Toward Minimum Size Self-Assembled Counters

Ashish Goel and Pablo Moisset de Espanés

Department of Management Science and Engineering and (by courtesy) Computer Science,
Stanford University, Terman 311, Stanford CA 94305
ashishg@stanford.edu
pmoisset@usc.edu

Abstract. DNA self-assembly is a promising paradigm for nanotechnology. In this paper we study the problem of finding tile systems of minimum size that assemble a given shape in the Tile Assembly Model, defined by Rothemund and Winfree [14]. We present a tile system that assembles an $N \times \lceil \log_2 N \rceil$ rectangle in asymptotically optimal $\Theta(N)$ time. This tile system has only 7 tiles. Earlier constructions need at least 8 tiles [7]. We managed to reduce the number of tiles without increasing the assembly time. The new tile system works at temperature 3.

The new construction was found by the combination of exhaustive computerized search of the design space and manual adjustment of the search output.

1 Introduction

Self-Assembly (SA) is the process by which autonomous components assemble into complexes following rules of local interaction only. SA is ubiquitous in Nature. Chemistry and Biology provide many examples, such as the formation of crystals and the growth of some organisms. SA is a promising paradigm for assembling shapes and patterns at molecular scale. The ability to construct many objects of intricate design may be useful in the fields of nano-electronics [9] and Material Sciences. The Watson-Crick law of pairing, together with the small size of bases, make DNA an attractive material to build self-assembled systems. There are numerous experimental results that support this approach [11,12,13,16,17,18,20].

Rothemund and Winfree proposed a theoretical model for DNA SA, the Tile Assembly Model (TAM. In the TAM, the DNA compounds are modeled as square tiles with glues on their sides. The individual tiles can stick to a growing assembly, as long as the glues on their sides provide enough sticking strength. Adleman *et al.* [1] added the notion of time complexity to the model. Some variants of the TAM have been explored in [2,3,8].

In [14], Rothemund and Winfree studied the problem of assembling an $N \times N$ square starting from a single tile. In their construction, they first built a rectangle from a base row by simulating a binary counter. Then, they completed the square by other means. Their counter construction required 12 tiles and needed $\Theta(N \cdot \log N)$ time to finish the assembly. Adleman *et al.* [1] presented a new counter that assembles in asymptotically optimal $\Theta(N)$ time, but requires 15 different tiles. Chen, Cheng, Goel and Moisset [7] improved the result by finding a counter that uses only 8 tiles and also achieves $\Theta(N)$ assembly time.

M.H. Garzon and H. Yan (Eds.): DNA 13, LNCS 4848, pp. 46–53, 2008.

Reducing the number of tiles to assemble a given shape has a practical motivation. The cost of materials, i.e. DNA, and the time to carry out an experiment is closely related to the number of tiles in the design. Also, finding a smaller, or the smallest number of tiles to accomplish a given task is a theoretical problem of independent interest. Tiles that assemble a given shape are analogous to a computer program that outputs that shape. Minimizing the number of required tiles is similar to minimizing the program size.

In some computational problems, there is a trade-off between program size and running time. The natural question to ask is if reducing the number of tiles to build a counter forces to increase the assembly time.

T he main results of the paper are: In Section 3, we show a set of 7 tiles that assembles an $N \times \lceil \log_2 N \rceil$ rectangle from an initial base row in asymptotically optimal $\Theta(N)$ time. It is the smallest counter known so far, and it does not incur an increased assembly time. In spite of the small number of tiles, the construction is more involved than those found in [1,14]. The proof of correctness of the new counter is non-trivial and it is outlined in Section 4. For a complete, formal proof, see [4]. The counter with 7 tiles was originally published in [5], but it was described informally and no proof of correctness was given.

The process of finding a working design with only 7 tiles is of independent interest. It relied an exhaustive computerized search. This search was not guaranteed to output a correct design. It was only meant to suggest a candidate set of tiles which had to be verified manually. In fact, the search program produced a set that was flawed, and had to be corrected by hand. Interestingly, the manual modification of the candidate set fell outside the design space our program searched. In any case, the resulting design is so involved that it is unlikely that it could have been found without using the computer-aided approach. Details of the search process will appear in [6].

2 Definitions

The Tile Assembly Model (TAM): The tile assembly model [14,1] extends the theoretical model of tiling by Wang [15] to include a mechanism for growth based on the physics of molecular SA. We will present a succinct definition, with minor modifications for ease of explanation.

A tile is an oriented unit square with the north, east, south and west edges labeled from some alphabet Σ of glues. For each tile $t \in T$, the labels of its four edges are denoted $\sigma_N(t)$, $\sigma_E(t)$, $\sigma_S(t)$, and $\sigma_W(t)$. Sometimes we will describe a tile t as the quadruple $(\sigma_N(t), \sigma_E(t), \sigma_S(t), \sigma_W(t))$. Consider the triple $\langle T, G, \tau \rangle$ where T is a finite set of tiles, $\tau \in \mathbb{Z}_{>0}$ is the *temperature*, and G is the *glue strength* function from Σ to $\mathbb{Z}_{\geq 0}$, where Σ is the set of glues.

Given $p = (x, y), p' = (x', y') \in \mathbb{Z}^2$, we say p and p' are *position adjacent* iff $|x - x'| + |y - y'| = 1$. A *shape* is a finite, connected (under the adjacency relation defined above) subset of $\mathbb{Z}^2$. Let $Dom(f)$ denote the domain of a function f. A *supertile* S of T is a partial function from $\mathbb{Z}^2$ to T such that $Dom(S)$ is a shape. For a supertile S, we will write $[S]$ to represent $Dom(S)$.

Let C and D be two supertiles. Suppose there exist some $t \in T$ and some $(x, y) \in \mathbb{Z}^2$ such that $(x, y) \notin Dom(C)$, $D(x, y) = t$ and $D = C$ except at (x, y). If $(x, y + 1) \in Dom(C)$ and $\sigma_N(t) = \sigma_S(C(x, y + 1))$, let $f_{N,C,t}(x, y) = G(\sigma_N(t))$ and let $f_{N,C,t}(x, y) = 0$ otherwise. Informally $f_{N,C,t}(x, y)$ is the strength of the bond between C and the north side of t. Define $f_{S,C,t}(x, y)$, $f_{E,C,t}(x, y)$ and $f_{W,C,t}(x, y)$ similarly. Then we say that tile t is *attachable* to C at position (x, y) iff $f_{N,C,t}(x, y) + f_{S,C,t}(x, y) + f_{E,C,t}(x, y) + f_{W,C,t}(x, y) \geq \tau$, and we write $C \rightarrow_{\mathbf{T}} D$ to denote the transition from C to D in attaching a tile to C at position (x, y). Informally, $C \rightarrow_{\mathbf{T}} D$ iff D can be obtained from C by adding a tile t such that the total strength of interaction between t and C is at least τ.

A *tile system* is a quadruple $\mathbf{T} = \langle T, s, G, \tau \rangle$, where T, G, τ are as above and s is a special supertile called the "seed". The notion of a *derived supertile* of a tile system $\mathbf{T} = \langle T, s, G, \tau \rangle$ is defined recursively:

1. The seed s is a derived supertile of $\mathbf{T}$, and
2. if $C \rightarrow_{\mathbf{T}} D$ and C is a derived supertile of $\mathbf{T}$, then D is also a derived supertile of $\mathbf{T}$.

Informally, a derived supertile is either just the seed (condition 1 above), or obtained by legal addition of a single tile to another derived supertile (condition 2).

A *terminal supertile* of the tile system $\mathbf{T}$ is a derived supertile A such that there is no supertile B for which $A \rightarrow_{\mathbf{T}} B$. Let $\rightarrow^*_{\mathbf{T}}$ denote the reflexive transitive closure of $\rightarrow_{\mathbf{T}}$. If there is a terminal supertile A such that for any derived supertile B, $B \rightarrow^*_{\mathbf{T}} A$, we say that the tile system *uniquely produces* A. A tile system $\mathbf{T}$ *uniquely produces a shape* W iff it uniquely produces some supertile Γ and $[\Gamma]$ is identical (up to translation) to W.

We will now add the notion of running time to this model. We associate with each tile $t \in T$ a non-negative probability $P(t)$, such that $\sum_{t \in T} P(t) = 1$. We assume that the tile system has an infinite supply of each tile, and $P(t)$ models the concentration of tile t in the system. Now SA of the tile system corresponds to a continuous time Markov process where the states are in a one to one correspondence with derived supertiles, and the initial state corresponds to the seed s. Suppose a single tile t can be added to a derived supertile C to produce supertile D. Then there is a transition from state C to D in the Markov chain, and the rate of the transition is $P(t)$. Suppose the tile system produces a unique terminal supertile $A_{\mathbf{T}}$. In the Markov chain, the time for reaching $A_{\mathbf{T}}$ from s is a random variable. The "running time" of the SA process is defined as the expected value of this random variable. Note that the Markov process modeling the SA process is inherently parallel. For details, see [1].

A supertile Γ is *full* iff for all $p, p' \in [\Gamma]$, if $p' = p + (1, 0)$ then $\sigma_E(\Gamma(p)) = \sigma_W(\Gamma(p'))$ and if $p' = p - (1, 0)$ then $\sigma_W(\Gamma(p)) = \sigma_E(\Gamma(p'))$ and if $p' = p + (0, 1)$ then $\sigma_N(\Gamma(p)) = \sigma_S(\Gamma(p'))$ and if $p' = p - (0, 1)$ then $\sigma_S(\Gamma(p)) = \sigma_N(\Gamma(p'))$. Intuitively, a supertile is full if there are no glue mismatches in the abutting edges of adjacent tiles.

General Purpose Counter (GPC): A quadruple $\langle T, T_s, G, \tau \rangle$, where T and T_s are finite sets of tiles with glues from some alphabet Σ, $G : \Sigma \rightarrow \mathbb{Z}_{\geq 0}$, and τ is a temperature, is a *general purpose counter* iff for all integers $h > 1$, for all integers $w \geq \lceil \log_2 h \rceil$, there exists a supertile $s_{h,w}$ of T_s such that:

1. $\langle T \cup T_s, s_{h,w}, G, \tau \rangle$ uniquely produces a supertile, denote it $\Gamma_{h,w}$, such that $[\Gamma_{h,w}] = \{0, -1, \ldots, -w+1\} \times \{0, 1, \ldots, h-1\}$.
2. $[s_{h,w}] = \{0, -1, \ldots, -w+1\} \times \{0\}$.
3. For all $(x, y) \in \{0, -1, \cdots, -w+1\} \times \{1, 2, \ldots, h-1\}$, $\Gamma_{h,w}(x, y) \in T$.

Informally, the seed row s_w has width w and is made out of tiles in T_s. The tiles in T will grow the rest of the $h \times w$ rectangle on top of s_w. The *size of a GPC* $\langle T, T_s, G, \tau \rangle$ is $|T|$.

The General Purpose Counter problem: Given a temperature τ, find the least positive integer m such that there exist an alphabet Σ and sets of tiles T and T_s with glues from Σ, and there exists $G : \Sigma \to \mathbb{Z}_{\geq 0}$, such that $\langle T, T_s, G, \tau \rangle$ is a GPC and $|T| = m$.

Informally, we would like to find the smallest set of tiles that assembles a rectangle whose size is determined solely by the initial supertile, i.e. the seed, of the SA process. We would like the size of this set of tiles to be independent of the size of the desired rectangle. We will also assume the shape of the seed has to be a horizontal line. Constructions with these properties were used by Rothemund and Winfree [14] and by Adleman *et al.* [1] as a "subroutine" to assemble squares. Since the techniques to assemble rectangles in [14,1,7] are based on repeated addition of binary numbers, we refer to these constructions as *counters*. Information theory imposes a logarithmic lower bound on the width of the counter. Hence, we impose the $w \geq \lceil \log_2 h \rceil$ constraint. Our choice of 2 as base of the logarithm is somewhat arbitrary.

3 A Counter of Size 7

In this section we present a GPC of size 7 that works at temperature 3. We also outline the proof of correctness. Before describing the counter, we introduce some notation.

A supertile Γ is said to be *rectangular* iff there are positive integers w and h such that $[\Gamma] = \{0, -1, \ldots, -w+1\} \times \{0, 1, \ldots, h-1\}$. We will call w and h the *width* and *height* of Γ, respectively.

Let Γ be a rectangular supertile, and let w and h be the width and height of Γ, respectively. For all $k \in \{0, 1, \ldots, w-1\}$, let $\mathcal{C}_{\Gamma,k} = (\Gamma(-k, 0), \Gamma(-k, 1), \ldots, \Gamma(-k, h-1))$. For all $k \in \{0, 1, \ldots, w-1\}$, we will refer to the restriction of Γ to $\{-k\} \times \{0, 1, \ldots, h\}$ as *the k-th column* of Γ. Similarly, for all $k \in \{0, 1, \ldots, h-1\}$, let $\mathcal{R}_{\Gamma,k} = (\Gamma(-w+1, k), \Gamma(-w+2, k), \ldots, \Gamma(0, k))$. For all $k \in \{0, 1, \ldots, h-1\}$, we will refer to the restriction of Γ to $\{0, -1, \ldots, -w+1\} \times \{k\}$ as the *k-th row* of Γ.

The set of tiles: We begin by giving a pictorial representation of the counter in Figure 1. Define $\mathcal{T} = \{T_1, T_2, \cdots, T_7\}$, and define the glue-strength function $G : \{a, b, c, d, e, f, g\}^2 \to \{0, 1, 2, 3\}$ so that $G(a) = G(b) = 3$, $G(c) = G(d) = G(e) = 2$ and $G(f) = G(g) = 1$.

The supertile $\mathcal{B}_w$: Given a sequence S, and a positive integer k, we will write S_k to denote the k-th element of S. For all positive integers k and l, define $S_{k,l}$ as the subsequence of S comprising all elements from S_k through S_l. Define the sequence concatenation operator $\bullet$ in the usual way. For all positive integers k, for all finite sequences S, we will write $k \times S$ to denote $S \bullet S \bullet \cdots \bullet S$, where S is concatenated k times.

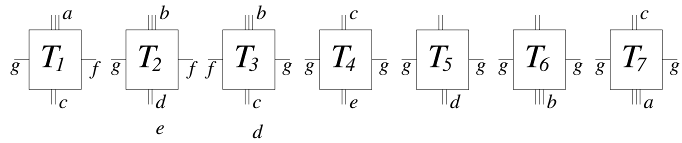

Fig. 1. Counter with 7 tiles

Define the following infinite sequences of tiles with period 6.

$$\bar{D} = (T_4, T_3, T_6, T_2, T_6, T_5, T_4, T_3, T_6, T_2, T_6, T_5, \ldots)$$
$$D = (T_6, T_5, T_4, T_1, T_7, T_3, T_6, T_5, T_4, T_1, T_7, T_3, \ldots)$$
$$\bar{E} = (T_4, T_1, T_7, T_3, T_6, T_5, T_4, T_1, T_7, T_3, T_6, T_5, \ldots)$$
$$E = (T_6, T_2, T_6, T_5, T_4, T_3, T_6, T_2, T_6, T_5, T_4, T_3, \ldots)$$

For all positive integers w, define the following w sequences of length 2^w:

1. $C^{(0,w)} = 2^{w-1} \times (T_1, T_7)$
2. For all odd and positive $k \leq w-1$, $C^{(k,w)} = 2^{w-k-1} \times (\bar{D}_{1,2^k} \bullet D_{1,2^k})$
3. For all even and positive $k \leq w-1$, $C^{(k,w)} = 2^{w-k-1} \times (\bar{E}_{1,2^k} \bullet E_{1,2^k})$

For all positive integers w, define the rectangular supertile $\mathcal{B}_w$ in such a way that the width of $\mathcal{B}_w$ is w, the height of $\mathcal{B}_w$ is 2^w and for all $(k,i) \in \{0,1,\ldots,w-1\} \times \{0,1,\ldots,h-1\}$, $\mathcal{B}_w(-k,i) = C^{(k,w)}_{i+1}$. Note that $\mathcal{C}_{\mathcal{B}_w,k} = C^{(k,w)}$. Define s_w as the 0-th row of $\mathcal{B}_w$. Figure 3 shows $\mathcal{B}_3$, as an example.

4 Results and Proof Outlines

We state now the main result of the paper:

Theorem 1. *For all positive integers w, the tile system $\mathbf{T}_w = \langle \mathcal{T}, G, s_w, 3\rangle$ uniquely produces $\mathcal{B}_w$.*

Note that the height of $\mathcal{B}_w$ is exactly 2^w. Minor modifications to s_w allow the assembly of rectangles of all heights up to 2^w. The details about the modifications are omitted.

For reasons of space, we present only an informal outline of the proof of correctness here. For details see [4].

The proof is constructive, showing that if s_w is the seed row and the temperature is 3, the tiles in $\mathcal{T}$ uniquely assemble $\mathcal{B}_w$. The first step is to show that $\mathcal{B}_w$ is a full supertile, i.e. there are no glue mismatches between adjacent tiles. This fact follows from the definition of $\mathcal{B}_w$. Then we prove that $\mathcal{B}_w$ can be derived from s_w. The proof of this fact is constructive, showing a particular derivation of $\mathcal{B}_w$ from s_w. The process is sketched in Figure 2.

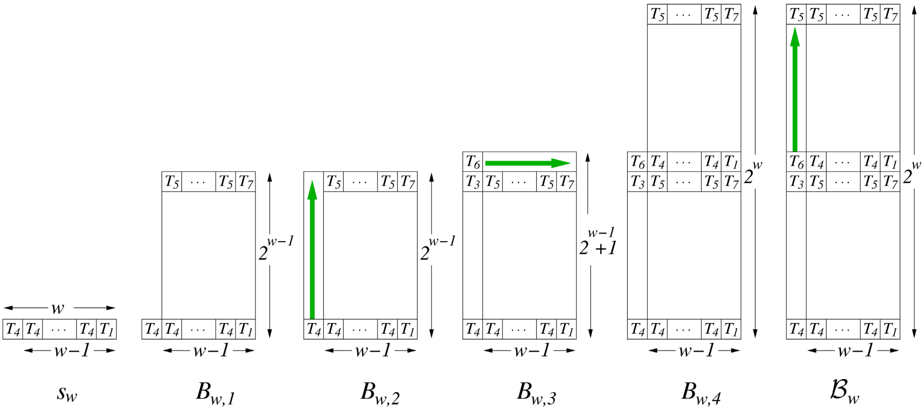

Fig. 2. The inductive step

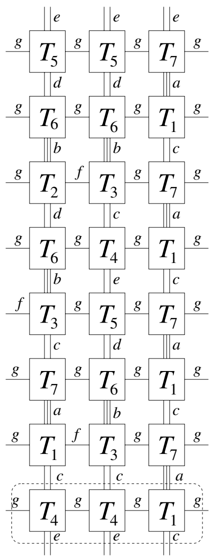

Fig. 3. The supertile $\mathcal{B}_3$. The dashed line encloses s_3.

We prove $\mathcal{B}_w$ can be derived from s_w by induction on w, exploiting the recursive structure of $\mathcal{B}_w$. Roughly speaking, $\mathcal{B}_w$ contains two copies of $\mathcal{B}_{w-1}$. Therefore, we use the inductive hypothesis to prove that we can derive the supertile $B_{w,1}$ from s_w. This follows from $B_{w,1}$ being $\mathcal{B}_{w-1}$ with an extra tile attached to its west side. It follows form $\mathcal{B}_w$ being full that we can start growing the westmost column of $\mathcal{B}_w$, deriving $B_{w,2}$ from $B_{w,1}$. Using a similar argument, we add one row to the northmost side of $B_{w,2}$ to obtain $B_{w,3}$. Now we use the inductive hypothesis, and grow another copy of $\mathcal{B}_{w-1}$ on top of $B_{w,3}$, yielding $B_{w,4}$. Finally, we use appeal to $\mathcal{B}_w$ being full to prove we can finish assembly the westmost column of $\mathcal{B}_w$.

We know that $\mathcal{B}_w$ is produced from s_w. Therefore, we just need to prove that production is unique, which is done through a case analysis.

We conclude by stating the time complexity of $\mathbf{T}_w$ is $\Theta(2^w)$, which follows from the derivation of $\mathcal{B}_w$ used to prove that $\mathcal{B}_w$ derives from s_w, and from results in [10]. The proof of the next theorem relies on Lemmas 3.3, 3.4, and 4.1, and Theorem 4.4 in [10].

Theorem 2. *There exists a concentration function $P : \mathcal{T} \to (0,1)$ such that for all positive integers w, the time complexity of $\mathbf{T}_w = \langle \mathcal{T}, G, s_w, 3 \rangle$ is $\Theta(2^w)$.*

Proof outline: Define P as a constant valued function with value 1/7. The bound $\Omega(2^w)$ is trivial. Call E_w the equivalent acyclic graph induced by the derivation of $\mathcal{B}_w$ described in Figure 2. E_w is identical to the DAG G_N defined in the proof of Lemma 3.3 in [10], if $N = 2^w$. The length of the longest path in E_w is $O(2^w)$. By Theorem 4.4 in [10], the time complexity of $\mathbf{T}_w$ is $O(2^w)$. □

5 Open Problems

Although the counter presented here uses fewer tiles than any other known counter, there are still some unanswered questions.

1. Our counter works at temperature 3, which is undesirable for lab implementations. Experience shows [12,19] that it is possible to obtain a reasonable approximation to the TAM at temperature 2 using DNA tiles. The question whether or not there exists a GPC of size 7 at temperature 2 remains open.
2. Our counter produces full supertiles. Dropping that constraint could potentially result in smaller counters that work at lower temperature. We are currently pursuing that goal.
3. The exhaustive exploration techniques we used to find the tile system do not scale up well past 7 or 8 tiles. This is consequence of the combinatorial explosion of the design space as the number of tiles grow. Perhaps it is possible to find an efficient algorithm that yields sub-optimal results.

Acknowledgments

We would like to thank Len Adleman, Ming-Deh Huang, Yuri Brun and Manoj Gopalkrishnan for useful discussion, and especially Dustin Reishus for his comments on the first manuscript.

References

1. Adleman, L., Cheng, Q., Goel, A., Huang, M.: Running time and program size for self-assembled squares. In: Proceedings of the thirty-third annual ACM symposium on Theory of computing, pp. 740–748. ACM Press, New York (2001)
2. Aggarwal, G., Goldwasser, M., Kao, M., Schweller, R.T.: Complexities for generalized models of self-assembly. In: Proceedings of symposium on discrete algorithms, ACM Press, New York (2004)

3. Cheng, Q., Moisset de Espanés, P.: Resolving two open problems in the self-assembly of squares. Technical Report 03-793, University of Southern California (2003)
4. Moisset de Espanés, P., Goel, A.: Toward minimum size self-assembled counters. In: 13th International Meeting on DNA Computing (2007)
5. Moisset de Espanés, P.: Computerized exhaustive search for optimal self-assembly counters. In: FNANO 2005: Proccedings of the 2nd Annual Foundations of Nanoscience Conference, pp. 24–25 (2005)
6. Moisset de Espanés, P.: Systems self-assembly: Multidisciplinary snapshots, N. Krasnogor, S. Gustafson, D. Pelta, J.L. Verdegay (eds.) Elsevier (2007)
7. Goel, A., Chen, H., Cheng, Q., Moisset de Espanés, P.: Invadable self-assembly, combining robustness with efficiency. In: Proceedings of symposium on discrete algorithms, ACM Press, New York (2004)
8. Kao, M.-Y., Schweller, R.: Reducing tile complexity for self-assembly through temperature programming (2006)
9. Rothemund, P., Cook, M., Winfree, E.: Self assembled circuit patterns. In: Proceedings of DNA Computing, Springer, Heidelberg (2003)
10. Goel, A., Cheng, Q., Moisset de Espanés, P.: Optimal self-assembly of counters at temperature two. In: Foundation of Nanoscience (2004)
11. Rothemund, P.: Theory and Experiments in Algorithmic Self-Assembly. PhD thesis, University of Southern California (2001)
12. Rothemund, P.W., Papadakis, N., Winfree, E.: Algorithmic self-assembly of dna sierpinski triangles. PLoS Biol. 2(12) (December 2004)
13. Rothemund, P.W.K.: Design of dna origami. In: ICCAD 2005: Proceedings of the 2005 IEEE/ACM International conference on Computer-aided design, pp. 471–478. IEEE Computer Society, Washington (2005)
14. Rothemund, P.W.K., Winfree, E.: The program-size complexity of self-assembled squares (extended abstract). In: Proceedings of the thirty-second annual ACM symposium on Theory of computing, pp. 459–468. ACM Press, New York (2000)
15. Wang, H.: Proving theorems by pattern recognition ii. Bell Systems Technical Journal 40, 1–42 (1961)
16. Winfree, E.: Algorithmic Self-Assembly of DNA. PhD thesis, California Institute of Technology, Pasadena (1998)
17. Winfree, E., Liu, F., Wenzler, L., Seeman, N.: Design and self-assembly of two-dimensional dna crystals (6 pages). Nature 394, 539–544 (1998)
18. Winfree, E., Yang, X., Seeman, N.: Universal computation via self-assembly of dna: Some theory and experiments. In: Proceedings of the Second Annual Meeting on DNA Based Computers, Princeton University (June 1996)
19. Winfree, E., Bekbolatov, R.: Proofreading tile sets: Error correction for algorithmic self-assembly. In: DNA, pp. 126–144 (2003)
20. Yurke, B., Turberfield, A., Mills Jr, A., Simmel, F., Neumann, J.: A dna-fuelled molecular machine made of DNA. Nature 406, 605–608 (2000)