One Dimensional Boundaries for DNA Tile Self-Assembly

Rebecca Schulman, Shaun Lee, Nick Papadakis, and Erik Winfree

Computer Science and Computation & Neural Systems
California Institute of Technology, Pasadena, CA 91125, USA

Abstract. In this paper we report the design and synthesis of DNA molecules (referred to as DNA tiles) with specific binding interactions that guide self-assembly to make one-dimensional assemblies shaped as lines, V's and X's. These DNA tile assemblies have been visualized by atomic force microscopy. The highly-variable distribution of shapes – e.g., the length of the arms of X-shaped assemblies – gives us insight into how the assembly process is occurring. Using stochastic models that simulate addition and dissociation of each type of DNA tile, as well as simplified models that more cleanly examine the generic phenomena, we dissect the contribution of accretion vs aggregation, reversible vs irreversible and seeded vs unseeded assumptions for describing the growth processes. The results suggest strategies for controlling self-assembly to make more uniformly-shaped assemblies.

1 Introduction

Self-assembly - the process by which monomer units come together to form a larger structure according to local, energetic rules - is of great theoretical and practical interest. Most natural self-assembling systems, including crystals, biological membranes, and virus capsids do not permit easy experimental variation of the specificity of binding between units. Control over binding specificity allows the investigation of both the theoretical possibilities of the self-assembly process as well as the practical goal of constructing more complex nano-fabricated patterns.

Control over binding specificity allows the self-assembly process to be *programmed*. The potential of self-assembly programming derives from Wang's investigations of the two-dimensional tiling problem [14], in which he showed that two-dimensional tiling systems can simulate a universal Turing machine [15]. The *(abstract) Tile Assembly Model* (aTAM) [17] is an extension of Wang's tiling systems to include a specific growth process motivated by physical considerations of crystallization. In aTAM, a program is the specification of the tile types, the bond types on their sides, the bond strengths and a threshold value. Assembly begins with a seed tile and proceeds by the non-deterministic addition of tiles at locations where the total strength of all bonds that would be formed is greater than the threshold. The abstract Tile Assembly Model has been shown

to be Turing-universal [16, 17], which implies that complex objects can be self-assembled from relatively small numbers of tile types [9]. We therefore say that aTAM supports *algorithmic* self-assembly.

Our goal is to implement the aTAM in chemistry by using DNA tiles – specifically, double crossover (DX) molecules [7] – making use of the complementary base-pairing between its strands to control the strength and specificity of molecular assembly. Each DX molecule consists of two parallel double helices; each strand first participates in one helix, then crosses to the other helix, holding the two helices together. There are four single-stranded *sticky ends* that can hydrogen bond to complementary sticky ends on other DX molecules, allowing fine control of binding specificity. Previous work has shown that the DX molecules R00 and S00 (shown in figure 1) will self-assemble into two-dimensional sheets in which R00 and S00 are arranged periodically in stripes, in agreement with their sticky-end interactions [18]. It remains to be shown experimentally that well-defined algorithmic patterns can result from two-dimensional self-assembly of DNA tiles. Additional forms of control are necessary to achieve this.

As an example of algorithmic self-assembly, consider the set of eight abstract tiles (and their DNA analogs) shown in figure 1 as *boundary tiles* and *rule tiles*[1]. Each DNA tile is designed to have the same bond types, with approximately the same binding strengths, as the corresponding abstract tile; in the abstract model, the **B** bond has twice the strength as the other bond types. See figure 1 for details.

Under the aTAM with the threshold for tile addition set at 2 units, these tiles set up boundary conditions and execute iterated XOR logic to construct Pascal's triangle modulo 2, the discrete analog of the fractal Sierpinski gasket [3]. Self-assembly begins with the *corner tile* RC as the seed tile: Due to the strong bond **B**, first SB, and then RB, can bind repetitively to either side of RC, creating a V-shaped boundary. As soon as a pair of SB tiles have bound to either side of RC, an R11 tile can be added directly above RC, thereby forming two strength-1 bonds and thus achieving the threshold. No other tile could have been added at that location, because all other tile types would form at most one strength-1 bond. As soon as the R11 tile and flanking RB tiles are present, two new locations become available for tile addition; this time, it is only S01 that can make two matching bonds. This process continues forever; at each location, a unique tile may be added, and thus the pattern generated is uniquely defined; an intermediate assembly is shown at the bottom of figure 1.

That the pattern is Pascal's triangle mod 2 derives from the fact that each rule tile corresponds to an entry of the truth table for the XOR function. Conceptually, the inputs to the tile are given in the two sticky ends at the bottom, and the tile's output is (repeated) in both sticky ends at the top. Because of the geometry of the DX molecules we used, and the fact that DNA strands are directed (by convention, from 5' to 3'), tiles in alternating rows have reversed strand orientation. This means that two tiles are required for each line

[1] Two additional tiles, RCxy and S00N, were needed as well for some experiments reported here.

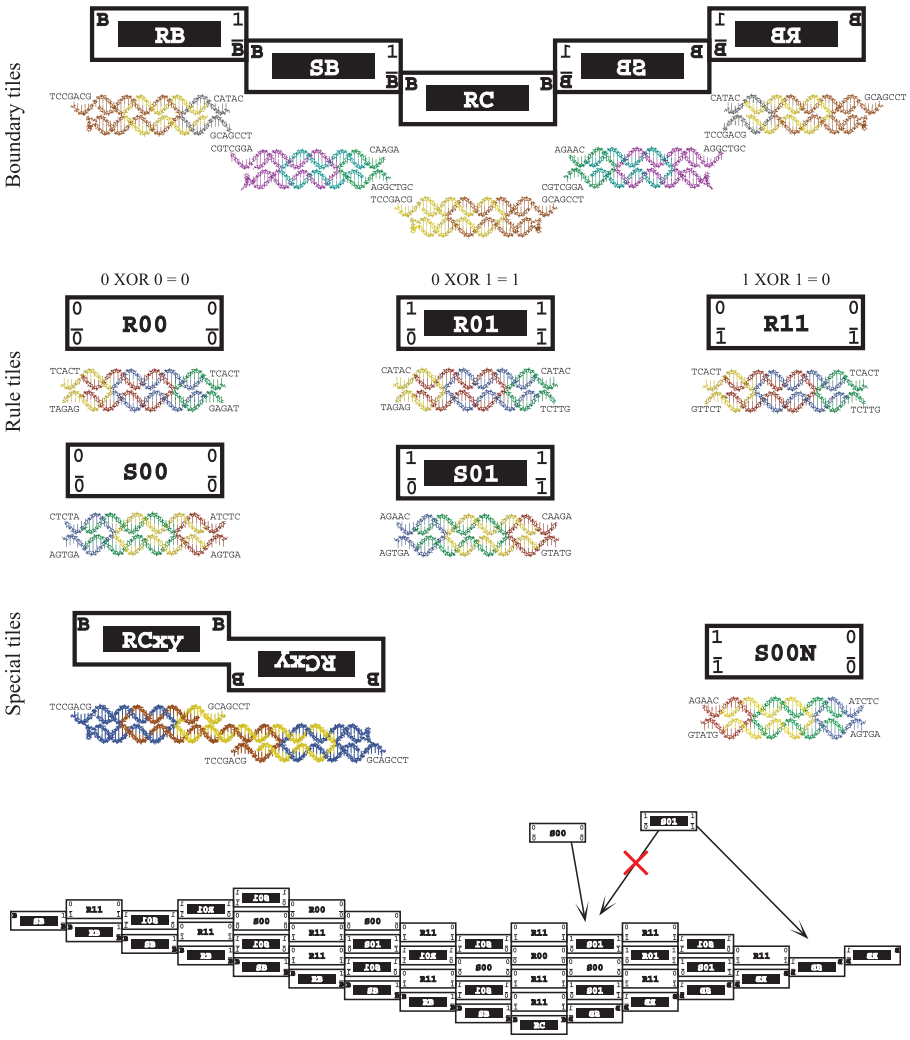


Fig. 1. Abstract tiles, DNA tiles, and their assembly according to the Tile Assembly Model. **Tile types.** Tiles are classified into boundary tiles and rule tiles. Boundary tile RCxy and rule tile S00N, although not part of the assembly process shown at the bottom, were used in experiments reported here. Boundary tiles RC and RCxy are also called corner tiles. A variant of S00N, called S00N-23J, contains hairpin sequences to enhance AFM contrast, as in [18]; it has the same sticky ends as S00N. Each DNA tile is approximately 4×12 nm. Abstract tiles may be flipped left-to-right (reflecting a symmetry present in the DNA molecule) as necessary. **Bond types.** Tiles whose names begin with R can bind only to tiles whose names begin with S (because R tiles have $5'$ sticky-end overhangs on top, while S tiles have $3'$ overhangs, necessitating different sticky end sequences). For the abstract tiles, matching bond types $\mathbf{B}$ and $\overline{\mathbf{B}}$ (which are implemented in DNA by GC-rich length-7 sticky ends) have a strength of 2, while the other matching bond types ($\mathbf{0}$ and $\overline{\mathbf{0}}$, and $\mathbf{1}$ and $\overline{\mathbf{1}}$, implemented in DNA as length-5 sticky ends) have a strength of 1, in some arbitrary units. **Assembly.** In the bottom half of the figure, tiles are added to a growing assembly either when a $\mathbf{B}$ bond can be formed or when two weaker bonds can be formed simultaneously. The red X indicates a location where mismatch of the sticky ends prevents tile addition according to the aTAM. Full sequences and sequence design procedures are available at `http://www.dna.caltech.edu/SupplementaryMaterial`.

in the truth table. For example, the R01 and S01 tiles have the same semantics (0 XOR 1 = 1), but R01 has 5' overhanging sticky ends on its output (top), while S01 has 3' overhangs. There is no need for an S11 tile, as it does not appear in the Sierpinski pattern.

In order to apply insights about abstract algorithmic self-assembly to real physical systems, we need to understand when a physical system is well-modeled by the aTAM. Initial steps in this direction were achieved in [17], which defined the *kinetic Tile Assembly Model* (kTAM) to include rules of reversible chemistry: any tile may be added at any site at a rate proportional to its concentration as free monomer, and any tile may leave the assembly at a rate exponentially related to the strength of its bonds with the rest of the assembly. In that work, it was argued that physical conditions can be achieved under which the Sierpinski tiles self-assemble correctly with high probability: growth from corner tiles proceeds with a low error rate; growth from rule tiles is very rare; and growth from boundary tiles quickly incorporates a corner tile, and then proceeds with a low error rate.

Although encouraging, the original kTAM model makes assumptions that are not appropriate for some circumstances. Specifically, it assumes (1) that the monomer tile concentration are held constant for the duration of the simulation. Therefore, we call it a *powered* model; in an *unpowered* model, monomer tile concentrations would be depleted as they are used. (2) that assemblies grow only by addition of a single tile at a time, a process called *accretion*. The alternative, *aggregation*, allows two large assemblies to come together and bind to each other. (3) that growth of assemblies is independent of one another, and therefore the fate of a single seed tile can be simulated in isolation. We call this a *(singly-)seeded* model. Assumption (3) is actually a consequence of assumptions (1) and (2); if either monomers are unpowered, or aggregation of assemblies is to be considered, then a *multiply-seeded* model is appropriate. Experiments where DNA molecules are passively assembled in a test tube are more likely to resemble a multiply-seeded, unpowered, aggregation model. Therefore, results from the singly-seeded, powered, accretion kTAM must be carefully interpreted, or – as we do here – enhancements of the original kTAM must be used.

In this work, we address these issues as they apply to the self-assembly of one-dimensional boundaries. Originally intended as a simple step toward demonstrating the Sierpinski pattern experimentally, construction of one-dimensional boundaries has turned out to be an interesting story in its own right. Although the individual DX tiles required for the Sierpinski tile set formed reliably and associated specifically in accord with the programmed interactions, several attempts to create uniformly V- and X-shaped boundaries produced, instead, a high-variance distribution of mostly asymmetrically-shaped assemblies. This turns out to be an excellent test of the various assumptions used in the kTAM. We show, first, that a multiply-seeded, unpowered, accretion variant of the kTAM gives simulation results qualitatively similar to the experimental results. This, however, does not identify which assumptions are valid for our experimental system, nor does it provide understanding. For that, we turn to

three simplified models that test the assumptions of accretion vs aggregation, reversible vs irreversible binding, and singly- vs multiply-seeded growth. Our experimental results are compatible with the reversible and irreversible aggregation models, and incompatible with the irreversible seeded accretion model. Furthermore, our understanding of the models suggests several approaches to solve the problem of creating uniformly-shaped boundaries.

2 Experiments with Seeded and Unseeded Assembly

Early attempts to produce the Sierpinski pattern by mixing the tiles described above were unsuccessful. AFM imaging revealed that some assembly occurred, but it was irregular and difficult to interpret, due to poor AFM resolution at the time. We have since begun a step-by-step process of debugging, testing components of the system one at a time and in simple combinations. During this process resolution has been improved to the point where we regularly can discern individual tiles.

First we tested the boundary tiles in isolation. RB and SB together make what we term a *single-layer* boundary. Figure 2(a) illustrates a typical AFM image of the long filaments formed. Images were difficult to obtain at lower tile concentrations, which precluded identifying individual assemblies. However, it was clear that filaments formed and that they were quite flexible, often forming loops, circles, or coils.

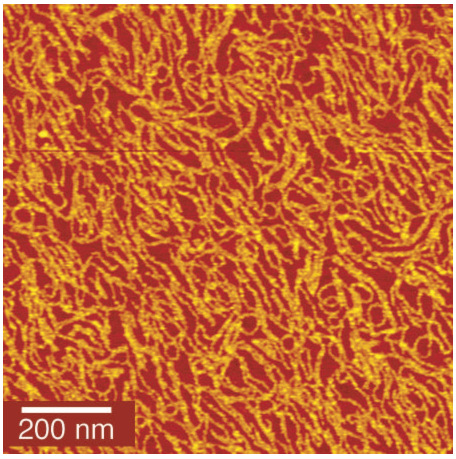
In order to create rigid, straight assemblies, a *double-layer* boundary was formed by adding the R11 and S00N tiles[2] to RB and SB, as shown in figure 2(b). Individual tiles can be distinctly recognized. We also observed long single-layer assemblies, both alone and as *tails* extending from double-layer assemblies (see arrow).

Unfortunately, the addition of small quantities of the RC tile to double-layer boundary did not create shallow V-shaped boundaries suitable for growing large assemblies. There was little evidence of the expected V-shape; images of these samples were essentially indistinguishable from double-layer boundary alone (data not shown). It is not clear whether the few V's that were observed were V-shaped tile arrangements, or simply two boundary assemblies aligned in a V shape by chance, or a boundary assembly folded into a V shape.

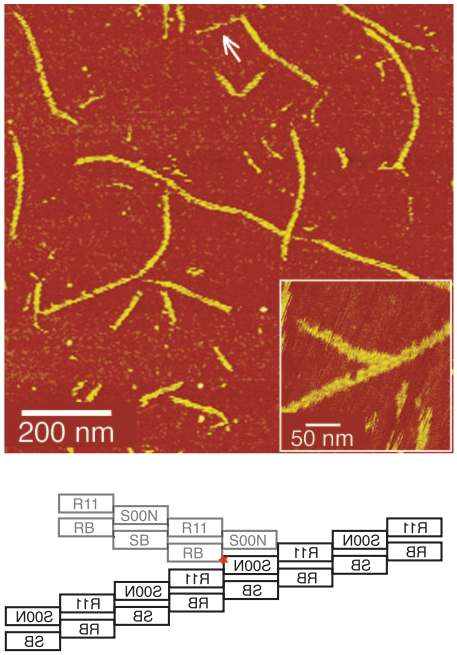
The construction of a four-tile assembly consisting of RC, R11, and two SB tiles called the *2x2 V*, verified that the RC tile was binding correctly to its immediate neighbors. Figure 3(a) shows that this construct forms as designed, although at this concentration, 2x2 assemblies appear to associate loosely with each other.

However, when the tiles needed to form the slightly larger V structure shown in figure 3(c) were combined, a *3x3 V* shape was almost never clearly observed. Instead, a great deal of double-layer boundary that may or may not have had a corner tile attached and very occasional shallow V's were visible.

[2] We used S00N here because the combination of R11, S01, RB, and SB tiles might be prone to growing additional layers with mismatches.

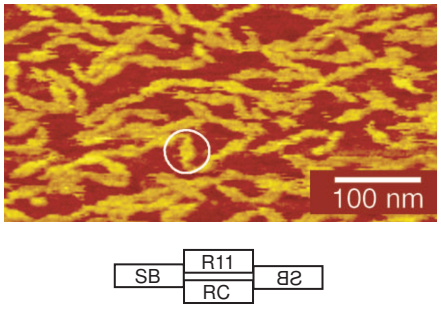


(a) Single-layer boundary at high concentration. RB and SB tiles at 1:1, 2.5μM each tile.

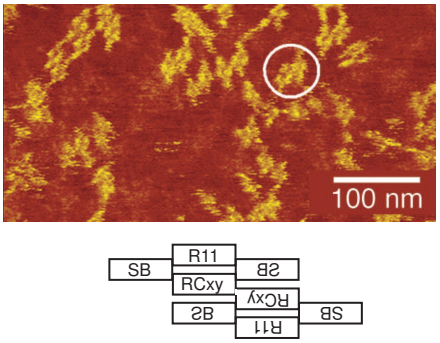


(b) Double-layer boundary. Bold portion of the diagram represents the arrangement of the double layer boundary assemblies. The bold tiles combination with the tiles in gray is a hypothesized structure for the motif shown in inset that is seen occasionally. Red star indicates mismatched sticky ends in the putative structure. RB, SB, S00N-23J, R11 at 1:1:1:1, 0.1μM each tile. Diluted 2:1 before imaging.

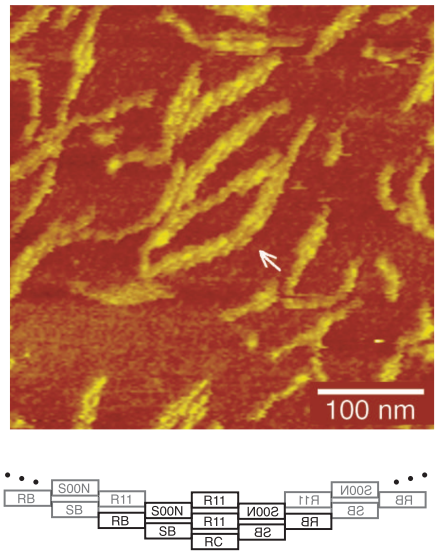
Fig. 2. Boundary assemblies imaged via AFM. **Sample prep:** Oligos were obtained PAGE-purified from Integrated DNA Technologies (www.idtdna.com), and quantified by UV-absorbance. The four strands comprising each tile were mixed at .4μM (each strand) in TAE/Mg^{++} buffer (40mM Tris-Acetate, 1mM Na_2EDTA, 12.5mM Magnesium Acetate) and annealed from 90 to 20 °C at 1 °/min. These tile stocks were mixed at room temperature ($\approx$ 25°C) in the specified ratios. Time between mixing and imaging was typically one hour, but varied from 20 min to several days. **AFM imaging:** 5μl of sample was deposited on freshly cleaved muscovite mica (Pella, www.tedpella.com) and allowed to adsorb for 3 min. After optional rinsing with TAE/Mg^{++} buffer, it was imaged under buffer using tapping mode on a Digital Instruments (www.di.com) Nanoscope IIIA with NP-S sharpened silicon nitride tips.



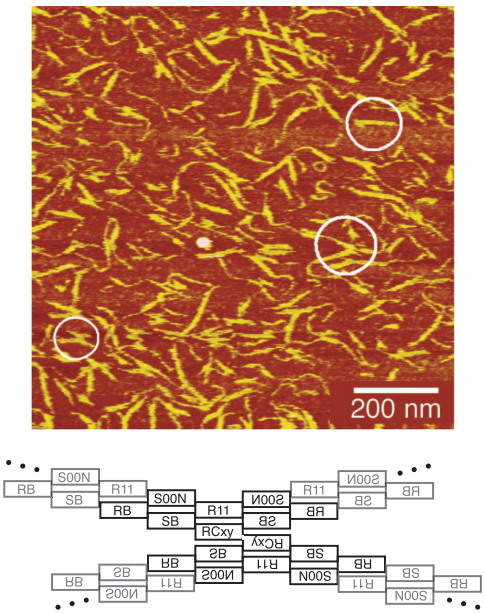
(a) 2x2 V. SB, R11, RC, 2:1:1, RC at 0.1μM



(b) 2x2 X. SB, R11, RCxy, 4:2:1, RCxy at 0.057μM



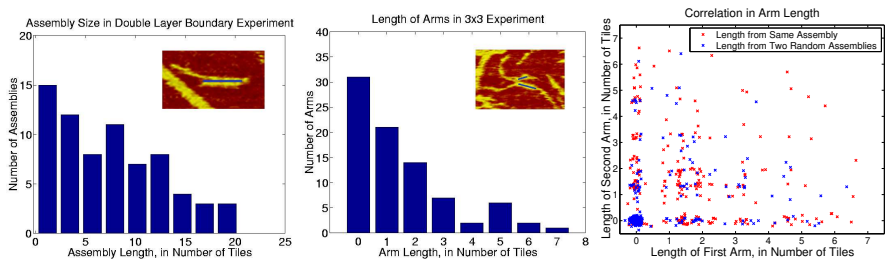
(c) 3x3 V. RB, SB, S00N-23J, R11, RC, 2:2:2:2:1, RC at 0.044μM. Arrow points to a V-shaped structure that is seen occasionally, and is believed to be the arrangement of tiles shown here in bold and gray.



(d) 3x3 X. Circles indicate examples of 1-, 3-, and 4-armed structures. RB, SB, S00N-23J, R11, RCxy, 4:4:4:2:1, RCxy at 0.027μM

Fig. 3. X and V Structures. In diagrams, black indicates the target assembly (according to stoichiometry); gray indicates other possible polymerization. Experimental methods are as shown in figure 2.

To address the theory that the corner tiles were present, but not visible in some of the previous experiments, a new corner tile that would form assemblies of a distinct X shape was created. This tile, RCxy, resembles an RC tile joined to a second RC tile that has been rotated. The *2x2 X*, a repetition of the 2x2 V experiment with the RCxy tile and adjusted stoichiometries of the other tiles, is shown in figure 3(b). The resulting motif formed without difficulty.



(a) (b) (c)

Fig. 4. (a) Measured sizes of double-layer boundary assemblies from an image similar to figure 2(b). (b) Measured sizes of double-layer arms extending from the RCxy corner tile from the image in figure 3(d) (c): A scatter plot of the length of two arms attached to the same corner tile and to different corner tiles, showing that the size of a given arm is independent of the size of the other arms attached to the same corner tile. Arms sizes are calculated from contour sizes of arms as shown in the insets in graphs (a) and (b). The measured contour length of each arm was converted to a length in tiles using the formula 1 tile = 12.5 nm. Simulated measurement noise was added in the scatter plot to avoid exact superposition of datapoints. Single-layer arms, single-layer tails, and assemblies only partially contained in the image were not counted.

Finally, the *3x3 X*, shown in figure 3(d), behaved analogously to the 3x3 V. Rather than forming the target assembly, some arms grew relatively long, while others didn't grow at all. The distribution of arm lengths found in an image of double-layer boundary and of 3x3 X is shown in figure 4. Two trends are clear: relative frequency decreases with arm length, and the lengths of arms attached to the same corner tile appear to be statistically uncorrelated.

These experiments confirm that the Sierpinski tiles form structures that reflect their programmed interactions. However, the frequency and shape of the structures that arose were sometimes unexpected.

3 Simulations of Tile Assembly

Our first approach to explaining these results considers a tile-based assembly model that incorporates basic aspects of the physical chemistry of DNA [1]:

1. The rate of association between tiles, accomplished by the hybridization of their sticky ends, is dominated by their respective concentrations. Therefore, the forward rate constant k_f is identical for all assemblies and all tiles.
2. The rate of dissociation of a tile from an assembly is based on the total free energy, ΔG, of all sticky-end bonds that must be broken. Therefore, the reverse rate constant $k_r = e^{\Delta G/kT}$.

Canonical reactions are illustrated in figure 5. If there were only one reaction involving tile T, assembly A, and assembly A', then

$$\frac{d[A]}{dt} = k_r[A'] - k_f[A][T].$$

The full model allows any tile to add to any assembly at any location, but tiles that do not match their neighbors will have weak bonds and hence dissociate quickly. In principle, the definition of the tile set, including the strengths of all pairwise interactions between tiles, uniquely determines the dynamics of self-assembly. Assembly concentrations evolve over time according to a set of ordinary differential equations, each being a sum of terms similar to the ones shown above. However, as there are an infinite number of different assemblies, an infinite number of ODE's are required. Solving this system explicitly is infeasible in general.

Therefore, we make use of the computationally tractable stochastic kinetic Tile Assembly Model (kTAM) described in [17]. In the original model, a single selected *seed tile* grows into a larger assembly by successive addition or dissociation of single tiles, as described above; it is a seeded, powered, accretion model. Fortunately, at steady-state (if it exists), the probability of observing a particular assembly in the simulation is proportional to the concentration of that same assembly in the full model at equilibrium, so long as in both cases the steady-state monomer tile concentrations are the same [12].

In order to use this model to simulate our experimental systems – where the set of tile types and their total initial concentrations are known, but the equilibrium concentration is not known – two enhancements of kTAM are necessary, which we call the *multiply-seeded, unpowered, accretion* kTAM (multi-kTAM).

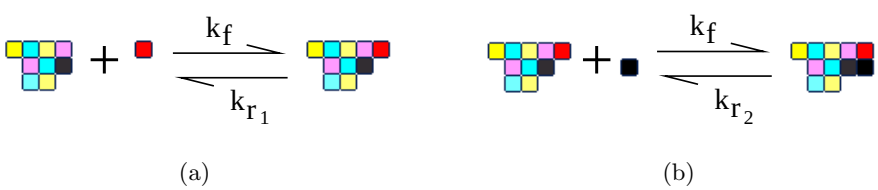


Fig. 5. Growth of tile assemblies based on (a) formation of a single bond, with energy ΔG_1 and reverse rate $k_{r_1} = e^{\Delta G_1/kT}$, and (b) formation of two bonds simultaneously, with energy ΔG_2 and reverse rate $k_{r_2} = e^{\Delta G_2/kT}$.

First, multiple assemblies are grown simultaneously; and second, the concentration of each monomer tile (shared by all assembly growth processes) is depleted with each tile addition and restored with each tile dissociation. Thus, there are two new parameters to the model: how many assemblies to simulate, N_A, and their "effective concentration", C_A, determining how much the global tile concentration changes with each monomer association or dissociation. Unfortunately, the choice of N_A and C_A can bias the steady-state distribution in the simulation, and it is at this point unclear how to optimally choose those parameters in order for the simulation to accurately reproduce the equilibrium distribution of assemblies defined by the full model. Therefore, although we believe we chose reasonable parameters for our simulation, the results of the multi-kTAM simulations must be considered qualitative until better understanding of the model is achieved.

We ran the multi-kTAM simulation for tile sets modeling each of the six experiments described in the previous section. The results, depicted qualitatively in figure 6, reproduce main features of AFM qualitatively: whereas the 1-layer, 2-layer, and V and X 2x2 tile sets all form the desired structures, but the 3x3 tile set polymerizes into assemblies with predominantly just one or two arms. At low simulation temperatures, the defect shown in figure 2(b) (inset), also occurs in the simulations.

This gives us confidence that the unexpected features we observe in the AFM images are not necessarily due to ill-formed tiles, bad DNA, old chemicals, or unknown physical or chemical effects, but rather, they may be due solely to the processes incorporated into our model. This is encouraging, because it implies that our model may provide the necessary insights required to fix the problem. However, because of the dependence on the parameters N_A and C_A, we are not confident that the distributions resulting from multi-kTAM simulations are the correct predictions of the general physical model of tile-based assembly. Furthermore, the simulations do not give us a clear intuitive understanding of *why* the model reproduces these effects. For that, we turn to simplified models that can be analyzed exactly.

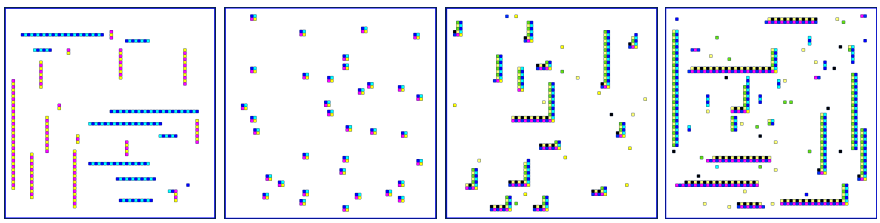
4 A Theory of Boundary Tile Assembly

Our experimental and simulation results can be summarized in the observation that small, well-defined assemblies form as expected, and that experiments in which the tiles could polymerize (*i.e.*, form arbitrarily long chains) produced a distribution of assemblies in which the target was a rarity.

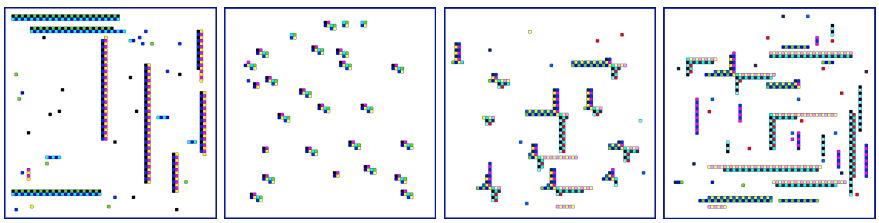
Therefore, with the goal of understanding the general features of boundary formation, we discuss a class of models that further idealizes the kTAM model and are simple enough that they can be analyzed and reasoned about intuitively.

Specifically, three models will be considered: (a) an irreversible seeded process, (b) reversible aggregation and accretion systems at equilibrium, and (c) an irreversible aggregation process. The experimentally measured arm lengths may seem to be more consistent with models (b) and (c).

In what follows, we reduce the formation of DNA tile boundaries to the formation of one dimensional heterogeneous polymers containing two types of monomers: the corner tile and the generic boundary tile. A corner tile may only occur once in the assembly and can bind up to four boundary tiles. Boundary tiles may attach to corner tiles and polymerize linearly. We refer to the boundary tile species as B and the corner tile species as C. Assemblies may consist of a line of boundary tiles, or a corner tile connected to four such lines (in this context called arms) forming an X shape. A boundary assembly is referred to as B_n, $n > 0$ and a four armed assembly that contains a corner tile and arms of lengths



(a) 1-layer boundary (b) 2x2 V (c) 3x3 V (d) 2-layer V



(e) 2-layer boundary (f) 2x2 X (g) 3x3 X (h) 2-layer X

Fig. 6. A selected sample of assemblies in the steady-state distribution of multi-kTAM simulations for each experiment shown in figures 2 and 3. Initial monomer tile concentrations were chosen exactly in correspondence to the experimental conditions, with the exception that concentration for 2-layer V and 2-layer X simulations were as in the corresponding 3x3 experiment, but with RC diluted 100:1. Each simulation was started with $N_A = 100$ seed tiles of each type, each with $C_A = [monomer]/1000$. Tile binding strengths ($\Delta G = 7.2$ kcal/mol for the **0** and **1** bond types) were chosen based on length-5 sticky ends at 25°C, as in [17], with the **B** bonds between boundary tiles RB and SB being treated as twice as strong as the bonds between other tiles. Mismatched bonds between tiles with non-complementary sticky ends were assigned $\Delta G = 0$. Simulation code and parameter files are available at `http://www.dna.caltech.edu/SupplementaryMaterial`.

m_1, m_2, m_3, and m_4 (with $m_i \geq 0$) is referred to as $CB_{m_1}B_{m_2}B_{m_3}B_{m_4}$. All the reactions that we consider involve growth or shrinkage of one of the four assemblies attached to the corner independent of the length of the others. For simplicity we will discuss models with one arm per corner tile, with the exception of the graphical depiction of assemblies produced by our various models (figures 8(d), 8(h), and 8(l)), and a discussion of the difficulty of creating an assembly where all arms are long. In all models discussed here, the dynamics of systems with X shaped corner tiles are identical to those of a one armed system, where each corner tile in the one armed system is replaced by a set of four corner tiles. The arm lengths of this group of four tiles corresponds to the four arms of a single X.

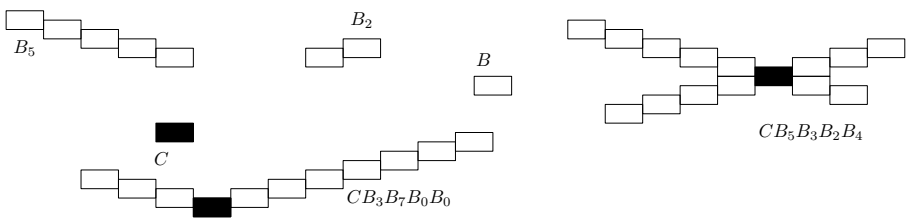


Fig. 7. Types of species in our theoretical models of boundary assembly include corner tiles attached to several boundary assemblies, and boundary and corner tiles alone.

4.1 Seeded Irreversible Tile Assembly by Accretion

The original kTAM model assumes that growth always begins from a seed tile. This supposition stems from the fact that the formation of a bond between two rule tiles is not energetically favored unless two such bonds can form at once. The assumption that growth begins from a seed nucleus is standard in crystal growth, where crystal formation is believed to be primarily governed by accretion.

Seeded irreversible accretion can be modeled by a single reaction:

$$B + CB_n \rightarrow CB_{n+1} \qquad n \geq 0 \tag{1}$$

Assembly formation through this reaction causes each boundary tile to randomly attach to a corner assembly, completing when every boundary tile has been attached to an assembly. The result is that, considered in isolation, the length of an arm is binomially distributed. Where L is a random variable representing the length of a single arm, n is a length of interest, c the number of corner tiles, and b the number of boundary tiles, and $p = \frac{1}{c}$,

$$Pr(L = n) = \binom{b}{n} p^n (1-p)^{b-n}$$

The length of the arms are approximately independent, so that the total distribution looks like a binomial distribution (figure 8(b); *ie* the expected number of arms with length n is approximately $c * Pr(L = n)$.

The variance of this distribution becomes small when the ratio of boundary to corner tiles and the total number of molecules grow large. Therefore, if this process were responsible for boundary assembly formation, corner and boundary tile experiments would reveal many evenly sized assemblies uniformly attached to corner tiles (figure 8(d)).

4.2 Reversible Tile Assembly

Another possibility was that the tile assemblies observed experimentally were at conditions approximating equilibrium. In this case, boundary tiles would be in equilibrium not only with corner assemblies but possibly also with assemblies made up solely of boundary tiles. The equilibrium conditions for two types of reaction systems are considered here: an accretion system, where only single tiles join assemblies with or without a corner tile, and an aggregation system, where boundary assemblies as well as corner assemblies can form, and boundary assemblies may interact as boundary tiles do. The reactions for the two models are (for $n > 0, m \geq 0$):

Accretion	Aggregation
$B + CB_m \rightleftharpoons CB_{m+1}$	$B_n + CB_m \rightleftharpoons CB_{m+n}$
$B + B_m \rightleftharpoons B_{m+1}$	$B_n + B_m \rightleftharpoons B_{n+m}$

The corner and boundary tiles have the same on and off rates, so we assigned each reaction a common forward rate constant k_{on}, backward rate constant k_{off}, and equilibrium constant $K_{\text{eq}} = \frac{k_{\text{on}}}{k_{\text{off}}}$.

These reactions are simple enough to calculate the equilibrium values in closed form. The accretion model has the same equilibrium as the aggregation model. [3] At equilibrium in the accretion model, for all $m, p \geq 0$ we have:

$$\frac{[B_{m+n}C]}{[B_m C][B_n]} = K_{\text{eq}} \qquad \frac{[B_{m+n}]}{[B_m][B_n]} = K_{\text{eq}} \tag{2}$$

Solving these gives

[3] The kinetics of boundary formation in a reversible accretion process are much slower than for a reversible aggregation process. Seeded reversible accretion growth also gives a geometric distribution of tiles at equilibrium. However, in an accretion model, the concentration of the single boundary species $[B]$ is much greater than in the aggregation model, where most boundary tiles can react with each other to form boundary assemblies in addition to connecting to corner tile assemblies. Thus the geometric distribution parameter $K_{eq}[B]$ for the equilibrium concentrations is much closer to one in the accretion case. In accretion, we see much more even but still geometric distributions of corner assembly sizes.

$$[B_n] = [B](K_{\text{eq}}[B])^{n-1} \qquad [B_m C] = [C](K_{\text{eq}}[B])^m \tag{3}$$

These equation predict that at equilibrium the sizes of assemblies should be geometrically distributed with parameter $(K_{\text{eq}}[B])$, where $[B]$ is the concentration of lone boundary tiles at equilibrium. This is a well known fact about polymerization [4, 5].

The equilibrium solution is entirely symmetric with respect to the four arms of an X shape. Thus for a four armed assembly the concentration of an assembly type is

$$[CB_{m_1}B_{m_2}B_{m_3}B_{m_4}] = [C](K_{eq}[B])^{(m_1+m_2+m_3+m_4)}$$

This implies that the total number of tiles connected to a corner X tile is geometrically distributed. Because tiles are distributed independently across each of the four arms of a corner X tile, we can conclude that corner tiles with four long arms are rare, for any reasonable definition of "long"[4]: the probability that all four arms are long is the fourth power of the probability that any single arm is long. Similarly, for corner V tiles, the probability of two long arms is the square of the probability of one. The arm length distribution for of a reversible process that approaches equilibrium is shown in figure 8(f). Most corner tiles do not have more than one or two nontrivial arms (figure 8(h)).

4.3 Irreversible Tile Assembly by Aggregation

Irreversible tile assembly would appear to be a limiting case of the reversible aggregation model, where binding reactions become so biased toward the forward reaction that they are essentially irreversible. The results (shown in figures 8(j), 8(l)) of a simulated irreversible aggregation process are similar to the reversible case, but have a more accentuated fraction of length-0 arms, and the distribution has longer tails. In irreversible aggregation tiles "freeze" when they are first attached to a corner tile, rather than finding a final equilibrium state. Therefore, the results of the irreversible tile assembly model are likely to resemble a kinetic intermediate rather than the equilibrium of a set of tiles where binding between tiles is very strong.

5 Discussion

As part of our work on the creation of complex self-assembled structures, we have designed and experimentally verified the binding specificities of many different DNA double crossover molecules. We have been able to form specific structures with these tiles, but not yet control their polymerization. Moreover,the kinetic

[4] For example, one could say an arm is "long" if it has at least half the average number of tiles.

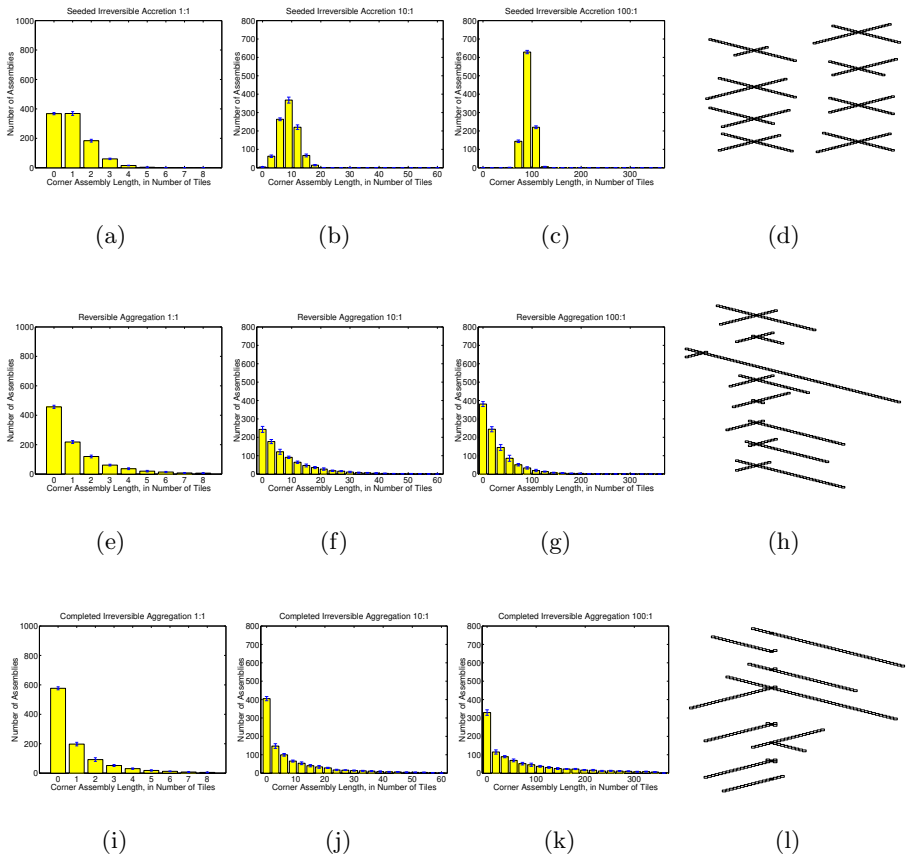


Fig. 8. Distributions of the size of assemblies attached to corner tiles for, (8(a), 8(b), 8(c)) the seeded irreversible accretion model at completion, (8(e), 8(f), 8(g)) reversible aggregation model close to equilibrium, and (8(i), 8(j), 8(k)) the irreversible aggregation model at completion. The column shows results for 1:1, 10:1, and 100:1 stoichiometric ratios of boundary tiles to corner tiles. The last column (8(d), 8(h), 8(l)) gives a visual depiction of eight assemblies randomly chosen from each set of results, for each of the three models when a 10:1 stoichiometric ratio is used. All simulations used c = 1,000 corner tiles and b = 1,000, 10,000 or 100,000 corner tiles. For irreversible models, at each step two assemblies were chosen from the collection, and if a reaction between them was allowed by the model being simulated, the assemblies were replaced by the reaction product. For the reversible model, each possible association or dissociation reaction between two assemblies was chosen randomly, weighted by rates K_{on} and K_{off} where $\frac{K_{on}b}{K_{off}[B]} = K_{eq}$. Assuming the concentration of boundary tiles was $[B] = 0.1\mu$M, the equilibrium coefficient in the reversible reaction was $K_{eq} = 1$ nM. Simulation code and scripts for these figures are available at `http://www.dna.caltech.edu/SupplementaryMaterial`.

tile assembly model's predictions correspond well with what is seen experimentally. We can further simplify this model to explore extremes in the behavior of tiles, and to provide simpler explanations for our experimental results.

The experimental statistics shown here represent an initial effort at quantification, and are not yet conclusive. Several sources of error may be present: A small number of images were analyzed, selected for clarity; they may not be representative of typical results. Sample preparation techniques, variations in the adhesiveness of the mica, or AFM tip interactions might affect the observed distribution of assemblies. The size or shape of an assembly might further affect its ability to adsorb.

Definitive and more substantial conclusions would require additional samples to be analyzed; in particular, experiments that vary the stoichiometric ratio of boundary to corner tiles, imaged with resolution than can distinguish between a boundary assembly and a corner assembly with one arm, would help discriminate between the growth models. It would also be instructive to compare to boundary formation at colder temperatures, which should be an essentially irreversible process, and to boundaries formed by slow annealing, which should yield the distribution of a reversible process.

Furthermore, the experimental results shown here are insufficient for extracting thermodynamic parameters (ΔG, or equivalently, K_{eq}) for tile binding events between our specific molecules. Are the parameters used in the simulations reasonable? Based on generic nearest-neighbor parameters for base stacking within a duplex [10], coaxial base stacking at nick sites [13], and dangling ends [2], we estimate[5] rule tile sticky ends to bind with K_{eq} between 2 and 200 μM, and boundary tile sticky ends to bind with K_{eq} between 10 and 100 nM at 25°C. The multi-kTAM simulations used 6 μM and 34 pM respectively, while the reversible aggregation simulations used 1 nM for the boundary tiles. Thermodynamic parameters for our DNA tiles should be measured experimentally.

Despite these difficulties, the results obtained here, when compared to the models, support qualitative conclusions, because the generic shape of the assembly shape distributions is not dependent upon the exact parameter values. Among the simplified models, the predictions of the irreversible aggregation (figure 8(i)) and reversible aggregation 8(e)) are closer to the experimental observations (figure 4) than the irreversible accretion model (figures 8(a)). These models also seem *a priori* more plausible than the irreversible accretion model, because boundary tiles have no mechanism to prevent aggregation of boundary assemblies prior to attachment to a corner assembly. Unfortunately, it will be difficult to build corner assemblies with long arms using irreversible or reversible aggregation processes.

Thus, the models suggest that one difficulty in creating large, uniformly-sized V and X assemblies may be the ability of boundary assemblies to form in the absence of corner tiles. In this case, neither increasing the concentration of bound-

[5] The wide range of the estimates is due to sequence dependence of the parameters (giving different sticky ends different K_{eq}'s), as well as being due to uncertainty about the reported parameters.

ary tiles in our reactions, nor increasing the binding strength between boundary tiles, will significantly increase the concentration of large, uniformly-sized corner assemblies. This also means that using ligase to lock the tiles together will be ineffective for making large, uniformly-size corner assemblies.

On the other hand, the seeded accretion model shows that if growth can be constrained to occur from the corner tile only, then large, relatively uniformly-sized corner assemblies will result. We are designing a new set of boundary tiles for a Sierpinski triangle that, even in an aggregation model, dramatically reduce the ability of boundary tiles to spontaneously assemble in the absence of a corner nucleus [11]. Preventing such spontaneous growth is also likely to reduce the effects of stoichiometry poisoning, discussed below.

Several other factors, other than the ones discussed above, may also be playing a role in shaping the experimentally observed distributions. For example, malformed DNA tiles could terminate growth, leading to truncated distributions. More subtle is the effect of stoichiometry poisoning. Ideally, the concentrations of RB and SB tiles in our experiments are equal, but in practice small pipetting errors cause differences in the amount of each type of tile present. These slight differences can greatly influence the size of assemblies that form, since in a polymerization reaction involving alternating monomer types, the more unequal the two concentrations are, the smaller the polymers become on average [6]. In the case where [RB] > [SB], most assemblies end up in the form RB-(SB-RB)$_n$. These assemblies cannot combine, so assembly slows down or halts. While individual pipetting steps are accurate to within 2 percent, multiple pipetting steps might increase the total error to as much as 5 or 10 percent. For a concentration difference between RB and SB tiles in this range, assemblies larger than 10 or 20 tiles would be rare.[6]

Previous work on DNA computation by self-assembly made use of DNA triple-crossover tiles that assemble into a seeded 2-layer one-dimensional array to compute a 4-bit cumulative XOR operation [8]. Attempts to extend that work to to longer inputs would likely be governed by the same principles discussed here.

The ultimate goal of one-dimensional boundary formation, in the context of algorithmic self-assembly, is to provide input for subsequent two-dimensional growth. However, the relationship between boundary formation and two dimensional crystal growth is not well understood. It is possible, for example, that problems with boundary formation would be ameliorated if rule tiles were simultaneously filling in the space between two arms on a V or X, thus reducing the troublesome statistical independence.

[6] As the concentrations of the two monomer types become more uneven, the distribution is limited by the ratio $r < 1$ rather than by the equilibrium constant between the monomers. Under irreversible binding, the size distribution is

$$[\text{RB-(SB-RB)}_n] = r^{(2n+1)} \frac{(1-r)^2}{1+r}. \tag{4}$$

In the case of 5-10 percent pipetting error $r \approx 0.9$.

Acknowledgments

We would like to thank Bernie Yurke for pointing out the effects of slight differences in stoichiometry, and for his very descriptive term "stoichiometry poisoning". Paul Rothemund, Rizal Hariadi, and other members of the DNA Lab provided helpful hints and stimulating conversation. This work was supported by NSF CAREER Grant No. 0093486, DARPA BioComputation Contract F30602-01-2-0561, NASA NRA2-37143, and GenTel.

References

1. V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr. *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, 2000.
2. S. Bommariro, N. Peyret, and J. SantaLucia, Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Research*, 28(9):1929–1934, 2000.
3. B. A. Bondarenko. *Generalized Pascal Triangles and Pyramids, Their Fractals, Graphs and Applications*. The Fibonacci Association, 1993. Translated from the Russion and edited by Richard C. Bollinger.
4. P. J. Flory. Molecular size distribution in linear condensation polymers. *Journal of the American Chemical Society*, 58:1877, 1936.
5. P. J. Flory. Fundamental principles of condensation polymerization. *Chemical Reviews*, 39:137, 1946.
6. P. J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, 1953.
7. T.-J. Fu and N. C. Seeman. DNA double-crossover molecules. *Biochemistry*, 32:3211–3220, 1993.
8. C. Mao, T. H. LaBean, J. H. Reif, and N. C. Seeman. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 407(6803):493–496, 2000.
9. P. W. K. Rothemund and E. Winfree. The program-size complexity of self-assembled squares. In *Symposium on Theory of Computing (STOC)*. ACM, May21-23 2000.
10. J. SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci. USA*, 95:1460–1465, 1998.
11. R. Schulman and E. Winfree. Controlling nucleation rates in algorithmic self-assembly. In preparation.
12. R. Schulman and E. Winfree. Relationships among tile assembly models. In preparation.
13. V. A. Vasiliskov, D. V. Prokopenko, and A. D. Mirzabekov. Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips. *Nucleic Acids Research*, 29(11):2303–2313, 2001.
14. H. Wang. Proving theorems by pattern recognition. II. *Bell System Technical Journal*, 40:1–42, 1961.
15. H. Wang. Dominoes and the AEA case of the decision problem. In J. Fox, editor, *Proceedings of the Symposium on the Mathematical Theory of Automata*, pages 23–55, Brooklyn, New York, 1963. Polytechnic Press.
16. E. Winfree. On the computational power of DNA annealing and ligation. In R. J. Lipton and E. B. Baum, editors, *DNA Based Computers: DIMACS Workshop, April 4, 1995*, volume 27, pages 199–221, Providence, RI, 1996. American Mathematical Society.

17. E. Winfree. Simulations of computing by self-assembly. Technical Report CS-TR:1998.22, Caltech, 1998. Originally appeared in the preliminary proceedings of the 4th DIMACS Meeting on DNA Based Computers, held at the University of Pennsylvania, June 16-19, 1998.
18. E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394:539–544, 1998.