

Beyond Watson and Crick: Programming the  
Self-Assembly and Reconfiguration of DNA  
Nanostructures Based on Stacking Interactions

Thesis by  
Sungwook Woo

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California  
2013  
(Defended May 23, 2013)



## Acknowledgements

I was extremely fortunate and privileged throughout my graduate years. Foremost, I became the first student, and for a significant time remained the only student, of my advisor Paul Rothmund. He was an exceptional mentor and a great friend. We had a lot of fun together—making a stop-motion movie of DNA origami on a sunny Californian beach, creating a trillion ROSENs together, swimming and snorkeling in a beautiful lake in New Hampshire, and dealing with my lawsuit (I can say it was fun now that it is over). But my best memories are of our scientific discussions, which were often 6-7 straight hours spent arguing about things, developing new ideas, and, when writing together, competing to see who was the real perfectionist. I will miss those intellectually challenging moments the most. My deepest gratitude goes to Paul for his excellent guidance, patience, stimulation, and encouragement.

I was also very lucky to be part of the greater Winfree lab. I am very grateful to Erik Winfree for all his guidance and encouragement since we first met and chatted in Seoul back in 2006 (and even before our Seoul meeting through emails). Being within a group co-led by the two geniuses was a one-of-a-kind experience. I learned rigor, persistence, and ways to approach difficult problems from them. Many thanks also go to the rest of the group—Lulu Qian, Damien Woods, Ashwin Gopinath, Dave Doty, Jongmin Kim, Niranjan Srinivas, Constantine Evans, as well as visiting members Cody Geary, Denis Selnihin, Pierre-Etienne Meunier, Craig Martin, Anu Thubagere and past members, Rizal Hariadi, Nadine Dabby, Si-ping Han, Elisa Franco, Dave Zhang, David Soloveichik, Joseph Schaeffer, Hareem Maune, Ho-Lin Chen, Rebecca Schulman, Georg Seelig, Bernie Yurke, Peng Yin, Kevin Young, Seung Woo Shin, Matt Cook, Mikhail Hanewich-Hollatz, and Jerzy Szablowski—for their friendship, constructive scientific comments, and for keeping the lab a stimulating and fun place. I also thank Lucinda Acosta and Karolyn Knoll for always being helpful with ordering supplies and for being good friends.

I was very fortunate to have Niles Pierce as a committee member—he always asked critical and thought-provoking questions. Niles was also a very supportive and helpful option chair, and an exceptional teacher. I am also grateful to David Tirrell for being on my committee (we went through an earthquake together in the middle of my qualifying exam!) and for his interesting and inspiring questions. Near the end of my graduate years, I was very lucky to have interacted with two great scientists—Deborah Fygenon and Ebbe Andersen—who were visiting scholars in the lab at the same time. I thank them for carefully reading my thesis and giving me insightful comments.

I am grateful to my friends in BE and BMB—Bill Dempsey, Alborz Mahdavi, Jonathan

Sternberg, Nakul Reddy, Niema Pahlevan, Indira Wu, and Geoffrey Lovely—for their friendship and help during all the hard times we experienced working on problem sets and studying for quals together. I also thank Linda Scott for her help with many administrative tasks during my years in graduate school.

I also thank my friends in the Korean community at Caltech—Seokmin Jeon, Junho Suh, Hansuek Lee, Sinchul Yeom, Young Shik Shin, Hee-Joong Chung, Min Seok Jang, Kiyoul Yang, Dong-Wook Kim, Myoung-Gyun Suh, Taesik Oh, Daejung Kwon, Young Hyun Kim, Juyoung Kim, Haekong Kim, Ji Hun Kim, In Ho Cho, Sanghee Park, Wonhee Lee, Jina Choi, Chang Ho Sohn, Dongwan Kim, Hugh Kim, Sang-Mok Lee, Daegyoun Kim—just to name a few—who were great mentors, brothers (and a sister), and life companions.

I am grateful to Chris Rivera for his help on my project with proteins. I also thank Ethan Garner and Jessica Polka, for sharing their expertise and adding the final pinch of spices that were necessary to the soup of my experiments.

I thank Jack and Robyn Lutz, Sergii Romanenko, Michelle Bobrow, and Dave Doty (again), for joining my class “DNA nanotechnology for biologists”, putting up with my insufficient teaching skills and actively participating in discussions.

I thank Caltech and the Benjamin M. Rosen Family Foundation for generous scholarships which funded my graduate study.

From my life before Caltech, I owe a great debt to Professors Joon Won Park and Hee Cheul Choi in Pohang—for their guidance and encouragement, and for teaching me the joy of research. I also thank my parents, Hansub Woo, and Hyeongseon Cho, and my parents-in-law, Jongki Lee and Sounok Lee for their endless support and trust.

Finally, and above all, I am sincerely grateful to my family. I first thank my wife, Soyoun, for her unlimited support, love and sacrifice, and for all the lunch boxes that she made (and often brought to campus) for me. Without her help, I would not have been able to complete my thesis. I also thank my son, Erik, for helping me survive at Caltech from the very moment of his birth—he waited one week past his expected due date until I finished my finals at the end of my first year—and for being my best friend ever since! I dedicate this thesis to Soyoun and Erik.

## Abstract

Life is the result of the execution of molecular programs: like how an embryo is fated to become a human or a whale, or how a person's appearance is inherited from their parents, many biological phenomena are governed by genetic programs written in DNA molecules. At the core of such programs is the highly reliable base pairing interaction between nucleic acids. DNA nanotechnology exploits the programming power of DNA to build artificial nanostructures, molecular computers, and nanomachines. In particular, DNA origami—which is a simple yet versatile technique that allows one to create various nanoscale shapes and patterns—is at the heart of the technology. In this thesis, I describe the development of programmable self-assembly and reconfiguration of DNA origami nanostructures based on a unique strategy: rather than relying on Watson-Crick base pairing, we developed programmable bonds via the geometric arrangement of stacking interactions, which we termed *stacking bonds*. We further demonstrated that such bonds can be dynamically reconfigurable.

The first part of this thesis describes the design and implementation of stacking bonds. Our work addresses the fundamental question of whether one can create diverse bond types out of a single kind of attractive interaction—a question first posed implicitly by Francis Crick while seeking a deeper understanding of the origin of life and primitive genetic code. For the creation of multiple specific bonds, we used two different approaches: binary coding and shape coding of geometric arrangement of stacking interaction units, which are called blunt ends. To construct a bond space for each approach, we performed a systematic search using a computer algorithm. We used orthogonal bonds to experimentally implement the connection of five distinct DNA origami nanostructures. We also programmed the bonds to control cis/trans configuration between asymmetric nanostructures.

The second part of this thesis describes the large-scale self-assembly of DNA origami into two-dimensional checkerboard-pattern crystals via surface diffusion. We developed a protocol where the diffusion of DNA origami occurs on a substrate and is dynamically controlled by changing the cationic condition of the system. We used stacking interactions to mediate connections between the origami, because of their potential for reconfiguring during the assembly process. Assembling DNA nanostructures directly on substrate surfaces can benefit nano/microfabrication processes by eliminating a pattern transfer step. At the same time, the use of DNA origami allows high complexity and unique addressability with six-nanometer resolution within each structural unit.

The third part of this thesis describes the use of stacking bonds as dynamically breakable bonds. To break the bonds, we used biological machinery called the ParMRC system extracted from bacteria. The system ensures that, when a cell divides, each daughter cell gets one copy of the cell's DNA by actively pushing each copy to the opposite poles of the cell. We demonstrate dynamically expandable nanostructures, which makes stacking bonds a promising candidate for reconfigurable connectors for nanoscale machine parts.

# Table of Contents

<b>Acknowledgements</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>v</b>
<b>Table of Contents</b> .....	<b>vi</b>
<b>I. Introduction</b> .....	<b>1</b>
I.1. DNA Nanotechnology .....	1
I.2. Towards the Development of Nanomachine Connectors.....	3
I.3. Survey of Studies on Base Stacking .....	5
I.4. Thesis Outline and Contributions .....	7
<b>II. Design and Implementation of Stacking Bonds</b> .....	<b>10</b>
II.1. Introduction .....	10
II.2. Design, Results and Discussion.....	12
II.2.1. Stacking of origami rectangles .....	12
II.2.1.1. DNA sequence at blunt ends .....	12
II.2.1.2. Global twist.....	13
II.2.1.3. Crossover geometry at blunt ends.....	14
II.2.2. Two approaches towards specific stacking bonds.....	19
II.2.2.1. Binary codes.....	19
II.2.2.2. Shape coding .....	30
II.2.2.3. Summary and comparison of the two approaches.....	38
II.2.3. Programming of cis-trans isomeric structures .....	41
II.2.4. Thermodynamic measurements .....	43
II.2.4.1. Measurement of the free energy of stacking bonds .....	43
II.2.4.2. First energy model: assuming loop-loop interactions are neutral.....	44
II.2.4.3. Second energy model: fitting with non-zero loop-loop interactions .....	47
II.3. Conclusion and Future Directions .....	49
II.4. Materials and Methods .....	52
II.4.1. General.....	52
II.4.2. Design of origami.....	52
II.4.3. Preparation of origami .....	53
II.4.4. Atomic force microscopy (AFM).....	54
II.4.5. Detail aspects in the design process.....	54
II.4.5.1. Searching for large orthogonal sets of sequences.....	54
II.4.5.2. Warnings: potential technical issues in designing DNA origami.....	56
II.4.6. Supplementary materials .....	60
<b>III. Two-Dimensional Crystallization of DNA Origami Checkerboards via Cation-Controlled Surface Diffusion</b> .....	<b>62</b>
III.1. Introduction .....	62
III.2. Results and Discussion.....	65
III.2.1. Two-dimensional assembly by surface diffusion .....	65

III.2.2. Factors that allow two-dimensional assembly .....	66	
III.2.3. Why checkerboards, not linear chains? .....	68	
III.2.4. Effects of the strength of stacking interaction .....	70	
III.2.5. Effects of the concentration of divalent cation.....	72	
III.2.6. Crystal size is limited by irregularity of mica.....	74	
III.3. Conclusion and Future Directions.....	76	
III.4. Materials and Methods .....	77	
III.4.1. Surface diffusion assembly .....	77	
III.4.2. Atomic force microscopy (AFM).....	77	
<b>IV. ParMRC and Expandable DNA Nanostructures.....</b>	<b>78</b>	
IV.1. Introduction .....	78	
IV.1.1. Using DNA origami to study biophysical questions .....	79	
IV.1.1.1. The ParMRC system.....	79	
IV.1.1.2. Open questions .....	81	
IV.1.1.3. DNA origami design: proposed and actual .....	83	
IV.1.2. Building expandable structures using growing biopolymers.....	86	
IV.2. Results and Discussion.....	88	
IV.2.1. Biophysical studies of the ParMRC system.....	88	
IV.2.1.1. AFM studies of ParR binding to <i>parC</i> on origami.....	88	
IV.2.1.2. Origami with multiple <i>parC</i> strands .....	95	
IV.2.2. Expandable nanostructures .....	101	
IV.2.2.1. Model system: origami as ParR/ <i>parC</i> cores stabilizing ParM filaments	101	
IV.2.2.2. Expanding origami chains.....	108	
IV.2.2.3. Control system: non-expanding DNA nanotubes .....	118	
IV.3. Conclusion and Future Directions .....	121	
IV.3.1. Expandable triangles.....	122	
IV.3.2. Effects of distance between ParR/ <i>parC</i> complexes on ParM filament	connection .....	123
IV.3.3. Effects of the number of <i>parC</i> sites on single origami structures on the	growth rate of ParM filament bundles .....	124
IV.3.4. Cup experiments .....	124	
IV.3.4.1. Is it just the geometry that really matters?.....	124	
IV.3.4.2. Would the active site in ParR alone stabilize filaments?.....	125	
IV.4. Materials and Methods .....	125	
IV.4.1. Preparation of DNA origami.....	125	
IV.4.1.1. DNA origami with a single <i>parC</i> .....	125	
IV.4.1.2. DNA origami with multiple <i>parC</i> strands .....	126	
IV.4.2. Preparation of microbeads with <i>parC</i> .....	127	
IV.4.3. Preparation of proteins .....	128	
IV.4.4. Preparation of ParM polymerization buffer.....	128	
IV.4.5. AFM of <i>parC</i> -origami, ParR, and ParM.....	129	
IV.4.6. Gel electrophoresis of <i>parC</i> -origami with ParR.....	129	
IV.4.7. Preparation of glass slides.....	129	
IV.4.8. Sample preparation for fluorescence microscopy.....	130	
<b>References .....</b>	<b>131</b>	

## *Chapter I*

# Introduction

*“Once we accept our limits, we go beyond them.”  
- Albert Einstein*

### **I.1. DNA Nanotechnology**

Many critical phenomena in biology are governed by the operation of molecular programs, written in the form of genetic code in DNA molecules. The programs are executed by a set of basic rules known as the central dogma: information in DNA is transcribed to RNA (by RNA polymerase), which is in turn translated into proteins (by ribosome), and then proteins perform complex tasks and determine phenotypes of the organism. There are complicated and exquisite machineries involved in each of these processes, but the software, the program itself is all – with the exception of RNA viruses, so far as is known to humans to date – in the molecule DNA.

DNA nanotechnology is a field that takes a direct shortcut from DNA to interesting and useful structures and functions without the help of protein machineries. We use the programming power of DNA to design and implement complex tasks *almost* purely out of DNA. We often borrow ideas from nature and try to mimic what nature does – from folding<sup>1</sup>, to neural networks<sup>2</sup>, to molecular walkers<sup>3-7</sup> – and we often go beyond, e.g., a smiley face<sup>1</sup> of ~100 nm scale does not exist in nature and no living organism with 4 neurons can play ‘read your mind’ games with a human<sup>2</sup>. DNA nanotechnology is in that sense a branch of synthetic biology, which aims to fabricate biological systems that do not exist in nature.

DNA nanotechnology is generally divided into two subcategories—structural DNA nanotechnology and dynamic DNA nanotechnology. Structural DNA nanotechnology deals with the design and construction of nanostructures based on the principle of DNA self-assembly. Structural information is transformed and encoded into the sequences of DNA such that in



thermodynamic equilibrium a collection of DNA molecules forms the desired structure. The focus of research has been on how one transforms information between the two domains: structure and sequence. The birth of the field originates from a very simple idea, in retrospect, that instead of two DNA strands with complementary sequences coming together and just forming a linear duplex as usual, one could design the sequences of four DNA strands such that a half of each strand has complementarity to only a half of another strand, so that they altogether form a crossbar-like immobile four-way junction<sup>8,9</sup>. By extending the ends of each duplex with short single-stranded domains<sup>8</sup>, termed ‘sticky ends’, and by introducing extended versions of the four-way junctions<sup>10</sup>, usually referred to as tiles, for rigidity and functionality, researchers have created lattices<sup>11</sup> with periodic patterns and implemented computation by structures formed by algorithmic self-assembly<sup>12-14</sup>. Then the breakthrough invention of DNA origami<sup>1</sup> changed the paradigm for methods of transformation between structure and sequence. It introduced the concept of folding: it uses a long single-stranded DNA as a scaffold and folds it into a desired shape with the help of many short strands, thereby enabling the creation of arbitrary (within limits) shapes and patterns at the ~100 nanometer scale. The technique, which was originally developed for two-dimensional shapes, was soon extended to enable the folding of complex three-dimensional shapes<sup>15-17</sup>. More recent advances in structural DNA nanotechnology include the development of a two dimensional molecular canvas<sup>18</sup> and three dimensional DNA bricks<sup>19</sup>, which significantly improved modularity<sup>20</sup> in creating shapes by using uniquely addressed single-stranded tiles.

The development and use of DNA nanostructures as structural templates for further functionalization and manipulation of other components can be largely considered within the boundary of structural DNA nanotechnology. For example, DNA origami was used to program the alignment of carbon nanotubes for the bottom-up construction of a nanoscale transistor<sup>21</sup>, to create a molecular analog of DNA chips that detect specific RNA targets<sup>22</sup>, and to monitor chemical reactions with single-molecule resolution<sup>23</sup>. DNA nanostructures have also been used as custom-designed tools for studying biological systems, e.g., as a structural template to study the effect of the distance of binding sites on multivalent ligand-protein binding<sup>24</sup>, as picture frame shaped architecture for direct observation and regulation of DNA-protein interactions<sup>25,26</sup>, and as a molecular rope for studying tug-of-war behavior between motor proteins<sup>27</sup>.

Dynamic DNA nanotechnology focuses on the change of transient states of DNA molecules, e.g., bound/unbound, rather than the structural information at the equilibrium state. While structural DNA nanotechnology tries to avoid kinetic traps for the best yield of the desired products, dynamic DNA nanotechnology takes control over kinetic traps and programs the pathways between one

metastable state and another. By exquisitely programming the conditional changes of the states of the molecules (and the rates of those changes), researchers have developed digital logic circuits<sup>28,29</sup>, DNA tweezers<sup>30</sup>, and molecular walkers<sup>3-7</sup>. One of the most widely used mechanisms for such programming is toehold-mediated strand displacement<sup>31</sup>. It is the mechanism by which a free strand of DNA and a partially-complementary duplex containing a small single-stranded region (“toehold”) can rearrange to form a duplex with greater complementarity, often displacing one of the two originally duplexed strands, without complete dissociation of the initial duplex via a random walk process called branch migration. Such reactions occur isothermally at rates somewhat well understood<sup>32</sup>. By embedding (often sequestering) toeholds and their complements in multiple molecular species, multi-level downstream reactions can be programmed to execute sequentially.

Naturally, there have been attempts to combine these two aspects of DNA nanotechnology and create nanostructures that conditionally change their configuration in response to input of other DNA strands. Yan et al. developed a tile-based nanodevice that switches between two states and rotates the structure attached to it by removal and addition of external strands<sup>33</sup>. Andersen et al. created a three-dimensional box that opens its lid upon input of key strands<sup>15</sup>. Han et al. devised a structure that can be partially cut by input of a set of strands and changes its configuration<sup>34</sup>. Douglas et al. developed a clamshell-like container that can hold molecular payloads inside the container and opens conditionally by the binding of specific antigens from cells to release the payloads<sup>35</sup>.

## **I.2. Towards the Development of Nanomachine Connectors**

Although there have been developments of dynamic nanostructures by combining structural and dynamic aspects of DNA nanotechnology, the level of complexity and exquisiteness is far from what nature has. For example, the bacterial flagellum that allows some bacteria, e.g., *E. coli*, to swim around the environment is driven at its base by a reversible rotary motor<sup>36</sup> (Figure I-1a). The motor is about 45 nm in diameter and is assembled from about 20 different kinds of parts. Some parts are embedded stationary in the cell membrane, while allowing other parts to freely rotate against them.

While we have the capability of creating almost any shape with DNA and of changing it conditionally, we seem to need much more in terms of functionality, connection methods, and interaction mechanisms between components. For example, while we need connectors that strongly hold components together, we would also need some kinds of connectors that allow the initial

assembly but are easily breakable by forces from neighboring components. We would also need some connectors that first initiate the assembly, then conditionally rotate by some mechanical interactions while maintaining the contact with the next components (see Figure I-1b). Chemically synthesized mechanically interlocked molecular architectures, such as rotaxanes and catenanes, have been used to create molecular shuttles or rotary motors that operate by chemical or photochemical stimuli<sup>37</sup>. However, for structural control and programmability of individual components over wide range of length scales, DNA is the most convenient and promising medium for designing nanomachines.

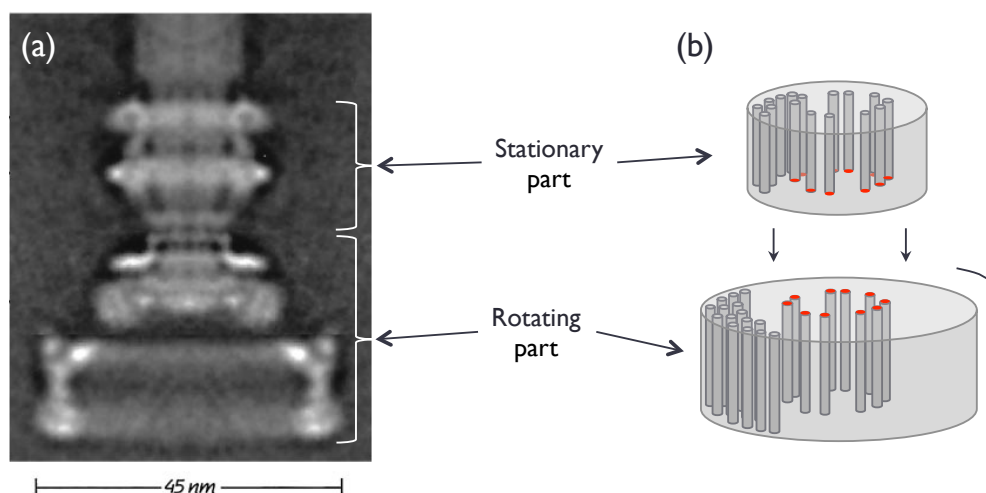


Figure I-1. **The rotary motor of bacterial flagella and stacking-bond-based rotation parts.** (a) An electron microscope (EM) image of the rotary part of a bacterial flagellum adopted from ref. 36. (b) A conceptual drawing of rotating machine parts using stacking bonds as the connector. Stacking bonds may be used as connectors that first guide specific self-assembly between parts, and then allow rotation by sliding past each other while maintaining contact, under external shear forces.

In this thesis we describe the development of specific and programmable bonds between DNA nanostructures based solely on stacking interactions, as potential connectors that are easily reconfigurable by forces. We achieve programmability by geometric alignment of the stacking interaction units, blunt ends—the termini of a DNA duplex. We believe that such bonds, termed *stacking bonds*, can achieve those functions described above. Stacking bonds only involve face-to-face attraction of blunt ends, unlike bonds based on base-pairing interactions that would require structural interdigitation of the connecting strands. We believe that stacking bonds would require much smaller forces to break than base-pairing bonds with similar energy, especially when the force is applied in the shear direction, because the potential stepwise breakage would require much smaller force for each step than the case of breaking the base pairs all at once after elastic extension.

Figure I-2 shows schematic force-distance curves adapted from the literature for the cases of pulling a DNA duplex in the axial direction<sup>38</sup> (Figure I-2a) and of “unzipping” a DNA duplex<sup>39</sup> (Figure I-2b). While a DNA duplex undergoes elastic extension and abrupt rupture when pulled in the axial direction, when unzipped, a DNA duplex exhibits a sawtooth-like curve involving stepwise breakage of base pairs with smaller force for each step. We hypothesize that the breakage of stacking bonds under shear force would exhibit a similar behavior to the case of unzipping a DNA duplex.

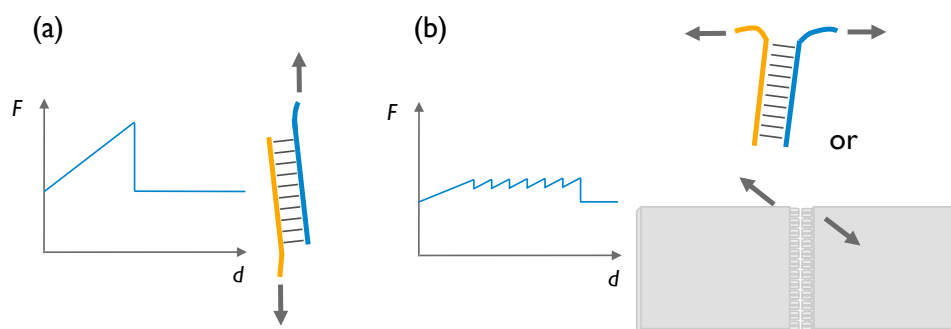


Figure I-2. **Schematic force-distance curves**, for the case of (a) a DNA duplex when pulled in the axial direction (adapted from ref. 38), and for the cases of (b) a DNA duplex when “unzipped” (adapted from ref. 39) and a pair of DNA origami nanostructures connected by stacking bonds when pulled in the shear direction (hypothetical). The direction of applied force is represented by arrows in each case.

With a similar mechanism of stepwise rearrangement requiring small force for each step, stacking bonds may serve as connectors that allow rotation while maintaining the contact—the kind of challenge described above, as inspired by nature’s nanomachine bacterial flagellum. The alignment of blunt ends on the same surface and the face-to-face configuration of stacking bonds would give a ‘smooth’ bond surface and would allow *sliding* of the bonded surfaces against each other.

This thesis represents a primitive step towards the development of such connectors for molecular machines. We present the design and implementation of stacking bonds as programmable and chemically stable bonds. We further demonstrate dynamic separation of stacking bonds that once connected DNA origami nanostructures by exploiting a force-generating filament protein system.

### I.3. Survey of Studies on Base Stacking

While the base pairing interaction of DNA seems to be well understood, it is hard to separate

the relative contributions of base stacking and hydrogen bonding. Theoretical and computational investigations on the chemical nature of base stacking have suggested that the interaction is mainly governed by dispersion attraction (instantaneous dipole – induced dipole interaction)<sup>40,41</sup>. But such studies are usually performed assuming a gas phase, i.e., no molecules other than DNA, to eliminate any potential interaction from other species, especially water, and characterize the ‘intrinsic’ energies. In real systems, in which DNA is usually surrounded by water molecules and ions, the base stacking interaction would have to be considered as a combination of multiple interactions, such as the intrinsic dispersion interaction, the hydrophobic interaction (an entropic effect for water molecules), and the ion-induced dipole-dipole interaction.

Experimentally, the base stacking interaction, independent of hydrogen bonding, can be observed by examining the binding between blunt ends. Blunt-end stacking between individual duplexes is negligible at the concentrations typically used in biochemistry experiments, but at extremely high DNA concentrations (~100 mM), it causes short DNA duplexes to connect end-to-end into long chains, which align to form liquid crystals<sup>42</sup>. DNA origami enabled the observation of blunt-end stacking at a much lower concentration (~1 nM)<sup>1</sup>, because each origami nanostructure carried multiple blunt ends in a linear arrangement along the edges that can bind cooperatively.

In the study of liquid crystals of DNA<sup>42</sup>, the energy of stacking was estimated, based on the average lengths of chains and the concentration of DNA, to be on the order of ~ -5 to -2 kcal/mol (in distilled water, ~20 °C). A molecular dynamics simulation study based on pulling simulations estimated the energy to be ~ -6 kcal/mol (1 M DNA in 120 mM NaCl, two DNA molecules in a box)<sup>43</sup>. An experimental attempt to directly measure the free energy of the stacking interaction in a condition more relevant to biochemical systems (Tris-Borate-EDTA buffer, pH 8.0, 37 °C) was carried out using a short duplex fragment that contained a nick (a break in the phosphate backbone) in the middle<sup>44</sup>. The free energy of stacking was estimated based on the equilibrium ratio between the bound and unbound states of the duplex at the nicked site. Each possible pair of blunt end sequences was examined and showed the free energy values between -2.17 and -0.19 kcal/mol. We also measured the stacking energy using our stacking bonds between DNA origami, and compared the value with their values, which showed a good agreement as will be discussed in Chapter II.

Certainly we were not the first ones who tried to employ stacking interactions for construction of DNA nanostructures. Wang et al. used three-helix tiles with blunt ends on each edge to create long chains and one-dimensional arrays of streptavidin<sup>45</sup>. Endo et al. created ‘jigsaw piece’ shaped DNA origami that connected together to form larger structures, where the specificity of binding was programmed by sticky ends<sup>46</sup>. They claimed that they achieved stable connections with

the help of the stacking interaction between the edges (although strictly speaking the edges did not have blunt ends: they used four-thymine loops to prevent stacking during the assembly of each pieces). Both of these studies exploited the stacking interaction merely as a source of stable connection, but did not attempt to achieve specificity or programmability.

## I.4. Thesis Outline and Contributions

The following chapters describe our development of stacking bonds and investigation of their potential for reconfiguration. Chapter II is based on a peer-reviewed publication, and Chapters III and IV present work that is not yet published.

Chapter II describes the design and implementation of stacking bonds. We first explore the stacking phenomenon of origami rectangles and flesh out some detail aspects and improve the performance of stacking between origami. We recognize the presence of the global twist in the original 2D origami, explain its origin, and create twist-corrected rectangles to improve bond fidelity and stability. By investigating the effect of the crossover geometry at blunt ends, we discover stacking polarity and find an optimal geometry for the creation of stacking bonds. For the design of diverse bond types, we introduce two different approaches: binary coding and shape coding of geometric arrangement of blunt ends. For each approach, we set up binding rules that derive from the stacking polarity and the nature of each approach. We define a set of design constraints, such as the number of active patches and the maximum number of allowed incorrect matches. We wrote computer programs to search and construct the bond spaces for each approach, and to explore the orthogonality between bond types. We experimentally demonstrate some of the bond types from each approach and present atomic force microscope (AFM) images for both origami chains based on single bond types and chains connected via orthogonal bond types. We further present data for stacking bonds programmed to control the cis/trans configuration between origami nanostructures. We estimate the stacking bond energy based on AFM results and we provide two different models: a nonlinear energy model with respect to the number of blunt ends on origami edges, and a linear energy model under the assumption that loops that were considered inert by design might contribute to the stability.

The original motivation for the project comes from Paul Rothmund's initial idea of programming shape matching between origami. While exploring some initial designs experimentally (and dealing with failures), we expanded our understanding of the properties of stacking *and* origami itself, e.g., global twist, edge geometry, etc., which helped us together to

develop the idea of binary coding in addition to the shape coding approach. All computer programs for the construction of bond space under various constraints for both approaches were written by me. All experiments for investigating the stacking properties between rectangle origami and for testing and characterizing the stacking bonds via binary and shape coding approaches were performed by me. Experiments with corner origami for the cis/trans configuration and the statistical analysis were performed by Paul Rothmund. All other statistical and numerical analyses for stacking bonds and the stacking energy estimation were done by me. This work was published in full as: Sungwook Woo & Paul W. K. Rothmund, “Programmable Molecular Recognition based on the Geometry of DNA nanostructures”, *Nature Chemistry*, 3, 620-627 (2011), doi: 10.1038/nchem.1070.

Chapter III describes large-scale self-assembly of DNA origami into two-dimensional checkerboard-pattern crystals based on a surface diffusion protocol. We describe our understanding of the surface diffusion phenomena and the protocol where the diffusion of DNA origami occurs on a mica substrate and is dynamically controlled by changing the cationic condition of the system:  $Mg^{2+}$  first fixes origami to a mica surface,  $Na^+$  makes them mobile while still keeping them on the surface, then  $Ni^{2+}$  fixes the resultant structure back onto the surface. The protocol also involves incubation under an elevated temperature. A series of systematic experiments suggested that  $Na^+$  and the surface are the essential factors, while heat merely facilitates the process. We found that the origami rectangles prefer the checkerboard pattern to a linear form because of their global twist and provide evidence based on control experiments carried out with twist-corrected origami. We explore the effect of the strength of the stacking bonds connecting origami on the assembly pattern: stronger stacking bonds mediated by larger contact area prevented the formation of checkerboard pattern crystals, again due to global twist. We also describe the discovery that the crystal growth on mica is limited by the nonuniform charge density of mica.

This work started and proceeded as a side project of the stacking bond project described in Chapter II. The idea of testing origami containing stacking patches at four corners was initiated by me during the early stage of investigating the stacking property of rectangle origami. The idea of testing different cation conditions emerged while chatting about the system with Si-ping Han. All the experiments for understanding the behavior of origami on surface and characterizing the effect of various factors were designed and performed by me. This work has a manuscript in its draft form to be submitted soon.

Chapter IV describes our studies of bacteria’s machinery called the ParMRC system from two perspectives. The ParMRC system ensures that, when a cell divides, each daughter cell gets one

copy of the cell's DNA by actively pushing each copy to the opposite poles of the cell. One perspective is to use DNA origami as a custom tool to study some unknown biophysical aspects of the machinery at the single-molecule level. We present a series of AFM studies that track the binding of components of the system on top of our origami template, and compare the results quantitatively with previously reported data from ensemble experiments. The second perspective we take is to exploit the dynamic property of the ParMRC system to engineer expandable nanostructures: we break stacking bonds between DNA origami by anchoring the binding sites of the protein system on each origami. We estimate the forces relevant to breaking the bonds with the protein system, and show preliminary fluorescence microscopy data for expanded structures, along with data from control experiments using DNA nanotubes.

This project initiated from discussion between Paul Rothemund and Dyche Mullins, who studies the protein system and wanted to use the programmability of DNA origami. The basic plan for the project was developed mostly by Paul Rothemund, and I participated in discussions at the planning stage. Proteins were provided by the Dyche Mullins lab, with the help of Chris Rivera. Most experimental details were designed by me, and all experiments were performed by me. All data presented in Chapter IV are unpublished data.



## Chapter II

# Design and Implementation of Stacking Bonds<sup>†</sup>

*“So beautiful it has to be true.”  
- James D. Watson*

### II.1. Introduction

DNA base pairing is unique among mechanisms for molecular recognition<sup>47</sup> since it allows the creation of *combinatorially diverse bonds*: the number of possible binding interactions scales exponentially with polymer length so that a polymer of just  $N$  nucleotides A, T, G, or C can express on the order of  $4^N$  binding interactions. In practice, many sequences are unusable, exhibiting undesired interactions with themselves or other sequences. But with attention to base composition, secondary structure, sequence homology, and thermodynamics, extremely large sets of orthogonal sequences ( $\sim 240,000$ ;  $N=25$ ) can be designed to bind within a narrow range of melting temperatures<sup>48</sup>. This ability to instantiate diverse binding interactions has driven the creation of a wide variety of DNA nanostructures<sup>1,3,8,11,16,49</sup>, molecular computers<sup>28,50,51</sup>, and nanomachines<sup>4,30,52</sup>.

While DNA hybridization is extremely powerful, researchers have developed numerous artificial systems for molecular and macroscopic recognition using simple interactions including hydrogen bonds<sup>53</sup>,  $\pi$ - $\pi$  stacking<sup>54,55</sup>, entropic depletion<sup>56,57</sup>, and capillary forces<sup>58,59</sup>. There has been little development, however, of general systems capable of creating combinatorially diverse

---

<sup>†</sup>This chapter is based on a peer-reviewed publication: Sungwook Woo & Paul W. K. Rothmund, “Programmable Molecular Recognition based on the Geometry of DNA nanostructures”, *Nature Chemistry*, 3, 620-627 (2011), doi: 10.1038/nchem.1070.

bonds; each new type of interaction typically requires unique design insights and considerable synthetic effort. Besides DNA, other information-bearing heteropolymers like proteins might be used for combinatorially diverse bonds. Yet even for well-understood systems such as protein coiled-coils<sup>60</sup>, no systematic method exists for designing multiple orthogonal bonds.

Towards understanding and recreating the recognition properties of DNA it is natural to ask, “What causes two complementary DNA strands to bind?” While the first answer that may spring to mind is the hydrogen bonding between base pairs, *base stacking* between adjacent base pairs is apparently the dominant stabilizing factor in DNA binding<sup>44,61,62</sup>. The dominance of stacking is consistent with the sequence-dependent strength of DNA binding because stacking is itself sequence-dependent: the strongest stacking occurs between a GC pair and a CG pair (-2.17 kcal/mol), the weakest stacking at AT/TA (-0.19 kcal/mol)<sup>44</sup>. The nearest neighbor model for the strength of DNA binding<sup>63-65</sup> owes its success over the base-composition model (simply counting AT and GC pairs) to the fact that it includes stacking. One consistent picture is this: stacking provides much of the binding enthalpy for the duplex, and base pairing enforces specificity.

The question arises whether stacking alone can be used to create specific molecular recognition. Pure base stacking can be studied by examining the binding between *blunt ends*, the termini of a DNA duplex. Blunt-end stacking between individual duplexes is directly observed only at extremely high (~100 mM) concentrations<sup>42</sup>, but DNA nanostructures such as 3-helix tiles<sup>45</sup> and 24-helix origami<sup>1</sup> form long chains at much lower (500 nM and 1 nM) concentrations due to cooperative binding of multiple blunt ends. In achieving specific bonds using stacking, one might expect some difficulties. First, in contrast to base pairing which involves heterophilic (“like-unlike”) attraction, blunt-end stacking involves homophilic (“like-like”) attraction. Whether combinatorially diverse bonds in homophilic polymers are even possible has been an open question since Crick first suggested that the primitive genetic code might have been homophilic<sup>66</sup>. Second, DNA uses two types of attraction (A-T and G-C), so how can combinatorially diverse bonds be constructed from just a single type of attraction?

Here we present two general approaches to programming combinatorially diverse and specific bonds in a scalable fashion from a single homophilic interaction. Both approaches use DNA origami to create *stacking bonds*—geometric arrangements of stacked blunt ends. We map these arrangements onto abstract sequences, use computer algorithms to find sets of sequences predicted to be orthogonal, and experimentally test sets for orthogonality. The first approach encodes *bond type* using a 16-bit binary code along the edges of an origami rectangle. It is *reprogrammable* and *efficient*: a single set of 16 strands can create  $2^{16}=65,536$  bond types. The

second approach encodes bond type using geometric complementarity between pairs of “Manhattan skyline” edge shapes, with  $N$  “skyscrapers” each having one of  $M$  heights; the case of  $N=4$  and  $M=3$  with  $3^4=81$  bond types is explored. This approach is not reprogrammable (each origami is “hard-coded”—a unique origami must be synthesized for each shape) but is experimentally simpler. In both systems, symmetry and mismatch constraints limit the number of bond types and size of orthogonal sets. Non-idealities like the flexibility of edges further decrease the number of usable bond types. We demonstrate the combination of origami using orthogonal sets of up to four distinct bond types. Finally, we use both systems to control the cis/trans geometry of multi-origami structures.

## II.2. Design, Results and Discussion

### II.2.1. Stacking of origami rectangles

We first explored stacking using a rectangular origami with 24 blunt ends along each edge (Figure II-1a). Approximately 200 *staple strands* (typically 32 nt) were used to fold a *scaffold strand* (~7000 nt) into the desired shape. An L-shaped pattern of *dumbbell hairpins*<sup>1</sup> was added to provide height contrast under atomic force microscopy (AFM). *Crossovers* are positions at which a strand jumps from one helix to another.

Previously<sup>1</sup>, we observed stacking of similar 24-helix rectangles into long chains (up to 5  $\mu\text{m}$ ). However, the quality of the chains was low: they exhibited complete breaks (as in Figure II-1b) or *dislocated bonds* (with edges in partial contact, as in Figure II-1e) and the bonds between origami occurred in all four possible orientations (as in Figure II-1e). We hypothesized that three factors might be responsible: (1) the sequence of blunt-end base pairs was random, (2) the origami had a large global twist, and (3) the blunt ends had a crossover geometry incompatible with B-form stacking. We reduced or eliminated all three factors, demonstrated that at least (2) and (3) were contributing to the low quality of chains, and obtained straight linear chains with only two orientations. Details of these three factors are discussed in the following sections.

#### II.2.1.1. DNA sequence at blunt ends

In principle, the strength of stacking bonds with random blunt-end sequences could vary by a factor of  $\sim 11$  (-2.17/-0.19) in the case of all ‘GC’ vs. all ‘AT’ pairs. Thus, in all experiments

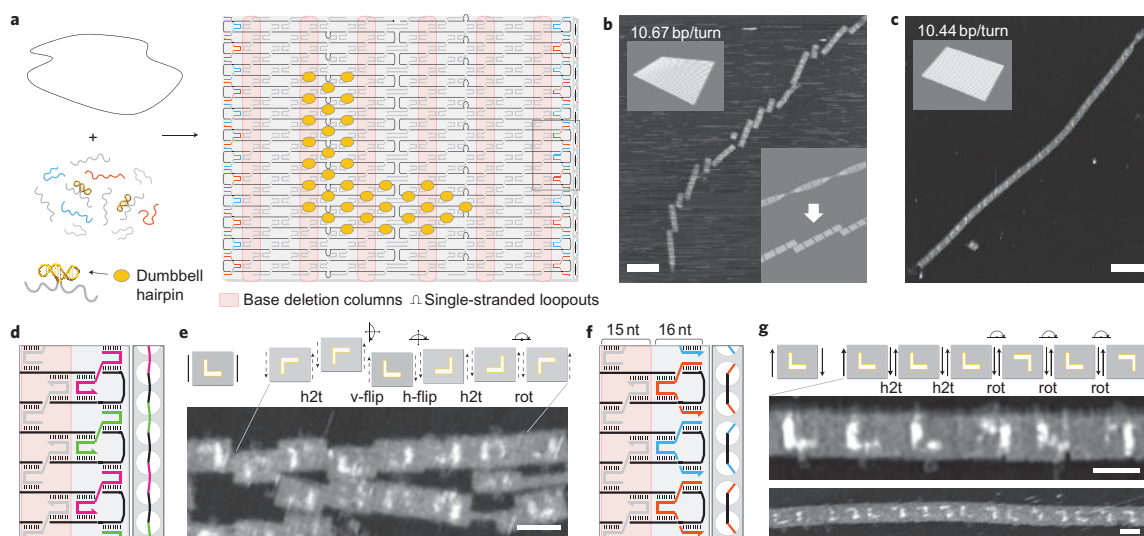


Figure II-1. **Stacking of rectangles.** (a) A long scaffold strand (black) is folded by multiple short staple strands to form a rectangle; features include edge staples (blue and red), interior staples (gray), dumbbell hairpins (orange ovals), and single-stranded loopouts (black bulges). The gray box indicates an area enlarged in (f). Each column of staples was originally 16 nt wide<sup>5</sup>; in twist-corrected rectangles, columns with base deletions (pink) are 15 nt wide. (b,c) AFM comparison of rectangle chains without (b) and with (c) twist correction, deposited on mica. Upper left insets show models of single rectangles. Lower right inset (b) models how periodic breaks arise in a twisted chain during deposition. (d) Proposed structure of a stressed edge. (e) Model and AFM of rectangles with stressed edges. Solid vertical bars indicate that no stacking polarity is expected. Dashed vertical arrows emphasize that edges do not bond in an exclusively antiparallel orientation, as exemplified by rectangles related by 180° horizontal or vertical flips (indicated by half-circle arrows with an in-plane axis of rotation). (f) Proposed structure of a relaxed edge. (g) Model and AFM of rectangles with relaxed edges. Vertical arrows label stacking polarity; only “antiparallel” bonds form. Half-circle arrows indicate 180° rotation (about an axis going into the plane through the center of a bond). Scale bars in (b,c), 500 nm; in (e,g), 100 nm.

presented here, we decreased potential variability by placing a ‘GC’ base pair at each blunt end. This was achieved by introducing *single-stranded loopouts* in the scaffold (Figure II-1a) to shift the scaffold sequence until a ‘GC’ occurred at the adjacent pair of blunt ends. A detailed procedure of generating single-stranded loopouts in the scaffold is described in Materials and Methods (see the section for “Design of origami”).

### II.2.1.2. Global twist

B-form DNA has a helical twist of 10.4 bp/turn<sup>67</sup>. The original rectangles<sup>1</sup> were designed

using a helical twist of 10.67 bp/turn, which turned out to induce a significant global twist (recently studied in detail<sup>17,68</sup>). Here, in order to achieve an average helical twist of 10.44 bp/turn, we deleted one base from every third column of staples (Figure II-1a) using our MATLAB design code or caDNAno<sup>69</sup>. Two AFM images show the difference in quality between chains formed by rectangles without (Figure II-1b) and with (Figure II-1c) twist correction, when deposited on mica. While chains of twisted origami break with a characteristic offset (with a chirality consistent with a right-handed superhelix) every 2–6 origami, chains of twist-corrected origami exhibit rare breaks. More wide-field AFM images for twisted (Figure II-3d-i) and twist-corrected (Figure II-3a-c) rectangles are shown in Figure II-3. Twisted and twist-corrected origami have factors (1), DNA sequence, and (3), crossover geometry, at blunt ends minimized by design.

### II.2.1.3. Crossover geometry at blunt ends

#### II.2.1.3.1. Stressed edges vs. relaxed edges

In general we expect deviations from B-form to weaken stacking<sup>61</sup>. The original rectangles were designed with a set of edge staples that result in a crossover at *every* available location between adjacent blunt ends (Figure II-1d). In such edges, a conflict may arise because (i) the simultaneous presence of the scaffold and staple crossovers pull the phosphates of scaffold and staples towards positions 180° away from each other, and (ii) the major-minor groove relationship of a B-form base pair would naturally place the phosphates 150° apart. Thus *stressed edges* seem incompatible with any geometry in which *all* of the blunt ends are in native form, and they might be expected to weaken or otherwise change stacking. Many different non-B-form geometries could resolve the stress at such edges, including breakage of the final base pair or a change in major/minor groove angles; accurately predicting what happens lies beyond the state of the art.

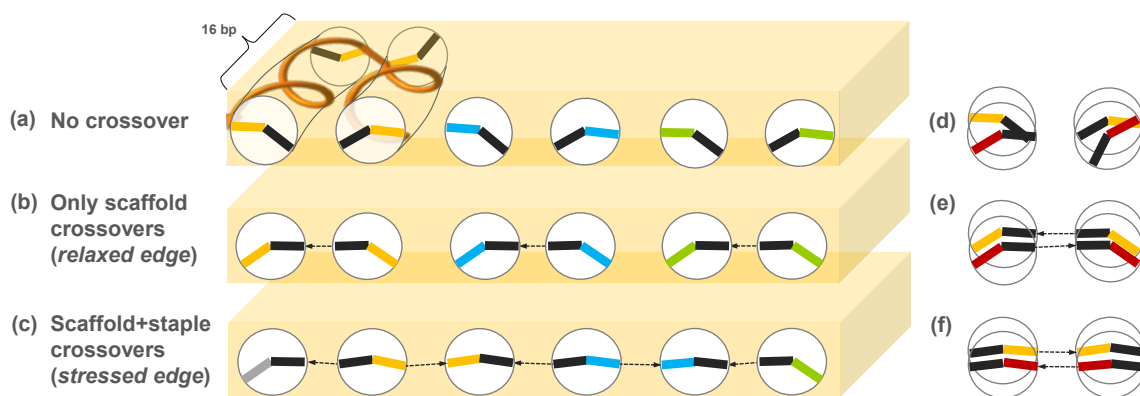
We propose that near-flattening of the major and minor grooves decreases the distinction between them (Figure II-1d) and creates a top-bottom *pseudosymmetry* that prevents stacking from exhibiting a strongly preferred orientation. Experimentally, just such a promiscuity of orientation is observed: rectangles bind in *head-to-tail* (34% of total bonds), *rotated* (44%), *horizontally flipped* (17%), and *vertically flipped* (5%) orientations ( $N=174$ , ‘h2t’, ‘rot’, ‘h-flip’, and ‘v-flip’ in Figure II-1e). Further, dislocated bonds are often observed (71%). One interpretation is that the bonds are well-aligned in solution but dislocate upon deposition. Another is that the dislocated bonds form in solution as kinetically trapped states—the stressed blunt ends might take on a corrugated geometry that, once bound, cannot easily slide to find a more stable state.

Edge staples were redesigned so that there are no staple crossovers (Figure II-1f). In such *relaxed edges* blunt ends are free to assume normal B-form groove angles. Rectangles with relaxed edges can bind via *near-B-form* stacking (see the subsection for “Models of edge structures” below) in head-to-tail and rotated orientations, but not in flipped orientations because of their strong top-bottom *asymmetry*. Our experiments are consistent with the hypothesis that near-B-form stacking is preferred; only head-to-tail (42% of total bonds) and rotated orientations (58%) are observed ( $N=318$ , Figure II-1g). Further, dislocations are exceedingly rare (1%). This is consistent with the idea that either (i) stacking bonds based on relaxed edges are stronger than those formed from stressed edges, or (ii) relaxed edges are more geometrically uniform (thus allowing any dislocated bonds that form as kinetic products to slide and become full bonds). Origami in Figure II-1e,g have factors (1), DNA sequence at blunt ends, and (2), global twist, minimized by design.

#### II.2.1.3.1.1. Models of edge structures

Because accurate models of origami edges backed by high resolution structural data do not exist, it is difficult to predict the exact structure and stacking configurations of the blunt ends on the edges of origami. Here we provide gross predictions based on the distance of the blunt ends from the nearest internal crossovers and the pattern of crossovers along the edge. We predict structures for three different edge models: (1) a *crossover-free edge* (Figure II-2a), (2) a *relaxed edge* with only scaffold crossovers (Figure II-2b), and (3) a *stressed edge* with both scaffold and staple crossovers (Figure II-2c). These predicted structures in turn imply predictions about the expected strength and behavior of the stacking bonds.

For all three models we are interested in the helical twist of the base pairs on the blunt ends at the edge, and for all three models we posit an internal crossover at a position 16 base pairs interior to the edge. Here we draw bars, separated by the major/minor groove angles, on the face of the blunt ends to indicate the helical twist of the base pair. To derive the orientation of these bars, we begin at the interior crossover and consider the strand that is “edgeward” of the crossover (e.g., the orange 3D strand in Figure II-2a). We model the two base pairs next to the crossover point as staggered up and down with respect to the midpoint of the crossover, having a helical twist angle that is rotated from the midpoint by half of the characteristic rotation/bp of B-DNA ( $\sim 34.6^\circ$  given 10.4 bp/turn)—that is approximately  $17^\circ$ . (Similar modeling is performed in ref. 70.) From these “first edgeward base pairs” the base pairs at the blunt end are 15 base pairs away. Thus the blunt end base pairs have a helical twist angle that is rotated  $\sim 519^\circ$  ( $15 \times 34.6^\circ$ ) relative to the bases of the edgeward strand in the crossover (in a clockwise direction when viewed from the blunt ends towards the crossover) for a total of  $\sim 537^\circ$  from the crossover midpoint. Given such a model, which is crossover-free at the



**Figure II-2. Comparative modeling of three different origami edge structures.** (a) *Crossover-free edges.* (b) *Relaxed edges* with only scaffold crossovers. (c) *Stressed edges* with both scaffold and staple crossovers. Each circle indicates a blunt end. In (a), both the black and colored bars (inside the circles) indicate the helical twist of bases belonging to *tile adapter strands*. In general, tile adapter strands are strands that extend from the edge of an origami to give it a geometry that is not possible using the canonical scaffold/staple geometry. Here, our intent is that the tile adapter strands create a crossover-free edge; we do not show the details and did not use them in our stacking experiments. Here we use them as part of a “thought experiment” concerning the geometry of stacking bonds, but we note that tile adapter strands have been used to create origami with a very similar crossover-free edges<sup>49</sup>. In (b) and (c), black bars indicate the helical twist of bases from the scaffold strand, and colored bars indicate the helical twist of bases from the staple strands. Bars of the same color indicate the same strand, e.g., the orange staple in (a) runs for 1.5 helical turns in one helix, switches between helices at a 16-bp-deep internal crossover, and runs back for a length of 1.5 helical turns in the adjacent helix, as depicted by the three-dimensional drawing. Black dotted arrows indicate crossovers at the edge. In all three models, colored strands are intended to make 16-bp-deep internal crossovers. The models in (a-c), predict that the blunt ends on the edge are either B-form (a), near-B-form (b), or have disrupted base pairs that are incompatible with B-form geometry (c). (d-f) show models of the juxtaposition that occurs when two different origami edges form a stacking bond; these bonding models correspond to the edge structure models in (a), (b), and (c), respectively. Models (d) and (e) make predictions for the relative helical twist between blunt ends across the bond. Model (f) suggests that the disturbance of the base pairs at the edge of the origami may decrease the distinction between the major and minor grooves enough to create a top-bottom pseudosymmetry. This pseudosymmetry could allow bonding between origami in one of the flipped orientations (not shown).

edge, the base pairs at the blunt end would be oriented like those depicted in Figure II-2a. We note that while we do not make such a structure in this work, origami with very similar crossover-free edges have been made before<sup>51</sup> using “tile adapters”, and so such structures can be experimentally synthesized.

Now consider a second origami with the same crossover-free edge structure, but with 15 base

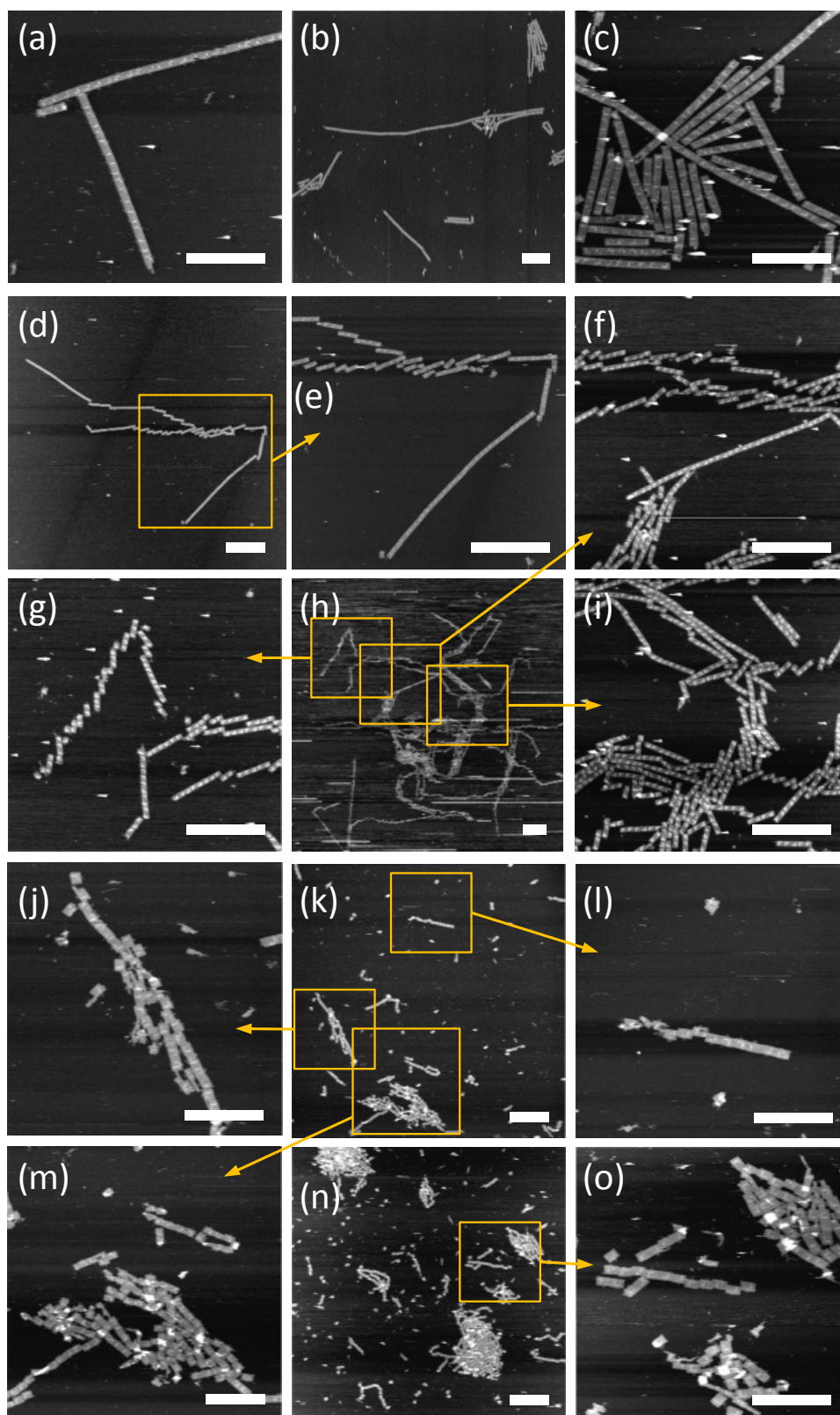
pairs between the edge and the crossover. When such an origami binds via a stacking bond to the origami described above, then the total number of base pairs between the first internal crossover points of the two origami will be  $15+16 = 31$ , or roughly 3 helical turns. This means that for such crossover-free origami, the blunt ends on opposite sides of the stacking bond are oriented with a relative twist angle of  $\sim 34.6^\circ$  (as depicted in Figure II-2d). Thus we would expect stacking of blunt ends between crossover-free edges to be native B-form stacking, and that it should be relatively strong.

Next consider our second edge structure, the relaxed edges (Figure II-2b), for which scaffold crossovers connect every other pair of helices. This is the edge structure that we use in all our work on stacking bonds, except for structures pictured in Figure II-1e and Figure II-3j-o. Because the scaffold crossovers act to pull the base pairs away from the helical twist angle that they would assume in a crossover-free edge, whatever structure forms at relaxed edges cannot be B-form DNA. However, because DNA can tolerate small deviations from B-form twist, we propose that the helices assume an amount of twist strain (roughly  $34.6^\circ$ , which is averaged over the 16 base pairs up to the crossover) and maintain native major/minor groove angles between the bases at the blunt end (depicted in Figure II-2b).

Given our model for relaxed edges, when two origami with relaxed edges bind via a stacking bond, their blunt ends will not be able to stack via B-form stacking; rather, they should bind with slightly different relative twist angles that are within approximately  $\pm 34.6^\circ$  of the natural twist angle in B-DNA (Figure II-2e). We call such stacking between relaxed edges “near-B-form stacking”, which we predict would be roughly as strong as B-form stacking. Since relaxed edges have a top-bottom asymmetry that is defined by the major and minor grooves, near-B-form stacking can only occur when two origami bind in either the head-to-tail or rotated orientations. This prediction agrees well with the distribution of observed bond orientations, as discussed above.

Finally, we consider the case of stressed edges. When staple crossovers are placed in opposition to scaffold crossovers along an edge, we propose that the balancing of the stresses they induce results in near-flattening of the major and minor grooves (Figure II-2c). The resulting decrease in distinction between the major and minor grooves should decrease the distinction between the top and bottom of the origami. Therefore, we would predict that blunt-end stacking between such stressed edges (Figure II-2f) should allow flipped bond orientations; this is, indeed, what is observed in experiments involving origami with stressed edges, such as those shown in Figure II-1e and Figure II-3j-o.





(Figure caption on next page)

Figure II-3. **Wide-field AFM images of rectangle origami systems.** (a-c) Twist-corrected origami with relaxed edges. Note that they form chains with lengths on the order of  $\sim 10$   $\mu\text{m}$ . Chains formed by these origami break in a way that suggests that the breakage occurs upon deposition since pieces lie close to each other. However, in contrast to the twisted origami shown in (d-i), twist-corrected origami break into long pieces and show no preferred direction for the shift between neighboring pieces. Note also that twist-corrected relaxed chains are straight with very rare dislocations, as opposed to the twist-corrected origami with stressed edges shown in (j-o). (d-i) Twisted origami (with relaxed edges). Chains break with a characteristic periodicity (2-6 origami) and directional offset. Note that some parts of the chains seem to unwind while depositing, especially near the ends (as suggested by the straight sections near the ends of twisted chains). (j-o) Origami with stressed edges (with twist-correction). Bonds are promiscuous: many dislocations occur and the bond orientations are random. Orange boxes and arrows show zoom-in areas. Scale bars in (a), (c), (j), (l), (m), (o) are 600 nm, and scale bars in (b), (d-i), (k), (n) are 1  $\mu\text{m}$ .

#### II.2.1.3.2. Anti-parallelism of stacking bonds

The asymmetry of relaxed edges gives stacking bonds an interesting and important property analogous to the antiparallel nature of DNA hybridization: if we label relaxed edges with arrows according to their asymmetry, we see that two edges form a bond only if the arrows point in opposite directions. Thus the arrows define an antiparallel *stacking polarity*. We label edges such that when the major grooves at an edge point up, the arrow's direction matches the 5' to 3' polarity of the scaffold at the edge (as shown in models in Figure II-1g). Stacking polarity allows stacking bonds to specify unique products by breaking the symmetry of otherwise symmetric bonds.

### II.2.2. Two approaches towards specific stacking bonds

#### II.2.2.1. Binary codes

##### II.2.2.1.1. Key concepts

Stacking between two origami edges can be largely abolished by omitting their edge staples<sup>1</sup>: each omitted staple prevents the formation of two blunt ends and leaves a 32-nt single-stranded *scaffold loop*. Sufficient complementarity between such loops could allow them to associate. For the M13mp18-based designs used so far, origami without edge staples do not aggregate, suggesting that scaffold loops may act primarily as entropic brushes with no affinity<sup>71</sup>.

The key idea, then, is to encode specificity using *stacking sequences*—binary sequences of

blunt ends and scaffold loops (abbreviated to “sequences” when clear). A stacking bond based on such sequences will have maximum strength when the sequences are aligned to maximize the number of blunt-end interactions (assuming  $\Delta G=0$  for potential loop-loop and loop-blunt end interactions; we revisit these assumptions later). We implement binary sequences by dividing edges into *patches*. Each patch is a 2-helix wide section of edge; it contains a single scaffold loop to represent ‘0’ or a doublet of blunt ends to represent ‘1’ (Figure II-4a). There is a one-to-one correspondence between each edge staple and a particular patch. Including an edge staple in the reaction mixture yields an *active patch* (‘1’) capable of stacking; omitting the same staple yields an *inactive patch* (‘0’). Here we explored 16-bit codes using a 32-helix-tall rectangle. Given the number of bits available, our goal was to construct a code for the largest set of specific bond types possible, and to use this set to connect the longest chain of distinct rectangles possible.

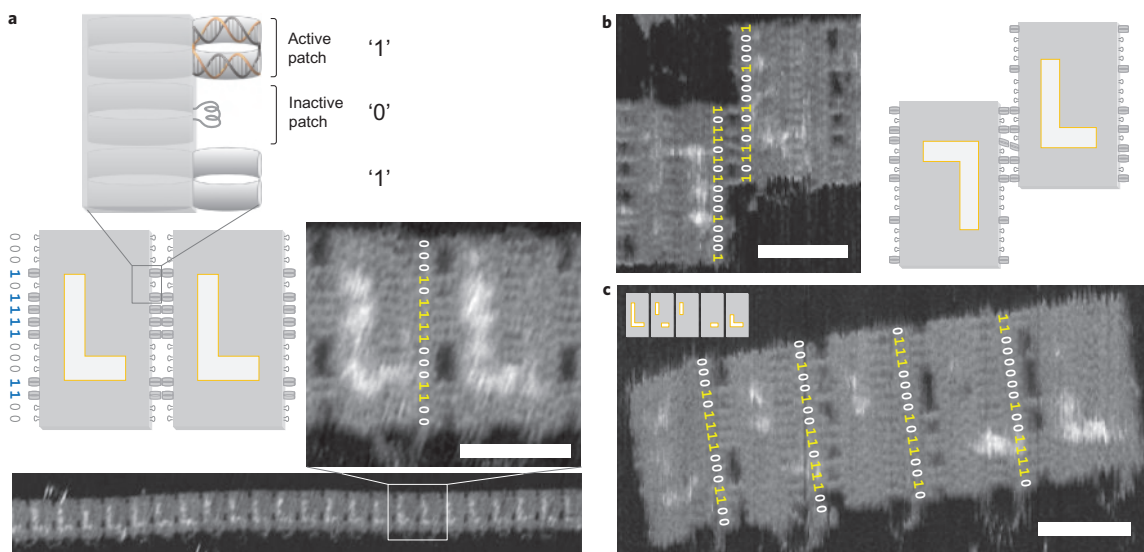


Figure II-4. **Recognition based on binary sequences of blunt ends and scaffold loops.** (a) Model and AFM images of a 32-helix tall rectangle that enables 16-bit binary codes. Addition of a staple at a specific edge site creates two blunt ends, which compose an active patch (‘1’); omission of the staple leaves a single-stranded loop that forms an inactive patch (‘0’). Use of an asymmetric sequence ‘000101110001100’ with seven active patches creates long chains with an exclusively head-to-tail orientation. Active patches can be clearly observed; each doublet of helices typically appears as a single gray bar across the bond. Scaffold loops have more variable appearance—sometimes invisible, sometimes appearing almost as prominently as an active patch—presumably due to variable conformation or potentially some loop-loop binding. (b) A bent-patch bond, a common error for binary-coded bonds. Here, helix bending allows a 5-patch bond that would otherwise not occur. (c) AFM image of a five-rectangle chain built using four orthogonal bond types. Inset shows dumbbell hairpin labels. Scale bars: 60 nm.

### II.2.2.1.2. Binding rules

To construct a code for sets of bond types, it is crucial to understand the binding rules of binary-coded stacking bonds. Stacking sequences are read in the direction of the stacking polarity. Just as the 5' to 3' polarity of DNA ensures that a DNA sequence is not, in general, equal to its *reverse* ('ATGC' ≠ 'CGTA'), stacking polarity ensures that a stacking sequence is not, in general, equal to its reverse ('1011' ≠ '1101'). But while the complementary sequence to a DNA sequence is its *reverse complement* ('ATGC' binds to 'GCAT'), the complementary sequence to a stacking sequence is simply its reverse ('1011' binds to '1101'). Another parallel distinction is that while a self-complementary DNA sequence must be a *reverse palindrome* ('AGCT' binds itself), a self-complementary stacking sequence must be a *palindrome* ('0110' binds itself). The last two differences are due to the homophilicity of blunt-end stacking, but they highlight the importance of stacking polarity as a symmetry breaker—without stacking polarity, nonpalindromic stacking sequences would also be self-complementary because an origami bearing *any* sequence could bind itself in a horizontally flipped orientation. Palindromic sequences are useful when a two-fold rotational symmetry is desired (as in Figure II-13c). If a nonpalindromic sequence, stripped of leading and lagging zeros, is still nonpalindromic (e.g., '0010110' → '1011'), it specifies a stacking bond with a unique head-to-tail orientation (Figure II-4a) and is *uniquely orienting*.

Like DNA sequences, stacking sequences can be *partially complementary*; thus origami can make *partial bonds* with a *strength* equal to the number of matching active patches. When origami with fully complementary stacking sequences match at every patch, they make a *full bond*. Partial bonds and other undesired bonds are *incorrect*; full bonds are *correct*. A set of stacking sequences will be defined to be *orthogonal* under mathematically-defined mismatch constraints similar to those used for DNA sequences. In practice, experimental claims of orthogonality must be accompanied by correct bond yields. DNA strands of the same base composition (percentage of A, C, G, and T) have similar bond energies; we say that stacking sequences with the same number of '1's have the same *weight* and assume that the corresponding stacking bonds will have similar bond energies, i.e. they are roughly *isoenergetic* (we provide evidence later).

### II.2.2.1.3. Design

#### II.2.2.1.3.1. Search process

We began our search for a code with the largest set of orthogonal sequences by first narrowing the search to nonorthogonal *candidate sets* of sequences that were of constant weight and

uniquely orienting, and that minimized undesired self-interactions. Candidate sets were parameterized by the number of active patches  $p$  in a correct bond, and the mismatch constraint  $i$ —the maximum strength of all possible partial bonds (considering all possible alignments between a sequence and itself as well as those between a sequence and its complement). A computer search was used to exhaustively determine candidate sets for values of  $p=5$  to 10 and  $i=2$  to 6, with  $i < p$ ; e.g.,  $(p, i)=(7, 4)$  contained 4614 sequences. Using the same mismatch constraint for interactions *between* sequences, we found the largest orthogonal subsets that we could by greedy search (see the note for ‘Searching for large orthogonal sets of sequences’ in Materials and Methods).

#### II.2.2.1.3.1.1. Design criteria & examples

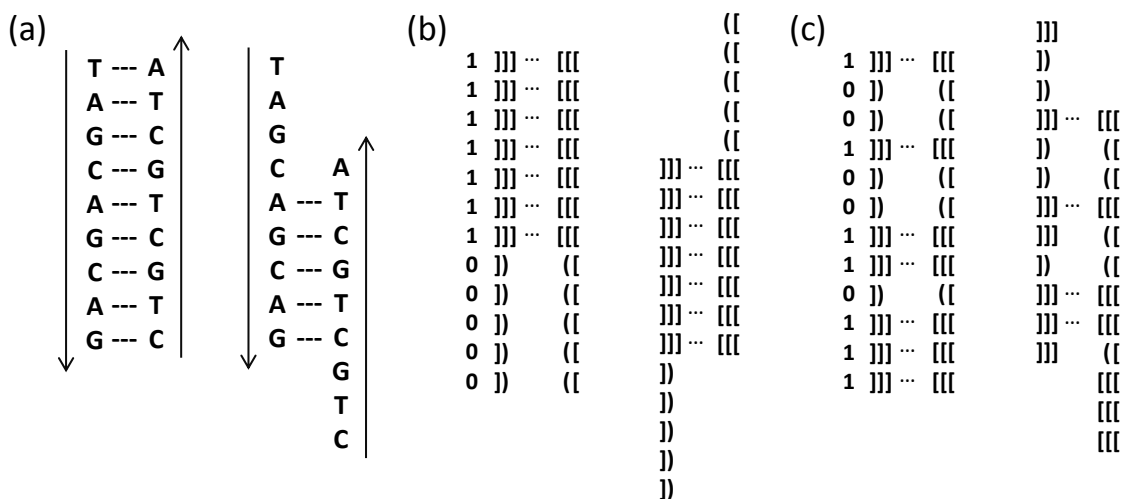


Figure II-5. **Issues in binary sequence design.** (a) DNA sequence design must deal with the problem of undesired partial complementarity. A desired bond is at left, an undesired partial bond at right. Binary sequence design is analogous, as explained in the text. (b) A simple binary sequence that allows a full-strength, self-complementary incorrect bond; this sequence, while nonpalindromic, is not uniquely-orienting. (c) A binary sequence whose strongest partial bonds are only of strength 4; an example is shown at right.

The basic design criteria for binary sequences can be understood by analogy to criteria used for DNA sequence design (Figure II-5a). Consider a DNA strand with the sequence 5'-TAGCAGCAG-3'; it is fully complementary to (and hence would bind most strongly with) a strand bearing the sequence 5'-CTGCTGCTA-3' (Figure II-5a-left). However, the two strands also have a partially complementary subsequence of length five, and could bind (albeit more weakly) via this partial interaction (Figure II-5a-right). In general, when DNA sequences are designed, they are designed to *minimize* such undesired interactions—with themselves, with their complements, and with any other strands that will be present in solution at the same time. Simple algorithms for

designing sequences use discrete criteria based on the maximum number of base pairs that occur in any partially complementary species. For example, an algorithm might be designed to find sequences that minimize this number. Partial bonds having the same number of base pairs but different sequences are not equal in strength, and so more sophisticated algorithms would minimize the sequence-dependent binding energy of undesired interactions. Still more sophisticated algorithms would use such binding energies to maximize the probability that the desired interactions form by considering the thermodynamic partition function.

Here, because we do not yet have a complete energy model for stacking bonds, we take a simple approach based on counting (and minimizing) the number of active patches involved in the strongest partial bonds. For example, consider an origami with the binary sequence ‘111111100000’; with its complementary partner it would form a stacking bond of strength 7 (Figure II-5b-left), but when rotated it can also form a self-complementary, undesired interaction of strength 7 (Figure II-5b-right). In contrast, the sequence ‘100100110111’ binds its complement (Figure II-5c, left) with a strength-7 bond, but the strongest possible partial bond that it can form has only strength 4 (Figure II-5c, right).

As for DNA, we are interested in minimizing such undesired interactions. For binary sequences of length  $l$  and number of active patches  $p$ , we wrote a program that enumerates sequences which have a maximum strength  $i$  for incorrect partial bonds (the mismatch constraint) with themselves and with their complements. Conceptually, the program compares each sequence to itself, and its complementary sequence, at all possible alignments by “sliding” the sequences relative to each other; the number of matches for each alignment is simply counted and the sequence is discarded if the number of matches exceeds  $i$  for any alignment.

The set of sequences enumerated for a given  $(p,i)$  constituted a *candidate set* from which we later attempted to construct maximal orthogonal subsets for use in making origami chains (see the note for ‘Searching for large orthogonal sets of sequences’ in Materials and Methods). It turns out that for  $p=7$ , and  $l=12$  or  $l=16$  (the length of the sequence applicable to the regular and tall rectangles used in our study, respectively), the candidate sets are empty for mismatch constraints  $i<3$ . That is, *however* we design a binary sequence with 7 active patches (for  $l=12$  or  $l=16$ ), such a sequence will have an undesired partial bond (with itself or its complement) involving at least 4 active patches. More generally, for all  $p$  there exists at least  $i$  for which candidate sequences can be found. As  $i$  is made larger, the size of the candidate set increases; this holds true for the size of the maximum orthogonal subsets as well. Thus, there is a tradeoff between the mismatch constraint  $i$  (our heuristic surrogate for the experimental specificity) and the number of distinct sequences

available as bond types. This can be seen in Table II-1 of the next section. Note that the minimum possible  $i$  is 2, since any pair of active patches in a binary sequence belongs to a partially self-complementary subsequence with at least two active patches.

For the 12-patch system with 7 active patches, a total of 98 different binary sequences were found to satisfy the mismatch constraint  $i=4$ ; for the 16-patch system with  $(p,i) = (7,4)$ , a total of 4614 sequences were obtained. We give some examples from each candidate set below. Full candidate sets are available upon request; alternatively, one can generate the sets easily using the program code (see Materials and Methods to learn how to obtain the code).

12-patch system (10 examples shown, out of a total of 98):	16-patch system (10 examples shown, out of a total of 4614):
0 1 0 0 1 0 1 1 0 1 1 1	0 0 0 0 1 0 0 0 1 1 1 1 0 1 1 0 1
0 1 0 0 1 1 0 0 1 1 1 1	0 0 0 0 1 0 0 1 0 0 1 0 1 1 1 1
0 1 0 1 0 1 0 0 1 1 1 1	0 0 0 0 1 0 0 1 0 1 0 0 1 1 1 1
1 1 0 0 0 1 0 1 1 0 1 1	0 0 0 0 1 0 0 1 0 1 1 0 0 1 1 1
1 1 0 0 0 1 1 1 1 0 1 0	0 0 0 0 1 0 0 1 0 1 1 0 1 1 1 0
1 1 0 0 1 0 0 0 1 1 1 1	1 0 0 1 1 0 0 0 1 1 0 1 0 0 0 1
1 1 0 0 0 0 1 1 1 1 0 1	1 0 0 1 1 0 0 0 1 1 0 1 0 1 0 0
1 1 0 0 0 1 1 0 1 1 0 1	1 0 0 1 1 0 0 1 0 0 0 0 0 1 1 1
1 1 0 0 1 0 0 0 1 1 1 1	1 0 0 1 1 0 0 1 0 0 0 0 1 1 0 1
1 1 0 0 1 0 1 1 0 1 0 1	1 0 0 1 1 0 0 1 0 0 0 0 1 1 1 0

#### II.2.2.1.3.1.2. Why use seven active patches with a mismatch constraint of four?

In Table II-1, we summarize the results of our computer search of bond types for different values for parameters  $p$ ,  $l$ , and  $i$ . We found that choosing the parameters  $(p, i)$  to be  $(7,4)$ ,  $(8,5)$ , or  $(9,6)$  with  $l=16$  yielded orthogonal subsets with more than ten sequences, while still maintaining a reasonably large energetic difference between full-strength correct bonds and partial incorrect bonds.

If one assumes the simplest model of binding energy for binary sequences (namely that the binding energy is linear in the number of active patches involved in a bond) then the energy of a full correct bond is  $p \cdot \Delta G_p$  (where  $\Delta G_p$  is the free energy of a bound active patch and is equal to two times  $\Delta G_s$ , the free energy of a stacked helix), the energy of the *strongest partial bond* is  $i \cdot \Delta G_p$  and the equilibrium ratio between the full correct bond and the strongest partial bond is:  $e^{-(p-i)\Delta G_p/kT}$ . A full treatment of the total error rate associated with a particular binary sequence would take into account not only the energy of the strongest partial bond, but also the number (multiplicity) of the different partial bonds having this energy, as well as the energies and multiplicities of all weaker partial bonds; such a treatment would calculate the full partition function for the system. Instead, here we simply assume that the multiplicity of the strongest partial bonds for different sequences is

roughly the same. Given these assumptions, then, the equilibrium error rates for sequences from the three different systems—(7,4), (8,5), and (9,6)—should be the same. However, because the fraction of correct bonds versus unbound origami should increase with increasing  $p$ , it would make sense to choose sequences from the system with full bonds of higher strength, i.e., a (9,6) system.

		Total # of available patches = 12 (regular rectangle)					Total # of available patches = 16 (tall rectangle)					
		# of active patches, $p$					# of active patches, $p$					
		5	6	7	8	9	5	6	7	8	9	10
# patches $i$ in partial bonds	2	4 (1)					320 (3)	0				
	3	214 (2)	0	0			1866 (27)	236 (6)	0	0		
	4		420 (15)	98 (2)	0			6520 (68)	4614 (12)	462 (2)	0	
	5			384	8 (1)	0			8322	2730 (13)	36 (2)	0
	6				328	14 (1)				6400	5870 (15)	496 (3)

Table II-1. Size of candidate sets and the largest orthogonal subsets found as a function of sequence length, number of active patches, and mismatch constraint. Numbers in parentheses indicate the size of the largest orthogonal subset found. Shaded areas indicate the systems with 3-patch difference between full-strength and partial bonds (corresponding to an equilibrium ratio of  $e^{-3\Delta G_p/kT}$ , where  $\Delta G_p$  is the free energy of a bound active patch and is equal to 2 times  $\Delta G_{st}$ , the free energy of a stacked helix). Blank spaces indicate that the search process was not performed for the corresponding parameters (because the result would either be meaningless [ $i \geq p$ ] or not useful, since either no candidate sequences would be found, or  $i$  was too close to  $p$  for bonds to be specific).

To check our assumptions about error rates, we measured the error rates for sample sequences from the (7,4), (8,5), and (9,6) candidate sets for length 12 sequences. Experiments analogous to those shown in Figure II-4a were conducted; Figure II-6 shows representative AFM images for each sequence tested. ‘L’-shaped labels on the origami made scoring correct head-to-tail bonds (L-L) easy; incorrect bonds included both bonds with rotated orientation and bonds with head-to-tail orientation that were misaligned. Surprisingly, the (7,4) sequence gave the best error rate, with the highest fraction of correct bonds out of total bonds—96.8% (N=344, for the sequence occurring in the bottom of Figure II-6a). The other systems performed considerably less well, with the (8,5) sequence having 77.7% correct bonds (N=358, Figure II-6b) and the (9,6) sequence having 52.7% correct bonds (N=277, Figure II-6c).

This surprising trend might not be a general phenomenon, since just a few sequences were examined, or it could be the case that our assumption about the multiplicity of partial bonds is wrong and that, for example, the (9,6) sequence observed simply had many more partial bonds than the other systems, all of them having the strongest possible strength ( $i$ ). However, given our thermodynamic experiments (see the “Thermodynamic measurements” section), another possibility suggests itself: that  $\Delta G_p$  is not constant as the number of active patches  $p$  increases and thus the total



stacking bond energy is not linear in the number of active patches. In particular, if  $\Delta G_p$  decreases with increasing  $p$  then our results make sense. Then the energy difference between a full correct bond and the strongest partial bond in the (9,6) system is not as large as the analogous energy difference for the (8,5) system, which in turn is not as large as that for the (7,4) system. Such sublinearity in stacking bond energy might be explained by steric interference or electrostatic repulsion between active patches, or it might be explained by a nonlinear bending energy term that increases as the use of more active patches results in them being more spread out, requiring them to overcome a large-scale deformation of the origami. Because of the trends we observed in our experiments using sequences with constant  $p$  and  $i$  (see the “Thermodynamic measurements” section;  $\Delta G_p$  seems to decrease as active patches are more spread-out along the origami edge) we suspect the latter hypothesis is a more likely explanation. Clearly performance measurements for many more sequences should be made, but based on these preliminary experiments, we chose to explore (7,4) sequences in the context of a longer, 16-patch system.

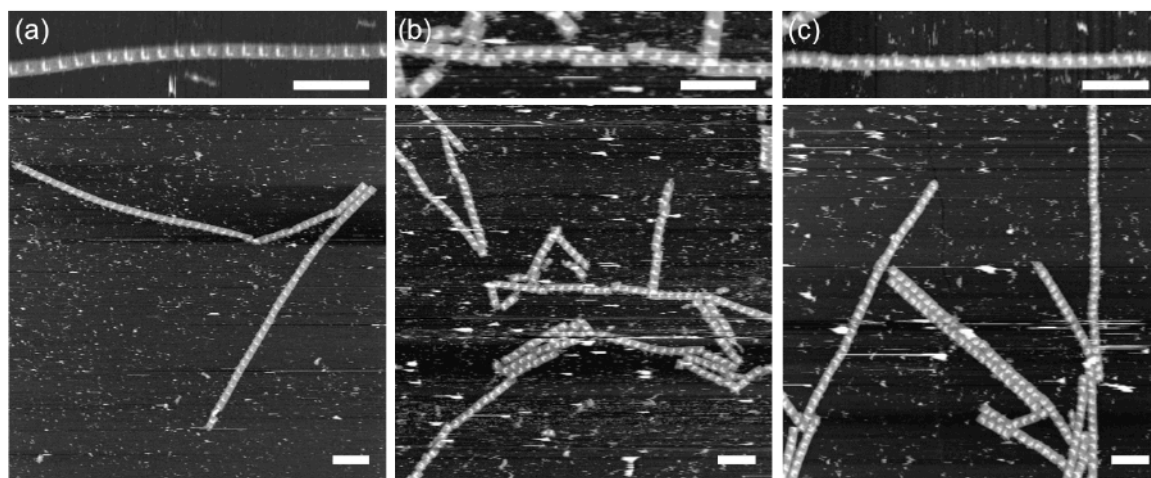


Figure II-6. **Comparison of sequence performance as a function of the number of active patches.** A binary sequence and its complement are placed on opposite edges of an origami such that it should form long chains; each origami carries the label “L”. Full-strength correct bonds are measured by counting the bonds with head-to-tail orientation (L-L). Partial bonds of all types are also counted; they usually involve origami bound in the rotated orientation. (a) One (7,4) system, ‘100100110111’ (top) and another ‘010111100011’ (bottom). Error rate data were taken for the bottom system; the top system is included to show a high-res image of a system of qualitatively similar error rate. (b) An (8,5) system, ‘100101011111’. (c) A (9,6) system, ‘110111001111’. Scale bars: 500 nm.

#### II.2.2.1.3.2. Quencher strands

When multiple origami containing different binary sequences on their edges are mixed together, interference can arise between the edge staples used to set the sequences on one origami

and the edges representing different sequences on another origami. This occurs because all origami in the binary coded system share the same basic design, and their edges share the same staple binding sites. For example, if one origami bears a sequence on its right edge that has a '1' in a particular position, then the staple that creates that active patch can bind to the *same* location on a different origami for which that location was intended to remain inactive, effectively "flipping" a '0' to a '1'. In the worst case, all of the origami would end up with exactly the same sequences, with the right edge of each origami encoding a sequence that represents the bitwise OR of all the sequences on the right edges of the original origami, and the left edge encoding an analogous bitwise OR of all the left-edge sequences. Prevention of such interference could be achieved by purifying the origami to remove excess staples *before* mixing the origami. But purification steps are usually accompanied by significant loss of the origami themselves and may incompletely remove staples. In particular, simple and fast methods such as spin filtration reduce excess staples by only a factor of 5- to 10- fold; more complete removal requires more stringent methods such as gel purification. Complete removal is important because, as we observed in tests of spin purification (data not shown), even sporadic changes to edge sequences (flipping of some '0's to '1's) caused by a relatively small amount of edge staples left can significantly increase error rate.

As an alternative approach, we introduced strands complementary to the edge staples, which we term *quenchers*. Quenchers bind to the excess free edge staples in solution and effectively prevent them from binding to the scaffold strand. Quenchers were designed so that they have complementary sequences to the corresponding edge staples (thus quenchers have sequences derived from scaffold strand subsequences), and extra two thymine bases were added to both the 5'-end and the 3'-end (so that the quencher sequence becomes 5'-TT-staple complement-TT-3'). The thymine addition was done to minimize the potential influence of (1) stacking interference from blunt ends that would be generated if simple complements were used and (2) breathing of the resulting quencher-staple duplex that might allow the edge staple strand to bind to a '0' location anyway, via a branch migration process.

The efficiency of the quenchers at blocking the free edge staples was not explicitly measured. However, the high molar excess of the quenchers used (10 times the concentration of edge staples) and the high free energy of binding between the quenchers and the edge staples (on the order of ~40 kcal/mol, calculated using Oligo Calc, <http://www.basic.northwestern.edu/biotools/oligocalc.html>) predicts the concentration of free edge staples, in the presence of the quenchers, to be extremely small — on the order of  $10^{-21}$  nM. The experimental protocol for using quenchers is described in the Materials and Methods section.

#### II.2.2.1.4. Results and discussion

##### II.2.2.1.4.1. AFM data and correct bond fractions of binary sequences

For edges of the tall rectangle system (which has 16 total available patches), there are 4614 different binary sequences with  $(p, i)$  of (7,4), as shown in Table II-1. Within the set of those binary sequences, subsets (codes) can be found for which every pair of sequences from the subset is mutually orthogonal with the same matching criterion (no partial match between any pair of sequences involves more than 4 active patches). Our computer-aided search generated several codes of size 11 and 12; size 12 was the largest obtained. One code of size 12 and another code of size 11 were chosen for more detailed investigation. Each binary sequence from these codes was tested by placing the binary sequence and its complement on opposite edges of the tall rectangle, such that the rectangles form bonds in the head-to-tail (h2t) orientation when the bonds are full-strength and correct. For each case, AFM data were analyzed and bond orientations were measured to obtain the ratio between the correct (h2t) bond orientation and the total number of bonds. The correct bond/total bond fractions for the 23 individual sequences ranged from 73% to 98% with an average of 87%, and are listed in Table II-2. Figure II-4a shows one of the better sequences (94.4%,  $N=659$ ). While predicted 4-active-patch incorrect bonds were observed, a significant source of error came from bonds not considered in design: the flexibility of active patches allows them to form *bent-patch bonds* encompassing 5 or more active patches (Figure II-4b).

set	sequence #	sequence	% correct bonds	N
set1	1	0000010111100011	90.07%	423
	2	1110000011000101	84.39%	538
	3	0010000010110111	79.73%	301
	4	1100000010111001	86.50%	941
	5	0110101000100101	87.96%	191
	6	0100010100011011	97.55%	245
	7	1000100011010011	82.25%	524
	8	0111100100001010	96.53%	346
	9	1001001001100101	72.54%	142
	10	1110101001000010	87.64%	259
	11	1000010110010101	83.68%	337
	12	1101000100001101	83.36%	559
set2	1	00010111110001100	94.39%	659
	2	0010010011011100	78.07%	456
	3	1010010000110110	95.24%	210
	4	1000100010101101	74.29%	210
	5	0111000010110010	95.41%	827
	6	1100000010011110	95.54%	112
	7	1101010010000011	94.20%	448
	8	1001100101000101	91.67%	168
	9	1101000000110101	81.68%	475
	10	0110010110001001	78.33%	300
	11	1001101000100011	93.81%	113

Table II- 2. **Fractions of correct bonds for binary sequences from two different sets (sequences within each set are orthogonal).** Each binary sequence was tested by placing the sequence and its complement on opposite edges of the tall rectangle origami, such that the rectangles form bonds in the head-to-tail (h2t) orientation when the bonds are full-strength and correct. The percentage of correct bonds (with h2t orientation) out of the total number of bonds analyzed (N) was recorded for each binary sequence.

#### II.2.2.1.4.2. Chain of multiple origami via orthogonal sequences

By closely examining error bonds in AFM data, we observed that contiguous runs of active patches were less likely to form bent-patch bonds than isolated active patches or pairs of active patches. Thus we verified the orthogonality of multiple sequences using a subset of the 11-sequence code for which the sequences all have runs of at least three active patches in a row ('111'). Figure II-4c shows the four sequences that were used, and a five-origami chain made with this code; all-'0' *null-bonds* were applied to the left edge of the first origami and the right edge of the fifth. 88% of total bonds ( $N=66$ ) observed were correct bonds, a rate similar to that observed for single bond types in isolation. Further, the rate of *monomer conversion* was significant: the fraction of origami found in correct length chains was 31% ( $N=192$ ). Wide-field AFM images are given in Figure II-7.

This result compares favorably with the best yield reported (24%) using DNA hybridization to construct chains of five origami<sup>46</sup>. Note also that the latter work requires purification of individual origami because excess copies of origami-connecting staples interfere with the coupling reaction. As discussed earlier, our approach has a similar difficulty: since all origami share the

same design, excess edge staples from one rectangle can bind inactive patches of a different rectangle, flipping ‘0’s to ‘1’s. Here, rather than purifying rectangles (which causes loss of origami), we added a 10-fold excess of complementary quencher strands to neutralize excess staples, which eliminates the cumbersome steps of purification and performed more effectively than purification in our tests (see the ‘Quencher strands’ section above).

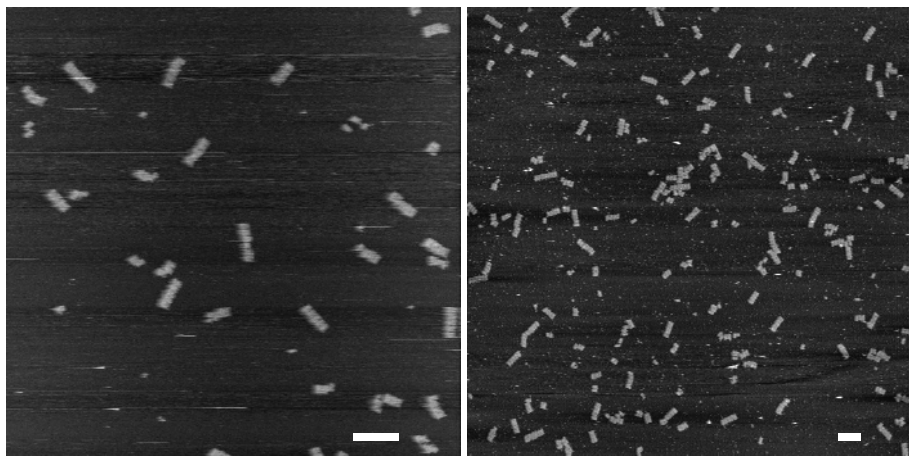


Figure II-7. **Wide-field AFM images of 5-origami chains with orthogonal bonds based on a binary code.** Full correct bonds between each pair of origami resulted in chains with five distinct origami. Due to mismatched bonds and small stoichiometric discrepancies, shorter chains, longer chains, and 5-origami chains containing incorrect bonds were also found. 88% of total bonds analyzed (N=66) were correct bonds, and the fraction of origami found in 5-origami chains was 31% (N=192). Scale bars, 500 nm.

## II.2.2.2. Shape coding

### II.2.2.2.1. Key concepts

We next encoded bond type using geometric complementarity between *shape pairs*, e.g., the right edge of origami A fits the left edge of B (Figure II-8a). Edges are again divided into patches, but with three differences. First, each patch has *four* helices rather than two. (Two-helix patches were too flexible and formed too many bent-patch bonds: see the warning on “length and width of a patch in shape design” in the Materials and Methods section.) Second, *all* patches are active. Third, each patch has one of  $d$  depths from 0 to  $(d-1)$ . Each depth corresponds to a physical width, measured in the  $x$ -direction in increments of three helical turns (e.g., depth-2 is six turns.)

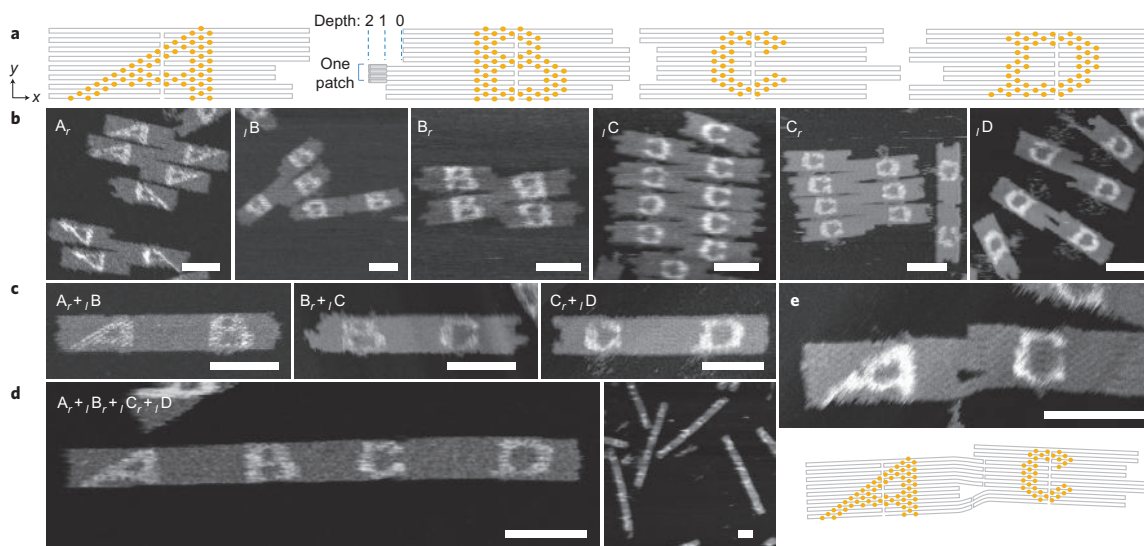


Figure II-8. **Recognition based on complementarity of origami edge shapes.** (a) Models of four origami: A, B, C and D. Orange dots mark positions of dumbbell hairpin labels. (b) Test of self-interactions for each edge shape. Subscripts ‘ $r$ ’ or ‘ $l$ ’ denote the edge tested. AFM shows common partial self-bonds that result in aggregation. (c) Tests of complementary edge shapes. AFM shows correct, full bonds. (d) AFM images of the four-origami chain, A-B-C-D. (e) AFM image and schematic representation of a 3-patch bent-patch bond. Scale bars, 100 nm.

#### II.2.2.2.2. Binding rules

Again, abstract sequences representing bonds are defined, e.g., the *shape sequence* on the right side of origami A is ‘2201’. Shape sequences are not necessarily unique, e.g., ‘0101’=‘1212’. Again, stacking polarity ensures that a shape sequence is not, in general, equal to its reverse; stacking polarity further ensures that *all* sequences except ‘0000’, including palindromes such as ‘0110’, are uniquely orienting. Complementarity for shape sequences is similar to that for DNA: a sequence  $n_1n_2n_3n_4$  binds a *reverse complement*  $\overline{n_4n_3n_2n_1}$  where  $\overline{n_k} = (d - n_k - 1)$ , e.g., ‘2201’ binds ‘1200’. All shape-pair bonds with the same number of patches  $p$  have the same number of blunt-end stacks, and so we assume they are roughly isoenergetic.

#### II.2.2.2.3. Design

##### II.2.2.2.3.1. Search process

As before, we constructed shape codes by starting with candidate sets of shape pairs with minimal self-interaction. Candidate sets were parameterized by  $p$ ,  $d$ , and the same mismatch constraint  $i$ . E.g., for  $(p, d, i)=(9, 5, 3)$ , 24,791 possible shape sequences were found. However, we

avoided large  $p$  because the size limit placed on origami by scaffold length prevented the use of numerous four-helix wide patches. Similarly, we avoided using large  $d$  because deeper patches are more flexible. Thus we started from the  $(p, d, i)=(4, 3, 2)$  candidate set which has 16 sequences (see below for full list). While  $p-i=2$  for such shape pairs is smaller than for binary sequences used above ( $p-i=3$ ), the *absolute* energy difference between a correct bond and a strongest possible partial bond is larger for the shape pairs ( $8\Delta G_{st}$  vs.  $6\Delta G_{st}$ ) and so we expect the fraction of correct bonds could be higher (ignoring the degeneracy of partial bonds or off-model interactions like bent-patch bonds). Using the same mismatch constraint  $i=2$ , we constructed maximal orthogonal sets of shape pairs exhaustively. Maximal sets with four shape pairs were found; we tried a four-shape-pair set and found that one of its shape pairs created numerous bent-patch bonds. Other four-shape-pair sets included the offending shape pair, or similar ones. Thus the largest orthogonal set achieved had three shape pairs (Figure II-8a).

#### II.2.2.2.3.1.1. Design criteria

As in the design of binary sequences, the goal of minimizing undesired bonds dictates design criteria for shape sequences. As before, fully self-complementary sequences (Figure II-9ab) or partially self-complementary sequences (Figure II-8cd) must be avoided (unless a homodimer of origami is desired.) Computer enumeration of candidate sets of sequences for shape codes is essentially similar to that for binary codes, with two important differences. First, unlike the case for DNA or binary sequences, *not all shape sequences encode physically distinct bonds*; we discuss this in the next section. Second, unlike the case for DNA or binary sequences, *the program must make an extra check for the self-complementarity of the shape sequence's complement*. For DNA base pairing or binary codes, to evaluate whether a sequence should be a candidate sequence, it is sufficient to check that sequence's self-complementarity, and to check for any partial complementarity that it might have with its complement. This is because, for a DNA or binary sequence, any self-complementary subsequence that occurs in the sequence implies the existence of a corresponding self-complementary subsequence in the sequence's complement, and vice versa. This is not the case for shape sequences: the shape sequence '100001' has a strongest self-complementary partial bond of strength 2, but its complement '011110' has a strongest self-complementary partial bond of strength 4. We note also that the minimum possible mismatch constraint  $i$  for shape sequences is 2 patches (as it is for binary codes), but this limit holds for a different reason in the case of shape sequences than for binary sequences. The reason is that, for an arbitrary shape sequence, the first two or last two patches both form self-complementary subsequences that can bind to themselves without steric hindrance from any of the other patches (if

the two origami carrying them are in a rotated orientation). The top two and bottom two examples in Figure II-9f demonstrate this phenomenon.

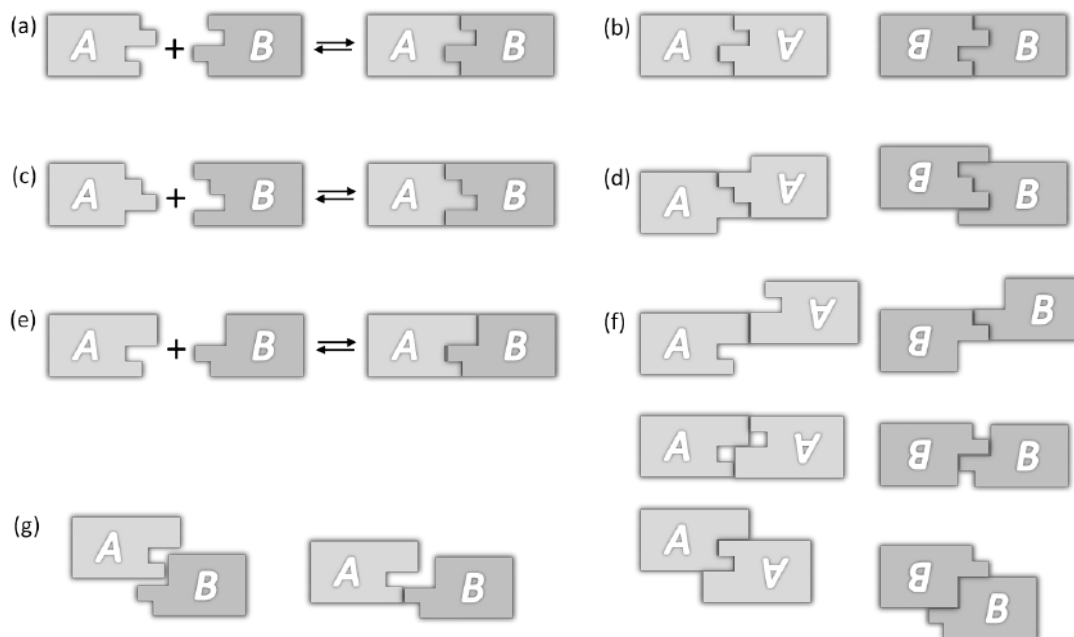


Figure II-9. **Examples of shape sequences and their partial bonds.** (a) A 4-patch shape sequence could form a 4-patch bond between distinct origami, but because it is fully self-complementary (b) it also allows full-strength undesired bonds. (c) Another shape sequence and its complement could form a 4-patch bond, but (d) it also allows partially self-complementary 3-patch bonds (3/4 the strength of a full bond). (e) A shape sequence and its complement that we used between the **A** and **B** origami. (f) Examples of partially self-complementary (homogeneous) bonds of strength 2 for the sequences in (e). (g) Examples of partial bonds of strength 1 between the two distinct origami (heterogeneous) for the sequences in (e).

Early on in the project we thought we might achieve a large number of specific bonds through the use of a long ( $l=6$  or  $l=9$ ) shape sequences. Fitting these high complexity sequences into the relatively small area of an origami necessitated using patches that were just two helices wide. These proved too flexible to prevent bent-patch bonds (see the warning on “Length and width of a patch in shape design” in the Materials and Methods section) so we decided to use four-helix wide patches to implement shape codes. This restricted the length of the shape sequences we could use to just four patches; similarly we restricted ourselves to just three depths to avoid long, flexible patches. Fortunately, even with these restrictions, the candidate set for the mismatch constrain  $i=2$  had 16 elements, which we discuss next.

#### II.2.2.2.3.1.2. Full list of candidate shape sequences for the (4,3,2) system

Given the number of patches ( $p$ ) and number of depths ( $d$ ), our program searches the entire



sequence space and examines the possible partial bonds for each shape sequence with itself, each sequence with its complement, and each complement with itself to see if they exceed the mismatch constraint ( $i$ ). For  $(p,d,i) = (4,3,2)$  the program generated a candidate set of 16 unique shape sequences, listed at left below. Note first that if a sequence appears, its complement does not appear: our desire is to make a set of candidate sequences for distinct bond types, and a sequence and its complement are *equivalent* with respect to the physical bond type that they encode. Similarly, sequences that are related by a simple *shift* in depth, such as ‘0010’ and ‘1121’ (Figure II-10ab), encode the same bond type and are thus equivalent. Only a single sequence from an equivalence class is included in the candidate set. Recall that due to stacking polarity, a shape sequence (e.g. Figure II-10a) and its reverse (Figure II-10e) are inequivalent, unless they are palindromic. (Shape sequences and their reverses are expected to be similar for some properties, e.g., flexibility.)

List of the 16 shape sequences:

- 1 ‘0010’ (= ‘1121’ = ‘1011’ = ‘2122’)
- 2 ‘0020’
- 3 ‘0021’
- 4 ‘0100’
- 5 ‘0102’
- 6 ‘0110’ (B-C bond)
- 7 ‘0200’
- 8 ‘0201’
- 9 ‘0211’
- 10 ‘0220’
- 11 ‘0221’
- 12 ‘1020’ (C-D bond)
- 13 ‘1120’
- 14 ‘1200’ (= ‘2201’, A-B bond)
- 15 ‘1220’
- 16 ‘2010’

(Equivalences are not exhaustively listed.)

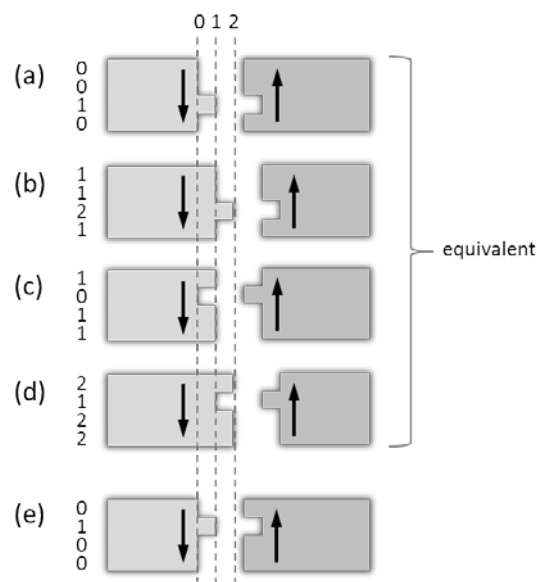


Figure II-10. **Equivalences between shape sequences.** (a), (b), (c), and (d) are all equivalent with respect to the bond type they encode. Vertical arrows denote stacking polarity. (a) and (b) are related by a simple shift in depth, (b) and (c) are complementary, (c) and (d) are related by a depth shift, and (a) and (d) are complementary. (e) is, however, distinct from the others because it is the reverse of (a), and has opposite stacking polarity.

#### II.2.2.2.3.1.3. Orthogonality graph for the (4,3,2) candidate shape sequences

Using our computer program, one can check the orthogonality between any two sequences in the set of 16 shapes listed in the previous section. By applying the same mismatch constraint for the strongest partial bonds ( $i=2$ ) between different sequences, the orthogonality relations can be determined for each pair of sequences (a total of 120 combinations). Figure II-11a shows the full orthogonality graph for all 16 shape sequences in the candidate set. Line segments between numbered circles indicate that the two shape sequences corresponding to the numbers are orthogonal to each other. (For the identity of each shape sequence, see list in the previous section.)

Sets of mutually orthogonal sequences correspond to *complete subgraphs* or *cliques* of the orthogonality graph. That is, a set of vertices for which *every* pair of vertices is connected by a line segment corresponds to a set of mutually orthogonal sequences. For example, one complete subgraph is the red triangle in Figure II-11b which corresponds to an orthogonal subset with three sequences, {6,12,14}. An exhaustive search confirmed that the size of the largest orthogonal subsets for the given system is 4—for example the subset indicated by the red subgraph in Figure II-11c: {3,6,7,16}. We attempted to construct origami chains based on a subset of size 4. Unfortunately, the

set we chose included sequence 5, which, along with its reverse sequence 16, turned out to be susceptible to the formation of bent-patch bonds. In the interest of saving time and money, and because we had demonstrated that sequence 6 worked well, we ended up choosing the 3-sequence subset  $\{6,12,14\}$  to explore as a shape code. A four-sequence orthogonal subset that does not contain sequences 5 or 16,  $\{1,11,12,14\}$  looks promising, but was not explored.

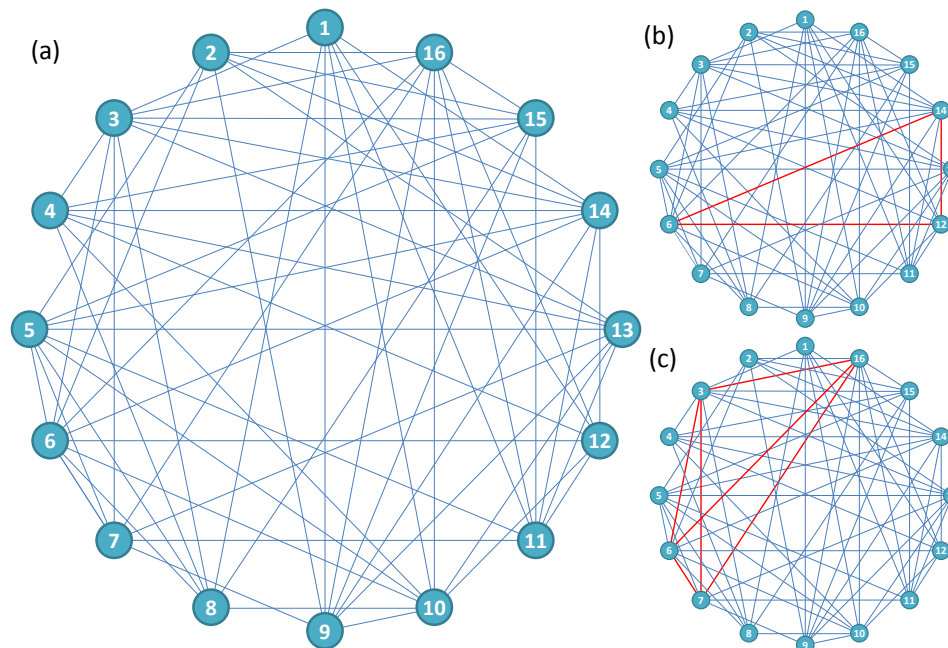


Figure II-11. **Orthogonality graph for the candidate set of 16 shape sequences.** (a) The full graph. Each line connecting two numbered circles indicates that the two shape sequences are orthogonal. (b) Red triangle highlights an orthogonal set of size three with sequences 6, 12, and 14. (c) Red subgraph highlights an orthogonal set of size four with sequences 3, 6, 7, and 16 (not chosen for tests because the sequence 16 turned out to be susceptible to error bonds).

#### II.2.2.2.3.1.4. Size of shape sequence spaces with other parameters

In addition to the 4-patch system with  $d=3$  and  $i=2$ , we explored other shape sequence spaces with different parameters, a couple of which were tested experimentally (some results are shown in the section “Length and width of a patch in shape design” in the Materials and Methods section). In Table II-3, we summarize the sizes of shape sequence spaces with various parameters (different numbers of patches, depths, and patches allowed in incorrect bonds). In the right-hand column we report the size of the largest orthogonal subset discovered over the course of multiple random searches, as described in Materials and Methods.

For parameters not included in this table, one can easily obtain the shape sequence space—not only the size but the entire list of candidate shape sequences—using the program code provided:

link to the zipped file can be found in Materials and Methods.

# patches ( $p$ )	# depths ( $d$ )	# patches allowed in incorrect bonds ( $i$ )	% patches in incorrect bonds	Size of shape sequence space	Size of largest orthogonal subsets found
4	2	2	50%	3	1
	3	2	50%	16	4
	4	2	50%	61	8
6	3	2	33%	5	1
		3	50%	49	7
	4	2	33%	145	6
		3	50%	423	22
9	3	2	22%	2	1
		3	33%	14	5
	4	2	22%	233	6
		3	33%	1113	25
	5	2	22%	6510	15
3		33%	24791	60	

Table II-3. **Sizes of shape sequence spaces and orthogonal subsets for different systems.** The purple-shaded parameters are those for the shape sequences most explored by this study.

#### II.2.2.2.4. Results and discussion

##### II.2.2.2.4.1. AFM data and correct bond fractions of shape sequences

*Half* of each shape pair was tested to measure its propensity for self-interaction: single origami were synthesized with just edge staples for the shape tested (Figure II-8b). When annealed from 90 °C to 20 °C without a complementary partner, such origami bind via expected 2-patch partial bonds and often form extended *zig-zags*. Simply mixing complementary origami at 20 °C gives poor results since they remain kinetically trapped in partial bonds. Full 4-patch bonds formed well (Figure II-8c) when complementary origami were annealed from 90 °C to 50 °C, mixed and held at 50 °C for 12 hours, and cooled to 20 °C over 6 hours: the fraction of correct bonds was 95 % for  $A_r+iB$  ( $N = 191$ ), 98 % for  $B_r+iC$  ( $N = 203$ ), and 97 % for  $C_r+iD$  ( $N = 179$ ) and the rate of monomer conversion into correct dimers was 91 % for  $A_r+iB$  ( $N = 397$ ), 90 % for  $B_r+iC$  ( $N = 442$ ), and 91 % for  $C_r+iD$  ( $N = 384$ ).

##### II.2.2.2.4.2. Chain of multiple origami via orthogonal sequences

All four complete origami were mixed together and subjected to the same protocol: each origami shape piece was annealed from 90 °C to 50 °C, mixed and held at 50 °C for 12 hours, and

cooled to 20 °C over 6 hours (Figure II-8d). 81% of total bonds ( $N=279$ ) observed were correct bonds and the rate of monomer conversion into correct four-origami chains was 44% ( $N=430$ ). Again, bent-patch bonds not considered in design were a significant source of error (Figure II-8e). Wide-field AFM images are shown in Figure II-12.

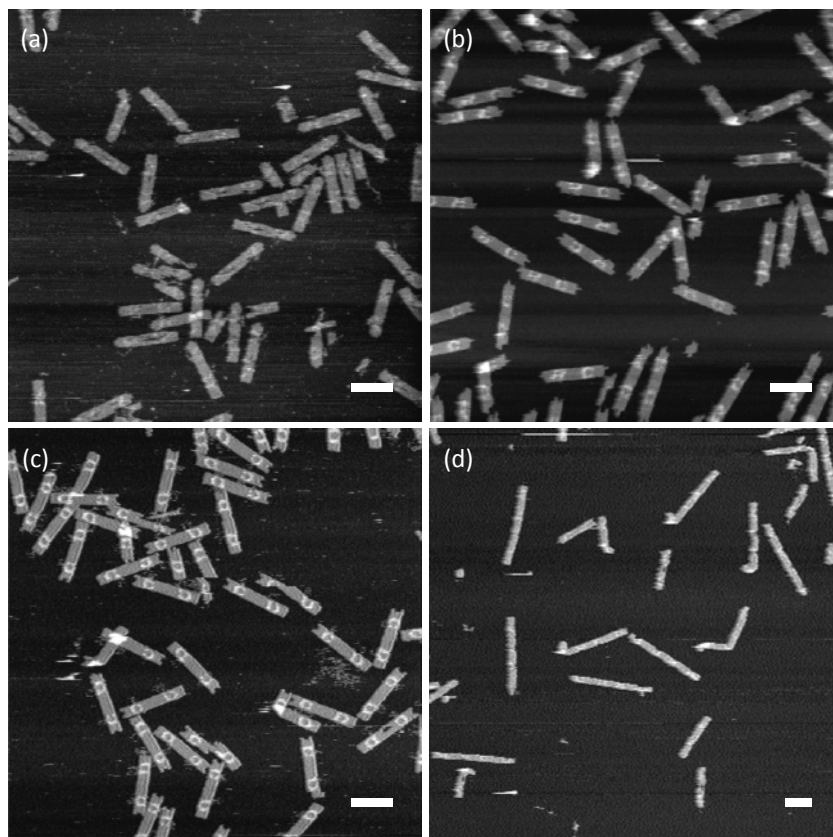


Figure II-12. **Wide-field AFM images of dimers and 4-origami chains with orthogonal bonds based on a shape code.** (a-c) AFM of dimers of (a)  $A_r + B_r$ , (b)  $B_r + C_r$ , and (c)  $C_r + D_r$ , respectively, show high yields of correct, full bonds. The fractions of correct bonds out of total bonds analyzed were 95% ( $N=191$ ), 98% ( $N=203$ ), and 97% ( $N=179$ ), respectively, and the fractions of origami found in correct dimers were 91% ( $N=397$ ), 90% ( $N=442$ ), and 91% ( $N=384$ ), respectively. (d) AFM of the **A-B-C-D** system show 4-origami chains with full, correct bonds (some chains shown are folded), and some shorter chains that may result from stoichiometric discrepancies or mismatched bonds. 81% of total bonds analyzed ( $N=279$ ) were correct bonds, and the fraction of origami found in 4-origami chains was 44% ( $N=430$ ). Scale bars, 200 nm.

### II.2.2.3. Summary and comparison of the two approaches

In Table II-4, we summarize the binding rules for both binary and shape coding approaches, and compare them with the corresponding rules of base sequences. The binding rules of stacking bonds emerge from the combination of two factors. One is the property of stacking polarity that

naturally occurs due to the distinction between major and minor grooves on each blunt end (see the section “Anti-parallelism of stacking bonds” above), which is analogous to the 5’ to 3’ directionality of a DNA sequence. The other is the nature of each coding approach: the complementary relation of each binding patch is defined in different ways in the two approaches. Shape coding generates a heterophilic complementarity relation—e.g., ‘2’ is a complement of ‘0’, and vice versa, and ‘1’ is a complement of ‘1’, the latter happening to be a special case—just like ‘A’ is a complement of ‘T’, and vice versa, and ‘C’ is a complement of ‘G’, and vice versa, in base sequences. On the contrary, the binary coding approach still utilizes the homophilic binding relation of stacking—‘1’ is a complement of ‘1’ itself, and ‘0’ is a complement of ‘0’ itself—and creates its unique set of binding rules. Thus, shape sequences show somewhat similar binding rules to those of base sequences—e.g., complementary sequence is a *reverse complement* of the original sequence—whereas binary sequences show their unique behaviors—complementary sequence is a simple *reverse* (see Table II-4).

	Base sequences	Binary sequences	Shape sequences
Directionality	5’ → 3’ (e.g. ‘ATGC’ ≠ ‘CGTA’)	Defined to match 5’ → 3’ of the scaffold strand (e.g. ‘1011’ ≠ ‘1101’)	
Complementary sequence	Reverse complement (e.g. ‘ATGC’ & ‘GCAT’)	Reverse (e.g. ‘1011’ & ‘1101’)	Reverse complement (e.g. ‘1200’ & ‘2201’)
Self-complementary sequence	Reverse palindrome (e.g. ‘AGCT’)	Palindrome (e.g. ‘0110’)	Reverse palindrome (e.g. ‘1201’)
Bond directionality	Antiparallel	Antiparallel	Antiparallel

Table II-4. Comparison of binding rules of different sequences.

Another interesting comparison among the sequences from different approaches can be drawn with respect to a few practical aspects (Table II-5). DNA base sequences and shape sequences are not reprogrammable in the sense that once a sequence is designed and synthesized, any change on the sequence would require a new synthesis in general. On the contrary, binary sequences are reprogrammable: one origami design can create a full set of binary sequences that can be theoretically constructed for the given number of patches along the edge. In terms of the theoretically reachable total bond space, when not considering any orthogonality nor sequence homology, shape sequence has the most flexibility, and has potentially the largest bond space in

case a depth ( $d$ ) larger than four and a reasonably large number of patches ( $N$ ) are used. However, as discussed in earlier sections (see also the “Length and width of a patch in shape design” section in Materials and Methods), the flexibility of DNA origami nanostructures, along with the size limit of DNA origami, restricts the number of patches and the depth that are practically usable. Even when large  $d$  and  $N$  could be achieved, off-design errors such as bent-patch bonds further limits the size of the practical bond space. With respect to the necessity of a purification step in experimental implementation, however, stacking sequences outperform base sequences. This comparison is for the particular case where these sequences are used to connect multiple origami together. When base sequences are used to connect origami together in the form of sticky ends, without purification those sticky ends would experience “saturation” by free strands in solution complementary to the sequences; the sticky ends added in excess amount along with regular staple strands during the assembly of origami would, upon mixing of multiple origami solutions, bind to their complementary sticky ends and leave the sticky ends on origami inaccessible to each other. As discussed earlier, binary sequences face the same problem. Without purification of the extra staple strands, free edge staples in solution can invade into areas that encode inactive patches and flip ‘0’s to ‘1’s. However, the property of binary sequences that the specificity arises from the alignment of active patches, not their base sequences, allows avoidance of purification. A system based on binary sequences can use even more excess amounts of complementary strands to the edge staples (“quencher strands”), effectively annihilating free edge staples in solution. In the case of shape sequences, on the other hand, where the specificity is encoded entirely by the geometry and there is no chance of interference between active and inactive patches (all patches are active), no purification is required whatsoever.

	Base sequences	Binary sequences	Shape sequences
Reprogrammability	Not reprogrammable	Reprogrammable	Not reprogrammable ('hard-coded')
Total bond space (theoretical, nonorthogonal)	$\sim 4^N$	$\sim 2^N$	$\sim d^N$
Usable bond space (practical, orthogonal)	$\sim O(10)$ with $N=4$	$\sim 4$ , with $N=16$	$\sim 3$ , with $N=4, d=3$
Purification (for connecting origami)	Purification needed.	Purification or quencher strands	No purification needed.

Table II-5. Comparison of some practical aspects of the different sequences.

### II.2.3. Programming of cis-trans isomeric structures

Next we show that stacking bonds based on the two approaches described so far can be used to control complex geometric arrangements of origami by exploring the multimerization of a  $60^\circ$  corner (Figure II-13). Such a corner, with straight edges of the same stacking polarity, can self-associate in two ways (Figure II-13a,f): in *cis* via a  $120^\circ$  rotation, or in *trans* via a  $180^\circ$  rotation. With all-cis bonds the corner would make *triangles*; with all-trans bonds the corner would make *zig-zags*. For all-‘1’ edges, a mixture of diastereomers results (Figure II-13k,p) with a cis:trans ratio that mildly favors cis bonds (68:32), and a relatively poor full bond yield (53%, cis+trans). The question is, how to use binary or shape codes to achieve high yields of a single diastereomer?

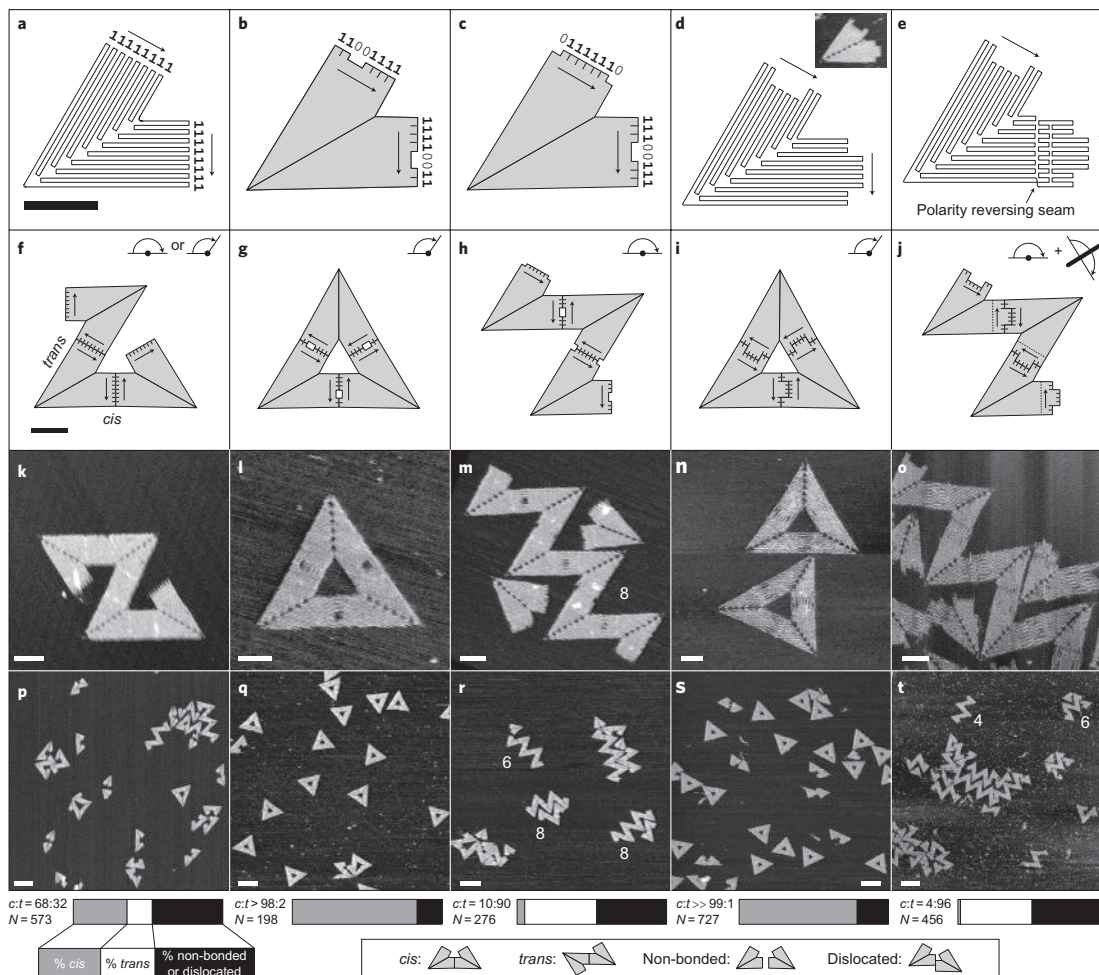
Using an asymmetric sequence ‘11001111’ on one edge and its reverse ‘11110011’ on the other specifies the creation of only cis bonds (Figure II-13b,g). Indeed, a high cis:trans ratio of 98:2 and cis full bond yield of 83% were observed. Conversely, use of two orthogonal palindromic sequences, ‘01111110’ and ‘11100111’ should create only trans bonds (Figure II-13c,h). A poorer cis:trans ratio of 10:90 and a lower trans full bond yield of 48% were observed.

Use of a simple centro-symmetric shape pair should create only cis bonds (Figure II-13d,i); it results in a very high cis:trans ratio of  $\gg 99:1$  (with only a single trans bond among 727 bonds analyzed) and cis full bond yield of 79%. For such centro-symmetric shape pairs, which isomer forms is specified entirely by the stacking polarity, and so switching from cis to trans isomers requires the addition of a polarity-reversing seam for one of the edges; here the trans bonds involve both a  $180^\circ$  rotation and an additional flip (Figure II-13e,j). The shape coding approach resulted in a better cis:trans ratio (4:96) than for the corresponding trans-specifying binary code, but gave a similarly low (48%) trans full bond yield.

Given that similar binary sequences or identical shapes were used for both cis- and trans-specifying systems, the performance of trans-specifying systems was unexpected. We hypothesize that the lower full bond yield and poorer cis:trans ratio of trans-specifying systems, as well as the cis preference for all-‘1’ bonds, are artifacts of deposition. Because triangles must break two bonds to fall apart, they may survive the deposition process better than zig-zag chains which need break only one bond to fall apart. Further, the absence of long zig-zag chains and the observed patterns of origami in trans-specifying experiments suggest that the origami may form *closed zig-zag loops* in solution: chains, closely associated pairs of chains, and other clusters for which the total number of origami is *even* are commonly observed (numbers in Figure II-13m,r,t; to have an odd length, closed loops would have to twist  $180^\circ$ ). In order to stick flat on mica, closed loops must break once to form chains or twice to form pairs of chains; small chains with several associated



singletons suggest that loops break more than twice.



**Figure II-13. Control of cis-trans isomerism.** (a) Scaffold path for a 60° corner, with straight edges. Arrows indicate stacking polarity, which allow corners to form two types of antiparallel bond: *trans* bonds (rotated 180°) or *cis* bonds (rotated 120°) as indicated in (f) by 180° or 120° arcs. (b) A corner with sequences ‘11001111’ and ‘11110011’, designed to specify all-*cis* bonds to create triangles (shown in (g)). (c) A corner with sequences ‘01111110’ and ‘11100111’, designed to specify all-*trans* bonds to create zig-zags (shown in (h)). (d) Scaffold path for a corner with the ‘0110’/‘1001’ shape pair used between B and C in Figure II-8. This shape pair specifies the formation of all-*cis* triangles (shown in (i)). (e) Scaffold path for a corner with the same shape pair, but with the polarity of one edge reversed. This specifies the formation of all-*trans* zig-zags (shown in (j)). (k–o) AFM of origami based on the designs in (a–e). (l) and (o) have been stretched and/or sheared to compensate for AFM drift. (p–t) Large field AFM corresponding to (k–o). Bar graphs indicate the fraction of *cis* (gray), *trans* (white), and disrupted (black, nonbonded or dislocated) bonds (percentages in text). The normalized *cis:trans* ratio (*c:t* such that  $c+t=100$ ) and number of origami counted (*N*) are given. White numbers next to zig-zag clusters in (m, r, t) give the number of origami they contain. Scale bars in (a,f,k–o), 50 nm; in (p–t), 200 nm.

## II.2.4. Thermodynamic measurements

To understand stacking bond energies, it is necessary to know the single blunt-end stacking energy,  $\Delta G_{st}$ . Because origami chains break and sometimes aggregate upon deposition, we measured  $\Delta G_{st}$  using a simpler system based on an equilibrium between monomers and dimers. A palindromic binary code with  $p=2, 4, \text{ or } 6$  ('000001100000', '000011110000', or '000111110000') was applied to only one edge of the 24-helix rectangle. We deduced equilibrium concentrations from the number of monomers and dimers in AFM images, calculated the free energy of binding for each bond, and interpreted these binding energies under two different models. Our first model assumes that loop-loop interactions at inactive patches are neutral (i.e.  $\Delta G_{ll}=0$ ). Our second model assumes  $\Delta G_{ll}$  is a potentially non-zero constant, independent of sequence. Because we did not observe loop-blunt end interactions we did not consider them.

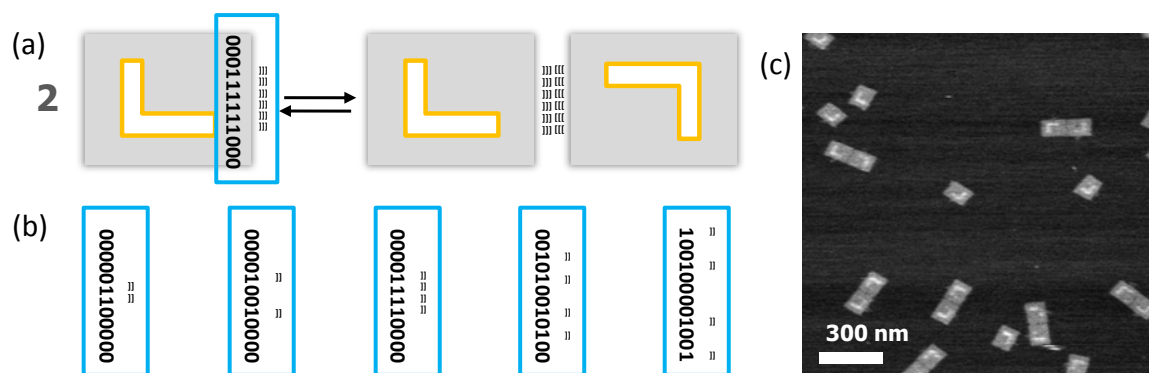


Figure II-14. **Thermodynamic measurements.** (a) Schematic of the monomer/dimer equilibrium for “one-sided” origami with six active stacking patches in the middle (binary sequence ‘00011111000’). (b) The number of active stacking patches and their locations were varied as shown and the free energy was measured in each case. (c) A representative AFM image showing the distribution of the monomers and dimers as well as the distribution of bond orientations.

### II.2.4.1. Measurement of the free energy of stacking bonds

The free energy of the stacking bonds was measured by assuming that monomers and dimers of “one-sided” rectangle origami (origami with edge staples on only one side, Figure II-14a) were at equilibrium. The initial monomer concentration equals the total origami concentration, which was assumed to be the initial scaffold concentration (assuming the yield of origami formation was  $\sim 100\%$ ). The equilibrium concentrations of monomers and dimers were measured by depositing the samples on mica and counting the numbers of each in AFM images (e.g. Figure II-14c). Here the relative ratio of monomers and dimers on surface was assumed to be representative of the ratio in

solution. At least two processes could invalidate this assumption: (1) origami dimers might break upon deposition, artifactually elevating the monomer count, or (2) origami monomers might land so close to each other that they would be scored as a dimer, artifactually elevating the dimer count. We did not try to estimate the frequency of these processes but we did dilute the origami 5-fold from their formation concentration before depositing them; this decreased the probability of a mismeasurement due to (2). In other experiments, dilution was performed on the mica surface by pipetting a sample onto a five times larger volume of buffer on the surface. Because origami stick so quickly to mica, this protocol would run the risk of depositing dimers and monomers before they had the chance to equilibrate at the new concentration. To decrease the potential for this effect, sample solutions were pre-diluted, left to equilibrate for ~5 hours (a longer equilibration time, e.g., 10 hours, was tested for the  $p=6$  system and did not show a statistically significant difference, so we assumed that a five hour equilibration time was long enough for the  $p=6$  and weaker bonds), and then deposited without further dilution. (To be completely free from the effects of surface deposition and dilution and to obtain more detailed thermodynamic parameters, e.g.,  $T_m$ ,  $\Delta H$ , and  $\Delta S$ , one could alternatively adopt solution-based measurement techniques such as real-time FRET analysis<sup>72</sup>.)

The free energy of the stacking bonds was calculated as follows: from the counts of monomers  $M$ , correct dimers,  $D$ , and incorrect dimers (misalignments or other orientations, *other*), and the concentration of origami, [*origami*], we calculated the monomer concentration,  $[M]$ , and dimer concentration,  $[D]$ :

$$[M] = \frac{M [\textit{origami}]}{M + 2D + 2\textit{other}} \text{ and } [D] = \frac{D [\textit{origami}]}{M + 2D + 2\textit{other}}$$

From  $[M]$  and  $[D]$  we calculated the equilibrium constant and the free energy of the bond

$$K = \frac{[D]}{[M]^2} \text{ and } \Delta G = -R T \ln K$$

where  $R$  is the gas constant (8.314 J/mol·K) and  $T$  is the temperature 295 K (22°C).

#### II.2.4.2. First energy model: assuming loop-loop interactions are neutral

For the first model, we simply divided the binding energy by the number of blunt end stacks per bond ( $2p$ ) to arrive at  $\Delta G_{st}$ . For short DNA complexes, the free energy of hybridization is linear in the number of base pairs; thus we assumed that the total stacking bond energy would be linear in  $p$ , and  $\Delta G_{st}$  would be roughly constant. Surprisingly, stacking bond energy appeared quite sublinear in  $p$  and  $\Delta G_{st}$  increased from -2.6 kcal/mol for  $p=2$ , to -1.8 kcal/mol for  $p=4$ , to -1.4

kcal/mol for  $p=6$ . While this range of values encompasses that measured for ‘GC’ blunt-end stacking elsewhere<sup>44</sup>, such sublinearity is predicted to decrease the performance of stacking bonds: it would result in a smaller energy difference between correct and incorrect bonds than is predicted by a linear energy model, and cause a correspondingly higher rate of incorrect bonds. It also suggests that stacking sequences with the same  $p$ - $i$  but higher  $i/p$  might give higher error rates; this is consistent with our observations for (7, 4), (8, 5) and (9, 6) sequences (see section “Why use 7 active patches with a mismatch constraint of 4?”).

Sublinear binding energies have been reported before in DNA tiling systems using sticky ends<sup>73</sup>. Here, we hypothesize that sublinearity might derive from deformation of the edge caused by residual local twist (twist correction sets only the global average twist), potential curvature induced because all breaks in the phosphate backbone lie on the same side of the origami, or a combination of both. If a few nearby patches bind, they would not have to bend or twist much; thus strain will contribute little to the stacking bond energy, and a large  $|\Delta G_{st}|$  that closely reflects the free solution stacking energy will be observed. Our data for  $p=2$  indeed match free solution values fairly well (see below). In contrast, if numerous patches bind, strain will make a large contribution to the stacking bond energy, and  $|\Delta G_{st}|$  will be underestimated. Our hypothesis further suggests that the distribution of ‘1’s in a stacking sequence might affect bond energy, so we tested an additional sequence (Table II-6 and Figure II-15) for  $p=2$  and two additional sequences for  $p=4$ , with more spread-out active patches (e.g., ‘100100001001’). The data support our earlier assumption that bonds with identical  $p$  are *roughly* isoenergetic: for  $p=2$  no significant difference was measured; for  $p=4$  small (up to 0.2 kcal/mol) but statistically significant differences were measured. The trend for  $p=4$  is that spreading out active patches weakens stacking bonds, in agreement with the deformation hypothesis.

System <sup>&amp;</sup>	binary code	[origami] (nM)	$\Delta G_{st}$ (kcal/mol hx)	N (origami count)
2patch-(6,7)	000001100000	0.424	-2.5889	362
2patch-(5,8)	000010010000	0.848*	-2.6738	276
4patch-(5,6,7,8)	000011110000	0.424	-1.7644	178
4patch-(3,5,8,10)	001010010100	0.424	-1.6593	566
4patch-(1,4,9,12)	100100001001	0.424	-1.5578	360
6patch-(4,5,6,7,8,9)	000111111000	0.212 <sup>#</sup>	-1.4223	442

Table II-6. Free energy of the stacking bond per helix for various systems.

<sup>&</sup> Numbers in parentheses indicate the locations of active stacking patches (the 1’s in the binary sequences).

\*A higher concentration was used because it was hard to find dimers for this system.

<sup>#</sup> A lower concentration was used because it was hard to find monomers for this system.

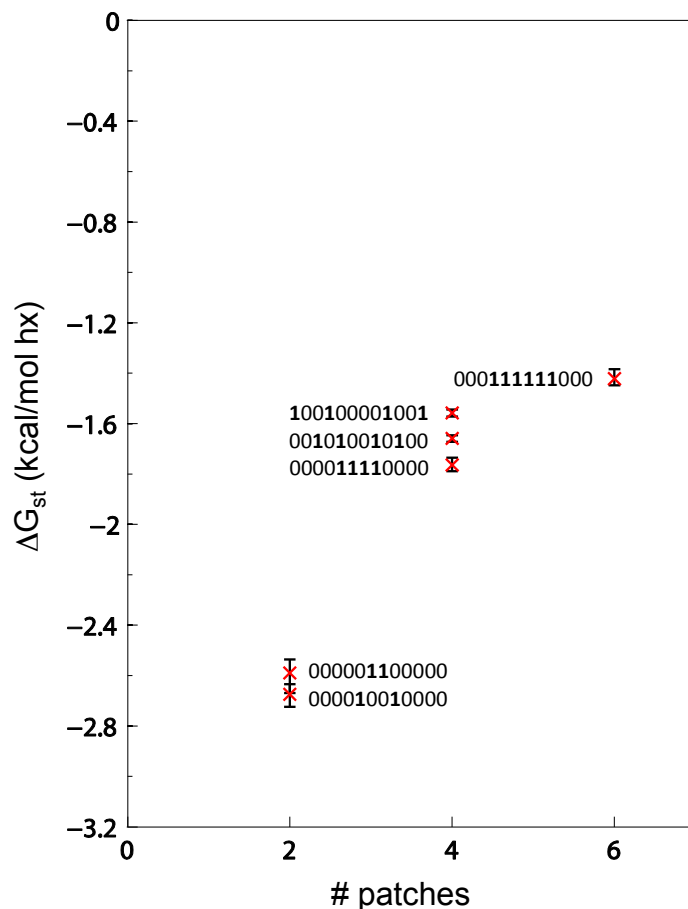


Figure II-15. **Free energy of the stacking bond per helix for various systems.** Energy values per helix vary depending on the number of patches, indicating a nonlinear relationship between the stacking energy and the number of helices. The overall trend (decreasing  $|\Delta G_{st}|$  as the number of patches increases) suggests that patches farther away from the middle of the edge must bend more (to counter some remnant global deformation) to bind; this hypothesis is consistent with the trend within the 4-patch systems. The binding energies for the 2-patch systems did not show a statistically significant difference (the error bars partially overlap). Error bars indicate standard error, obtained by bootstrapping the count data and propagating errors through the equations.

#### II.2.4.2.1. Comparison with free energy of stacking in literature

Because we hypothesize that non-stacking factors are all destabilizing, we suggest that the average energy obtained for the 2-patch systems,  $-2.63$  kcal/mol ( $1\times$  TAE with  $12.5$  mM  $Mg^{2+}$ ,  $22^\circ C$ ), is most reflective of a pure stacking interaction. One literature value<sup>44</sup> for the energy of GC/CG stacking, measured under a condition closest to ours, is  $-2.17$  kcal/mol ( $1\times$  TBE solution at  $37^\circ C$ ). The same group later reported temperature-dependent data under the same experimental setup<sup>62</sup>. While buffer conditions between our and their experiments differ, we did our best to make

the measurements comparable by correcting the literature value using the temperature-dependent data. Figure II-16 shows a plot that we reproduced based on experimental data given in the literature<sup>62</sup>: from their Figure 3a and Supplementary Table 2. Data were taken for five different temperatures (32°C, 37°C, 42°C, 47°C, and 52°C). Assuming that the temperature dependence of the enthalpy and the entropy of blunt-end stacking is negligible for the given temperature range, it is appropriate to make a linear fit to  $\Delta G_{st}$  as a function of temperature. A regression line and its equation ( $R^2 = 0.8943$ ) are shown in Figure II-16. Linear extrapolation to the y-axis ( $T=22^\circ\text{C}$ ) gives an energy of -2.42 kcal/mol at 22°C, which is very close to the value we obtained.

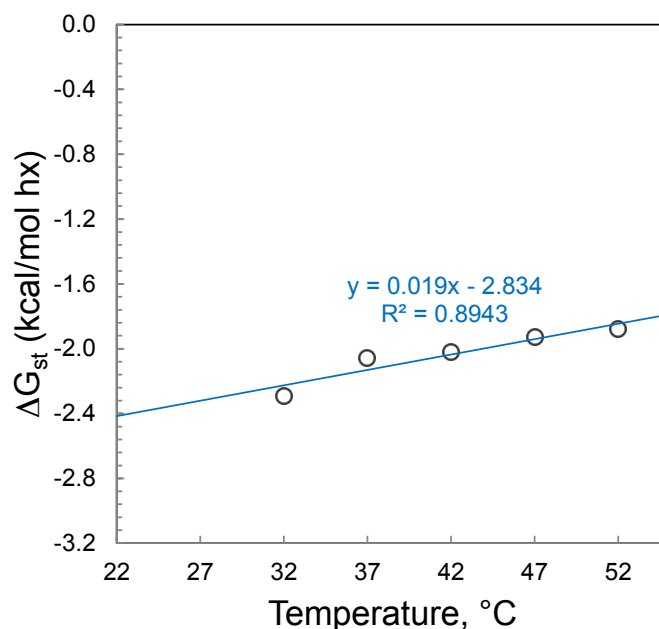


Figure II-16. **Temperature dependence of the stacking free energy (data taken from ref. 62).** A linear fit and its extrapolation gives a stacking free energy of -2.42 kcal/mol at 22°C, which is very close to the value we obtained, -2.63 kcal/mol, at the same temperature.

### II.2.4.3. Second energy model: fitting with non-zero loop-loop interactions

Under the second energy model, we examined the hypotheses that  $\Delta G_{ll} > 0$  (due to entropic brush interactions) or  $\Delta G_{ll} < 0$  (due to some hydrogen-bonding or stacking between loops), assuming a linear relationship between the free energy of binding and both the number of active patches (each contributing  $\Delta G_p = 2\Delta G_{st}$ ) and the number of inactive patches (each contributing  $\Delta G_{ll}$ ). For simplicity, for a given  $p$ , we averaged the measured free energies for different arrangements of active patches. This resulted in a single free energy for each of three values of  $p$  (2, 4, and 6) which allowed us to write the following three equations:

$$(1) 2\Delta G_p + 10\Delta G_{ll} = -10.53$$

$$(2) 4\Delta G_p + 8\Delta G_{ll} = -13.28$$

$$(3) 6\Delta G_p + 6\Delta G_{ll} = -17.07$$

(all in kcal/mol)

Here we have more equations than unknowns. In principle, for this system of linear equations to be consistent, the intersection of each pair of equations should coincide exactly. In practice, because of experimental error and potential sequence-dependent effects, we expect that the intersections should lie in close proximity to each other. Figure II-17 shows a plot of the three equations. The intersections occur at  $(\Delta G_p, \Delta G_{ll}) = (-2.03, -0.65)$ ,  $(-2.24, -0.60)$ , and  $(-2.37, -0.48)$ , for equations (1) and (2), equations (1) and (3), and equations (2) and (3), respectively, and they all give very similar estimates for  $\Delta G_p$  and  $\Delta G_{ll}$ . Least squares analysis gives a solution of  $(\Delta G_p, \Delta G_{ll}) = (-2.23, -0.59)$  with a root mean square error of 0.24 (which can be roughly interpreted as the average distance of the solution from each intersection in the plot of  $\Delta G_{ll}$  vs.  $\Delta G_p$ ). This solution yields  $\Delta G_{st} = -1.12$  kcal/mol and  $\Delta G_{ll} = -0.59$  kcal/mol.

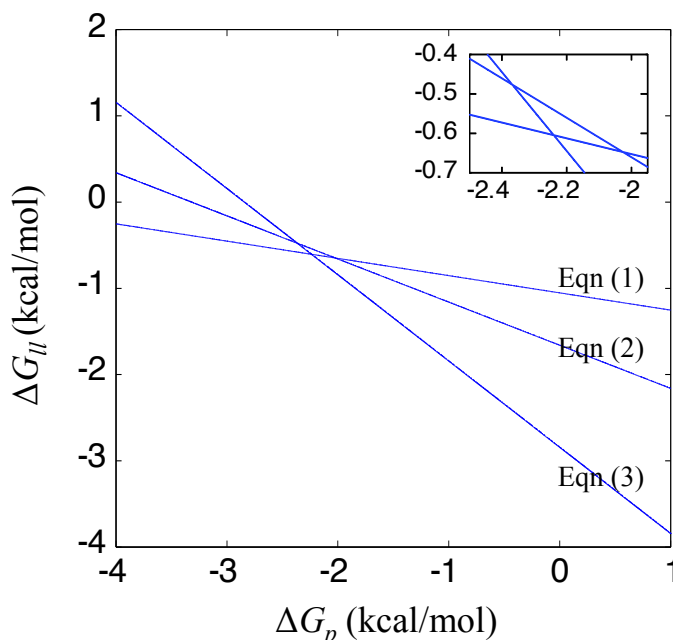


Figure II-17. **Plots of the three equations given for  $\Delta G_p$  and  $\Delta G_{ll}$ .** Least squares analysis gives a solution of  $(\Delta G_p, \Delta G_{ll}) = (-2.23, -0.59)$  with a root mean square error of 0.24. Inset shows a zoom-in view of the plot near the intersections of the three lines.

Thus we find that the average free energy of loop-loop interactions ( $\Delta G_{ll}$ ) is negative (suggesting that loop-loop interactions contribute favorably to the binding) but small—less than half the average free energy of a single base pair,  $-1.41$  kcal/mol (nearest neighbor model<sup>65</sup>). It would be

interesting to ask whether the average loop-loop interaction is typical, or whether most loop-loop interactions are neutral and just a few inactive patches contribute most of the binding energy. Answering this question will require more experiments, in particular measurements of the binding energy for stacking bonds that have the same stacking sequence, but loops of different base sequence. Another observation is that  $\Delta G_{ll}$ , the binding energy of a pair of inactive patches, is one-fourth the free energy for an active patch  $\Delta G_p$ . Its effect on stacking bond strength will depend not just on this ratio, but the number of inactive patches used. For a 16-patch stacking sequence with 7 active patches, the 9 inactive patches will make a contribution to the binding energy that is roughly equivalent to two active patches, and about one-fourth (coincidentally) of the total free energy of the bond. With respect to predicting the ratios of correct vs. incorrect bonds, the contribution of loop-blunt end interactions (which occur frequently in mismatch incorrect bonds) will likely have to be included; so far we have no quantitative data that address such interactions. Finally, we observe that if this model is correct then we must reconcile the relatively small  $|\Delta G_{st}|$  observed (1.12 kcal/mol) with much higher literature values (2.42 kcal/mol). It may be partially due to the difference in measurement methods, or it might suggest that the near-B-form stacks (see Section “Crossover geometry at blunt ends”) which occur in “relaxed” edges do, in fact, have a somewhat smaller free energy than the regular B-form stacks. Future experiments using “crossover-free” edges may address this question.

In sum, this analysis suggests that assigning a small attractive energy for loop-loop interactions eliminates the need to interpret stacking bond energies as nonlinear; on the other hand loop-loop interactions are likely to be highly sequence specific, and assigning a single average energy to all of them is unsatisfying. While the  $|\Delta G_{st}|$  estimated by our second model is smaller than stacking energies measured elsewhere, the data suggest that, on average, the bond strength for an active patch ( $2\Delta G_{st}$ ) is significantly ( $\sim 4$  times) stronger than the loop-loop interaction of an inactive patch. Neither of our energy models is perfect, but together they highlight the extent to which the behavior of real stacking bonds might depart from our assumptions.

### II.3. Conclusion and Future Directions

Our goal was to develop new systems in which we could create large sets of orthogonal and isoenergetic bonds. Starting from relatively large sequence spaces, we achieved the creation of relatively small sets of bonds: the flexibility of DNA helices forced us to pick bond types that were both simple and rigid, and thus less numerous. Can one do better? To solve this problem in the



binary code system, inactive patches could be implemented by more rigid non-stacking “steric blockers” (which would prevent neighboring active patches from bending) or replaced by “weakly active patches” (with ‘AT’ on their blunt ends). Stacking polarity has revealed itself to be a strong symmetry breaker. This suggests that inactive patches could also be replaced by active patches with an opposite stacking polarity, creating a stacking polarity-based binary code. For both binary and shape coding systems, the flexibility of active patches could be directly addressed by increasing the density of crossovers near the edge, or using multi-layer 3D origami<sup>16,68</sup>, rather than single-layer origami (for such different architectures, strand polarity might have to be redefined). Hybrid codes<sup>59</sup>, which simultaneously use binary coding and shape coding, offer another possible route to greater bond diversity without directly addressing flexibility. Above all, it will be important to have better energy models so we can maximize the difference between correct and incorrect bonds.

More immediately, stacking bonds offer a couple of practical advantages over DNA hybridization for the hierarchical assembly of origami into larger, more complex structures. When origami are joined by DNA strands such as “sticky-ends”, each new origami-origami interaction requires the design and synthesis of unique sequences. Our binary-coding approach allows bond type to be reprogrammed easily and cheaply, post-synthesis, merely by pipetting a different subset of the edge staples. Further, neither of our approaches requires the purification usually needed for sticky-end-based approaches: the binary coding approach requires quenchers, but the shape coding approach allows the direct coupling of origami without additional steps. One disadvantage of stacking bonds is that the total binding energy is limited by the size of the origami: a greater range of binding energies might be achieved in a sticky-end-based approach by changing sticky-end lengths.

So far, we have concentrated on replicating the combinatorial nature of DNA hybridization, and ignored its other useful properties. In particular we have ignored *strand displacement*<sup>30,31</sup>, the dynamic mechanism by which a partially-complementary duplex and a third, free strand of DNA can rearrange (without initial dissociation of the duplex) to form a duplex with greater complementarity, thus freeing one of the two originally duplexed strands (Figure II-18a). This mechanism is the foundation for a large number of DNA nanomachines and circuits, both providing a fuel source to drive non-equilibrium reactions and enabling the order in which reactions occur to be programmed. An analogous displacement mechanism for stacking bonds might allow programming of large-scale rearrangements of origami, in the context of much larger DNA nanomachines (Figure II-18b).

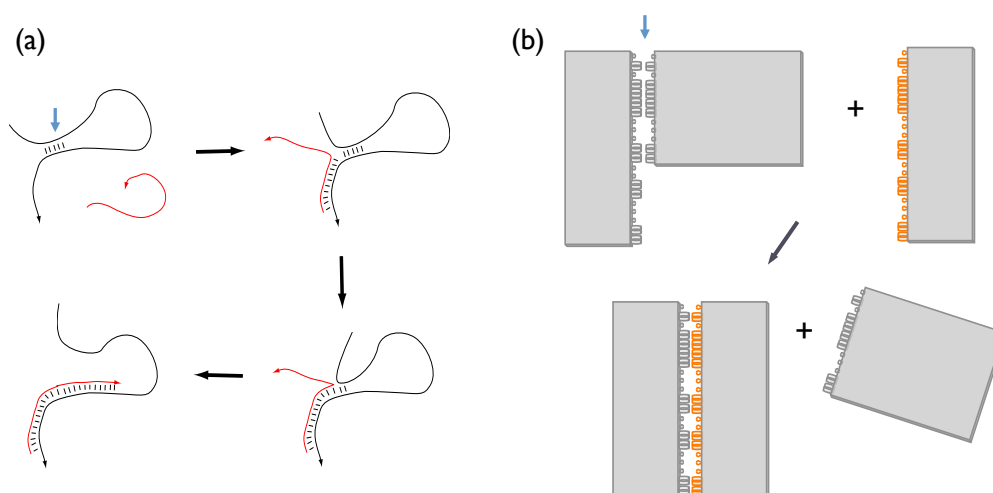


Figure II-18. **Dynamic displacement mechanisms.** (a) For DNA strands, a strand with a longer binding region (red) can displace a strand with a shorter binding region (indicated by a blue arrow, where partial self-binding occurs within the black strand). Figure adapted from ref. 1. (b) For origami nanostructures connected by stacking bonds, a similar displacement mechanism could be achieved based on competition between a structure with a longer binding region (orange) and an initial bond with a shorter binding region (indicated by a blue arrow), which might allow programming of large-scale rearrangements of DNA nanostructures.

Looking ahead, we believe that stacking bonds will offer important new mechanisms for force-induced bond rearrangement that are difficult or impossible to implement with DNA hybridization. In order for DNA structures linked by DNA hybridization to rearrange, DNA helices must unwind and then rewind; this could involve a high activation energy, depending on the number and strength of the links involved. Stacking bonds, on the other hand, may exhibit low activation energies for sliding under shear forces. If so, we hope to use stacking bonds between mechanical parts that must both self-assemble initially, and then slide past each other (while maintaining contact) under the shear forces applied by molecular motors—many biological nanomachines meet this challenge<sup>36,74</sup> and such a capability is fundamental if we are to build truly complex nanomachines.

Finally, we return to the question, “What causes two complementary DNA strands to bind?” One answer is that base stacking is the dominant stabilizing force and the specificity derives from base pairing. But base pairing has a couple of components: in addition to the interaction between hydrogen bond donors and acceptors, there is the geometric complementarity of the base pairs. These factors are difficult to disentangle since, without geometric fit, hydrogen bonds could not form. However, in certain contexts it appears that geometry alone underlies specificity: geometrically complementary base analogs can be incorporated into DNA by a polymerase in the

*absence* of hydrogen bonding<sup>75</sup>. Hence perhaps the role of base pairing is mostly to provide a geometric framework—aligning the two bases in a plane—that encourages and allows stacking if and only if the bases are complementary, and the molecular recognition of DNA hybridization may be thought of as mostly “stacking interactions given specificity by geometric complementarity”. Here we have used DNA origami as a geometric framework to align complementary sequences of blunt-end stacking interactions. Thus, in a sense, DNA hybridization and our systems work on a very similar principle—the geometric relationships between stacking interactions in our systems just operate at a *ten-times* larger scale.

## II.4. Materials and Methods

### II.4.1. General

Single-stranded M13mp18 DNA (scaffold strand) was purchased from New England Biolabs (Catalog #N4040S) and staple strands were obtained unpurified from Integrated DNA Technologies in water at 150  $\mu$ M each.

### II.4.2. Design of origami

Design of the rectangle systems (regular and tall) was performed based on the procedure described in the original DNA origami paper<sup>1</sup>, using variations of the original MATLAB code. Origami with edge shapes (**A**, **B**, **C**, and **D** origami) and corner origami were designed using a modified version of the “square lattice” version of caDNAno<sup>69</sup>: to facilitate the process of designing origami with uniform-sequence (specifically ‘GC’) blunt ends, caDNAno was customized to allow, (1) the creation of single stranded loopouts in the scaffold strand, and (2) the highlighting of the scaffold strand with a user-specified sequence. In our modified version of caDNAno, the “loop tool” which is normally used to generate double-stranded loops (loops involving both the scaffold and staple strands) has been changed to a tool that creates single-stranded loops only in the scaffold strand. The highlighting feature has been integrated into caDNAno’s existing “add sequence” function, with which a user can select the scaffold sequence to use; now a user can specify a sequence to highlight using the same dialog box (Figure II-19c). Color is fixed for each highlighted base: G-Green, C-Cyan (light blue), A-Amber (roughly orange), T-Tomato (red). The length limit for a highlighted sequence is set to be the same as caDNAno’s existing length limit for the scaffold

sequence (20,000), so a user can highlight the entire scaffold sequence if desired. A few other minor modifications include: (1) skipping the step of waiting for the user to click on the 5' end of the scaffold during an “add sequence” operation if there is only one 5' scaffold end in the design, and (2) updating the scaffold and staple sequences automatically after creating a new loopout (this automatic update works only after the scaffold sequence has been defined). For general instructions and the original version of caDNAno, visit <http://www.cadnano.org>.

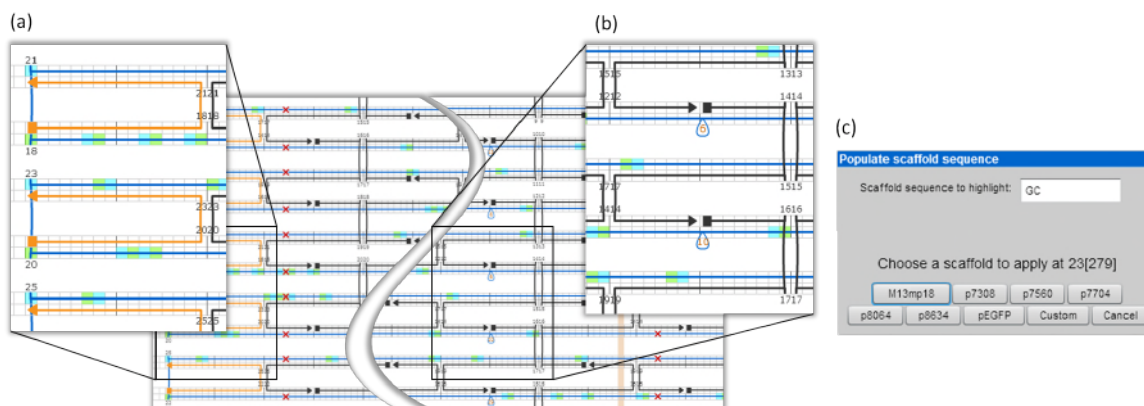


Figure II-19. Screenshots of a version of caDNAno modified to allow placement of ‘GC’ on origami edges. All occurrences of the sequence ‘GC’ are highlighted along the scaffold strand in the diagram by green and cyan for G and C, respectively. (a) One can count the number of bases from a blunt end on the edge to the next occurrence of ‘GC’ and (b) make a single-stranded loopout inside the structure to *shift* the scaffold sequences. In the example design shown, the scaffold strand has been already aligned to have ‘GC’ along the edge (as shown in the zoom-in at (a)). (c) Here we have shown highlighting of the sequence ‘GC’. One can enter any sequence to highlight at the step where user selects the scaffold strand.

### II.4.3. Preparation of origami

Individual origami structures that were not destined to be mixed with other structures were prepared by a protocol similar to that presented in earlier work. Scaffold strands and staple strands for each design were mixed together to target concentrations of  $\sim 2$  nM and  $\sim 75$  nM, respectively, in  $1\times$  Tris-Acetate-EDTA (TAE) buffer with 12.5 mM magnesium acetate (TAE/Mg<sup>2+</sup>). The mixtures were kept at 90°C for 5 min and annealed from 90°C to 20°C with a constant rate of  $-1^\circ\text{C}/\text{min}$ .

To create origami chains with multiple bonds based on binary sequences (as shown in Figure II-4c), constituent origami were first annealed separately from 90°C to 20°C. Next, corresponding quencher strand mixtures for each origami were added (at  $10\times$  the edge staple concentration) to each origami mixture. Each of the solutions was kept at room temperature for 1 hr to ensure

complete hybridization, and then they were mixed together, heated to 50°C, kept for 12 hr at 50°C, and then cooled to 20°C at a rate of -5°C/hr.

For the origami chain (**A-B-C-D**) and dimers (**A-B**, **B-C**, **C-D**) with shape complementarity (as shown in Figure II-8c,d), each origami mixture (scaffold + corresponding staples) was annealed separately from 90°C to 50°C (with a rate of -1°C/min), mixed together at 50°C, and kept at 50°C for 12 hr, then cooled to 20°C at a rate of -5°C/hr. The mixing operation was performed inside a temperature-controlled chamber (Coy Laboratory Products Inc.), to maintain the temperature at 50°C while the samples were transferred between test tubes.

#### II.4.4. Atomic force microscopy (AFM)

Samples for AFM imaging were prepared by depositing 5  $\mu\text{l}$  of the origami solution with 20  $\mu\text{l}$  of TAE/Mg<sup>2+</sup> buffer onto freshly-cleaved mica (Ted Pella). In most cases, clean buffer solution was deposited first and the origami solution was added on top of it. (For concentrated samples we felt this procedure minimized spatial variation in the density of origami on the mica.) In cases wherein we were concerned that this procedure might distort data (i.e., for thermodynamic data, see the section “Thermodynamic Measurements”) we pre-diluted the origami solution by 5-fold, and then deposited 25  $\mu\text{l}$  onto mica. AFM images were taken under TAE/Mg<sup>2+</sup> buffer in Tapping Mode with a Nanoscope III Multimode AFM (Veeco Metrology Group, now Bruker AXS). Typically, we used silicon nitride cantilevers with 2 nm radius silicon tips as AFM probes (the “short, fat” A cantilever on SNL probes from Veeco, now Bruker AFM Probes).

#### II.4.5. Detail aspects in the design process

##### II.4.5.1. Searching for large orthogonal sets of sequences

For the construction of non-orthogonal candidate sets of sequences, for both binary sequences and shape sequences, we have largely discussed the symmetry and mismatch constraints that are required of a sequence for it to be useful as an individual stacking bond in isolation—constraints such that, with high probability, the sequence will bind its complement by a full correct bond rather than binding to itself or binding its complement by a partial bond. Given a set of parameters including the length of the sequence, number of active patches, the number of depths (if appropriate), and a mismatch constraint, we have shown that it is straightforward to enumerate all

sequences that individually satisfy the constraints. Our two most studied examples are the 4614 binary sequences for  $(p,i) = (7,4)$  and the 16 shape sequences for  $(p,d,i) = (4,3,2)$ . In order for such candidate sequences to be used in a multiple-bond system, one must find a subset that is orthogonal—that is, all pairs of sequences must satisfy the mismatch constraint. A diagram of the orthogonality relation between all 16 sequences in the  $(4,3,2)$  shape sequence candidate set and example orthogonal subsets are discussed and diagrammed in the section “Orthogonality graph for the  $(4,3,2)$  candidate shape sequences”.

An important goal is to find the *maximal* orthogonal subset of a candidate set, to find a code of sequences that can support the largest diversity of stacking bond types. For candidate sets containing relatively few sequences, such as the  $(4,3,2)$  shape sequences, an exhaustive search through all possible orthogonal subsets is possible. But for bigger candidate sets, the combinatorially large number of subsets makes finding the maximal orthogonal subset by exhaustive search a computationally intractable task.

More specifically, the problem of finding the maximal orthogonal subset is a trivial rephrasing of the well-known *Max Clique* problem in computer science. Max Clique is known to be NP-hard, and here two facts about NP-hardness are relevant: (1) most computer scientists believe that NP-hard problems can only be solved exactly by using an amount of time that is an *exponential* function of the size of the problem—this is what is meant by “computationally intractable”, (2) NP-hard problems can sometimes be approximated—thus, while it might be computationally intractable to find the maximal orthogonal subset, it may be possible to find large subsets, which are close in size to the maximal orthogonal subset, quickly.

There is a large literature on approximating NP-hard problems, but we did not take advantage of such approximation techniques here. Instead, to quickly get large orthogonal subsets that we could use for multiple stacking bonds, we implemented a simple, randomly seeded, greedy search procedure. For many of the smaller candidate sets, the orthogonal sets we typically obtain are probably maximal; in the case of the  $(4,3,2)$  shape code system we verified that this was the case. For larger candidate sets it is highly likely that the orthogonal sets we have obtained are not maximal; it might be possible that there exist 13-sequence orthogonal sets for the  $(7,4)$  binary code system. For the moment, the orthogonal sets that we have found are satisfactory since we obtained orthogonal sets whose size roughly matches the maximum number of origami that we can handle easily, or can afford in the lab. However, if the need should arise, e.g., for much larger codes, many relatively fast algorithms for finding large cliques (and hence finding large orthogonal subsets) are available. One example is *Cliquer*, a set of C routines that are available for download from:

<http://users.tkk.fi/pat/cliquer.html>

Our program for discovering large orthogonal subsets constructs them in a “greedy” fashion starting from single sequences from the candidate set. Let the size of the candidate set be  $N$ . Each run of our program constructs  $N$  different orthogonal subsets, by sequentially using each one of the elements of the candidate set as a seed for a different orthogonal subset. The details of our program are as follows:

(1) The program first picks one candidate sequence as a seed; that candidate sequence becomes the first element of the orthogonal set under construction.

(2) The program randomly picks another sequence from the candidate set and checks its orthogonality with the existing element(s), with respect to the mismatch constraint  $i$ .

(3) If the newly picked sequence is orthogonal to the existing element(s), it is added to the set; otherwise it is discarded.

(4) The program repeats steps (2) and (3) until all candidate sequences have been tested for the orthogonality with the growing set. After all candidates have been tested, the orthogonal set is output.

(5) The program repeats steps (1) through (4) until all candidate sequences have been used as a seed for an orthogonal set.

Since the construction of an orthogonal set is sensitive to the order of addition of candidate sequences (a different order results in a different set), each run of the program results in  $N$  potentially distinct orthogonal sets. Typically, we ran the program multiple times; we did not keep track of the number of runs performed. The largest orthogonal subsets found are recorded in Table II-1 (for binary sequences) and Table II-3 (for shape sequences).

#### II.4.5.2. Warnings: potential technical issues in designing DNA origami

In case one wants to repeat or adopt some of our experiments, we give warnings that describe some difficulties which we have encountered and suggest some potential problems that we did not discuss so far.

##### II.4.5.2.1. Length and width of a patch in shape design

Besides the 4-patch design in the shape code system, we have tried other designs with higher complexity (6-patch and 9-patch systems) that we expected to give higher specificity. But as the number of patches increased, we had to design each patch with less material, yielding patches with a smaller number of helices. The flexibility of DNA, coupled with the strength of the stacking

interactions, caused these “narrow-patch” systems to be more vulnerable to bent-patch bonds. Figure II-20 briefly summarizes the two systems.

In the 6-patch design (Figure II-20a) we introduced physical gaps between each adjacent pair of 2-helix patches, to minimize any effect of the electrostatic repulsion. (It seemed possible that electrostatic repulsion between adjacent patches might decrease binding energy. This hypothesis has yet to be adequately tested.) The introduction of physical gaps made the patches narrower and longer, allowing various kinds of bent-patch bonds ( $\sim 30\%$  of all bonds) as shown in Figure II-20d,e. Because we used 3 helical turns for each depth increment, the length of the longest protruding patch was 9 helical turns, which is about 30% of the persistence length of double-crossover DNA tiles ( $\sim 30$  helical turns,  $\sim 100$  nm)—the most similar structure to the 2-helix patch structure for which the persistence length is known<sup>76</sup>. We chose to use a (6,4,3) shape code so that, in principle, the maximum-strength partial bond would have three active patches (half of the full strength). To our dismay, 5-patch bent-patch bonds formed; if the bending energy were small, these bonds would have a binding energy comparable to that of full 6-patch correct bond.

In the 9-patch system (without physical gaps and with much shorter patches), a significant fraction of the bonds ( $\sim 20\%$ ) were still bent-patch partial bonds (Figure II-20i,j). We had chosen to use a (9,5,2) shape code so we had expected high binding specificity—the strongest expected incorrect bonds would have a binding energy  $2/9$  of a full correct bond. To decrease the flexibility of patches, we used a single helical turn for each depth increment, so that the longest protruding patch was just 4 helical turns in length. Thus it was to our further dismay that 8-patch bent-patch bonds formed, which were again potentially very close in energy to full-strength bonds. As a point of interest we note that the 6- and 9-patch systems were not twist-corrected, so the chains in Figure II-20g show the characteristic breaking pattern that is similar to that shown in Figure II-1b. The global twist might be playing a role in encouraging bent-patch bonds in these systems, but we have not done any experiments to test this possibility.

To decrease flexibility, our final “successful” shape code system employed only four patches that were 4-helices wide and protruded at most 6 helical turns. Many questions remain: How many patches are optimal for this kind of study? How wide (in terms of number of helices) should each patch be? How long can they be? What is the bending energy of the patches under the buffer condition used? We do not yet have answers to these questions, but it is certain that there is a trade-off between the complexity (and hence the potential specificity in ideal case without helix bending) and the bond reliability. Of course, this problem is limited to “soft” systems like DNA, thus might be avoided in a system with sufficient rigidity.



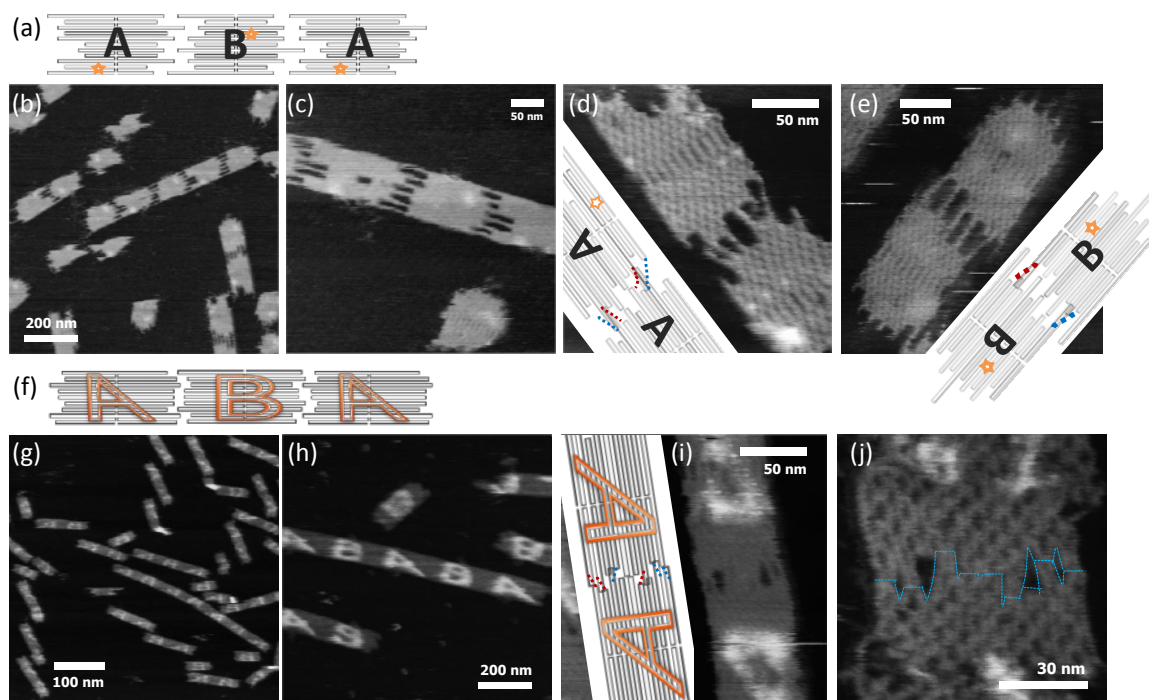


Figure II-20. **Performance of 6-patch and 9-patch shape-coded systems.** (a) Models of the 6-patch system. The edge shapes of the **A** origami and **B** origami were designed such that the origami form continuous alternating **AB** chains. Shape sequences were ‘132120’ for **A-B** bonds and ‘011310’ for **B-A** bonds. Stars indicate the locations of dumbbell hairpins, which serve as topographic labels for AFM. (b) & (c) Typical AFM images of the system that show full-strength correct bonds. (d) & (e) Typical bent-patch bonds which manage to bind via 5 active patches. The red dotted lines on the models depict bent patches coming from the origami on the top, and the blue dotted lines depict bent patches coming from the origami on the bottom. (f) Models of the 9-patch system. Shape sequences were ‘034222043’ for **A-B** bonds and ‘340224301’ for **B-A** bonds. (g) & (h) Typical AFM images. Note that the chains in (g) show the characteristic breaking pattern of the twisted origami chains described in Figure II-1. (i) & (j) Typical bent-patch bonds with 8-patch bond strength. The red and blue dotted lines in (i) are used in the same way as in (d) or (e). The narrow blue dotted line in (j) roughly follows the blunt ends and helical sidewalls of the edge structures.

#### II.4.5.2.2. Potential interference from the remainder staples

Since the length of the scaffold strand is fixed to be 7249 bases, a DNA origami design that uses fewer than 7249 bases will leave a *remainder* in the form of unfolded single-stranded scaffold—in most designs the remainder takes the form of a single loop. To avoid potential interactions of such a single-stranded remainder on one origami with the remainder of another origami, it is usual to add a set of *remainder staple strands* which have the function of hybridizing to the remainder and turning it into an unreactive double-stranded loop.

When multiple origami which do not share the same underlying design are mixed together, e.g., as in our **A-B-C-D** chains with shape complementarity, there is the possibility of interference between the remainder staple strands of one origami and single-stranded regions of the other origami. In general, a subset of staple strands from one origami may bind single-stranded loopouts on other origami via partial complementarity. (Such loopouts are common in our system because they are used to enforce the ‘GC’ sequence constraint at the blunt ends of helices.) Binding of staple strands to loopouts does not, in general, seem to affect the origami, but in certain cases remainder strands may have complementarity to surrounding scaffold sequence outside of the loopout. In such cases the remainder strands can begin to displace nearby staples. Because the remainder staples are designed to be “continuous” complements to the remainder loop, unlike any other regular staple strands, each successive remainder staple that displaces a regular staple potentially opens up a site for another remainder to bind. Remainder staples may thus sequentially unfold the local structure of another origami. This process may be energetically favorable because the remainder strands make a continuous duplex which likely has a lower energy than origami structure (because of its crossovers and twist strain). In some of our initial experiments on shape complementary origami, we experienced this problem: individual origami folded well, but when mixed together remainder staples from one origami caused large structural disruptions in other origami (data not shown).

Two potential solutions exist: (1) one can avoid the use of remainder staples—in most cases single-stranded remainder sections of the scaffold will cause no further problem. (2) One can design the remainder loops of different origami to coincide (have almost the same sequence), so that the remainder staples of one origami will not bind and invade loopouts of another origami. The latter approach was used successfully in our **A-B-C-D** chain system.

#### II.4.5.2.3. Possible collisions between edge staples

When designing an origami system with uniform edge sequences (e.g., ‘GC’) as in our system, if one takes the same approach as ours—generating loopouts to shift the scaffold sequences—one should note that doing so limits the number of possible edge staple strands. In the 7249-base sequence of the M13mp18 scaffold strand, there are 393 occurrences of ‘GC’ (occurring on average every 18.4 bases, see schematic figure in Figure II-21). Hence, ideally, there are 393 different positions at which edge staples can be located. Given a particular geometric design for an origami, one has some choice in terms of which edge staple positions to use: one can change the edge staple sequence at a particular geometric position in the origami by changing the length of the loopouts and/or changing the position at which the scaffold sequence starts in the design. However,

when designing multiple origami that are large and use up all the sequence, or further, when the origami designs share a similar “start position” for the scaffold sequence (as occurs in our shape-coded **A**, **B**, **C**, **D** system), there is a high probability that some of the edge staples from different designs will share subsequences or have identical sequence. For the shape code system we explored, this does not cause any difficulties since all edge staple positions are occupied, by design. However, in some potential systems (say a hybrid shape code/binary code system), it might be possible for an edge staple present in one origami to fill in an empty edge staple position in another origami and give unexpected results. For example, in some of our initial experiments (not shown), edge staple collisions resulted in unintended aggregation. We note that taking an adapter strand approach to controlling edge sequences (as suggested earlier for potentially achieving crossover-free edges) would obviate this problem.

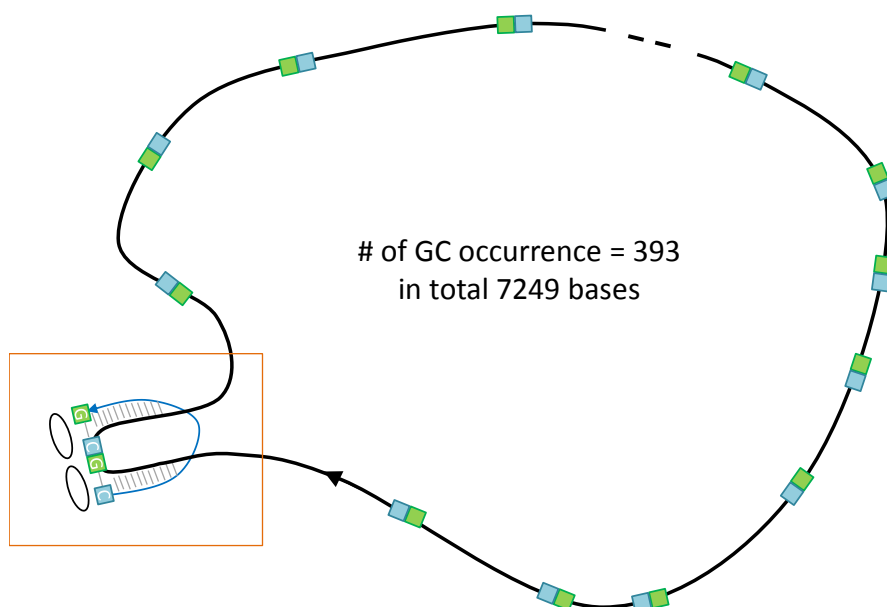


Figure II-21. **The limited number of ‘GC’ occurrences in the scaffold strand constrains the number of usable edge staple strands.** In case of the M13mp18 scaffold strand, with 7249 bases in total, there are 393 occurrences of ‘GC’. The black circular strand represents the scaffold. The boxed area shows how an edge staple strand (blue) binds to the scaffold and forms two ‘GC’ blunt ends (depicted by ellipses).

## II.4.6. Supplementary materials

Additional supplementary materials listed below are available for download at:  
<http://www.nature.com/nchem/journal/v3/n8/full/nchem.1070.html#supplementary-information>,  
 or by request to [woo@dna.caltech.edu](mailto:woo@dna.caltech.edu) for the most up-to-date versions.

- Sequence lists and diagrams of DNA origami structures
- caDNAno design files for shape systems (origami A,B,C, and D) and corner origami designs
- Computer program codes for designing bond types (MATLAB files)
- Installer for modified caDNAno program, “caDNAnoSQ\_SW”

## *Chapter III*

# Two-Dimensional Crystallization of DNA Origami Checkerboards via Cation-Controlled Surface Diffusion

*"Nothing happens until something moves."  
- Albert Einstein*

### **III.1. Introduction**

DNA allows programming of nanoscale shapes and patterns<sup>1</sup>, whose resolution cannot yet be met with current top-down micro/nanofabrication methods. Transferring self-assembled DNA nanostructures onto substrates is a promising approach towards patterning surfaces with nanoscale resolution for potential technological applications. However, depositing large-scale self-assembled DNA nanostructures on surfaces in a reliable way is challenging because the sample transfer process from solution to surface can cause breakage or distortion of the product structures. Assembling DNA nanostructures directly on substrate surfaces can eliminate the sample transfer process. There have been studies where small DNA motifs were allowed to self-assemble on substrate surfaces to form large lattices<sup>77-79</sup>, but technologically more interesting systems would be at a scale large enough to be reached by the lithographic regime, which was proved to be able to direct the orientation of DNA origami nanostructures on substrates<sup>80</sup>. At the same time, the use of DNA origami allows higher complexity and access to unique addressability in each structural unit. Here, we report large-scale self-assembly of DNA origami into two-dimensional checkerboard

pattern crystals based on surface diffusion, where the diffusion of DNA origami occurs on the substrate surface and is dynamically controlled by changing the cationic condition of the system. Unlike previous surface-based self-assembly studies of DNA nanostructures<sup>77-79</sup>, our protocol can operate at an isothermal condition without the need of a thermal annealing process, which gives a further practical advantage. The bonds between origami were mediated by blunt ends at four corners that are capable of connecting through stacking interactions for the following two reasons: first, to allow potential rearrangements during the surface assembly process via a putative *sliding* mechanism (as opposed to using sticky ends which may require unwinding and rewinding of the strands), and second, to achieve symmetry between the corners (each corner would be able to form the same kind of bond regardless of the orientation). We designed and tested origami that have a defined number ( $N \in \{3, 5, 7\}$ ) of blunt ends at each of the four corners (Figure III-1).

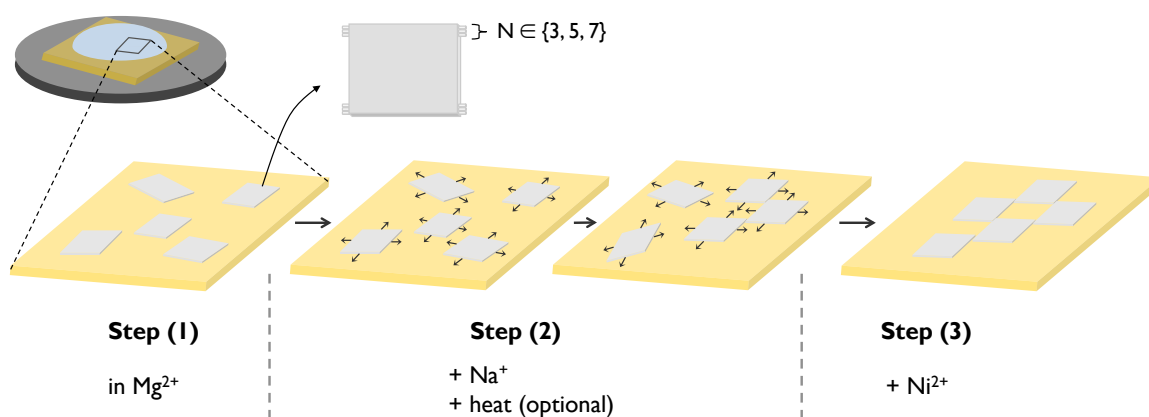


Figure III-1. **Schematic of origami design used for this study and the surface diffusion protocol.**

For the two-dimensional assembly, (1) we first deposit origami on a mica surface, then (2) change the buffer condition to allow them to diffuse around within the confined space near the surface, and then (3) change the buffer condition again to fix them back on the surface (Figure III-1). For the first step of depositing origami on a mica surface, we use a buffer with divalent cation  $Mg^{2+}$ , the condition under which origami are formed and typically imaged by AFM. For allowing origami to diffuse around near the surface, we exchange the buffer to a solution that contains mostly the monovalent cation  $Na^+$ . Monovalent cations weaken the binding of DNA onto mica surfaces<sup>81,82</sup> and have often been observed to release DNA previously adsorbed onto mica<sup>83,84</sup>. At this step we also apply heat (40°C, with ~4 hr incubation), but it turns out that heat is not an essential factor although it facilitates the process. Then, in the next step, we exchange the buffer back to a divalent cation condition,  $Ni^{2+}$  instead of  $Mg^{2+}$  at this time, to stop diffusion and “freeze” the product structures onto the surface:  $Ni^{2+}$  is known to bind DNA onto mica more strongly than  $Mg^{2+}$  because

$\text{Ni}^{2+}$ , as a transition metal ion, can form more stable coordination complexes with surrounding negatively charged ligands<sup>85,86</sup>. It has been shown that  $\text{Ni}^{2+}$  ions enable a permanent binding of DNA to mica even at high  $\text{Na}^+$  concentration<sup>83</sup>.

Our protocol is based on the previous findings in the literature that the surface mobility change and reversible binding of long duplex DNA on mica can be achieved by changing the cationic conditions of the system, such as the concentration of a divalent cation ( $\text{Zn}^{2+}$ )<sup>87</sup> and the relative concentrations of divalent and monovalent cations<sup>83,84</sup>. Mica is a layered mineral with a negative surface charge, and DNA is a negatively charged polymer. Thus, divalent cations are thought to mediate the binding between DNA and mica by bridging the negative charges on the DNA and the mica<sup>81-83,86,88</sup>. Such direct binding of DNA to mica can be considered a *short-range* attraction<sup>82</sup>, which induces strong adsorption of DNA onto mica and practically allows stable imaging of DNA by AFM. The surface diffusion phenomenon, during which our origami rearrange to form checkerboard crystals, is considered to be governed by a *long-range* interaction<sup>82</sup>: though the monovalent cation mostly neutralizes the charges on the DNA and the mica, weakening the binding interaction, the divalent cation pre-existing and pre-adsorbed on the mica surface effectively inverts the charge of the mica surface and prevents the DNA from completely leaving the surface<sup>82</sup>. In other words, the addition of  $\text{Na}^+$  into the  $\text{Mg}^{2+}$  solution drives the system to the regime of a *non-adsorbing*<sup>84</sup> but long-range-attractive<sup>82</sup> interaction between the DNA origami and the mica surface. Adsorbed  $\text{Mg}^{2+}$  is believed to persist on the mica surface during step (2) for two reasons. First, when exchanging the buffer, the original  $\text{Mg}^{2+}$  buffer is not removed, but only diluted. We leave  $\sim 5\mu\text{l}$ , to which we add 40  $\mu\text{l}$  of the  $\text{Na}^+$  solution. Second, Mg ions have higher valence than Na ions which would exert a stronger electrostatic attraction to the negatively charged centers on mica, and the ionic radius of  $\text{Mg}^{2+}$  ( $\sim 0.65 \text{ \AA}$ ) allows the ions to “fit” into cavities near the negatively charged centers in the atomic structure of mica, further ensuring stronger binding (the cavities are compatible with ions with radii  $< \sim 0.74 \text{ \AA}$ ;  $\text{Na}^+$  has a  $\sim 1 \text{ \AA}$  radius)<sup>86</sup>.

We discuss in detail the factors that allow the surface diffusion and two-dimensional crystallization of DNA origami rectangles. We also examine why the crystals would take the checkerboard form, rather than a linear chain form. We explore the effects of the strength of the stacking interaction between corners and of the concentration of  $\text{Mg}^{2+}$  during the surface diffusion step. Finally, we present a limitation of our approach, which is that the crystal size is limited by the irregularity of mica surfaces.

## III.2. Results and Discussion

### III.2.1. Two-dimensional assembly by surface diffusion

We first explored the case of  $N=3^{\ddagger}$ . Figure III-2 shows the self-assembly of rectangular origami into checkerboard crystals by surface diffusion. Figure III-2a compares large representative regions of surface-adsorbed origami before (left) and after (right) our protocol. Note that the three-blunt-end stacking patches at each of the four corners do not allow reliable connections between origami while in solution: these stacking patches are based on the stressed edges, which we believe to have weaker binding energy than the energy we estimated for “relaxed edges”. By the surface diffusion protocol, origami rectangles move around and form two-dimensional crystals via stacking through the corners: confining origami within the narrow space near the surface is expected to increase the chance of collision between the stacking patches in the right configuration, and the interaction with the surface should contribute to the stability, too (the entropy of confinement has already been paid).

Figure III-2b shows two consecutive AFM images during the step (2) of the assembly process that capture some attachment events of rectangles to crystals (indicated by orange circles), while some detachment events are also witnessed (indicated by green triangles). The images reveal the movement of some individual origami rectangles as well: there are single rectangles (or nucleating crystals of size  $\sim 2$ ) whose locations are not correlated between the two images—those that either moved away from the original location after the first scan or were “caught” in the second scan at a location where there were no origami previously (indicated by blue squares). The images also reflect the increased mobility of some individual origami rectangles: several linear artifacts appear by some origami getting dragged by the AFM tip (indicated by yellow arrows), suggesting that origami are mobile while still loosely bound to the surface and do not completely leave the surface (more evidence provided later).

---

<sup>‡</sup>The origami rectangle used for this study did not have the considerations discussed in Chapter II applied, such as the uniform edge sequences, twist correction, and relaxed edge geometry, for a simple reason—this system was explored before those properties were discovered and understood. Some of those factors—edge sequences and twist—were considered in a later stage of this study and are discussed in the text.



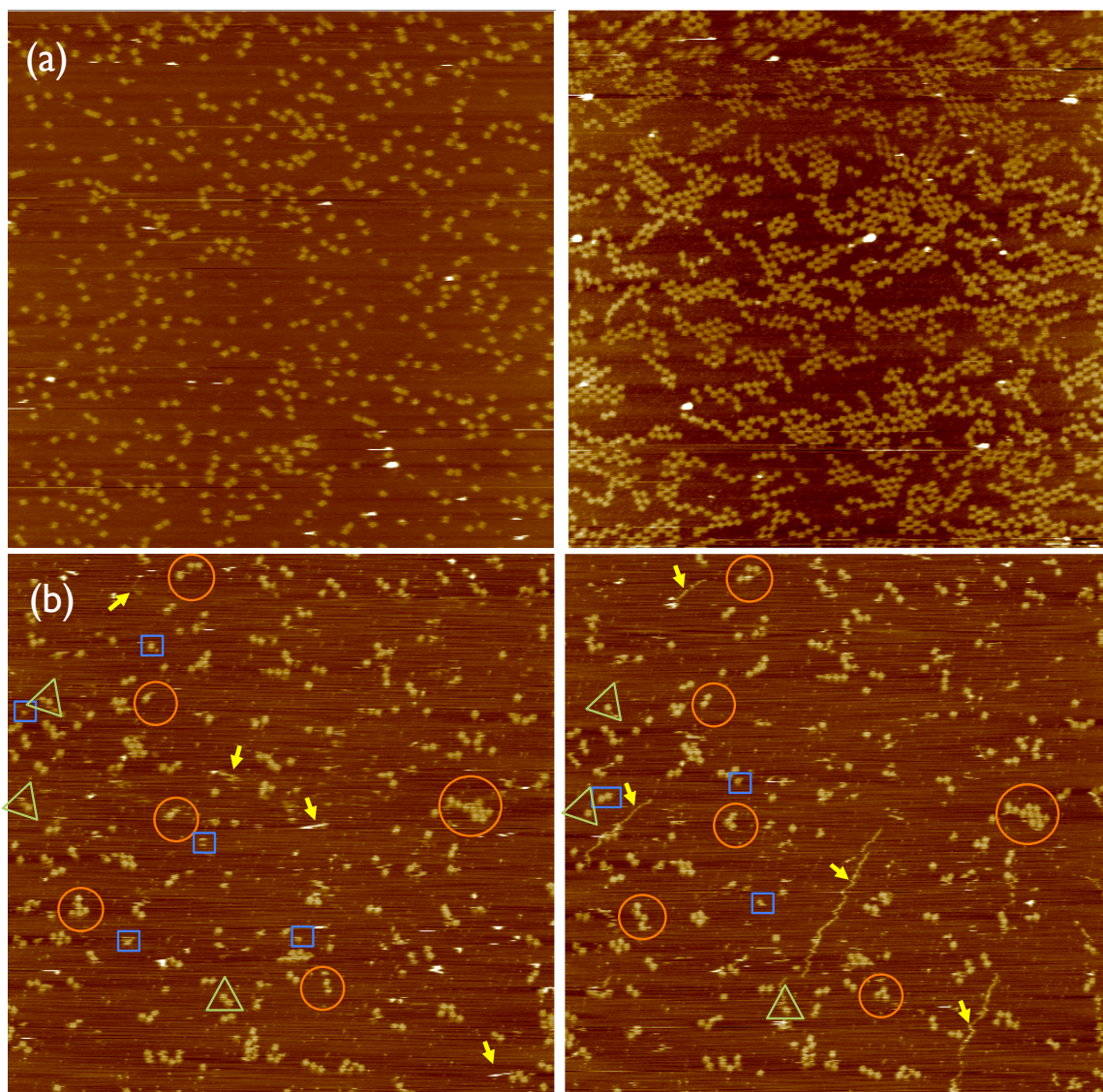


Figure III-2. **DNA origami checkerboard crystals self-assembled by the surface diffusion protocol.** (a) Representative AFM images of before (left) and after (right) the surface diffusion protocol. (b) Consecutive AFM images that show attachment events of origami to crystals (orange circles), detachment events (green triangles), and movements of individual origami or small crystals (appearance or disappearance between frames; blue rectangles, and structures dragged by the AFM tip; yellow arrows). All images are  $10\text{ }\mu\text{m} \times 10\text{ }\mu\text{m}$ .

### III.2.2. Factors that allow two-dimensional assembly

In order to understand the factors that enable the surface diffusion assembly, we carried out a series of systematic control experiments by removing one factor at a time (Figure III-3). When we

removed the “surface” factor, i.e., when we incubated a sample in the buffer condition of step (2) with monovalent cations at the same temperature for the same amount of time, but in a test tube, we could not see any crystal formation (Figure III-3a). As discussed earlier, it seems that the bond through the three-helix corner is not strong enough in solution with the equilibrium biased towards the unbound state, but the surface seems to pay some of the entropic cost by confining the structures into two-dimensional space. When we incubated the DNA solution deposited on a mica sheet, but in  $Mg^{2+}$  solution instead of  $Na^+$ , again we did not see any crystal formation (Figure III-3b). However, when we removed the factor “heat”, i.e., when we incubated the sample deposited on a mica sheet under NaCl solution for the same amount of time, but at room temperature, without heating it to  $40^{\circ}C$ , we still saw the formation of checkerboard pattern crystals, even though the size of crystals was somewhat smaller (Figure III-3c). Therefore, we concluded that the surface and  $Na^+$  ions are the most important factors, while it is suggested that heat might be just facilitating the process kinetically.

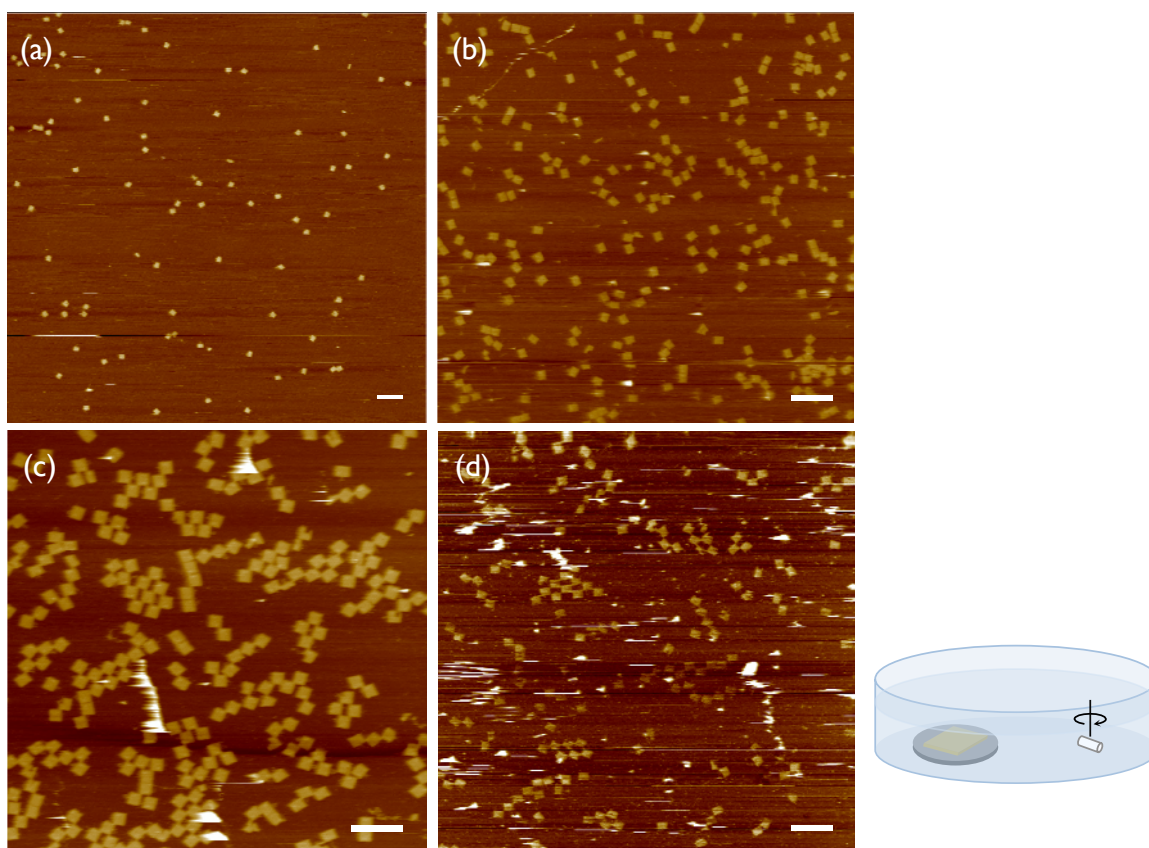


Figure III-3. **Control experiments that revealed the essential factors: the surface and  $Na^+$  ions.** (a) “No surface” condition. (b) Under  $Mg^{2+}$  only (“no  $Na^+$ ”). (c) “No heat” condition. (d) “Bath” experiment, where the entire mica substrate was immersed in a large reservoir of  $Na^+$  solution during the assembly step. Scale bars: 500 nm.

Next, a question arises whether the origami rectangles are really moving within the two-dimensional space, or whether they might actually leave the surface, then somehow assemble with each other, and then fall back down. It is somewhat suggested from the control experiment where we incubated the sample in a test tube, i.e., removed the “surface” factor, that the origami should not leave the surface to form the crystals, but in order to further prove that the assembly actually occurs in the two-dimensional space, we carried out an experiment where an entire mica sheet onto which origami was deposited was immersed in a big “bath” of NaCl solution in a petri dish (6 ml, ~150 times the usual amount used for the surface assembly process; see model in Figure III-3d). We also agitated the solution by a magnetic stirring bar throughout the incubation period to ensure the solution was well circulated. Under such a condition, we expected that if origami rectangles do leave the surface during the surface assembly process, we would lose most of the origami from the mica sheet into the solution and have significantly less number of them left on the surface. The result shown in Figure III-3d suggests that the origami rectangles do not leave the surface much and yet move around to form checkerboard crystals. This result is consistent with the prediction that  $Mg^{2+}$  pre-adsorbed to the surface (during step (1)) would effectively maintain the inverted charge on mica and prevent DNA from leaving the surface.

### III.2.3. Why checkerboards, not linear chains?

Why do the origami rectangles form checkerboard crystals, rather than linear chains? Since each origami rectangle has stacking patches with three blunt ends at each of the four corners, they could form linear chains with two corners on the same side participating in a bond with a single origami partner. Such a bond should be thermodynamically more favorable because it involves twice as many blunt ends per single bond between two origami rectangles as a single diagonal connection in checkerboard patterns does.

We had initially thought that it might be attributable to the non-uniform sequences of the bases on the blunt ends. The binding energy of the stacking interaction has strong dependence on the base sequences of the blunt ends<sup>44,62</sup>, so different corners of a rectangle may not have the same free energy of binding with each other. Irregular bond strength on different corners may result in irregular preference in bonding orientation (e.g., bias towards bonding between pairs of corners with sequences that would give strongest base stacking interaction). When we looked into the sequences at the blunt ends at each corner, there was indeed one corner that was rich in ‘GC’, which would give the highest stabilization when bound to itself in a ‘rotated’ orientation, yielding a bond

in diagonal direction. The preference for those diagonal bonds perhaps guides the remaining corners to form bonds along the diagonal direction rather than in a linear configuration along the helical direction.

To eliminate such a potential factor, we designed and tested an origami rectangle with a uniform sequence ('GC') along the edge, as described previously in Chapter II. It turns out that the rectangles still predominantly form checkerboard crystals even with uniform sequences (data not shown; but they show the typical pattern—similar to Figure III-2a, right), which led us to move our focus to the structural point of view.

The original origami design with average 10.67 base pairs per turn was later found to have global twist in the actual structure<sup>17,68,89</sup>, as also discussed in Chapter II. Even after confinement onto a surface, such global twist may still influence and distort the structures of individual origami rectangles, especially under the regime of non-adsorbing, long-range attraction during step (2). Perhaps, a pair of corners in one diagonal direction in a single origami rectangle is at a more elevated height from the surface than the pair of corners in the other diagonal direction, due to the global twist, preventing two edges from meeting in a linear configuration along the helical direction (see Figure III-4a).

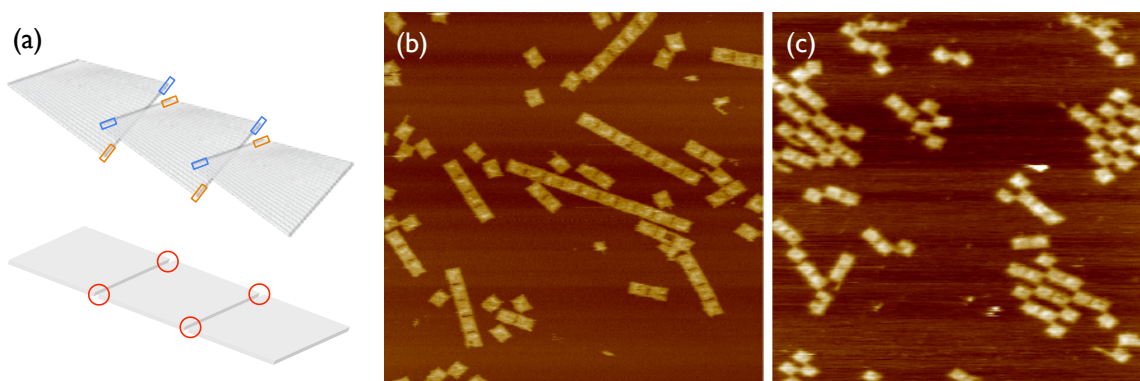


Figure III-4. **Twist of origami affects the assembly behavior.** (a) Models that illustrate how the twist might inhibit linear chain formation on surface (top; a pair of corners in one diagonal direction may be in a more elevated height – indicated by blue boxes – than the other corners – indicated by orange boxes, preventing bonds in linear configuration) while twist-free origami would not have such an issue (bottom). The twisted structure is compatible with checkerboard formation (see Figure III-6a). (b) Twist-reduced origami form linear chains in solution. (b) After surface diffusion, some remain as linear chains, while some form checkerboards and others form *double* checkerboards (see text). All AFM images are 3  $\mu\text{m}$   $\times$  3  $\mu\text{m}$ .

To test this hypothesis, we used an origami rectangle with reduced global twist, described in Chapter II. Interestingly, twist-corrected origami already formed long linear chains *in solution* (Figure III-4b), which was never achieved from the uncorrected origami. The twist may have

induced severe distortion along the edges, such as curvature, preventing the edges of the uncorrected origami from meeting in the linear configuration even in solution. After the surface diffusion protocol, some of the rectangles still formed checkerboard patterns, and some of them remained as linear chains. Also, an interesting feature that appeared was a new kind of crystal arrangement—*double* checkerboards—whose unit structure now associates two rectangles rather than one (Figure III-4c, upper left and lower right). The observation that checkerboard and double checkerboard patterns appear (and dominate) upon the surface diffusion suggests the following two things. One is that the long linear chains that formed in solution seem to partially break and rearrange during the diffusion stage. The other is that, under surface confinement, bonds through one corner may be kinetically more favorable than bonds through two corners on the same side; subsequent formation of stacking bonds in remaining corners and propagation of crystallization seem to quickly stabilize the rectangles in the crystal because each side has achieved an energy state equivalent to a side forming a six-helix bond in a linear chain (at least in terms of the number of blunt ends associated). Also, the appearance of the double checkerboard patterns reflects the influence of the global twist in origami on the bonding pattern on surface, i.e., with reduced twist, the dominance of checkerboard has decreased, allowing double checkerboards to form. It again supports our hypothesis that the preference of checkerboard formation to linear chain formation is, in part, due to the global twist in each origami rectangle.

### III.2.4. Effects of the strength of stacking interaction

We varied the number of stacking blunt ends at each corner to see the effect of the strength of the stacking interaction. We tested  $N \in \{3, 5, 7\}$ , where  $N$  is the number of helices with blunt ends at each corner of a rectangle.  $N=3$  is the case that has been discussed. Even with global twist, rectangles with larger numbers of stacking patches ( $N=5$  or  $7$ ) may have higher tendency to overcome the penalty from the twist and form linear chains. Results of the origami in solution were consistent with the prediction; right after the formation of origami in solution, before surface diffusion, rectangles with  $N=5$  or  $7$  tended to form linear chains, with an increasing tendency with  $N$  (Figure III-5a,b).

After surface diffusion, although some origami remained in the linear chain form, a large portion of structures observed were checkerboards and diagonal chains (Figure III-5c,d), and a few interesting observations could be made. First, some of the linear chains that formed in solution, again, seem to break during the diffusion stage, especially in the case of  $N=5$ . Second, even with the

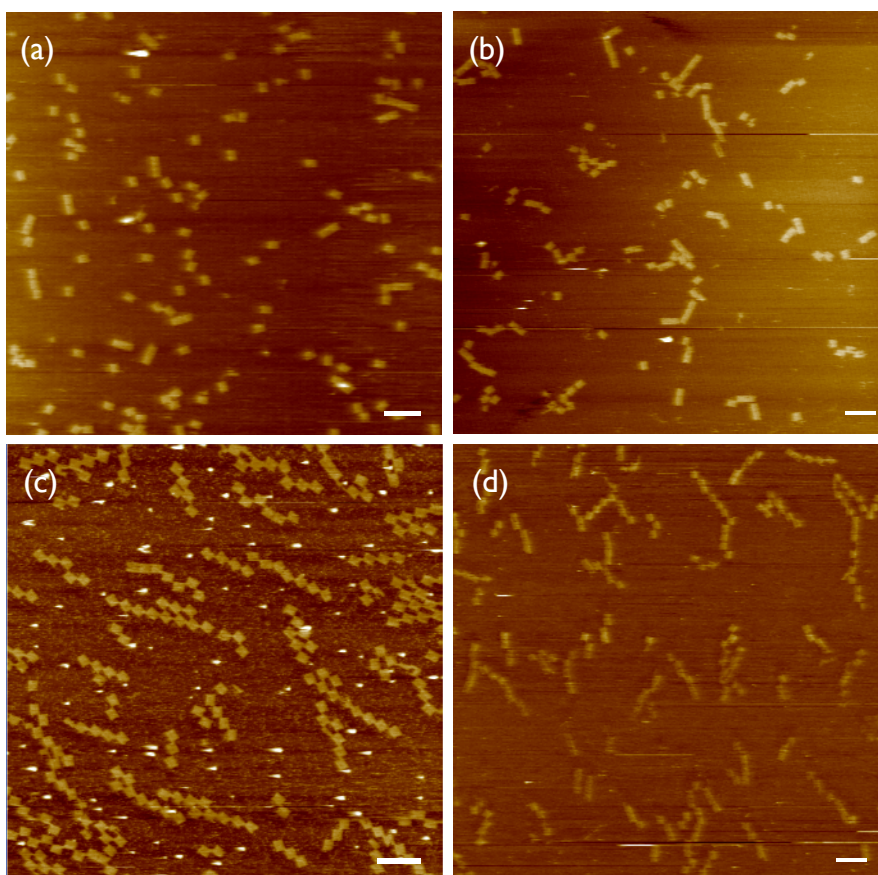


Figure III-5. **Different behaviors of origami with stacking patches with different strengths.** (a)  $N=5$ , before surface diffusion: short linear chains form. (b)  $N=7$ , before surface diffusion: linear chains form which are longer than the case of  $N=5$ . (c)  $N=5$ , after surface diffusion: some checkerboard crystals form while diagonal chains dominate. (d)  $N=7$ , after surface diffusion: linear chains and diagonal chains appear with a similar frequency. Scale bars: 500 nm.

higher strength of stacking interaction, the global twist of each origami seems to still prevent them from forming chains in linear configuration while confined on the surface (these origami are the original *twisted* origami). Also, propagation of linear chains, i.e., attachment of a rectangle to an existing linear chain in linear configuration, would become even more difficult because the global twist would accumulate with the number of origami in the linear chain form. Third, a more interesting behavior, especially found in the case of  $N=5$ , is that the most predominantly observed pattern was diagonal chains (Figure III-5c). This appears to be again due to the global twist in the origami structure. When one edge forms bonds with two other origami in each diagonal direction (see Figure III-6b), the angles of the other edges in the two origami may get ‘locked’ into the angles that the bonded edges define by the stronger stacking interaction with the wider overlap areas than in the case of  $N=3$ . This makes the non-bonded corners unavailable for a bond in checkerboard

configuration, while still allowing diagonal bonds that involve only one corner to form. In comparison, in the case of  $N=3$ , the corners with narrower overlap areas would have allowed bonds in the checkerboard configuration with minimal deformations on the edge geometry (Figure III-6a).

In the case of  $N=7$ , both diagonal chains and linear chains are found with a similar chance (Figure III-5d). Presumably, with the strength of the bond with  $N=7$ , the linear chains that already formed in solution may have less tendency to break on the surface, hence preserving the linear form. Diagonal chains may be those that newly formed during the surface diffusion process—and similarly to the  $N=5$  case, they did not proceed much to checkerboard patterns.

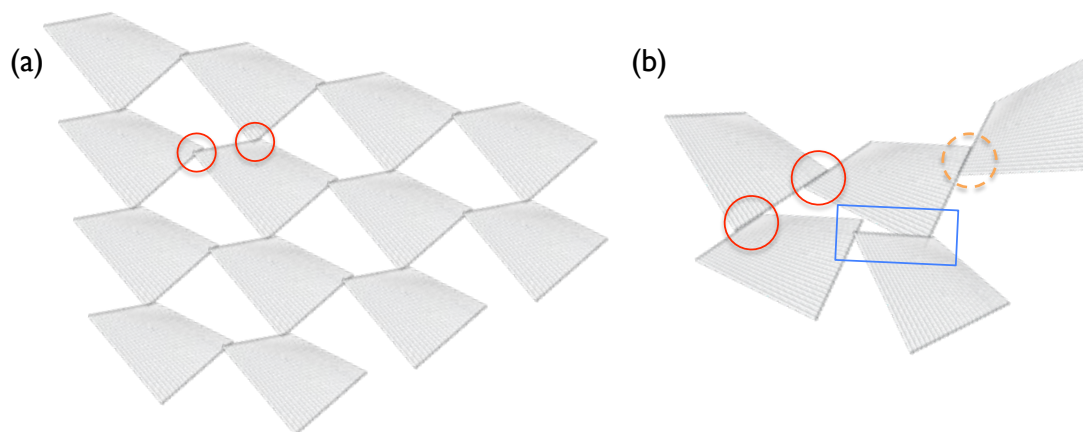


Figure III-6. **Models that illustrate the hypothetical effect of the size of stacking patches.** (a) With small overlap area for stacking bonds (e.g., 3 blunt ends), checkerboard bonds can form with minimal deformation on the edge geometry (two of which are indicated by circles). (b) With larger overlap area for stacking bonds (e.g., 7 blunt ends shown), bonds (red full circles) guide the angles of next layer stacking patches, making checkerboard bonds hard to form (blue rectangular box), while still allowing diagonal bonds to form (orange dotted circle).

### III.2.5. Effects of the concentration of divalent cation

As discussed earlier, during step (2) of the protocol, we still have Mg ions in the system, especially pre-adsorbed on the mica surface. The deposition of DNA origami in the  $Mg^{2+}$  condition during step (1) also acts effectively as “pre-treatment” of mica with  $Mg^{2+}$ . We asked the question of what would happen if we completely eliminate  $Mg^{2+}$  from the system. Might origami diffuse better and form larger crystals? Our prediction, based on our understanding so far, is that the lack of long-range attraction would poorly confine origami near the surface, impeding the assembly.

In order to answer this question, we synthesized the origami rectangle in a 0.75 M NaCl

solution, directly deposited the sample onto mica and followed the regular incubation procedure. When examined under AFM, origami turned out to form successfully in the  $\text{Mg}^{2+}$ -free condition (as also separately reported recently<sup>90</sup>), and more interestingly, no large crystals were found. Those that occasionally formed checkerboard patterns remained very small (Figure III-7a).

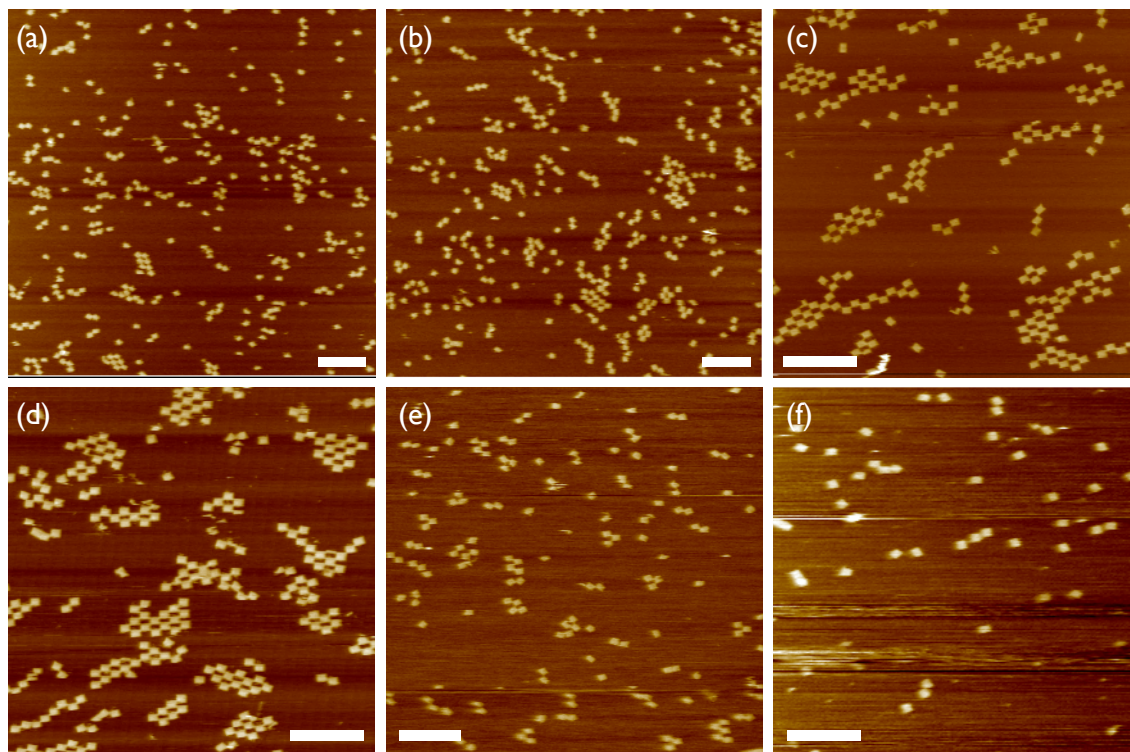


Figure III-7. **Effect of the concentration of  $\text{Mg}^{2+}$  in the  $\text{Na}^+$  solution during the surface diffusion step.** (a) No  $\text{Mg}^{2+}$  (only  $\text{Na}^+$ ) condition. (b) 0.28 mM  $\text{Mg}^{2+}$  in the  $\text{Na}^+$  solution. (c) 1.4 mM  $\text{Mg}^{2+}$  in the  $\text{Na}^+$  solution. (d) 2.8 mM  $\text{Mg}^{2+}$  in the  $\text{Na}^+$  solution. (e) 4.2 mM  $\text{Mg}^{2+}$  in the  $\text{Na}^+$  solution. (f) 5.6 mM  $\text{Mg}^{2+}$  in the  $\text{Na}^+$  solution. Scale bars, 1  $\mu\text{m}$ .

We also varied the concentration of  $\text{Mg}^{2+}$  in the DNA origami solution before deposition to see its effect on the assembly, by mixing origami solutions formed in the  $\text{Mg}^{2+}$  condition and in the  $\text{Na}^+$  condition with appropriate relative amounts (strictly, this is not the same as the standard protocol, where  $\text{Mg}^{2+}$  ‘pre-treats’ the surface in the absence of  $\text{Na}^+$ , but the higher valence and the smaller ionic size<sup>86</sup> of  $\text{Mg}^{2+}$  are expected to allow preferential adsorption onto the negatively charged sites on the surface). Under the standard protocol, the final concentration of  $\text{Mg}^{2+}$  during step (2) is  $\sim 1.4$  mM. As we increased the concentration of  $\text{Mg}^{2+}$  starting from  $0.2\times$  (0.28 mM) to  $1\times$  (1.4 mM), and  $2\times$  (2.8 mM), we found that the size of crystals after the protocol become larger (Figure III-7b-d) up to the point of  $2\times$   $\text{Mg}^{2+}$  concentration. The observation is that in order to confine the origami near the surface for the two-dimensional assembly, the system needs some



amount of  $\text{Mg}^{2+}$ , consistent with our prediction, and the trend is that with more  $\text{Mg}^{2+}$  content the crystal size gets bigger. However, when we increased the concentration of  $\text{Mg}^{2+}$  further, we could observe a decrease in the crystal size at  $3\times$  (4.2 mM) of  $\text{Mg}^{2+}$ , and almost no formation of crystals at  $4\times$  (5.6 mM) of  $\text{Mg}^{2+}$ , which suggests loss of mobility of origami rectangles under higher  $\text{Mg}^{2+}$  concentrations (Figure III-7e,f).

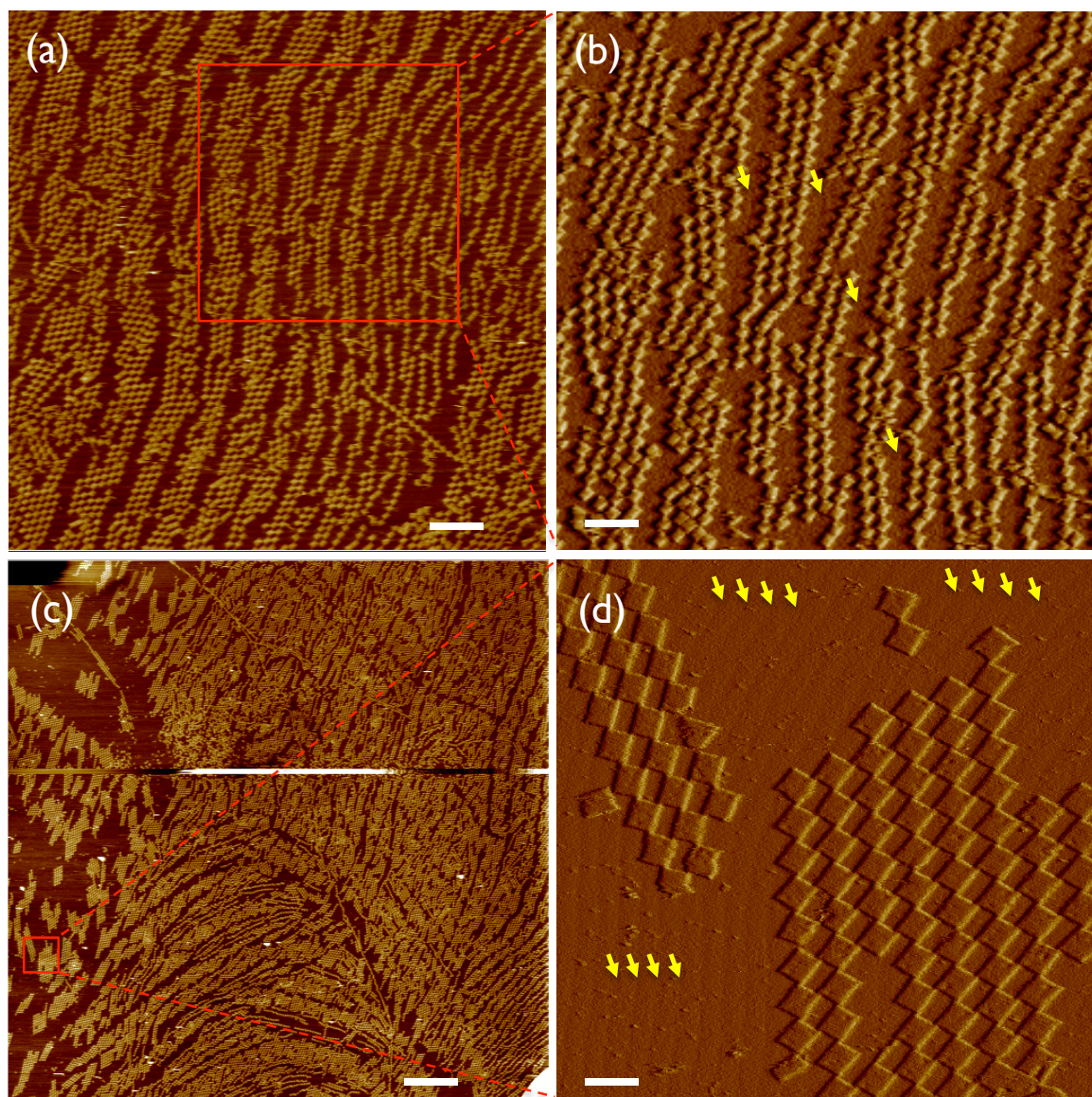
### III.2.6. Crystal size is limited by irregularity of mica

While it may seem possible to optimize the concentration of cations to achieve very large crystals, we encountered a problem that potentially prevents it and is putatively inherent to the protocol: the variation and irregularity of mica surfaces. We hypothesize that a structurally flat mica surface is not atomically uniform at least in terms of the charge distribution, and that such irregularity of a mica surface affects the distribution of origami checkerboard crystals (Figure III-8), although we have not carefully excluded other possible hypotheses such as flow of the solution during the surface diffusion step possibly generating long-range ordering of crystals. We observe that there are some areas where mica seems to *avoid* DNA origami and hence no crystals can stay. We hypothesize that those areas are where the boundaries lie (shown as line patterns on the mica surface and indicated by yellow arrows in Figure III-8b,d) between domains within which the charge distribution is uniform and compatible with origami. It is an interesting observation that with varying periods of the putative charge distribution patterns on mica, one can clearly see that different sizes of crystals are made. When the period of the patterns on mica is much larger than the size of individual origami, origami are able to form large checkerboards within each domain (Figure III-8a,b). With smaller periods of the patterns on mica, origami tend to stay as narrow crystals. Interestingly, when the period roughly matches the size of origami, origami formed packed, relatively large crystals (Figure III-8d).

Even if the observed patterns of origami crystals are due to irregularity of mica as we hypothesize, we do not understand the cause of such irregularity. However, we note that mica is a mineral with varied chemical composition<sup>91</sup>, e.g., Si and Al share a common atomic site and another site mostly occupied by Al can also host Mg or Fe<sup>92</sup>. In addition, mica crystallites were found to exhibit different growth rates depending on the direction of crystal growth and generate fibrous growth patterns, yielding long linear domains of crystallites<sup>93</sup>.

We believe that such irregularity of mica influences our system in general and limits the size of crystals, although the boundaries could not always be clearly observed. It is consistent with the

results of longer incubation experiments where the crystal size did not get much larger (data not shown).



**Figure III-8. Formation of origami checkerboard crystals with various patterns, putatively depending on the variation of mica surface properties.** (a, b) Origami checkerboard crystals that follow the pattern of a mica surface. The zoom-in amplitude image (b) of the boxed area in (a) reveals some of the putative domain boundaries on the mica surface (indicated by arrows). (c) A large-scan (30  $\mu\text{m}$ ) AFM image that shows regions with distinct patterns of mica. (d) A zoom-in amplitude image of the boxed area in (c), which clearly shows narrow (close to origami size), long domain patterns on mica (boundaries shown as line patterns on mica and indicated by arrows). Scale bars: (a) 1  $\mu\text{m}$ , (b) 500 nm, (c) 3  $\mu\text{m}$ , (d) 200 nm.

Since the surface property of mica is not something that can be easily controlled by, for

example, lithographic techniques, it would be interesting to try a similar protocol on chemically functionalized Si wafers<sup>79,80</sup>, which are technologically more useful as well. Given the observation from Figure III-8d that parallel patterns of putative charge distribution whose period roughly matches the size of individual origami allow large crystal formation, it seems that just simple parallel patterning of Si substrates with alternating chemical patches may be effective, if the period is optimally designed.

### III.3. Conclusion and Future Directions

In summary, we described a novel protocol for the assembly of large-scale DNA nanostructures based on surface diffusion. The diffusion of DNA nanostructures was dynamically controlled by changing the cationic condition of the system, and thus changing the mobility of DNA nanostructures on mica substrates. We used monovalent cations to weaken the attraction between origami-based nanostructures and the mica surface and relied on residual divalent cations to confine the DNA nanostructures close to the surface and promote assembly. Although accelerated by elevated temperatures, this protocol is operable isothermally as well. We show that global twist of our origami building blocks favored the formation of two-dimensional checkerboard pattern crystals, over, for example, one-dimensional linear chains. However, the crystal size turned out to be limited by the irregularity of mica surface properties. It would be interesting and technologically useful to develop a similar protocol on chemically functionalized Si wafers. Optimizing the chemical functionality such as hydrophobicity, charge distribution, etc., and the periodicity, might allow wafer-scale self-assembly of quasi-uniform crystals of DNA nanostructures.

Although we exclusively employed stacking patches to allow potential rearrangements during the assembly process and to take advantage of the symmetry, investigating the possibility of using sticky-end based connectors might be interesting. The symmetry constraint would be achieved by the use of simple palindromic sequences, e.g., ‘AGCT’ or even ‘AATT’. The use of sticky ends would allow more control over the strength of the bond, without the restriction from the global twist—with the stacking patches, stronger bonds required larger contact area, which made the system fall into problems caused by the twist, preventing the growth of large crystals.

## III.4. Materials and Methods

### III.4.1. Surface diffusion assembly

Surface diffusion of DNA origami rectangles proceeded with the following three steps. Step (1): we first deposit rectangles that were formed in  $1\times$  TAE/Mg<sup>2+</sup> (40 mM Tris acetate, 1 mM EDTA, 12.5 mM magnesium acetate) solution by the standard protocol for origami formation<sup>1</sup> onto a freshly cleaved mica substrate. Typically we deposit 20 ul of the buffer solution first, to which we add 5 ul of the origami solution. Step (2): we then exchange the buffer on the mica substrate by pipetting out  $\sim 20$  ul of the TAE/Mg<sup>2+</sup> buffer on top (we expect the majority of origami are strongly bound to the surface), leaving  $\sim 5$ ul of the buffer spread on the mica surface. We then add 40 ul of 0.75 M NaCl solution. We put the substrate in a humidity chamber (usually a small petri dish with a piece of water-soaked paper tissue) inside a temperature-controlled box (Coy Laboratory Products Inc.) and incubate it at a constant temperature of 40 °C for  $\sim 4$  hr. Step (3): after the incubation, we add 5 ul of 10 mM nickel acetate solution onto the substrate while still keeping the substrate within the temperature-controlled box at 40 °C, to “freeze” the product structures onto the surface at that state. Then we exchange the majority buffer from NaCl back to TAE/Mg<sup>2+</sup> and take out the substrate for analysis.

### III.4.2. Atomic force microscopy (AFM)

AFM images were taken in Tapping Mode with a Nanoscope III Multimode AFM (Veeco Metrology Group, now Bruker AXS). Typically, we used silicon nitride cantilevers with 2 nm radius silicon tips as AFM probes (the “short, fat” A cantilever on SNL probes from Veeco, now Bruker AFM Probes).

## *Chapter IV*

# ParMRC and Expandable DNA Nanostructures

*"Life is definitely always about expanding. That's how we grow."  
- Jessica Simpson*

### **IV.1. Introduction**

How does a cell make sure that each copy of replicated DNA gets into each daughter cell when it divides? Here a mother cell's main task is to move each copy of its DNA towards one of its two opposite poles, so that when it divides in the middle its DNA is partitioned equally between the daughter cells. Eukaryotic cells, like our own, have developed very complicated machinery based on microtubules for such a task. Before a cell divides, microtubules grow from each pole, reach out to pairs of just-replicated chromosomes that have aligned along the cell's midplane, then precisely grab one copy of each type of chromosome and pull them towards the poles—thus achieving DNA segregation. Though simple to describe, the whole process is maintained by several different multi-subunit protein machines and is under the exquisite control of signaling networks<sup>94</sup>. In prokaryotes, like bacteria, a variety of different mechanisms are known to exist, depending on the type of DNA<sup>95</sup>. Some types of plasmid DNA, in particular ones which maintain a high copy number, rely on random distribution of the plasmids within the cell; for them, free diffusion throughout the cell suffices for DNA segregation between daughter cells. On the contrary, low copy number plasmids cannot depend on diffusion, and have developed partitioning machinery that actively pushes (or pulls) the replicated copies to opposite poles of the cell. Those active partitioning systems are usually very simple, because they are encoded and contained within the plasmid's own DNA. One of the most well-understood such partitioning systems is called ParMRC<sup>96,97</sup>.

We, as DNA nanotechnologists, look at the bacterial partitioning system, ParMRC, from two different perspectives. One is to use DNA origami as a custom tool to study some biophysical aspects of the ParMRC machinery. Though it is relatively well-understood, there are still many open questions as to the details of structural organization of the components and mechanisms of operation at the molecular level. We try to tackle some of these problems using DNA origami as a single-molecule custom tool. With a length scale relevant to the protein components, and programmability of the shape and locations of binding functionality, DNA origami is poised to be a useful tool for single-molecule biophysics studies.

Our second perspective is to exploit the dynamic plasmid segregation system to engineer *expandable* nanostructures that grow from the nanometer scale to the micrometer scale. While DNA origami allows the creation of various shapes at the nanometer scale, it is challenging to programmably self-assemble objects of much larger length scales, e.g., at the length scale similar to eukaryotic cells (~10-100  $\mu\text{m}$ ), which would be potentially interesting for engineering artificial cells. We hope to be able to program the construction of objects of such length scales, by first creating some nanoscale objects with a predesigned basic architecture, and then dynamically expand them to a much larger length scale. Operating with a relatively small number of components, the ParMRC system exhibits various dynamic behaviors, such as search and capture of the plasmids and steady extension until the plasmids reach the opposite poles. We try to couple the dynamic behaviors of the ParMRC system with DNA origami nanostructures, and to demonstrate dynamically expandable nanostructures.

### IV.1.1. Using DNA origami to study biophysical questions

#### IV.1.1.1. The ParMRC system

ParMRC is an active partitioning system for a low copy number plasmid R1 found in *E. coli*<sup>96,97</sup>. The system involves three components: *parC*, ParR and ParM (Figure IV-1). *parC* is a DNA sequence in the plasmid which acts as a centromere during partitioning. ParR is a DNA-binding protein that binds to *parC* and forms a ParR/*parC* complex which subsequently binds to filaments composed of ParM proteins. ParM polymerizes into an actin-like filament that exhibits dynamic instability: a filament repeatedly grows and shrinks until the ends get stabilized either by binding to a ParR/*parC* complex or by bundling with another filament<sup>98</sup>. Once stabilized, the

ParR/*parC* capped end of the ParM filament continues to grow. This moves the ParR/*parC* complexes, and hence each of the two plasmid copies, to opposite poles of the cell.

*parC* is a DNA sequence in the plasmid, and has a promoter region for the genes *parR* and *parM* at the center of its ~170 long sequence (Figure IV-1a). The binding regions for ParR on *parC* is composed of 10 repeating units of homologous sequence of DNA, each of ~11 bps (called “iterons”).

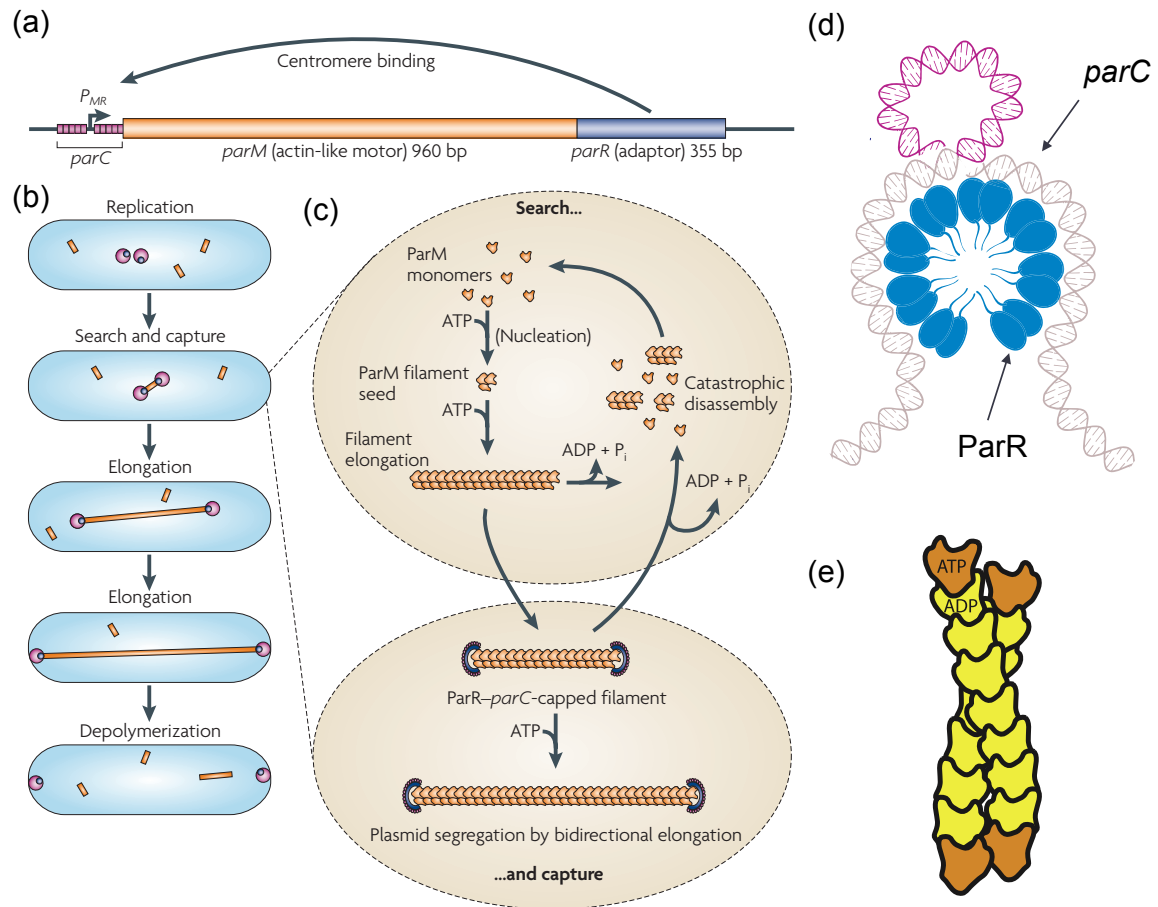


Figure IV-1. **The ParMRC system.** (a) The genes for ParM and ParR follow the gene *parC*, which contains the promoter region between two sets of five binding sites (iterons) for ParR. ParR protein binding to *parC* acts as self-suppression for the ParMRC genes. (b) Model showing the process of plasmid segregation by the polymerization of ParM. (c) The search process involves the dynamic instability of ParM filaments maintained by the balance of ATP-ParM and ADP-ParM. The capture process involves the stabilization of filament ends by ParR/*parC* complexes. Figures (a-c) adapted from ref. 96. (d) Schematic model of a *parR*/*parC* complex, which shows the promoter part looped out (magenta). Figure adapted from ref. 100. (e) Schematic model of a ParM filament, which is left-handed. ParM is an ATPase and converts ATP to ADP. Figure adapted from ref. 101.

ParR binds to the iterons in dimeric form. Each dimer binds to a single iteron, hence the total of 20 ParR monomers bind to the *parC* sequence and form the ParR/*parC* complex. The binding is found to be cooperative<sup>99</sup>. Binding of ParR to *parC* also serves the role of repressing the promoter in *parC*, so the expression of ParR is self-regulated by negative feedback.

ParM, in the presence of ATP, forms ATP-ParM and assembles into a polar filamentous form, similar to actin filaments, but in a left-handed double helical form with two 12-subunit chains (“protofilaments”) as a repeating unit, whose length<sup>98</sup> is  $\sim 566.9$  Å. ATP bound to ParM then undergoes hydrolysis by ATPase contained in ParM. ADP-ParM has significantly higher  $K_d$  ( $\sim 2.4$   $\mu$ M,  $\sim 60$ -fold higher) than ATP-ParM ( $K_d = \sim 42$  nM)<sup>97</sup>, and readily dissociates from the filament under physiological conditions. Repeats of the growing phase of ATP-ParM and the shrinking phase of ADP-ParM exhibit dynamic instability, which enables plasmid segregation by a “search and capture” mechanism. Once captured and capped by a ParR/*parC* complex, each stabilized end of a filament keeps growing by an insertional polymerization mechanism (i.e., monomers add to the ends, not into the middle), and pushes the plasmids to the opposite poles of a cell.

#### IV.1.1.2. Open questions

Although the ParMRC system is one of the simplest and most-studied plasmid partitioning mechanisms involving just three components, there are still many aspects of the system which are not well understood. For example, there is contradictory evidence, depending on the method of observation regarding the shape of a ParR/*parC* complex, and several structural models have been proposed, ranging from a simple ring<sup>102</sup> to a helix<sup>100</sup>, or a U-shape<sup>103</sup> (Figure IV-2). Also, the relative orientation and binding mode of a ParM filament to a ParR/*parC* complex is not known: whether a ParM filament attaches to a ParR/*parC* complex in a parallel or perpendicular orientation (Figure IV-3a), and whether the ParR/*parC* complex binds the filament on its end just as an “open clamp”, or wraps around the filament’s sidewalls<sup>96</sup> (Figure IV-3b). Another unknown aspect is whether a ParR/*parC* complex is polarized in its function (Figure IV-3c). The sequence directionality of the *parC* DNA might impose some structural or functional polarity in the ParR/*parC* complexes. Or, even if the ten repeating units in *parC* may hold ParR protein molecules in a symmetric fashion, the complex may be asymmetric in terms of the alignment of binding domains of ParR for the ParM filaments. Further, the asymmetry of the polar ParM filament itself may also impose potential polarity in the ParR/*parC* complex. Thus, it is uncertain whether both plus and minus ends of a filament bind to a ParR/*parC* complex, and if so, whether or not the binding happens in the same fashion. This question may be differently asked depending on the



model for the binding mode. The “wrap around” model is compatible with either polar or symmetric binding because the complex may wrap around filament ends in the same (parallel) direction with respect to the filament axis. If the “clamp” model is assumed, one could ask whether the complex clamps filament ends universally in the same structural manner, regardless of whether it is a plus or minus end. If the ParR/*parC* complex does not allow universal binding, the complex would only bind to one end of a filament, and the bipolar segregation mechanism would require association of two antiparallel filaments. A recent report<sup>98</sup> provides some suggestion that it may indeed be the case that the ParR/*parC* complex is stabilizing only one end and the ParMRC system requires bundling of two antiparallel filaments. But contradictory evidence that supports binding of ParR/*parC* to both ends of a filament<sup>104</sup> still exists and has not been ruled out.

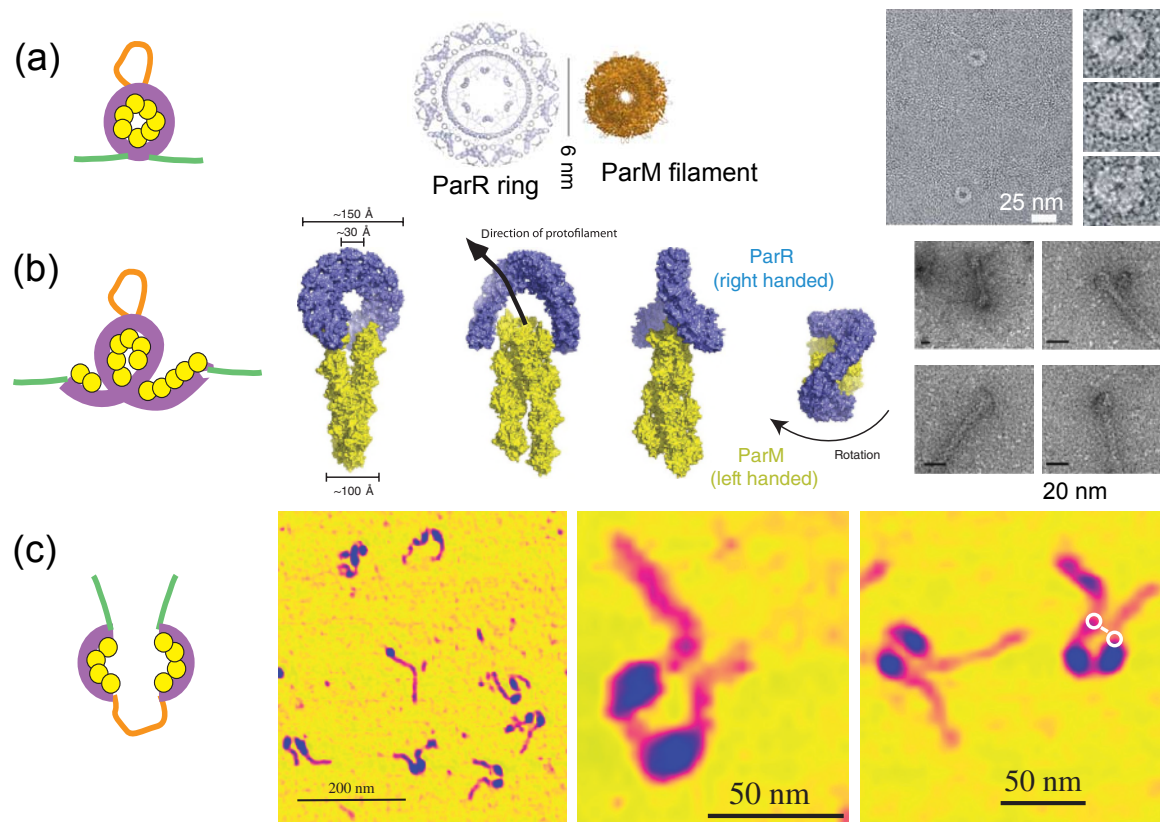


Figure IV-2. **Contradictory models and supporting evidence for the geometry of the ParR/*parC* complexes.** (a) “Ring” model and corresponding EM data. The outer diameter of a ParM filament matches the inner diameter of the ring created by a ParR/*parC* complex. Molecular model and EM data adapted from ref. 102. (b) “Helix” model and corresponding EM data. ParR/*parC* rotation mechanism is proposed based on the helical geometry. Molecular model and EM data from ref. 100. (c) “U-shape” model and corresponding AFM data. ParR-bound *parC* was observed as an open U-shaped complex under AFM. AFM data from ref. 103.

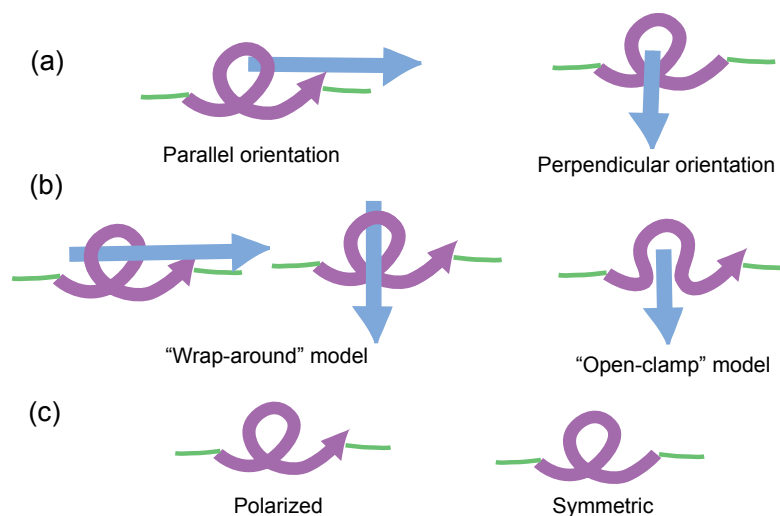


Figure IV-3. **Schematic models illustrating different hypotheses**, (a) for the orientation of a ParM filament relative to a ParR/*parC* complex, (b) for the binding mode of a ParM filament end to a ParR/*parC* complex, and (c) for the polarity of a ParR/*parC* complex.

#### IV.1.1.3. DNA origami design: proposed and actual

Our initial goal in this investigation of biophysics of the ParMRC system was to try to answer some questions in regards to the structure, orientation, and binding mode of a ParM filament with a ParR/*parC* complex, described in the previous section, by using DNA origami as a custom single-molecule template. We can create a DNA origami structure that has a *parC* sequence incorporated into its structure (Figure IV-4ab). By changing the subset of staple strands, we can easily vary parameters such as the position of the *parC* strand, the angle of the *parC* strand relative to the origami frame, and the distance between the ends of the *parC* sequence, therefore changing the tension of the *parC* strand. Then, we hoped we could study the interaction of the geometrically-constrained *parC* component with ParR and ParM. For example, if a ParR/*parC* complex can stay rigid enough on an origami template, we could measure the angle distribution of a ParM filament relative to the ParR/*parC* complex on the origami by taking AFM images, which may allow us to tell if ParM filaments associate with ParR/*parC* complexes at a preferred relative angle, perhaps either perpendicular or parallel (Figure IV-4c). Also, if we combine two origami structures together each containing a ParR/*parC* complex, controlling the bonds between the origami structures in such a way that we have multiple combinations of the orientation of the ParR/*parC* complexes (Figure IV-4d), we may be able to see different growth directions of a ParM filament, if there indeed exists polarity in the ParR/*parC* complexes. Further, depending on the orientation and alignment of the

ParR/*parC* complexes relative to the ParM filament growth direction, and on the force generated by the filament growth, we may be able to see an interesting behavior that the growing filament may push the two combined origami rectangles apart, which leads us to the idea of building ‘expandable’ nanostructures described in the next section.

For our investigation, to achieve a DNA origami structure that contains a single *parC* sequence in it, we first created an M13-variant scaffold strand by genetically engineering M13 (via Genestitute) to include the 169-nucleotide *parC* sequence in it (Figure IV-4a; we call this “*parC*-M13” below). For a DNA origami design, as a first trial, we used a pre-existing set of staples that had been originally designed for rectangular origami (twist-corrected, GC-ended). Since the *parC* sequence is just an addition to the basic scaffold sequence with little (~60 nucleotides) extra difference, the staple strands for the rectangle design should still fold the *parC*-M13 scaffold strand, leaving the *parC* addition as a single-stranded loop. Sequence comparison between the regular M13 and the *parC*-M13 revealed where the loop would fall (Figure IV-5a). We also added a complementary sequence to the *parC* sequence to make it double-stranded (as is in a plasmid).

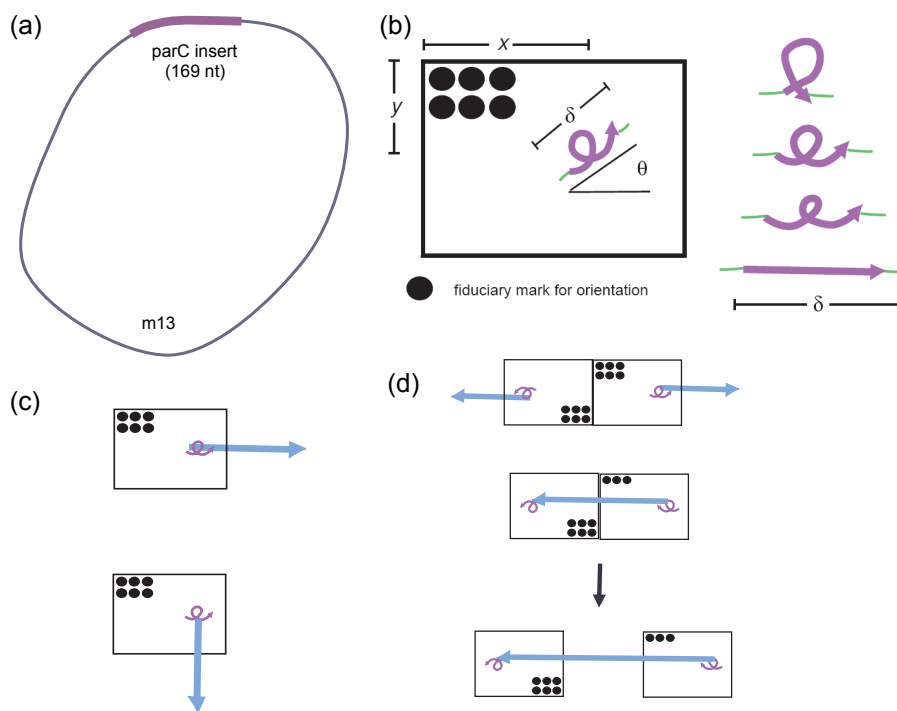


Figure IV-4. **Design of DNA origami initially proposed for this study.** (a) Custom-engineered M13 scaffold for origami with a *parC* sequence insert. (b) Origami rectangle design that can be tuned to control the distance and angle between the ends of *parC* strands. (c) Proposed idea of using DNA origami to study the orientation of a ParM filament relative to a ParR/*parC* complex. (d) Proposed idea of using DNA origami to study the polarity of a ParR/*parC* complex, which may lead to an expanding origami structure.

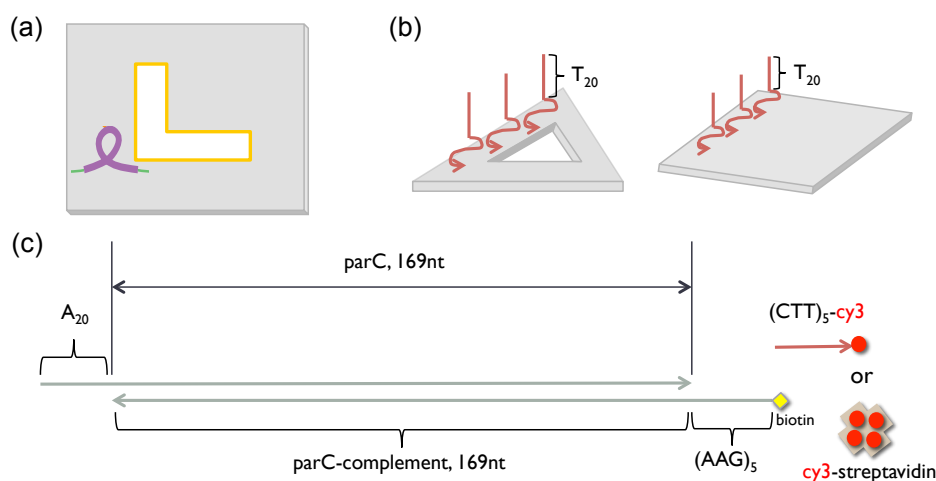


Figure IV-5. **Design of DNA origami used for this study.** (a) The design of rectangle origami, used for studies with single-*parC* origami, which shows the location of the *parC* sequence (purple) and the hairpin label for contrast in AFM images (“L” mark). (b) Schematic diagrams of the anchors on triangle and rectangle origami for *parC* strands, used for studies with multiple-*parC* origami. Each staple strand was extended with T<sub>20</sub> at the 5'-end. (c) The design of *parC* strands. The 5'-end of one of the *parC* strands (top) was extended by A<sub>20</sub>, which was bound to a 5'-end T<sub>20</sub> extension of staple strands on origami. The 5'-end of the bottom strand was extended by (AAG)<sub>5</sub> for fluorescence labeling.

Using this DNA origami structure containing a single *parC* sequence, we tried to first observe the binding of ParR proteins onto their binding sites of *parC* on origami, then the binding of ParM filaments to the ParR/*parC* complexes that would have formed on origami. We first chose to use AFM as our primary measurement tool. Besides being most familiar and available to us, AFM is an ideal instrument for obtaining the details of structures and potentially some interactions at the nanometer scale. We will present AFM results where we could see some interesting behaviors of ParR, depending on its concentration. However, we could also see some limitation of the approach, especially the difficulty in observing ParM, the filament-forming protein, under AFM.

For later investigations, we decided to use other instruments such as gel electrophoresis and fluorescence microscopy as well, which would be more compatible with DNA origami structures containing more than one *parC* strands. Thus we designed DNA origami that contains multiple *parC* strands in each origami structure. First, separate *parC* strands were designed (shown in Figure IV-5c). For incorporation of multiple *parC* strands onto origami, pre-existing DNA origami designs, such as a triangle and a rectangle, were modified such that each staple strand has a 5'-end extension of 20 T's (Figure IV-5b), to which *parC* sequences were designed to link (with an extension of 20 A's). The top strand of the double-stranded *parC* was extended at the 5' end with an

overhang of 20 A's that can bind to the 20-T anchors on origami. In addition to the 169-base sequence that is complementary to the corresponding sequence in the top strand, the bottom strand had (AAG)<sub>5</sub>-biotin extension at its 5'-end, to which we can incorporate either (CTT)<sub>5</sub>-Cy3 DNA strand or Cy3-streptavidin for fluorescent labeling.

Using the origami with multiple *parC* strands, we obtained AFM movies of progressive ParR binding onto *parC* sites on origami, and gel electrophoresis data that show concentration-dependent binding of ParR to *parC* strands both on origami and free in solution. We also present fluorescence microscope data in a later part of this chapter.

#### IV.1.2. Building expandable structures using growing biopolymers

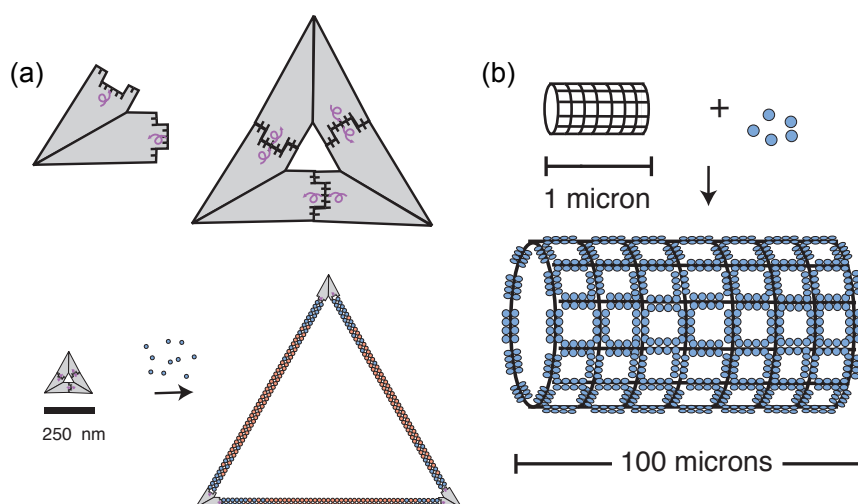


Figure IV-6. **Initially proposed expandable structures from nanoscale to microscale.** (a) Triangle structure created by connecting three *corner* origami nanostructures via stacking bonds, each of which origami containing *parC* sites near the bond edges. The triangle may expand by the growth of ParM filaments between each corner origami. (b) The same principle can be applied to a nanotube structure expandable to a microscale tube.

In order to build nanoscale machines, some mechanism of converting chemical energy to mechanical energy is essential. ParM polymerization is a naturally occurring mechanism where growing filaments based on chemical interactions exert force on plasmid DNA to push two copies of them to opposite poles of a cell. We planned to use such a mechanism to build *expandable structures* that grow from the nanometer scale to the micrometer scale. Our initial plan was to use DNA origami with a 60° corner shape which, with properly designed bonds, fit together and create an equilateral triangle, each edge near the bonds having a *parC* strand (Figure IV-6a), and to let the

triangle expand by the polymerization of ParM and the stabilization of the growing filaments by ParR/*parC* complexes at each edge near the bonds of origami. Also, we intended to extend the idea to tile-based nanotubes connected by weak bonds, each tile containing *parC* strands, and to create nanotubes expanding to micrometer scale, in the presence of the ParM and ParR proteins (Figure IV-6b).

Our initial investigations taught us several things that are against our proposed ideas. First, observing ParM filaments under AFM turned out to be difficult. Although it was expected to be hard to capture the dynamic behaviors of the filaments by AFM, AFM would have been a useful tool to obtain static snapshots at different stages of the system, e.g., at the beginning where the three-origami triangles are intact, and at a later stage where majority of them are expanded by filaments, etc. However, the nature of the ParM system that it needs a certain concentration of monomers around to ensure filament growth and dynamic instability, and the difficulty of finding an AFM-compatible condition where grown filaments can stick to a substrate surface while still enabling their dynamics and repelling monomers from the surface, prevented us from continuing to use AFM for investigations of ParM filaments. Second, with fluorescence microscopy, which is a less surface-dependent technique, it turned out to be difficult to observe a single filament because of various issues, with the primary one being the low signal-to-background ratio for a single filament. Third, total internal reflection fluorescence (TIRF) microscopy would be helpful to increase the signal-to-background ratio, but it involved other problems. Our initial surface treatment that later turned out to be inappropriate caused fluorescent monomers to bind to the surface and did not improve the signal-to-background ratio. Our current optimal surface treatment, which prevents direct binding of any protein species, requires higher crowding agent concentration than usual, to push filaments toward the surface within the TIRF regime (~200 nm), which turned out to cause unwanted, spontaneous formation and stabilization of ParM filament bundles without their ends being capped by ParR/*parC* complexes.

Thus, we used DNA origami that contains multiple *parC* strands and tried to achieve optimal imaging conditions under a fluorescence microscope. We successfully reconstituted growing ParM filaments with their ends stabilized by DNA origami containing ParR/*parC* complexes. Then, we demonstrated expandable nanostructures by using DNA origami rectangles connected by stacking bonds, and by breaking the stacking bonds and separating origami via the growth of ParM filaments.

## IV.2. Results and Discussion

### IV.2.1. Biophysical studies of the ParMRC system

In this section, we describe some AFM results obtained using DNA origami containing a single *parC* strand, where we could see some interesting behavior of ParR, the *parC*-binding protein, depending on its concentration. We could also see some limitation of the approach, especially the difficulty in observing ParM, the filament-forming protein, under AFM. In a later part, we show more AFM data and some gel data taken for DNA origami containing multiple *parC* strands.

#### IV.2.1.1. AFM studies of ParR binding to *parC* on origami

##### IV.2.1.1.1. ParR binds to non-recognition sequence DNA at high concentration

We tried to observe the binding of ParR proteins onto their binding sites of *parC* on origami, where each origami rectangle contained a single *parC* strand (based on the engineered *parC*-M13 scaffold mentioned above). We also added ParM and ATP in later steps. To clearly visualize the changes made by the binding of the proteins, we took AFM images of the same area after each step—from first depositing only the origami onto mica, to exchanging the buffer from TAE/Mg<sup>2+</sup> to a buffer that is more compatible with the ParR and ParM proteins (ParM polymerization buffer), to adding ParR proteins, and finally to adding ParM and ATP. We used chains of origami connected by stacking bonds instead of individually separated origami rectangles, initially to use the chains for fluorescence microscopy (after proper labeling) and to be able to easily identify them under a fluorescence microscope. Using chains also proved to be helpful in preventing origami chains from leaving the surface upon buffer exchange due to the lower divalent cation concentration (1 mM Mg<sup>2+</sup>) of the ParM polymerization buffer; the larger the surface contact, the stronger the structure is held by the surface.

Figure IV-7 shows the same part of an origami chain after each step. Two blue vertical dotted lines across the whole figure connect roughly the same part on each panel. Figure IV-7a shows the first step when only origami chains were deposited onto mica in TAE/Mg<sup>2+</sup> buffer. The model in Figure IV-7a shows the dumbbell hairpin pattern (“L”: shown as flipped to match the data, which indicates this particular chain is facing down, with the hairpin side pointing towards the mica surface), the rough position of the *parC* strand (purple loop) on the origami, and the double-stranded

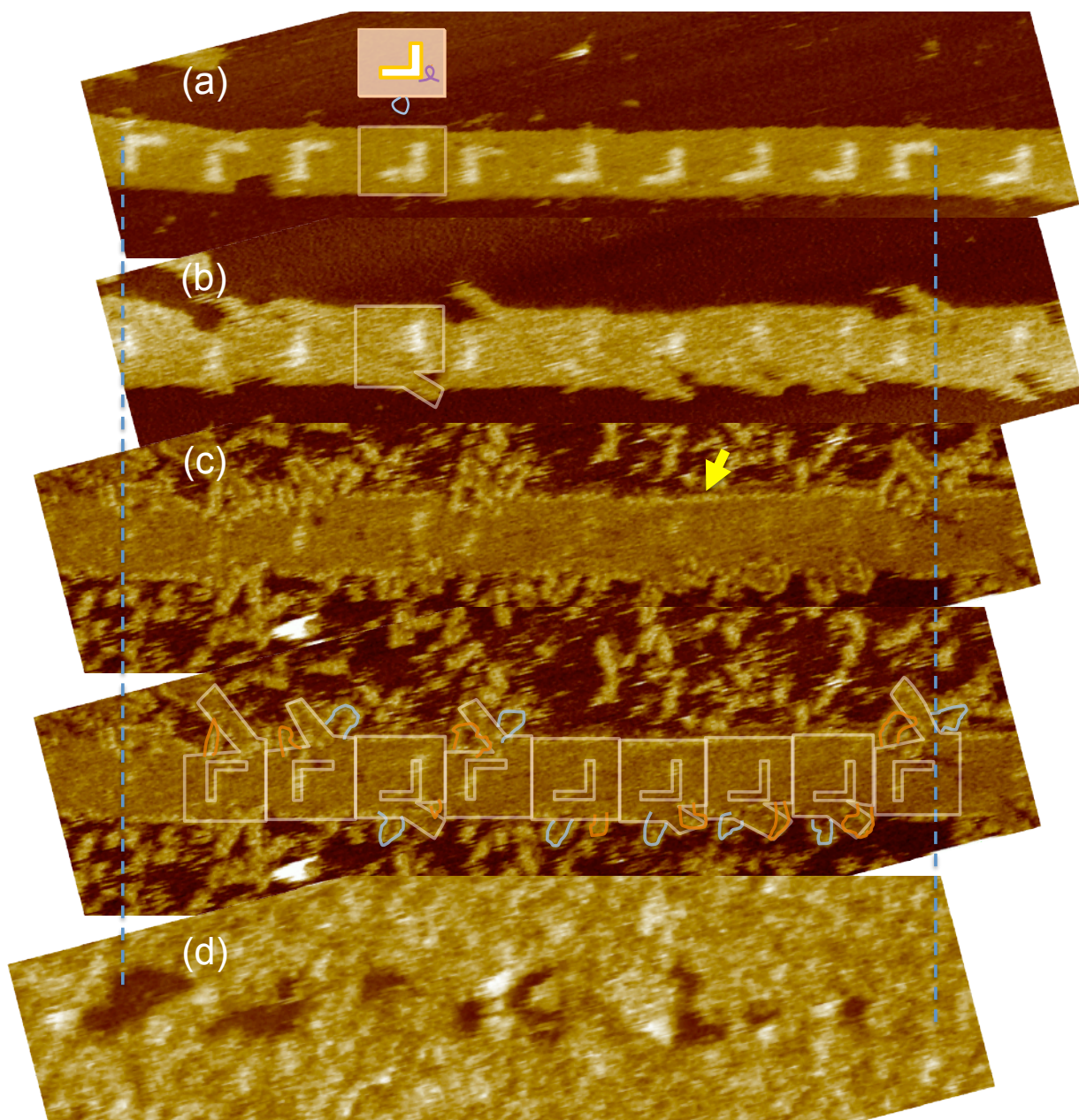


Figure IV-7. **Binding of ParR and ParM to DNA origami, each containing a single *parC*, tracked by AFM.** (a) AFM of DNA origami chains in TAE/Mg<sup>2+</sup> solution, with a model for a single origami rectangle, showing the hairpin label (“L”) for AFM contrast, the *parC* loop (purple) and the “remainder” strand (blue). (b) The same area as in (a), in ParM polymerization buffer, with a representative “jaw-dropped” origami highlighted. (c) The same area after the addition of ParR. Top: AFM image with an arrow indicating ParR binding along the sidewall helices of origami. Bottom: The same AFM image with relevant models superimposed (see text). (d) The same area after the addition of ParM. Scale: the longer side of each rectangle origami is ~100 nm.



loop generated by the scaffold part left unused by the rectangular shape ('remainder loop'; blue loop at the bottom of the rectangle). One can note that the *parC* strands are not visible by AFM at this step, most likely because they are relatively floppy (compared to the dumbbell hairpin structure tightly held down close to the origami surface) and can easily escape the sweep of the AFM tip. The remainder loop at the bottom of each rectangle is visible, although somewhat faint, because those loops can be held down directly to the mica surface. A wider area view of the same sample is shown in Figure IV-8. The area captured in Figure IV-7 is highlighted as a box in Figure IV-8a.

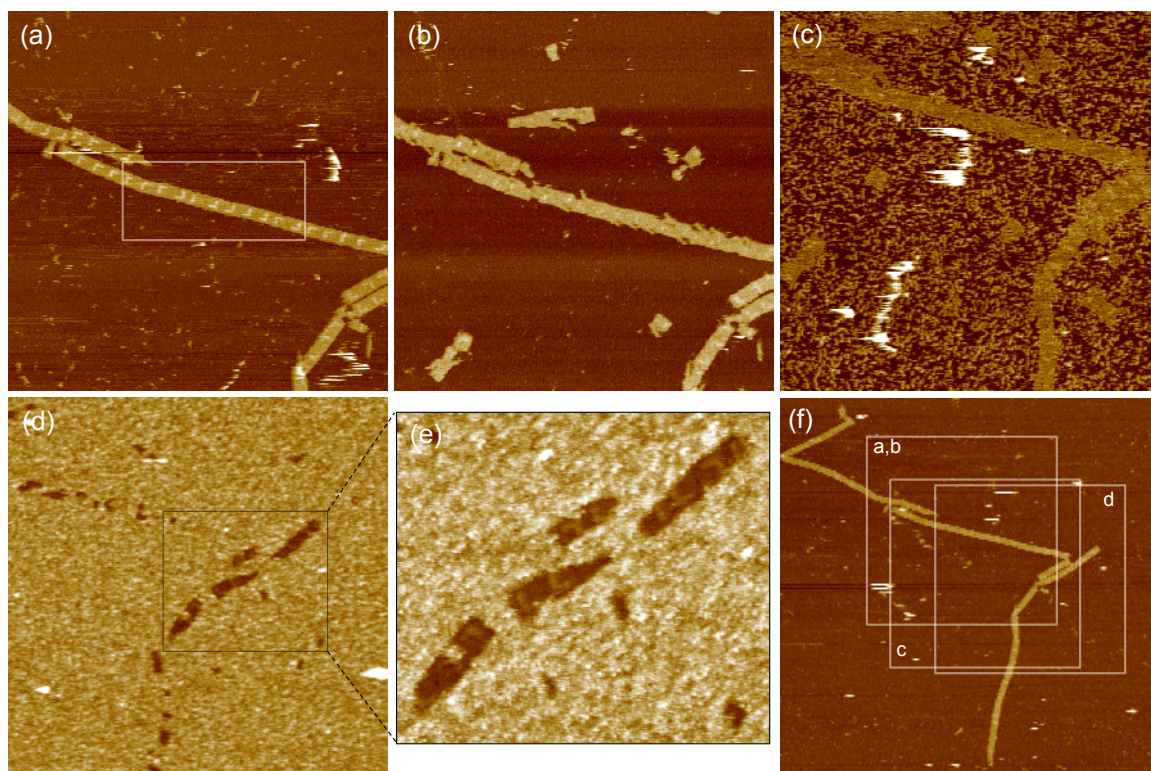


Figure IV-8. **Binding of ParR and ParM to DNA origami, each containing a single *parC*, tracked by AFM (wider view).** (a) DNA origami chains in TAE/Mg<sup>2+</sup> solution. (b) The same area as in (a), in ParM polymerization buffer. (c) The same area after the addition of ParR. (d,e) The same area after the addition of ParM. (f) A wider-view AFM that locates the areas shown in (a-d). Scale: the longer side of each rectangle origami is ~100 nm.

Figure IV-7b shows the chain after exchanging the buffer. The lower divalent cation concentration affects the system in several different ways. First, it makes the origami less tightly bound to the mica surface. It can be noticed by the generally less crisp AFM image while the AFM tip was still the same as the one used in the previous step; the edges of the origami chain is less well-resolved because the more loosely bound origami, especially the helices near the edges, would

react to the AFM tip movement in a more elastic fashion (attracted earlier and dragged longer) than the case where the origami are more tightly bound. Second, the weak divalent cationic strength also makes the dumbbell hairpins loosely held to the origami body. One can notice that the “L” patterns are now not as well visible as the previous step. Third, a more interesting phenomenon is that each rectangle has “jaw-dropping” deformation near the location of the *parC* strand (a representative one is contoured). This happens because the small discrepancy between the regular M13mp18 scaffold and the *parC*-M13 scaffold (engineered with M13mp8 as the basis; note the “8”) occurs

near the *parC* region, and because we used a staple set for the regular M13mp18 scaffold, which left ~25 bases unpaired making the region a weak point. Upon buffer exchange, the weak points started to break off and created this deformation. Fourth, appearance of relatively small structures, e.g., origami monomers and short (3-4 origami) chains in Figure IV-8b on the area where there were no such structures in Figure IV-8a, may indicate that those small structures may have broken off of another chain or aggregate, or deposited as those small size chains somewhere else from the beginning, and traveled to this area, potentially by the weakened interaction with the surface mediated by divalent cations.

To this sample we added ParR proteins, in high excess: ~1000× the number of origami rectangles, and hence ~1000× the number of *parC* strands (Figure IV-7c and Figure IV-8c). Since each *parC* strand has 10 binding sites for *dimers* of ParR, the “effective” ratio of the number of ParR binding sites to the number of ParR dimers in this case was ~1:50. Interestingly, ParR proteins not only bound to the *parC* strands, now making them visible (indicated by the orange contours in the bottom image in Figure IV-7c; the AFM images themselves are identical between the top and bottom images in Figure IV-7c), but they also decorated other “relatively accessible” double-stranded DNA helices—such as the helices along the edge of the origami chain (indicated by the yellow arrow in the top image) and the remainder loop at each rectangle (indicated by the light blue contour lines in the bottom image). As a DNA-binding protein, in order for ParR to find its recognition sequence (iterons in *parC*), it may have to first bind and search. At this regime of high ParR concentration, ParR may saturate any double-stranded DNA regions, except for those within the origami body, where helices are tightly interwoven together by crossovers of high density (every 1.5 turns, 15 or 16 bps for each helix); the gap between helices—estimated<sup>1</sup> to be ~1 nm on average—may not give enough room for ParR molecules, whose size is ~3 nm in the dimeric form<sup>100</sup>. The binding of ParR to the edges of origami might be partially facilitated by some interaction between the protein and the mica surface, somewhat suggested by molecules bound to the background area; they may be ParR proteins, aggregates of ParR proteins, or free staple strands

bound by ParR proteins that would not bind to the mica surface in the common AFM condition, although the majority of the free staple strands would have been removed through the buffer exchange step. Such interaction with the surface seems to be a kinetic trap that is irreversible even by extra washing.

The observation that ParR binds to non-recognition sequence DNA is somewhat contrary to a previous study in the literature<sup>99</sup>, where they measured the gel shift patterns depending on the relative concentrations of DNA and ParR, both for *parC* DNA and DNA with no sequence homology to the *parC* recognition sequences. Interestingly, they reported no gel shift for non-recognition DNA even at high concentration of ParR, but the range of the concentration ratio they tested remained somewhat narrow; they showed data for ParR concentrations up to 400× (800 nM) of that of *parC* (2 nM), which would give effectively 1:20 ratio between the recognition sites in *parC* and ParR dimers, 2.5-fold lower than the test condition in our assay. We further performed a series of titration experiments varying the concentration of ParR, which will be discussed in the next section.

Also, the observation that ParR binds to non-recognition sequence DNA lead us to pose a question: would ParM filaments bind to a loop created by ParR binding to non-*parC* DNA as well? In other words, is it just the geometry that really matters, rather than the DNA sequence? Perhaps the mere role of the *parC* sequence is simply facilitate ParR to form a structurally well-defined loop by providing repeated recognition sites (ten 11-bp iterons) and promoting cooperative binding<sup>99</sup>. Could we create a loop-like geometry similar to the ParR/*parC* complex, with non-*parC* DNA, and let ParR bind and stabilize ParM filaments? We plan to test this hypothesis by creating a cylindrical origami shape with dimension similar to the ParR/*parC* complex, which will be discussed in more detail in the Future Directions section.

It would actually be, in part, a test of such a hypothesis to add ParM to the origami chain sample where we already have ParR bound to non-*parC* DNA and see how ParM filaments would engage with the ParR proteins. Would ParM filaments have preferable binding to the ParR/*parC* complexes, compared to ParR bound to non-*parC* regions, such as the remainder loops or the edges of origami? To the sample of origami chains with ParR, we added ParM proteins to final concentration of 2.4 uM (chosen to be slightly above the critical concentration, 2.3 uM, to allow filament formation) and ATP to final concentration of 10 mM. Unfortunately, we could not see any differentiation between the ParR/*parC* complexes and other ParR-bound areas; rather, ParM would coat the entire mica, except part of the area where origami chains were deposited (Figure IV-7d and Figure IV-8d,e). It is interesting that the inner area of DNA origami structures repelled the protein,

which may reflect the fact that ParM does not interact directly with DNA<sup>99</sup>, but only when complexed with ParR. ParM binding onto the background area may have been mediated by the ParR bound to the background area. ParM proteins may have grown filaments and bundles of filaments, stabilized by ParR on and around origami and on mica, some of which may have grown on top of origami, covering some areas of origami chains.

#### IV.2.1.1.2. ParR still binds *parC* specifically at the right concentration regime

Since ParR was found to bind to non-*parC* DNA as well as *parC* DNA at high concentration, we performed a series of experiments varying the concentration of ParR, to see if ParR can specifically recognize the right sequence at a lower concentration regime. As briefly discussed in the previous section, contrary to our results, a previous study in the literature<sup>99</sup> reported that ParR does not interact with non-*parC* sequences. They used a gel shift assay varying the concentration of ParR (between 0 and 800 nM) with a fixed DNA concentration (2 nM). The highest concentration that they tested, 800 nM, 400× of the concentration of DNA, was 2.5-fold lower than the condition used in our experiment described in the previous section. Hence, we further tested lower ParR concentrations, with the ratios of *parC* to ParR concentrations 1:10, 1:100, and 1:400.

Figure IV-9 summarizes the results of tests with different *parC*:ParR ratios. First, in all cases, the characteristic “jaw-dropping” by the weak divalent cationic strength is observed. At the 1:10 ratio, which corresponds to effective ratio of 1:0.5 between the binding sites in *parC* and ParR dimers, there was no extensive coating of origami edges or loops by ParR that was observed in the 1:1000 case; yet, some *parC* loops were found to be bound by some objects of length scale similar to ParR (indicated by yellow arrows in Figure IV-9a). As the determination of the size, number and location of the ParR-like objects can become ambiguous, no attempt for quantification was made. A similar behavior was observed in the case of 1:100 ratio, which would correspond to effective ratio of 1:5 between the binding sites in *parC* and ParR dimers (Figure IV-9b). Binding of ParR-like objects was also observed (indicated by yellow arrows), while no extensive binding along the origami edges occurred. In the case of 1:400 ratio (effective ratio 1:20), the *parC* loops became more pronounced by putatively more binding by ParR, while there is still no extensive binding along the origami edges or the extra scaffold loops (Figure IV-9c,d). ParR shows very specific binding onto the *parC* sites. It is interesting that the 1:400 ratio is close to where ParR binding to the *parC* recognition sequence saturates in the gel data (between their data points 1:250 and 1:500) while the ratio is where non-*parC* sequence still does not interact with ParR in the study mentioned above<sup>99</sup>.

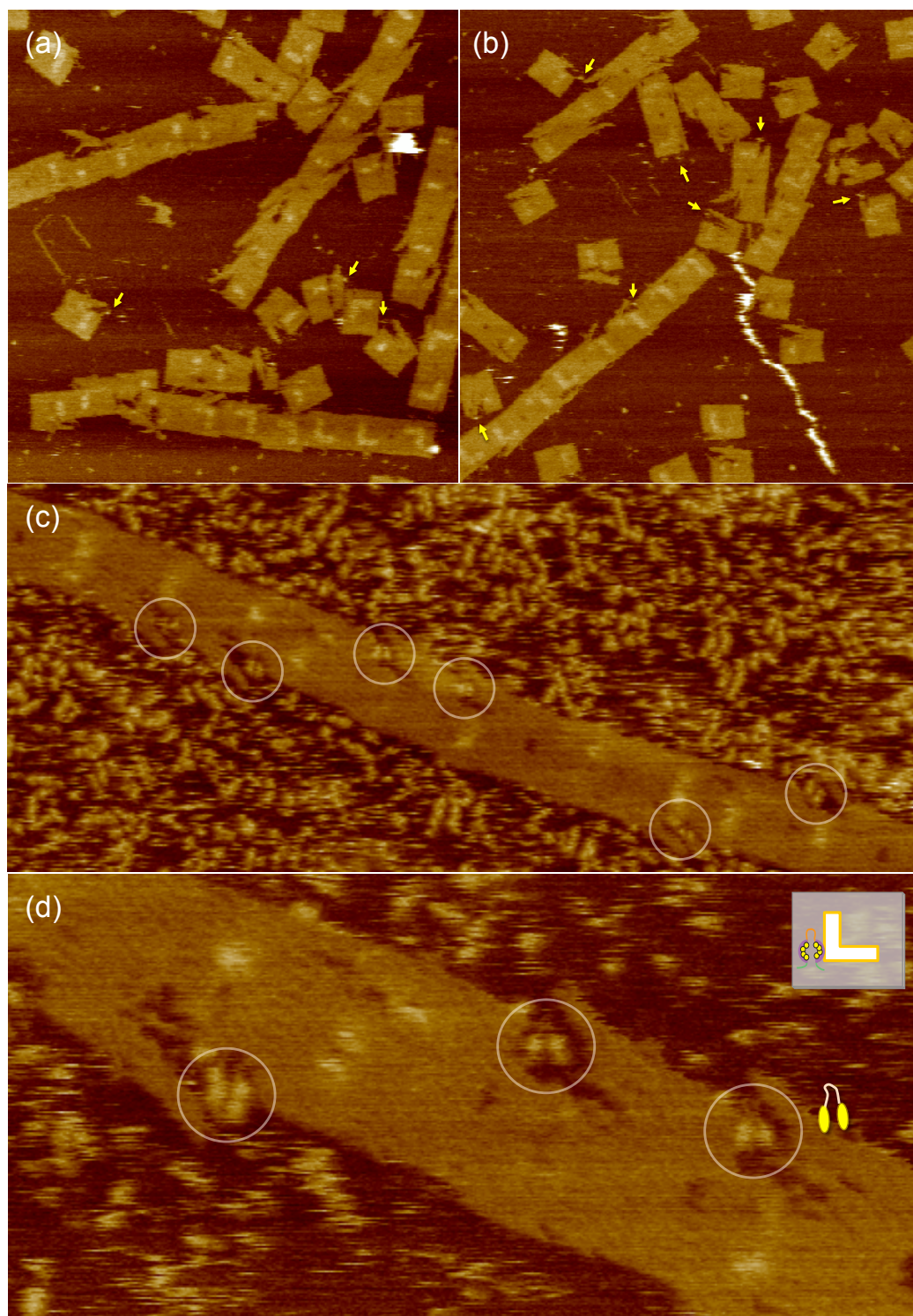


Figure IV-9. **Binding of ParR to *parC*-containing DNA origami, with varying concentrations of ParR.** Number ratios between *parC* strands and ParR monomers were (a) 1:10, (b) 1:100, and (c,d) 1:400, respectively. At the ratio of 1:400, ParR specifically binds to the *parC* sequence. Scale: the longer side of each rectangle origami is ~100 nm.

A detailed look at the images roughly reveals the structure of the ParR/*parC* complexes (Figure IV-9d). Each complex shows slightly different structure but has in common two oval parts and loops connecting the two. The loops have roughly the same height as that of the rectangle origami (based on the color profile), so it is highly likely that they are single DNA double strands, and in particular, the promoter part (~59 bps) between the two sets of five iterons (ParR binding sites) in the *parC* sequence. This structure is, in fact, in very good agreement with previous AFM data for short DNA fragments in the literature (shown in Figure IV-2c), where they claimed the ParR/*parC* complex to be “U-shaped” based on their AFM data<sup>103</sup>. However, one needs to be careful in interpreting AFM data, as the interaction between molecules and the mica surface can create various kinds of artifacts. For example, the ParR/*parC* complex may simply get squashed onto the surface, while adopting, for example, a helical or circular form in solution.

In summary, using our single-molecule DNA origami templates, we were able to reveal that ParR can bind to non-recognition sequence DNA, in addition to the *parC* sequences, under high stoichiometric conditions of ParR. But in the right concentration regime, ParR did specifically bind to *parC* and discriminated non-recognition sequences. The ParR-*parC* interaction observed in our system in the recognition regime is consistent with previous studies in literature in terms of both stoichiometry<sup>99</sup> and structure<sup>103</sup>.

#### IV.2.1.2. Origami with multiple *parC* strands

Although the approach of using a single *parC* sequence on each origami rectangle was illuminating to us in confirming the interaction between *parC* and ParR, it was limiting in many ways. First, AFM was found to be not as effective as we had hoped in observing the interaction between ParR/*parC* complexes and ParM filaments. ParR seems to bind to bare mica surfaces to some extent, and ParM seems to get stabilized by ParR on the background area. If we passivate the surrounding area to prevent ParR binding (not tested), then it may block ParM filaments from binding as well. But then it will make it harder to scan the filaments with AFM tips, because the filaments would not stick to the surface well. Overall, the nature of AFM measurement relying on molecules having to stick to surface turns out unsuitable to studying these protein molecules. Second, the single-*parC* approach is not compatible with other measurement methods, e.g., gel electrophoresis or fluorescence microscopy. The mass difference caused by ParR binding to *parC* on origami would be too small to be resolved by agarose gel. Also, fluorescence imaging may not be effective to resolve single ParM filaments attaching to single ParR/*parC* complexes on origami.

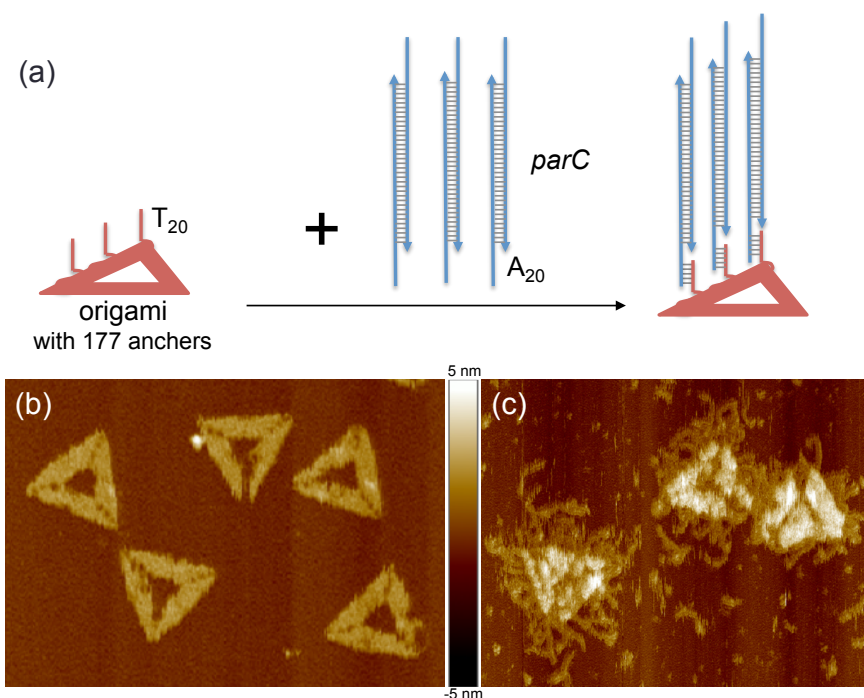


Figure IV-10. **DNA origami triangle with multiple *parC* strands.** (a) Model diagram of the triangle with anchors ( $T_{20}$ ) for *parC* strands, without and with *parC* strands bound. (a) AFM of DNA origami triangles without *parC* strands. (b) AFM of DNA origami triangles with *parC* strands. Scale: one side of a triangle origami is  $\sim 130$  nm.

We decided to test DNA origami with multiple *parC* strands, to be able to examine more bulk interactions between *parC* strands, ParR and ParM, at the regime where distortion by surface to measurements is minimal. We created DNA origami with multiple *parC* strands by incorporating linear *parC* strands onto an origami structure via linkers extended from 5' ends of staple strands. *parC* strands were further modified for fluorescence labeling. The detailed design of the *parC* strand and the linker for origami are described in the Materials and Methods section. We used triangle origami and rectangle origami for the investigation.

Figure IV-10 shows the model diagram (Figure IV-10a) and AFM images of the triangle origami in the absence (Figure IV-10b) and the presence (Figure IV-10c) of the *parC* strands. As can be seen in Figure IV-10c, *parC* strands (199 bps, double stranded part, including the linker to origami) stretch out of the triangular shape of the origami. Since the *parC* strands were linked to staple strands that compose the body of origami, they make the origami body thicker, as compared to the single double-strand layer of *parC* strands, as shown by the colored height profile.

#### IV.2.1.2.1. AFM movies of ParR-*parC* binding

Although the use of origami with multiple *parC* strands was intended for characterization using methods other than AFM, AFM is still a very powerful tool to reveal and confirm details of structures and phenomena at the molecular level. Using the triangle origami functionalized with multiple *parC* strands, and using a high-speed AFM, we could obtain AFM movies that show progressive binding of ParR onto *parC* strands on origami. Figure IV-11 shows the model diagram and frames taken from two independent movies of *parC*-functionalized origami, to which ParR molecules were binding progressively. ParR was added into a large (~100 ul) buffer bath *in situ* while scanning, so the progression of binding also reflects the diffusion of ParR molecules in solution, in addition to the potential binding, searching and unbinding process.

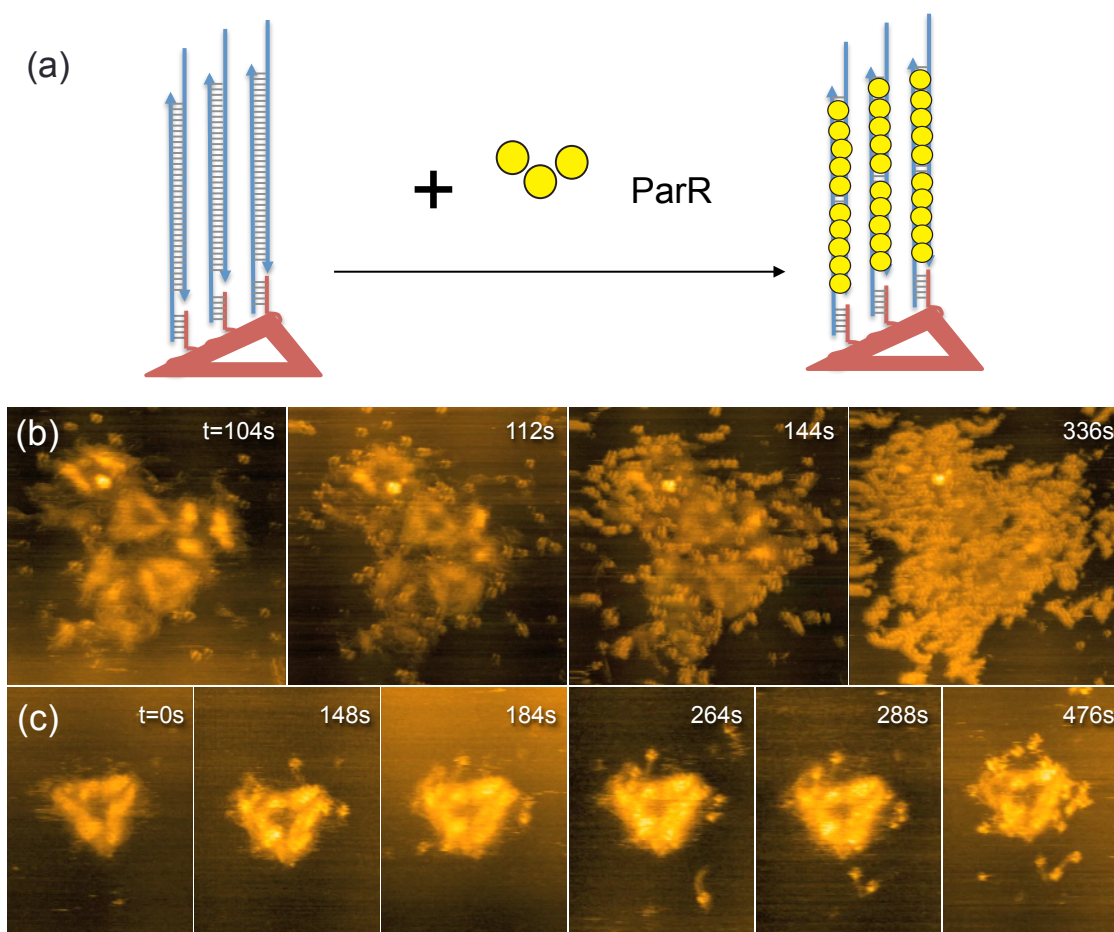


Figure IV-11. AFM movies of ParR binding to *parC* strands on origami. (a) Model diagram of the triangle with multiple *parC* strands, before and after binding of ParR. (b,c) Representative frames from two independent AFM movies that show the progressive binding of ParR to *parC* strands on origami. Time points are indicated at the top of each frame. Note that in (b) the AFM tip was doubled. Scale: one side of a triangle origami is ~130 nm.



The ratio between the total numbers of molecules *parC* and ParR was between  $\sim 1:10$  and  $\sim 1:100$ , the stoichiometric ratios where binding was unsaturated in the single-*parC* origami experiments. These movies are in good agreement with the observation for single-*parC* origami in that the *parC* strands do not seem to be saturated and form loop-like structures. Although it looks as if a large number of ParR molecules are bound to origami in a single imaging field, it may be due to the high local density of *parC* strands (177 per origami). Since these high-speed AFM experiments were carried out during demo sessions at a conference away from the lab (see Materials and Methods), it was difficult to perform more experiments with careful control over the concentrations due to the lack of time and resources.

Nonetheless, these movies provide interesting footage of the dynamic behavior of the DNA-binding protein, including some transient binding of ParR molecules to the *parC* strand. The movie files can be downloaded from <http://dna.caltech.edu/Woo-thesis-movies>. The movies were obtained at a rate of 1 frame / 4 seconds (by taking an AFM image every 4 seconds), and played at a rate of 4 frames/sec (16 $\times$  faster).

#### IV.2.1.2.2. Gel electrophoresis study of concentration dependence of ParR binding

Binding of ParR to the multiple *parC* sites on origami was also confirmed by gel electrophoresis. Different extents of gel mobility shift depending on the concentration of ParR revealed concentration dependence of ParR binding to *parC* sites on origami. Figure IV-12 shows the gel data. There are a few additional interesting points to note.

First, the ParR concentration dependence of ParR-*parC* binding can be recognized in a few different ways. The blue dotted boxes highlight the different gel mobility shifts of the origami that contain multiple *parC* strands. From the relatively fast migrating origami-only band (lane 2)<sup>§</sup>, the origami band loses its mobility when bound by multiple *parC* strands (lane 3) and subsequently by ParR (lanes 4-7). The mobility decreases progressively with increasing ParR concentrations from 10 $\times$ , to 20 $\times$ , 50 $\times$ , and 100 $\times$ . Near the highest concentration of ParR (50 $\times$  and 100 $\times$ ), the mobility reduced significantly and a lot of materials were left unmigrated into the gel in this particular experiment. Those materials might have migrated a bit if the gel had been kept running longer, but it may be the case that either the origami-*parC*-ParR complexes with increasing occupancy of ParR

---

<sup>§</sup> In lane 2, there are two bands for origami within the blue dotted box, the higher band of which was later found to be structurally disrupted origami because of the absence of Mg<sup>2+</sup> in the gel-running buffer. Those disrupted origami are believed to have failed to overcome the high strain in origami crossover structures, thus, at least for this gel mobility experiment purposes, the *parC* strands are expected to be still incorporated and function similarly to those on normal origami.

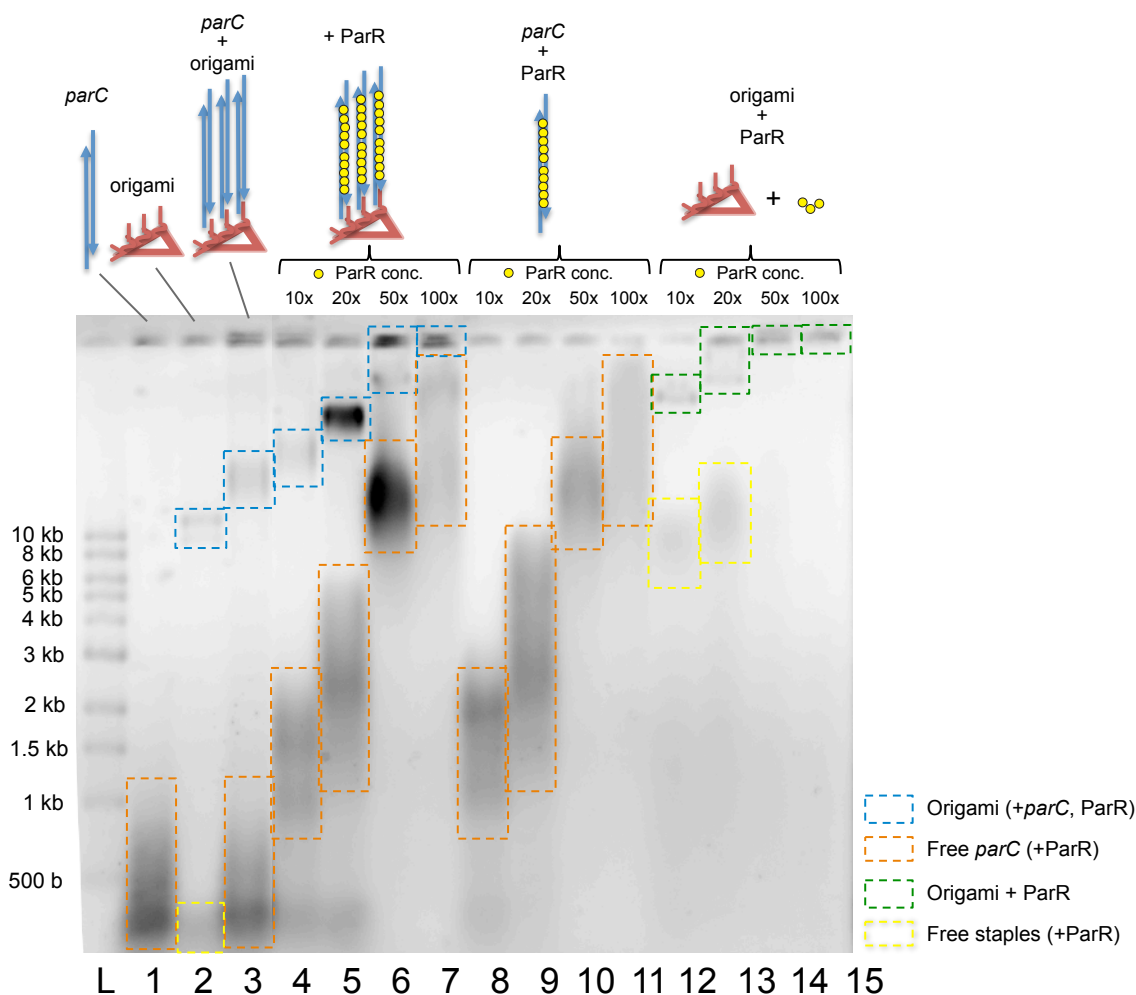


Figure IV-12. Gel electrophoresis data of ParR binding to *parC* strands on origami, which revealed both the concentration-dependent binding of ParR to *parC* and nonspecific binding of ParR to non-recognition DNA. (L: 1kb DNA ladder) The ratios given for ParR are relative to the *parC* concentration of each lane. For lanes 12-15, the same amounts of ParR were added as in the cases with *parC*.

on each origami just became too bulky to go through the gel (1% agarose), or ParR may be positively charged and may have neutralized DNA, substantially reducing the mobility in gel electrophoresis.

The ParR concentration dependence can also be observed in the free *parC* bands. The wide band in the lower area (highlighted by the orange box) in lane 3 corresponds to free *parC* strands that did not bind to origami (initially added in  $\sim 1600\times$  excess of the origami scaffold,  $\sim 2\times$  of total staple concentration), as verified by comparing it with the band in lane 1 (free *parC* only band). Since *parC* strands were added in  $\sim 2\times$  concentration of total staples, about half of the population in the band in lane 3 are bound to free staple strands, but the change in molecular weight is small (52

nucleotides) compared to the weight of *parC* (total ~370 nucleotides) and does not get resolved in agarose gel. The mobility of those *parC* strands get shifted by ParR, with the extent increasing with increasing concentration of ParR (lanes 4-7). The same pattern can be observed for free *parC* strands mixed with ParR in lanes 8-11, with the relative gel heights between corresponding bands agreeing quite well.

An interesting observation is made for the lanes for origami that do not contain any *parC* strands (lanes 12-15). ParR was found to bind to the non-recognition sequences on the origami (higher bands, highlighted by green boxes) and on the free staples as well (lower bands, highlighted by yellow boxes; note that free staples in lane 2 appear in a similar position to free *parC* strands because of the low resolution of agarose gel), consistent with our earlier observations on origami containing single *parC* under AFM. Origami bands appear to migrate slower by the binding of ParR, and such non-recognition binding also occurs in a concentration-dependent manner (lower mobility with higher ParR concentration). Note that the ratios of ParR given are relative to *parC* strands as if there *were* *parC* strands in the system, to keep the condition the same as the cases with *parC* strands (lanes 4-7). The concentration ratios between *origami* and ParR become ~16,000 $\times$ , ~32,000 $\times$ , ~80,000 $\times$ , and 160,000 $\times$  for 10 $\times$ , 20 $\times$ , 50 $\times$ , and 100 $\times$ , respectively, all of which are enormously excessive amounts from origami's point of view. These ratios between origami and ParR were still the same for the cases with *parC* (lanes 4-7). Note also that the bands of nonspecific complexes, origami-ParR (green boxes) and staple-ParR complexes (yellow boxes) are located at higher positions than their counterparts of specific complexes, origami-*parC*-ParR (blue boxes) and *parC*-ParR (orange boxes)—compare lane 12 with lane 4, and lane 13 with lane 5, etc., despite the absence of *parC* in each complex. Ironically, this is in fact because of the absence of *parC*; the ParR molecules, which would have been bound to the *parC* strands both anchored on origami and free in solution in the presence of *parC* (lanes 4-7) and hence would have been “consumed”, do not have the specific targets and instead would extensively bind to the bare origami and staple strands. Note that *parC* strands were added at 2 $\times$  the concentration of staples in the cases of lanes 4-7, so the origami and staple strands in lanes 12-15 have roughly double the amount of ParR to “accommodate” per each origami or staple, compared to the cases in lanes 4-7. In addition to that, the high excess amount of ParR and extensive binding to origami and free staples might cause them to aggregate, thereby yielding larger and less mobile structures. Also, potential spurious interactions between free staple strands via partial complementarity may allow transient formation of duplex regions, which may get bound and stabilized by the high excess amount of ParR, which then may propagate into large network structures.

## IV.2.2. Expandable nanostructures

With confirmation of the ParR-*parC* interaction with single *parC* embedded or multiple of them anchored on origami, we proceeded to testing the idea of expandable structures based on DNA origami. We first started with reconstructing the dynamic ParMRC system, where growing ParM filaments push ParR/*parC* complexes apart (in the absence of any bond to break). We used DNA origami coated with *parC* as ParR/*parC* cores, where a number of ParR/*parC* complexes would altogether stabilize bundles of ParM filaments. We used fluorescence microscopy to observe their behavior. Then, later we tested if growing ParM filaments or bundles of filaments could push apart origami associated in a larger structure, e.g. chains, by stacking bonds.

### IV.2.2.1. Model system: origami as ParR/*parC* cores stabilizing ParM filaments

To create nanostructures that can be expanded by dynamically growing protein filaments, we first wanted to reconstruct a model system with essential components that work well with DNA origami while keeping their native functions. In 2007, Garner et al.<sup>105</sup> isolated the three essential components of the ParMRC system and successfully reconstructed the native behavior using micron-sized beads coated with *parC* strands, where the *parC* strands on the beads become coated by ParR and stabilize the ends of growing ParM filaments, thereby being pushed by the filaments. We initially tried to use the same components, except for the beads; in the place of beads, we used fluorescently labeled DNA origami containing multiple *parC* strands as ParR/*parC* cores, and tried to observe ParM filament dynamics under fluorescence microscope, also using fluorescently labeled ParM proteins. Details are described in the following subsections.

#### IV.2.2.1.1. Dynamic behavior of ParM filaments

As probing ParM filaments using AFM turned out to be impractical, we switched to fluorescence microscopy for observing the behavior of ParM filaments. ParM filaments exhibit dynamic instability; filaments grow by association of ATP-incorporating ParM monomers, and shrink by dissociation of ADP-ParM monomers that are less stable in the filament form, after hydrolysis of ATP by ParM. A single ParM filament grows at a rate of  $\sim 5$   $\mu\text{M}/\text{s}$  and shrinks at a rate of  $\sim 64$   $\mu\text{M}/\text{s}$  (ref. 97). As part of reconstruction of the model system based on DNA origami, we first wanted to confirm the dynamics of ParM-only system in our lab, using our instruments. However, initial attempts were not successful. In retrospect, the main reasons were photo-damaging of filaments and improper surface treatment. To be able to see filaments under fluorescence

microscope, the fluorescent dyes (Alexa488) need to be excited by light source of proper wavelength ( $\sim 488$  nm). Upon light irradiation, while we could see a number of filaments that had been grown in that area, we observed the filaments suddenly shrinking, breaking and disappearing out of the field (Figure IV-13). When moved to a nearby area, there were again a number of filaments, which again soon disappeared. We knew that the filaments had grown, but we just could not capture the moment of growing, because as soon as we shined light, the grown filaments broke and disappeared. It seemed to be photo-damaging rather than pure photo-bleaching, because instead of the whole field uniformly darkening, the filaments tended to break in the middle and shrunk from their ends, with small pieces quickly diffusing away. To overcome the problem, reducing the light intensity by using neutral density filters was essential, but it sacrificed the signal as well—there was a tradeoff between signal and photo-damage. A less destructive imaging system, total internal reflection fluorescence (TIRF) microscopy, which uses a narrow window ( $\sim 200$  nm) of evanescent wave rather than direct light irradiation, hence exciting molecules only near the surface and minimizing disturbance to the remaining system, was also tried. However, now that the depth of view was limited near the surface, the surface background became more of an issue. Even with a narrower window of excitation, the signal-to-noise ratio did not improve, perhaps because surface-bound fluorescent monomers provided high background. Proper surface treatment that can discriminate filaments above a certain length from monomers needed to be devised.

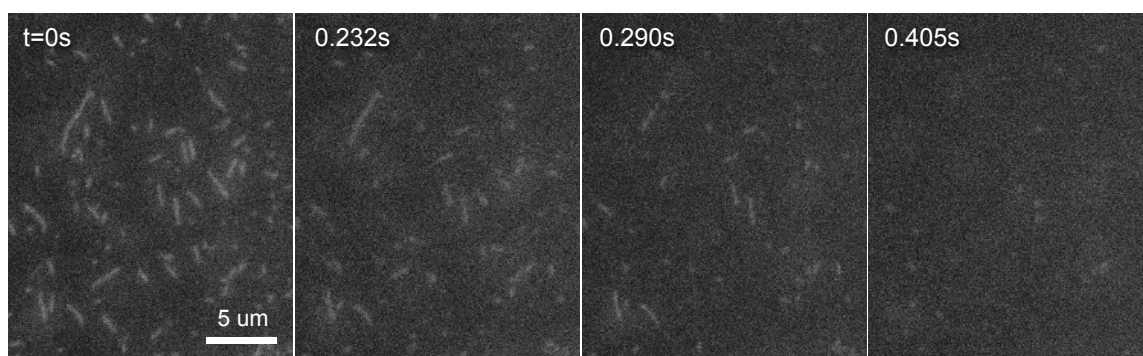


Figure IV-13. **ParM filaments that break, shrink and disappear from the field of imaging by photo-damaging.** Time points are indicated at the top of each frame. Note the rapid time scale. The longest filament ( $\sim 5$   $\mu\text{m}$ ) shown in this field dissociates completely within  $\sim 0.4$  s, which corresponds to a  $k_{off}$  of  $\sim 5 \times 10^3$  /s, about 100-fold higher than normal.

Nonetheless, despite those limiting conditions, we were able to capture the growing moment of some filaments. It was still a very rare event, which required analysis of multiple movies of filaments, the majority of which were merely shrinking. The signal-to-noise ratio was still low. One representative filament that grew for a while then shrunk is shown in Figure IV-14. From the

time lapse images, we could obtain the length change curves for the growth and shrinkage phases for this particular filament (Figure IV-14b). From linear fits to the curves, we estimated the rate constants:  $k_{on} = \sim 9.9$  / $\mu\text{M}/\text{s}$ , and  $k_{off} = \sim 27.7$  /s. These values are within the same order as the known values (5 / $\mu\text{M}/\text{s}$  and 64 /s, respectively<sup>97</sup>), but the on rate is higher and the off rate is lower, each by a factor of  $\sim 2$ . This might suggest that the surface is giving extra stabilization to ParM monomers, which is consistent with the high background observed.

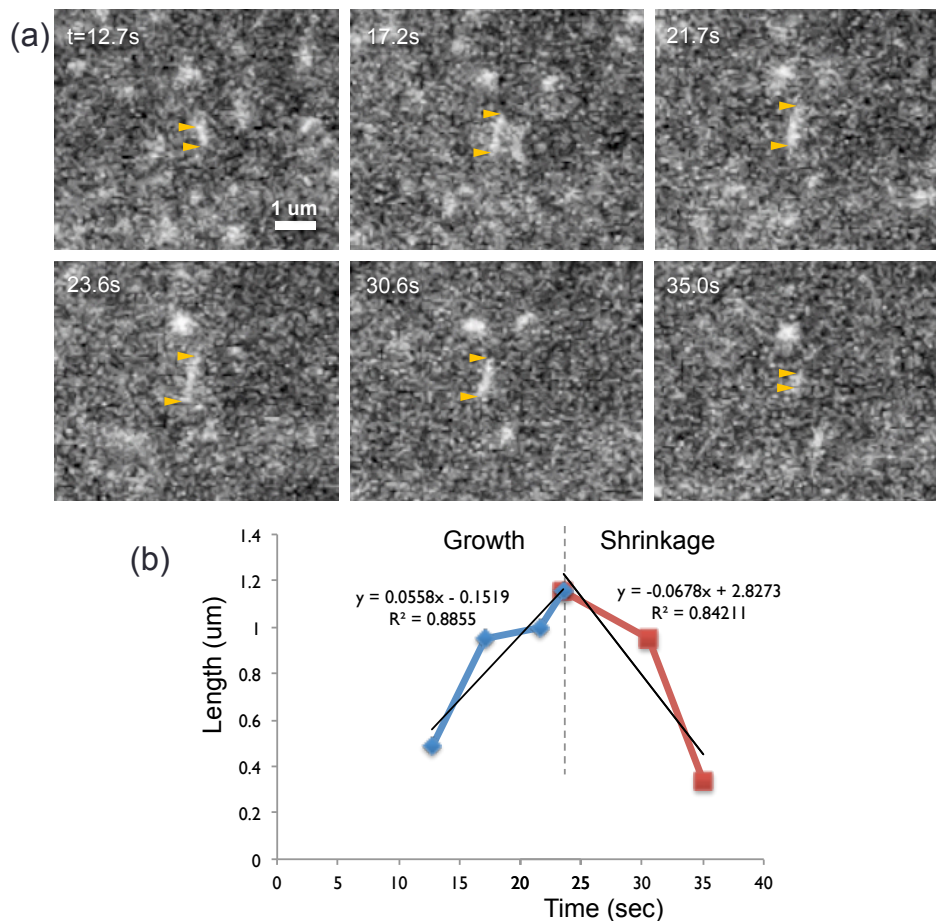


Figure IV-14. **Dynamic instability of a ParM filament.** (a) Representative frames from a movie that captured the growth and shrinkage of a ParM filament. (b) Plot showing the length change and the linear fits for the growth phase and the shrinkage phase.

#### IV.2.2.1.2. Colocalization of ParM filaments with DNA origami containing ParR/*parC* complexes

After we confirmed that ParM behaves as expected, exhibiting dynamic instability, and we could observe the behavior under our instrument (with some limitations), we moved forward to mixing ParM filaments and DNA origami that contain *parC* strands, along with ParR and ATP. We

labeled DNA origami with Cy3 and ParM with Alexa488. Although we still struggled capturing dynamic behavior of ParM filaments, we were able to observe some colocalization of DNA origami containing ParR/*parC* complexes and the ends of ParM filaments in static images (Figure IV-15), suggesting that the ParR/*parC* complexes do engage with the ends of ParM filaments.

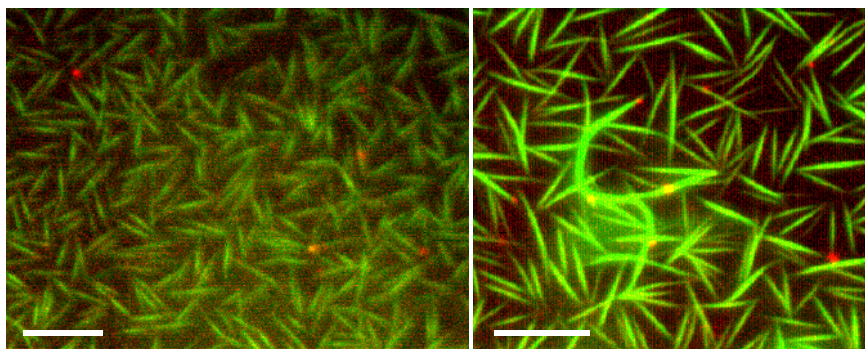


Figure IV-15. **Colocalization of the ends of ParM filaments and DNA origami containing ParR/*parC* complexes.** Majority of the origami with ParR/*parC* complexes (labeled red) appear at the ends of ParM filaments (labeled green), suggesting the binding of ParM filaments to ParR/*parC* complexes. Scale bars: 5  $\mu\text{m}$ .

#### IV.2.2.1.3. Reference system with microbeads in place of origami

While we could get some evidence that individual components work properly—ParR binding to *parC* sites on origami with specificity, and ParM filaments dynamically growing and shrinking—and we could witness colocalization of origami containing ParR/*parC* complexes with the ends of ParM filaments, as discussed in several previous sections, we could not successfully get any dynamic behavior of the system when they were mixed altogether. Could it be the DNA origami structure, e.g., the size, aspect ratio, etc., that prevents the system from working? To test the hypothesis that the DNA origami might cause some problem and not other components, we decided to take a step back and reconfirm that all other components worked fine by using micron-sized beads in the place of origami, which is closer to what Garner et al. tested in their first reconstruction system<sup>105</sup>. We used fluorescent beads (Nile Red—whose emission spectrum is wide and lies across both the Alexa488 channel and Cy3 channel) with  $\sim 1 \mu\text{m}$  diameter, coated with streptavidin. We had biotin at the end of each *parC* strand (see Materials and Methods), so we were able to coat the beads with the *parC* strands via the streptavidin-biotin interaction.

Interestingly, the use of microbeads revealed the most important problem—surface treatment. Even with the beads, our initial attempts to get dynamic behavior of ParM filaments failed. Meanwhile, we often observed that the beads were stuck to a glass surface, and even

witnessed the cases where beads and filaments were stuck on a different surface (either the top—coverslip—or the bottom—glass slide). This led us to the realization that the surface might have been the main issue—not only creating high background signal (which was noted earlier), but more importantly, being too sticky to the molecules that it distorted the concentrations of all the components. We had been following the standard protocol for surface passivation suggested by the group that provided the proteins, which is to add bovine serum albumin (BSA) into the system such that it dominates the surface and blocks other proteins from binding. But that protocol did not seem to be optimal at least for our system. Alternatively, we tried to use the older protocol that was originally used in the 2007 Garner et al. reconstruction work, which is chemically treating the surface to make it highly hydrophobic (silanization using diethyl-dichloro-silane; see Materials and Methods). Using the chemically treated glass slides and coverslips, we could finally observe a dynamic behavior of ParM filaments, pushing apart microbeads coated with ParR/*parC*. Some frames are captured in Figure IV-16, and the full movie is available for download at <http://dna.caltech.edu/Woo-thesis-movies>.

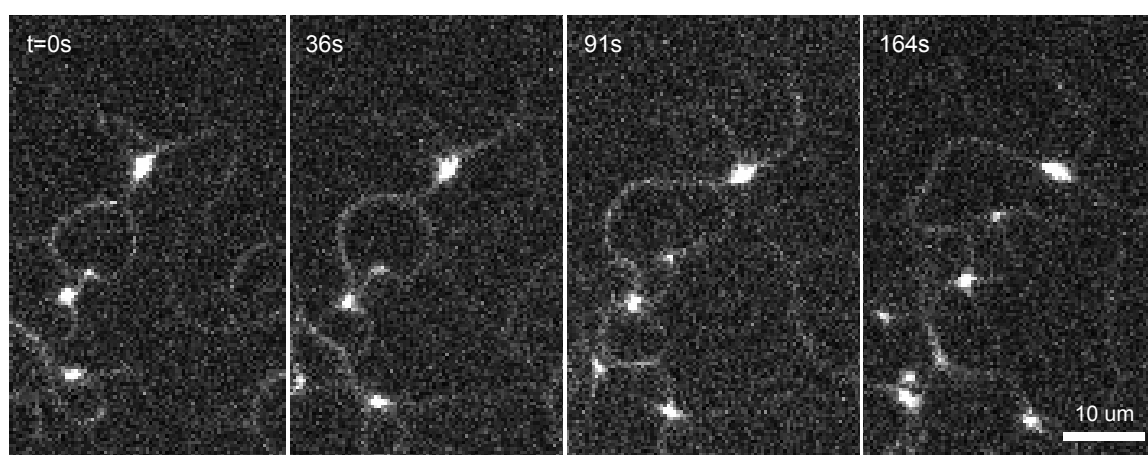


Figure IV-16. **Frames from a movie that shows networks of ParM filaments/bundles growing and pushing microbeads apart.** Time points are indicated at the top of each frame. The beads show up as dots that are brighter than the filamentous parts.

#### IV.2.2.1.4. Large scale expanding network structures in the high origami concentration regime

Now, with the surface problem solved, we went back to testing the system with DNA origami. Using triangular origami at  $\sim 400$  pM, each containing 177 *parC* strands, we were able to capture dynamically expanding ParM filaments (Figure IV-17, the full movie is available at <http://dna.caltech.edu/Woo-thesis-movies>). Although each origami core is not identifiable because



they were not fluorescently labeled, the fact that the expanding filaments were stabilized by origami is supported by a control experiment where, in the absence of origami, ParM filaments that potentially form are not observable under the same setup, perhaps due to their transient nature and narrow thickness.

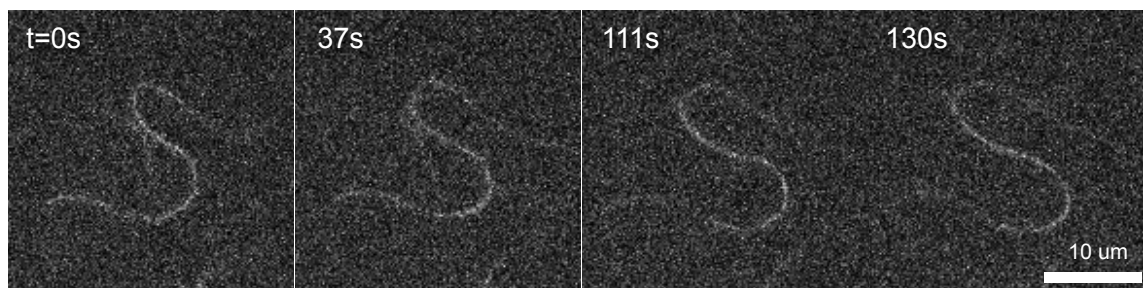


Figure IV-17. **Frames from a movie that shows networks of ParM filaments/bundles which grow and expand in the presence of DNA origami with ParR/*parC* complexes.** Time points are indicated at the top of each frame. DNA origami were not fluorescently labeled, so they do not appear.

#### IV.2.2.1.5. Distinct filament bundle stabilization in the low origami concentration regime: Exploring the vast parameter space, especially the underestimated role of the crowding agent

Although the expansion of ParM filament bundles stabilized by DNA origami ParR/*parC* cores described in the previous section was promising, the system is not ready to be used for expandable structures because the filament bundles tended to form large scale network structures with one another. We wanted to find a regime where filaments/bundles do not form extensive networks and are stabilized by one origami containing clusters of ParR/*parC* complexes at each end.

Finding the right regime required exploring a large, multi-dimensional parameter space of concentrations of all different components—much larger than we initially anticipated. We initially thought that the space would be defined by the concentrations of just the three essential components—*parC*, ParR, and ParM. In fact, even within just the three-dimensional space, there were a large number of cases to try, to find a regime that works reliably and in a somewhat expected way. Yet, there were more variables that we had not considered important.

The most influential variable was the concentration of the crowding agent, methylcellulose in our system. Methylcellulose is often used for microscopy samples that contain molecules with tubular shapes to help confine them near slide/cover slip surfaces<sup>106</sup>, while not letting them permanently stick to the surfaces. Methylcellulose can also cause molecules with tubular shape to

bundle together<sup>107</sup> by volume exclusion. The effect of methylcellulose on bundle formation also applies to ParM filaments; the higher the concentration of methylcellulose, the more bundle formation is promoted. But we had not realized that the concentration of methylcellulose was a critical parameter for even “observing” the filament bundles under the fluorescence microscope. We later learned that without using methylcellulose, ParM filament bundles do not get stabilized and the *in vitro* reconstructed ParMRC system does not work<sup>108</sup>. A small increase of methylcellulose (by ~0.4% of total solution volume) induced too much stabilization to the bundles, causing spontaneous bundle formation and stabilization of the filaments even without ParR/*parC* complexes at the ends.

In addition, the fraction of fluorescent ParM among the total ParM monomer population was another important variable. Adding fluorescently labeled ParM is essential for the ability to observe their behaviors, and the fraction of fluorescent ParM needed to be optimized to obtain the best signal-to-noise ratio and a signal level comparable to that from the other channel for origami or beads. However, the fluorescent tags may somehow affect the behavior of the individual protein, so using high fractions of them may change the overall behavior of the system. This turned out to be true in our experiments; with 100% labeled ParM, filaments seemed to become overly stabilized even without any other stabilizing factors (e.g., ParR/*parC* or AMPPNP, a non-hydrolyzable ATP analog), perhaps suggesting the fluorescence labeling might have disabled the ATPase activity of some portion of the ParM monomers, or otherwise changed the monomers to favor the configuration in filaments and bundles.

After exploration of the large parameter space, we found a regime where distinct filament bundles are stabilized by ParR/*parC* complexes on origami at each end. To get a smaller number of “stabilization cores”, a lower concentration of DNA origami was used (~50 pM, with total [*parC*] = ~10 nM). Also, the concentration of methylcellulose was reduced (from 0.8% to 0.4 %) to suppress spontaneous bundle stabilization. The concentration of ParM monomer was increased (~7 uM) to increase the thickness of bundles that do get specifically stabilized by ParR/*parC* and hence improve the signal-to-noise ratio. Some frames from a representative movie are shown in Figure IV-18, and the full movie is available at <http://dna.caltech.edu/Woo-thesis-movies>.

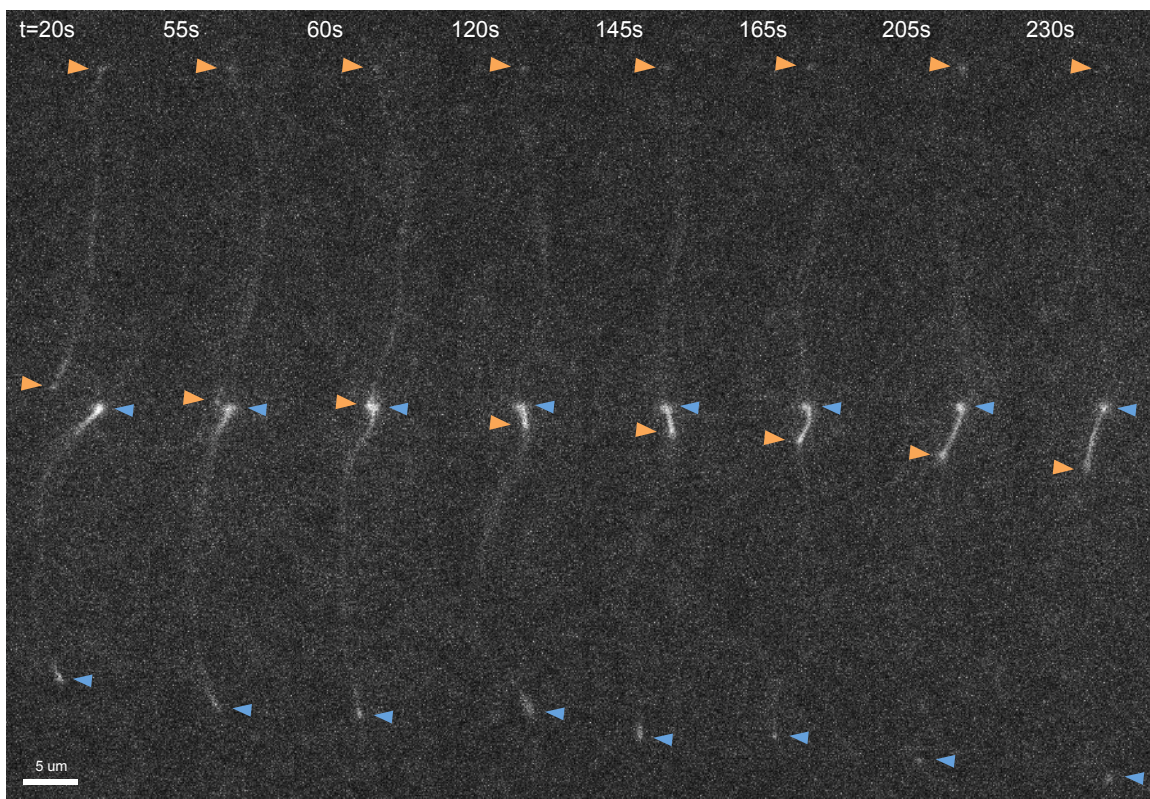


Figure IV-18. **Frames from a movie that shows distinct ParM filaments/bundles whose ends are stabilized by DNA origami with ParR/*parC* complexes.** Time points are indicated at the top of each frame. At the regime of lower concentration of origami and higher concentration of ParM monomers (than in the case of Figure IV-17), more filaments are anchored at each origami, forming thick bundles of filaments. In this series of images, two independent filament bundles appear, whose ends are visible. The ends of the top bundle are indicated by orange arrowheads and the ends of the bottom bundle are indicated by blue arrowheads. The two ends meet and the top bundle continues to grow along the bottom bundle. The bottom bundle also grows, which can be observed by the relative movement of the bottom end of the bundle.

#### IV.2.2.2. Expanding origami chains

Now that we found the working regime for the ParMRC system mediated by DNA origami, we moved on to testing expandable structures—once connected and later dynamically expanded by actively growing protein filaments. We first tested chains of rectangular origami connected by stacking bonds, because (1) they are easy to create—something that we know works very well and we are very familiar with, (2) stacking bonds we believe are a good candidate for reconfigurable bonds because they may have lower activation energy to mechanical deformation such as shear force, and (3) it is easy to detect the change in structure under fluorescence microscope due to its

large size in length.

#### IV.2.2.2.1. Design of origami chains

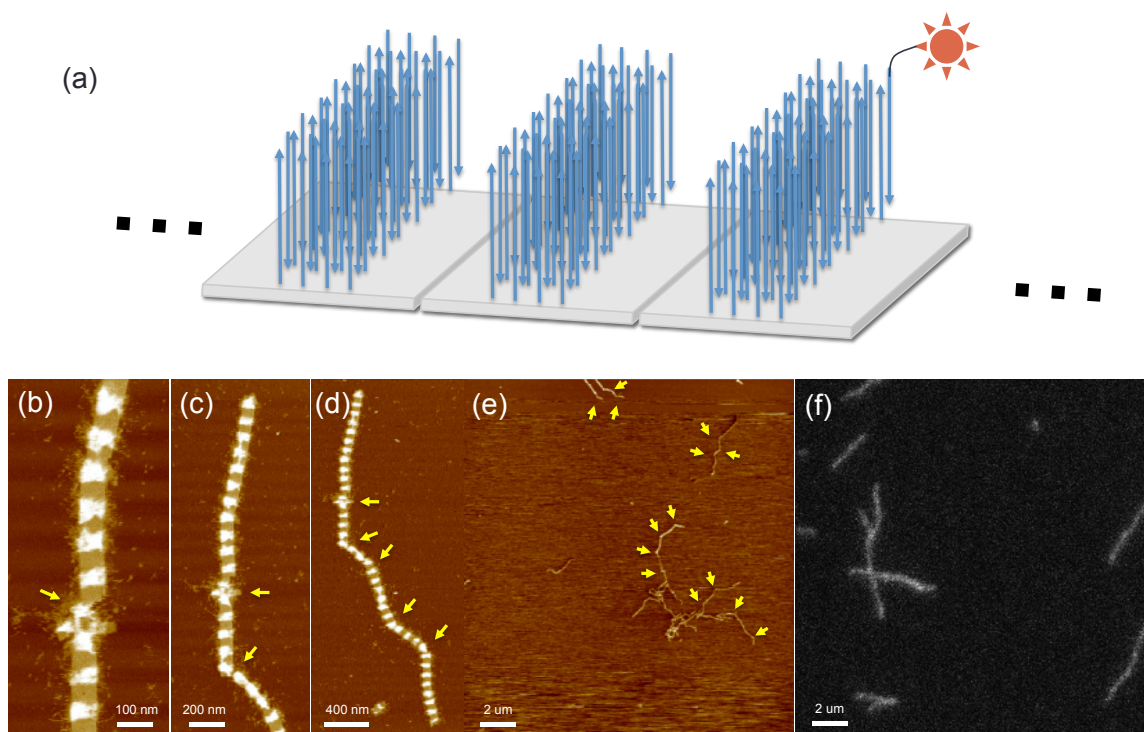


Figure IV-19. **Design and microscope images of DNA origami rectangle chains.** (a) Schematic diagram of the DNA origami rectangle chains used for the study. 96 *parC* strands (depicted as vertical stretches of DNA, with smaller number than actual, for simplicity) were anchored in the central half area of each rectangle. Each of the *parC* strands was fluorescently labeled with Cy3. (b-e) AFM images and (f) a fluorescence microscope image that show the rectangle chains. Densely packed *parC* strands seem to induce twist over the length of the chains, generating somewhat periodic kinks, folds, and breaks when deposited on a surface (indicated by arrows).

We designed staple strands with 5'-end extension ( $T_{20}$ ) to anchor *parC* strands for the rectangle origami. By choosing a subset of staple locations, we can control the total number and locations of the *parC* strands. For initial tests of expandable origami chains, we used 50% of the staples containing the anchor for *parC* strands (96 staples) at the center half area, far from the edges (Figure IV-19a). Each of the *parC* strands were fluorescently labeled with Cy3 (by hybridizing  $(CTT)_5$ -Cy3 with the end part of the *parC* “bottom” strand). All rectangles were first made without any edge staples to prevent stacking, gel purified, then put through a second annealing step with edge staples to allow chain formation. Figure IV-19b-e show the AFM images of the origami chains where each origami rectangle contains *parC* strands in the central half area. While the rectangle

used was a twist-corrected design (the same as that used in Chapter II), the introduction of *parC* strands seem to have induced some global twist or curvature into each rectangle and accumulated along the chains (noted by somewhat periodic kinks, folds or breaks in the chain structures, some of which are marked by arrows in Figure IV-19b-e), perhaps due to the repulsion between the densely packed, long *parC* strands. Figure IV-19f shows a fluorescence microscope image of the origami chain sample, whose field-of-view size is the same as the AFM image shown in Figure IV-19e. Although with lower resolution due to the diffraction limit, the chains in the fluorescence microscope image show similar structural features to those in the AFM images, e.g., kinks and rough length distribution.

#### IV.2.2.2.2. Can ParM filaments separate DNA origami? Estimation and comparison of force generated by ParM filament growth and force required to separate origami in a chain

Now we ask: can growing ParM filaments separate the DNA origami rectangles in chains at all? Are the forces or energies comparable? In this section, we estimate the force generated by the growth of ParM filaments and the force required to break stacking bonds and separate origami in a chain, and compare them. The ambiguity in determining the length scale of stacking bond breakage makes it difficult to estimate the relevant force from the bond energy, and just comparing the chemical energies involved in the two systems—ParM polymerization and stacking bonds—might give a useful insight. Nonetheless, here we make attempts to estimate the forces based on two different assumptions about the relevant distance, one from a simulation-based stacking energy potential curve and the other by simply taking the distance of separation created by a ParM polymer at each insertion step. In fact, the second assumption makes the comparison effectively the same as comparing at the energy level.

Also, a more careful analysis than a mere force comparison would require additional considerations about the lifetimes of the bonds, i.e., the off rate of ParM monomers associated in filaments and the off rate of stacking bonds. For example, if a ParM monomer at a filament end dissociates at a rate much faster than the dissociation of stacking bonds, the filaments in a bundle might not be able to collectively exert enough force to separate stacking bonds. Conversely, if the off rate of stacking bonds is reasonably higher than the ParM monomer dissociation rate, the growth of ParM filaments will be able to successfully break the bonds within the lifetime of the associated monomers. We do not have a good estimate for the off rates of stacking bonds. (Although we have an estimate for the bond energy and hence the equilibrium constant, and a rough sense of the initial

rate of bond formation—they are observed to have formed stacking bonds within the order of minutes—we do not have an estimate for the on/off rates of the bond at its equilibrium state.) Also, the off rate of a ParM monomer from a filament end while it is capped by a ParR/*parC* complex is not known, although the off rate of an ATP-bound ParM (the state of a ParM monomer at a capped end) at *free* filament ends can be estimated from the  $K_d$  and the on rate of ATP-ParM: the  $K_d$  can be treated the same as the critical concentration of the monomers for polymerization systems<sup>109,110</sup>, and hence, using the critical concentration, 0.6  $\mu\text{M}$ , and the on rate,  $\sim 5/\mu\text{M}/\text{s}$ , of ATP-ParM<sup>97</sup>, we can obtain the off rate of ATP-ParM at free ends to be  $\sim 3/\text{s}$ . However, once filament ends are stabilized by ParR/*parC* complexes, ParM filaments seem to grow continuously at least over the course of seconds to minutes, as shown in the continuous travel of ParR/*parC* complexes pushed by ParM polymerization from one pole to the other within a cell<sup>111</sup> and in the steady growth of filament bundles in the *in vitro* reconstitution assay<sup>105</sup>, unless they break and shrink from the middle by buckling or photodamage. Hence, it seems reasonable to treat ParM filaments as a quasi-infinitely growing thermal “ratchet”<sup>109,110,112</sup>, as long as ParM monomers in solution maintain the steady state concentration, and thus, we limit our estimation and comparison to the amounts of relevant forces in our system.

#### IV.2.2.2.2.1. Estimation of the force generated by ParM filament growth

The dynamics of a filament end, e.g. the on/off rates of a monomer, is likely to be affected by the presence of a ParR/*parC* cap, especially when the filament is exerting force<sup>109,110</sup> to push ‘loads’, such as plasmids or DNA origami structures. For example, the monomer addition is likely to be inhibited by the presence of the cap and the load, decreasing the on rate, and the reaction force against pushing would likely destabilize the monomer at the end, increasing the off rate. The change, especially the increase, in  $K_d$ , can be considered to correspond to an extra free energy cost that arises from the work that the filament growth has to do against its load<sup>110</sup>. Hence, the dissociation constant in the presence of force,  $K_d(F)$  can be expressed by the following equation<sup>110</sup>:

$$K_d(F) = K_d(0)e^{\Delta\Delta G/k_B T},$$

where  $K_d(0)$  is the native dissociation constant in the absence of force,  $\Delta\Delta G$  is the extra free energy cost in the presence of force,  $k_B$  is the Boltzmann constant, and  $T$  is the absolute temperature. Since the dissociation constant is the same as the critical concentration of monomers in polymerization systems, the  $K_d$  under force can be interpreted as the critical concentration *required* to balance the force (below the critical concentration, the filament would not grow). Hence, the *maximum* force,  $F_{max}$ , that a filament end can ‘withstand’, or by reaction, ‘exert’ against a load over a distance  $\delta$  at a monomer concentration  $M$  can be expressed by the following equation<sup>110</sup>:

$$M = C_{crit} e^{F_{max} \delta / k_B T},$$

where  $C_{crit}$  is the critical concentration of the monomer in the absence of force. Rearranging the equation gives:

$$F_{max} = \frac{k_B T}{\delta} \ln \left( \frac{M}{C_{crit}} \right).$$

The critical concentration of ATP-ParM monomers ( $C_{crit}$ ) at free filament ends in the absence of force is 0.6  $\mu$ M. The concentration of ParM monomers in solution is maintained at the steady state concentration, 2.3  $\mu$ M, by the dynamic instability of free filaments (after hydrolysis of ATP to ADP by ParM, monomers dissociate from filaments and get recycled). We can assume that the majority of the monomers at the steady state are in the ATP-bound form, given that under the high abundance of ATP in the system,  $\sim 10$  mM, ATP-ParM coupling is a much faster step ( $\sim 2000$ /s, based on  $k_{on} = 2.32 \times 10^5$  M/s) than the dissociation of ADP-ParM from a filament ( $\sim 64$ /s) and the dissociation (recycling) of ParM from ADP ( $\sim 0.56$ /s)—all numbers are from ref. 97. Thus, with  $M = 2.3$   $\mu$ M, and  $\delta = 2.36$  nm (the length increment corresponding to the addition of a single monomer to a filament, calculated based on a recently reported model<sup>98</sup> for ParM filaments obtained from a fit to cryo-EM data), the maximum force generated by the growth of a single filament at  $T = 295$  K (22  $^{\circ}$ C) can be obtained to be  $\sim 2.32$  pN.

#### IV.2.2.2.2. Estimation of the force required to break stacking bonds

Now, let us estimate the force required to break the stacking bonds between origami rectangles. The exact mechanism by which growing ParM filaments could take apart origami rectangles is hard to predict; perhaps stepwise sliding of stacking bonds caused by some shearing mechanism could require less force at each step, by dividing the area of contact, and thus the energy, than the case of breaking all the bonds at once. A recent molecular dynamics simulation study<sup>43</sup> suggests that an end-to-end stacked pair of DNA duplexes break by sliding of one terminal base pair relative to the other under an external force, even when the force was applied in the axial direction (involving tilting of the terminal base pairs). Thus, considering the case of directly breaking all the bonds at once along the direction of the helical axis without any sliding would give the upper-bound estimation for the force required.

In the system we tested, 24 blunt ends on each rectangle edge were used for the stacking bonds. In Chapter II, the free energy of the stacking bonds was measured based on the equilibrium between monomers and dimers of one-sided origami rectangles with different numbers of blunt ends, 4, 8, and 12 (each corresponding to and described as 2, 4, and 6 active patches, respectively). When no loop-loop interaction was assumed (where “loop” means scaffold loop in inactive patches,

i.e., locations where there were no blunt ends), the free energy values obtained showed non-linearity with respect to the number of blunt ends; with an increasing number of blunt ends, the absolute value of the free energy of stacking bonds *per helix* tended to decrease, potentially revealing the effect of the residual global twist along the edge. Following the trend, with 24 blunt ends, the free energy of stacking bonds per helix should be smaller (less negative) than the case with 12 blunt ends; i.e.,  $\Delta G$  per helix of stacking bonds for 24 blunt ends  $> -1.42$  kcal/mol, but we do not have a good way to estimate a more concrete value for the 24-blunt-end case under this energy model. In the second thermodynamic model from Chapter II, where loop-loop interaction was assumed to be non-neutral, the free energy of stacking bonds per helix was obtained consistently across the systems tested, as  $\Delta G = -1.12$  kcal/mol. Assuming this value can be applied to the 24-blunt-end case (which actually uses full stacking patches, with no loop-loop interactions), the free energy of stacking bonds for the 24-blunt-end bonds can be calculated to be  $\Delta G = -26.88$  kcal/mol. Now, based on this estimation for the bond energy and two different assumptions on the relevant distance, let us estimate the force to break this bond.

#### IV.2.2.2.2.1. Assumption on distance from energy potential curve

Estimating the force required to separate two chemical entities connected by the given free energy would require more knowledge on the energy potential; the depth of the energy well should correspond to the free energy of the bond, and the *width* of the energy well should determine how far one should take the entities apart for the two entities to be considered to be two separate ones. The distance should be the lateral distance between the bottom of the well (energy minimum) and the top of the well where it plateaus near 0 (theoretically it approaches 0 in infinity). There is no experimental measurement on such information, but the recent molecular dynamics simulation study<sup>43</sup> suggests that the energy minimum occurs near 3.5 Å, close to the distance between consecutive base pairs in a DNA helix, which agrees with a notion about the distance by Kool<sup>61</sup> and with our experimental data where stacking bonds show no gaps and appear as continuous helices under AFM, and that the energy plateau near 0 occurs at separation of ~10-11 Å. With this distance of ~7 Å, the force required to break the stacking bond of free energy -26.88 kcal/mol is estimated to be on the order of  $3 \times 10^2$  pN, which corresponds to a maximum force generated by a bundle of on the order of  $\sim 1 \times 10^2$  ParM filaments.

Each of our origami rectangles in the chain system was designed to contain 96 *parC* strands, and in the ideal case, at maximum each half of them could anchor 48 ParM filaments in either direction to push apart neighboring origami. Assuming the recent report<sup>98</sup> that each *parR/parC* complex stabilizes only one end of a ParM filament and the mechanism that enables bipolar plasmid



segregation is antiparallel bundling of filaments is correct, in the case that all of those 48 filaments form antiparallel bundling with another set of 48 filaments from a neighboring rectangle, a maximum of 96 ParM filaments might be able to exert a force comparable to the necessary force requiring a bundle of on the order of  $1 \times 10^2$  filaments.

This estimation might deviate from reality, perhaps due to simplifications made by the simulation. The simulation was carried out based on an approximation of the true (quantum mechanical) behavior of the system through classical equations of motions and pairwise potentials<sup>113</sup>. Based on their pulling ‘experiments’ using artificial harmonic springs, the free energy of stacking that they obtained for end-to-end stacking of ‘AT’ blunt-end pairs was as large as  $\sim -6.3$  kcal/mol, under 120 mM NaCl condition (in the absence of divalent cations). This value is an overestimate, compared to a previous experimental study<sup>44</sup> where the stacking energy between ‘AT’ blunt ends was measured to be between  $-0.19$  and  $-1.34$  kcal/mol, and also compared to our own measurement (Chapter II) where the free energy of stacking between ‘GC’ blunt ends was obtained to be between  $-1.12$  and  $\sim -2.63$  kcal/mol, which is known to be more stable than stacking between ‘AT’ blunt ends<sup>44</sup>. Moreover, the maximum force they measure during the process of separation was  $\sim 150$  pN (!) for a *single* pair of ‘AT’ blunt ends. (In fact, if a correct potential energy curve had been given, the maximum force required to break the bond would have to be obtained by differentiating the curve—the maximum value of the derivative. The estimation above is for the average force throughout the distance.)

#### IV.2.2.2.2.2. Assumption on distance from ParM monomer insertion step size

A somewhat more simplistic consideration for estimating the force required to separate adjacent origami rectangles would be to assume that the *effective width* of the energy well for the stacking interaction is comparable to the length increment generated by ParM monomer insertion in growing filaments, i.e.,  $\sim 2.36$  nm, and to assume (again) that the force required is averaged throughout the distance. The assumption of the effective width being the same as the distance created by ParM monomer addition seems reasonable, because (1) we are considering the case where the origami rectangles are separated by the growth of ParM filaments, hence by ‘discrete’ separation, (2) given the persistence length of ParM filaments<sup>104</sup>,  $\sim 10$   $\mu$ m, and the distance between the *parC* sites on adjacent origami,  $\sim 50$  nm, it is unlikely that the filaments would bend in the length scale relevant to the initial separation of the rectangles, and (3) when multiple ParM filaments are pointing towards the same direction, connecting two adjacent origami, they would likely bundle, so it is also less likely that some subpopulation of the filaments would be in a far too off-axis angle, effectively reducing the distance of movement along the chain axis.

With the same separation of  $\sim 2.36$  nm, between adjacent origami rectangles, created by an insertion of ParM monomer to a growing filament, now the comparison is effectively between the free energy of the stacking bond and the work done by the growing filament over that distance. Nonetheless, if we calculate the force corresponding to the free energy of the stacking bond with 24 helices,  $-26.88$  kcal/mol, and the distance,  $2.36$  nm, the force of  $\sim 79$  pN is obtained. Using the force value  $2.32$  pN, obtained above for ParM filament growth, the number of filaments in a bundle required to generate this force is estimated to be  $\sim 34$ . Given the number of *parC* sites per origami, 96, with each half of them (48) pointing to opposite directions in the ideal case, and again assuming the case of ParM filaments forming antiparallel bundles, the maximum number of filaments between any adjacent pair of origami would be  $\sim 96$ . Hence, the number obtained,  $\sim 34$  filaments per bundle to separate adjacent origami, seems plausible.

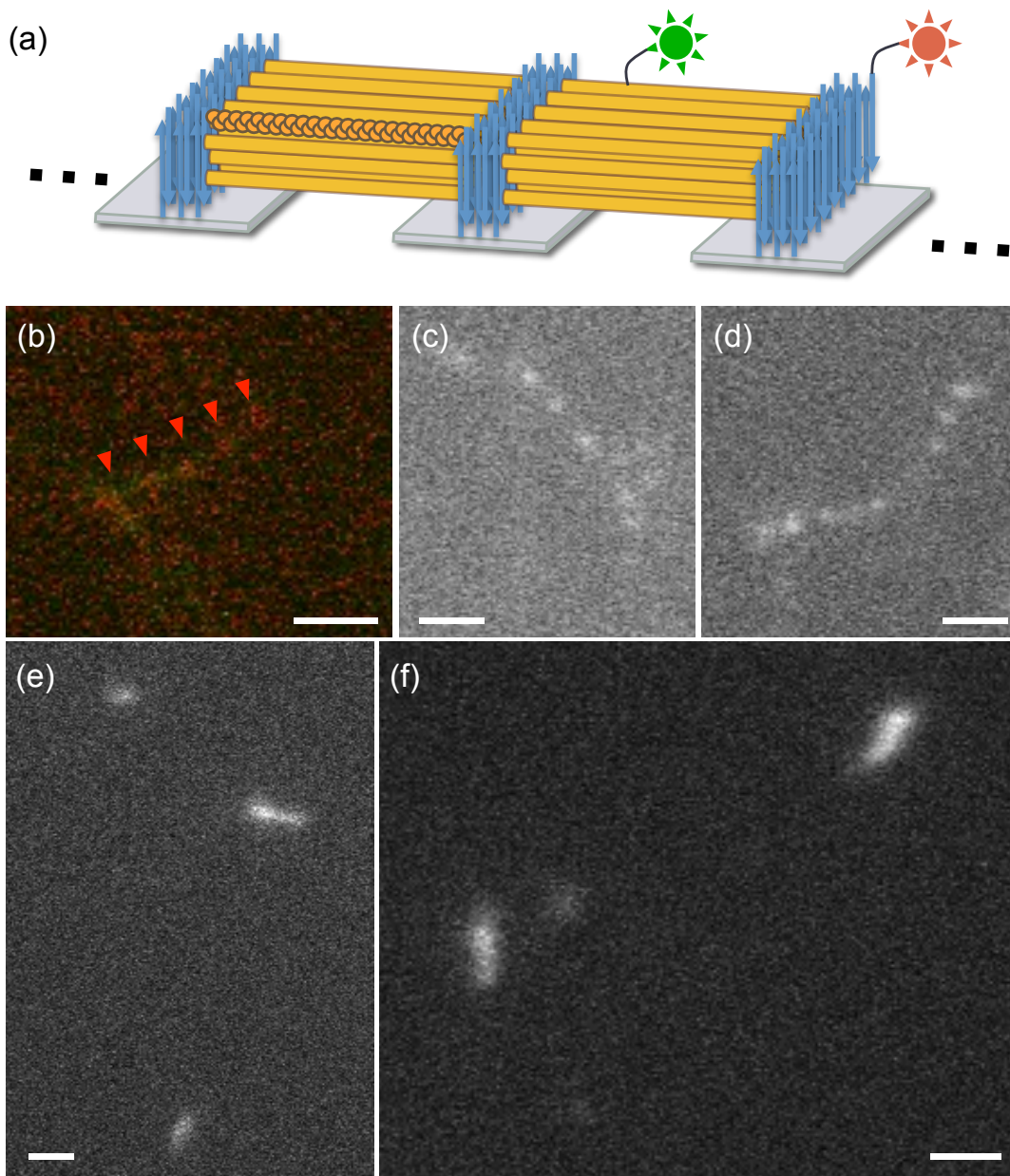
However, this estimation contains a few caveats. First, the free energy of the stacking interaction was estimated based on measurements carried out under high  $\text{Mg}^{2+}$  condition ( $12.5$  mM), whereas the actual condition for the ParMRC assay contains less divalent cation ( $\sim 1$  mM  $\text{Mg}^{2+}$ ). This would actually overestimate the force required to separate origami, given that the stacking bonds may get weaker under the low divalent cation condition, in a similar manner to the case of base pairs, although the strength of the bonds would also depend on the concentration of monovalent cations (our system has  $100$  mM  $\text{K}^+$ ). Second, as mentioned at the beginning of this section, this estimation is based on the assumption of the case where all the stacking bonds are breaking at once along the axial direction. In reality, the actual mechanism might be different, potentially involving some stepwise shearing of the bonds.

As a future work, by systematically controlling the number of *parC* sites on each origami and the strength of the stacking bond by changing the number of active patches, we would be able to carry out “force calibration” experiments, and get a better understanding on the balance of the forces.

#### IV.2.2.2.3. Expanding structures

Although the results are still somewhat preliminary, we were able to implement expandable nanostructures using the rectangle origami chains. When we mixed the origami chains with ParR and ParM, we could observe a unique behavior of the ParM filaments. The filament bundles in general seemed to behave normally, i.e., grow and expand, but a new interesting feature that showed up was linear arrays of dots (in the Cy3 channel) decorating the positions of bundles (shown in the Alexa488 channel) in a manner somewhat like “threaded beads”, each presumably

indicating the locations of Cy3-labeled origami rectangles (Figure IV-20b-d).



**Figure IV-20. DNA origami chains broken and expanded by the growth of ParM filaments.** (a) Model and (b-d) fluorescence microscope images of DNA origami chains putatively broken and expanded by the growth of ParM filaments. In (a), ParM filaments are depicted as linear rods with one of them shown as non-helical filamentous assembly of monomers, both for simplicity. Subpopulations of ParM monomers were fluorescently labeled with Alexa488 (green) and each *parC* strand on origami was labeled with Cy3 (red). The image in (b) is a composite image of the two channels (red and green), which reveals both the origami and ParM filament bundles between them. (e,f) Intact DNA origami chains in the absence of ParM and ParR proteins. All microscope images except (b) are from the Cy3 channel. Scale bars: 2  $\mu\text{m}$ .

To rule out the possibility that the low divalent cation condition in the ParM polymerization buffer might have weakened and broken the stacking bonds between origami, even in the absence of the growing filaments, we examined the sample of origami chains in the ParM polymerization buffer (Figure IV-20e,f). While the lengths in general seem to have shrunk to some extent, compared to the chains in the TAE/Mg<sup>2+</sup> buffer (Figure IV-19), one can clearly see that those chains still retain the chain form (note that the lower density than in Figure IV-19 is due to dilution). One should also note that some of the long and large features observed in the TAE/Mg<sup>2+</sup> buffer (Figure IV-19) are due to lateral aggregation caused by the relatively high divalent cation condition—often witnessed in AFM images with molecular detail and in fluorescence microscopy images, as well, by nonuniform fluorescence intensities, which would have separated into smaller individual chains under the ParM polymerization buffer condition.

Does the length scale of the expanded origami chains make sense? For some representative microscope images that show the origami (bright red dots) relatively clearly and discretely (as in Figure IV-20b-d), we measured the average distance between the origami, which was obtained to be  $\sim 1.08$   $\mu\text{m}$  ( $N=13$ ). Under the condition used, where the initial ParM monomer concentration was 7  $\mu\text{M}$ , and the concentration of individual origami rectangle (as if they were separate) was  $\sim 800$  pM (the concentration of origami chains was much lower), each containing 96 *parC* sites, we can estimate how many ParM monomers, and hence how many filaments, are associated in each gap between adjacent origami rectangles. However, it is not easy to determine at what stage the filaments are in these images—still in growing phase or in saturation phase where they no longer grow; technically it is not easy to measure the distance change between adjacent origami over time for the same origami rectangles because under the particular setup we had the Cy3 fluorophores on origami required long ( $\sim 2$ -5 s) exposure time to be properly observed with comparable signal to the Alexa488 (ParM) channel. Nonetheless, assuming that the filaments reached the saturation phase and that all the monomers above the ATP-ParM critical concentration (0.6  $\mu\text{M}$ ) were consumed and associated within the gap between origami, we would be able to get an upper-bound estimation for the number of ParM monomers and the number of filaments in each bundle. Based on these assumptions, with  $\sim 6.4$  ( $=7-0.6$ )  $\mu\text{M}$  of ParM ( $\sim 5.8 \times 10^{12}$  ParM monomers), we estimate the number of filaments in each gap between adjacent origami to be  $\sim 17$ .

Comparing the number of filaments required to break stacking bonds between origami,  $\sim 1 \times 10^2$  or 34, depending on the assumptions, estimated in the previous section, the number obtained above, 17 filaments, seems much fewer than that required. Perhaps this indicates that the stacking bonds between origami may indeed get weaker in the low divalent cation condition. Or it

may indicate that the breaking does involve some mechanism that requires much lower force, e.g., by shearing.

For more convincing evidence that the growing ParM filament bundles are pushing origami apart, we plan to devise a fluid chamber system and proper surface treatment protocol such that origami chains can be loosely pre-fixed on a substrate, and other components are introduced at a later point under the microscope, to allow capturing the expansion from the very early moments. Alternatively, we could use photo-activatable ATP, in place of ATP, with all other conditions the same, then activate ATP to start the reaction. We also plan to use other origami systems that can show more dramatic structural change, e.g., triangles created by three corner origami structures connected together via stacking bonds, to further demonstrate dynamically expandable structures.

#### IV.2.2.3. Control system: non-expanding DNA nanotubes

What would happen to a nanostructure created with DNA that has similar structural features to origami chains but is not easily breakable, when mixed with the ParMRC system? At earlier stages of investigation, we tested a sample of DNA nanotubes that are functionalized with *parC* strands around their peripheries in the presence of ParM and ParR, to see if we can find any interesting behavior potentially controllable by the programmability of DNA. But in retrospect, in comparison to the cases with the origami chains, we believe that the nanotube cases can serve as a good control system, because the nanotubes have similar structural features to the origami chains, but have slightly lower density of *parC* strands (~30-70% of that of origami chains, per unit length), while having stronger bonds connecting the subunits than the stacking bonds in the origami chains case. Each tile has four sticky ends of 5 nucleotides length. Along the circumference, where there are typically 4-10 tiles<sup>114</sup>, there are 8-20 sticky-end-based bonds, with a total of 40-100 base pairs. The free energy of the bonds can be estimated to be  $\sim -141$  to  $-56.4$  kcal/mol (using an average base pairing energy value based on the nearest neighbor model<sup>65</sup>  $-1.41$  kcal/mol, neglecting the detail sequences), which compares to  $\sim 210$ - $525\%$  of the stacking bonds that connect origami rectangles together in the chains. The best control system to the expanding origami chains with reconfigurable stacking bonds would be the same origami chain system with sticky-end-based bonds in the place of stacking bonds, but we believe that the DNA nanotube system provides an insight as a control system with reasonably hard-to-break bonds.

Figure IV-21 shows the structure of the nanotube design that we used. It was adopted from an earlier work by Rothmund et al.<sup>114</sup>, and the center strand (strand numbered '3' in Figure IV-21a) was modified to anchor a *parC* strand (the same 5'-end extension made to the staple strands in

origami designs) and also the 5'-end of the extension was modified with Cy3 for fluorescence. Figure IV-21b shows an AFM image of the nanotubes decorated with *parC* strands. (Note the hairy features around the nanotubes.) Figure IV-21c shows a fluorescence microscope image of the same nanotube solution, with the AFM image shown in Figure IV-21b as an inset whose size is adjusted to match the scale of the fluorescence microscope image. One can note that the lengths of the nanotubes in the two images are comparable, and also that the lengths of the nanotubes are quite comparable to those of the origami chains shown in Figure IV-20 in the previous section.

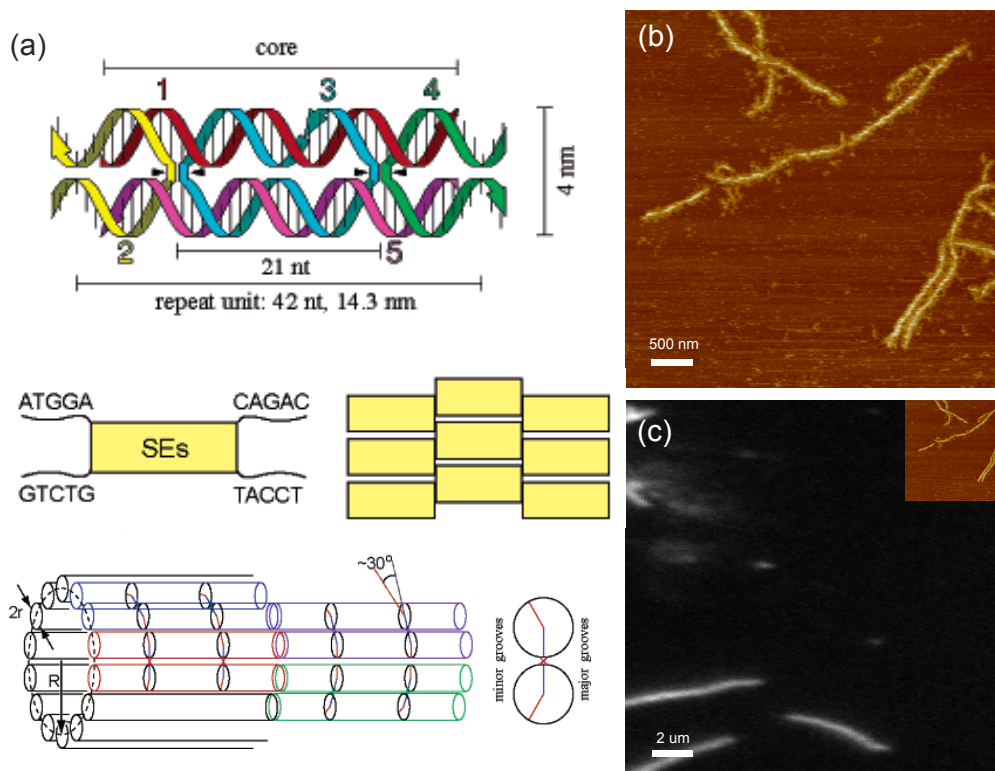


Figure IV-21. **Design and microscope images of DNA nanotubes.** (a) Schematic diagram of the DNA nanotubes used for our experiments. The strand numbered “3” was extended with a  $T_{20}$  tail at the 5'-end to allow binding of *parC* strands. The sticky ends allow the single tile to extend to a lattice that can curve and form a tube. Figures adapted from ref. 114. (b) AFM image of the DNA nanotubes with *parC* strands. Note the hairy features around the tubes. (c) Fluorescence microscope image of the DNA nanotubes with *parC* strands. Both the *parC* strands and the strand “3” contained Cy3 dyes.

How would ParM filaments behave in the presence of these nanotubes? The expectation was three-fold: (1) ParM filaments grown in solution may find and get stabilized by ParR/*parC* complexes along a nanotube and form “long green (Alexa488) hairs” around the red (Cy3) nanotube, (2) those filaments that have grown may collapse the nanotube into a ball by lateral bundling between the filaments, or (3) before anything like those mentioned in (1) or (2) happens,

ParM filaments may bundle with the nanotubes and create some hybrid structures. We tested two cases: when the two structures (ParM filaments and DNA nanotubes) were pre-formed and mixed, and when the ParM was allowed to grow from the beginning in the presence of the nanotubes.

#### IV.2.2.3.1. Bundling of pre-formed filaments and nanotubes

ParM filaments pre-formed and stabilized by AMPPNP (non-hydrolyzable ATP analog) was mixed with DNA nanotubes with *parC*, in the absence of ParR. As shown in Figure IV-22, the filaments and nanotubes bundled together (there was also a crowding agent, methylcellulose, present at 0.4 %). Because there was no ParR, no specific interaction was expected, and any interaction would be based mostly on the structural relations. This observation led us to believe that they may indeed bundle together, even under the dynamic condition.

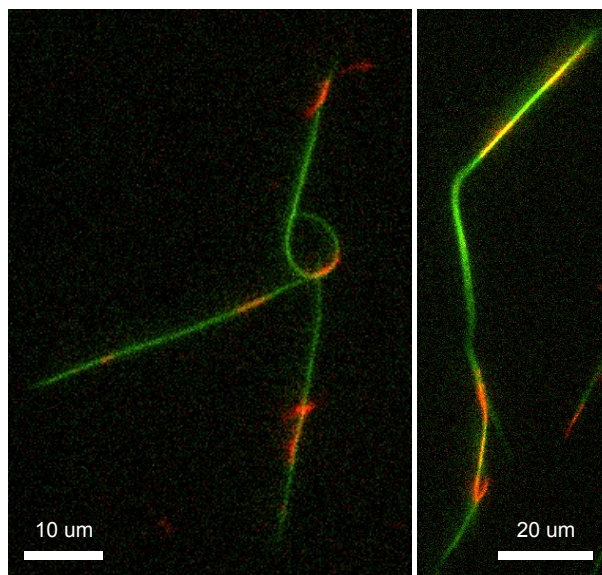


Figure IV-22. **Fluorescence microscope images of static ParM filaments mixed with DNA nanotubes.** The composite images of the two channels—Cy3 (red) for DNA nanotubes and Alx488 (green) for AMPPNP-stabilized ParM filaments—show that the two species bundle together. The absence of free monomers and the static nature of the objects allow lower background and longer exposure time, and hence, clearer images.

#### IV.2.2.3.2. Static hybrid structure formation of filaments and nanotubes

When ParM was allowed to grow in the presence of the nanotubes, filaments seemed to form complicated hybrid structure with the nanotubes and, interestingly, became completely static. While in the origami chain case, the filaments appeared dynamic, with some of them clearly expanding especially at the early stage, the hybrid structure formed between the filaments and nanotubes looked like rigid strings and did not change the conformation at all (Figure IV-23). It seems

something close to the scenario (3) suggested above is happening.

Although only speculative at this stage, there might be multiple mechanisms involved. Some short filament bundles may get captured at both ends by ParR/*parC* complexes within a single nanotube, and stop to grow any further after reaching the length allowed by the distance along the nanotube. Some filaments may extend out of a nanotube with only one end anchored, but as soon as they reach some critical length, they may bundle with the nanotube, as bundling seemed quite favorable between the two objects as shown in the previous section. Even if some ParM filaments manage to grow longer, the free sidewalls may get bound and coated by other nanotubes. All of these mechanisms seem to “lock” the filaments by extensive stabilization by bundling with nanotubes and prevent them from growing any further, yielding the giant, static hybrid structures.

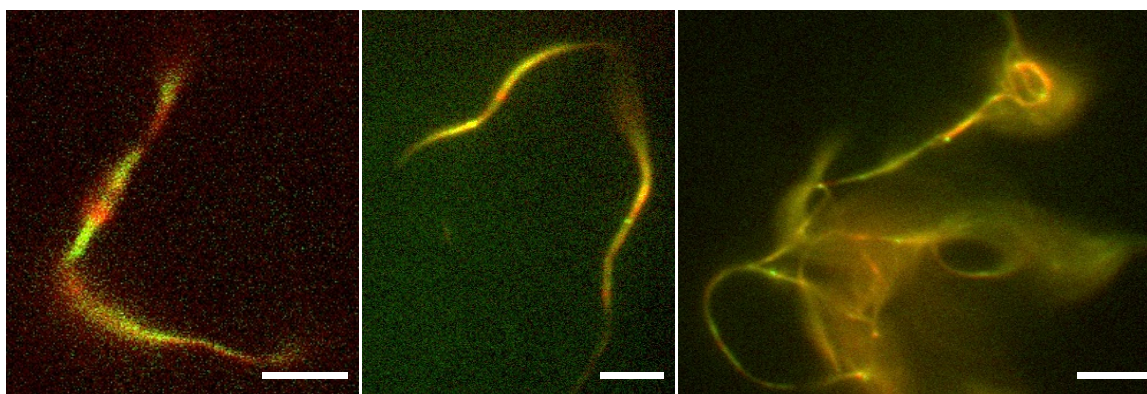


Figure IV-23. **Fluorescence microscope images of ParM filaments grown in the presence of DNA nanotubes containing *parC* strands.** When ParM was subject to the dynamic environment (with ATP and ParR), mixed with nanotubes, filaments formed hybrid structures with the nanotubes, and became static. Scale bars: 5  $\mu\text{m}$ .

### IV.3. Conclusion and Future Directions

We used DNA origami nanostructures as templates in attempt to study some structural aspects of the ParMRC system. We were able to reveal that ParR at high concentration binds to non-recognition sequence DNA, in addition to the *parC* sequences, both through AFM studies and gel electrophoresis analysis. In the right concentration regime, we observed that ParR recognizes *parC* and discriminates non-recognition sequences. The specific ParR-*parC* interaction observed in our system was consistent with previous studies in literature in terms of both stoichiometry<sup>99</sup> and structure<sup>103</sup>.

Then we used the dynamic filament growing mechanism of the ParMRC system to engineer expandable nanostructures. We first placed *parC* sequences on origami structures and successfully



reconstituted the ParMRC system: using fluorescence microscopy, we observed colocalization of origami and the ends of ParM filaments, and dynamically growing filaments whose ends are stabilized by ParR/*parC* complexes anchored on origami. By using chains of rectangular origami connected via stacking bonds, we demonstrated that bundles of growing ParM filaments could break stacking bonds and dynamically expand the structure.

This work is in progress, and we discuss potential future directions in subsections below.

### IV.3.1. Expandable triangles

As briefly discussed earlier in the discussion, we are planning to test expandable triangle structures, each made with three corner shape origami, connected via stacking bonds, to amplify the dramatic change in the overall structure; from a small single bright dot (due to three corners being unresolvable by light) to a large, growing triangle with red dots at each corner. Figure IV-24a shows the schematic of the triangle design. Within the shaded area, there are ~100 staple strands, each of which we can extend to anchor a *parC* strand. The number of *parC* strands per origami (a single corner) is comparable to the number we had per origami in the origami chain case, so this triangle design would be appropriate for proof-of-concept experiments and for comparison with the chain case.

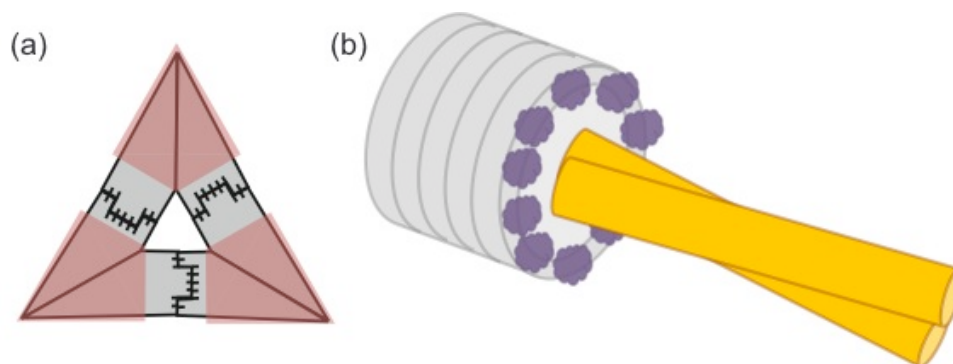


Figure IV-24. **Schematic diagrams** of (a) a three-corner triangle as a substrate for expandable nanostructure with the shaded area indicating the anchor sites for *parC* strands and (b) a cup-shaped origami design that might allow the creation of an artificial (*parC*-free or ParR/*parC*-free) ParMRC system.

Also, as mentioned earlier, to be able to capture the early moments in the dynamic system (ideally from  $t=0$ ), construction of a fluid chamber system that enables mixing of the components while under the microscope would be helpful. However, careful control over the concentration of the components would be a challenge, and proper surface treatment protocol that can hold origami *loosely* in place on the substrate while preventing binding of protein components should be

developed. Or, introduction of photo-activatable ATP might be a useful alternative. With photo-activatable ATP, all other the components can be added into the closed system without any change from the systems we have tested, and with an external stimulus in the form of light irradiation of specific wavelength, we may be able to watch the reaction from  $t=0$ .

### IV.3.2. Effects of distance between ParR/*parC* complexes on ParM filament connection

Besides engineering expandable nanostructures, we would still pursue the direction of studying biophysics of the ParMRC system, using DNA origami as a custom tool. As it has been the case that it is hard to study the system at the single *parC* regime (mostly because of technical reasons, e.g., difficulty of taking AFM of filaments, whether dynamic or not, and too low signal-to-background ratio of a single filament under fluorescence microscopes), we seek other questions that we could answer about the biophysical behavior of the ParMRC system using our DNA origami designs.

One question is: how would the distance between ParRC complexes affect the efficiency of ParM filament connection between the two positions? In a living bacterial cell, the actual distance relevant to the search process of ParM filaments for two ParRC complexes seems to be on the order of about hundred nanometers; within a cell whose size is  $\sim 1$   $\mu\text{m}$  in diameter and  $\sim 2-3$   $\mu\text{m}$  in length, the plasmids do not diffuse much and show a confinement radius of  $\sim 0.28$   $\mu\text{m}$ , in the absence of (i.e., before getting pushed by) any ParM filaments<sup>111</sup>. Examining how the efficiency of the search process changes depending on controlled distances between ParRC complexes would reveal interesting insights and expand our understanding on the dynamics of the system.

We can precisely control the positions of *parC* sites on a single origami, and hence the distance between *parC* sites on neighboring origami structures in rectangle chains or the corner-origami triangles. For longer range distances, say, larger than  $\sim 100$  nm, the length of a single origami rectangle, we can program the stacking bonds based on our orthogonal bond sets to create origami multimers with defined arrangement, and put *parC* sites at distances larger than a single origami. We also plan to examine much larger distance ranges, e.g., several microns to a few dozens of microns, using our origami surface placement technology<sup>80</sup>, varying the distance between the binding patches on patterned substrates. With surface-immobilized origami, we could use the total internal reflection microscopy (TIRF) technique, which might allow us to resolve a small number of filaments, perhaps down to the single-filament regime.

### IV.3.3. Effects of the number of *parC* sites on single origami structures on the growth rate of ParM filament bundles

Another question that we are interested in is how the number of *parC* sites on a single grouped structure (in our case, origami) would affect the growth rate of the ParM filament bundles. In living cells, it is proposed that multiple plasmids often form clusters and segregate together as a single unit<sup>104,111</sup>, so it would be interesting to examine how the behaviors change depending on the number of *parC* sites grouped in a single cluster.

Again, we can precisely control the number of *parC* sites on a single cluster using our origami structures. We can use discrete origami structures, e.g., single-unit triangles, instead of chains connected by stacking bonds, and observe the dynamic expansion driven by growing filament bundles. By taking time lapse images, and measuring the rate of change in length of filament bundles between origami foci, we could potentially draw some interesting relationships between the number of *parC* sites and the growth rate.

### IV.3.4. Cup experiments

#### IV.3.4.1. Is it just the geometry that really matters?

As discussed earlier, we observed strong binding of ParR to non-recognition sequences of dsDNA on origami under high ParR concentrations. That observation led us to ask the question whether a ParR loop created by artificial shapes with non-recognition sequences would stabilize ParM filaments. This is a question of whether it is just the geometry that really matters, or whether the *parC* sequence has any important role other than facilitating the cooperative binding of ParR into a complex.

For such a test, we can design a three-dimensional origami in the shape of a cup or cylinder, where helices are curved and aligned perpendicular to the cylindrical axis (Figure IV-24b). At high concentrations of ParR, ParR will bind to the helices that are exposed and relatively free, as found by AFM studies discussed earlier. Then we can detect and measure the association of ParM filaments with these origami-mediated ParR loops. If these “artificial” ParR caps do the job of the real ParR/*parC* complexes, even in the absence of *parC*, we would be able to see continuous growth of the ParM filaments, with their ends in association with the ParR-origami complex.

Since we do not know the exact shape of the ParR/*parC* complexes and the binding mode of

ParM filaments—ring vs. helix shape, parallel vs. perpendicular binding, wrap-around vs. open-clamp binding, etc.—the geometry of the origami would have to be carefully designed to accommodate various possibilities. For example, the cylinder shape could be designed to be reconfigured by changing the subset of the staple strands to allow a helical contour of the DNA helices. Or we could leave out some staple strands at part of the loop to allow potential open-clamp type binding.

#### IV.3.4.2. Would the active site in ParR alone stabilize filaments?

In fact, the active site in ParR that is responsible for stabilizing ParM filaments was identified to be the C-terminal domain (~17-33 residues) of the protein<sup>98,100</sup>, and it was hinted that the active site alone could stabilize ParM filaments; the presence of synthetic peptide with the C-terminal amino acid sequence in high excess (~100× of ParM) interfered with the ParM filament binding to regular ParR/*parC* complexes<sup>100</sup>.

This allows us to pose a question: could we construct a system just using the C-terminal domain that stabilizes ParM filaments? Beyond eliminating *parC*, could we also remove ParR, except the C-terminal domain, from the ParMRC system? The cup design may be useful for such a test as well. We can synthesize a hybrid of a DNA strand and the C-terminal domain, and place them on the inside of the cup. By placing a large number of them, we can create a high local concentration of the domain inside the cup, and the active domain peptides may successfully stabilize ParM filaments without the need of the well-defined ParR loop structure, much less the *parC* strand. If this approach is realized, based on a synthetic peptide, and a designed DNA origami structure, we will be the first to have created a *nearly* totally artificial ParR/*parC* complex—just except for the scaffold strand in origami.

## IV.4. Materials and Methods

### IV.4.1. Preparation of DNA origami

#### IV.4.1.1. DNA origami with a single *parC*

The DNA origami structure that contains a single *parC* sequence was created by using an M13-variant scaffold strand that was genetically engineered (by Genestitute) to include the 169-

nucleotide *parC* sequence in it. The complementary sequence to the *parC* sequence was designed and ordered (from IDT, PAGE purified) as two separate strands (85 bases and 84 bases, respectively) to avoid synthesis of long DNA, which often suffers from low yield. The complementary strands would leave a nick on the double-stranded *parC* strand where the promoter part is located between the two sets of five iterons (to which ParR proteins bind), so the introduction of a nick is not expected to cause any inhibition to the desired binding of ParR to the binding sites on *parC*.

#### IV.4.1.2. DNA origami with multiple *parC* strands

We used two DNA origami designs—triangle and rectangle—that contain multiple *parC* strands in each origami structure. For incorporation of multiple *parC* strands, pre-existing DNA origami designs were modified such that each staple strand has a 5'-end extension of 20 T's, to which *parC* sequences were designed to link (with an extension of 20 A's). *parC* sequence design is shown in Figure IV-5c. The top strand of the double-stranded *parC* was extended at the 5' end with an overhang of 20 A's that can bind to the 20-T anchors on origami. In addition to the 169-base sequence that is complementary to the corresponding sequence in the top strand, the bottom strand had (AAG)<sub>5</sub>-biotin extension at its 5'-end, to which we can incorporate either (CTT)<sub>5</sub>-Cy3 DNA strand or Cy3-streptavidin for fluorescent labeling.

For triangle origami, 177 staple strands out of the total 228 were replaced with those containing the 20-T tail. The staples that did not contain the 20-T tail were edge staples, bridge staples (left to not include the 20-T tail to prevent potential disruption to the structure) and dumbbell hairpin staples that were embedded for labeling (to tell which side is up) under AFM. In other words, almost all available staple strands, with minimal exceptions for structural and functional purposes, were replaced with those with 20-T overhangs for maximum number of *parC* strand incorporation.

For rectangle origami, all staple strands except those at the edges (192 out of 216) were designed to have their 20-T tail counterparts, and different subsets were used for different sets of experiments. For rectangle chains, central half staples (96) were replaced with those with the 20-T tail to prevent long *parC* strands from interfering with the stacking bonds between origami via their edges.

For sample preparation, staple mixtures for corresponding structures (8× excess of the scaffold concentration), *parC*-top strand (1.2× of the staples), *parC*-bottom strand (1.2× of the *parC*-top strand), and (CTT)<sub>5</sub>-Cy3 DNA strand (1.2× of the *parC*-bottom strand, when fluorescence

labeling was needed) were mixed together with M13 scaffold (from either New England Biolabs or Bayou Biolabs; both show similar performance) in  $1\times$  TAE/Mg<sup>2+</sup> solution. Since a purification step was essential to remove extra *parC* strands in solution (procedure described in the next section), as high concentration of M13 as possible was used to yield a high origami concentration, and the concentration of staple strands and other strands were calculated accordingly. The mixtures were annealed from 90 °C to 50 °C at a rate of  $-1^\circ/\text{min}$ , held at 50 °C for 3 hr, and slowly cooled from 50 °C to 20 °C at a rate of  $-10^\circ/\text{hr}$ , and then were subsequently treated for purification.

#### IV.4.1.2.1. DNA origami purification

DNA origami containing multiple *parC* strands were purified away from extra strands, mainly to remove the extra *parC* strands, because, if not properly removed, they would interact with ParR proteins and form independent ParR/*parC* complexes in solution stabilizing ParM filaments, apart from DNA origami.

For the purification method, gel purification was used, where DNA origami samples were run through agarose gel electrophoresis and then extracted from gel by the pellet-pestle homogenization method using the “Freeze N’ Squeeze” kit (Biorad). Spin filtration, a simpler purification method, could not be applied to this case because *parC* double strands (top + bottom + staples + Cy3-DNA) were too big (~130 kD) to be filtered out by common spin filters (100 kD cut off). Another method, dialysis, with a 300k cutoff membrane was tried, but it turned out to leave *parC* strands in significant amount when checked by re-running through gel electrophoresis.

For gel purification, DNA origami samples were run through 1% agarose gel, which had been cast in  $1\times$  TAE/Mg<sup>2+</sup> buffer, in the same buffer for ~2 hr under 100 V, with the gel box immersed in an ice bath. To keep the origami bands away from DNA stain chemical (SYBR Gold), reference lanes with a lower origami load were used next to the purification lanes; reference lanes were cut, stained and used as guides for cutting out bands of interest under UV irradiation. Cut bands were put into individual Freeze N’ Squeeze tubes, and squashed (one can use special pestles; I used 1 ul pipette tips). Those tubes were placed in a freezer for 5 min and then spun down at 4 °C with centrifugal force of 16000 g for 10 min.

#### IV.4.2. Preparation of microbeads with *parC*

For experiments with micro-sized beads, we used Streptavidin-coated fluorescent (Nile Red) beads with a mean diameter of 0.59  $\mu\text{m}$  (SPHERO from Spherotech, CAT No. SVFP-0556-5).

Coupling of the microbeads and the *parC* strands were performed via streptavidin-biotin interaction. Each *parC* strand was designed to contain a biotin at the 5'-end of the "bottom" strand. The bead solution and the *parC* solution were mixed at the ratio of 1 mg of the beads to 1.40 nmole of *parC*, following the manufacturer's guideline. The mixed solution was vortexed and incubated at room temperature for ~12 hr, with three more intermittent mixes (vortexing and spinning down). The solution was then subjected to a dialysis assay with a 300k cutoff membrane in a two-liter reservoir of the 1× TAE/Mg<sup>2+</sup> buffer containing a stirring bar for ~12 hr, to remove extra *parC* strands if any. The final concentrations estimated of the beads and *parC* were ~1.56 pM and 485 nM, respectively.

#### IV.4.3. Preparation of proteins

ParM (wild type and Alexa488-labeled) and ParR proteins were purified and supplied by Chris Rivera in Dyche Mullins' group at UCSF. Wild type ParM and ParR solutions come frozen in ParM polymerization buffer containing 20 % (v/v) glycerol. All proteins were stored in a -80 °C freezer before use. For wild type ParM, the ParM solution was thawed and buffer-exchanged into ParM polymerization buffer to remove glycerol. G-25 size-exclusion columns (GE illustra MicroSpin G-25 Columns, CAT No. 27-5325-01) were used for buffer exchange. 150 ul of ParM polymerization buffer was first run through a column to pre-treat the resin mixture for 1 min at 750 g at room temperature, then 150 ul of ParM solution was run through the column for 2 min at 750 g at room temperature. ParM protein was then quantified by UV spectrophotometer at wavelength 280 nm, with extinction coefficient 34039. Alexa488-labeled ParM came in ParM polymerization buffer containing 0.2 M sucrose, with the ParM concentration quantified during the step of coupling with the fluorescent dye (by Mullins' group). Alexa488 ParM was used directly from the sucrose-containing solution, as mixture with wild type ParM to yield 1 % to 10 % Alexa488-labeled ParM.

#### IV.4.4. Preparation of ParM polymerization buffer

ParM polymerization buffer contained the following compounds: 100 mM KCl, 30 mM Tris-HCl (pH 7.5), 1 mM MgCl<sub>2</sub>, and 1 mM Tris(2-carboxethyl)phosphine (TCEP). 10× stock of KCl, Tris-HCl and MgCl<sub>2</sub> was made and stored in a refrigerator. TCEP stock solution was made at 100× concentration and divided into small aliquots and stored frozen at -80°C. The 10× stock solution was diluted to 1× and mixed with TCEP freshly before use.

#### IV.4.5. AFM of *parC*-origami, ParR, and ParM

AFM for DNA origami with *parC* strands (both single and multiple) was taken under the same condition as for usual DNA origami (described in previous chapters). For AFM experiments where the same spot was followed after each step, the buffer condition was changed accordingly. When exchanging the buffer (e.g., from TAE/Mg<sup>2+</sup> to ParM polymerization buffer), the existing buffer was taken out as much as possible by using a pipette with the smallest (P2) pipette tip (usually leaving ~5 ul of buffer), then the sample was rinsed with the new buffer 2-3 times (adding ~20 ul and removing the same amount), and finally ~20 ul of the new buffer was added. When adding ParR protein in later steps, the number of molecules (assuming all the DNA origami molecules were bound to the mica surface at the initial deposition) were used as the measure for stoichiometry.

High-speed AFM movies were taken during demo sessions lead by Research Institute of Biomolecule Metrology Co., Ltd. (RIBM) of Japan during the Foundations of Nanosciences (FNANO) conference in 2012. Because of the unique design of their AFM, where the sample stage with a tiny (1.5 mm diameter) mica piece is placed upside down on top of a (relatively) large buffer bath of ~100 ul, a small amount of DNA origami sample (2 ul) was deposited, and ParR protein sample (2 – 8 ul) was added *in situ* into the buffer bath while scanning. Because of the large buffer volume, it took ParR some time (on the order of minutes) to diffuse and bind to the *parC* sites on origami, which was another factor that enabled AFM movies (other than the high speed scanning). AFM movies were obtained at a rate of 1 frame / 4 seconds.

#### IV.4.6. Gel electrophoresis of *parC*-origami with ParR

Agarose gels were prepared for analysis of *parC*-origami and its interaction with ParR proteins. The gel running condition was the same as described above for origami purification, except for the step of gel extraction (not needed). See above.

#### IV.4.7. Preparation of glass slides

Glass slides and coverslips for fluorescence microscopy were silanized with diethyl-dichloro-silane to make the surface hydrophobic. First, glass slides and coverslips were rinsed thoroughly under flowing water, then sonicated in a 0.1 M KOH solution for 20 min, then rinsed under water,



and then sonicated in a 95% EtOH solution for 20 min. Silanization was then performed by sonicating the slides and coverslips in isopropanol containing 5% (v/v) water and 2% (v/v) diethyl-dichloro-silane, for 1 hr. The slides and coverslips were then rinsed in isopropanol containing 5% water by sonicating for 20 min, and then dried in an oven (65 °C) overnight. After drying, the slides and coverslips were sonicated in a 95% EtOH solution for 20 min, and then finally rinsed in double-distilled water (MilliQ) by immersing them one by one into a big reservoir of water (~500 ml). Slowing pulling the slides and coverslips out of the water reservoir can show how well the silanization was done (water should completely repel the surface and the slides and coverslips should come out dry). The slides and coverslips were dried again in an oven overnight, and then they were sealed and stored.

#### IV.4.8. Sample preparation for fluorescence microscopy

For final mixtures of components for microscopy studies, ParM and *parC* solutions were added to the desired concentrations as described in each experiment. ParR was added to be 200-300 nM (usually ~20× to 100× of *parC* concentration). In addition, methyl cellulose was added at 0.4 – 0.5% (v/v), and BSA was added at 0.25 – 0.5 % (v/v). All components except ATP solution were mixed and deposited onto a silanized glass slide, to which ATP solution was added to be at the final concentration of 10 mM. A coverslip was placed as soon as adding ATP, and then sealed with a mixture of Vaseline and paraffin (~1:1).

## References

- 1 Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006).
- 2 Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368-372 (2011).
- 3 Yin, P., Choi, H. M. T., Calvert, C. R. & Pierce, N. A. Programming biomolecular self-assembly pathways. *Nature* **451**, 318-322 (2008).
- 4 Lund, K., Manzo, A. J., Dabby, N., Michelotti, N., Johnson-Buck, A., Nangreave, J., Taylor, S., Pei, R., Stojanovic, M. N., Walter, N. G., Winfree, E. & Yan, H. Molecular robots guided by prescriptive landscapes. *Nature* **465**, 206-210 (2010).
- 5 Yin, P., Yan, H., Daniell, X. G., Turberfield, A. J. & Reif, J. H. A unidirectional DNA walker that moves autonomously along a track. *Angew. Chem. Int. Ed.* **43**, 4906-4911 (2004).
- 6 Green, S. J., Bath, J. & Turberfield, A. J. Coordinated chemomechanical cycles: A mechanism for autonomous molecular motion. *Phys. Rev. Lett.* **101**, 238101 (2008).
- 7 Omabegho, T., Sha, R. & Seeman, N. C. A bipedal DNA brownian motor with coordinated legs. *Science* **324**, 67-71 (2009).
- 8 Seeman, N. C. Nucleic-acid junctions and lattices. *J. Theor. Biol.* **99**, 237-247 (1982).
- 9 Kallenbach, N. R., Ma, R.-I. & Seeman, N. C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305**, 829-831 (1983).
- 10 Fu, T. J. & Seeman, N. C. DNA double-crossover molecules. *Biochemistry* **32**, 3211-3220 (1993).
- 11 Winfree, E., Liu, F. R., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539-544 (1998).
- 12 Winfree, E. Algorithmic self-assembly of DNA: Theoretical motivations and 2D assembly experiments. *J. Biomol. Struct. Dyn.* **2**, 263-270 (2000).
- 13 Rothemund, P. W. K., Papadakis, N. & Winfree, E. Algorithmic self-assembly of DNA sierpinski triangles. *PLoS Biol.* **2**, e424 (2004).
- 14 Barish, R. D., Rothemund, P. W. K. & Winfree, E. Two computational primitives for algorithmic self-assembly: Copying and counting. *Nano Lett.* **5**, 2586-2592 (2005).
- 15 Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L. P., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V. & Kjems, J. Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73-76 (2009).
- 16 Douglas, S. M., Dietz, H., Liedl, T., Hogberg, B., Graf, F. & Shih, W. M. Self-assembly of DNA

- into nanoscale three-dimensional shapes. *Nature* **459**, 414-418 (2009).
- 17 Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725-730 (2009).
  - 18 Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623-626 (2012).
  - 19 Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177-1183 (2012).
  - 20 Rothemund, P. W. K. & Andersen, E. S. Nanotechnology: The importance of being modular. *Nature* **485**, 584-585 (2012).
  - 21 Maune, H. T., Han, S.-p., Barish, R. D., Bockrath, M., Goddard, W. A., Rothemund, P., W. K. & Winfree, E. Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nature Nanotech.* **5**, 61-66 (2010).
  - 22 Ke, Y., Lindsay, S., Chang, Y., Liu, Y. & Yan, H. Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays. *Science* **319**, 180-183 (2008).
  - 23 Voigt, N. V., Topping, T., Rotaru, A., Jacobsen, M. F., Ravnsbaek, J. B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, A., Besenbacher, F. & Gothelf, K. V. Single-molecule chemical reactions on DNA origami. *Nature Nanotech.* **5**, 200-203 (2010).
  - 24 Rinker, S., Ke, Y., Liu, Y., Chhabra, R. & Yan, H. Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding. *Nature Nanotech.* **3**, 418-422 (2008).
  - 25 Endo, M., Katsuda, Y., Hidaka, K. & Sugiyama, H. A versatile DNA nanochip for direct analysis of DNA base-excision repair. *Angew. Chem. Int. Ed.* **49**, 9412-9416 (2010).
  - 26 Endo, M., Katsuda, Y., Hidaka, K. & Sugiyama, H. Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *J. Am. Chem. Soc.* **132**, 1592-1597 (2010).
  - 27 Derr, N. D., Goodman, B. S., Jungmann, R., Leschziner, A. E., Shih, W. M. & Reck-Peterson, S. L. Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science* **338**, 662-665 (2012).
  - 28 Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585-1588 (2006).
  - 29 Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196-1201 (2011).
  - 30 Yurke, B., Turberfield, A. J., Mills, A. P., Jr., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605-608 (2000).
  - 31 Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chem.* **3**, 103-113 (2011).

- 32 Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303-17314 (2009).
- 33 Yan, H., Zhang, X. P., Shen, Z. Y. & Seeman, N. C. A robust DNA mechanical device controlled by hybridization topology. *Nature* **415**, 62-65 (2002).
- 34 Han, D., Pal, S., Liu, Y. & Yan, H. Folding and cutting DNA into reconfigurable topological nanostructures. *Nature Nanotech.* **5**, 712-717 (2010).
- 35 Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831-834 (2012).
- 36 Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19-54 (2003).
- 37 Balzani, V. Nanoscience and nanotechnology: The bottom-up construction of molecular devices and machines. *Pure Appl. Chem.* **80**, 1631-1650 (2008).
- 38 Strunz, T., Oroszlan, K., Schäfer, R. & Güntherodt, H.-J. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11277-11282 (1999).
- 39 Krautbauer, R., Rief, M. & Gaub, H. E. Unzipping DNA oligomers. *Nano Lett.* **3**, 493-496 (2003).
- 40 Šponer, J., Jurečka, P., Marchan, I., Luque, F. J., Orozco, M. & Hobza, P. Nature of base stacking: Reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem. Eur. J.* **12**, 2854-2865 (2006).
- 41 Šponer, J., Riley, K. E. & Hobza, P. Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.* **10**, 2595-2610 (2008).
- 42 Nakata, M., Zanchetta, G., Chapman, B. D., Jones, C. D., Cross, J. O., Pindak, R., Bellini, T. & Clark, N. A. End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. *Science* **318**, 1276-1279 (2007).
- 43 Maffeo, C., Luan, B. & Aksimentiev, A. End-to-end attraction of duplex DNA. *Nucleic Acids Res.* **40**, 3812-3821 (2012).
- 44 Protozanova, E., Yakovchuk, P. & Frank-Kamenetskii, M. D. Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* **342**, 775-785 (2004).
- 45 Wang, R., Kuzuya, A., Liu, W. & Seeman, N. C. Blunt-ended DNA stacking interactions in a 3-helix motif. *Chem. Commun.* **46**, 4905-4907 (2010).
- 46 Endo, M., Sugita, T., Katsuda, Y., Hidaka, K. & Sugiyama, H. Programmed-assembly system using DNA jigsaw pieces. *Chem. Eur. J.* **16**, 5362-5368 (2010).
- 47 Suckling, C. J. Molecular recognition — a universal molecular science? *Cell. Mol. Life Sci.* **47**, 1093-1095 (1991).
- 48 Xu, Q., Schlabach, M. R., Hannon, G. J. & Elledge, S. J. Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2289-2294 (2009).

- 49 He, Y., Ye, T., Su, M., Zhang, C., Ribbe, A. E., Jiang, W. & Mao, C. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198-201 (2008).
- 50 Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021-1024 (1994).
- 51 Barish, R. D., Schulman, R., Rothmund, P. W. K. & Winfree, E. An information-bearing seed for nucleating algorithmic self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6054-6059 (2009).
- 52 Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202-205 (2010).
- 53 Zimmerman, S. C. & Corbin, P. S. Heteroaromatic modules for self-assembly using multiple hydrogen bonds. *Struct. Bond.* **96**, 63-94 (2000).
- 54 Claessens, C. G. & Stoddart, J. F.  $\pi$ - $\pi$  interactions in self-assembly. *J. Phys. Org. Chem.* **10**, 254-272 (1997).
- 55 Klosterman, J. K., Yamauchi, Y. & Fujita, M. Engineering discrete stacks of aromatic molecules. *Chem. Soc. Rev.* **38**, 1714-1725 (2009).
- 56 Sacanna, S., Irvine, W. T. M., Chaikin, P. M. & Pine, D. J. Lock and key colloids. *Nature* **464**, 575-578 (2010).
- 57 Zhao, K. & Mason, T. G. Directing colloidal self-assembly through roughness-controlled depletion attractions. *Phys. Rev. Lett.* **99**, 268301 (2007).
- 58 Bowden, N., Terfort, A., Carbeck, J. & Whitesides, G. M. Self-assembly of mesoscale objects into ordered two-dimensional arrays. *Science* **276**, 233-235 (1997).
- 59 Rothmund, P. W. K. Using lateral capillary forces to compute by self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 984-989 (2000).
- 60 Woolfson, D. N. The design of coiled-coil structures and assemblies. *Adv. Protein Chem.* **70**, 79-112 (2005).
- 61 Kool, E. T. Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1-22 (2001).
- 62 Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564-574 (2006).
- 63 DeVoe, H. & Tinoco, I., Jr. The stability of helical polynucleotides: Base contributions. *J. Mol. Biol.* **4**, 500-517 (1962).
- 64 Crothers, D. M. & Zimm, B. H. Theory of the melting transition of synthetic polynucleotides: Evaluation of the stacking free energy. *J. Mol. Biol.* **9**, 1-9 (1964).
- 65 SantaLucia, J. & Hicks, D. The thermodynamics of DNA structural motifs. *Ann. Rev. Biophys. Biomol. Struct.* **33**, 415-440 (2004).

- 66 Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367-379 (1968).
- 67 Wang, J. C. Helical repeat of DNA in solution. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 200-203 (1979).
- 68 Ke, Y., Douglas, S. M., Liu, M., Sharma, J., Cheng, A., Leung, A., Liu, Y., Shih, W. M. & Yan, H. Multilayer DNA origami packed on a square lattice. *J. Am. Chem. Soc.* **131**, 15903-15908 (2009).
- 69 Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M. & Shih, W. M. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001-5006 (2009).
- 70 Sherman, W. B. & Seeman, N. C. Design of minimally strained nucleic acid nanotubes. *Biophys. J.* **90**, 4546-4557 (2006).
- 71 Kegler, K., Konieczny, M., Dominguez-Espinosa, G., Gutsche, C., Salomo, M., Kremer, F. & Likos, C. N. Polyelectrolyte-compression forces between spherical DNA brushes. *Phys. Rev. Lett.* **100**, 118302 (2008).
- 72 Saccà, B., Meyer, R., Feldkamp, U., Schroeder, H. & Niemeyer, C. M. High-throughput, real-time monitoring of the self-assembly of DNA nanostructures by FRET spectroscopy. *Angew. Chem. Int. Ed.* **47**, 2135-2137 (2008).
- 73 Nangreave, J., Yan, H. & Liu, Y. Studies of thermal stability of multivalent DNA hybridization in a nanostructured system. *Biophys. J.* **97**, 563-571 (2009).
- 74 Capaldi, R. A. & Aggeler, R. Mechanism of the F1F0-type ATP synthase, a biological rotary motor. *Trends Biochem. Sci.* **27**, 154-160 (2002).
- 75 Smirnov, S., Matray, T. J., Kool, E. T. & de los Santos, C. Integrity of duplex structures without hydrogen bonding: DNA with pyrene paired at abasic sites. *Nucleic Acids Res.* **30**, 5561-5569 (2002).
- 76 Sa-Ardyen, P., Vologodskii, A. V. & Seeman, N. C. The flexibility of DNA double crossover molecules. *Biophys. J.* **84**, 3829-3837 (2003).
- 77 Sun, X., Ko, S. H., Zhang, C., Ribbe, A. E. & Mao, C. Surface-mediated DNA self-assembly. *J. Am. Chem. Soc.* **131**, 13248-13249 (2009).
- 78 Hamada, S. & Murata, S. Substrate-assisted assembly of interconnected single-duplex DNA nanostructures. *Angew. Chem. Int. Ed.* **48**, 6820-6823 (2009).
- 79 Lee, J., Kim, S., Kim, J., Lee, C.-W., Roh, Y. & Park, S. H. Coverage control of DNA crystals grown by silica assistance. *Angew. Chem. Int. Ed.* **50**, 9145-9149 (2011).
- 80 Kershner, R. J., Bozano, L. D., Micheel, C. M., Hung, A. M., Fornof, A. R., Cha, J. N., Rettner, C. T., Bersani, M., Frommer, J., Rothmund, P. W. K. & Wallraff, G. M. Placement and orientation of individual DNA shapes on lithographically patterned surfaces. *Nature Nanotech.* **4**, 557-561 (2009).
- 81 Bezanilla, M., Manne, S., Laney, D. E., Lyubchenko, Y. L. & Hansma, H. G. Adsorption of DNA to mica, silylated mica, and minerals: Characterization by atomic force microscopy. *Langmuir* **11**, 655-659 (1995).

- 82 Pastré, D., Piétrement, O., Fusil, S., Landousy, F., Jeusset, J., David, M.-O., Hamon, L., Le Cam, E. & Zozime, A. Adsorption of DNA to mica mediated by divalent counterions: A theoretical and experimental study. *Biophys. J.* **85**, 2507-2518 (2003).
- 83 Piétrement, O., Pastré, D., Fusil, S., Jeusset, J., David, M.-O., Landousy, F., Hamon, L., Zozime, A. & Le Cam, E. Reversible binding of DNA on NiCl<sub>2</sub>-treated mica by varying the ionic strength. *Langmuir* **19**, 2536-2539 (2003).
- 84 Pastré, D., Hamon, L., Landousy, F., Sorel, I., David, M.-O., Zozime, A., Le Cam, E. & Piétrement, O. Anionic polyelectrolyte adsorption on mica mediated by multivalent cations: A solution to DNA imaging by atomic force microscopy under high ionic strengths. *Langmuir* **22**, 6651-6660 (2006).
- 85 Bezanilla, M., Drake, B., Nudler, E., Kashlev, M., Hansma, P. K. & Hansma, H. G. Motion and enzymatic degradation of DNA in the atomic force microscope. *Biophys. J.* **67**, 2454-2459 (1994).
- 86 Hansma, H. G. & Laney, D. E. DNA binding to mica correlates with cationic radius: Assay by atomic force microscopy. *Biophys. J.* **70**, 1933-1939 (1996).
- 87 Thomson, N. H., Kasas, S., Smith, B. L., Hansma, H. G. & Hansma, P. K. Reversible binding of DNA to mica for AFM imaging. *Langmuir* **12**, 5905-5908 (1996).
- 88 Shao, Z., Mou, J., Czajkowsky, D. M., Yang, J. & Yuan, J.-Y. Biological atomic force microscopy: What is achieved and what is needed. *Adv. Phys.* **45**, 1-86 (1996).
- 89 Woo, S. & Rothmund, P. W. K. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nature Chem.* **3**, 620-627 (2011).
- 90 Martin, T. G. & Dietz, H. Magnesium-free self-assembly of multi-layer DNA objects. *Nat. Commun.* **3**, 1103 (2012).
- 91 Ted Pella, [http://www.tedpella.com/vacuum\\_html/micragradev1properties.htm](http://www.tedpella.com/vacuum_html/micragradev1properties.htm).
- 92 U. S. Geological survey open-file report 01-041, <http://pubs.usgs.gov/of/2001/of01-041/htmldocs/clays/illite.htm>.
- 93 Güven, N. Mica structure and fibrous growth of illite. *Clays Clay Miner.* **49**, 189-196 (2001).
- 94 Bloom, K. & Joglekar, A. Towards building a chromosome segregation machine. *Nature* **463**, 446-456 (2010).
- 95 Ghosh, S. K., Hajra, S., Paek, A. & Jayaram, M. Mechanisms for chromosome and plasmid segregation. *Annu. Rev. Biochem.* **75**, 211-241 (2006).
- 96 Salje, J., Gayathri, P. & Löwe, J. The ParMRC system: Molecular mechanisms of plasmid segregation by actin-like filaments. *Nat. Rev. Microbiol.* **8**, 683-692 (2010).
- 97 Garner, E. C., Campbell, C. S. & Mullins, R. D. Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science* **306**, 1021-1025 (2004).
- 98 Gayathri, P., Fujii, T., Møller-Jensen, J., van den Ent, F., Namba, K. & Löwe, J. A bipolar spindle

- of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* **338**, 1334-1337 (2012).
- 99 Møller-Jensen, J., Borch, J., Dam, M., Jensen, R. B., Roepstorff, P. & Gerdes, K. Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol. Cell* **12**, 1477-1487 (2003).
- 100 Salje, J. & Lowe, J. Bacterial actin: Architecture of the ParMRC plasmid DNA partitioning complex. *EMBO J.* **27**, 2230-2238 (2008).
- 101 Gayathri, P., Fujii, T., Namba, K. & Löwe, J. Structure of the ParM filament at 8.5 Å resolution. *J. Struct. Biol.* (2013).
- 102 Møller-Jensen, J., Ringgaard, S., Mercogliano, C. P., Gerdes, K. & Lowe, J. Structural analysis of the ParR/*parC* plasmid partition complex. *EMBO J.* **26**, 4413-4422 (2007).
- 103 Hoischen, C., Bussiek, M., Langowski, J. & Diekmann, S. Escherichia coli low-copy-number plasmid R1 centromere *parC* forms a U-shaped complex with its binding protein ParR. *Nucleic Acids Res.* **36**, 607-615 (2008).
- 104 Choi, C. L., Claridge, S. A., Garner, E. C., Alivisatos, A. P. & Mullins, R. D. Protein-nanocrystal conjugates support a single filament polymerization model in R1 plasmid segregation. *J. Biol. Chem.*, M803833200 (2008).
- 105 Garner, E. C., Campbell, C. S., Weibel, D. B. & Mullins, R. D. Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* **315**, 1270-1274 (2007).
- 106 Murrell, M. P. & Gardel, M. L. F-actin buckling coordinates contractility and severing in a biomimetic actomyosin cortex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20820-20825 (2012).
- 107 Köhler, S., Lieleg, O. & Bausch, A. R. Rheological characterization of the bundling transition in F-actin solutions induced by methylcellulose. *PLoS ONE* **3**, e2736 (2008).
- 108 Personal communication with E. C. Garner.
- 109 Theriot, J. A. The polymerization motor. *Traffic* **1**, 19-28 (2000).
- 110 Phillips, R., Kondev, J. & Theriot, J. *Physical biology of the cell.* (Garland Science, 2009).
- 111 Campbell, C. S. & Mullins, R. D. In vivo visualization of type II plasmid segregation: Bacterial actin filaments pushing plasmids. *J. Cell. Biol.* **179**, 1059-1066 (2007).
- 112 Peskin, C. S., Odell, G. M. & Oster, G. F. Cellular motions and thermal fluctuations: The brownian ratchet. *Biophys. J.* **65**, 316-324 (1993).
- 113 Personal communication with C. Maffeo.
- 114 Rothmund, P. W. K., Ekani-Nkodo, A., Papadakis, N., Kumar, A., Fygenson, D. K. & Winfree, E. Design and characterization of programmable DNA nanotubes. *J. Am. Chem. Soc.* **26**, 16344-16353 (2004).