairing of strands of different length in the assembly of DNA-tiles due to Watson-Crick interacting sequences. The drawback in sequential assembly of DNA-tiles is due to the complex combinatorics of sequences which are needed to construct objects with discrete asymmetric shapes, or aperiodic assemblies. The search for multiple sequences which pair in a controlled way and avoid unwanted interactions is far from being

an obvious task. Alternative approaches concern besides tile design, self-assembly algorithms and protocols (Section 7).

A sequential process might either be *commutative*, if the order of the assembly steps can be interchanged along the pathway leading to the final assembly, or it might be non-commutative, if the intermediates need to interact in a fixed consecutive manner. DNA-based computations, such as the assembly of graphs [34] are commutative: a series of branched junctions can come together in any order to yield the final product (as discussed in Section 6 for 3-color graphs). An example of a non-commutative process is the construction of DNA tiles along the assembly of a periodic 2D array: single stranded DNA sequences are put in a pot at once, and since the tiles melt at a temperature higher than the intermolecular interactions, tiles are "prepared" first, before the 2D assembly takes place. Even if indirectly, these physical conditions imply non-commutativity. Later on, the 2D lattice can assemble with gaps that can later be filled in from the 3rd direction. Commutativity, in this latter step, may create irregularities when 3D arrays are considered instead, since gaps might get sealed in as a defect. Any hierarchical construction, such as solid-support-based DNA object synthesis [58] is non-commutative. Another example of a non-commutative assembly is a frame-based construction [32], wherein an assembly is templated by a "frame" that surrounds it: tiles assemble within the boundaries of the frame and they are guided by the code of the tiles forming the frame. It is non-commutative, in that the frame has to be available first.
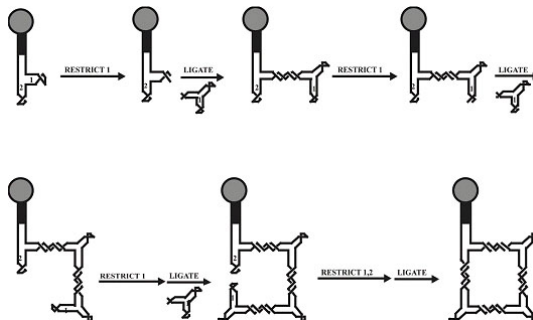

Fig. 1. *Protocol for the Synthesis of a Quadrilateral.* The intermolecular additions of corners is repetitive, but a different route leads to intramolecular closure.

Another characteristic of a molecular self-assembly is that the *hierarchical build-up* of complex assemblies, allows one to intervene at each step, either to suppress the following one, or to orient the system towards a different pathway. For example, the construction of a square from identical units using the solid-support method entailed the same procedures to produce an object with 2, 3, or 4 corners. Once the fourth corner was in place, a different pathway was taken to close the square [58], as shown in Figure 1. A pentagon, hexagon or higher
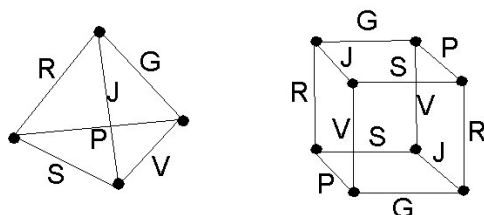
Fig. 2. Triplet junctions GPV, JRV, JGS and PSR can combine in different configurations. The two smallest ones are a tetrahedron and a cube.

polygon could have been made by the same procedure, just by choosing to close it after adding more units.

Instructions might be strong but still allow for different objects to appear. The same set of tiles might assemble into objects with different geometrical shapes and different sizes, that satisfy the same designed combinatorial coding. For instance, consider chemical "three-arm junction nodes" (name them GPV, JRV, JGS, PSR) accepting 6 kinds of "rods", called G, P, V, J, S and R. Several geometrical shapes can be generated from these junctions and rods, in such a way that all junctions in a node are occupied by some rod. Two such shapes are illustrated in Figure 2. In general, there is no way to prevent a given set of strands from forming dimers, trimers, etc. Dimers are bigger than monomers, trimers bigger than dimers, and so on, and size is an easy property for which to screen. However, as a practical matter, entropy will favor the species with the smallest number of components, such as the tetrahedral graph in Figure 2; it can be selected by decreasing the concentration of the solution. If, under convenient conditions, a variety of products results from a non-covalent self-assembly, it is possible to obtain only one of them by converting the non-covalent self-assembly to a covalent process (e.g., [16]). Selecting for specific shapes having the same number of monomers though, might be difficult. It is a combinatorial question to design a coding for a set of tiles of fixed shape, that gives rise to an easily screenable solution set.

4   Molecular Tiling: A Mathematical Formulation

Attempts to describe molecular assembly, and in particular DNA self-assembly, in mathematical terms have been made in [2,6]. Here, we discuss some algorithmic and combinatorial aspects of self-assembly keeping in mind the physics behind the process.

Tiles and self-assembly. Consider a connected subset T (tile) in $\mathbb{R}^3$, for example a convex polyhedron, with a distinguishable subset of mutually complementary (possibly overlapping) non-empty domains on the boundary, denoted $D_b, D'_b \subset \partial T$, where b runs over a (possibly infinite) set B. We are interested in assemblies generated by T, that are subsets A in the Euclidean space, decomposed into a
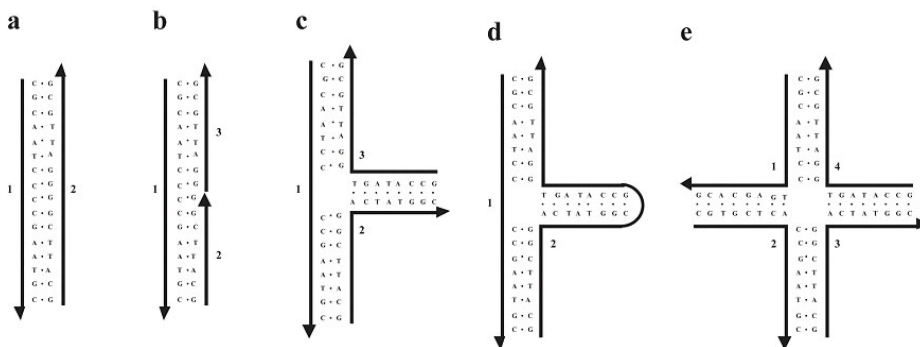
Fig. 3. *A Variety of Complements to a Single Strand.* Panel (a) illustrates a conventional Watson-Crick duplex, where strand 2 complements strand 1. Panels (b-e) illustrates a variety of complements to strand 1.

union of congruent copies of T, where two copies may intersect only at their boundaries and have a "tendency" to meet across complementary domains on the boundary. It is important to recognize that in the case of DNA, there are many forms of complementarity, as a function of motif structure [41]. Figure 3 illustrates a DNA strand (named 1) complementary to a variety of other DNA strands; more complex types of complementarity exist, such as complementarity in the PX sense [60,46] or in the anti-junction sense [12,60].

We want to consider a biological macromolecule T (e.g., a protein or a nucleic acid motif), with complementary binding sites D_b, D_b' such that different copies of T bind along complementary domains and self-assemble into complexes. In the geometric context we specify the binding properties by introducing (binding) isometries $b : \mathbb{R}^3 \to \mathbb{R}^3$ to each $b \in B$ such that T and $b(T)$ intersect only at the boundary, and $b(D_b) = D_b'$. From now on B is understood as a subset in the Euclidean isometry group $Iso(\mathbb{R}^3)$.

Accordingly, we define an *assembly A associated with* (T, B) by the following data:

1. a connected graph $G = G_A$ with the vertex set $1 \dots N$,
2. subsets T_i in $\mathbb{R}^3$, where $i = 1 \dots N$, which may mutually intersect only at their boundaries,
3. an isometry $b_{k,l} : \mathbb{R}^3 \to \mathbb{R}^3$ moving T_k onto T_l, for each edge (k, l) in G, such that there exists an isometry $a_{k,l}$ which moves T_k to T and conjugates $b_{k,l}$ to some binding isometry b in B. Notice that this b is uniquely determined by $b_{k,l}$ up to conjugation.

Given a graph G_A and a tile T, the assemblies described by G_A and T might not be unique. The assembly is unambiguously described by the isometries associated to the edges of G_A (i.e. condition (3) above). See Figure 4 for an example.
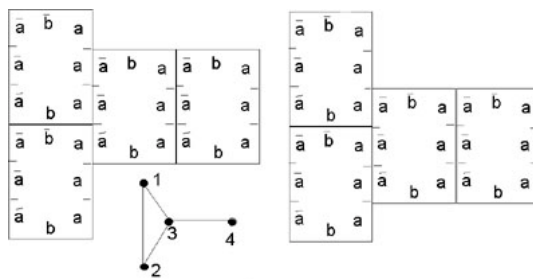
Fig. 4. Four copies of the same tile are arranged in two different assemblies that correspond to the same graph G_A. The labels $a, \bar{a}, b, \bar{b}$ correspond to codes for edges.

Several tiles. If we start with several different tiles $T^1, \ldots, T^n$ rather than with a single T, we consider the sets of pairs of binding isometries $B^{i,j} \subset Iso(\mathbb{R}^3) \times Iso(\mathbb{R}^3)$ such that $b_1^{i,j}(T^i)$ and $b_2^{i,j}(T^j)$ intersect only at their boundaries and their intersection is non-empty. The definition of an assembly associated to $(\{T^i\}, \{B^{i,j}\})$ goes as above with the following modifications: the graph G has vertices colored by the index set $1 \ldots n$, the corresponding subsets in $\mathbb{R}^3$ are denoted T_k^i where $i = 1 \ldots n$ and $k = 1 \ldots N_i$, and finally, we forfeit the isometries $b_{k,l}$ and for each edge (k^i, l^j) we emphasize an isometry of $\mathbb{R}^3$ which moves T_k^i to $b_1^{i,j}(T^i)$ and T_l^j to $b_2^{i,j}(T^j)$.

In what follows, we refer to the union of tiles defined above, as an assembly.

Qualities of an assembly. The *tightness* of the tiling is one quality that chemists appreciate. This can be measured by the number of cycles in the graph G, or equivalently by the negative Euler characteristic of the graph.

The *imperfection* of a tiling is measured by the "unused" areas of the boundaries of the tiles. First define the *active domain* $\partial_{act}(T) \subset \partial T$ as the union of the intersections of ∂T with $b(T)$ for all $b \in B$. Then define the "unused boundary" $\partial_{un}(A = \cup T_i)$ as the union $\cup_{i=1}^N \partial_{act}(T_i)$ minus the union of the pairwise intersections $\cup_{(k,l) \in G} T_k \cap T_l$. An assembly is called *perfect* if the area of the imperfection equals zero. We say that an assembly contained in a *given* subset $X \subset \mathbb{R}^3$ is *perfect* with respect to ∂X, if $\partial_{un}(A) \subset \partial X$.

The *uniqueness* refers to the uniqueness of an assembly subject to some additional constraints. For example, given an $X \subset \mathbb{R}^3$, one asks first if X can be tiled by (T, B) and then asks for the uniqueness of such a tiling. We say that (T, B) generates an *unconditionally unique* assembly if every imperfect assembly uniquely extends to a perfect assembly.

The essential problem of tiling engineering is designing a relatively simple tile or a few such tiles which assemble with high quality into large and complicated subsets in $\mathbb{R}^3$. Here is a specific example for the unit sphere S^2 rather than S^3, where one uses the obvious extension of the notion of tilings to homogeneous spaces. Given $\epsilon, \delta > 0$, consider triangulations of the sphere into triangles Δ with $Diam(\Delta) \leq \epsilon$ and $area(\Delta) \geq \delta Diam^2(\Delta)$. It is easy to see that the number of

mutually non-congruent triangles in such a triangulation, call it $n(\epsilon, \delta)$, goes to ∞ for $\epsilon \to 0$ and every fixed $\delta > 0$. The problem is to evaluate the asymptotic behavior of $n(\epsilon, \delta)$ for $\epsilon \to 0$ and for either a fixed δ or $\delta \to 0$.

Complementarity of the domains. Two tiles T_1 and T_2 have complementary sites, D_b^1, D_b^2, when they can bind along their boundaries to each other forming a connected subset of $\mathbb{R}^3$. In physical terms, the two overlapping parts D_b^1, D_b^2 can have complementary geometrical shape (e.g. think of the concave surface of a protein and of the convex surface of a ligand binding to it, much as a classical 'lock and key'), but might also correspond to Watson-Crick complementary sequences (e.g. $5' - ATTCGA - 3'$ and $3' - TAAGCT - 5'$, where A is complementary to T and C to G as discussed before; see Figure 3).
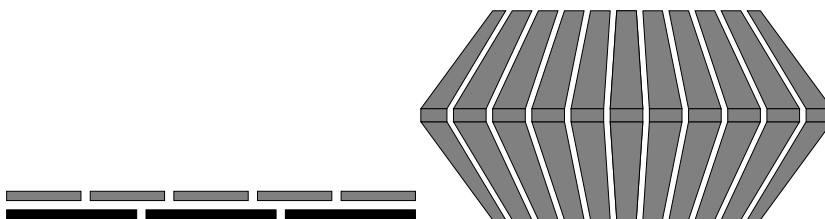

Fig. 5. Left: Rodlike tiles differing in length form an assembly that grows until the ends exactly match. Right: polymeric structure growing until the energy required to fit new subunits becomes too large.
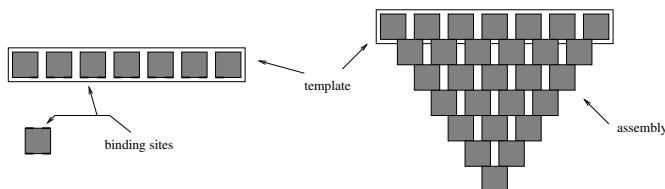

Fig. 6. A tile is stable in the assembly only if it binds at two adjacent binding sites. The stability of the whole assembly is insured by the enforced stability of the template. The formal description of this example is not completely captured by our model.

Real life examples. It remains unclear, in general, how cells control the size of (imperfect, with some unused boundary) assemblies, but certain mechanisms are understood. For example, out of two rod-like molecules of length three and five, one gets a double rod of length 15 as illustrated in Figure 5 (left). Another strategy is starting an assembly from a given template (see Figure 6 for a specific
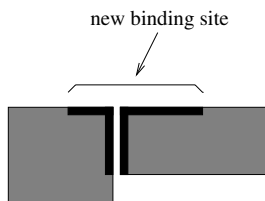
Fig. 7. Tiles which differ in shape and binding sites. Their binding generates a new contiguous binding site.

design). Sometimes, tiling is non-isometric: tiles distort slightly in order to fit, and the assembly terminates when the bending energy becomes too costly or when the accumulated distortion deforms and deactivates the binding sites (see Figure 5 (right)). Also, the binding of a ligand to an active site might change the shape of the molecule and thus influence the binding activity of other sites. Another possibility is the creation of a new binding site distributed over two or more tiles bound together at an earlier stage of the assembly (see Figure 7). These mechanisms may produce a non-trivial dynamics in the space of assemblies in the presence of free-energy. In particular, one may try to design a system which induces a periodic motion of a tile over a template, something in the spirit of RNA-polymerase cycling around a circle of DNA [11].
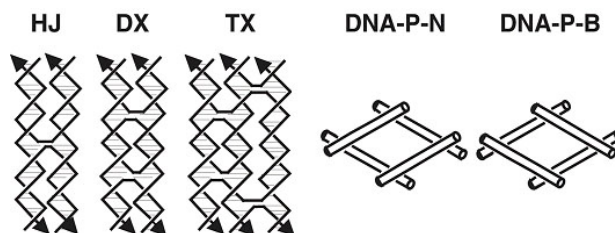

Fig. 8. Key Motifs in Structural DNA Nanotechnology. On the left is a Holliday junction (HJ), a 4-arm junction that results from a single reciprocal exchange between double helices. To its right is a double crossover (DX) molecule, resulting from a double exchange. To the right of the DX is a triple crossover (TX) molecule, that results from two successive double reciprocal exchanges. The HJ, the DX and the TX molecules all contain exchanges between strands of opposite polarity. To the right of the TX molecule are a pair of DNA parallelograms, DNA-P-N [29], constructed from normal DNA, and DNA-P-B [43], constructed from Bowtie junctions, containing 5', 5' and 3', 3' linkages in their crossover strands.

DNA tiles and tensegrity. Molecules of some nanometer size, made out of DNA strands, have been proposed in a variety of different shapes. See Figure 8 for

a representative collection of shapes. Algorithms have been developed successfully to produce that self-assemble into these and other motifs [36]. *Branched molecules* are tiles constituted by several single strands which self-assemble along their coding sequences in a "star-like" configuration, where a tip of the star is a branch [36,51,38] (Figure 3a,c,e illustrate 2, 3 and 4-arm branched molecules). Theoretically, one might think to construct k-armed branched molecules, for any $k > 2$, where each strand is paired with two other strands to form a pair of double-helical arms; in practice, molecules with 6 arms have been reported, but larger ones are under construction. The angles between the arms are known to be flexible in most cases. If one adds *sticky ends* to a branched molecule, i.e. single stranded extensions of a double helix, a cluster is created that contains specifically addressable ends [36]. This idea is illustrated in Figure 9, where a 4-arm branched junction with two complementary pairs of sticky ends self-assembles to produce a quadrilateral.
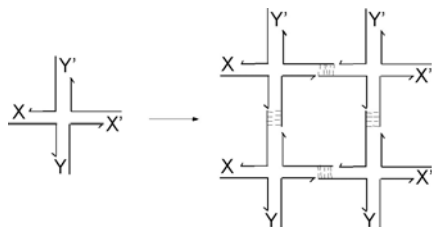

Fig. 9. Formation of a 2-dimensional lattice (right) from a 4-arm branched junction (left). X is a sticky end and X' is its complement. The same relationship holds for Y and Y'. X and Y are different from each other.

The search for motifs based on DNA branched junctions that behave as though they are "rigid" while in the test tube, led to the design of several DNA-molecules, and some are illustrated in Figure 8. Rigid shapes impose strong limitations on the design of suitable molecular tiles; roughly speaking, a *rigid*, or *tense*, object is a 3-dimensional solid that does not undergo deformations: we ask that if its 1-dimensional faces do not undergo deformation, then no deformation exists. For a tetrahedron or any convex deltahedron, it is easy to see that no change of the angles between edges (edges are 1-dimensional faces for the tetrahedron) can take place without the edges be deformed. On the other hand, a cube is an example of a non-tense object since we can fix the edges (1-faces) of the cube not to undergo any deformation and still be able to deform the angles between them.

Geometry of the boundaries: smooth deformations of tiles. It might be appropriate to consider assemblies which are affected by an ϵ-deformation in the shape of the tiles after binding. More precisely, a tile $T \subseteq \mathbb{R}^3$ is mapped in $\mathbb{R}^3$ by some ϵ-*deformation* as follows: there is an embedding $\epsilon : T \subseteq \mathbb{R}^3 \mapsto T' \subseteq \mathbb{R}^3$ such that for all points $x \in T$ there is a point $y \in T'$ such that the Euclidean distance

$d(x, y) < \epsilon$. The definitions of isometry and binding site given at the beginning of Section 4 need to be adjusted accordingly into new notions of ϵ-isometry and ϵ-binding site, which intuitively correspond to the original notions up to some ϵ-variation. One needs to establish whether an ϵ-deformation affects a binding site or not, and give thresholds on the amount of deformation which is accepted to affect non-empty domains of the boundary.

The growth of the assembly affected by ϵ-deformation asks for the estimation of bounds in the size of the construction. The instability of the system comes from a narrow range of conditions on which the assembly takes place. The formation of singularities and of bifurcation points between different assemblies, might lead to the disruption of the assembly, but might also lead to variety in the complexity.

Physical considerations on the shape of tiles. In addition to the need to observe appropriate solution conditions that encourage self-assembly, it is important to realize that there are physical constraints on the assembly of real tiles that do not affect virtual tiles. For example, the helicity of random-sequence DNA is ≈ 10.5 nucleotide pairs per turn in solution. This value makes it easy to make TX molecules (Figure 8) whose three helix axes form an angle of 120°, but 90° is much harder, unless one is able (perhaps through sequence variation) to change the repeat to 10.4 nucleotide pairs per turn [45].

In a similar vein, a likely form of shaped 3D arrays will entail polyhedra whose edges contain DX molecules (Figure 8) [40]. It might appear that a tetrahedron would be a good polyhedron to use as the basis of such a 3D tile. However, although the edges of a tetrahedron obviously span 3-space, there is no group of three edges to which one can attach a single extra helix (i.e. to make those edges DX molecules instead of single DNA helices, with the extra helices outside the helices defining the tetrahedron) to produce the needed vectors: their diameters would cause them to clash stereochemically when extended beyond the boundaries of the tetrahedron. Notice that extra-hedral domains on adjacent edges inherently clash, and there is no group of three edges in a tetrahedron that does not include an adjacent pair.

5   An Abstract Model to Describe the Dynamics of Self-assembly

A formal description of the dynamics of a self-assembly on a space S, where S can be either $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$ or any discrete approximation of those, can be formulated by a simple iterative process as follows. Consider n tiles $T_1, \ldots, T_n$, and take a finite number of copies of each T_i, for all $i = 1 \ldots n$. At stage 1, randomly assign to each physical tile a specific position in S in such a way that no two tiles overlap and that only tiles lying side-by-side and having complementary boundary stuck together. The set of complexes containing more than one tile with their position in S, define a *configuration* on S; single tiles are removed from S and used to re-iterate the random assignment on the next stage: the configuration of tiles lying in S which one reaches at stage i, is filled up further by new non-overlapping complementary adjacent tiles at stage $i + 1$. The process is repeated until all

tiles are used or when a sufficiently large connected area in S is filled (e.g. area $> N$, for some large N).

Different outputs might result from this random process: they go from very tight assemblies, to assemblies with several unfilled regions, to disconnected surfaces, and so on. The resulting configurations and the time for reaching a configuration strongly depend on the coding hypothesis, e.g. whether new binding sites can appear or not by the combination of several tiles, whether "holes" can be filled or not, how many different competing boundary sites are in the system, how many tiles are in the system, whether connected regions can undergo translations in S while the process takes place, whether connected tiles might become disconnected, etc.

The process could start from a specific configuration of S instead of using the first iteration step to set a random configuration. Such an initial configuration, if connected, would play the role of a *template* for the random process described by the iteration steps. Figure 6 illustrates an example of templating, where a 1-dimensional array of n molecules disposed in a row is expected to play the role of a template and interact with single molecules following a schema coded by the boundary of the tiles. Another example is the "nano-frame" proposed in [32], a border template constraining the region of the tiling assemblies.

6   Closed Assemblies and Covering Graphs

Closed tiling systems are assemblies whose binding sites have been all used. More formally, this amounts to saying that the graph G_A underlying the assembly A is such that all tiles T_i corresponding to its vertices, where $i = 1 \ldots N$, intersect on all their boundary sites T_i. This means also that the degree of connection of each node i of G_A corresponds to the number of available interaction sites of T_i, and that each edge (i, l) departing from i corresponds to an isometry $b_{i,l}$ moving T_i onto T_l which fixes some binding site. Many graphs might be locally embeddable in G_A, and we call them *covering graphs* of G_A: a graph G is a covering graph of G_A if there is a map $p \ : \ G \to G_A$ such that

1. p is surjective, i.e. for all nodes $y \in G_A$ there is a node $x \in G$ such that $p(x) = y$,
2. if $x \to y$ in G then $p(x) \to p(y)$ in G_A,
3. $degree(x) = degree(p(x))$, for all nodes x in G,
4. $b_{x,y} = b_{p(x),p(y)}$, for each edge $x \to y \in G$,
5. $\{b_{x,y} \ : \ x \to y \in G\} = \{b_{p(x),z} \ : \ p(x) \to z \in G_A\}$, for each node $x \in G$.

Condition (4), saying that the binding site between T_x and T_y is the same as the binding site between $T_{p(x)}$ and $T_{p(y)}$, and condition (5), saying that the binding sites of T_x are the same as the binding sites of $T_{p(x)}$, ensure that G and G_A underlie tiling systems for the same set of tiles. The graph on the left hand side of Figure 10 does not satisfy (5) and provides a counterexample. If G_A represents a *closed* tiling system for a set of tiles $\mathcal{S}$, then each covering graph of G_A represents a closed tiling system for $\mathcal{S}$ also.
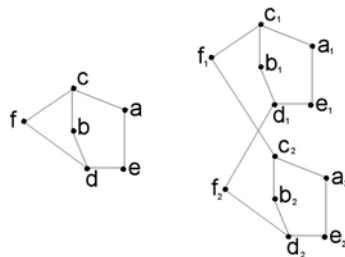
Fig. 10. Given a set of six distinct tiles whose binding sites are specific to each pair of tile interaction described by edges in the graph G_A (left), notice that G (right) is not a covering graph for G_A since it satisfies conditions $(1) - (3)$ but it does not satisfy (5) (see text). To see this, consider the mapping p between nodes of G and G_A which is suggested by the labels of the nodes. We want to think of f in G_A as representing a tile T_f with two *distinct* binding sites, one interacting with T_c and the other with T_d. Node f_1 is linked to two copies of c and node f_2 is linked to two copies of d; this means that T_{f_1} (T_{f_2}), having the same binding sites as T_f, should bind to T_{c_1}, T_{c_2} (T_{d_1}, T_{d_2}). But this is impossible because the binding would require the existence of two *identical* sites in T_{f_1} (T_{f_2}).

Given a set of tiles one would like to characterize the family of closed assemblies, or equivalently, of covering graphs, if any. An important application is in the solution of combinatorial problems.

Example 1. [22]. A graph $G = (V, E)$ is said to be *3-colorable* if there is a surjective function $f : V \rightarrow \{a, b, c\}$ such that if $v \rightarrow w \in E$, then $f(v) \neq f(w)$. Imagine constructing the graph G with two kinds of molecules, one coding for the *nodes* and one for the *edges*. Node-molecules are branched molecules, where the number of branches is the degree of the node, and edge-molecules are two-branched molecules. Each branch of a node-molecule has a sticky end whose code contains information on the node of the graph that connects to it and on a color for the node. The n branches of a same node-molecule are assumed to have the same code. Edge-molecules have two sticky ends and their code contains information on the origin and target nodes as well as on the colors of such nodes. The two colors are supposed to be different.

To consider three colors in the physical realization of the graph G, one constructs a node-molecule for each one of the three colors, together with all possible combinations of pairs of different colors for edge-molecules.

By combining several identical copies of these molecules and ligating them, open and possibly closed assemblies will form. Open assemblies are discharged (this can be done with the help of exonuclease enzymes that digest molecules with free ends) and closed assemblies, if any, ensure that the graph is 3-colorable. The only closed assemblies that can be formed in the test tube are covering graphs.

7   A Random Model for Sequential Assembly

The random model introduced in Section 5 needs to be adjusted slightly to simulate *sequential assembly*. Sequentiality presupposes the formation of specific intermediates, i.e. complexes, at specific moments along the assembly process. This means that one can start from a random configuration in S, let the input tiles form complexes at random, remove from S isolated tiles and use them as input tiles to re-iterate the process until a sufficiently large number of specific intermediates is formed. This will provide one step of the sequential assembly, and in the *simplest* case, this step will be re-iterated to model the next steps of the sequential process, until all steps are realized. Different types of tiles might be used as input tiles to perform different steps of the sequential process.

In more *complicated* cases, the above model, might need to integrate new kinds of steps. It might be that some of the steps of the sequential process require the intervention of specific enzymes, cleaving or ligating DNA tiles. Such operations are random and their effect on tiles and complexes can be described rigorously. Also, one might need to consider that tiles forming a complex at step i, disassemble in step $i + 1$ because of the interaction with new molecular tiles. This process is also random and can be formally described.

As mentioned in Section 3, the difficulty in inducing a sequential assembly comes from the complex combinatorics needed to realize objects of irregular but well-defined shape or aperiodic assemblies. A number of solutions have been proposed to overcome these combinatorial difficulties; they concern tile design (1)-(2), the algorithm for self-assembly (3)-(4) and the engineering protocol (5):

1. a variety of different forms of cohesion have been proposed, such as *sticky ended cohesion*, where single-stranded overhangs cohere between molecules [10]; *PX cohesion*, where topologically closed molecules cohere in a double-stranded interaction [60]; *edge-sharing*, where the interactions are characterized by lateral interactions [46]; *tecto-RNA*, where RNA domains pair laterally through loop osculations [21];
2. one can allow different forms of coding within the same molecule, which can involve the Watson-Crick sequences as well as the geometry of the molecule [9];
3. one can use "instructed gluing" elements together with DNA-tiles [7]; the idea is to add structural sub-units, as gluing elements between tiles, along the self-assembly process; in many cases, the use of such sub-units decreases the complexity of the protocol: the number of elementary molecules becomes smaller and the assembly algorithms becomes more specific;
4. the use of protecting groups, through which some of the potential interaction sites of the molecules are momentarily inhibited, is inherently a sequential protocol, applicable both to DNA objects [58] and to fractal assemblies [9];
5. the solid-support methodology in DNA nanotechnology [58] is an example of sequential assembly; it was used to construct the most complex DNA object to date, a truncated octahedron [59]; the step-wise synthesis of a square is illustrated in Figure 1 – here, enzymes intervene in some of the sequential steps.

8   Hierarchical Tiling

A set of tiles $\{T_1, \ldots, T_n\}$ is *self-consistent* if for each T_i with binding site D_i there is a tile T_j with binding site D_j such that $b(D_i) = D_j$, for some isometry b. Notice that i need not be different from j. In particular, a single tile T is *self-consistent with itself* if it has at least two binding sites which are complementary to each other.

Let $\{T_1, \ldots, T_n\}$ be basic elementary tiles which assemble in a variety of *tile complexes* $S_1, \ldots, S_l$, i.e. finite assemblies S_i with *unused* binding sites. A set of tile complexes is *self-consistent* if for each S_i with binding site D_i there is a tile complex S_j with binding site D_j such that $b(D_i) = D_j$, for some isometry b defined on tile complexes. New binding sites D_i generated from the assembly of a tile complex (as in Figure 7) are allowed.

A *hierarchical tiling* is an assembly X of tiles $\{T_1, \ldots, T_n\}$ that is obtained by successive steps of assembly generating intermediary sets of tile complexes $\mathcal{F}_0, \ldots, \mathcal{F}_m$ such that:

1. $\mathcal{F}_0 = \{T_1, \ldots, T_n\}$;
2. $\mathcal{F}_i = \{S_{i,1}, \ldots, S_{i,l_i}\}$, for $i = 1 \ldots m$, where each $S_{i,j}$ is a tile complex in $\mathcal{F}_{i-1}$;

3. $\mathcal{F}_i$ is a self-consistent set of tile complexes;
4. X is an assembly of $S_{m,1}, \ldots, S_{m,l_m}$.

The value of m is called *order* of the hierarchical tiling. A hierarchical tiling is *non-trivial* if for each family $\mathcal{F}_i$ there is at least one tile complex $S_{i,j}$ which is not in $\mathcal{F}_{i-1}$ already. Notice that not all assemblies can be defined as hierarchical assemblies of order m, for $m > 1$.

A dynamical model of hierarchical tiling. It can be defined by a repeated iteration of the random model for self-assembly presented in Section 5, where the tile complexes used as input tiles at step $i + 1$ are the complexes formed in S at the end of step i. In general, a hierarchical assembly is *not* a sequential assembly. It might happen though, that certain assembly processes are defined by a combination of sequential steps during the hierarchical self-assembly.

Some concrete examples of hierarchical assembly. Suitable selection of structural units allows the design of molecular entities undergoing self-organisation into well-defined architectures, which subsequently may self-assemble into supramolecular *fibrils* and *networks*. The assembly of "infinite" *tubes* and *spheres* has been realized many times and in many laboratories with different kinds of molecules. A basic approach is to design a rod-like molecule with an hydrophobic end and a hydrophilic one. Then, one puts the molecules in different media and observes the formation of spheres, where the hydrophilic side of the molecules lies either inside or outside the sphere, depending on the properties of the medium. Alternatively, one might observe the formation of a long tube where, again, on the surface one finds sides with identical hydrophilic/hydrophobic properties. The formation of spheres and tubes leads us to ask how these shapes might assemble

among themselves into supramolecular periodic or aperiodic structures. What other shapes do allow for the assembly of 1D, 2D and 3D arrays of such tile complexes?

Besides spheres, tubes and networks, chemists work on the design of synthetic molecules which lead to *helical* architectures of both molecular and supramolecular nature by hierarchical self-organization, or again to the formation of *mushroom-like* shapes and to a consequent assembly of such complexes into 3D arrays [18] (these arrangements are not *regular*, in the sense that they are not crystals). Mimicking nucleic-acid sequences, specific sequences of hydrogen bonding residues are led to act as structure-inducing codons, and such structural coding allows for the spontaneous but controlled generation of organized materials, e.g. [18].
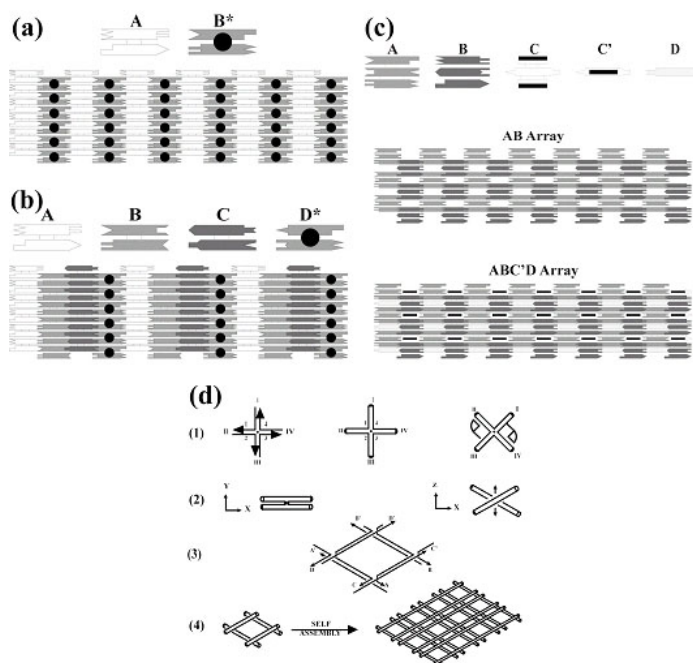

Fig. 11. *A variety of two-dimensional arrays that have been formed from DNA tiles.* Panels (a) and (b) illustrate 2D arrays composed of DX and $DX + J$ molecules. Panel (c) illustrates patterns obtained from TX molecules. Panel (d) illustrates an array made of DNA parallelograms.

At a different scale, for nanoscale molecules, like DNA, a broader range of possibilities can be explored since all of the contacts can be forced to be of a Watson-Crick form, although many other types of interaction are possible (e.g.,

[60]). The ifferent shapes of tiles introduced at the end of Section 4, enabled the assembly of several different kinds of periodic 1D and 2D arrays (see Figure 11). These hierarchical assemblies have order 2: single strands make the starting set of tiles, which assemble into specific intermediary molecular tiles (described in Section 4), and finally these molecular tiles self-assemble into a 2D array.

Periodic assemblies in 3 dimensions are still an open problem. Protocols for the assembly have been proposed, but highly ordered programmed 3D arrangements have not yet been realized to resolutions below 1 nm in the laboratory for DNA tiles. Aperiodic arrangements, typically harder to assemble and analyze than periodic assemblies, present an even greater challenge, because their characterization cannot rely on diffraction analysis in a simple fashion.

Example 2. Fractal assemblies. [8,9] Fractal constructions are a special case of aperiodic assemblies. The algorithm here is simple: from a starting molecular shape, which can look like a square or a triangle, and is designed to interact with copies of itself, one constructs a molecule with the same shape but a larger size, and re-iterates the process to get larger and larger assemblies of the given shape. The difficulty lies in the design of a set of basic shapes which can self-assemble into new self-similar shapes of larger sizes, and whose binding sites are coded by self-similar coding. An appropriate coding is important to ensure that tile complexes will self-assemble and that undesired binding is avoided. The order of this hierarchical tiling, corresponding to the number of iterations of the algorithm, is m, for potentially any value of m. In practice, a chemist would be happy with $m = 4, 5$.

These examples lead to some questions: within the set of feasible shapes and interactions, can we *classify* potential complexes? Once a complex is formed, can it be used as a building block to construct larger 1D, 2D or 3D arrays?

9   Size of the Assembly

How can the size of an assembly be controlled?

Rough termination is easy to induce. An obvious way is to limit the number of molecules in the solution. Another way is to use *protecting groups*, i.e. DNA molecules, which might be single strands for instance, whose binding sites are complementary to the binding sites of the tiles used in the assembly. The idea being that protecting groups might be added to the solution during the process of self-assembly to prevent new tiles from binding to available sites.

Exact termination is a consequence of the coding for the termination. If a synthesis or an assembly is templated, it is always possible to limit growth, by leaving out the constituent that is coded at the terminal position, for instance. The algorithmic synthesis of triangular circuits illustrated in Figure 6, provides another example where this is done [7]. In general, exact size control of a DNA self-assembly is hard to achieve. A few protocols have been presented so far.

In theory, DNA tiles can be used to "count" by creating boundaries with programmable sizes for 1D, 2D and possibly 3D periodic assemblies. The idea

is 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 [54,56]. If this schema can be physically implemented, then self-assembly of precisely-sized nanoscale arrays will be possible.

Fractal assemblies [8,9] can be thought of as a way to generate fixed geometrical shapes of controlled size. Besides the rectangular shapes of $n \times m$ arrays, one would like to have a way to grow assemblies with other shapes such as triangles, hexagons, etc. Fractal assembly allows us to do so by constructing objects with 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.

10   Algorithmic Assembly

The combination of different instructions in a "molecular program" has been used to design self-assembly systems which follow specific assembly pathways. This idea has its mathematical analogue in the work of Wang [48,49,50], who proposed a finite set of tiles mimicking the behavior of any Turing Machine.

Wang tiles are squared tiles in $\mathbb{R}^2$ whose binding sites are the four sides of the square, and whose interaction is possible on binding sites labelled by the same color. If $T_1, T_2, \ldots, T_n$ are Wang tiles, then one asks that $\{T_1, T_2, \ldots, T_n\}$ be a self-consistent set. Once a set of Wang tiles is given, one asks whether the plane can be tiled with it, and what are the properties of the tiling, namely if the set generates periodic tiling only, or both periodic and non-periodic tiling, or aperiodic tiling only.

The molecular realization of Wang tiles (where a square becomes a 4-arm branched molecule with Watson-Crick complementary sticky ends as binding sites) can, theoretically, be used to make computations [54]. This notion has not yet been realized experimentally in more than one dimension [30]. A three-dimensional framework for computing 2D circuits and constructing DNA-objects with given shapes, has been suggested [7], where again, DNA tiles mimic Wang tiles. It is important to stress that molecular tiles are not conceived to generate only uniform tiling of the plane, but on the contrary, they can be used to induce the assembly of objects of arbitrary shapes.

Combinatorial optimisation problems: fixing a single shape. Two combinatorial problems have been stated in [3]. The first concerns minimum tile sets, i.e. given a shape, find the tile system with the minimum number of tile types that can uniquely self-assemble into this shape. The second concerns tile concentration, i.e. given a shape and a tile system that uniquely produces the given shape, assign concentrations to each tile-type so that the expected assembly time for the shape is minimized. The first combinatorial problem is NP-complete and the second is conjectured to be #P [3]. These problems have been formulated for any given shape even though only square tiles, i.e. Wang tiles, have been studied until now.

Templates and fixed shapes. Can one find a small set of relatively simple tiles such that, starting from a template supporting a linear code (that may be a DNA

or RNA molecule incorporated into a macromolecular complex), the assembly process will create a given three dimensional shape in the space? We think here of interacting tiles performing a transformation from labeled templates into three dimensional structures and we ask what kind of transformations can be realized in this way [7]. Also, one wants to understand how much the complexity of the construction depends on the complexity of the tiles, where the latter can be measured by the number of the binding sites of the tiles, the size of the sets $B_{i,j}$, etc.

Combinatorial optimisation problems: fixing a "family" of shapes. Fractal assembly provides an example of an iterative algorithm for self-assembly which generates fractals of arbitrary dimension and not just a single shape with a given size. For each dimension, the building blocks necessary to build the corresponding fractal shape need to satisfy the same self-similar properties, and the design of a tile set which satisfies these properties is not obvious. For instance, given a Sierpinski square fractal and an iterative algorithm that produces arbitrarily large instances of this shape, is there a set of Wang tiles that can uniquely assemble into any fractal size? It is not at all clear that a set of Wang tiles with self-similar coding exists. In [9] a set of tiles, whose boundaries are characterized by both a coding sequence and a *geometrical* shape, is proposed. Does geometry have to be included in the coding of the tile boundaries to impose extra control on the assembly? What is the minimum number of tiles necessary to generate a *family* of shapes?

In general, let an algorithm for self-assembly be fixed. What are the properties of the tiles which are necessary to realize the algorithm?

Dynamic tiling. A molecular feature that has been used in algorithmic self-assembly is the possibility to program and change the status of a molecule. This means that the molecule passes in time through several possible physical conformations, i.e. geometrical shapes. In DNA nanotechnology, this has been done by using "template" molecules (programmable tiles) that interact with DNA single strands [47,46]: the pairing of the single stranded DNA present in the solution to a single strand subsequence of the tile induces this latter to change its conformation. Because of these conformational changes, tiles get a different status during the assembly, with the effect that one is able to control the dynamics of the algorithm and the direction of the assembly. As a result, one can generate different architectures out of the same set of tiles by varying their conformations.

Example 3. One can imagine a basic molecular system that is fundamentally a layer of *programmable* tiles which can guide the assembly of multiple layers of tiles above it [7]. In the 2-dimensional case this device can compute tree-like boolean circuits, and in 3D, it can induce finite regular and less regular shapes. *Multiple regular layers* are obtained by completely filling up the template board: a new layer of tiles will cover-up the old one and will play the role of a new board in allowing the creation of a third layer, and so on. *"Walls"* with specified "height", or *discrete irregular shapes* are obtained by *partially* filling-up the board, and this can be achieved by inserting appropriate coding in the programmable tiles

that form the template board. The coding will discriminate what are the tiles that will interact with new ones and what are those that will avoid interaction.

In the example, a change in the programming of the board induces the formation of different shapes out of the same input set. This suggests that a formal notion of *complexity* describing self-assembly of molecular systems cannot be based merely on the variety of shapes that potentially can be assembled, but rather on the much larger variety of algorithms that allow their assembly.

DNA computing. Last, we want to mention the effort in designing algorithms for DNA-computation. The landmark step is in [1], where DNA is used to solve an instance of the Hamiltonian Path problem, asking to establish whether there is a path between two cities, given an incomplete set of available roads. A set of strands of DNA is used to represent cities and roads (similar to the description of the 3-coloring problem in Section 6), and the coding is such that a strand representing a road would connect (according to the rules of base-pairing) to any two strands representing a city. By mixing together strands, joining the cities connected by roads, and weeding out any wrong answers, it has been shown that the strands could self-assemble to solve the problem.

The first link between DNA-nanotechnology and DNA-computation was established in [54] with the suggestion that short branched DNA-molecules could be "programmed" to undergo algorithmic self-assembly and thus serve as the basis of computation. Other work has followed as [30,34,25].

11   Discussion

Most examples in this paper were based on Watson-Crick interactions of DNA molecules. Other kinds of interaction, usually referred to as *tertiary interactions*, can be used to lead a controlled behavior in the assembly of DNA molecules, for example, DNA triplexes [15], tecto-RNA [21] and G-quartet formation [42]. In the combinatorial DNA constructions that we presented, tertiary interactions were carefully avoided with the goal of maximizing control on the dynamics of the assembly. Tertiary interactions are not as readily controlled as Watson-Crick interactions. The next generation of structural DNA nanotechnologists will be likely to exploit this wider range of structural possibilities and it appears possible that new combinatorics might arise from these options.

References

1. L.M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266:1021-4, 1994.
2. L.M. Adleman. Toward a mathematical theory of self-assembly. Technical Report 00-722, Department of Computer Science, University of Southern California, 2000.
3. L.M. Adleman, Q. Cheng, A. Goel, M-D. Huang, D. Kempe, P. Moisset de Espanés, P.W.K. Rothemund. Combinatorial optimisation problems in self-assembly. *STOC'02 Proceedings*, Montreal Quebec, Canada, 2002.
4. A. Aggeli, M. Bell, N. Boden, J.N. Keen, P.F. Knowles, T.C.B. McLeish, M. Pitkeathly, S.E. Radford. Responsive gels formed by the spontaneous self-assembly of peptides into polymeric beta-sheet tapes, *Nature* 386:259-62, 1997.
5. P. Ball Materials Science: Polymers made to measure, *Nature*, 367:323-4, 1994.
6. A. Carbone, M. Gromov. Mathematical slices of molecular biology, *La Gazette des Mathématiciens*, Numéro Spéciale, 88:11–80, 2001.
7. A. Carbone, N.C. Seeman. Circuits and programmable self-assembling DNA structures. *Proceedings of the National Academy of Sciences USA*, 99:12577-12582, 2002.
8. A. Carbone, N.C. Seeman. A route to fractal DNA-assembly, *Natural Computing*, 1:469-480, 2002.
9. A. Carbone, N.C. Seeman. Coding and geometrical shapes in nanostructures: a fractal DNA-assembly, *Natural Computing*, 2003. In press.
10. S.N. Cohen, A.C.Y. Chang, H.W. Boyer, R.B. Helling. Construction of biologically functional bacterial plasmids in vitro, *Proceedings of the National Academy of Science USA*, 70:3240-3244, 1973.
11. A.M. Diegelman, E.T. Kool. Generation of circular RNAs and trans-cleaving catalytic RNAs by rolling transcription of circular DNA oligonucleotides encoding hairpin ribozymes, *Nucleic Acids Research*, 26, 3235-3241, 1998.
12. S.M. Du, S. Zhang, N.C. Seeman. DNA Junctions, Antijunctions and Mesojunctions, *Biochemistry*, 31:10955-10963, 1992.
13. I. Duhnam, N. Shimizu, B.A. Roe et al. The DNA sequence of human chromosome 22, *Nature*, 402:489–95, 1999.
14. B.F. Eichman, J.M. Vargason, B.H.M. Mooers, P.S. Ho. The Holliday junction in an inverted repeat DNA sequence: Sequence effects on the structure of four-way junctions, *Proceedings of the National Academy of Science USA*, 97:3971-3976, 2000.
15. G. Felsenfeld, D.R. Davies, A. Rich. Formation of a three-stranded polynucleotide molecule, *J. Am. Chem. Soc.*, 79:2023-2024, 1957.
16. T.-J. Fu, B. Kemper and N.C. Seeman. Endonuclease VII cleavage of DNA double crossover molecules, *Biochemistry* 33:3896-3905, 1994.
17. B. Grünbaum, G.C. Shephard. *Tilings and Patterns*, W.H. Freeman and Company, 1986.
18. J.D. Hartgerink, E. Beniash, and S.I. Stupp. Self-assembly and mineralization of peptide-amphiphile nanofibers, *Science*, 294:1684, 2001.
19. I. Huck, J.M. Lehn. Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly, *Proceedings of the National Academy of Sciences USA*, 94:2106-10, 1997.
20. S. Hussini, L. Kari, S. Konstantinidis. Coding properties of DNA languages, *Theoretical Computer Science*, 290:1557-1579, 2003.
21. L. Jaeger, E. Westhof, N.B. Leontis. TectoRNA: modular assembly units for the construction of RNA nano-objects, *Nucleic Acids Research*, 29:455-463, 2001.

22. N. Jonoska. 3D DNA patterns and Computing, *Pattern formation in Biology, Vision and Dynamics*, edited by A. Carbone, M. Gromov, P. Prusinkiewicz, World Scientific Publishing Company, 310-324, 2000.

23. G. von Kiedrowski. Personal communication, February 2003.

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

25. T.H. LaBean, E. Winfree, J.H. Reif. Experimental progress in computation by self-assembly of DNA tilings, *Proc. DNA Based Computers V*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, E. Win- free and D.K. Gifford (eds), American Mathematical Society, Providence, RI, vol.54:123-140, 2000.

26. J.M. Lehn. Sopramolecular Chemistry, *Science*, 260:1762-3, 1993

27. J.M. Lehn. Toward complex matter: Supramolecular chemistry and self-organisation. *Proceedings of the National Academy of Science USA*, 99(8):4763-4768, 2002.

28. F. Liu, R. Sha and N.C. Seeman. Modifying the surface features of two-dimensional DNA crystals, *Journal of the American Chemical Society* 121:917-922, 1999.

29. C. Mao, W. Sun and N.C. Seeman. Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy, *Journal of the American Chemical Society*, 121:5437-5443, 1999.

30. C. Mao, T. LaBean, J.H. Reif, N.C. Seeman. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature* 407:493-496, 2000; Erratum: *Nature* 408:750-750, 2000.

31. H. Qiu, J.C. Dewan and N.C. Seeman. A DNA decamer with a sticky end: The Ccystal structure of d-CGACGATCGT, *Journal of Molecular Biology*, 267:881-898, 1997.

32. J.H. Reif. Local parallel biomolecular computation, *DNA Based Computers, III*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol 48 (ed. H. Rubin), American Mathematical Society, 217-254, 1999.

33. J.H. Reif. Molecular assembly and computation: from theory to experimental demonstrations, *29-th International Colloquium on Automata, Languages, and Programming(ICALP)*, Málaga, Spain (July 8, 2002). Volume 2380 of Lecture Notes in Computer Science, New York, pages 1-21, 2002.

34. P. Sa-Ardyen, N. Jonoska, N.C. Seeman. Self-assembling DNA graphs, *DNA Computing, 8th International Workshop on DNA-based Compters*, LNCS 2568:1-9, 2003.

35. J.M. Schnur. Lipid tubules: a paradigm for molecularly engineered structures, *Science*, 262:1669-76, 1993.

36. N.C. Seeman. Nucleic acid junctions and lattices, *J. Theor. Biol.*, 99:237-247, 1982.

37. N.C. Seeman, N.R. Kallenbach. Design of immobile nucleic acid junctions, *Biophysical Journal*, 44:201-209, 1983.

38. N.C. Seeman, N.R.Kallenbach. Nucleic-acids junctions: a successfull experiment in macromolecular design, *Molecular Structure: Chemical Reactivity and Biological Activity*, J.J. Stezowski, J.L. Huang, M.C. Shao (eds.), Oxford University Press, Oxford, 189-194, 1988.

39. N.C. Seeman. DNA engineering and its application to nanotechnology, *Trends in Biotech.*, 17:437-443, 1999.

40. N.C. Seeman. DNA nanotechnology: from topological control to structural control, *Pattern formation in Biology, Vision and Dynamics*, edited by A. Carbone, M. Gromov, P. Prusinkiewicz, World Scientific Publishing Company, 310-324, 2000.

41. N.C. Seeman. In the nick of space: Generalized nucleic acid complementarity and the development of DNA nanotechnology, *Synlett*, 1536-1548, 2000.

42. D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and applications to meiosis. *Nature*, 334:364-366, 1988.

43. R. Sha, F. Liu, D.P. Millar and N.C. Seeman. Atomic force microscopy of parallel DNA branched junction arrays, *Chemistry & Biology* 7:743-751, 2000.

44. Z. Shen. DNA Polycrossover Molecules and their Applications in Homology Recognition. Ph.D. Thesis, New York University, 1999.

45. W.B. Sherman, N.C. Seeman. (in preparation)
Abstract "The design of nucleic acid nanotubes" appeared in *Journal of Biomolecular Structure & Dynamics*, online at `http://www.jbsdonline.com/index.cfm?search=seeman&d=3012&c=4096&p11491&do=detail`, 2003.

46. H. Yan, X. Zhang, Z. Shen, N.C. Seeman. A robust DNA mechanical device controlled by hybridization topology. *Nature*, 415:62-5, 2002.

47. B. Yurke, A.J. Turberfield, A.P.Jr. Mills, F.C. Simmel, J.L. Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, 406:605-608, 2000.

48. H. Wang. Proving theorems by pattern recognition, *Bell System Tech. J.*, 40:1-42, 1961.

49. H. Wang. Dominos and the AEA case of the decision problem. In *Proceedings of the Symposium on the Mathematical Theory of Automata*, 23-56, Polytechnic, New York, 1963.

50. H. Wang. Games, logic and computers, *Scientific American*, November, 98–106, 1965.

51. Y. Wang, J.E. Mueller, B. Kemper, and N.C. Seeman. The assembly and characterization of 5-arm and 6-arm DNA junctions, *Biochemistry* 30:5667-5674, 1991.

52. G.M. Whitesides, J.P. Mathias, C.T. Seto. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures, *Science*, 254:1312-9, 1991.

53. W.R. Wikoff, L. Liljas, R.L. Duda, H. Tsuruta, R.W. Hendrix, J.E. Johnson. Topologically linked protein rings in the bacteriophage HK97 caspid. *Science*, 289:2129-2133, 2000.

54. E. Winfree. On the computational power of DNA annealing and ligation, *DNA based computers, Proceedings of a DIMACS workshop, April ???4, 1995, Princeton University*, eds. R.J. Lipton and E.B. Baum, AMS Providence, 199-219, 1996.

55. E. Winfree, F. Liu, L.A. Wenzler, N.C. Seeman. Design and self-assembly of two-dimensional DNA crystals, *Nature*, 394:539–44, 1998.

56. E. Winfree. Algorithmic self-assembly of DNA: theoretical motivations and 2D assembly experiments. *J. Biol. Mol. Struct. Dynamics Conversat.*, 2:263-270, 2000.

57. S. Zhang, T. Holmes, C. Lockshin, A. Rich. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proceeding of the National Academy of Sciences USA*, 90:3334-8, 1993.

58. Y. Zhang, N.C. Seeman. A solid-support methodology for the construction of geometrical objects from DNA, *J. Am. Chem. Soc.*, 114:2656-2663, 1992.

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

60. X. Zhang, H. Yan, Z. Shen and N.C. Seeman. Paranemic cohesion of topologically-closed DNA molecules, *J Am. Chem. Soc.*, 124:12940-12941, 2002.