Nat Comput (2009) 8:589–612
DOI 10.1007/s11047-008-9093-9

Error suppression mechanisms for DNA tile self-assembly and their simulation

Kenichi Fujibayashi · David Yu Zhang · Erik Winfree · Satoshi Murata

Published online: 9 July 2008

Abstract Algorithmic self-assembly using DNA-based molecular tiles has been demonstrated to implement molecular computation. When several different types of DNA tile self-assemble, they can form large two-dimensional algorithmic patterns. Prior analysis predicted that the error rates of tile assembly can be reduced by optimizing physical parameters such as tile concentrations and temperature. However, in exchange, the growth speed is also very low. To improve the tradeoff between error rate and growth speed, we propose two novel error suppression mechanisms: the Protected Tile Mechanism (PTM) and the Layered Tile Mechanism (LTM). These utilize DNA protecting molecules to form kinetic barriers against spurious assembly. In order to analyze the performance of these two mechanisms, we introduce the hybridization state Tile Assembly Model (hsTAM), which evaluates intra-tile state changes as well as assembly state changes. Simulations using hsTAM suggest that the PTM and LTM improve the optimal tradeoff between error rate ϵ and growth speed r, from $r \approx \beta\epsilon^{2.0}$ (for the conventional mechanism) to $r \approx \beta\epsilon^{1.4}$ and $r \approx \beta\epsilon^{0.7}$, respectively.

Keywords Algorithmic self-assembly · Assembly errors · Branch migration · DNA self-assembly · Protecting molecules

K. Fujibayashi · S. Murata (✉)
Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Midori-ku, Yokohama 226-8502, Japan
e-mail: murata@dis.titech.ac.jp

K. Fujibayashi
e-mail: fuji@mrt.dis.titech.ac.jp

D. Y. Zhang · E. Winfree
Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: dzhang@dna.caltech.edu

E. Winfree
Department of Computer Science, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: winfree@caltech.edu

1 Introduction

Molecular self-assembly of synthetic components is a powerful method to construct nanoscale objects (Whitesides et al. 1991; Seeman 2005). DNA is one of the most suitable molecules for designing such self-assembling building blocks for the following reasons: First, it is a locally stiff molecule in its double-helix conformation, and its rigidity can be increased if several double helices are coupled together. Second, intra- and inter-molecular interaction among DNA molecules are highly predictable from Watson–Crick base pairing, therefore interactions can be programmed with sequence composition of DNA. Third, DNA can be synthesized by an automatic synthesizer based on phosphoramidite chemistry. Fourth, sophisticated off-the-shelf tools for DNA biotechnology are available.

In this article, we focus on self-assembly of nanoscale building blocks called DNA "tiles." The standard DNA tile is composed of a pair of interwoven double helices of DNA. Both ends of each double helix are extended by single-stranded DNA "sticky ends" that adhere specifically to other tiles that have Watson–Crick complementary sticky ends. In this manner, a set of DNA tiles forms implementations of Wang tiles (Whitesides et al. 1991; Seeman 2005), which can be used to generate many intricate aperiodic patterns, such as the Sierpinski fractal (Wang 1961, 1963), or the binary counting pattern (Winfree 1998; Rothemund et al. 2004) An assembly of DNA tiles can be regarded as an emulation of a one-dimensional (1-D) cellular automaton, with each new layer corresponding to the state of the cellular automaton at a new time step. Since 1-D cellular automata are capable of simulating Turing machines, self-assembly of DNA tiles is thus capable of universal computation (Winfree 1996).

DNA tile-based self-assembly has been proposed as a bottom-up technique for fabricating complex molecular structures (Rothemund and Winfree 2000; Adleman et al. 2001; Cook et al. 2004; Soloveichik and Winfree 2004). Because these works focus on the logic of programming self-assembly, they consider the situation where sticky-end binding specificity is infallible. Realistically, however, correctness of matching between tiles cannot be guaranteed, because the thermodynamics and kinetics of DNA tile self-assembly results in occasional erroneous assembly steps. The number of assembly errors increases with the number of tile types, and accruing errors render large scale complex computation practically infeasible.

The first quantitative analysis on the error rate of DNA tile self-assembly was given by Winfree (1998). He developed a kinetic model of DNA tile self-assembly, and simulated the tile assembly process under various conditions. The simulations suggested that by optimally decreasing the temperature and tile concentrations, the error rate ϵ decreases without bound, where the error rate ϵ represents the percentage of tiles in the assembly which have at least one mismatched sticky end. As a tradeoff, the growth speed r is slowed by a factor of ϵ^2. This can provide very low error rates, but only at a large cost in growth speed. Therefore, we are interested in error reduction techniques that have an improved tradeoff between error rate and growth speed.

Assembly errors can be classified into three types: "growth errors," "facet errors," and "nucleation errors." Growth and facet errors are the errors that occur on the growth front of an existing assembly, while nucleation errors deal with the spurious initiation of assemblies. A growth error occurs when a DNA tile with one or more mismatched sticky ends is embedded in the assembly. A facet error occurs when two DNA tiles attach on a growth front (facet) side by side, and thus stabilize each other's binding. This is considered an error because the identity of these tiles may not be correct with respect to the computation being performed. Nucleation errors are similar to facet errors in that a number of tiles

spontaneously assemble a cluster by stabilizing each other's binding. This then seeds ordinary but meaningless computation. Growth errors, facet errors, and nucleation errors all occur because tiles that make only weak contacts with the assembly—and must fall off in order for correct growth to proceed—might first be stabilized by the arrival of another tile that secures it in place.

In order to suppress different types of errors, several methods have been proposed based on the idea of increasing the amount of time required to lock in erroneously assembled tiles (Reif 1999; Chen et al. 2004; Winfree and Bekbolatov 2004; Chen and Goel 2005; Reif et al. 2005; Baryshnikov et al. 2006; Sahu and Reif 2006). The proofreading tile model (Winfree and Bekbolatov 2004) The proofreading tile model (Chen and Goel 2005) suppresses growth errors by replacing each original tile with a set of $n \times n$ different sub-tiles. The snaked proofreading tile model (Chen and Goel 2005). extended the proofreading tile model by restricting the order of association of the sub-tiles, suppressing facet errors in addition to growth errors (Chen et al. 2007). However, they require many different sub-tiles and thus more orthogonal sticky end sequences, which complicates the sequence design process and increases the size of the tile set and the size of the assembled object by a factor of n^2. The error resilient tile model (Reif et al. 2005). avoids the scale-up of the assembled object by using a larger tile set with longer and more information-intensive sticky ends, and performs comparably to the 2×2 proofreading tile model (also only suppressing growth errors). A related tile set modification for reducing nucleation errors has been proposed and demonstrated by Schulman and Winfree (2005, 2007).

In this article, we propose two mechanisms for error suppression based on the idea of decreasing the amount of time a mismatched stays bound to the assembly: the Protected Tile Mechanism (PTM) and the Layered Tile Mechanism (LTM). When kinetically simulated, both mechanisms can suppress all three error types, while making the tiles each only slightly larger. In these mechanisms, we alter the implementation of the DNA tiles by introducing new structural motifs called "protection strands" and "protection tiles." These molecules cover (protect) the exposed sticky ends of the DNA tile to minimize spurious interactions involving other tiles. Tiles on the growth front of proper assemblies become de-protected, and thus correct growth proceeds at a much accelerated rate as compared to spurious growth. Our proposals can be seen as a 2-D generalization of the hybridization chain reaction (Dirks and Pierce 2004), along the general lines suggested by Turberfield et al. (2000).

In Sect. 2, we review the principles of DNA tile-based assembly and assembly errors. In Sect. 3, we show the design and function of the PTM and the LTM. In Sect. 4, we evaluate the performance of these mechanisms by simulation results. We conclude our article in Sect. 5.

2 The tile assembly mechanism

We refer to the DNA tile assembly mechanism originally presented in (Winfree, 1998) as the Original Tile Mechanism (OTM). There are two models of the behavior of OTM: the abstract Tile Assembly Model (aTAM) and the kinetic Tile Assembly Model (kTAM).

The aTAM is based on Wang's tiling problem (Wang 1961, 1963). Wang's tiling problem uses a finite set of square tiles with labeled sides that determine binding between tiles. In the aTAM, each label also possesses a predetermined strength. The strength b of association for each tile is the sum of the strengths of matching labels between the tile and the assembly. Tiling in this model proceeds as follows: A new tile can be added to a "seed structure" (composed of one or several tiles) only if b is greater than or equal to the value of a threshold τ. The growth process is asynchronous; any tile can associate with or

dissociate from the assembly at any time. Only certain tile sets with properly designed labels are capable of producing a uniquely defined structure. For example, regardless of the order in which valid tile additions occur, the Sierpinski tile set forms an assembly with a triangular fractal pattern. This set includes four "rule tiles" and a "preformed border structure" (seed structure) shown in Fig. 1a. The rule tiles are used to grow the Sierpinski pattern. Each tile has specific input and output labels and implements the exclusive-or (XOR) function. White semicircles represent '0', while the black semicircles represent '1'. In this construction, all labels possess strength 1. The preformed border structure is used to provide initial conditions for growth and has initial binding sites that accept free rule tiles. As the correct tiles bind to the assembly, new properly oriented binding sites are created for more tiles. As illustrated in Fig. 1b, at $\tau = 2$, input edges of each tile uniquely attach to output edges of a growth front only if two input labels of the tile match. Consequently, a desired fractal pattern is formed at $\tau = 2$. At $\tau = 1$, the pattern cannot be reliably formed because uniqueness of tile association and binding site orientation does not hold.

In order to describe the process of DNA tile self-assembly, we use the kTAM (Winfree 1998). This model was developed to provide a more realistic treatment of the reversible self-assembly. In the kTAM, a monomer tile can be added to the assembly with some association (forward) rate, or removed from the assembly with some dissociation (reverse) rate.[1] These rates are denoted by r_f and $r_{r,b}$, respectively. At every available site, every possible monomer tile can associate to the assembly, regardless of whether the monomer is correct or not. The forward rate depends only on the monomer tile concentration, $[monomer]$:

$$r_f = k_f[monomer] = k_f e^{-G_{mc}},$$

and $G_{mc} > 0$ is the non-dimensional entropic cost of associating to the growth front. For simplicity, we assume that tile concentrations are kept constant at $[monomer] = e^{-G_{mc}}$. Since the forward rate constant k_f is a constant, forward rate r_f is also constant. The reverse rate depends exponentially on b, the binding strength of the tile to the assembly:

$$r_{r,b} = k_{r,b} = k_f e^{-bG_{se}}.$$

$G_{se} > 0$ is the non-dimensional free energy cost of breaking a bond. The ratio of $\frac{G_{mc}}{G_{se}}$ corresponds to the threshold τ used in the aTAM; if $b < \tau$, then a tile falls off faster than it attaches.

Wang's tiling problem has many possible physical implementations. Following Winfree (1996), we consider DNA tiles as an implementations of the problem. We focus on a specific type of DNA tile called the double-crossover (DX) molecule (Fu and Seeman 1993; Winfree et al. 1998). Each DX molecule is comprised of four or five DNA oligomers depending on the design of the tile. Figure 2a illustrates a DX molecule made of four DNA oligomers. Structurally, it is two double-stranded helices cross-linked at two crossover junctions. The ends of each double helix remain single-stranded, and are known as sticky ends. The sticky ends correspond to the labels of each rule tile, and a pair of sticky ends matches when their sequences are complementary. Figure 2b shows a process of DNA tile self-assembly via hybridization of complementary sticky ends.

For DX molecules, we assume a forward rate constant of $k_f = 10^6$/M/s.[2] The length of the sticky ends correspond to the bonding strength at each edge. The reverse rate depends

[1] We consider only association and dissociation reactions between monomer tiles and the assembly. Interactions between assemblies are not considered.

[2] The value of k_f has been experimentally measured for hybridization between DNA oligomers and is sequence dependent (Wetmur 1991). It has not been experimentally measured for DNA tiles yet.

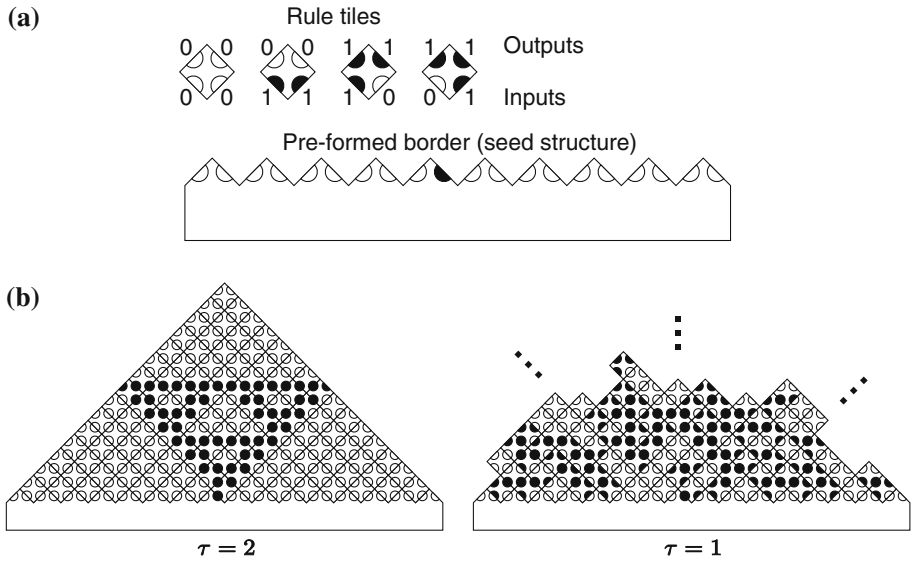


Fig. 1 (**a**) Sierpinski tile set. Four rule tiles represent the XOR function ($0 \oplus 0 = 0, 1 \oplus 1 = 0, 0 \oplus 1 = 1$ and $1 \oplus 0 = 1$). The preformed border structure provides the initial conditions for growth. (**b**) Sample growths of the Sierpinski tile set at $\tau = 2$ and 1. Tiles may not be rotated

on b, which is expressed as the number of matching sticky ends. Rates of association and dissociation for various tiles on the growth front are shown in Fig. 3.

The parameter G_{se} considers both entropy and enthalpy, and thus decreases as the temperature rises.[3] The parameters G_{mc} and G_{se} are determined by the physical conditions of the environment. Large G_{mc} means that the DNA tiles are at low concentrations in solution, while small G_{mc} denotes high concentrations. Large G_{se} denotes low temperature or longer sticky ends on the DNA tiles.

Figure 4 shows simulated assembly growth seeded by the preformed border. When τ is small, DNA tiles assemble regardless of binding rules. Tiles grow quickly, but no coherent assembly is formed (Fig. 4a). As τ becomes gradually larger, assemblies' growth becomes slower. The patterns on the assembly are mostly correct, but it contains some mismatched tiles (Fig. 4b). When τ approaches 2, tile addition is even slower, and fewer mismatched tiles are observed (Fig. 4c).[4]

As mentioned before, assembly errors can essentially be classified into three types: "growth errors," "facet errors," and "nucleation errors." Figure 5a shows a typical

[3] G_{se} depends on temperature T and sticky end length s, where it is approximated by $G_{se} = (\frac{4000}{T} - 11)s$ (Winfree 1998). We neglect the initiation entropy for hybridization, as in Winfree and Bekbolatov (2004).

[4] In reality, it is difficult to achieve such near perfect assemblies that the simulation suggests. This discrepancy could be due to overly simplified assumptions of the kinetic model:

(1) The concentrations of all the monomer tiles are not kept constant through the entire duration of the assembly process, and also are not in perfect stoichiometry with one another.
(2) The binding strengths of all the different sticky ends are not identical.
(3) The macroscopic interaction between separate assemblies is ignored, when in reality assemblies join and/or hinder each other's growth. In addition, there are technical problems such as (a) the nucleating scaffold is so floppy that makes many defects in the assembly, and (b) synthesis of DNA oligonucleotide is not perfect.

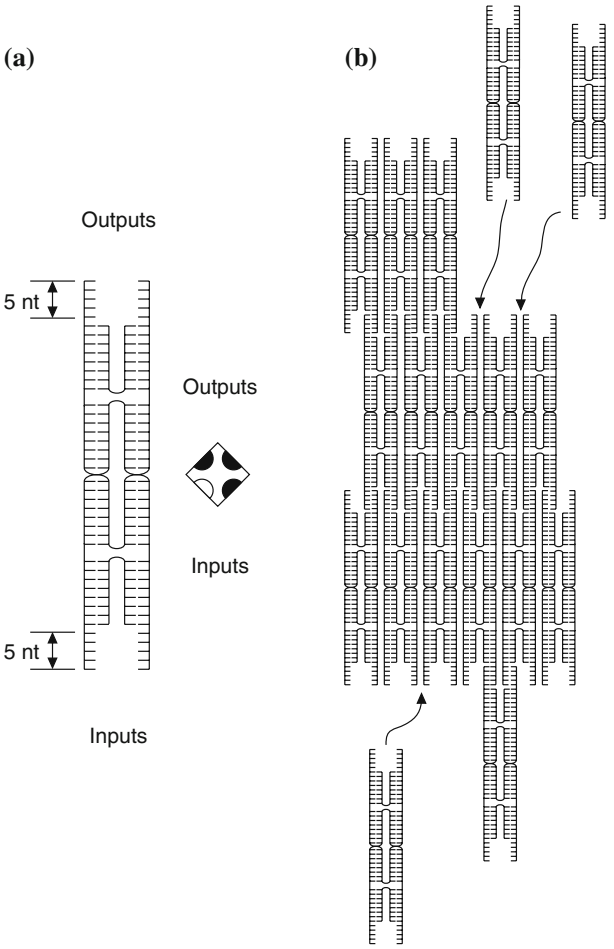


Fig. 2 (**a**) A DNA tile (DX molecule) and its corresponding abstract tile (a diamond). (**b**) Self-assembling of DNA tiles

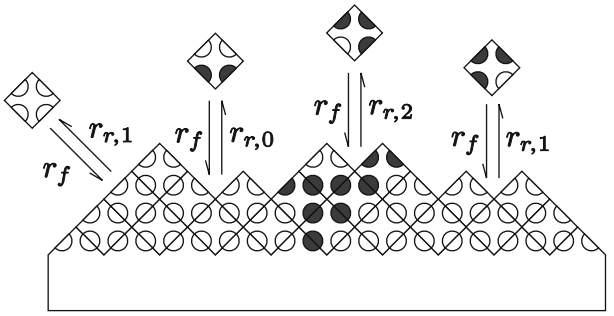


Fig. 3 Rates for tile addition and tile dissociation in the kinetic model. Forward rates r_f have the same value at any sites, but reverse rates $r_{r,b}$ depends on total bonding strength b of each tile

instance of a growth error. A DNA tile with at least one mismatched sticky end is embedded in the assembly temporarily. These tiles usually dissociate from the growth front very quickly, but occasionally another tile will correctly associate to an adjacent position before the first tile dissociates. When this happens, both the mismatched tile and the next tile are bound to the assembly with strength $b = 2$. Practically, this means that the mismatched tile is locked into the assembly and cannot rapidly dissociate from it. As a result,

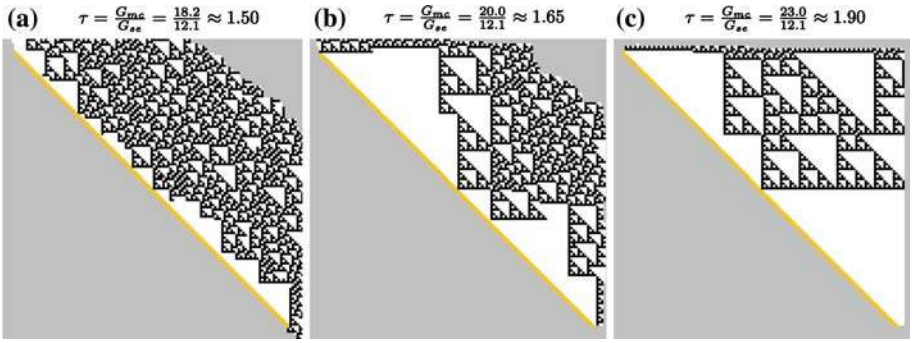


Fig. 4 Simulation results of growth in the OTM. These results were computed on a 128 × 128 array. A preformed border is given as the initial seed structure (orange). We assume that tiles of this border are fixed and not subject to dissociation. Starting from the border, the assembly grows to the upper right forming a triangular region. The numbers of erroneous tiles were (**a**) 109 tiles, (**b**) 22 tiles, and (**c**) 6 tiles. Simulated duration (in simulated seconds) to complete these assemblies was (**a**) about 1.3×10^4 s, (**b**) about 8.8×10^4 s, and (**c**) about 6.9×10^6 s. These are rotated 45° clockwise from Fig. 1

the assembly pattern of tiles in the shading area is locally correct, but incorrect with respect to the original seeded computation. Facet errors occur on a flat surface (facet) of the assembly as illustrated in Fig. 5b. Normally during assembly growth, each tile can only attach to concave sites on the assembly because only here it is possible to bind with strength $b = 2$. However, occasionally a single tile will temporarily attach to a facet, and then be locked in by the locally correct addition of another tile. Then both tiles are attached with strength $b = 2$, and will not dissociate rapidly. A new layer on the facet begins to grow from the pair, and it eventually causes inconsistency with the pattern grown from the seed if the tile was incorrect. Nucleation errors denote when tiles spontaneously assemble without a preformed border, as illustrated in Fig. 5c. The error begins when several monomers attach to form a small cluster. Under supersaturated conditions, i.e., when $\tau < 2$, there is a threshold cluster size above which clusters are more likely to grow than to shrink (Schulman and Winfree 2005, 2007). The threshold size becomes larger as τ approaches 2. Nucleation errors deplete the concentration of DNA tile monomers and also decrease the fractional yield of the correct assemblies.

Figure 6 shows simulation results for varying G_{mc} and G_{se} (Winfree 1998; Winfree and Bekbolatov 2004). These diagrams show sizes of assembly and numbers of errors evaluated by simulations. Figure 6a shows the conditions of growth when seeded by a preformed border. The tiles form good assemblies at $\tau = 2-\delta$ for small δ and large G_{se}. In the region of $\tau > 2$, assemblies do not grow at all because forward rates are smaller than reverse rates. Figure 6b shows the conditions of growth when seeded by a single rule tile. This plot shows that nucleation errors occur when monomer concentrations are high (i.e., when G_{mc} is low), while it hardly occurs near the optimal condition (i.e., $\tau = 2-\delta$ and large G_{mc}).

3 Error suppression mechanisms with new motifs

Here, we propose two mechanisms for error-correction, which we call the PTM and the LTM. Both suppress all three mentioned types of errors. Controlling hybridization of the sticky ends is the crux of the function of these mechanisms. We introduce two new motifs called protection strands and protection tiles, which respectively function with the PTM

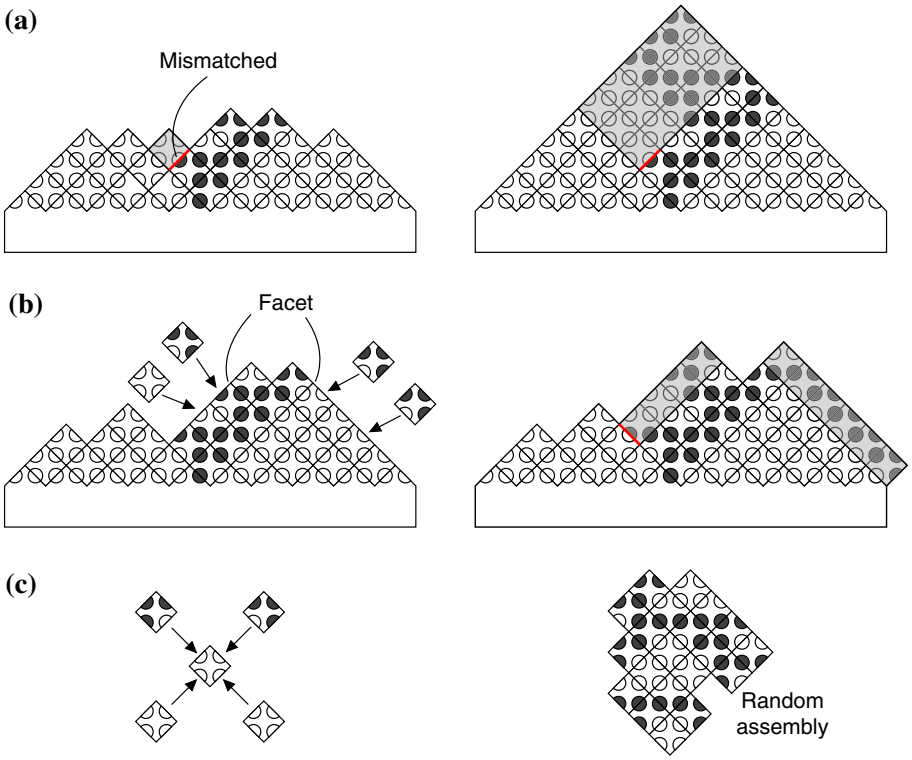


Fig. 5 Growth processes of three types of errors: (**a**) growth errors, (**b**) facet errors, and (**c**) nucleation errors. Red lines indicate mismatched sides. In each case, the initial assembly error is shown on the left, while the result after subsequent perfect $\tau = 2$ growth is shown on the right; except in (**c**) where growth of the random assembly results from additional insufficient attachment errors as well

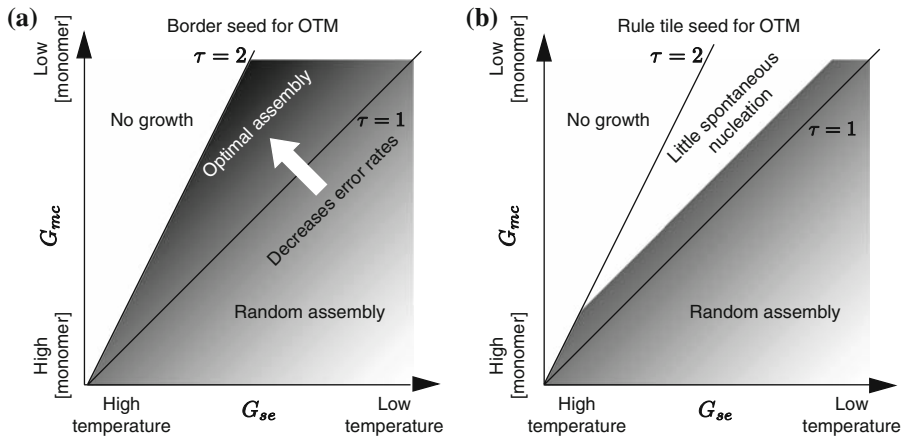


Fig. 6 Phase diagrams for the OTM of (**a**) assemblies seeded by the preformed border and (**b**) assemblies seeded by a rule tile, i.e., spontaneous nucleation. Lines with slopes 1 and 2 represent $\tau = \frac{G_{mc}}{G_{se}} = 1$ and 2. Darker shading denotes lower error rates of assemblies

and the LTM. These motifs, added to the original tile set, aim to reduce unnecessary association of monomer tiles by covering exposed sticky ends. In the PTM, a protection strand covers two exposed sticky ends at the input side of each DNA tile, while in the LTM, a protection tile covers both input and output sticky ends. The LTM was originally proposed by Fujibayashi and Murata (2005) but the analysis here is more fully developed.

Both the PTM and the LTM make use of a dynamic process known as 3-strand branch migration as part of their mechanism for detecting and preventing erroneous assembly steps. Branch migration takes place when two strands with identical subsequences, say A and B, are each partially bound to the same Watson–Crick complementary strand, C. The branch point is the junction between the A:C duplex region and the B:C duplex region. By pairing more with A and less with C, or visa versa, the branch point can move easily without any significant change in energy. Single base-pair steps take on the order of 10 μs, so in the case where one strand has a longer subsequence complementary to C, this strand will quickly and selectively replace the other strand (Panyutin and Hsieh 1994; Biswas et al. 1998; Reynaldo et al. 2000). Such strand displacement reactions can be engineered by designing a single-stranded region on one strand, called a toehold, that serves as a binding domain to initiate branch migration. The speed of the overall strand displacement reaction is strongly dependent on the length of the toehold, with a maximum around 6 nt, and depends quadratically on the length of the branch migration region, although for short domains (less than 50 bp) this time is on the order of seconds and is therefore usually negligible (Yurke et al. 2000; Yurke and Mills Jr. 2003). In the following, branch migration and strand displacement will be used to control the removal of protection strands and protection tiles.

3.1 The Protected Tile Mechanism

We introduce the PTM to reduce the errors in self-assembly. The key feature of the PTM is that each tile has a "protection strand" covering most of the input sticky ends, and that this protection strand is likely to be removed only if *both* inputs correctly match. Otherwise the protection strands prevent the erroneous binding of the sticky ends.

Figure 7a shows one instance of a DNA tile used in the PTM. The protection strand is a single oligomer that *covers* the input side of the tile. Each sticky end remains uncovered and it works as the toehold for initiating a branch migration process that removes the protection strand. The combination of a tile and a protection strand is called a "protected tile," or just "tile" when it is clear from context, and "foundation tile" refers to the unprotected foundation tile. The output sides of all tiles are unprotected, thus the growth front always displays unprotected sticky ends.

The length of the sticky ends is chosen in consideration of thermodynamics and geometrical constraints for the DX molecule design.[5] In our design, the DNA tile has four sticky ends with 14 nt. (Hereafter, 14 nt is regarded as unit bonding strength 1.) The protection strand is a single 22-nt oligomer that covers most of each sticky end on the input side of the tile, leaving 3 nt of each sticky end uncovered. Our assumption is that a newly arrived tile will bind to both 3 nt toeholds simultaneously, acting roughly equivalent to a

[5] In the OTM, the length of the sticky ends is 5 nt. If we follow this design, the length of the protection strands must be shorter than 10 nt. However, this length is not enough to prevent the protection strand from spontaneously dissociating from the foundation tile even at room temperature. This indicates that we need longer sticky ends. We also have to consider the geometrical constraints of maintaining the planarity of the assembly made of DNA tiles. Both the intra- and inter-molecular spacing of the crossover points must be an integer number of half-turns of the B-form DNA double helix. The 14-nt sticky ends we have chosen here satisfies this constraint.

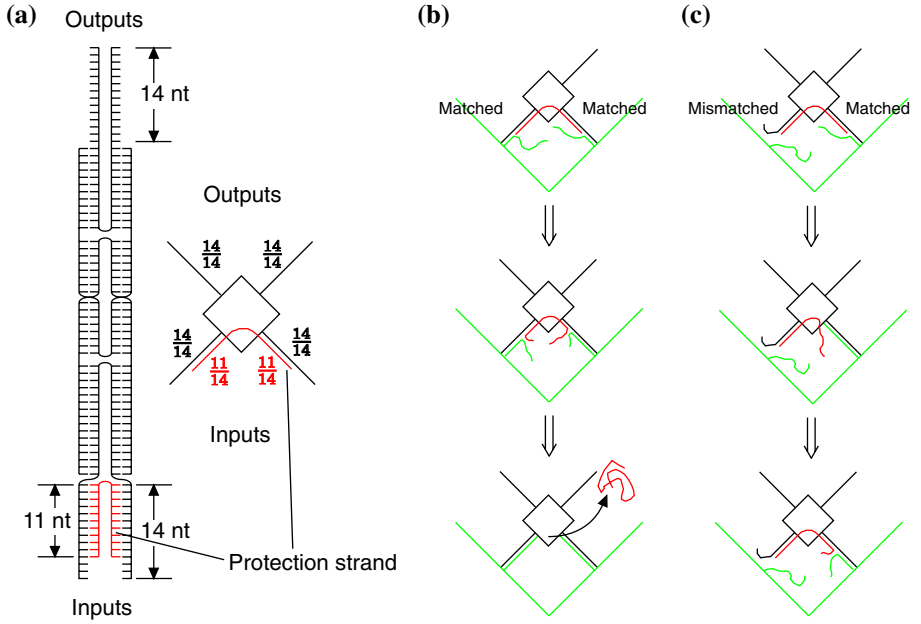


Fig. 7 (**a**) Schematic of the protected tile in the PTM. The protection strand (red) covers two input strands of a foundation tile. Growth processes in cases of a tile matching (**b**) input sticky ends (T_{11} to T_{00}* in Fig. 8a) and (**c**) one input sticky end ($T_{11} \rightleftharpoons T_{10}$ in Fig. 8b). Only the correct tile can release the protection strand and lock in the attachment to the growth front

single 6 nt toehold, which has been shown to accelerate toehold-mediated branch migration and strand displacement reactions to nearly the same rate as the hybridization of unprotected single-stranded DNA (Yurke and Mills Jr. 2003). Using a longer toehold would be expected to result in faster growth rates but higher error rates. Figure 7b,c shows the intended growth process in the PTM. The protected tiles associate to the growth front by the exposed 3-nt sticky ends first, then branch migration results in strand displacement on each of the matching input arms; if both arms are matched, the protection strand is completely displaced, and it dissociates. In the case of a tile matching by two input sticky ends as illustrated in Fig. 7b, both input sticky ends attach to their respective complements on the growth front with strength $\frac{3}{14}$. Branch migration takes place on both input sticky ends and eventually the protection strand is displaced, leaving the tile bound with strength 2. If a tile has one matched (e.g., right) input and one mismatched (e.g., left) input as illustrated in Fig. 7c, the matched sticky end can undergo branch migration, but not the mismatched one. For this partially mismatched tile to dissociate quickly from the growth front, we require $\tau > \frac{6}{14}$, so that breaking the total 6 bp of toehold is energetically favored. On the other hand, we need $\tau < \frac{11}{14}$ so that it is energetically favorable for the half-displaced protection strand to remain attached to the foundation tile. Finding the optimal conditions for achieving the lowest error rates and fastest assembly will be explored by simulation.

Figure 8 shows the hybridization state Tile Assembly Model (hsTAM), which is a continuous-time Markov process for self-assembly that models intermediate states of a tile associating to the assembly. This transition diagram is applied at each site on the growth front. It has three cases: a tile matching by (a) two input sticky ends, (b) one input sticky end, and (c) neither input sticky end. In the hsTAM, a (protected or foundation) monomer

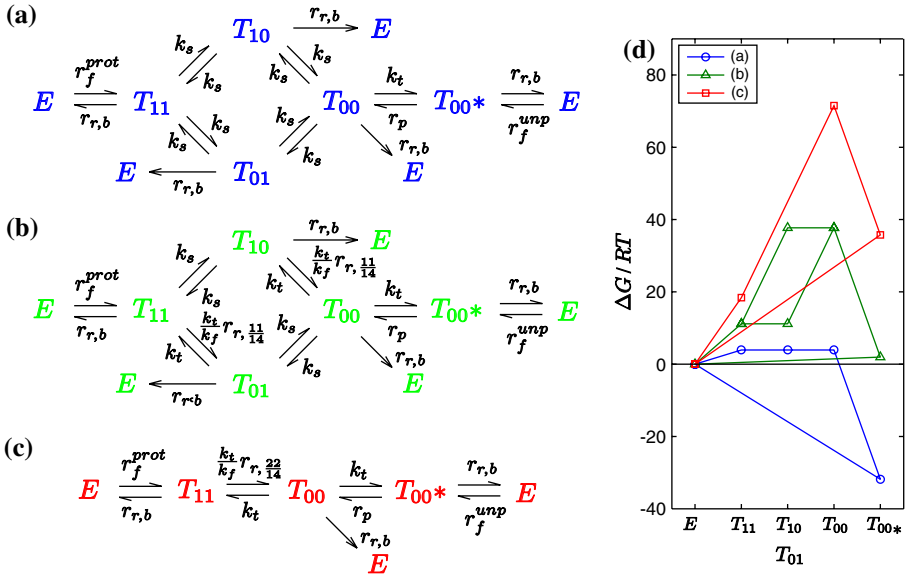


Fig. 8 Markov process for the PTM in cases of a tile matching by (**a**) two inputs, (**b**) one input (e.g., the right input), and (**c**) neither input. Two diagrams have six states (E, T_{11}, T_{10}, T_{01}, T_{00} and T_{00}*). Subscript of T denotes whether the protection strand is detached (0) or attached (1) to the foundation tile. (**d**) The relative free energies of the substate at each case when $\tau = G_{mc}/G_{se} = 18.4/33.8 \approx 0.54$

tile can have different substates. To represent them, we have introduced two kinds of states: In E, the site is empty; the tile is not yet localized to the site. In $T = \{T_{11}, T_{10}, T_{01}, T_{00}, T_{00}^*\}$, the tile is located at the site. T_{11} denotes when the protected tile is physically at the site, but both its inputs are covered by the protection strand. It is bound only by the 3-nt sticky ends. In T_{10}, the left half of the protection strand covers the left input of the foundation tile, and similarly for T_{01} and the right half. In T_{00} the protection strand detaches from the tile, but is still in the vicinity of the tile. It still has some chance of re-association in this state. Finally in T_{00}^*, the protection strand is fully released from the tile.

When a tile associates with the assembly, it may or may not have a protection strand depending on the environment. When $\tau < \frac{22}{14}$, most associating tiles have the protection strand attached ($E \rightarrow T_{11}$). This is the desired process for error suppression to work. On the other hand, when $\tau > \frac{22}{14}$, it associates with an assembly as a foundation tile because the dissociation of the 22-nt protection strand is thermodynamically favored ($E \rightarrow T_{00}^*$). Both association rates of the *protected* tile and *unprotected* (foundation) tile to the crystal depend on their respective concentrations; $r_f^{prot} = k_f[FP]$, and $r_f^{unp} = k_f[F]$, where F and P are the foundation tile and the protection strand, and FP is the protected tile. Concentrations $[F]$ and $[FP]$ can be estimated as follows. First, the reaction to form either structure can be written:

$$F + P \rightleftharpoons FP.$$

As mentioned before, we assume that for each logical tile species, the total monomer concentration has the same value and is constant.[6] The constant should be the total of the

[6] In practice, the concentration $[P]$ of released protection strands or tiles increases while assemblies are formed, and it may affect the growth of the assemblies. This can be solved by external control of concentrations, for instance, by using microfluidic devices (Somei et al. 2006).

concentration of the foundation tile (or equivalently, the protection strand) and the protected tile:

$$[monomer] = e^{-G_{mc}} = [F] + [FP] = [P] + [FP] \qquad ([F] = [P]).$$

Concentrations $[F]$, $[P]$ and $[FP]$ can be calculated via the equilibrium constant:

$$K = \frac{[F][P]}{[FP]} = \frac{[F]^2}{e^{-G_{mc}} - [F]} = e^{\Delta G/RT} = e^{-\frac{22}{14}G_{se}}, \tag{1}$$

where $R = 2$ cal/mol/K and ΔG (in cal/mol) is the difference in free energy between the tile with and without the protection strand. Thus, we can solve for $[F]$ and $[FP]$.

Each of five substates (T_{11}, T_{10}, T_{01}, T_{00} and T_{00}*) has a specific rate $r_{r,b}$ to spontaneously revert to the empty state E. Each substate possesses a different value of $r_{r,b}$ depending on the strength b of association between the tile and the growth front of the assembly. For example, in case of a tile matching by two inputs, the rates $r_{r,b}$ from T_{11}, T_{10}, T_{01}, T_{00} and T_{00}* to E are $r_{r,\frac{6}{14}}$, $r_{r,\frac{17}{14}}$, $r_{r,\frac{17}{14}}$, $r_{r,\frac{28}{14}}$ and $r_{r,\frac{28}{14}}$ respectively.[7]

In the analysis of the PTM, we need the values of rates of intra-molecular reactions, namely, reactions between the foundation tile and the protection strand. Assume a protected tile associates with the crystal on the growth front. Since the foundation tile and protection strand are colocalized within the size of a tile ($\sim$10 nm), hybridization between the protection tile and the foundation tile can be regarded as an intra-molecular reaction. The rate of this reaction is proportional to the "effective concentration," which is estimated as $C_{eff} = \frac{1\,\text{molecule}}{1{,}000\,\text{nm}^3} \approx 6 \times 10^{-2}$ M. We apply the rate $k_t = k_f C_{eff} \approx 6 \times 10^4$/s to these colocalized reactions among T_{11}, T_{10}, T_{01} and T_{00}. In the case of a tile matching by two inputs, branch migration starts at either the left or right input of the foundation tile. Namely, half of the protection strand is detached (e.g., $T_{11} \rightleftharpoons T_{10}$). Then, branch migration continues on the other side ($T_{10} \rightleftharpoons T_{00}$).[8] Branch migration is a random process and has no preference for direction; therefore we apply the same rate constant $k_s = 10^3$/s in both directions (Panyutin et al. 1995).[9] In the case of a tile matching by one input (e.g., the left input of the protected tile does not match), the protection strand is removed by branch migration on the right, but sticks to the tile on the left strand ($T_{11} \rightarrow T_{10}$). When the state of the tile is T_{10}, it has two possibilities of transition. Most likely, branch migration reverses the tile to its initial state with rate k_s ($T_{10} \rightarrow T_{11}$). Occasionally though, the 11 nt of the left side of the protection strand spontaneously dissociates from the tile with rate

[7] There are 2^4 situations of possible neighborhoods. All of such possibilities are simulated. b is therefore the sum of attachment strengths on all four sticky ends. The cases where tiles are attached on the output is discussed later in this section.

[8] Although the branch migration may occur at both ends of the protection strand simultaneously, for simplicity we have assumed that they occur sequentially, which may give unrealistic transition rates.

[9] k_s is essentially treated as a constant value. However at large G_{mc} or G_{se}, this results in a dramatic and unnecessary slowdown in the simulation. This is because all rates except k_s become extremely small (e.g., $T_{11} \rightleftharpoons T_{10}$ in Fig. 8b) and therefore the majority of steps are spent redundantly simulating branch migration that has already reached equilibrium. To avoid this, we use the smallest value of k_s that ensures rapid equilibration on the timescale of the other reactions:

$$k_s = \begin{cases} min(10^3, 10 \times max(r_f^{prot}, r_f^{unp}, r_{r,b})) & \text{(a tile matching by two inputs)} \\ min(10^3, 10 \times max(r_f^{prot}, r_f^{unp}, r_{r,b}, \frac{k_t}{k_f} r_{r,\frac{11}{14}})) & \text{(a tile matching by one input)} \end{cases}$$

Since the strand replacement by the branch migration is isoenergetic, this modification is effective in accelerating the simulation with the guarantee that internal substate probabilities achieve the same pseudo-equilibrium probabilities as they would with the faster rates.

$\frac{k_t}{k_f} r_{r,\frac{11}{14}} = k_t e^{-\frac{11}{14}G_{se}}$ ($T_{10} \rightarrow T_{00}$). In the case of a tile matching by neither input, most likely, $T_{11} \rightarrow E$ or $T_{00}^* \rightarrow E$, but the protection strand spontaneously dissociates from the tile with rate $\frac{k_t}{k_f} r_{r,\frac{22}{14}} = k_t e^{-\frac{22}{14}G_{se}}$. The transition $T_{00} \rightleftharpoons T_{00}*$ denotes the removal of the protection strand from the vicinity of the foundation tile and its reverse reaction. Since T_{00}^* denotes a bimolecular state, the forward rate k_t ($=k_f C_{eff}$) among these states depends on the effective concentration ($T_{00} \rightarrow T_{00}^*$), while the reverse rate r_p ($=k_t\ [P] = k_f C_{eff}\ [P]$) depends on the product of the effective concentration and the concentration of the protection strand ($T_{00}^* \rightarrow T_{00}$). One concern is the case when another tile attaches on the output of the foundation tile before the protection strand is released (i.e., T_{11}, T_{10}, T_{01} and T_{00}). In such a case, b in $r_{r,b}$ must be changed. However, this does not affect the reaction between the protection strand and the foundation tile, since the protection strand has nothing to do with the output side of the tile.

We can assume that all the elementary reactions in the diagram satisfy the detailed balance (Schulman and Winfree 2005),[10] and thus we can calculate the relative free energies of each state as shown in Fig. 8d by the principle of detailed balance. For instance, along the reaction path $E \rightleftharpoons T_{11} \rightleftharpoons T_{10} \rightleftharpoons T_{00}$ in Fig. 8a, assuming no tile is bound on the output side,

$$\frac{E}{T_{00}} = \frac{E}{T_{11}}\frac{T_{11}}{T_{10}}\frac{T_{10}}{T_{00}} = \frac{r_{r,\frac{6}{14}}}{r_f^{prot}}\frac{k_s}{k_s}\frac{k_s}{k_s} = e^{-\frac{6}{14}G_{se}-\ln[\mathrm{FP}]}$$

holds. From Eq. 1, we have $\ln[FP] = 2\ln[P] + \frac{22}{14}G_{se}$, and the free energy of T_{00} is calculated as

$$\Delta G/RT = -2G_{se} - 2\ln[P] = -2G_{se} - 2\ln[F].$$

By the similar calculation, we can estimate the free energy of T_{00} in Fig. 8b, c as $-G_{se}-2\ln[F]$ and $-2\ln[F]$, respectively. The binding between the protected tile and the assembly is thermodynamically more favorable than that of the foundation tile and the assembly under typical experiment conditions (e.g., $T = 25°\mathrm{C}$: $G_{se} \approx 33.8$, and $[monomer] = e^{-G_{mc}} \approx 10$ nM: $G_{mc} \approx 18.4$), or equivalently $r_f^{prot} > r_f^{unp}$. This implies that there exists a high kinetic barrier for the incorrect tiles to reach the substate T_{00}^*.

3.2 The Layered Tile Mechanism

The purpose of the LTM is to prevent spurious interactions of exposed sticky ends not only at the inputs but also at the outputs. In the LTM, all four sticky ends are protected by the "protection tile." Hereafter, the combination of the "foundation (lower) tile" and "protection (upper) tile" is called a "layered tile" (Fig. 9a). When a layered tile attaches to the growth front, its input sticky ends are displaced by the same process as in the PTM. Figure 9b and c shows the growth processes of the LTM under correct tile addition and incorrect tile addition. In case of a tile matching by two inputs as illustrated in Fig. 9b, both sticky ends of the protection tile are replaced by the sticky ends of the growth front. The protection tile then is only bound with the short length of the output sticky ends, and it will be thermodynamically favorable to spontaneously dissociate, leaving only the foundation tile at the site to allow further association. Figure 9c shows the case of a tile matching by the right input. In this case, the total bonding strength of the protection tile is not high

[10] Reactions $T_{10} \rightarrow E$, $T_{01} \rightarrow E$ and $T_{00} \rightarrow E$ are one-directional reactions, and thus they do not comply strictly with the detailed balance. We neglected them here for the simplicity of the analysis.

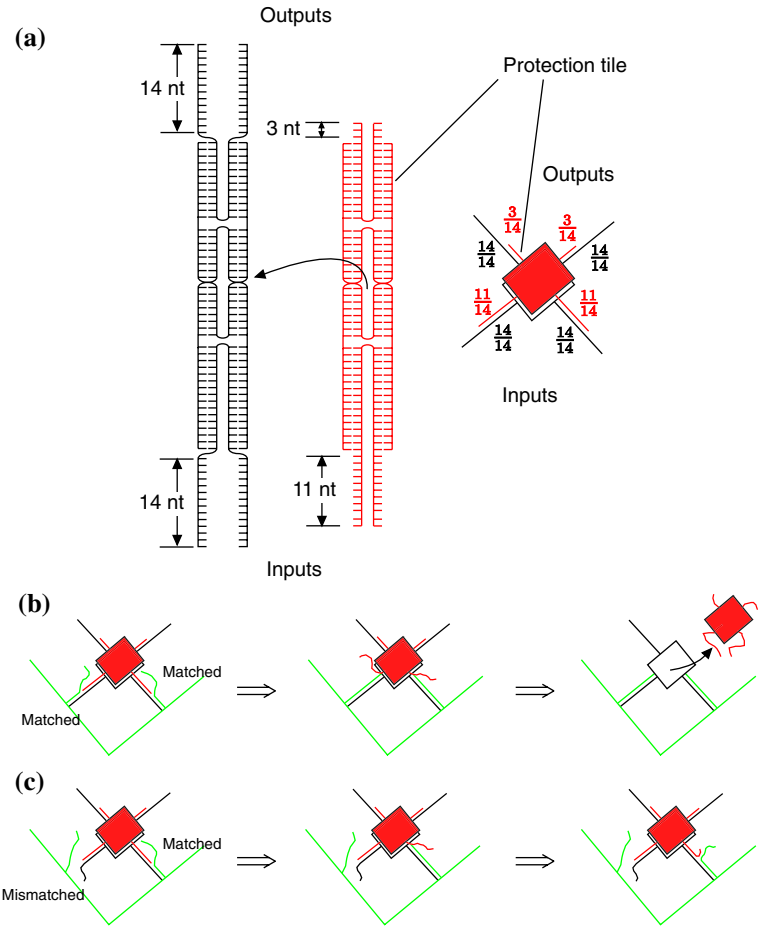


Fig. 9 (**a**) Schematic of the layered tile in the LTM. A red (protection) and a black (foundation) tile comprise a layered tile. The protection tile protects the foundation tile's sticky ends at both input and output. Growth processes in cases of a tile matching (**b**) both input sticky ends (T_{11}^{11} to $T_{00}^{00}*$ in Fig. 10a), and (**c**) one input sticky end ($T_{11}^{11} \rightleftharpoons T_{10}^{11}$ in Fig. 10b). The protection tile intercepts the association of incoming tiles. It is removed only when both input sticky ends of the foundation tile match

enough to rapidly release the protection tile. As a result, the branch migration of the matched sticky end reverses itself, and the layered tile is released.

One implementation of the protection tile is another DX molecule that rests on top of the lower tile. We designed the molecules for the LTM based on the same foundation tile used in the PTM. By following similar considerations as for the PTM, we chose the length of covered sticky ends to be 11 nt for the input sticky ends and 3 nt for the output sticky ends as shown in Fig. 9a.

These lengths of the sticky ends satisfy the following: The sum of the input protected bases and the output protected bases should not have strength > 1. In case of a tile matching by two inputs, after branch migration the protection tile is bound with strength $\frac{6}{14}$, and it will be energetically favorable to dissociate at $\tau > \frac{6}{14}$, leaving the foundation tile

at the site. In case of a tile matching by the right input, the total bonding strength of the protection tile, b, is $\frac{17}{14}$, which makes releasing it unfavorable at $\tau < \frac{17}{14}$. Thus, we expect simulations to show optimal performance somewhere in this range.

This implementation requires further considerations, for instance, we have to take the 3D geometrical constraints with molecular structure of the DX molecule into account. Other implementations of protection tiles are possible, but they are not discussed further in this article.

Next, we show the hsTAM for the LTM in Fig. 10. This diagram is derived in the same manner as the PTM. There are six substates in this diagram: $T = \{T^{11}_{11}, T^{11}_{10}, T^{11}_{01}, T^{11}_{00}, T^{00}_{00}, T^{00}_{00}*\}$. These substates represent the intermediate states of local reactions between the foundation tile and the protection tile. Subscripts denote coverage state of input sticky ends by the protection tile (1: covered, 0: uncovered), while superscripts denote that of output sticky ends. T^{00}_{00} denotes the state where the protection tile is about to fall off, but it is in the vicinity of the foundation tile, thus still has some chance of reconnection. In $T^{00}_{00}*$, the released protection tile is no longer colocalized and is unlikely to return to the growth front.

The diagram is also classified by cases of a tile matching by (a) two input sticky ends, (b) one input sticky end, and (c) neither input sticky end as in that of the PTM. Rates of $E \rightarrow T^{11}_{11}$ and $E \rightarrow T^{00}_{00}*$ depends on concentrations of the layered tile and the foundation tile: $r^{prot}_f = k_f[FP]$ and $r^{unp}_f = k_f[F]$, where F and P are the foundation tile and the protection tile, and FP is the layered tile. Concentrations $[F]$, $[P]$, and $[FP]$ can be calculated similarly as for the PTM via the equilibrium constant:

$$K = \frac{[F][P]}{[FP]} = e^{\Delta G/RT} = e^{-\frac{28}{14}G_{se}}, \tag{2}$$

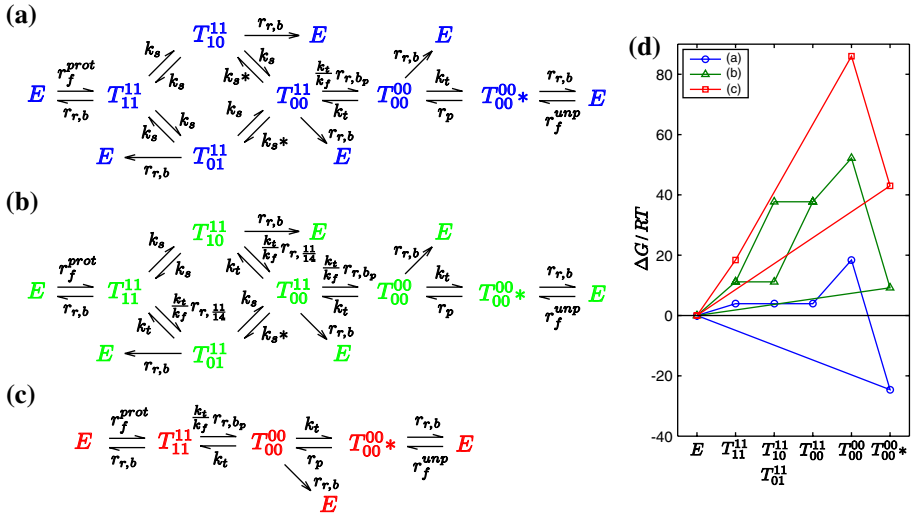


Fig. 10 Markov process for the LTM in cases of a tile matching by (**a**) two inputs, (**b**) one input (e.g., the right input), and (**c**) neither input. Each diagram has seven states ($E, T^{11}_{11}, T^{11}_{10}, T^{11}_{01}, T^{11}_{00}, T^{00}_{00}$ and $T^{00}_{00}*$). Subscript and superscript of T represent whether the input and output sticky ends of the protection tile are detached from (0) or attached to (1) the foundation tile. (**d**) The relative free energies of the substate at each case when $\tau = G_{mc}/G_{se} = 18.4/33.8 \approx 0.54$

where ΔG (in cal/mol) is the difference in free energy between the tile with and without the protection tile. Thus, we can solve for $[F]$ and $[FP]$.

The first steps $T_{11}^{11} \rightleftharpoons T_{10}^{11} \rightleftharpoons T_{00}^{11}$, are the same as those of the PTM. Even if both inputs of the protection tile are detached, they have a high probability of returning to the original state E because its output strands are still bound to the lower tile. Thus we assume that the rate of reverse reactions $k_s* = k_s e^{-G_b}$ ($T_{00}^{11} \rightarrow T_{10}^{11}$ and $T_{00}^{11} \rightarrow T_{01}^{11}$). A free parameter $G_b \geq 0$ represents an anti-cooperative free energy, which means that the protection tile *pops half open* and e^{-G_b} is related to closing it. This value depends on the rigidity of the DNA tile and electrostatic repulsion between DNA backbones. Here we simulate under $G_b = 0$; simulations with larger values of G_b yield higher error rates (data not shown); to obtain better performance in this case, it may suffice to re-design the protection tile so that it attaches to the output sticky ends by more than 3 nt, thus offsetting G_b.

The protection tile dissociates from the foundation tile at rate $\frac{k_t}{k_f} r_{r,b_p} = k_t e^{-b_p G_{se}}$. This rate depends on the total bonding strength b_p of sticky ends between the protection tile and the foundation tile. If other tiles with uncovered input sticky ends attach on one or two output sides, the corresponding output sticky ends of the protection tile dissociate, and the foundation tile associates with the other tile. Thus, in case of a tile matching by two inputs and by one input,

$$b_p = \begin{cases} 6/14 & \text{(No tile attaches on output sides.)} \\ 3/14 & \text{(The other tile attaches on either output side.)} \\ 0/14 & \text{(Two other tiles attach on both output sides.)} \end{cases}$$

and in case of a tile matching by neither input, the value of b_p contains strengths of input sticky ends:

$$b_p = \begin{cases} 22/14 + 6/14 & \text{(No tile attaches on output sides.)} \\ 22/14 + 3/14 & \text{(The other tile attaches on either output side.)} \\ 22/14 + 0/14 & \text{(Two other tiles attach on both output sides.)} \end{cases}$$

Finally, the protection tile diffuses away from the assembly position at rate k_t ($T_{00}^{00} \rightarrow T_{00}^{00}*$), and the reverse reaction has rate $r_p = k_t[P] = k_f C_{eff}[P]$ ($T_{00}^{00} * \rightarrow T_{00}^{00}$).

In the same manner as for the PTM, we can calculate the relative free energies of each state in the LTM by the principle of detailed balance as shown in Fig. 10d. For instance, along the reaction path $E \rightleftharpoons T_{11}^{11} \rightleftharpoons T_{10}^{11} \rightleftharpoons T_{00}^{11} \rightleftharpoons T_{00}^{00}$ in Fig. 10a, assuming no tile is bound on the output side,

$$\frac{E}{T_{00}^{00}} = \frac{E}{T_{11}^{11}} \frac{T_{11}^{11}}{T_{10}^{11}} \frac{T_{10}^{11}}{T_{00}^{11}} \frac{T_{00}^{11}}{T_{00}^{00}} = \frac{r_{r,\frac{6}{14}}}{r_f^{prot}} \frac{k_s}{k_s} \frac{k_s*}{k_s} \frac{k_t}{\frac{k_t}{k_f} r_{r,b_p}} = e^{-\ln[\mathrm{FP}]}$$

holds. From Eq. 2, we have $\ln[FP] = 2\ln[P] + \frac{28}{14} G_{se}$, and the free energy of T_{00}^{00} is calculated as

$$\Delta G/RT = -2G_{se} - \ln[P] = -2G_{se} - 2\ln[F].$$

By the similar calculation, we can estimate the free energy of T_{00} in Fig. 10b, c as $-G_{se} - 2\ln[F]$ and $-2\ln[F]$, respectively. The binding between the protected tile and the assembly is thermodynamically more favorable than that of the foundation tile and the assembly under typical experiment conditions (e.g., $T = 25^\circ\mathrm{C} : G_{se} \approx 33.8$, and $[monomer] = e^{-G_{mc}} \approx 10$ nM: $G_{mc} \approx 18.4$), or equivalently $r_f^{prot} > r_f^{unp}$.

4 Simulations

4.1 Simulation details for the PTM and the LTM

We simulated the behavior of the PTM and the LTM with algorithms based on the "Xgrow" simulator. The original Xgrow was developed to simulate the behavior of the kTAM of the OTM (Winfree 1998).

In the Xgrow simulation, we assume that an assembly grows starting from a given seed structure. Reaction rates are expressed as probability rates for a Poisson process: the association or dissociation of a monomer from the current assembly. A 2-D array stores the information on the arrangement of tiles in the current assembly. A seed structure is pre-set on the array at initial time $t = 0$. One of three events is chosen stochastically according to the rates of all possible reactions: An *on*-event, where a new tile is added to a site of the array, an *off*-event, where a tile in the assembly is removed, or a *change*-event, where the current state of a protected or layered tile changes states. On-events occurs at rate r_f, taken to be a constant. All possible rule tiles are equally likely to be chosen to attempt to bind to the site. Off-events are chosen by reverse rate $r_{r,b}$, which depends on the total bonding strength b of all matched edges. As shown in Table 1, strength at each edge depends on whether each tile's protection status.

The original Xgrow does not consider the existence of the protection strand or the protection tiles. In order to take them into account, we have to use the hsTAM. We introduce the "*change*-events" for this purpose, which expresses the change of intermediate states. The change-events are chosen by change rate $r_{c,s_1 \to s_2}$. For example in case of a tile matching by one input in Fig. 8b, a tile in state T_{11} for the PTM has $r_{c,T_{11} \to T_{10}} = k_s$ and $r_{c,T_{11} \to T_{01}} = \frac{k_s}{k_f} r_{r,\frac{11}{14}}$. Once the event is chosen and effected, all rates are recalculated according to the new status of the assembly to determine the next event.[11]

4.2 Suppression of growth errors, facet errors and nucleation errors

We simulated OTM-kTAM, PTM-hsTAM, and LTM-hsTAM to evaluate the error-reducing effect of the protection strand and protection tile. Figure 11 shows simulation results of (a) OTM, (b) PTM, and (c) LTM. All simulations were performed at the same simulated temperature ($\sim$25°C) and concentration ($\sim$10 nM). (Unlike for the PTM and LTM, these conditions are not near to optimal for the OTM, but the rank comparison remains the same if this is corrected for.)

The PTM and the LTM are evaluated for all the combinations of G_{se} and G_{mc} ($1 \leq G_{se}$, $G_{mc} \leq 60$). The first and second row of Fig. 12 shows the results as a phase plot for (a) the PTM and (b) the LTM seeded by a preformed border and a single foundation tile, respectively. These results tell us which part of the regime is required to obtain a crystal with few errors by using the PTM or the LTM.

The third row of Fig. 12 shows the abstract diagrams of simulation results by a preformed border seed. In the PTM's diagram, there are two border lines $G_{mc} = \frac{22}{14} G_{se} - 28$ and $G_{mc} = \frac{3}{14} G_{se} + 9$. When $\frac{22}{14} G_{se} - 28 < G_{mc} < 2G_{se}$, spontaneous dissociation of a 22-nt protection strand from the foundation tile is thermodynamically favorable, as $r_f^{prot} < r_f^{unp}$. Thus, foundation tiles associate with the assembly ($E \to T_{00}$*). In this region, behavior of the PTM is mostly similar to that of the OTM. When $\frac{3}{14} G_{se} + 9 < G_{mc} < \frac{22}{14} G_{se} - 28$, a protected tile associates with the assembly because the protection strand keeps binding to

[11] The actual computer code is optimized to remove redundant calculations.

Table 1 Bonding strength of sticky ends

Mechanism	Input	Output	
		Covered	Uncovered
OTM	Covered	–	–
	Uncovered	–	5/5
PTM	Covered	–	3/14
	Uncovered	–	14/14
LTM	Covered	0	3/14
	Uncovered	11/14	14/14

The value of strength in the OTM is always 5/5. The value of strength in the PTM and the LTM is different depending on whether the sticky end is covered or not

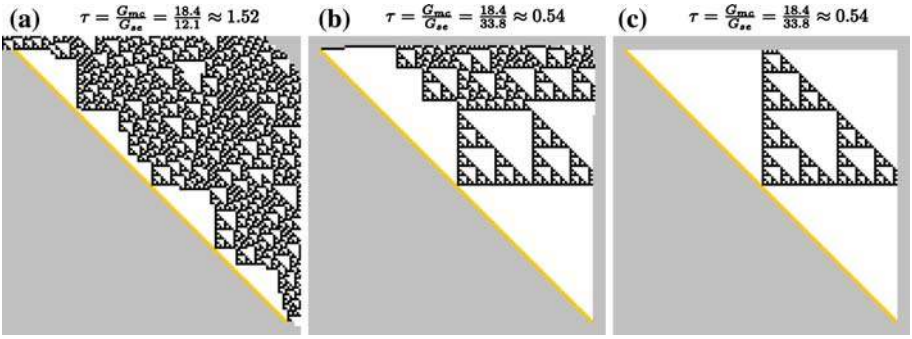
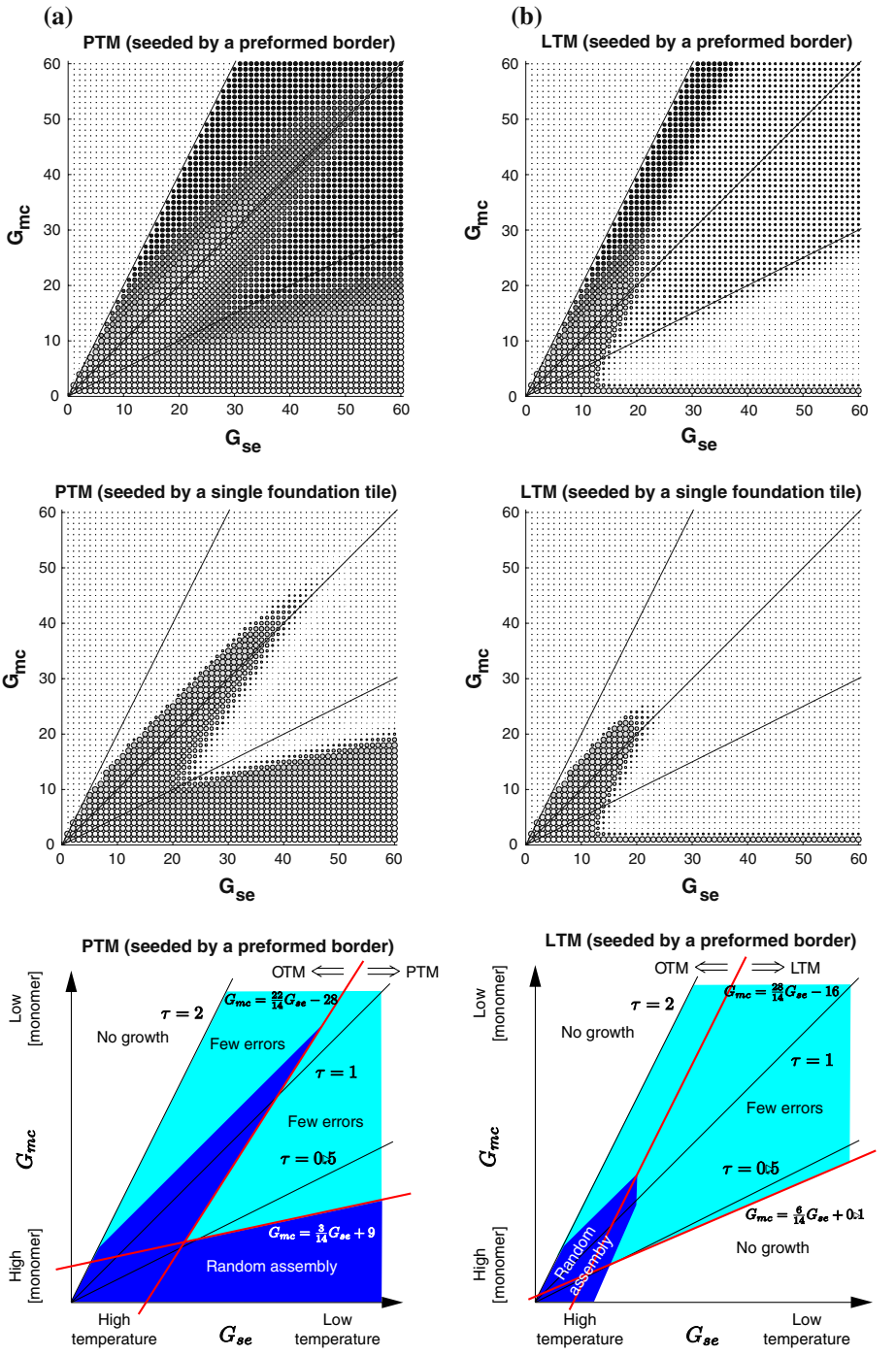


Fig. 11 Simulation results of growth in (**a**) the OTM, (**b**) the PTM, and (**c**) the LTM. These results were computed on a 128 × 128 array. A preformed border is given as the initial seed structure (orange). All simulations were performed at the same condition: $[monomer] = e^{-G_{mc}} \approx 10$ nM, $T = \frac{4000s}{G_{se}+11s} \approx 25+273.15$ K (OTM: $s = 5$ nt, PTM & LTM: $s = 14$ nt) (Winfree 1998). The numbers of erroneous tiles were 89 tiles for the OTM, 3 tiles for the PTM, and 0 tile for the LTM. Note that while these conditions are near-optimal in the PTM and the LTM for these growth rates, much lower error rates (5–48 tiles) could be achieved in the OTM by increasing to temperatures such that G_{se} is comparably close to optimal (cf. Figs. 4 and 6). Simulated duration (in simulated seconds) to complete these assemblies was about 2.1×10^4 s for the OTM, about 2.5×10^4 s for the PTM, and about 7.3×10^4 s for the LTM

inputs of the tile as $r_f^{prot} > r_f^{unp}(E \rightarrow T_{11})$. When $G_{mc} < \frac{3}{14}G_{se} + 9$, random assemblies are generated because 3 nt of the uncovered sticky end is enough to bind to another tile. In this region, the protected tile is not enough to inhibit spontaneous formation of assemblies. If several monomer protected tiles cohere at the same time, then the protection strand can be removed by branch migration and random assemblies form.

The abstract diagram of the LTM shows more unique results. It also has two border lines $G_{mc} = \frac{28}{14}G_{se} - 16$ and $G_{mc} = \frac{6}{14}G_{se} + 0.1$. When $\frac{28}{14}G_{se} - 16 < G_{mc} < 2G_{se}$, behavior of the LTM are mostly similar to that of the OTM because a protection tile with $b = \frac{28}{14}$

Fig. 12 Phase plot for (**a**) the PTM and (**b**) the LTM seeded by a preformed border (top) and a single ▶ foundation tile (middle). Black lines are with slopes 2, 1, and 0.5. Each disc represents results of a single simulation; the disc size represents the size of the assembly; the shading tone represents the error rate (solid black indicates few or zero errors). Each simulation was computed on a 128 × 28 array, and it was run until on-event = 250,000 times. The assembly sizes were limited to a maximum of 10,000 tiles. The third diagrams are abstract diagrams of simulation results by a border seed. A blue region represents random assemblies, and a light blue region represents assemblies with few errors


(a)
PTM (seeded by a preformed border)
(b)
LTM (seeded by a preformed border)
PTM (seeded by a single foundation tile)
LTM (seeded by a single foundation tile)
G_{mc}
G_{se}
PTM (seeded by a preformed border)
OTM ⟸ ⟹ PTM
Low [monomer]
High [monomer]
$\tau = 2$
No growth
$G_{mc} = \frac{22}{14}G_{se} - 28$
Few errors
$\tau = 1$
Few errors
$\tau = 0.5$
$G_{mc} = \frac{3}{14}G_{se} + 9$
Random assembly
High temperature
Low temperature
LTM (seeded by a preformed border)
OTM ⟸ ⟹ LTM
$G_{mc} = \frac{28}{14}G_{se} - 16$
$\tau = 2$
No growth
$\tau = 1$
Few errors
$\tau = 0.5$
$G_{mc} = \frac{6}{14}G_{se} + 0.1$
No growth
Random assembly
High temperature
Low temperature

spontaneously dissociates from the foundation tile (i.e., $r_f^{prot} < r_f^{unp}$). When $\frac{6}{14} G_{se} + 0.1 < G_{mc} < \frac{28}{14} G_{se} - 16$, a layered tile associates with the assembly as $r_f^{prot} > r_f^{unp}$. Self-assembly with the LTM possessed lower error rate (though smaller crystals) than with the OTM. When $G_{mc} < \frac{6}{14} G_{se} + 0.1$, almost no assembly is observed. Dissociation of the protection tile is thermodynamically unfavorable because it gains $b = \frac{6}{14}$, which is higher than τ in this regime, even if both of its input sticky ends are detached. Therefore, tiles seldom form random assemblies in this mechanism.

To look into the suppression effect for facet errors, we simulated assemblies seeded by a "preformed horizontal border" for (a) the OTM, (b) the PTM, and (c) the LTM at the same temperature ($\sim$25°C) and concentration ($\sim$10 nM) as in Fig. 13. The border is composed of many identical border tiles bound irreversibly to each other. On the facet, the tiles have sticky ends of strength 1. Under the OTM, when two tiles simultaneously attach onto the facet, both tiles are bound with strengths $b = 2$, and thus stabilize each other's binding.

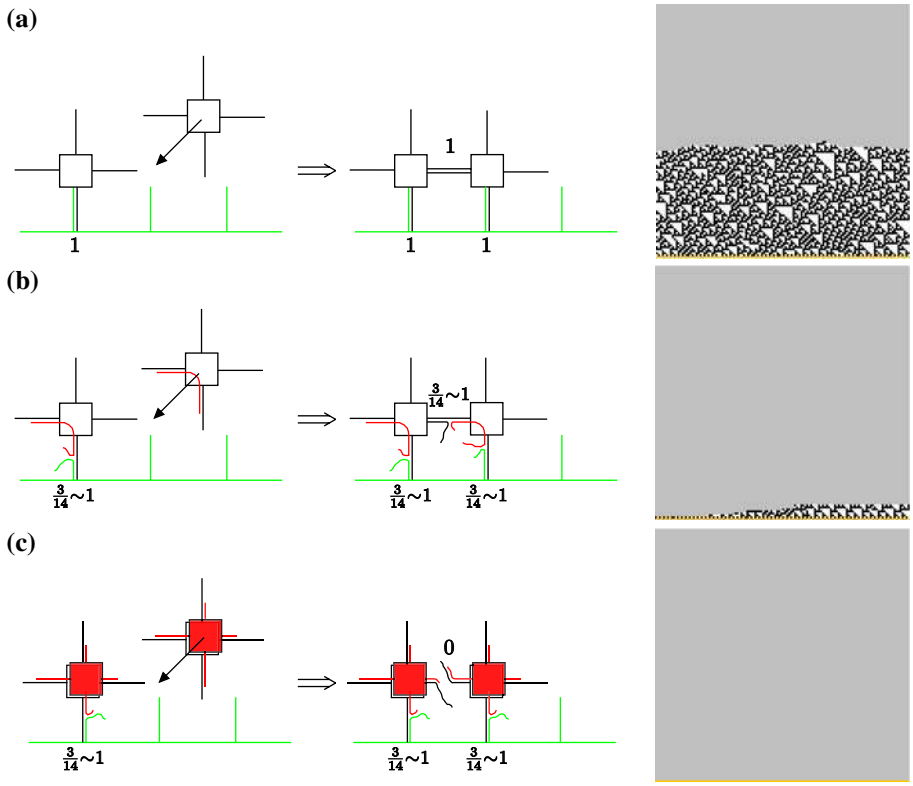


Fig. 13 Growth processes on a facet under (**a**) the OTM, (**b**) the PTM, and (**c**) the LTM at (G_{se}, G_{mc}) = (12.1, 18.4), (33.8, 18.4), and (33.8, 18.4), respectively. As in Fig. 11, this corresponds to [*monomer*] $\approx$ 10 nM and $T \approx 25 + 273.15$ K. (While these conditions are near-optimal in the PTM and the LTM for these growth rates, much lower facet growth rates (by 3- to 12-fold) could be achieved in the OTM by increasing to temperatures such that G_{se} is comparably close to optimal (cf. Figs. 4, 6, and 11). Growth starts from a preformed horizontal border (represented by the orange line). These results were computed on a 128 $\times$ 128 array. Simulated duration (in actual time) to complete these assemblies was about 5 $\times$ 10^4 s for each mechanism

Once this happens, concave sites appear, and on the sides of the pair, a new layer can be formed via normal growth. This type of errors constructs another layer ad infinitum and eventually makes the assembly completely random.

Consider facet errors under the PTM. One protected tile attaches on the facet, and branch migration starts on its matched edge. The facet only has three unique bases compared to the 11 from the protection strand, and thus dissociation is likely. Therefore, the PTM can suppress facet errors compared to the OTM. However, if another protected tile simultaneously binds to the assembly adjacent to the first tile before the first tile detaches, then the output sticky end of the second tile may branch-migrate and displace the protection strand, and thus stabilize each other, and allow growth of a row. In simulations, several layers are formed, but fewer than that of the OTM. (The conditions used in Fig. 13 are far from optimal for the OTM at the given growth rate, but the qualitative conclusions still apply even when this is corrected for.)

In case of the LTM, the strength between two tiles attaching on the facet is zero because the protection tile covers both input and output sticky ends. When two layered tiles simultaneously attach to adjacent positions on the facet, the output sticky end of one tile does not stabilize the input sticky end of the other. Hence, simulations show few facet errors (none in the particular simulation run shown) compared to the OTM and the PTM.

Nucleation errors are basically initiated by the roughly simultaneous association of more than four tiles. In the OTM, many clusters are easily produced by cohesion of tiles with exposed sticky ends. On the other hand, the PTM and the LTM can suppress such nucleation errors as well as the facet errors. The middle diagrams in Fig. 12 ("seeded by a foundation tile") indicate the size and error rates in such spontaneously nucleated assemblies. The PTM suppresses nucleation errors when $\frac{3}{14}G_{se} + 9 < G_{mc} < \frac{22}{14}G_{se} - 28$. Random assemblies are formed by protected tiles in the PTM when $G_{mc} < \frac{3}{14}G_{se} + 9$, but their region is smaller than that of the OTM. The LTM dramatically decreases the number of random assemblies all over the region.

4.3 Relationship between growth speed and error rates

We compared the growth speed r to the error rates ϵ of the three mechanisms. Self-assembly seeded by a preformed diagonal border was simulated. The obtained error rate includes the growth errors and the facet errors, but not the nucleation errors. It is calculated as the total number of mismatches divided by the total number of tiles in the assembly after a predetermined number of on-events have occurred. The results are shown in Fig. 14.

Each reddish or blueish curve represents simulation results with varying G_{mc} fixed G_{se} in the three mechanisms. In the PTM and the LTM, the growth speed and error rates decrease for larger G_{mc}. And then however, error rates of these curves gradually get worse. The reason is that the protection strand or the protection tile easily dissociate from the tile at larger τ.

Green plots show estimated Pareto-optimal boundaries which achieve the highest speed and the lowest error rates of each mechanism, and black lines show their fitted envelopes. The OTM's line occurs for $G_{mc} = 2\ G_{se} - 0.2$ (Winfree 1998; Winfree and Bekbolatov 2004). The lines for the PTM and the LTM maintain $G_{mc} = 0.62\ G_{se} - 2.8$ and $G_{mc} = 0.63\ G_{se} - 5.1$, respectively. From the slopes for these lines $\frac{d(\log \epsilon)}{d(\log r)}$, we obtained $r \approx \beta_{OTM} \times \epsilon^{2.0}$ for the OTM, $r \approx \beta_{PTM} \times \epsilon^{1.4}$ for the PTM, and $r \approx \beta_{LTM} \times \epsilon^{0.7}$ for the LTM. The multiplicative constants are estimated to be $\beta_{OTM} \approx 1.2 \times 10^4$ /M/s, $\beta_{PTM} \approx 4.4 \times 10^2$ /M/s and $\beta_{LTM} \approx 3.6 \times 10^2$ /M/s. Thus we see that the LTM achieves

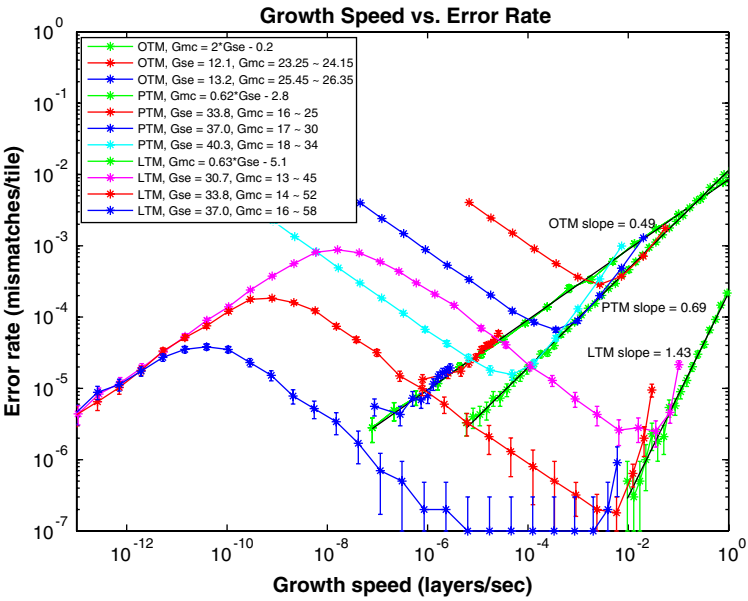


Fig. 14 Growth speed versus error rate. Self-assembly of 1,000,000–50,000,000 tiles represent one point. The error rate is given as 10^{-7} if the simulation results in no errors. Error bars of these plots show two standard deviations: $\sigma = \sqrt{\epsilon}/\sqrt{m}$ where m is the total number of tiles grown in all simulation runs under a given G_{se} and G_{mc} and ϵ is the total number of mismatches divided by the total number of tiles. Red, blue, magenta, and light blue curves show each run at varying concentrations under fixed temperature (i.e., fixed G_{se}). Green plots show simulations along estimated Pareto-optimal boundaries for each mechanism, and black lines show their approximations

higher speed and lower error rate compared with the other two mechanisms. For instance, in order to obtain assemblies with an error rate of less than 10^{-4} mismatches per tile, the growth speed must be slowed down to 10 layers/day by the OTM. It can be significantly accelerated to 4 layers/h by the PTM, or 34 layers/min by the LTM.

5 Conclusions

In this article, we proposed two novel error suppression mechanisms for DNA tile self-assembly. The first is the PTM, which utilizes short additional DNA strands called protection strands that cover the input sticky ends of monomer tiles. The second is the LTM, which utilizes additional DNA tiles called the protection tiles that cover both the input and the output sticky ends of monomer tiles.

In order to verify the effect of these error suppression mechanisms, we conducted simulations to evaluate their performance in terms of growth errors, facet errors, nucleation errors, and growth speed. We found that both the PTM and the LTM produce an error suppression effect. Furthermore, we revealed relations between the growth speed r and the error rates ϵ for the two mechanisms under optimal conditions. In the conventional mechanism (OTM), $r \approx \beta\epsilon^{2.0}$ was optimal, whereas in the PTM and the LTM, this relation is dramatically improved to $r \approx \beta\epsilon^{1.4}$ and $r \approx \beta\epsilon^{0.7}$ respectively. By comparison, 2×2 and 3×3 proofreading yield $r \approx \beta\epsilon^{1.0}$ and $r \approx \beta\epsilon^{0.7}$ respectively, but at the cost of using

four or nine times as many tiles. It is somewhat surprising that the LTM performs so much better than the PTM. It may be possible that adjusting the design parameters (e.g., sticky end length and protection strand length) could improve the performance of the PTM, but we do not expect this to change the rank ordering of performance.

As a practical matter, it may be difficult to directly implement the PTM or the LTM in a one-pot reaction. To understand this problem, one may assume the initial solution in a tube contains 10 nM ($G_{mc} \approx 18.4$) of each species and it is annealed from 90 to 20°C ($G_{se} \approx$ 0.2–33.8). The annealing process corresponds to a horizontal traverse on the phase plot of Fig. 12, which must pass through the region of random assembly to reach the optimal region. This could be solved by introducing the microfluidic device (Somei et al. 2006), in which we can form necessary covered complexes (i.e., protected tiles or layered tiles) in independent chambers and then mix them under optimal temperatures and concentrations. Alternatively, this problem could be circumvented by pre-annealing the tiles individually and then mixing them together at room temperature.

Acknowledgements This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (No. 17059001) from MEXT and Grant-in-Aid for Scientific Research (A) (No. 19200023) from JSPS to SM, JSPS Research Fellowships for Young Scientists (No. 05697) to KF, with additional support from NSF Grant (No. 0523761) to EW, and the Fannie and John Hertz Foundation to DYZ.

References

Adleman L, Cheng Q, Goel A, Huang M-D (2001) Running time and program size for self-assembled squares. In: STOC'01: Proceedings of the 33rd annual ACM symposium on theory of computing. ACM Press, New York, NY, pp 740–748

Barish RD, Rothemund PWK, Winfree E (2005) Two computational primitives for algorithmic self-assembly: copying and counting. Nano Lett 5:2586–2592

Baryshnikov Y, Coffman E, Seeman N, Yimwadsana T (2006) Self-correcting self-assembly: growth models and the Hammersley process. In: Carbone A, Pierce NA (eds) DNA Computing 11, vol. 3892 of LNCS. Springer-Verlag, Berlin, pp 1–11

Biswas I, Yamamoto A, Hsieh P (1998) Branch migration through DNA sequence heterology. J Mol Biol 279:795–806

Chen H-L, Goel A (2005) Error free self-assembly using error prone tiles. In: Ferretti C, Mauri G, Zandron C (eds) DNA Computing 10, vol 3384 of LNCS. Springer-Verlag, Berlin, pp 1–11

Chen H-L, Cheng Q, Goel A, Huang M-D, de Espanes PM (2004) Invadable self-assembly: combining robustness with efficiency. In: SODA '04: Proceedings of the 15th annual ACM-SIAM symposium on discrete algorithms. SIAM, Philadelphia, PA, pp 890–899

Chen H-L, Schulman R, Goel A, Winfree E (2007) Reducing facet nucleation during algorithmic self-assembly. Nano Lett 7:2913–2919

Cook M, Rothemund PWK, Winfree E (2004) Self-assembled circuit patterns. In: Chen J, Reif JH (eds) DNA Computing 9, vol 2943 of LNCS. Springer-Verlag, Berlin, pp 91–107

Dirks RM, Pierce NA (2004) Triggered amplification by hybridization chain reaction. Proc Natl Acad Sci USA 101:15275–15278

Fu T-J, Seeman NC (1993) DNA double-crossover molecules. Biochemistry 32:3211–3220

Fujibayashi K, Murata S (2005) A method of error suppression for self-assembling DNA tiles. In: Ferretti C, Mauri G, Zandron C (eds) DNA Computing 10, vol 3384 of LNCS. Springer-Verlag, Berlin, pp 113–127

Panyutin IG, Hsieh P (1994) The kinetics of spontaneous DNA branch migration. Proc Natl Acad Sci USA 91:2021–2025

Panyutin IG, Biswas I, Hsieh P (1995) A pivotal role for the structure of the Holliday junction in DNA branch migration. EMBO J 14:1819–1826

Reif JH (1999) Local parallel biomolecular computation. In: Rubin H, Wood DH (eds) DNA based computers III, vol 48 of DIMACS. AMS Press, Providence, RI, pp 217–254

Reif JH, Sahu S, Yin P (2005) Compact error-resilient computational DNA tiling assemblies. In: Ferretti C, Mauri G, Zandron C (eds) DNA Computing 10, vol 3384 of LNCS. Springer-Verlag, Berlin, pp 293–307

Reynaldo LP, Vologodskii AV, Neri BP, Lyamichev VI (2000) The kinetics of oligonucleotide replacements. J Mol Biol 297:511–520

Rothemund PWK, Winfree E (2000) The program-size complexity of self-assembled squares (extended abstract). In: STOC'00: Proceedings of the thirty-second annual ACM symposium on theory of computing. ACM Press, New York, NY, pp 459–468

Rothemund PWK, Papadakis N, Winfree E (2004) Algorithmic self-assembly of DNA Sierpinski triangles. PLoS Biol 2: 2041–2053

Sahu S, Reif JH (2006) Capabilities and limits of compact error resilience methods for algorithmic self-assembly in two and three dimensions. In: Mao C, Yokomori T (eds) DNA Computing 12, vol 4287 of LNCS. Springer-Verlag, Berlin, pp 223–238

Schulman R, Winfree E (2005) Programmable control of nucleation for algorithmic self-assembly. In: Ferretti C, Mauri G, Zandron C (eds) DNA Computing 10, vol 3384 of LNCS. Springer-Verlag, Berlin, pp 319–328. Extended abstract in DNA Computing 10; preprint of the full paper is cond-mat/0607317 on http://arXiv.org

Schulman R, Winfree E (2007) Synthesis of crystals with a programmable kinetic barrier to nucleation. Proc Natl Acad Sci USA 104:15236–15241

Seeman NC (2005) The challenge of structural control on the nanoscale: Bottom-up self-assembly of nucleic acids in 3D. Int J Nanotechnol 2:348–370

Soloveichik D, Winfree E (2004) Complexity of self-assembled shapes. SIAM J Comput 36:1544–1569, 2007. Extended abstract in LNCS 3384: 344-354 (2004); preprint is cs.CC/0412096 on http://arXiv.org

Somei K, Kaneda S, Fujii T, Murata S (2006) A microfluidic device for DNA tile self-assembly. In: Carbone A, Pierce NA (eds) DNA Computing 11, vol 3892 of LNCS. Springer-Verlag, Berlin, pp 325–335

Turberfield AJ, Yurke B, Mills AP Jr (2000) DNA hybridization catalysts and molecular tweezers. In: Winfree E, Gifford DK (eds) DNA based computers V, vol 54 of DIMACS. AMS Press, Providence, RI, pp 171–182

Wang H (1961) Proving theorems by pattern recognition II. Bell Syst Tech J 40:1–42

Wang H (1963) Dominoes and the AEA case of the decision problem. In: Fox J (ed) Proceedings of the symposium on the mathematical theory of automata. Polytechnic Press, Brooklyn, NY, pp 23–55

Wetmur JG (1991) DNA probes: applications of the principles of nucleic acid hybridization. Crit Rev Biochem Mol Biol 26:227–259

Whitesides GM, Mathias JP, Seto CT (1991) Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. Science 254:1312–1319

Winfree E (1996) On the computational power of DNA annealing and ligation. In: Lipton RJ, Baum EB (eds) DNA based computers, vol 27 of DIMACS. AMS Press, Providence, RI, pp 199–221

Winfree E (1998) Simulations of computing by self-assembly. CaltechCSTR:1998.22. California Institute of Technology

Winfree E, Bekbolatov R (2004) Proofreading tile sets: error correction for algorithmic self-assembly. In: Chen J, Reif JH (eds) DNA Computing 9, vol 2943 of LNCS. Springer-Verlag, Berlin, pp 126–144

Winfree E, Liu F, Wenzler LA, Seeman NC (1998) Design and self-assembly of two-dimensional DNA crystals. Nature 394:539–544

Yurke B, Mills AP Jr (2003) Using DNA to power nanostructures. Genet Program Evol Machines 4: 111–122

Yurke B, Turberfield AJ, Mills AP Jr, Simmel FC, Nuemann JL (2000) A DNA-fuelled molecular machine made of DNA. Nature 406:605–608