Thermodynamic Simulations of DNA Tile Self-Assembly

Kenichi Fujibayashi
Department of Computational Intelligence and
Systems Science
Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama
226-8503, Japan
fuji@mrt.dis.titech.ac.jp

Satoshi Murata
Department of Computational Intelligence and
Systems Science
Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama
226-8503, Japan
murata@dis.titech.ac.jp

ABSTRACT

Self-assembling DNA complexes have been intensively studied in recent years aiming to achieve bottom-up construction of nanoscale objects. Among them, a DNA complex called DNA tile is known for its high programmability. By using a set of DNA tiles, we are able to self-assemble two-dimensional crystals with programmable patterns. This is called algorithmic self-assembly. In order to create a wide range of complex objects by this self-assembly process, we need a methodology to predict its behavior. Especially, the relationship between the error rates and growth speed is of our interest. To estimate these properties, we use thermodynamic simulations based on the Monte Carlo method. However, conventional simulation models assume some much simplified conditions, therefore cannot explain the results of crystal growth experiments. Here, we propose Realistic Tile Assembly Model (R-TAM), in which we model the detailed conditions of the experimental protocol. We will show that the simulation can explain growth process of DNA tile crystals in experiment.

Categories and Subject Descriptors

I.6.5 [**Simulation and Modeling**]: Model Development—*modeling methodologies*

1. INTRODUCTION

In the last decade, aiming to realize bottom-up construction of structures with nanoscale precision, the self-assembly of engineered DNA molecules has been intensively studied. Varieties of two-dimensional (2-D) complexes made of DNA strands called *DNA tiles* have been proposed as building blocks for self-assembled lattices (often referred as *DNA crystals*) [13]. For instance, a DNA tile called Double Crossover (DX) molecule [4] is made of four or five short single-stranded DNA, which has a single-stranded dangling end, called the *sticky end*, at each of its four corners. A well-designed set of DNA tiles can form a large 2-D crystal with a programmed pattern such as Sierpinski triangle

by the specific connective relationship defined at the sticky ends [5, 7]. In principle, we can implement any program by encoding rules in sequences at the sticky ends of DNA tiles. We call this *algorithmic self-assembly*. The algorithmic self-assembly is equivalent to the time course evolution of the *1-D cellular automaton*, and thus is Turing universal [14]. We can theoretically exploit fabrication tasks such as constructing the patterns that define certain digital circuits [3].

Recently, the algorithmic self-assembly of DNA tile with a relatively simple tile set have been experimentally verified. For example, Rothemund et al. successfully made crystals with a Sierpinski fractal pattern using a set of tiles corresponding to XOR logic calculation [7]. A set of binary counting tiles was implemented by Barish et al. [1]. In these experiments, error rates in the range of $1 \sim 10\%$ are observed. Suppression of errors is crucial to make the algorithmic self-assembly come into practical use.

The errors may due to inaccurate control of temperature in the solution and the concentration of DNA tiles. Different DNA strands are annealed in temperature-controlled bath. They form the DNA tiles at first, and then the tiles form 2-D crystals by algorithmic self-assembly. Desired algorithmic self-assembly occurs during annealing, however, it is difficult to determine the exact temperature and the concentration of DNA tiles when that happens. In order to build complex nanostructures by self-assembly, we need a methodology to predict the behavior of the self-assembly process.

Winfree developed Tile Assembly Simulator (called Simple Tile Assembly Model (S-TAM) hereafter), to model the process of algorithmic self-assembly [12]. It is based on thermodynamics and kinetics of DNA tile self-assembly and is able to predict properties such as error rates and growth speed. Four conditions are assumed in S-TAM: (1) The temperature of the DNA solution is always constant during the growth. (2) The concentration of each tile species is also constant. (3) The flip and rotation of the tiles are negligible. (4) The bonding strength of every matching pair of sticky ends has a uniform value and the bonding strength for undesired pairs of sticky ends is nil.

S-TAM tells us what condition is suitable for obtaining a large and errorless DNA tile crystal. While this is significant, it still oversimplifies the actual process in the test tube. For a more precise prediction we need to take the following conditions into account: (1) The temperature of the DNA
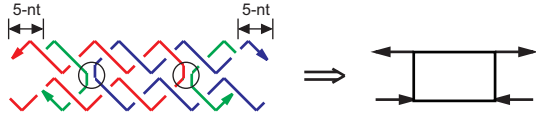
Figure 1: The DAO-E tile and corresponding abstract tile.

solution is not constant in the experiments. A mixture of all the component DNA strands is annealed from high temperature (∼90°C) down to room temperature in one test tube. (2) The concentration of monomer DNA tiles does not remain the same. It decreases over time during the annealing. (3) The flips and rotations of DNA tiles are disregarded in S-TAM, however, they do affect the self-assembling process. (4) The bonding strength of different sticky ends has different values, but this is not taken into consideration in S-TAM.

Proceeding sections are organized as follows: In Section 2, a conventional simplified model of DNA tile self-assembly is reviewed. In Section 3, we propose a more realistic model of DNA tile self-assembly by considering several conditions such as annealing temperature profile and varying concentrations of each DNA tile species. Section 4 compares simulation results of conventional and proposed models, demonstrating the effectiveness of the latter as a prediction and design tool for DNA tile self-assembly. Section 5 gives the conclusion.

2. SIMPLE TILE ASSEMBLY MODEL

The Simple Tile Assembly Model (S-TAM) is the first simulation model of DNA tile self-assembly [1, 7, 13]. We assume the readers are familiar with the formal details of the S-TAM (see [12]). By using S-TAM, we can simulate self-assembly of DX molecules. In Figure 1, DAO-E (Double-crossover, Antiparallel junction, Odd intramolecular spacing, Even intermolecular spacing) type is shown [4]. This tile has 5-nt sticky end at both ends of each helix. These tiles can bind with other tiles when their sticky ends are complementary. In S-TAM, only matching pair of sticky ends possess bonding strength, otherwise there is no interaction between them. The orientation and the flip of the tiles are also not considered.

An appropriately designed set of DNA tiles can generate self-assembled DNA crystals with a specific pattern. For example, a set of four species of DNA tiles that represents XOR (exclusive OR) function generates a Sierpinski triangle pattern (Figure 2). To give initial values for the algorithmic self-assembly, we need some *seed* structures. It is a row of tiles such as '··· 0001000 ···' which defines initial bit sequence of self-assembly. We assume the temperature of the solution and the concentration of the DNA tiles are controlled so that only the tiles that have two matching sticky ends can attach to the seed structure. Tiles can attach themselves not only to the seed structure but to other tiles that already have been attached to the seed structure. These attachments occur *asynchronously*, and as a consequence the crystal grows. Only when all the tiles are correctly assembled, we have the desired pattern.
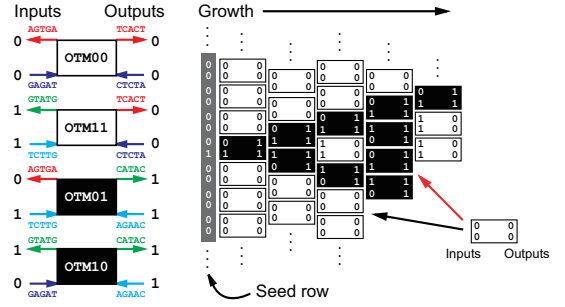
Figure 2: Sierpinski triangle pattern generated by XOR tile set.
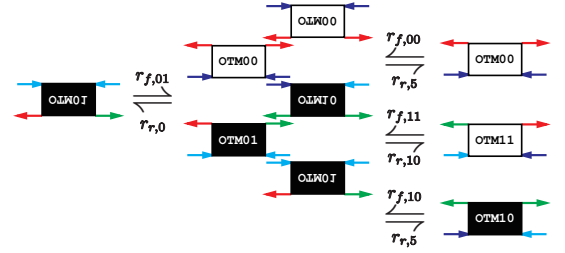
Figure 3: Rates for tile association and tile dissociation.

In the following models (both in S-TAM and R-TAM explained later), any interaction among the clusters of aggregated tiles is not taken into account. We consider only association and dissociation between a monomer tile and the crystal. Balance between the association and dissociation rate determines whether the attachment occurs or not (Figure 3).

The forward (association) rate $r_{f,x}$ of tile T_x (x: a tile type) is given by

$$r_{f,x} = k_f[T_x] \text{ /sec.} \qquad (1)$$

This means the forward rate is dependent only on the concentration $[T_x]$. The forward rate constant depends on the temperature: $k_f = A_f e^{-E_f/RT}$ /M/sec where $A_f = 5 \times 10^8$ /M/sec, $R = 2$ cal/mol/K and $E_f = 4000$ cal/mol is the activation energy for the reaction. In S-TAM, the concentration is assumed to be constant.

On the other hand, the reverse (dissociation) rate $r_{r,b}$ of tile T_x depends exponentially on b, the sum of perfectly matched base pairs of four sticky ends, and an absolute temperature T. We do not count b for partially matched sticky ends in S-TAM, therefore, $b = \{0, 5, 10, 15, 20\}$ in cases of 0-, 1-, 2-, 3- and 4-match. We also assume that the temperature T is always constant. As a result, we have,

$$r_{r,b} = k_f e^{\Delta G_b^\circ/RT} = k_f e^{(11-\frac{4000}{T})b+3} \text{ /sec.} \qquad (2)$$

The standard free energy ΔG_b° cal/mol can be calculated from the standard enthalpy ΔH_b° kcal/mol and the standard entropy ΔS_b° cal/mol/K: $\Delta G_b^\circ = \Delta H_b^\circ - T\Delta S_b^\circ$. ΔH_b° and ΔS_b° are average values estimated by the nearest neighbor (NN) model [9]: $\Delta H_b^\circ \approx -8b$ kcal/mol and $\Delta S_b^\circ \approx -22b - 6$ cal/mol/K.

When the concentration of the tiles is increased (or equivalently, the temperature is decreased), the forward rate becomes larger than the reverse rate. This means that the crystal grows faster, but at the same time, more erroneous tiles are embedded in the crystal. On the contrary, when the concentration is decreased (or temperature is increased), the crystal growth slows down, but fewer errors occur. It is certain that some optimal concentration and temperature do exist.

TAM is a thermodynamic simulation model based on the Monte Carlo method [12]. The tiles are repeatedly appended or deleted from the aggregation according to the association or dissociation rate at the site (i, j). A 2-D array stores the information on the arrangement of the tiles in the current assembly. A seed structure is pre-set on the array at an initial time $t = 0$ sec. At every iteration step, one of the two events is stochastically chosen to simulate the reaction between a monomer and a crystal: an *on*-event, where a new tile is added to the array, and an *off*-event, where a tile in the assembly is removed.

$$r(i,j) = \begin{cases} \sum_x r_{f,x} & \text{(Empty sites adjacent to a crystal.)} \\ r_{r,b} & \text{(Tiles composing a crystal.)} \\ 0 & \text{(otherwise.)} \end{cases}$$
(3)

We can calculate the total rate of association or dissociation at all sites, $r_{total} = \sum_m \sum_x r_{f,x} + \sum_n r_{r,b}$, based on the number of empty sites around the crystal (denoted by m) and the number of tiles in the crystal (denoted by n). On-event will be chosen with the probability of $\Pr(on) = \sum_x r_{f,x}/r_{total}$, and in this case, tile T_x is attached to the crystal at (i, j) with the probability of $r_{f,x}/\sum_x r_{f,x}$. Similarly, off-event will be chosen with the probability of $\Pr(off) = r_{r,b}/r_{total}$, and in this case, tile T_x is removed from the array. Single time step for each event is estimated by

$$\Delta t = -\ln([0.0, 1.0))/r_{total}.$$
(4)

Once an event is chosen and executed, all the rates are recalculated according to the new status of the crystal.

3. REALISTIC TILE ASSEMBLY MODEL

We propose Realistic Tile Assembly Model (R-TAM), in which we model the detailed conditions of DNA tile self-assembly. This model well-corresponds with the experimental results and suggests useful information for DNA tile design. Here, we take four conditions into account that were neglected in S-TAM.

Decreasing temperature by annealing. Actual annealing schedule of the experiment such as

$$T = 73e^{-0.001t/60} + 22 \ °C$$
(5)

is used in R-TAM. In this case, the tube was annealed in a hot water bath insulated by Styrofoam container [5].

Decreasing tile concentration. The material in a tube is finite. Therefore, the concentration of each tile species decreases as the crystal grows. Consequently, the speed of growth gradually becomes slower or suspended. To evaluate how the change in concentration affects the rate of growth,
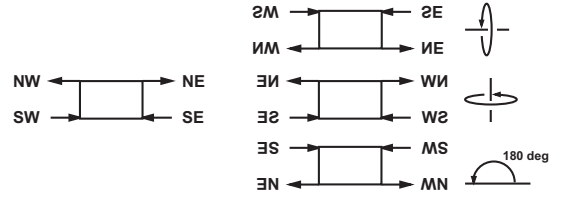

Figure 4: Horizontally or vertically flipped tiles and the rotated tile in $180°$.
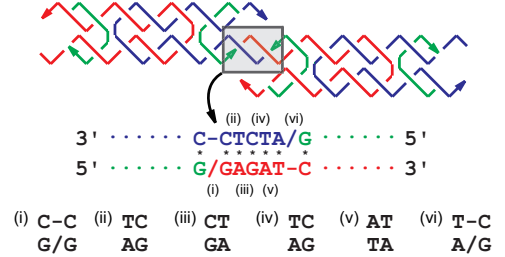

Figure 5: Estimation of standard free energy between specific pair of sticky ends by the NN model.

we calculate the concentration of each tile species by counting the number of tiles forming the crystal:

$$[T_x] = [T_x]_{init} - \frac{n_{used,x}}{N}[Sd],$$
(6)

where $[Sd]$ is the concentration of the seed structure, $n_{used,x}$ is the number of tile composing each crystals, and N is the number of seed structures in the simulation. Thus, we can calculate forward rate in the next step.

Flips and Rotations of tiles. DNA tiles approach the growth front of crystals from random directions. In R-TAM, flipped and rotated tiles are considered. There are three possible orientations of a tile as shown in Figure 4.

Estimation of $\Delta G°$ by the NN model. When two tiles match by a sticky end, the standard free energy can be estimated by the NN model. The calculation of $\Delta G°_{total}$ can be divided into six segments of nearest base pairs in the sticky end as shown in Figure 5. Standard free energies of segment (ii) to (v) can be calculated by normal NN data [8, 10]. We must be cautious about segments (i) and (vi), because they include a *nick*. Data of *coaxial stacking* must be used for these segments [6, 10]. The standard enthalpy and the standard entropy are calculated:

$$\begin{aligned} \Delta H°_{total} &= \Delta H°_{(i)} + \Delta H°_{(ii)} + \Delta H°_{(iii)} + \Delta H°_{(iv)} \\ &\quad + \Delta H°_{(v)} + \Delta H°_{(vi)} + \Delta H°_{init} \\ &= -61.0 \ \text{kcal/mol}, \end{aligned}$$
(7)

and

$$\begin{aligned} \Delta S°_{total} &= \Delta S°_{(i)} + \Delta S°_{(ii)} + \Delta S°_{(iii)} + \Delta S°_{(iv)} \\ &\quad + \Delta S°_{(v)} + \Delta S°_{(vi)} + \Delta S°_{init} \\ &= -176.2 \ \text{cal/mol/K}. \end{aligned}$$
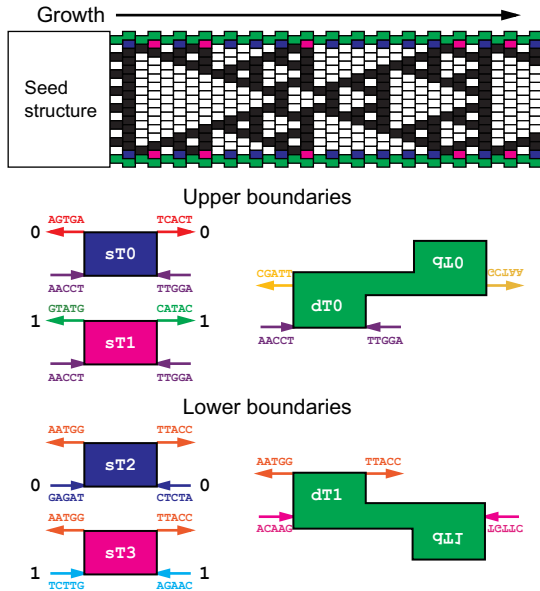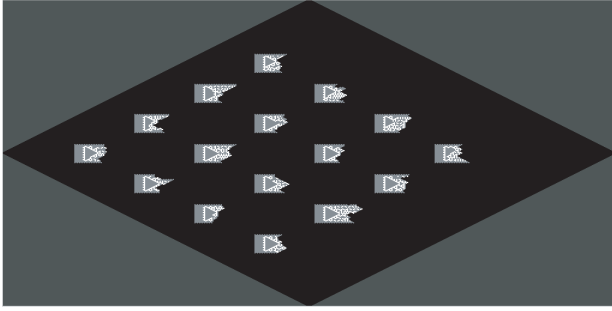(8)

Figure 6: DNA ribbon and added boundary tiles.

Figure 7: Simulated self-assembly of DNA ribbons.

Therefore,

$$\begin{aligned} \Delta G^{\circ}_{total} &= \Delta H^{\circ}_{total} - T\Delta S^{\circ}_{total} \\ &= -61000 + 176.2T \text{ cal/mol.} \end{aligned} \qquad (9)$$

We can estimate the dissociation rate of the DNA tiles with specific sticky ends by this method.

4. SIMULATIONS

In order to demonstrate the ability of R-TAM, we have simulated a complicated experiment of DNA tile self-assembly called DNA ribbons that is a fixed-width self-assembled structure made of DNA tiles (see Figure 6). To make the ribbon structure, six kinds of boundary tiles are added on the original tile set of four XOR tiles. The ribbon structure is suitable for measuring the error rate on actual samples. The details of this experiment are to be reported [5].

We have compared the simulation results of S-TAM and R-TAM for the DNA ribbon. We have simulated 16 ribbons for ($t =$) 75 hours at a time, and repeated the simulation 50 times (800 ribbons in total). Figure 7 shows a snapshot of R-TAM. The concentration of the DNA tiles gives the initial condition for both models: [seed structures] = 1 nM,
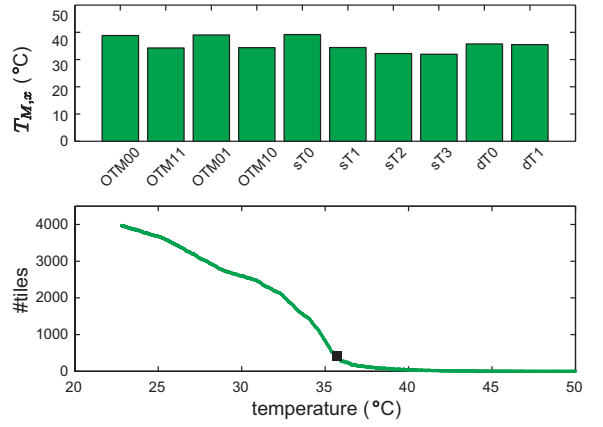

Figure 8: The predicted melting temperature for each tile (top) and the number of tiles (16 ribbons) plotted against the temperature in R-TAM (bottom).
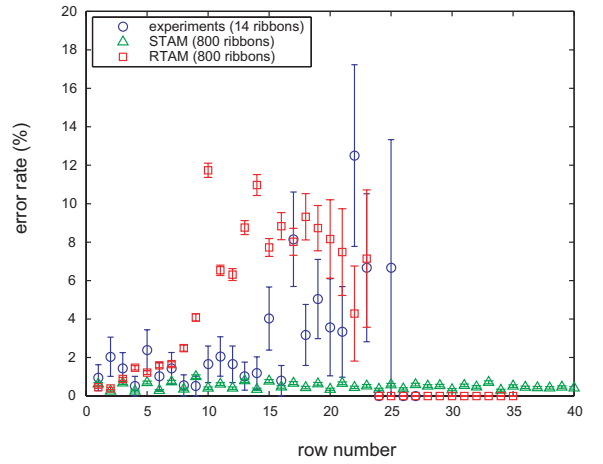

Figure 9: Error rates in each row of ribbons.

[4 XOR tiles] = 50 nM, [6 boundary tiles] = 10 nM. In S-TAM, the concentration of the tiles is kept at constant. The temperature is determined by Equation 5 in R-TAM, and kept at 30°C in S-TAM.

The *melting temperature* can be also calculated by sticky-end sequences of the tile set [10] (see Figure 8, top):

$$T_{M,x} = 1000\Delta H^{\circ}_x/(\Delta S^{\circ}_x + R\ln([\text{T}_x]/4)) - 273.15 \text{ °C}, \quad (10)$$

where ΔH°_x and ΔS°_x are the standard enthalpy and the standard entropy in the case of two matching sticky ends, respectively. The average of $T_{M,x}$ gives the melting temperature of ribbon crystals: $T_M \approx 35.5$°C. On the other hand, Figure 8 bottom shows the increase of the number of tiles plotted against the temperature. The growth started around the calculated melting temperature.

Figure 9 shows the results of the simulation. An error rate (%) is evaluated by counting the number of mismatched tiles in each row of ribbons. Experimental results obtained by atomic force microscopy (AFM) are also shown in the
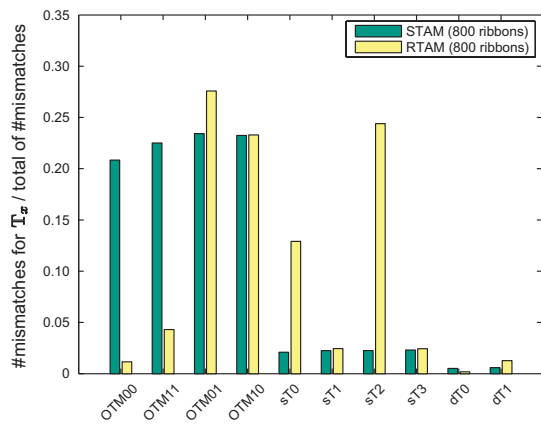
Figure 10: Frequencies of errors in each tile type.

same plot. The error rates for S-TAM stayed at $\sim 0.5\%$ because of the constant temperature and tile concentrations. S-TAM gives an infinite growth of ribbons. On the other hand, R-TAM can explain the growth process of the experiment. When the temperature gradually decreases, error rates increase. The average error rates from the first row to the 8th row was close to the experimental results ($e_{1\sim 8}^{R-TAM} \approx 1.3\%$ vs. $e_{1\sim 8}^{exp} \approx 1.3\%$). However, error rates from the 9th row to the 16th row did not correspond with experimental data.

More detailed analysis is possible by the simulation. Figure 10 shows the frequencies of errors for each tile type x. S-TAM does not consider any specific sequences on the sticky ends, thus there is no difference in error rates for the tile type (error rates for XOR tiles are higher than boundary tiles because of higher concentration). On the contrary, with R-TAM the mismatch rates depends on tile type x. Especially, one of the boundary tiles sT2 has a high error rate. It is difficult to explain why such an uneven distribution of error rate is observed. Many factors can be suspected such as sticky end sequences, initial concentrations and their ratio, and so on. There are implications that the errors are caused by the boundary tiles in the experiment (these errors cannot be detected by AFM).

5. CONCLUSIONS
We have developed a simulator called Realistic Tile Assembly Model (R-TAM), which is able to reproduce the practical conditions of algorithmic self-assembly of DNA tiles: annealing profile, varying concentration of each DNA tile species, simulation of rotated or flipped tiles, and bonding strength dependent on sticky end sequence. The performance of the simulator is evaluated by comparison with the results of real experiments, and the properties of the experiments were well described by the simulation.

In order to improve the simulator, we have to model two unknown parameters, which have not been implemented in R-TAM: additional interaction by hairpin structures protruding from the tile body for AFM imaging and modeling of spurious nucleation without seed structures. Also, we need to evaluate the simulator by data taken by other experiments [1, 2, 7, 11, 13].

The R-TAM can be extended to predict and design various types of DNA tile self-assembly. We think this kind of simulation tool is indispensable for future DNA nanotechnology.

7. REFERENCES
[1] Barish, R.D., Rothemund, P.W.K. and Winfree, E. Two computational primitives for algorithmic self-assembly: Copying and counting. *Nano Lett.*, 5:2586–2592, 2005.
[2] Chen, H.-L., Schulman, R., Goel, A. and Winfree, E. Preventing facet nucleation during algorithmic self-assembly. *Nano Lett.*, 2007 (to appear).
[3] Cook, M., Rothemund, P.W.K. and Winfree, E. Self-assembled circuit patterns. In Chen, J. and Reif, J.H., editors, *DNA Computing 9*, volume 2943 of *LNCS*, pages 91–107, Berlin, 2004. Springer.
[4] Fu, T.-J. and Seeman, N.C. DNA double-crossover molecules. *Biochemistry*, 32:3211–3220, 1993.
[5] Fujibayashi, K., Hariadi, R., Park, S.H., Winfree, E. and Murata, S. Toward reliable algorithmic self-assembly of DNA tiles: a fixed-width cellular automaton pattern. 2007 (submitted).
[6] Peyret, N. *Prediction of Nucleic Acid Hybridization: Parameters And Algorithms*. PhD thesis, Wayne State University, 2000.
[7] Rothemund, P.W.K., Papadakis, N., and Winfree, E. Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biol.*, 2:2041–2053, 2004.
[8] SantaLucia Jr., J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Nat. Acad. Sci. USA*, 95:1460–1465, 1998.
[9] SantaLucia Jr., J., Allawi, H.T. and Seneviratne, P.A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35:3555–3562, 1996.
[10] SantaLucia Jr., J. and Hicks, D. The thermodnamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.
[11] Schulman, R. and Winfree, E. Synthesis of crystals with a programmable kinetic barrier to nucleation. *Proc. Nat. Acad. Sci. USA*, 2007 (to appear).
[12] Winfree, E. Simulations of computing by self-assembly. Caltech CSTR:1998.22, California Institute of Technology, 1998. http://www.dna.caltech.edu/Papers/simulation.pdf.
[13] Winfree, E., Liu, F., Wenzler, L.A. and Seeman, N.C. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394:539–544, 1998.
[14] Winfree, E., Yang, X. and Seeman, N.C. Universal computation via self-assembly of DNA: Some theory and experiments. In Landweber, L.F. and Baum, E.B., editors, *DNA Based Computers II*, volume 44 of *DIMACS*, pages 191–213, Providence, RI, 1998. AMS Press.