Appendix A

Reducing Algorithmic Assembly Error Rates using Polyamides

Abstract

The use of DNA self-assembly to perform molecular computation is a promising field of research. DNA-based computation has the potential to perform large numbers of computations in parallel in a single solution. To this end, DX crystal growth can be designed to follow a series of algorithmic instructions. In these types of arrays, a nucleating strand functions to give a series of logical inputs, and crystal growth is determined by a set of designed logic rules implemented by a series of DX tiles with varying sticky ends. The error rate for tile incorporation is extremely important for these applications, as a single incorrectly incorporated tile can result in a cascade of errors, as subsequent binding events will be determined based on the erroneous tile. Prior work has required a subset of the DX tiles to be modified with dsDNA hairpins for AFM visualization, which may inadvertently affect the kinetics of binding and thus the error rate. By redesigning this system so that none of the tiles contain DNA hairpins, it may be possible to reduce the error rate. Polyamide-biotin conjugates targeted to specific sequences could be used to visualize a subset of tiles after array formation resulting in lower error rates in array assembly.

A.1 Introduction

The use of 2-dimensional DNA arrays for the creation of nanoscale assemblies was detailed in Chapter 2.[1-4] DX arrays, made up of individual DNA tiles that repeat in a predictable manner through a series of programmed sticky end interactions can be used to create 2-dimensional assemblies. However, the periodic nature of these arrays limits the amount of complexity that can be built into these assemblies. One method for breaking the periodicity of these structures is algorithmic assembly. In algorithmic assembly, the interactions that govern crystal formation are determined by a series of designed algorithms, which results in non-periodic arrays. A set of inputs or starting conditions result in an "assembly program" being executed by the various DNA tiles, resulting in programmed array formation. The ability to perform copy and count functions using algorithmic assembly has recently been demonstrated.[5] It is likely that assembly programs governed by DNA interactions will be important for applications such as DNA-based computing.

The algorithmic assembly concept has been demonstrated by the Winfree lab using a series of DX tiles in combination with a single stranded DNA input strand.[6] Individual DX tiles function as logical operators with binding of each tile being determined by a set of logic rules. Each DX tile acts as a logic gate, where two of the sticky ends act as inputs, and the other two act as outputs. The binding of the tile results in a single logical computation. For example, to compute the XOR function for the inputs 0 and 1, a tile would be designed with 2 input strands that correspond to the 0 and 1 and two output strands that correspond to 1. Appropriate tiles are designed for each set of input and output conditions. Tiles that output 1 are modified to contain a dsDNA hairpin

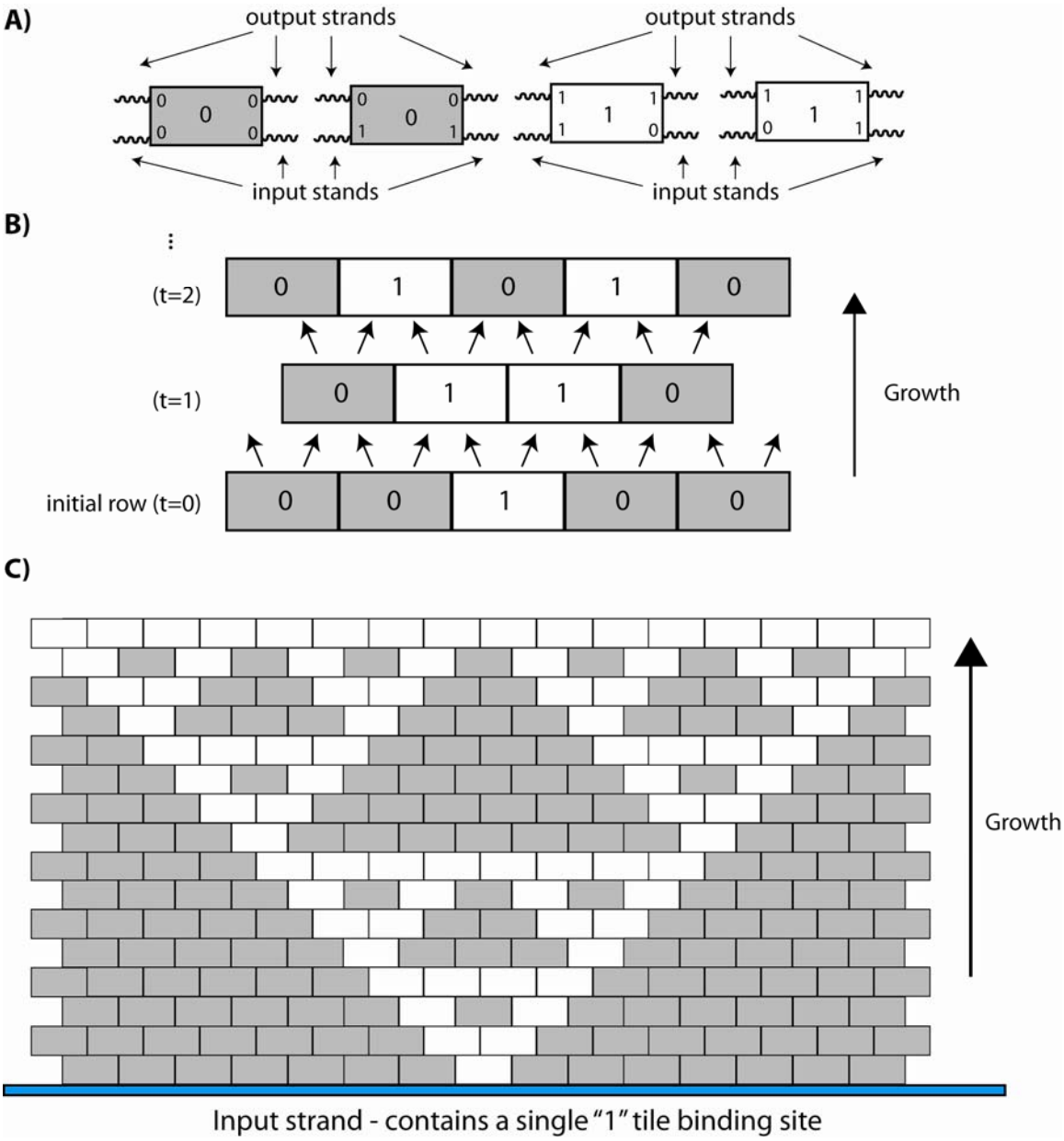

Figure A.1 Overview of algorithmic assembly. A) DX tiles corresponding to the logic function XOR are shown. Each tile contains two input and two output strands that determine binding in a growing array. B) Crystal growth proceeds from an input strand in a manner consistent with the designed logic rules. At each step, future binding events are determined by the previously bound tiles. C) The XOR tile set gives rise to the Sierpinski triangle. The input condition is set by a ssDNA template containing a single "1" binding site surrounded by "0" binding sites.

while tiles that output 0 are left unmodified. The resulting differences in height between tiles can be observed using an atomic force microscope (AFM). An input strand consisting of ssDNA with a binding site for a "1" tile surrounded by a large number of

"0" tiles is used to seed crystal growth. When DX tiles designed to compute the XOR function are used, the algorithm results in the pattern known as the Sierpinski triangle.

One of the hurdles in this field is the error rates for tile binding. In periodic arrays, the sticky ends of tiles are designed such that a mismatch would occur at both ends of any tile which attempted to bind at the incorrect site. However, in the case of algorithmic assembly, many tiles may contain single match and single mismatch domains. If the rate of dissociation for these single match / single mismatch tiles is sufficiently slow, this can lead to incorporation of incorrect tiles in a growing array. Unfortunately, due to the nature of algorithmic growth, a single error can propagate throughout the array, as an incorrect tile will influence future binding events leading to a cascade effect.[6] One of the factors that could possibly influence the kinetics of binding is the incorporation of dsDNA loops into specific tiles which have been used as a visual marker to distinguish logical ones from zeroes. If these dsDNA loops were removed, it may improve the error rates by eliminating any artificial disparities in the binding kinetics between labeled and unlabeled tiles. The use of dsDNA hairpins also means that in order to visualize different tiles in an array, something that is often desirable in algorithmic assembly, it would require constructing structurally different versions of the array, with the appropriate modification transferred from one tile to another.

Given our previous success in non-covalently labeling DX tiles,[7, 8] we proposed that polyamide labeling could provide a method for reducing the error rate in algorithmic assembly. Instead of the dsDNA hairpin modification, a simple polyamide binding site sequence would be incorporated into the tiles that contain a logical 1 output, while the tiles that have a logical zero are designed not to contain the binding site sequence. The

assembly should take place with a reduced error rate since all of the DX tiles used lack any structural modification and should be incorporated at nearly equal rates. After the assembly has formed, visualization is performed by incubation with a polyamide-biotin conjugate and streptavidin and AFM imaging.

A.2 Results and Discussion

The initial design of the project was based upon the originally reported arrays using a DAO-E type tile.[6] Polyamide-biotin conjugates **1** and **2** were previously used for DX array labeling[8] and they are shown along with their target sequences in Figure A.2A. The use of two polyamide-biotin conjugates makes it possible to visualize either the "1" or the "0" tiles simply by changing the polyamide that is used. If successful, the polyamide targeted to the "1" tiles would give the Sierpinski triangle, while the "0" tile polyamide would give an inverse image. Next, the six DX tiles used to create the original array were redesigned. First, the hairpins were removed and the tiles were examined to eliminate any unintentional polyamide binding sites. Three match sites were present in the original tiles for polyamide **1** but their locations at the junctions would prevent any spurious binding. Polyamide **3** had two match sites in the original tiles which were subsequently removed in the newly designed tiles. Each of the new tiles contains two match sites for the desired polyamide in order to maximize potential binding and labeling yields. The sites are designed to face opposite faces to ensure that at least one site is always accessible independent of whether the array lands face up or face down. Nondenaturing polyacrylamide gel electrophoresis showed that each of the new tiles formed properly. Finally, the ssDNA template which acts as a nucleating complex was synthesized as previously described.[6]

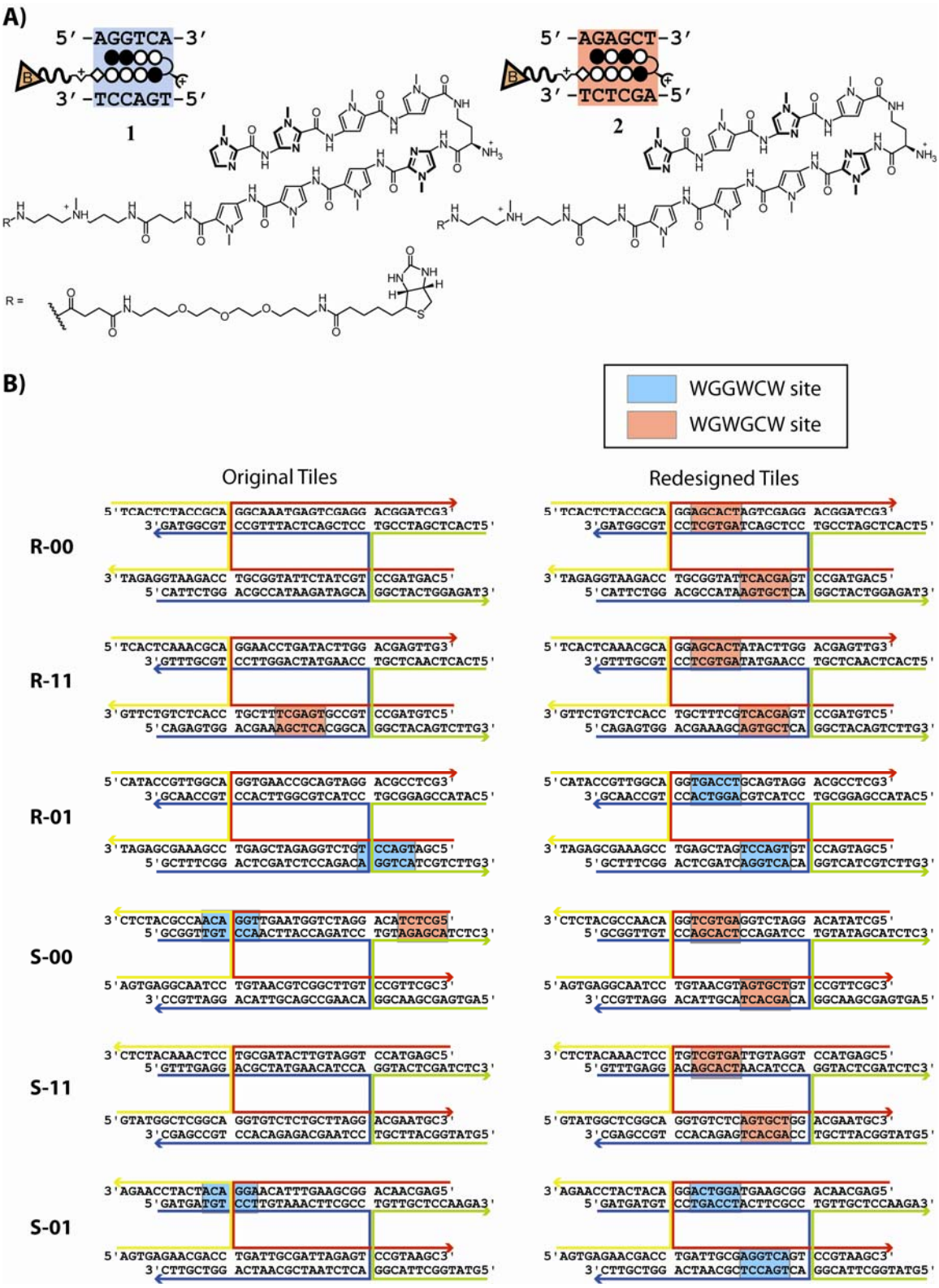

Figure A.2 Polyamide-biotin conjugates and DX tiles for algorithmic assembly. A) Two polyamide-biotin conjugates were designed to be able to visualize either the "1" or "0" tiles in the array. B) The original tiles

were redesigned to contain two binding sites for a single polyamide. Tiles that output a "1" have sites for **1** while tiles that output a "0" have sites for **2**.

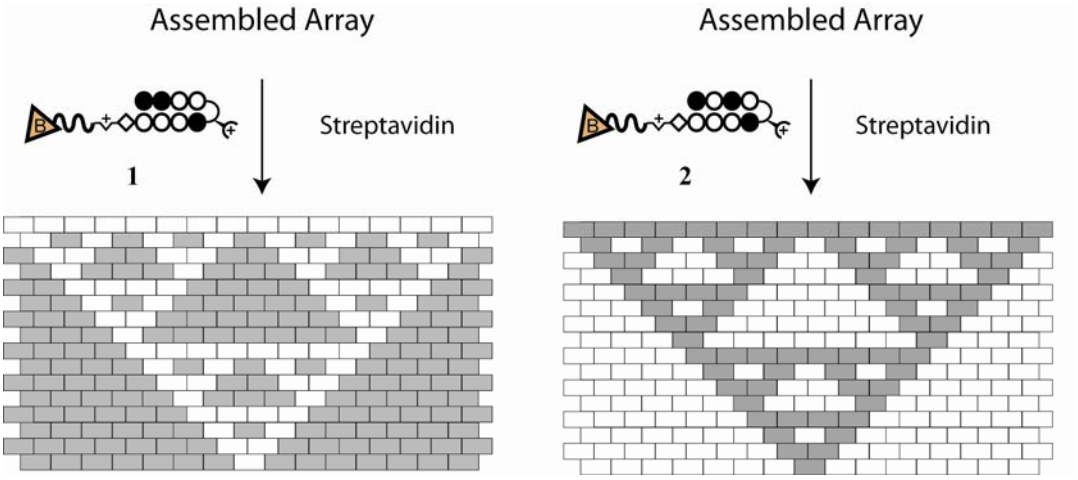

Figure A.3 Labeling an array formed by algorithmic self-assembly with polyamide-biotin conjugates. After the array is formed it is treated with streptavidin and a polyamide-biotin conjugate targeted to a subset of tiles, either those that output a 1 or a 0. The height difference between the unmodified tile and the tiles to which streptavidin has been recruited is visible by AFM.

AFM Imaging

The annealed array was incubated with polyamide **1** and streptavidin. AFM images of the assembly are shown in Figure A.4. Although somewhat difficult to distinguish, it appears as though the expected Sierpinski triangle pattern is visible. However, the large amount of labeling observed exceeds what would be expected and makes absolute determinations of the binding pattern difficult. One possible explanation is that the polyamide is binding to a second mismatch site or nonspecifically to tiles that it should not. However, this should be unlikely considering the design of the tiles and the efforts taken to eliminate potential mismatch sites. Another potential explanation is that the nucleating DNA strand contains too many initiation sites that are resulting in the formation of overlapping triangles. This is consistent with the large number of binding events observed near the edges of the arrays. Finally, a large number of structures

corresponding to “spontaneous nucleation” are visual as well. These structures, rather than forming in a controlled manner from an input template strand result from spontaneous crystal growth. The arrays from this sort of growth result in a seemingly random pattern although the tiles may still obey the programmed binding rules.

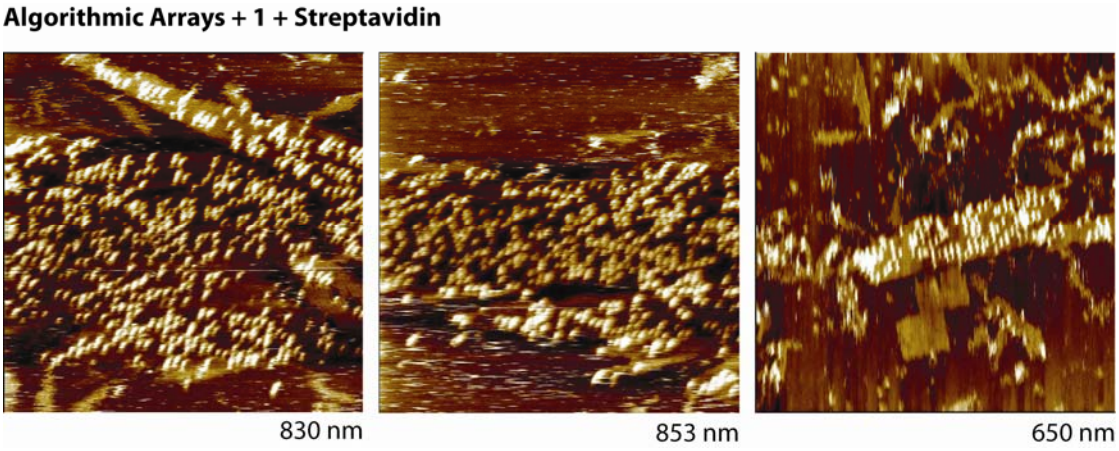

Figure A.4 AFM images of algorithmic arrays labeled with polyamide **1**. Images on the left and middle were taken after the arrays were incubated with 200 nM **1.** 2.5 µL of sample was spotted and 2 µL of 1 µM streptavidin was added 1 min prior to imaging. The image on the far right was taken of annealed arrays that were incubated with 100 nM **1** and 200 nM streptavidin prior to being spotted on mica.

A.3 Conclusions

This initial study into polyamide use for reducing error rates in algorithmic assemblies demonstrated several challenges that will need to be met before this methodology could become practical. First, extremely high levels of labeling are essential, as any site which is unbound by polyamide will result in an apparent error, despite successful incorporation of the correct tile into the array. Similarly, polyamide binding must be specific for one tile versus the other. If a polyamide binds to the wrong tile, this will again result in an apparent error, despite successful incorporation of the correct tile into the growing array. For this method to have a significant advantage over the use of dsDNA hairpins, polyamide labeling would likely have to approach >95%

efficiency and specificity, as the error rates reported in previous arrays have been between 1 – 10%.[5, 6] Based upon the initial observed results in this study, as well as those in Chapter 2 and 3, it is unlikely that the current generation of pyrrole-imidazole polyamides are capable of meeting these stringent requirements.

It should be noted that one of the limitations in this study was the input strand used to generate the assemblies. The process used to generate the input strands uses a mixture of strands present in different concentrations to limit the incorporation of initiation sites. (See ref. 6, Supporting Information). The observation of a large number of spontaneous growth arrays as well as what appeared to be multiple initiation sites in the templated arrays indicates that the quality of the input strand may have hindered the growth of high quality arrays. It is worth noting that more recent studies have switched to the use of DNA origami as a programmable input for crystal growth.[9] Reported error rates in these assemblies were just 1.4% per tile. Although polyamide-biotin conjugates may eventually find use for the labeling purposes outlined here, it is more likely that error-correcting assemblies and the evolving platforms for algorithmic assemblies may effectively eliminate the need for error correction in this fashion. Despite this, the results of this study highlight the continued need for the generation of increasingly high affinity polyamides with improved selectivity. With the recent development of the (R)–3,4-Diaminobutyric Acid turn unit that has been shown to have increased affinities for DNA,[10] labeling efficiencies closer to what are required for these challenging applications may soon be within reach.

A.4 Experimental Details

Abbreviations. Trishydroxymethylaminomethane (TRIS), ethylenediaminetetraacetic acid (EDTA).

Materials. Boc-*β*-Ala-PAM resin was purchased from Peptides International. Trifluoroacetic acid (TFA) was purchased from Halocarbon. Methylene Chloride (DCM) was obtained from Fisher Scientific and *N,N*-dimethylformamide (DMF) from EMD. EZ-Link TFP-PEO_3-Biotin was purchased from Pierce. Streptavidin was purchased from Rockland. DNA oligomers were purchased from IDT Technologies. Ultra Pure TRIS was purchased from ICN. Magnesium Acetate 4-hydrate was obtained from J.T. Baker. Water (18 MΩ) was purified using a aquaMAX-Ultra water purification system. Biological experiments were performed using Ultrapure Water (DNase/RNase free) purchased from USB. The pH of buffers was adjusted using a Thermo Orion 310 PerpHect Meter. All buffers were sterilized by filtration through a Nalgene 0.2 μm cellulose nitrate filtration.

Polyamide Synthesis. Polyamide monomers were prepared as described previously.[11] Synthesis was performed using established protocols and all polyamides were characterized by MALDI-TOF and analytical HPLC. The synthesis of polyamides **1** and **2** is described in section 2.4 and in the literature.[8]

DNA Oligomers. The following DNA oligomers, described previously,[6] were used for the assembly of the nucleation strand.

cpBr1: 5'-GTTGATGGAGTATAGTGTATTGGATGAAATGTTATGT-3'

A1S: 5'-TCACTGCTGAAGGCAGAGGACTGTGCTGGACTTGGTC-3'

A2: 5'-TGGTAATGTAAGGACCTCTGCCTTCAGC-3'

A4SV: 5'-CATACGACCAAGTGGATTTGTAGGAT-3'

A4_S00: 5'-TCACTGACCAAGTGGATTTGTAGGAT-3'

A3_nick: 5'-GGTTGAATGACCAGCACAGT-3'

Annealing Algorithmic Assembly Arrays. The long input scaffold strand was made using the published protocol.[6] Individual tiles were annealed in TAEMg Buffer (40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 1 mM EDTA, 12.5 mM magnesium acetate) at a final concentration of 1 μM, heated to 90°C and allowed to cool to RT over several hours. The array was created by mixing the input scaffold strand (~.004 μM), Strands A1S (.2 μM), A2 (.2 μM), A3-nick (.2 μM), A4-S00 (.2 μM), cpBr1 (.2 μM), A4SV (.002 μM), and each of the tiles (.2 μM) in TAEMg buffer at 90°C and allowing it to cool to RT over several hours.

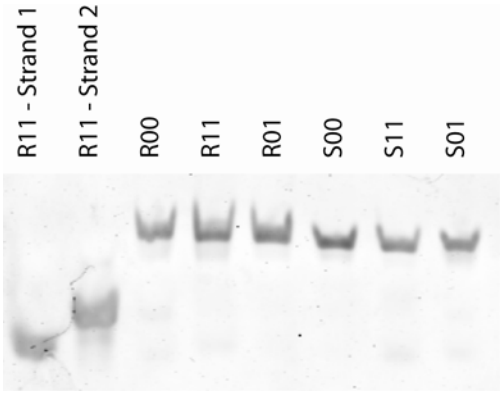

Figure A.5 Non-denaturing PAGE of the redesigned DX tiles for algorithmic assembly. As a control, the ssDNA component strands from tile R11 were run in the first two lanes. Clean formation of each tile is observed.

AFM Sample Preparation. The annealed origami was diluted 1:100 and 10 μL and incubated with polyamide (200 nM). 5 μL of sample was spotted on the mica, and 2 μL of 1 μM streptavidin was added for 1 min prior to imaging.

A.5 References

1. Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide *Science* **2008,** 321, 1795-1799.

2. Seeman, N. C., DNA in a material world *Nature* **2003,** 421, 427-431.

3. Lin, C. X.; Liu, Y.; Rinker, S.; Yan, H., DNA tile based self-assembly: Building complex nanoarchitectures *ChemPhysChem* **2006,** 7, 1641-1647.

4. Seeman, N. C.; Lukeman, P. S., Nucleic acid nanostructures: bottom-up control of geometry on the nanoscale *Rep. Prog. Phys.* **2005,** 68, 237-270.

5. Barish, R. D.; Rothemund, P. W. K.; Winfree, E., Two computational primitives for algorithmic self-assembly: Copying and counting *Nano Lett.* **2005,** 5, 2586-2592.

6. Rothemund, P. W. K.; Papadakis, N.; Winfree, E., Algorithmic self-assembly of DNA Sierpinski triangles *PLoS Biol.* **2004,** 2, 2041-2053.

7. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Addressing single molecules on DNA nanostructures *Angew. Chem.-Int. Edit.* **2007,** 46, 7956-7959.

8. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Programming multiple protein patterns on a single DNA nanostructure *J. Am. Chem. Soc.* **2008,** 130, 402-403.

9. Fujibayashi, K.; Hariadi, R.; Park, S. H.; Winfree, E.; Murata, S., Toward reliable algorithmic self-assembly of DNA tiles: A fixed-width cellular automaton pattern *Nano Lett.* **2008,** 8, 1791-1797.

10. Dose, C.; Farkas, M. E.; Chenoweth, D. M.; Dervan, P. B., Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit *J. Am. Chem. Soc.* **2008,** 130, 6859-6866.

11. Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids *J. Am. Chem. Soc.* **1996,** 118, 6141-6146.